

# Improved Link Prediction Technique Using Community Detection Algorithm

Snigdha Luthra, Gursimran Kaur, Dilbag Singh

**Abstract:** As social technology has connected substantial number of users together to interact and share of millions of information across the network. It is essential to foresee the frequent future links of the users which are connecting together and has the future possibility to connect. The social hub network is dynamic as it changes the structure at different timestamps. The network obtained at time  $t$  is varied at time  $t+1$ . In order to predict the ongoing changes on network, graph embedded techniques are used to obtain an unsupervised graph with different parameters of nodes and edges which can be used in machine learning methods. In this paper, we devise a community detection algorithm with edge betweenness, closeness, betweenness, degree, hubs and authority parameters to predict the efficiency of the model with the dataset and visualize a graph network to determine the centrality of the network model.

**Keywords:** Link prediction, Community Detection, Betweenness, Closeness, Edge Betweenness, Hubs and Authority.

## I. INTRODUCTION

In order to present partnership in the different network between the links and nodes, we consider nodes as people and edge displays a relationship between various nodes. Edge data can end up being shed anticipated due to many reasons such as if information gathered is missing or ambiguous. The modifications in brief period cause many concerns and generate a large number of complicated inquiries such as two nodes will be associate together in how many periods of time and Do the link that occur between two nodes are devised by many others, Nodes which can be not linked, maybe it expected that they will not receive any connection in the future [1]. The different examples that we all address through this formulate study should be to predict the potential relationship between two nodes, seeing that there is no linked relationship between the different people present in the existing derived state. Hence, to forecast such fluctuations with great correctness is obviously significant to obtain the prospect of a sociable network [1]. Consider a network which may be envisaged as a chart in which nodes represent factors and businesses match links or human relationships. The bringing up availability of over the internet public sites is generally beneficial not only in sociable network examination, such as community detection or perhaps web page rumors but also used in exhortation devices also.

Manuscript published on 30 June 2019.

\* Correspondence Author (s)

**Snigdha Luthra** Apex Institute of Technology, Chandigarh University, Gharuan, India

**Gursimran Kaur** Apex Institute of Technology, Chandigarh University, Gharuan, India

**Dilbag Singh** Apex Institute of Technology, Chandigarh University, Gharuan, India

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Devices are effective depiction and are employed in numerous jobs including machine learning and data exploration [3]. Developing editions to analyze these kinds of relational info has attracted a significant quantity of interest, wherever webpage link prediction could be a significant process. Hyperlink conjecture can be described as a task of tracing and forecasting hidden links with a particular cultural guide. The massive prognosis of losing links or the links which will turn out to be formed in the permanent based in photographs on the networking which can be formally described although a good web page link conjecture concern and link through conjecture is a vital job in link research. While building entities are routine practice, building links as in the custom are generally more complex. Therefore, growing a model that can control with this sort of information is normally a demanding task, which will normally reverse to persona, and obtain latent characteristic [5]. One of the outstanding features of the cultural network website link opinion is the continuous change in proportions of the network which boosts and lowers links and organizations over a period of time. In addition, a large dynamic network may be challenging by the variable relational information which changes at different time intervals. Largely, network space is heterogeneous and coping with multiple links. The sparsity of connected data presents another challenge. When a network is unquestionably thin, it may be delicate to sound and noise within a heterogeneous and energetic network.[9] At regular intervals, non-linear transformations are generally seen in strong networks within season diversities. On the other hand, selecting a suited technique can get demanding because there are issues in community network hyperlink prediction. With esteem for this, we mention a triplex analytical construction, which as in the custom classifies link conjecture approaches, was introduced. Furthermore, the suggested framework examines each category based on the presented efficient measure. Conventionally, social networks also include a pair of entities for certain precedent known as network entities. The other imperative elements for the network are building relations and new interactions. A data structure in a graph depends on a number of links between two nodes. Although the comprehensive network structure is normally a graph, a consequence of the complexity and diversity of social networks, it develops simply because of a separate field.

A graph of homogeneous network is defined as:

$$\begin{aligned} \text{Graph} &= (V, E) \text{ where} \\ V &= \{v \in V \mid \forall i, j \text{ vi identical to } v j\} \\ E &= \{e \in E \mid \forall i, j \text{ ei identical to } e j\} \end{aligned}$$

(1)

where  $V$  is the node in the node set and  $E$  is the link in the link set.



# Improved Link Prediction Technique Using Community Detection Algorithm

A discordant network is defined in an identical way, but it may contain many kinds of link and nodes:

$$Graph = (V, E)$$

$$V = \cup i Vi$$

(2)

where  $V$  is defined as the node set which contains the collection of many non identical node type of sets and  $E$  is stated as the union of many heterogeneous different link sets [3]. Link prediction in various social networks using semantics in deep learning comprises of different steps which we apply on the twitter dataset against the positive, negative and neutral comments and dividing the user into categories to predict the relation between them [8]. The more positive or optimistic post on the page, the page is more positive [4]. We also identify the net conjecture strategies structured about their particular complex strategy and discuss the strength and weak spot of different strategies. Because of this importance, Link prediction and semantic analysis in social networking have gained a lot of attention from the researchers. In data mining link prediction is widely used. The main direct of link prediction is to predict future connections in the network which is not present in the current network. As social media is a dynamic object; they change and grow at the different timestamps [7]. Appearing of new edges, nodes and different paths are the changes which occur at different time period and predicting the new edge which can occur in future is our primary task. The network is mainly studied by expert mathematicians so it resulted in a very prominent theory of networks using graphs.

## II. METHODOLOGY

Community Detection algorithm is used to predict the succeeding links in the network. The dataset is from Stanford university social network data in the form of nodes and various edges of the graph. The unique algorithm is devised to cluster nodes together in one community. The nodes belonging to same cluster has more possibility to combine in future than the nodes from another cluster. Various parameters help to predict the accuracy of the graph which is obtained by cluster edge betweenness method. The graph showing hubs describes the node which is following more people is visualized with bigger size and graph showing authority parameter describes the node which is followed by maximum number of nodes.

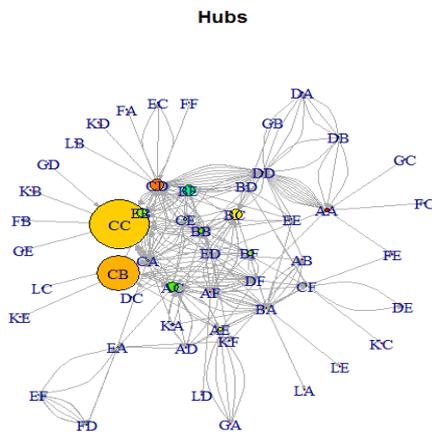


Fig 1. Network graph of hubs

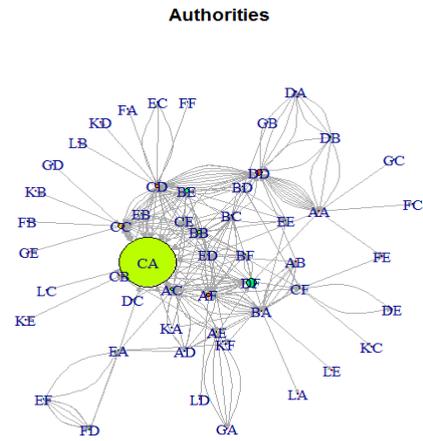


Fig 2. Network graph of authorities

By calculating hub score we obtain the values that depicts CC node is following maximum number of people and authority score shows CA node as the popular node followed by maximum number of users in the network as shown in the figure.

NODES	HUB SCORE	AUTHORITY SCORE
CA	4.494456E-02	1.000000E+00
CC	1.000000E+00	1.121441E-01
CB	7.145888E-01	3.660444E-03
AC	2.093518E-01	6.262068E-02
LE	1.167140E-17	1.133971E-03

This algorithm compares the parameters such as Closeness centrality to calculate the number of different steps required to visit or access every vertex from the given specified vertex. Betweenness parameter calculates the shortest path which is going through every vertex or edge in the graph.

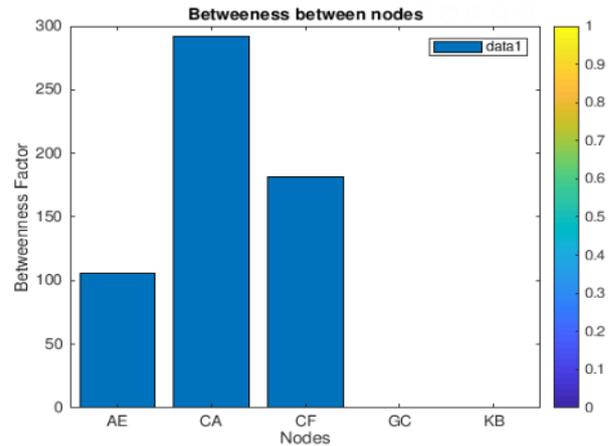


Fig 3. Betweenness parameter values

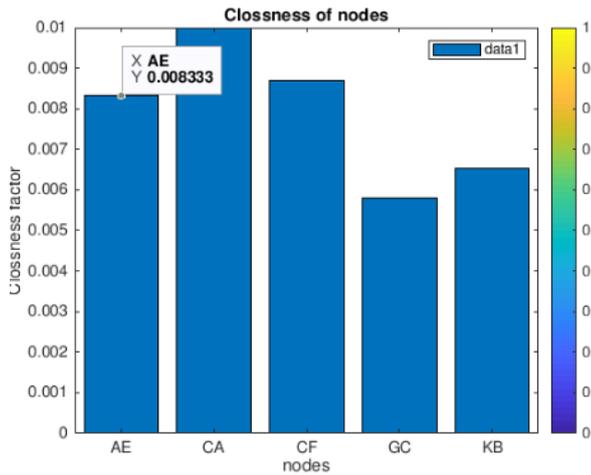


Fig 4. Closeness parameter values

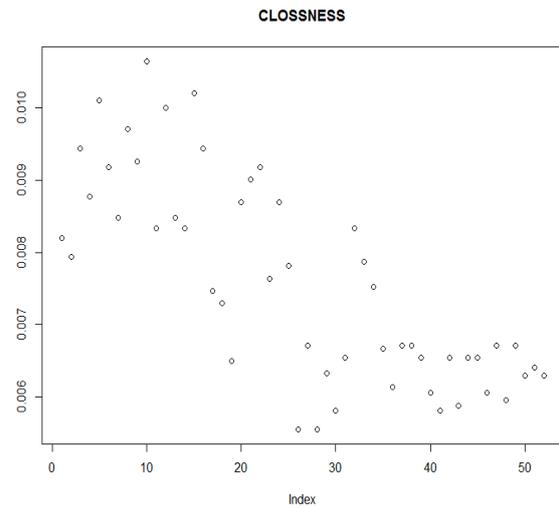


Fig 6. Scatter values of closeness

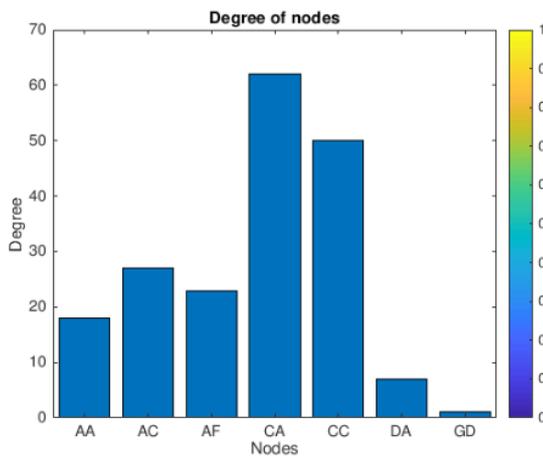


Fig 5. Degree parameter values

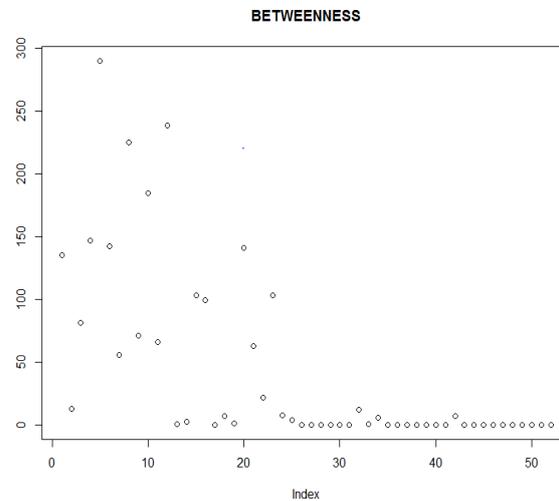


Fig 7. Scatter values of Betweenness

The graph shows CA node has maximum betweenness, closeness and degree which verifies that CA node has the maximum centralization value. The accuracy and centralization value of CA with cluster edge betweenness is found as 0.9969834 which shows that the network is 99% centralized around CA

NODES	CLOSSNESS	BETWEENNESS
CA	0.01000000	292.157605
AE	0.00833333	105.368424
CF	0.00869565	181.814414
GC	0.00581395	0.000000
KB	0.00653594	0.000000

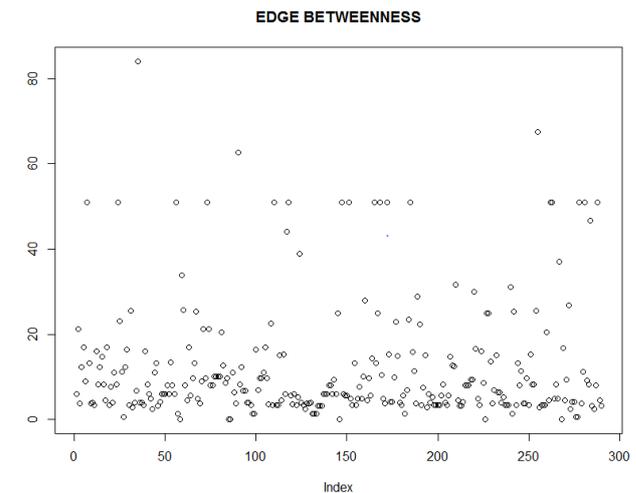
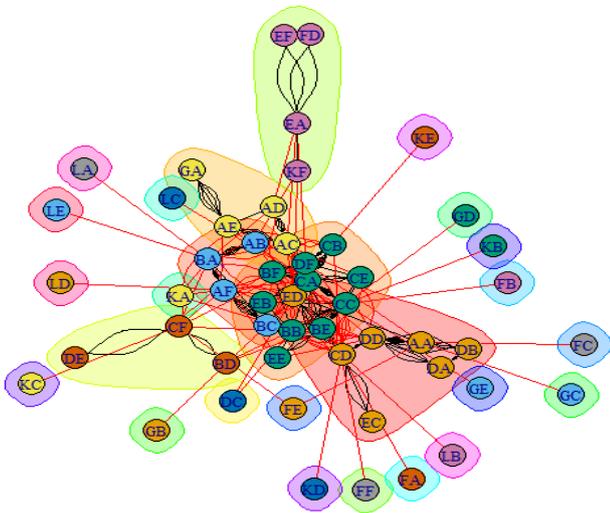


Fig 8. Scatter values of Edge Betweenness

### III. RESULTS AND DISCUSSIONS

The edge betweenness is used in clustering the nodes together to predict whether the nodes will combine in future or not. The edge density of the graph is 0.2187029 which is 21% of the network. The edge density is edge count divided by vertex count and vertex count-1. The methodology of link prediction in social networks using semantics in deep learning comprises of different steps which we apply on the twitter dataset against the positive, negative and neutral comments and dividing the user into categories to predict the relation between them. By giving the snapshot of the network consisting of nodes and edges, it can be predicted very easily that which nodes have the higher probability to combine in future. In order to achieve this we have to build a classifier and an improved model to utilize all the features of node attributes and predict which pair of nodes is likely to collaborate in the future and to design a framework which consider both direct and head-on indirect links and relations among different users in the same network and also predict the interactivity between users using likes, comments, tweets, mention and re-tweets to predict the future links.

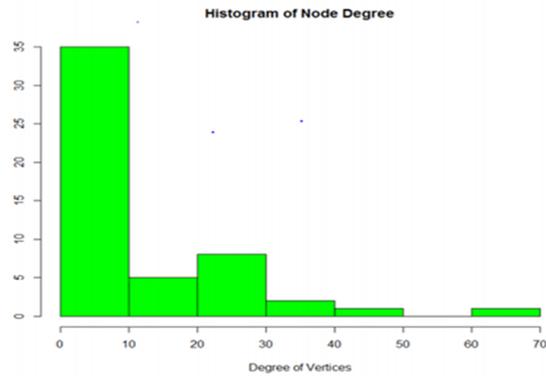


**Fig 9. Community detection clusters**

The communities are divided which shows maximum centralization occurs around CA node. The nodes in same cluster has more possibility to combine in future than the nodes belonging to another cluster. Link prediction is used to recommend new friends in the online social network. It also suggests the interactivity between the members belonging to the company and the organization and also used to predict the relationship between the members of the terrorist organization. It also suggests the collaborations between the researchers based on the authorship and co-authorship and used in websites like tinder to find matches. It is also used in E-commerce application like Amazon to purchase the items and understanding the behavior of the customers.

### IV. PERFORMANCE ANALYSIS

A histogram is basically an accurate description and portrayal of the different dispersal of various numerical data present. It is a rough estimate of the different probability distribution which is of a continual variable also known as a quantitative variable and which was introduced by Karl Pearson [10].



**Fig 10. Histogram of node degree**

This histogram shows that the maximum nodes has less connections and has less degree while there are only few nodes which are having maximum connections.

### V.CONCLUSION

The primary motivation of this work is to forecast the links which can occur in future at different timestamps of directed and undirected graphs. The nature of the relationships can be efficiently measured by graph terminology to discover the different patterns of the nodes linkage. For this experiment community detection algorithm performs clustering technique to group nodes together in same cluster which has similar edge betweenness factor. The proposed framework decides which further connections can be established on the basis of nodes belonging to similar cluster. Furthermore links can also be initialized with weights to calculate their interaction in do future work.

### REFERENCES

1. Ijaz, M., Ferzund, J., Suryani, M. A., & Sardar, A. (2018). Social Network Link Prediction using Semantics. (IJACSA) International Journal of Advanced Computer Science and Applications , Vol. 9, No. 1.
2. Bengio, Y., Courville, A., & Vincent, P. (24 Jun 2012). Unsupervised Feature Learning and Deep Learning: A Review and New Perspectives. Department of computer science and operations research .
3. Ahmed, C., & ElKorany, A. (25-28 Aug. 2015). Enhancing link prediction in Twitter using semantic user attributes. 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). Paris, France: IEEE.
4. Bahabadi, M. D., Golpayegani, A. H., & Esmaeili, L. (July 2014). A Novel C2C E-Commerce Recommender System Based on Link Prediction: Applying Social Network Analysis. International Journal of Advanced Studies in Computer Science & Engineering (ijascse) , Vol 3, Issue 7.
5. Behnaz, M., & Meybodi, M. R. (2018). Link prediction in weighted social networks using learning automata. Department of Computer Engineering, Amirkabir University of Technology, Elsevier .
6. Colgrove, C., Neidert, J., & Chakoumakos, R. (December 11, 2011). Using Network Structure to Learn Category Classification in Wikipedia.
7. Ermis, B., Acar, E., & Cemgil, A. T. (3 0 A u g 2 0 1 2). Link Prediction via Generalized Coupled TensorFactorisation. a r X i v : 1 2 0 8 . 6 2 3 1 v 1 [ c s . L G ] .
8. Gao, F., Musial, K., Cooper, C., & Tsoka, S. (2015). Link Prediction Methods and Their Accuracy for Different Social Networks and Network Metrics. Scientific Programming .
9. Haghan, S., & Keyvanpour, M. R. (2017). A systemic analysis of link prediction in social network. springer .



10. Hasan, M. A., & Zaki, M. J. (17 March 2011). A Survey of Link Prediction in Social Networks. *springer* , pp 243-275.
11. Hasan, M. A., & Zaki, M. J. (March 2011). A Survey of Link Prediction in Social Networks. *Social Network Data Analytics* .
12. Ijaz, M., Ferzund, J., Suryani, M. A., & Sardar, A. (2018). Social Network Link Prediction using Semantics. (*IJACSA*) *International Journal of Advanced Computer Science and Applications* , Vol. 9, No. 1.
13. Lü, L., & Zhou, T. (15 March 2010). Link prediction in complex networks: A survey. *ScienceDirect* , 1150-1170.
14. Manjula, R., & Srilatha, P. (August 2016). User behavior based link prediction in online social networks. *International Conference on Inventive Computation Technologies (ICICT)*.
15. Nowell†, D. L., & Kleinberg‡, J. (January 8, 2004). The Link Prediction Problem for Social Networks. *1 ACM International Conference*, (pp. pp. 556–559).
16. Papadimitriou, A., Symeonidis, P., & Manolopoulos, Y. (September 2012). Fast and accurate link prediction in social networking systems. *Journal of Systems and Software* .
17. Sachan, M., & Ichise, R. ( August 2010). Using Semantic Information to Improve Link. *IACSIT International Journal of Engineering and Technology* , , Vol.2, No.4,.
18. Tang, J., Chang, S., Aggarwal, C., & Liu, H. (8 Dec 2014). Negative Link Prediction in Social Media. *Social and Information Networks* .
19. Zhang, C., Zhang, H., Yuan, D., & Zhang, M. (04 May 2017). Deep Learning Based Link Prediction with Social Pattern and External Attribute Knowledge in Bibliographic Networks. *2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*. Chengdu, China: IEEE.
20. Zhao, P., Aggarwal, C., & He, G. (23 June 2016). Link Prediction in Graph Streams. *IEEE 32nd International Conference on Data Engineering (ICDE)*. Helsinki, Finland: IEEE.

### AUTHORS PROFILE



**Snigdha Luthra** is pursuing her master of engineering in computer science and applications in big data specialization from Chandigarh University, Gharuan. She has done her Bachelor of engineering from Haryana engineering college affiliated under Kurushetra university (2013). Currently, she is working as teaching assistant at Chandigarh University, Gharuan, Mohali, Punjab, India. She has published 4 research papers in scopus international conferences. Her research interest is in machine learning, networking and data analysis.



**Gursimran kaur** received her MTech degree from Thapar Institute of Engineering and technology, Patiala(2018). She did her Btech degree from Chandigarh Group of colleges, Landran (2016). Currently, she is working as an Assistant Professor in Chandigarh University, Gharuan. She has published 5 research papers in scopus conferences. Her research interest includes data mining, data analysis and machine learning.



**Dilbag Singh** received his PhD degree from Department of Computer Science and Engineering, Thapar Institute of Engineering and Technology, Patiala. He has done his Master in Technology (Computer Science and Engineering) from Guru Nanak Dev University, Amritsar, Punjab (2012). Currently he is working as Assistant professor at Chandigarh University, Gharuan, Punjab, India. He has published more than 27 research papers in well known reputed SCI indexed journals and international conferences. His research interest includes wireless sensor networks, digital image processing and meta heuristic techniques.