DYNAMIC NETWORK LINK PREDICTION THROUGH INFORMATION PROPAGATION

B.V.S.P Pavan Kumar[1], M. Sai Thanmai[2], M. Nandini[2], M. Ashwini[2], K. Soujanya[2]

[1]Associate Professor, [2]UG Scholar, [1,2]Department of Computer Science and Engineering

[1,2]Malla Reddy Engineering College for Women (A), Maisammaguda, Medchal, Telangana.

ksoujanya976@gmail.com, medinandini062@gmail.com, mattathanmai@gmail.com, ashwinimoota@gmail.com

ABSTRACT

Link prediction is an important issue in graph data mining. In social networks, link prediction is used to predict missing links 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.

Keywords: Link prediction, Evolving networks, Information propagation.

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. 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.

2. PROPOSED SYSTEM

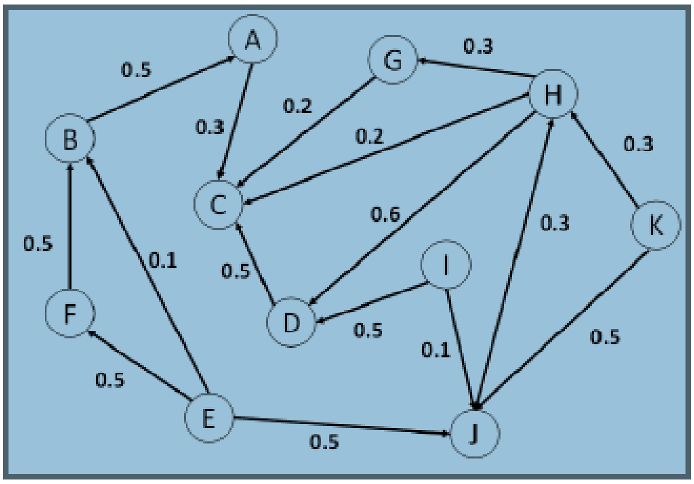
Links prediction involves two primary methods: namely, structural and feature-based. Most of the structural-based link prediction methods use network structure to measure node similarities. For example, in a social network, two individuals with many common friends are more likely connect in future. Lada and Adar 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 used the number of paths between two nodes and their length, producing reasonable results. Liu 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 Lu proposed a method that use a local random walk to estimate the probability of the existence of a link between two nodes. Wang et al proposed a method that uses a clustering based collaborative ltering approach, including both topological and node attributes. Xu et al proposed a method that use path entropy as similarity index to measure nodes similarity. Shang et al rst proposed a method for using past links to predict the future links. In Shang et al. found that if a pair of nodes are connected, they are more likely to connect to the common nodes in the future networks, and they rst use the past links and future links for link prediction. In Shang et al. proposed the metric Precision for the evolving networks. Lee and Tukhvatov proposed a topology-based similarity measure to predict future friends.

3. SYSTEM TEST SCREENSHOTS

Link Prediction in Evolving Networks Based on Information Propagation

In this paper author is describing concept to predict missing link in a network and new links for future network. Link prediction can be useful in product recommendation or friend recommendation in social network where a user is not connected or not have link to other user in network but by identifying influence and common influence nodes with high similarity we can predict link for missing link user and we can suggest that predicted link as future predicted link that can be connected with missing links. Influence nodes are those nodes which has high connectivity with other nodes or have high influence for information passing. For example if one influence user connected to so many users then the message will be propagate or pass to all users who has connectivity with influence user.

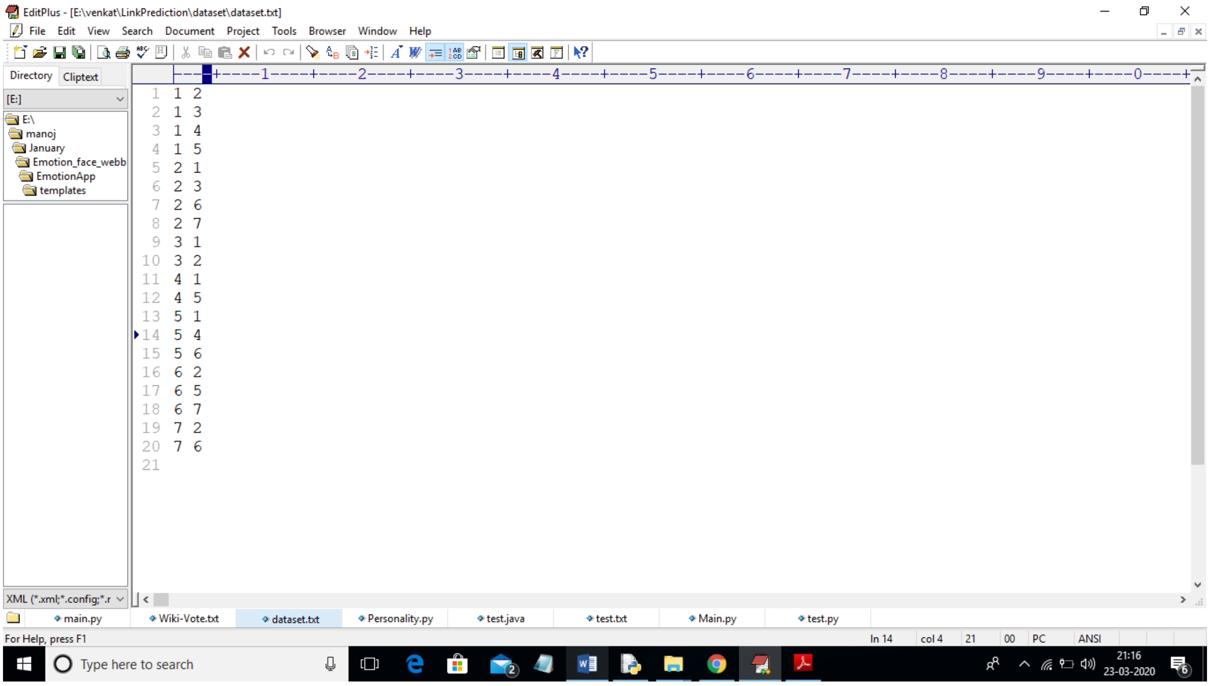
Common Influence Set (CIS) are those multiple nodes which has high connectivity with other nodes. By finding common nodes between two CIS and by calculating common nodes similarity we can predict that those common nodes with high similarity can be connected with missing links.



In above paper example graph we can see ‘C’ is an influence node as it has connectivity with many nodes. In paper TABLE1 we can see all nodes connection with C based on similarity and connectivity. In above graph all circle represents Vertex and line between two vertex indicate edges or connectivity and numeric value represents weight or distance between two vertex or nodes.

In above graph C is connected with A, B, K, D, E and I connected with K, E, F, D and from two C and I influence set we can see common influence set as K, D and E. Based on common nodes similarity score will be calculated and if score > 0.1 then those common nodes can be predicted for missing links. All influence nodes can be calculated by using dijkstra distance function and by applying algorithm 1 from paper. To calculate common influence nodes and its similarity author is using Naïve Algorithm2 from paper. In algorithm 2 to find common nodes its using INTERSECTION function and SORTING function which will consume lots of execution time. To reduce that execution time author is using Algorithm 3 to calculate UPPERBOUND common nodes value by using single FOR LOOP and IF condition to check common nodes availability and not using SORTING or INTERSECTION function to reduce execution time. Upper Bound values pass to Advance Common Influence Link Prediction (Algorithm 4) to find similarity score and if similarity score > 0.1 then that Upper Bound link will be predicted for missing link.

To implement this project author is using WIKI VOTE dataset which consists of VERTEX and EDGES (VERTEX refers to election person and EDGES refers to person who vote that election person). To understand missing link see simple dataset example below

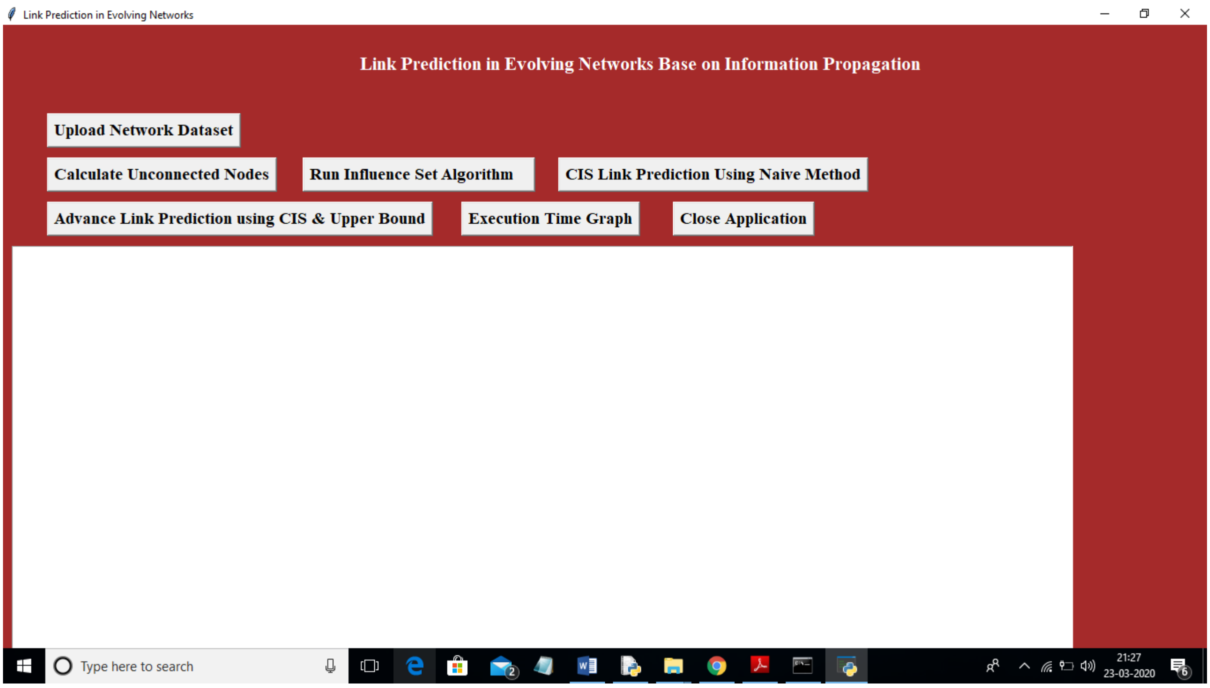


In above dataset two values are there with space separated where first value is the vertex and second value is the edge. In above screen 1 is connected to 2, 3, 4, 5 and 1 is having missing link with 7 and 6 and by finding influence nodes and common influence nodes we can predict node which can allow 7 to connect with 1 in future. By running application we can see that output.

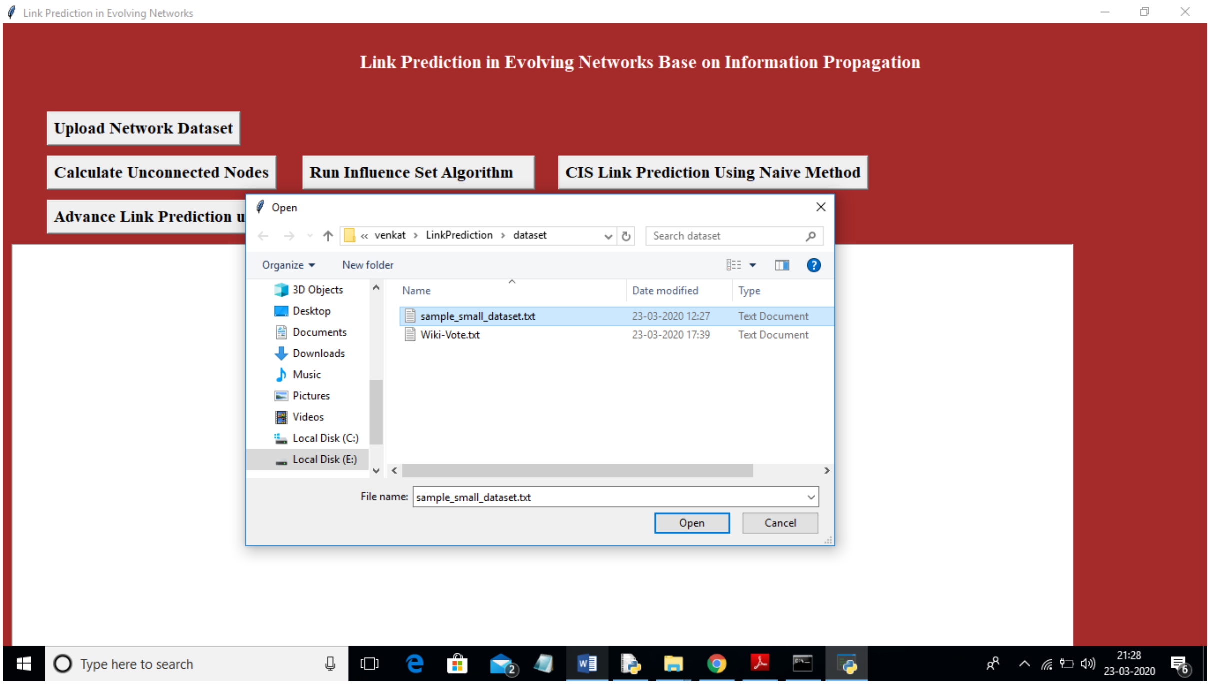
Here wiki vote dataset contains 500 users details and u may not clearly see the output so I am giving sample small dataset with 20 users also. Execution time comparison will not suitable for small sample dataset so u used execution time comparison graph for wiki vote large dataset. Execution time comparison graph is calculated between NAÏVE Algorithm 2 and Advance Algorithm 4.

Screen shots

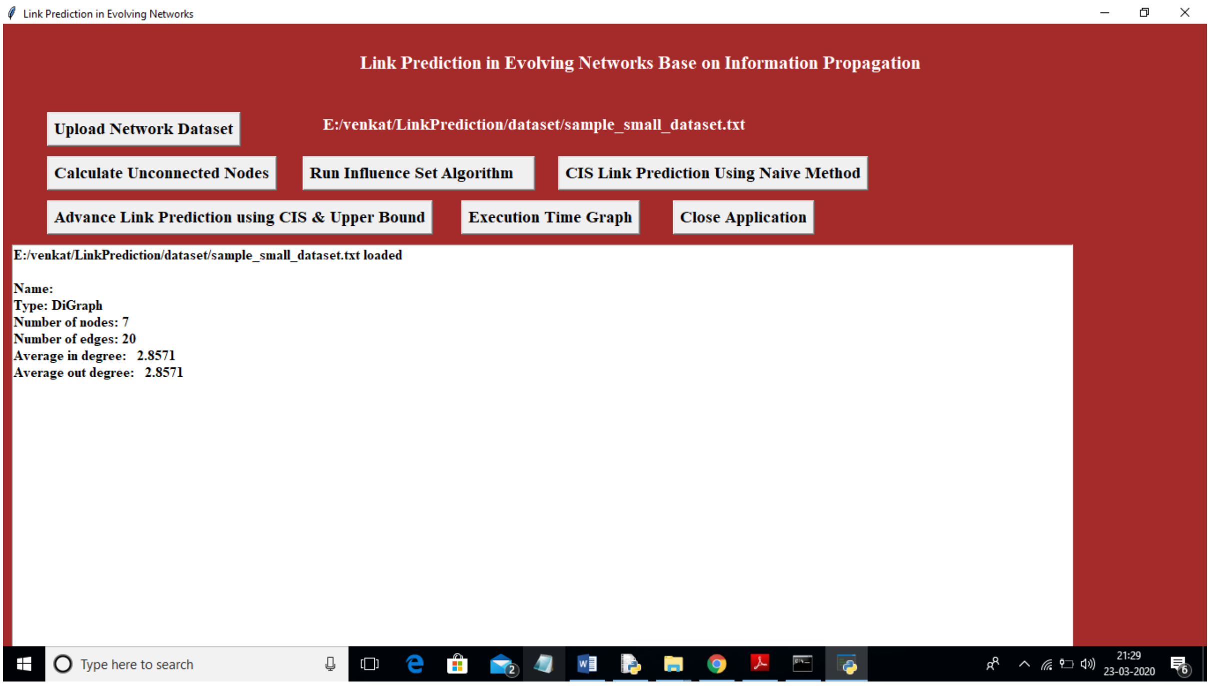
To run this project double click on ‘run.bat’ file to get below screen



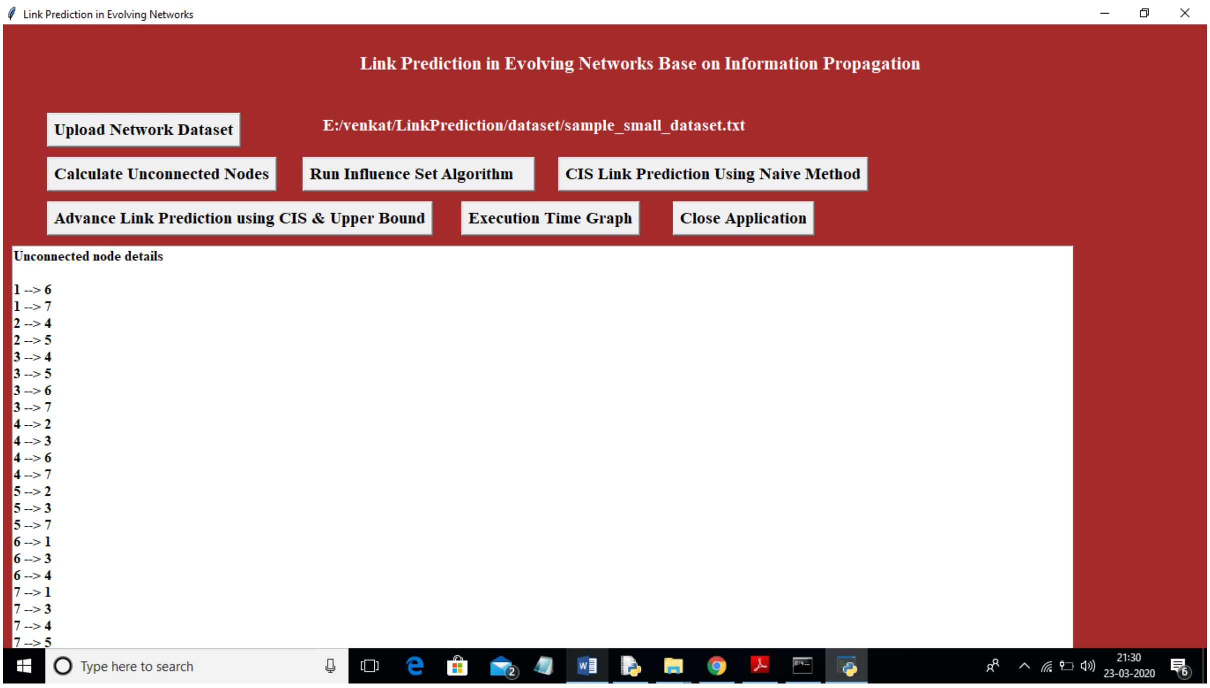
In above screen click on 'Upload Network Dataset' button to upload dataset



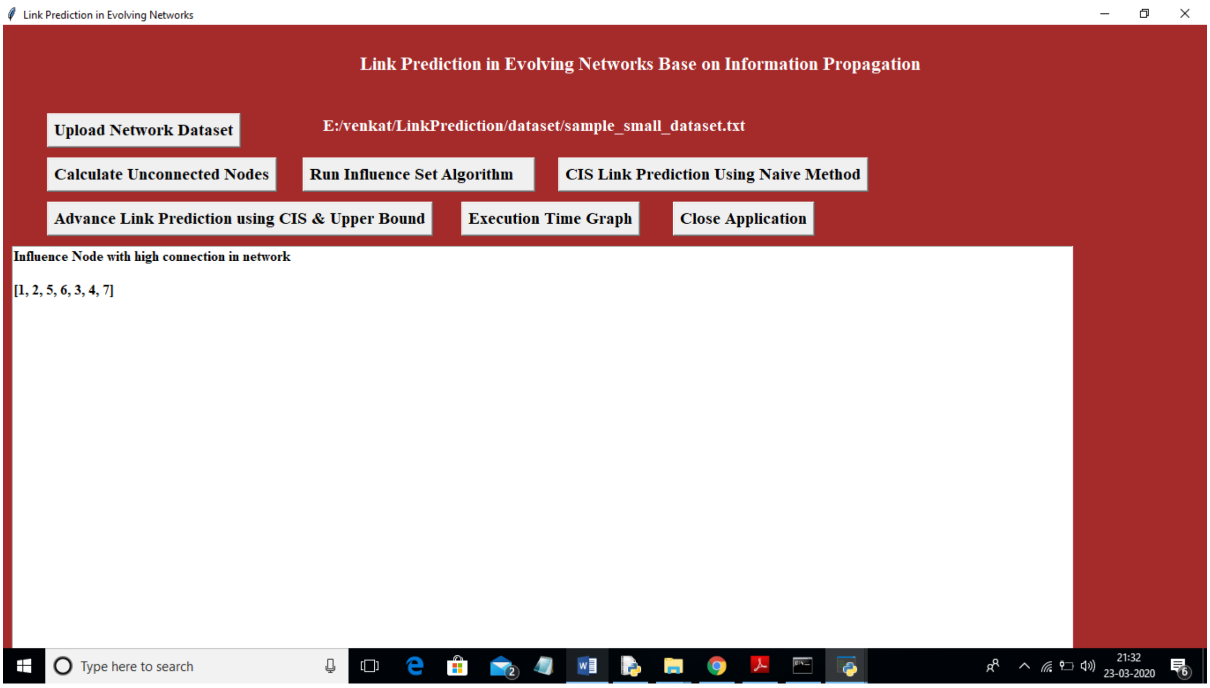
In above screen I am uploading sample_small_dataset.txt' file and after uploading dataset will get below screen



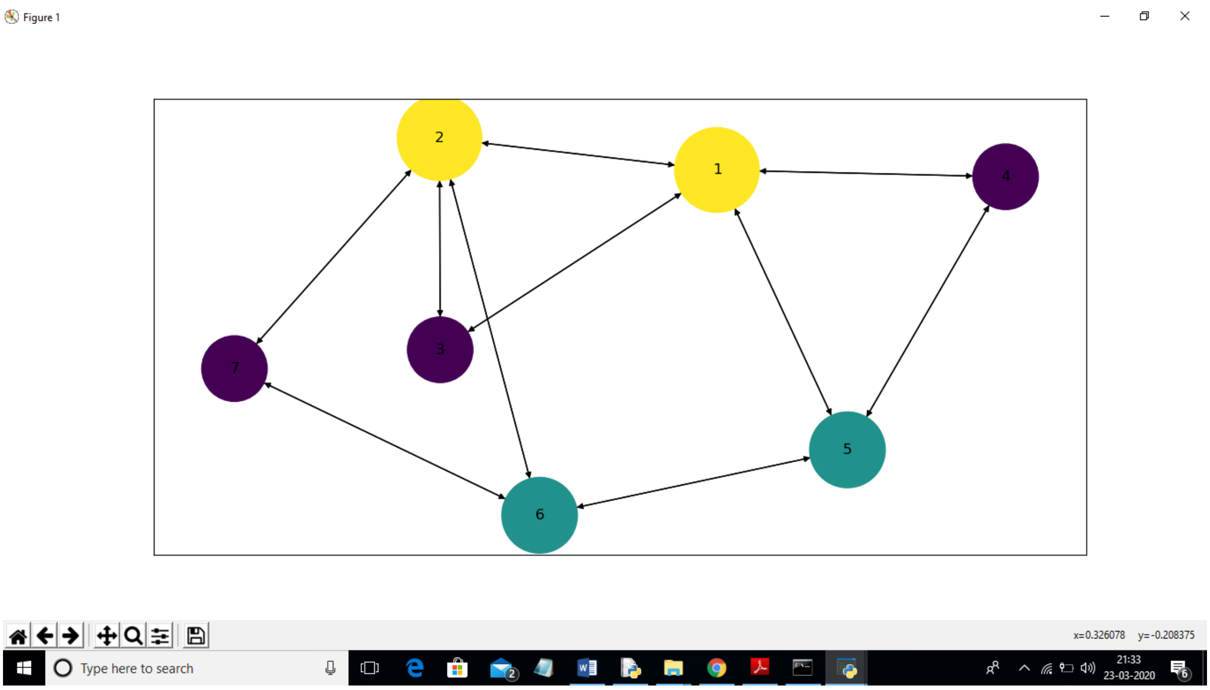
In above screen we can see network contains how many nodes, edges and its in and out degree connectivity. Now click on 'Calculate Unconnected Nodes' button to find all those nodes which has missing links.



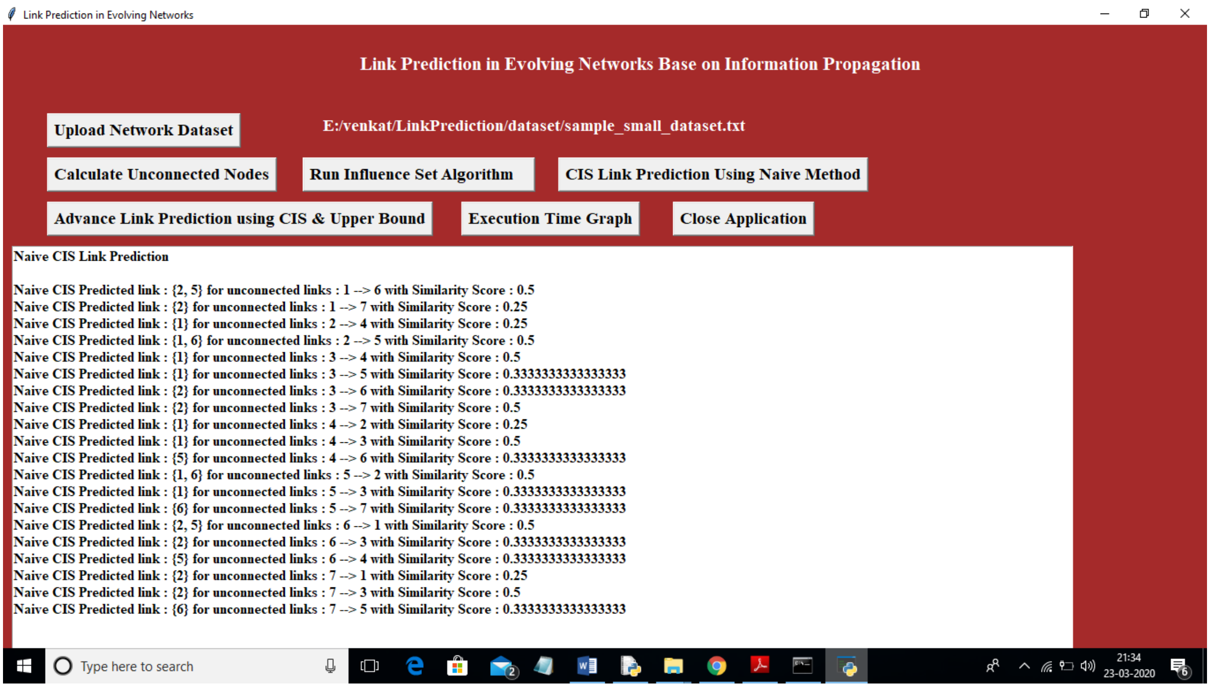
In above screen we got all missing links as 1 and 6 has no connectivity so that link is missing and we can predict future link for 1 and 6. To predict first click on ‘Run Influence Set Algorithm’ button to find all influence nodes or nodes which has high connectivity.



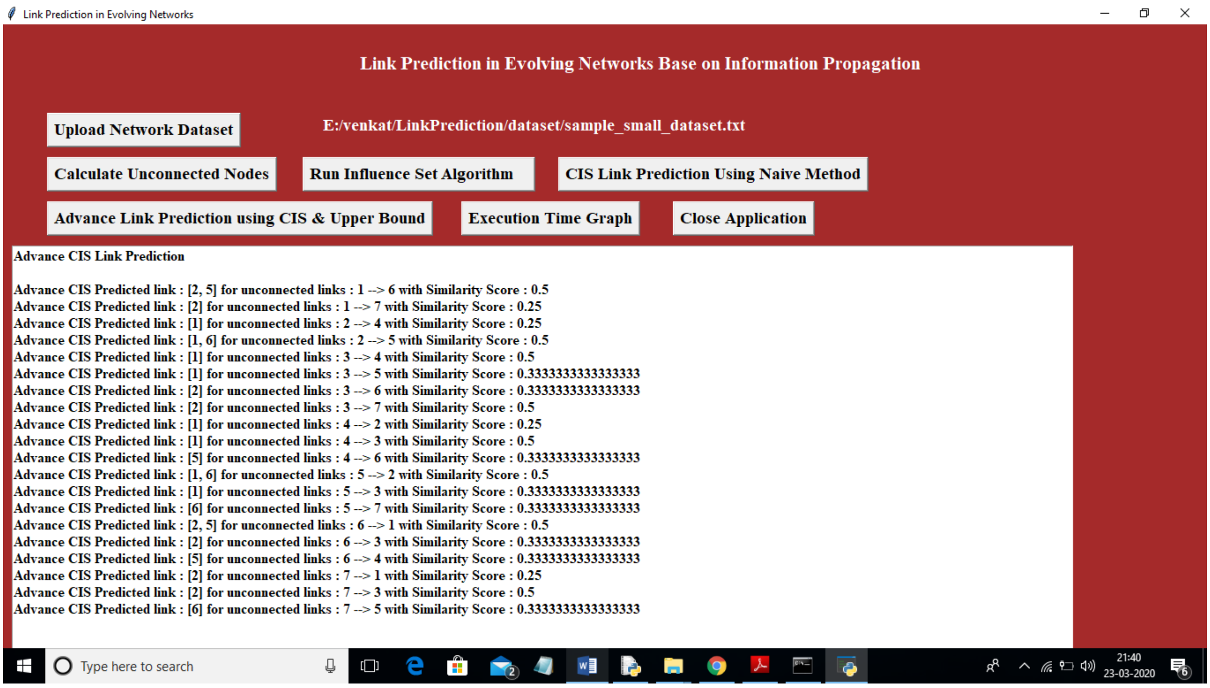
In above screen we can see 1, 2, 5, 6, 3 4 and 7 are the influence nodes which has high connection with one and other. To understand see below graph



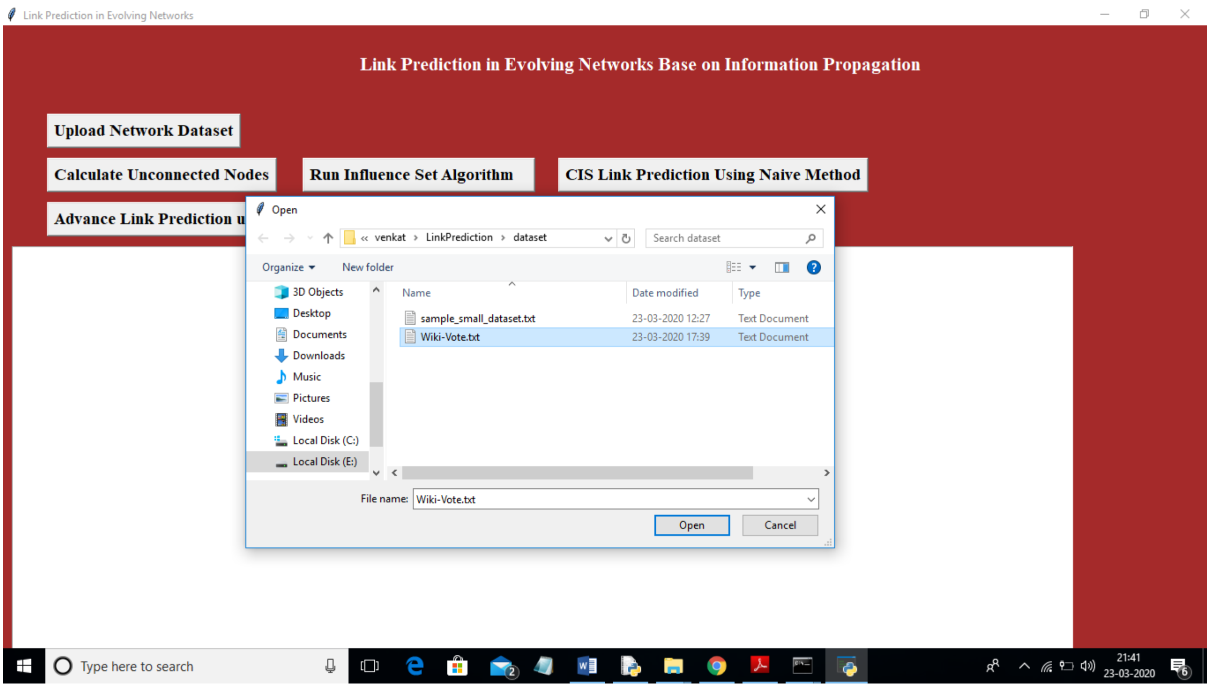
In above graph all nodes are connected with one and other so all nodes become influence node and this may not happen for large dataset. Now click on 'CIS Link Prediction Using Naive Method' button to predict link for missing link nodes.



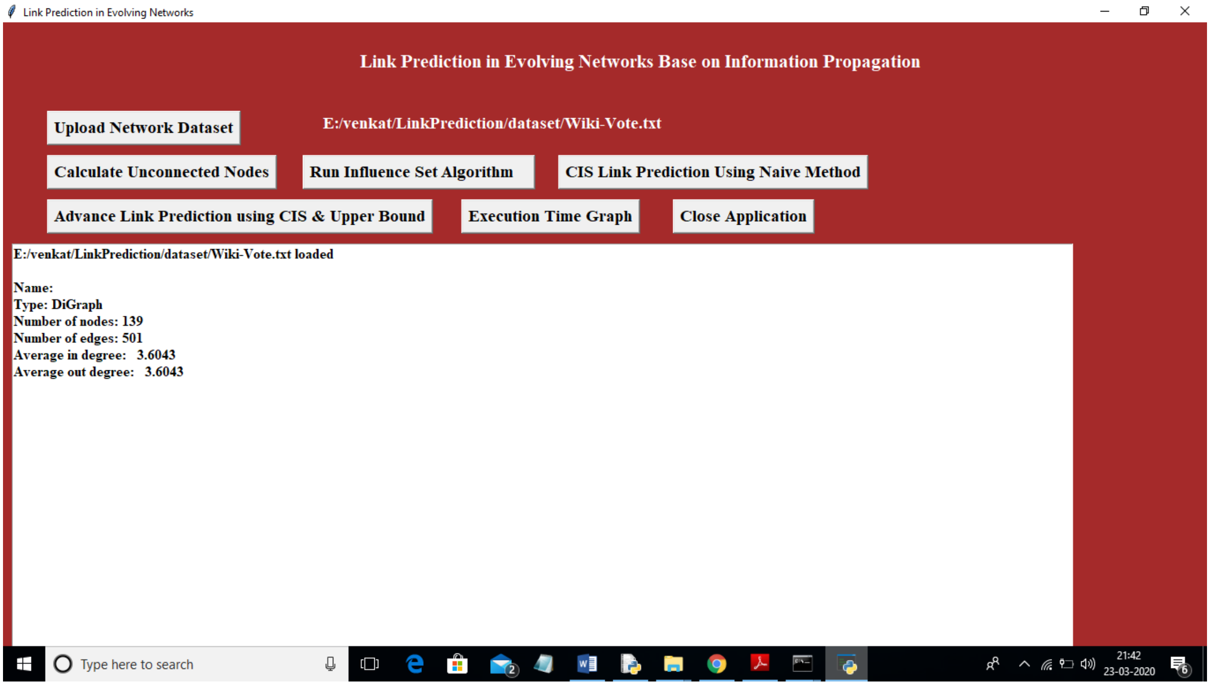
In above screen for unconnected node 1 -> 6 we can see predicted node links are 2 and 5 which means in future 6 can connect with 1 by using link nodes called 2 or 5. I am displaying similarity score between common nodes set found.Similarly I am displaying predicted links for all missing or unconnected 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. Below screen with Advance Algorithm



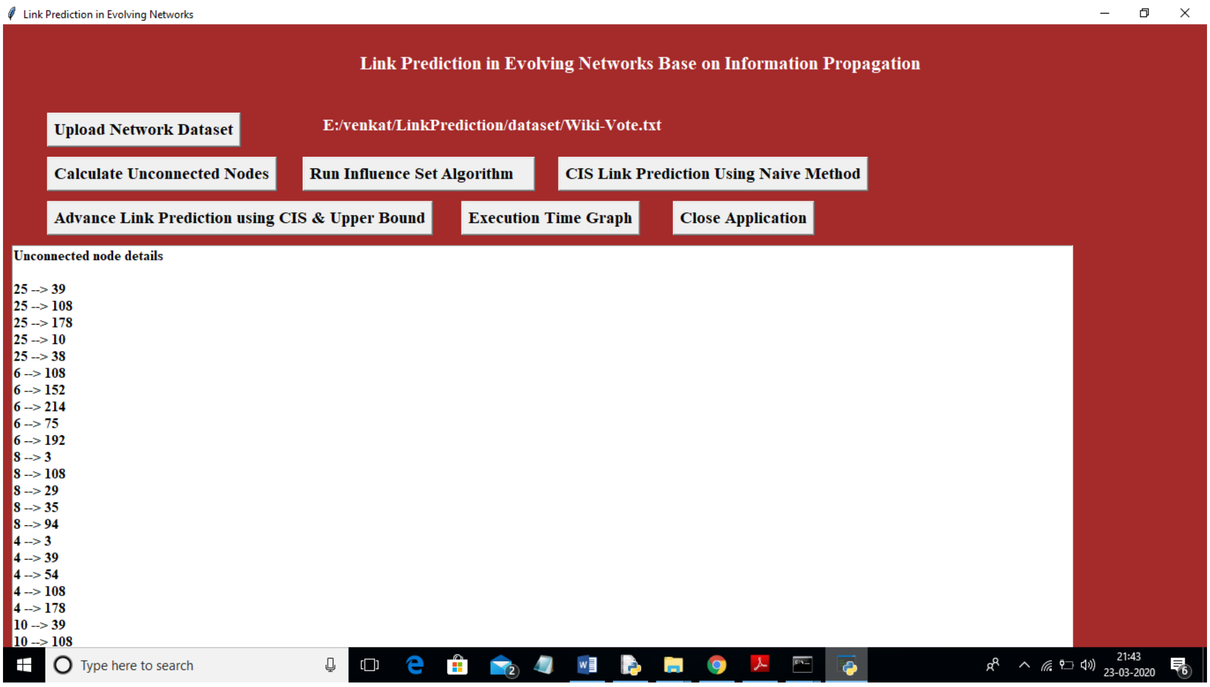
In above screen we saw output with sample small dataset and now we will upload wiki vote dataset and see output.



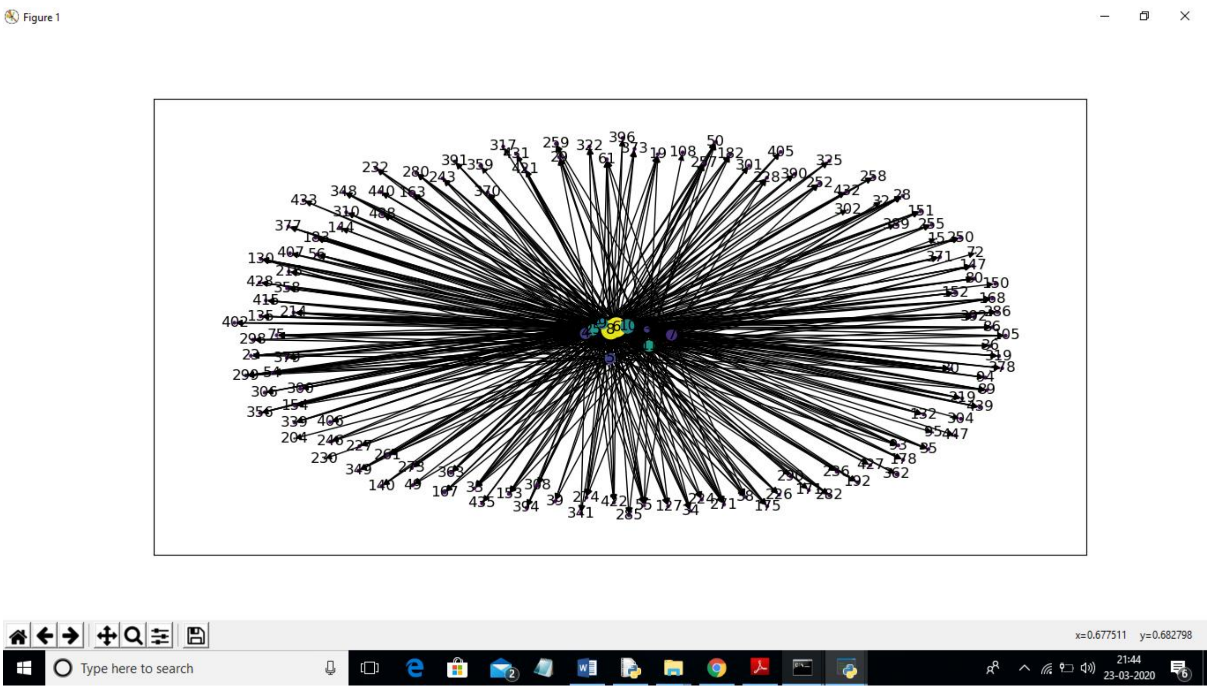
In above screen I am uploading 'Wiki-Vote.txt' file and after upload will get below screen



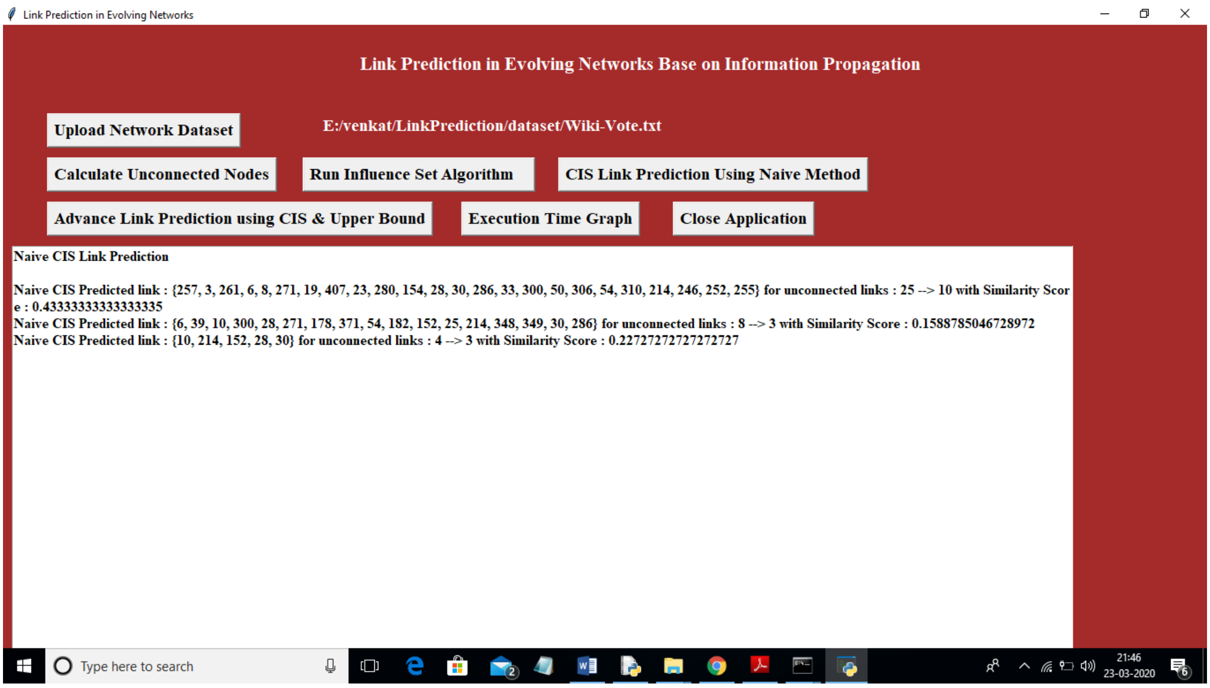
In above screen we can see network details with total nodes and edges. Now click on ‘Calculate Unconnected Nodes’ button to get missing links for wiki dataset



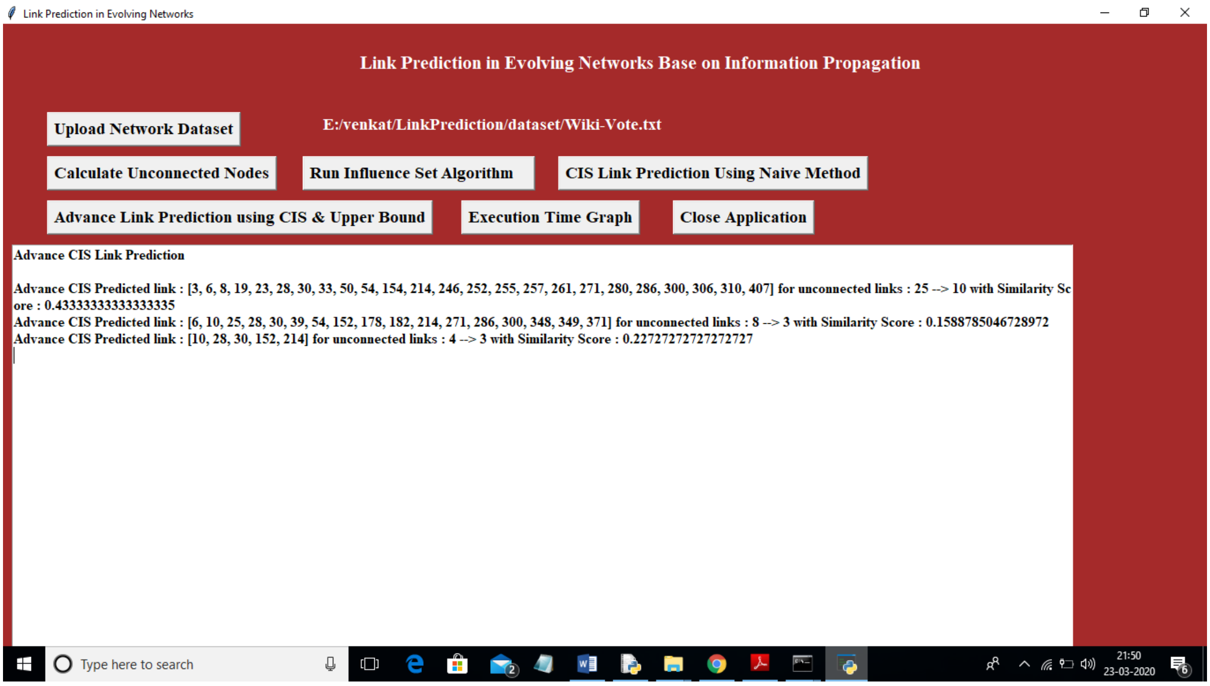
In above screen we can see missing links for above nodes and scroll down text area to view all missing link nodes. Now click on ‘Run Influence Set Algorithm’ button to find all influence node using dijkstra distance function.



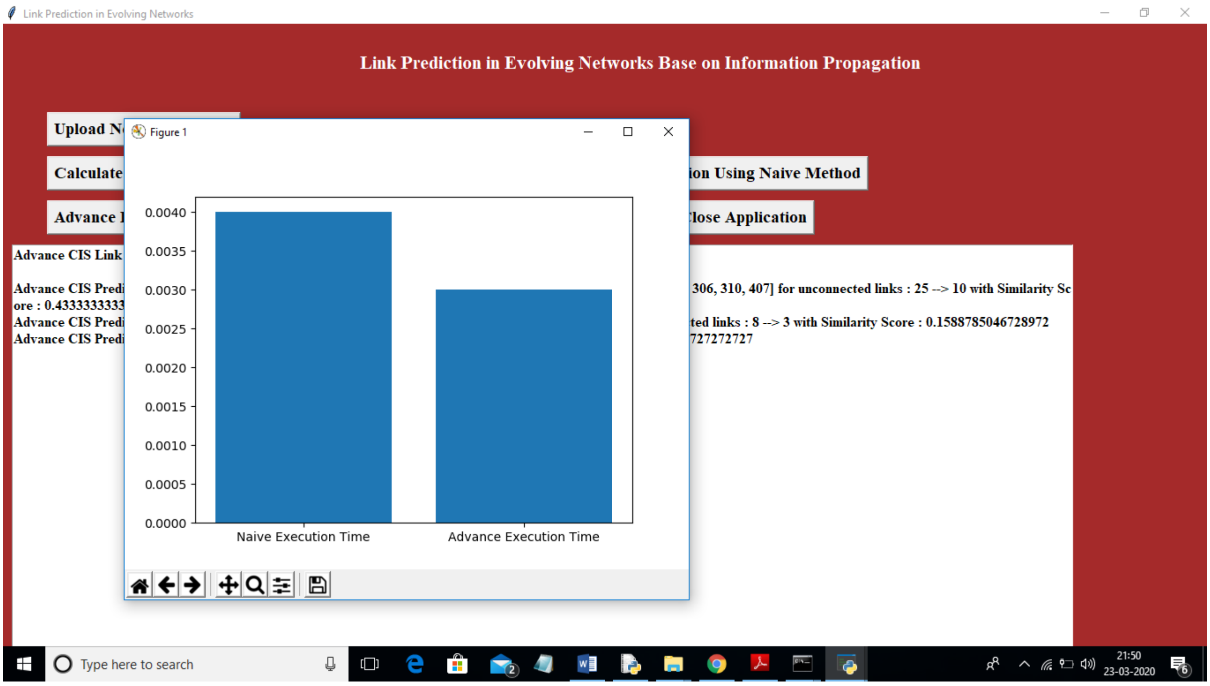
In wiki vote dataset there are lots of users so graph will be dense. In above graph all colour nodes are the influence node which has connectivity with many users. Now close above graph and click on 'CIS Link Prediction Using Naive Method' button to find common nodes set and link prediction for missing links.



In above screen for unconnected missing links 25 → 10 there are lots of predicted links using which 10 can connect to 25. For example 10 can use 257 or 3 or 261 etc to connect to 25. All this 257, 3, 261, 6, 8, 271, 19, 407, 23, 280, 154, 28, 30, 286, 33, 300, 50, 306, 54, 310, 214, 246, 252, 255 are the predicted links for 25 and 10. Similarly we can see for other missing link 8 and 3, 4 and 3. Now click on 'Advance Link Prediction using CIS & Upper Bound' button to find missing links using advance algorithm. Output will be same the only difference is execution time.



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

4. CONCLUSION

This project addressed the crucial issue of link prediction in graph data mining, with a focus on social networks. Link prediction is a valuable tool for predicting missing connections in existing networks as well as predicting new links in future networks. The project introduced a new model called Common Influence Set to calculate node similarities, which is used in their proposed link prediction algorithm. This work evaluated the performance of their algorithm and compared it to previous link prediction algorithms based on similarity using the area under the ROC curve (AUC) as the evaluation metric. The experimental results demonstrated that the proposed algorithm outperformed the previous methods, suggesting its effectiveness in predicting future node similarity.

The future scope of this project involves several avenues for further improvement and applicability. Firstly, enhancing the algorithm's scalability to handle larger and more complex social networks is crucial. Additionally, incorporating temporal information to capture the dynamic nature of social connections could improve accuracy in predicting future links. Exploring hybrid approaches that combine various link prediction techniques may yield even better results. Expanding the algorithm's application to different domains and datasets would provide valuable insights into its generalizability. Moreover, ensuring the interpretability of the model and making it robust to noisy data are important considerations. Lastly, developing techniques for online link prediction would enable real-time link predictions as networks evolve dynamically. Addressing these aspects will contribute to advancing the field of link prediction in graph data mining and its widespread applications.

REFERENCES

[1] Porselvi, Pradeep Kumar A.V Hemakumar S.Manoj Kumar,"Artificial Intelligence Based Price Negotiating E- commerce Chat Bot System" International Research Journal of Engineering and Technology (IRJET)

[2] Tingwei Liu, Zheng Zheng,"Negotiation Assistant Bot of Pricing Prediction Based on Machine Learning" International Journal of Intelligence Science > Vol.10 No.2, April 2020

[3] Shubham Pingale Prasad Kulkarni Rushikesh Ambekar Vanita Babanne," Implementing E-Negotiator Chatbot for E-commerce Website" International Research Journal of Engineering and Technology (IRJET)

[4] Eleni Adamopoulou and Lefteris Moussiades,"An Overview of Chatbot Technology" Springer

[5] Rushikesh Khandale, Shashank Sombansi, Siddharth Mishra, Mohd Fahad Shaikh, Prof. Pooja Mishra," E- Negotiator Chatbot for E-commerce Websites: Implementation" Journal of Applied Science and Computations

[6] Rushikesh Khandale, Shashank Sombansi, Siddharth Mishra, Mohd Fahad Shaikh, Prof. Pooja Mishra," E- Negotiator Chatbot for E-commerce Websites" Journal of Applied Science and Computations

[7] How to Negotiate with a Chatbot – and Win! https://online.hbs.edu/blog/post/how-to-negotiate- with-a-chatbot-and-win

[8] Facebook teaches bots how to negotiate. They learn to lie instead https://www.wired.co.uk/article/facebook-teaches- bots-how-to-negotiate-and-lie