Journal of Informetrics 13 (2019) 616–634

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Regular article

Ranking scientific articles based on bibliometric networks with a weighting scheme

Yu Zhang[a,*], Min Wang[b], Florian Gottwalt[a], Morteza Saberi[a], Elizabeth Chang[a]

[a] *School of Business, University of New South Wales, Canberra, Northcott Dr, Campbell ACT 2612, Australia*
[b] *School of Engineering and Information Technology, University of New South Wales, Canberra, Northcott Dr, Campbell ACT 2612, Australia*

ARTICLE INFO

Article history:
Received 20 June 2018
Received in revised form 24 March 2019
Accepted 25 March 2019

Keywords:
Scientific article ranking
Ranking algorithm
Citation analysis
Heterogeneous network

ABSTRACT

As the volume of scientific articles has grown rapidly over the last decades, evaluating their impact becomes critical for tracing valuable and significant research output. Many studies have proposed various ranking methods to estimate the prestige of academic papers using bibliometric methods. However, the weight of the links in bibliometric networks has been rarely considered for article ranking in existing literature. Such incomplete investigation in bibliometric methods could lead to biased ranking results. Therefore, a novel scientific article ranking algorithm, W-Rank, is introduced in this study proposing a weighting scheme. The scheme assigns weight to the links of citation network and authorship network by measuring citation relevance and author contribution. Combining the weighted bibliometric networks and a propagation algorithm, W-Rank is able to obtain article ranking results that are more reasonable than existing PageRank-based methods. Experiments are conducted on both arXiv hep-th and Microsoft Academic Graph datasets to verify the W-Rank and compare it with three renowned article ranking algorithms. Experimental results prove that the proposed weighting scheme assists the W-Rank in obtaining ranking results of higher accuracy and, in certain perspectives, outperforming the other algorithms.

1. Introduction

Ranking scientific publications has always been an attractive topic in the community of scientometric research as it can assist researchers in locating important studies with critical contribution amongst ever growing literature (Wang, Tong, & Zeng, 2013). However, obtaining ranking lists for the great volume of academic papers is extremely challenging predominantly due to the complex relationship amongst the bibliometric entities and the intricate links in heterogeneous bibliometric networks. Research on this topic has been undertaken for decades, yet there are still issues to be revealed and addressed.

Earlier article ranking methods, including PageRank and its variants, focused too much on measuring citations of publications, which led to results being biased to papers published earlier due to their longevity allowing for the possibility of more citations over the years (Sayyadi & Getoor, 2009). CiteRank preliminarily addressed this issue (Walker, Xie, Yan, & Maslov, 2007), FutureRank (Sayyadi & Getoor, 2009) and P-Rank (Yan, Ding, & Sugimoto, 2011) proposed to seek more information or indicators for ranking publications equitably and efficiently, such as publication time (time indicator), author-paper information (authorship) and journal-paper information. Later on, the concept of heterogeneous networks and propaga-

[*] Corresponding author.

https://doi.org/10.1016/j.joi.2019.03.013
1751-1577/© 2019 Elsevier Ltd. All rights reserved.

tion algorithms were introduced to integrate bibliometric information to a heterogeneous network (Yan et al., 2011) and calculate a ranking score for each paper by considering the impact coming from every layer of the network (Wang et al., 2013). However, the influence coming from the weight of the links in the networks has rarely been considered. Although the factors of topic and time have been used to quantify the weight of the links in citation networks (Tang, Jin, & Zhang, 2008; Walker et al., 2007), the reason of using these factors is weak in terms of interpreting the root cause of the dynamic nature of bibliometric networks.

Accordingly, a link weighting scheme is proposed in order to interpret the relationships between the nodes in bibliometric networks and assign weight to the links based on the relationships. For instance, citation relevance varies amongst articles because references are cited for various reasons (Garfield, 1979). Similarly, co-authors contribute differently to a paper.

Based on the weighting scheme, a scientific article ranking algorithm, W-Rank, is proposed which takes into account the weight of the links in citation and authorship networks attempting to obtain reliable ranking results. A combination of network-based method and semantic-based method is adopted for measuring the citation relevance, and a harmonic counting method is used for measuring author contribution. Integrating the HITS algorithm (Kleinberg, 1999) with a heterogeneous bibliometric network, the W-Rank is shaped and able to rank scientific articles.

The contribution of this work is three-fold.

- A weighting scheme is proposed to take into account the influence coming from citation relevance and author contribution to scientific article ranking methods. It succeeds in promoting the traditional heterogeneous bibliometric networks to weighted topology by adding weight to the links in citation and authorship networks. Moreover, the citation relevance is quantified by using a combination of network-based and semantic-based methods for the first time.
- A scientific article ranking algorithm, W-Rank, is proposed based on the improved heterogeneous bibliometric networks and a propagation algorithm. According to the results of experiments, the W-Rank shows promising and competitive performance compared with its predecessors.
- This study introduces a new perspective to explore the meaning of the links in bibliometric networks. This perspective interprets the links based on the practical meaning of the relationships between the nodes in the networks, which could bring deeper thought to improving the structure of bibliometric networks and scholarly ranking algorithms.

The remaining sections are arranged as follows. The second section narrates the literature in the field of academic paper ranking and critical review pointing out research gaps. The next section explains the proposed methods in length including the link weighting scheme and improved heterogeneous bibliometric networks. The experiments are designed and conducted in the fourth section explaining the dataset, evaluation procedures and detailed experiment process. Results, discussion and important findings are included in the fifth section. Lastly, the sixth section concludes this study and forecasts the future work for this research field.

2. Related work

2.1. Literature review on scholarly ranking methods

Bibliometric information has been adopted for investigating scholarly analysis and ranking, such as citation analysis, ranking academic papers and measuring the prestige of authors and journals. For instance, Impact Factor was the earliest method proposed to estimate the significance of journals using citation count (Garfield, 1965). Following this study, several journal ranking methods were proposed, including Source Normalised Impact per Paper (SNIP), Impact per Publication (IPP) (Moed, 2010; Waltman, van Eck, van Leeuwen, & Visser, 2013) and SCImago Journal Rank (SJR) (González-Pereira, Guerrero-Bote, & Moya-Anegón, 2010; Guerrero-Bote & Moya-Anegón, 2012). These methods are widely used in various scientific search engines and academic websites, which indicates the importance of this research. Therefore, measuring Impact Factor was followed up and turned to measuring the prestige of other entities, such as ranking co-authors (Garfield, 1986) and academic paper (Page, Brin, Motwani, & Winograd, 1999).

PageRank was the first algorithm introduced to consider academic paper citation networks as Internet web page links, and then integrate the citation information with a paper ranking procedure (Page et al., 1999). Later research adopted PageRank as an underlying algorithm in experiments. For example, some studies proposed to utilise PageRank for analysing citation networks and ranking academic papers, such as Y-Factor (Bollen, Rodriquez, & Van de Sompel, 2006), CiteRank (Walker et al., 2007), FutureRank (Sayyadi & Getoor, 2009) and Eigen-factor (Bergstrom & West, 2008). Some research, such as AuthorRank (Amjad, Daud, Akram, & Muhammed, 2016; Liu, Bollen, Nelson, & Van de Sompel, 2005; Nykl, Campr, & Ježek, 2015), used the PageRank to investigate author ranking. Some applied it to analyse the metadata of particular research subjects (Chen, Xie, Maslov, & Redner, 2007; Ennas, Biggio, & Di Guardo, 2015). Although the PageRank had shown advantages, it revealed several defects in the following aspects. Firstly, the PageRank method only concentrated on the citation numbers of academic papers, which led paper ranking results biased to early published articles because new articles were less likely to have sufficient citations. Secondly, extending its methodology to other scientific aspects revealed limitations. For instance, it was difficult for the PageRank to measure the impact contributed by authors or journals to the articles (Ma, Guan, & Zhao, 2008). Thirdly, all the PageRank-like studies focused on ranking scientific entities in homogeneous networks instead of integrating related information and calculating scores through heterogeneous networks (Yan et al., 2011).

In order to address the above issues, new techniques were proposed. With regards to the biased ranking results, much subsequent research has presented algorithms to capture the dynamic nature of bibliometric information. For instance, CiteRank was proposed to take the publication time of academic articles into account, which helped promote higher citation scores for recently published papers, on some level improving the bias ranking problem (Walker et al., 2007). In the same year, the Co-Rank approach was introduced to rank scientific papers and authors simultaneously by adding the social networks of authors to its algorithm (Zhou, Orshanskiy, Zha, & Giles, 2007). FutureRank was another famous method which integrated citation networks with authorship networks, predicting future citations for articles by means of the interaction between the networks (Sayyadi & Getoor, 2009). In addition, some paper ranking methods adopted a propagation algorithm to solve the expending limitations of PageRank (Kleinberg, 1999). It claimed the distinction between hubs and authorities within citation networks and interclass networks between articles and other bibliometric entities, then calculated their prestige in a mutually reinforcing way (Kleinberg, 1999). Researchers also adopted authorship and propagation algorithms to obtain better ranking results for scientists and their publications (Gollapalli, Mitra, & Giles, 2011). Moreover, the concept of heterogeneous networks was introduced in paper ranking methodologies for the purpose of utilising citation, authorship and journal information together to obtain more comprehensive and balanced results. A random walk method called P-Rank was proposed to consider citation networks, authorship networks and paper-journal networks together as a heterogeneous scholarly network for the first time. P-Rank performed a random walk algorithm on a paper-author network and paper-journal network, then conducted a propagation algorithm on the networks until convergence was encountered (Yan et al., 2011). Following the idea of heterogeneous networks, a PageRank+HITS framework was presented which deployed the HITS reinforcement algorithm together with the traditional PageRank to calculate the authority scores of academic papers. The contribution of the PageRank+HITS method was that it combined HITS and PageRank algorithms, constructed a looped propagation amongst papers, authors and journals so that paper ranking results were more reasonable (Wang et al., 2013).

All of the above studies have either directly utilised paper citations to evaluate the influence or importance of papers, or imperceptibly taken advantage from citation networks. For instance, based on the existing research, it was a common technique to grant credit from authors to their papers. That is to say, papers tended to obtain higher scores if their authors held higher authority, but it was difficult to measure the reality of an author's authority solely based on public information. Therefore, in these studies, the authority of an author was determined by the quality and quantity of the papers written by the author, which means author's authorities eventually derived from the paper citation network. However, the PageRank methodology was adopted in the above studies for manipulating citation networks, which led to biased ranking results. In addition, although the heterogeneous network concept was introduced, the underlying networks were simply represented as unweighted bipartite or direct graphs, which neglected the weight of the edges in bibliometric networks.

2.2. *Attempt on weighted bibliometric networks*

Researchers have attempted to address the biased ranking issue coming from the unweighted bibliometric networks. Associating topic modelling techniques with scholarly ranking methods was the first try. An Author-Conference-Topic (ACT) model was proposed to represent the inter-dependencies amongst authors, papers, and publication entities by using a probabilistic model (Tang et al., 2008). ACT was the first attempt to combine topic modeling with random walk in bibliometric networks for improving scientific searches, and it achieved promising experimental results. Following the ACT study, topic-based PageRank was introduced to apply the ACT method in an information retrieval test field and represented authors for different topics at different time phases (Ding, 2011). The topic-based PageRank formed weighted vectors for the PageRank algorithm by using the probability of a topic for a given author in order to rank authors on a topic level. Next, the inventor of the topic-based PageRank method applied weighted PageRank to author citation networks hoping to measure two newly defined terms in science and social science journals which were 'popularity' and 'prestige' of scholarly authors (Ding & Cronin, 2011). The term of weighted citation was also defined as an indicator of an article's prestige in which the weight of citations was differentiated in two ways: the prestige of cited journals and citation time interval (Yan & Ding, 2010). Topic-based heterogeneous ranking was presented to measure the impact of a scholarly entity with respect to a given topic in a heterogeneous scholarly network containing authors, papers and journals (Amjad, Ding, Daud, Xu, & Malic, 2015). Further, topic modelling was used for exploring topic level expertise searches (Tang et al., 2011) and dynamic research interest findings (Daud, 2012). Another inspiring study proposed RALEX, a random-walk based document ranking framework for scientific literature, which ranked scientific papers in a decreasing order of potential usefulness taking both topical content and age of the document into consideration (Xu, Martin, & Mahidadia, 2014).

In addition to topics, time was used as a factor to weigh citations in some studies. For instance, the weight of a citation was simply decreased exponentially with citation age using a base (decay rate) of 0.5 (Yu, Li, & Liu, 2004), and the age of publications was associated with the PageRank algorithm by favouring citations of more recent articles (Walker et al., 2007). Time-aware PageRank was later proposed to weigh the citations according to authors depending on a number of factors such as the number of common publications and whether or not they were published before a citation was made (Fiala, 2012; Yu, Wang, Zhang, Zhang, & Liu, 2017).

Although weight systems were introduced to associate with heterogeneous bibliometric networks, the above methods focused less on ranking academic papers, instead they measured the authority of authors based on the level of topics using topic modelling techniques. Additionally, quantifying topics or time could be one way to assign weight to the links of bibliometric networks. This makes less sense in real life since the factor of time does not have influence in the weight of

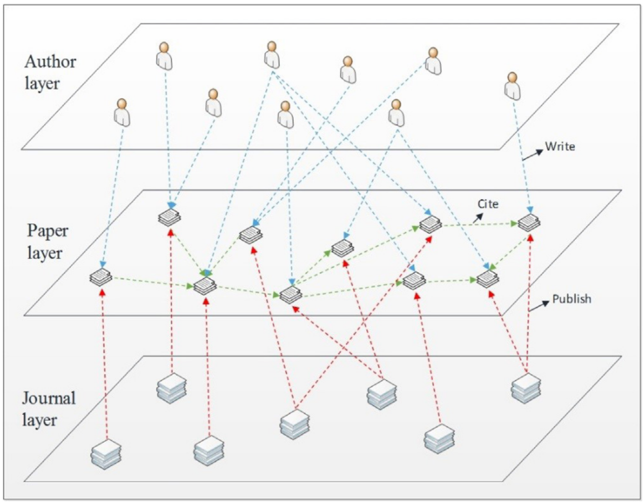


Fig. 1. An illustration of a bibliometric network.

the citation between two papers. In fact, the only relationship between citation and time was that the number of citations increases with time goes on, rather than the weight of the citations. Topic modelling is an inspiring technique to grant weight to citations, but it cannot be used on the other links of bibliometric networks. Therefore, considering all the above issues, this study tries to further improve the combination of weight systems and bibliometric networks, and to develop a more reasonable and reliable academic paper ranking algorithm.

3. Methodology

This section introduces the proposed scientific article ranking method. Firstly, a heterogeneous bibliometric network is introduced which defines three layers (i.e., author layer, paper layer, and journal/conference layers) and how the entities in the layers are linked and interacted. Secondly, a weighting scheme is proposed to capture and embed the information of citation relevance and author contribution into the bibliometric network by calculating and assigning weight to the corresponding links. Finally, a graph-based propagation algorithm is developed to calculate the scores of the articles in the weighted bibliometric network.

3.1. Heterogeneous bibliometric network

A bibliometric network is a scholarly network that integrates the information of papers, authors and journals/conferences into one heterogeneous unit that allows them to interact with each other via subnetworks (Yan et al., 2011). A bibliometric network is composed of three layers, including paper layer, author layer and journal layer. The heterogeneous graph of authors, journals and papers can be represented as follows:

$$G = (V, E) = (V_P \cup V_A \cup V_J, E_P \cup E_{PA} \cup E_{PJ}) \tag{1}$$

where V_P, V_A and V_J represent the vertices (nodes) in the paper layer, author layer and journal layer respectively, and E_P refers to the links amongst the vertices inside the paper layer while E_{PA} denotes the links between papers and authors with E_{PJ} being the links between papers and journals. The complete heterogeneous network can be demonstrated in Fig. 1.

In Fig. 1, the citation network is an unweighted direct graph located in the paper layer which can be presented as $G_P = (V_P, E_P)$. The author-paper (authorship) network and journal-paper network are two unweighted bipartite graphs presented respectively as $G_{PA} = (V_P \cup V_A, E_{PA})$ and $G_{PJ} = (V_P \cup V_J, E_{PJ})$ in which E_{PA} refers to the edges between papers and corresponding authors while E_{PJ} is between papers and the journals in which they are published.

In this study, we assign weight to the links of the bibliometric network so that the traditional bibliometric network can be updated to a weighted one. Particularly the weight is assigned in the citation network and authorship network. The paper citation graph is then updated to $G_P = (V_P, E_P, W_P)$ while the paper-author graph $G_{PA} = (V_P \cup V_A, E_{PA}, W_{PA})$, where both W_P and W_{PA} refer to the weight of the links in the graphs.

3.2. Link weighting scheme

The idea of integrating a weighting scheme to academic article ranking algorithms has been rarely considered in past literature. Such inadequate investigation leads to insufficient understanding of bibliometric networks, thereby resulting in biased ranking results. Specifically, most existing article ranking algorithms consider the relevance of the links in a

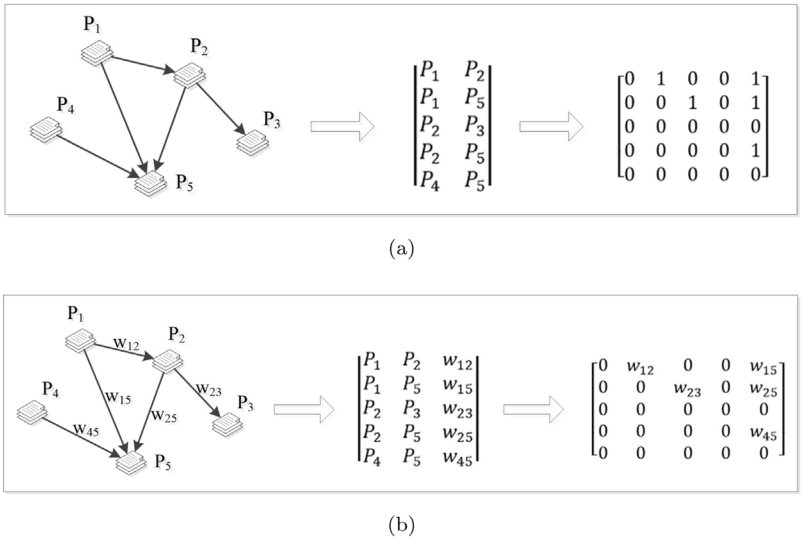


Fig. 2. Comparison between unweighted and weighted citation network.

bibliometric network as equivalent, which means the difference between the links is neglected. Missing this information simplifies the calculations for the existing algorithms by assigning the same value to all the links. Meanwhile, this could generate biased ranking results. For example, if all citations carry same effect, a paper could deliberately cite more references to increase its impact regardless of whether they are actually related to the paper. Including references and co-authors with higher reputation could also boost the impact of a paper if the weight of the links in bibliometric networks is neglected. These activities could bring about critical influence to scientific ranking algorithms and make the ranking results biased to the papers that maliciously cite unrelated references or add fake authorship.

In order to remedy the above issue, a link weighting scheme is proposed in this section to assign weight to the links of bibliometric networks based on their practical meaning. For instance, the scheme assigns weight to citation links explaining the citation relevance. A citation link carrying higher weight means the papers connected by this citation are more relevant from the perspective of topics, methods or other aspects. Integrating the weighting scheme to the networks brings a deeper understanding of bibliometric methods from a new perspective, which takes into account the difference of the links in the networks. Such improvement could pull the article ranking closer to reality, thereby making the ranking results more reasonable.

This study proposes to assign weight to the links in citation networks and authorship networks. Accordingly, the assumptions for the scheme are listed as following:

- Important articles tend to be cited by other important articles, and articles are important if they are cited by other important articles.
- Prestigious authors tend to write articles of higher quality, and authors become prestigious if other important articles cite their work.
- Journals (or conferences) with high authority tend to publish higher quality articles, and generate higher impact if their articles are cited by important articles.
- References are cited with varying degrees of citation relevance.
- Co-authors of papers contribute in varying degrees to their papers.

The first three conditions have been discussed in existing scholarly ranking research. In this study, two new assumptions are added to help further explore the nature of bibliometric networks and improve existing scientific article ranking algorithms.

3.2.1. Link weighting for citation

Most existing article ranking algorithms only use article citations in a binary way (either cited or not cited) without considering the relevance of the citations. However, we believe that the citation relevance is important information for evaluating scientific articles and should be embedded in citation networks. To address this issue, we assign weight to the links in a citation network based on citation relevance in order to extend the binary network. The differences between the two cases are illustrated in Fig. 2. In the citation networks, each node refers to a unique article in the collection and each link represents a citation with the arrow indicating the citation direction. For example, a link from P_4 to P_5 means article P_5 is cited by P_4. The graph is a two-column matrix where each column indicates source nodes and destination nodes respectively. In a

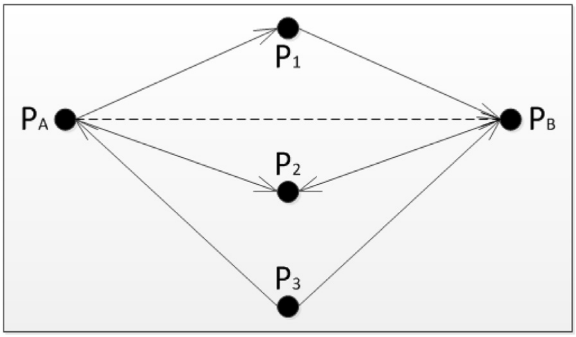


Fig. 3. Types of one-hop paper connections.

weighted graph, there will be a third column indicating the weight of each link, which, in our case, is the citation relevance. Another way of presenting a network is to use the $N \times N$ adjacency matrix A, where N is the number of articles in the network. The non-zero entry of the adjacency matrix $a_{i,j}(i, j \in N)$ denotes a link from node i to node j. In the binary case, the entries are either 0 or 1; and in the weighted case, the entries $w_{i,j}(i, j \in N)$ are the weight of the corresponding links in the network.

In this study, we define the citation relevance between two articles based on their semantic similarity and mutual links in a citation network. The relevance is then assigned to the links of the network as weight. The hypothesis is that a citation is relevant when the two articles are semantically similar or share many mutual links in the network. Specifically, if article i cites j, and their content is semantically similar, we believe the citation i to j is more relevant. Meanwhile, if the two articles are simultaneously linked by more common article nodes in the citation network (i.e., i and j cite or are cited by the same article(s), or i cites j through one-hop article node(s)), meaning the citation i to j is more relevant. The inverse of the hypothesis also holds true. A more relevant citation from article i to j means the two articles are more likely to be similar in content and are more likely to share common nodes in the citation network. We refer to these two similarities as semantic-based similarity and network-based similarity.

Regarding the network-based similarity, this study assumes that academic papers are connected into a generic citation network with directions, which means there are three possible types of one-hop connections between two individual papers as shown in Fig. 3. Say paper P_A cites paper P_B, then the three types of two-step connections are: (1) bibliographic coupling (P_A and P_B reference a common paper P_2), (2) co-citation coupling (P_A and P_B are cited by a common paper P_3), and (3) directed citation (P_A cites P_1 who cites P_B).

When considering both the in- and out-neighbourhoods of nodes, Jaccard index (Jaccard, 1901) and cosine similarity (Salton, 1970) have been commonly used. Besides, there were other methods proposed to give weight to the citation links based on the number of in- and out-neighbourhoods, such as Combined Linkage (Small, 1997) and weighted direct citations (WDC) (Persson, 2010). In this study, cosine similarity (Salton, 1970) is adopted since it has been widely used for citation networks (Li, Luo, & Wu, 2014), and yields better measuring results when it comes to the situation where two nodes share fewer neighbours (Steinert & Hoppe, 2016). The cosine similarity between two nodes is calculated as:

$$S_1(P_i, P_j) = CosSim(P_i, P_j) = \frac{|L_{P_i} \cap L_{P_j}|}{\sqrt{|L_{P_i}| \times |L_{P_j}|}} \tag{2}$$

where L_{P_i} denotes the number of links to and from node P_i, and $|L_{P_i} \cap L_{P_j}|$ denotes the number of nodes that connect to both P_i and P_j regardless of the direction.

However, measuring paper similarities based on network topology greatly relies on the quantity of common neighbours that nodes are sharing, which may not realistically be suitable. For instance, research about bibliometric networks may cite the articles in the field of graph theory or mathematics, which means two completely different papers in different research fields could be linked by multiple connections. Additionally, two semantically similar papers could be less connected when one or both are freshly published. Therefore, using only a network-based method to calculate paper similarities does not hold true in many situations.

To address the above issue, measuring semantic-based similarities amongst articles is also considered. According to analysis in science and technology fields, an abstract consists of the significant information elements of an article (Weissberg & Buker, 1990) which includes background information, principle activity, method of the study, results and conclusion (Cargill & O'Connor, 2013). Since the title and abstract contain important information, the sense-level semantic similarity of this information between two articles can reflect their similarity in content. To be specific, the semantic-based similarity applied in this study measures the degree of similarity between the titles and abstracts of two articles in terms of the content meaning. Therefore, we focus on semantic similarity at the sense (meaning) level, rather than lexical or text levels. A strong citation happens when the two articles are addressing similar issues or focusing on similar techniques, in other words, the citation relevance is high.

A sense-based semantic similarity measure named 'align, disambiguate and walk (ADW) (Pilehvar, Jurgens, & Navigli, 2013)' is adopted in this study to measure the semantic similarity between the titles and abstracts of two articles. The advantages of using ADW for semantic similarity in this study are two folds. Firstly, ADW focuses on the semantic similarities at the sense level, which allows comparison of the meaning (topics/issues and methods/techniques) of two articles. Secondly, ADW is able to handle texts in different sizes and has been demonstrated to be effective on semantic similarity at different levels, such as textual, word and sense levels.

ADW defines a unified semantic representation (referred to as signature) of any lexical item as a multinomial distribution generated from the random walks over a set of word senses in WordNet 3.0 where the nodes in the network represent senses present in the item. Following alignment-based disambiguation, ADW computes the semantic similarities of these signatures that are, essentially, weighted ranking of the importance of senses in WordNet for each lexical item. Considering the different sizes of these lexical items, we use a nonparametric similarity based on weighted overlap of senses in both signatures. Let S denote the overlapping senses with non-zero probability in both signatures and r_i^j denote the rank of sense $s_i \in S$ in signature j. The weighted overlap, semantic-based similarity, is calculated as:

$$S_2(P_i, P_j) = SemSim(P_i, P_j) = \sum_{i=1}^{|S|} (r_i^1 + r_i^2)^{-1} \tag{3}$$

then normalised by its maximum value, $\sum_{i=1}^{|S|}(2i)^{-1}$, to restrict the similarity value in [0, 1]. An example of calculating semantic similarity using the ADW approach is illustrated in Appendix A.

Combining the above network-based and semantic-based similarities, we finally define the weight of citation network in Eq. (4). The citation relevance is quantified by leveraging both network-based and semantic-based similarity measures as follows:

$$w_{i,j} = \alpha \cdot S_1(P_i, P_j) + \beta \cdot S_2(P_i, P_j) \tag{4}$$

where $w_{i,j}$ denotes the final weight for the citation from paper i to paper j; S_1 and S_2 denote the network-based and semantic-based similarities between two papers respectively; and α and β are their corresponding coefficients. The coefficients α and β are defined by the following exponential functions,

$$\alpha = e^{\lambda(S_1(P_i, P_j) - \tau_1)} \tag{5}$$

$$\beta = e^{\lambda(S_2(P_i, P_j) - \tau_2)} \tag{6}$$

where λ is a parameter to control the shape of the exponential function, τ_1 and τ_2 are thresholds set to be the median values of $S_1(P_i, P_j)$ and $S_2(P_i, P_j)$ respectively. We use exponential functions to calculate α and β since it is positive, monotonic, and more importantly, is able to adjust the contribution of network-based similarities and semantic-based similarities according to their values. In this study, we set $\lambda = 6$ to have a relatively steep curve in order to favour those similarity values that exceed the threshold. We normalise α and β so that $\alpha + \beta = 1$. An example of how the weight of a citation network are computed is illustrated in Fig. 4.

For a collection of N papers, the adjacency matrix of the citation network can then be represented by an $N \times N$ matrix A, where the entry $a_{i,j}$ is obtained as follows:

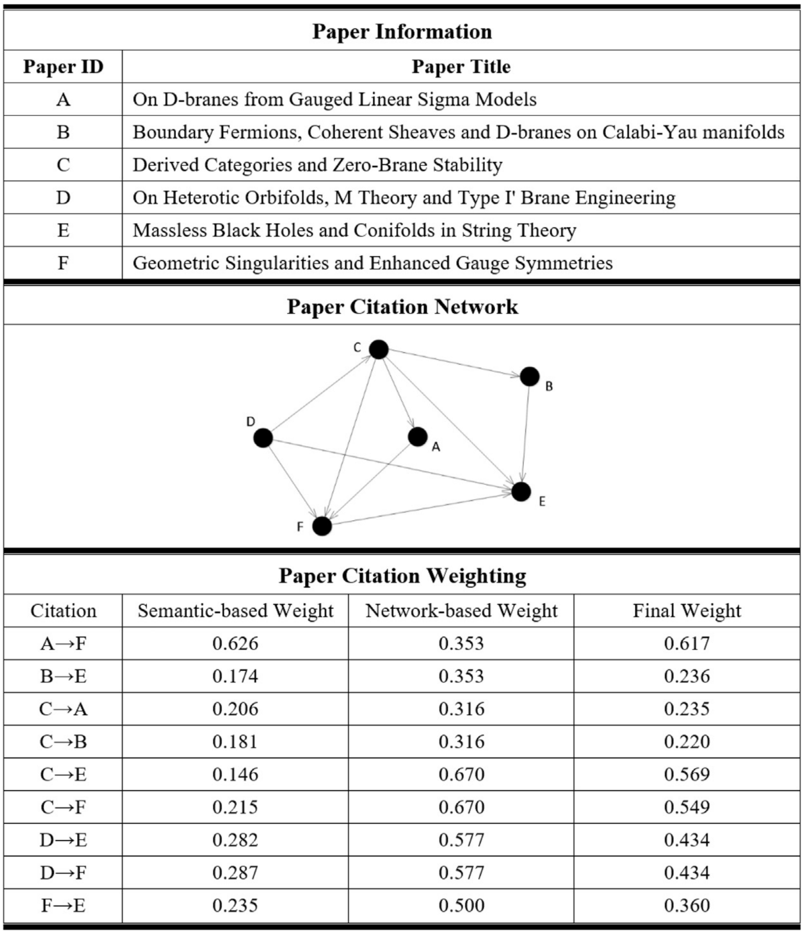
$$a_{i,j} = \{ \begin{matrix} w(i,j) & \text{if paper i cites paper j} \\ 0 & \text{otherwise} \end{matrix} \tag{7}$$

3.2.2. Link weighting for authorship

The link weighting system is also applicable to authorship networks because most scientific papers are multi-authored (Wuchty, Jones, & Uzzi, 2007) and in many cases, the co-authors contribute unequally (Waltman, 2012). Thus, the weight of the links in authorship networks can be considered as the level of contribution devoted by authors to their papers.

The contribution list of a paper is mostly implicit and hidden behind the paper. Although bylines of papers reveal the order of authors, pursuing harmonious unity between a byline and contribution list for publications is a difficult topic (He, Ding, & Yan, 2012). Some guidelines and standards were suggested by organisations to regulate authorship, such as the "Ethical Guidelines for Journal Publication" and the "Criteria of the International Committee of Medical Journal Editors", but co-author positions in a byline were not defined (Zbar & Frank, 2011). In addition, much scientific research has tried to address this issue by proposing different methods to fairly assign and allocate credit to each author (Rahman, Mac Regenstein, Kassim, & Haque, 2017; Xu, Ding, Song, & Chambers, 2016). Given the fact that journals employ different types of authorship order for bylines, no existing guidelines or methods can, in practice, provide an universal procedure to determine the contribution of co-authors (Yang, Wolfram, & Wang, 2017). Nevertheless, three routine authorship credit counting methods have been widely discussed and applied, namely, full (or inflated) counting, fractional counting and authorship-weighted counting (Vavryčuk, 2018). The first two methods have proven to generate biased results (Hagen, 2008) while the authorship-weighted counting method attempts to distribute credit to the authors more appropriately than the others (Vavryčuk, 2018).

In this study, one of the authorship-weighted counting methods, harmonic credit counting (Hodge, Greenberg, & Challice, 1981), is adopted to quantify authors' contribution to each paper. The harmonic counting method proposes to allocate the



Paper Information

Paper ID	Paper Title
A	On D-branes from Gauged Linear Sigma Models
B	Boundary Fermions, Coherent Sheaves and D-branes on Calabi-Yau manifolds
C	Derived Categories and Zero-Brane Stability
D	On Heterotic Orbifolds, M Theory and Type I' Brane Engineering
E	Massless Black Holes and Conifolds in String Theory
F	Geometric Singularities and Enhanced Gauge Symmetries

Paper Citation Network

Paper Citation Weighting

Citation	Semantic-based Weight	Network-based Weight	Final Weight
A→F	0.626	0.353	0.617
B→E	0.174	0.353	0.236
C→A	0.206	0.316	0.235
C→B	0.181	0.316	0.220
C→E	0.146	0.670	0.569
C→F	0.215	0.670	0.549
D→E	0.282	0.577	0.434
D→F	0.287	0.577	0.434
F→E	0.235	0.500	0.360

Fig. 4. An example of assigning weight to citation network.

ratio of credit to author r and $(r+1)$ at a fixed ratio $[(r+1):r]$, regardless of the total number of authors (Hagen, 2008). It calculates the contribution credit of co-authors according to the following equation:

$$AC_i = \frac{\frac{1}{i}}{[1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{N}]} \tag{8}$$

where AC_i denotes the authorship credit of author i, i the position of author i in the author list, and N the total number of authors of the paper.

The harmonic counting is employed in this study due to its relatively high accuracy and robustness in processing practical data (Hagen, 2009, 2013). In addition, its applicability has been verified by the empirical experiments conducted in a wide range of scientific fields, such as psychology, economics, marketing, medicine, chemistry, bioscience, aquaculture, physics, software engineering and information retrieval (Fernandes, 2014; Hagen, 2008, 2009; Kim & Diesner, 2014; Kim & Kim, 2015; Vinkler, 1993; Wren et al., 2007; Xu, Ding, & Malic, 2015).

Since the contribution of this study does not lie in authorship credit allocation, only a majority of co-author byline cases are considered. Specifically, the policies of the journals are first checked, and if co-authors are listed alphabetically, credit will be granted equally to them. Afterwards, co-authors' citation numbers are explored, and the co-author in the last byline position will be relocated to the first position when his or her citation number is found to far exceed the first authors. The harmonic counting method is then applied to calculate credit for each author, and the credit will be used as the weight in authorship networks. Finally, we update the binary paper-author network into a weighted network according to the author credits.

Note that the weighting scheme cannot be applied to paper-journal networks since there only exist two types of relationships in these networks, namely, publish and not publish. Therefore, in this study, we only consider the weighting scheme in citation network and paper-author network.

3.3. Ranking algorithm

A graph-based propagation algorithm is developed to calculate the authority scores of the papers in bibliometric networks by leveraging the heterogeneous nature of the layered networks. Inspired by the HITS algorithm (Kleinberg, 1999), the iteration lies in the interaction between the layered networks, which enables the paper nodes (authorities) to obtain scores from the author nodes and journal nodes (hubs). Meanwhile, the hubs accumulate scores by achievement from the authorities. Furthermore, the HITS algorithm is improved by taking the proposed weighting scheme into account in each iteration process, and the improvement will be discussed in the following hub score and authority score calculation equations. The overall propagation algorithm is operated through the following steps:

1 Initialise the authority score of each paper equally as $1/N_p$, where N_p refers to the number of all the papers in the collection.
2 Calculate the hub scores for authors, journals and papers in paper-author, paper-journal and citation networks, respectively, by collecting authority scores from the corresponding papers.
3 Update the authority score of each paper using the paper authority scores and hub scores of all the papers, authors and journals/conferences, as well as publication time.
4 Repeat steps 2 and 3 until convergence is achieved.

3.3.1. Hub scores

Hub scores represent the quality of the hubs (authors, journals and papers) in the bibliometric networks. In this algorithm, the hub scores are measured by integrating the authority scores of the corresponding papers, and then normalised by the sum of the scores contributed from each corresponding node.

In the paper-author network, the hub score of author A_i is defined as:

$$HS(A_i) = \frac{\sum_{P_j \in Out(A_i)} w(A_i, P_j) AS(P_j)}{\sum_{P_j \in Out(A_i)} w(A_i, P_j)} \tag{9}$$

where $AS(P_j)$ is the authority score of paper P_j, $Out(A_i)$ denotes all the paper nodes that author A_i points to (i.e. the papers that are written by author A_i), $w(A_i, P_j)$ represents the weight of the links between author A_i and paper P_j in the paper-author network.

Similarly, in the paper-journal network, the hub score of journal J_i is defined as:

$$HS(J_i) = \sum_{P_j \in Out(J_i)} \frac{AS(P_j)}{|Out(J_i)|} \tag{10}$$

where $Out(J_i)$ denotes all the papers published in journal J_i. Note that the paper-journal network is a binary network with no weight in the equation.

In the citation network, the hub score of paper P_i is defined as:

$$HS(P_i) = \frac{\sum_{P_j \in Out(P_i)} w_{i,j} AS(P_j)}{\sum_{P_j \in Out(P_i)} w_{i,j}} \tag{11}$$

where $Out(P_i)$ denotes all the paper nodes to which paper P_i points (i.e., papers cited by P_i), $w_{i,j}$ represents the weight of link from node P_i to node P_j.

3.3.2. Authority scores

The authority score of a paper is affected by four entities, namely, the quality of the papers that link to it, the authors who wrote it, the journals that published it, and its publication time. The authority score can be updated by a linear combination of the four entities as following:

$$\begin{aligned} AS(P_i) = \ & \alpha \cdot PageRank'(P_i) \\ & + \beta \cdot Paper(P_i) \\ & + \gamma \cdot Author(P_i) \\ & + \delta \cdot Journal(P_i) \\ & + \theta \cdot Time(P_i) \\ & + (1 - \alpha - \beta - \gamma - \delta - \theta) \cdot \frac{1}{N_p} \end{aligned}$$

where α, β, γ, δ and θ are parameters of the algorithm. $(1-\alpha-\beta-\gamma-\delta-\theta)\frac{1}{N_p}$ denotes the probability of a random jump, where N_p is the number of all the papers in the collection. In the experiment, the probability of a random jump is set to 0.15, which means, $\alpha+\beta+\gamma+\delta+\theta=0.85$. Therefore there are five variations and the parameters are set to be optimal.

PageRank(P_i) refers to the authority score of paper P_i achieved from the paper citation network. Traditional PageRank algorithm calculates the scores of papers as:

$$PageRank(P_i) = \sum_{P_j \in In(P_i)} \frac{1}{|Out(P_j)|} AS(P_j) \tag{12}$$

where $In(P_i)$ and Out (P_j) denote the papers that cite and are cited by paper P_i, respectively. $AS(P_j)$ is authority score of paper P_j.

Considering the citation relevance, our algorithm needs to update PageRank by taking the weight of links in citation networks into account. The PageRank is updated as:

$$PageRank'(P_i) = \sum_{P_j \in In(P_i)} \frac{w_{i,j}}{\sum_{P_m \in Out(P_j)} w_{i,j}} AS(P_j) \tag{13}$$

where $w_{i,j}$ denotes the weight of link from paper P_i to paper P_j.

Paper(P_i) is the authority score of paper P_i collected from the hub cores of papers through the citation network, as follows,

$$Paper(P_i) = \frac{1}{Z(P)} \sum_{P_j \in In(P_i)} HS(P_j) w_{i,j} \tag{14}$$

where $HS(P_j)$ denotes the hub score of paper P_j, $In(P_i)$ denotes the papers that point to paper P_i (i.e., papers cite P_i), $w_{i,j}$ represents the weight of link from P_i to P_j, and $Z(P)$ is a normalised value, which is the sum of scores, transferred from the hub papers in the collection.

Similarly, *Author*(P_i) is the authority score of paper P_i propagated from the hub authors through the paper-author network:

$$Author(P_i) = \frac{1}{Z(A)} \sum_{A_j \in In(P_i)} HS(A_j) w(A_j, P_i) \tag{15}$$

where $HS(A_j)$ denotes the hub score of author A_j, $In(P_i)$ denotes the authors of the paper P_i, $w(A_j, P_i)$ represents the weight of the links between author A_j and paper P_i. $Z(A)$ is a normalised value, which is the sum of scores transferred from the hub authors.

Journal(P_i) is the authority score of paper P_i transmitted from the hub journals through paper-journal networks, as follows,

$$Journal(P_i) = \frac{1}{Z(J)} \sum_{J_j \in In(P_i)} HS(J_j) \tag{16}$$

where $HS(J_j)$ refers to the hub score of journal J_j, $In(P_i)$ denotes the journal of paper P_i, and $Z(J)$ is a normalised value, which is the sum of scores transferred from the journal hubs to all the papers.

Time(P_i) is a time-aware value for paper P_i that is used to balance the bias to earlier published papers. Specifically, new papers tend to be underestimated because previous algorithms mostly rank scientific papers based on citations even though they may hold great contributions and importance. Therefore, we follow the time-aware method of FutureRank (Sayyadi & Getoor, 2009) to promote the prestige of newly published papers as follows:

$$Time(P_i) = e^{-\rho \times (T_{Current} - T_{P_i})} \tag{17}$$

where ρ is a constant value assigned as 0.62 in this paper based on FutureRank (Sayyadi & Getoor, 2009). T_{P_i} refers to the publication time of paper P_i, $T_{Current}$ is the current time of evaluation, so $T_{Current} - T_{P_i}$ represents the number of years since the paper P_i was published. The sum of *Time*(P_i) scores for all the papers is normalised to 1.

For initialisation, the authority score of each paper is assigned as $1/N_p$. After each iteration, the sum of authority score of all the papers is kept to be 1. The iteration process ends when convergence is encountered. The rule for meeting the convergence requires the difference between the current authority score and last authority score is less than a given threshold which is set as 0.0001 in this paper. For the papers which do not cite any other articles, we assume that they have links to all the other papers, so that the sum of the authority scores of all the papers will remain at 1 in each iteration.

The integrated calculation procedure of the proposed W-Rank algorithm is listed in Algorithm 1.

Algorithm 1. W-Rank algorithm

```
Input     : citation network $C$, paper-author network $A$,
            paper-journal network $J$, publication time list $T$
Parameter: $\alpha, \beta, \gamma, \delta, \theta, \tau, \rho$
Output    : paper authority score $AS$
Steps     :
update citation network: $C_w \leftarrow C$
update paper-author network: $A_w \leftarrow A$
compute time score: $Time \leftarrow Normalize(exp(-\rho \times (\tau - T)))$
initialise paper authority score: $AS \leftarrow \{\frac{1}{N_p}, \frac{1}{N_p}, ..., \frac{1}{N_p}\}$, where $N_p$ is the
  number of papers in collection.
while not converging do
    update Pagerank': $PageRank' \leftarrow Pagerank'(C_w, AS)$
    update hub scores:
    $HS(A) \leftarrow GetHubScore(A_w, AS)$
    $HS(J) \leftarrow GetHubScore(J, AS)$
    $HS(P) \leftarrow GetHubScore(C_w, AS)$
    update authority scores:
    $Author \leftarrow GetAuthScore(A_w, HS(A))$
    $Journal \leftarrow GetAuthScore(J, HS(J))$
    $Paper \leftarrow GetAuthScore(C_w, HS(P))$
    update paper authority scores:
    $AS \leftarrow$
      $Integrate(\alpha PageRank', \beta Paper, \gamma Author, \delta Journal, \theta Time, \frac{1}{N_p})$
end
```

4. Experimental design

4.1. Dataset and pre-processing

ArXiv hep-th and Microsoft Academic Graph (MAG) are adopted for our experiments. These two datasets are chosen for three main reasons. Firstly, arXiv hep-th is one of the most famous datasets used for scholarly bibliometric analysis and research, which indicates its validity and richness of citations. Secondly, MAG is a recently published dataset and it contains bibliometric information that is up-to-date and huge size in every field of research. This dataset provides an opportunity to obtain recent data and extract data for any desired topic. Thirdly, applying and testing the proposed algorithm on two different datasets can examine the flexibility and robustness of the algorithm.

The public arXiv hep-th dataset was released for the KDD cup 2003 so it contains bibliometric data from year from 1992 to 2003. Approximately 29,000 papers and 350,000 citations are covered in this dataset predominantly in the research fields of physics and mathematics. In this dataset, complete papers, abstracts, papers' publication time and citation are included. Further information extraction is needed to extract authors and journals from the dataset. After mining the required data, 14909 authors and 428 journals are found and stored. Few mistakes are found during the information extraction due to style inconsistencies or typographical errors, so they are deleted from the dataset.

The MAG is selected as an additional dataset for the purpose of verifying the flexibility and robustness of the proposed algorithm. This dataset is obtained upon application with specific request that it can only be used for academic use. The obtained MAG dataset contains bibliometric data up to year 2017, and the data is stored in the format of JSON in TXT files with the size of 283 gigabytes (GB) in total. For the purposes of saving computing memories and fast calculation, we extract the bibliometric information of the papers that are in the research field of Intrusion Detection in Cyber Security. In this research field, 6428 papers, 94,887 citations, 18,890 authors and 6428 journals are collected from the MAG dataset.

After pre-processing the original datasets, five tables are set as follows:

1 PaperID_Title_Abstract. This table stores the title and abstract of each paper. It consists of three columns, namely the paper IDs, the corresponding titles, and the abstracts, respectively.
2 PaperID_Author. This table represents the author-paper network by two columns, namely the paper IDs and the author names of the corresponding papers.
3 PaperID_Journal. This table represents the journal-paper network by two columns, namely the paper IDs and the names of journals where the papers were published.
4 SourceID_TargetID. This table summarises the citation network, where each row represents a citation link from a source paper in the first column 'SourceID' to a target paper in the second column 'TargetID'.

5 PaperID_Time. This table stores the publication time of each paper. It consists of two columns, namely the paper IDs and the corresponding time.

In the proposed W-Rank, we take into consideration the citation relevance and author contribution using a weighting scheme. Therefore, the citation network and author-paper network, i.e., the table SourceID_TargetID and table PaperID_Author, will be further updated into weighted networks. The calculation of weight for the citation network and author-paper network is based on the link weighting scheme proposed in Section 3.2.

4.2. Evaluation procedures

In this paper, several evaluation experiments are conducted. We compare the proposed W-Rank algorithm with baselines including PageRank, FutureRank and PageRank+HITS (HITS for short) algorithms using a proposed ground truth, that is, weighted future citations. Moreover, Spearman's ranking correlation and receiver operating characteristic (ROC) curve are adopted to measure the accuracy of the algorithms. The following sections describe the evaluation procedures required in experiments.

4.2.1. Evaluation criterion

The evaluation criteria for academic publication rankings has always been a critical and controversial issue since it is fairly difficult to define the ground truth. PageRank has been used as the evaluation criterion (Sayyadi & Getoor, 2009). However, using PageRank as the ground truth to evaluate and compare different ranking algorithms is inherently controversial, because PageRank itself is a ranking algorithm which is biased to earlier papers. An evaluation criterion based on future citation numbers is proposed (Wang et al., 2013). This criterion is based on fact instead of any algorithm, therefore, is more objective than criterion based on PageRank. On top of the future citation numbers, we define the "weighted future citation" evaluation criterion, as in Eq. (18), which is an improved version of the future citations by considering citation relevance.

$$S(P_i) = \sum_{P_j \in In(P_i)} w(P_j, P_i) \tag{18}$$

where $In(P_i)$ refers to the papers that cites paper P_i, and $w(P_j, P_i)$ is the weight of the link between paper P_i and P_j based on the method proposed in Section 3.2.1. Defining the evaluation criterion is a part of the problem definition. In our study, we define the concept of Citation Relevance, and consider it as an important attribute to each citation. Therefore, this concept should be well-incorporated in the evaluation criteria. Using weighted citations as ground truth is important as it can help restrain the factors which could influence the judging criteria, such as malicious or meaningless citations, enabling the ground truth to be more reasonable and legitimate.

4.2.2. Baselines

The baselines used in experiments are three existing scientific article ranking algorithms including PageRank, FutureRank and HITS. PageRank is one of the early paper ranking algorithms which is widely used as a basic ranking algorithm as it deals well with obtaining scores based on paper citations. FutureRank integrates the entity of time with a paper ranking method for the first time in order to improve the bias from the time factor. HITS for the first time applies a propagation algorithm to heterogeneous bibliometric networks. We compare W-Rank with these baselines based on the proposed ground truth hoping to highlight our method and, in the meantime, estimate the validity of the proposed ground truth.

4.2.3. Spearman's ranking correlation

We firstly evaluate our W-Rank and the comparison methods by calculating the Spearman's ranking correlation between the estimated ranks and the ground truth proposed in Section 4.2.1. The Spearman's correlation between two ranks is defined by the following equation Myers, Well, and Lorch (2013).

$$\rho = \frac{\sum_i (R_1(P_i) - \bar{R_1})(R_2(P_i) - \bar{R_2})}{\sqrt{\sum_i (R_1(P_i) - \bar{R_1})^2 (R_2(P_i) - \bar{R_2})^2}} \tag{19}$$

where $R_1(P_i)$ refers to the position of paper P_i in the first rank list which is the result of the experiment conducted in the first period P_1, R_2 (P_i) is the position of the paper P_i in the second rank list conducted in period P_2. $\bar{R_1}$ and $\bar{R_2}$ are the average rank positions of all papers in the first and second rank list respectively.

The corresponding 0.95 confidence intervals (CI) are also calculated to show the differences between methods. The CI for Spearman correlation is calculated by Fisher transformation (Fisher, 1915) as follows.

$$CI = \tanh(\operatorname{arctanh} \rho \pm z_{\alpha/2} / \sqrt{n-3}) \tag{20}$$

where ρ is the estimated Spearman's correlation, n is the sample size which equals to 14521 in arXiv dataset and 2252 in MAG dataset, and $z_{\alpha/2} = 1.96$ is the two-tailed critical value of the standard normal distribution with $\alpha = 0.05$.

4.2.4. Receiver operating characteristic (ROC) curve

Receiver operating characteristic (ROC) curve is adopted for evaluation purposes. The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at different threshold settings (Fawcett, 2006). In this study, the TPR and FPR can be calculated based on article ranking algorithms and the ground truth previously proposed. By doing so, the ranking list of the ground truth is split into two sets of papers, based on which the papers in the estimated ranking lists obtained by the ranking algorithms will be judged and categorised into true positive (TP), false positive (FP), true negative (TN) and false negative (FN). The TPR and FPR are obtained according to the following formulas.

$$TPR = \frac{TP}{TP + FN} \tag{21}$$

$$FPR = \frac{FP}{FP + FN} \tag{22}$$

4.3. Experiments

Based on the proposed ground truth and baselines, three sets of experiments are conducted. Using both arXiv hep-th and MAG datasets, we evaluate the effects of different configurations, compare W-Rank with state-of-the-art methods using Spearman's ranking correlation and ROC curve analysis, and verify the robustness of each method.

4.3.1. Evaluating the effect of configurations

Configuration experiments include a series of test trials estimating various configurations of bibliometric networks, that is to say, combinations of different scholarly data in the networks are examined. The configurations consist of the permutation and combination between PageRank and the information of citation, authorship, journal and time. The aim of this experiment is to determine which type of information brings greater influence to the accuracy of ranking results.

Spearman's ranking correlation is adopted in the experiments to obtain the accuracy of each configuration. The weighting scheme is added to each configuration in order to observe the influence of weight to the ranking results.

4.3.2. Comparison with state-of-the-art methods

The comparison experiments are designed based on the approaches of Spearman's ranking correlation and receiver operating characteristic (ROC) curve respectively.

In the experiments based on Spearman's ranking correlation, the proposed W-Rank is compared with three well-known methods including PageRank, FutureRank and HITS. These experiments are designed to prove the statistical advantage of W-Rank in terms of accuracy compared to the other methods. Four scientific article ranking methods have been tested based on two datasets including arXiv hep-th and MAG.

In addition, the approach of the receiver operating characteristic (ROC) curve is adopted to evaluate the accuracy of the estimated algorithms for comparison. Specifically, we set a threshold in order to split the ranking list of ground truth into two sets of papers and mark them as positive and negative so that a binary classification problem is initiated. Then the papers ranked by the four ranking algorithms will be judged and categorised into true positive (TP), false positive (FP), true negative (TN) or False Negative (FN). Based on the categorising results and formulas (21) and (22), a ROC curve is plotted to further compare the accuracy of the algorithms.

4.3.3. Robustness

The robustness of a method embodies the consistency of its performance. In this experiment, the robustness of four paper ranking methods is tested. We first set up a historical time point in datasets, separating the whole time slot into two historical periods. The first period before the historical time point is denoted as P_1 while the whole time period is referred to as P_2. Then we rank the papers in the historical period P_1 based on the bibliometric information of P_1 and P_2 respectively. Lastly, the correlation of the two sets of ranking scores is measured. The method obtaining better correlation result shows greater robustness, as it achieves paper rankings that vary less over time.

5. Results and discussion

5.1. Evaluating the effect of configurations

The results of evaluating the effect of different configurations are listed in Table 1. In this table, the configurations are tested as trials and listed in the left column, while the datasets and attribute of weight are set on the top two rows. *PR* refers to PageRank, *PA* denotes the paper-author network, *PJ* paper-journal network and *PC* paper citation network. The correlation results are listed accordingly in the body of the table.

According to the results, the best correlation results from both datasets are obtained by deploying all types of bibliometric data jointly. This means that W-Rank experimentally exceeds accuracy of the other configuration combinations. Observing each dataset, the weighted correlation results are higher than unweighted ones to varying degrees. This verifies that the proposed weighting scheme is able to promote the accuracy of scientific ranking algorithms. Compared across the datasets,

Table 1
Spearman's ranking correlation of different configurations.

Dataset	arXiv hep-th		MAG	
Config	Unweighted	Weighted	Unweighted	Weighted
PR	0.420	0.439	0.472	0.491
PR+PA	0.454	0.458	0.461	0.459
PR+PJ	0.521	0.529	0.440	0.448
PR+PC	0.520	0.562	0.467	0.464
PR+PA+PJ	0.520	0.530	0.453	0.476
PR+PC+PJ	0.527	0.538	0.467	0.479
PR+PA+PC	0.529	0.595	0.471	0.468
PR+PA+PC+PJ	0.541	0.582	0.503	0.557
PR+PA+PC+PJ+T	0.612	**0.682**	0.545	**0.627**

Table 2
Spearman's ranking correlation and the corresponding confidence intervals.

Dataset	NumCitation	PageRank	FutureRank	HITS	W-Rank
arXiv hep-th	0.517	0.420	0.572	0.612	**0.684**
0.95 CI	0.505, 0.528	0.407, 0.433	0.561, 0.582	0.602, 0.622	**0.676, 0.692**
MAG	0.523	0.472	0.441	0.546	**0.623**
0.95 CI	0.493, 0.552	0.440, 0.503	0.408, 0.473	0.517, 0.574	**0.598, 0.647**

the results from arXiv hep-th generally outnumber the ones from the MAG, except the PageRank method shows a rise in the MAG. The algorithms perform slightly better in the arXiv hep-th dataset because this dataset includes more papers with a larger number of citations and connections. Although the MAG dataset provides limited connections, PageRank still achieves correlation result higher than most of the configuration combinations. This means PageRank can be used as a fast and efficient article ranking method when resources are inadequate. In addition, a drop can be found in the configurations of $PR+PJ$, $PR+PA+PJ$ and $PR+PC+PJ$ from the weighted results but this situation is not clear in the unweighted ones. This demonstrates that the paper-journal network is bringing down the correlation as the weighting scheme is not used in the paper-journal network.

5.2. Comparison with state-of-the-art methods

5.2.1. Comparison based on Spearman's ranking correlation

In this experiment, we compare the proposed W-Rank with state-of-the-art methods in terms of the Spearman's ranking correlation. The corresponding 0.95 confidence intervals (CI) are also calculated to show the differences between methods. The results are summarised in Table 2.

According to the results, the proposed W-Rank algorithm outperforms the other ranking methods, which again confirms that integrating weight into bibliometric networks is beneficial for improving scientific article ranking. The confidence interval represents values for the population parameter for which the difference between the parameter and the observed estimate is not statistically significant at the 0.05 level. Therefore, it can be inferred that the Spearmans ranking correlation of W-Rank is significantly better than that of PageRank, FutureRank, HITS, and NumCitation. In addition, the W-Rank scores well in both the arXiv hep-th and the MAG dataset. This verifies not only its reliability in obtaining accurate ranking results but also its capability for handling different datasets regardless of dataset quality.

Another interesting finding is that the PageRank outperforms the FutureRank in the MAG dataset, although it performs poorly in the arXiv hep-th dataset. This is demonstrated more clearly in Fig. 5 where the horizontal axis lists several configuration combinations while the vertical axis shows the correlation scores. In this figure, PR, $PR+PA+T$, $PR+PA+PC+PJ+T$ and $PR+PA+PC+PJ+T+W$ refer to the PageRank, FutureRank, HITS and W-Rank respectively. The algorithms perform better in the arXiv hep-th dataset than in the MAG except a reverse situation occurs in the PageRank. It reveals the advantage of the PageRank in terms of relying less on resources. Moreover, the FutureRank also shows an obvious gap in the two datasets, which illustrates that the quality of bibliometric datasets could influence article ranking algorithms.

5.2.2. Comparison based on ROC curve

In addition to Spearman's ranking correlation, we also evaluate the performance of each ranking algorithm using ROC curve. We first order all the articles in the collection according to the evaluation criterion, that is, the weighted future citation. We then apply a moving threshold to have different numbers of highly ranked articles. With each threshold, the articles are divided into two classes, including highly ranked articles and lower ranked articles. These are considered to be positive and negative entities respectively. The true positive rate (TPR) represents the ratio of the number of true positives to the total number of positives. In other words, TPR reflects how many top ranked articles are correctly predicted to be highly ranked articles. The false positive rate (FPR) calculates the ratio of the number of false positives to the total number of negatives, in other words, FPR shows how many low ranked articles are incorrectly predicted as highly ranked articles. ROC

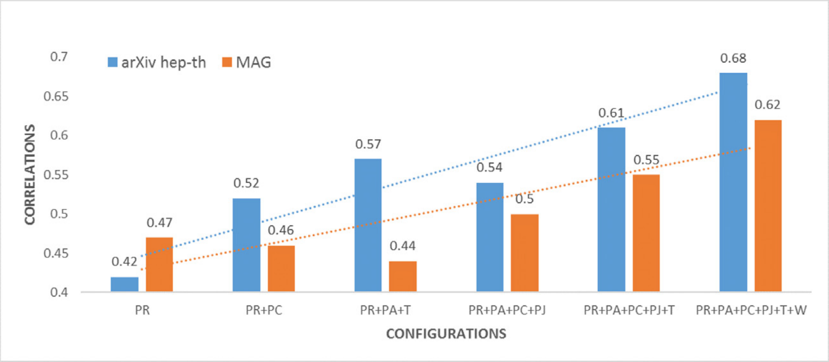
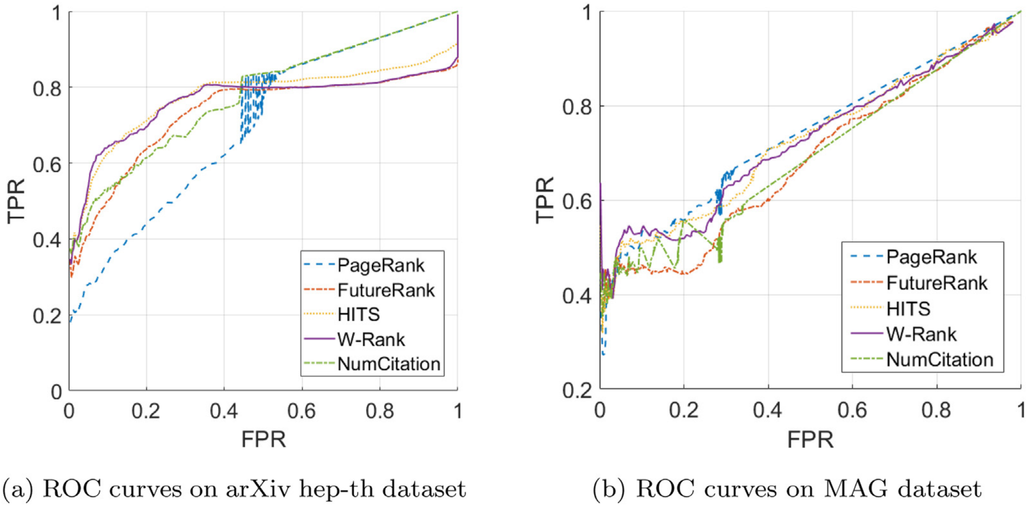


Fig. 5. Correlation trend chart of significant configurations.



(a) ROC curves on arXiv hep-th dataset

(b) ROC curves on MAG dataset

Fig. 6. ROC curves of different scientific article ranking algorithms on both datasets.

curve is defined as the TPR against the FPR at various threshold settings, therefore it reflects how the performance evolves. The performance of the different approaches (PageRank, FurtureRank, HITS, and the proposed W-Rank) and plain sum of citations (NumCitation) are compared by plotting the ROC curves. Both of the datasets are tested.

Fig. 6 shows the ROC curves of the PageRank, FutureRank, HITS, W-Rank and NumCitation using both the arXiv hep-th and the MAG dataset respectively, where the X axis refers to the FPR and the Y axis is the TPR. According to Fig. 6a, W-Rank and HITS outperform FutureRank and PageRank to different degrees when the FPR is less than 0.5, and the leading two curves are increasing at a similar pace. The W-Rank curve rises faster than the HITS curve during the period where the FPR goes from 0 to 0.2, which means W-Rank performs better when the error tolerance is rather limited. After that, the curves of W-Rank and HITS are rising in almost the same speed. The performance of FutureRank is similar to the plain sum of citations before the FPR reaches 0.5, and after that, the NumCitation rises at the same pace as PageRank. On the other hand, the curves on the MAG dataset present different styles compared to those on the arXiv hep-th dataset. The performance of all the algorithms drops when the FPR is between 0.1 and 0.3 except PageRank which shows a higher level than most of the other algorithms. This again verifies that the quality of datasets has less influence on PageRank algorithm compared to other more complicated ones. Moreover, the W-Rank curve still rises at the fastest pace at the beginning, which further confirms that the proposed W-Rank algorithm is more suitable when the tolerance for ranking errors is strict. The NumCitation curve shows an overall upward trend while a distinctive fluctuation occurs when the FPR is less than 0.4, which is possibly cause by the quality of the dataset.

Table 3 shows the values of area under curve (AUC) performance of the four algorithms and NumCitation on both datasets respectively. The AUC value is commonly used to represent the performance of a classifier (Fawcett, 2006). In this study, AUC

Table 3
The values of AUC performance.

Dataset	NumCitation	PageRank	FutureRank	HITS	W-Rank
arXiv hep-th	0.753	0.690	0.726	**0.771**	0.764
0.95 CI	0.747, 0.759	0.684, 0.696	0.720, 0.732	**0.766, 0.776**	0.759, 0.769
MAG	0.693	0.712	0.681	0.720	**0.733**
0.95 CI	0.678, 0.708	0.697, 0.727	0.666, 0.696	0.705, 0.735	**0.718, 0.748**

Table 4
Robustness.

Dataset/Method	PageRank	FutureRank	HITS	W-Rank
arXiv hep-th	0.892	0.778	0.977	0.982
MAG	0.733	0.585	0.721	0.779

value measures the performance of threshold that splits the ranking list of ground truth. Confidence intervals of AUC are also calculated to show the differences between methods using the Hanley's approach (Hanley & McNeil, 1982), as follows.

$$CI = AUC \pm z_{\alpha/2} \cdot se$$
$$se = \sqrt{\frac{q_0 + (n_1 - 1)\cdot q_1 + (n_2 - 1)\cdot q_2}{n_1 \cdot n_2}}$$

where $q_0 = AUC(1 - AUC)$, $q_1 = \frac{AUC}{2-AUC} - AUC^2$, and $q_2 = \frac{2AUC^2}{1+AUC} - AUC^2$. In our experiments, we calculate 0.95 CI, i.e., $\alpha = 0.05$, therefore $z_{\alpha/2} = 1.96$.

According to the results, on both datasets the W-Rank and HITS algorithms achieve similar AUC values, while the values obtained by the PageRank and FutureRank algorithms are lower overall, although the PageRank shows a rise on the MAG dataset. It seems that the predefined threshold benefits the NumCitation performance in the arXive hep-th dataset but does less in the MAG dataset. The confidence intervals indicate the AUC performance of W-Rank is significantly better than that of PageRank, FutureRank, and NumCitation, however, the difference between W-Rank and HITS is not significant.

5.3. Robustness

The results of robustness experiments for each paper ranking method are listed in Table 4.

According to the results, all the algorithms perform better using the arXiv hep-th dataset compared to the MAG dataset, which means a dataset with richer bibliometric information is more helpful for any algorithm to achieve better ranking results. On both datasets, the W-Rank and HITS show their outstanding advantage in terms of robustness. In addition, the PageRank retains its performance level that is higher than average on both datasets, which again proves its capability of handling various datasets.

6. Conclusion

Scientific article ranking has been studied to sort academic papers and highlight the well-established ones with high authority and influence. Although many ranking models and algorithms have achieved promising performance, the influence coming from the weight of the links in bibliometric networks has not been thoroughly explored. To remedy this gap, a new scientific article ranking algorithm is proposed in this study, which promotes existing ranking algorithms by taking link weight into account. A link weighting scheme is developed for this purpose to assign weight to the links of citation and authorship networks so that the traditional bibliometric networks are updated to weight directed graphs. Citation relevance and author contribution are quantified for assigning weight to the links. Combining a propagation algorithm, W-Rank is able to calculate scores for scientific papers based on a weighted bibliometric network.

The contribution of this study lies in both its conceptual framework and experimental results. The proposed link weighting scheme explains the practical meaning of the links in bibliometric networks, and further quantifies citation relevance for the first time using a combination of network-based and semantic-based methods. The results of our experiments also verify the proposed weighting scheme and W-Rank algorithm. Both arXiv hep-th and MAG datasets are adopted for these purposes, and the experimental results demonstrate the advantage of W-Rank in several perspectives compared to the other existing ranking algorithms (PageRank, FutureRank and HITS). Therefore, it is confirmed that the link weight of bibliometric networks plays a significant role in improving scientific article ranking algorithms.

Some limitations of this study are acknowledged. This study uses weighted future citation as an evaluation criterion, which is inspired by future citations Wang et al. (2013) and improved based on the same idea as the proposed link weighting scheme. As explained in Section 4.2.1, simply applying future citation as an evaluation criterion shows drawbacks, and integrating the weight to the future citation would address this issue according to the idea of this study. However, it seems the proposed evaluation criterion could be biased in favour of the W-Rank from a certain point of view. Therefore, in terms of future work, establishing universal evaluation criteria for evaluating scientific paper ranking results will be an attractive and critical topic for the community of scholarly management. In addition, the proposed W-Rank algorithm is rather complex and lacks transparency. It is difficult for users to understand how it accomplishes tasks in steps and generates the ranking results, which limits the W-Rank in terms of complexity and practicability. This also gives room for the proposed W-Rank algorithm to be further explored.

Author contribution

Yu Zhang: Conceived and designed the analysis, Collected the data, Performed the analysis, Wrote the paper.
Min Wang: Conceived and designed the analysis, Collected the data, Performed the analysis.
Florian Gottwalt: Contributed data or analysis tools.
Morteza Saberi: Conceived and designed the analysis, Other contribution.
Elizabeth Chang: Other contribution.

Appendix A. An example of calculating semantic similarity using ADW approach

An example of the calculation procedure of ADW approach for semantic similarity between two papers is illustrated in this appendix as follows. Note that in this example, due to the limitations of the visual graphics, only the titles of the two papers are used for demonstration purposes.

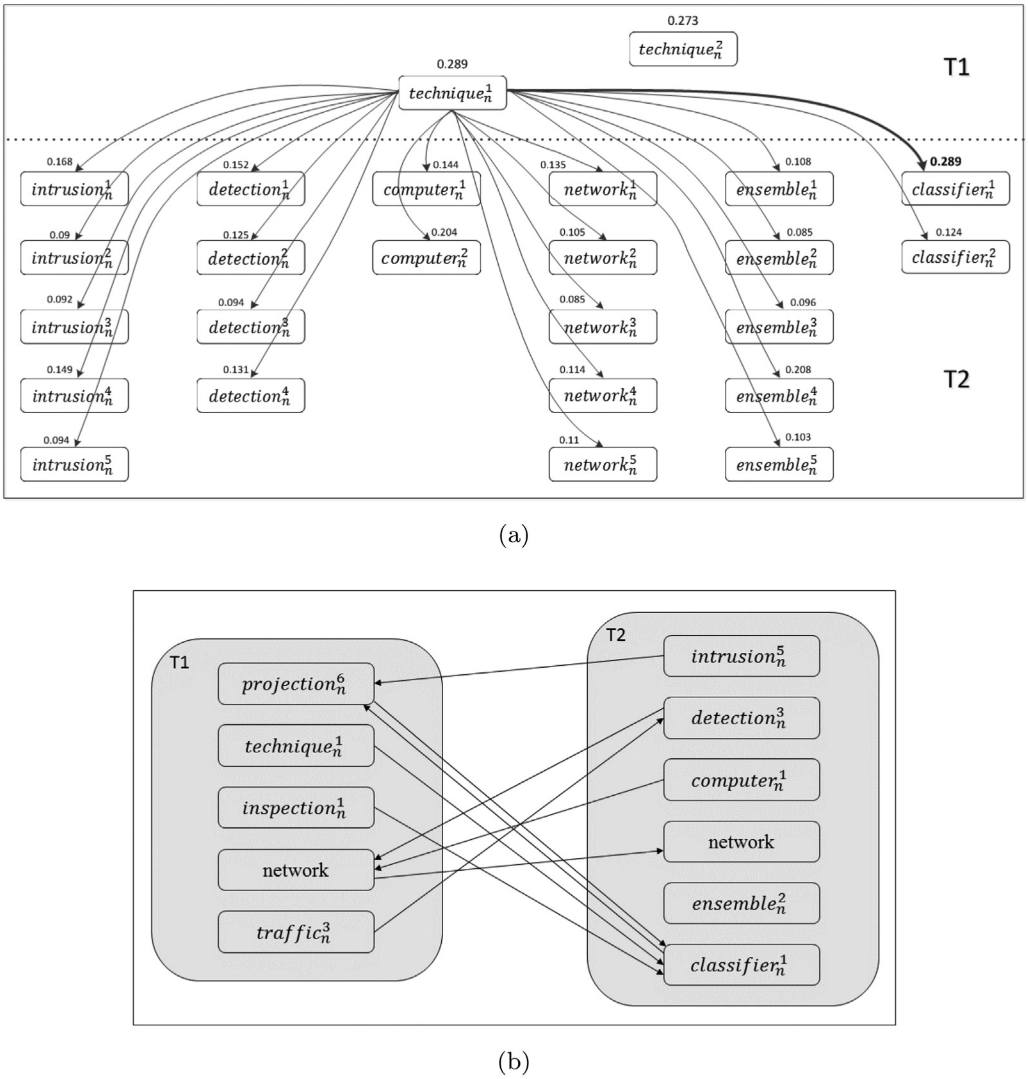


Fig. A1. (a) Example alignments of the sense of technique (in sentence T_1) to the senses of the words in sentence T_2, along with the similarity of the senses. (b) The alignments which maximize the similarities across words in T_1 and T_2.

The titles of the two papers are “Neural projection techniques for the visual inspection of network traffic” and “Intrusion detection in computer networks by a modular ensemble of one-class classifiers”, denoted as T_1 and T_2 respectively. After tokenisation, parsing and other word cleaning processing, words ‘projection’, ‘technique’, ‘inspection’, ‘network’ and ‘traffic’ are obtained from title T_1, while ‘intrusion’, ‘detection’, ‘computer’, ‘network’, ‘ensemble’, ‘classifier’ are returned from title T_2.

According to ADW, each word has one or more senses (semantic meanings) stored in WordNet. For example, the word ‘technique’ has two senses in WordNet, including ‘$technique_n^1$’ and ‘$technique_n^2$’.

ADW aligns each word in title T_1 to the words in title T_2 and find the maximal semantic similarity amongst these alignments. Fig. A1a shows the alignments between ‘$technique_n^1$’ to the senses of all the words in T_2, and the maximal similarity is the alignment between ‘$technique_n^1$’ and ‘$classifier_n^1$’ which is 0.289.

After aligning all the words between the title T_1 and T_2, the alignments which maximise the similarities across the words are found as shown in Fig. A1b. The alignments can be represented as:

$$\begin{aligned} P_{T1} &= \left\{projection_n^6, technique_n^1, inspection_n^1, network_n^1, traffic_n^3\right\} \\ P_{T2} &= \left\{intrusion_n^5, detection_n^1, computer_n^1, network_n^1, ensemble_n^2, classifier_n^1\right\} \end{aligned}$$

where vector P_x denotes the corresponding set of senses of sentence x.

In order to calculate the similarity between the title T_1 and T_2, the weighted overlap approach proposed by ADW is adopted. Let S denote the intersection of all senses with non-zero probability in both signatures and r_i^j denote the rank of sense $s_i \in S$ in signature j, where rank 1 denotes the highest rank. The similarity between the two sets of senses, i.e., the semantic similarity between the title T_1 and T_2, is defined as:

$$Sim(P_{T_1}, P_{T_2}) = \frac{\sum_{i=1}^{|S|}(r_i^1 + i_i^2)^{-1}}{\sum_{i=1}^{|S|}(2i)^{-1}} \tag{23}$$

where $\sum_{i=1}^{|S|}(2i)^{-1}$ is the maximum value set to bound the similarity value in [0, 1]. This maximum value occurs when each sense has the same rank in both sense sets.

The measure first sorts the two sense sets according to their values and then harmonically weights the overlaps between them. The minimum value is zero and occurs when there is no overlap between the two sense sets, i.e., $|S|= 0$. The measure is symmetric and satisfies the top-weightedness property, i.e., it penalizes the differences in the higher rankings more than it does for the lower ones. Note that r_i^1 is the rank of the sense s_i in the original vector P_{T_1} and not that in the corresponding vector truncated to the overlapping senses.

References

Amjad, T., Daud, A., Akram, A., & Muhammed, F. (2016). Impact of mutual influence while ranking authors in a co-authorship network. *Kuwait Journal of Science*, *43*(3).

Amjad, T., Ding, Y., Daud, A., Xu, J., & Malic, V. (2015). Topic-based heterogeneous rank. *Scientometrics*, *104*(1), 313–334.

Bergstrom, C. T., & West, J. D. (2008). Assessing citations with the eigenfactor metrics. *Neurology*, *71*(23), 1850–1851.

Bollen, J., Rodriquez, M. A., & Van de Sompel, H. (2006). Journal status. *Scientometrics*, *69*(3), 669–687.

Cargill, M., & O'Connor, P. (2013). *Writing scientific research articles: Strategy and steps*. John Wiley & Sons.

Chen, P., Xie, H., Maslov, S., & Redner, S. (2007). Finding scientific gems with googles pagerank algorithm. *Journal of Informetrics*, *1*(1), 8–15.

Daud, A. (2012). Using time topic modeling for semantics-based dynamic research interest finding. *Knowledge-Based Systems*, *26*, 154–163.

Ding, Y. (2011). Topic-based pagerank on author cocitation networks. *Journal of the Association for Information Science and Technology*, *62*(3), 449–466.

Ding, Y., & Cronin, B. (2011). Popular and/or prestigious? Measures of scholarly esteem. *Information processing & management*, *47*(1), 80–96.

Ennas, G., Biggio, B., & Di Guardo, M. C. (2015). Data-driven journal meta-ranking in business and management. *Scientometrics*, *105*(3), 1911–1929.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognition Letters*, *27*(8), 861–874.

Fernandes, J. M. (2014). Authorship trends in software engineering. *Scientometrics*, *101*(1), 257–271.

Fiala, D. (2012). Time-aware pagerank for bibliographic networks. *Journal of Informetrics*, *6*(3), 370–388.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, *10*(4), 507–521.

Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics*, *1*(4), 359–375.

Garfield, E. (1986). The 250 most-cited authors in the arts-and-humanities citation index, 1976–1983. *Current Contents*, (48), 3–10.

Garfield, E. (1965). *Can citation indexing be automated*. pp. 189–192. *Statistical association methods for mechanized documentation, symposium proceedings* (Vol. 269) Washington, DC: National Bureau of Standards, Miscellaneous Publication 269.

Gollapalli, S. D., Mitra, P., & Giles, C. L. (2011). Ranking authors in digital libraries. In *Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries* (pp. 251–254).

González-Pereira, B., Guerrero-Bote, V. P., & Moya-Anegón, F. (2010). A new approach to the metric of journals scientific prestige: The SJR indicator. *Journal of Informetrics*, *4*(3), 379–391.

Guerrero-Bote, V. P., & Moya-Anegón, F. (2012). A further step forward in measuring journals scientific prestige: The sjr2 indicator. *Journal of Informetrics*, *6*(4), 674–688.

Hagen, N. (2009). Harmonic publication and citation counting: Sharing authorship credit equitably-not equally, geometrically or arithmetically. *Scientometrics*, *84*(3), 785–793.

Hagen, N. T. (2008). Harmonic allocation of authorship credit: Source-level correction of bibliometric bias assures accurate publication and citation analysis. *PLoS One*, *3*(12), e4021.

Hagen, N. T. (2013). Harmonic coauthor credit: A parsimonious quantification of the byline hierarchy. *Journal of Informetrics*, *7*(4), 784–791.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), 29–36.

He, B., Ding, Y., & Yan, E. (2012). Mining patterns of author orders in scientific publications. *Journal of Informetrics*, *6*(3), 359–367.

Hodge, S. E., Greenberg, D. A., & Challice, C. (1981). Publication credit. *Science*, *213*, 950.

Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des alpes et des jura. *Bull Soc Vaudoise Sci Nat*, *37*, 547–579.

Kim, J., & Diesner, J. (2014). A network-based approach to coauthorship credit allocation. *Scientometrics*, *101*(1), 587–602.

Kim, J., & Kim, J. (2015). Rethinking the comparison of coauthorship credit allocation schemes. *Journal of Informetrics*, *9*(3), 667–673.

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, *46*(5), 604–632.

Li, Y., Luo, P., & Wu, C. (2014). *A new network node similarity measure method and its applications*. , arXiv preprint arXiv:1403.4303.

Liu, X., Bollen, J., Nelson, M. L., & Van de Sompel, H. (2005). Co-authorship networks in the digital library research community. *Information Processing and Management*, *41*(6), 1462–1480.

Ma, N., Guan, J., & Zhao, Y. (2008). Bringing pagerank to the citation analysis. *Information Processing and Management*, *44*(2), 800–810.

Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of informetrics*, *4*(3), 265–277.

Myers, J. L., Well, A. D., & Lorch, R. F., Jr. (2013). *Research design and statistical analysis*. Routledge.

Nykl, M., Campr, M., & Ježek, K. (2015). Author ranking based on personalized pagerank. *Journal of Informetrics*, *9*(4), 777–799.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The pagerank citation ranking: Bringing order to the web. Technical report*. Stanford InfoLab.

Persson, O. (2010). Identifying research themes with weighted direct citation links. *Journal of Informetrics*, *4*(3), 415–422.

Pilehvar, M. T., Jurgens, D., & Navigli, R. (2013). Align, disambiguate and walk: A unified approach for measuring semantic similarity. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers), vol. 1*, 1341–1351.

Rahman, M. T., Mac Regenstein, J., Kassim, N. L. A., & Haque, N. (2017). The need to quantify authors relative intellectual contributions in a multi-author paper. *Journal of Informetrics*, *11*(1), 275–281.

Salton, G. (1970). Automatic text analysis. *Science*, *168*(3929), 335–343.

Sayyadi, H., & Getoor, L. (2009). Futurerank: Ranking scientific articles by predicting their future pagerank. In *Proceedings of the 2009 SIAM International Conference on Data Mining* (pp. 533–544).

Small, H. (1997). Update on science mapping: Creating large document spaces. *Scientometrics*, *38*(2), 275–293.

Steinert, L., & Hoppe, H. U. (2016). A comparative analysis of network-based similarity measures for scientific paper recommendations. In *2016 Third European Network Intelligence Conference (ENIC)* (pp. 17–24).

Tang, J., Jin, R., & Zhang, J. (2008). A topic modeling approach and its integration into the random walk framework for academic search. In *Eighth IEEE International Conference on Data Mining, 2008, ICDM'08* (pp. 1055–1060).

Tang, J., Zhang, J., Jin, R., Yang, Z., Cai, K., Zhang, L., & Su, Z. (2011). Topic level expertise search over heterogeneous networks. *Machine Learning*, *82*(2), 211–237.

Vavryčuk, V. (2018). Fair ranking of researchers and research teams. *PLoS One*, *13*(4), e0195509.

Vinkler, P. (1993). Research contribution, authorship and team cooperativeness. *Scientometrics*, *26*(1), 213–230.

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.

Waltman, L. (2012). An empirical analysis of the use of alphabetical authorship in scientific publishing. *Journal of Informetrics*, *6*(4), 700–711.

Waltman, L., van Eck, N. J., van Leeuwen, T. N., & Visser, M. S. (2013). Some modifications to the snip journal impact indicator. *Journal of informetrics*, *7*(2), 272–285.

Wang, Y., Tong, Y., & Zeng, M. (2013). Ranking scientific articles by exploiting citations, authors, journals, and time information. *Twenty-seventh AAAI Conference on Artificial Intelligence*.

Weissberg, R., & Buker, S. (1990). *Writing up research*. Englewood Cliffs, NJ: Prentice Hall.

Wren, J. D., Kozak, K. Z., Johnson, K. R., Deakyne, S. J., Schilling, L. M., & Dellavalle, R. P. (2007). The write position: A survey of perceived contributions to papers based on byline position and number of authors. *EMBO Reports*, *8*(11), 988–991.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036–1039.

Xu, H., Martin, E., & Mahidadia, A. (2014). Contents and time sensitive document ranking of scientific literature. *Journal of Informetrics*, *8*(3), 546–561.

Xu, J., Ding, Y., & Malic, V. (2015). Author credit for transdisciplinary collaboration. *PLoS One*, *10*(9), e0137968.

Xu, J., Ding, Y., Song, M., & Chambers, T. (2016). Author credit-assignment schemas: A comparison and analysis. *Journal of the Association for Information Science and Technology*, *67*(8), 1973–1989.

Yan, E., & Ding, Y. (2010). Weighted citation: An indicator of an article's prestige. *Journal of the Association for Information Science and Technology*, *61*(8), 1635–1643.

Yan, E., Ding, Y., & Sugimoto, C. R. (2011). P-rank: An indicator measuring prestige in heterogeneous scholarly networks. *Journal of the Association for Information Science and Technology*, *62*(3), 467–477.

Yang, S., Wolfram, D., & Wang, F. (2017). The relationship between the author byline and contribution lists: A comparison of three general medical journals. *Scientometrics*, *110*(3), 1273–1296.

Yu, D., Wang, W., Zhang, S., Zhang, W., & Liu, R. (2017). A multiple-link, mutually reinforced journal-ranking model to measure the prestige of journals. *Scientometrics*, *111*(1), 521–542.

Yu, P. S., Li, X., & Liu, B. (2004). On the temporal dimension of search. In *Proceedings of the 13th International World Wide Web Conference on Alternate Track Papers & Posters* (pp. 448–449).

Zbar, A., & Frank, E. (2011). Significance of authorship position: An open-ended international assessment. *The American Journal of the Medical Sciences*, *341*(2), 106–109.

Zhou, D., Orshanskiy, S. A., Zha, H., & Giles, C. L. (2007). Co-ranking authors and documents in a heterogeneous network. In *Seventh IEEE International Conference on Data Mining, 2007, ICDM 2007* (pp. 739–744).