IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 16, 2019, accepted July 1, 2019, date of publication July 5, 2019, date of current version August 2, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2927016

Quantifying a Paper's Academic Impact by Distinguishing the Unequal Intensities and Contributions of Citations

MINGYANG WANG, JINJIN REN, SHI LI, AND GUANGSHENG CHEN
College of Information and Computer Engineering, Northeast Forestry University, Harbin 150040, China

Corresponding authors: Mingyang Wang (wmynefu@163.com) and Guangsheng Chen (kjc_chen@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 71473034, in part by the Heilongjiang Provincial Natural Science Foundation of China under Grant LH2019G001, and in part by the Financial Assistance from the Postdoctoral Scientific Research Developmental Fund of Heilongjiang Province under Grant LBH-Q16003.

ABSTRACT Citation count is a widely-used indicator for calculating the academic impact of scientific papers, but it is limited because it assumes all citations are of similar value and weights each equally. By examining the influence changes in papers' citation distribution and the cited papers' unequal contributions to the citing ones, this study aims to distinguish citations and, on this basis, evaluate the academic impact of the papers. Three indices of time-weighted citation count, citation width and citation depth are proposed to distinguish citations and perform the evaluation task. The experimental results show that papers exhibit different influence intensity characteristics in different periods of citation life. Those papers got larger citations in recent years are more influential and more active to gain new citations. The papers show the different scope of influence in their citing environment, although they were originally published in the same journal and the same year. In addition, the different frequency of mentions and the different subject similarities with the citing works suggest that the papers have different importance and usefulness for subsequent research. These results suggest that these three indices do help to distinguish citations and reveal the different intensity and contribution of influences in citations. Finally, the three indices are integrated into the overall evaluation of the academic impact of the paper, and the weight of each index is calculated by the entropy weight method. Quite different overall impacts in the paper are shown due to their different performances in the three indices, even though they have the same total number of citations.

INDEX TERMS Academic impact, time window, knowledge flow, mentioned frequency, entropy weight method.

I. INTRODUCTION

The academic research and policy agenda increasingly seeks to measure and use "impact" as a means of determining the value of different items of published research [1], [2]. Within academia, it is generally accepted that the number of times a paper is cited offers the most useful and easily measured guide to its impact [1]–[4]. The underlying assumption is that the cited work has influenced the citing work in some way. Furthermore, there are many citation-based metrics, such as the h-index family of indicators [5]–[14], the R-index [15], the ha-index [16], the IQP-index [17], the Eigenfactors scores [18]–[20], and many more, the w-index [21] and the q2-index [22].

The associate editor coordinating the review of this manuscript and approving it for publication was Eunil Park.

Inspired by the success of Google's ranking system for web pages, the popular algorithm PageRank as well as some of its variants have been used to show the prestige in citation networks of journals [23], [24], scientists [25]–[27], and publications [28]–[42]. The local and global structure of citation graph [27], [35], [37], [41], [42], the heterogeneity of the citation networks [32], the "virtual node" which considers the citations not included in the current systems [33], and the age of the publications [28], [31], [35] are taken into account in these researches.

However, in many of the above citation-based metrics, all the citations are treated equally; and they simply count the number of citations. However, using citation count alone to measure the impact of a paper is a limited approach. The citations in a paper are not equally important and some citations are more important than the others. Indeed, the idea that some citations are more important than others began to be explored long ago[43], which was later famously adapted by the Google search engine [44]. Nowadays, there are real life implementations of such indicators that do not only consider citation counts, but also the source of the citations. For instance, the SJR indicator [45], [46] and the Eigenfactor Metrics [18]–[20] have been implemented by Clarivate Analytics in its Journal Citation Reports. The search engine Semantic Scholar also makes a distinction for citing papers that have been strongly influenced by certain cited papers. Although considerations are given to weight citations in the network-based metrics, measuring the structure of citation networks is particularly complex because there is first the need to establish a citation network in order to establish its structural properties. And, the science is a cumulative effort where every contribution is published at a certain time, the resulting citation network is a dynamic and open-ended process. More specifically, citations happen at a given time, when a certain citation intensity exists among papers, which could be different at another point in time. Yet the most established evaluating systems for papers are static methods that disregard the dynamic nature of the citation process [47]. Furthermore, many studies [48]–[50] stated that the academic impact of a paper is the extent to which it is useful to other researchers. So, when the citation occurs and how it has influenced the following research are relevant factors to consider when exploring its potential impact.

The aim of this study was to evaluate papers' academic impact by distinguishing the citations: to test the feasibility of creating a "citation profile" for a single paper by exploring and quantifying its academic impact based on when the paper is cited and how the paper was being used, rather than simply whether it was cited. First, as scientific knowledge and contributions are dynamic and quickly changing, it is important to acknowledge nuanced factors of an article's influence, including its changing impact over time. An analysis of the temporal evolution of citations can discriminate between similarly rated papers that in fact have quite different citation histories. In other words, understanding how a paper's citation intensity evolves could help differentiate the impact of papers even when they have roughly the same number of citations [51]. Second, the number of citations cannot show how a paper influences the citing work. From the perspective of knowledge flow, the citation behavior of the paper can reflect the diffusion properties of its academic influence in its citing environment. When one citation activity occurs, the knowledge is diffused from the cited paper to the citing one. An earlier work by the authors found that the wider citation distribution of a paper in various journals, subjects, countries, and institutions had a greater influence on the paper's citation impact [52], [53]. Such a citation distribution of one paper in its citing environment reflects the scope of the knowledge diffusion of this paper, indicating the range of its contribution on the scientific environment in some extent. Finally, the extent to which a paper is useful to another scientific work is also an indicator of the influences of the cited work on the citing ones. The number of citations cannot tell us how a piece of research was used in practice, only that it is known and cited. The frequency an article is cited within a publication has been used to demonstrate its usefulness [2], [51], [54]. This consideration on the frequency of citation therefore offers an easily measurable, potentially useful metric for gauging the importance of the cited work to the citing ones. Recently, an unsupervised framework, Doc2vec, that learns continuous distributed vector representations for pieces of texts has achieved great success in various areas due to its strong ability of feature learning and modeling [55]. The great power for Doc2vec in capturing the semantics of phrases and sentences makes it possible and easy to detect the information on how a reference is useful to the citing work in the semantic level. By realizing the distributed representation of the cited and the citing works using Doc2vec framework, this paper calculated the subject similarities between the cited and the citing works to detect the usefulness of the cited papers.

Based on the above discussion, this paper proposed three indices of: 1) the time-weighted citation count, 2) the citation width, and 3) the citation depth, to distinguish the unequal intensities and contributions in citations. In order to make a universal evaluation on the papers' academic impact, the entropy weight method was introduced to determine weights for each index; then a holistic quantified impact was calculated by integrating these indices to realize the evaluation task on papers' academic impact.

The remainder of this paper is organized as follows. Section II discusses the related work. Section III describes the detailed process for measuring a paper's academic impact using the three indices. Section IV shows the data used in the experiments. Section V presents the experimental results. Section VI gives the overall discussion of the present work.

II. RELATED WORK

Evaluation on the academic impact of papers has been studied for a long time and became a popular and significant research direction [56], [57]. Generally, there are two major kinds of methods for examining the academic impact of scholars or publications, i.e., citation-based methods and network-based methods.

A. CITATION-BASED METHODS

The citation count is the most widely used indicator to gauge the influence of scholars and articles in citation-based method. Based on citation counts, a series of metrics were put forward to measure the impact of scientific entities [27]. Initially, the journal impact factor is proposed for evaluating the quality of journal [58]. Continually, the h-index [5] is proposed to measure scholar's impact by considering the productivity and the quality of their research work. Moreover, the AIF was proposed to depict the dynamics of scholars' impact by considering the ever-increasing characteristic of h-index [59]. However, there exist critical shortcomings of using citation counts to evaluate the impact of scientific entities.

The first problem is that citation counts aggregate with time. Articles published for a long period have the advantage of occupying more time for citations than newly published articles. Similarly, using the same time interval to evaluate the academic impact is unfair for young researchers comparing to senior researchers. Considering the above facts, researchers proposed several methods to alleviate the effects of publishing time. Several age-based citation growth models were proposed along with measures such as ''preferential attachment'', ''fitness'', ''quality'' and ''perceived novelty'' [60]–[63]. The effect of citations, time information, and the combination of PageRank and HITS algorithm were explored to quantify the scientific impact of scholars [61]. Based on information distribution, a scheme was designed to calculate the citation weight of each journal and conference in different years and used the weighting cited credits of every journal and conference to classify publications [64]. An earlier work by the authors found that the citations obtained in the recent two years are best to forecast articles' future referential ability by detecting the time characteristics in the articles' citation distributions [65]. These works show that the citations obtained in different time periods have diverse influences, which can't be treated equally. In the present work, the evolving properties of citations were investigated and, on this basis, the intensity characteristics of influence in different time periods were quantified and weighted to be the first impact index when examining papers' universal academic impact.

In fact, treating all citations equal and assigning the same values to all the citations is the second problem existing in most of citation-based indicators. Researchers realized the unequal value of different citations and performed various works to help measure and distinguish citations [66]. The idea that some citations are more important than others began to be explored long ago[43], [67]. Researchers addressed the issue of treating all citations equally by analyzing citation contribution based on its location within the introduction, methodology, discussion, or conclusion sections of a citing article. They concluded that the contribution of a citation should be based on both its frequency and its location within the citing article [43], [67]. Then the citation relevance was similarly categorized based on how citations were mentioned in the citing article [68], and frequency was used to calculate a citation's contribution to the citing article [69]. Some researchers explored the citations sourced from prestigious affiliations and measured the academic impact of the articles based on these prestigious citations [70]. The polarity of citations, i.e. positive and negative citations, were taken into account when evaluating the impact of scientific outputs [71]. The ways of self-citation, including direct, co-author, collaborative and coercive induced self-citation, were discussed to distinguish the citation [72]. Practices in identifying in anomalous citations were also considered in evaluating the academic impact of paper [61]. Researchers also proposed that a citation's value should be addressed by interpreting each one based on its context at both the syntactic (e.g., where the reference has been mentioned in a citing article) and semantic levels (e.g., why the reference has been cited in a citing article) [66]. Subsequently, the significance of citations was determined based on their appearing sections [73]. Researchers stated that the number of citations that a publication is cited in the other publications can represent the formal quality of the publication [54]. They examined about how many times each reference was really mentioned within the citing publications to evaluate the impact of references. In addition to distinguishing citations based on location and mention frequencies, the machine learning or natural language processing method were also used to calculate the semantic similarities between the cited and the citing works. The ''Jaccardized Czekanowski index'' was proposed to measure the similarity between the cited and citing journal list for a given journal [74]. The InText citation-based and vector space-based measures was used to calculate the similarity of research papers [75]. An ontology-driven knowledge black summarization approach was proposed to compute documents similarities [76]. A PBC (Passage-based Bibliographic Coupling) technique was designed to estimate inter-article similarity with information collected from context passages around important out-link citations in each article [77]. The performance of various methodologies of co-word analysis, Subject-Action-Object (SAO) structures, bibliographic coupling, co-citation analysis, and self-citation links were compared in describing semantic similarities [78]. Researchers found that the content-based methodologies such as co-word analysis could be better in extracting semantic information than those citation-based methodologies. But they also stated that the semantic analyses encounter substantial difficulties in comparing the content of documents because a simple co-word analysis is only able to recognize a superficial level of similarity since it is limited to the exact type of words used. Recently, an unsupervised framework of Doc2vec was proposed to easily realize the distributed expressions of words, sentences, paragraphs and even entire texts [55], and is widely used in various semantic analysis applications such as sentiment analysis [79], syntactic parsing [80], text summarization [81] and many other tasks [82]. In the present work, both the mention frequencies of one paper within its citing publications, and also the subject similarities between the paper and its citing publications were investigated to describe the impact of this paper to the following research. And this investigation would constitute the third index, citation depth, when we made a universal evaluation on papers' academic impact. And, the unsupervised framework of Doc2vec was introduced in this paper to realize the vectorization of the documents to facilitate the calculation of subject similarity.

Furthermore, in the authors' previous work on extracting typical features for predicting articles' future citation impact, we found that the features describing the scope of the citations diffused in its citing environment, especially the four dimension features of {countries, subjects, journals, institutions}, are the vital predictive indicators [52], [53]. In fact, the citation diffusion properties indicate the size of the scope of cited article's influence in the citing environment. This influence reveals the width of the impact of the article, which in turn provides valuable visibility for the article, laying an important foundation for its higher academic impact in future. Therefore, we believe that the size of the scope of cited article's influences could also be one of important aspects to distinguish articles' citations, which constitutes the second index, citation width, when we examined the universal academic impact of papers.

B. NETWORK-BASED METHODS

Apart from the citation-based methods, researchers also utilize the academic networks to measure the scientific impact. The PageRank [44] and HITS algorithms [40] are the most commonly used ones to rank the importance of scholarly entities in academic networks. The PageRank algorithm was intended to rank webpages by importance using the link structure of the web, and this recursive technique quickly gained popularity and found numerous other applications. The citation networks of research papers were particularly well suited for the usage of methods based on PageRank because they could be easily modelled as directed graphs [28], [30], [83], [84]. And the two most eminent academic databases have taken PageRank-based metrices to assess journal's impact [81]. They are: Web of Science as Eigenfactor Score [18]–[20] and Scopus as SJR indicator [45], [46]. Considering the distinct importance of different entities and relationships in academic networks, researchers have proposed a number of weighting schemes, together with the variants of PageRank or HITS algorithm, to evaluate the scientific impact of articles [42], [85]–[88], journals [24], and scientists [89]–[91]. In addition, researchers also combine the citation and network-based evaluation metrics to measure the impact of scholars. A revised application of PageRank was presented via the differentiation of citation types to facilitate a multidimensional evaluation of 39 selected management journals [92]. A weighted PageRank algorithm was proposed by considering citation and co-authorship network topology to measure the author's impact [93].

In the present work, we mainly perform to evaluate the academic impact of papers by distinguishing citations. The evolving intensity characteristics of influence, the diverse citation diffusion properties and the different usefulness of cited papers were detected and integrated to evaluate papers' universal academic impact. The network-based analysis was not involved in the present work.

III. METHODOLOGY

A. THREE INDICES FOR IMPACT ASSESSMENT

1) THE TIME-WEIGHTED CITATION COUNT

The number of citations a paper gets is a cumulative effort where every contribution is achieved at a certain time, and the resulting citation impact is a dynamic and open-ended index. Papers would have different intensities of their influence at different times, and, as a result, when the citations occurred would be important to distinguish citations and to evaluate papers. We examined the correlation between a paper's past citations and its future citations, with the aim of investigating the evolving influence intensities in different time periods.

Taking a year as the unit, two series of time windows were generated from the reference year to the papers' publication year to perform this task. Because the citations for all papers were counted up to December 31, 2016, the year 2016 is regarded as the reference year in this paper.

a) Independent time windows: A time parameter t is used to represent the different time window: $t = 0$ denotes the reference year 2016, $t = 1$ denotes 2015, and so on. The rest are calculated in the same manner.

b) Continuously increasing time windows: Time parameter T is used to represent the size of the continuously increasing time window. $T = 1$ denotes 2015; the previous year, $T = 2$ denotes the most recent two years: years $\{2015, 2014\}$. The rest are calculated in the same manner.

Using these formulas, we collected the citations obtained in each time window. In order to establish how a paper's past citations influence on the acquisition of new citations, its citation distribution data was divided into two periods. 1) The time from each paper's publication year to 2015, and the citation data collected in this period was used to model the paper's dynamic past citations. 2) The 2016 citation data for these papers. This data set was used to determine the paper's future citation performance, which is the capacity for it to be cited again in the near future. The Pearson's correlation coefficient was used to calculate the dependence of papers' future citation possibilities on their past citations.

Different correlation coefficient would be got in different time windows. And a higher correlation in one time window means that a higher contribution of papers' citations obtained in this time period on their future citation activities. As a result, the citations obtained in this time window would be more active and have larger influence intensity. The assessment of the overall impact of one paper needs to take account of these different intensities. From this perspective, the correlation coefficient calculated in different independent windows were used to weight the citations obtained in the same time period. By accumulating the weighted past citations, we get the time-weighted citation count, which can better reveal the influence evolution characteristics in the citation distribution compared with the total citation counts.

2) THE CITATION WIDTH

The citation relations among papers relates to the knowledge flow among them. When one citation activity occurs, the knowledge would be propagated from the cited article to the citing one. By considering the journals publishing the articles, the language and the document type of the articles, the subject category of the articles and the affiliation attributes of the authors, every citation activity would then be treated as a single knowledge diffusion process from one space (subject, country, institution, journal, language, document type, etc.) in the cited environment to the other space (subject, country, institution, journal, language, document type, etc.) in the citing environment. For the cited papers, the citation diffusion distribution in this feature space could indicate the size of the scope of their influences in the citing environment. And as a result, a larger size of the scope of one paper's influence should be one indicator representing a wider diffusion character of its impact. In the authors' previous work, we detected the articles' knowledge diffusion properties and concluded that the wider citation distribution in the four dimensions of {subjects, journals, countries, and institutions} had greater influences on increasing an article's citation impact [52], [53]. Here, we constructed the same feature space of F = {Subject, Journal, Country, Institution} to describe the influence diffusion characteristics of papers.

In order to make a quantified examination of the citation width of each paper basing on feature space F, the theory of mutual information was introduced to generate weights for each dimension in F by calculating the dependency of the paper's total citation count on each of the feature dimensions. Mutual information is a statistical measure of interactions between variables and can access linearly and nonlinearly their dependency [94], [95]. The mutual information between X and Y is defined by the following equation:

$$I(X;Y) = \sum_{\mathrm{Y}\in R}\sum_{\mathrm{X}\in S} p(X,Y)\log_2 p(X,Y)/p(X)p(Y) \quad (1)$$

where X is the citations from one dimension in feature space F; Y is the total citations of this paper. In order to facilitate the calculation process, the papers were divided into three categories, highly-cited, medium-cited and low-cited, according to their total citations. Detailed information on this division process is shown in Section IV. Therefore, Y actually represents the three categories of papers. $p(X)$ and $p(Y)$ are probability density functions, and $p(X, Y)$ represents the joint probability. The mutual information can determine the reduction of uncertainty of the values of Y once X is known. Mutual information is a nonnegative concept, with $0 \leq I(X; Y) \leq 1$, the value $I = 1$ means the highest dependency, and a value close to zero denotes no inter-correlation.

The dependency provides important information for understanding the contributions of articles' citation diffusions to their total citations, which is the weight for each feature dimension. Then the citations from each feature dimension were weighted and accumulated to evaluate papers' influence width.

3) THE CITATION DEPTH

The citation depth gives one evaluation on how the paper is useful for the subsequent research. As discussed in the citation-based methods in Section II, counting the mention frequencies and calculating the subject similarities could both benefit to measure the contribution of one paper to the citing works. This paper investigated the mention frequencies, that is, how many times the paper was cited within each citing publication to examine the contribution in the syntactic level. And a Doc2vec derived subject similarities between cited and citing works were calculated to evaluate the contribution of the cited work in the semantic level.

A scientific paper usually cites tens or hundreds of reference articles. It is time-consuming to count the citation frequencies from all of the citing publications for one paper, so we focused on the high-quality citing publications. Based on the "Create Citation Report" tool on the webpage for citing articles, it is easy to find the h-index of citing articles as a whole. For one paper, all the citing articles that were cited at least the same number of times as the h-index were collected, which were labeled as HQCPs (the abbreviation for high-quality citing publications). These collected citing articles were regarded as higher quality, and were used to gather the mention frequencies and to calculate the subject similarities with the cited paper.

For the task of calculating subject similarity, we proposed to use the Doc2vec method, which builds a distributed vector representation at the document level using an unsupervised approach [94]. Suppose that there are N papers in the corpus comprising all the cited papers and all the HQCPs citing to them, and we want to learn the distributed document vector such that each paper is mapped to a fixed dimension. There are two models of Distributed Memory Model of Paragraph Vectors (PV-DM) and Distributed Bag of Words version of Paragraph Vector (PV-DBOW) in Doc2vec method [94]. In our experiment, each document vector is a combination of these two vectors: one learned by the paragraph vector with PV-DM and one learned by the PV-DBOW, which is the same with that in [96]. And the learned document vector representations have 50 dimensions both in PV-DM and in PV-DBOW, which means that each paper is mapped to a distribution vector with 100 dimensions.

Suppose p_i and q_j denote the document vector representations of the ith paper and the jth article in the HQCPs of p_i. We calculated the subject similarity SS_{ij} between p_i and q_j with a cosine similarity:

$$SS_{ij} = (p_i \cdot q_j)/(||p_i||||q_j||) \quad (2)$$

Then, both the mention frequencies and the subject similarities were incorporated to obtain a unitary evaluation on citation depth for one paper. And the value of citation depth for paper p_i is the accumulation of the total usefulness of p_i in its HQCPs:

$$CD_i = \sum\nolimits_{j=1}^{n} (MF_{ij} * SS_{ij}) \quad (3)$$

where n is the number of the HQCPs for paper p_i; MF_{ij} denotes the mention frequencies for p_i within q_j; and SS_{ij} is the subject similarity between p_i and q_j. Obviously, a higher mention frequency and a larger subject similarity can both benefit to generate a larger contribution or usefulness, which would characterize a higher supporting extent of one paper for the follow-up research.

B. EVALUATING PAPERS' ACADEMIC IMPACT WITH THE ENTROPY WEIGHT METHOD

After quantifying the three indices, each paper was transformed into a vector with three dimensions, with each dimension representing one of the three indices. Figure 1 shows a sketch of the paper's vector matrix, where p_i denotes the ith paper ($i = 1, 2, \ldots n$); d_j denotes the jth index ($j = 1, 2, 3$); x_{ij} denotes the value of paper p_i on index d_j. Then, the entropy weight method was introduced to aggregate the three indices to make a holistic evaluation of the paper's academic impact. The entropy weight method determines weights through quantifying the disorder extent of a particular system [97], [98]. Because the weighting factors are purely dependent on the value of indices rather than human subjective assessment, this was recognized as an objective method for weight calculation.

$$
\begin{array}{c c}
 & \begin{array}{ccc} \mathbf{d_1} & \mathbf{d_2} & \mathbf{d_3} \end{array} \\
\begin{array}{c} \mathbf{p_1} \\ \vdots \\ \mathbf{p_i} \\ \vdots \\ \mathbf{p_n} \end{array} &
\left(\begin{array}{ccc} \mathbf{x_{11}} & \mathbf{x_{12}} & \mathbf{x_{13}} \\ \vdots & \vdots & \vdots \\ \mathbf{x_{i1}} & \mathbf{x_{i2}} & \mathbf{x_{i3}} \\ \vdots & \vdots & \vdots \\ \mathbf{x_{n1}} & \mathbf{x_{n2}} & \mathbf{x_{n3}} \end{array} \right)
\end{array}
$$

FIGURE 1. **The paper's vector matrix.**

The main steps for evaluation of a paper's academic impact with the entropy weight method are as follows:

Step 1: Normalization of the matrix. In order to ensure the uniformity of indices' units or value range, the normalization of all indices is performed as:

$$y_{ij} = \frac{x_{ij} - \min_j(x_{ij})}{\max_j(x_{ij}) - \min_j(x_{ij})}(\max_{new} - \min_{new}) + \min_{new} \tag{4}$$

where $[\min_{new}, \max_{new}]$ is the new value range for all the indices, which is set as $[\min_{new}, \max_{new}] = [0.001, 0.998]$. After this step, the normalized vector matrix $Y = (y_{ij})_{n\times 3}$ is obtained.

Step 2: Calculation of weighting coefficients. The information entropy of each index is calculated by:

$$E_j = -(\ln n)^{-1} \sum\nolimits_{i=1}^{n} p_{ij} \ln p_{ij} \tag{5}$$

where $E_j$is the information entropy of each index, p_{ij} can be calculated by $p_{ij} = {y_{ij}}/{\sum_{i=1}^{n} y_{ij}}$.

Entropy is a measure of the uncertainty of the system. A larger entropy means that the system is in the status with largest uncertainty. For example, if all papers have the same value in one indicator, the information entropy calculated by this indicator is the largest. In this status, the indicator is the most uncertain and can't provide any information for distinguishing papers. Vice versa, if the information entropy of one indicator is smaller, it means that there's smaller uncertainty so that this indicator could provide more useful information on differentiate these papers. Then, the index that can create a larger inconsistency degree among papers, in other words, which has a larger capacity to discriminate papers, would have a larger weighting coefficient. The idea of entropy-based weighting method has been widely used in the assessment task in various fields [99], [100].

Based on the value of information entropy E_j, the weighting factor of each index is calculated by:

$$w_j = \frac{1 - E_j}{\sum_{j=1}^{3}(1 - E_j)} = \frac{1 - E_j}{3 - \sum_{j=1}^{3} E_j} \tag{6}$$

where $\sum_{j=1}^{3} w_j = 1$ and $0 \leq w_j \leq 1$.

$1 - E_j$ indicates the inconsistency degree of each paper under the jth index from the theory of information entropy. In this way, the index with a smaller entropy, in other words, with a larger capacity to differentiate papers, would have a larger weighting coefficient.

Step 3: Calculation of the comprehensive evaluation value on a paper's academic impact:

$$V_i = \sum\nolimits_{j=1}^{3} w_j x_{ij} \tag{7}$$

Following the steps discussed above, a holistic assessment of a paper's academic impact can be achieved.

IV. DATA

Papers published in year 2003 in *Journal of Applied Physics (JAP)* on the web version of the Science Citation Index (SCI) were used as the data source for our experiments. We collected papers with the same number of total citations to make a clear comparison of their different academic impacts. The approach of selecting articles with the same number of total citation counts has been used in Yu et al.'s work [101]. In their work, four articles with the similar published time and the same total citations were analyzed to make difference in the features describing the citations. In order to verify the universality of our method, papers were chosen from different categories of highly-cited, medium-cited and low-cited, and their citation data were gathered from the paper's publication year till to 2016.

The method for dividing different categories of articles is similar to that applied in a study of highly cited papers by Norwegian authors [102]. That is:

a) Highly-cited papers (HCPs): a paper was considered highly cited if the number of citations it received was at least seven times the mean citation rate among all papers published in *Journal of Applied Physics* in 2003.

b) Medium-cited papers (MCPs): a paper was considered medium-cited if the number of citations received was in the range of 1-7 times the mean citation rate among all papers.

TABLE 1. The original information about the four HCPs.

ID	Name of the article	First author	Citation counts	h-index
HCP-1	Internal structure and expansion dynamics of laser ablation plumes into ambient gases	Harilal	291	29
HCP-2	Temperature dependence of the free-exciton transition energy in zinc oxide by photoluminescence excitation spectroscopy	Wang	290	36
HCP-3	Extraordinary magneto elasticity and lattice softening in bcc Fe-Ga alloys	Clark	289	26
HCP-4	Comparing organic to inorganic photovoltaic cells: Theory, experiment, and simulation	Gregg	289	64

TABLE 2. The original information about the eight MCPs.

ID	Name of the article	First author	Citation counts	h-index
MCP-1	Chemisorption on nickel nanoparticles of various shapes: Influence on magnetism	Cordente	73	24
MCP-2	Mathematical description of atomic layer deposition and its application to the nucleation and growth of HfO2 gate dielectric layers	Alam	73	20
MCP-3	Current limiting mechanisms in indium- tin- oxide / poly3-hexylthiophene/aluminum thin film devices	Chiguvare	73	17
MCP-4	Synthesis and aging effect of spherical magnetite (Fe3O4) nanoparticles for biosensor applications	Gee	73	25
MCP-5	Determination of the refractive indices of AlN, GaN, and AlxGa1-xN grown on (111)Si substrates	Antoine-Vincent	73	17
MCP-6	Nucleation, growth, and graphitization of diamond nanocrystals during chlorination of carbides	Welz	73	17
MCP-7	Calorimetric measurements of phase transformations in thin films of amorphous Te alloys used for optical data storage	Kalb	73	19
MCP-8	Characterization and modeling of the nitrogen passivation of interface traps in SiO2/4H-SiC	McDonald	73	22

c) Low-cited papers (LCPs): this included the rest of publications, in other words, those where the number of citations received was less than the mean citation rate.

It should be mentioned that the choice of multiple 7 has nothing to do with the subject of the journal, mainly to ensure that an appropriate number of articles are selected as highly-cited papers. According to this standard, only 41 papers published in 2003 in *JAP* were selected as highly-cited papers, accounting for 1% of the total number of papers published in that year, which is consistent with the criteria in *Web of Science* to select highly-cited papers. In *Web of Science*, papers received enough citations to place it in the top 1% in the same subject area and in the same publication year are classified as highly-cited papers.

Four HCPs, eight MCPs, and twenty-eight LCPs were finally selected as the experimental samples for impact evaluation. Tables 1 , 2, and 3 show the original information about these papers, as well as the h-index of citing articles for each paper. It is important to note that it was difficult to find a considerable number of HCPs with the same number of citations. Thus, Table 1 shows four HCPs with the similar citation counts.

V. EXPERIMENTAL RESULTS

A. EXPERIMENTAL RESULTS FOR THE TIME-WEIGHTED CITATION COUNT

Figure 2 shows the dependency of papers' future citation performance on their past citations obtained in different time windows. The subgraphs in Figure 2(a) and (b) show the sketch for dividing the independent and increasing time windows, respectively. The correlation in the independent time windows is continually decreasing, exponentially, from the nearest time window ($t = 1$) to the furthest one ($t = 13$) as shown in Figure 2(a). The past citations, obtained in different time periods, do generate different influences on the papers' future citation abilities. Obviously, recent citations have made larger contributions for paper to gain new citations. Figure 2(b) gives the results for determining the special time period in which the citations can be the most predictive on papers' future citation activity. The correlation reaches the peak when $T = 2$ and then decreases continuously in an exponential way. It shows that the citations obtained in the last two years contribute the most for predicting papers' future citations. The perfect linear dependence of papers' citations in 2016 on citations obtained in the last two years $\{2015, 2014\}$ in Figure 2(c) provides the direct proof for this strong correlation. All of these results provide an evidence that one paper will have different influence intensities in different time periods. The total citations of the paper is only the accumulation of citation behaviors from each of the time periods. Only by decomposing it and exploring the vitality characteristics of the paper in various time periods, can we reveal the change of the influence of the paper in its citation life.

In the present work, all the past citations were considered when evaluating the total influence intensity of papers.

TABLE 3. The original information about the twenty-eight LCPs.

ID	Name of the article	First author	Citation counts	*h*-index
LCP-1	Laser propagation and energy absorption by an argon spark	Bindhu	28	11
LCP-2	Near theoretical microwave loss in hot isostatic pressed (hipped) polycrystalline yttrium iron garnet	Nazaro	28	11
LCP-3	Role of carbon atoms in the remote plasma deposition of hydrogenated amorphous carbon	Benedikt	28	13
LCP-4	Physics of thermionic dispenser cathode aging	Longo	28	10
LCP-5	Contact angle-based predictive model for slip at the solid-liquid interface of a transverse-shear mode acoustic wave device	Ellis	28	12
LCP-6	Leakage current and relaxation characteristics of highly (111)-oriented lead calcium titanate thin films	Tang	28	8
LCP-7	Investigation of charging phenomena in silicon nanocrystal metal-oxide-semiconductor capacitors using ramp current-voltage measurements	Ioannou-Sougleridis	28	11
LCP-8	Persistent photoconductivity and optical quenching of photocurrent in GaN layers under dual excitation	Ursaki	28	8
LCP-9	Characterization of tellurium layers for back contact formation on close to technology treated CdTe surfaces	Kraft	28	8
LCP-10	Extension of the semiclassical theory of current-perpendicular-to-plane giant magneto resistance including spin flip to any multilayered magnetic structures	Strelkov	28	10
LCP-11	Band gap engineering of SiCN film grown by pulsed laser deposition	Park	28	9
LCP-12	Equation of state of polytetrafluoroethylene	Bourne	28	11
LCP-13	Origin and effect of impurity incorporation in plasma-enhanced ZrO2 deposition	Cho	28	13
LCP-14	X-ray studies of magnetic nanoparticle assemblies	Anders	28	18
LCP-15	Magnetic clusters in Nd1-xSrxMnO3 ($0.3 \leq x \leq 0.5$): An electron-spin resonance study	Angappane	28	8
LCP-16	Design of Curie point written magnetoresistance random access memory cells	Daughton	28	14
LCP-17	Magnetic properties of IV-VI compound GeTe based diluted magnetic semiconductors	Fukuma	28	7
LCP-18	Effect of Mo content on the structure stability of R-3(Fe,Co,Mo)(29)	Li	28	6
LCP-19	Current-induced switching in low resistance magnetic tunnel junctions	Liu	28	11
LCP-20	Vortex pinning at individual defects in magnetic nanodisks	Rahm	28	12
LCP-21	Effect of interfacial coupling on the magnetic ordering in ferro-antiferromagnetic bilayers	Tsai	28	12
LCP-22	Transparent ohmic contacts of oxidized Ru and Ir on p-type GaN	Jang	28	10
LCP-23	Growth of uniformly a-axis-oriented ferroelectric lanthanum-substituted bismuth titanate films on silicon substrates	Lee	28	9
LCP-24	Relation between microstructure and stress in titanium nitride films grown by plasma immersion ion implantation	Lim	28	13
LCP-25	Effects of substrate annealing and post-crystallization thermal treatments on the functional properties of preferentially oriented (Pb,Ca)TiO3 thin films	Poyato	28	9
LCP-26	Effect of implant temperature on defects created using high fluence of helium in silicon	David	28	9
LCP-27	Optical properties of nanocrystalline diamond films by prism coupling technique	Sharda	28	9
LCP-28	Ge/Si(100) islands: Growth dynamics versus growth rate	Capellini	28	11

The citations obtained in different independent time windows were weighted by the correlation coefficient calculated in the corresponding time windows. And the weighted past citations were accumulated to be as the time-weighted citation count index for the universal impact evaluation.

B. EXPERIMENTAL RESULTS FOR THE CITATION WIDTH

In order to give a clearly discussion on the performance of different papers on the citation diffusion characteristics, we detected the evolving of the numbers of citing subjects, citing journals, citing countries, and citing institutions in different time intervals. Figure 3 shows the results of this discussion. The time interval was divided by taking each biennial period from papers' publication year till to 2016. Obviously, papers show diverse scope of their influences in different diffusion dimensions and in different time intervals. Although they were published in the same journal and in the same year, some papers still achieved more citations from different subjects, journals, countries and institutions from their publication year till to the reference year 2016. The advantages of citation diffusion in different dimensions help these articles get more total citations and eventually grow into HCPs. The reason is that the diffusion characters of papers' influence may have brought valuable visibilities for these papers, which is ultimately directly or indirectly converted into the citations to them.

Figure 4 shows the distribution of mutual information in different citation diffusion dimensions. The time interval is divided in the same manner as in Figure 3. The value of the mutual information gives the dependence between the frequency of citations papers obtained in one time period

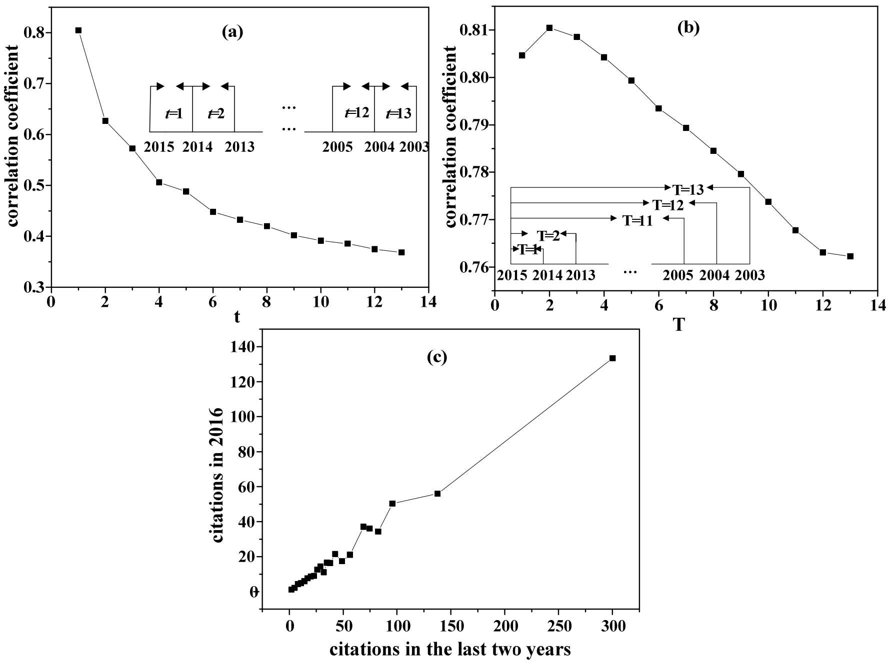

FIGURE 2. Correlation between papers' future citation counts and their past citations. (a) the correlation in the independent time windows. (b) the correlation in the continuous increasing time windows. (c) the dependence of the citation in 2016 (representing the future citation possibilities) on the citations obtained in the last two years.

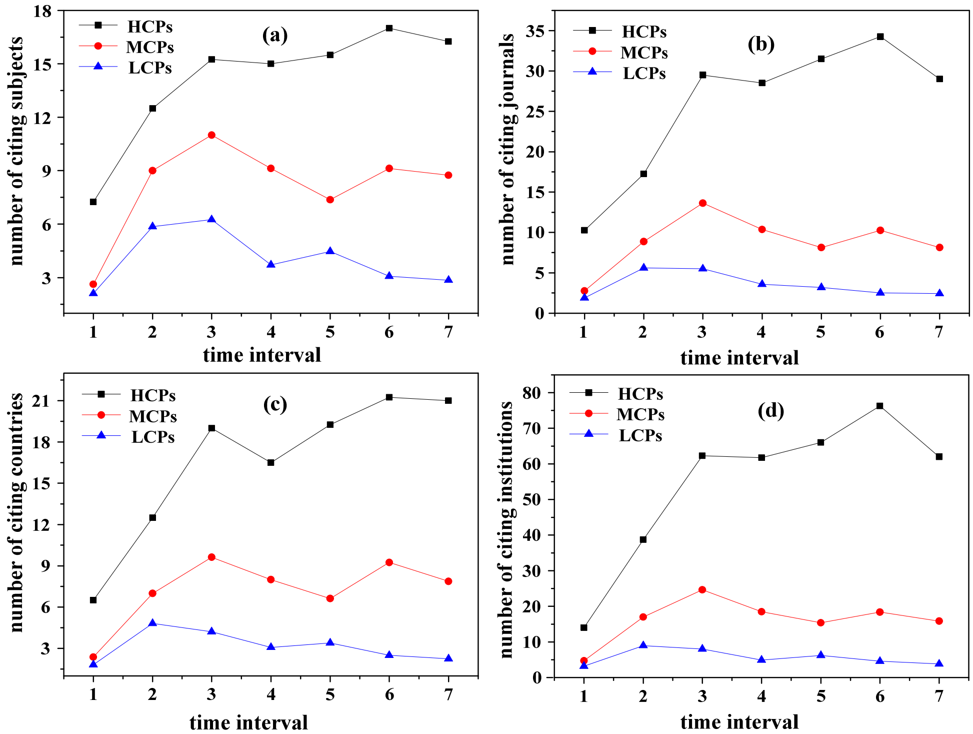

FIGURE 3. The distribution of the number of citing subjects (a), citing journals (b), citing countries (c), and citing institutions (d), in different time intervals.

and its citation diffusion dimensions in the same time period. Obviously, there is a strong dependency in almost all of the time intervals. The values of mutual information continually rise from the publication year, and reach a small peak in the third time intervals (5-6 years after publication). In the subsequent time intervals, the value of mutual information calculated in the four dimensions almost all exceeded 0.7. In some time intervals, the value of mutual information even exceeds 0.9. The results are largely consistent with the conclusion in the authors' previous work [52], [53]. We found that the citation diffusion performances on the four dimension features of {countries, subjects, journals, institutions} in the first five years are the kernel predictive indicators on papers' future citation impact. The results indicate that the citation diffusion characteristics, reflecting the range of the influence of one paper in its citation environment, has played great role in attracting citations for the paper. From the perspective of knowledge flow, the diverse range of influence characteristics of one paper indicates the degree of diffusion of the knowledge contained in the paper within different academic entities. This degree of diffusion provides a way to measure the usefulness or contribution of the paper. Thus, in the present work, the scope of papers' influence in their citing environment was taken as the citation width index in evaluating papers' academic impact.

For the papers with the same or similar total citations shown in Tables 1 , 2, and 3, we calculated the mutual information to describe the dependence of the papers' total citation counts on each diffusion dimensions. Then, the values of mutual information were normalized and assigned as weights to each dimension. The value of paper's citation width was quantified by accumulating the weighted citations from each dimension in feature space F.

Tables 4, 5, and 6 show the citation data in each dimension of feature space F, as well as the final quantified citation width for each paper. The different characteristics of citation diffusion in different dimensions have resulted in the diverse values of citation width of papers, showing the distinctive width characteristics of their total influence in the scientific environment. Obviously, HCPs achieved citations from the most countries, institutions, subjects, and journals compared with MCPs and LCPs. And papers with the same or similar total citations also exhibit diverse exhibition in the width characters of influence. A larger value on citation width for one paper indicates a wider influence of this paper on academic entities through the occurrence of citation activities, reflecting the larger contribution of this paper to the academic community in some extent.

C. EXPERIMENTAL RESULTS FOR THE CITATION DEPTH

Figure 5 shows the distribution of mentioned frequency of papers within their HQCPs. There is a significant power law distribution relationship between the mentioned frequencies and the number of citing papers for HCPs (a), MCPs (b) and LCPs (c). The result is consistent with that of [54]. Most of the cited papers were mentioned less than three times. But there are still a few papers be mentioned more times in the citing works. The largest mentioned frequency for HCPs, MCPs, and LCPs is eighteen, twelve, and seven, respectively. Some of the MCPs and LCPs, have gained more mentioned frequencies in several citing works than HCPs. It shows that there is a great difference between the number of mentioned frequencies and the number of total citations. Usually, the mentioned frequency of one paper in its citing works is better than its total citations to examine the usefulness of it, because the total citations is a simple accumulation of the number of citing works, regardless of how many times the cited paper were mentioned in the citing works. More mentioned frequencies is actually a represent for the higher contribution of one paper on the citing work. And these conclusions or phenomena could never be got from only counting the number of total citations.

When examining the citation depth for one paper, both the mentioned frequency of one cited paper within each of its HQCPs, and also the subject similarity of this paper with each of its HQCPs were counted and calculated, and finally integrated by using (3) to quantify the citation depth of this paper. The experimental results for the final citation depth for the HCPs, MCPs and LCPs can be found in the column "Citation depth" in Tables 7 , 8, and 9.

Obviously, papers have shown diverse depth properties even though they have the same or similar number of citations. In Table 7, HCP-4 shows the best depth property among all the four HCPs. HCP-4 has the largest number of HQCPs, the moderate subject similarities with its HQCPs, and also the largest mention frequency of eighteen times among all the HCPs. All these advantages help HCP-4 achieve the largest citation depth. For the similar reasons, MCP-7 and LCP-20 in Tables 8 and 9 show the largest citation depth among eight MCPs and twenty-eight LCPs. In the HQCPs of MCP-7, two citing articles mentioned MCP-7 for eight times and the other two citing articles mentioned MCP-7 for five times. MCP-7 also has moderate subject similarities with its HQCPs. LCP-20 gets the first rank in LCPs is also because its outstanding performance in mention frequencies. It is mentioned for five times in one citing publication and four times in the other three citing works.

D. EXPERIMENTAL RESULTS FOR EVALUATING PAPERS' ACADEMIC IMPACT WITH THE ENTROPY WEIGHT METHOD

Tables 7, 8, and 9 show the experimental results for the final integral academic impact for the three categories of papers. The entropy weight method was used to integrate the three indices to make a holistic evaluation on a paper's academic impact. The weights endowed to each dimension by the entropy weight method are also shown in Tables 7, 8, and 9. A larger weight means that the corresponding dimension has a larger capacity for differentiating papers. The dimension of citation depth is more useful to distinguish between the four HCPs; the three dimensions play a comparative role in differentiating MCPs; and the citation depth is more helpful to distinguish the LCPs. With these different "time-weighted citation count", "citation width", and "citation depth", papers have shown entirely diverse integrated academic impacts. Among the four HCPs in Table 7 , HCP-4 achieves the greatest academic impact, with its excellent performance in citation width and

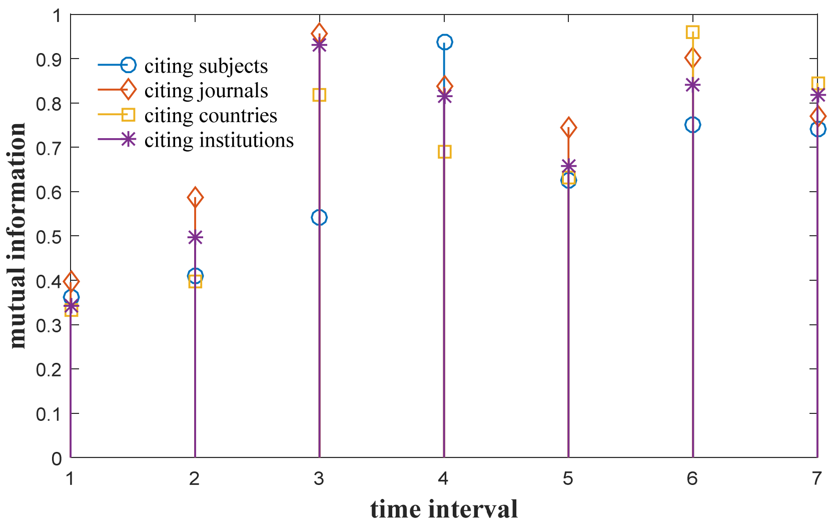

FIGURE 4. The distribution of mutual information in different citation diffusion dimensions.

TABLE 4. The citation width properties for the HCPs.

ID	Number of citing countries	Number of citing institutions	Number of citing subjects	Number of citing journals	Citation width
HCP-1	43	266	23	81	109.214
HCP-2	46	312	25	91	125.499
HCP-3	29	204	17	60	82.023
HCP-4	32	266	32	219	141.78

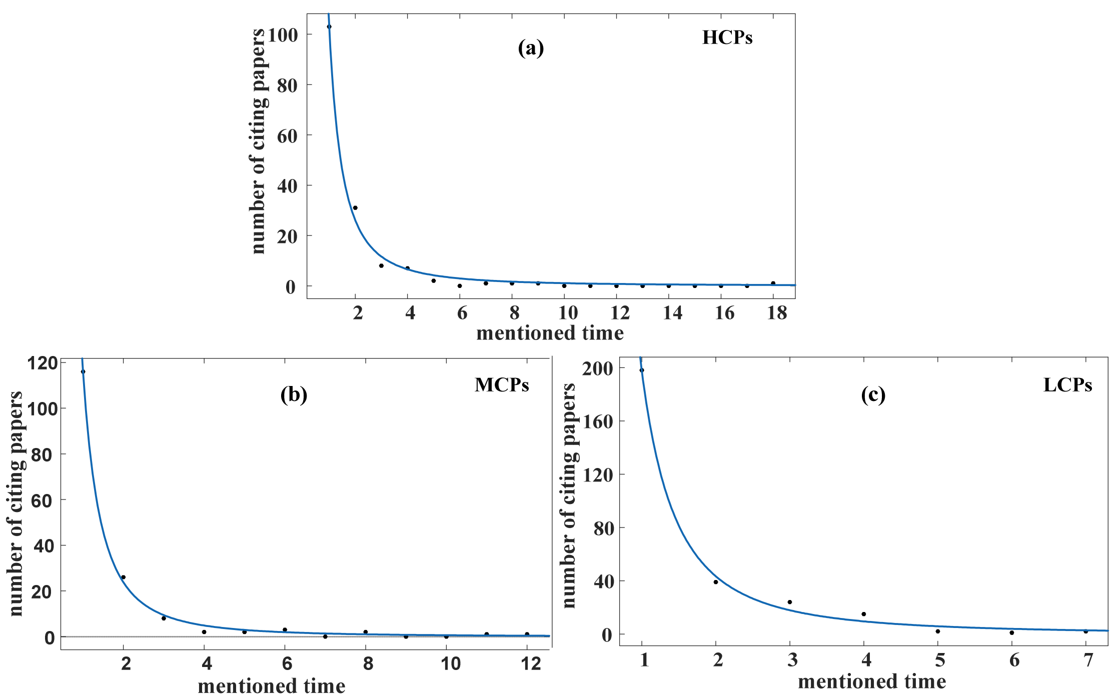

FIGURE 5. The distribution of mentioned frequencies within the HQCPs.

citation depth. Among the eight MCPs in Table 8, MCP-7 gets the greatest academic impact, with its considerable exhibition in citation width and citation depth. Among the twenty-eight LCPs in Table 9, LCP-20 has the greatest academic impact, with its perfect representation in all the three indices.

TABLE 5. The citation width properties for the MCPs.

ID	Number of citing countries	Number of citing institutions	Number of citing subjects	Number of citing journals	Citation width
MCP-1	21	98	25	49	49.786
MCP-2	18	73	20	27	35.699
MCP-3	29	99	17	41	48.607
MCP-4	22	102	21	46	49.583
MCP-5	18	73	10	30	34.331
MCP-6	29	87	26	49	49.159
MCP-7	20	94	18	30	42.311
MCP-8	20	78	13	29	36.632

TABLE 6. The citation width properties for the LCPs.

ID	Number of citing countries	Number of citing institutions	Number of citing subjects	Number of citing journals	Citation width
LCP-1	10	28	13	18	17.52
LCP-2	15	38	9	15	20.05
LCP-3	11	28	15	17	17.968
LCP-4	5	29	12	13	15.002
LCP-5	13	30	18	21	20.877
LCP-6	8	28	12	18	16.765
LCP-7	10	29	11	14	16.398
LCP-8	15	36	7	13	18.594
LCP-9	9	25	11	16	15.517
LCP-10	11	28	6	10	14.37
LCP-11	12	44	13	16	21.962
LCP-12	5	16	11	11	10.741
LCP-13	11	27	15	17	17.694
LCP-14	18	45	10	19	23.972
LCP-15	12	38	11	14	19.409
LCP-16	11	37	9	15	18.687
LCP-17	16	44	15	19	24.203
LCP-18	3	20	10	16	12.303
LCP-19	11	26	7	15	15.252
LCP-20	13	30	8	14	16.859
LCP-21	16	33	9	15	18.952
LCP-22	5	24	13	16	14.574
LCP-23	13	36	8	18	19.48
LCP-24	13	36	13	16	20.042
LCP-25	11	29	10	15	16.704
LCP-26	13	40	16	18	22.256
LCP-27	11	33	13	20	19.651
LCP-28	10	30	11	19	17.892

TABLE 7. The academic impact for the HCPs.

ID	Time-weighted citation counts (weight:0.276)	Citation width (weight:0.24)	Citation depth (weight:0.484)	Integrated academic impact	Citation counts	Paper rank
HCP-1	119.85	109.2125	40.585	78.949	291	2
HCP-2	121.39	125.499	31.26	78.77	290	3
HCP-3	123.78	82.023	26.15	66.529	289	4
HCP-4	118.02	141.780	84.185	107.35	289	1

Thus, all the papers in our experiments have shown diverse exhibition in the three influence indices, as well as the final universal academic impact, even if they have the same or a similar number of total citations. It shows that the total number of citations could not be a good indicator when referring to the papers' academic impact. Revealing the changing

TABLE 8. The academic impact for the MCPs.

ID	Time-weighted citation counts (weight:0.321)	Citation width (weight:0.365)	Citation depth (weight:0.314)	Integrated academic impact	Citation counts	Paper rank
MCP-1	31.51	49.786	18.735	34.161	73	3
MCP-2	30.56	35.699	24.585	30.56	73	6
MCP-3	32.04	48.607	20.2	34.361	73	2
MCP-4	31.82	49.583	18.385	34.076	73	4
MCP-5	27.15	34.331	13.07	25.342	73	8
MCP-6	31	49.159	18.16	33.588	73	5
MCP-7	28.91	42.311	31.03	34.466	73	1
MCP-8	26.99	36.632	25.705	30.104	73	7

TABLE 9. The academic impact for the LCPs.

ID	Time-weighted citation counts (weight:0.331)	Citation width (weight:0.239)	Citation depth (weight:0.43)	Integrated academic impact	Citation counts	Paper rank
LCP-1	11.208	17.52	10.885	12.572	28	14
LCP-2	12.099	20.05	9.59	12.913	28	13
LCP-3	12.669	17.968	15.16	15.003	28	3
LCP-4	9.661	15.002	12.52	12.163	28	18
LCP-5	11.429	20.877	12.535	14.155	28	6
LCP-6	12.683	16.765	6.99	11.205	28	21
LCP-7	9.939	16.398	8.23	10.742	28	24
LCP-8	9.032	18.594	10.07	11.756	28	19
LCP-9	10.625	15.517	7.335	10.374	28	26
LCP-10	11.287	14.37	13.565	13.002	28	11
LCP-11	10.981	21.962	9.54	12.976	28	12
LCP-12	8.839	10.741	16.27	12.49	28	15
LCP-13	11.607	17.694	10.135	12.423	28	17
LCP-14	10.078	23.972	19.305	17.358	28	2
LCP-15	12.392	19.409	8.64	12.448	28	16
LCP-16	10.449	18.687	12.75	13.401	28	9
LCP-17	13.101	24.203	8.015	13.556	28	8
LCP-18	10.061	12.303	4.615	8.251	28	28
LCP-19	10.618	15.252	7.955	10.575	28	25
LCP-20	13.273	16.859	21.465	17.653	28	1
LCP-21	11.547	18.952	11.335	13.219	28	10
LCP-22	10.261	14.574	9.095	10.786	28	23
LCP-23	10.521	19.48	6.845	11.072	28	22
LCP-24	11.641	20.042	13.255	14.336	28	4
LCP-25	9.739	16.704	5.74	9.676	28	27
LCP-26	12.154	22.256	10.48	13.839	28	7
LCP-27	11.818	19.651	6.66	11.463	28	20
LCP-28	11.069	17.892	14.815	14.306	28	5

characteristics of the influence, detecting the spreading property of the influence and investigating the contribution of one paper to the follow-up research, can benefit to express the academic impact of the paper more effectively than the total number of citations.

In order to clarify that the proposed method can still be used to measure the academic impact of papers published in different times, with different number of citations and from other disciplines, we further analyzed papers published in the field of *Information Science & Library Science*. Taking journal *Scientometrics* as an example, four highly-cited articles, twenty medium-cited articles and forty low-cited articles were selected from the articles published in 2003 and 2004 respectively. The academic impact of these papers was measured by the method proposed in this paper. The experimental results show that the method proposed in our work can measure the academic impact of the papers with different publishing ages and different number of citations, and the experimental conclusions are similar to those in the field of *Physics*. These papers exhibit different citation intensity characteristics in different citation years, and the recent citation behavior has relatively higher intensity. These papers have different characteristics of citation width, showing the different range of their influence in the scientific environment. A few papers have been mentioned many times in their citing documents, but most of them have been mentioned less, showing an obvious power law distribution between the mentioned frequencies and the number of citing documents. After integrating the three dimensions into the overall academic impact, it shows that the performance of these papers in academic impact is different. Some of the papers with the similar number of citations have shown great differences in academic impact. Some of the papers with lower citation counts even have higher academic impact than those with higher citation counts. However, due to the length of the paper, these experimental data and the related experimental results are all put in the Supplementary Materials.

VI. CONCLUSIONS

In this paper, we mainly perform to distinguish citations. Three indices of time-weighted citation counts, citation width, and citation depth were established to investigate the evolving characteristics of influences and the diverse contributions of citations, with the aim of revealing the different impact of papers even they have the same or similar total citations. Some interesting phenomena and useful results have been discovered in this paper.

Firstly, the influence intensities of papers are constantly decreasing from near to far, indicating that the citations obtained in different time periods have the unequal importance. And the citations obtained recently, especially those obtained in the last two years, have the largest vitality and can best predict the future citation activities of papers. In the present work, all the past citations were considered and weighted by the value of correlations calculated in different independent time windows to make a universal evaluation on papers' influence intensity. Recently obtained citations are given greater weights, and the older ones have less weights. In this way, we distinguish and characterize the contribution of different citations in measuring the overall impact of papers. Detecting the influence intensity of papers in different time windows could help to solve the time biased dilemma of evaluating academic entities in reality. For example, it's usually hard to compare papers if they published in the same journal but in different years. As it needs time to accumulate citations, the papers published much longer tend to gain more citations. The conclusion achieved in this paper provides a solution to differentiate the papers by comparing the weighted total citations to verify which paper would have higher total vitality or influence. Comparing the vitality by weighting past citations could also benefit to eliminate the dilemma of how to evaluate scientists with different research ages.

Secondly, papers show diverse scope of their influences in the scientific environment. These influences have made considerable contributions for paper to accumulate citations proved by the high mutual information between papers' total citations and their citation diffusion properties in different dimensions. In the previous work, the authors have established a machine learning framework and found that the citation distribution in the early years in the feature space of $F = \{$subjects, journals, countries, and institutions$\}$ is good indicator for predicting papers' future citation impact [52], [53]. The present work further verified the important dependence of papers' citations on this feature space by using the statistics method. From the perspective of knowledge flow, every citation activity is accompanied with a spread of knowledge from the cited paper to the citing ones. Therefore, the citation diffusion characteristics actually reflect the scope of the knowledge diffusion. A larger influence range represents a larger scope of knowledge diffusion. In this way, the citation diffusion of papers provides an important index to evaluate papers' contributions in the width level.

Thirdly, papers are mentioned with different times in different citing works, showing the different usefulness of them within the citing publications. Although MCPs and LCPs have smaller total citation counts than HCPs, there are still some MCPs and LCPs get more mentions in the citing works than HCPs, showing the great difference between papers' total citation counts and their mention frequencies. The total citation is only the accumulation number of citing publications, regardless of how many times one cited paper is mentioned in the citing works. The mention frequencies could be better to investigate the contribution of one paper to the follow-up research. A paper with more mention frequencies should be more valuable for the citing work than the ones less mentioned. However, the mention frequencies mainly address the usefulness of citations from the syntactic level, but not refer to any semantic. This paper further calculated the subject similarity between cited and citing work to distinguish the contribution of citations from the semantic level. An unsupervised framework, Doc2Vec, was introduced to vectorize the unstructured documents in order to facilitate the similarity calculation. Finally, both the mention frequencies and the subject similarities were integrated to describe the depth index in evaluating papers' academic impact.

In conclusion, calculating academic impact using these three indices has provided a novel way to evaluate the value of publications. Previously, a publication's accumulated influence was calculated using a citation-count-based assessment strategy. Here, we contribute to the impact evaluation by considering various citation patterns to detect the different citation intensities and the different contributions of cited papers to the citing works. This consideration should not only be helpful for evaluating publications, but also be useful for decision-makers to evaluate the academic performance of different researchers, or different institutions. It is undeniable that the three impact indices and the method of calculating weights can be adjusted depending on the application. Decision-makers can select or focus on different aspects of the influence to meet the needs of actual decision-making tasks, and formulate a reasonable weight system in line with the actual situation, so as to complete the evaluation task from a more realistic point of view rather than just looking at the number of citations. And the method proposed in this paper is not only applicable to the measurement of academic impact of journal papers, but also to the evaluation of academic impact of other types of academic achievements, such as conference papers, academic monographs, etc. For conference papers and academic monographs, the total number of citations is also accumulated by citations distributed in different years, and there are also differences in citation intensity in each citation year. In addition, the conference papers and academic monographs will also show different influence width and depth on the follow-up research through the occurrence of citation behavior. By examining the performance in the above three citation dimensions, we can also reveal a more comprehensive academic impact for conference papers and academic monographs. However, there may be differences in academic resource databases that provide article indexing, which will lead to differences in data acquisition processes. The data in this paper are all from *Web of Science* provided by *Clarivate Analytics*, which provides tools such as "*Analysis Results*" and "*Create Citation Report*" to help complete the collection of all experimental data. For the conference papers and academic monographs, it may be difficult to collect the data used in the three citation dimensions, but as long as the technical limitations of data collection are overcome, the academic impact of them can be measured.

Furthermore, the work of this paper can only be used as the beginning of the scientific evaluation of academic value of academic achievements, and there are still many work to be improved to apply to a wider range of impact evaluation issues. For example, the proposed method can only measure and distinguish the academic impact of articles published in the same subject area. For academic papers in different fields, it is difficult to make a comparative analysis of the academic impact of them because of the differences in their domain characteristics. At this time, it is necessary to find an appropriate technology to measure the differences of domain features and to normalize the domain features according to these differences, so as to analyze the academic impact of articles from different fields. This kind of exploration is very important for completing the evaluation of the influence of cross-domain articles and for realizing the evaluation of academic entities based on it. These explorations will be part of our future work.

REFERENCES

[1] J. Wilsdon, L. Allen, E. Belfiore, P. Campbell, S. Curry, S. Hill, R. Jones, and R. Kain, "The metric tide," in *Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. Bristol, U.K.: HEFCE, 2015.

[2] C. Carroll, "Measuring academic research impact: Creating a citation profile using the conceptual framework for implementation fidelity as a case study," *Scientometrics*, vol. 109, no. 2, pp. 1329–1340, 2016.

[3] C. C. Sarli, E. K. Dubinsky, and K. L. Holmes, "Beyond citation analysis: A model for assessment of research impact," *J. Med. Library Assoc.*, vol. 98, no. 1, pp. 17–23, 2010.

[4] B.-C. Björk and D. Solomon, "Article processing charges in OA journals: Relationship between price and quality," *Scientometrics*, vol. 103, no. 2, pp. 373–385, 2015.

[5] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 46, pp. 16569–16572, 2005.

[6] J. E. Hirsch, "Does the *h* index have predictive power?" *Proc. Nat. Acad. Sci. USA*, vol. 104, no. 49, pp. 19193–19198, 2007.

[7] M. Kosmulski, "A new Hirsch-type index saves time and works equally well as the original h-index," *ISSI Newslett.*, vol. 2, no. 3, pp. 4–6, 2006.

[8] A. Sidiropoulos, D. Katsaros, and Y. Manolopoulos, "Generalized Hirsch *h*-index for disclosing latent facts in citation networks," *Scientometrics*, vol. 72, no. 2, pp. 253–280, 2007.

[9] T. R. Anderson, R. K. S. Hankin, and P. D. Killworth, "Beyond the Durfee square: Enhancing the h-index to score total publication output," *Scientometrics*, vol. 76, no. 3, pp. 577–588, Sep. 2008.

[10] L. Bertoli-Barsotti and T. Lando, "The *h*-index as an almost-exact function of some basic statistics," *Scientometrics*, vol. 113, no. 2, pp. 1209–1228, 2017.

[11] L. Egghe, "Theory and practise of the g-index," *Scientometrics*, vol. 69, no. 1, pp. 131–152, 2006.

[12] R. S. J. Tol, "A rational, successive *g*-index applied to economics departments in Ireland," *J. Informetrics*, vol. 2, no. 2, pp. 149–155, Apr. 2008.

[13] J. Y. Lee, "A proposal on modified g-index for evaluating research performance," *Scientometrics*, vol. 34, no. 3, pp. 209–228, 2017.

[14] P. D. Batista, M. G. Campiteli, and O. Kinouchi, "Is it possible to compare researchers with different scientific interests?" *Scientometrics*, vol. 68, no. 1, pp. 179–189, 2006.

[15] B. Jin, L. Liang, R. Rousseau, and L. Egghe, "The *R*- and *AR*-indices: Complementing the *h*-index," *Chin. Sci. Bull.*, vol. 52, no. 6, pp. 855–863, Mar. 2007.

[16] N. J. van Eck and L. Waltman, "Generalizing the *h*- and *g*-indices," *J. Informetrics*, vol. 2, no. 4, pp. 263–271, 2008.

[17] J. Antonakis and R. Lalive, "Quantifying scholarly impact: IQp versus the Hirsch *h*," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 59, no. 6, pp. 956–969, Apr. 2008.

[18] C. Bergstrom, "Eigenfactor: Measuring the value and prestige of scholarly journals," *College Res. Libraries News*, vol. 68, no. 5, pp. 314–316, 2007.

[19] A. Fersht, "The most influential journals: Impact Factor and Eigenfactor," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 17, pp. 6883–6884, 2009.

[20] J. D. West, T. C. Bergstrom, and C. T. Bergstrom, "The Eigenfactor Metrics: A network approach to assessing scholarly journals," *College Res. Libraries*, vol. 71, no. 3, pp. 236–244, 2010.

[21] Q. Wu, "The w-index: A measure to assess scientific impact by focusing on widely cited papers," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 61, no. 3, pp. 609–614, 2010.

[22] F. J. Cabrerizo, S. Alonso, E. Herrera-Viedma, and F. Herrera, "q^2-index: Quantitative and qualitative evaluation based on the number and impact of papers in the Hirsch core," *J. Informetrics*, vol. 4, no. 1, pp. 23–28, Jan. 2010.

[23] J. Bollen, M. Rodriquez, and H. Van de Sompel, "Journal status," *Scientometrics*, vol. 6, no. 3, pp. 669–687, Dec. 2006.

[24] F. Zhang, "Evaluating journal impact based on weighted citations," *Scientometrics*, vol. 113, no. 2, pp. 1155–1169, 2018.

[25] F. Radicchi, S. Fortunato, S. Markines, and A. Vespignani, "Diffusion of scientific credits and the ranking of scientists," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 80, no. 5, 2009, Art. no. 056103.

[26] Y. Ding, E. Yan, A. Frazho, and J. Caverlee, "PageRank for ranking authors in co-citation networks," *J. Assoc. Inf. Sci. Technol.*, vol. 60, no. 11, pp. 2229–2243, 2009.

[27] J. Zhang, Y. Hu, Z. Ning, A. Tolba, E. Elashkar, and F. Xia, "AIRank: Author impact ranking through positions in collaboration networks," *Complexity*, vol. 6, Jun. 2018, Art. no. 4697485.

[28] D. Walker, D. Xie, K.-K. Yan, and S. Maslov, "Ranking scientific publications using a model of network traffic," *J. Stat. Mech., Theory Exp.*, vol. 6, Jun. 2007, Art. no. P06010.

[29] S. Maslov and S. Redner, "Promise and pitfalls of extending Google's PageRank algorithm to citation networks," *J. Neurosci.*, vol. 28, no. 44, pp. 11103–11105, 2008.

[30] S. Ma, J. Guan, and Y. Zhao, "Bringing PageRank to the citation analysis," *Inf. Process. Manage.*, vol. 44, no. 2, pp. 800–810, 2008.

[31] J. Li and P. Willett, "ArticleRank: A PageRank-based alternative to numbers of citations for analysing citation networks," *Aslib Proc.*, vol. 61, no. 6, pp. 605–618, 2009.

[32] E. Yan, Y. Ding, and C. R. Sugimoto, "P-Rank: An indicator measuring prestige in heterogeneous scholarly networks," *J. Assoc. Inf. Sci. Technol.*, vol. 62, no. 3, pp. 467–477, Mar. 2011.
[33] C. Su, Y. Pan, Y. Zhen, Z. Ma, J. Yuan, H. Guo, Z. Yu, C. Ma, and Y. Wu, "PrestigeRank: A new evaluation method for papers and journals," *J. Inform.*, vol. 5, no. 1, pp. 1–13, 2011.
[34] H. Qiao, Y. Wang, and Y. Liang, "A value evaluation method for papers based on improved PageRank algorithm," in *Proc. 2nd Int. Conf. IEEE Comput. Sci. Netw. Technol. (ICCSNT)*, Dec. 2012, pp. 2201–2205.
[35] C. Chen, "Predictive effects of structural variation on citation counts," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 63, no. 3, pp. 431–449, Mar. 2012.
[36] C. Chen, "Eugene Garfield's scholarly impact: A scientometric review," *Scientometrics*, vol. 114, no. 2, pp. 489–516, 2018.
[37] E. Fragkiadaki, G. Evangelidis, N. Samaras, and D. A. Dervos, "F-value: Measuring an article's scientific impact," *Scientometrics*, vol. 86, no. 3, pp. 671–686, 2011.
[38] L. Waltman and E. Yan, "PageRank-related methods for analyzing citation networks," in *Measuring Scholarly Impact*. New York, NY, USA: Springer, 2014, pp. 83–100.
[39] E. Fragkiadaki and G. Evangelidis, "Review of the indirect citations paradigm: Theory and practice of the assessment of papers, authors and journals," *Scientometrics*, vol. 99, no. 2, pp. 261–288, 2014.
[40] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *J. ACM*, vol. 46, no. 5, pp. 604–632, 1999.
[41] X. Bai, H. Liu, F. Zhang, Z. Ning, X. Kong, I. Lee, and F. Xia, "An overview on evaluating and predicting scholarly article impact," *Information*, vol. 8, no. 3, p. 73, 2017.
[42] X. Bai, F. Zhang, J. Hou, I. Lee, X. Kong, A. Tolba, and F. Xia, "Quantifying the impact of scholarly papers based on higher-order weighted citations," *PLoS ONE*, vol. 13, no. 3, 2018, Art. no. e0193192.
[43] G. Pinski and F. Narin, "Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics," *Inf. Process. Manage.*, vol. 12, no. 5, pp. 297–312, 1976.
[44] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," *Comput. Netw. ISDN Syst.*, vol. 30, nos. 1–7, pp. 107–117, Apr. 1998.
[45] B. González-Pereira, V. P. Guerrero-Bote, and F. Moya-Anegón, "A new approach to the metric of journals' scientific prestige: The SJR indicator," *J. Informetrics*, vol. 4, no. 3, pp. 379–391, Jul. 2010.
[46] V. P. Guerrero-Bote and F. Moya-Anegón, "A further step forward in measuring journals' scientific prestige: The SJR2 indicator," *J. Informetrics*, vol. 6, no. 4, pp. 674–688, Oct. 2012.
[47] J. Guan, J. Yan, and J. J. Zhang, "The impact of collaboration and knowledge networks on citations," *J. Informetrics*, vol. 11, no. 2, pp. 407–422, May 2017.
[48] R. Rousseau, C. García-Zorita, and E. Sanz-Casado, "The *h*-bubble," *J. Informetrics*, vol. 7, no. 2, pp. 294–300, 2013.
[49] E. Garfield, "Is citation analysis a legitimate evaluation tool?" *Scientometrics*, vol. 1, no. 4, pp. 359–375, 1979.
[50] L. Leydesdorff and L. Bornmann, "Integrated impact indicators compared with impact factors: An alternative research design with policy implications," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 62, no. 11, pp. 2133–2146, Nov. 2011.
[51] C. Lu, C. Ding, and C. Zhang, "Understanding the impact change of a highly cited article: A content-based citation analysis," *Scientometrics*, vol. 112, no. 2, pp. 927–945, 2017.
[52] M. Wang, G. Yu, S. An, and D. Yu, "Discovery of factors influencing citation impact based on a soft fuzzy rough set model," *Scientometrics*, vol. 93, no. 3, pp. 635–644, 2012.
[53] M. Wang, G. Yu, J. Xu, H. He, D. Yu, and S. An, "Development a case-based classifier for predicting highly cited papers," *J. Informetrics*, vol. 6, no. 4, pp. 586–599, 2012.
[54] S. Pak, G. Yu, and G. Wang, "A study on the citation situation within the citing paper: Citation distribution of references according to mention frequency," *Scientometrics*, vol. 114, no. 3, pp. 905–918, Mar. 2018.
[55] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 26, 2013, pp. 3111–3119.
[56] D. Wang, C. Song, and A.-L. Barabási, "Quantifying long-term scientific impact," *Science*, vol. 342, no. 6154, pp. 127–132, 2013.
[57] F. Xia, X. Su, W. Wang, C. Zhang, Z. Ning, and I. Lee, "Bibliographic analysis of *nature* based on Twitter and Facebook altmetrics data," *PLoS ONE*, vol. 11, no. 12, 2016, Art. no. e0165997.
[58] E. Garfield, "The history and meaning of the journal impact factor," *J. Amer. Med. Assoc.*, vol. 295, no. 1, pp. 90–93, 2006.
[59] R. K. Pan and S. Fortunato, "Author Impact Factor: Tracking the dynamics of individual scientific impact," *Sci. Rep.*, vol. 4, May 2014, Art. no. 4880.
[60] K. B. Hajra and P. Sen, "Aging in citation networks," *Phys. A, Stat. Mech. Appl.*, vol. 346, nos. 1–2, pp. 44–48, 2005.
[61] Y. Wang, Y. Tong, and M. Zeng, "Ranking scientific articles by exploiting citations, authors, journals, and time information," in *Proc. 27th AAAI Conf. Artif. Intell.*, 2013, pp. 933–939.
[62] M. Wang, G. Yu, and D. Yu, "Measuring the preferential attachment mechanism in citation networks," *Phys. A, Stat. Mech. Appl.*, vol. 387, no. 18, pp. 4692–4698, Jul. 2008.
[63] M. Wang, G. Yu, and D. Yu, "Effect of the age of papers on the preferential attachment in citation networks," *Phys. A, Stat. Mech. Appl.*, vol. 388, no. 19, pp. 4273–4276, 2009.
[64] Z. Yang, S. Zhang, W. Shen, X. Xing, and Y. Gao, "Artificial intelligence related publication analysis based on citation counting," *IEEE Access*, vol. 6, pp. 52205–52217, 2018.
[65] M. Wang, S. Li, and G. Chen, "Detecting latent referential articles based on their vitality performance in the latest 2 years," *Scientometrics*, vol. 112, no. 3, pp. 1557–1571, Sep. 2017.
[66] Y. Ding, G. Zhang, T. Chambers, M. Song, X. Wang, and C. Zhai, "Content-based citation analysis: The next generation of citation analysis," *J. Assoc. Inf. Sci. Technol.*, vol. 65, no. 9, pp. 1820–1833, 2014.
[67] H. Voos and K. S. Dagaev, "Are all citations equal? Or, did we op. cit. your idem?" *J. Acad. Librarianship*, vol. 1, no. 6, pp. 19–21, 1976.
[68] S. Bonzi, "Characteristics of a literature as predictors of relatedness between cited and citing works," *J. Amer. Soc. Inf. Sci.*, vol. 33, no. 4, pp. 208–216, Jul. 1982.
[69] B. C. Peritz, "A classification of citation roles for the social sciences and related fields," *Scientometrics*, vol. 5, no. 5, pp. 303–312, 1983.
[70] F. Luo, A. Sun, M. Erdt, A. S. Raamkumar, and Y.-L. Theng, "Exploring prestigious citations sourced from top universities in bibliometrics and altmetrics: A case study in the computer science discipline," *Scientometrics*, vol. 114, no. 1, pp. 1–17, 2018.
[71] X. Bai, J. Zhang, H. Cui, Z. Ning, and F. Xia, "PNCOIRank: Evaluating the impact of scholarly articles with positive and negative citations," *Proc. 25th Int. Conf. Companion World Wide Web, Int. World Wide Web Conf. Steering Committee*, 2016, pp. 9–10.
[72] J. P. A. Ioannidis, "A generalized view of self-citation: Direct, co-author, collaborative, and coercive induced self-citation," *J. Psychosomatic Res.*, vol. 78, no. 1, pp. 7–11, 2015.
[73] M. Valenzuela, M. Ha, and O. Etzioni, "Identifying meaningful citations," in *Proc. Workshops 29th AAAI Conf. Artif. Intell.*, 2015, pp. 1–26.
[74] A. Schubert, "Measuring the similarity between the reference and citation distributions of journals," *Scientometrics*, vol. 96, no. 1, pp. 305–313, 2013.
[75] Q. Mahmood, M. A. Qadir, and M. T. Afzal, "Application of COReS to compute research papers similarity," *IEEE Access*, vol. 5, pp. 26124–26134, 2017.
[76] Y. Ma, P. Zhang, and J. G. Ma, "An ontology driven knowledge block summarization approach for Chinese judgment document classification," *IEEE Access*, vol. 6, pp. 71327–71338, 2018.
[77] R.-L. Liu, "Passage-based bibliographic coupling: An inter-article similarity measure for biomedical articles," *PLoS ONE*, vol. 10, no. 10, 2015, Art. no. e0139245.
[78] C. Sternitzke and I. Bergmann, "Similarity measures for document mapping: A comparative study on the level of an individual scientist," *Scientometrics*, vol. 78, no. 1, pp. 113–130, 2009.
[79] R. Socher, D. Chen, C. D. Manning, and A. Ng, "Reasoning with neural tensor networks for knowledge base completion," in *Proc. 26th Int. Conf. Neural Inf. Process. Syst.*, vol. 1, 2013, pp. 926–934.
[80] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2013, pp. 1631–1642.
[81] C. Zhang, L. Zhang, C.-J. Wang, and J.-Y. Xie, "Text summarization based on sentence selection with semantic representation," in *Proc. IEEE 26th Int. Conf. Tools Artif. Intell.*, Nov. 2014, pp. 584–590.
[82] L. Yang, Y. Zheng, X. Cai, H. Dai, D. Mu, L. Guo, and T. Dai, "A LSTM based model for personalized context-aware citation recommendation," *IEEE Access*, vol. 6, pp. 59618–59627, 2018.

[83] D. Fiala and G. Tutoky, "Pagerank-based prediction of award-winning researchers and the impact of citations," *J. Informetrics*, vol. 11, no. 4, pp. 1044–1068, 2017.
[84] P. Chen, H. Xie, S. Maslov, and S. Render, "Finding scientific gems with Google's PageRank algorithm," *J. Informetrics*, vol. 1, no. 1, pp. 8–15, 2007.
[85] M. S. Mariani, M. Medo, and Y.-C. Zhang, "Identification of milestone papers through time-balanced network centrality," *J. Informetrics*, vol. 10, no. 4, pp. 1207–1223, Nov. 2016.
[86] P. Kazi, M. Patwardhan, and P. Joglekar, "Towards a new perspective on context based citation index of research articles," *Scientometrics*, vol. 107, no. 1, pp. 103–121, Apr. 2016.
[87] H. Sayyadi and L. Getoor, "FutureRank: Ranking scientific articles by predicting their future PageRank," in *Proc. SIAM Int. Conf. Data Mining (SDM)*, Sparks, NV, USA, 2009, pp. 533–544.
[88] Y.-B. Zhou, L. Lü, and M. Li, "Quantifying the influence of scientists and their publications: Distinguishing between prestige and popularity," *New J. Phys.*, vol. 14, Mar. 2012, Art. no. 033033.
[89] J. Zhou, A. Zeng, Y. Fan, and Z. Di, "Identifying important scholars via directed scientific collaboration networks," *Scientometrics*, vol. 114, no. 3, pp. 1327–1343, Mar. 2018.
[90] U. Senanayake, M. Piraveenan, and A. Zomaya, "The pagerank-index: Going beyond citation counts in quantifying scientific impact of researchers," *PLoS ONE*, vol. 10, no. 8, 2015, Art. no. e0134794.
[91] D. Fiala, L. Šubelj, S. Žitnik, and M. Bajec, "Do PageRank-based author rankings outperform simple citation counts?" *J. Informetrics*, vol. 9, no. 2, pp. 334–348, 2015.
[92] B. Cheang, S. K. W. Chu, C. Li, and A. Lim, "A multidimensional approach to evaluating management journals: Refining PageRank via the differentiation of citation types and identifying the roles that management journals play," *J. Assoc. Inf. Sci. Technol.*, vol. 65, no. 12, pp. 2581–2591, 2014.
[93] E. Yan and Y. Ding, "Discovering author impact: A PageRank perspective," *Inf. Process. Manage.*, vol. 47, no. 1, pp. 125–134, 2011.
[94] J. Lee and D.-W. Kim, "Feature selection for multi-label classification using multivariate mutual information," *Pattern Recognit. Lett.*, vol. 34, no. 3, pp. 349–357, 2013.
[95] M. Han, W. Ren, and X. Liu, "Joint mutual information-based input variable selection for multivariate time series modeling," *Eng. Appl. Artif. Intell.*, vol. 37, pp. 250–257, Jan. 2015.
[96] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *Proc. 31st Int. Conf. Mach. Learn.*, 2014, pp. 1188–1196.
[97] S. Huang, J. Chang, G. Leng, and Q. Huang, "Integrated index for drought assessment based on variable fuzzy set theory: A case study in the Yellow River basin, China," *J. Hydrol.*, vol. 527, pp. 608–618, Aug. 2015.
[98] S. Huang, B. Ming, Q. Huang, G. Leng, and B. Hou, "A case study on a combination NDVI forecasting model based on the entropy weight method," *Water Resour. Manage.*, vol. 31, no. 11, pp. 3667–3681, 2017.
[99] C.-L. Gao, S.-C. Li, J. Wang, L.-P. Li, and P. Lin, "The risk assessment of tunnels based on grey correlation and entropy weight method," *Geotech. Geol. Eng.*, vol. 36, no. 3, pp. 1621–1631, Jun. 2018.
[100] T. Xie, M. Wang, W. Chen, and C. Su, "Evaluation of the natural attenuation capacity of urban residential soils with ecosystem-service performance index (EPX) and entropy-weight methods," *Environ. Pollut.*, vol. 238, pp. 222–229, Jul. 2018.
[101] G. Yu, T. Yu, and L. Wang, "Assessing influence of scientific articles based on feature spaces of citations," in *Proc. Int. Conf. Manage. Sci. Eng. (ICMSE)*, 2016, pp. 41–48.
[102] D. W. Aksnes, "Characteristics of highly cited papers," *Res. Eval.*, vol. 12, no. 3, pp. 159–170, 2003.

MINGYANG WANG received the Ph.D. degree from the School of Economics and Management, Harbin Institute of Technology, China, in 2011. She is currently an Associate Professor with the College of Information and Computer Engineering, Northeast Forestry University. She has a number of publications in international journals and conferences. Her current research interests include text mining, knowledge management, and scientometrics. She is the Reviewer of several international journals.

JINJIN REN is currently pursuing the B.S. degree with the College of Information and Computer Engineering, Northeast Forestry University, China. Her current research interests include machine learning, social networks, and citation networks.

SHI LI received the Ph.D. degree from the School of Economics and Management, Harbin Institute of Technology, China, in 2012. She is currently an Associate Professor with the College of Information and Computer Engineering, Northeast Forestry University. Her current research interests are in the areas of artificial intelligence, natural language processing, and bibliometrics.

GUANGSHENG CHEN is currently a Professor with the College of Information and Computer Engineering, Northeast Forestry University. He is also a member of the National Innovation Method Research Association and also the Executive Director of the Education Informationization Council, Ministry of Education. His current research interests include the technological innovation methods (TRIZ theory), information retrieval, social networks, and data mining.