The MIT Press

RESEARCH ARTICLE

The relationship between journal citation impact and citation sentiment: A study of 32 million citances in PubMed Central

Erjia Yan (iD), **Zheng Chen, and Kai Li** (iD)

College of Computing and Informatics, Drexel University, 3675 Market Street, Philadelphia, PA 19104

ABSTRACT

Citation sentiment plays an important role in citation analysis and scholarly communication research, but prior citation sentiment studies have used small data sets and relied largely on manual annotation. This paper uses a large data set of PubMed Central (PMC) full-text publications and analyzes citation sentiment in more than 32 million citances within PMC, revealing citation sentiment patterns at the journal and discipline levels. This paper finds a weak relationship between a journal's citation impact (as measured by CiteScore) and the average sentiment score of citances to its publications. When journals are aggregated into quartiles based on citation impact, we find that journals in higher quartiles are cited more favorably than those in the lower quartiles. Further, social science journals are found to be cited with higher sentiment, followed by engineering and natural science and biomedical journals, respectively. This result may be attributed to disciplinary discourse patterns in which social science researchers tend to use more subjective terms to describe others' work than do natural science or biomedical researchers.

1. INTRODUCTION

Journal citation impact is no stranger to bibliometrics and science of science research. A journal's citation impact can be measured by several indicators, such as journal impact factor, CiteScore, field normalized citation scores, and citation network-based indicators. Among these, journal impact factor is perhaps the most studied and debated citation-based indicator. Researchers tend to be cautious about its use because they are aware of the many cases in which it can be misused or abused (Hicks, Wouters, et al., 2015). Publishers, and sometimes evaluators, possess a more positive attitude toward this indicator because it is a straightforward way to show the audience a journal's citation impact.

The goal of this paper is not to delve into the heated discussion of the strengths and weakness of journal impact factor and related indicators, but rather to compare journal citation impact with citation sentiment. The motivation behind this goal can be traced back to the early studies of citation functions by Garfield (Garfield, 1965; Garfield & Merton, 1979), in which the authors argued that citations serve difference functions in scholarly communication and proposed several citation functions for theoretical study. Citations also function as a symbolic language of science, reflecting the underlying substance of and relationship among scientific documents (Small, 1978). However, to understand the functions of citations, we must go beyond

the binary presence or absence of a citation to the "comparison of the cited text with its context of citation in the citing texts" (Small, 2004, p. 76).

The contextual information required—that is, the text surrounding each of the cited references within the text—is called *citances* (Nakov, Schwartz, & Hearst, 2004). Foundational work in citation context analysis depends upon the classification of citations by their function. A significant number of studies have been carried out in which citations have been classified manually (Chubin & Moitra, 1975; Erikson & Erlandson, 2014; Moravcsik & Murugesan, 1975; Zhang, Ding, & Milojević, 2013). Teufel, Siddharthan, and Tidhar (2006) proposed an automated classification of citations in which a distinction is made between (a) citations indicating a weakness of the cited work, (b) citations comparing the citing and the cited work, (c) citations expressing a positive sentiment toward the cited work, and (d) citations providing a neutral description of the cited work. The significance of their work is not just the automation of citation classification but also in incorporation of citation sentiment in citation context analysis. Citation sentiment, operationalized as the opinion expressed by the citing author of a cited work, is an important element in citation context analysis, as it can be used to verify the status of claims. There are two types of citation sentiment analysis: One is the measure of citation sentiment of a citing entity (e.g., author, journal) and the other is the measure of citation sentiment of a cited entity. One example of the former type is the measure of the sentiment of an author's citing citances made in all of his or her publications to establish this author's sentiment baseline—some authors are more generous in accolading others' work while others are less liberal. An example of the second type is to measure the sentiment in which other authors cited an author's work (see a study of this type by Yan, Chen, and Li [2020]). The current study belongs to the latter type. For a journal, it measures the sentiment made by authors who cited papers published in this journal.

Sentiment analysis has been made possible for a number of text genres—mainly online reviews and social media, where polarized opinions abound (Agarwal, Xie, et al., 2011; Pang & Lee, 2008). In scholarly publications, however, expressions of opinion are subtle, and strong opinions are rare (Athar, 2011; Athar & Teufel, 2012). In addition, certain technical terms contain words that may signify a strong sentiment (e.g., *support vector machine*, *loss aversion*) but should be regarded as nonpolarized for purposes of sentiment analysis. To address these genre-specific issues, this study uses a high-performing sentiment detection method for citation context analysis. Developed by our team (Yan et al., 2020), this method relies on advanced natural language processing techniques. Here, we apply the method to a large full-text corpus of 1.68 million publications retrieved from PubMed Central (PMC).

One central hypothesis for this paper is that higher impact journals also tend to be cited with more positive sentiment. This hypothesis is conceived based on the observations that higher impact journals publish more papers of higher importance. When authors cited those papers, they may be more likely to use affirming words to describe the cited work in citances. We developed one research question here:

- Do papers published in a high-impact journal (defined as one with higher CiteScores) tend to be cited more positively (i.e., do citances to this journal's publications have higher citation sentiment)?

Because of the inclusion of a large multidisciplinary corpus of full-text data, the second hypothesis is that disciplines have different norms of expressing opinions in citances: Citances made in natural sciences publications are more observation based and thus less polarized, while citances made in social sciences publications are more argument based

and more polarized. This hypothesis is brought forth based on earlier linguistic studies of disciplinary norms of writing conventions, in that writings in "hard" sciences are fact-based and impersonal whereas writings in "soft" sciences are metaphorical and interpretive (Becher & Trowler, 2001; Parry, 1998). We developed one research question for this hypothesis:

- What are the disciplinary patterns of citation sentiment as demonstrated through journals' citances in which they were cited?

By answering these questions, this paper makes distinctive contributions to studies of impact assessment, as it illustrates key aspects of citation sentiment—a key concept in scholarly communication, yet insufficiently studied due to previous technical challenges—and its relationship with journal citation impact. The results inform our understanding of citation contexts and provide evidence to support ongoing advocacy for context-aware science evaluation.

2. MATERIALS AND METHODS

2.1. Data

PubMed Central Open Access Subset (PMC OA) was selected as the data source. PMC OA is the largest open-access repository for scientific publications, making it possible to evaluate citation sentiment at the sentence level. Although most journals included in PMC OA are related to biomedical research, a wide range of science, technology, engineering, and mathematics (STEM) and social science disciplines are also represented.

We collected PMC OA data in March 2018. After a simple preprocessing step of removing papers without a citation, the resulting data set contains 1.68 million publications between 1989 and 2018 and 68 million citations for these publications. The numbers of publications and citations before 2000 are quite limited; over 85% of papers and over 93% of citations were made after 2010. About 15% of the citances are negative while the rest are nonnegative.

2.2. Methods

We rely on the XML tag <xref> to track down citances. To identify the exact sentence where the citance appears, the <xref> nodes were first replaced by a special token, then the full paragraph text was tokenized and further split by recognized punctuation marks. Apart from the standard use of <xref> in denoting citations, there are also nonstandard usages, complex tag embeddings, and even typos, as detailed in Yan et al. (2020). Due to the size of this data set, it is impossible to enumerate all scenarios of nonstandard usages; thus, we only considered the standard usage of the XML tag in this research.

After the citances were extracted, we applied the sentiment detection method described below. As noted, the expression of sentiments in scientific literature tends to be subtle, and the majority of citations are nonnegative (Case & Higgins, 2000; Jurgens, Kumar, et al., 2016). Meanwhile, technical terms can sometime introduce noise to sentiment classification (e.g., *discriminative* models, *support* vector machines). To minimize the noise brought by technical terms when conducting sentiment analysis, we processed the full-text data by removing scientific terms: First, we extracted scientific terms from full texts using a term extraction method developed in our previous work (Chen & Yan, 2017; Yan, Williams, & Chen, 2017); second, we screened out technical words unlikely to express a sentiment (e.g., *system*, *injection*, *neuron*). These words tend to have a high "uniqueness" score in our term extraction

method (Chen & Yan, 2017; Yan et al., 2017), an indication that they are more likely to be scientific terms.

With this process, we effectively reduced the noise introduced by technical terms. We then fed the processed full-text data into SenticNet, a state-of-the-art concept-level sentiment analysis program (Cambria, Poria, et al., 2016). SenticNet is built using natural language and statistical methods over a word network; its merit is that it is weakly supervised, requiring minimal training compared to supervised methods (e.g., Athar & Teufel, 2012; Jha, Jbara, et al., 2017; Xu, Zhang, et al., 2015). The output of SenticNet for a citance is a sentiment score between –1 (most negative) and 1 (most positive), with 0 indicating a neutral sentiment. Because most citations are made in a nonnegative way, when aggregating citations to journals, the average sentiment score is about 0.17. Therefore, in essence, when conducting sentiment analysis for citances, what we measure is the occurrence of words of approval such as "novel" and "important" that tend to have high sentiment scores. Our sentiment analysis method performed well on nonnegative citances, reaching a precision level in excess of 0.9 (Yan et al., 2020). We provide two citances here, one positive and the other negative. Scores in parentheses show the sentiment score of the term.

- Positive example: The first(0.649) evidence(0.115) of the existence(0.57) of this mechanism(0.053) appeared(0.574) in 1998, when Fire et al. [(CITE [9486653])] observed (0.343) in Caenorhabditis elegans that double-stranded RNAs (dsRNAs) were the basis (0.03) of sequence-specific inhibition(0.066) of protein expression.
- Negative example: Another serious(-0.51) problem(-0.62) is the gene set itself that is used(0.271) for the induction of pluripotency [(CITE [17554338])]

Lastly, we aggregated citances to journals and calculated the average sentiment score for each journal. To properly obtain journal citation impact data, we downloaded the 2018 version of the journal metrics report by Scopus. The 2018 report contains journal-level citation data for more than 20,000 journals indexed in the Scopus database. A journal's CiteScore in the 2018 report is calculated as the number of citations the journal received in 2017 for its articles published between 2014 and 2016 divided by the number of documents the journal published between 2014 and 2016. Another indicator used in this study—percentage of documents cited—is the percentage of documents published between 2014 and 2016 that received citations in 2017. We further grouped the journals based on their CiteScore quartiles, domain, publisher, and open access status, all of which were provided by the journal metrics report. We then matched the journals in PMC with those included in the journal metrics report. Slightly more than 3,700 journals were matched, with an aggregated number of 32 million citances. These are used as the final data set in the analysis[1]. The data set is available for downloading at Yan (2019).

3. RESULTS

We first calculated the Spearman rank correlation between journal citation sentiment scores and journal citation impact. Journal citation impact is represented by two indicators, CiteScore and percentage of documents cited, both of which are taken from the 2018 journal metrics report. In Table 1, we show the distributions between citation sentiment and impact for all

---

[1] One limitation of the match is that CiteScore reflects the performance of a journal in a particular year, whereas citances in PMC span more than a decade. However, considering that more than half of the citances in PMC were made after 2015, the average sentiment score of a journal should signify its recent citation sentiment, thus making it a meaningful point of comparison with CiteScore.

Table 1. Spearman correlation coefficients between citation sentiment scores and two journal citation impact metrics

	CiteScore	Percentage cited
All journals ($n = 3{,}728$)	0.0764 ($p < 0.01$)	0.0852 ($p < 0.01$)
Journals with 10k or more citations ($n = 400$)	0.2379 ($p < 0.01$)	0.2519 ($p < 0.01$)
Closed-access journals ($n = 2{,}732$)	0.0055 ($p = 0.77$)	0.0186 ($p = 0.33$)
Open-access journals ($n = 996$)	0.2674 ($p < 0.01$)	0.2667 ($p < 0.01$)

journals ($n = 3{,}728$) and for those with more than 10,000 citances ($n = 400$). Separately, we show the distributions for closed-access journals ($n = 2{,}732$) and open-access journals ($n = 996$).

Table 1 shows that the strongest correlation between citation sentiment and CiteScore occurs for open-access journals, with Spearman correlation coefficient at the 0.27 level ($p < 0.01$), whereas there is no correlation between citation sentiment scores and journal citation impact for closed-access journals. Meanwhile, there is a moderate correlation between citation sentiment and CiteScore for journals with more than 10,000 citations, with Spearman correlation coefficient at the 0.24 level ($p < 0.01$). About a third of open-access journals ($n = 293$) also belong to this category. For this group of journals, CiteScore has a correlation coefficient of 0.28 with citation sentiment. Overall, the correlation between citation sentiment and percentage cited is statistically similar to that between sentiment and CiteScore.

Next, we grouped journals into four CiteScore quartiles and compared their CiteScore with citation sentiment (Figure 1).

The box plot in Figure 1 provides clear evidence that journals in upper CiteScore quartiles have higher median citation sentiment scores. Quartile 1 journals have a median sentiment score of 0.18, followed by quartiles 2 (0.178), 3 (0.17), and 4 (0.16). The results show that individual journals' CiteScore and citation sentiment score may not follow a strong linear pattern; nonetheless, when journals are grouped into broad categories based on their citation impact, the relationship between citation impact and citation sentiment is evident, with a journal from upper citation quartile groups more likely to have a higher citation sentiment score. This pattern is further confirmed by Figure 2, in which journals were grouped into 10 CiteScore deciles. Journals in the top decile for CiteScore have the highest median citation sentiment score (0.185), and journals in the bottom decile have the lowest (0.165). The relationship
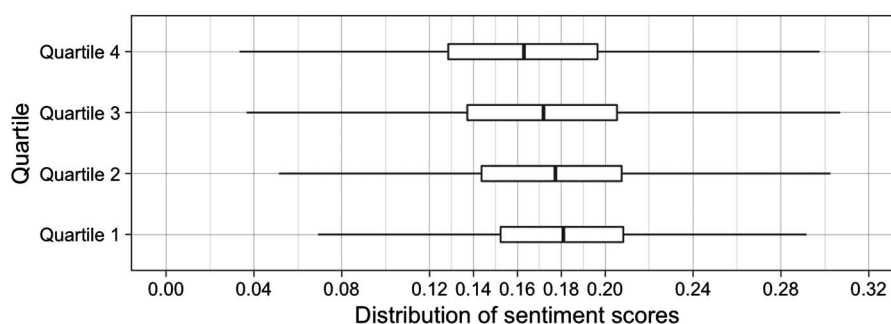

Figure 1. Box-plot visualization of sentiment scores over four quartiles of CiteScore.
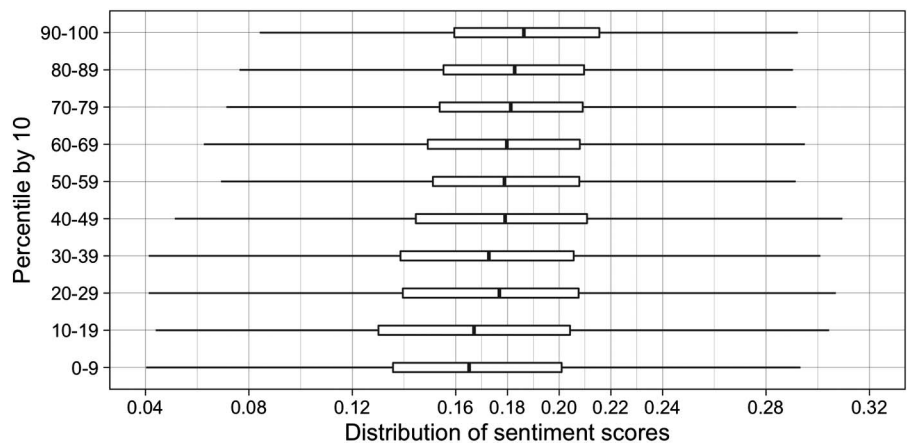
Figure 2. Box-plot visualization of sentiment scores over 10 sections of CiteScore percentiles.

between CiteScore percentile and sentiment score is almost linear, with the exception of two percentile groups (the 20–29th and 30–39th percentile groups).

Another key variable to examine is a journal's open-access status: Do open-access journals tend to have higher citation sentiment scores? As before, we use a box-plot visualization to reveal the relationship between open-access status and citation sentiment (Figure 3).

Closed-access journals, measured by median, have higher citation sentiment scores for each of the four quartiles than open-access journals (Figure 3). The median sentiment score of all closed-access journals is 0.18, compared with the open-access median of 0.172. The results suggest that the median citation sentiment score of open-access journals is comparable with the sentiment scores of the 30–39th percentile or the third quartile (by CiteScore) of all journals. For open-access journals, there is a clear, consistent increase of sentiment scores for journals in the fourth quartile to journals in the first quartile. For closed-access journals, the increase is much less noticeable, although it still exists. For journals in the first CiteScore
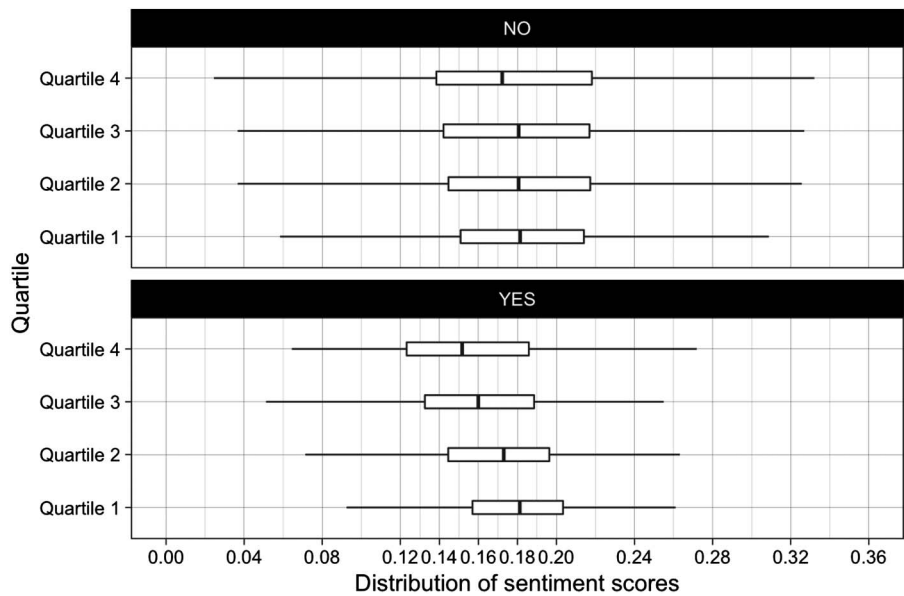

Figure 3. Box-plot visualization of sentiment scores for closed- and open-access journals in four quartiles.
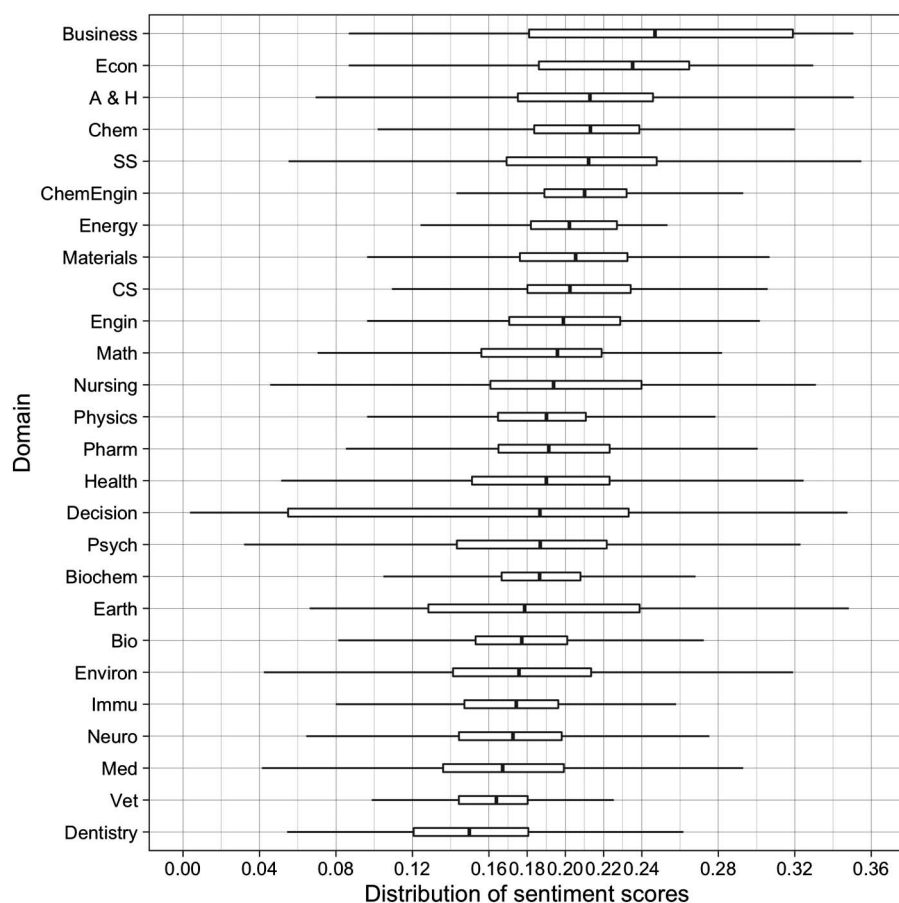
Figure 4. Box-plot visualization of sentiment scores for 26 subjects (SS: Social Sciences, A & H: Arts and Humanities, CS: Computer Science).

quartile, there is no difference between closed-access and open-access journals' median sentiment scores.

Scopus assigns a journal into one of 27 domains. Except for a multidisciplinary domain, we visualize the sentiment scores for the other 26 domains in Figure 4.

As shown in Figure 4, Business has the highest median sentiment score (0.25). A few related social science domains also report high sentiment scores, including Economics (0.23) and Social Sciences (0.21). Followed by the social science domains, a few chemistry domains also have high sentiment scores, including Chemistry (0.21), Chemistry Engineering (0.21), and Materials Science (0.20). The third group include a few physics and engineering domains, including Energy (0.20), Computer Science (0.20), Engineering (0.198), Mathematics (0.195), and Physics (0.19). Other domains shown in Figure 4, mostly biomedical domains, reported low sentiment scores. The results reveal differences in disciplinary discourse: Social science researchers seem to embed more favorable views when citing other works than do chemists and physicists and engineers, who in turn cite more favorably than biomedical researchers.

Finally, we examine publisher-level citation sentiment scores. The eight publishers with the largest number of journals in our data set are included in Figure 5.

Publishers that largely publish social science journals (e.g., Sage and Wiley-Blackwell) tend to have higher sentiment scores, whereas publishers that publish journals in the biomedical
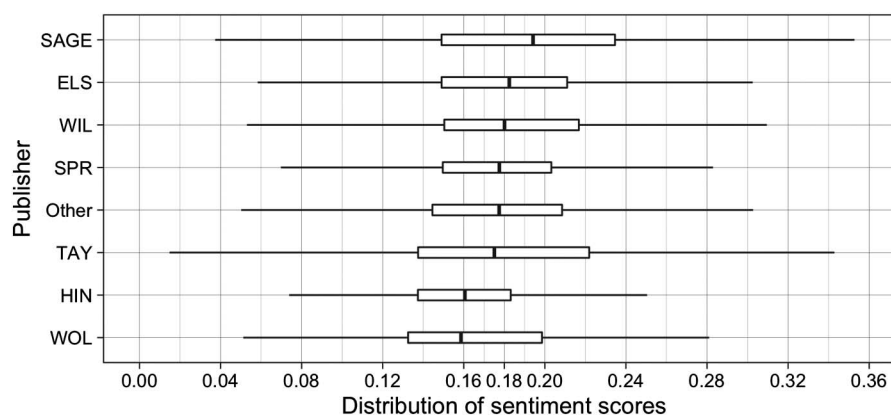
Figure 5. Box-plot visualization of sentiment scores for seven major publishers (SAGE: Sage, ELS: Elsevier, WIL: Wiley-Blackwell, SPR: Springer Nature, TAY: Taylor and Francis, HIN: Hindawi, WOL: Wolters Kluwer Health).

domains (e.g., Hindawi and Wolters Kluwer Health) tend to have low median sentiment scores.

4. DISCUSSION

4.1. Citation Impact and Sentiment

The results show that at the individual journal level, there is a weak relationship between a journal's citation sentiment score and its citation impact as measured by CiteScore. When CiteScore is replaced with percentage of papers cited, the relationship becomes stronger, although citation sentiment is still not strong enough to predict a journal's exact citation impact. However, when journals are aggregated by CiteScore at higher levels, such as quartiles or deciles, we find a noticeable relationship between citation impact and citation sentiment: A journal in a top quartile or decile is, in general, more likely to have a higher citation sentiment than a journal from a lower ranking group. The phrase "in general" is important here: The box plots in Figures 1 and 2 show that the distributions of sentiment scores for different quantiles overlap. This distribution pattern suggests that a journal's citation impact can provide only limited information about its citation sentiment. These findings partially support the first hypothesis on the relationship between journal citation impact and sentiment score. Although, probabilistically, journals with higher citation impact are likely to have higher citation sentiment, this statistic certainly cannot be applied to individual journals, and it would be even more problematic to use the macrolevel statistics to characterize individual papers. The best way to understand a paper's citation sentiment is, of course, to collect all citances to this paper and use a sentiment classifier to detect their citation sentiment. Any journal-level or quartile-level statistics should not be assumed to apply at the paper level.

The results provide evidence to support discussions about the misuse of journal-level indicators, including journal impact factor and CiteScore, in evaluation of individual papers and authors. This paper shows that citations do not have uniform importance, nor do they comply with the same sentiment polarity. When aggregating paper-level citations to journals, we are confronted with even fewer options to discern citation context. Granted, most citations are not polarized and nonnegative, as the current paper shows. Meanwhile, the normative theory paved the way for quantitative use of citations (Merton, 1973). Classic and contemporary work in this area flourished and advanced our understanding of the science of science; however, the limitations of the operationalization of citations as a proxy for productivity, impact, or any

evaluative metric should be clearly communicated. Regrettably, in one of his own early publications, the lead author of this study used journal impact factor as a proxy for individual paper impact without sufficient discussion of the caveats of such approximation (Yan & Ding, 2010). Previous research has provided concrete evidence that a few highly cited papers can significantly change a journal's citation impact, even as those highly cited papers have no bearing on the impact of other papers in the same journal (PLOS Medicine Editors, 2006). Moreover, impact does not directly quantify quality. This is why, for instance, many Nobel-Prize-winning works are published in reputable, domain-specific journals but not in multidisciplinary journals with the highest citation-based scores (Yan et al., 2020). It is an important reminder that one should not trade accuracy for convenience, but always use a paper's context and contents for more meaningful science evaluation.

4.2. Differences in Disciplinary Discourse

The results showed that social science domains tend to be cited with higher sentiment, followed by engineering and natural science domains and lastly biomedical related-domains. This work is the first to report such disciplinary differences in citation sentiment. The preliminary evidence presented here requires further in-depth review and analyses, using other sentiment classifiers on a cross-disciplinary corpus, to confirm the findings obtained in this study. One plausible interpretation of the results is that social science researchers may use more subjective terms to describe others' work (Demarest & Sugimoto, 2015), and those terms tend to carry more polarity. On the other hand, biomedical sciences are more clinical and objective; researchers in these fields tend to stick to facts and use fewer subjective terms. Because most subjective terms used in scientific writing are nonnegative, the higher the occurrence of such terms, the higher the citation sentiment will tend to be. The findings support the second hypothesis about the patterns of discipline-level citation sentiment as signified by disciplinary discourse characteristics.

Disciplinary discourse has been employed by linguists and social scientists to understand the social, cognitive, and epistemological cultures of disciplines (Parry, 1998). Scholars used small samples of full texts, typically a few journal articles, to reveal the nature of disciplinary writing conventions (Bazerman, 1981). One key finding is that social sciences writing is persuasive because scholars in social sciences do not necessarily share the same methodological or theoretical frameworks, whereas science writing is accretive and tacit knowledge is shared and embedded within the science communities (Bazerman, 1981; Parry, 1998). Becher and Trowler (2001) grouped disciplines into four categories based on their nature of knowledge: hard-pure (e.g., physics), soft-pure (e.g., humanities and anthropology), hard-applied (e.g., engineering and clinical medicine), and soft-applied (e.g., education and law). The nature of knowledge for hard-pure disciplines is atomistic and the nature of writing conventions is impersonal and value-free. Conversely, the nature of knowledge for soft-pure disciplines is organic and the nature of writing conventions is personal and value-laden. The nature of the applied disciplines lies between hard-pure and soft-pure disciplines.

It is clear from prior linguistic studies that disciplines possess different writing norms, attributed to "the nature of the knowledge bases concerned and to identifiable cultural traditions" (Parry, 1998, p. 275). This study adds a new dimension of citation sentiment to the understanding of disciplinary discourse and writing norms. The results obtained from this study also provide large quantitative evidence that confirmed these early observations of disciplinary writing conventions. Prior research in this area has largely relied on small samples of full texts, as evidenced in Parry's and Bazerman's works, or bibliographic records—for instance, by using

cocitation and coauthorship relationships (Newman, 2004; Small, 1973, 1978). Demarest and Sugimoto (2015, p. 1376) coined the term *discourse epistemetrics*, which describes the conduct of "large-scale quantitative inquiries into the socio-epistemic basis of disciplines." Abstracts, titles, and keywords are typically employed to study discourse epistemetrics. However, as argued by Montgomery (2017, p. 3), "communicating is doing science"; thus, we need to use large-scale full texts to understand disciplinary science making. Such full texts have become increasingly accessible to the public; PMC in particular has been extensively used by researchers to conduct research in text mining, knowledge discovery, and bibliometrics. Using PMC, we have identified patterns of disciplinary vocabulary use (Yan et al., 2017) and software and data citation practices (Pan, Yan, et al., 2015; Yan & Pan, 2015). The current work on citation sentiment further extends the techniques of large full-text data analysis, as well as our understanding of the epistemological cultures of disciplines.

5. CONCLUSION

Using a large full-text data set, this study analyzed the citation sentiment of more than 32 million citances and revealed citation sentiment patterns at the journal and discipline levels. We found that at the individual journal level, there is a weak relationship between a journal's citation impact (measured by CiteScore) and the citation sentiment, measured as the average sentiment score of citances to its publications. When journals were aggregated into quartiles based on their citation impact, we found that journals in higher quartiles tended to be cited more favorably than those in the lower quartiles. We also found that social science journals tended to be cited with higher sentiment, followed by engineering and natural science journals and then biomedical journals. This result may be attributed to disciplinary discourse patterns: Social science researchers may use more subjective terms to describe others' work than do biomedical researchers, and those terms tend to carry more polarity.

One limitation of this research is that the sentiment classifier is designed to deal with generic citances at scale and is not fine-tuned to treat the discipline-specific citances mentioned in Teufel (2009). Future research will benefit from designing a citation sentiment classifier that is capable of dealing with discipline-specific citances while achieving high efficiency in processing large citance corpora.

AUTHOR CONTRIBUTIONS

Erjia Yang: Conceptualization, Funding acquisition, Methodology, Project administration, Writing—original draft preparation, Writing—review and editing. Zhang Chen: Methodology, Writing—original draft preparation, Writing—review and editing. Kai Li: Visualization, Writing—original draft preparation, Writing—review and editing.

COMPETING INTERESTS

The authors declare no competing interests.

FUNDING INFORMATION

DATA AVAILABILITY

The data used in this paper is accessible via PubMed Central Open Access Subset at https://www.ncbi.nlm.nih.gov/pmc/tools/ftp/.

REFERENCES

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of Twitter data. In *Proceedings of the Workshop on Language in Social Media (LSM 2011)* (pp. 30–38). USA: Association for Computational Linguistics.

Athar, A. (2011). Sentiment analysis of citations using sentence structure-based features. Paper presented at the Proceedings of the ACL 2011 student session.

Athar, A., & Teufel, S. (2012). Context-enhanced citation sentiment detection. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 597–601). USA: Association for Computational Linguistics.

Bazerman, C. (1981). What written knowledge does: Three examples of academic discourse. *Philosophy of the Social Sciences*, *11*(3), 361–387.

Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines*. London: McGraw-Hill Education.

Cambria, E., Poria, S., Bajpai, R., & Schuller, B. (2016). SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics* (pp. 2666–2677). Osaka: The COLING 2016 Organizing Committee.

Case, D. O., & Higgins, G. M. (2000). How can we investigate citation behavior? A study of reasons for citing literature in communication. *Journal of the Association for Information Science and Technology*, *51*(7), 635–645.

Chen, Z., & Yan, E. (2017). Domain-independent term extraction & term network for scientific publications. *iConference 2017 Proceedings* (pp. 171–189). Illinois: IDEALS. http://hdl.handle.net/2142/96671

Chubin, D. E., & Moitra, S. D. (1975). Content analysis of references: Adjunct or alternative to citation counting? *Social Studies of Science*, *5*(4), 423–441.

Demarest, B., & Sugimoto, C. R. (2015). Argue, observe, assess: Measuring disciplinary identities and differences through socio-epistemic discourse. *Journal of the Association for Information Science and Technology*, *66*(7), 1374–1387.

Erikson, M. G., & Erlandson, P. (2014). A taxonomy of motives to cite. *Social Studies of Science*, *44*(4), 625–637.

Garfield, E. (1965). Can citation indexing be automated? In *Statistical Association Methods for Mechanized Documentation. Symposium Proceedings*, Washington 1964. Washington, DC: US Government Printing Office.

Garfield, E., & Merton, R. K. (1979). *Citation indexing: Its theory and application in science, technology, and humanities* (Vol. 8). Wiley: New York.

Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature*, *520*(7548), 429–431.

Jha, R., Jbara, A.-A., Qazvinian, V., & Radev, D. R. (2017). NLP-driven citation analysis for scientometrics. *Natural Language Engineering*, *23*(1), 93–130.

Jurgens, D., Kumar, S., Hoover, R., McFarland, D., & Jurafsky, D. (2016). Citation classification for behavioral analysis of a scientific field. *arXiv:1609.00435*.

Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. Chicago: University of Chicago Press.

Montgomery, S. L. (2017). *The Chicago guide to communicating science*. Chicago: University of Chicago Press.

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

Nakov, P. I., Schwartz, A. S., & Hearst, M. (2004). Citances: Citation sentences for semantic analysis of bioscience text. In *Proceedings of the SIGIR '04 workshop on Search and Discovery in Bioinformatics*. https://biotext.berkeley.edu/papers/citances-nlpbio04.pdf

Newman, M. E. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, *101*(suppl 1), 5200–5205.

Pan, X., Yan, E., Wang, Q., & Hua, W. (2015). Assessing the impact of software on science: A bootstrapped learning of software entities in full-text papers. *Journal of Informetrics*, *9*(4), 860–871.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, *2*(1–2), 1–135.

Parry, S. (1998). Disciplinary discourse in doctoral theses. *Higher Education*, *36*(3), 273–299.

PLOS Medicine Editors. (2006). The impact factor game. *PLOS Med*, *3*(6), e291. https://doi.org/10.1371/journal.pmed.0030291

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, *24*(4), 265–269.

Small, H. (2004). On the shoulders of Robert Merton: Towards a normative theory of citation. *Scientometrics*, *60*(1), 71–79.

Small, H. G. (1978). Cited documents as concept symbols. *Social Studies of Science*, *8*(3), 327–340.

Teufel, S. (2009). Citations and sentiment. Paper presented at the Workshop on Text Mining for Scholarly Communications and Repositories, University of Manchester, UK.

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* (pp. 103–110). Sydney: Association for Computational Linguistics.

Xu, J., Zhang, Y., Wu, Y., Wang, J., Dong, X., & Xu, H. (2015). Citation sentiment analysis in clinical trial papers. In *AMIA Annual Symposium Proceedings* (p. 1334). Bethesda, MD: American Medical Informatics Association.

Yan, E. (2019). *Journal sentiment data for 3728 Scopus journals*. Retrieved from: https://doi.org/10.6084/m9.figshare.10247069.v1

Yan, E., Chen, Z., & Li, K. (2020). Authors' status and the perceived quality of their work: Measuring citation sentiment change in Nobel articles. *Journal of the Association for Information Science and Technology*, *71*(3), 314–324.

Yan, E., & Ding, Y. (2010). Weighted citation: An indicator of an article's prestige. *Journal of the American Society for Information Science and Technology*, *61*(8), 1635–1643.

Yan, E., & Pan, X. (2015). A bootstrapping method to assess software impact in full-text papers. In *Proceedings of the 15th International Conference of the International Society for Scientometrics and Informetrics*. Leuven: ISSI.

Yan, E., Williams, J., & Chen, Z. (2017). Understanding disciplinary vocabularies using a full-text enabled domain-independent term extraction approach. *PLOS ONE*, *12*(11), e0187762.

Zhang, G., Ding, Y., & Milojević, S. (2013). Citation content analysis (CCA): A framework for syntactic and semantic analysis of citation content. *Journal of the American Society for Information Science and Technology*, *64*, 1490–1503.