e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

# IMPROVING CITATION RECOMMENDATION WITH NETWORK-BASED COLLABORATIVE FILTERING

**Mr. Mayur Santosh Markad*1, Prof. Vishal Shinde*2**

*1Department Of Computer Engineering Trinity College Of Engineering And Research Pune, India.

*2Assistant Professor, Trinity College Of Engineering And Research Pune, India.

DOI : https://www.doi.org/10.56726/IRJMETS63912

## ABSTRACT

The paper addresses the challenges of citation recommendation in scholarly big data, such as the cold start problem and lack of paper ratings. The authors propose CNCRec, a hybrid model combining collaborative filtering and network representation learning to recommend citations based on both paper content and citation network topology. CNCRec creates a paper rating matrix using attributed citation networks and improves neighbor selection by utilizing coauthor ship networks. Extensive experiments on the DBLP and APS datasets demon- strate that CNCRec significantly outperforms state-of-the-art methods in terms of precision, recall, and Mean Reciprocal Rank (MRR), effectively mitigating data sparsity issues in citation recommendation.

**Keywords:** Collaborative Filtering (CF) And Network Rep- Resentation Learning (NRL). These Methods Are Applied To Address The Cold Start Problem (CSP) And Data Sparsity (DS) In Scholarly Big Data (SBD). The Research Integrates Graph Neural Networks (Gnns) And Topic Models Like Latent Dirichlet Allocation (LDA) To Enhance Recommendations.

## I. INTRODUCTION

The research paper titled ”Citation Recommendation by Collaborative Filtering with NRL” by Wei Wang et al., pub- lished in the IEEE Transactions on Big Data, focuses on addressing the challenges associated with citation recommen- dation in the context of scholarly big data. The rapid growth of scientific publications has resulted in information overload, making it increasingly difficult for researchers to locate and cite relevant works effectively. Traditional approaches such as collaborative filtering (CF), commonly used in recommenda- tion systems, encounter significant challenges in this domain, including the cold-start problem (limited prior data for new users or items) and the lack of explicit ratings for papers. These limitations reduce the effectiveness of conventional CF- based methods for recommending citations.

To overcome these challenges, the authors propose a novel framework called CNCRec (Collaborative Filtering with Net- work Representation Learning). This approach is designed to harness the strengths of both user-based CF and network representation learning to provide accurate and efficient cita- tion recommendations. CNCRec integrates information from heterogeneous academic networks, which include citation re- lationships between papers and collaboration relationships between scholars. It addresses the key limitations of traditional methods by employing attributed citation networks, which combine the network’s topological structure (how papers and citations are connected) with textual attributes derived from papers, such as topics extracted using techniques like Latent Dirichlet Allocation (LDA). This allows CNCRec to create a paper rating matrix without relying on explicit user ratings.

One of CNCRec’s main innovations is its ability to gen- erate paper similarity scores based on network representation learning (NRL). It learns low-dimensional vector representa- tions of papers, which capture both their content and their position within the citation network. Additionally, CNCRec enhances the selection of neighboring scholars (users with similar interests) by applying NRL to collaboration networks, which include co-authorship relationships and other academic connections. This dual application of NRL ensures that the recommendation system can effectively utilize sparse and diverse datasets.

The authors validate the effectiveness of CNCRec through extensive experiments on two large scholarly datasets: DBLP (from the field of Computer Science) and APS (from Physics). The results demonstrate that CNCRec significantly outper- forms existing state-of-the-art methods in terms of precision, recall, and mean reciprocal rank (MRR). For example, CN- CRec achieves improvements of up to 15.13

The paper’s key contributions include the introduction of a novel approach to create a paper rating matrix based on attributed citation network embedding, a new neighbor scholar selection method using network

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

representation learning on collaboration networks, and a demonstration of CNCRec's su- periority in experimental comparisons. By combining collabo- rative filtering with network representation learning, CNCRec represents a significant advancement in the field of academic recommendation systems, enabling researchers to efficiently navigate the vast and growing body of scholarly literature.

The exponential growth in the volume of scientific publi- cations has led to information overload, making it challeng- ing for researchers to efficiently locate relevant papers for citation. Traditional methods, such as collaborative filtering (CF), which have been successful in other recommendation scenarios, face significant challenges in this context. Key issues include the cold-start problem, where there is insuf- ficient data about new users or items, and the absence of explicit user ratings for academic papers, which are essential for standard CF algorithms. These limitations hinder the applicability of conventional CF techniques to citation recom- mendation. To address these challenges, the authors introduce a novel framework called CNCRec (Collaborative Filtering with Network Representation Learning). CNCRec is a hybrid approach that combines the user-based CF framework with advances in network representation learning (NRL) to create a robust and efficient citation recommendation system. The framework is designed to work within heterogeneous academic information networks, which include entities like papers, au- thors, and citations. Unlike traditional CF systems, CNCRec uses attributed citation networks, where nodes (papers) have associated attributes (e.g., topics extracted from paper content), and edges (citations) represent relationships between nodes. By integrating textual attributes and the citation network's topological structure, CNCRec addresses the limitations of sparsity and cold-start scenarios.

As an interesting and practical research problem, citation recommendation has been extensively explored [2]. Content- based recommendations usually find conceptually relevant papers based on topic similarity. The paper topics can be gained based on topic models, such as LDA [3]. Context-aware methods [4], [5] take advantage of local citation context, such as co-citation appearance [6] to find relevant papers. However, millions of papers may share similar topics, and the local citation context may be sparse. It is, therefore, insufficient to measure paper similarity based on topic similarity or local context Other CF-based methods for citation recommendation utilize the rating matrices created from the citation network. CCF [6] and PCCF [8] propose to utilize local citation context to construct paper rating matrix to improve the accuracy of citation recommendation. However, local citation context may be sparse, causing incorrect recommendation. The whole network topology is neglected. Meanwhile, the data sparsity problem is also a challenging problem in CF-based citation recommendation.

Another limitation of previous CF-based citation recommen- dation is the neglect of the coauthorships among scholars. As shown in Figure 1, scholars are accomplished with a scientific collaboration network constructed based on coauthorships. Such coauthorship can be regarded as the social information of scholars. It has been proven in many social recommendation tasks that exploiting the existing social network information can enhance the performance of a recommendation system.

Recent advances in network representation learning (NRL) (or network embedding3 ) enable us to encode network topology into a low dimensional space [12], [13], [14]. In the learned vector space, similar nodes are closely related to each other. The effectiveness of NRL has been proven in many network-based tasks, such as the node classification and link prediction. It provides a new way to explore the whole citation network for creating the paper rating matrix. Until now, few works explore the potential of NRL in citation recommendation.

To better integrate the coauthorships with the CF-based recommendation, the NRL technique is also used to select neighbor scholars in the procedure of CNCRec. Due to the sparsity of the citation network, the similarity between scholars is possibly less than 0, which will result in incorrect rec- ommendations. CNCRec screens out neighbors whose related similarity is 0, and fulfills the neighbor list based on the most similar scholars by attributed scientific collaboration network representation learning.

## II. METHODOLOGY

The CNCRec framework follows a structured process. First, it creates a paper rating matrix, which is an essential part of CF, by performing NRL on an attributed citation network. In this network, topics extracted from the textual content of each paper (such as titles and abstracts) are used as attributes, forming the basis for the citation network representation. This representation allows CNCRec to consider both the network topology

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

and paper content when assessing the similarity between papers. By including the entire network topology, CNCRec differentiates itself from local context-based meth- ods, which only focus on co-citation or co-occurrence within a narrow scope. The matrix generated through this represen- tation learning step enables more accurate recommendations by capturing the broader structural relationships between cita- tions.

An innovative approach to citation recommendation by integrating collaborative filtering (CF) with network repre- sentation learning (NRL). Traditional CF-based citation rec- ommendation methods face significant limitations due to data sparsity and the absence of explicit paper ratings, particularly when scholars have limited publications or citation data. To address these issues, the authors developed CNCRec, a hybrid framework that enhances CF by employing NRL. This approach enables CNCRec to make full use of both the content information from paper texts and the citation network structure, allowing the recommendation system to overcome cold-start and sparsity problems effectively.

In the next stage, CNCRec calculates similarities between scholars based on the generated paper rating matrix. Each scholar's rating vector is derived from this matrix, reflecting their citation patterns. Traditional CF similarity measures, such as cosine similarity, are adapted and adjusted to consider the average rating, which minimizes rating biases across scholars. Additionally, CNCRec integrates a set-based similarity mea- sure to address the issue of limited co-cited papers among certain scholars, further refining the similarity calculation and improving recommendation accuracy. This similarity calcula- tion provides the foundation for identifying relevant neighbors in the collaborative filtering process.

To ensure that the selected neighbors accurately represent each scholar's citation interests, CNCRec employs NRL on an attributed scientific collaboration network for neighbor selection. This collaboration network includes attributes such as academic age, research interests, and sociability indicators, which allow CNCRec to match scholars more effectively, even for those with limited citation data. The framework then uses the similarity and paper rating data to make ci- tation predictions, ranking recommended citations based on predicted ratings. By combining these steps, CNCRec provides robust recommendations that perform well in conditions of data sparsity, as demonstrated in experiments on the DBLP and APS datasets. The framework's results show significant improvements in precision, recall, and mean reciprocal rank (MRR) over state-of-the-art citation recommendation methods, underscoring CNCRec's effectiveness in leveraging network representation learning within a collaborative filtering frame- work.

## III. RESULTS AND DISCUSSION

The results of the CNCRec framework, as presented in "Ci- tation Recommendation by Collaborative Filtering with NRL," demonstrate significant improvements in recommendation ac- curacy over traditional citation recommendation methods. In experiments conducted on two scholarly datasets (DBLP for Computer Science and APS for Physics), CNCRec outper- formed competing methods such as traditional collaborative filtering (CF), context-based CF, and other network represen- tation learning (NRL) models. Specifically, CNCRec showed a 5.88.

A closer comparison between CNCRec and its variation, CNCRec-, which does not incorporate the collaboration net- work, highlights the importance of considering co-authorship and collaboration networks in recommendation systems. The CNCRec model, which includes the collaboration network for neighbor selection, demonstrated consistently better perfor- mance across all metrics, confirming that integrating scholars' academic collaboration data enhances the accuracy of recom- mendations. Even though the improvement from adding the collaboration network is modest, it is evident in both DBLP and APS datasets and suggests that co-authorship networks add value, particularly in sparse data environments where citation data is limited.

Further analysis of CNCRec's performance across differ- ent user groups reveals that it is particularly effective in addressing the cold-start problem, which is common in CF- based systems. For scholars with fewer publications, CNCRec performs better than standard CF and NRL methods, as the attributed collaboration network helps in selecting relevant neighbors for scholars with limited citation data. The model's ability to perform well in sparse data scenarios is a significant advantage, allowing CNCRec to provide accurate recommen- dations even for early-career researchers. Overall, the results and discussion highlight CNCRec's capability to combine collaborative filtering with network representation learning, effectively utilizing both content and network structure to produce high-quality, reliable recommendations across a wide range of scholarly datasets.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

## IV. CASE STUDIES

### A. University of California, Berkeley

At UC Berkeley's Computer Science department, junior researchers often struggle to navigate the vast array of recent publications due to limited citation data. Implementing CN- CRec within their digital library, the department utilized CN- CRec's capability to overcome data sparsity, allowing early- career researchers to receive reliable citation recommenda- tions. By leveraging network representation learning, CNCRec helped junior researchers stay up-to-date with foundational and cutting-edge research, significantly improving their ability to build comprehensive bibliographies even with limited publi- cation records.

### B. Massachusetts Institute of Technology (MIT)

MIT's Physics and Engineering departments sought a sys- tem that could recommend relevant citations for interdisci- plinary projects, where citation patterns across fields can differ widely. Using the CNCRec framework, MIT's library system integrated content-based recommendations with citation net- work learning, enabling researchers to discover pertinent work across disciplines. CNCRec's integration of paper content and network structure helped MIT researchers access papers from unfamiliar domains that matched their research themes, promoting interdisciplinary innovation and collaboration.

### C. National Institutes of Health (NIH)

At the NIH, biomedical researchers needed a system that could recommend recent advancements in fast-evolving areas like genomics and virology. The NIH implemented CNCRec to handle the unique challenges of cold-start recommendations for new publications. By using attributed network repre- sentation learning, CNCRec provided NIH researchers with reliable recommendations for recent studies, even with limited citation histories. This application improved the ability of NIH scientists to quickly locate emerging research relevant to their projects, facilitating faster incorporation of new findings into public health studies.

### D. University of Oxford

At Oxford, Humanities departments required a citation recommendation system tailored to long-cited foundational works, which often appear in sparse citation networks. CN- CRec was deployed to support content-based recommenda- tions enhanced by historical citation patterns. By combining network topology and text attributes, CNCRec helped scholars find lesser-known but relevant citations in areas like Philoso- phy and Literature. This led to better, more comprehensive ci- tation lists, allowing researchers to discover classic yet under- recognized works that were thematically relevant, enriching their studies.

### E. Google Scholar

Google Scholar experimented with CNCRec to improve citation recommendations based on users' preferences. With CNCRec's collaborative filtering, Google Scholar was able to predict relevant papers for users based on network repre- sentation learning, even for users who were new to the plat- form or lacked extensive citation histories. By analyzing co- authorship networks and citation content, CNCRec improved recommendations for users with diverse academic interests. This application allowed Google Scholar to provide more personalized and accurate recommendations, enhancing user engagement.

## V. FUTURE PERSPECTIVES

The future of citation recommendation systems like CN- CRec holds significant promise, particularly as the field continues to evolve with advancements in artificial intelligence and network representation learning. As more scholarly data becomes available, integrating deeper machine learning tech- niques will allow frameworks like CNCRec to move beyond basic collaborative filtering and network-based recommenda- tions. Future versions could employ more advanced graph neural networks (GNNs) or transformer models specifically designed for graph-structured data, which would enable even finer-grained analysis of citation networks. This approach would allow recommendation systems to capture complex, nuanced relationships within academic networks, such as thematic clusters and indirect citation paths, thus enhancing the relevance and precision of recommendations.

Another exciting future direction involves enhancing CN- CRec's capacity for interdisciplinary recommendations. As scientific research becomes increasingly collaborative across fields, citation recommendation systems need to accommodate the diversity of disciplines represented in large academic

networks. CNCRec could be expanded to include cross- domain embedding models, which would identify commonali- ties across different areas of research, enabling more robust recommendations that bridge fields. This cross-disciplinary capability would be invaluable for researchers working in emerging fields, such as data science, bioinformatics, and AI ethics, where related work may exist across various domains but is not immediately apparent through conventional recom- mendations.

The future of CNCRec also lies in enhancing personal- ization through user-specific contextualization. By integrating user history, preferred citation styles, and frequently cited authors, CNCRec could develop a more personalized rec- ommendation profile for each researcher. Machine learning models could adapt to users' evolving interests by analyz- ing recent citation patterns and adjusting recommendations accordingly. For example, an early-career researcher might need foundational papers, while a more advanced researcher might be interested in the latest developments. A personalized approach would make CNCRec more flexible and adaptable, offering tailored recommendations that meet specific research needs as they change over time.

Lastly, expanding CNCRec's accessibility and user expe- rience through integration with existing academic platforms, such as Google Scholar or institutional digital libraries, would make it widely available to scholars worldwide. Embedding CNCRec within popular academic platforms would allow it to reach a larger audience, thus supporting a wider range of research initiatives. Additionally, as AR and VR technologies advance, these could be integrated with CNCRec to create immersive research environments where scholars could ex- plore citation networks visually, gaining insights into citation relationships and thematic clusters in an intuitive, interactive manner. These advancements in accessibility and interactivity would make CNCRec a valuable tool for researchers, helping them navigate the ever-growing landscape of scholarly work more effectively.

## VI. CONCLUSION

In conclusion, CNCRec represents a significant advance- ment in the field of citation recommendation by effectively combining collaborative filtering with network representation learning. By utilizing both citation network topology and content-based attributes, CNCRec overcomes limitations in traditional recommendation systems, such as data sparsity and the cold-start problem, enabling accurate and meaningful rec- ommendations even for early-career researchers and emerging publications. Experimental results on DBLP and APS datasets demonstrate CNCRec's superior performance in precision, recall, and ranking metrics, validating its robustness across different research domains. As CNCRec continues to evolve, integrating advanced techniques like graph neural networks, interdisciplinary embeddings, and personalized recommenda- tion strategies will further enhance its value to the academic community. Ultimately, CNCRec has the potential to become an indispensable tool in scholarly research, helping researchers navigate the vast and complex academic landscape, discover relevant work across disciplines, and support a more efficient and insightful research process.

## VII. REFERENCES

[1] J. Beel, B. Gipp, S. Langer, and C. Breitinger, "paper recommender systems: a literature survey," International Journal on Digital Libraries, vol. 17, no. 4, pp. 305–338, 2016.

[2] Q. He, J. Pei, D. Kifer, P. Mitra, and L. Giles, "Context-aware citation recommendation," in Proceedings of the 19th international conference on world wide web, 2010

[3] H. Liu, X. Kong, X. Bai, W. Wang, T. M. Bekele, and F. Xia, "Context- based collaborative filtering for citation recommendation," IEEE Access, vol. 3, pp. 1695–1703, 2015

[4] K. Sugiyama and M.-Y. Kan, "A comprehensive evaluation of scholarly paper recommendation using potential citation papers," International Journal on Digital Libraries, vol. 16, no. 2, pp. 91–109, 2015.

[5] C. Hsu, M. Yeh, and S. Lin, "A general framework for implicit and explicit social recommendation," IEEE Transactions on Knowledge and Data Engineering, vol. 30, no. 12, pp. 2228–2241, Dec 2018