International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET

Link Prediction in Evolving Networks Based on Information Propagation

T. Yuvana[1], B. Shriya[2], MD. Samreen[3], Dr. K. Kranthi Kumar[4]
[1,2,3]Students, IT Department Sreenidhi Institute of Science and Technology, Yamnampet, Hyderabad
[4]Associate Professor, IT Department Sreenidhi Institute of Science and Technology, Yamnampet, Hyderabad

Article Info
Volume 9, Issue 3

Page Number : 412-417

Publication Issue :
May-June-2022

Article History
Accepted : 01 June 2022
Published: 10 June 2022

ABSTRACT
In graph data mining, link prediction is a major problem. Link prediction is used in social networks to anticipate lost linkages in present networks as well as new ties in future networks. This method has a wide range of applications, including recommender systems, spam mail categorization, and domain expert identification in a variety of research fields. We present a novel model, Common Influence Set, to calculate node similarities in order to forecast future node similarity. The suggested link prediction method calculates a similarity score between two unconnected nodes using the shared influence set of the two nodes. We compared the performance of our method to that of earlier link prediction algorithms based on similarity across a variety of parameters using the area under the ROC curve (AUC).
Index Terms: Link Prediction, Common Influence, Similarity Index.

I. INTRODUCTION

Social networks are complex, with a huge number of nodes and linkages, and a continually changing network structure. Links between nodes may be broken or re-established over time. Developments in information are directly linked to these changes. A wide number of research and assessments of link prediction in complex networks reveal that network structure and information at various moments can aid in the prediction of link presence. Connection prediction refers to the knowledge obtained by examining network data to forecast how the link will change the next time it is used. Link prediction is a key component of social network analysis; it may be used to a variety of facets of the field, including friend recommendations in social networks and the prediction of possible biological relationships. In most cases, one of two approaches is used to forecast links: structural methods or feature methods. The analysis and summary of the network structure is done using structural approaches, which include node analysis, neighbour node analysis, path analysis between nodes, link analysis, and similarity analysis of relationships between neighbouring links. Consider the case of two person in a social network, u and v. If u and v don't know one other or have a lot of mutual acquaintances, it's probable that they'll be introduced. The structural technique is not the same as the feature method. In this situation, two researchers who have published articles on link prediction and community clustering, for example, will have a higher chance of cooperating. Because generic node attribute information is not readily available and the legitimacy of the data acquired cannot be verified, this study concentrates on network structure analysis.

II. Literature Survey

There are two main strategies for predicting links: structural and feature-based. The majority of structural-based link prediction algorithms estimate node similarity using network structure. In a social network, two people who have a lot of mutual friends are more likely to interact in the future. Lada and Adar [1] established a strategy for predicting associations between individuals based on shared neighbours. Murata and Moriyasu devised a method for estimating the similarity of the presence of a connection between two nodes in weighted networks using a directed action graph. Liu et al. developed a similarity score based on the previously described common neighbour approach and LBN (local naive Bayes), which outperforms common neighbours. Link prediction may also be done using pathways between nodes; Katz utilised the number of paths between two nodes and their length to get satisfactory results. Lü et al. suggested a local path index as a method for estimating the probability of the existence of a link between two nodes that was both effective and efficient. Liu and Lü presented a method for estimating the likelihood of a connection between two nodes using a local random walk. To quantify node similarity, Xu et al suggested a technique that uses route entropy as a similarity index. Shang et al. were the first to suggest a strategy for predicting future linkages based on past ties. By developing object identification systems based on photos and video, link prediction may be applied not just to typical social networks, but also to predict the associations between objects in films. A web-based recommendation system can also forecast a user's future relationship with an item. Users' dynamic interests and themes can be utilised to propose goods that will be of interest to the user in the future in a complicated network. Similarly, Li presented an impact-based collaborative filtering algorithm for proposing events of interest to users in event-based social networks (EBSNs).

III. Proposed System Architecture

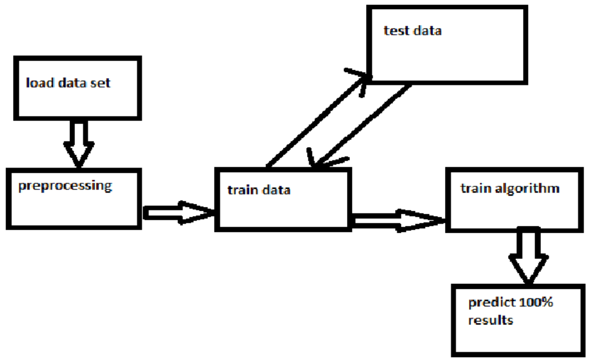

Fig-1: System Architecture

A. Materials and Methods

People join groups in various social networks because they share shared interests and activities. In previous studies, the likelihood of propagation between nodes in a network was seldom employed for link prediction. When calculating influence between nodes, the propagation probability can more accurately reflect the relationships between nodes. This metric can more correctly estimate the similarity between nodes than prior techniques. To accomplish link prediction, we must first identify a common group and assess the group's effect on two unconnected nodes. The estimated results are compared to two unconnected nodes to see how similar they are. We utilize an approximation model to quickly determine the impact of a node set on a single node since accurate calculation of the influence between nodes is time-consuming. We devised a series of algorithms to handle the challenge of swiftly calculating the most potential future top-k connections. In this research, we offer a similarity calculation technique based on influence propagation. The key principle is that the similarity of two disconnected nodes is determined by the impact of nodes in their shared influence set. To compute the similarity of two nodes, we propose a new similarity index and increase the performance of the similarity method.

The main contributions are as follows:

1. To calculate similarity, we propose a similarity index based on impact propagation. Finding the shared influence set of two disconnected nodes yields the similarity index.
2. We offer an approach for doing offline indexing of each node in a graph, and we utilize this off-line index to compute the upper limit of each two unconnected node pairs in order to calculate the similarity of two unconnected node pairs effectively.
3. We present a pruning technique that takes use of the upper bound of node similarity to increase efficiency while calculating the top-k similar nodes.

B. CIS Model

Firstly, In a weighted graph G (V , E , W), the influence set of node u can be denoted as

$$Infset(u) = \{v |\inf(u, v) > 0\} - - - (1)$$

In Equation (1), The influence value from node u to v in network G is denoted by inf (u, v), and is a threshold value for determining the size of node u's influence set. In a high-connectivity network, a node can be impacted by many nodes, but most of them contribute little to the final outcome, therefore we establish a threshold to decrease the amount of useless processing. So, in order to create a trade-off between time and accuracy, we set a threshold to remove nodes with little contributions.

Common Influence Set:

The common impact set of node u and node v in a weighted graph G(V,E,W) is a collection of nodes that may influence both node u and node v. It can be expressed as

$$CIS(u, v) = \{w |\inf(w, v) > 0 \; and \; CIS(u, v) > 0\}$$
$$Infset(u) \cap \mathrm{Infset(v)} - - - (2)$$

In Equation (2) The Common Influence Set is a collection of nodes drawn from node u's and node v's influence sets. We need to calculate the common influence set in the link prediction process because we need to calculate the similarities of each unconnected pair of nodes. Each pair of disconnected nodes must be estimated as part of the common influence set calculation. To quickly compute the common influence set of each pair of disconnected nodes, we first compute the influence set of each node and store it. We utilise the influence sets of each node when computing the common influence set.

$$\boldsymbol{simScore(u, v)} = \inf(CIS(u, v), u * \inf(u, v), v) — (3)$$

In Equation 3, The influence value from seed set S to node u is denoted by inf(S,u). The Common Influence Set of nodes u and v is denoted by CIS(u, v). The influence value from CIS(u,v) to node u and the influence value from CIS(u,v) to node v are both included in the formula. The similarity score is calculated by multiplying each part's product.

C. Algorithm

The working of the Naive Method for Link prediction using CIS Similarity Index is explained below:

Require: Graph G (V , W ,E), inf set , k,

Ensure: top-k missing edges scores S

Step 1: Let S=0.

Step 2: For each and every edge e(u,v) in E do

IS_u = infset(u)

IS_v = infset(v)

Step 3: Then C= $IS_u \cap IS_v$

Step 4: Continue the process for every edge in
E S(u,v) = inf (c, u)*inf (c, v)
Step 5: End for loop and Sort S in descending order
Step 6: Return S.top(k)

IV. Result Analysis

To run this application double click on 'run.bat' files to get below screen.

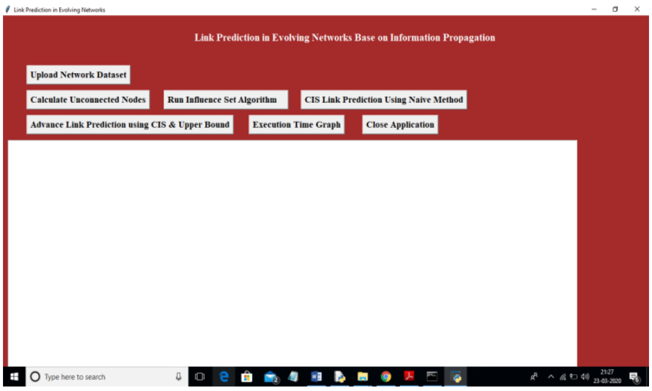

Fig-2: Initial Configuration of Application

In above screen click on 'Upload Network Dataset' button to upload dataset.

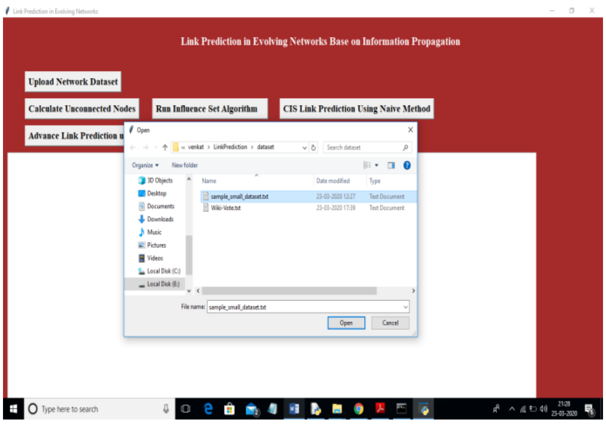

Fig-3: Upload Network Dataset

In the above screen, I am uploading a sample dataset file and after uploading the dataset will get below screen.

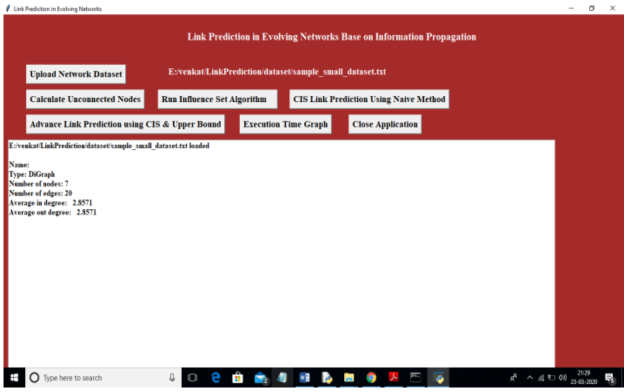

Fig-4: Showing description of dataset

In above screen, we can observe how many nodes, edges, and degrees of connectedness there are in the network. To locate all the nodes with missing links, click the 'Calculate Unconnected Nodes' button.

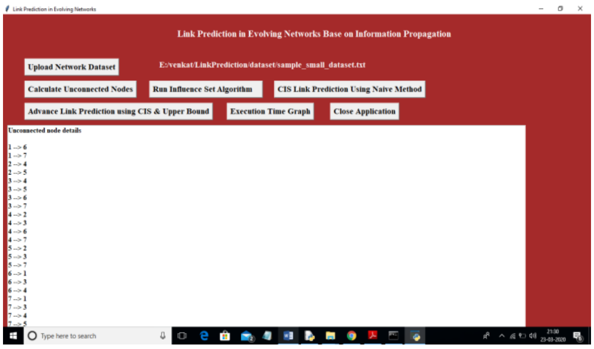

Fig-5: Calculate Unconnected Nodes

In the above screen, we received all the missing connections since 1 and 6 aren't connected, thus that link isn't there, and we can forecast the future link for 1 and 6. To anticipate, first select the 'Run Influence Set Algorithm' button to locate all influence nodes or high-connected nodes.

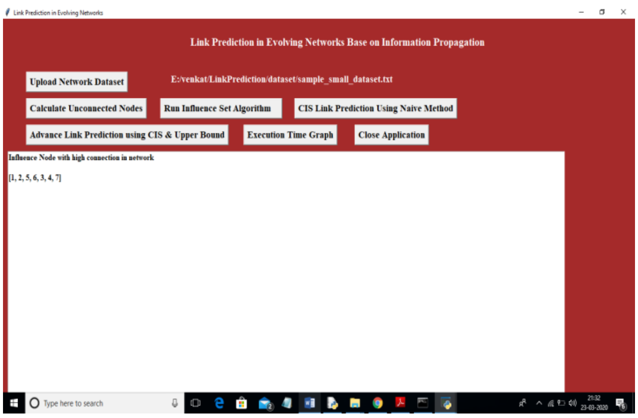

Fig-6: Nodes with high connectivity

In the above screen, we can observe that impact nodes 1, 2, 5, 6, 3, 4, and 7 have a strong link with one another. See the graph below to have a better understanding.

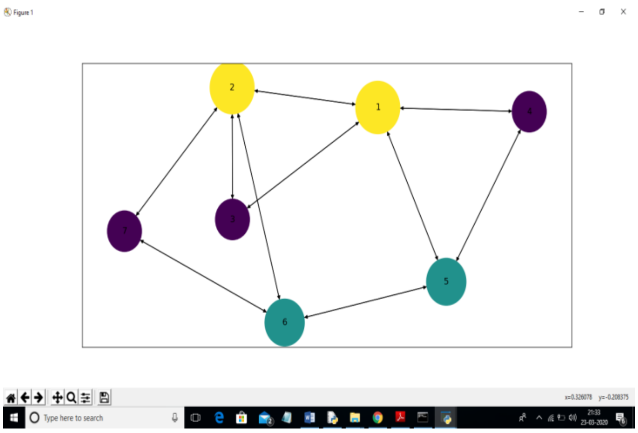

Fig-7: influence nodes graph

In the above graph, because all nodes are related to one another, they all become impact nodes, which may not be the case for a huge dataset. To forecast missing link nodes, click the 'CIS Link Prediction Using Naive Method' button.

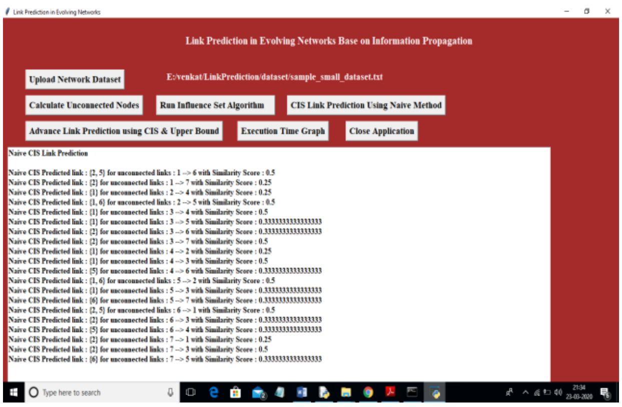

Fig-8: Similarity score between common nodes set

In the above screen, we observe that the anticipated node linkages for disconnected node 1 -> 6 are 2 and 5, implying that 6 will be able to connect with 1 in the future utilizing link nodes 2 or 5. I'm presenting the similarity score between the common node-set that was discovered. For every missing or disconnected nodes, I'm also providing anticipated linkages. To acquire projected link output, click the 'Advance Link Prediction utilising CIS & Upper Bound' button. We achieve the same result here as well; the only variation is the algorithm and execution time.

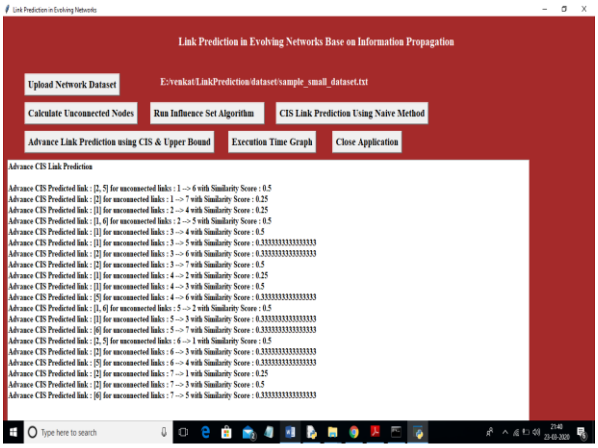

Fig-9: Advanced Link Prediction using CIS & Upper Bound algorithm results

In the above screen also, we expected missing connections in the range of 25 to 10. To get the graph below, click the 'Execution Time Graph' button.

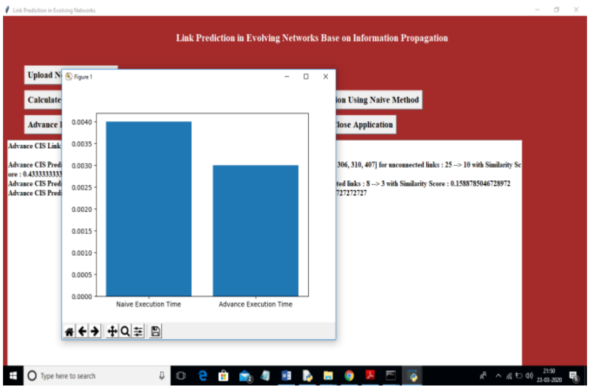

Fig-10: Time Graph

In the above graph the algorithm name is on the x-axis, while the execution time is on the y-axis. We may deduce from the graph above that the advanced approach takes less time to execute than the old Nave technique. As a result, the advanced method outperforms the traditional Nave algorithm.

V. Conclusion

We present a new similarity index for link prediction in this research. Experiments have shown that our similarity index outperforms other popular similarity indexes. We presented an enhanced approach to compute the similarity score efficiently due to the time cost of CIS's similarity score computation. In the future, we'll concentrate on solving the challenge in dynamic graphs, whose structures alter over time.

VI. REFERENCES

[1]. L. A. Adamic and E. Adar, "Friends and neighbors on the Web," Soc. Netw., vol. 25, no. 3, pp. 211–230, 2003.

[2]. L. Backstrom and J. Leskovec, "Supervised random walks: Predicting and recommending links in social networks," in Proc. ACM Int. Conf. Web Search Data Mining, 2010, pp. 635–644.

[3]. L. Berton, J. Valverde-Rebaza, and A. de Andrade Lopes,"Link prediction in graph construction for supervised and semi-supervised learning," in Proc. Int. Joint Conf. Neural Netw., 2015, pp. 1–8

[4]. B. Du and L. Zhang, "A discriminative metric learning based anomaly detection method," IEEE Trans. Geosci. Remote Sens., vol. 52, no. 11, pp. 6844–6857, Nov. 2014.

[5]. B. Du and L. Zhang, "Target detection based on a dynamic subspace," Pattern Recognit., vol. 47, no. 1, pp. 344–358, 2014.

[6]. B. Du, S. Cai, C. Wu, L. Zhang, and D. Tao, "Object tracking in satellite videos based on a multi-frame optical flow tracker," CoRR, vol. abs/1804.09323, 2018. [Online]. Available: https://arxiv.org/ abs/1804.09323

[7]. B. Du, Y. Zhang, L. Zhang, and D. Tao, "Beyond the sparsity-based target detector: A hybrid sparsity and statistics-based detector for hyperspectral images," IEEE Trans. Image Process., vol. 25, no. 11, pp. 5345–5357, Nov. 2016.

[8]. L.Gao, J. Wu, C. Zhou, and Y. Hu, "Collaborative dynamic sparse topic regression with user profile evolution for item recommendation," in Proc. 21st Conf. Artif. Intell. (AAAI), 2017, pp. 1–7.

[9]. M. Jiang, Y. Chen, and L. Chen, "Link prediction in networks with nodes attributes by similarity propagation," CoRR, vol. abs/1502.04380, Feb. 2015. [Online]. Available: https://arxiv.org/abs/1502.04380

[10]. L. Katz, "A new status index derived from sociometric analysis," Psychometrika, vol. 18, no. 1, pp. 39–43, 1953.

Cite this Article

T. Yuvana, B. Shriya, MD. Samreen, Dr. K. Kranthi Kumar, "Link Prediction in Evolving Networks Based on Information Propagation ", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 9 Issue 3, pp. 412-417, May-June 2022.

Journal URL : https://ijsrset.com/IJSRSET2293167