Hindawi
Complexity
Volume 2018, Article ID 4697485, 16 pages
https://doi.org/10.1155/2018/4697485

Research Article

AIRank: Author Impact Ranking through Positions in Collaboration Networks

Jun Zhang,[1] Yan Hu,[1] Zhaolong Ning,[1] Amr Tolba,[2,3] Elsayed Elashkar,[4,5] and Feng Xia[1]

[1] *Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, School of Software, Dalian University of Technology, China*
[2] *Computer Science Department, Community College, King Saud University, Riyadh 11437, Saudi Arabia*
[3] *Mathematics and Computer Science Department, Faculty of Science, Menoufia University, Shebin-El-Kom 32511, Egypt*
[4] *Administrative Sciences Department, Community College, King Saud University, Riyadh 11437, Saudi Arabia*
[5] *Applied Statistics Department, Faculty of Commerce, Mansoura University, Mansoura 35516, Egypt*

Correspondence should be addressed to Yan Hu; wohuyan@gmail.com

Received 30 November 2017; Revised 16 March 2018; Accepted 12 April 2018; Published 11 June 2018

Academic Editor: Xiuzhen Zhang

Citation is a universally acknowledged way for scientific impact evaluation. However, due to its easy manipulability, simply relying on citation cannot objectively reflect the actual impact of scholars. Instead of citation, we utilize the academic networks, in virtue of their available and abundant academic information, to evaluate the scientific impact of scholars in this paper. Through the collaboration among scholars in academic networks, we notice an interesting phenomenon that scholars in some special positions can access more kinds of information and connect researchers from different groups to promote the scientific collaborations. However, this important fact is generally ignored by the existing approaches. Motivated by the observations above, we propose the novel method AIRank to evaluate the scientific impact of scholars. Our method not only considers the impact of scholars through the mutual reinforcement process in heterogeneous academic networks, but also integrates the structural holes theory and information entropy theory to depict the benefit that scholars obtain via their positions in the network. The experimental results demonstrate the effectiveness of AIRank in evaluating the impact of scholars more comprehensively and finding more top ranking scholars with interdisciplinary nature.

1. Introduction

The development of modern research technologies allows researchers to get access to the plentiful scholarly data timely and facilitates the academic cooperation among scholars with diverse backgrounds. The easy access to the various scholarly data and the diverse data analysis technologies make researchers conduct their work more efficiently [1, 2]. However, due to the large volume of scholarly data, it is time-consuming to filter the influential and related scholars or references from the massive data. The evaluation of scientific impact not only sheds light on the above problem, but also provides basis for academic awards applications, faculty employments, fund decisions, etc. [3]. Therefore, evaluating the scientific impact is of great significance, and our primary concern is on measuring the impact of scholars in this paper.

The existing evaluation methods generally prefer using the qualities and quantities of scholars' papers to measure the scientific impact. For a long time, citation has been widely used to gauge the influence of scholars and articles, such as h-index [4], g-index [5], and the journal impact factor [6]. However, some crucial shortcomings exist with such approaches that heavily rely on citation counts. The first problem is that the accumulation process of citation counts is involved with time. Therefore, previously published papers obviously have the advantage of having longer time cited by other literature than newly published papers. Another existing problem is that the citation counts can be easily manipulated through self-citations or citations via acquaintanceships. As a consequence, citation counts cannot accurately reflect the qualities of scholarly articles to some extent.

Apart from the citation-based methods, researchers also utilize the academic networks to measure the scientific impact. Typical academic networks include various kinds of entities and relationships, such as papers, authors, venues, citation relationship, and coauthorship. Therefore, by considering the above-mentioned attributes of heterogeneous networks, it is obvious that using heterogeneous network topology [7] to depict the academic networks is more suitable than applying homogeneous network topology. The PageRank [8] and HITS algorithms [9] are the most commonly used ones to rank the importance of scholarly entities in academic networks. 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 in academic networks [10].

Academic networks have been widely employed for scientific impact evaluation in the above-mentioned network-based methods. It not only provides plentiful information about scholarly entities, but also explicitly indicates relationships among them [11]. Under the coauthor network structure, we find that scholars that possess some special positions can access diverse information from various kinds of scholars and act as bridges that connect different groups of scholars. These scholars can benefit from the various information, and consequently their research capacities can be improved. In addition to the gains that these positions bring to the scholars themselves, they also accelerate the dissemination of knowledge among scholars in different fields. Simultaneously, the communications between scholars also promote the interdisciplinary collaborations and, furthermore, propel the development of science. Therefore, the effect of scholars' positions is of great significance for the evaluation of scholars' impact.

Although current works have proposed many solutions on evaluating the scientific impact, they mainly ignore the vital effect of scholars' positions on their impact. In this paper, we propose the AIRank to evaluate scholars' impact. In order to measure the overall scientific impact of scholars, our method considers the scholar's impact in heterogeneous academic networks through the mutual influence mechanism among academic entities and combines this with the effects of scholars' positions in the network.

To investigate the effects of scholars' positions in the network on their impact, we look into this question from the angle of sociology. In sociology, the structural holes theory [12] indicates that the positions of individuals in the networks are closely related to their benefits. The structural holes theory suggests that individuals can access richer information and let the disconnected people know each other through them if they are in the positions that act as bridges between different groups of individuals. Figure 1 shows an illustration of the structural holes theory; the nodes represent scholars from different domains in computer science area. It is obvious that the red node in the center can connect and cooperate with scholars from different domains. Therefore, when facing problems, researchers can apply ideas and techniques obtained from other groups to solve them if they span structural holes. Several studies have indicated that the social success is positively correlated with the structural holes [13]. Thus we apply the structural holes theory to depict the importance of scholars' positions and their abilities on both accessing rich information and connecting different researchers.

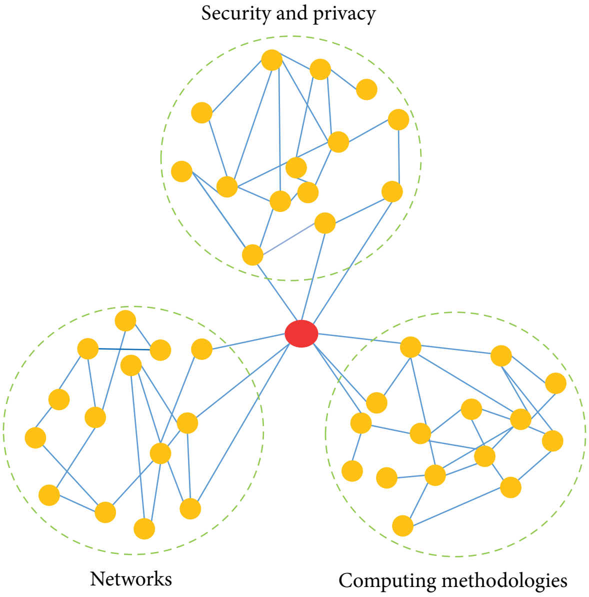


Figure 1: Illustration of structural holes.

To explore the diversity of information that researchers obtained, we solve this issue by considering the diverse backgrounds of coauthors. Researchers can directly acquire information or ideas from their coauthors due to the close cooperation in publishing scholarly articles. Hence the varieties of coauthors' backgrounds can indicate researchers' abilities to acquire diverse information. Besides acquiring information through the direct connections with coauthors, another way is the attendance of academic activities. Researchers can encounter other scholars and may further establish cooperation relationship through attending academic activities [14]. Scholars publishing articles in conferences have the opportunities to make acquaintance with other people through the attendance. Therefore, the quantities and qualities of articles published in conferences can represent the diverse information researchers acquire to some extent, and we utilize them to represent the diversity of information that researchers can acquire.

Generally speaking, we make the following contributions in this paper.

(i) *New Insight into Scientific Impact Evaluation.* We creatively provide a new solution to solve the impact evaluation issues from the angle of scholars' network positions for the first time, to the best of our knowledge.

(ii) *Novel Features for Evaluating Scholars.* We present three new indicators through utilizing the structural holes theory and information entropy theory to depict the effects of scholars' positions in collaboration networks and furthermore integrate the interplay among diverse scholarly entities in heterogeneous academic networks together to quantify scholars' scientific impact.

(iii) *Effectiveness in Identifying Outstanding Interdisciplinary Scholars.* The experiments on real datasets verify the significant role of scholars' positions in their impact, and our method outperforms the state-of-the-art methods in evaluating scholars' impact more comprehensively and identifying more outstanding interdisciplinary scholars.

The rest of the paper is organized as follows. Related work is discussed in the next section. Section 3 formulates the studied problem of scholar's scientific impact evaluation. Section 4 introduces our proposed method. Section 5 presents the experimental results of our method, followed by a section dedicated to the conclusion.

2. Related Work

The problem of scientific impact evaluation has been studied for a long time and became a popular and significant research direction [15–17]. The evaluation of scientific impact can assist scholars in diffusing their work and maximizing the academic influence [18, 19]. Generally, there are two major kinds of methods for measuring scholars' scientific impact, i.e., citation-based methods and network-based methods. In this section, we survey the existing literature in the above areas, respectively.

2.1. Citation-Based Methods. The achievements of scholars are often represented by their articles; therefore, the qualities of articles are usually used to measure the scientific impact of scholars. To measure the qualities of articles, the citation counts are one of the most widely used indicators. A series of metrics has been put forward to measure the scientific impact according to citations. Initially, the journal impact factor is proposed for evaluating the quality of journals [6]. Continually, the h-index [4] is proposed to measure scholar's impact by considering the productivity and the quality of their research work. Moreover, Pan and Fortunato [20] proposed the AIF to depict the dynamics of scholars' impact by considering the ever-increasing characteristic of h-index. These works all successfully depict the scientific impact and are commonly used due to the uncomplicated calculation process.

However, there exist critical shortcomings of using citation counts to evaluate the impact of scholars. The first problem is citation counts aggregate with time. Therefore, it is obvious that 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 scientific impact is unfair for young researchers comparing to senior researchers. Considering the above facts, researchers have proposed several methods to alleviate the effects of publishing time [21]. In addition, citations take time to happen; therefore, it cannot reflect the current impact of scholars timely.

Another problem existing in citation counts metrics is that citation counts can be distorted by self-citations or citations from colleagues, etc. Therefore, some researchers argue that the diverse citations should be considered disparately instead of regarding them equally [22]. Motivated by this observation, scholars have proposed diverse methods to differentiate the importance of citations. Valenzuela et al. [23] determined the significance of citations based on their appearing sections. Bai et al. [24] proposed a COIRank method to distinguish the conflict of interest citation relationship when measuring the impacts of articles. Other researches considered different aspects, such as citation distribution and coercive induced self-citation, to assess the qualities of citations.

2.2. Network-Based Methods. Considering the drawbacks of citation-based metrics, another way of measuring the impact of scholars is the network-based methods. Typically, the academic networks contain several main entities and relationships, e.g., articles, authors, venues, citing relationship, and coauthorship. Researchers have proposed a variety of ranking algorithms to gauge scholars' impact based on academic networks [25, 26].

A series of network-based methods has been proposed through calculating the degrees of scholars in academic networks by different methods to measure the impact of scholars. For instance, degree centrality, closeness centrality, Katz-Bonacich centrality, and eigenvector centrality are the commonly used measures to calculate the degrees of scholars based on different network structures [27, 28]. In addition, due to the merits of different measurements, researchers also integrate them together to quantify the scientific impact [29, 30].

Except for the above-mentioned centrality measurements, researchers also apply the commonly known ranking algorithms, i.e., the PageRank algorithm and HITS algorithm, to evaluate the scientific impact of scholars [31]. Previous researches utilize the PageRank and HITS algorithms to quantify the impact of scholars in homogeneous network. While the real academic networks contain various kinds of entities and links, diverse evaluation metrics have been proposed using different heterogeneous academic networks because of their topological merits. Figure 2 shows an illustration of a heterogeneous academic network; articles can be linked through citation relationship; authors can be linked to the articles they write; articles can be linked to the venues they published on; and authors can be related through the coauthorship.

Based on the above-mentioned heterogeneous academic network structures, researchers have proposed a series of the PageRank and HITS algorithms based methods to evaluate the impact of scholars. Considering the various kinds of relations that might exist among different entities, researchers have constructed distinct academic networks that contain novel relationships to measure the impact of scholars. A major kind of network-based methods is extending the original PageRank and HITS algorithms, which primarily focus on exploring new weights of the entities and links in the networks by considering the diverse importance of

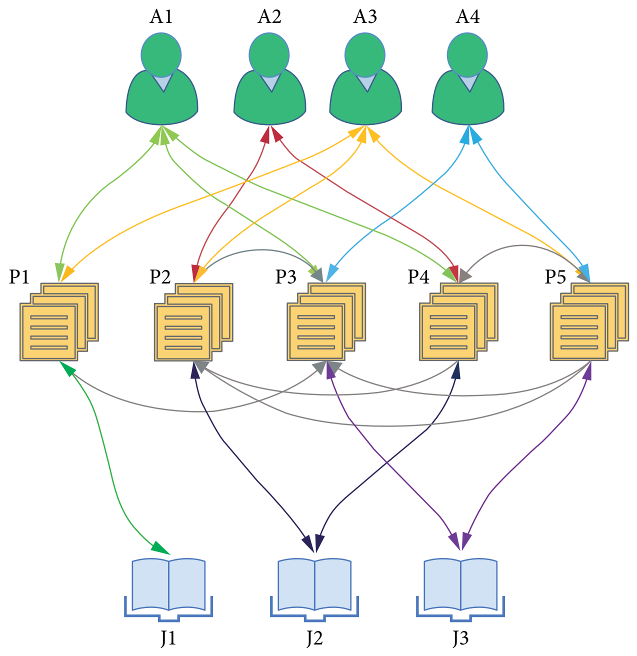


Figure 2: Illustration of a heterogeneous academic network, where P1 to P5 represent the papers, A1 to A4 indicate their corresponding authors, and J1, J2, and J3 represent the venues.

them. There also exist some works that utilize the PageRank algorithm and the HITS algorithm in the meantime to find the more appropriate one [32]. Instead of applying single kind of algorithm, researchers also combine the PageRank and HITS algorithm together to measure the scientific impact in order to utilize the advantages of both algorithms.

The primary mechanism of PageRank and HITS algorithm is that nodes would have higher influence value if the nodes that point to them are influential through the iterative process. Therefore, there exist mutual effects among the entities in the networks through the links. For instance, papers would become influential if they are cited by other articles with high qualities in the citation network, while the corresponding authors would be ranked high in the paper-author network, respectively. Several studies have been carried out on jointly evaluating the impact of scholars, articles, and venues according to specific academic networks. Based on the journals' impact, Nykl et al. [10] proposed an author ranking system through utilizing the PageRank algorithm. Considering the diverse research topics, Amjad et al. [33] measured the impact of scholarly entities by the topic-based heterogeneous rank in academic networks.

In addition, researchers also combine the citation and network-based evaluation metrics together to measure the impact of scholars because using single type of indicators is unable to capture the impact of scholars comprehensively. Wang et al. [34] explored the effect of citations, time information, and the combination of PageRank and HITS algorithm to quantify the scientific impact of scholars. Furthermore, Wang et al. [26] proposed the MRCoRank, which integrates the text features and HITS algorithm to determine the impact of scholars.

However, one important fact has been ignored by the existing approaches, that is, the effects of scholars' positions in the network and their abilities to acquire multiplicities of information via the existing relationships on their own impact. It is universally acknowledged that scholarly articles commonly represent the cooperation achievements of several coauthors; therefore, scholars can be influenced through the coauthorship. Although some researchers have investigated that scholars' impact can be affected by their coauthors' abilities [35, 36], no prior work exists to explore the influence of researchers' positions in the network and their capacities of obtaining diverse information on evaluating the scientific impact.

3. Problem Formulation

Generally, the task of scientific impact evaluation is formulated as statistical analysis problems or importance ranking algorithms. However, such existing approaches tend to evaluate scholars within the same disciplines and may be incapable of capturing the increasing interdisciplinary collaborations among researchers. Meanwhile, some scholars have noticed that the interactions among researchers can promote the quality and quantity of scientific achievements. Inspired by this interesting phenomenon, we propose a novel method which can identify influential scholars with interdisciplinary nature, thus formulating the following task: given the detailed information of scholars' publications, we evaluate the scientific impact of scholars with our proposed indicators implying their interdisciplinary collaborations in heterogeneous academic networks.

To solve our task, we decompose it into three subtasks. We first extract the coauthor network according to the information of scholars' publications. Let $G_c(V_{a_i}, E_{a_{ij}})$ denote the coauthor network, where V_{a_i} represents the node, and $E_{a_{ij}}$ exists if a_i has cooperated with a_j. Under the coauthor network, we then define and calculate several indicators $\{x_1, x_2, x_3, \ldots, x_n\}$ of scholars. Based on the above analysis, the first subtask can be formalized as follows: given that an undirected graph $G_c(V_{a_i}, E_{a_{ij}})$ represents the cooperation relationships among researchers and given a set of factors $\{x_1, x_2, x_3, \ldots, x_n\}$ of scholars, a function $f(a_i)$ that calculates the benefits of scholars through their positions in the network can be obtained.

Considering the overall task is to quantify scholars' scientific impact, we then compute the importance of scholars in heterogeneous academic networks. In order to fulfil this subtask, three academic networks need to be built. Let $G_{\mathrm{cit}}(V_{p_i}, E_{p_{ij}})$ indicate the citation network, where V_{p_i} represents the node and $E_{p_{ij}}$ exists if p_i has cited p_j. From the citation network, the importance of scholars' corresponding papers can be obtained. Based on the values of papers, the importance degrees of corresponding venues and scholars can be calculated in paper-venue network $(G_{pv}(V_{p_i} \cup U_{v_j}, E_{p_i v_j}))$ and paper-author network $(G_{pa}(V_{p_i} \cup U_{a_j}, E_{p_i a_j}))$, respectively. V_{p_i} represents the paper, U_{v_j} is the publishing venue of papers, and $E_{p_i v_j}$ exists if p_i has been published on v_j. Similarly, U_{a_j} is the author of papers, and $E_{p_i a_j}$ exists if

p_i was written by a_j. In this part, we study scholars' importance in heterogeneous academic networks, formally defined as follows: given directed graphs $G_{\text{cit}}(V_{p_i}, E_{p_{ij}})$, $(G_{pv}(V_{p_i} \cup U_{v_j}, E_{p_i v_j}))$, and $(G_{pa}(V_{p_i} \cup U_{a_j}, E_{p_i a_j}))$ and a set of intermediate results $\{r_1, r_2, r_3, \ldots, r_n\}$ obtained from the above-mentioned networks, a function $g(a_i)$ that calculates the importance of scholarly entities in heterogeneous academic networks can be obtained.

Our main purpose is to gauge the scientific impact of scholars. According to the above-mentioned subtasks, the final scientific impact can be obtained and formalized as follows.

Input. This includes the results obtained from functions $f(a_i)$ and $g(a_i)$.

Output. This includes the overall scientific impact of scholars.

The scientific impact evaluation problem we solve in this paper is formulated to be distinct from the traditional problem of simply relying on citation counts or network-based evaluation metrics. We explore the effect of scholars' network positions on the scientific impact. The primary advantage of our formulation is transforming the complex problem into three subtasks with low computational complexity, so that the efficiency of our method can be improved.

4. Design of AIRank

In most previous works, scholars are evaluated in the same time interval and their academic ages are commonly ignored. However, it is unfair for young researchers to be evaluated in the same time period compared to senior researchers. As a consequence, we choose scholars with the same academic age for evaluation to alleviate the effects of different research lengths. The real academic networks include various kinds of entities and relationships; therefore, we employ the heterogeneous network topology to represent academic network in order to depict it more appropriately.

The structural holes theory can indicate scholars' abilities to connect different people; therefore we utilize it in our method to depict scholars' positions in the network. To capture the multiplicities of information that researchers acquire through their relationships with other people, we measure these multiplicities from two aspects, which are the diversity of their coauthors and the quantity and quality of academic conferences they attend. In addition, we also consider the mutual effects among different academic entities in the networks together to quantify scholars' scientific impact.

Our proposed method consists of three main steps, the architecture of which is shown in Figure 3. The first part is calculating scholar's structural index (SI) value which captures the effect of scholars' positions in the networks. Three factors are proposed and the structural holes theory is employed in SI. In addition, we also consider the impact of scholars in academic networks through our proposed network index (NI). We apply the PageRank and HITS algorithms to measure scholars' impact in the three constructed academic networks. Finally, considering the above two parts, the overall impact of scholars is calculated according to the final formula. The calculation procedure of our proposed AIRank is shown as follows.

Step 1. Calculate the value of SI, which consists of the three proposed indicators and will be introduced in detail in the following.

Step 2. Calculate the value of NI, which utilizes the PageRank and HITS algorithms together to measure the impact of scholars in the networks.

Step 3. Calculate scholar's final score according to the above two steps.

4.1. Calculation Procedure of SI. With the development of research techniques, researchers nowadays can easily trace the studies of scholars from related areas and keep up with the research trends. Due to the convenience of the Internet, scholars can establish cooperation relationships even though they may never meet before in reality. Consequently, interdisciplinary cooperation happens more frequently than in the past, and the positions of scholars in the network play an important role in promoting the collaborations. The academic collaborations among diverse domains accelerate the advancements of science; meanwhile, researchers can also obtain information or techniques through the collaborations with diverse researchers.

4.1.1. Scholars' Structural Holes Measurements. To depict scholars' positions in the network, we first apply the structural holes theory. The main principle of structural holes is that people would benefit more if they are in the positions that can link people from different groups. Typically, there are several ways of measuring the structural holes; we apply the most commonly used measurements which are the bridge counts and the betweenness centrality. To find the appropriate measures of structural holes for our algorithm, we apply the above methods, respectively, in the calculation of SI to evaluate their performances. The specific calculation processes are illustrated as follows.

Bridge Counts. It is an intuitively appealing measure. The link between two people is a bridge if there are no indirect connections between the two people. Equation (1) indicates the calculation formula:

$$\text{BrC}(a_i) = \sum_{a_j=1}^{n-1} b_{ij} \tag{1}$$

where $\text{BrC}(a_i)$ is the total number of bridges between authors a_i and a_j; n is the number of authors in the network. If there exists a bridge between a_i and a_j, the value of b is 1; otherwise, the value of b_{ij} is 0.

Betweenness Centrality. The betweenness centrality is the count of the structural holes to which a person has monopoly access. Given that a network contains n nodes, the maximum possible value for node is degree which is $n - 1$, and the maximum possible value for its betweenness centrality equals

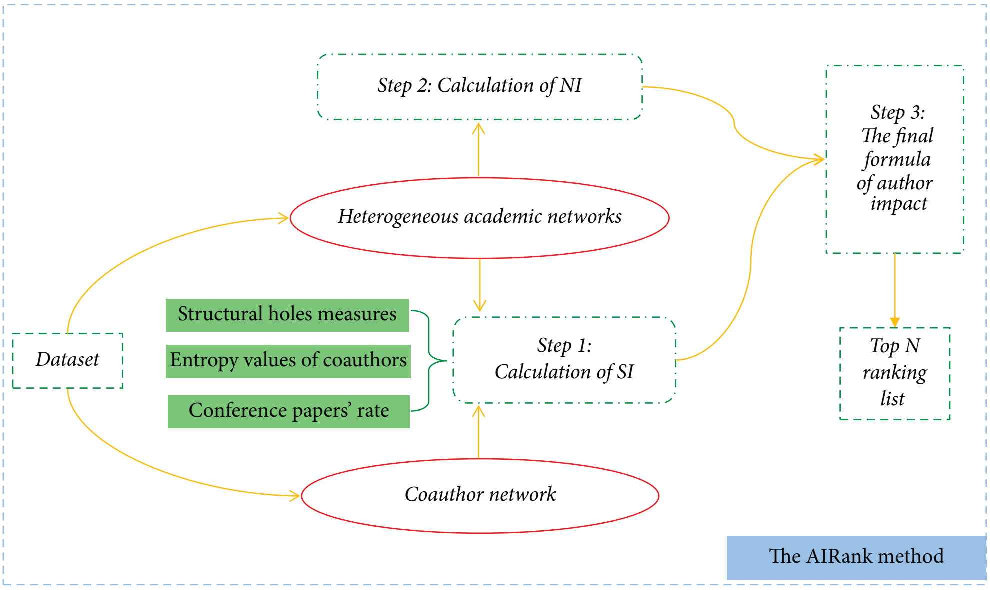

Step 2: Calculation of NI
Step 3: The final formula of author impact
Heterogeneous academic networks
Structural holes measures
Entropy values of coauthors
Conference papers' rate
Dataset
Step 1: Calculation of SI
Top N ranking list
Coauthor network
The AIRank method

Figure 3: Architecture of AIRank.

the hub node which is betweenness centrality value in a star network. More specifically, the shortest path between all the other node pairs is unique and definitely via the hub node. Therefore, node is betweenness centrality value this is the sum of all the above-mentioned shortest paths which equals the following formula:

$$\frac{(n-1)(n-2)}{2}=\frac{n^2-3n+2}{2} \tag{2}$$

Based on the above equation, the normalized betweenness centrality is defined as follows:

$$\mathrm{BeC}(a_i)=\frac{2}{n^2-3n+2}\sum_{a_i\neq a_s\neq a_t}\frac{N_{a_{st}}^{a_i}}{g_{a_{st}}} \tag{3}$$

where $\mathrm{BeC}(a_i)$ is the betweenness centrality value of author a_i, n is the size of the network, $g_{a_{st}}$ is the number of the shortest paths from a_s to a_t, and $N_{a_{st}}^{a_i}$ is the number of shortest paths that go through author a_i.

4.1.2. Diversity of Cooperators. The scholarly articles usually are the collective efforts of several coauthors, and researchers can benefit a lot from their coauthors through the cooperation relationship. Research ideas or techniques can be exchanged among coauthors through the collaboration process; as a consequence, scholars' academic achievements can be affected by the information they acquired and the people they interact with. Previous studies have investigated that researchers' academic level can be influenced by their coauthors' impact; however, the effect of the diversity of information and scholars that researchers accessed still needs to be explored.

To capture the variety of information, we consider two apparent information sources that researchers directly contact with. The first one is acquiring information through their cooperators. Ideas, problems, or techniques can be discussed and shared through the collaborative working towards publishing scholarly articles among coauthors. Therefore, the background of a scholar can represent the variety of information he or she commands. As in our previous work [37], the theory of entropy is utilized in measuring the diverse backgrounds of cooperators which only considers the differences between institutions, while in this work we not only think about the differences of institutions, but also take the distinctions of research interests into consideration. The calculation process is as follows:

$$\mathrm{Div}(a_i)_{\mathrm{inst}}=-\sum_{m=1}^{r}w_m\log_2(w_m) \tag{4}$$

$$\begin{gathered}\mathrm{Div}(a_i)_{\mathrm{key}}=-\sum_{\rho=1}^{q}k_\rho\log_2(k_\rho)\\ \mathrm{Div}(a_i)=\mathrm{Div}(a_i)_{\mathrm{inst}}+\mathrm{Div}(a_i)_{\mathrm{key}}\end{gathered} \tag{5}$$

where $\mathrm{Div}(a_i)_{\mathrm{inst}}$ and $\mathrm{Div}(a_i)_{\mathrm{key}}$ represent the diversities of cooperators' institutions and their papers' keywords of author a_i, and $\mathrm{Div}(a_i)$ is the overall cooperators' diversities of author a_i. w_m is the frequency of occurrences of word m in the combination of words extracted from the institutions' information of a_i's collaborators, and r is the total amount of word m in (4). k_ρ is the frequency of occurrences of word ρ in all the papers' keywords of a_i's collaborators, and q is the sum of words ρ.

4.1.3. Benefit Obtained via Academic Conferences. Another universal way of getting information is through attending academic conferences. Researchers publishing articles in the same conference commonly have similar research interests, and they can share their ideas or exchange information

```
Step 1 SI ($r, w_m, q, k_\rho, S(C), \mathrm{Num}_{a_i}^{\mathrm{conf}}, \mathrm{Num}_{a_i}^{p}, b$)
for $m \leftarrow 1$ to $r$ do
    $\mathrm{Div}(a_i)_{\mathrm{inst}} \longleftarrow -(\mathrm{Div}(a_i)_{\mathrm{inst}} + w_m \log_2(w_m))$
end for
for $\rho \leftarrow 1$ to $q$ do
    $\mathrm{Div}(a_i)_{\mathrm{key}} \longleftarrow -(\mathrm{Div}(a_i)_{\mathrm{key}} + k_\rho \log_2(k_\rho))$
end for
$\mathrm{Div}(a_i) \longleftarrow \mathrm{Div}(a_i)_{\mathrm{inst}} + \mathrm{Div}(a_i)_{\mathrm{key}}$
for $v \leftarrow 1$ to $t$ do
    $\mathrm{temp} \longleftarrow \mathrm{temp} + S(C_v)$
end for
$\mathrm{Bene}(a_i) \longleftarrow \frac{\mathrm{Num}_{a_i}^{\mathrm{conf}}}{\mathrm{Num}_{a_i}^{p}} \mathrm{temp}$
for $a_j \leftarrow 1$ to $n-1$ do
    $\mathrm{BrC}(a_i) \longleftarrow \mathrm{BrC}(a_i) + b$
end for
$\mathrm{SI}_{a_i}^{\mathrm{BrC}} = \frac{1-\chi-\psi-\varphi}{n} + \chi Z_{\mathrm{Div}(a_i)} + \psi Z_{\mathrm{Bene}(a_i)} + \varphi Z_{\mathrm{BrC}(a_i)}$
$\mathrm{SI}_{a_i}^{\mathrm{BeC}} = \frac{1-\tau-\lambda-\varepsilon}{n} + \tau Z_{\mathrm{Div}(a_i)} + \lambda Z_{\mathrm{Bene}(a_i)} + \varepsilon Z_{\mathrm{BeC}(a_i)}$
```

Algorithm 1

through attending the conference unlike publishing journal articles. Therefore, researchers can benefit a lot through participating in academic conferences, and the benefit that researchers get is captured by the following equation:

$$\mathrm{Bene}(a_i) = \frac{\mathrm{Num}_{a_i}^{\mathrm{conf}}}{\mathrm{Num}_{a_i}^{p}} \times \sum_{v=1}^{t} S(C_v) \quad (6)$$

where $\mathrm{Bene}(a_i)$ represents a_i's benefit obtained through attending academic conferences, $\mathrm{Num}_{a_i}^{\mathrm{conf}}$ is the number of conference papers that author a_i published, and $\mathrm{Num}_{a_i}^{p}$ is the total number of published papers of author a_i. $S(C_v)$ is the impact value of the conferences (C_v) that author a_i published papers in, and t is the total number of v. The value of $S(C_v)$ equals its PageRank value in the paper-venue network.

4.1.4. Final Formula of SI. In this paper, we propose three factors to measure the effect of scholars' positions in the networks, which are scholars' structural holes values, the diversity of coauthors, and the benefits obtained via academic conferences. The pseudocode of SI is shown in Algorithm 1, and its specific calculation procedure is illustrated as follows.

Step 1. Calculate scholars' structural holes values, which exist with two ways of calculation (bridge counts and betweenness centrality).

Step 2. Calculate the diversity of coauthors, which utilizes the concept of information entropy to measure the diversity of scholars' cooperators.

Step 3. Calculate the benefits researchers obtained through attending academic conferences.

Step 4. Calculate scholar's final SI values, which exist in two ways ($\mathrm{SI}^{\mathrm{BrC}}$ and $\mathrm{SI}^{\mathrm{BeC}}$), according to the normalized above-mentioned factors.

The calculation procedure of $\mathrm{Div}(a_i)$, $\mathrm{Bene}(a_i)$, $\mathrm{BeC}(a_i)$, and $\mathrm{BrC}(a_i)$ can be obtained based on the above equations. While these three indicators cannot be arithmetically operated directly due to their different scales, therefore, we need to normalize them before the calculation process. The normalization process is shown as follows:

$$Z_i = \frac{v_i - \min_A}{\max_A - \min_A}\left(\mathrm{new}_{\max_A} - \mathrm{new}_{\min_A}\right) + \mathrm{new}_{\min_A} \quad (7)$$

where A is the set of scholars' attributes, which includes the $\mathrm{Div}(a_i)$, $\mathrm{Bene}(a_i)$, $\mathrm{BeC}(a_i)$, and $\mathrm{BrC}(a_i)$. $\max_A$ is the maximum value and $\min_A$ is attribute A's minimum value. v_i is attribute A's original value, and Z_i is the normalization value of v_i in the range of $[\mathrm{new}_{\min_A}, \mathrm{new}_{\max_A}]$, which equals $[0, 1]$.

To find the appropriate measures of structural holes for our algorithm, we apply $\mathrm{BeC}(a_i)$ and $\mathrm{BrC}(a_i)$, respectively, in the SI method to find the most efficient one. Therefore, the overall assessment of scholars' abilities to acquire diverse information and their positions in the networks can be interpreted in two ways ($\mathrm{SI}_{a_i}^{\mathrm{BrC}}$ and $\mathrm{SI}_{a_i}^{\mathrm{BrC}}$) through the following equations:

$$\begin{aligned} \mathrm{SI}_{a_i}^{\mathrm{BrC}} &= \frac{1-\chi-\psi-\varphi}{n} + \chi Z_{\mathrm{Div}(a_i)} + \psi Z_{\mathrm{Bene}(a_i)} \\ &\quad + \varphi Z_{\mathrm{BrC}(a_i)} \\ \mathrm{SI}_{a_i}^{\mathrm{BeC}} &= \frac{1-\iota-\lambda-\varepsilon}{n} + \iota Z_{\mathrm{Div}(a_i)} + \lambda Z_{\mathrm{Bene}(a_i)} \\ &\quad + \varepsilon Z_{\mathrm{BeC}(a_i)} \end{aligned} \quad (8)$$

where $\chi, \psi, \iota, \lambda, \varepsilon$, and φ are parameters; $\mathrm{SI}_{a_i}^{\mathrm{BrC}}$ and $\mathrm{SI}_{a_i}^{\mathrm{BeC}}$ represent the value of SI_{a_i} which utilize BrC and BeC, respectively, to measure the positions of scholars in the network. $Z_{\mathrm{Div}(a_i)}$, $Z_{\mathrm{Bene}(a_i)}$, $Z_{\mathrm{BrC}(a_i)}$, and $Z_{\mathrm{BeC}(a_i)}$ are the normalization value of $\mathrm{Div}(a_i)$, $\mathrm{Bene}(a_i)$, $\mathrm{BrC}(a_i)$, and $\mathrm{BeC}(a_i)$ according to (7).

4.2. Calculation Procedure of NI. The next procedure of our method is measuring the influence of scholars in heterogeneous academic networks through utilizing the PageRank and HITS algorithms. Considering the mutual influence among academic entities through different relationships in the networks, we construct three academic networks to evaluate the scientific impact, i.e., the citation network, the paper-venue network, and the paper-author network.

(i) Citation network: it contains one type of entities and relationships, i.e., papers, and the citation relationship among them.

(ii) Paper-venue network: it composes two kinds of nodes and one kind of relationships. The nodes in the network are the papers and venues, and the publication relationship links the papers and their corresponding venues.

(iii) Paper-author network: it consists of two kinds of entities, which are papers and their corresponding authors. Only one type of relationships is included in this network which depicts the writing relationship between papers and their authors.

We first apply the original PageRank algorithm to evaluate the importance score of articles in the citation network. According to this, the initial importance of papers in the citation network can be obtained. Then we calculate the impact of venues and authors in the constructed paper-venue network and paper-author network, respectively, through using the HITS algorithm, and we set the initial value of the entities in the networks accordingly. The pseudocode of NI is shown in Algorithm 2, and its specific calculation procedure is conducted as follows:

(1) The initial value of publications is set as $1/N$, where N is the total number of articles in the network.

(2) Calculate the scores of papers through utilizing the PageRank algorithm in the citation network.

(3) Calculate the scores of papers and the corresponding venues in the paper-venue network by HITS algorithm; the initial values of papers are set according to their PageRank scores obtained in the above step.

(4) Calculate the scores of scholars in the paper-author network through the HITS algorithm; the initial values of papers are set according to their values obtained from Step (3).

(5) Repeat Steps (2)–(4) until convergence is encountered.

```
Step 2 NI ($S, U, \mathrm{Pr}, \alpha, h$)
$G \longleftarrow \alpha S + \frac{1-\alpha}{n} U$
for $i \leftarrow 0$ to $n$ do
    $\mathrm{pr\_next} \longleftarrow G\mathrm{Pr}$
    $\mathrm{Pr} \longleftarrow \mathrm{Pr\_next}$
end for
$a \longleftarrow \mathrm{copy}(\mathrm{Pr})$
for $i \leftarrow 0$ to $n$ do
    for $i \leftarrow 0$ to $n$ do
        $h_i \longleftarrow h_i + a_i$
        $h_i \longleftarrow \frac{h_i}{\max(h_i)}$
    end for
    for $i \leftarrow 0$ to $n$ do
        $a_i \longleftarrow a_i + h_i$
        $a_i \longleftarrow \frac{a_i}{\max(a_i)}$
    end for
end for
return $a$
```

Algorithm 2

4.2.1. Article's Score in Citation Network. Initially, the PageRank algorithm is proposed to evaluate and rank the importance of webpages since there may pop up many searching results and it is time-consuming for users to discover the useful one. The fundamental principle of the PageRank algorithm is that the webpages would be ranked high if it is pointed by high-rank webpages, and top ranking webpages are more likely to be pointed to than lower ranked webpages. Other than ranking the importance of websites, researchers nowadays also use it to measure the importance of diverse entities in a variety of networks, such as ranking the importance of scholars in academic networks. The PageRank values of articles can be obtained by the following formula:

$$\mathrm{PR}(p_i) = \frac{1-d}{N} + d \sum_{j=1}^{m} \frac{\mathrm{PR}(p_j)}{L(p_j)} \tag{9}$$

where p_i represents the paper, N is the total amount of the articles, p_j is the node that links to p_i, and $L(p_j)$ is p_j's total outgoing links. $\mathrm{PR}(p_i)$ and $\mathrm{PR}(p_j)$ indicate the importance values of p_i and p_j correspondingly. d is the damping factor which controls the visiting probability of node p_i that can be visited by the link directed to it. A variety of researches have studied the influence of damping factor's different values, and they all believe that it is more suitable for the whole calculation process when set as 0.85. Therefore, in our paper, the values of damping factor are all set as 0.85 as mentioned above. Since the PageRank calculation procedure is iterated, we update each paper's value at every step of the computations based on (9). When the values of all the papers are converged to a steady state, the calculations are stopped, and finally the PageRank value of each article is obtained.

4.2.2. Updated Scores of Papers and Venues in the Paper-Venue Network. Next, the undirected paper-venue network is constructed to calculate the importance of papers and venues considering the mutual influence among them by using the

HITS algorithm. Because the qualities of papers are different originally, we take the PageRank scores of them that are obtained from the last step as their initial value in the step. The major function of HITS algorithm is similar to the PageRank algorithm, which also calculates the importance of entities in the networks. In HITS algorithm, each node possesses two values, which are the authority and hub values. The hub value indicates the value of node's links to other nodes, and the authority represents the quality of node itself. If a node is widely known as a hub, it can guide the users to the nodes with high authority values. On the contrary, if a node's authority value is high, it can be regarded as the node with important content. The authority and hub values of nodes can be calculated as follows:

$$\begin{aligned} \operatorname{auth}(a_k) &= \sum_{i=1}^{s} \operatorname{hub}(l_i) \\ \operatorname{hub}(a_k) &= \sum_{i=1}^{v} \operatorname{auth}(p_i) \end{aligned} \tag{10}$$

where a_k is the node, auth(a_k) is the authority value of it, and we apply it to represent its impact in the network. l_i is the node links to a_k in the network, and s is the sum of l_i. p_i indicates the node that a_k points to, and v is the total number of p_i. At the beginning, if a_k is a venue, its initial authority and hub values are set as 1; otherwise, its initial authority and hub values are set equal to its PageRank score that is obtained from the last step.

4.2.3. Scores of Scholars in the Paper-Author Network. In this part, the paper-author network is established to evaluate scholars' impact. Other than the PageRank algorithm, we also utilize the HITS algorithm to measure the importance of scholars based on the paper-author network. To obtain scholars' authority values, the above-mentioned calculation equations are still applied; however, we set the initial values differently. If the node is a paper, we set its initial values equal to its value obtained from the last step; else the values of the node are set equal to 1. The overall measurement of scholars' impact in heterogeneous academic networks (NI) is calculated as follows:

$$\operatorname{NI}(a_i) = \operatorname{auth}(a_i)\left\{\sum_{p=1}^{n} \operatorname{PR}(p_i) \operatorname{auth}(j_k)\right\} \tag{11}$$

where n is a_i's total amount of scholarly articles, PR(p_i) is the PageRank value of a_i's paper in the citation network, auth(a_i) is author a_i's authority value in the paper-author network, and auth(j_k) is the authority value of p_i's corresponding venue j_k in the paper-venue network.

With the above analysis and the applications of three heterogeneous academic networks, the mutually reinforced procedure of scholarly entities can be explored. In addition, the hybrid of the PageRank and HITS algorithms also can highlight their different advantages in adapting different network topologies and improve the ranking results of scholarly entities in the networks.

4.3. Final Calculation of Scholars' Impact. After finishing the calculation of the above two parts, we then come up with the final formula for evaluating the impact of scholars. In our proposed AIRank method, it consists of two major parts, which are scholars' positions in the network and the hybrid importance values of scholarly entities in the above-mentioned three subnetworks. The theory of structural holes can indicate scholars' abilities to connect different people; therefore we utilize it in our method to depict scholars' positions in the network. To capture the multiplicities of information that researchers acquire through their relationships with other people, we measure these multiplicities from two aspects, which are the diversity of their coauthors and the quantity and quality of academic conferences they attend. In addition, we also consider the mutual effects among different academic entities in the networks together to gauge the scientific impact of scholars. As a consequence, we calculate scholar's final score according to the following formula:

$$F(a_i) = \frac{1-\xi-\varpi}{n} + \xi Z_{\mathrm{SI}_{a_i}} + \varpi Z_{\mathrm{NI}_{a_i}} \tag{12}$$

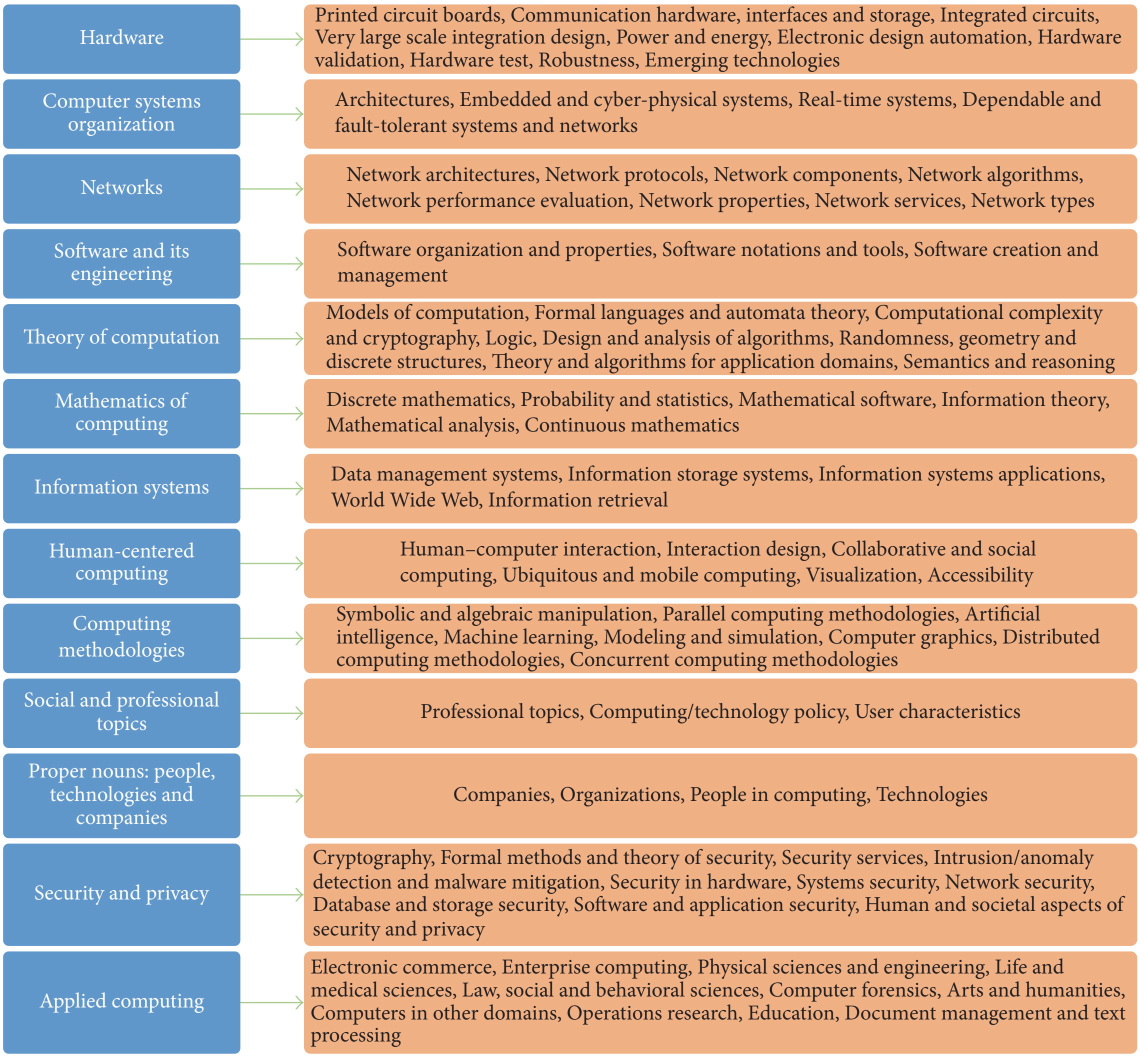
where $F(a_i)$ represents the final impact score of author a_i, $Z_{\mathrm{SI}_{a_i}}$ and $Z_{\mathrm{NI}_{a_i}}$ are the normalization values of SI_{a_i} and NI_{a_i} according to (7), and ξ and ϖ are parameters.

With the above descriptions, we propose a scholars' impact evaluation method which measures the scientific impact from two aspects. Our method not only considers the impact of scholars in heterogeneous academic networks through the mutual influence mechanism among academic entities, but also integrates the positions of scholars in the networks and their abilities to access various kinds of information and researchers to measure their overall scientific impact.

5. Experimental Results

In this section, we explore the performance of AIRank in the real dataset. Since there is no ground truth for the evaluation of scholars' impact, the citation counts are applied as the ground truth to validate their performance. In academia, it is commonly acknowledged that if one scholar is outstanding, he or she has higher citation counts comparing to other researchers. To explore the effectiveness of the AIRank in selecting high-impact scholars with interdisciplinary nature, we first compare each method's top ranking scholars' average citation counts, common members with citation's ranking lists, and ranking positions of scholars. To specifically show the detailed information of the top researchers selected by each method, we then list the detailed citation counts and cross-domain citations of top 10 scholars in each method to prove the efficiency of our AIRank. In addition, the Pearson Correlation Coefficient between the citation counts and each ranking list is also calculated to show the correlations.

5.1. Dataset and Experimental Setup. The subdataset used for our experiments is acquired from the Microsoft Academic Graph (MAG) datasets. It provides the detailed information of each article. To improve the efficiency of our experiments, the dataset needs to be extracted. In order to alleviate



Hardware	Printed circuit boards, Communication hardware, interfaces and storage, Integrated circuits, Very large scale integration design, Power and energy, Electronic design automation, Hardware validation, Hardware test, Robustness, Emerging technologies
Computer systems organization	Architectures, Embedded and cyber-physical systems, Real-time systems, Dependable and fault-tolerant systems and networks
Networks	Network architectures, Network protocols, Network components, Network algorithms, Network performance evaluation, Network properties, Network services, Network types
Software and its engineering	Software organization and properties, Software notations and tools, Software creation and management
Theory of computation	Models of computation, Formal languages and automata theory, Computational complexity and cryptography, Logic, Design and analysis of algorithms, Randomness, geometry and discrete structures, Theory and algorithms for application domains, Semantics and reasoning
Mathematics of computing	Discrete mathematics, Probability and statistics, Mathematical software, Information theory, Mathematical analysis, Continuous mathematics
Information systems	Data management systems, Information storage systems, Information systems applications, World Wide Web, Information retrieval
Human-centered computing	Human–computer interaction, Interaction design, Collaborative and social computing, Ubiquitous and mobile computing, Visualization, Accessibility
Computing methodologies	Symbolic and algebraic manipulation, Parallel computing methodologies, Artificial intelligence, Machine learning, Modeling and simulation, Computer graphics, Distributed computing methodologies, Concurrent computing methodologies
Social and professional topics	Professional topics, Computing/technology policy, User characteristics
Proper nouns: people, technologies and companies	Companies, Organizations, People in computing, Technologies
Security and privacy	Cryptography, Formal methods and theory of security, Security services, Intrusion/anomaly detection and malware mitigation, Security in hardware, Systems security, Network security, Database and storage security, Software and application security, Human and societal aspects of security and privacy
Applied computing	Electronic commerce, Enterprise computing, Physical sciences and engineering, Life and medical sciences, Law, social and behavioral sciences, Computer forensics, Arts and humanities, Computers in other domains, Operations research, Education, Document management and text processing

Figure 4: The ACM Computing Classification System.

the effect of different research areas and years of entering academia, we choose scholars that are from the same area and whose academic careers ages are the same for scientific impact evaluation. The academic age in our paper refers to the years between scholar publishing his or her first article and the last article in the database. The final dataset includes 79,321 scholars and 105,123 publications.

When calculating the values of NI, we apply both the PageRank and HITS algorithms to rank the importance of scholars in heterogeneous academic networks. The operation mechanism of these two algorithms is similar in which they both need a sufficient number of iterations to converge. In our case, we set the iteration numbers as 500 times, and the difference value of the sum of all the scholars' values obtained from two successive iterations is smaller than a threshold (set as 0.000001).

5.2. The CCS Classification.

To measure the cross-domain citations of articles and their authors, we adopt the ACM Computing Classification System (CCS) from the website https://www.acm.org/. It is a subject classification system for computing, to classify the articles into the related areas according to their keywords. The specific classification criteria are shown in Figure 4. As it shows, there are several major domains and a set of keywords is included in each main kind. According to the keywords listed above, articles can be sorted to the corresponding domains.

5.3. Baseline Methods.

In order to investigate the effectiveness of our proposed AIRank method, we employ the different variants of our method, the PageRank algorithm, and *h*-index for comparison. The details of the above methods are as follows:

(i) $\mathrm{SI}^{\mathrm{BrC}}$: it represents the value of SI which utilizes BrC (see (1)) to measure the positions of scholars in the network.

(ii) $\mathrm{SI}^{\mathrm{BeC}}$: it represents the value of SI which utilizes BeC (see (3)) to measure the positions of scholars in the network.

(iii) NI: it is part of our proposed AIRank, which only considers the combination results through applying PageRank and HITS algorithm under heterogeneous academic networks to evaluate the impact of scholars.

(iv) AIRank$^{\mathrm{BrC}}$: it is our proposed method, which utilizes BrC to measure the positions of scholars in the network.

(v) AIRank$^{\mathrm{BeC}}$: it is our proposed method, which utilizes BeC to measure the positions of scholars in the network.

(vi) PageRank: it applies the PageRank algorithm to evaluate the impact of each scholar.

(vii) h-index*:* it is the h-index value of each scholar.

To start the research work, scholars often need to review the existing literature from related areas. Therefore, it is commonly recognized that scholars may be inspired by articles in areas other than articles within the same area. As a consequence, the citations of articles may be not only from a single area, but also from other disciplines due to their impact on other areas. To understand the interdisciplinary nature of citations, we first investigate the citation distributions of different domains in MAG dataset.

As shown in Figure 5, it is a chord graph, which indicates the proportions of articles from each domain in the MAG dataset. Different domains are represented with different colors, and the citation distributions of articles in each domain can be easily observed. The diagram displays that the total numbers of papers in applied computing and computing methodologies areas are larger than the numbers of articles in other domains. Furthermore, papers in these two areas also shed light on the scientific inventions of other areas due to their citation distributions. Generally, it is obvious that almost every article cites papers from other areas. The areas in computer science correlate with each other closely and promote the development of computer science together.

With the above analysis, the tendency of citation distributions is apparently showing an increasing trend of interdisciplinary collaborations, i.e., the number of cross-domain citations. To further explore the effect of cross-domain citations on the scientific impact of scholars, we then list the cross-domain citations of top ranking scholars by citation counts and h-index. As shown in Figure 6(a), the top percentile ranking scholars with bigger citation counts also obtain higher cross-domain citations. The same phenomenon is also observed in Figure 6(b), where the higher the h-index values of scholars, the more the average cross-domain citation counts that they will get. The trends of these two figures are alike; however, their concrete average cross-domain citation counts of top ranking scholars appear to be different. There exists a great numerical difference of top 10% scholars' average cross-domain citation counts between Figures 6(a) and 6(b), and the numerical differences decrease with the increase of top percentile ranking scholars. The reason behind this phenomenon correlates closely with the principle of calculating scholar's h-index. Although there exist some numerical differences, the overall trend of these figures is similar, which validates the fact that high-impact scholars also gain high reputations in other domains.

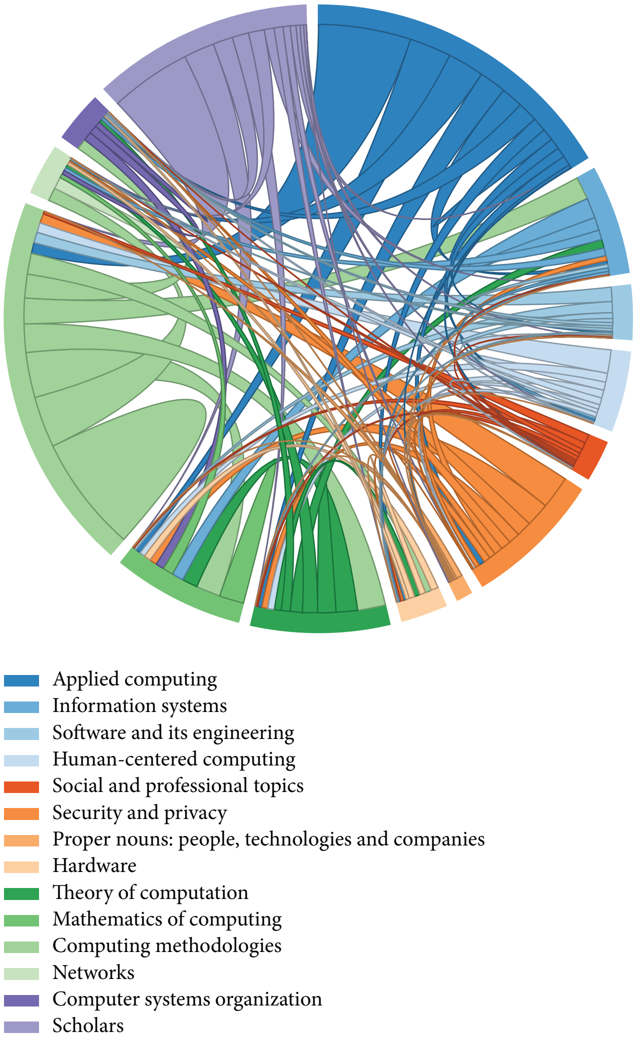
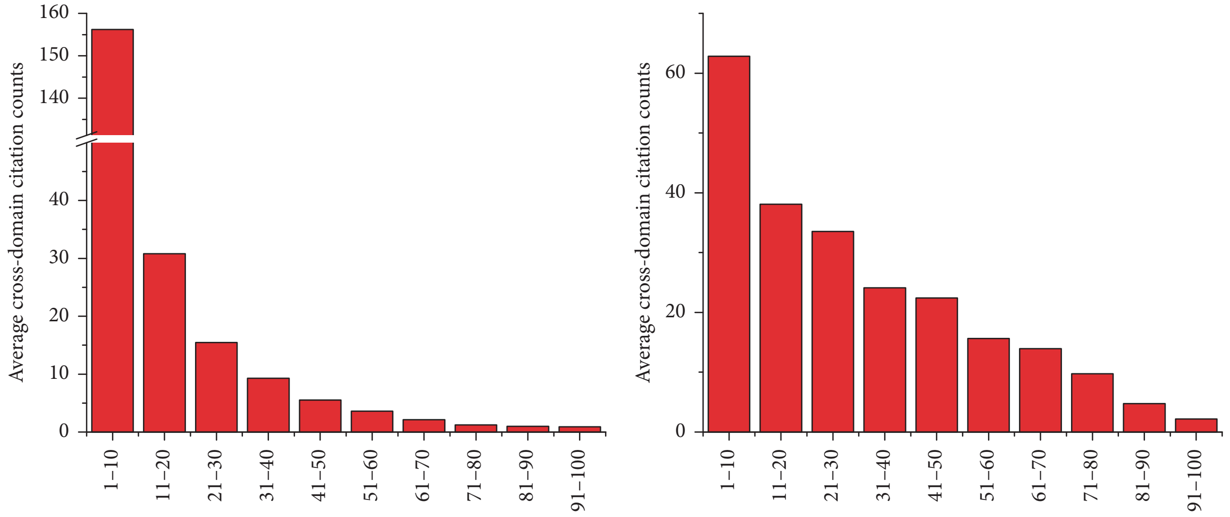


Figure 5: The interdisciplinary citations among articles in computer science area.

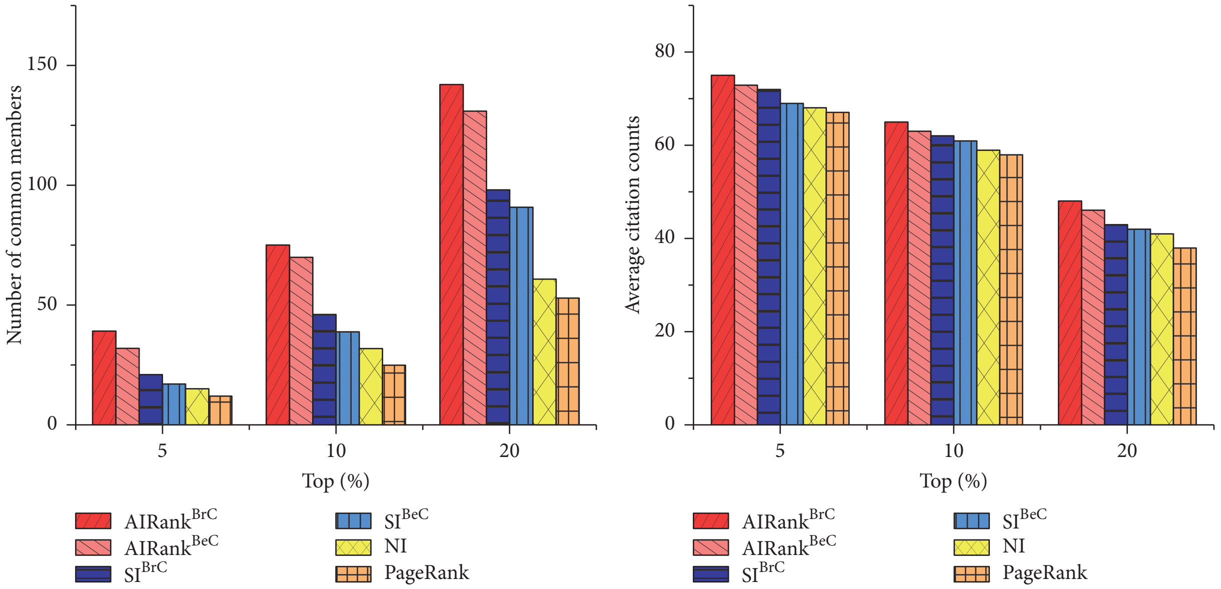
In order to investigate each method's ability to identify influential scholars more exquisitely and in convincible manner, we first compare the number of common members between each methods ranking list and citation rankings. A ranking list of scholars can be obtained through their final scores by each method. As shown in Figure 7(a), the SI shows a better result than the performance of NI. Meanwhile, the overall performances of AIRank variants are better than other methods. Our proposed AIRank$^{\mathrm{BrC}}$ method can get the most common members with the citation counts rankings when comparing the top 5%, top 10%, and top 20% ranking lists by each method. Furthermore, we then compare each method's average citation counts of top ranking scholars. As shown in Figure 7(b), the number of the average citation counts of top scholars according to our AIRank method is the highest among other methods, while the AIRank$^{\mathrm{BrC}}$ method still achieves the best performance comparing with other methods. Through Figures 6(a) and 6(b), we find that



(a) Citation counts

(b) *h*-index

Figure 6: The average cross-domain citation counts of top ranking scholars.



(a) Common member with top lists

(b) Average citation counts of top lists

Figure 7: The performance of top ranking scholars by SI^{BrC}, SI^{BeC}, NI, $\text{AIRank}^{\text{BrC}}$, and $\text{AIRank}^{\text{BeC}}$.

the more influential the scholars, the more the cross-domain citation counts that they will obtain. We then specifically show each method's top 10 researchers' citation counts and cross-domain citation counts. As shown in Table 1, it is clear that performance of our method is better than the PageRank method. Due to the mechanism of PageRank algorithm, the higher value of PageRank score indicates the more citations from influential scholars; therefore, the top 3 scholars' citation counts according to PageRank algorithm are high while the rest decrease distinctly. As shown in Tables 1 and 2 and Figures 7(a) and 7(b), the results demonstrate that the performance of our method is better than other approaches when comparing top ranking scholars' overall average citation counts and cross-domain citations. These results also confirm the findings displayed in the above tables. Generally, the AIRank method has a better performance when applying the bridge counts to measure the positions of scholars in the network.

The ranking positions of the top 100 scholars according to the citation counts in our proposed methods are also investigated. Since the specific calculation process of each method is different, scholars' ranking positions by each method are distinct either. In this paper, the number of citation counts is chosen as the ground truth; hence we assume

Table 1: Top 10 scholars of each method.

AIRank$^{\mathrm{BrC}}$			AIRank$^{\mathrm{BeC}}$			PageRank		
Top 10	Citations	Cross-domain citations	Top 10	Citations	Cross-domain citations	Top 10	Citations	Cross-domain citations
7F2CEC81	1127	1011	8023C793	846	818	80EB57FC	613	596
7FB76008	857	793	7E2B1F64	783	726	7F2CEC81	1127	1011
7E2B1F64	783	726	7FB76008	857	793	8023C793	846	818
8173CEDE	68	34	80EB57FC	613	514	7D6A4BFF	187	179
7D6A4BFF	187	179	7F2CEC81	37	23	80AD9709	98	87
8023C793	846	818	7D6A4BFF	98	98	7F680B0B	98	98
80EB57FC	613	514	7DE7A740	485	409	756F9F32	23	14
7DE7A740	485	409	0838B97F	87	80	4899EC1B	79	53
80D1979B	134	126	78322C72	126	113	78322C72	97	80
7BB5A93A	137	122	7F78CE41	112	102	7FC94B6B	89	81

Table 2: Top 10 scholars of each method.

SI$^{\mathrm{BrC}}$			SI$^{\mathrm{BeC}}$			NI		
Top 10	Citations	Cross-domain citations	Top 10	Citations	Cross-domain citations	Top 10	Citations	Cross-domain citations
7DE7A740	485	409	7DE7A740	485	409	7DE7A740	485	409
80EB57FC	613	514	80EB57FC	613	514	80EB57FC	613	514
802E02C5	168	161	802E02C5	69	61	802E02C5	69	61
0857BCE0	286	286	0857BCE0	286	286	0857BCE0	286	286
7ED3570E	228	213	7ED3570E	228	213	7ED3570E	228	213
7FF53EE6	89	79	7FF53EE6	89	79	7FF53EE6	89	79
80FE41D4	112	112	80FE41D4	112	112	80FE41D4	112	112
8173CEDE	68	34	8043DB84	45	30	7EFAE119	24	18
7F2CEC81	37	23	7F2CEC81	34	27	7F2CEC81	34	27
113BBABC	63	45	75ADB28C	28	21	4769E8AE	16	11

that the more effective in identifying influential scholars of the above-mentioned method it is, the higher the ranking positions of the top 100 scholars by citation counts are. For instance, one scholar ranks the first by citations counts while in other methods he or she, respectively, ranks the 4th, 10th, and 3rd; then it is obvious that the method which ranks this scholar the 3rd achieves the best performance among others. The top 100 scholars' ranking positions by each method are shown in Figure 8, and the ranking differentials can be directly obtained. It is apparent that the AIRank method achieves the best performance, whose range of the ranking positions for top scholars is the smallest. Among these methods, it is obvious that the AIRank$^{\mathrm{BrC}}$ still performs the best in scholars' ranking positions.

Other than the efficiency in identifying high-impact scholars, the performance of evaluating the overall scientific impact of scholars still needs to be explored. We first examine the performance from the angle of distinguishing scholars with different scientific impact. According to scholars' citation counts, the higher ranked scholars are considered as positive entities, and authors that ranked low are deemed as negative entities. The above-mentioned methods are used as classifiers to evaluate their ranking results. In general, the classification results can have four types: top ranking scholar is classified as higher ranked (true positive); the scholar is higher ranked but is considered as top ranking scholar (false positive); lower ranked scholar is classified as lower ranked (true negative); lower ranked scholar but classified as top ranking scholar (false negative). With these four kinds of classification results, the four rates can be calculated. The true positive rate (TPR) can be calculated as $\sum \text{truepositive}/\sum \text{conditionpositive}$, the false positive rate (FPR) can be calculated as $\sum \text{falsepositive}/\sum \text{conditionnegative}$, the true negative rate (TNR) can be calculated as $\sum \text{truenegative}/\sum \text{conditionnegative}$, and the false negative rate (FNR) equals $\sum \text{falsenegative}/\sum \text{conditionpositive}$.

The Receiver Operating Characteristic (ROC) curves of each method can be obtained through the above-mentioned rates. As shown in Figure 9, the ordinate is the Sensitivity = TPR/(TPR + FNR), and the abscissa is the 1 − Specificity = TNR/(TNR + FPR). The ROC curves in Figure 9 indicate

Table 3: AUC of each method.

	$\mathrm{SI}^{\mathrm{BrC}}$	$\mathrm{SI}^{\mathrm{BeC}}$	NI	$\mathrm{AIRank}^{\mathrm{BrC}}$	$\mathrm{AIRank}^{\mathrm{BeC}}$	PageRank
AUC	0.64504	0.66962	0.61812	0.73476	0.80749	0.59133

Table 4: Comparison of Pearson Correlation Coefficient.

	$\mathrm{SI}^{\mathrm{BrC}}$	$\mathrm{SI}^{\mathrm{BeC}}$	NI	$\mathrm{AIRank}^{\mathrm{BrC}}$	$\mathrm{AIRank}^{\mathrm{BeC}}$
Citation counts	0.453	0.437	0.496	0.538	0.522
h-index	0.230	0.220	0.098	0.231	0.222
PageRank	0.738	0.782	0.305	0.785	0.742

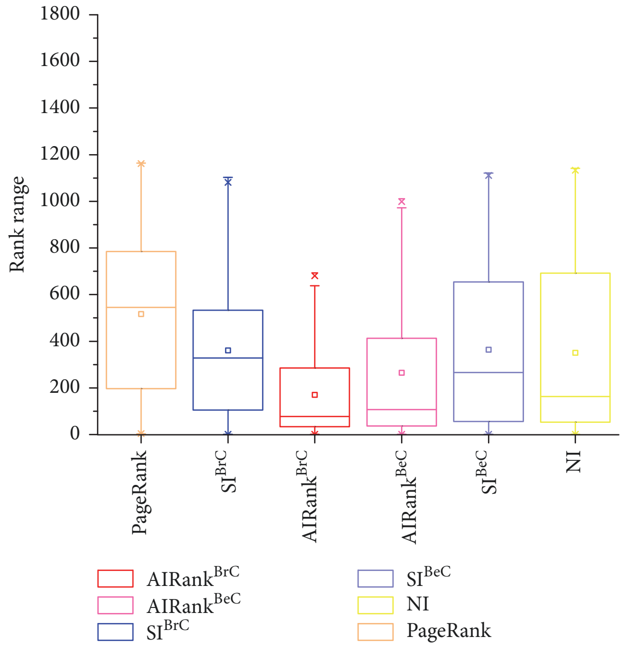


Figure 8: Boxplots of ranking positions for top scholars.

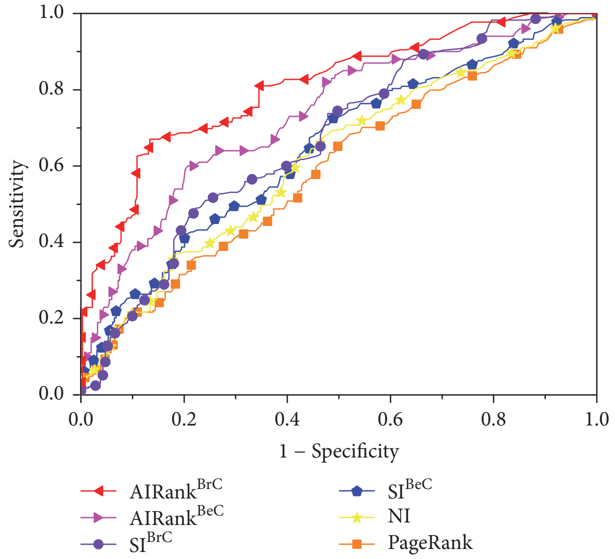


Figure 9: ROC curves of $\mathrm{SI}^{\mathrm{BrC}}$, $\mathrm{SI}^{\mathrm{BeC}}$, NI, $\mathrm{AIRank}^{\mathrm{BrC}}$, and $\mathrm{AIRank}^{\mathrm{BeC}}$.

that our AIRank method can classify different scholars with the best performance. Moreover, we calculate the area that the ROC curves cover (AUC) which indicates the classifying accuracy rate. It is clear that our AIRank method has the highest accuracy rate according to Table 3. Through the above results, we can observe that the AIRank method performs better than other methods in classifying the scholars.

We adopt the universally acknowledged citation counts and h-index values to evaluate the performance of each method. The Pearson Correlation Coefficient is commonly used to measure the correlation between two sets of data. The value of it ranges from −1 to 1, which represents the fact that the correlations of two sets of data are from the most negative to the most positive ones. We apply the Pearson Correlation Coefficient to calculate the correlation among all the baseline methods ($\mathrm{SI}^{\mathrm{BrC}}$, $\mathrm{SI}^{\mathrm{BeC}}$, NI, $\mathrm{AIRank}^{\mathrm{BrC}}$, and $\mathrm{AIRank}^{\mathrm{BeC}}$) with the citation counts, h-index, and the PageRank algorithm. As shown in Table 4, the results indicate that the AIRank method outperforms other methods with higher values, and it makes a great improvement comparing to applying the SI and NI measurements alone. Meanwhile, the $\mathrm{AIRank}^{\mathrm{BrC}}$ method still achieves the best performance compared to other methods.

Generally, we examine the performance of each method from two main aspects: the ability to identify influential scholars and the comprehensiveness of evaluating the overall impact of scholars. We compare the cross-domain citations, ranking positions, common members, and average citations of the top ranking scholars in each method to investigate the capacity of identifying influential scholars. The results indicate that our AIRank method, specifically the $\mathrm{AIRank}^{\mathrm{BrC}}$ method, shows the best performance among all the other methods in identifying influential scholars. In addition, the ROC curve, the value of AUC, and the Pearson Correlation Coefficient are utilized to measure each method's efficacy in evaluating the overall impact of scholars. Similarly, the $\mathrm{AIRank}^{\mathrm{BrC}}$ method still prevails over all the other methods.

6. Conclusion

In this paper, our primary concern is to quantify scholars' scientific impact by utilizing the heterogeneous academic network topology. The positions of scholars in the coauthor network are taken into consideration to measure the scientific impact of scholars and their effects as well. We depict it from three aspects, which are the diversity of coauthors, the qualities of conference papers that scholars published, and their measurements of structural holes. Besides, we also integrate the interplay between different scholarly entities in heterogeneous academic networks through the random walk algorithms. Based on these indicators and scholars' impact in heterogeneous academic networks, we propose the AIRank method.

We construct the experiments on MAG dataset to prove the efficiency of AIRank and select the appropriate measurements on the positions of scholars in the network. Through the experiments on the real dataset, we find that influential scholars in some specific areas also obtain high reputation in other domains. The results also demonstrate that our algorithm performs better than other methods in selecting top ranking scholars with more cross-domain citation counts and measuring scholars' scientific impact more comprehensively. Furthermore, there still exists room for further modifications; e.g., the effects of the interplay and relationships between scholars on their scientific impact should be mined deeper. Our method is conducted only on literature from computer science area; the results obtained from more datasets on other disciplines could be examined, so that exploring other scientific disciplines for the same observed phenomena could further prove the effectiveness of our work.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at King Saud University for funding this work through Research Group no. RG-1438-027.

References

[1] F. Xia, W. Wang, T. M. Bekele, and H. Liu, "Big Scholarly Data: A Survey," *IEEE Transactions on Big Data*, vol. 3, no. 1, pp. 18–35, 2017.

[2] Z. Ning, X. Wang, X. Kong, and W. Hou, "A Social-aware Group Formation Framework for Information Diffusion in Narrowband Internet of Things," *IEEE Internet of Things Journal*, vol. PP, no. 99, pp. 1-1, 2017.

[3] Y. Dong, R. A. Johnson, and N. V. Chawla, "Can Scientific Impact Be Predicted?" *IEEE Transactions on Big Data*, vol. 2, no. 1, pp. 18–30, 2016.

[4] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 102, no. 46, pp. 16569–16572, 2005.

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

[6] E. Garfield, "The history and meaning of the journal impact factor," *Journal of the American Medical Association*, vol. 295, no. 1, pp. 90–93, 2006.

[7] C. Shi, Y. Li, J. Zhang, Y. Sun, and P. S. Yu, "A survey of heterogeneous information network analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 1, pp. 17–37, 2017.

[8] L. Page, S. Brin, R. Motwani, and T. Winograd, *The pagerank citation ranking: bringing order to the web*, 1999.

[9] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM*, vol. 46, no. 5, pp. 604–632, 1999.

[10] M. Nykl, K. Ježek, D. Fiala, and M. Dostal, "PageRank variants in the evaluation of citation networks," *Journal of Informetrics*, vol. 8, no. 3, pp. 683–692, 2014.

[11] Z. Ning, X. Hu, Z. Chen et al., "A cooperative quality-aware service access system for social internet of vehicles," *IEEE Internet of Things Journal*, vol. PP, no. 99, pp. 1-1, 2017.

[12] R. S. Burt, *Structural hole*, Harvard Business School Press, Cambridge, MA, USA, 1992.

[13] T. Lou and J. Tang, "Mining structural hole spanners through information diffusion in social networks," in *Proceedings of the the 22nd international conference on World Wide Web*, pp. 825–836, Rio de Janeiro, Brazil, May 2013.

[14] X. Su, W. Wang, S. Yu, C. Zhang, T. M. Bekele, and F. Xia, "Can Academic Conferences Promote Research Collaboration?" in *Proceedings of the the 16th ACM/IEEE-CS*, pp. 231-232, Newark, New Jersey, USA, June 2016.

[15] L. Li and H. Tong, "The Child is Father of the Man," in *Proceedings of the the 21th ACM SIGKDD International Conference*, pp. 655–664, Sydney, NSW, Australia, August 2015.

[16] D. Wang, C. Song, and A.-L. Barabási, "Quantifying long-term scientific impact," *Science*, vol. 342, no. 6154, pp. 127–132, 2013.

[17] 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, Article ID e0165997, 2016.

[18] R. Sinatra, D. Wang, P. Deville, C. Song, and A.-L. Barabási, "Quantifying the evolution of individual scientific impact," *Science*, vol. 354, no. 6312, Article ID aaf5239, 2016.

[19] A. Clauset, D. B. Larremore, and R. Sinatra, "Data-driven predictions in the science of science," *Science*, vol. 355, no. 6324, pp. 477–480, 2017.

[20] R. K. Pan and S. Fortunato, "Author impact factor: Tracking the dynamics of individual scientific impact," *Scientific Reports*, vol. 4, article no. 4880, 2014.

[21] S. Xiao, J. Yan, C. Li et al., "On modeling and predicting individual paper citation count over time," in *Proceedings of the 25th International Joint Conference on Artificial Intelligence, IJCAI 2016*, pp. 2676–2682, usa, July 2016.

[22] X. Wan and F. Liu, "Are all literature citations equally important? Automatic citation strength estimation and its applications," *Journal of the Association for Information Science and Technology*, vol. 65, no. 9, pp. 1929–1938, 2014.

[23] M. Valenzuela, V. Ha, and O. Etzioni, "Identifying meaningful citations," in *Proceedings of the in Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[24] X. Bai, F. Xia, I. Lee, J. Zhang, and Z. Ning, "Identifying anomalous citations for objective evaluation of scholarly article impact," *PLoS ONE*, vol. 11, no. 9, Article ID e0162364, 2016.

[25] R. Liang and X. Jiang, "Scientific ranking over heterogeneous academic hypernetwork," in *Proceedings of the in Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pp. 20–26, 2016.

[26] S. Wang, S. Xie, X. Zhang, Z. Li, P. S. Yu, and Y. He, "Coranking the future influence of multiobjects in bibliographic network

through mutual reinforcement," *ACM Transactions on Intelligent Systems and Technology*, vol. 7, no. 4, article no. 64, 2016.

[27] Y. Li, C. Wu, X. Wang, and P. Luo, "A network-based and multiparameter model for finding influential authors," *Journal of Informetrics*, vol. 8, no. 3, pp. 791–799, 2014.

[28] J. D. West, M. C. Jensen, R. J. Dandrea, G. J. Gordon, and C. T. Bergstrom, "Author-level eigenfactor metrics: Evaluating the influence of authors, institutions, and countries within the social science research network community," *Journal of the Association for Information Science and Technology*, vol. 64, no. 4, pp. 787–801, 2013.

[29] X. Cao, Y. Chen, and K. J. Ray Liu, "A data analytic approach to quantifying scientific impact," *Journal of Informetrics*, vol. 10, no. 2, pp. 471–484, 2016.

[30] J. Zhang, F. Xia, W. Wang et al., "Cocarank: A collaboration caliber-based method for finding academic rising stars," in *Proceedings of the International Conference Companion on World Wide Web*, pp. 395–400, Montral, Qubec, Canada, April 2016.

[31] D. Yu, W. Wang, S. Zhang, W. Zhang, and R. Liu, "A multiple-link, mutually reinforced journal-ranking model to measure the prestige of journals," *Scientometrics*, vol. 111, no. 1, pp. 521–542, 2017.

[32] D. Fiala, L. Šubelj, S. Žitnik, and M. Bajec, "Do PageRank-based author rankings outperform simple citation counts?" *Journal of Informetrics*, vol. 9, no. 2, pp. 334–348, 2015.

[33] T. Amjad, Y. Ding, A. Daud, J. Xu, and V. Malic, "Topic-based heterogeneous rank," *Scientometrics*, vol. 104, no. 1, pp. 313–334, 2015.

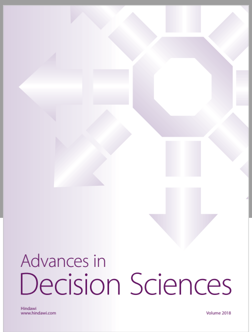
[34] Y. Wang, Y. Tong, and M. Zeng, "Ranking scientific articles by exploiting citations, authors, journals, and time information," in *Proceedings of the in Twenty-Seventh AAAI Conference on Artificial Intelligence*, pp. 933–939, 2013.

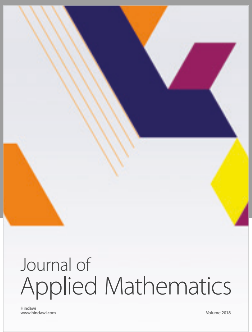
[35] T. Amjad, Y. Ding, J. Xu et al., "Standing on the shoulders of giants," *Journal of Informetrics*, vol. 11, no. 1, pp. 307–323, 2017.

[36] W. Wang, S. Yu, T. M. Bekele, X. Kong, and F. Xia, "Scientific collaboration patterns vary with scholars' academic ages," *Scientometrics*, vol. 112, no. 1, pp. 329–343, 2017.

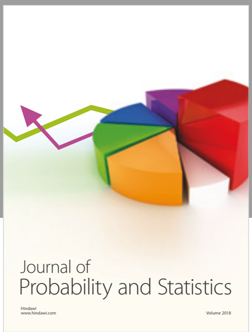
[37] J. Zhang, Z. Ning, X. Bai, W. Wang, S. Yu, and F. Xia, "Who are the Rising Stars in Academia?" in *Proceedings of the the 16th ACM/IEEE-CS*, pp. 211-212, Newark, New Jersey, USA, June 2016.

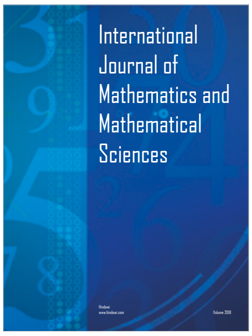
Advances in
Operations Research


Advances in
Decision Sciences


Journal of
Applied Mathematics

The Scientific
World Journal

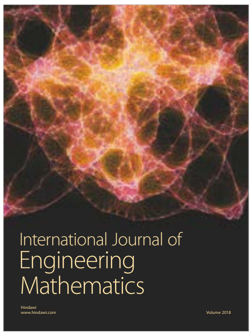

Journal of
Probability and Statistics


International
Journal of
Mathematics and
Mathematical
Sciences

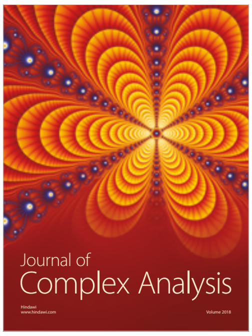
Hindawi

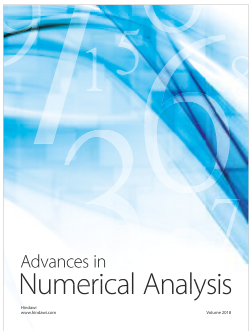
Submit your manuscripts at
www.hindawi.com

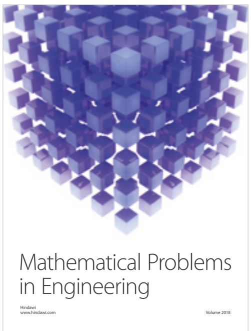
Journal of
Optimization

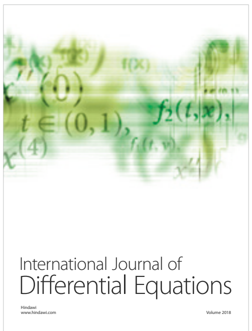

International Journal of
Engineering
Mathematics

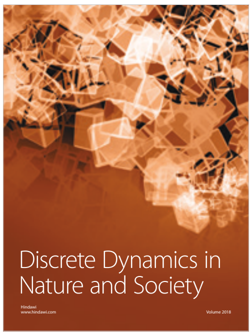
International Journal of
Analysis

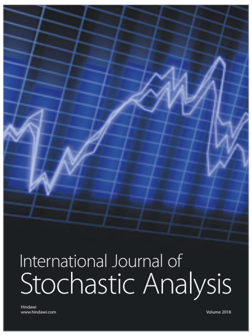

Journal of
Complex Analysis


Advances in
Numerical Analysis


Mathematical Problems
in Engineering

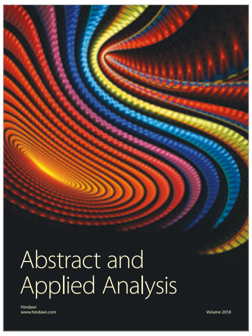

International Journal of
Differential Equations

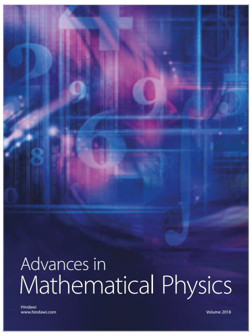

Discrete Dynamics in
Nature and Society


International Journal of
Stochastic Analysis

Journal of
Mathematics

Journal of
Function Spaces


Abstract and
Applied Analysis


Advances in
Mathematical Physics