Improved PageRank and New Indices for Academic Impact Evaluation Using AI Papers as Case Studies

Journal Title
XX(X):2–19

Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Rui Wang[1], Shijie Li[1], Qing Yin[1], Ji Zhang[2], Rujing Yao[1] and Ou Wu[1]

Abstract

Evaluating academic papers and groups is important in scholar evaluation and literature retrieval. However, current evaluation indices, which pay excessive attention to the citation number rather than the citation importance and unidirectionality, are relatively simple. This study proposes new evaluation indices for papers and groups. First, an improved PageRank (PR) algorithm introducing the citation importance is proposed to obtain a new citation-based paper index (CPI) via a pre-ranking and fine-tuning strategy. Second, to evaluate the paper influence inside and outside its research field, the focus citation-based paper index (FCPI) and diversity citation-based paper index (DCPI) are proposed based on topic similarity and diversity, respectively. Third, aside from the statistical indices for academic papers, we propose foreign academic degree of dependence (FAD) to characterize the dependence between two academic groups. Finally, Artificial Intelligence (AI) papers from 2005 to 2019 are utilized for a case study.

Keywords

PageRank, Paper Evaluation, Group Evaluation

Introduction and related work

Evaluation of papers, researchers, and groups (countries, institutes, and journals) has received much attention from government agencies, research organizations, individual researchers, and so on[1]. Scientists and institutions still measure the capabilities of scientific research through quantitatively[2], which is also a scientific evaluation based on indicators. Well-known evaluation indices include h-index[3] (for researchers), impact factor[4] (for journals), and others[5–9].

The existing academic evaluating methods, which are called citation-based paper indices (CPIs) in this paper, evaluate each paper according to the citation count[10,11]. However, those methods ignore some useful information such as the importance of citing papers and the evaluation results are easily affected by invalid citations[12]. Gradually, PageRank (PR), which is proposed by Page et al.[13] to evaluate the importance of webpages on the basis of the web link structure, is used to evaluate the influence of papers based on the citation network. Chen et al.[14] applied PR to assess the relative importance of all publications about physic. Ma et al.[15] used PR to evaluate the importance of scientific papers based on the citation network.

Many studies have focused on the improvement of PR. Du et al.[12] did not use direct citations between papers, but applied relativity measures to retrieve indirect relationships between papers. Ding et al.[16] proposed weighted PR, an approach that considers the citations and publications of authors as reference weights. Yan et al.[17] allocated a different weight to each reference by taking into account the impact of citing journals and citation time intervals. Wang et al.[18] used the time factor to ensure that the newly published papers have higher evaluation scores. Fiala[19] utilized the author network and citation year to improve PR. Nykl et al.[20] found that combining publish network with author network would evaluate a research better. Other methods of evaluating papers were discussed by Hirsch[21], Schubert et al.[3], Waltman et al.[22], Ye et al.[23], and Ding et al.[24].

Although PR is the most widely used algorithm in paper evaluation, this method has two main limitations. First, edge weights (i.e., citation importance)

[1]Tianjin University, China
[2]Zhejiang Lab, China

Corresponding author:
Ou Wu, Center for Applied Mathematics, Tianjin University, Tianjin, China.
Email: wuou@tju.edu.cn

are ignored during the construction of the citation network in almost all exiting studies, weakening some highly important citations. Second, the average CPI for each year shows a downward trend, meaning the influence of papers decreasing year by year, which is not reasonable.

Simultaneously, Min et al.[25] investigated the multilevel influence of papers in terms of depth. However, for multilevel depth influence, even if there is only a three-level citation cascades (e.g., paper $\mathcal{A}$ cites paper $\mathcal{B}$ and paper $\mathcal{B}$ cites paper $\mathcal{C}$), it is difficult to evaluate the influence degree that each end point (such as paper $\mathcal{C}$) has on the starting point (such as paper $\mathcal{A}$) in the reference chain. Gerrish et al.[9] and Gotelli et al.[26] studied the out-of-discipline influence of the paper in terms of width. A discipline, such as AI, still consists of many distinct fields. Papers cited by many other fields in the same discipline should have high values in the width dimension.

In addition, some studies have been conducted on the evaluation of academic groups. For the evaluation of journals, Garfield[27] used impact factor, which is the yearly average citation number of the papers published in a given journal and is the most common metric. Pajíc[28] used the 5-year impact factor, which is calculated based on 5-year citation windows. Moed et al.[29] used the Source Normalized Impact per Paper (SNIP) to measure a journal's contextual citation impact. González[30] used the SCImago Journal Rank (SJR), which takes into account not only the prestige of the citing scientific journal but also its relevance to the cited journal. When evaluate the universities, researchers usually directly use two indices that indicate the number of articles published by the group and the per capita citation rate (e.g. Times Higher Education World University Rankings[*] and QS World University Rankings[†]). Meyérs et al.[31] directly used the h-index to rank academic institutes.

Based on the above researches, this study makes the following improvements. For academic papers, this study introduces the citation importance and proposes an improved CPI method that inspired by the ideas of pre-training and fine-tuning in deep learning to improve the existing PR toward the citation unidirectionality. To measure the influence of a paper in depth and in width, an in-field focus citation-based paper index (FCPI) and an out-of-field diversity citation-based paper index (DCPI) are proposed, respectively. For academic groups, inspired by the foreign trade degree of dependence in international trade,

[*] http://www.timeshighereducation.co.uk/world-university-rankings/
[†] https://www.topuniversities.com/qs-world-university-rankings

a measurement named foreign academic degree of dependence (FAD) is proposed to quantify the academic dependence among groups.

Methodology

Limitations of the existing CPIs

We suppose that the academic database $\mathcal{P} = \{p_1, p_2, \ldots, p_n, \ldots, p_N\}$ contains N papers. The intersection of the reference set of p_n and $\mathcal{P}$ is denoted as $Out(n)$. Let $In(n)$ denote the intersection of the citation set of paper p_n and $\mathcal{P}$. Given p_n, suppose $p_m \in In(n)$, then the citation importance of paper p_n in p_m is $w_{m,n}$.

Existing CPI methods construct a citation network and apply PR to calculate the PR score (i.e., CPI) of each node (i.e., paper). In PR, each paper is assigned an initial value. Through iterations, PR updates the CPI value of each node until convergence. For a paper p_n, the iteration of PR is expressed as following:

$$CPI(p_n) = \frac{d}{N} + (1-d) \sum_{p_m \in In(n)} \frac{1}{|Out(m)|} CPI(p_m), \tag{1}$$

where d is the damping factor, which represents the probability that one paper randomly cites the other.

Existing CPI calculation methods have two main limitations. First, the construction of the citation network ignores the edge weight reflecting citation importance and assumes that each reference has the same influence on the citing paper. When the CPI of p_n is calculated, all the associated weights are $1/|Out(m)|$. However, the citation importance of the cited papers varies significantly in a citing paper (e.g., refs.[8] and[12] in this paper).

Second, theoretically, the average scientific values of papers in each year should be roughly identical. However, according to PR, papers published earlier usually have higher PR scores than those published later because the citation network is time unidirectional (citations only exist from later to earlier papers). We conduct simulated experiments on three randomly generated citation networks. In the simulated experiments, 20, 000 nodes and about 110, 000 directed edges are randomly generated to construct the citation network. Nodes represent papers. Directed edges refer to citation. The average CPI (i.e., PR) scores of papers published in each year are reported in Figure 1. The scores decrease over the years, which illustrates that the time unidirectionality leads the indices vary.

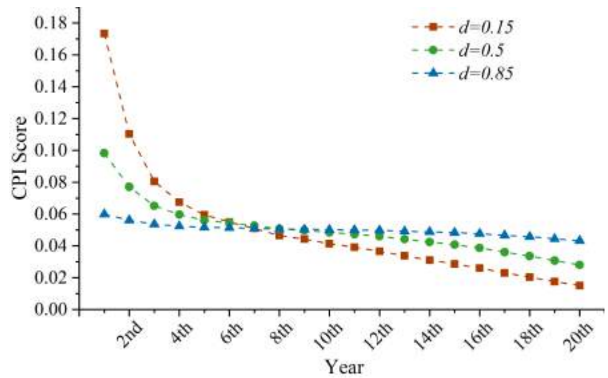

Figure 1. Results of simulation experiments. The horizontal axis represents the year of simulation, from the first year to the twentieth year. With different damping factors, the average CPI score change of papers published in each year.

The proposed CPI

For the first limitation, inspired by Lu et al.[32], Ding et al.[33], and Hu et al.[34], the following factors are used to measure citation importance: the average length of citation description, the number of citation mentions, and the position of citation mentions. In general, the importance of references mentioned in the method and experiment parts of AI papers is higher. The formula of citation importance is

$$w_{m,n} = \sum_{h=1}^{\mathcal{H}_{m,n}} loc_h l_h Pr_h, \tag{2}$$

where $\mathcal{H}_{m,n}$ represents the total number of times that p_n is mentioned in p_m, h denotes that p_n is mentioned for the hth time in p_m, loc_h is the score of the position of the hth mention, l_h is the citation description length of the hth mention, and Pr_h is the proportion of the hth mention in the description. For example, if three documents are cited in the current sentence, $Pr_h = \frac{1}{3}$.

For the second limitation, inspired by the pre-training and fine-tuning strategy[35] in deep learning, a new strategy called pre-ranking and fine-tuning is proposed[‡], as shown in Figure 2. This strategy is based on the assumption

[‡]In deep learning tasks, when the datasets are not big enough, the pre-training and fine-tuning strategy can help train a better model. Pre-ranking is to obtain initial PR scores in a citation subnetwork within a given time interval, which can weaken the negative impact of time unidirectionality. Fine-tuning is to obtain the final PR scores in the overall citation network based on the initial PR scores. These two steps can use local citation information as well as global citation information to improve the ability to measure the impact of a paper.

that the citation count for the papers published in a given year generally remains stable after $\mathscr{Y}$ years. In the pre-ranking step, for papers published in the mth year, the citations in the next $\mathscr{Y}$ years are collected to form a citation subnetwork (with intervals). The CPIs of nodes in this interval network are obtained and only the CPIs of the papers published in the mth year are retrained. For the last $\mathscr{Y}$ years, the CPI values of all papers are calculated and retained. Then, the CPI values are linearly adjusted to ensure the equality of the average values of the papers in each year. The adjusted CPI values are marked as CPI_0. In the fine-tuning step, CPI_0 is used as the initial value, and the final CPI values are obtained based on the citation network in all years. When normalizing the rows of the network matrix, we give a relatively small weight to all other uncited papers in the database. Finally, the improved CPI is calculated as following:

$$CPI(p_n) = \frac{dCPI_0(p_n)}{\sum_{p_j \in \mathcal{P}} CPI_0(p_j)} + (1-d) \sum_{p_m \in In(n)} w_{m,n} CPI(p_m), \tag{3}$$

where $CPI_0(p_j)$ is obtained by pre-ranking.

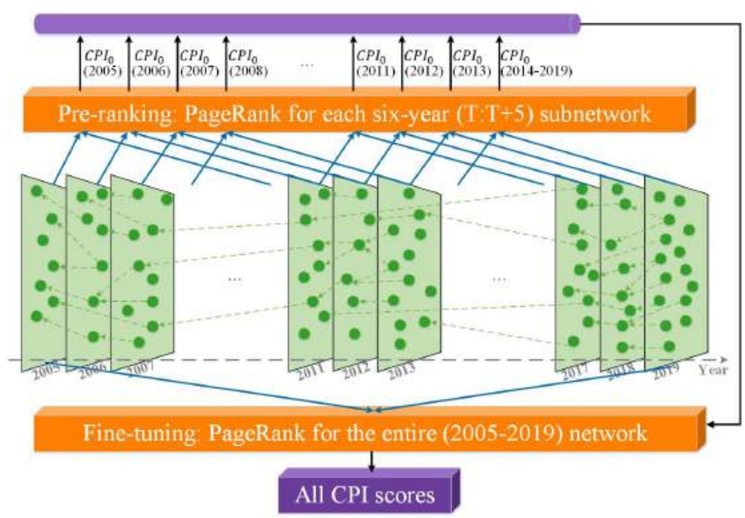

Figure 2. Schematic diagram of pre-ranking and fine-tuning.

Figure 3 is a comparison chart of the CPIs before and after the improvement. With the increase of d, CPI values by conventional PR are forced to be equivalent in different years. Our improved method can better keep the average values of papers published in years between 2015 and 2019 relatively fixed. Thus, the negative impact of unidirectional time is mitigated.

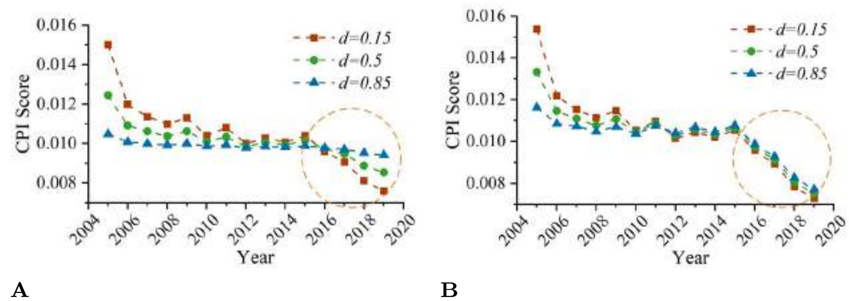

Figure 3. The results of the conventional PR with formula (1) are presented in A; Those of our improved PR with formula (3) are presented in B.

Focus and diversity citation-based paper indices

Focus and diversity citation-based paper indices rely on Latent Dirichlet Allocation (LDA)[36] analysis. Given a (paper) document set $\mathcal{D}$, LDA assumes that each paper is composed of several topics in a topic set denoted as $\mathcal{T} = \{t_1, t_2, \ldots, t_z\}$. Each topic consists of several keywords, denoted as $t_i = \{t_{i_1}, t_{i_2}, \ldots, t_{i_m}\}$. LDA generates a topic vector for each topic. Topic similarity is used in index calculation. However, the topic similarity calculated directly by LDA topic vectors are small in practice due to the sparsity of vectors. Therefore, deep learning is used to improve the generation of topic vectors. First, a 300-dimensional vector is obtained for each word by GloVe[37]. Then, the average word vector of the topic words contained in a topic is used as the topic vector.

The first index FCPI measures the influence of a paper on its research field. Research fields can be regarded as a more fine-grained division for science than disciplines. Even in a discipline, several research fields exist. To avoid artificially defining the research fields, this study defines FCPI based on the probability of two papers belonging to the same field:

$$FCPI(p_n) = \sum_{p_m \in In(n)} P(f(p_m) = f(p_n) | p_n, p_m) w_{m,n} CPI(p_m), \qquad (4)$$

where $f(p_n)$ represents the field of p_n, and $P(f(p_m) = f(p_n)|p_n, p_m)$ is the probability that papers p_m and p_n belong to the same research field. As p_m and p_n can be regarded as being composed of many different topics, the probability of the two in the same field can be calculated as following:

$$\begin{aligned}
&P(f(p_m) = f(p_n)|p_m, p_n) \\
&= \sum_{i,j\in\{1,2,\ldots,z\}} P(f(p_m) = f(p_n)|t_i, t_j)P(t_i, t_j|p_m, p_n) \\
&= \sum_{i,j\in\{1,2,\ldots,z\}} P(f(p_m) = f(p_n)|t_i, t_j)P(t_i|p_m)P(t_j|p_n) \\
&= \sum_{i,j\in\{1,2,\ldots,z\}} Sim(t_i, t_j)P(t_i|p_m)P(t_j|p_n),
\end{aligned} \tag{5}$$

where $P(t_i|p_m)$ is the probability that the topic t_i is in p_m. $Sim(t_i, t_j)$ is the topic similarity of topics t_i and t_j. z refers to the total number of topics. In LDA, $P(t_i|p_m) = \frac{n_{p_m,t_i}}{n_{p_m}}$, where n_{p_m,t_i} is the number of words that belong to topic t_i in p_m, and n_{p_m} is the total number of words in p_m.

In summary, the FCPI of the paper is:

$$FCPI(p_n) = \sum_{p_m\in In(n)} \sum_{i,j\in\{1,2,\ldots,z\}} Sim(t_i, t_j)P(t_i|p_m)P(t_j|p_n)w_{m,n}CPI(p_m). \tag{6}$$

The second is DCPI. Intuitively, if a paper p_n has a stronger impact on the papers outside of its research field and simultaneously the number of fields in $In(n)$ is greater, then the DCPI of the paper is higher. Accordingly, DCPI is defined as following:

$$DCPI(p_n) = \sum_{p_m\in In(n)} (1 - P(f(p_m) = f(p_n)|p_n, p_m))w_{m,n}CPI(p_m)exp(d_{f_n}), \tag{7}$$

where $exp(d_{f_n})$ is diversity factor, characterizing the diversity of the field referred in $In(n)$ and $exp(d_{f_n}) \in [1, e]$. d_{f_n} is defined as following:

$$\begin{aligned}
d_{f_n} &= \frac{2}{v_n(v_n-1)} \sum_{m_1=1}^{v_n} \sum_{m_2>m_1}^{v_n} \\
&P(f(p_{m_1}) \neq f(p_{m_2})|f(p_{m_1}) \neq f(p_n), f(p_{m_2}) \neq f(p_n), p_{m_1}, p_{m_2}, p_n) \\
&\approx \frac{2}{v_n(v_n-1)} \sum_{m_1=1}^{v_n} \sum_{m_2>m_1}^{v_n} \sum_{i_1,i_2,j\in\{1,2,\ldots,z\}} P(t_{i_1}|p_{m_1})P(t_{i_2}|p_{m_2})P(t_j|p_n) \\
&\frac{(1 - Sim(t_{i_1}, t_{i_2}))(1 - \frac{1}{2}Sim(ti_2, t_j))}{1 - Sim(t_{i_2}, t_j)},
\end{aligned} \tag{8}$$

where v_n is the number of papers citing paper p_n, other symbols are the same to the Section **Limitations of the existing CPIs**. When $v_n = 1$, let $exp(d_{f_n}) = 1$. When $v_n \geq 2$, the greater the probability that two arbitrary citing papers do not belong to the same field, the wider the fields p_n diverges into. That is, the diversity factor $exp(d_{f_n})$ is larger.

In the formula above, we use approximate calculation in mathematics. Let A be $f(p_{m_1}) \neq f(p_{m_2})$, B be $f(p_{m_1}) \neq f(p_n)$, and C be $f(p_{m_2}) \neq f(p_n)$. Let $P(A) = 1 - S_1$, $P(B) = 1 - S_2$, $P(C) = 1 - S_3$, then we have

$$\begin{aligned} &P(AB) \geq P(ABC) \geq P(A)P(B)P(C) \\ &P(AB) = (1 - S_1)(1 - S_2) \\ &P(A)P(B)P(C) = (1 - S_1)(1 - S_2)(1 - S_3). \end{aligned} \tag{9}$$

Then we have

$$P(ABC) \approx \frac{1}{2}(P(AB) + P(A)P(B)P(C)) = (1 - S_1)(1 - S_2)(1 - \frac{1}{2}S_3). \tag{10}$$

Evaluation indices for academic groups

Statistical values (such as average) of CPI, FCPI, and DCPI are usually chosen for group evaluation. However, they are inappropriate to evaluate the dependence degree between two groups. A comparison for two groups is usually necessary. Economics uses the foreign trade degree (FTD) of dependence to characterize the dependence of an economy on imports and exports of another economy within a certain period. The import dependence is calculated by the ratio of imports to GDP. The export dependence is calculated by the ration of exports to GDP. Inspired by this idea, FAD is proposed to characterize the degree of citation dependence between two different academic groups.

Given two groups $\mathscr{A}$ and $\mathscr{B}$, the other groups except $\mathscr{A}$ are recorded as $\mathscr{A}_{others}$. The FAD of $\mathscr{A}$ on $\mathscr{B}$ is defined as the ratio of the influence of $\mathscr{B}$ on $\mathscr{A}$ to the influence of $\mathscr{A}_{others}$ on $\mathscr{A}$ within a certain period of time. The FAD of $\mathscr{A}$ on $\mathscr{B}$ is

$$FAD_{\mathscr{A},\mathscr{B}} = \frac{\sum_{p_m \in \mathscr{A}} \sum_{p_n \in \mathscr{B}} w_{m,n}}{\sum_{p_m \in \mathscr{A}} \sum_{p_n \in \mathscr{A}_{others}} w_{m,n}}, \tag{11}$$

where $p_m \in \mathscr{A}$ represents paper p_m is published by group $\mathscr{A}$. Obviously, if more papers in $\mathscr{A}$ cite papers in $\mathscr{B}$, the greater the impact of the cited papers on the citing papers, the greater the value of the $FAD_{\mathscr{A},\mathscr{B}}$.

Materials and methods

Pre-process for the paper documents

The GROBID[§] is utilized to convert PDF format papers into XML format with identified titles, authors, references, and formulas. A citation network can be constructed according to the XML data.

Among the five characteristics, paper type is determined according to the themes of the conferences/journals; topic popularity is introduced in the **Method** part; international cooperation is determined according to whether authors come from one country.

Data statistics

Figure 4 shows the published papers from different countries in 2005–2019. The number of papers with different characteristics is shown in Figure 5.

The papers are extracted form 53 conferences/journals, which are AAAI, NeurIPS, TPAMI, ACL, IJCV, CVPR, JMLR, ICCV, ICML, IJCAI, COLT, EMNLP, AAMAS, ECAI, ECCV, ICCBR, COLING, KR, TASLP, UAI, TFS, Journal of Automated Reasoning, Machine Learning, AISTATS, ACCV, Applied Intelligence, ACML, BMVC, NLPCC, CONLL, ICTAI, ALT, ICANN, FG, ICDAR, ILP, KSEM, ICONIP, ICPR, IJCNN, PRICAI, NAACL, IJDAR, Machine Translation, Machine Vision and Applications, Natural Computing, NCA, NPL, PAA, and Soft Computing.

Methods

Citation importance calculation Inspired by Altmetrics[38] assigning values to different indicators, we set the scores of the method part, related-work part, and the rest parts to 30, 5, and 16, respectively. If chapter names are not obvious, the body part is divided into three parts, and the starting, middle, end parts are set to 5, 30, and 16, respectively.

LDA implementation A fast and lightweight LDA, called LightLDA[39] is utilized to model the thesis data. First, the contents of the abstract, title, and keywords extracted from the database are converted into binary format data as input. In the training process, according to the calculation formula of the experience topic number, $\sqrt{N}/3$, which is the training topic number, is set to 115; the

[§]https://grobid.readthedocs.io/en/latest/

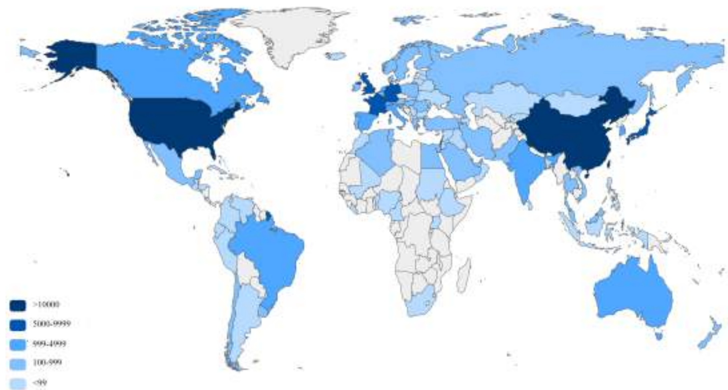

Figure 4. Numbers of papers in different countries. The top five countries in terms of number are the US, China, the UK, Germany, and Japan. The countries which are almost white in this figure publish few papers.

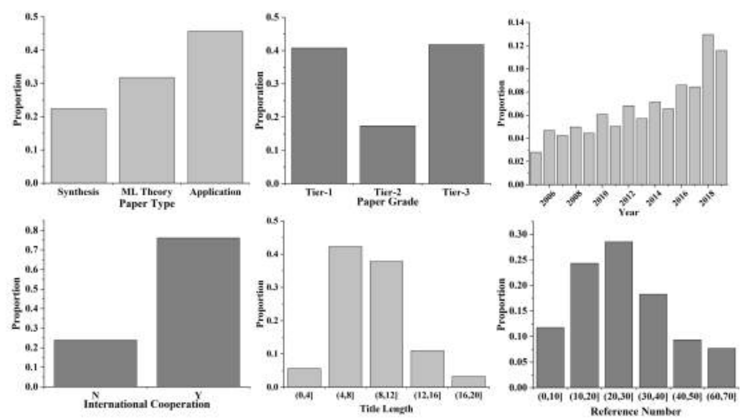

Figure 5. Distributions of papers on six characteristics.

iteration number is set to 1,000 times considering the time cost; the remaining parameters are the same as those used in LightLDA[39].

PageRank implementation This article refers to the analysis of[40], and d is set as 0.5. In the pre-ranking step, the data statistical result shows that most of the citations of a paper are concentrated within six years after its publication. Therefore, $\mathscr{Y} = 5$ is selected, that is, the subnetworks of every six years are utilized for pre-ranking. The value of $\mathscr{Y}$ is disciplinary dependent. For other disciplinary, $\mathscr{Y}$ need to be reset.

Topic popularity calculation The popularity of each topic is calculated by the formula provided by Pennington et al.[37]. For p_n, the topic popularity is obtained from the probability of the topics contained in the paper and the average topic popularity during five years before the paper.

Given a paper p_n, the "weighted popularity" of topics it contains is regarded as the paper's topic popularity. Assuming that the topics and associated probabilities in p_n are $\{t_1 : P(t_1|p_n), t_2 : P(t_2|p_n), \ldots, t_r : P(t_r|p_n)\}$, where r is the number of topics contained in p_n. Let the average popularity of each topic during 5 years before p_n published be $\{\overline{TP(t_1)}, \overline{TP(t_2)}, ..., \overline{TP(t_r)}\}$. Then the topic popularity of p_n is

$$TP_n = P(t_1|p_n)\overline{TP(t_1)} + P(t_2|p_n)\overline{TP(t_2)} + ... + P(t_r|p_n)\overline{TP(t_r)}. \qquad (12)$$

Results

Using the list of AI journals and conferences in China Computer Federation (CCF) ranks [¶] (Tier-1, Tier-2, and Tier-3), we collect a total of 122, 525 papers published from 2005 to 2019. According to the above indices, the following results are obtained.

Data distribution of different indices

Figures 6A, 6B, and 6C show the distributions of the three indices for papers with citation count greater than four (for convenience, all index values are expanded by 1, 000 times). CPI, FCPI, and DCPI generally follow the long tail distribution.

Order difference of different indices

In this section, the relationships of the order changes among different indices and five characteristics of academic papers are analyzed to effectively understand the above three indices. The five characteristics are paper grade according to CCF rankings (Tier-1, Tier-2, and Tier-3), paper type (namely, synthesis, machine learning theory, application such as computer vision and natural language processing), international cooperation, topic popularity, and citation count. Three order differences are rank (CPI) versus rank (PR), rank (FCPI) versus rank (CPI), and rank (DCPI) versus rank (CPI). For example, rank (CPI) versus

[¶] CCF compiled a list of AI journals and conferences with different ranks.

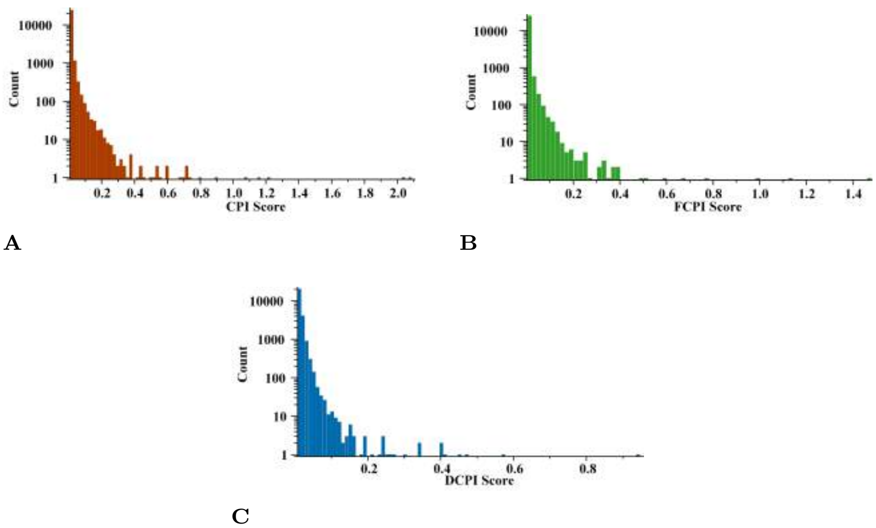

Figure 6. A. Distribution histogram of CPI. B. Distribution histogram of FCPI. C. Distribution histogram of DCPI.

rank (PR) of a paper is the paper's rank in the improved CPI minus its rank in PR. Figure 7 shows the average order differences of papers ($|In(p_n)| >= 5$) with different characteristics.

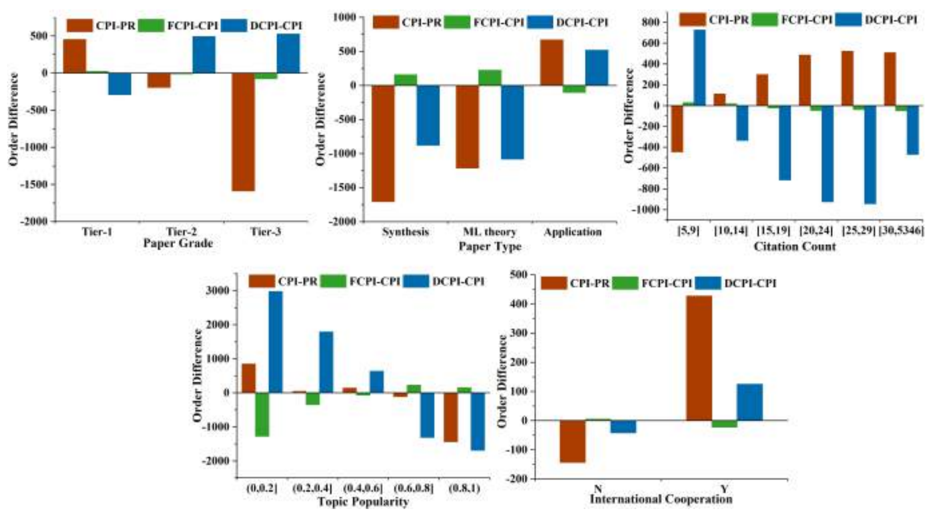

Figure 7. Average net increase of three sequences under five characteristics. Under various characteristics, the order of the paper changes significantly.

The ranking of papers in the improved CPI has the following changes: the average ranking of Tier-1 papers increases, the ranking of applied papers increases, the average ranking of papers with international cooperation increases, and the ranking of papers with lower topic popularity increases. The bigger the citation number is, the greater increasement the ranking of the paper gains. When the citation count is greater than 30, the increase rate drops partially because when the citation count increases, the ranking of papers is increased, but the change in order and the increasement are reduced.

Compared with CPI, the ranking results of the papers in FCPI have the following changes: Tier-3 papers have the highest increase in FCPI order. Machine learning (ML) papers have the highest increasement in FCPI order. The FCPI order of papers without international cooperation is higher than papers with international cooperation; with the rise of topic popularity, the FCPI order continues to increase. The order change of DCPI is basically the opposite of that of FCPI: Tier-1 papers have the highest increase in DCPI order; among various types, application papers have the highest increase in DCPI order; papers with international cooperation have a higher increase in DCPI order; as the popularity of the paper increases, the DCPI order decreases. As the citation count rises, both rankings decrease, but the influence of citation count on the DCPI order is much greater than that on FCPI order.

Indices of different groups

The average FCPI and DCPI of the papers in different groups can show the average focus and diversity level of the group. This section shows the average CPI, FCPI, and DCPI of different countries, and various journals and conferences. The annual change of conferences/journals are shown in the supplementary material.

Figures 8A, 8B, and 8C report the average CPI, FCPI, and DCPI of the top five countries including the United States, China, the United Kingdom, Germany, and Japan. As of 2015, China's CPI, DCPI, and FCPI have shown an upward trend, indicating that China's development is faster and the gap with other countries is decreasing.

Figure 9D shows the rank of 15 conferences and journals according to the average CPI, average FCPI, and average DCPI. We can observe that computer vision (CV) papers are more influential than others. In fact, CV leads the development trend in AI development.

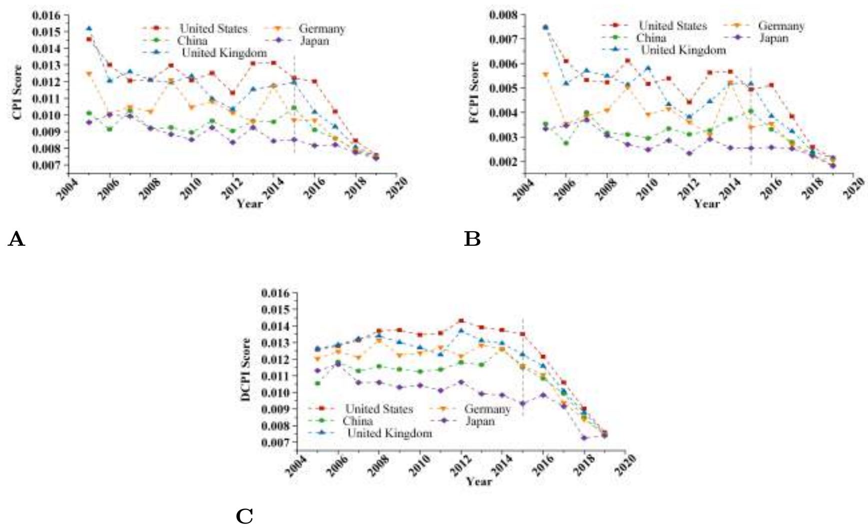

Figure 8. A. Average CPI changes in the top 5 countries by year. As of 2015, the average CPI value in the United States is higher and that in Japan is lower. The figure for the United States is basically unchanged, and the volatility for the United Kingdom, Japan, and Germany has declined. B. Annual average FCPI changes in top 5 countries. C. Annual average DCPI changes in top 5 countries. As of 2015, the average DCPI score for the United States, the United Kingdom, and Germany have remained unchanged.

Results of FAD

Comparative studies on several typical countries and conferences have been conducted. Five countries with the most published papers are considered. Top 15 journals and conferences according to ranks of average CPI are considered.

As shown in Figure 9A, other countries all have higher FAD on the US based on the limit data set. The US has the highest FAD on China compared with other countries. See the supplementary material for more results. Figure 9B is an example of FADs between China and the US. The FAD of the US on China is getting higher, but China dependents on the US more in general.

As shown in Figure 9C, between International Conference on Machine Learning (ICML) and Neural Information Processing Systems (NeurIPS), ICML depends on NeurIPS more. Similarly, Annual Meeting of the Association for Computational Linguistics (ACL) depends on NeurIPS more, and ACL depends on ICML more. See the supplementary material for more results.

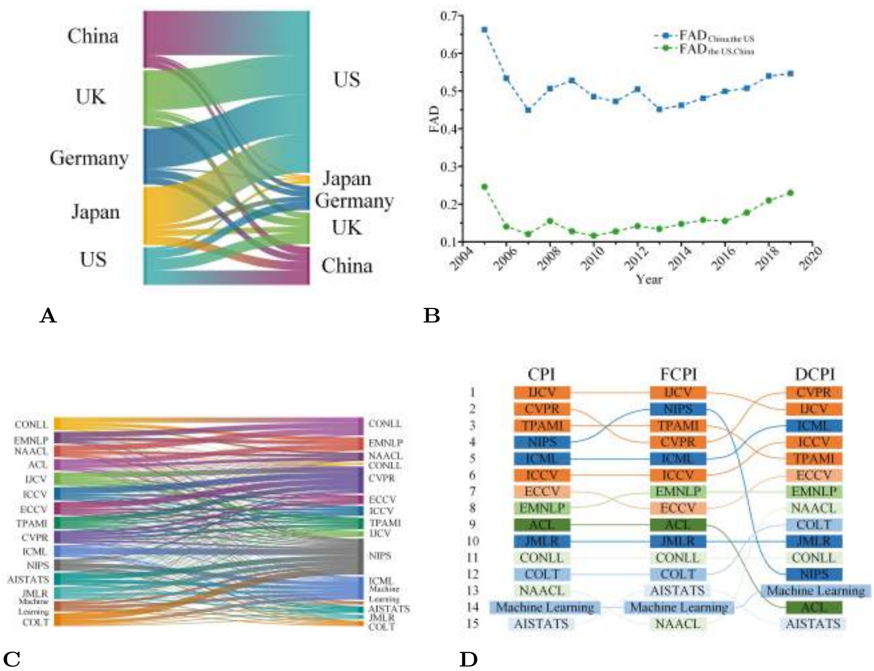

Figure 9. A. FADs of the left countries on the right countries. B. FADs between China and the US. C. FADs of conferences/journals on the left on conferences/journals on the right. D. The top 15 journals and conferences are ranked in descending order of average CPI, FCPI, and DCPI.

Conclusion

This study considers citation importance and introduces deep learning ideas to improve the standard PR algorithm in citation-based paper evaluation. FCPI and DCPI are proposed to measure paper's influence inside and outside its research field. Instead of using the average of the single literature index for group evaluation, FAD is proposed, which can accurately and comprehensively describe the influence dependence between two groups.

The results of this study on the limited data set show that according to CPI values, the average ranking of CCF Tier-1 papers is indeed higher than papers in other tiers. Compared with the previous CPI, our CPI can more naturally overcome the negative effects caused by the time unidirectionality of the citation network. The CPI proposed by this paper is better than PR when measuring the influence of papers. Simultaneously, FCPI and DCPI can further measure the influence of papers inside and outside the field. The result of the dependence of the academic groups shows that the US still leads the AI research and other countries have high dependence on the US. Though this study uses AI papers as a case study, the indices in this paper can be applied to other disciplines as well. The disciplinary complication is a challenge to a feasible use of the current method.

References

1. Zhang F and Wu S. Predicting future influence of papers, researchers, and venues in a dynamic academic network. *Journal of Informetrics* 2020; 14(2): 101035.
2. Chu JS and Evans JA. Slowed canonical progress in large fields of science. *PNAS* 2021; 118(41).
3. Schubert A. Using the h-index for assessing single publications. *Scientometrics* 2009; 78(3): 559–565.
4. Garfield E. The impact factor. *Current contents* 1994; 25(20): 3–7.
5. Garfield E. Citation indexing for studying science. *Nature* 1970; 227: 669–671.
6. Garfield E. Citation indexes for science: A new dimension in documentation through association of ideas. *Science* 1955; 122: 108–111.
7. Siudem G, Żogała-Siudem B, Cena A et al. Three dimensions of scientific impact. *PNAS* 2020; 117(25): 13896–13900.
8. Singh AP, Shubhankar K and Pudi V. An efficient ranking algorithm for scientific research papers. In *2011 3rd Conference on Data Mining and Optimization (DMO)*. pp. 88–95.
9. Gerrish S and Blei DM. A language-based approach to measuring scholarly impact. *ICML* 2010; : 375–382.
10. Jiang X, Sun X, Yang Z et al. Exploiting heterogeneous scientific literature networks to combat ranking bias: Evidence from the computational linguistics area. *Journal of the Association for Information Science and Technology* 2016; 67(7): 1679–1702.
11. Wang S, Xie S, Zhang X et al. Coranking the future influence of multiobjects in bibliographic network through mutual reinforcement. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2016; 7(4): 1–28.
12. Du M, Bai F and Liu Y. Paperrank: A ranking model for scientific publications. *2009 World Congress on Computer Science and Information Engineering* 2009; : 277–281.
13. Page L, Brin S, Motwani R et al. The pagerank citation ranking: Bringing order to the web. *Technical Report* 1998; .

14. Chen P, Xie H, Maslov S et al. Finding scientific gems with google's pagerank algorithm. *Journal of Informetrics* 2007; 1(1): 8–15.
15. Ma N, Guan J and Zhao Y. Bringing pagerank to the citation analysis. *Information Processing Management* 2008; 44(2): 800–810.
16. Ding Y, Yan E, Frazho A et al. Pagerank for ranking authors in co-citation networks. *Journal of the American Society for Information Science and Technology* 2009; 60(11): 2229–2243.
17. Yan DY Erjia. Weighted citation: An indicator of an article's prestige. *Journal of the American Society for Information Science and Technology* 2010; 61(8): 1635–1643.
18. Wang Y, Tong Y and Zeng M. Ranking scientific articles by exploiting citations,authors, journals, and time information. *AAAI* 2013; : 800–810.
19. Fiala D. Time-aware pagerank for bibliographic networks. *Journal of Informetrics* 2012; 6(3): 370–388.
20. Nykl M, Ježek K, Fiala D et al. Pagerank variants in the evaluation of citation networks. *Journal of Informetrics* 2014; 8(3): 683–692.
21. Hirsch JE. An index to quantify an individual's scientific research output that takes into account the effect of multiple coauthorship. *Scientometrics* 2010; 85(3): 741–754.
22. Waltman L and Van Eck NJ. The inconsistency of the h-index. *J Am Soc Inf Sci* 2012; 63(2): 406–415.
23. Ye FY and Leydesdorff L. The "academic trace" of the performance matrix: A mathematical synthesis of the h-index and the integrated impact indicator (i3). *J Assoc Inf Sci Technol* 2014; 65(4): 742–750.
24. Ding J, Liu C and Kandonga GA. Exploring the limitations of the h-index and h-type indexes in measuring the research performance of authors. *Scientometrics* 2020; 122: 1303–1322.
25. Min C, Sun J and Ding Y. Quantifying the evolution of citation cascades. *Proceedings of the Association for Information ence Technology* 2017; 54(1): 761–763.
26. Gotelli NJ and Chao A. Measuring and estimating species richness, species diversity, and biotic similarity from sampling data. *Encyclopedia of Biodiversity* 2013; 5: 195–211.
27. Garfield E. Citation analysis as a tool in journal evaluation. *Science* 1972; 178(4060): 471–479.
28. Pajíc D. On the stability of citation-based journal rankings. *Journal of Informetrics* 2015; 9(4): 990–1006.

29. Moed HF. Measuring contextual citation impact of scientific journals. *Journal of Informetrics* 2010; 4(3): 265–277.
30. Borja González-Pereira FMA Vicente PGuerrero-Bote. A new approach to the metric of journals' scientific prestige: The sjr indicator. *Journal of Informetrics* 2010; 4(3): 379–391.
31. Meyérs MA and Quan H. The use of the h-index to evaluate and rank academic departments. *Journal of Materials Research and Technology* 2017; 6(4): 304–311.
32. Lu C, Ding Y and Zhang C. Understanding the impact change of a highly cited article: a content-based citation analysis. *Scientometrics* 2017; 112: 927–945.
33. Ding Y, Liu X, Guo C et al. The distribution of references across texts: Some implications for citation. *Journal of Informetrics* 2013; 7(3): 583–592.
34. Hu Z, Chen C and Liu Z. Where are citations located in the body of scientific articles? a study of the distributions of citation locations. *Journal of informetrics* 2013; 7(4): 887–896.
35. Devlin J, Chang MW, Lee K et al. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:181004805* 2018; .
36. Blei DM, Ng AY and Jordan MI. Latent dirichlet allocation. *J Mach Learn Res* 2003; 3: 993–1022.
37. Pennington J, Socher R and Manning C. GloVe: Global vectors for word representation. *Proc Conf Empirical Methods Natural Language Processing* 2014; 14: 1532–1543.
38. Priem J, Taraborelli D, Groth P et al. Altmetrics: A manifesto 2010; .
39. Yuan J, Gao F, Ho Q et al. Lightlda: Big topic models on modest computer clusters. In *Proceedings of the 24th International Conference on World Wide Web*. pp. 1351–1361.
40. Xu J, Ding Y, Bu Y et al. Interdisciplinary scholarly communication: an exploratory study for the fiel. *Scientometrics* 2019; 119: 1597–1619.