

Deriving the Impact of Scientific Publications by Mining Citation Opinion Terms

Sofia Stamou

Nikos Mpouloumpasis

Lefteris Kozanidis

Computer Engineering and Informatics Department, Patras University, 26500 GREECE
{stamou, mpouloum, kozanid}@ceid.upatras.gr

ABSTRACT

One of the most important measures for estimating the impact of scientific publications is the number of citations they have received. Today, there exist several tools and metrics to evaluate the relative importance of individual papers and publication venues based on their citation distribution. Despite their acknowledged usefulness, most of the existing techniques rely on quantitative rather than qualitative aspects of the citation analysis and thus they are inherently limited in conveying any specific information about the author opinions towards the papers they cite. In this paper, we introduce a method that combines text mining and lexical analysis in order to elucidate the authors' attitude towards the works they cite in their publications. We have applied our method on a set of 4,520 citations that span to 40 publications and tried to shed some light on the following issues. How often authors express an opinion about the papers they cite? Do all authors who cite the same publication share a common understanding on that publication's impact? Does the citations' context influence people's perception of the referred papers and how we can take context into consideration? Can we define a qualitative measure for estimating the impact of scientific publications? Our evaluation shows that although authors do not always express their personal opinions about the papers they cite, their judgments (when articulated) have a great influence on the papers' perceived importance.

Categories and Subject Descriptors

H.3.1 [Content Analysis and Indexing]: Linguistic Processing; I.2.7 [Natural Language Processing]: Text Analysis; H.2.8 [Database Applications]: Data Mining.

General Terms

Measurement, Experimentation, Human Factors.

Keywords

Opinion mining, citation impact, sentiment analysis, text mining.

1. INTRODUCTION

The most widely employed measure for quantifying the impact of scientific publications is the number of citations they receive, i.e. how often they are refereed by other publications. As of today, there exist numerous citation indexing systems, such as Google Scholar¹, CiteSeer², Scopus³, which store scientific publications interlinked to their citations and enable users to search their contents for obtaining the complete citation record of a given research article or author.

Over the last decade, citation analysis has attracted the interest of many researchers who are in need of some perceptible evidence for quantifying the impact of scientific publications, scientists and publication venues. In this respect, there exist several methods that rely on the citations' distribution in order to: (i) assess the merits and popularity of research material (e.g. bibliometrics [5]), (ii) evaluate the scientific productivity of individuals or working groups (e.g. the *h-index* [9]), and (iii) approximate the importance of a publication venue (e.g. impact factor [1]). The commonality in all existing citation analysis tools and methods is that they mainly examine the statistical distribution of paper citations for quantifying their impact and they are less concerned about the underlying authors' attitude towards the work they cite. By author attitude we refer to the stance that the author takes towards the contents of a prior work that she cites in her publications.

In this paper, we introduce a hybrid method that combines text mining and lexical analysis in order to derive from the citations of a scholar paper the opinions of the authors who cite it towards its content. To make our discussion concrete, we distinguish two citation orientations: (i) supportive, i.e. references to publications that reveal the author's positive stance towards the publications' contents and (ii) unsupportive, i.e. references to publications that reveal the author's negative stance towards the publications' contents. Based on the author opinions mined from the contextual elements of their references to prior works, we

¹ <http://scholar.google.com/>

² <http://citeseer.ist.psu.edu/>

³ <http://www.scopus.com/scopus/home.url>

present a new method for assessing the impact of scientific publications based on both the citations' distribution and orientation.

In the present work, we focus on conference publications pertaining to the discipline of computer science and we demonstrate the value of our method using a dataset of 40 scientific publications and 4,520 citations they have received altogether. We show that nearly 28% of the citations that research papers receive do express author opinions about the cited works. Moreover, we observe that most of the opinions expressed are supportive in nature; implying some reluctance on behalf of the authors towards pointing out negative aspects about the papers they cite. Experimental results reveal that there is considerable agreement in the author opinions towards the cited papers and that our method can capture pretty accurately the citations' orientation.

The remainder of the paper is organized as follows. We first briefly discuss previous attempts to evaluate the impact of scientific publications. In Section 3, we present our methodology that combines lexical analysis and text mining for processing the citations' contents in order to derive the author opinions for the cited works. We also show how the derived opinions can be encapsulated in an objective measure for quantifying the impact of scientific publications. In Section 4, we present the experimental evaluation of our proposed measure. In Section 5, we conclude the paper and discuss avenues for future work.

2. RELATED WORK

Related work falls into two directions, i.e. citation analysis and sentiment analysis. Citation analysis has attracted the interest of many researchers who try to quantify the impact of research works in an objective manner [14] [17]. Citation analysis basically involves counting the number of times a scientific paper or scientist is cited and it operates on the assumption that influential scientists and important works will be cited more frequently than others. In this direction, there exist quite a few citation analysis tools (e.g. CiteSeer, Scopus, etc.) that allow automatic extraction and grouping of citations for online research documents. However, citation analysis methods have been criticized about giving credit to works that are refereed in a negative way [13], i.e. their citing articles point out limitations and insufficiencies.

Another measure widely used in citation analysis is the impact factor [1], which counts the citations from articles of thousands of journals in order to estimate the journals' quality. Again the impact factor has several weaknesses such as the fact that a few highly cited articles bias the scores of impact factors or that impact factor scores do not take into account articles that were used but not cited in a publication. To get around some of the above limitations, the *h-index* metric [9] has been introduced that is based on the set of a scientist's most cited research papers and the number of citations that they have received in other people's publications. Again the main drawback of the *h-index* is that it counts negative citations to incorrect works and that it is insensitive to highly cited works [4]. To tackle such problems, several variations of the *h-index* have been proposed, the most known of which are the *a-index* [3] and the *g-index* [6]. For a detailed overview on the different citation analysis measures and approaches, we refer the interested reader to the work of [13].

Citation analysis has also been addressed in the context of publications' impact prediction. Recently, [20] introduced a new approach for selecting which scientific papers should be published in a venue based on their expected citations, as the latter are determined via citation auctions. The idea is that the number of citations an author has received from her previously published papers can serve as a predictor of the impact (i.e. expected citations) of that her currently unpublished paper. In a different approach, [21] introduced a method for using citations among technical documents in a collection of scientific publications in order to evaluate the quality of the discovered document relations, obtained results showed that the proposed method is capable of evaluating the following aspects of document relations: soft/hard scoring, direct/indirect citations, relative quality over expected validity and comparison to human judgments.

The inherent limitation, though, in all existing methods is that they ignore why a scientific publication was cited in the first place and because of that they cannot really assess the true impact of research works. Our proposed method tries to overcome this limitation by exploring the contextual elements that surround references to prior works in order to detect the underlying authors' attitude towards the works they cite. In other words, our approach suggests the integration of citation and sentiment analysis into a single method for assessing the publications' impact.

Sentiment analysis refers to a complex process that leverages text mining and natural language processing techniques in order to determine the attitude of a writer with respect to some topic she writes about. Existing works on sentiment analysis concentrate on the identification of people's opinions regarding online products [2] or news media [12], the determination of opinion polarity in online reviews [16], as well as on analyzing the correlation between online mentions of a product and sales of that product [11].

While most of existing works on sentiment analysis share a common goal with our current study, i.e. to decipher the writers' opinion about the subjects of their writings, nevertheless our approach does differ in several significant ways. First and foremost, our method tries to identify the authors' opinions by relying on the analysis of scientific publications, which are much different in both language usage and structure from online product reviews. That is, in online product reviews customers express their personal opinions without any space constraints. On the other hand, authors who express an opinion about a cited work have to consider space limitations as they would like to keep most of their publication space available for presenting their own work. Secondly, scientists scrutinize their writings as these are reviewed by field experts before being accepted for publication. Therefore, authors are more selective in the vocabulary they use for articulating their opinions and they are often reluctant in

giving negative citations, especially if they are young researchers or if the articles they cite are deemed of high impact. As a consequence the sentiment analysis of the text surrounding a citation becomes more difficult given the limited size of explicit author opinions.

The only existing work that addresses the issue of citation opinion mining is [15]. However, the model that the authors describe is rather generic and does not reveal any specific approach towards sentiment analysis. Moreover, the proposed model has not been evaluated in terms of effectiveness or usage. Motivated by that work, in our study we have designed and implemented a method for identifying the authors' opinions in their scientific writings. Most importantly, we suggest a novel measure to evaluate the impact of scientific publications based on both the sentiment orientation and the distribution of their citations. The details of our method are discussed next.

3. MINING CITATIONS CONTEXT

In this section, we describe our approach for assessing the true impact of scientific publications based on the opinions of the researchers who cite them. In Section 3.1, we present our approach for identifying the opinions of the authors towards the prior works to which they refer in their publications. Then, in Section 3.2 we show how to encapsulate the author opinions in the evaluation of the publications' impact.

3.1 Identifying Author Opinions

Given a number of citations that a particular paper has received, we firstly download the citations' contents and we pre-process them in order to extract the text nuggets that surround references to the article under consideration. Citations' pre-processing accounts to converting their textual content into plain text and applying tokenization in order to identify the boundaries of the text sentences. We then take the sentences containing citations to the article under examination and we process them in order to evaluate the sentiment of the citations that the paper has received. At this point we should note that for clarity issues we focus our discussion on how to derive the citations' sentiment for a single publication. However, as we discuss later our method can be applied for mining citation opinions for multiple publications.

Based on the set of sentences that refer to a given article, we want to identify the authors' opinions about the article's impact. To do so, we rely on the notion of words' *semantic orientation*⁴, introduced in [8]; a factor used to discriminate between words of positive and negative sentiments. The main idea is that not all words in a text serve as good indicators of the text's polarity. Thus, we rely on the observations of [8] and [18] that people use adjectives to evaluate an item or verbalize an opinion and we try to infer the orientation of the citation sentences as follows.

We firstly pass a publication's citation sentences through a Part-of-Speech tagger in order to extract the adjectives that the authors used for expressing their opinions about the work they cite. Having collected all the adjectives used by all the authors who have referred to a given work, we map them to their corresponding WordNet [19] nodes in order to derive their semantic associations. WordNet encodes two basic semantic relations for the adjectives it contains, namely *similar to* and *antonymy*. Based on these two relation types, we group the adjectives extracted from the citation sentences into three clusters. The first cluster groups pairs of adjectives that according to WordNet are similar to each other. For instance adjective pairs like "*interesting/ intriguing*", "*erroneous/ incorrect*" are listed in the first cluster. The second cluster groups pairs of adjectives that are antonyms to one another. As example pairs consider "*high/ low*" and "*increased/ decreased*". Finally, the third cluster groups the remaining adjectives, i.e. those which are neither similar nor opposite to any other adjective in the citation sentences. Note that, under our approach a given adjective might belong to one or two clusters, e.g. "*low*" might belong to cluster one as similar to "*low-level*" and to cluster two as antonym of "*high*". Note also that in order for an adjective to be listed in any of the first two clusters its similar or opposite adjective has to appear in the citation sentences as well, otherwise it is listed in the third cluster.

Having grouped adjectives in the above clusters, the next step is to investigate which of the adjectives communicate some implicit information about the authors' opinions and which do not. In this respect, we firstly rely on FrameNet [7] in order to obtain the semantic frame of every adjective extracted from the citation sentences. Based on the identified frames, we exclude from the clusters those adjectives that have been assigned the frames: *Increment* (e.g. *another, additional, further*), *Relative Time* (e.g. *recent, previous, early*), or *Similarity* (e.g. *different, like, variant*), based on the intuition that these are not indicative of the authors' opinions about the work they cite. Note that once an adjective is removed from either cluster one or two, its counterpart, i.e. similar or opposite adjective, is also eliminated.

Following on from that, we manually assign an orientation label, either positive (+) or negative (-), to each of the remaining adjectives in our clusters. The criterion under which the adjectives' labeling takes place is that a positive adjective is one that gives praise to the cited work, whereas a negative adjective is one that criticizes some or all aspects of the cited work. Note that the manual labeling of adjectives is a fairly easy task since there are only a few distinct adjectives that express the authors' opinions about the cited works in their publications, and it is also objective in the sense that there is extremely high inter-annotations agreement for the semantic orientation of adjectives [8].

The result of the above method is a set of positive and negative adjectives that correspond to the opinions of the authors who cite a paper towards its content. We refer to these adjective sets as *opinion terms*. Opinion terms help us deduce the reasons why scientists cite a given work but most importantly they help us estimate the true impact of scientific papers. Having deter-

⁴ It is also known as valence in the linguistics literature.

mined the polarity of opinion terms that surround references to a given publication, the next step is to assess the strength of the author opinions about the works they cite. In this direction, we rely on the citations of a paper that contain opinion terms and we compute for every citation that expresses an opinion about the cited work, the strength of that opinion, i.e. $Strength(O_t)$ as:

$$Strength(O_{t_i}) = \frac{1}{|O_t(C_s)|} \sum_{i=1, i \in O_t(C_s)}^{|O_t(C_s)|} |O_{t_i}(C_s)| \quad (1)$$

Where $O_{t_i}(C_s)$ is the number of times an opinion term t_i appears in the citation sentences (C_s) that refer to a paper and $O_t(C_s)$ is the number of all opinion terms contained in the citation sentences of the considered paper. Based on the above formula, we can estimate for every opinion term that appears in the contextual elements of a citation, the strength of the underlying author opinion about the cited work. Opinion strength scores are normalized taking values between 1 and 0; with values close to one indicating that a given opinion is globally shared by the authors who cite a given work and values close to zero indicating that the underlying opinion is only representing a few individuals' stance towards the cited work. At the end of this process, we represent the sentiment of every citation that a paper has received as a set of one or more polarity adjectives each one accompanied by a *strength* value indicating the degree to which the opinions expressed are commonly shared across researchers.

As a final step, we rely on the opinion terms that are contained in a citation in order to derive the citation's polarity. To that end, we examine the polarity of the identified opinion terms and work as follows. If all opinion terms in the citation of a paper have the same polarity, that polarity is attributed to the citation's orientation. Conversely, if the opinion terms in a citation have distinct polarity labels, then we compute the degree to which the citation has a positive and a negative orientation as follows. We rely on the average strength values of the positive opinion terms to deduce the degree to which the citation has a positive orientation and on the average strength values of the negative opinion terms to deduce the degree to which the citation has a negative orientation. We then select the polarity of the highest average strength value to indicate the orientation of the citation. This way, we end up with the sentiment orientation (either positive or negative) of every citation that refers to a given work. Next, we discuss how we encapsulate the derived sentiment orientations of a paper's citations while measuring the paper's impact.

3.2 A Quality Measure of the Publications' Impact

Based on the polarity and the average strength scores of the authors' opinion terms about the publications they cite, we assign to each citation a positive or negative orientation accordingly, associated with an overall strength value that indicates the degree to which authors who cite the same paper share a common opinion (verbalized via opinion terms) about that paper's value. Then, we rely on the identified citations' orientation to measure the true impact of the papers. Before describing our impact measure, let's point out that in our work we deem the citations that have neither positive nor negative orientation as *neutral*. Neutral citations are references to related papers that are not revealing through opinion terms the authors' stance towards the papers' contents.

The first step towards measuring the citations' impact is to exclude from our estimation self-citations. Self-citations are references that an author provides in a paper to other papers written by herself. According to [10] self-citation is a neutral form of emphasizing a writer's personal contribution to a piece of research and strengthening her research credibility. Thus, self-citations express biased author opinions for their refereed works and as such we remove them from the measurement of the publications' impact. Now, based on the orientation of the citations that a scientific publication p has received from the works of other researchers we estimate the publication's impact, denoted as $I(p)$, as follows:

$$I(p) = \frac{|C(p)| + |CPos(p)| \cdot Avg.Strength(CPos(p))}{|C(p)| + |CNeg(p)| \cdot Avg.Strength(CNeg(p))} \quad (2)$$

Where $C(p)$ denotes the total number of citations that publication p has received, $CPos(p)$ denotes the number of citations to p with a positive orientation, $CNeg(p)$ denotes the number of citations to p with a negative orientation and $Avg.Strength$ indicates the degree to which each orientation is globally perceived to be indicative of the paper's impact by the authors who have cited p in their works. $Avg.Strength$ values are computed by taking the average strength scores of the citations' opinion terms.

According to our formula, the higher the $I(p)$ value of a paper, the increased its perceived impact. Note that $I(p) > 1$ indicates that a publication has received more positive than negative citations, while $I(p) < 1$ indicates that negative citations are more than the positive ones. Finally, $I(p) = 1$ suggest that either a publication has received an equal number of positive and negative citations or that it has only received neutral citations.

Based on the above formula, we compute the overall impact of a paper based not only on quantitative criteria but by considering also the strength of the author opinions expressed in their citations of that paper. Note that under our formula, a paper with more positive than negative or with positive only citations has an increased impact compared to a paper with the same number of only neutral citations. The intuition behind our formula is that a work praised in the publications of others should be valued higher than a work which is simply referred as being relevant to a particular study. Based on our citations' impact measure, we

can derive the true value of a scientific paper as this is determined not only from the citations' distribution but also by their orientation.

4. EXPERIMENTAL EVALUATION

In this section, we validate our technique on a dataset of scientific publications drawn from Google Scholar. We first describe our dataset and how it was processed for identifying the author opinions towards the cited papers. In Section 4.2 we report obtained statistics concerning the sentiment of the citations examined for our experimental papers. We discuss obtained results about the authors' citing trends in an attempt to shed some light on how people verbalize their references to relevant works. In Section 4.3 we present the details of a human study we carried out in order to validate our method's accuracy in determining the polarity of the citations considered in our previous experiment. Finally, in Section 4.4 we compare the impact that our $I(p)$ metric computes for each of the examined papers to their baseline impact that existing methods estimate. By cross-examining obtained results, we draw some conclusions concerning the functionality and the underlying contribution of our proposed paper impact metric.

4.1 Experimental Setup

To collect our experimental data we relied on Google Scholar out of which we obtained the citations of 40 research papers that relate to web search personalization. To gather our dataset, we issued the query "web search personalization" to Google Scholar and we picked 40 research publications from the returned results (as of July 2008). Our selection of the research publications took place on the sole requirement that all selected works should be conference or workshop papers published after 2000⁵. Out of all the returned papers that meet the above requirement, we randomly picked 40 papers and we collected all their citations. Having decided on the research papers the citations of which we would examine for deriving their impact, we performed data cleaning in order to retain a single version for every examined paper and eliminate paper citations that concern online books, demos, Ph.D. thesis, patents, technical reports, self-citations as well as citations in languages other than English. After cleaning our experimental data, we ended up with a total set of 4,520 citations that our 40 experimental papers have received altogether.

Thereafter, we downloaded the above set of citations for each of our experimental papers; we converted their files to plain text format and processed their contents in order to extract their citation sentences. In this respect, we used the Biblio-Citation-Parser-1.10⁶ and linked each of the sentences containing references to prior works to their corresponding referred paper from the 40 papers considered. Note that in case a citation sentence contains consecutive references to multiple works, we assume that its content is equally attributed to all referring papers. On the other hand, if the references in a given sentence discuss prior works sequentially (i.e. a few sentence elements refer to one prior cited work and right after the citation there are some more elements referring to another cited work) then we consider that each citation sentence ends at the point the reference to a prior work appears.

Following on, we passed the extracted citation sentences through a Part-of-Speech tagger in order to annotate sentence word tokens with the tags that correspond to their grammatical categories. We then relied on the sentences' adjectives to which we applied our method presented in Section 3.1 in order to identify the adjectives that correspond to opinion terms. Identified opinion terms were manually labeled with polarity tags, indicating their sentiment orientations. Based on the polarity of opinion terms, we deduced the orientation of their corresponding citation sentences as previously discussed and we estimated for every citation sentence that communicates some author opinion about a cited work, the strength of that opinion. Recall, that citation sentences with no opinion terms are deemed to be neutral in the sense that they simply refer to related works without commenting their contents and/ or findings. At the end of this processing step, we deduced the sentiment of every citation considered for each of the 40 papers and we measured the degree to which researchers agree on the citations' derived sentiments.

4.2 Sentiment Analysis of Paper Citations

To elucidate the perceived impact of our 40 experimental papers, we relied on their citations' derived polarity and we estimated for every paper the fraction of its neutral, positive, negative and mixed citations. Obtained results, listed in Table 1, help us understand how authors verbalize their references to related works. Note that the increased number of citation sentences compared to the total citations examined is because a paper might be referred more than once in the same publication.

⁵ This is in order to assess the impact of relatively current works.

⁶ <http://search.cpan.org/~mjewell/Biblio-Citation-Parser-1.10/>

Table 1. Statistics on the experimental data

Examined papers	40
# of citations	4,520
# of citation sentences	8,136
# of citations with opinion terms	1,266
# of opinion terms	1,285
# of distinct opinion terms with positive polarity	56
# of distinct opinion terms with negative polarity	14
% of neutral only citations	72%
% of positive only citations	17.8%
% of negative only citations	1.2%
% of mostly positive citations	3.9%
% of mostly negative citations	5.1%

Obtained results indicate that the majority, i.e. 72%, of the citations examined are neutral in nature as they do not contain any opinion terms about the cited works. This primarily indicates that authors refer to the works of others simply to summarize relevant literature without intending to reveal their personal stance towards the contents of the cited works. However, when they express their opinions, they mostly highlight positive aspects of the referred works. According to the obtained results 21.7% of the total citations considered and 77.7% of the citations that do express some author opinions have a positive orientation. This indicates that when authors decide to *comment* a cited paper, they usually do so in order to underline some positive aspects of the paper. On the other hand, the fraction of citations with negative orientation is relatively low and amounts to 6.3% of the total citations examined and to 22.3% of the citations that communicate some author opinions about the referred works. This confirms our intuition that authors are generally reluctant in criticizing the work of others in their scientific writings.

Another observation is that 9% of the paper citations contain opinion terms of both positive and negative polarity. This suggests on the one hand that different authors cite the same paper for different reasons and on the other that not all authors who cite the same publication share a common understanding of that publication's impact. Nevertheless, the majority of the citations that do express some author opinions about the cited papers (i.e. 67.8% of the citations that contain opinion terms) demonstrate a clear orientation, either positive or negative and this orientation is globally shared by the authors who refer to the same papers in their writings.

Based on the above findings, we observe that 28% of the citations that scholar papers receive communicate some author opinions about the contents of the cited works. This justifies the need for a qualitative paper impact metric that accounts not only for their citations' distribution but also for their citations' sentiment orientation.

Another observation is that there is little variation in the terminology (i.e. adjectives) that authors use for verbalizing their opinions about the papers they cite. From the Table we can see that there are only 56 distinct adjectives with positive polarity from all opinion terms contained in the citations examined and 14 distinct adjectives with negative polarity. These figures suggest that: (i) either there is great author agreement in the adjectives they select for expressing their opinions about the papers they cite; agreement that reaches up to 95.6% for positive and up to 98.9% for negative polarity adjectives, (ii) or that the opinions of some authors about a cited paper have a significant influence on the way other researchers perceive the value of that paper and this influence is attested in the overlapping adjectives used for verbalizing author opinions about a cited work. Whatever the reasons, unraveling the specific criteria under which authors make their terminological selections when citing a paper goes beyond the scope of our study, which focuses on deriving the papers' impact as this is attested in the contextual elements of their received citations. A close look at the adjectives used for verbalizing author opinions shows that the most commonly used adjectives for praising a cited work are: *efficient*, *improved* and *promising*, whereas the mostly used adjectives for criticizing some aspects of a cited work are: *expensive*, *complex* and *limited*.

Besides quantifying the degree of overlapping opinion terms in the citations' contextual elements, we also estimated the average strength values of the citations with negative and positive orientations respectively. To that end, we applied our opinion strength formula (cf. Section 3.1) to estimate the strength of every opinion term appearing in a citation sentence and then based on the average strength values of all opinion terms with the same polarity label within a citation we derived the overall strength of that citation's sentiment. Table 2 reports the average strength values of all the citations that have a positive or negative orientation respectively. That is, the figures below are computed for the 1,266 citations with some polarity orientation in our dataset.

Table 2. Opinion agreement in citations' sentiment

Avg.Strength CPos	Avg.Strength CNeg
0.89	0.91

As Table 2 shows there is a remarkable level of opinion agreement among the authors who cite common papers. A comparative evaluation of the results obtained so far reveals that there is not only considerable agreement in author opinions about the sentiment of a paper they cite in their own publications, but also that there is great agreement in the terminology they select for casting their opinions. Having quantified the sentiment orientation of the citations that our experimental papers have received, we performed a user study in which we assessed the accuracy of the orientation labels that have been assigned to the opinion terms contained in the examined paper citations.

4.3 Accuracy of Opinion Terms Polarity

In this section, we present the results of a human study we carried out in order to evaluate the accuracy of the orientation labels that have been assigned to the citation sentences of our experimental papers. The motivation for this study is the dependence of our paper impact measure $I(p)$ on the correct polarity labeling of the citations of our experimental papers. To carry out our study, we recruited 10 volunteers from our school to whom we showed a set of citation sentences and asked them to annotate every sentence with a suitable polarity label. Our volunteers were selected on the ground that they have published at least 10 research papers each pertaining to the discipline of computer science⁷. This way we ensure that our participants are familiar with verbalizing and interpreting references to relevant works.

Having formulated the group of our study participants, we showed them the 1,266 citation sentences of our dataset that contain opinion terms and asked them to read the sentences and indicate for every sentence whether in their opinion it expressed a positive, a negative or a neutral author stance towards the cited work. Sentences were displayed in a random order and they were not revealing any details (e.g. author names, title, etc.) about the source or the referred papers. Before the experiment, we instructed our participants to annotate a citation sentence only if they felt confident about its orientation and leave it unlabelled otherwise. Participants were given ample of time to complete their task, they were allowed to go back in a previous sentence and alter a label they had already assigned, but they were not given any Internet access, they were not allowed to communicate with each other during the experiment and they were not given any further information about the nature or the scope of our study.

Out of the 1,266 citation sentences; 1,179 (93.1%) were annotated with polarity labels by at least six different participants, while for 19 (1.5%) of them none of our volunteers indicated a polarity label. Therefore, we relied on these 1,179 citation sentences that have been examined by the majority of our users in order to assess the validity of the annotations that we have previously assigned to those sentences. For our assessment, we essentially compared the labels that our volunteers assigned to each of these 1,179 sentences to the labels that had been assigned to them based on the process described in Section 3.1. To eliminate errors or idiosyncratic annotations, we considered an annotation to be valid if at least 4 of the study participants used the same polarity label to indicate the citation's orientation.

Obtained results, listed in Table 3, indicate that in 1,175 of the 1,179 (i.e. 99.6%) of the examined sentences at least 4 individuals indicated the same polarity label. Moreover, the comparison of those commonly agreed labels to the ones that had been considered for the same sentences in our experimental evaluation, shows 98.8% agreement between the citations' orientation that we had derived in our experiment and that our participants had indicated. Therefore, based on the results of our human study we may safely conclude that the process we have adopted towards deriving the polarity of the citation sentences is both valid and accurate.

Table 3. Accuracy of citation orientations

# of examined citation sentences	1,179
Inter-annotators' agreement	99.6%
Accuracy of experimental annotations	98.8%

Having assessed the accuracy of the examined citations' orientation, we can exploit them in measuring the impact of their referred papers as we describe next.

4.4 Publications' Quality Impact

As a final evaluation step, we estimated and compared the relative impact of our 40 experimental papers with and without considering their citations' orientation. This is in order to assess how much the citations' contextual elements influence the referred papers' perceived importance. To begin our evaluation, we firstly relied on the process followed by existing impact metrics in order to derive the importance of our publications irrespectively of their citations' polarity. Recall that existing impact measures perceive the publications' importance to be analogous to the number of citations they have received. Obtained results deliver what we call **baseline impact** and are computed simply by counting the number of citations each of the considered papers has received, so that the mostly cited paper is deemed the most influential of all.

Then, we applied our $I(p)$ measure to the same set of 40 publications in order to estimate their importance when accounting for both their citations' polarity and orientation. In this respect, we relied on the number of citations considered for each of the pa-

⁷ This is because our experimental papers relate also to the discipline of computer science.

pers, the citations' orientation and the average strength values of every orientation, in order to derive what we call the **qualitative impact** of each paper considered. Based on our measure, the most influential paper of those considered is the one that has many citations, the majority of which have a positive orientation.

Having computed the relative importance of our 40 experimental papers under the baseline and the qualitative impact measures, we organized the papers in two ranked lists. The first list maintained the papers ordered according to their baseline impact (i.e. the paper with the most citations is ranked first) and the second list maintained the papers ordered according to their qualitative impact (i.e. the paper with many positive citations is ranked first). Finally, we compared the ordering of the papers across the ranked lists in order to assess how much the citations' orientation influences the perceived value of the referred papers. For our comparison, we examined the difference in rankings between the baseline and the qualitative impact measures for each of the papers considered. Figures 1 and 2 illustrate the average difference in the paper rankings between the baseline and our qualitative impact measures; with each figure emphasizing on how positive and negative citations respectively contribute to the ranking differences.

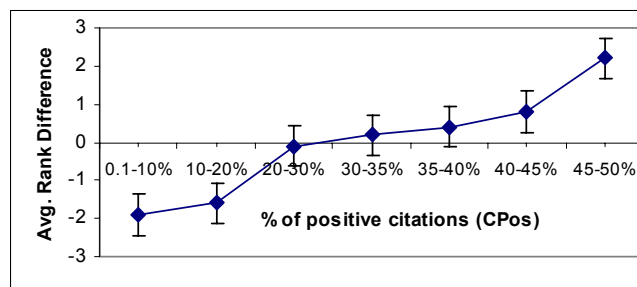


Figure 1. Comparison of differences in paper rankings with respect to the fraction of positive citations.

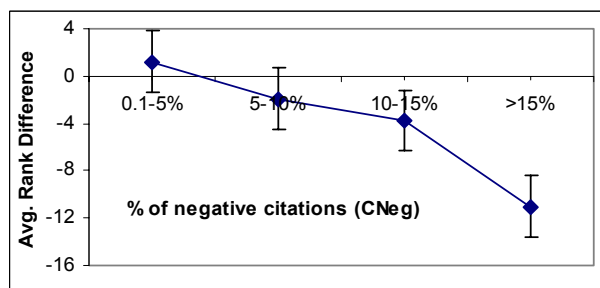


Figure 2. Comparison of differences in paper rankings with respect to the fraction of negative citations.

In both figures the baseline ranking of the examined papers corresponds to the rank difference zero on the y-axis; suggesting that values above zero correspond to higher ranking positions computed by our measure for the papers with a given fraction of positive citations compared to the rankings estimated for the same papers by the baseline method. Conversely, values below zero correspond to lower ranking positions computed by our measure for the papers with a given fraction of negative citations compared to the baseline rankings.

As the Figures show, our method values the impact of research papers differently compared to existing techniques that rely exclusively on the paper citations' distribution. In particular, we observe that the difference in the considered papers' impact is analogous to their citations' orientation, i.e. the increased the fraction of positive citations a paper has, the higher its perceived impact and vice versa.

To elucidate the correlation between the examined papers qualitative impact and their citations' orientation, we report on Tables 4 and 5 respectively the distribution of our experimental papers to the considered fractions of positive and negative citations.

Table 4. Distribution of papers according to the fraction of their positive citations

% of positive citations	# of papers
>0 -10%	13
10-20%	13
20-30%	8
30-35%	2
35-40%	1
40-45%	1
>45%	2

Table 5. Distribution of papers according to the fraction of their negative citations

% of negative citations	# of papers
>0 -5%	8
5-10%	11
10-15%	4
>15%	1

An interesting observation from the figures reported in Tables 4 is that all of the papers considered have received at least one positive citation. Conversely, as Table 5 indicates 60% of the examined papers (i.e. 16 out of 40) have not received any negative citation. The above findings further support our argument that existing impact metrics cannot accurately capture the true value of the papers they examine, since they do not account for the supportive opinions that other researchers express about the cited works.

A comparative evaluation of the obtained results indicates that the more positive citations a paper gets the more influential it is considered (i.e. the higher it is ranked) compared to that paper's influence when its citations' orientation is not accounted. Likewise, as the number of negative citations to a paper increases, the paper's perceived value generally decreases in contrast to the paper's baseline impact that remains unchanged at specific levels of incoming references. Results obtained so far demonstrate not only that our publications' impact metric differs from existing techniques but also that it appreciates the value of a paper depending on how other researchers refer to it in their own publications. Therefore, under our approach a publication with many positive citations is valued higher than a publication with a comparable number of neutral citations.

Based on our findings, we claim that existing impact measures manage to capture the papers' *popularity* pretty well, but they are less effective in capturing the papers' impact. Considering that in scientific publications popularity does not always entail quality, it is imperative that citations' orientation is also accounted when measuring the contribution of research papers. Although, further experimentation is needed before we can obtain a clear picture about the merits of our impact measure, nevertheless we believe that it can serve as a good complement to existing techniques that assess the importance of scholar papers by relying exclusively on statistical evidence.

5. CONCLUDING REMARKS

We have presented a method for mining author opinions towards the papers they cite in order to objectively assess the impact of scientific publications. The major novelty of our study is that we suggest a qualitative measure for evaluating the impact of scientific publications based on both their citation distribution and orientation. Our preliminary experiment reveals some interesting trends in the authors' citing behavior and demonstrates how our method captures the true value of research papers.

In the future, we plan to extend our model in order to account for other lexical elements and combination of elements in the citation sentences, such as conjunctions, verbs, verb-adverb combinations, etc. in order to identify the author opinions for the cited works. Another area for future study is to detect the particular aspects of a paper that the authors who cite it evaluate. This would help us deduce not only how authors refer to related works but also to which specific aspects of the prior works they focus their references. Moreover, we intend to apply our impact measure to scientific publications pertaining to research fields other than computer science, in order to comparatively analyze the author citing trends across disciplines. Last but not least, we consider the examination of alternative semantic tools for deciphering the citations' sentiment. Considering the lack of any other method for mining author opinions about their cited works, it is difficult to compare the findings of our study with relevant methods. Nevertheless, we hope that our method will inspire researchers towards coming up with qualitative publication impact metrics.

6. REFERENCES

- [1] Amin, M. and Mabe, M. 2000. Impact Factors: Use and Abuse. Perspectives in Publishing.

- [2] Archak, N., Ghose, A. and Ipeirotis, P. 2007. Show me the money! Deriving the pricing power of product features by mining consumer reviews. In Proceedings of the 13th ACM SIGKDD Intl. Conference on Knowledge Discovery and Data Mining, San Jose, CA, USA.
- [3] BiHui, J., LiMing, L., Rousseau, R. and Egghe, L. 2007. The R- and AR-indices: Complementing the h-index. In Chinese Science Bulletin, 52 (6): 855-863.
- [4] Bornmann, L. and Daniel, H.D. 2007. What do we know about the h-index? Journal of the American Society for Information Science and Technology, 58(9): 1381-1385.
- [5] Donohue, J.C. 1974. Understanding scientific literatures; a bibliometrics approach. Cambridge: MIT Press, xiii, 101.
- [6] Egghe, L. 2006. Theory and practice of the g-index. In Scientometrics, 69 (1): 131-152.
- [7] FrameNet: <http://framenet.icsi.berkeley.edu>
- [8] Hatzivassiloglou, V. and McKeown, K.R. 1997. Predicting the semantic orientation of adjectives. In Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and the 8th Conference of the European Chapter of the ACL. pp. 174-181.
- [9] Hirsch, J.E. 2005. An index to quantify an individual's scientific research output. In Proceedings of the National Academy of Science, 102 (46): 16569-72.
- [10] Hyland, K. 2003. Self-citation and self-reference: credibility and promotion in academic publications. Journal of the American Society for Information Science Technology, 54 (3): 251-259.
- [11] Gruhl, D., Guha, R., Kumar, R., Novak, J. and Tomkins, A. 2005. The predictive power of online chatter. In Proceedings of the 11th ACM SIGKDD Intl. Conference on Knowledge Discovery and Data Mining, pp. 78-87.
- [12] Kim, S.M. and Hovy, E. 2006. Extracting opinions, opinion holders and topics expressed in online news media text, In Proceedings of the ACL/COLING Workshop on Sentiment and Subjectivity in Text. Sydney, Australia.
- [13] Meho, L.I. 2007. The rise and rise of citation analysis. In Physics World. CoRR abs/physics/0701012.
- [14] Moed, H.F. 2005. Citation analysis in research evaluation. Dordrecht, the Netherlands, Springer.
- [15] Piao, S., Ananiadou, S., Tsuruoka, Y., Sasaki, Y. and McNaught, J. 2007. Mining opinion polarity relations in citations. In Proceedings of the 7th Intl. Workshop on Computational Semantics, Tilburg, the Netherlands, pp. 366-371.
- [16] Popescu, A.M. and Etzioni, O. 2005. Extracting product features and opinions from reviews. In the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, pp. 339-346.
- [17] Rahm, E. and Thor, A. 2005. Citation analysis of database publications. In SIGMOD Record, 34(4): 48-53.
- [18] Turney, P.D. and Littman, M.L. 2003. Measuring praise and criticism: inference of semantic orientation from association. ACM Transactions on Information Systems 21 (4): 315-346.
- [19] WordNet <http://wordnet.princeton.edu/>
- [20] De La Rosa, J.L. and Szymanski, B. 2008. Citation auctions as a method to improve selection of scientific papers. In The Journal of Digital Information Management, vol. 6, no. 5, pp. 414-420.
- [21] Sriphaew, K. and Theeramunkong, T. 2007. Quality evaluation for document relation discovery using citation information. In IEICE Transactions on Information Systems. Vol. E-90-D, no. 8, pp. 1225-1234.