Link Prediction on Social Media: Survey

Aditi Jangid[1], Ipshita Deb[2], Shiva Duddupudi[3], Aditya Bhardawaj[4], Dr. Nitin N Pise[5]

[1, 2, 3, 4]*Dept. of computer science, Maharashtra institute of technology, Pune, Maharashtra, India*
[5]*Assistant Professor, Dept. of computer science, Maharashtra institute of technology, Pune, Maharashtra, India*

Abstract: The term "link prediction" is a task for analyzing social networks which also has applications in other domains like, information retrieval, bioinformatics and e-commerce. Social network information is the key sources that are used to design the similarity function between entities. Link prediction is mainly based on measures for analyzing the "proximity" of nodes in a network. In this survey paper we are discussing about link prediction concept and the main focus is on the different related work done by different researchers. Finally we have briefly describes social network and link prediction with its advantages.
Keywords: Link prediction, network evolution model, social network analysis.

I. INTRODUCTION

In today's world the rapid development, online social network has been a part of people's life. The plenty of information systems ,sociology and biology, can use the network that can describe, which nodes represent particular and edges represent the relationships between individuals or the interaction between individuals. Link prediction is to predict whether there will be links between two nodes that are based on its attribute information. Link prediction has been applied in many relation networks such as collaborations in co authorship networks, prediction of being actor, underground relationships between terrorists, and so on. A single person is the node of the network while edges, that link nodes and are called also "connections", "links", correspond to relationships between people as represented in following Figure. There are several examples of social network services, which includes;

A. My Space;
B. Twitter;
C. Live Journal, etc.

Link prediction provides potential applications in automatically inferring real-world connections and discovering, labeling, and characterizing communities. The proposed system extracts the user's homepage data from their twitter account. By using NLP they can extract the text categories.

II. LITERATURE REVIEW

Lada A. Adamic et al [1] introduce techniques and tools to mine this information in order to extract social networks and the exogenous factors underlying the networks' structure. In an analysis of two data sets, from Stanford University and the Massachusetts Institute of Technology (MIT), we show that some factors are better indicators of social connections than others, and that these indicators vary between user populations.

The techniques provide potential applications in automatically inferring real world connections and discovering, labeling, and characterizing communities.

William H. Hsu et al [5] describes prediction task for previously unobserved links in social networks was discussed in the project. The concept of social network was defined as well as social graph representation.

The related works in research area were analyzed. Existed approaches were compared: supervised vs. unsupervised, single table data representation as feature vector vs. relational data mining etc. In case of interpretation link prediction task as classification task the application of range of induces were digested such as: J48, OneR, IB1, Logistic, NaiveBayes as well as other available approaches for resolving given task e.g. SVM, GP, BN, PRMs etc.

Hasti Akbari Deylami et al [7] focuses to boost the performance of similarity based link prediction methods by using community information. This information is derived from the structure of the graph, based on the number of community levels that two vertices have in common, in a hierarchical representation of communities. To evaluate the performance of the proposed method, four datasets are used as benchmark. The results suggest that the information of communities often increases the efficiency and accuracy of link prediction.

Md. Shafiur Rahman et al [10] introduce concept of the link in the social network has been utilized to discover the hidden meaning of different fields such as e-commerce, bioinformatics and information retrieval. The prediction of a new link between two nodes in

the social network is normally accomplished based on the nature of the topology and the similarity function among the nodes is defined with the help of the number of common friends.

In this paper, two link prediction algorithms: Local Link Prediction Algorithm and Global Link prediction by taking into consideration of user's activities as well as the common friends. We apply two formulas called correlation based cScore and influential score based iScore to measure the similarity between the two predicted nodes. Finally, analyze the performance of the proposed algorithms by using DBLP, PPI, PB, and USAir data sets and the experimental result attests that our link predicted algorithm outperforms over the existing algorithms.

The supervised algorithms often have better performance than unsupervised algorithms; although in some cases, the supervised algorithms, due to their complexity and time-consumption in training phase, cannot be used in large scale graphs and may only be practical for networks of a few thousand nodes. Probabilistic models and maximum likelihood methods are some examples of supervised link prediction methods.

There is a more detailed review of these approaches in [11].

In [12], the structure of communities is used to improve the performance of link prediction methods. To do so, community information of common neighbors of a pair of nodes has been added to the similarity based link prediction methods. Our method is different from this paper; instead of using community information of common neighbors of a pair of nodes, we use community information of the pair of nodes, themselves.

Mohammad Al Hasan et.al [14] largely discusses about three types of models: first, are the traditional models which extract a set of features to train a binary classification model. Second, includes the probabilistic approaches which model the joint-probability among the entities in a network by Bayesian graphical models. And, finally is the linear algebraic approach which computes the similarity between the nodes in a network by rank-reduced similarity matrices.

III. EXISTING SYSTEM

The existing systems provide the One of the first large scale web applications was the serving of individual homepages. These generally autobiographical pages reflect a user's interests and experiences. Homepages are not free-floating in the Web, but point to and are pointed at by other users, our "friends and neighbors" on the Web. These links can represent anything from friendship, to collaboration, to general interest in the material on the other user's homepage. In this way individual homepages become part of a large community structure.

A. Social Network And Link Prediction

The link prediction is an important research field in data mining. It has a wide range of scenarios. Many data mining tasks involve the relationship between the objects. Link prediction can be used for recommendation systems, social networks, information retrieval, and many other fields.

The link prediction task is divided into two categories;
1) The initial category is to predict the new link will appear in future time.
2) The second category is to forecast hidden unknown link in the space.

The easiest framework of link prediction algorithm is based on the pair of node x and node y respectively. This node is a function Similarity (x, y), this function is defined as the similarity function between nodes x and y.

Then the pair of nodes are sorted according to the function values that may range from the largest to smallest, the greater the value of the similarity function, the greater the probability of the link in the nodes respectively.

IV. ADVANTAGES
A. The whole system can be implemented in very low cost and provides better accuracy.
B. Predict the links between users.

V. CONCLUSION

This work survey initially an attempt is made to introduce the concept of social networks and describe the basic nature of social networks. Then we introduce the link prediction definition.

This survey paper also includes various literature work done by different researchers till now. Last section of this paper includes brief discussion about social network and link prediction with its advantages.

REFERENCES

[1]   Adamic, L.A., 1999. The small world Web. In: Proceedings of the European Conference on Digital Libraries, Springer, Berlin.

[2]   Adamic, L.A., Adar, E., 2000. Frequency of Friendship Predictors, 2003 Elsevier http://www .hpl.hp.com/shl/papers/fnn/.

[3]   Albert, R., et al., 1999. The diameter of the World Wide Web. Nature 401, 130.

[4]   Batagelj, V., Mrvar, A., 2000. Some analyses of Erdös collaboration graph. Social Networks 22, 173–186.

[5]   Prof. William H. Hsu and Svitlana Volkova "LINK PREDICTION IN SOCIAL NETWORKS" https://www.researchgate.net/publication/226566834

[6]   Bollobas, B., 1985. Random Graphs. Academic Press, London.

[7]   Hasti Akbari Deylami and Masoud Asadpour "Link Prediction in Social Networks UsingHierarchical Community Detection" 2015 IEEE.

[8]   Carley, K., 1991. A theory of group stability. American Sociological Review 56 (3), 331–354.

[9]   Davis, J.A., 1967. Clustering and structural balance in graphs. Human Relations 20, 1181–1187.

[10] Md. Shafiur Rahman, Leema Rani Dey, Sajal Halder, Md. Ashraf Uddin, Manowarul Islam" Link Prediction by Correlation on Social Network" 2017 IEEE.

[11] L. Lu and T. Zhou, "Link prediction in complex networks: A survey," Physical A: Statistical Mechanics and its Applications, 2010.

[12] S. Soundarajan and J. Hopcroft,"Using community information to improve the precision of Link prediction methods," ACM press, 2012: 607-608

[13] Diaz, S., 2000. Cell Phone Signals Touted to Fight Traffic Wars, San Jose Mercury News, and http://www.Mercurycenter.com/svtech/news/indepth /docs/ traf012100.htm.

[14] Mohammad Al Hasan and Mohammed J. Zaki "LINK PREDICTION IN SOCIAL NETWORKS"