International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

RELATIONSHIP PREDICTION IN GROWING NETWORKS BASED ON INFORMATION BROADCAST

Ch. Vaishnavi , Ch. Achyuth Reddy, N. Akhilesh, And P. Varun Reddy[1] , Dr. Deena Babu Mandru[2]

[1]UG Student, Department of Information Technology, Malla Reddy Engineering College, Hyderabad , India.
[2]Professor, Department of Information Technology, Malla Reddy Engineering College, Hyderabad , India.

ABSTRACT

Link prediction is asignificantconcern in graph data mining. In social networks, link predictionis used to predict lostassociations in current networks and new links in future networks. This process has a wide range of applications including recommender systems, spam mail classification, and the identification of domain experts in various research areas. In order to predict future node similarity, we propose a new model, Common Influence Set, to calculate node similarities. The proposed link prediction algorithm uses the common influence set of two unconnected nodes to calculate a similarity score between the two nodes. We used the area under the ROC curve (AUC) to evaluate the performance of our algorithm and that of previous link prediction algorithms based on similarity over a range of problems. Our experimental results show that our algorithm outperforms previous algorithms.

Index Terms: *Link prediction, common influence, similarity index.*

1. INTRODUCTION

Social networks are complex, and usually have a large number of nodes and links, and the network structure is constantly changing. With the passage of time, links between nodes may disappear or be re-established. These changes are closely related to changes in information. A large number of studies and analyses of link prediction in complex networks show that network structure and information at different times can help predict the existence of links. The information gained by analyzing network information at it changes the next time the link is called link prediction. Link prediction is an important element in social network analysis, it can be applied to many aspects of social network analysis, such as friend recommendations in social networks, prediction of potential links in biological protein networks, or the prediction the potential relationships in collaborative networks. Link prediction generally involves one of two methods: structural methods and feature methods. Structural methods involve the analysis and summarization of the network structure, including the analysis of nodes, neighbor node analysis, analysis of paths between nodes, link analysis and similarity analysis of relationships between adjacent links. For example, consider two people u and v in a social network. If u and v do not know each other, or have a lot of friends in common, it is likely that u and v will be introduced to each other. The feature method differs from the structural method. In this case, two scholars who have, for example, published papers relating to link prediction and community clustering, will have a greater probability of cooperating. This study focuses on the analysis of network structure, because general node attribute information is not readily available, and the authenticity of the data obtained cannot be guaranteed.

2. LITERATURE SURVEY

Links prediction involves two primary methods: namely, structural and feature-based. Most of the structural-based link prediction methods use network structure to measure node similarities. For example, in a social network, two individuals with many common friends are more likely connect in future. Lada and Adar [1] proposed a method based on common neighbors to predict relationships between individuals. Murata and Moriyasu proposed a link prediction method which constructed a directed action graph to estimate the similarity of the existence of a link between two nodes in weighted networks. Liu et al. proposed a similarity score based on a common neighbor method mentioned before and LBN(local naive Bayes) which performs better than common neighbors. Paths between nodes may also be used for link prediction, Katz [10] used the number of paths between two nodes and their length, producing reasonable results. Lü et al. proposed approach which had high effectiveness and efficiency, a local path index, to estimate the probability of the existence of a link between two nodes. Liu and Lü [13] proposed a method that use a local random walk to estimate the probability of the existence of a link between two nodes. Xu et al. [25] proposed a method that use path entropy as similarity index to measure nodes' similarity. Shang et al. first proposed a method for using past links to predict the future links. Link prediction can be applied not only to traditional social networks, but also predict the relationships of objects in videos by the development of object recognition techniques based on images and video. A web-based recommendation system can also predict the user's connection to an item in the future. In a complex network, users' dynamic interests and topics can be used to recommend products that will be of interest to the user in the future. Similarly, in event-based social networks (EBSNs), Li proposed an impact-based collaborative filtering algorithm for recommending events of interest to users.

3. PROPOSED SYSTEM ARCHITECTURE

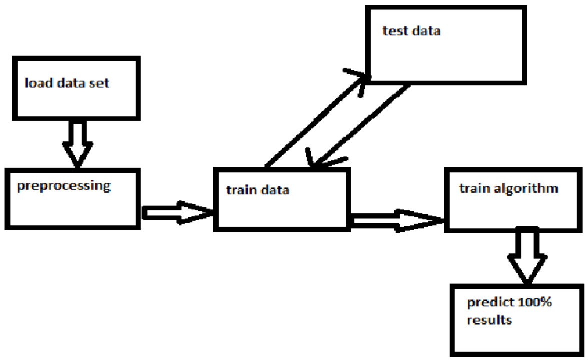

Fig-1: System Architecture

A) Materials and Methods

In many social networks, people connect because of their common interests and hobbies, forming a group. In past studies, researchers have rarely used the probability of propagation between nodes in a network for link prediction. Using the propagation probability to calculate influence between nodes can more reliably reflect the relationships between nodes. Compared to previous algorithms, this measure can more accurately measure the similarity between nodes. In order to achieve link prediction, we first need to find a common group, and calculate the influence of the group on two unconnected nodes. The calculated results are taken as the similarity of two unconnected nodes. Since the precise calculation of the influence between nodes is time-consuming, we use an approximate model to quickly calculate the influence of a node set on a single node. In order to be able to quickly calculate the most possible future top-k links, we developed a set of algorithms to solve this problem. In this paper, we propose an algorithm based on the propagation of influence for calculating similarity. The main idea is that the similarity of two unconnected nodes is the product of the influence of nodes in the common influence set of these two nodes. We proposea new similarity index to compute the similarity of two nodes, and improve the efficiency of the similarity algorithm.

The main contributions are as follows:

1. We propose a similarity index based on influence propagation to calculate similarity. The similarity index is calculated by finding the common influence set of two unconnected nodes.

2. In order to calculate the similarity of two unconnected node pairs efficiently we propose an algorithm to perform offline indexing of each node in a graph, and we use this off-line index to calculate the upper bound of each two unconnected node pairs.

3. In order to be able to calculate the top-k similar nodes efficiently, we propose a pruning algorithm which makes use of the upper bound of nodes similarity to improve efficiency.

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), inf (u , v) denotes the influence value from node u to v in network G , and θ is a threshold value to determine the size of node u's influence set. We set a threshold is to reduce the amount of unnecessary computation, in a high connectivity network, a node can be influenced by many nodes, but most of them contribute little to the final result. So we set a threshold to reduce nodes which have tiny contributions, in order to make a trade-off between time and accuracy.

Common Influence Set:

In a weighted graph G(V,E,W),the common influence set of node u and node v is a set of nodes which can influence both node u and node v. It can be denoted 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 Setis a set of nodes which from both node u's influence set and node v's influence set. In the link prediction process, we need to calculate the similarities of each unconnected pair of nodes, and therefore need to calculate the common influence set. In the process of calculating the common influence set, each pair of unconnected node needs to be calculated. In order to efficiently compute the common influence set of each two unconnected node pairs, we calculate the influence set of each node at the beginning, and store the influence set of each node. In the process of calculating the common influence set, we use the influence set of each node.

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

In Equation 3, inf(S,u) denotes the influence value from seed set S to node u. CIS(u, v) denotes the Common Influence Set of node u and node v. The formula consists of two parts: the influence value from CIS(u,v) to node u and the influence value from CIS(u, v) to node v. We use the product of each part as the similarity score.

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)

4. RESULT ANALYSIS

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

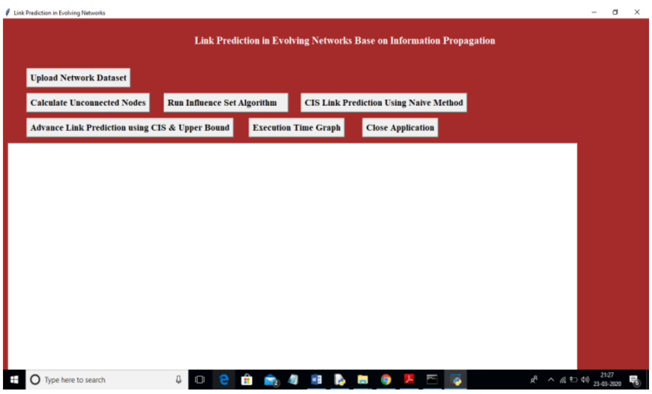

Fig-2: Initial Configuration of Application

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

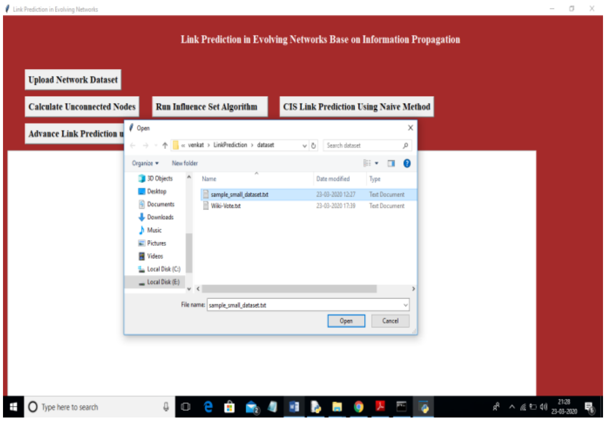

Fig-3: Upload Network Dataset

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

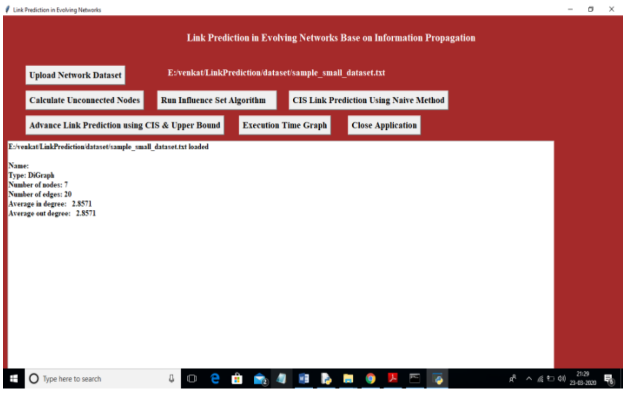

Fig-4: Showing description of dataset

In above screen we can see network contains how many nodes, edges and it's in and out degree connectivity. Now click on 'Calculate Unconnected Nodes' button to find all those nodes which has missing links.

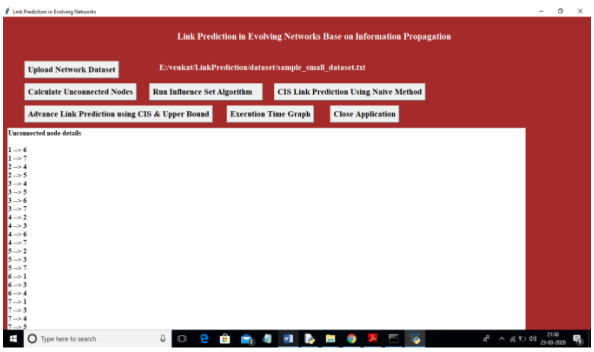

Fig-5: Calculate Unconnected Nodes

In above screen we got all missing links as 1 and 6 has no connectivity so that link is missing and we can predict future link for 1 and 6. To predict first click on 'Run Influence Set Algorithm' button to find all influence nodes or nodes which has high connectivity.

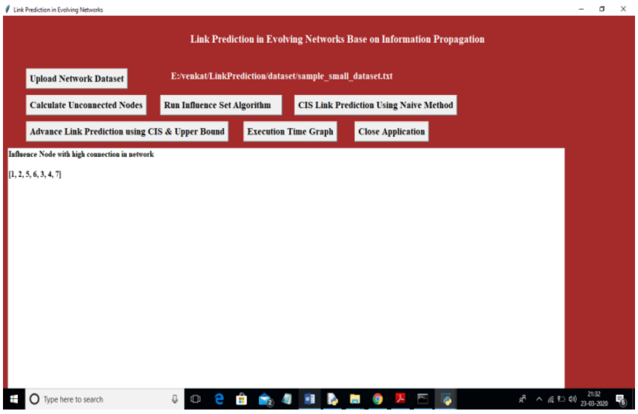

Fig-6: Nodes with high connectivity

In above screen we can see 1, 2, 5, 6, 3 4 and 7 are the influence nodes which has high connection with one and other. To understand see below graph,

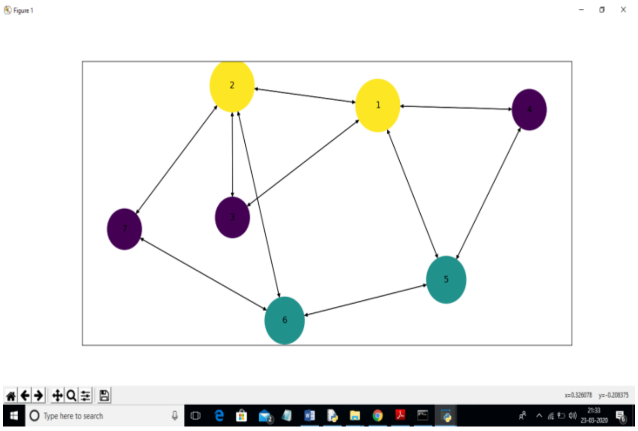

Fig-7: influence nodes graph

In above graph all nodes are connected with one and other so all nodes become influence node and this may not happen for large dataset. Now click on ‘CIS Link Prediction Using Naive Method’ button to predict link for missing link nodes.

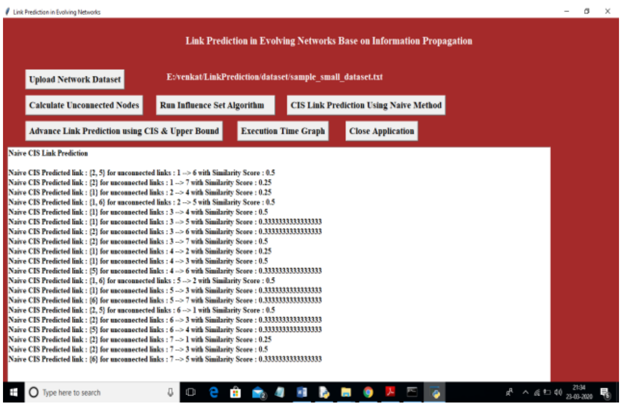

Fig-8: Similarity score between common nodes set

In above screen for unconnected node 1 -> 6 we can see predicted node links are 2 and 5 which means in future 6 can connect with 1 by using link nodes called 2 or 5. I am displaying similarity score between common nodes set found. Similarly, I am displaying predicted links for all missing or unconnected nodes. Now click on ‘Advance Link Prediction using CIS & Upper Bound’ button to get predicted link output. Here also we get same output the only difference is algorithm and execution time.

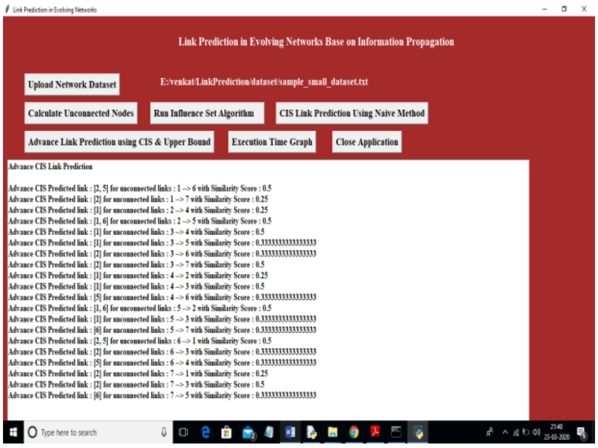

Fig-9: Advanced Link Prediction using CIS & Upper Bound algorithm results

In above screen also and 25 -> 10 we predicted missing links. Now click on 'Execution Time Graph' button to get below graph.

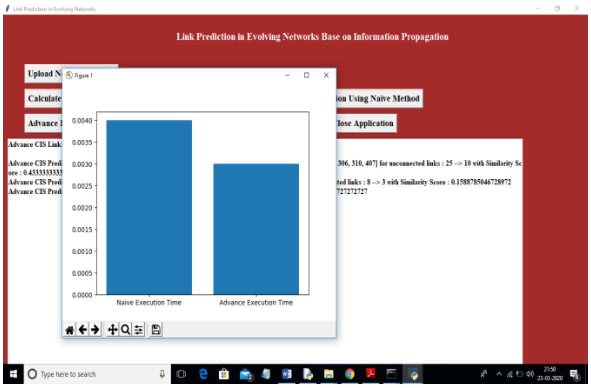

Fig-10: Time Graph

In above graph x-axis represents algorithm name and y-axis represents execution time. From above graph we can conclude that advance algorithm taking less execution time compare to old Naïve algorithm. So advance algorithm is better than old Naïve algorithm.

5. CONCLUSION

In this paper, we propose a new similarity index for link prediction. Experiments showed that our similarity index performs better than other main stream similarity indices. Due to the time cost of the similarity score calculation of CIS, we proposed an advanced method to calculate the similarity score efficiently. For future work, we will solve the problem in the dynamic graphs whose structures will change along over time.

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.