International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 03 | Mar-2018 www.irjet.net p-ISSN: 2395-0072

A Survey on Link Prediction Techniques

Sisira C Sunny[1], Dr. Anishin Raj M. M[2]

[1]PG Student, Dept. of CSE, Viswajyothi College of Engineering and Technology, Muvattupuzha, Kerala, India
[2]Assoc.Professor, Dept. of CSE, Viswajyothi College of Engineering and Technology, Muvattupuzha, Kerala, India

---

Abstract - *Link prediction is one of the interesting problems in the area of networks. Link prediction is the process of inferring new links in a network that may form in the future. This is applicable to social media and other kind of networks. Several models are used for link prediction which comprises of several algorithms. The major problem with link prediction is the massive size of networks. Network sparsity is also a challenge. This paper presents a literature survey on different link prediction methods that are commonly used and it focus on different link prediction algorithms.*

***Key Words*: Link prediction, Index, Social networks, Friendship network, Network.**

1. INTRODUCTION

Link prediction is a good method for identifying the hidden relationship in networks. It also involves predicting missing links. It is also known as link inference. The renowned applications of link inference involve friend suggestions in Facebook and recommendations in YouTube. Link prediction can be applied to different kinds of networks such as friendship network, scientific network etc. These involve finding the similarity between nodes in the network and use that similarity for predicting links. There are mainly two kind of link prediction techniques. There are supervised and unsupervised. Supervised techniques are similar to classification problem. They are also more accurate. Unsupervised method uses proximity and other methods for link prediction.

2. RELATED WORKS

In this section, different methods for link prediction are described. There are many techniques. Some of them are discussed below.

Lada A. Adamic and Eytan Adar proposed a technique for predicting links between individuals [1].This is a popular unsupervised method. It uses neighborhood measurement for predicting links. They focus mainly on individual homepages. Personal homepages provide a glimpse of user communities. Homepages represent a user's interest and behavior. Their study include text, out-link, in-link and mailing list as data sources. In this paper, they clearly describe the structure of homepage link. For predicting friendship between users, they used a score. This score is used for ranking users and ranks represent the closeness between users. Similarity between two users can be found by taking the summation of the reciprocal of logarithm of frequency of shared items between them.

Tao Zhou, Linyuan Lu and Yi-Cheng Zhang proposed another method for link prediction [5]. They mainly focus on missing link prediction problem. It is also another neighborhood based method. This method also ranks non-existing links based on their score. In this paper, they used six real-world networks namely PPI (Protein-protein interaction network), NS (Network of co-authorships between scientists), Grid (Electrical power grid of the western US), PB (Network of the US political blogs), INT- (Router-level topology of the Internet), and USAir (Network of the US air transportation system). Comparisons between nine similarity indices are also provided in this paper. The basic idea behind resource allocation is that a node can send some information to another node via their common neighbours. So the similarity between two nodes a and b is the amount of resource b received from a. This method works well for networks with high average degree.

Jaewon Yang and Jure Leskovec proposed an overlapping community detection method called BIGCLAM [4]. A community is a group of related nodes. Overlapping community refers to the interlinked communities. Their proposed method works well for massive networks. In this paper, they presented certain observations about community networks. They used six datasets in this paper. They proposed a cluster affiliation model as well as community detection.

Leo Katz proposed a new status index [2] for popularity contests. It not only tracks the number of votes, but also identify who chooses them. He described a matrix-representation of sociometric data which shows who vote for whom. Rows represent chooser and column represent chosen. Cell values are always zero or one indicating the voting status. Row sum for a particular chosen shows the number of people who votes for him. Leo Katz later proposed a method to find column sum. From those findings, he developed a new status index.

Lars Backstrom and Jure Leskovec described another technique called supervised random walk [3] for link recommendation. They also addressed the sparsity problem and network size. Supervised random walk finds links by combining network structure with attributes of nodes and edges. They tries to assign strength to upcoming links based on these features. There are two kind of nodes-positive nodes and negative nodes. Positive nodes are nodes that may be a

part of a future link. All other nodes are negative. Since this is a supervised method, one method is to use a training dataset that contain positive and negative nodes for every node in the network. Then the algorithm starts learning. The problems with this method are class imbalance and feature extraction. Second approach is to rank nodes based on probabilities. Algorithm proposed by Lars Backstrom and Jure Leskovec combine these two techniques effectively. That is, node and edge features are used to find strength and probabilities.

3. CONCLUSIONS

Social networks are nowadays emerging rapidly. User experience is a major factor in such networks. To enhance user experience and to increase the popularity of networks link prediction is widely used. Since size of networks increases day by day, there is a high need for efficient link prediction algorithms. In this paper, several such methods are described. Most of them are effective in massive networks. They provide unique techniques for solving real-world link inference problem. Therefore it contribute much to the area of networking.

REFERENCES

[1] L. Adamic and E. Adar, "Friends and neighbors on the web. Social Networks," 25:211–230, 2001.

[2] L. Katz, "A new status index derived from sociometric analysis," Psychometrika, 18(1):39–43, 1953.

[3] L. Backstrom and J. Leskovec, "Supervised random walks: predicting and recommending links in social networks," In WSDM, 2011.

[4] J. Yang and J. Leskovec, "Overlapping community detection at scale: A nonnegative matrix factorization approach," In WSDM, 2013.

[5] T. Zhou, L. L¨u, and Y. Zhang, "Predicting missing links via local information," The European Physical Journal B, 71:623–630, 2009.