

# Towards a new perspective on context based citation index of research articles

ParvezAhmad Kazi<sup>1</sup>  · Manasi Patwardhan<sup>1</sup> · Pushkar Joglekar<sup>1</sup>

Received: 10 July 2015 / Published online: 2 February 2016  
© Akadémiai Kiadó, Budapest, Hungary 2016

**Abstract** Citation index measures the impact or quality of a research publication. Currently, all the standard journal citation indices are used to measure the impact of individual research article published in those journals and are based on the citation count, making them a pure quantitative measure. To address this, as our first contribution, we propose to assign weights to the edges of citation network using three context based quality factors: 1. Sentiment analysis of the text surrounding the citation in the citing article, 2. Self-citations, 3. Semantic similarity between citing and cited article. Prior approaches make use of PageRank algorithm to compute the citation scores. This being an iterative process is not essential for acyclic citation networks. As our second contribution, we propose a non-iterative graph traversal based approach, which uses the edge weights and the initial scores of the non-cited nodes to compute the citation indices by visiting the nodes in topologically sorted order. Experimental results depict that rankings of citation indices obtained by our approach are improved over the traditional citation count based ranks. Also, our rankings are similar to that of PageRank based methods; but, our algorithm is simpler and 70 % more efficient. Lastly, we propose a new model for future reference, which computes the citation indices based on solution of system of linear inequalities, in which human-expert's judgment is modeled by suitable linear constraints.

**Keywords** Citation index · PageRank · Sentiment analysis · System of linear inequalities · Topological sorting

---

✉ ParvezAhmad Kazi  
parvezahmed.kazi@gmail.com

Manasi Patwardhan  
manasi.patwardhan@vit.edu

Pushkar Joglekar  
pushkar.joglekar@vit.edu

<sup>1</sup> Computer Engineering Department, Vishwakarma Institute of Technology, Pune, India

## Introduction

In past decade, quality of research is used as one of the major parameter by academic and research institutes as well as industries to evaluate the performance of scientists, faculties and researchers, for the purpose of hiring, promoting and funding the research. It is also used to compare the academic impact of an individual or groups. Initially, number of citations received was used as a measure for quality of a research article (Eugene 1955). Later, Garfield proposed impact factor (IF), a journal evaluation tool using citation analysis (Garfield 1972); which is a measure of impact of journal in terms of average number of citations received per article published in that journal, during the two preceding years. As these measures are easily obtainable, they are widely used today. However, these citation indices have some limitations (Amin and Mabe 2003; Casal 2004; Saha et al. 2003), because of which it is questionable whether they can serve to be a true qualitative measure for analysis.

IF is a measure for journal, not for an individual article published in that journal. Measuring research quality of an individual based on the citation indices of journals, where his or her research articles are published is actually not an apt method to evaluate that individual. Moreover, IF considers citations irrespective of the quality of the citing article. This makes IF liable to manipulation and misuse.

Prior approaches to address this problem have made use of Google's PageRank algorithm to compute citation index for an article using citation network (Li and Willett 2009). Many of these approaches don't take into account the quality of the citing articles while computing the citation index of the article. They give equal weights to all the edges of the citation network. In our approach we have defined a context based citation index for an article that is derived from providing distinct weights to the citing articles based on their quality. We mainly consider three quality factors: first, sentiment analysis of the text surrounding the citation in the citing article; second, use of weighing scheme to provide lower weights for self-citations; and third, higher weights given to semantic similarity between citing and cited article. Using these three factors along with the quality scores of the citing articles, we produce a citation index for a research article.

All the existing PageRank based citation index computation approaches use original Google's PageRank algorithm (Haddadene et al. 2012; Li and Willett 2009; Ma et al. 2008; Qiao et al. 2012; Sayyadi and Getoor 2009; Singh et al. 2011; Su et al. 2011; Walker et al. 2007; Yan et al. 2011), which considers virtual weighted edges from certain nodes to address the problem of rank sink. However, addition of such false edges violates the properties of citation network giving false results for citation score computations.

The edges of the citation network follow time ordering. An older article can never cite a newer article. This makes the underlying directed graph of the citation network acyclic and thus there is no need for an iterative algorithm for computations of scores of individual nodes. The PageRank algorithm originally developed to list the web-pages in the network in the order of their relevance and the popularity. Since both the relevance and the popularity are more quantitative and measurable parameters, human intervention is not necessary in this case and the ranking can be computed in fully automated manner. All the existing PageRank based algorithms for computing citation indices as well do not involve any kind of human-expert intervention or supervised learning.

In case of citation networks we are more interested in ordering the articles based on their quality rather than the popularity. Determining absolute quality of the article in fully automated manner without any kind of supervised learning apriori is extremely difficult

task. So we believe that certain amount of human-expert intervention would help in the process of ranking the articles.

In our approach we assign weights to the edges using the above mentioned three quality factors and scores of the individual nodes are computed one by one in their topologically sorted order (the ordering can be found using a simple Depth First Search traversal) as appose to prior iterative PageRank based algorithms used for the same. This makes our algorithm more efficient in terms of time-complexity.

This paper also proposes an innovative approach for the problem based on finding solution for system of linear inequalities. We express the score corresponding to any node in the network as a linear form in the scores corresponding to the source nodes using context based quality factors. We can incorporate expert-advice by introducing suitable linear constraints on the linear-forms associated with the nodes. Thus, ultimately we get a system of linear inequalities in which we want to maximize number of satisfied constraints. So essentially, the computation of citation indices reduces to finding solution for Maximum Feasible Subsystem Problem (MaxFS) for which several good heuristics are known (Amaldi and Kann 1995).

The remainder of this paper is organized in the following way. “[Related work](#)” section discusses the related work. “[PageRank for citation network](#)” section describes the link network’s PageRank algorithm used for citation network and its limitations. In “[Methodology](#)” section we define the overall methodology of our approach. The Context based quality factors for the edge weight calculations are presented in the “[Computation of edge weights](#)” section. “[Citation index using graph traversal](#)” section describes our approach of computing citation index using graph traversal method. “[Data](#)” section describes the dataset used in this work. “[Experimental results and discussion](#)” section presents the experimental results. As a future scope for research, we propose another innovative approach in “[Citation index using solution for MaxFS problem: a model proposed for future directions](#)” section, which incorporates expert-advice in computation of citation indices and solves this problem by reducing it to maximum feasible subsystem problem. In “[Conclusion](#)” section we discuss conclusions.

## Related work

The current research evaluation methods such as citation analysis use only the citation count of the research article to evaluate the impact of article; it does not consider the quality of citing article. The link network’s PageRank algorithm ranks the webpages based on the importance of the webpages that point to it. Citation network with articles as nodes and citations as edges have similar characteristics of link network. So many researchers applied PageRank algorithm on citation network (Li and Willett 2009; Yan et al. 2011; Singh et al. 2011; Walker et al. 2007; Sayyadi and Getoor 2009; Su et al. 2011; Ma et al. 2008; Qiao et al. 2012).

Li and Willett (2009) presented an ArticleRank algorithm, which is an extension of the PageRank algorithm. This approach has removed the bias caused due to more contribution made by the citing articles having fewer references towards the computation of the rank of the cited article than the contribution made by the citing articles with more references. Yan et al. (2011) applied P-rank, a PageRank based algorithm on the heterogeneous scholarly network. Authors define the heterogeneous scholarly network as the network which allows authors to interact with papers, journals to interact with papers and papers to interact with

other papers. It assigns fixed lower weights to the self-citations and takes into account the importance of citing article, journal and authors. Singh et al. (2011) have considered article publication time in their algorithm. The same publication time factor has also been considered by Walker et al. (2007). They have derived the PageRank algorithm and named it as the CiteRank method, which assigns more weight to recent publications.

Sayyadi and Getoor (2009) have proposed the PageRank based algorithm FutureRank, which calculates the future PageRank score based on citations that would be received in future. Su et al. (2011) have addressed the problem of missing citations in the dataset and proposed PrestigeRank, a PageRank based algorithm. They introduced the concept of ‘virtual node’ which receives all the citations which are missing in the dataset. The similar problem of missing data is also addressed by Ma et al. (2008). They have introduced concept of ‘internal citations’, which is the number of citations made only by the papers from the selected dataset.

Qiao et al. (2012) proposed a PageRank based algorithm using text mining approach. They consider the correlation between two papers as the citation weight and IF of the journal in which that paper is published as the weight corresponding to the paper. Haddadene et al. (2012) have used similar methods in their algorithm; instead they calculate the similarity between two research articles using Jaccard index. Use of sentiment analysis of citations to rate citations as positive, negative or neutral to compute the impact of research publications has been widely studied (Cavalcanti et al. 2011; Piao et al. 2007; Stamou et al. 2009).

## PageRank for citation network

Current literature demonstrates that the PageRank algorithm, which is used for the link network to compute the rank of the web pages, is applied on the citation network to measure the impact of the research. Though link network and citation network have similarities, applying PageRank algorithm as it is on the citation network violates some of the properties of the citation network. In this section, we discuss a citation network, a PageRank algorithm and its limitations on citation network.

### Citation network

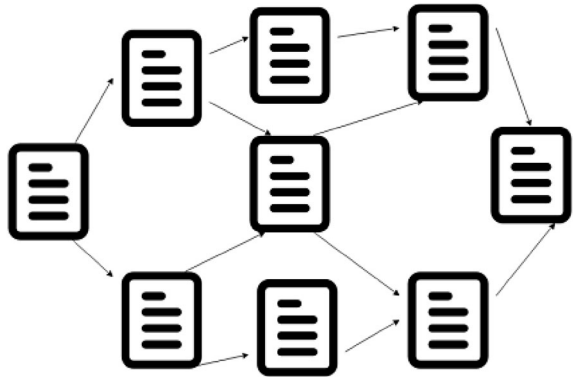
A citation network consists of a directed graph having the research articles as its nodes. There is a directed edge from the node  $D_i$  to the node  $D_j$  if the article  $D_i$  cites the article  $D_j$ . Clearly the edges in the citation network follow the time ordering of the articles corresponding to the nodes, i.e. newer articles can cite only the older articles. A simple form of citation network is shown in Fig. 1.

### PageRank algorithm

The basic idea behind PageRank algorithm can be summarized as follows: A page is considered to be important if it is cited by other important pages (Page et al. 1999). The basic PageRank formula defined by Langville and Meyer (2011) is shown in (1):

$$\text{PR}(u) = \sum_{v \in I(u)} \frac{\text{PR}(v)}{|O_v|} \quad (1)$$

**Fig. 1** Simple citation network



where  $u, v$  are the web pages,  $I(u)$  is the set of web pages pointing to page  $u$ ,  $O(v)$  is the set of web pages pointed by page  $v$ , and  $PR(x)$  is the PageRank value for the page  $x$ .

PageRank algorithm follows the random surfer model on the web. It starts by picking a random page. Once the page is picked it follows the link corresponding to one of the URLs on the page to go to the next page and in this way the algorithm visits successive pages and traverses through the network mimicking a random walk on the underlying graph. The formula (1) fails when the network has dangling nodes or the cycles. As dangling nodes don't have any outgoing links, they can't share the rank with others and nodes in the cycle can't distribute rank with others, leading to dangling pages trapping the random surfer forever. This problem is called as rank sink problem. The obvious solution to this problem is, pick another URL at random and continue surfing again (Langville and Meyer 2011) [9]. To address this issue a stochastic adjustment is done in the basic formula. A parameter  $d$  is introduced called 'damping factor',  $d$  which is typically set to 0.85. It allows a random surfer to visit any random node with a probability of  $0.15/N$  even after encountering a sink node, where  $N$  are the total number of nodes. To be fair with pages that are not sinks, these random transitions are added to all the nodes in the Web, with a probability  $(1 - d) = 0.15$ . The new formula for computing PageRank of nodes is as below:

$$PR(u) = \frac{(1 - d)}{N} + d \times \sum_{v \in I(u)} \frac{PR(v)}{|O_v|} \tag{2}$$

where  $d$  is a damping factor,  $d \in [0, 1]$  and  $N$  is total number of pages in the network.

Iterative method of PageRank computation computes PageRank score for one page at a time, so Langville and Meyer (2011) presented same PageRank formula using matrices which computes a PageRank vector ' $\pi$ ' per iteration. It is shown in (3):

$$\pi^{(k+1)T} = \pi^{(k)T} G \tag{3}$$

where  $G$  is stochastic, irreducible, periodic and dense matrix called as Google Matrix and is calculated as shown in (4):

$$G = d \cdot H + (d \cdot \alpha + (1 - d)e) \frac{1}{n} e^T \tag{4}$$

where  $H$  is a row normalized hyperlink matrix,  $d$  is damping factor,  $\alpha$  is binary dangling node vector,  $n$  is the number of nodes in the network,  $e$  is a row vector with all values to 1.

From (3), we can see that PageRank vector is nothing but the eigenvector of the Google matrix for eigenvalue 1. Hyperlink matrix  $H$  is computed as shown in (5):

$$H_{ij} = \begin{cases} 1/|P_i|, & \text{if } i \text{ points to } j \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

### Limitations of PageRank for citation networks

Apart from some similarities with link network, citation network has its own properties which should be considered while applying PageRank algorithm on it. While calculating Google matrix of PageRank in (4), we can see that PageRank adds more than  $N^2$  virtual edges in the network to overcome the problem of rank sink; where  $N$  is total number of nodes in the network. From each dangling node, it adds edges to all the nodes in the network with  $d/N$  weight and from each node it adds edges to every other node in the network with  $(1 - d)/N$  weight. Applying PageRank algorithm as it is would add such virtual edges in the citation graph leading to violation of timely order property of citation network.

A directed graph is said to be irreducible if any pair of nodes in the graph is connected via a directed path. For applying PageRank algorithm we need underlying directed graph of the network to be irreducible. If the graph is not irreducible then we need to add virtual edges to make it irreducible before application of the PageRank algorithm. In case of citation networks, any article can cite only the articles published in the past. So the underlying directed graph for the citation network is acyclic. Every edge follows the ordering on the time of publication of the articles. So clearly, the graph corresponding to the citation network is not irreducible. The solution opted in the case of web is not valid in this context as adding virtual edges violates the timely ordering on the nodes of the citation network.

Since the underlying graph is acyclic, using iterated PageRank algorithm to update the scores successively is not a “natural” approach. A natural approach would be to visit the nodes in the topologically sorted order (this ordering can be computed by simple Depth First Search traversal on the citation network) and compute their scores in a single pass as appose to an iterative PageRank based algorithm. This also makes our algorithm more efficient in terms of time-complexity.

The PageRank algorithm originally developed to list the web-pages in the network in the order of their relevance and the popularity. Since both the relevance and the popularity are more quantitative and measurable parameters, human intervention is not necessary in this case and the ranking can be computed in fully automated manner. All the existing PageRank based algorithms for computing citation indices as well do not involve any kind of human-expert intervention or supervised learning. In case of citation networks we are more interested in ordering the articles based on their quality rather than the popularity. Determining absolute quality of the article without any kind of supervised learning is extremely difficult task. So we believe that certain amount of human-expert intervention would be helpful in the process of ranking the articles.

Because of the above mentioned dissimilarities in the citation network and the link network; applying PageRank algorithm as it is for the citation networks is not a valid solution. In this work we propose a new approach based on solution of system of linear inequalities to calculate citation index of research articles. The basic idea of our approach is similar to the PageRank, i.e. a research article has higher score, if it has received

citations by other high scored research articles; but our approach is simpler and natural. We don't need to add any non-existing or virtual edges in the network which violate the time ordering of the nodes.

We express the score corresponding to any node in the network as a linear form in the scores corresponding to the source nodes. Context based quality factors discussed in the “[Computation of edge weights](#)” section are used for defining the linear forms corresponding to the nodes. Our approach is amenable to easy incorporation of human-expert advice. We can get randomly chosen pairs of nodes judged by the domain-experts and add appropriate linear constraints to our system of linear inequalities to reflect these judgments. Our goal is to compute suitable values for scores of the source nodes so as to maximize number of satisfied inequalities. This is precisely the Maximum Feasible Subsystem Problem (MaxFS). There are several good heuristics for solution of MaxFS (Amaldi and Kann 1995) which can be used to finally compute the citation indices.

### Methodology

The proposed research problem is divided into two parts, first is computing edge weights using context based quality factors and second is our perspectives towards computing citation index. The general architecture of our approach is shown in Fig. 2.

A possible input to our system is a citation database having full text articles; but as per our study, there is no such citation database readily available, which provides full text articles. We crawl the ACM digital library and download the full text articles along with their metadata. We build a citation network on the citation database we generated. Next step is to compute the edge weights using different text mining techniques. Final step is our perspective towards computing the node weights i.e. citation indices of articles using

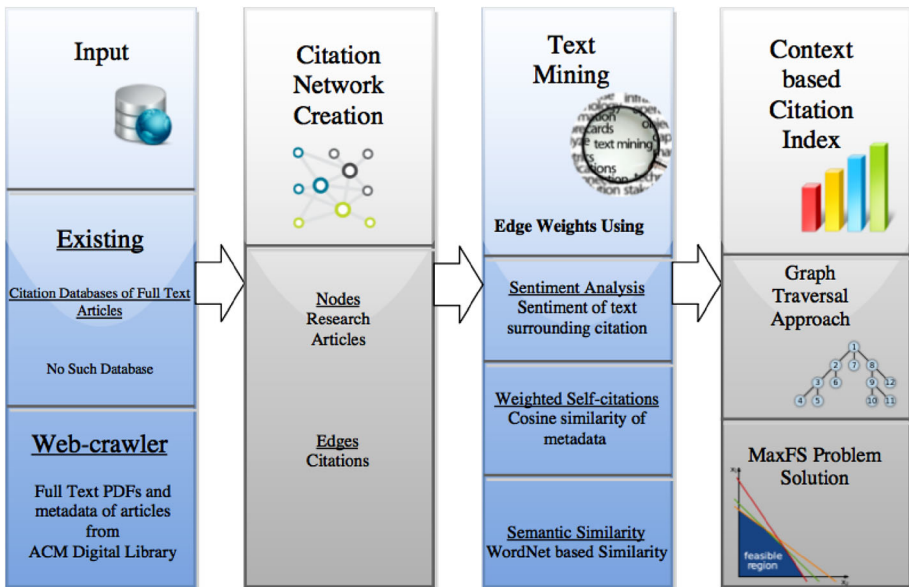


Fig. 2 Architecture of our perspective towards citation index for research articles

Graph Traversal approach and MaxFS Problem Solution. Edge weights computation is discussed in “[Computation of edge weights](#)” section and node weights computation is presented in “[Citation index using graph traversal](#)” and “[Citation index using solution for MaxFS problem: a model proposed for future directions](#)” sections.

## Computation of edge weights

We propose to calculate the edge weights of citation network using three different factors: (1) sentiment analysis of citations; (2) weighted self-citations; (3) the semantic similarity between the citing and the cited article. These factors help us in taking into account the context in which a paper is cited. We discuss these factors in details below:

### Sentiment analysis of citation

Sentiment refers to the attitude, opinion or feeling toward some entity (Liu 2012). Sentiment analysis is the process of determining the polarity of that sentiment, i.e. positive, negative or neutral (Liu 2012). Sentiment analysis has been widely used in the field of e-commerce to get the polarity of the product reviews or user feedbacks and to plan the further strategies and decision making. In scientific literature, authors refer to other author’s work for surveying, criticizing, agreeing, recommending etc. (Nicolaisen 2007). So, any recommending or agreeing citations should have more weight than any criticized citations. In conventional citation analysis methods, all citations have equal weights. Here we assign different weights to citations based on their citing purpose. We propose a method which derives the sentiment from the cited text and classify them into three classes, i.e. positive, negative and neutral. Obviously positive citations will have more weight as compare to negative and neutral citations, whereas negative citations will have lowest weight than positive and neutral citations.

We propose an unsupervised sentiment analysis method which uses SentiWordNet, which is a lexical resource having synsets with their sentiment scores (Esuli and Sebastiani 2006). We use the latest SentiWordNet 3.0 version developed by Esuli and Sebastiani (2006) which contains synsets from WordNet 3.0. Each synset is assigned three different scores for three polarity classes, i.e. positive, negative and objective. The total score of each synset is 1, i.e.  $Pos(s) + Neg(s) + Obj(s) = 1$ .

Cited text in the citing article is the input to the method. Cited text can be in two forms: one is explicit citation where the cited paper is cited explicitly by using authors name or serial number of that reference and second is implicit citation. Implicit citations are the citations which indirectly or implicitly refer to the paper (Athar and Teufel (2012)). Such implicit citations may refer to the cited paper by using the nouns appeared in the explicit citations. We extract an explicit citation, then nouns in that explicit citation are extracted and these nouns are searched in next two paragraphs. All the sentences that contain these nouns are referred as implicit citations. We extract all explicit and implicit citations from citing article and pass it to the Stanford POS tagger written by Toutanova et al. (2003), to retrieve all the adjectives and verbs in the text surrounding the citation. We make use of adjectives and verbs for finding polarity, because they represent the sentiment in the sentence. SentiWordNet contains number of synsets for a single word; each such synset for a word is assigned a sense number that indicates its meaning. We select all senses for the



adjectives and verbs without finding correct sense number. We build a term set from extracted adjectives and verbs; then we compute the positive score for each term using (6).

$$\text{SynsetScore}(t) = \sum_{\forall \text{sense number } s} \frac{(\text{Pos}(t, s) - \text{Neg}(t, s))}{s} \tag{6}$$

where  $\text{Pos}(t, s)$  is the positive score of term ‘ $t$ ’ for sense number ‘ $s$ ’ and  $\text{Neg}(t, s)$  is the negative score of term ‘ $t$ ’ for sense number ‘ $s$ ’. The final SentiScore of term is a weighted average, where synsets are assigned weights according to their sense number. It is computed using (7):

$$\text{Score}(t) = \frac{\text{SynsetScore}(t)}{\sum_{\forall \text{sense number } s} 1/s} \tag{7}$$

The average of SentiScores of each term in the cited text gives the total sentiment score of the citation between  $A$  and  $B$ . It is computed using (8):

$$\text{SentiScore}(A, B) = \frac{\sum_{\text{each term } t} \text{Score}(t)}{N} \tag{8}$$

where  $N$  is total number of adjectives and verbs in the cited text.

The final SentiScore of citation takes values between  $[-1, +1]$ , so values between  $[-1, 0)$  indicates the negative polarity, values between  $(0, +1]$  indicates the positive polarity and 0 score indicates the objectivity. We normalize this final SentiScore values to  $[0, 1]$ , so that positive citations receive scores greater than 0.5, objective citations receive 0.5 score and negative citations receive  $<0.5$  score. We use that normalized score as sentiment score for that citation.

### Weighted self-citations

Traditional citation indices are based on the number of citations received to the article. It makes them a pure quantitative measure and liable to manipulation and misuse. They can be manipulated in terms of falsely inflating it by fake citations or self-citations. Self-citation can be of three different types: (1) Author self-citation, (2) Research Group self-citation, (3) Publication self-citation.

The purpose of self-citations cannot be always to inflate the citation number, but they can be made for genuine purpose also. Considering all these facts, we propose a weighing scheme for self-citations using text mining approach. We assign the weights to self-citations according to the level of self-citation. We refer the level of self-citation as how strongly self-citation is made. If self-citation is made by same author, same research group and same publication, then the level of self-citation will be highest. The higher the level of self-citation, lower the weight will be assigned to that citation.

Our proposed approach for weighted self-citation using text mining is described below:

1. Metadata such as names and IDs of authors, institutional affiliation of authors and IDs of institutional affiliation, publication name and publication ID is extracted for each pair of articles in the dataset. The extracted metadata is represented by vectors where coordinates correspond to term frequency (TF) of the term which is calculated using (9):

$$\text{TF}(t, d) = \frac{f(t, d)}{|d|} \tag{9}$$

where  $f(t, d)$  is number of times term  $t$  appears in a document  $d$  and  $|d|$  is total number of terms in the document  $d$ .

2. Then the similarity between these metadata documents is computed using vector cosine similarity shown in (10).

$$\text{Sim}(A, B) = \frac{\sum_{i=1}^N a_i b_i}{\sqrt{\sum_{i=1}^N a_i a_i} \times \sqrt{\sum_{i=1}^N b_i b_i}} \quad (10)$$

Here the value of  $\text{Sim}(A, B)$  indicates the level of self-citation.

3. We calculate the self-citation weight of citation between article  $A$  and  $B$  using (11):

$$\text{SelfCitScore}(A, B) = \begin{cases} (1 - \text{Sim}(A, B)), & \text{self-citation} \\ 1, & \text{otherwise} \end{cases} \quad (11)$$

All self-citations are assigned  $(1 - \text{Sim}(A, B))$  score, therefore higher level of self-citations get assigned lower weights and all non-self-citations is assigned maximum score of 1.

### Semantic similarity

It is often observed that there is a direct semantic relationship between the citing and cited article (Harter et al. 1993). This relationship between research articles is useful in many applications like: recommender systems, information retrieval and document categorization. This motivates us to give some more weights to the links between the article pairs which has semantic similarity. This leads to the weight of the citing article getting distributed amongst its cited articles unevenly, such that the article which is semantically closer gets higher contribution. Semantic similarity is a confidence score that reflects the semantic relation between the meanings of two articles. To calculate this semantic similarity score, we use WordNet lexical database. WordNet is a large lexical database which is available online and provides a large repository of English lexical items that are linked together by their semantic relationships (Miller 1995).

As we are going to find semantic similarity between two document vectors representing the citing and cited research articles, the traditional vector cosine formula doesn't work. Because there are certain features such as words, which are different in VSM model but they can be semantically similar. For example, words like 'play' and 'game' are different words and thus they are mapped to different dimensions in VSM, but it is obvious that they are semantically similar.

The abstracts of two articles depict the level of similarity between two articles. Therefore, we use terms from the abstract to find the semantic similarity between two articles. As stated above, we use WordNet hierarchy to find relatedness between two terms. We make use of the WordNet based Leacock–Chodorow (LCH) similarity (Leacock and Chodorow 1998) to measure the semantic similarity between two terms as shown in (12). As LCH similarity measure considers words having part of speech (POS) type of noun and verb, we consider only nouns and verbs from each abstract for our similarity calculation

$$\text{LCH}(w_1, w_2) = \max \left[ -\log \frac{L(w_1, w_2)}{2 \times \max \text{Depth}(\text{POS})} \right] \quad (12)$$

where  $L(w_1, w_2)$  is the length of the shortest path between  $w_1$  and  $w_2$  and  $\max \text{Depth}(\text{POS})$  is the maximum depth of taxonomy.

The articles’ semantic similarity depends upon the semantic correlation of the individual features of the article vectors. To reflect this semantic similarity we use the Soft Cosine measure proposed by Sidorov et al. (2014). It is an extension of vector cosine measure. It is calculated as shown in (13).

$$\text{Soft Cosine}(A, B) = \frac{\sum \sum_{i,j}^N s_{ij} a_i b_j}{\sqrt{\sum \sum_{i,j}^N s_{ij} a_i a_j} \times \sqrt{\sum \sum_{i,j}^N s_{ij} b_i b_j}} \tag{13}$$

We take the union of the terms in the abstracts of both the articles, which forms  $N$  dimensionality vectors  $a$  and  $b$  for articles  $A$  and  $B$  respectively, whose coordinates represents the frequencies  $a_i$  and  $b_j$  of the terms in the feature vector and  $s_{ij}$  is the similarity measure between  $i$ th word in  $a$  and  $j$ th word in  $b$ .

Our aim is to find semantic similarity, so in our approach:

$$s_{ij} = LCH(f_{a,i}, f_{b,j}) \tag{14}$$

where  $f_{a,i}$  is the  $i$ th feature in vector  $a$  and  $f_{b,j}$  is the  $j$ th feature in the vector  $b$ . So the semantic similarity score between article  $A$  and  $B$  is defined as shown in (15):

$$\text{SimScore}(A, B) = \text{Soft Cosine}(A, B) \tag{15}$$

Obviously, if we consider that there is no similarity between features, i.e.  $s_{i,i} = 1$  and  $s_{i,j} = 0$  for  $i \neq j$ , then (13) is equivalent to the original cosine measure (10).

### Computation of edge weights

For a given citation network, we assign the weights to edges using the three factors computed in (8), (11) and (15). First, sentiment analysis of the text surrounding the citation in the citing article; second, use of weighing scheme to provide lower weights for self-citations; and third, higher weight given to semantic similarity between citing and cited article.

These three factors are weighed according to their importance for the citation. To assign a single qualitative weight to the edge, we take a weighted sum of all three scores obtained by that edge. It is shown below in (16):

$$W_{A,B} = w_{\text{Senti}} \times \text{SentiScore}(A, B) + w_{\text{Self}} \times \text{SelfCitScore}(A, B) + w_{\text{Sim}} \times \text{SimScore}(A, B) \tag{16}$$

where  $A$  is citing article and  $B$  is cited article,  $\text{SentiScore}(A, B)$  is the sentiment score with importance weight of  $w_{\text{Senti}}$ ;  $\text{SelfCitScore}(A, B)$  is the self-citation score between  $A$  and  $B$  with importance weight of  $w_{\text{Self}}$  and  $\text{SimScore}(A, B)$  is semantic similarity between  $A$  and  $B$  with importance weight of  $w_{\text{Sim}}$ .

We define the values of these importance weight factors such that  $w_{\text{Senti}} > w_{\text{Self}} > w_{\text{Sim}}$ . We give more weight to sentiment analysis because that mainly decide the context with which the article cited in terms of positive or negative reference. The next in line is the self-citation, whereas lowest weight is assigned to the similarity as it is applicable to only the citations where citing and cited articles are from the same domain. The values for these weight factors are set to:

$$\begin{aligned}w_{\text{Senti}} &= 0.4 \\w_{\text{Self}} &= 0.35 \\w_{\text{Sim}} &= 0.25\end{aligned}$$

This combined weight of an edge is then normalized with respect to all outgoing edges of the source of that edge using (17):

$$C_{AB} = \begin{cases} \frac{W_{AB}}{\sum_{D \in O(A)} W_{AD}}, & \text{if } A \text{ cites } B \\ 0, & \text{otherwise} \end{cases} \quad (17)$$

where  $O(A)$  is the set of articles cited by  $A$ .

## Citation index using graph traversal

In this section we describe a simple graph traversal based algorithm to compute rank of articles in the citation network using context based weight assignment for the edges defined in (17). As discussed earlier, citation networks are directed acyclic graphs. There are three different types of nodes in DAG: first, source nodes which don't have any incoming edges; second, sink nodes which don't have any outgoing edges and third, internal nodes which have at least one incoming edge and outgoing edge. Source nodes in the citation network are the articles which don't have any citations from other articles. It is clear that we cannot draw any inference pertaining to the quality of the research article corresponding to the source nodes solely based on the structure of the citation network as these articles don't have any citations. So while computing the scores of the nodes in the network we start with the assumption that all source nodes have same score  $\alpha$ .

Every internal and sink node is assigned a score based on the context based weights of its incoming edges and the scores of the nodes from where those incoming edges originate. We compute the scores of nodes using (18).

$$\text{Score}(P) = \begin{cases} \sum_{Q \in I(P)} \text{Score}(Q) \times C_{QP}, & \text{if } P \text{ is non-source article} \\ \alpha, & \text{if } P \text{ is source article} \end{cases} \quad (18)$$

where  $P$  is an article,  $I(P)$  is the set of articles which cite  $P$  and  $C_{QP}$  is the weight of the citation between article  $Q$  and  $P$  computed using (17).

Note that when scores of all source nodes are same then the relative ordering of the scores of the other nodes as well as ratios of scores of any pair of nodes is independent of the actual choice of the value of score for source nodes. When we discuss our proposed new approach based on incorporation of the expert's advice while computing the citation indices, we allow source nodes to have flexible scores in the moderate range.

There are two important features of our algorithm that differ from prior works. Firstly, we use a novel context based edge weights scheme defined in “[Computation of edge weights](#)” section while computing the scores of the nodes. Secondly, we exploit acyclic nature of underlying graph of the citation network to give a simple graph traversal based non-iterative algorithm contrasting with earlier iterative PageRank based approaches.

For a directed graph  $G(V, E)$ , the topological ordering of its vertices is an ordering such that for any edge  $(u, v)$  in  $E$ , vertex  $u$  appears prior to vertex  $v$  in the ordering. We compute topological ordering of nodes of the citation network using Depth-First search traversal as follows. In the depth first search traversal of a directed graph, we begin with a single source

node in the stack. At any step of the algorithm we observe node A on the top of the stack. We mark node A as “visited”. If all the neighboring nodes of A are marked “visited” we pop A and mark it as “explored”, otherwise we pick a node B such that there is an edge between A and B and B is not yet visited and push B on the stack. We continue till stack is empty. If all nodes are not explored we repeat the above procedure with a new unvisited node on the stack. We continue till all the nodes are marked explored. To get a topological sorted order of nodes we simply take the reverse ordering in which the nodes are marked as “explored” in DFS traversal. It is easy to check from standard properties of DFS traversal that the ordering obtained in this way is indeed a topological ordering of the nodes. Once this ordering of nodes is obtained, we visit the nodes of the citation network in that order and compute their scores using formula (18). Note that while computing the score of a node, we would have already computed scores of the nodes from which the node has an incoming edge as we are visiting the nodes in the topological order. This gives us  $O(|V| + |E|)$  algorithm to compute scores for all the nodes in the network.

For an example, consider the simple network shown in the Fig. 3.

One of the possible topological order of nodes of above network is [D, F, B, A, C, E]. Here, D and F are source nodes, so they are initialized with same score ‘ $\alpha$ ’. The scores of the nodes are computed in topological order in the sequence as shown below:

$$\begin{aligned}
 D &= F = \alpha \\
 B &= \frac{1}{2}\alpha + \frac{1}{2}\alpha \\
 A &= \frac{1}{2}\alpha + \frac{1}{3}B \\
 C &= \frac{1}{3}B \\
 E &= \frac{1}{2}\alpha + \frac{1}{3}B + 1A
 \end{aligned}$$

We use  $\alpha = 0.05$  and compute the scores for nodes as shown in Table 1.

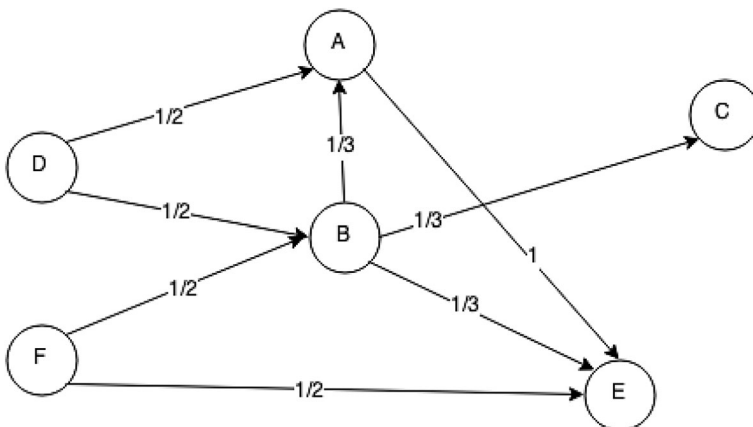


Fig. 3 Example of citation network

**Table 1** Citation index using graph traversal

Node	Score	Rank
D, F	0.05	2
B	0.05	2
A	0.042	3
C	0.0167	4
E	0.084	1

## Data

We choose the ACM digital library as the data source. We randomly select an article, and downloaded all the articles citing to it and cited by it. Then we downloaded all the articles which cite to and cited by the articles that we have downloaded, and so on. Along with these full text PDF articles, we have created a metadata dataset which contains information like article ID, title of the article, year of publication, authors and their affiliation, publication name, publication ISBN and DOI of article. Then we built a citation network with 541 nodes and 659 edges. This citation network contains 516 source nodes.

## Experimental results and discussion

For every citation, we compute three different weights for the three different quality factors using (8), (11) and (15) and then compute final normalized weight for that citation using (17). After computing weights for edges of the network, we generate separate equation for each non-source node using (18). The citation network built here has 25 non-source nodes and 516 source nodes which are initialized with same score ‘ $\alpha$ ’.

Table 2 shows the scores and ranks of top 15 articles obtained by our graph traversal based approach. It also shows the scores and the ranks computed using PageRank algorithm with our proposed qualitative edge weights.

From Table 2, we can see that the rankings obtained by our approach and the PageRank based approach are almost similar. We computed how correlated is the ranking of articles returned by proposed approach and the earlier PageRank based approach. This correlation coefficient is computed using Spearman Rank Correlation Coefficient (SRCC) (Spearman 1904), which is calculated using (19):

$$SRCC = 1 - \frac{6 \times \sum PRR_i - LPR_i}{n(n^2 - 1)} \quad (19)$$

where  $PRR_i$  and  $LPR_i$  are the PageRank based ranking and the ranking obtained by our approach of article  $i$  respectively and  $n$  is number of count of rankings.

The correlation coefficient computed using SRCC (19) is 0.995. It is very close to 1, so it indicates that these two rankings are positively correlated. This shows that, computing citation index of research articles using our algorithm is simpler and more efficient than the prior approaches that used PageRank algorithm for the same.

Table 3 shows the scores and ranks of top 15 articles obtained by graph traversal based approach and their ranks according to their citations counts.

From Table 3, we can see that in most cases citation count influences the proposed citation index, i.e. higher citation count gives higher citation index. But there are three different cases apart from this:

**Table 2** Top 15 results of proposed citation index and their ranks according to weighted PageRank scores

Article ID	Weighted PR		Graph traversal	
	Score	Rank	Score	Rank
1772751	0.0583	1	4.6476	1
1557077	0.0575	2	4.5998	2
1718520	0.0441	3	3.4996	3
1531710	0.0440	4	3.3755	4
1519089	0.0429	5	3.2705	5
1600175	0.0418	6	3.0574	6
1526806	0.0376	7	2.8639	7
1644900	0.0320	8	2.4262	8
1935845	0.0278	9	2.0277	9
1963504	0.0270	10	1.9471	10
1835933	0.0254	12	1.8945	11
1866311	0.0255	11	1.8476	12
1753543	0.0164	13	1.1855	13
1566421	0.0159	14	1.1693	14
1520584	0.0122	17	1.0479	15

**Table 3** Top 15 results of proposed citation index and their ranks according to citation count

Article ID	Citation count		Graph traversal	
	Count	Rank	Score	Rank
1772751	54	1	4.6476	1
1557077	49	4	4.5998	2
1718520	51	2	3.4996	3
1531710	51	2	3.3755	4
1519089	50	3	3.2705	5
1600175	45	7	3.0574	6
1526806	47	6	2.8639	7
1644900	43	8	2.4262	8
1935845	48	5	2.0277	9
1963504	48	5	1.9471	10
1835933	35	10	1.8945	11
1866311	38	9	1.8476	12
1753543	17	12	1.1855	13
1566421	27	11	1.1693	14
1520584	2	13	1.0479	15

1. Some articles having same citation count may have different citation indices
2. Some articles with higher citation count may have lower citation indices
3. Some articles with lower citation count may have higher citation indices

We consider the reason for each case and we find that it is because of the weights for quality factors received by these articles. Table 4 shows all weights received by these articles.

**Table 4** Types of citations received by articles

Paper ID	Citation count	Sim > 0.75	Positive citations	Negative citations	Self-citations
1531710	51	46	17	10	1
1718520	51	40	21	16	5
1963504	48	39	14	18	9
1557077	49	46	17	16	12

In first case, articles 1531710 and 1718520 have same citation count of 51; but in proposed algorithm, they are ranked 3rd and 4th respectively. From Table 4, we can see that article 1531710 received more number of citations having semantic similarity greater than 0.75 with this article than those with article 1718520. Also article 1531710 received less number of negative citations and self-citations. So despite of having same citation count, they are ranked differently based on the number of quality citations received by them.

Article 1963504 comes under 2nd case. According to citation count, it is ranked 5th, but it has 10th rank according to our proposed citation index. We find the reason for this from Table 4. Article 1963504 received less similar citations, 29 % positive citations and 38 % negative citations and also got 19 % self-citations out of 48 citations.

Third case shows the actual benefit of our proposed citation index. Despite of having less citation count, the articles receives higher citation index because of the qualitative factors. Article 1557077 is the example of this case. According to citation count, it has 4th rank, but it is ranked 2nd in our proposed citation index. This is happened because article 1557077 has received 94 % citations from similar articles, 35 % positive and 32 % negative citations. Though it has received 24 % self-citations, it is ranked higher than the article 1963504 because it has two citations from articles which are in top 15.

In our experiments, we computed the citation index using traditional weighted PageRank with our edge weights and our graph traversal based approach. The execution time taken by traditional weighted PageRank approach is 1111 ms, whereas the execution time taken up by our graph traversal approach is 330 ms. This shows 70 % improvement in the performance. With large publication datasets this gain in efficiency would be even more evident and valuable.

From above analysis, we can see that citation count as impact of research article does not consider the actual quality of citations received. Factors like semantic similarity between citing and cited article, sentiment analysis of citations and weighted self-citations determines the actual quality of the citation received. Proposed citation index overcomes problem of quantitative citation index by assigning three different weights to citations and gives the qualitative citation index to the articles.

## Citation index using solution for MaxFS problem: a model proposed for future directions

In this section we propose a new approach for computation of citation indices based on the solution of maximum feasible sub-system problem. The important aspect of this approach is: it can easily incorporate domain-expert's opinion regarding relative quality of pairs of research articles and update the scores of nodes in the network suitably, so as to reflect the expert-judgment, as far as possible. All the existing PageRank based algorithms for computing citation indices do not involve any kind of human-expert intervention or



supervised learning. We believe that certain amount of human-expert intervention would be helpful (in fact needed) to improve the process of ranking the articles. Note that, in case of citation networks we want to arrange the articles based on their quality; unlike in case of link-networks where the popularity of a page is the main focus. Determining absolute quality of an article in fully automated manner without any kind of supervised learning a priori is extremely difficult task. So, we believe that human-expert intervention would be helpful in the initial stage of the process.

Let  $\alpha_i$  be the score associated with the  $i$ th source node for  $i = 1$  to  $k$ . In the previous section we express score associated with a node as a linear form in the scores of nodes from which the node has incoming edges [using formula (18)]. By visiting all the nodes in topologically sorted order, we update these linear forms and express score of each node as a linear form in the scores of source nodes. So every node  $v$  in the network is associated with a linear form  $L_v = a_{v,1}\alpha_1 + a_{v,2}\alpha_2 + \dots + a_{v,k}\alpha_k$  for rational numbers  $a_{v,j}$ 's.

There are several plausible ways in which one can get domain-expert's opinion regarding the quality of the article. Expert can be asked to rate the article on scale of 1–10 based on the quality of the article. This method or some variant of it is usually followed by various journals to decide their accepted articles. Another possible way, which is more suitable for our approach is: Expert can be asked to compare a pair of articles ( $u, v$ ) from the same domain and decide which article is better in terms of the quality. Suppose article  $u$  is better than  $v$  as per the expert's judgment, then we model this by linear inequality  $L_u > L_v$ . Comparing relative quality of two research articles would also be easier for experts rather than giving absolute score on scale of 1–10.

We get several randomly chosen pairs of articles ( $u, v$ ) evaluated from some experts and introduce appropriate linear inequalities corresponding to experts' judgments. Experts' judgment is typically directly helpful in improving the system when one of the articles in the pair corresponds to a source node in the citation network, as we cannot draw any inference about quality of source nodes solely from the structure of the network. In our approach we are expressing scores of all other nodes as linear forms in the scores of source nodes. If we have incorrect relative scores for some source nodes it is going to affect scores of all other nodes in the network.

We additionally add constraints  $\alpha_i > 0$  for  $i = 1$  to  $k$ , where  $k$  is the number of source nodes in the network. So ultimately we have system of linear inequalities in the variables  $\alpha_i$  for  $i = 1$  to  $k$ . Note that this system of inequalities can be inconsistent. Our goal is to find  $(\alpha_1, \alpha_2, \dots, \alpha_k)$  such that every  $\alpha_i > 0$  and out of the remaining inequalities we want to maximize number of satisfied inequalities. This is a variant of the Maximum Feasible Subsystem (MaxFS) Problem (Amaldi and Kann 1995). The Maximum Feasible Subsystem of linear inequalities has been extensively studied in the literature. Different variants of this combinatorial problem occur in various fields such as pattern recognition (Duda and Hart 1973; Warmack and Gonzalez 1973), operations research (Greenberg and Murphy 1991) and artificial neural networks (Marchand and Golea 1993; Hoffgen et al. 1995). Whenever a system of linear equations or inequalities is consistent, it can be solved in polynomial time using an appropriate linear programming method (Renegar 1988). In general, if the system is inconsistent, standard algorithms provide solutions that minimize the least mean squared error. But, such solutions, which are appropriate in linear regression, are not satisfactory in the context of our approach, where the objective is to maximize the number of relations that can be simultaneously satisfied. The general MaxFS problem is NP-hard, in fact even hard to approximate within polynomial factor (Amaldi and Kann 1995) but there are several good heuristic algorithms known for the problem (Chinneck 2001) which we can use to find suitable scores for the source nodes (maximizing number of satisfied inequalities), which in turn can be used to compute scores of all the nodes in the network.

Our future work includes the implementation of this proposed approach based on solution of Maximum Feasible Subsystem problem which can incorporate expert's judgments by introducing suitable linear constraints. We plan to use suitable heuristic for Maximum Feasible Subsystem Problem to ultimately compute ranking of research articles in the citation network.

## Conclusion

Use of citation count for measuring the impact of individual research articles is not an apt method to compute citation indices. In the existing literature PageRank based algorithms are used to compute citation indices of individual articles, which take into account the importance of the citing article. We take a step further in this direction and introduce three context based quality factors: first, sentiment analysis of the text surrounding the citation in the citing article; second, use of weighing scheme to provide lower weights for self-citations; and third, higher weights given to semantic similarity between citing and cited article. These factors we take into account while weighing the edges and further scoring the articles.

By exploiting the fact that the directed graph underlying a citation network is acyclic, we define a simple graph-traversal based single pass algorithm, which visits nodes in the topologically sorted order and computes their scores. Our algorithm is more efficient in terms of time-complexity compared to prior PageRank based algorithms showing 70 % improvement in the efficiency. From the experimental results, we see that our algorithm gives similar rankings as that of iterative PageRank algorithm, but better rankings (in terms of quality) than the traditional citation count based approach.

## References

- Amaldi, E., & Kann, V. (1995). The complexity and approximability of finding maximum feasible subsystems of linear relations. *Theoretical Computer Science*, 147(1), 181–210.
- Amin, M., & Mabe, M. A. (2003). Impact factors: Use and abuse. *Medicina (Buenos Aires)*, 63(4), 347–354.
- Athar, A., & Teufel, S. (2012). Detection of implicit citations for sentiment detection. In *Proceedings of the Workshop on Detecting Structure in Scholarly Discourse* (pp. 18–26). Association for Computational Linguistics.
- Casal, G. B. (2004). Assessing the quality of articles and scientific journals: Proposal for weighted impact factor and a quality index. *Psychology in Spain*, 8, 60–76.
- Cavalcanti, D. C., Prudêncio, R. B. C., Pradhan, S. S., Shah, J. Y., & Pietrobon, S. (2011). Good to be bad? Distinguishing between positive and negative citations in scientific impact. In *Proceedings of 23rd IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*, 2011 (pp. 156–162).
- Chinneck, J. W. (2001). Fast heuristics for the maximum feasible subsystem problem. *INFORMS Journal on Computing*, 13(3), 210–223.
- Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis* (Vol. 3). New York: Wiley.
- Esuli, A., & Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC* (Vol. 6, pp. 417–422).
- Eugene, G. (1955). Citation indexes for science: A new dimension in documentation through association of ideas. *Science*, 122(3159), 108–111.
- Garfield, E. (1972). *Citation analysis as a tool in journal evaluation*. American Association for the Advancement of Science.
- Greenberg, H. J., & Murphy, F. H. (1991). Approaches to diagnosing infeasible linear programs. *ORSA Journal on Computing*, 3(3), 253–261.

- Haddadene, H. A., Harik, H., & Salhi, S. (2012). On the PageRank algorithm for the articles ranking. In *Proceedings of the World Congress on Engineering* (Vol. 1, pp. 4–6).
- Harter, S. P., Nisonger, T. E., & Weng, A. (1993). Semantic relationships between cited and citing articles in library and information science journals. *Journal of the American Society for Information Science*, 44(9), 543–552.
- Hoffgen, K. U., Simon, H. U., & Vanhorn, K. S. (1995). Robust trainability of single neurons. *Journal of Computer and System Sciences*, 50(1), 114–125.
- Langville, A. N., & Meyer, C. D. (2011). *Google's PageRank and beyond: The science of search engine rankings*. Princeton: Princeton University Press.
- Leacock, C., & Chodorow, M. (1998). Combining local context and WordNet similarity for word sense identification. *WordNet: An Electronic Lexical Database*, 49(2), 265–283.
- Li, J., & Willett, P. (2009). ArticleRank: A PageRank-based alternative to numbers of citations for analysing citation networks. In *Aslib Proceedings* (Vol. 61, No. 6, pp. 605–618). Emerald Group Publishing Limited.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1–167.
- Ma, N., Guan, J., & Zhao, Y. (2008). Bringing PageRank to the citation analysis. *Information Processing and Management*, 44(2), 800–810.
- Marchand, M., & Golea, M. (1993). An approximation algorithm to find the largest linearly separable subset of training examples. In *IEEE World Congress on Neural Networks* (Vol. 3).
- Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM*, 38(11), 39–41.
- Nicolaisen, J. (2007). Citation analysis. *Annual Review of Information Science and Technology*, 41(1), 609–641.
- Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the Web*. Stanford InfoLab: Technical Report.
- 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).
- Qiao, H., Wang, Y., & Liang, Y. (2012). A value evaluation method for papers based on improved PageRank algorithm. In *Computer Science and Network Technology (ICCSNT), 2012 2nd International Conference on* (pp. 2201–2205). IEEE.
- Renegar, J. (1988). A polynomial-time algorithm, based on Newton's method, for linear programming. *Mathematical Programming*, 40(1–3), 59–93.
- Saha, S., Saint, S., & Christakis, D. A. (2003). Impact factor: A valid measure of journal quality? *Journal of the Medical Library Association*, 91(1), 42.
- Sayyadi, H., & Getoor, L. (2009). FutureRank: Ranking scientific articles by predicting their future PageRank. In *SDM* (pp. 533–544).
- Sidorov, G., Gelbukh, A., Gómez-Adorno, H., & Pinto, D. (2014). Soft similarity and soft cosine measure: Similarity of features in vector space model. *Computación y Sistemas*, 18(3), 491–504.
- Singh, A., Shubhankar, K., & Pudi, V. (2011). An efficient algorithm for ranking research papers based on citation network. In *Data Mining and Optimization (DMO), 2011 3rd Conference on* (pp. 88–95). IEEE.
- Spearman, C. (1904). The proof and measurement of association between two things. *The American journal of psychology*, 15(1), 72–101.
- Stamou, S., Mpouloumpasis, N., & Kozanidis, L. (2009). Deriving the impact of scientific publications by mining citation opinion terms. *JDIM*, 7(5), 283–289.
- Su, C., Pan, Y., Zhen, Y., Ma, Z., Yuan, J., Guo, H., Yu, Z., Ma, C., & Wu, Y. (2011). PrestigeRank: A new evaluation method for papers and journals. *Journal of Informetrics*, 5(1), 1–13.
- Toutanova, K., Klein, D., Manning, C. D., & Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology* (Vol. 1, pp. 173–180). Association for Computational Linguistics.
- Walker, D., Xie, H., Yan, K. K., & Maslov, S. (2007). Ranking scientific publications using a model of network traffic. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(06), P06010.
- Warmack, R. E., & Gonzalez, R. C. (1973). An algorithm for the optimal solution of linear inequalities and its application to pattern recognition. *Computers, IEEE Transactions on*, 100(12), 1065–1075.
- Yan, E., Ding, Y., & Sugimoto, C. R. (2011). P-Rank: An indicator measuring prestige in heterogeneous scholarly networks. *Journal of the American Society for Information Science and Technology*, 62(3), 467–477.