International Journal For Technological Research In Engineering
Volume 1, Issue 9, May-2014 *ISSN (Online): 2347 - 4718*

LINK PREDICTION IN SOCIAL MINING

Yogesh Vaghela[1] (ME scholar), M.B.Chaudhari[2] (HOD)
Computer Science and Engineering Department
Government Engineering College, Gandhinagar, Gujarat, India.

Abstract: Online social networking sites have become very popular over the last few years. So, new researchers are interested to work in social network. Different methods are applied to network containing millions of user. Unfortunately, links between individuals may be missing either due to online network. The primary goal of link prediction technique is extracting structural features required for classify links. In this paper, we focus on link prediction in social networks. We provide an efficient and effective link prediction method, which consists of three steps as follows: (1) we locate the similar nodes of a target node; (2) we identify candidates that the similar nodes link to; and (3) we rank candidates using weighing schemes. One of the methods used to uncover missing information in social networks is referred to as link prediction.

Index Terms: Link prediction; Social network; Social mining;

I. INTRODUCTION

Link prediction, introduced by Liben - Nowell and Kleinberg refers to a basic computational problem underlying social network evolution in time. Given a snapshot of a social network at time t and a future time t′, the problem is to predict the new friendship links that are likely to appear in the network within the time interval [t, t′]. As Liben-Nowell and Kleinberg state, the link prediction problem is about to what extent the evolution of a social network can be modeled using features intrinsic to the network itself. Indeed, in their framework, they consider only the features that are based on the link structure of the network. The Below Figure illustrate simple link prediction [1]

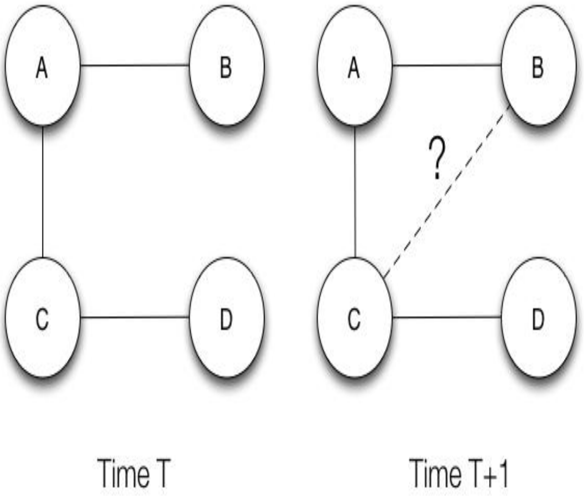

Fig.1: Link prediction simple definition

As part of the recent surge of research on large, complex networks and their properties, a considerable amount of attention has been devoted to the computational analysis of social networks structures whose nodes represent people or other entities embedded in a social context, and whose edges represent interaction, collaboration, or influence between entities. Natural examples of social networks include the set of all scientists in a particular discipline, with edges joining pairs who have co-authored papers; the set of all employees in a large company, with edges joining pairs working on a common project; or a collection of business leaders, with edges joining pairs who have served together on a corporate board of directors. The increased availability of large, detailed datasets encoding such networks has stimulated extensive study of their basic properties, and the identification of recurring structural features. In this Paper, we focus on the problem of link prediction particularly in the context of evolving co-authorship networks. This has been a hotbed of recent research activity where much of the focus has been on encapsulating the topological and/or semantic information embedded in such networks to address the link prediction problem. In contrast in this article we explore the realm of probabilistic models derived from frequency statistics and use the resulting predictions from the probabilistic models as additional features to further enhance predictions made by topological-based and semantic-based link prediction algorithms. In addition to its role as a basic question in social-network evolution, the link-prediction problem could be relevant to a number of interesting current applications of social networks. Increasingly, for example, researchers in artificial intelligence and data mining have argued that a large organization, such as a company, can benefit from the interactions within the informal social network among its members; these ties serve to supplement the official hierarchy imposed by the organization itself. Effective methods for link prediction could be used to analyze such a social network to suggest promising interactions or collaborations that have not yet been identified within the organization. In a different vein, research in security has recently begun to emphasize the role of social-network analysis, largely motivated by the problem of monitoring terrorist networks; link prediction in this context allows one to conjecture that particular individuals are working together even though their interaction has not been directly observed. The link-prediction problem is also related to the problem of inferring missing links from an observed network: in a number of domains, one constructs a network of interactions based on observable data and then tries to infer additional links that, while not directly visible, are likely to exist. This line of work differs from our problem formulation in that it works with a static snapshot of a network, rather than considering network evolution; it also tends to take into account specific attributes of the nodes in the network, rather than evaluating the power of prediction methods that are based purely on the graph structure.

II. RELATED WORK

The link prediction problem consists of a family of prediction problems. While most classification of the link prediction problem sub-divides it into two or three sub-problems [3], we give more comprehensive classifications of the problem as follows:

Link Disappearance Prediction. (Will a current link disappear?) [4]

Link Classification. (What is the nature of the link?)[5]

Anomalous Link Discovery. (What are the unexpected links?) [6]

Link Weight Prediction (Predict the change in the weight of link)

Time Series Link Prediction (Prediction which links will reoccur over time)

Link Regression. (How does a user rate an item?)[7]

A number of topology based measures have been used for the link prediction tasks, these include Newmans common neighbors [8], Jaccard’s Index, Adamic/Adar metric [9] etc. Murata and Moriyasu [10] extend these metrics for weighted graphs and use for link prediction in a Q&A system. Huang [11] proposed a graph topology based method which generalizes the clustering coefficient and defines the problem of link prediction as that of cycle completion in graphs. It should also be noted that the topology based formulation of the problem can also be described as the problem of matrix completion which can be accomplished by matrix factorization .Topology based temporal metrics were employed by Potgieter et al to increase the performance of link prediction techniques. Most existing methods of link prediction assume that these links in network are undirected. However, examples of directed networks are numerous: the web is made up of directed hyperlinks; the food webs consist of directed links from predators to praise and users form links to their opinion leaders in micro blog. Modeling links as directed networks introduce complexity but offer significant analytical benefits. When a link is symmetric, there are only two states: the link is present or absent. When links are asymmetric, there are four states between two nodes: node links to node, links to, and are mutually connected, or the absence of a link between and. If there exists a directed link from to, we might say that has a power or status advantage over, since is more important to than is to [2]. The directed link is an indicator of the direction in which attention flows. To the best of our knowledge, quantitative approaches in directed networks are few. To fill this gap, we focus on link prediction in directed networks in this paper. We propose link prediction method, which can provide efficient and effective link prediction in directed network. We conduct experiment to evaluate the accuracy of proposed method using real-world micro blog data. [13]

III. PROXIMITY FEATURES

Proximity features are characteristics that represent some form of proximity between the pair of nodes. For instance, it is highly possible that a user could follow another with whom they share a mutual friend. Additionally, proximity features are usually cheap to be computed. Denotations used in this section are:

Γin (u) denotes the followers of u; Γout (u) denotes the followees of u.

Common Followers Count -- Measures the overlap of the followers between u and v. $CFER (u,v) = | \Gamma in(u) \cap \Gamma in(v)|$

Common Followees Count -- Measures the overlap of the followees between u and v. $CFEE (u,v) = | \Gamma out(u) \cap \Gamma out(v)|$

Common Friends Count -- Measures the overlap of the friends between u and v, where bi-directional edge denotes friendship. It can be formularized as:

$CF (u,v) = | \Gamma out(u) \cap \Gamma in(u)| \cap | \Gamma out(v) \cap \Gamma in(v)|$

Common Neighbours Count -- Measures the overlap of local networks of u and v by ignoring the directions of links. This feature is included to evaluate the impact of direction on link creation.

$CN (u,v) = | \Gamma out(u) \cup \Gamma in(u) | \cap | \Gamma out(v) \cup \Gamma in(v) |$ [12]

IV. THE PROPOSED METHOD

In this section, we propose a link prediction method in directed network. He idea of the proposed method is that a node tends to link to the nodes which its similar nodes link to. So, for a given node, the method we present consists of three steps: (1) we locate similar nodes of a target node; (2) we identify candidates that the similar nodes link to; and (3) we rank candidates using weighing schemes. To describe the proposed method, we construct a directed graph G(V, E),where V represents a set of nodes in directed network and E represents a set of links among these nodes. A directed link $\langle u, v\rangle \in E$ exists between nodes u and v if u links to v. The set of our neighbors of node u is $\Gamma out(u) = \{v \in V|(u, v) \in E\}$, and the out-degree of u is $|\Gamma out (u)|$, where $|\cdot|$ denotes the size of the set. Similarly, $\Gamma in(u) = \{v \in V|(v,u) \in E\}$, represents the set of in neighbors of u and in-degree of u is $|\Gamma in(u)|$.he input to our problem is the directed network G and a target node u. Our task is to predict the likelihood of the existence of the link from u to other unlinked nodes in terms of observed topology of the directed network. In the remaining subsections, we, respectively, provide detailed descriptions of these three steps that essentially constitute the proposed method.

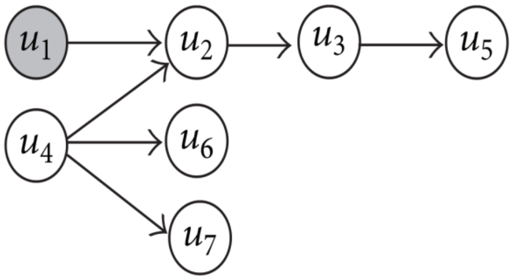

Fig 2 (a). An Example of Directed Graph

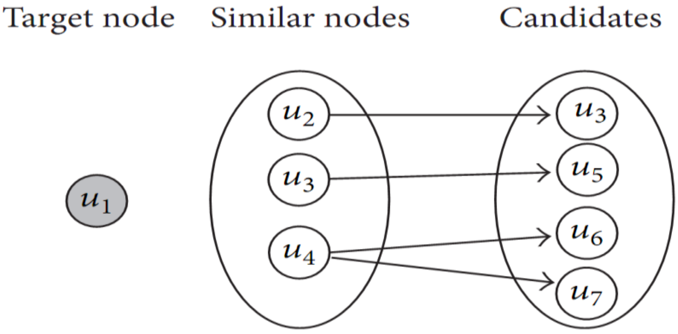

Fig 2 (b). Similar nodes and candidates of u_1 in Fig 2(a)
Fig 2. Example of Proposed method

V. CONCLUSION

Link prediction has important theoretical and practical value. Recently, many link prediction algorithms have been proposed. However, most studies of link prediction assumed that links of network are undirected. In this paper, we focus on link prediction in directed networks, which provide efficient and effective link prediction in directed network. The method we present consists of three steps as follows: (1) we locate the similar nodes of a target node; (2) we identify candidates that the similar nodes link to; and (3) we rank candidates using weighing schemes. We conduct experiment in micro blog to evaluate the accuracy of proposed algorithm by using real micro blog data. The experimental results show that the proposed approach is promising, which indicates that our proposed method is effective in user recommendation in micro blog. First, aggregating three categories of similar nodes with different weights is effective because they contain more useful information to recommend followees that a target may be interested in. Second, considering similarity of similar users and target user can improve the accuracy performance. In light of our future study, we would like to explore an efficient and effective method to determine the required parameters, and we are planning to include other directed networks to carry out more experiment.

REFERENCES

[1] David Liben-Nowell, Jon Kleinberg “The Link-Prediction Problem for Social Networks”, Journal of the American Society for Information Science and Technology.

[2] Chao Wang, Venu Satuluri, Srinivasan “Local Probabilistic methods for link prediction”,

[3] Xiang, Evan W., A Survey on Link Prediction Models for Social Network Data (2008).

[4] Sharan, U., Neville, J., Exploiting Time-Varying Relationships in Statistical Relational Models. SNA-KDD 2007.

[5] Rattigan, Jensen. The case for anomalous link discovery. ACM SIGKDD, 2005

[6] Tylenda, T., Angelova, R., Bedathur, S.,Towards Time-aware Link Prediction in Evolving Social Networks, KDD-SNA 2009

[7] Muhammad Aurangzeb Ahmad, Zoheb Borbora, Jaideep Srivastava , Noshir Contractor, “Link Prediction Across Multiple Social Networks”, 2010 IEEE

[8] Newman, M. E., Clustering and Preferential Attachment in Growing Networks, Physical Review Letters E, Vol.64 (025102), 2001.

[9] Murata, Tsuyoshi, Moriyasu, Sakiko. Link Prediction of Social Networks Based on Weighted Proximity Measures. Web Intelligence 2007: 85-88

[10] Adamic, L. A., Adar, E., Friends and Neighbors on the Web, Social Networks, Vol.25, No.3, pp.211-230, 2003.

[11] Swati Jain1, Naveen Hemrajani2,” Detection and mitigation techniques of black hole attack in MANET: An Overview”, International Journal of Science and Research (IJSR), India Online ISSN: 2319-7064

[12] Huang, Z., Link Prediction Based on Graph Topology: The Predictive Value of the Generalized Clustering Coefficient, Workshop on Link Analysis: Dynamics and Static of Large Networks, the 12th ACM SIGKDD, Philadelphia, PA, 2006.

[13] T. Zhou, J. Ren, M. Medo, and Y. C. Zhang, “Bipartite network projection and personal recommendation,” Physical Review E, vol. 76, no. 4, Article ID 046115, 2007.