International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Cloud-Enabled Social Network Mining for Advanced Recommendation Systems: An Integrated Data Mining and Social Network Analysis Approach

Jaishree Jain, Santosh Kumar Upadhyay, Sharvan Kumar, Neerja Arora

Submitted: 05/02/2024 **Revised**: 13/03/2024 **Accepted**: 19/03/2024

Abstract: Recommendation systems play a crucial role in assisting users in discovering relevant and personalized content in social networks. For recommendation system data mining has a profound impact on several domains, including databases, artificial intelligence, machine learning, and social networks. It plays a crucial role in driving significant research advancements in the field. In today's fast-paced world, where data is rapidly expanding and information retrieval poses complex challenges, users increasingly demand valuable insights from their vast datasets. Social networks have emerged as a fascinating domain that has made substantial contributions to data mining research, ushering in a new era of possibilities. To determine the intrusion index based on the source address of the network security alarm, a simulation test is run. The findings demonstrate that this strategy can successfully implement cloud network security situation awareness as the related window attack index drops as soon as the security event is cancelled. You can accurately detect changes in network security circumstances using the suggested technique.

***Keyword**: Cloud Computing, Social Network, Data mining, Recommendation System, Normalization.*

1. Introduction

In recent years, social networks have become a prominent platform for individuals to connect, interact, and share information. By combining data mining techniques with social network analysis, we aim to leverage the rich context and social connections within the network to improve the quality of recommendations [1,2]. This research will investigate various methodologies, algorithms, and models that effectively utilize social network information and user behavior data to generate personalized recommendations.

Furthermore, this study will address several challenges and considerations associated with mining social networks for recommendation systems [3]. These challenges include dealing with the scale and complexity of social network data, handling data sparsely, accounting for evolving network dynamics, and ensuring privacy and security of user information. By addressing these challenges, we can develop robust and efficient algorithms that effectively mine social networks to deliver accurate and reliable recommendations.

Link Mining is focused on links between objects rather than objects themselves and emerging as technical approach for finding useful information and hidden pattern in social networks [4, 5]. Link mining has number of tasks associated with it in which Link prediction is the one having the objective of predicting future relationships among individuals in a social network.

A straightforward binary classification problem can be seen as the link prediction problem [6,7]: Determine whether L_{ij} is 0 or 1 for any two objects O_i and O_j that might be related. Many scholars have proposed many effective link prediction algorithms. Some of the approaches mentioned in [8, 9] took topological features, but they neglected to take these features' normalization aspects into account to increase predictor efficiency [10, 11].

2. Related Work

Multidimensional clustering was suggested as an addition to a collaborative filtering recommendation model by Li et al. In the first stage, the suggested technique was used to gather and cluster background data in the form of person and object profiles. After that, the subpar clusters with comparable attributes were eliminated, and the suitable clusters were chosen once more via cluster pruning. A weighted average of the departures from the neighbor's mean was used to make an item prediction in the third stage. Such a strategy would probably compromise on keeping the recommendations' accuracy while broadening their range. [12], [13]

Zhou et al. considered the composite relation between input, output, and semantic relations between them in order to represent Data-Providing (DP) service in terms of vectors. The C-means algorithm was enhanced and fuzzy to cluster the vectors. The capacity of the service search engine was

Department of CSE, Ajay Kumar Garg Engineering College Ghaziabad
jaishree3112@gmail.com,
kumarsatntosh@akgec.ac.in,kumarsharvan@akgec.ac.in,
aroraneerja@akgec.ac.in

greatly enhanced by grouping related services into a single cluster, particularly in big Internet-based service repositories. But this method assumes the existence of domain ontology to provide semantic interoperability [14.15]. Furthermore, this method is inappropriate for certain services that lack parameters.
In order to identify the linkages among the users, Simon et al. employed a high-dimensional parameter free, divisive hierarchical clustering algorithm that only requires implicit feedback on previous user purchases. Users were suggested high-interest products based on the clustering results. Implicit feedback, however, does not always offer precise details regarding the user's preferences [16, 17].

3. Methodology

Our proposed approach incorporates various elements from existing link prediction methodologies while introducing a significant perspective known as normalization. Our target social network(s) for recommendation system development is Co-Authorships Network data Taken from DBLP website database [12, 13]. Although some existing link prediction methods have explored attributive features [14, 15], they often overlook the importance of normalization, which is crucial for enhancing predictor efficiency. The systematic approach of our link prediction method is illustrated inthe figure presented below, outlining the step-by-step process of predicting links within a social network. By integrating normalization and leveraging topological patterns, our approach aims to improve the accuracy and reliability of link prediction in social networks as shown in Fig 1.

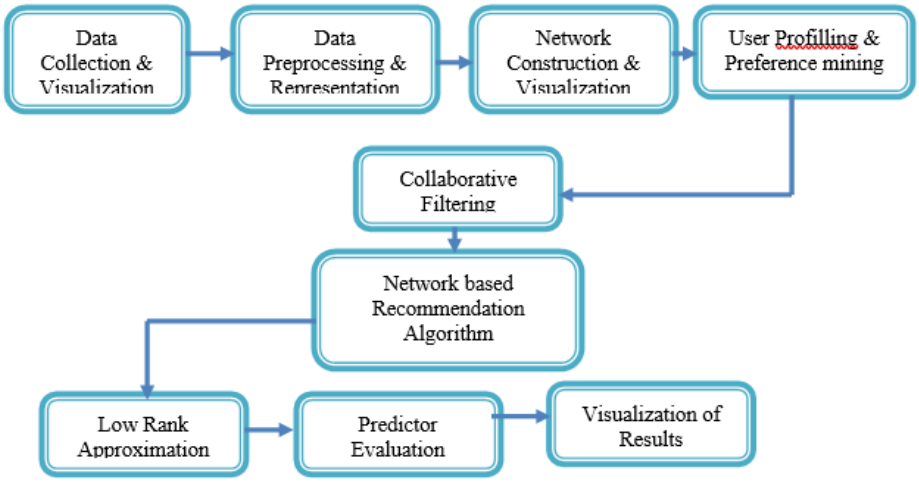

Fig. 1. Methodology for Link Data Mining

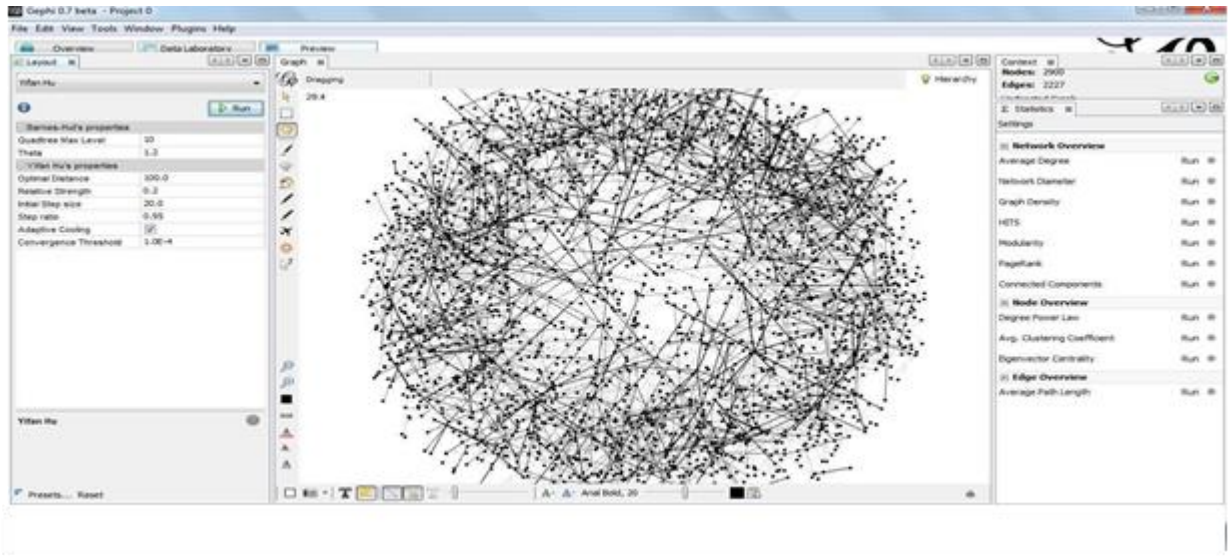

Fig. 2. Co-authorship Network

Network construction comes after data has been preprocessed and transformed into an appropriate format. In a built network, two people are connected if they have a relationship. For instance, coauthor ship networks, in which two scholars are connected if they coauthored a publication, are built using bibliographic data. As seen in the image

below, one co-authorship network was created using bibliographic data that was obtained from the DBLP website. An open source network visualization tool called GEPHI is used to show this network as shown in Fig 2.

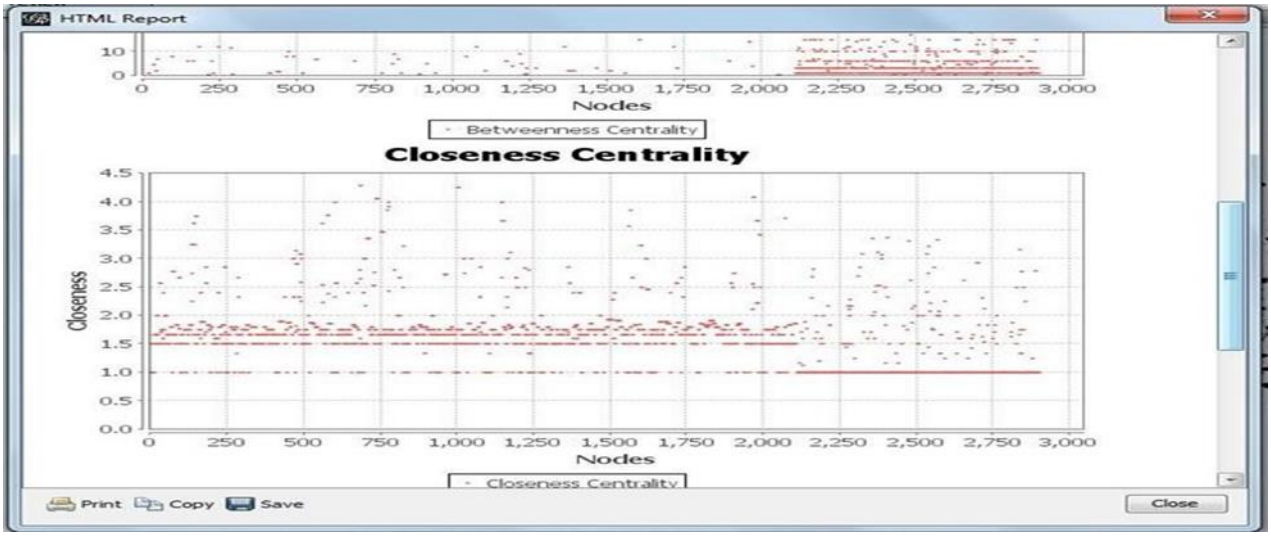

Fig. 3. Closeness Network

This attribute is frequently used in collaboration networks to affirm the number of x and y's shared neighbors.

If Vi and Vj are two nodes in a network, then their common neighbors can be calculated as

$$\text{Score } (Vi,Vj) = |\, Vi \cap Vj \,| \qquad (1)$$

In link prediction, the Jaccard's coefficient can be used to identify potential connections between nodes in a social network as shown in Figure 3. By computing the Jaccard's coefficient between the neighborhoods of two nodes, it is possible to identify nodes that have similar neighbors or shared connections. This information can be utilized to predict missing links or recommend new connections in the social network. Normalized score of common neighbors is given by,

$$\textbf{Normalized Score (Vi,Vj)} = \frac{|\, \mathbf{V_i} \cap \mathbf{V_j} \,|}{|\, \mathbf{V_i} \cup \mathbf{V_j} \,|} \qquad (2)$$

4. Predictor Evaluation and Visualization of Results

This attribute is frequently used in collaboration networks to affirm the number of x and y's shared neighbours.

If Vi and Vj are two nodes in a network, then their common neighbors can be calculated as:

$$\text{Recall} = \frac{|TP|}{|TP+FN|} \qquad (3)$$

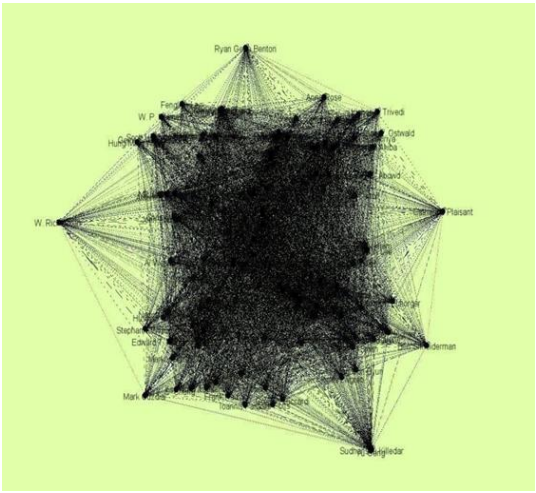

Fig. 4. Snap shot of clustered network

According to its definition, precision is the percentage of positive future forecasts that come true out of all positive future predictions.

$$\text{Precision} = \frac{|TP|}{|TP+FP|} \qquad (4)$$

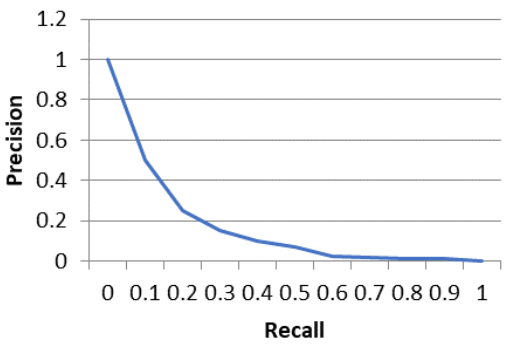

Fig. 5. Precision-Recall Curve

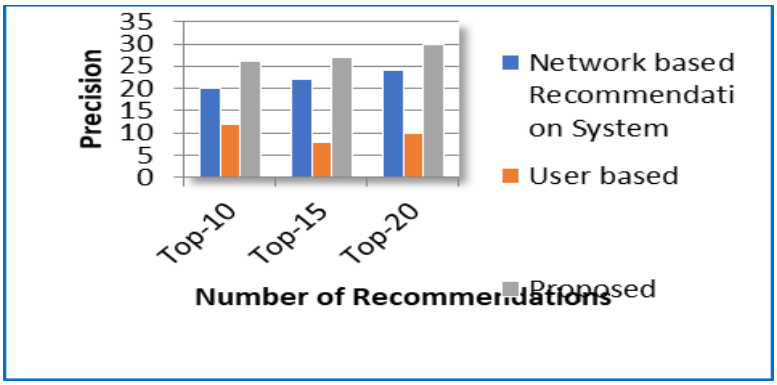

Fig.6. Precision comparison between the proposed approach and other recommendation approaches.

These metrics can be used to calculate the final values of a predictor, which represents the likelihood of future author collaboration. Author clusters aid in the visualization of results as well. Based on the likelihood that authors will work together, clusters are created.

5. Conclusions

Our study effort provided a framework that builds on previous work by incorporating a number of new network properties, and then a matrix-based approach is used to get successful outcomes. Based on numerous network metrics provided in this research effort, it is anticipated that partnerships will endure the secret to surviving in a cutthroat society. Our suggested methodology framework lays the groundwork for projecting future cooperation and takes important aspects of social networks' diversity and feature diversity into consideration. Although some of the existing approaches have investigated these traits, they have not taken into account a crucial factor known as normalization. To increase the predictor's effectiveness, the normalization component of these features is necessary. The suggested approach will produce useful results in more accurately anticipating future ties and potential cooperation between different scholars. Future research in this area of data mining is very active. In today's competitive environment, teamwork is the important concept, as was already stated before. Collaboration is a process by which two or more individuals can exchange ideas and expertise in order to produce prospective outcomes.

References

[1] Loyalty Square: Social Network Analysis https://www.indiamart.com/loyaltysquare/aboutus.html ".

[2] Liben-Nowell, D., Kleinberg, J.: "The link prediction problem for social Networks": Twelfth international conference on Information and Knowledge Management, ACM Press, New York, pp. 556-559(2003).

[3] Pavlov, M., Ichise, R.: "Finding experts by link prediction in co-authorship networks": 2nd International Workshop on Finding Experts on the Web with Semantics, pp. 42-55(2007).

[4] Mrinmaya Sachan, Ryutaro Ichise. "Using Abstract information and Community Alignment Information for Link Prediction": Second International Conference on Machine Learning and Computing (ICMLC), ISBN: 978-1-4244- 6006-9, pp. 61-65(2009).

[5] Jingfeng Guo, hongwei Guo."Mullti-features Link Prediction Based on Matrix": International Conference on ComputerDesign and Applications (ICCDA) ISBN: 978-1-4244-7164-5, V1-357 - V1-361(2010).

[6] Victor Stroele, Jonice Oliveria et. Al."Mining and Analyzing Multi Relational Social Networks": International Conference on Computational Science and Engineering, Print ISBN: 978- 1-4244-5334-4, pp. 711-716(2009).

[7] Getoor, L., Diehl, C.P.: "Link Mining: A Survey ": ACM-SIGKDD Explorations, Volume 7, Issue 2(2005).

[8] Evan Wei Xiang, Qiang Yang.: "A Survey on Link prediction Model for Social Network Data". Science and Technology (2008).

[9] Tarun Kumar, O. P. Vyas.: "Harnessing Social Network with link Data Mining for Predictive Analytics: An Extended Approach": 2nd IIMA International Conference on Advanced Data Analysis, Business Analytics and Intelligence (ICADABAI) (2011).

[10] Christopher D. Manning, Prabhakar Raghavan & Hinrich Schütze.: "Introduction to Information Retrieval", CambridgeUniversity Press (2008).

[11] Pavan, Rytaro Ichise, O.P. Vays.:"LiDDM: A Data Mining System for Linked Data":WWW 2011 workshop: LinkedData on the web (LDOW)(2011).

[12] Jennifer Neville, Foster Provost." Predictive modeling using social networks":14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2008).

[13] Jiawei Han, Yizhou Sun, Xifeng Yan, Philip S. Yu.:"Mining Knowledge from Databases: An information Network Analysis Approach": ACM SIGMOD Conference tutorial, Indianapolis, Indiana (2010).

[14] Knowledge Discovery Laboratory Dataset. DBLP (http://www.informatik.uni-trier.de/-Iey/db/index.html/). Retrieved Sep. 13, 2010 from DBLP database (2010).

[15] Aydoğdu, Ş. (2020). EDUCATIONAL DATA MINING STUDIES IN TURKEY: A SYSTEMATIC REVIEW. Turkish Online Journal of Distance Education, 21 (3), 170-185 . DOI: 10.17718/tojde.762046.

[16] Jain J.,Sahu S. and Sahu N., "Sentiment Analysis based on NLP using Learning Techniques", 2nd International Conference on Advance data driven computing, Scopus Indexed CRC Press Taylor & Francis,12-13 January 2024.

[17] Jain J.,Sahu S., and Dixit A.," Brain Tumor Detection model based on CNN and Threshold Segmentation" International Journal of Experimental Research & Review, Vol. 32: 358-364 30 Aug 2023.

https://doi.org/10.52756/ijerr.2023.v32.031