Scientometrics (2019) 119:987–1008
https://doi.org/10.1007/s11192-019-03058-3

In quest of new document relations: evaluating co-opinion relations between co-citations and its impact on Information retrieval effectiveness

Maryam Yaghtin[1] · Hajar Sotudeh[1] · Mahdieh Mirzabeigi[1] · Seyed Mostafa Fakhrahmad[2] · Mehdi Mohammadi[3]

Received: 10 September 2018 / Published online: 2 March 2019

Abstract
Document relational network has been effective in retrieving and evaluating papers. Despite their effectiveness, relational measures, including co-citation, are far from ideal and need improvements. The assumption underlying the co-citation relation is the content relevance and opinion relatedness of cited and citing papers. This may imply existence of some kind of co-opinionatedness between co-cited papers which may be effective in improving the measure. Therefore, the present study tries to test the existence of this phenomenon and its role in improving information retrieval. To do so, based on CITREC, a medical test collection was developed consisting of 30 queries (seed documents) and 4823 of their co-cited papers. Using NLP techniques, the co-citances of the queries and their co-cited papers were analyzed and their similarities were computed by 4 g similarity measure. Opinion scores were extracted from co-citances using SentiWordnet. Also, nDCG values were calculated and then compared in terms of the citation proximity index (CPI) and co-citedness measures before and after being normalized by the co-opinionatedness measure. The reliability of the test collection was measured by generalizability theory. The findings suggested that a majority of the co-citations exhibited a high level of co-opinionatedness in that they were mostly similar either in their opinion strengths or in their polarities. Although anti-polar co-citations were not trivial in their number, a significantly higher number of the co-citations were co-polar, with a majority being positive. The evaluation of the normalization of the CPI and co-citedness by the co-opinionatedness indicated a generally significant improvement in retrieval effectiveness. While anti-polar similarity reduced the effectiveness of the measure, the co-polar similarity proved to be effective in improving the co-citedness. Consequently, the co-opinionatedness can be presented as a new document relation and used as a normalization factor to improve retrieval performance and research evaluation.

Keywords Opinion mining · Co-citations · Co-opinion · Information retrieval

✉ Hajar Sotudeh
sotudeh@shirazu.ac.ir

Extended author information available on the last page of the article

Introduction

Relational networks of documents have been effectively applied in their evaluation and retrieval. Co-citations, co-words, bibliographical couples and co-classes are among the well-known content and citation relations. The assumption underlying the co-citation-based information retrieval (IR) is that co-cited papers are similar and related in their subjects (Eto 2015; White 2016). The assumption has its roots in the selectivity of authors when citing articles (Hamedani et al. 2016) and reflecting their viewpoints about the relevance of the cited contents (Yoon et al. 2016). The co-citation has been shown effective in retrieving relevant documents missed by conventional indexing methods (Eto 2013). Co-citedness, i.e. the frequency of being co-cited, is the first and simplest measure related to the co-citation concept (Eto 2014). However, it has been criticized for ignoring co-cited contents (Martyn 1964) and equating all citations (Dabrowska and Larsen 2015; Boyack et al. 2013).

New horizons have been opened to understand the nature of citation networks, thanks to fulltext availability along with advancement in content and linguistic analysis techniques. Two approaches are dominant in content-based citation analyses. "Syntactic analysis" or "citation proximity analysis (CPA)" (Gipp and Beel 2009), which takes into consideration the proximity of co-cited papers within the citing articles and "semantic analysis" of co-citances, i.e., the co-citation contexts (see Fig. 1). This itself can be divided into two groups, based on the type of information the citances carry: the citing authors' objective narration or viewpoints about the cited article and its features (Sendhilkumar et al. 2013). While the objective narration is studied using citance semantic similarity at conceptual as well as textual, i.e. distributional levels (Jeong et al. 2014; Ding et al. 2014), the authors' viewpoints are investigated using opinion mining techniques. In this way, authors' opinions, which used to be limitedly studied through interviews or surveys on citation motivations (Lipetz 1965; Moravcsik and Murugesan 1975), gained new breakthrough (Ding et al. 2014). This has given rise to the new research area called "citation sentiment analysis" or "citation opinion mining" devoted to discovering how citing authors feel or think about their cited papers based on the

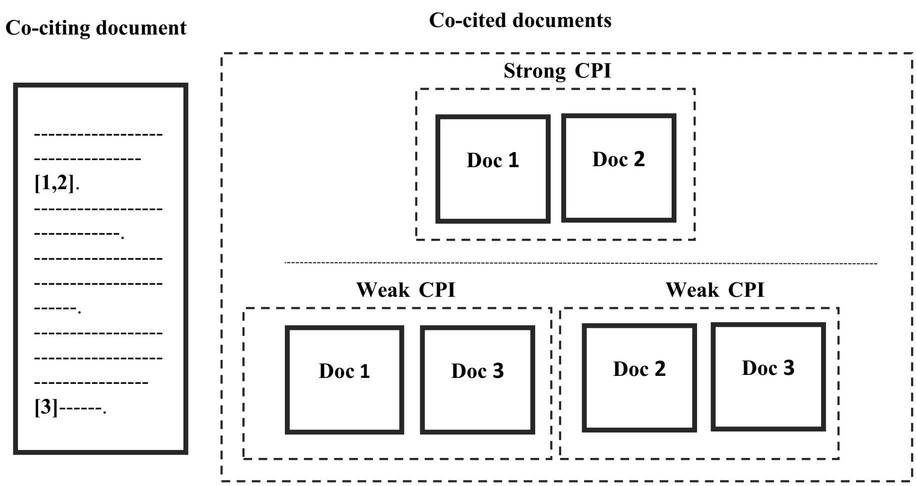

Fig. 1 Co-citations and citation proximity index (CPI)

polarities and strengths of their opinions extracted from citances (Teufel et al. 2006; Schafer and Spurk 2010; Small 2011; Athar and Teufel 2012).

According to citation motivation studies, citing authors usually exhibit a collective dominant trend in their citation behavior, e.g. they use confirmative or persuasive citations that reinforce their findings. They usually avoid criticizing previous papers, or mention negative credits alongside positive ones when they have to do so. Mostly have also been shown to have a perfunctory citation behavior which is reporting the features of the cited papers with no commenting or digging deep into their results (Bornmann and Daniel 2008; MacRoberts and MacRoberts 1989; Brooks 1985; Hanney et al. 2005).

On the other hand, co-cited articles are largely similar in their contents and subjects (Eto 2015; White 2016; Small 1973; Smith 1981). On this basis, co-citing authors are expected to be, to a large extent, similar in their opinions about the same co-cited pair. Consequently, "co-opinionatedness" could be presented as a new indicator in semantic analysis of co-citances to measure the opinion similarity between co-citing papers about their co-cited articles.

The newly-proposed concept of co-opinionatedness of the co-cited papers can be defined by their opinion similarity measured by two dimensions including the strengths and polarities of the opinions they receive. Therefore, the opinion similarity scores range on a continuum from slightly to absolutely co-opinionated. Obviously, at the two extremes of the continuum that represent complete accordance and discordance, there exist co-citing papers that have respectively the minimum and maximum distances in their opinion strengths towards their co-cited pairs.

Based on the aforementioned points regarding the relatively conservative citation behavior of scholars, we expect to see the co-opinionated co-citances to be dominated by co-polarity with a majority being positive. They are also expected to be more similar in their opinion scores compared to the anti-polar ones. Furthermore, based on the co-citation assumptions regarding the content similarity of co-citations, we expect that the co-opinionatedness is effective in retrieving textually similar citances. Therefore, by concentrating on a sample of co-cited medical papers, the present communication, first, endeavors to test the extent of co-opinionatedness between co-citing papers about their co-cited articles, and thereby to introduce a new relation in the scholarly document network. Eventually, it tries to analyze the new relation's role in improving co-citation-based retrieval of relevant documents.

Literature review

Co-citation

Since its invention, co-citation has increasingly been the focus of many studies due to its theoretical strength and operational success. The effectiveness of the co-citation relation in improving information retrieval has been verified frequently (Egghe and Rousseau 1990; Bichteler and Eaton 1980; Zhao 2014; Badran 1984). Co-citation-based IR commonly use co-citedness as a criterion to judge the relatedness of co-cited documents. However, the quantitative measure did not prove to return remarkable results, since it fails to take into consideration the content of the co-cited articles. This may be reflected in the insignificant association observed between the measure and content similarity of co-cited papers (Elkiss et al. 2008).

Content-based citation analysis tries to bridge the gap by analyzing the similarities of co-cited papers in terms of their contents or their citances within the co-citing papers. This is based on the assumption that co-cited papers are similar in their subject coverage (Small 1973; Bichteler and Eaton 1980; Hamedani et al. 2016; Yoon et al. 2016), which has also been experimentally confirmed by the similarity between citances and abstracts of co-cited papers (Elkiss et al. 2008).

Citation proximity analysis (CPA), or citation syntactical similarity, is one of the approaches in content-based citation analysis, which differentiates between citations based on their distance within the text, e.g. within one sentence, within one paragraph but in different sentences, between different paragraphs or sections, etc.)Gipp and Beel 2009). Distinguishing the co-citations based on their proximity within the citing document has been found to be effective in improving IR performance and decreasing the number of irrelevant results (Gipp and Beel 2009; Callahan et al. 2010; Eto 2012, 2013, 2014, 2015; Boyack et al. 2013; Liu et al. 2014). The closer the co-citations within the citances, the more similar are the co-cited papers in their contents (Elkiss et al. 2008).

The syntactical similarity of co-citances has proved to be capable of successfully normalizing the co-citedness (Eto 2013). Further, the semantic similarity of co-citances has proved to be more effective in comparison with co-citedness in reflecting intellectual structure of a discipline and its subdomains (Jeong et al. 2014).

Citance opinion mining

Citance opinion mining has been attracting a great deal of attention in Natural Processing Language (NLP) as a new emerging research area. Using machine learning and dictionary-based approaches, opinion mining studies have pursued three essential targets including developing or testing tools and techniques, discovering essential features and patterns, and finally evaluating documents.

For instance, support vector machine classifiers proved to be more effective in classifying opinions extracted from citances, compared to naïve Bayesian classifiers (Athar 2014). There are ongoing discussions on the reliability of opinion mining of content-based citation analysis of scientific papers, because the previous results have not been comparable due to employing different corpuses (Hernández-Alvarez and Gomez 2016). However, several opinion analyses have been conducted to discover patterns and trends governing scientific articles, whose results have been promising so far. For example, they have uncovered the effective role of certain text features in determining the opinion polarity of citations. These include certain linguistic features (e.g. modal auxiliaries, main verbs, verb tenses and voices, first- and third-person pronouns, comparative and superlative conjunctions), the overall position of a sentence, and parts-of-speech tagging (Teufel et al. 2006; Athar 2011, 2014; Jochim and Schütze 2012; Abu-Jbara et al. 2013; Athar and Teufel 2012).

In addition, previous studies have confirmed the correlation between citation functions and opinion polarity of citances at sentence, paragraph, and section levels (Teufel et al. 2006, 2009; Dong and Schäfer 2011; Jochim and Schütze 2012). Opinion citation analyses were found to differently rank cited papers compared to quantitative citation analyses (Cavalcanti et al. 2011; Sendhilkumar et al. 2013; Amadi 2014), where highly cited papers descend to lower ranks (Amadi 2014) since they do not necessarily gain positive citations (Cavalcanti et al. 2011). Furthermore, polarity of opinions has been found to be associated with citation quantity and locations. The locations, where the citations occur, are found to have a considerable impact on the citation polarities, such that introductory sections reflect more positive, while discussion section reveals more negative opinions (Amadi 2014). In addition, the total number of citations has proven to be more strongly correlated to positive citations than negative ones (Abu-Jbara et al. 2013).

Consequently, it is believed that by analyzing citation opinions, we can devise new bibliometric measures to evaluate cited papers by distinguishing between and differently weighing negative and positive citations (Abu-Jbara et al. 2013; Hernández-Alvarez and Gomez 2016), thereby helping scientometricians and research managers to gain insight into the contribution and actual impact of cited works (Yu 2013). Meanwhile, the credibility of cited papers can also be judged by classifying the citing authors who have commented on their cited papers (Parthasarathy and Tomar 2014). Citation opinion analysis is also believed to have many applications in automatically evaluating individuals and journals' impacts and in interpreting knowledge maps (Athar and Teufel 2012; Small 2011). In this way, qualitative features can be added to citation analysis which may identify citation polarities and purposes (Abu-Jbara et al. 2013). Although citances often consist of positive (Mahalakshmi et al. 2015), rather than negative citation sentences (Athar 2014), ignoring the latter may result in losing sentiments and criticisms about the cited papers (Athar and Teufel 2012).

As can be inferred from the literature, citances have been found to be influential in improving co-citation-based retrieval and reducing the number of irrelevant results returned. They are also believed to be useful in research evaluation. Citance proximity, opinions, and textual similarities are among the measures tested and approved. However, as far as our literature review goes, there exist no studies dealing with opinion analysis of co-citation contexts or their relatedness in terms of opinions they express about their co-cited papers. It is, therefore, necessary to test whether the co-opinion concept exists in co-cited papers and if so, whether it can improve the effectiveness of successful co-citation measures i.e. co-citedness and CPI in retrieving textual similar co-cited papers.

Research questions

1 To what extent are the co-citances similar in terms of the strengths of opinions they receive from their co-citing papers?
2 To what extent are the co-citances similar in terms of the polarities of opinions they receive from their co-citing papers?
3 Are the co-citances with different opinion polarities also different in their opinion strengths?
4 Is there any significant difference between rankings of cocited documents retrieved by co-citedness before and after normalization by co-opinionatedness?
5 Is there any significant difference between rankings of cocited documents retrieved by CPI before and after normalization by co-opinionatedness?

Methodology

Using a content analysis method with citation analysis approach, the present study analyzes co-cited works and the opinions expressed by their co-citing papers about them. To do so, it analyzes co-citances as the representations of the contents of co-cited papers and carriers of co-citing authors' opinions (Ritchie et al. 2008; Agarwal et al. 2010; Doslu and Bingol 2016; Sendhilkumar et al. 2013).

To calculate co-citance textual similarity, all words, except for stop words, were analyzed. However, opinion analyses were performed on nouns, adjectives, adverbs, and verbs that are believed to carry opinion loads in a sentence (Esuli and Sebastiani 2007; Mahalakshmi et al. 2015).

Experimental framework

Test collection building

CITREC[1] covers 255,339 papers. A test collection was built up consisting of 30 seed documents and their co-cited papers. The 30 seed documents were selected randomly from those CITREC documents that had at least 100 co-citations. The reason for applying the criterion was to ensure the inclusion of a high number of co-cited papers in the test collection. The 30 seed documents served as queries. Their co-cited papers were identified to be 5394, such that each of the seed documents had at least 109 and at most 284 co-cited documents.

Variables definition and measurement

Citedness

According to Small (1973), co-citation patterns represent the linkage between the cocited documents in terms of key concepts, methods or experiments. He defines the strength of co-citation as the frequency with which two co-cited papers are cited together. In the present study, cocitedness values were extracted from Citrec's Sim_cocit measure.

Co-citances

A citation context, or citance as coined by Nakov et al. (2004), is a set of sentences surrounding a reference marker within a citing document stating the important facts, features or contribution of the cited paper. In the present paper, citances are derived from CoLil (Comments on Literature in Literature)[2] database. In the database, a citance is defined as the sentence containing the reference marker, or the sentence containing the reference marker plus its succeeding one if the marker is located at the end of the former. Each citance is consisted of 240 characters before and up to 240 characters after the reference marker (Fujiwara and Yamamoto 2015). Co-citances are, thus, defined as the citances about papers cocited within a cociting article (see Fig. 2). The verification of the test collection showed that within the co-citing papers, there were 38,178 citances about the 5394 co-cited papers.

Using KNIME (the Konstanz Information Miner) platform, the contents of the co-citances were processed via a workflow consisting of appropriate modules of linguistic

[1] https://www.isg.uni-konstanz.de/projects/citrec/.

[2] http://colil.dbcls.jp/browse/papers/.

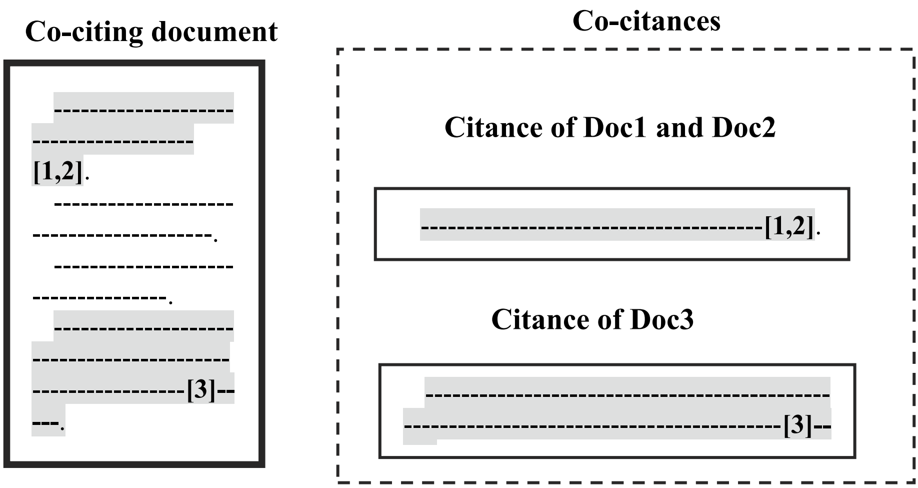

Fig. 2 Co-citances about cocited documents within a cociting article

preprocessing; erasing trivial words and characters (i.e. stop words, numbers, punctuations), PoS tagging, and lemmatization. Eventually, 4 g similarity between the co-citances of seed documents and their co-cited papers was calculated using KNIME's string similarity function. To do so, 4 g to 9 g similarity measures as well as TF-IDF was calculated and their correlation were tested using Pearson correlation analysis. The 4 g index was selected and tested, given its colinearity with other similarity measures i.e. 5 g ($r=0.99$, Sig. = 0.000), 6 g ($r=0.98$, Sig. = 0.000), 7 g ($r=0.97$, Sig. = 0.000), 8 g ($r=0.96$, Sig. = 0.000), 9 g ($r=0.97$, Sig. = 0.000) and TF-IDF ($r=0.86$, Sig. = 0.000).

Citation proximity index (CPI)

The index is proposed by Gipp and Beel (2009) to measure the relatedness between two cited documents based on their closeness within a citing document. To measure the closeness, their situation within the citing document is considered depending on whether they appear as enumeration, within a sentence, within a paragraph, or across two paragraphs (Eto 2013). In the present study, the CPI was extracted from CITREC using Sim_cpa_simple index that is calculated based on Gipp and Beel (2009). They considered the probability of similarity between two citations to be 100% if they are located in the same sentence (CPI = 1), 50% if they are situated in two sentences within the same paragraph (CPI = 1/2), 25% if within the same chapter and so on (see Fig. 1).

Opinion

Opinion is defined as views, judgments, attitudes, emotions or appraisals formed in one's mind about a particular matter (i.e. entities, individuals, issues, events, topics and their attributes). In the field of opinion mining, people's opinions are computationally extracted from their written texts, especially from the subjective sentences expressing subjective views and opinions (Shuy 2003; Liu and Zhang 2012; Liu 2012). In the present study, opinion is measured using the scores given to subjective words in SentiWordNet.[3] Since every word may have different meanings, in SentiWordNet there are Synsets collecting different senses of a given word. Every Synset may have positive, negative and neutral scores which totally amount to one. Given the existence of several Synsets for a given word, it was necessary to assign a single score to each word. To do so, using the methods applied by Cavalcanti et al. (2011), for any given word within a specific PoS, all opinion scores of its Synsets were averaged after subtracting negative scores from the positive ones.

The opinion score of each citance was, then, calculated based on the summation of opinion scores of its words. Of the 38,178 citances extracted, a scarce number (accounting for 1.3%) did not gain opinion scores.

The list of negation words including "no, not, n't, never, neither, nor, none, nobody, nowhere, nothing, cannot, cannot, without, no one, no way" were identified from Wilson et al. (2005) and completed by their contractions as well as other negation words including "aren't, can't, can't didn't, couldn't daren't, doesn't, don't, hadn't, hasn't, haven't, isn't, mayn't, mightn't, mustn't, needn't, oughtn't, wouldn't shan't, shouldn't wasn't, weren't, won't". The examination of the negation words in the citances revealed that they rarely emerged (less than 1% of the total number of the total words of citances analyzed). Consequently, we did not deal with the negation words, because of their scarce number in our sample and their insignificant improvement in identifying citance opinions as confirmed by Athar (2014).

After processing the citances and assigning their opinion scores, the number of opinionated co-cited papers reached 4823 pairs. Also, the number of co-citing papers reduced to 8296, due to unavailability of some of the co-citances.

Co-opinionatedness

Co-opinionatedness is a new indicator proposed and tested for the first time in the present research. It is defined as the similarity between two opinion scores (see Fig. 3). Similarity between two numbers is expressed according to their distance (Andrejko and Bieliková 2012; Segaran 2007). Minkowski metric is among the popular ones (Saraçoglu et al. 2007):

$$\text{Minkowski}_p\left(x_i, x_j\right) = \left(\sum_{k=1}^{m} \left|x_{ik} - x_{jk}\right|^p\right)^{1/p}$$

where $p \geq 1$. A special case of Minkowski metric, is Euclidean distance, with $p = 2$, that is one of the most common ones (Su and Chou 2001). In the present study, the inverse of the Euclidean distance is used to measure "opinion similarity" between two co-cited papers. As the distance is calculated between two cocited papers within each pair, then $k = m = 1$. As a result, the similarity of two opinion scores equals:

$$\text{Sim}\left(\text{Opinion}_1, \text{Opinion}_2\right) = \frac{1}{1 + \left|\text{Opinion}_1 - \text{Opinion}_2\right|}$$

[3] http://sentiwordnet.isti.cnr.it/.

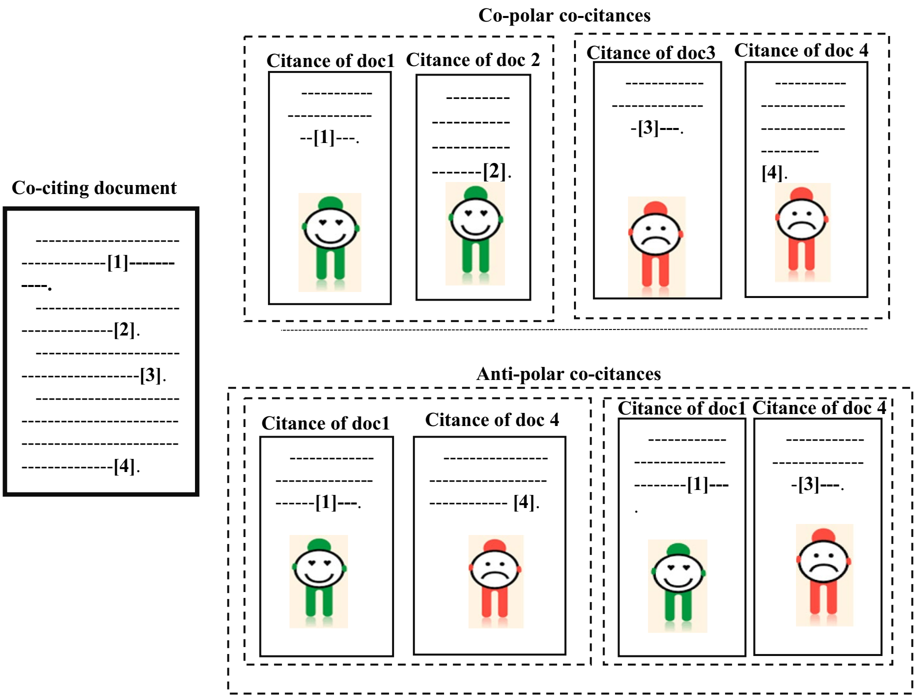

Fig. 3 Co-polar and anti-polar co-citations

To avoid zero in the denominator, 1 is added to the formula. In this way, each co-cited pair may have opinion similarity based on their absolute opinion degrees. As an instance, a co-citation pair, one with +1 and the other with +2 opinion scores, has 0.5 degree of opinion similarity. The more the value tends to 1, the more dissimilar are the two opinion scores.

As mentioned previously, the co-opinionatedness concept can be measured by two dimensions including the papers' opinion strengths and polarities. As a result, the concept can be measured at two levels. At a general level, the "overall opinion similarity" of two co-cited papers can be measured by calculating opinion similarity of all their co-citing papers. At a specific level, for each co-cited pair, its co-citances can be divided into two categories based on their polarity (dis)similarity. Co-polar co-citances share the same polarity either positive, negative or neutral, while anti-polar co-citances are different in their polarities. Inside the categories, the co-opinionatedness can be graded by the opinion similarity. Obviously, anti-polar co-citances have no co-opinionatedness regarding their polarities. However, they, too, may expand on a continuum from slightly to completely accordant depending on their opinion score similarities. Accordingly, anti-polar co-citances with shorter distances in their strengths may show some extent of opinion similarity, compared to those with farther opinion strengths. The logic underlying the calculation is that in a purposeful IR environment, users may be interested in finding co-cited papers in line with their results or similar in the opinions they receive from their co-citing papers.

Baseline and benchmark measures

As previous literature confirmed the effectiveness of co-citedness in the retrieval of relevant documents (Egghe and Rousseau 1990; Bichteler and Eaton 1980; Badran 1984; Zhao 2014) and CPI (Gipp and Beel 2009; Callahan et al. 2010; Eto 2012, 2013, 2014, 2015; Boyack et al. 2013; Liu et al. 2014), in the present study, the measures have been used as baselines. The distributional semantic similarity of co-citances, i.e. their textual similarity, representing the similarity of co-cited papers (Jeong et al. 2014), served as benchmark. The baselines were, then, normalized by opinion similarity scores and evaluated as described in the following section.

Evaluation metric

nDCG was selected as evaluation metric. The rationale for the selection is that the measure is useful for graded relevance (Kekäläinen 2005). Besides, it takes into consideration just top ranked papers that users care much more than others (Wang et al. 2013). It was calculated for the co-cited documents in terms of their textual similarities of co-citances at three precision points including $p@10$, 20, and 50. Applying normalized Discounted Cumulative Gain (nDCG) as an evaluation metric (Järvelin and Kekäläinen 2002), we compared the results of cocitedness and CPI measures before and after being normalized by the opinion similarity scores. Devised as a useful measure for graded relevance (Kekäläinen 2005), nDCG can compare the two methods minutely (Eto 2013). nDCG is the ratio of DCG to the ideal DCG (IDCG) computed based on:

$$\mathrm{DCG}_p = \sum_{i=1}^{p} \frac{2^{\mathrm{rel}_i} - 1}{\log_2(i+1)} \text{ and } \mathrm{IDCG}_p = \sum_{i=1}^{\mathrm{REL}} \frac{2^{\mathrm{rel}_i} - 1}{\log_2(i+1)}$$

where p is the precision point, rel_i is the relevance score of ith document ranked and REL is the relevance of documents ordered by the best possible relevance degrees up to position p. The more nDCG approximates 1, the nearer the relevance ranking yielded by the tested method to the perfect one. In the present study, the IDCG is measured by the ranking yielded by the co-citance similarities used as our benchmark.

Test collection reliability

Based on the generalizability theory (GT) proposed by Bodoff and Li (2007), we tested the reliability of the test collection using gt4ireval package.[4] GT is based on two coefficients yielded in its two phases including generalizability study (G-study) and decision study (D-study) (Urbano et al. 2013). Generalizability coefficient ($E\rho^2$) is dedicated to measuring the stability of differences between systems based on the ratio of system variance to the variance in relative nDCG scores (i.e. in system rankings). The latter is the summation of system variance and relative error variance:

$$E\rho^2 = \frac{\partial_s^2}{\partial_s^2 + \frac{\partial_e^2}{n'_q}}$$

4 https://rdrr.io/cran/gt4ireval/src/R/gstudy.R.

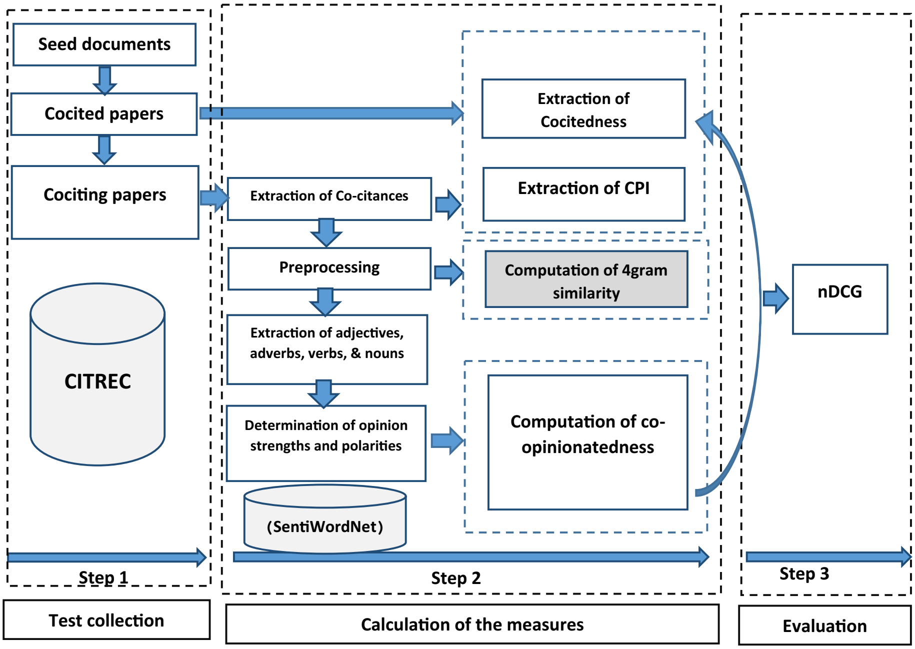

Fig. 4 Research process flow

Dependability coefficient (φ), which is the ratio of system variance to the total variance (i.e. the summation of system variance and absolute error variance), measures the stability of the absolute effectiveness scores:

$$\varphi = \frac{\partial_s^2}{\partial_s^2 + \frac{\partial_q^2 + \partial_e^2}{n'_q}}$$

Both of the measures should tend to one in order to have a reliable test collection, and thus a reliable ranking. The coefficients are believed to be advantageous due to their potentials to estimate the reliability of any query set with an arbitrary size, with no need to follow the traditional methodologies based on random what if scenarios and extrapolation (Urbano et al. 2013). The variance components yielded from the set of queries at hand (30 queries in the present study) are used by the coefficients to project how reliable the scores would be if query sets of any size (Kanoulas and Aslam 2009). The generalizability theory has been used in several studies to test the reliability of the test collections (see e.g. Kanoulas and Aslam 2009; Hasanain et al. 2018).

Data analysis method

The data were then analyzed using SPSS 22. Chi Square test was used to compare the frequency of co-citations. Given the non-normality of the opinion similarity data distributions, Wilcoxon test was used to compare the co-polar and anti-polar opinion similarity.

Table 1 The frequency of co-cited papers in opinion similarity categories

Opinion similarity level	Similarity degree	Frequency	Percent
The least similar	0–0.20	3	0.06
Less similar	0.21–0.40	291	6.03
Moderately similar	0.41–0.60	1224	25.38
Similar	0.61–0.80	1962	40.68
Highly similar	0.81–1	1343	27.85
Total		4823	100.00

Table 2 The frequency of co-polar and anti-polar co-citing papers

Polarity	Frequency	Percent
Co-polar		
Positive	3384	40.79
Negative	1413	17.03
Neutral	3	0.04
Sum	4800	57.86
Anti-polar		
Positive–negative	3431	41.36
Positive–neutral	28	0.34
Negative–neutral	37	0.45
Sum	3496	42.15
Total	8296	100

However, as the distribution of the nDCG values was normal, they were compared using paired samples T test (Fig. 4).

Results

Opinion similarity of co-cited papers in terms of opinion strengths

To gain an insight into the opinion similarity of the co-cited papers, they were categorized into five groups based on their opinion similarity degrees. The results presented in Table 1 suggest that the co-cited documents are mostly "similar" or "highly similar" to their related queries in terms of the opinions they received. The results of Chi^2 reveal that the groups are significantly different in their frequency ($\chi = 2678.513$, $df = 4$, Sig. = 0.000)

Polarity of the co-citances

The verification of the opinions of co-citances indicated that 57.86% of the co-citing papers were co-polar in their opinions about their co-cited pairs, with a majority being positive (40.79%). Anti-polar contexts were relatively lower in number (42.15%). The results of Chi^2 test showed that the two groups were significantly different in their numbers ($\chi = 204.968$, Sig. = 0.000) (Table 2).

Table 3 Comparison of opinion strengths of co-polar and anti-polar co-citing papers

Polarity type	Opinion similarity				Comparison	
	Min	Max	Mean	Mean rank	Chi square	Sig.
Co-polar	0	1	0.47	1.60	181.262	0.000
Anti-polar	0	1	0.32	1.40		

Table 4 nDCG means for baselines before and after being normalized by opinion measures

Baselines	nDCG	
	Mean	SD
Co-citedness		
Before normalization	0.48	0.20
After normalization by		
Co-cited opinion	0.34	0.20
Overall opinion similarity	0.52	0.19
Co-polar opinion similarity	0.51	0.20
Anti-polar opinion similarity	0.44	0.21
CPI		
Before normalization	0.78	0.12
After normalization by		
Co-cited opinion	0.34	0.20
Overall opinion similarity	0.82	0.11
Co-polar opinion similarity	0.77	0.15
Anti-polar opinion similarity	0.54	0.20

Comparison of co-polar and anti-polar co-citations in terms of their opinion strengths

According to the results of the Wilcoxon test, summarized in Table 3, co-polar co-citing papers were more similar in terms of their opinion strengths (Mean rank = 1.60), compared to the anti-polar ones (Mean rank = 1.40).

Effectiveness of co-opinionatedness in retrieving textually similar co-cited papers

Table 4 summarizes nDCG mean values for CPI and co-citedness before and after being normalized by opinion measures. As seen, CPI is more powerful in retrieving textually similar co-citations (mean nDCG = 0.78) compared to the co-citedness which is about average (mean nDCG = 0.48). Both measures were found to be weakened after being normalized by co-cited opinion (mean nDCG = 0.34 for both of them), but boosted after being normalized by overall opinion similarity (mean nDCG = 0.82, 0.52). While co-citedness was observed to be improved by co-polar opinion similarity (mean nDCG = 0.51), CPI slightly worsened (mean nDCG = 0.77).

As mentioned previously, in order to test the effectiveness of the co-opinionatedness in improving the results of information retrieval, we first calculated the power of co-citedness in ranking textually similar co-cited papers by nDCG. We, then, normalized the

Table 5 Comparing the effectiveness of co-citedness in retrieving textual similar co-citances before and after being normalized by opinion-based measures

Normalization factor	P	Mean difference (raw nDCG mean – normalized nDCG mean)	T	Sig.
Co-cited opinion	@10	– 0.162	– 4.552	0.000
	@20	– 0.149	– 4.986	0.000
	@50	– 0.133	– 5.047	0.000
Opinion similarity				
Overall	@10	0.018	2.138	0.041
	@20	0.027	4.083	0.000
	@50	0.037	3.854	0.001
Co-polar	@10	0.013	2.020	0.053
	@20	0.019	3.067	0.005
	@50	0.049	4.426	0.000
Anti-polar	@10	– 0.028	– 1.474	0.151
	@20	– 0.031	– 1.596	0.121
	@50	– 0.037	– 1.658	0.108

co-citedness values based on the opinion similarity scores of co-cited papers and re-calculated the nDCG values and finally compared the raw and normalized nDCG values using paired *T* test.

According to the results reported in Table 5, although the sheer values of opinion scores of co-citations significantly reduce the textual similarity of co-citedness measure, opinion similarity has significantly improved the power of co-citedness in retrieving textually similar co-cited papers. According to the nDCG mean difference values, the "overall opinion similarity" has improved the retrieval of relevant results from 0.012 (for *P*@10) to 0.037 (for *P*@ 50). Co-polar opinion similarity was also found to be effective in improving the results with almost the same power for *p*@10 and *P*@20, but relatively higher for *p*@50. It signifies that the opinion similarity effectiveness has generally its roots in the co-polar opinion similarity.

On the other hand, anti-polar opinion similarity has negatively influenced the co-citedness power, though it is not significant.

We also tried to test co-opinionatedness power in augmenting CPI power in retrieving textually similar citances. The results, provided in Table 6, reveal that the "overall opinion similarity" can increase CPI power for *P*@10 = 0.039, *P*@20 = 0.041, and *P*@50 = 0.035. Although co-polar opinion similarity was not found significantly effective, anti-polar opinion similarity significantly reduced the effectiveness of the CPI in retrieving similar co-citances.

Reliability of the ranking

The results of the generalizability study are illustrated in Table 7. As seen, all of the expected $E\rho^2$ values tend to one and confirm the reliability of the ranking. According to the estimation of the required queries, the test collection is expected to have 30 queries for the co-citedness-based system, 15 queries for the CPI-based system and 12 ones for the

Table 6 Comparing the effectiveness of CPI in retrieving textual similar co-citances before and after being normalized by opinion-based measures

Normalization factor	P	Mean difference (raw nDCG mean – Normalized nDCG Mean)	T	Sig.
Co-cited opinion	@10	−0.474	−9.911	0.000
	@20	−0.475	−12.285	0.000
	@50	−0.428	−12.753	0.000
Opinion similarity				
Overall	@10	0.039	2.332	0.027
	@20	0.041	2.664	0.012
	@50	0.035	3.011	0.005
Co-polar	@10	0.010	0.387	0.701
	@20	0.002	0.077	0.939
	@50	−0.022	−1.039	0.307
Anti-polar	@10	−0.206	−4.853	0.000
	@20	−0.240	−6.202	0.000
	@50	−0.253	−6.853	0.000

combined system. According to the results of the D-study phase, the co-citedness-based system, is the lowest in terms of the stability of the ranking ($\varphi=0.70$). It is expected to have 241 queries in order to reach 0.95 level of stability. However, the CPI-based system ($\varphi=0.96$) and the combined system ($\varphi = 0.96$) show a highly stable ranking.

Discussion and conclusion

To evaluate document relations, various measures have been devised so far, each partially reflecting the relation features depending on its weaknesses and strengths. As one of the document relevance measures, co-citation has been found to be powerful in retrieving relevant documents (Janssens and Gwinn 2015; Yoon et al. 2016; Eto 2013). It is, therefore, necessary to improve the measure in order to calibrate relation networks of scientific papers. In early studies of co-citations, only the number of references at the end of papers mattered and all of them had equal binary weights. Consequently, content-based citation analysis, especially on citation contexts, was proposed (Small 1982) to obtain more desirable results that the quantitative co-citation failed to achieve, due to ignoring the differences of citations in their importance and usefulness for the citing papers. Citance analysis has been found influential in improving IR results compared to the raw citation (Nakov et al. 2004; Eto 2013). As papers do not always receive affirmative citations, one of the solutions is to differentiate between them in terms of their citation functions based on the citing authors' opinions (Hernández-Alvarez and Gomez 2016; Parthasarathy and Tomar 2014). Given the content affinity between co-cited papers, it seems that a new relation can be found between co-cited papers in the co-opinionatedness of their co-citing papers. To test the existence of the relation and then its impact on improving traditional co-citation measures in IR, the present research attempted to mine and explore co-citing authors' opinions about their co-cited papers.

The results of the first step indicated that the studied co-citances were mainly co-opinionated in that they were mostly similar either in their opinion strengths (Table 1) or in

Table 7 The stability ($E\rho^2$) and dependability (φ) coefficients

System	G-study						D-study					
	$E\rho^2$			Required number of queries			φ			Required number of queries		
	Expected	Lower	Upper	Expected	Lower	Upper	Expected	Lower	Upper	Expected	Lower	Upper
Co-citedness	0.95	0.83	0.99	30	3	111	0.70	0.36	0.97	241	17	972
CPI	0.97	0.91	0.99	15	2	53	0.96	0.87	0.99	23	2	80
Both	0.98	0.95	0.99	12	3	28	0.96	0.89	0.99	25	6	66

their opinion polarities (Table 2). These suggest that co-cited papers that have been previously proven to be similar in their subjects (Small 1973; Bichteler and Eaton 1980; Hamedani et al. 2016; Yoon et al. 2016) are largely similar in the opinions they receive from their co-citing papers.

Further, the co-opinionated co-citances are generally positive in their polarity. The positivity results from either expressing positive opinions or reporting positive results, which can be itself considered a kind of confirmation, though not so explicit (Jia 2018). The finding is consistent with previous studies confirming the dominance of positive citations, which is attributed to authors' avoidance of directly criticizing or rejecting previous papers (MacRoberts and MacRoberts 1984; Athar and Teufel 2012; Mahalakshmi et al. 2015; Sendhilkumar et al. 2013). This may also be attributed to confirmation biases pushing authors to selectively cite those papers confirming their ideas, findings or hypotheses (Chubin and Moitra 1975; Leung et al. 2017). In addition, there is a wide-spread tendency among authors and journals to report and accept positive results (MacRoberts and MacRoberts 1984; Athar and Teufel 2012; Matosin et al. 2014; Mahalakshmi et al. 2015).

The dominance of the opinion similarity among the co-citations implies the existence of the co-opinionatedness relation proposed by the present study as a new relation in the scholarly document network. However, the other side of the coin, i.e. the anti-polarity, should not be ignored. As the results showed, a considerable, though relatively lower, number of co-citing papers are anti-polar with farther opinion strengths compared to the co-polar ones (Table 3). In particular, a large number of the anti-polar co-citations are negative–positive in their polarity. This would seem unexpected since negative citations are believed to be mostly avoided. However, as mentioned before, in NLP-based opinion mining, reporting negative results that are objective in nature cannot be distinguished from negative opinions (Jia 2018). As a result, negative values do not necessarily represent negative opinions, but possibly negative results cited. Therefore, it could be argued that a co-citing author may not hesitate to contrast negative and positive results or contrast those papers about which s/he has positive and negative opinions, when necessary.

As the earliest and simplest ready-made, quantitative, and thus efficient measure, co-citedness and CPI are widely used in operational and research milieus. In the present study, the nDCG values obtained from the measures re-confirm their effectiveness in retrieving similar co-citations, with the latter being considerably more powerful in comparison (Table 4). This is in line with previous studies confirming the effectiveness of the co-citedness (Egghe and Rousseau 1990; Small 1973; Bichteler and Eaton 1980; Badran 1984; Zhao 2014) and CPI (Gipp and Beel 2009; Callahan et al. 2010; Eto 2012, 2013, 2014, 2015; Boyack et al. 2013; Liu et al. 2014) in retrieving relevant results. This is due to the fact that co-citances involve representations of the contents of co-cited papers (Doslu and Bingol 2016; Agarwal et al. 2010; Ritchie et al. 2008). However, according to previous studies, co-citedness is not highly powerful in retrieving relevant documents (Eto 2012, 2013, 2014, 2015; Gipp and Beel 2009; Callahan et al. 2010; Boyack et al. 2013; Liu et al. 2014) and requires to be improved by content-based tools.

In our endeavors to improve the measure, we tested the co-opinionatedness of co-citances. The "overall opinion similarity" computed based on the opinion similarity of all co-citing papers was observed to successfully improve the co-citedness power in retrieving textually similar co-citances (Table 5). Furthermore, CPI, which has previously been proven to effectively improve IR results was also found to get further improved after being normalized by the overall opinion similarity (Table 6).

When the opinion similarity score was separately calculated based on the co-polarity or anti-polarity of the opinions of co-citances, it was observed that co-polar opinion similarity significantly improves the effectiveness of co-citedness. However, it was not found effective in boosting CPI's. Instead, anti-polar co-citations were revealed to strongly, though inversely, impact its effectiveness. It means that co-opinionatedness is present and effective in improving retrieval of similar texts. However, it is not consistently scattered all over a citing paper, where the more similar co-citances in closer locations display more divergent opinions. It seems that when co-citing authors decide to directly—i.e. in closer locations-contrast two co-cited papers, they considerably differ in their results or in their opinion strengths. Accordingly, IR systems based on CPI and co-citedness, can boost their effectiveness by applying the overall opinion similarity. However, in co-citedness-based systems, employing co-polar opinion similarity may find both more textually similar and more accordant co-cited papers, thereby supporting a purposeful IR environment, where users need to find co-cited papers that are similar to their findings, or attract the same feelings and attitudes from other researchers. Further, anti-polar opinion similarity may be useful in CPI-based IR systems, when diversifying the retrieval results is desired, so that, users could find similar co-cited papers with divergent opinions or findings on the top of the results.

Previous literature has emphasized the importance of distinguishing between negative and positive opinions since it can help in improving IR measures (Piao et al. 2007), reviewing literature (Yu 2013), summarizing scientific papers and discovering the gaps in current studies (Athar and Teufel 2012), and evaluating scholarly papers (Yu 2013; Abu-Jbara et al. 2013; Hernández-Alvarez and Gomez 2016; Athar and Teufel 2012; Parthasarathy and Tomar 2014, 2015). Although confirmatory citation has been revealed to be more effective in retrieving relevant documents (Cavalcanti et al. 2011), the role of negative credits should also be taken into account, since many users are willing to ensure the validity of a research method, tool, or finding, and to do so they need to compare and judge the pros and cons clustered by the opinion algorithms (Eto 2012).

The findings of the present study suggested that the strength of the opinions of documents has the same function as the anti-polarity, in that it is inversely effective in retrieving similar documents. However, it cannot be helpful in discovering the opinion affinity between the retrieved documents and the seed document. Instead, opinion similarity may be helpful in retrieving opinion relations of co-citations, which may be desired in an interactive IR environment and research evaluation.

Overall, the co-opinionatedness concept proved to be effective in improving IR measures. This may be attributed to the fact that CPI and co-citedness measure the text similarity rather indirectly. Although CPI is believed to reflect the content of the citance (Gipp and Beel 2009), it involves no content elements in its calculation. As a result, adding co-opinionatedness that has a double function, i.e. both content and opinion similarity element, boosts their powers in retrieving similar documents. A more accurate ranking regarding relevance equipped with the added value of the opinion similarity may result in a more effective retrieval of scholarly documents which in turn facilitates accurate and quick access to knowledge (Zhao 2014). The concept can be a new contribution both in theoretically as well as practically enhancing the performance of IR systems and in research evaluation by scrutinizing distinctions between citations, and thereby, boosting the bibliometric indicators.

In spite of the positive impact of co-opinionatedness, it is necessary to conduct further studies to delve into the implicit layers of text semantics beyond lexical surface. The reasons is that due to the social nature of citation behavior, negative citations are not completely explicit and openly disclosed and hence are not easy to be lexically discovered (Athar 2014). Consequently, there could exist indirect, negative opinions that the dictionary-based NLP fails to grasp. Advanced techniques such as machine learning should be tried to see if it gains a clearer image of the co-opinion concept at implicit and explicit levels, and probably achieving a higher degree of improvement. The techniques can also be helpful in distinguishing between negative results and negative opinions. Given the importance of discovering new relations to improve IR effectiveness, in our ongoing study we are trying to test the role of co-opinionatedness in improving co-citation-based IR by boosting semantic similarity measures.

References

Abu-Jbara, A., Ezra, J., & Radev, D. R. (2013). Purpose and polarity of citation: Towards NLP-based bibliometrics. In *HLT-NAACL* (pp. 596–606).

Agarwal, S., Choubey, L., & Yu, H. (2010). Automatically classifying the role of citations in biomedical articles. In *Proceedings of American Medical Informatics Association fall symposium (AMIA)*, Washington, DC (pp. 11–15).

Amadi, U. P. (2014). *Exploiting the role of polarity in citation analysis*. Baltimore County: University of Maryland.

Andrejko, A., & Bieliková, M. (2012). Comparing instances of ontological concepts for personalized recommendation in large information spaces. *Computing and Informatics, 28*(4), 429–452.

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. *Technical Report, University of Cambridge, Computer Laboratory,(UCAM-CL-TR-856).*

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

Badran, O. M. (1984). An alternative search strategy to improve information retrieval. In *Proceedings of the 47th ASIS annual meeting* (pp. 137–140).

Bichteler, J., & Eaton, E. A. (1980). The combined use of bibliographic coupling and cocitation for document retrieval. *Journal of the American Society for Information Science, 31*(4), 278.

Bodoff, D., & Li, P. (2007, July). Test theory for assessing IR test collections. In *Proceedings of the 30th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 367–374). ACM.

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.

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

Brooks, T. A. (1985). Private acts and public objects: An investigation of citer motivations. *Journal of the American Society for Information Science, 36*(4), 223–229.

Callahan, A., Hockema, S., & Eysenbach, G. (2010). Contextual cocitation: Augmenting cocitation analysis and its applications. *Journal of the American Society for Information Science and Technology, 61*(6), 1130–1143.

Cavalcanti, D. C., Prudêncio, R. B., Pradhan, S. S., Shah, J. Y., & Pietrobon, R. S. (2011). Good to be bad? Distinguishing between positive and negative citations in scientific impact. In *2011 23rd IEEE international conference on tools with artificial intelligence (ICTAI)* (pp. 156–162). IEEE.

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.

Dabrowska, A., & Larsen, B. (2015). Exploiting citation contexts for physics retrieval. In *Second workshop on bibliometric-enhanced information retrieval* (pp. 14–21).

Ding, Y., Zhang, G., Chambers, T., Song, M., Wang, X., & Zhai, C. (2014). Content-based citation analysis: The next generation of citation analysis. *Journal of the association for information science and technology, 65*(9), 1820–1833.

Dong, C., & Schäfer, U. (2011). Ensemble-style self-training on citation classification. In *IJCNLP* (pp. 623–631).

Doslu, M., & Bingol, H. O. (2016). Context sensitive article ranking with citation context analysis. *Scientometrics, 108,* 653–671.

Egghe, L., & Rousseau, R. (1990). *Introduction to informetrics: Quantitative methods in library, documentation and information science*. Amsterdam: Elsevier.
Elkiss, A., et al. (2008). Blind men and elephants: What do citation summaries tell us about a research article? *Journal of the American Society for Information Science and Technology, 59*(1), 51–62.
Esuli, A., & Sebastiani, F. (2007). SentiWordNet: A high-coverage lexical resource for opinion mining. Technical Report ISTI-PP-002/2007, Institute of Information Science and Technologies (ISTI) of the Italian National Research Council (CNR). http://nmis.isti.cnr.it/sebastiani/Publications/2007TR02.pdf.
Eto, M. (2012). Spread co-citation relationship as a measure for document retrieval. In *Proceedings of the fifth ACM workshop on research advances in large digital book repositories and complementary media* (pp. 7–8). ACM.
Eto, M. (2013). Evaluations of context-based co-citation searching. *Scientometrics, 94*(2), 651–673.
Eto, M. (2014). Document retrieval method using random walk with restart on weighted co-citation network. *Proceedings of the American Society for Information Science and Technology, 51*(1), 1–4.
Eto, M. (2015). Combination effects of word-based and extended co-citation search algorithms. In *Proceedings of the 15th ACM/IEEE-CS joint conference on digital libraries* (pp. 245–246). ACM.
Fujiwara, T., & Yamamoto, Y. (2015). Colil: A database and search service for citation contexts in the life sciences domain. *Journal of biomedical semantics, 6*(1), 38.
Gipp, B., & Beel, J. (2009). Citation proximity analysis (CPA)-A new approach for identifying related work based on co-citation analysis. In *Proceedings of the 12th international conference on scientometrics and informetrics (ISSI'09)* (Vol. 2, pp. 571–575). Rio de Janeiro (Brazil): International Society for Scientometrics and Informetrics.
Hamedani, M. R., Kim, S. W., & Kim, D. J. (2016). SimCC: A novel method to consider both content and citations for computing similarity of scientific papers. *Information Sciences, 334,* 273–292.
Hanney, S., Grant, J., Jones, T., & Buxton, M. (2005). Categorising citations to trace research impact. In *Proceedings of the 10th international conference of the international society for scientometrics and informetrics*. Stockholm: Karolinska University Press.
Hasanain, M., Suwaileh, R., Elsayed, T., Kutlu, M., & Almerekhi, H. (2018). EveTAR: Building a large-scale multi-task test collection over Arabic tweets. *Information Retrieval Journal, 21*(4), 307–336.
Hernández-Alvarez, M. Y. R. I. A. M., & Gomez, J. M. (2016). Survey about citation context analysis: Tasks, techniques, and resources. *Natural Language Engineering, 22*(03), 327–349.
Janssens, A. C. J., & Gwinn, M. (2015). Novel citation-based search method for scientific literature: Application to meta-analyses. *BMC Medical Research Methodology, 15*(1), 1.
Järvelin, K., & Kekäläinen, J. (2002). Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems (TOIS), 20*(4), 422–446.
Jeong, Y. K., Song, M., & Ding, Y. (2014). Content-based author co-citation analysis. *Journal of Informetrics, 8*(1), 197–211.
Jia, M. (2018). Citation function and polarity classification in biomedical papers. Electronic Thesis and Dissertation Repository, 5367.
Jochim, C., & Schütze, H. (2012). Towards a generic and flexible citation classifier based on a faceted classification scheme. In *Proceedings of COLING 2012* (pp. 1343–1358).
Kanoulas, E., & Aslam, J. A. (2009). Empirical justification of the gain and discount function for nDCG. In *Proceedings of the 18th ACM conference on Information and knowledge management* (pp. 611–620). ACM.
Kekäläinen, J. (2005). Binary and graded relevance in IR evaluations—Comparison of the effects on ranking of IR systems. *Information Processing and Management, 41*(5), 1019–1033.
Leung, P. T., Macdonald, E. M., Stanbrook, M. B., Dhalla, I. A., & Juurlink, D. N. (2017). A 1980 letter on the risk of opioid addiction. *New England Journal of Medicine, 376*(22), 2194–2195.
Lipetz, B. A. (1965). Improvement of the selectivity of citation indexes to science literature through inclusion of citation relationship indicators. *Journal of the Association for Information Science and Technology, 16*(2), 81–90.
Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies, 5*(1), 1–167.
Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In *Mining text data* (pp. 415–463). Boston, MA: Springer.
Liu, S., Chen, C., Ding, K., Wang, B., Xu, K., & Lin, Y. (2014). Literature retrieval based on citation context. *Scientometrics, 101*(2), 1293–1307.
MacRoberts, M. H., & MacRoberts, B. R. (1984). The negational reference: Or the art of dissembling. *Social Studies of Science, 14*(1), 91–94.

MacRoberts, M. H., & MacRoberts, B. R. (1989). Problems of citation analysis: A critical review. *Journal of the American Society for information Science, 40*(5), 342–349.

Mahalakshmi, G. S., Siva, R., & Sendhilkumar, S. (2015). Context based retrieval of scientific publications via reader lens. In *Computational intelligence in data mining* (Vol. 3, pp. 583–596). Springer India.

Martyn, J. (1964). Bibliographic coupling. *Journal of Documentation, 20*(4), 236.

Matosin, N., Frank, E., Engel, M., Lum, J. S., & Newell, K. A. (2014). Negativity towards negative results: A discussion of the disconnect between scientific worth and scientific culture. *Disease Models & Mechanisms*, *7*:171–173.

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

Nakov, P. I., Schwartz, A. S., & Hearst, M. (2004). Citances: Citation sentences for semantic analysis of bioscience text. In *Proceedings of the SIGIR'04 workshop on search and discovery in bioinformatics* (pp. 81–88).

Parthasarathy, G., & Tomar, D. C. (2014). Sentiment analyzer: Analysis of journal citations from citation databases. In *2014 5th international conference- confluence the next generation information technology summit (confluence)* (pp. 923–928). IEEE.

Parthasarathy, G., & Tomar, D. C. (2015). A survey of sentiment analysis for journal citation. *Indian Journal of Science and Technology*. https://doi.org/10.17485/ijst/2015/v8i35/55134.

Piao, S., Ananiadou, S., Tsuruoka, Y., Sasaki, Y., & McNaught, J. (2007). Mining opinion polarity relations of citations. In *International workshop on computational semantics (IWCS)* (pp. 366–371).

Ritchie, A., Robertson, S., & Teufel, S. (2008). Comparing citation contexts for information retrieval. In *Proceedings of the 17th ACM conference on Information and knowledge management* (pp. 213–222). ACM.

Saraçoğlu, R., Tütüncü, K., & Allahverdi, N. (2007). A fuzzy clustering approach for finding similar documents using a novel similarity measure. *Expert Systems with Applications, 33*(3), 600–605.

Schafer, U., & Spurk, C. (2010). TAKE scientist's workbench: semantic search and citation-based visual navigation in scholar papers. In *2010 IEEE fourth international conference on semantic computing (ICSC)* (pp. 317–324). IEEE.

Segaran, T. (2007). *Programming collective intelligence: Building smart web 2.0 applications*. Beijing: O'Reilly Media, Inc.

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).

Shuy, R. W. (2003). 22 Discourse analysis in the legal context. *The Handbook of Discourse Analysis, 18,* 437.

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

Small, H. (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.

Smith, L. C. (1981). Citation analysis. *Library Trends, 30*(1), 83–106.

Su, M. C., & Chou, C. H. (2001). A modified version of the K-means algorithm with a distance based on cluster symmetry. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23*(6), 674–680.

Teufel, S., Siddharthan, A., & Tidhar, D. (2006). Automatic classification of citation function. In *Proceedings of the 2006 conference on empirical methods in natural language processing* (pp. 103–110). Association for Computational Linguistics.

Teufel, S., Siddharthan, A., & Tidhar, D. (2009). An annotation scheme for citation function. In *Proceedings of the 7th SIGdial workshop on discourse and dialogue* (pp. 80-87). Association for Computational Linguistics.

Urbano, J., Marrero, M., & Martín, D. (2013). On the measurement of test collection reliability. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval* (pp. 393–402). ACM.

Wang, Y., Wang, L., Li, Y., He, D., Chen, W., & Liu, T. Y. (2013, April). A theoretical analysis of NDCG ranking measures. In *Proceedings of the 26th annual conference on learning theory (COLT 2013)* (Vol. 8).

White, H. D. (2016). Bag of works retrieval: TF* IDF weighting of co-cited works. In *BIR@ ECIR* (pp. 63–72).

Yoon, S. H., Kim, S. W., & Park, S. (2016). C-Rank: A link-based similarity measure for scientific literature databases. *Information Sciences, 326,* 25–40.

Yu, B. (2013). Automated citation sentiment analysis: What can we learn from biomedical researchers. *Proceedings of the American Society for Information Science and Technology, 50*(1), 1–9.

Zhao, H. (2014). Sharding for literature search via cutting citation graphs. In *2014 IEEE international conference on Big Data (Big Data)* (pp. 77–79). IEEE.

Affiliations

Maryam Yaghtin[1] · Hajar Sotudeh[1] · Mahdieh Mirzabeigi[1] · Seyed Mostafa Fakhrahmad[2] · Mehdi Mohammadi[3]

Maryam Yaghtin
yaghtin.maryam@gmail.com

Mahdieh Mirzabeigi
mmirzabeigi@gmail.com

Seyed Mostafa Fakhrahmad
mfakhrahmad@yahoo.com

Mehdi Mohammadi
m48r52@gmail.com

1 Department of Knowledge and Information Sciences, Faculty of Education and Psychology, Eram Campus, Shiraz University, Shiraz, Iran

2 Department of Computer Science and Engineering, School of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran

3 Department of Educational Management and Planning, Faculty of Education and Psychology, Eram Campus, Shiraz University, Shiraz, Iran