



Revolutionizing scholarly impact: advanced evaluations, predictive models, and future directions

Xiaomei Bai^{1,2} · Fuli Zhang³ · Jiaying Liu⁴ · Xiaoxia Wang⁵ · Feng Xia⁶

Accepted: 25 June 2025 / Published online: 15 July 2025
© The Author(s) 2025

Abstract

Artificial intelligence (AI) is revolutionising scholarly impact evaluation and prediction. By integrating AI and machine learning techniques, researchers can leverage diverse academic networks and multiple sources of academic big data. This integration transforms traditional evaluation methods that rely on structured measurements such as citation counts and journal impact factors, into more comprehensive and objective evaluations. In this paper, we dive deep into latest advancements in scholarly impact evaluation and prediction within the context of AI. We categorize existing models, highlighting their similarities and distinctions, with a particular emphasis on AI-enabled approaches. Building upon the analysis, we discuss the ongoing challenges in scholarly impact research and outline future directions in this field.

Keywords Scholarly impact · Science of science · Artificial intelligence for science · Evaluation · Prediction

1 Introduction

With the help of information and artificial intelligence (AI) technologies, human beings' ability to master and process data has achieved a qualitative leap (Yu et al. 2023; Xia et al. 2022; Liu et al. 2018). The way humans obtain knowledge has undergone major changes, from paper to browsers, making the dissemination of knowledge more flexible, diverse, and timely. Scholarly big data refers to the large-scale and complex datasets generated from various scholarly activities and sources within the academic domain (Xia et al. 2017). Mining scholarly data helps scholars extract valuable knowledge and insights (Porwal and Devare 2024b; Liu et al. 2019, 2022). As an important research direction in scholarly big data, the evaluation and prediction of the impact of published papers can provide important data support for individual promotion, awards, and funding allocation (Bai et al. 2018; Dunaiski et al. 2019; Ramani et al. 2022). The wave of digitalisation drives the continued growth of scholarly big data (Xia et al. 2017; Dong et al. 2017), which also increases the difficulty of evaluating and predicting scholarly paper impact (Dunaiski et al. 2016; Vaccario et al. 2017;

Brito and Rodriguez-Navarro 2019; Siudem et al. 2020). Since the quantification of scholarly paper impact is the basis for evaluating other academic entities, including scholars, journals, and institutions, the evaluation of scholarly paper's impact has attracted significant attention of scholars and achieved remarkable results (Zhang et al. 2022; Bai et al. 2018; Radicchi et al. 2017).

Two important issues in scholarly paper impact research are the evaluation and prediction of paper impact (Zhou et al. 2021; Shaikh et al. 2023; Wang et al. 2023b). From a macro perspective, scholarly impact evaluation and prediction can guide national policy-makers in making reasonable decisions to maximize the utility of limited resources. From a micro perspective, scholarly impact evaluation and prediction provide important data support for researchers' promotions and transfers. The evaluation of scholarly paper impact measures past success, whereas the prediction of scholarly paper impact estimates a paper's future success. Evaluating and predicting the impact of scholarly papers is fundamental for evaluating the impact of academic entities, including researchers, institutions, journals, and country. Just as with citation counts, this single-dimensional measure of impact can no longer meet the needs of the rapid development of scholarly big data. A large number of self-citations and irrelevant citations may lead to the inflation of scholarly paper impact. The rise of academic media platforms and the dynamic evolution of academic social network have introduced new challenges to the research on evaluating and predicting academic impact. The most core research objective in this area is to scientifically and reasonably evaluate scholarly impact.

The evaluation of scholarly paper impact is an important research area in the discipline of science of science. Most previous evaluation methods of scholarly paper impact do not distinguish the citation weights (Wan and Liu 2014). In other words, the academic networks used for paper impact evaluation are unweighted. However, citations in academic papers are not equally important, and it is more appropriate to use weighted academic networks to evaluate the impact of academic papers (Bai et al. 2018; Cai et al. 2019). Factors such as the journals in which the paper is cited, the authors who cite it, and the papers in which it is cited are all taken into account as weights (Yan et al. 2011; Khatoon et al. 2024). Some researchers have also explored time-related weights to capture the evolution of academic networks (Wang et al. 2019). Nonlinearity is integrated into the algorithm to improve the robustness of the PageRank algorithm against malicious citations. The nonlinear ranking method favors papers cited by high-score papers but penalizes those cited by low-score papers (Yao et al. 2014). From a traditional one-dimensional perspective, citations are the most representative evaluation indicator of a scholarly paper impact. Based on the development of deep learning, especially deep graph learning, researchers have achieved significant results in evaluating the impact of scholarly papers from a multi-dimensional perspective (Bu et al. 2021).

Artificial Intelligence for Science (AI4Science) has gained consensus in the academic community as a new scientific research paradigm since it was first proposed in 2018 (Wang et al. 2023). Artificial intelligence, as a tool, can help scientists in various fields solve problems that are difficult to address with traditional methods, such as in physics, chemistry, and biology. "AI4Science" reflects the importance of interdisciplinary collaboration. Similarly, artificial intelligence also plays a significant role in the field of science of science. In the past ten years, with the AI technologies, the evaluation of scholarly paper impact has undergone major changes: (1) from using homogeneous academic networks to heterogeneous

academic networks, and then to heterogeneous academic hypernetworks (Liu et al. 2024; Zhang and Wu 2018; Kanellos et al. 2019; Bu et al. 2021a; (2) from unweighted academic network to weighted academic network (Zhang et al. 2019; Cai et al. 2019; (3) from no time-aware evaluation to time-aware evaluation (Zhang et al. 2019; (4) from linear evaluation to nonlinear evaluation (Yao et al. 2014; Pech and Delgado 2020; (5) from single-dimensional evaluation to multi-dimensional evaluation. The PageRank algorithm is introduced to the citation networks to evaluate the scholarly paper impact, realizing the transformation from unstructured evaluation to structured evaluation (Chen et al. 2007; Qiu and Han 2024). In recent years, researchers have continuously explored methods for evaluating scholarly paper impact from heterogeneous academic networks, and have made breakthrough progress (Zhao et al. 2019; Zhang and Wu 2021).

The process of predicting the scholarly paper impact can help identify those that are likely to be high-quality scholarly papers and provide timely assistance to researchers (Thelwall et al. 2023; Yu et al. 2014; Bai et al. 2019; Ruan et al. 2020). Over the past ten years, with the help of machine learning, deep learning, and large language models, paper impact prediction research has also undergone significant changes: (1) from feature-based predictive methods to generative prediction methods; (2) from simple features to complex features; (3) from machine learning models to deep learning models, and then to large language models. Predicting the impact of a scholarly paper requires exploring truly effective features, which is very challenging (Zhang and Wu 2024; Porwal and Devare 2024a; He et al. 2023). One challenge is identifying features related to the scholarly paper impact. Another challenge is combining these relevant features in a unified predictive model. Among the various metrics used to predict scholarly paper impact, citations are often considered the primary indicator (Ruan et al. 2020; Yan et al. 2011b; Huang et al. 2022a).

Paper impact evaluation and prediction have attracted significant interest over the past decade (Xiao et al. 2016; Mercier et al. 2020), although the methods and models used have changed over time (Kenton and Toutanova 2019). The features used in scholarly paper impact prediction model can be specifically divided into the following aspects: paper-related features, author-related features, venue-related features, institution-related features, network-related features, and Altmetrics-related features (Bhat et al. 2015; Zhu and Ban 2018; Thelwall and Nevill 2018). When these features are integrated into AI models, paper impact prediction can be achieved. Representative AI models for predicting paper impact include support vector machines, neural networks, and XGBoost (Ruan et al. 2020; El Mohadab et al. 2019). Recently, some scholars have leveraged deep learning techniques to solve the problem of paper impact prediction (Zhao and Feng 2022; Jiang et al. 2023; Ji et al. 2024; Feng et al. 2023). In addition, ChatGPT has been used to predict citation counts (de Winter 2024).

Several recent works review research related to the evaluation and prediction of paper impact (Bai et al. 2017; Hou et al. 2019; Cai et al. 2019; Bai et al. 2020a; Xia et al. 2023). The early work mainly focuses on the introduction of key techniques for paper impact research (Bai et al. 2017), involving less relevant work for evaluating and predicting the impact of scholarly papers, and only includes 30 references. Recent work concentrates on the weighted citation solutions for the evaluation of scholarly entities including articles, scholars, and journals (Cai et al. 2019). Among these works, the citation weighting solutions for the article assessment section cover 13 relevant references. Another work (Bai et al. 2020a) emphasizes the factors influencing paper impact and evaluation methods, com-

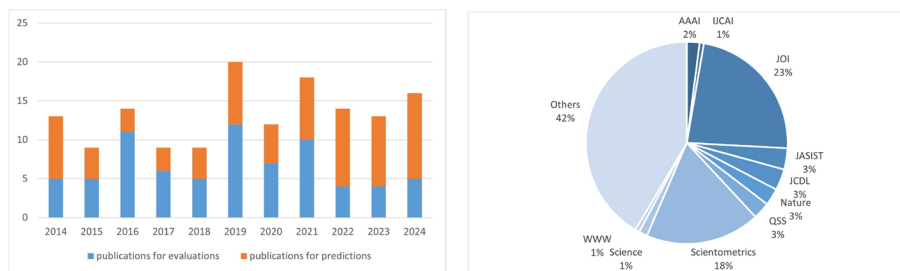
pletely excluding research related to the paper impact prediction. Two papers reveal the scientific impact prediction, focusing on the classification of features and methods (Hou et al. 2019; Xia et al. 2023). Overall, the gaps of these reviews lie in the limited scope of research coverage, insufficient number of references, and a lack of discussion related to large language models for scholarly impact evaluation and prediction.

The objective of this paper is to deliver a comprehensive overview of the revolutionary changes that artificial intelligence has introduced to the realm of scholarly impact evaluation and prediction. It is designed to equip those newly entering the research field with a profound comprehension, facilitating informed decisions about their future research direction. Furthermore, by discussing future directions in scholarly impact research, the paper aspires to offer valuable insights to the researchers. Compared to previous surveys, the contributions of this work are listed as follows:

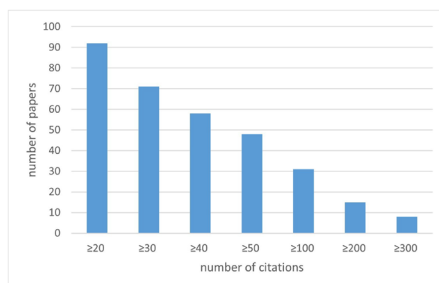
- We provide a comprehensive overview of the topic of academic paper impact assessment and prediction, compiling and organizing over 140 relevant papers for analysis.
- We supplement existing reviews at various levels, including the evolution, advantages, and limitations of different methodologies (Bai et al. 2017; Hou et al. 2019; Cai et al. 2019; Bai et al. 2020a; Xia et al. 2023).
- We focus on emerging technologies within the realm of AI for Science, particularly emphasizing extensive feature-driven approaches and generative models.

Figure 1 provides a summary of the statistics of the publications analyzed in this paper. Figure 1a shows the number of publications for paper impact evaluation and prediction over the last decade. It can be observed that research on the prediction of scholarly impact is increasingly drawing attention. Figure 1b illustrates the distribution of these publications across different journals. The Journal of Informetrics (JOI) accounts for 23% of the total, closely followed by Scientometrics. From Fig. 1c, it is evident that there are 92 publications that have been cited 20 times or more, 71 of which have been cited at least 30 times, and a total of 4 documents that have achieved a citation count of 300 or more. We conducted searches using relevant keywords such as 'academic impact,' 'paper impact evaluation,' 'paper impact prediction,' 'artificial intelligence for science,' etc. Based on this, we located the literature that cited these papers as well as the literature these paper cited. Subsequently, we filtered out the highly relevant papers to serve as references for this paper. The literature for this paper comes from high-ranking journals or conferences related to scientometrics such as Nature, Science, Journal of Informetrics (JOI), Scientometrics, Journal of the Association for Information Science & Technology (JASIST), Quantitative Science Studies (QSS), ACM/IEEE Joint Conference on Digital Libraries (JCDL), the Web Conference (WWW), and the International Joint Conference on Artificial Intelligence (IJCAI). More than 40% of the publications come from other high-quality journals and conferences such as IEEE Transactions on Big Data, Information Process and Management, and the International Conference on Information and Knowledge Management (CIKM).

The rest of this paper is structured as follows. Section 2 provides a detailed analysis of the evolution, advantages, and limitations of methods for paper impact evaluation. Section 3 introduces a variety of paper impact prediction models, covering feature-driven models, generative models, deep learning models and large language models. In Sect. 4, we discuss



(a) Number of publications for paper impact evaluation and prediction over the last decade. (b) Distribution of publication venues for paper impact evaluation and prediction.



(c) The relationships between the number of publications and the number of citations.

Fig. 1 Summary statistics for the corpus of publications analyzed. **a** Annual counts of papers on impact evaluation (blue) and impact prediction (orange) between 2014 and 2023. **b** Distribution of publication venues **c** Relationship between the number of publications and their citation counts

open issues and challenges in the field. The last section provides a summary and future directions for our research.

2 Paper impact evaluation

Scholarly paper impact evaluation has always attracted scholars' attention (Moed and Hal- evi 2015; Jiang et al. 2016; Waltman 2016; Wang et al. 2024). Researchers continuously explore methods for evaluating paper impact from the perspectives of network structure, citation weights, linear relationships, and time awareness by utilizing AI technology (Wang et al. 2014; Ye et al. 2015; Zhu et al. 2015; Bornmann and Williams 2020). These studies are all efforts to more reasonably evaluate the impact of scholarly paper. The following section discusses network-based evaluation methods, focusing on the composition of academic networks, whether homogeneous or heterogeneous, weighted or unweighted, linear or non-linear, time-aware or not time-aware, as well as their advantages and disadvantages. These will be followed by knowledge-based evaluation methods focusing on knowledge flow and knowledge graphs (Tu et al. 2023). Finally, the deep learning-based evaluation methods are discussed.

2.1 Network-based evaluation

We compare the network-based methods for scholarly paper impact evaluation in Table 1. Through a comparative analysis that encompasses method and reference, scholarly networks, homogeneity or heterogeneity, weighting, linearity, and time-awareness aspects, we are able to trace the evolution of network-based evaluation methods. Network-based evaluation methods are based on the scholarly network structure. These methods have evolved from homogeneous networks to heterogeneous networks, from first-order to higher-order

Table 1 Comparing the representative network-based methods for scholarly paper impact evaluation

Method and Reference	Scholarly network	Homogeneous or heterogeneous	Weighted or unweighted	Linear or nonlinear	Time-ware or no time-ware	strengths or weaknesses
PageRank Liu et al. (2014)	Citation Paper-author Author-venue Paper-venue	Homogeneous Heterogeneous	Weighted	Linear	No time-ware	Simultaneously rank academic entities in diverse networks in heterogeneous networks
Nonlinear-Rank Yao et al. (2014)	Citation	Homogeneous	Unweighted	Nonlinear	No time-ware	Significantly enhances the ranking effectiveness
SPRank Zhou et al. (2016)	Citation	Homogeneous	Unweighted	Linear	No time-ware	Improve ranking effectiveness, against malicious manipulations
PageRank Liang and Jiang (2016)	Hypernetwork	Heterogeneous	Unweighted	Linear	No time-ware	Use heterogeneous academic hypernetwork
COIRank Bai et al. (2016)	Citation Paper-author	Homogeneous Heterogeneous	Weighted	Linear	Time-ware	Discover the anomalous citations
PANDORA Bai et al. (2017a)	Citation Paper-author Paper-journal	Homogeneous Heterogeneous	Weighted	Linear	Time-ware	Against malicious manipulations
Quantum PageRank Bai et al. (2018)	Citation	Homogeneous	Weighted	Linear	No time-ware	Consider higher-order relationship
W-Rank Zhang et al. (2019)	Citation Paper-author Paper-journal	Homogeneous Heterogeneous	Weighted	Linear	Time-ware	Could be biased in favour of the W-Rank
CPRank Wang et al. (2019)	Citation	Homogeneous	Unweighted	Linear	No time-ware	Is still biased
IPRank Bai et al. (2020b)	Citation Paper-institute	Homogeneous Heterogeneous	Unweighted	Linear	No time-ware	Rank institution and paper
BTPR Dayeh et al. (2022)	Citation	Homogeneous	Unweighted	Linear	Time-ware	Reduce the bias in favor of old publications
ASP Chen et al. (2023)	Citation	Homogeneous	Unweighted	Linear	No time-ware	Use multi-disciplined citation network

networks, from single-dimensional to multi-dimensional evaluation, from unweighted networks to weighted networks, and from time-unaware evaluation to time-aware evaluation. Next, we will introduce the evolution, advantages, and limitations of different methods in detail.

In previous network-based evaluation methods, malicious citations are not taken into account such as FutureRank (Sayyadi and Getoor 2009), CiteRank (Jomsri et al. 2011), and P-Rank (Zhao et al. 2009). To address this, Bai et al. (2016) propose the COIRank algorithm, in which weakens the impact of conflicts of interest. Their experimental results show that COIRank significantly outperforms the previous method (Wang et al. 2013b) and COIRank_Notime. On this basis, Bai et al. (2017a) present a PANDORA algorithm, which determines citation weights based on COI relationships in citation networks. The experimental results demonstrate PANDORA outperforms the previous method (Wang et al. 2013b) and FutureRank (Sayyadi and Getoor 2009). However, PANDORA algorithm has limitations, such as identifying COI relationships through co-authors and the same institution, while ignoring other COI relationships. Similarly, Zhang et al. (2019) propose the W-Rank method, which considers citation relevance and author contribution to allocate weights. Their results show that W-Rank outperforms PageRank (Brin and Page 1998), HITS (Kleinberg 1999) and FutureRank (Sayyadi and Getoor 2009) in accuracy. limitation of W-Rank is its dependency on parameters.

Zhang et al. (2019) propose an intelligent time-aware mutual reinforcement ranking (TAMRR) model to evaluate scholarly paper impact, in which temporal factors and the distribution of co-authors' contributions are considered. Moreover, the enhanced mechanism is mainly implemented by the PageRank and HITS algorithms. A bias-free time-aware PageRank algorithm called BTPR is developed to evaluate scholarly paper impact, which considers the citation average rate of change to address bias issues such as assuming that recent scholarly papers are more valuable than older ones or that a reputable author can publish valuable papers. BTPR is a linear evaluation approach and it outperforms PageRank in ranking scholarly papers.

Although Liu et al. (2014) propose a Tri-Rank ranking framework based on the heterogeneous academic networks to rank scholarly publications, scholars, and venues, they do not consider time-aware weights of edges. Their experimental results show that the proposed Tri-Rank is more effective than PageRank (Brin and Page 1998), HITS (Kleinberg 1999), and Co-Rank (Meng and Kennedy 2013). Zhou et al. (2016) use the similarity-preferential mechanism to improve the ranking effectiveness of the PageRank algorithm. However, the proposed SPRank algorithm is not very satisfactory for small nodes or zero-degree nodes. Wang et al. (2019) propose the CPRank algorithm, in which the aging characteristics are introduced to the PageRank algorithm by considering the first 10-year citations. Their experimental results show that the CPRank algorithm outperforms PageRank (Brin and Page 1998) and CiteRank (Jomsri et al. 2011). The difference between CPRank and PageRank lies in the adjacency matrix. In the PageRank algorithm, as long as there are citations between papers, the citation weight is set to 1, otherwise, it is 0. However, in the CPRank algorithm, the citation weight is set to 1 only if the citation between the citing paper and the cited paper does not exceed 10 years, otherwise, it is 0. Bai et al. (2020b) propose the IPRank model. They first construct a heterogeneous institution-citation network and then introduce PageRank to the heterogeneous academic network. Their experimental results demonstrate that IPRank outperforms IRank (Massucci and Docampo 2019) in retrieving

the Nobel Prize papers among top N papers. Zhang and Wu (2021) propose a scholarly paper impact ranking approach, called WCCMR, in which the citation weight between the citing paper and the cited paper is calculated by relying on its strength and topical similarity. Their experimental results show that WCCMR algorithm outperforms MutualRank (Jiang et al. 2016), Tri-rank (Liu et al. 2014), and GraLap (Chakraborty and Narayanam 2016) in terms of NDCG.

It is well established that the evaluation results of scholarly papers' impact based on PageRank depend on the parameter α . The evaluation outcomes are influenced by the choice of α , and varying this value will lead to different results. The quantum PageRank algorithm is independent of the parameter α . This algorithm effectively illuminates the structure of secondary hubs within scale-free networks, discerning the intrinsic hierarchical arrangement. To address the shortcomings of the PageRank algorithm, Bai et al. (2018) explored higher-order dependencies in citation networks and developed a higher-order weighted quantum PageRank algorithm to rank scholarly papers' impact. Their results show that this approach better distinguishes scholarly papers' impact compared with the weighted quantum PageRank and PageRank. The advantage of utilising higher-order dependencies is that it reduces the effect of maliciously manipulated citation activities. Following progress with the linear PageRank algorithm, scholars have turned their attention to its nonlinearity. Yao et al. (2014) confirm that nonlinearity can improve PageRank's effectiveness.

PageRank on simple scholarly networks, whether homogeneous or heterogeneous, may limit their ranking performance. Researchers have explored evaluating scholarly paper impact from the perspective of hypernetwork. A representative ranking algorithm based on heterogeneous academic hypernetworks demonstrates that PageRank on author citation hypernetworks is more effective than on simple author citation network (Liang and Jiang 2016).

2.2 Knowledge-based evaluation

Wang et al. (2019a) evaluate the impact of scholarly papers based on the knowledge flow. Firstly, they extract three knowledge flow patterns, including knowledge flow intensity, knowledge diffusion capacity, and knowledge transfer capacity. On this basis, they quantify the three knowledge flow patterns. Then, they introduce the entropy weight method to these three knowledge flow patterns. There are two limitations to knowledge-flow-based evaluation. First, the evaluation method cannot be directly leveraged to quantify the scholarly paper impact in different disciplines. Second, the citations of scholarly papers cannot be distinguished by this method. Chatzopoulos et al. (2021) predict the impact of the most recent publications based on the Artsim+ approach. They compute the paper similarity by using scholarly knowledge graphs. It can be seen that compared to network-based evaluation methods, knowledge-based evaluation methods can better improve the accuracy of evaluation. Sun et al. (2022) propose an innovative method for the automated evaluation of scholarly paper impact by constructing a knowledge graph. The explicit features, implicit features, and higher-order relationships among scholarly entities are used in this method. In addition, a knowledge graph embedding method is used to learn the representations of scholarly entities and generate evaluation scores for scholarly papers by utilizing these learned vectors. Zaeem et al. (2023) introduce an innovative framework for the formation of an intelligent article knowledge graph, which employs the BM25 probabilistic retrieval

model to construct a knowledge graph that encapsulates the conditional dependency relationships among academic articles. Stoian et al. (2024) develop an advanced intelligent recognition model for identifying high-quality scholarly papers, relying on the knowledge-based metasegmentic networks. Based on the networks, the structural features of knowledge elements are mined. Their experimental outcomes demonstrate the efficacy of the proposed method.

In summary, knowledge-based evaluation methods for scholarly paper impact have both advantages and disadvantages. They offer detailed and potentially more accurate assessments, leveraging innovative techniques like knowledge graphs. However, they face limitations in cross-disciplinary application.

2.3 Deep learning-based evaluation

Kanellos et al. (2021) introduce a novel method based on the estimated short-term impact of scholarly papers for evaluating scholarly paper impact, called AttRank. The process of random reading of literature by researchers is modeled, and then, an attention mechanism is introduced to identify popular papers. Finally, they use a time-based mechanism to enhance recent publications. Their experimental results show that the AttRank outperforms CiteRank (Jomsri et al. 2011), FutureRank (Sayyadi and Getoor 2009), Retained Adjacency Matrix (Ghosh et al. 2011), Effective Contagion Matrix (Ghosh et al. 2011), and WSDM cup's 2016 winner (Feng et al. 2016) in terms of NDCG. It can be seen that deep learning-based evaluation methods are superior to some classic network-based evaluation methods such as FutureRank (Sayyadi and Getoor 2009) and Retained Adjacency Matrix (Ghosh et al. 2011). The two methods evaluate scholarly paper impact by considering citation networks.

Huang et al. (2021) evaluate scholarly paper impact by combining citation polarity and purpose. They first identify citation polarity and purpose by relying on Word2Vec (Church 2017) and convolutional neural network(CNN) (Gu et al. 2018). Then they explore how the type of paper affects the citation purpose and polarity. Finally, they explain the actual impact of scholarly paper by mining knowledge cited in a scholarly paper. Although citations are used as one of the most important metrics to measure the impact of scholarly papers, citations may be gamed in different ways. Therefore, Liu et al. (2022) propose a deep graph learning model, in which they incorporate text semantic mining to network representation learning to identify anomalies in citation networks. Their experimental results demonstrate the effectiveness of this model for identifying anomalous citations.

Deep learning-based evaluation methods for scholarly paper impact offer sophisticated analysis, improved performance over traditional methods, and the ability to detect anomalies like citation manipulation. However, they are complex, prone to overfitting, dependent on data quality.

3 Paper impact prediction

3.1 Datasets

Widely used datasets are listed in Table 2 to predict the impact of scholarly papers. These datasets contain rich information such as papers, authors, journals, and institutions. In addition to the datasets listed in Table 2, Web of Knowledge (Klimek et al. 2016), ACL Anthology website (Huang et al. 2022; Xiao et al. 2022), PubMed Central (Huang et al. 2022a; Li et al. 2022), MAG (Ji et al. 2024) and Semantic Scholar's Open Research Corpus (Kleśniński et al. 2021) are used for predicting scholarly paper impact.

3.2 Feature-based predictive models

Feature extraction plays an important role in predicting scholarly paper impact. Various features and AI techniques have been used in this research field, analyzing the relationship between citation structure and scholarly paper impact, and constructing predicting models. In this section, we first discuss the features used in scholarly paper impact prediction, which are categorized into four groups: attribute-based features, content-based features, network-based features, and temporal features (see Table 3). Through a comparative analysis that includes features, type of features, reference aspects, we are able to discern the various features utilized by different literatures for predicting the impact of scholarly papers. Then we compare the feature-based predictive models. We perform statistical analysis on the main features of the literature involved in this paper to predict the scholarly paper impact, as shown in Fig. 2. The larger the font size is, the higher the popularity of the feature is. In terms of attribute-based features, citation count is the most frequently used feature, followed by h-index and the number of papers (see Fig. 2a). These attribute-based features can be divided into two categories: one is explicit features such as citation count, title length,

Table 2 Representative datasets in paper impact prediction literature

Dataset	Reference
Scopus (https://www.scopus.com)	Thelwall and Nevill (2018) Abramo et al. (2019) Shahmandi et al. (2020) Fronzetti Colladon and 'Angelo CA, Gloor PA, (2020) Ma et al. (2021) Fronzetti Colladon and 'Angelo CA, Gloor PA, (2020) Croft and Sack (2022)
Web of Science (http://www.webofscience.com)	Revesz (2014) Stegehuis et al. (2015) Robson and Mousquès (2016) Abrishami and Aliakbary (2019) Abramo et al. (2019) El Mohadab et al. (2019) Abramo et al. (2019) Wang and Zhang (2020) Wang et al. (2021) Hu et al. (202) Chen et al. (2023)
American Physical Society (https://www.aps.org)	Shen et al. (2014) Bai et al. (2019) Jiang et al. (2021) Zhao and Feng (2022)
AMiner (https://www.aminer.cn)	Li et al. (2019) Jiang et al. (2021) Xiao et al. (2022)
DBLP (https://dblp.org)	Zhang and Wu (2024)
arXiv dataset(http://www.cs.cornell.edu/projects/kddcup/datasets.html)	Zhang and Wu (2024)
Semantic Scholar's Open Research Corpus (https://www.semanticscholar.org)	Kleśniński et al. (2021)

Table 3 An example of different features for scholarly paper impact prediction

Features	Type of features	Reference
Centrality	Network-based features	Sarigöl et al. (2014)
Author productivity h-index citations	Attribute-based features	Chakraborty et al. (2014)
Topic diversity keyword diversity	Content-based features	Chakraborty et al. (2014)
Number of authors citations journal impact factor	Attribute-based features	Yu et al. (2014)
Degree closeness betweenness	Network-based features	Fronzetti Colladon and 'Angelo CA, Gloor PA, (2020)
Complexity diversity	Content-based features	Fronzetti Colladon and 'Angelo CA, Gloor PA, (2020)
Citations number of authors abstract length	Attribute-based features	Fronzetti Colladon and 'Angelo CA, Gloor PA, (2020)
Sentiment	Content-based features	Fronzetti Colladon and 'Angelo CA, Gloor PA, (2020)
Temporal cues	Temporal features	Jiang et al. (2021)
Citations number of authors number of institutions	Attribute-based features	Thelwall et al. (2023)
Title abstract	Attribute-based features	Thafar et al. (2024)
Degree PageRank Betweenness Centrality	Network-based features	Thafar et al. (2024)

paper length, etc.; the other is implicit features such as h-index, g-index, e-index, and so on. A small amount of content-based features are used to predict the impact of scholarly paper such as diversity, topic rank and sentiment (see Fig. 2b). For a detailed introduction, refer to Section 3.2. The frequently used network-based features are shown in Fig. 2c. Network-based features fall into two distinct categories: those that capture the global network structure, including metrics like centrality, sociality and authority, and those that focus on the local network structure, encompassing measures such as closeness and betweenness. Section 3.2.3 provides a detailed introduction to the specific details.

3.2.1 Attribute-based features

Attribute-based features have been among the most important features in scholarly paper impact prediction (Akella et al. 2021). Attribute features related to papers include citations (Stegehuis et al. 2015), the reciprocal of the first-cited age of the paper, the total citations in the first two years after publication, and the number of countries citing this paper in the first two years after publication (Yu et al. 2014), the year of publication, abstract length, title length, figure count, table count, differential equations, DOI date (Robson and Mousquès 2016), firstPA (first order of preferential attachment), secPA (second order of preferential attachment), firstRef (first order of reference), secRef (second order of reference) (Xiao

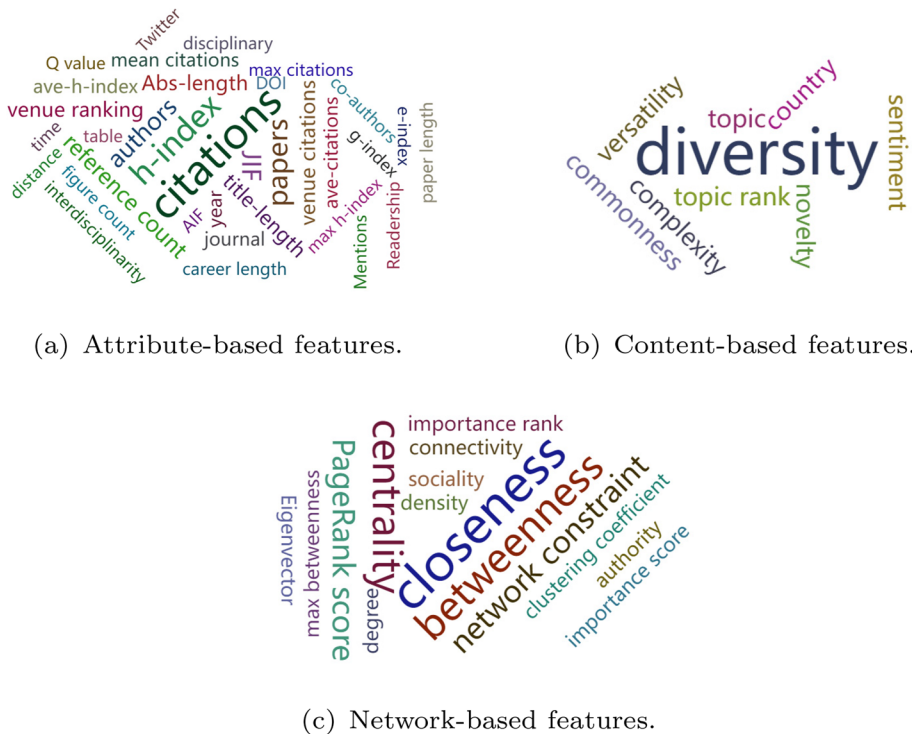


Fig. 2 Popular features in scholarly paper impact prediction models **a** Attribute-based features **b** Content-based features **c** Network-based features

et al. 2016), author_papers (number of papers published), author_papers_delta (number of the paper published in last two years) and so on (Weihs and Etzioni 2017).

Attribute features related to authors cover author count, author name (Robson and Mousquès 2016), h-index (h-index of author), authorrank (rank of author), noca (number of co-author), insitrank (rank of author's institute), producibility (number of publications by author), teamsize (number of authors of papers), insitnum (number of institute) (Xiao et al. 2016), author rank (citations of author), past influence of authors (maximum citations of an author or total citations) (Singh et al. 2017), sum(Q), max(Q), min(Q), avg(Q), med(Q), sum(H-index), max(H-index), min(H-index), avg(H-index), med(H-index), dev(H-index), sum(AIF), max(AIF), min(AIF), avg(AIF), med(AIF), dev(AIF) (Bai et al. 2019).

Attribute features related to venue consist of page count, venue prestige (Robson and Mousquès 2016), venurank (rank of venues), venueaut (number of authors in venues), venuepub (number of publications of venues) (Xiao et al. 2016), past influence of venues (max citations or total citations of a venue) (Singh et al. 2017). These research show that factors such as the length of researchers' careers and the age of papers play an important role in determining the accuracy of scholarly paper impact prediction.

In summary, we find that the existing attribute-based features used to predict the impact of scholarly papers include explicit and implicit features. Previous research suggests that there is still room for new features to improve the accuracy of prediction, which requires further exploration.

3.2.2 Content-based features

Content-based features are considered to be intrinsic drivers and one of the key features to improve prediction accuracy (Hassan et al. 2020). The content-based features cover novelty (measuring the similarity between papers), topic rank, diversity (measuring the breadth of a scholarly paper inferred from its topic distribution) (Singh et al. 2017), filedhot (topic hotness of the paper) (Xiao et al. 2016), sentiment (measure the positivity or negativity of the language used in the abstract of a scholarly paper), complexity (lexical complexity of an abstract is computed by the standard deviation of the frequency distribution of words), lexical diversity (the ration between different unique word with the total number of words), commonness (Fronzetti Colladon and 'Angelo CA, Gloor PA, 2020), paper metadata features (Ma et al. 2021).

To sum up, the content-based features have proven to be the qualified predictors. Many existing studies predict scholarly paper impact by analyzing the relationships between the content of the scholarly paper and their impact, including title, abstract, keyword, topic and so on (Singh et al. 2017; Xiao et al. 2016). The internal content-based features driving the scholarly paper impact are worth further exploration.

3.2.3 Network-based features

Network-based features have been studied extensively and can be categorized as author-wise, venue-wise, and paper-wise. The author-wise network features span centrality of authors (Xiao et al. 2016), sociality (compute based on the co-authorship network), authority (obtain based on citation network) (Singh et al. 2017), degree constraint, closeness, betweenness in author network (Fronzetti Colladon and 'Angelo CA, Gloor PA, 2020). The venue-wise network features include the centrality of venues (Xiao et al. 2016; Singh et al. 2017). The paper-wise network features cover the document centrality (Klimek et al. 2016), degree constraint, closeness, and betweenness in publication network (Fronzetti Colladon and 'Angelo CA, Gloor PA, 2020).

Despite the importance of network-based features on scholarly paper impact, the influence of these features remains largely unclear, with neither of the standard network-based feature selection methods for improving the accuracy of prediction performance.

3.2.4 Temporal features

Castillo et al. (2007) leverage author reputation and time to construct a model for predicting citations. This model integrates three types of features, such as a priori author-based features, priori link-based features, and a posteriori features, into a linear regression model. The empirical findings indicate that incorporating the author's reputation and temporal features enhances the predictive performance of the model. Davletov et al. (2014) introduce a new method for predicting the future impact of scholarly papers based on temporal and topological features from citation networks. By clustering the citation patterns over time and treating impact prediction as a regression problem within each cluster, the proposed technique outperforms existing approaches in predicting the number of citations a paper will receive. Recent studies suggest that temporal features in scholarly paper impact prediction are important (Singh et al. 2017; Jiang et al. 2021). Singh et al. (2017) use the recent feature

to describe the temporal proximity of a scholarly paper. Jiang et al. (2021) use temporal cues to predict scholarly paper impact based on the content of metadata and their temporal trends. Yan et al. (2024) present CoDy, a novel framework designed to model the temporal dynamics in the citation network. Experimental validations confirm CoDy's efficacy in predicting citation counts, with in-depth analyses supporting its reliability and robustness.

Temporal features in scholarly paper impact prediction offer the advantages of enhanced predictive performance. However, they also come with limitations such as reliance on data quality and the potential for proximity bias favoring recent publications.

3.2.5 Predictive models

The core challenge of feature-based scholarly paper impact prediction models is feature engineering. Table 4 presents the comparison of the feature-based models for scholarly paper impact prediction. By contrasting method and reference, features, datasets, comparing algorithms, evaluation metrics and performance, we can discern the commonalities and distinctions among various feature-based models for predicting the impact of scholarly paper. As time progresses, feature-based models have undergone continual enhancements. Next, we will introduce the evolution, advantages, and limitations of different models in details.

Sarigöl et al. (2014) compare the difference between random forest and Naive Bayes (Frank et al. 2000) in predicting scholarly paper impact. The results indicate that the random forest can obtain better performance than naive Bayes in terms of recall, precision, and *F*-score. In addition, they also expect their model to be used for predicting the success of actors based on the features of the actors position in the collaboration network. Robson and Mousquès (2016) compare the importance of different variables in the random forests. The experimental results indicate that venue, paper length, abstract length, topic, and number of authors are important for predicting scholarly paper impact. Chakraborty et al. (2014) propose a two-stage predictive model, in which the support vector machine is used to categorize the scholarly papers in the first stage. In the second stage, a support vector regression is leveraged to predict the future citations of scholarly papers. Experimental results show that classification in the training phase can improve prediction performance (about 50%) compared to well-known baseline systems. It is worth mentioning that this paper conducted an experimental analysis on the robustness of the model, indicating that further systematic analysis for category is needed. Based on this work, Singh et al. (2017) further explores the influence of early citations on the long-term impact of scholarly papers. In this model, attribute-based features, content-based features, and network-based features are incorporated in linear regression (LR) (Su et al. 2012), gaussian process regression (GPR) (Schulz et al. 2018), classification and regression trees (CART) (Loh 2011), and support vector regression (SVR). The experimental results show the proposed model outperforms all three baselines (Yan et al. 2012; Chakraborty et al. 2014; Singh et al. 2015), indicating that the early citations can be used as an important factor to predict the long-term impact of scholarly papers.

Compared with previous research, Yu et al. (2014) use the stepwise multiple regression analysis to analyze the relationships between various features and citation impact for selecting good features and construct a linear regression model to predict scholarly paper impact. Compared to the previous two papers (Sarigöl et al. 2014; Chakraborty et al. 2014), this paper does not conduct an analysis of generalization ability and robustness.

Table 4 Comparing feature-based models for scholarly paper impact prediction

Method and Reference	Features	Datasets	Comparing algorithms	Evaluation metrics	Performance
Random forest McNamara et al. (2013)	Attribute-based network-based	Scopus	Linear regression decision trees	RMSE	Random forest has the best prediction performance.
Random forest Sarigöl et al. (2014)	Attribute-based network-based	Microsoft Academic Search	Naive Bayes	Recall Precision F-score	The random forest performs better than a naive Bayes.
Two-stage prediction model Chakraborty et al. (2014)	Attribute-based network-based	Microsoft Academic Search	Baseline system	R^2	The proposed model can remarkably improve performance.
Regression model Yu et al. (2014)	Attribute-based	Thomson ISI	Citations	RMSE MAPE	Revealing interesting relationship between various features and citation impact.
Quantile Regression Steghuis et al. (2015)	Attribute-based	Web of Science	Only IF Only c_1	Empirical conditional distribution	Predict a probability distribution.
Random forest Robson and Mousquès (2016)	Attribute-based content-based	Web of Knowledge	Citations	MSE	All factors predict 16-29% of the variation in citations.
LR GPR CART SVR Singh et al. (2017)	Attribute-based content-based network-based	Microsoft Academic Search	References Yan et al. (2012) Singh et al. (2015)	R^2, ρ	The early citers play an important role.
WMR-Rank Zhang and Wu (2020)	Attribute-based	ACL	Mutual-Rank Jiang et al. (2016) MRCoRank Wang et al. (2016) PageRank Zhou et al. (2016) TAPRank Kong et al. (2015) Tri-Rank Liu et al. (2014)	ρ	WMR-Rank is more effective than comparison algorithms.
Instance-based learning model Zhang and Wu (2024)	Attribute-based	DBLP	CitingNum citedNum Sim	Accuracy F-Score	Using all three features indicates best performance.
FutureCite Thafar et al. (2024)	Attribute-based	Self-collected data	SVM Naive Bayes RF XGBoost	Precision Recall F1-Score	XGboost is more effective than other classifiers.

Different from previous studies, Steghuis et al. (2015) predict a probability distribution for citations of a scholarly paper. In their model, the journal impact factor and early citations are integrated into the quantile regression model, indicating the two factors are important in predicting the probability distribution.

Zhang and Wu (2020) leverage seven types of relationships including paper citation, author citation, co-authorship, venue citation, paper-author, paper-venue, and author-venue to predict the impact of scholarly papers, researchers, and venues based on a heterogeneous scholarly network. The proposed WMR-Rank outperforms state-of-the-art competitors including FutureRank (Sayyadi and Getoor 2009), MutualRank (Jiang et al. 2016), MRCo-Rank (Wang et al. 2016), PageRank (Zhou et al. 2016), TAPRank (Kong et al. 2015), and Tri-Rank (Liu et al. 2014). However, this paper does not distinguish citation weights. In addition to the models listed in Table 4, other models are used to predict scholarly paper impact such as MPNN model (Li et al. 2022) and NCFM model (Wang et al. 2021).

In conclusion, SVR and random forest are frequently used to predict the scholarly paper impact. Researchers have attempted to leverage the explicit features in their predictive models. To improve the capability of predictive models, further investigation into implicit features is necessary.

3.3 Generative predictive models

The citation activity of scholarly papers can be formulated as an event sequence in the continuous temporal domain. This citation activity can be characterized by probabilistic statistical generative approaches. Among these, the point process is frequently used in the models for predicting scholarly paper impact. Shen et al. (2014) leverage the Bayesian treatment by adding a conjugate prior to improve the predictive ability of this model (Wang et al. 2013). The experimental results show that the optimized model is more robust and accurate for predicting popularity dynamics. Wang et al. (2013) propose a generative probabilistic predictive model. This model uses the reinforced poisson process to predict the long-term impact of publications. The model includes the three key ingredients: (1) the relative fitness for capturing the importance of a scholarly paper relative to other papers; (2) the temporal decay function indicating the decay rate of its popularity, and (3) the reinforcement mechanism for capturing the fact that researchers tend to cite highly cited papers.

The two methods (Wang et al. 2013; Shen et al. 2014) mentioned above ignore the time-stamp and aggregate all past citations to model the citation intensity, which may lead to poor representations of citation dynamics. To complement the shortcoming, Xiao et al. (2016) use a self-triggering Hawkes process to favor the recent citations and apply a Laplace Kernel to control the effect time-decaying window. In their model, the citation intensity of paper p is defined as:

$$\lambda_p(t) = \beta s_p(t) e^{-\omega_1 p^t} + \alpha_p \sum_{i, t_i < t} e^{-\omega_2 p^{(t-t_i)}} \quad (1)$$

where $s_p(t)$ is a row vector, indicating the quality of scholarly paper, and β is a column vector for the coefficients. $e^{-\omega_1 p^t}$ represents the time-varying decay function. $\alpha_p \sum_{i, t_i < t} e^{-\omega_2 p^{(t-t_i)}}$ is the triggering strength of each citation. Based on previous research, Bai et al. (2019) model the Paper Potential Index (PPI) by considering early citations and early citers' impact and the experimental results show that the PPI model leads to better performance in terms of RMSE.

Generative models for scholarly paper impact prediction do not require a large number of features and mainly rely on time-series data. However, these models rely heavily on the assumptions of fixed parameters, limiting the model's expressiveness. In addition, one drawback of generative predictive models is that they fail to model the structural information of the scholarly networks, which would be helpful in understanding the citation dynamics. Therefore, generative models for predicting scholarly paper impact have significant room for improvement in making precise predictions.

3.4 Deep learning predictive models

Existing deep learning predictive models for scholarly paper impact can be categorized as (1) models based on temporal sequences, e.g., the citation behavior in scholarly networks, and these models depend on Recurrent Neural Networks (RNN) (Rumelhart et al. 1986; Zhou et al. 2022) or sequence-to-sequence model (Sutskever et al. 2014; (2) models based on the content of scholarly papers, e.g., paper metadata text, and these models utilize natural language processing technology to predict the future citations; (3) models based on graphs, e.g., global graph or local graph, and these models mainly deal with graph representation learning, which purpose is to learn the effective embedding representations (Jiang et al. 2021; Xiao et al. 2022).

Abrishami and Aliakbary (2019) leverage the sequence pattern of the early citations and deep learning techniques such as RNN and sequence-to-sequence model to learn scholarly paper impact prediction model, called NNCP. The NNCP model outperforms the two comparative models AVR and GMM (Cao et al. 2016) concerning both efficiency and accuracy.

In comparison with the model proposed by Abrishami and Aliakbary (Abrishami and Aliakbary 2019), the model proposed by Ma et al. (2021) captures more accurate prediction results. Ma et al. (2021) design a two-stage model to predict the future citations of scholarly papers. They first exploit Doc2Vec to extract the semantic features from the paper metadata text, and then utilize Bi-LSTM with an attention mechanism to further refine more meaningful features based on the previous step. Finally, they apply two fully-collected layers to improve the learning ability of the predictive model. The experimental results indicate that the proposed model outperforms other comparative models such as GBRT (Friedman 2001), XGBoost (Chen and Guestrin 2016), Bi-LSTM (Graves 2012), NNCP (Abrishami and Aliakbary 2019) and BP (Ruan et al. 2020).

Compared with traditional citation prediction, the fine-grained citation count prediction task (FGCCP) predicts not only citations but also citation location (Huang et al. 2022a). To complete FGCCP, Huang et al. (2022a) propose the MTAT model, which uses the attention mechanism. Their experimental results show that MTAT surpasses several comparative models such as XGBoost (Chen and Guestrin 2016), NNCP (Abrishami and Aliakbary 2019), and RNN (Rumelhart et al. 1986).

Applying representation learning to impact prediction, Xiao et al. (2022) propose ESMR. A heterogeneous network embedding model is designed for scholarly papers to learn the semantic information and local structure simultaneously and then use multivariate random walk algorithm to predict scholarly impact. Experiments demonstrate that the ESMR model outperforms the comparative models such as MRcoRank (Wang et al. 2016), Future-Rank (Sayyadi and Getoor 2009), PageRank (Zhou et al. 2016), LINE+CoRank (Tang et al. 2015), and EOE+CoRank (Xu et al. 2017).

Deep learning-based predictive methods for scholarly paper impact offer advanced analysis capabilities, improved performance, and the ability to make fine-grained predictions. However, these predictive methods have their own limitations, such as reliance on data quality, relatively complex models, and a tendency to overfit. Despite outperforming some traditional methods, there is ongoing room for enhancement.

3.5 LLMs for predicting paper impact

In recent years, Large Language Models (LLMs) have showcased remarkable proficiency in long-form text modelling, finding extensive application across a diverse range of NLP tasks. de Winter (2024) uses ChatGPT-4 to analyze the abstracts of scholarly papers, revealing that ChatGPT-4's scores exhibited a higher degree of correlation with the number of citations than traditional readability scores. This indicates that LLMs are very promising within the field of scientometrics. Zhao et al. (2024) leverage LLMs to predict the impact of scholarly papers by considering titles and abstracts. Restricting the analysis to internal information alone, their experimental findings reveal that the LLaMA-3 model outperforms earlier models with respect to both Mean Absolute Error (MAE) and NDCG@20 metrics. Hao et al. (2024) propose the HLM-Cite model to retrieve the core citations and leverage LLMs to rank scholarly papers, revealing the underlying relationships among scholarly papers. Their model has the following advantages: (1) It excels in extensive candidate sets while maintaining a high level of precision; (2) It exhibits good scalability. A major limitation of this model is LLMs' illusion problem. Under certain circumstances, LLMs may produce unreliable analyses and affect contaminate specific samples. Therefore, the reliability and scalability of large language models in predicting the impact of research papers warrant further investigation by more researchers.

Large Language Models (LLMs) offer advanced text analysis capabilities that correlate with citation metrics, outperforming traditional methods in predicting the impact of scholarly papers with precision and scalability. However, they have some limitations, as they can suffer from the illusion problem, producing unreliable analyses that may contaminate data, necessitating further research to ensure their reliability.

3.6 Discussion on predictive models

In scholarly papers on impact prediction models, more researchers tend to focus on the accuracy of predictions in their papers (Sarigöl et al. 2014; Bai et al. 2019). It is worth mentioning that the Reinforced Poisson Process performs well for predicting the long-term impact of papers, while for predicting the short-term impact, the self-triggering Hawkes process and deep learning models are the top performers. Typically, the evaluation of a paper's impact encompasses long-term forecasts ranging from 1 to 10 years, whereas short-term spans are generally confined to a period of 0.5 to 3 years (Jiang et al. 2023; Abramo et al. 2019; Akella et al. 2021). However, as is widely acknowledged, an effective model for predicting the impact of scholarly papers should not only consider the accuracy of predictions but also take into account factors such as generalization ability (Sarigöl et al. 2014; Hao et al. 2024), robustness (Chakraborty et al. 2014), and interpretability (Wang et al. 2013; Xiao et al. 2016; Bai et al. 2019). These previous studies will have a positive impact on the research field of scholarly paper impact, expanding the intellectual scope for researchers.

There is an anticipation that researchers will delve into more in-depth conversations about the generalization capabilities, robustness, interpretability, and the complexity of their models. It is important to highlight that the development of scholarly impact prediction models necessitates meticulous design and rigorous monitoring. Without such precautions, they may indeed reinforce biases and disparities in the scientific field. When developing predictive models, researchers ought to consider factors such as data diversity, fairness assessment, interpretability, multi-dimensionality, and supervision, all in an effort to diminish and alleviate potential biases.

4 Emerging directions in scholarly impact research

4.1 Network-based evaluation integrating attention mechanisms

The citation weights between scholarly papers are not equally important. Distinguishing citation weights has always been an urgent problem that researchers want to solve. Although researchers have made some progress in this field (Zhu et al. 2015; Bai et al. 2016, 2018), there is still a lot of room for exploration in network-based evaluation methods. Attention mechanisms can be used to distinguish these weights (Veličković et al. 2018). Therefore, the network-based scholarly paper impact evaluation integrated with attention mechanisms might solve this issue.

4.2 Multimodal evaluation based on large models

Although researchers have attempted network-based evaluation, knowledge-based evaluation, and deep learning-based evaluation, there have been few studies on scholarly paper impact evaluation methods based on the multimodal large models. However, with the development of large model technology and scholarly big data, based on different types of scholarly big data, such as text, images, videos, etc., a multimodal evaluation model can be constructed to achieve correlation analysis and fusion understanding of different types of data, providing a new evaluation approach for the impact of scholarly paper.

4.3 Prediction with content-rich graph neural networks

The previous models that used deep learning to predict the impact of academic papers mainly covered time series-based models, content-based models, and graph-based models. Among existing models for predicting scholarly paper impact, the content-based predictive models lack consideration for graph structure, and the graph-based models lack consideration for attention mechanisms. Therefore, important information can easily be lost in graph-based predictive models. To address this, it is worth considering the content-rich graph neural networks with attention. In other words, the nodes in this model can fully consider the context of their neighbors with an attention mechanism.

4.4 Prediction based on self-attentive temporal point process

Previous generative models for characterizing citation dynamics mainly leverage stochastic point process models such as Poisson processes and self-exciting Hawkes point process. One drawback of these models is that they cannot capture the long-term dependencies between citations well. To address this issue, attention mechanisms can be introduced into the paper impact prediction task. In addition, the neural marked temporal point process model has advantages for the continuous-time distribution of citation (Gupta and Bedathur 2022).

4.5 Prediction with reducing bias or debiasing large language models

The application of large language models in predicting the impact of research papers offers certain benefits, including enhanced precision in predictions, alongside drawbacks such as the potential for bias (de Winter 2024; Hao et al. 2024; Algaba et al. 2024). Identifying and mitigating bias in large language models when predicting impact of scholarly paper is an extremely challenging issue. To address this challenge, researches can leverage algorithms to identify biases within dataset, such as analyzing community dynamics within citation networks or investigating author collaboration relationships. Based on this, researchers can construct a model for predicting the impact of academic papers that retains the advantages of large language models while also addressing the shortcomings inherent in these models.

5 Conclusion

In this paper, we present a comprehensive survey on the evaluation and prediction of scholarly paper impact under the AI4Science background, aiming to provide researchers with an advanced understanding of this field. We categorize and analyze the relevant literature related to deep learning techniques, delving into the nuanced examination of the evolution, advantages, and limitations of evaluation models and predictive models in scholarly impact evaluation and prediction. Furthermore, we address the open research issues and challenges that researchers face in this domain. Given the vast amount of existing literature and the rapid growth of research in this area, our paper strives to cover recent and representative works comprehensively, especially those related to AI4Science. By doing so, we aim to offer readers an overview of the scholarly paper impact evaluation and prediction landscape in scholarly big data and AI4Science, equipping them with valuable insights into the current state of research in this field.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. Open Access funding enabled and organized by CAUL and its Member Institutions. This work was partially supported by National Natural Science Foundation of China under Grant No. 72204037, Humanities and Social Science Fund of Ministry of Education (Project No. 22YJC870009), and Natural Science Foundation of Liaoning Province under Grant No. 2023-BSBA-025.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abramo G, D'Angelo CA, Felici G (2019) Predicting publication long-term impact through a combination of early citations and journal impact factor. *J Informet* 13(1):32–49
- Abramo G, D'Angelo CA, Reale E (2019) Peer review versus bibliometrics: which method better predicts the scholarly impact of publications? *Scientometrics* 121:537–554
- Abrishami A, Aliakbary S (2019) Predicting citation counts based on deep neural network learning techniques. *J Informet* 13(2):485–499
- Akella AP, Alhoori H, Kondamudi PR et al (2021) Early indicators of scientific impact: predicting citations with altmetrics. *J Informet* 15(2):101128
- Algaba A, Mazijn C, Holst V, et al (2024) Large language models reflect human citation patterns with a heightened citation bias. *arXiv preprint arXiv:2405.15739*
- Bai X, Xia F, Lee I et al (2016) Identifying anomalous citations for objective evaluation of scholarly article impact. *PLoS One* 11(9):e0162364
- Bai X, Lee I, Ning Z et al (2017) The role of positive and negative citations in scientific evaluation. *IEEE Access* 5:17607–17617
- Bai X, Liu H, Zhang F et al (2017) An overview on evaluating and predicting scholarly article impact. *Information* 8(3):73
- Bai X, Zhang F, Hou J et al (2018) Quantifying the impact of scholarly papers based on higher-order weighted citations. *PLoS One* 13(3):e0193192
- Bai X, Zhang F, Lee I (2019) Predicting the citations of scholarly paper. *J Informet* 13(1):407–418
- Bai X, Pan H, Hou J et al (2020) Quantifying success in science: an overview. *IEEE Access* 8:123200–123214
- Bai X, Zhang F, Ni J et al (2020) Measure the impact of institution and paper via institution-citation network. *IEEE Access* 8:17548–17555
- Bhat HS, Huang LH, Rodriguez S, et al (2015) Citation prediction using diverse features. In: 2015 IEEE International Conference on Data Mining Workshop (ICDMW), IEEE, pp 589–596
- Bornmann L, Williams R (2020) An evaluation of percentile measures of citation impact, and a proposal for making them better. *Scientometrics* 124(2):1457–1478
- Brin S, Page L (1998) The anatomy of a large-scale hypertextual web search engine. *Comput Netw ISDN Syst* 30(1–7):107–117
- Brito R, Rodriguez-Navarro A (2019) Evaluating research and researchers by the journal impact factor: is it better than coin flipping? *J Informet* 13(1):314–324
- Bu Y, Lu W, Wu Y et al (2021) How wide is the citation impact of scientific publications? a cross-discipline and large-scale analysis. *Inform Proc Manage* 58(1):102429
- Bu Y, Waltman L, Huang Y (2021) A multidimensional framework for characterizing the citation impact of scientific publications. *Quant Sci Stud* 2(1):155–183
- Cai L, Tian J, Liu J et al (2019) Scholarly impact assessment: a survey of citation weighting solutions. *Scientometrics* 118:453–478
- Cao X, Chen Y, Liu KR (2016) A data analytic approach to quantifying scientific impact. *J Informet* 10(2):471–484
- Castillo C, Donato D, Gionis A (2007) Estimating number of citations using author reputation. In: International Symposium on String Processing and Information Retrieval, Springer, pp 107–117
- Chakraborty T, Narayanam R (2016) All fingers are not equal: Intensity of references in scientific articles. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp 1348–1358
- Chakraborty T, Kumar S, Goyal P, et al (2014) Towards a stratified learning approach to predict future citation counts. In: IEEE/ACM Joint Conference on Digital Libraries, IEEE, pp 351–360
- Chatzopoulos S, Vergoulis T, Kanellos I et al (2021) Further improvements on estimating the popularity of recently published papers. *Quant Sci Stud* 2(4):1529–1550

- Chen P, Xie H, Maslov S et al (2007) Finding scientific gems with google's pagerank algorithm. *J Informet* 1(1):8–15
- Chen T, Guestrin C (2016) Xgboost: a scalable tree boosting system. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp 785–794
- Chen Y, Koch T, Zakiyeva N et al (2023) Article's scientific prestige: Measuring the impact of individual articles in the web of science. *J Informet* 17(1):101379
- Church KW (2017) Word2vec. *Nat Lang Eng* 23(1):155–162
- Croft WL, Sack JR (2022) Predicting the citation count and citescore of journals one year in advance. *J Informet* 16(4):101349
- Davletov F, Aydin AS, Cakmak A (2014) High impact academic paper prediction using temporal and topological features. In: *Proceedings of the 23rd ACM international conference on conference on information and knowledge management*, pp 491–498
- Dayeh MA, Sartawi B, Salah S (2022) A bias-free time-aware pagerank algorithm for paper ranking in dynamic citation networks. *Intell Inf Manag* 14(2):53–70
- Dong Y, Ma H, Shen Z, et al (2017) A century of science: globalization of scientific collaborations, citations, and innovations. In: *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pp 1437–1446
- Dunaiski M, Visser W, Geldenhuys J (2016) Evaluating paper and author ranking algorithms using impact and contribution awards. *J Informet* 10(2):392–407
- Dunaiski M, Geldenhuys J, Visser W (2019) On the interplay between normalisation, bias, and performance of paper impact metrics. *J Informet* 13(1):270–290
- El Mohadab M, Bouikhalene B, Safi S (2019) Predicting rank for scientific research papers using supervised learning. *Appl Comput Inform* 15(2):182–190
- Feng MH, Chan K, Chen HY, et al (2016) An efficient solution to reinforce paper ranking using author/venue/citation information-the winner's solution for wsdm cup 2016. *WSDM Cup*
- Feng X, Zhao Q, Zhu R (2023) Towards popularity prediction of information cascades via degree distribution and deep neural networks. *J Informet* 17(3):101413
- Frank E, Trigg L, Holmes G et al (2000) Naive bayes for regression. *Mach Learn* 41:5–25
- Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *Ann Stat* 29(5):1189–1232
- Fronzetti Colladon A, D'Angelo CA, Gloor PA (2020) Predicting the future success of scientific publications through social network and semantic analysis. *Scientometrics* 124:357–377
- Ghosh R, Kuo TT, Hsu CN, et al (2011) Time-aware ranking in dynamic citation networks. In: *2011 IEEE 11th international conference on data mining workshops, IEEE*, pp 373–380
- Graves A (2012) Long short-term memory. *Studies in Computational Intelligence*, pp 37–45
- Gu J, Wang Z, Kuen J et al (2018) Recent advances in convolutional neural networks. *Pattern Recogn* 77:354–377
- Gupta V, Bedathur S (2022) Proactive: Self-attentive temporal point process flows for activity sequences. In: *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp 496–504
- Hao Q, Fan J, Xu F, et al (2024) Hlm-cite: Hybrid language model workflow for text-based scientific citation prediction. *arXiv preprint arXiv:2410.09112*
- Hassan SU, Aljohani NR, Idrees N et al (2020) Predicting literature's early impact with sentiment analysis in twitter. *Knowl-Based Syst* 192(105):383
- He G, Xue Z, Jiang Z et al (2023) H2cgl: Modeling dynamics of citation network for impact prediction. *Inform Proc Manage* 60(6):103512
- Hou J, Pan H, Guo T et al (2019) Prediction methods and applications in the science of science: a survey. *Comput Sci Rev* 34(100):197
- Hu Z, Cui J, Lin A (2023) Identifying potentially excellent publications using a citation-based machine learning approach. *Inform Proc Manage* 60(3):103323
- Huang H, Zhu D, Wang X (2021) Evaluating scientific impact of publications: combining citation polarity and purpose. *Scientometrics*, pp 1–25
- Huang S, Huang Y, Bu Y et al (2022) Fine-grained citation count prediction via a transformer-based model with among-attention mechanism. *Inform Proc Manage* 59(2):102799
- Huang S, Qian J, Huang Y et al (2022) Disclosing the relationship between citation structure and future impact of a publication. *J Am Soc Inf Sci* 73(7):1025–1042
- Ji T, Self N, Fu K et al (2024) Citation forecasting with multi-context attention-aided dependency modeling. *ACM Trans Knowl Discov Data* 18(6):1–23
- Jiang S, Koch B, Sun Y (2021) Hints: Citation time series prediction for new publications via dynamic heterogeneous information network embedding. *Proc Web Conf* 2021:3158–3167
- Jiang X, Gao C, Liang R (2016) Ranking scientific articles in a dynamically evolving citation network. 2016 12th International Conference on Semantics. *IEEE, Knowledge and Grids (SKG)*, pp 154–157

- Jiang X, Sun X, Yang Z et al (2016) Exploiting heterogeneous scientific literature networks to combat ranking bias: evidence from the computational linguistics area. *J Am Soc Inf Sci* 67(7):1679–1702
- Jiang Z, Lin T, Huang C (2023) Deep representation learning of scientific paper reveals its potential scholarly impact. *J Informet* 17(1):101376
- Jomsri P, Sangsantikul S, Choochaiwattana W (2011) Citerank: Combination similarity and static ranking with research paper searching. *Int J Internet Technol Secur Trans* 3(2):161–177
- Kanellos I, Vergoulis T, Sacharidis D et al (2019) Impact-based ranking of scientific publications: a survey and experimental evaluation. *IEEE Trans Knowl Data Eng* 33(4):1567–1584
- Kanellos I, Vergoulis T, Sacharidis D, et al (2021) Ranking papers by their short-term scientific impact. In: 2021 IEEE 37th International Conference on Data Engineering (ICDE), IEEE Computer Society, pp 1997–2002
- Kenton JDMWC, Toutanova LK (2019) Bert: Pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of naacL-HLT*, Minneapolis, Minnesota, p 2
- Khatoun A, Daud A, Amjad T (2024) Categorization and correlational analysis of quality factors influencing citation. *Artif Intell Rev* 57(3):70
- Kleinberg JM (1999) Authoritative sources in a hyperlinked environment. *J ACM (JACM)* 46(5):604–632
- Kleśniński R, Kazienko P, Kajdanowicz T (2021) Where should I publish? heterogeneous, networks-based prediction of paper's citation success. *J Informet* 15(3):101200
- Klimek P, Jovanovic AS, Egloff R et al (2016) Successful fish go with the flow: citation impact prediction based on centrality measures for term-document networks. *Scientometrics* 107:1265–1282
- Kong X, Zhou J, Zhang J, et al (2015) Taprank: A time-aware author ranking method in heterogeneous networks. In: 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity), IEEE, pp 242–246
- Li M, Xu J, Ge B et al (2019) A deep learning methodology for citation count prediction with large-scale biblio-features. 2019 IEEE International Conference on Systems. IEEE, Man and Cybernetics (SMC), pp 1172–1176
- Li X, Tang X, Cheng Q (2022) Predicting the clinical citation count of biomedical papers using multilayer perceptron neural network. *J Informet* 16(4):101333
- Liang R, Jiang X (2016) Scientific ranking over heterogeneous academic hypernetwork. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, pp 20–26
- Liu J, Kong X, Xia F et al (2018) Artificial intelligence in the 21st century. *Ieee Access* 6:34403–34421
- Liu J, Xia F, Wang L et al (2019) Shifu2: a network representation learning based model for advisor-advisee relationship mining. *IEEE Trans Knowl Data Eng* 33(4):1763–1777
- Liu J, Xia F, Feng X et al (2022) Deep graph learning for anomalous citation detection. *IEEE Trans Neural Netw Learn Syst* 33(6):2543–2557
- Liu J, Bai X, Wang M et al (2024) Anomalous citations detection in academic networks. *Artif Intell Rev* 57(4):103
- Liu Z, Huang H, Wei X, et al (2014) Tri-rank: An authority ranking framework in heterogeneous academic networks by mutual reinforce. In: 2014 IEEE 26th International Conference on Tools with Artificial Intelligence, IEEE, pp 493–500
- Loh WY (2011) Classification and regression trees. *Wiley Interdiscip Rev Data Min Knowl Discov* 1(1):14–23
- Ma A, Liu Y, Xu X et al (2021) A deep-learning based citation count prediction model with paper metadata semantic features. *Scientometrics* 126(8):6803–6823
- Massucci FA, Docampo D (2019) Measuring the academic reputation through citation networks via pagerank. *J Informet* 13(1):185–201
- McNamara D, Wong P, Christen P, et al (2013) Predicting high impact academic papers using citation network features. In: *Trends and Applications in Knowledge Discovery and Data Mining: PAKDD 2013 International Workshops: DMAApps, DANTh, QIMIE, BDM, CDA, CloudSD, Gold Coast, QLD, Australia, April 14–17, 2013, Revised Selected Papers 17*, Springer, pp 14–25
- Meng Q, Kennedy PJ (2013) Discovering influential authors in heterogeneous academic networks by a co-ranking method. In: *Proceedings of the 22nd ACM international conference on information & knowledge management*, pp 1029–1036
- Mercier D, Rizvi STR, Rajashekar V, et al (2020) Impactcite: An xlnet-based method for citation impact analysis. *arXiv preprint arXiv:2005.06611*
- Moed HF, Halevi G (2015) Multidimensional assessment of scholarly research impact. *J Am Soc Inf Sci* 66(10):1988–2002
- Pech G, Delgado C (2020) Assessing the publication impact using citation data from both scopus and wos databases: an approach validated in 15 research fields. *Scientometrics* 125(2):909–924
- Porwal P, Devare MH (2024) Citation count prediction using weighted latent semantic analysis (wlsa) and three-layer-deep-learning paradigm: a meta-heuristic approach. *Multim Tools Appl* 83(11):32245–32276

- Porwal P, Devare MH (2024) Scientific impact analysis: unraveling the link between linguistic properties and citations. *J Informet* 18(3):101526
- Qiu J, Han X (2024) An early evaluation of the long-term influence of academic papers based on machine learning algorithms. *IEEE Access* 12:41773–41786
- Radicchi F, Weissman A, Bollen J (2017) Quantifying perceived impact of scientific publications. *J Informet* 11(3):704–712
- Ramani RS, Aguinis H, Coyle-Shapiro JA (2022) Defining, measuring, and rewarding scholarly impact: mind the level of analysis. *Acad Manage Learn Educ* 21(3):470–486
- Revesz PZ (2014) A method for predicting citations to the scientific publications of individual researchers. In: *Proceedings of the 18th international database engineering & applications symposium*, pp 9–18
- Robson BJ, Mousquès A (2016) Can we predict citation counts of environmental modelling papers? fourteen bibliographic and categorical variables predict less than 30% of the variability in citation counts. *Environ Model Softw* 75:94–104
- Ruan X, Zhu Y, Li J et al (2020) Predicting the citation counts of individual papers via a bp neural network. *J Informet* 14(3):101039
- Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. *Nature* 323(6088):533–536
- Sarigöl E, Pfützner R, Scholtes I et al (2014) Predicting scientific success based on coauthorship networks. *EPJ Data Sci* 3:1–16
- Sayyadi H, Getoor L (2009) Futurerank: Ranking scientific articles by predicting their future pagerank. In: *Proceedings of the 2009 SIAM international conference on data mining*, SIAM, pp 533–544
- Schulz E, Speckenbrink M, Krause A (2018) A tutorial on gaussian process regression: modelling, exploring, and exploiting functions. *J Math Psychol* 85:1–16
- Shahmandi M, Wilson P, Thelwall M (2020) A new algorithm for zero-modified models applied to citation counts. *Scientometrics* 125:993–1010
- Shaikh AR, Alhoori H, Sun M (2023) Youtube and science: models for research impact. *Scientometrics* 128(2):933–955
- Shen H, Wang D, Song C, et al (2014) Modeling and predicting popularity dynamics via reinforced poisson processes. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, pp 1–9
- Singh M, Patidar V, Kumar S, et al (2015) The role of citation context in predicting long-term citation profiles: An experimental study based on a massive bibliographic text dataset. In: *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*, pp 1271–1280
- Singh M, Jaiswal A, Shree P, et al (2017) Understanding the impact of early citers on long-term scientific impact. In: *2017 ACM/IEEE joint conference on digital libraries (JCDL)*, IEEE, pp 1–10
- Siudem G, Żogała-Siudem B, Cena A et al (2020) Three dimensions of scientific impact. *Proc Natl Acad Sci* 117(25):13896–13900
- Stegehuis C, Litvak N, Waltman L (2015) Predicting the long-term citation impact of recent publications. *J Informet* 9(3):642–657
- Stoian MC, Tardios JA, Samdanis M (2024) The knowledge-based view in international business: a systematic review of the literature and future research directions. *International Business Review*, p 102239
- Su X, Yan X, Tsai CL (2012) Linear regression. *Wiley Interdiscip Rev Comput Stat* 4(3):275–294
- Sun K, Qiu Z, Salinas A, et al (2022) Assessing scientific research papers with knowledge graphs. In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp 2467–2472
- Sutskever I, Vinyals O, Le QV (2014) Sequence to sequence learning with neural networks. *Adv Neural Inf Process Syst* 27:1–8
- Tang J, Qu M, Wang M, et al (2015) Line: Large-scale information network embedding. In: *Proceedings of the 24th international conference on world wide web*, pp 1067–1077
- Thafar MA, Alsulami MM, Albaradei S (2024) Futurecite: predicting research articles' impact using machine learning and text and graph mining techniques. *Math. Comput. Appl.* 29(4):59
- Thelwall M, Nevill T (2018) Could scientists use altmetric. com scores to predict longer term citation counts? *J Informet* 12(1):237–248
- Thelwall M, Kousha K, Wilson P et al (2023) Predicting article quality scores with machine learning: The uk research excellence framework. *Quant Sci Stud* 4(2):547–573
- Tu H, Yu S, Saikrishna V et al (2023) Deep outdated fact detection in knowledge graphs. *IEEE Int Conf Data Mining Workshops (ICDMW)*. IEEE, Shanghai, China, pp 1443–1452
- Vaccario G, Medo M, Wider N et al (2017) Quantifying and suppressing ranking bias in a large citation network. *J Informet* 11(3):766–782
- Veličković P, Cucurull G, Casanova A, et al (2018) Pgraph attention networks. In: *Proceedings of International Conference on Learning Representations*, pp 1–12
- Waltman L (2016) A review of the literature on citation impact indicators. *J Informet* 10(2):365–391

- Wan X, Liu F (2014) Are all literature citations equally important? automatic citation strength estimation and its applications. *J Am Soc Inf Sci* 65(9):1929–1938
- Wang D, Song C, Barabási AL (2013) Quantifying long-term scientific impact. *Science* 342(6154):127–132
- Wang H, Fu T, Du Y et al (2023) Scientific discovery in the age of artificial intelligence. *Nature* 620(7972):47–60
- Wang K, Shi W, Bai J et al (2021) Prediction and application of article potential citations based on nonlinear citation-forecasting combined model. *Scientometrics* 126:6533–6550
- Wang M, Zhang J, Jiao S et al (2019) Evaluating the impact of citations of articles based on knowledge flow patterns hidden in the citations. *PLoS One* 14(11):e0225276
- Wang M, Zhang X, Zhong H, et al (2023b) Airank: An algorithm on evaluating the academic influence of papers based on heterogeneous academic network. *Journal of Information Science*, p 01655515231151406
- Wang R, Li S, Yin Q et al (2024) Improved pagerank and new indices for academic impact evaluation using ai papers as case studies. *J Inf Sci* 50(3):690–702
- Wang S, Xie S, Zhang X, et al (2014) Future influence ranking of scientific literature. In: *Proceedings of the 2014 SIAM international conference on data mining*, SIAM, pp 749–757
- Wang S, Xie S, Zhang X et al (2016) Coranking the future influence of multiobjects in bibliographic network through mutual reinforcement. *ACM Trans Intell Syst Technol (TIST)* 7(4):1–28
- Wang X, Zhang Z (2020) Improving the reliability of short-term citation impact indicators by taking into account the correlation between short-and long-term citation impact. *J Informet* 14(2):101019
- Wang Y, Tong Y, Zeng M (2013b) Ranking scientific articles by exploiting citations, authors, journals, and time information. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, pp 933–939
- Wang Y, Zeng A, Fan Y et al (2019) Ranking scientific publications considering the aging characteristics of citations. *Scientometrics* 120:155–166
- Weihls L, Etzioni O (2017) Learning to predict citation-based impact measures. In: *2017 ACM/IEEE joint conference on digital libraries (JCDL)*, IEEE, pp 1–10
- de Winter J (2024) Can chatgpt be used to predict citation counts, readership, and social media interaction? an exploration among 2222 scientific abstracts. *Scientometrics*, pp 1–19
- Xia F, Wang W, Bekele TM et al (2017) Big scholarly data: a survey. *IEEE Tran Big Data* 3(1):18–35
- Xia F, Wang L, Tang T et al (2022) CenGCN: Centralized convolutional networks with vertex imbalance for scale-free graphs. *IEEE Trans Knowl Data Eng* 35(5):4555–4569
- Xia W, Li T, Li C (2023) A review of scientific impact prediction: tasks, features and methods. *Scientometrics* 128(1):543–585
- Xiao C, Sun L, Han J et al (2022) Heterogeneous academic network embedding based multivariate random-walk model for predicting scientific impact. *Appl Intell* 52(2):2171–2188
- Xiao S, Yan J, Li C, et al (2016) On modeling and predicting individual paper citation count over time. In: *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pp 2676–2682
- Xu L, Wei X, Cao J, et al (2017) Embedding of embedding (eoe) joint embedding for coupled heterogeneous networks. In: *Proceedings of the tenth ACM international conference on web search and data mining*, pp 741–749
- Yan E, Ding Y, Sugimoto CR (2011) P-rank: an indicator measuring prestige in heterogeneous scholarly networks. *J Am Soc Inform Sci Technol* 62(3):467–477
- Yan P, Kang Y, Jiang Z, et al (2024) Modeling scholarly collaboration and temporal dynamics in citation networks for impact prediction. In: *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp 2522–2526
- Yan R, Tang J, Liu X, et al (2011b) Citation count prediction: learning to estimate future citations for literature. In: *Proceedings of the 20th ACM international conference on Information and knowledge management*, pp 1247–1252
- Yan R, Huang C, Tang J, et al (2012) To better stand on the shoulder of giants. In: *Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries*, pp 51–60
- Yao L, Wei T, Zeng A et al (2014) Ranking scientific publications: the effect of nonlinearity. *Sci Rep* 4(1):6663
- Ye Z, Gao C, Jiang X et al (2015) Ranking scientific articles over heterogeneous academic network. *2015 11th International Conference on Semantics. IEEE, Knowledge and Grids (SKG)*, pp 268–271
- Yu S, Xia F, Li S et al (2023) Spatio-temporal graph learning for epidemic prediction. *ACM Trans Intell Syst Technol* 14(2):1–25
- Yu T, Yu G, Li PY et al (2014) Citation impact prediction for scientific papers using stepwise regression analysis. *Scientometrics* 101:1233–1252
- Zaeem JM, Garg V, Aggarwal K, et al (2023) An intelligent article knowledge graph formation framework using bm25 probabilistic retrieval model. In: *Iberoamerican Knowledge Graphs and Semantic Web Conference*, Springer, pp 32–43

- Zhang F, Wu S (2018) Ranking scientific papers and venues in heterogeneous academic networks by mutual reinforcement. In: Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries, pp 127–130
- Zhang F, Wu S (2020) Predicting future influence of papers, researchers, and venues in a dynamic academic network. *J Informet* 14(2):101035
- Zhang F, Wu S (2021) Measuring academic entities' impact by content-based citation analysis in a heterogeneous academic network. *Scientometrics* 126(8):7197–7222
- Zhang F, Wu S (2024) Predicting citation impact of academic papers across research areas using multiple models and early citations. *Scientometrics*, pp 1–30
- Zhang L, Fan Y, Zhang W et al (2019) Measuring scientific prestige of papers with time-aware mutual reinforcement ranking model. *J Intell Fuzzy Syst* 36(2):1505–1519
- Zhang Y, Wang M, Gottwalt F et al (2019) Ranking scientific articles based on bibliometric networks with a weighting scheme. *J Informet* 13(2):616–634
- Zhang Y, Wang M, Saberi M et al (2022) Analysing academic paper ranking algorithms using test data and benchmarks: an investigation. *Scientometrics* 127(7):4045–4074
- Zhao F, Zhang Y, Lu J et al (2019) Measuring academic influence using heterogeneous author-citation networks. *Scientometrics* 118:1119–1140
- Zhao P, Han J, Sun Y (2009) P-rank: a comprehensive structural similarity measure over information networks. In: Proceedings of the 18th ACM conference on Information and knowledge management, pp 553–562
- Zhao P, Xing Q, Dou K, et al (2024) From words to worth: Newborn article impact prediction with llm. *arXiv preprint arXiv:2408.03934*
- Zhao Q, Feng X (2022) Utilizing citation network structure to predict paper citation counts: a deep learning approach. *J Informet* 16(1):101235
- Zhou J, Zeng A, Fan Y et al (2016) Ranking scientific publications with similarity-preferential mechanism. *Scientometrics* 106:805–816
- Zhou Y, Li Q, Yang X et al (2021) Predicting the popularity of scientific publications by an age-based diffusion model. *J Informet* 15(4):101177
- Zhou Y, Wang R, Zeng A (2022) Predicting the impact and publication date of individual scientists' future papers. *Scientometrics* 127(4):1867–1882
- Zhu X, Turney P, Lemire D et al (2015) Measuring academic influence: not all citations are equal. *J Am Soc Inf Sci* 66(2):408–427
- Zhu XP, Ban Z (2018) Citation count prediction based on academic network features. In: 2018 IEEE 32nd international conference on advanced information networking and applications (AINA), IEEE, pp 534–541

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Xiaomei Bai^{1,2} · Fuli Zhang³ · Jiaying Liu⁴ · Xiaoxia Wang⁵ · Feng Xia⁶

✉ Xiaoxia Wang
spice.wang@rmit.edu.au

Xiaomei Bai
xiaomeibai@outlook.com

Fuli Zhang
zfuli@outlook.com

Jiaying Liu
jiaying_liu@outlook.com

Feng Xia
f.xia@ieee.org

¹ School of Artificial Intelligence, Anshan Normal University, Anshan 114007, China

- ² Big Data Research Institute, Anshan Normal University, Anshan 114007, China
- ³ Information Center, Anshan Normal University, Anshan 114007, China
- ⁴ School of Economics and Management, Dalian University of Technology, Dalian 116024, China
- ⁵ Centre for Educational Innovation and Quality, RMIT University, Melbourne, VIC 3000, Australia
- ⁶ School of Computing Technologies, RMIT University, Melbourne, VIC 3000, Australia