Authors' status and the perceived quality of their work: measuring citation sentiment change in Nobel papers

Erjia Yan, Zheng Chen, Kai Li
College of Computing and Informatics, Drexel University, Philadelphia, PA, U.S.A.

Abstract
Prior research in status ordering has used numeric indicators to examine the impact of a status change on the perception of a scientist's work. This study measures the perception change directly as reflected in citation sentiment, with the attainment of a Nobel Prize in Chemistry or a Nobel Prize in Physiology or Medicine considered as the status change. The paper identifies 12,393 citances to 25 Nobel papers in PubMed Central and includes a control paper set of 75 papers with 30,851 citances. Results show a moderate increase in citation sentiment toward Nobel papers post-award. Dynamically, for Nobel papers, there is a steady sentiment increase, and a Nobel Prize seems to co-occur with this trend. This trend, however, is not evident in the control paper set.

Introduction
Status ordering is a fundamental area of sociological research. Humans are hardwired to understand their position within a group, from children in playgrounds (Bothner, Godart, & Lee, 2010), to pupils in classrooms (Patterson, Kupersmidt, & Griesler, 1990), to employees in workplaces (George, Dahlander, Graffin, & Sim, 2016; Harley, 1999). This near-obsession with knowing one's status and advancing in the status ordering may be attributed to the availability of resources: the scarcer the resources, the more desirable they are to us. In this competition for resources, higher status confers an advantage, thus making such status more appealing.

In science, the disproportionate concentration of resources is known as the Matthew Effect. Coined by Robert Merton (Merton, 1968), the term describes the phenomenon whereby eminent scientists often get more scientific credit compared with those who are relatively unknown. Thus, a change in status will inevitably result in a marked change in one's access to resources, social capital, and self-assurance, among other characteristics. Past research has established the existence of the Matthew Effect in a variety of scientific and scholarly settings, including scientific collaboration (Barabâsi et al., 2002; Newman, 2001a, 2001b, 2001c, 2004), information networks (Clauset, Shalizi, & Newman, 2009), distribution of citations (Bensman, 2008; Redner, 1998; Seglen, 1992; Stringer, Sales-Pardo, & Amaral, 2010), and word frequency distributions (Newman, 2005), among many other types of empirical data (Perc, 2014).

Prior studies have largely relied on numeric indicators to approximate resources when examining status ordering. Number of collaborators, for instance, has been used as a proxy for social capital (Barabâsi et al., 2002; Newman, 2001a, 2001b, 2001c, 2004), and number of citations has stood in for the concept of scientific rewards (Bensman, 2008; Redner, 1998; Seglen, 1992; Stringer et al., 2010). Although these studies have provided quantitative evidence of Matthew Effect, we have made little headway in understanding the impact of status change on subjective perceptions of one's work. This lack of progress is likely due to the lack of proper instruments with which to measure perceptions. In the present study, we aim to address this problem by tackling the complexity of citation sentiments, as used to measure the impact of authors' status change on how others cite their work. Among all the possible status changes in a scientist's career, winning a Nobel Prize is unparalleled. Thus, this study uses Nobel Prizes as the status change and measures the citation sentiment of articles citing a Nobel paper before and after its author's investiture.

As the most highly esteemed science award, the Nobel Prize has been studied in the context of scholarly communication from several perspectives, as have its laureates. Topics of such research include prize prediction (Ashton & Oppenheim, 1978; Garfield, 1986; Gingras & Wallace, 2010), sources of funding of Nobel papers (Tatsioni, Vavva, & Ioannidis, 2010), Nobel Prize effects on citation impact (Farys & Wolbring, 2017; Frandsen & Nicolaisen, 2013), patterns of collaboration (Chan, Önder, & Torgler, 2015), collaboration networks (Wagner, Horlings, Whetsell, Mattsson, & Nordqvist, 2015), and the likelihood of laureates to obtain more awards after their Nobel Prize (Chan, Gleeson, & Torgler, 2014).

Because of the Nobel Prize's unmatched prestige, scholarly communication researchers have been curious to find out whether Nobel laureates behave differently from average scientists, and whether their characteristics can be identified and used to predict future laureates. It has been found that although Nobel laureates differ from average scientists in their citation impact, bibliometric indicators alone are not able to predict prize winners; rather, such indicators simply identify a group of elite scientists (Ashton & Oppenheim, 1978; Garfield, 1986; Gingras & Wallace, 2010). In addition, there seems to be a cascading effect: a Nobel Prize affects not only the citation of the laureate's work, but also that of its cited references, as shown in a case study of Robert J. Aumann (Frandsen & Nicolaisen, 2013). However, when citations are normalized based on the papers' publication year, such an effect is no longer observed (Farys & Wolbring, 2017). Apart from citation impact, studies have also shown that Nobel laureates have fewer coauthors compared with a control group and tend to possess higher social capital, by bridging different communities of researchers (Wagner et al., 2015).

Particularly relevant to the scope of this research are the studies examining pattern changes following prize recognition. Patterns examined to date include collaboration patterns (Chan et al., 2015), citation profiles (Azoulay, Stuart, & Wang, 2013), and the ability to obtain awards (Chan et al., 2014). Studies have found that Nobel laureates form fewer collaborations with new coauthors post-award than pre-award (Chan et al., 2015), and their likelihood to obtain awards drops after they win a Nobel Prize (Chan et al., 2014). In a similar vein, a study showed a post-award citation increase to articles published before their authors earned a Howard Hughes Medical Institute (HHMI) Investigator recognition (Azoulay et al., 2013). This confirms the theory that people change their perceptions of the quality of others' work when there is a shock to the authors' status (such as a highly esteemed award).

This study builds on prior work by using citation sentiment to directly measure the change in perception of a scientific paper after a change to the author's status. The change in this case is the receipt of either of two Nobel Prizes: the Nobel Prize in Chemistry or the Nobel Prize in Physiology or Medicine. The set of citation sentences (i.e., "citances") from citing articles in PubMed Central (PMC) to Nobel laureates' work is extracted and processed for sentiment analysis. We compare the sentiment of citances from before and after the work's author won a Nobel Prize and use a set of control group papers to benchmark the sentiment change. This approach allows us to address one central research question in this paper: to what extent does a change in authors' status—such as winning a Nobel Prize—change the perception of the quality of their work as measured by citation sentiment?

Literature review

Citation analysis is strongly based on the relationship between citing and cited documents. The classic explanation of citation behavior is the normative theory, which regards citations as a means of paying an intellectual debt to the authors being cited (Garfield & Merton, 1979; Kaplan, 1965). Despite its importance in the history of citation analysis, subsequent generations of researchers increasingly questioned the normative assumptions behind this theory (e.g., Bornmann & Daniel, 2008) and accepted a new approach to citation data which focuses on the content and context of citation relation.

This newer program of research was officially named *content and context analysis* in the early 1980s (Small, 1982), defined as including those studies examining the "particular message or statement within the citing document containing the reference" (p. 288). This type of analysis can be traced back to Lipetz's pioneering work (Lipetz, 1965), in which the author identified four categories of relationship between the document pair: original scientific contribution of citing paper, contribution other than original scientific contribution, identity or continuity relationship between papers, and disposition of the scientific contribution of the cited paper to the citing paper. A large number of studies under this category have been conducted since then (Case & Higgins, 2000; Chubin & Moitra, 1975; Duncan, 1981; Frost, 1979; McCain & Salvucci, 2006; Moravcsik & Murugesan, 1975; Spiegel-Rösing, 1977), most of which developed a classification scheme based on a set of scientific documents.

Cronin (1984) has correctly observed that most of these studies fail to form a "cumulative endeavor" (p. 35), despite the fact that some regularities are evident in these studies. His observation is supported by work which identifies and summarizes the facets underlying these classification schemes. An article by Zhang, Ding, & Milojević (2013), for example, identifies six principles embedded in all these studies, including the type of motivation, level of importance, type of resource, function of citing, type of disposition/sentiment, and location of mentioning (p. 21).

Despite its status as one of the central principles of content and context analysis, the sentiment polarity expressed in texts is a relatively late-coming research topic. Although sentiment polarity has been explored by researchers from the perspectives of citation analysis (e.g., MacRoberts & MacRoberts, 1984; Ziman, 1968) and computational linguistics (e.g., Hearst, 1992; Wilks & Bien, 1983), it did not become a popular topic in its own right until the beginning of the 20th century, a phenomenon partly driven by the rise of machine learning methods and the increasing availability of data sets (Pang & Lee, 2008).

As a response to these new developments in computational data science, data-driven sentiment analysis was soon introduced to the field of citation analysis. Awais Athar and colleagues discussed the difficulties of applying sentiment analysis to citation statements and proposed a support vector machine (SVM) method based on different aspects of sentence structure (e.g., Athar, 2011, 2014, Athar & Teufel, 2012a, 2012b). Machine learning techniques are a dominant methodological element in citation sentiment studies. In addition to SVM (see also Hernández-Alvarez & Gómez, 2015; Kim & Thoma, 2015; Xu, Martin, & Mahidadia, 2013), notable examples include random forest (Abu-Jbara, Ezra, & Radev, 2013; Parthasarathy & Tomar, 2014), naïve Bayes (Butt et al., 2015; Sula & Miller, 2014), and neural network methods (Lauscher, Glavaš, Ponzetto, & Eckert, 2017). Within this category, SentiWordNet, a lexical resource for opinion mining that is partly based on a semi-supervised machine learning method, has also been used in a number of studies (Goodarzi, Mahmoudi, & Zamani, 2014; Sendhilkumar, Elakkiya, & Mahalakshmi, 2013).

It should be noted that most of the above-mentioned studies focus primarily on establishing and testing a methodological framework for citation sentiment analysis. With more citation analyses taking the sentiment into consideration, we are also gaining more insights about how social factors help to construct scientific knowledge. For example, before the advancement of computational citation semantic analysis, Small (2011) concluded that there is a correlation between prominent sentiments and competing knowledge claims. On the other hand, a more recent study (Ma, Nam, & Weihe, 2016) has suggested that an author's reputation is a reliable predictor of the sentiment towards one's works. Although these findings contributed to a deeper understanding of citations, we are still far from fully understanding the relationship between citation sentiment and the authors who receive it. Thus, in this paper, we examine the impact of a change to authors' status on the perception of their work as measured by citation sentiment.

Data

Two sets of papers are distinguished, and both are introduced in this section. The first set are the so-called "Nobel papers", written by Nobel Prize laureates and thought to be closely connected with the conferral of their respective Nobel Prizes. The Advanced Information section of each award at www.nobelprize.org served as our data source for identifying Nobel papers. We focused on two awards, the Nobel Prize in

Chemistry and the Nobel Prize in Physiology or Medicine, because papers that cite these Nobel papers are more likely to be included in PubMed Central (PMC)—a large full-text repository freely accessible to the public. Nobel papers, as we define them, are typically referred to as "landmark" papers in the Advanced Information section or were published within a three-year window of the landmark papers by the same teams of authors.

Most papers in PMC were published after 2008; thus, to monitor the impact of an award on citation sentiment change, we narrowed Nobel Prizes to those conferred between 2010 and 2015. This time frame provides a minimum three-year citation window for the oldest (2008) paper in PMC to cite a Nobel paper at the time of data collection. To effectively link a Nobel paper with its citing papers, the Nobel paper should be indexed in PubMed and have a PubMed ID (PMID), but it does not itself need to be included in PMC. After removing those Nobel papers that lack a PMID and those with fewer than 20 citations before their respective Nobel Prize, we included in the final data set 14 papers for Physiology or Medicine (hereafter Medicine) and 11 for Chemistry. The 2018 October version of PMC was used for data collection. In all, these papers received 12,393 citances, of which 4,677 took place before award. The data set used in this study is uploaded to Figshare (Yan, 2019).

The second set contains control-group papers with which to benchmark the sentiment change in the Nobel papers. For each Nobel paper, we selected three control papers that each had a similar citation count to the Nobel paper and were published within two years of the Nobel papers. These selection criteria ensure that the control papers are sufficiently similar to the Nobel papers that the major distinguishing factor is the Nobel Prize. In total, there are 75 papers in the control group, receiving a total of 30,851 citances.

To calculate the number of citations a paper has received in PMC, one can count the number of times that the cited paper's PMID occurred in other papers' cited reference sections. A full-text search does not need be conducted; thus, we refer to this process as citation metadata matching. Tracking down citances within a paper's full text, however, is less straightforward and presents several challenges. The first challenge is to identify the exact sentence where the citance appears and extract all words that surround the citance in order to evaluate sentiment. To achieve this, for every paragraph in the abstract and the document body, the XML nodes for citations (<xref> nodes) were first replaced by a special token, and then the full paragraph text were tokenized and further split by the recognized punctuations. Initially, when only considering the standard XML nodes (<xref> nodes), for our Nobel paper set, this returned less than half of all citations Nobel papers received in PMC as returned by a metadata matching.

Thus, we need to deal with the second challenge which is the lack of consistency in XML citation coding. We provide some examples here.

1) Different encodings are used for the dash character '-' inside a multi-citation instance such as "[11-17]". It may be just simply '-', but can also be '—', or other similar-looking characters with actually different encodings. In addition, a document might use HTML codes to represent the dash, like "-" or "—", "–", "−", etc.
2) Sometimes, an <xref> node is inside another XML node such as <sub> or <sup>. Consequently, after replacing <xref> nodes by tokens, these tokens are still embedded within an XML node. This situation needs additional treatment; otherwise, the citations will be removed when extracting citances.
3) Many documents use incorrect citation types. The correct citation type should be "bibr" or "ref", but many documents put it as "other".

By repeated checking of the discrepancies, experiments and improvement, we were able to reduce the gap between the metadata matching and full-text matching to less than 10%. We further examined the remaining differences and identified four categories of issues that make it impractical to track citances:

1) Older publications that come with PDF but not XML documents (e.g. PMID: 6203731);

2) Citations embedded in tables or figure captions (e.g. the citations of Nobel paper PMID: 957439 by papers PMID: 17428344, PMID: 15960807, and PMID: 21789187);
3) Papers cited in references but were not actually cited in the body of citing documents (e.g. the article PMID: 21619615 has Nobel paper PMID: 17952055 in its references, but no actual citation in the body of the full text); and
4) Unusual XML citation coding, likely due to lack of version control, typos, or other technical issues (e.g. PMID: 27481189 has XML format errors, PMID: 22022248 cited Nobel paper PMID: 17962520 but a wrong PMID is used).

Detailing these obstacles allows us to highlight the challenges of full-text citation analysis while suggesting tentative ways to mitigate some of the issues.

Methods

Citation sentiment classification is a complex task for two main reasons. First, within the genre of scholarly writing, the expression of sentiments tends to be subtle and implicit. Secondly, technical terms can sometime introduce noise to sentiment classification (e.g., *discriminative* models, *support* vector machines). For this research, we based our method upon the state-of-the-art sentiment classifier SenticNet, which we tailored for citation sentiment classification.

SenticNet is a publicly available and widely used semantic and affective resource for concept-level sentiment analysis (Cambria, Poria, Bajpai, & Schuller, 2016). It is built using natural language and statistical methods over a word network. In our analysis, we applied the sentiment polarity values of SenticNet version 4 (Cambria et al., 2016) to process sentences containing citations to Nobel papers, regarding the sentiment of the sentence as the citation sentiment. This version of SenticNet uses two word networks to disambiguate words: one for nouns, called "AffectNet," and one for verbs, called "AffectiveSpace." In AffectNet, nouns used in similar contexts are classified together. AffectNet is built upon a knowledge base called ConceptNet which is a semantic network of concepts and their relationships (e.g., a tabletop *is part of* a table). For instance, cats and dogs are classified as animals and tend to have positive sentiments, whereas sharks and snakes tend to have negative sentiments. AffectiveSpace is also predicated upon ConceptNet; thus, verbs such as "gain" and "increase" are classified together and tend to have similar sentiments. In addition, SenticNet also considers negation, flipping the polarity of terms which are preceded by negation words: "sufficient," for instance, has polarity value 0.63, while "not sufficient" has polarity –0.63. When a negative verb is used together with a negative noun, the outcome is a positive expression, such as "decrease pain" (Cambria et al., 2016). SenticNet's chief merit for our purposes is that it is weakly supervised, requiring minimal training compared to supervised methods (e.g., Athar & Teufel, 2012; Jha, Jbara, Qazvinian, & Radev, 2017; Xu et al., 2015).

We performed additional preprocessing by removing frequently used scientific terms: we first extracted potential scientific terms from the publication texts using a term extraction method developed in our previous work (Chen & Yan, 2017; Yan, Williams, & Chen, 2017), then counted word frequencies in the extracted terms and screened out high-frequency words which were very unlikely to have a sentiment polarity (e.g., "system," "injection," "neuron"). These technical terms 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 screening process, we effectively reduced the noise introduced by technical terms (e.g., "discriminative" and "support" in the contexts mentioned above).

The output of SenticNet consists of a sentiment score between –1 (most negative) and 1 (most positive), with 0 indicating a neutral sentiment. For example, words like "discovery," "pioneer," and "exciting" have high sentiment value (close to 1); when they appear alongside a citation, they serve as evidence of positive citation sentiment. Words such as "however," "only," "not observed," and "not satisfying," in contrast, provide evidence of negative citation sentiment. For evaluation purposes, we consider any score between –1 and –0.2 to reflect a negative sentiment, –0.2~0.2 to be neutral, and 0.2~1 to be positive. We evaluated the performance of the sentiment classification method by randomly selecting 50 citances for each sentiment (positive, neutral, and negative) as classified by the method. We established the following rubrics:

- For citances classified as positive: if human coders identify a citance as simply mentioning a paper, it receives a score of 1 (i.e., 50% accurate). If human coders identify a positive sentiment toward the cited paper, the score is 2. Otherwise, the score is 0.
- For citances classified as negative: if human coders identify a citance as showing a negative sentiment toward a cited paper, it receives a score of 2. Otherwise, the score is 0.
- For citances classified as neutral: if human coders identify a citance as mentioning a paper, it receives a score of 2. Otherwise, the score is 0.

Based on this scoring scheme, the accuracy of positive classification is 71%, while those of negative and neutral classification are 38% and 90% respectively. The average for all three classifications is 67%. The performance of the negative classification is a limitation resulting from our use of a weakly supervised method, but considering the number of negative citations (<4% of all citances coded by the classifier), we do not expect the negative classifications to have a major impact on overall accuracy. Because of the inherent blurriness between positive citances and citances that simply mention one's work, we also tested the performance of a binary classification: non-negative (positive and mention citances) and negative. The accuracy remains at 38% for negative classification, while the accuracy for non-negative classification is 96%. The combined two-class accuracy is 77%. We regard this performance as comparable with that achieved by supervised methods (e.g., Athar, 2011; Athar & Teufel, 2012). The unit of analysis is each citance: if paper A cited paper B *n* times, there will be *n* citances with sentiments for paper B.

Results

Distributions of sentiment scores

In this section, we report the distributions of sentiment scores pre- and post-award. We use violin plots with embedded box plots to visualize the distributions of sentiment scores for aggregated citances to Nobel papers and control-group papers (Figure 1) as well as the distributions of individual Chemistry and Medicine papers (Figure 2 and Figure 3). For these figures, the width of a violin shows the density of a value on the x-axis. In the embedded box plot, the middle vertical bar indicates the median, while the two sides of the box show the 25th and 75th percentiles of the distribution. The ends of the horizontal lines depict the 95% confidence interval.

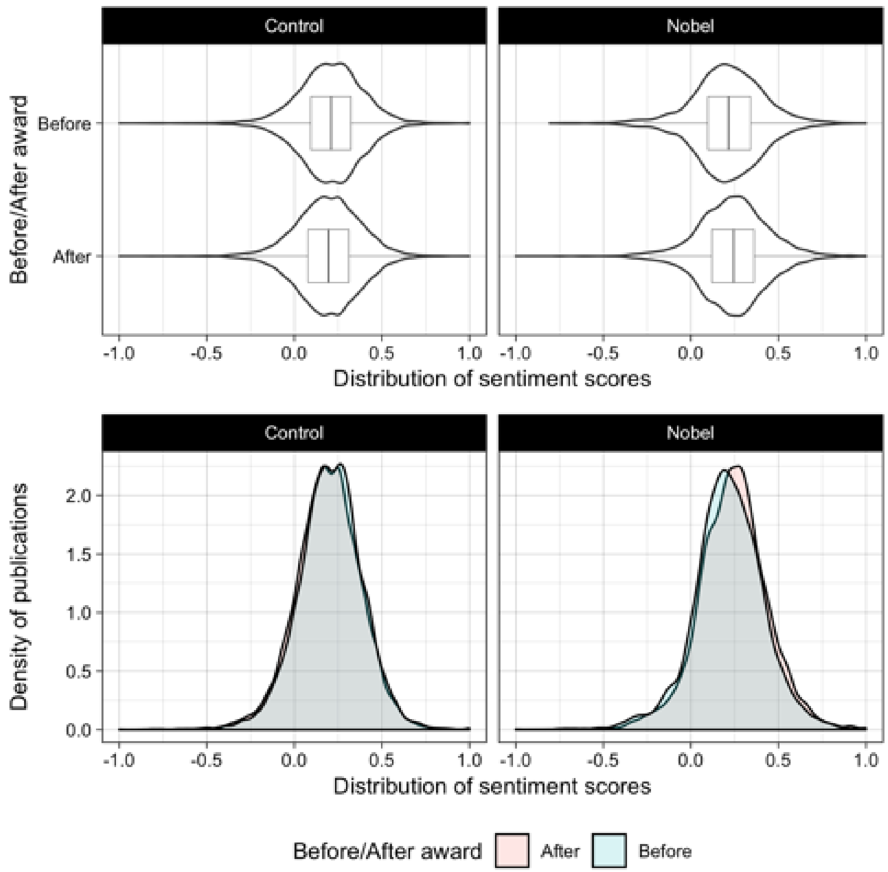

Figure 1. Distribution of sentiment scores for aggregated citances of Nobel and control-group papers before and after award

For control-group papers, the pre- and post-award timeline was the same as their respective Nobel papers'. This allows for direct examinations of the impact of Nobel Prizes on citation sentiment by controlling for award year as a potential artifact. The violin plots in Figure 1 show that for Nobel papers and control group papers, the distributions of sentiment scores before and after award have comparable shapes. The median sentiment score of the 4,677 before-award citances in the Nobel papers group is 0.21; for the 7,716 after-award citances, the median is 0.24 (a 13% increase). Meanwhile, the after-award sentiment scores for the 25th and 75th percentiles (0.12, 0.36) are higher than their before-award counterparts (0.09, 0.34). For the control group, the median sentiment scores of the 14,109 before-award and 16,742 after-award citances are 0.21 and 0.19, respectively (an 8% decrease). The after-award sentiment scores for the 25th and 75th percentiles (0.07, 0.31) are each lower than the corresponding before-award sentiment (0.09, 0.32). The results suggest that there is a moderate sentiment gain for the Nobel papers after their authors win a Prize. The Nobel paper density plot shows a noticeable rightward shift in the distribution of the after-award sentiments (values increased); for control-group papers, there seems instead to be a leftward shift (values decreased).

We employed a two-sample Kolmogorov-Smirnov (K-S) test to verify the sentiment change. The test found a statitical difference between the before- and after-award sentiment for the Nobel papers at the 0.01 level. Another two-sample K-S test indicated a statistical difference between the two distributions for the control-group papers, again at the 0.01 level. The findings suggest an increase in citation sentiment toward Nobel papers after Prize conferral but a decrease in citation sentiment for control-group papers.

We next present violin plots depicting the sentiment-score distribution of individual papers in Chemistry (Figure 2) and Medicine (Figure 3).

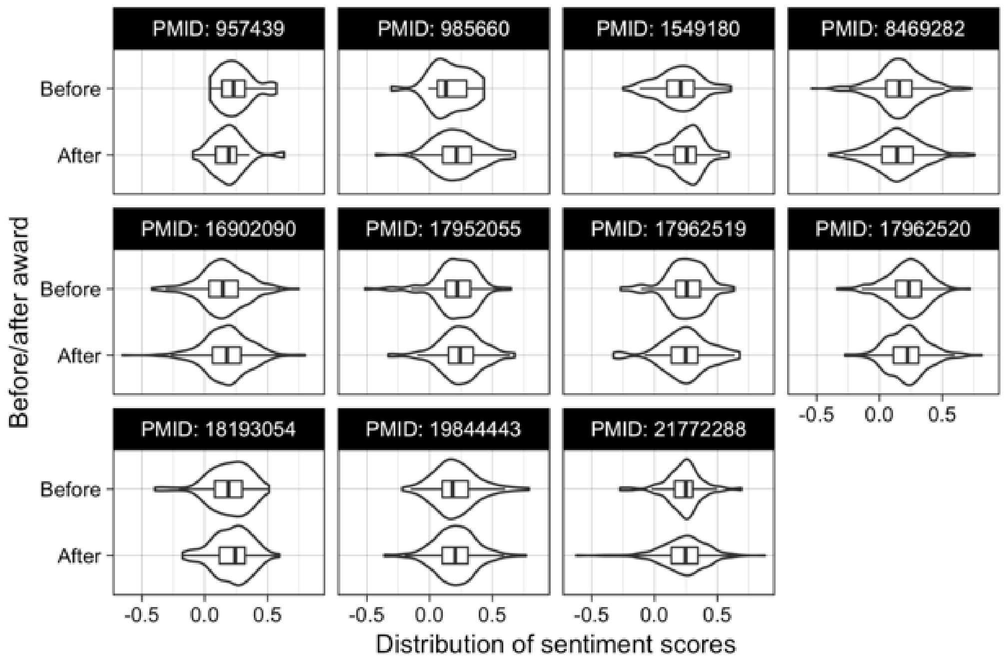

Figure 2. Distribution of sentiment scores for individual Chemistry papers before and after award (papers ranked by PMID)

At the individual paper level, the density of sentiment scores varies between before- and after-award groups. When we examine the shape of the violins for after-award sentiment, there seems to be a higher density of high sentiment scores (PMID: 985660, PMID: 1549180), lower density of low sentiment scores (PMID: 8469282, PMID: 17962520), or in some cases both (PMID: 17952055, PMID: 18193054). There are, however, a few papers whose after-award density has shifted in the low-sentiment direction (PMID: 957439, PMID: 17962519, PMID: 19844443). As for median sentiment score, seven papers out of 11 have an after-award median sentiment higher than the before-award median (PMID: 985660, PMID: 1549180, PMID: 16902090, PMID: 17952055, PMID: 18193054, PMID: 19844443, and PMID: 21772288). We employed a two-sample K-S test and found a non-significant difference between before- and after-award sentiment distributions for all Chemistry papers ($p > 0.05$).

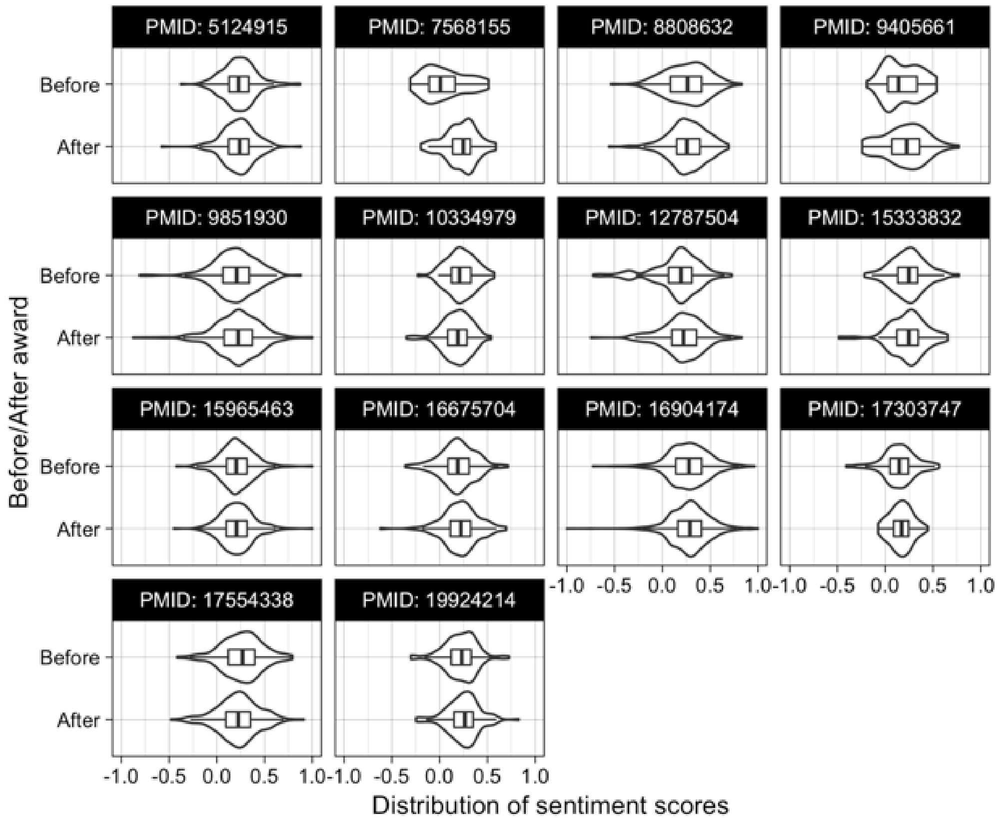

Figure 3. Distribution of sentiment scores for Medicine papers before and after award (papers ranked by PMID)

For Medicine papers, too, we witnessed a change in sentiment between before- and after-award groups: some papers obtained a higher density of high sentiment scores (PMID: 8808632, PMID: 16675704), some had a lower density of low sentiment scores (PMID: 17303747, PMID: 19924214), and some displayed both trends (PMID: 7568155, PMID: 9405661). Eleven of the 14 Medicine papers had an after-award median sentiment higher than the before-award median (all but PMID: 10334979, PMID: 15333832, PMID: 17554338). A two-sample K-S test revealed a statistical difference between before- and after-award sentiment distributions for two papers at the 0.01 level (PMID: 7568155, and PMID: 17554338) and one at the 0.05 level (PMID: 12787504).

Dynamic change of sentiment scores

We now turn to the dynamic change in sentiment scores for Nobel and control-group papers. To reduce the noise from outliers, we used smoothed sentiment scores as the *y*-axis: for year *n*, this is the mean of sentiment scores in years $n - 1$, n, and $n + 1$. Because papers have different award years, we used citation-award year difference as the *x*-axis in Figure 4, where 0 is the award year, –1 is one year before the award year, 1 is one year after the award year, and so forth. We trimmed the time span by only including years with more than 100 citances.

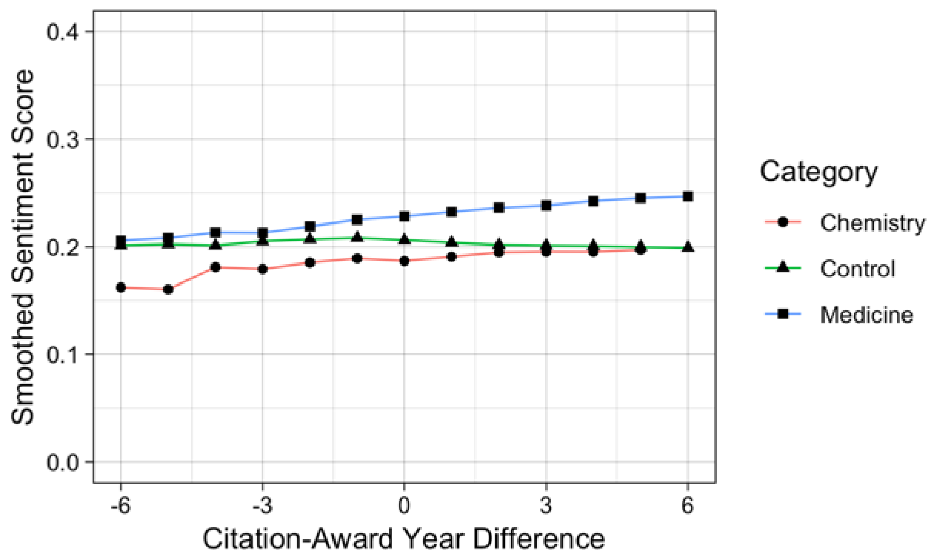

Figure 4. Smoothed citation sentiment score over time for aggregated Nobel and control-group papers, where year 0 is award year

For Nobel papers, Figure 4 shows a near-consistent gain in sentiment scores within both Chemistry and Medicine. Among Chemistry papers, the sentiment score increased from 0.15 to 0.2; for Medicine, the sentiment score increased from 0.2 to 0.25. The increase seems to be independent of the award event itself, as there is no noticeable burst in sentiment immediately post-award. This trend indicates that scientists viewed the Nobel papers increasingly favorably in the decade prior to the award, the conferral of which co-occurred with this dynamic. Overall, the sentiment scores for Medicine papers are consistently higher than the scores for Chemistry papers. Using the citation-award year difference as the independent variable and the smoothed sentiment score as the dependent variable, we estimated the slope of the two regression lines. Medicine has a slightly higher slope than does Chemistry (3.6e–3 vs. 3.1e–3), with both significant at the 0.01 level. Meanwhile, the sentiment for control group papers did not seem to change much (slope = –3e–4, p = 0.18). The differences in citation behaviors between Nobel and control groups suggest that Nobel papers possess certain intrinsic characteristics that propel their sentiment gain, while there is no such drive for control-group papers, whose citation sentiment remains largely unchanged.

We also visualized the sentiment change for individual papers in Chemistry (Figure 5) and Medicine (Figure 6). We used the natural citation year in both figures and highlighted the award year in green. Using the natural citation year as the independent variable and the smoothed sentiment score as the dependent variable, we estimated the slope for individual papers.

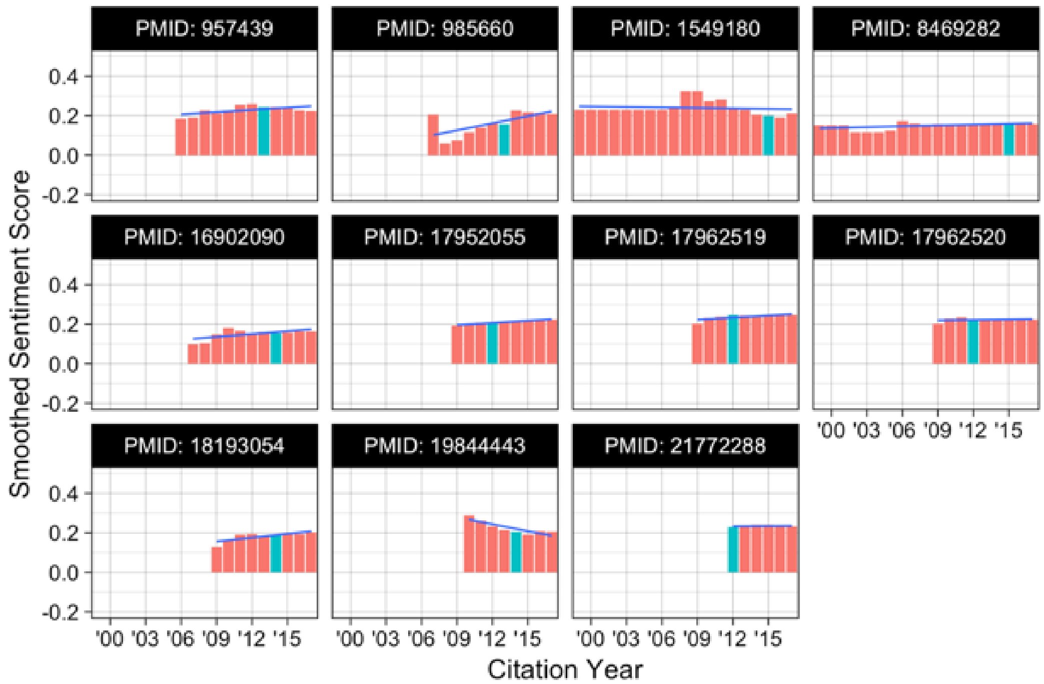

Figure 5. Smoothed citation sentiment score over time for Chemistry papers (green bars indicate award year)

Two of 11 papers in the Chemistry group have negative slopes (PMID: 1549180, PMID: 19844443), whereas the other nine papers have positive slopes. Only one paper has slope greater than 0.01: namely, PMID: 985660 (slope = 0.012, $p < 0.05$). Six papers' slopes are greater than 0.001 but smaller than 0.01 (in descending order): PMID: 18193054, PMID: 16902090, PMID: 957439, PMID: 17952055, PMID: 17962519, and PMID: 8469282, all significant at the 0.05 level.

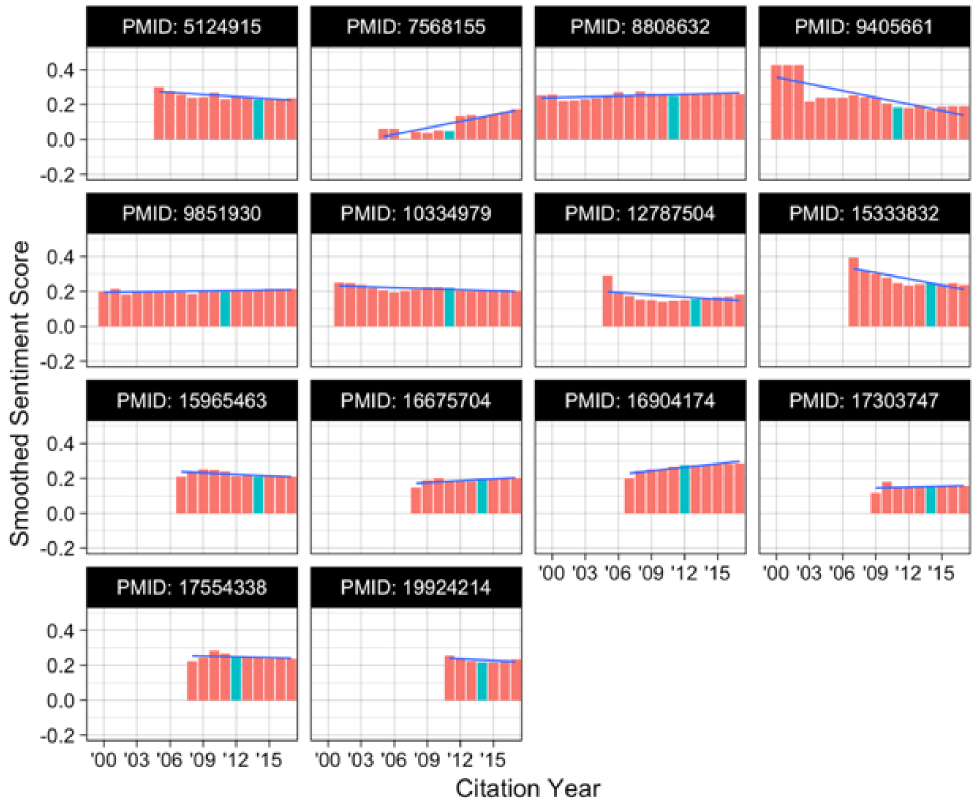

Figure 6. Smoothed citation sentiment score over time for Medicine papers (green bars indicate award year)

Eight of 14 papers in the Medicine group have negative slopes (in descending order): PMID: 17554338, PMID: 10334979, PMID: 19924214, PMID: 5124915, PMID: 12787504, PMID: 15333832, PMID: 9405661, and PMID: 15965463. The first five have slope larger than –0.004. Two papers have slope greater than 0.006; these are PMID: 7568155 (slope = 0.013, $p < 0.01$) and PMID: 16904174 (slope = 0.007, $p < 0.01$). We ran a Spearman rank correlation between slope and a few variables, including number of citations before and after award, year of publication, and year of award. We found a weak, non-significant correlation between slope and year of award ($r = –0.14$). The result suggests that the longer the award-to-citation window is, the greater the citation sentiment increase. This result should be interpreted with caution, because the award may have co-occurred with the upward sentiment trend without causing it, as Figure 4 suggests.

Discussion and conclusion

In this paper, we studied citation sentiments as a function of authors' status change. A full-text data set of PubMed Central publications was used to identify citation sentiments from texts. The receipt of a Nobel Prize in Chemistry, or in Physiology or Medicine, was considered as the status change. In total, 12,393 citances were identified from papers that cited 25 Nobel papers. To benchmark the sentiment change in citations to the Nobel papers, we also selected 75 control-group papers. These papers were cited in 30,851 citances.

We found a moderate increase in citation sentiment toward Nobel papers after their authors won a Prize. At the individual paper level, 7 of the 11 Chemistry papers and 11 of 14 Medicine papers had an after-award median sentiment higher than their before-award median sentiment. Using a two-sample K-S test, we found that the difference was statistically significant for three Medicine papers, but no Chemistry paper. Overall, there was a steady increase in sentiment toward these papers over time; the conferral of a Nobel Prize seemed to co-occur with this trend. We found a weak and non-significant association between the year of award and the citation sentiment toward an award paper: the wider the citation-to-award window is, the higher the observed citation sentiment. One possible explanation is that award papers are of intrinsically high quality and impact, and such attributes lead to two related events: 1) scientists cite those papers with increasing favorability, and 2) the award itself is conferred because of these attributes. These two events co-occur, which is likely the reason that pre- and post-award sentiment growth is rather smooth (as seen in Figure 4). Thus, a Nobel Prize may not directly contribute to growth, but appears alongside it.

The results corroborate the work of Azoulay et al. (2013) and Stuart, Hoang, and Hybels (1999) regarding perception change and uncertainty; these researchers argue that the magnitude of perception change is proportional to the level of uncertainty about the quality of work. In other words, the author's status advancement will have a small impact on the perception of an already well-known piece, because audiences are faced with less uncertainty about the work. On the contrary, for an obscure piece, a status advancement on the part of the author would lead to a more noticeable perception change, because the validation conferred by the award carries more weight. Arguably, because of the lifetime-achievement nature of the Nobel Prize, the level of uncertainty is low and thus the perception change is moderate. Nobel laureates receive other awards before the Nobel Prize; those awards may serve as more direct "shocks" in changing people's perceptions of their work.

One major limitation is the small number of Nobel papers included in this paper. We see this as a largely technical limitation due to the use of PMC as our data source. PMC is one of the most accessible full-text publication repositories, but its coverage is mainly restricted to biomedical publications within the past decade. It is our interest to include Nobel Prizes in other domains and across longer time periods when we have access to a more comprehensive full-text source. We also plan to include junior-level awards as status shocks to investigate their impact in shaping people's opinions about awardees' work.

Acknowledgement
This project was made possible in part by the Institute of Museum and Library Services (Grant Award Number: RE-07-15-0060-15), for the project titled "Building an entity-based research framework to enhance digital services on knowledge discovery and delivery".

References

Abu-Jbara, A., Ezra, J., & Radev, D. (2013). Purpose and polarity of citation: Towards nlp-based bibliometrics. In Proceedings of the 2013 conference of the North American chapter of the association for computational linguistics: Human language technologies (pp. 596–606).

Ashton, S. V., & Oppenheim, C. (1978). A method of predicting Nobel prizewinners in chemistry. Social Studies of Science, 8(3), 341-348.

Athar, A. (2011). Sentiment analysis of citations using sentence structure-based features. In Proceedings of the ACL 2011 student session (pp. 81–87). Association for Computational Linguistics.

Athar, A. (2014). Sentiment analysis of scientific citations. University of Cambridge, Computer Laboratory.

Athar, A., & Teufel, S. (2012a). 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). Association for Computational Linguistics.

Athar, A., & Teufel, S. (2012b). Detection of implicit citations for sentiment detection. In Proceedings of the Workshop on Detecting Structure in Scholarly Discourse (pp. 18–26). Association for Computational Linguistics.

Azoulay, P., Stuart, T., & Wang, Y. (2013). Matthew: Effect or fable? Management Science, 60(1), 92-109.

Barabâsi, A.-L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. Physica A: Statistical Mechanics and its Applications, 311(3), 590-614.

Bensman, S. J. (2008). Distributional differences of the impact factor in the sciences versus the social sciences: An analysis of the probabilistic structure of the 2005 journal citation reports. Journal of the American Society for Information Science and Technology, 59(9), 1366-1382.

Bornmann, L., & Daniel, H.-D. (2008). What do citation counts measure? A review of studies on citing behavior. Journal of Documentation, 64(1), 45–80.

Bothner, M. S., Godart, F. C., & Lee, W. (2010). What is social status? Comparisons and contrasts with cognate concepts. Retrieved from

Butt, B. H., Rafi, M., Jamal, A., Rehman, R. S. U., Alam, S. M. Z., & Alam, M. B. (2015). Classification of research citations (CRC). ArXiv Preprint ArXiv:1506.08966.

Cambria, E., Poria, S., Bajpai, R., & Schuller, B. (2016). SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives. Paper presented at the Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers.

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.

Chan, H. F., Gleeson, L., & Torgler, B. (2014). Awards before and after the Nobel Prize: A Matthew effect and/or a ticket to one's own funeral? Research Evaluation, 23(3), 210-220.

Chan, H. F., Önder, A. S., & Torgler, B. (2015). Do Nobel laureates change their patterns of collaboration following prize reception? Scientometrics, 105(3), 2215-2235.

Chen, Z., & Yan, E. (2017). Domain-independent term extraction & term network for scientific publications. iConference 2017 Proceedings.

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.

Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. SIAM review, 51(4), 661-703.

Cronin, B. (1984). The citation process. The role and significance of citations in scientific communication. London: Taylor Graham, 1984, 1. Retrieved from http://adsabs.harvard.edu/abs/1984cprs.book

Duncan, E. B. (1981). Qualified Citation Indexing: Its Relevance to Educational Technology.

Farys, R., & Wolbring, T. (2017). Matched control groups for modeling events in citation data: An illustration of nobel prize effects in citation networks. Journal of the Association for Information Science and Technology, 68(9), 2201-2210.

Frandsen, T. F., & Nicolaisen, J. (2013). The ripple effect: Citation chain reactions of a nobel prize. Journal of the Association for Information Science and Technology, 64(3), 437-447.

Frost, C. O. (1979). The use of citations in literary research: A preliminary classification of citation functions. The Library Quarterly, 399–414.

Garfield, E. (1986). Do Nobel-Prize winners write citation-classics. Current contents(23), 3-8.

Garfield, E., & Merton, R. K. (1979). Citation indexing: Its theory and application in science, technology, and humanities (Vol. 8). Wiley New York. Retrieved from http://www.garfield.library.upenn.edu/cifwd.html

George, G., Dahlander, L., Graffin, S. D., & Sim, S. (2016). Reputation and status: Expanding the role of social evaluations in management research. Academy of Management Journal, 59(1), 1-13.

Gingras, Y., & Wallace, M. L. (2010). Why it has become more difficult to predict Nobel Prize winners: a bibliometric analysis of nominees and winners of the chemistry and physics prizes (1901–2007). Scientometrics, 82(2), 401-412.

Goodarzi, M., Mahmoudi, M. T., & Zamani, R. (2014). A framework for sentiment analysis on schema-based research content via lexica analysis. In Telecommunications (IST), 2014 7th International Symposium on (pp. 405–411). IEEE.

Harley, B. (1999). The myth of empowerment: work organisation, hierarchy and employee autonomy in contemporary Australian workplaces. Work, Employment and Society, 13(1), 041-066.
Hearst, M. A. (1992). Direction-based text interpretation as an information access refinement. Text-Based Intelligent Systems: Current Research and Practice in Information Extraction and Retrieval, 257–274.
Hernández-Alvarez, M., & Gómez, J. M. (2015). Citation impact categorization: for scientific literature. In Computational Science and Engineering (CSE), 2015 IEEE 18th International Conference on (pp. 307–313). IEEE.
Jha, R., Jbara, A.-A., Qazvinian, V., & Radev, D. R. (2017). NLP-driven citation analysis for scientometrics. Natural Language Engineering, 23(1), 93-130.
Kaplan, N. (1965). The norms of citation behavior: Prolegomena to the footnote. American Documentation, 16(3), 179–184.
Kim, I. C., & Thoma, G. R. (2015). Automated classification of author's sentiments in citation using machine learning techniques: A preliminary study. In Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2015 IEEE Conference on (pp. 1–7). IEEE.
Kuhn, T. S. (2012). The structure of scientific revolutions. University of Chicago press.
Lauscher, A., Glavaš, G., Ponzetto, S. P., & Eckert, K. (2017). Investigating convolutional networks and domain-specific embeddings for semantic classification of citations. In ACM international conference proceedings series: AICPS (p. tba). ACM; Mannheim Univ. Press.
Lipetz, B.-A. (1965). Improvement of the selectivity of citation indexes to science literature through inclusion of citation relationship indicators. American Documentation, 16(2), 81–90. https://doi.org/10.1002/asi.5090160207
Ma, Z., Nam, J., & Weihe, K. (2016). Improve sentiment analysis of citations with author modelling. In Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (pp. 122–127).
MacRoberts, M. H., & MacRoberts, B. R. (1984). The negational reference: Or the art of dissembling. Social Studies of Science, 91–94.
McCain, K. W., & Salvucci, L. J. (2006). How influential is Brooks' law? A longitudinal citation context analysis of Frederick Brooks' The Mythical Man-Month. Journal of Information Science, 32(3), 277–295.
Merton, R. K. (1968). The Matthew effect in science. Science, 159(3810), 56-63.
Moravcsik, M. J., & Murugesan, P. (1975). Some results on the function and quality of citations. Social Studies of Science, 5(1), 86–92.
Newman, M. E. (2001a). Scientific collaboration networks. I. Network construction and fundamental results. Physical Review E, 64(1), 016131.
Newman, M. E. (2001b). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Physical Review E, 64(1), 016132.
Newman, M. E. (2001c). The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences, 98(2), 404-409.
Newman, M. E. (2004). Coauthorship networks and patterns of scientific collaboration. Proceedings of the National Academy of Sciences, 101(suppl 1), 5200-5205.
Newman, M. E. (2005). Power laws, Pareto distributions and Zipf's law. Contemporary physics, 46(5), 323-351.
Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2(1–2), 1–135.
Parthasarathy, G., & Tomar, D. C. (2014). Sentiment analyzer: Analysis of journal citations from citation databases. In Confluence The Next Generation Information Technology Summit (Confluence), 2014 5th International Conference- (pp. 923–928). IEEE.
Patterson, C. J., Kupersmidt, J. B., & Griesler, P. C. (1990). Children's perceptions of self and of relationships with others as a function of sociometric status. Child development, 61(5), 1335-1349.

Perc, M. (2014). The Matthew effect in empirical data. Journal of The Royal Society Interface, 11(98), 20140378.
Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. The European Physical Journal B-Condensed Matter and Complex Systems, 4(2), 131-134.
Seglen, P. O. (1992). The skewness of science. Journal of the American Society for Information Science, 43(9), 628.
Sendhilkumar, S., Elakkiya, E., & Mahalakshmi, G. S. (2013). Citation semantic based approaches to identify article quality. In Proceedings of international conference ICCSEA (pp. 411–420).
Small, H. (1982). Citation context analysis. Progress in Communication Sciences, 3, 287–310.
Small, H. (2011). Interpreting maps of science using citation context sentiments: a preliminary investigation. Scientometrics, 87(2), 373–388.
Spiegel-Rösing, I. (1977). Science studies: Bibliometric and content analysis. Social Studies of Science, 97–113.
Stringer, M. J., Sales-Pardo, M., & Amaral, L. A. N. (2010). Statistical validation of a global model for the distribution of the ultimate number of citations accrued by papers published in a scientific journal. Journal of the Association for Information Science and Technology, 61(7), 1377-1385.
Stuart, T. E., Hoang, H., & Hybels, R. C. (1999). Interorganizational endorsements and the performance of entrepreneurial ventures. Administrative Science Quarterly, 44(2), 315-349.
Sula, C. A., & Miller, M. (2014). Citations, contexts, and humanistic discourse: Toward automatic extraction and classification. Literary and Linguistic Computing, 29(3), 452–464.
Tatsioni, A., Vavva, E., & Ioannidis, J. P. (2010). Sources of funding for Nobel Prize-winning work: public or private? The FASEB journal, 24(5), 1335-1339.
Wagner, C. S., Horlings, E., Whetsell, T. A., Mattsson, P., & Nordqvist, K. (2015). Do Nobel Laureates create prize-winning networks? An analysis of collaborative research in physiology or medicine. PloS one, 10(7), e0134164.
Wilks, Y., & Bien, J. (1983). Beliefs, points of view, and multiple environments. Cognitive Science, 7(2), 95–119.
Xu, H., Martin, E., & Mahidadia, A. (2013). Using heterogeneous features for scientific citation classification. In Proceedings of the 13th conference of the Pacific Association for Computational Linguistics.
Xu, J., Zhang, Y., Wu, Y., Wang, J., Dong, X., & Xu, H. (2015). Citation sentiment analysis in clinical trial papers. Paper presented at the AMIA Annual Symposium Proceedings.
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.
Yan, E. (2019). Nobel Prize papers citation sentiment (Version 2). figshare. doi:10.6084/m9.figshare.7837517.v2
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(7), 1490–1503.
Ziman, J. M. (1968). Public knowledge: An essay concerning the social dimension of science (Vol. 519). CUP Archive.