Chapter 61
Link Prediction in Complex Networks

Manisha Pujari
Université Paris 13, France

Rushed Kanawati
Université Paris 13, France

ABSTRACT

This chapter presents the problem of link prediction in complex networks. It provides general description, formal definition of the problem and applications. It gives a state-of-art of various existing link prediction approaches concentrating more on topological approaches. It presents the main challenges of link prediction task in real networks. There is description of our new link prediction approach based on supervised rank aggregation and our attempts to deal with two of the challenges to improve the prediction results. One approach is to extend the set of attributes describing an example (pair of nodes) calculated in a multiplex network that includes the target network. Multiplex networks have a layered structure, each layer having different kinds of links between same sets of nodes. The second way is to use community information for sampling of examples to deal with the problem of class imbalance. All experiments have been conducted on real networks extracted from well-known DBLP bibliographic database.

1. INTRODUCTION

Link prediction has attracted the attention of many researchers from different research fields. It consists of estimating the likelihood of existence or appearance of an edge between two unlinked nodes, based on observed links and attributes that contain information about the nodes, edges or the entire graph. It has important applications in many fields including social, biological and information systems etc. Link prediction has been widely used in biological networks like protein interaction network, metabolic networks, food web. It is used for finding missing links and thereby helps in reducing the experimental cost if the predictions are accurate. In social interaction and academic or commercial collaboration networks they can play an important role to predict new associations (new edges). This further has utility in

DOI: 10.4018/978-1-7998-2460-2.ch061

recommendation task: a service provided by almost all social networks and majorly used in e-commerce sites. Link prediction can also be helpful in finding hidden links in criminal networks, which is another critical field of research. Link prediction can be basically of two types: structural and temporal.

- Structural link prediction refers to the problem of finding missing or hidden links which probably exist in a network (Liben-Nowell & Kleinberg, 2007; Menon & Eklan, 2011; Taskar et al., 2003; Yin et al., 2011). It focuses on inferring the existence of links that are not directly visible, by using observable data of the network. It has direct application to find unobserved patterns of genes, protein interactions for the medical studies on various diseases like cancer, HIV, Alzheimer etc. (Airoldi et al., 2006; Eronen & Toivonen, 2012). It can also help to find existing criminal links which often remain hidden in a network (Clauset et al., 2008; Fire et al., 2013).
- Temporal link prediction refers to the problem of finding new links by studying the temporal history of a network (Benchettara, Kanawati, & Rouveirol, 2010; Hasan, Chaoji, Salem, & Zaki, 2006; Berlingerio et al., 2009; Hasan et al., 2006; Huang & Lin, 2008; Liben-Nowell & Kleinberg, 2007). So here we have information about the network till time t and the goal will be to predict a new link that may appear at some point of time in future say t+k. It has its application primarily in recommendation systems that are being used widely in e-commerce websites for product recommendations, in any search engines to help users with probably relevant terms they might be searching, for recommendation of tags in social resources sharing websites like Flickr[1], YouTube[2], De.li.ci.ous[3] etc. and very commonly used for recommendation of friends in many social networks like Facebook[4] and Twitter[5]. It has another significant use in predicting future collaborations between researchers for academic purposes (Benchettara et al., 2010a,b; Kunegis et al., 2010).

Rest of this chapter continues as follows. In section 2, there is a formal description of the problem of link prediction. Details about different evaluation methods for link prediction approaches is given in section 3. Section 4, has a detailed description of various link prediction approaches focusing mainly on topological and temporal link prediction methods. Section 5 summarizes some important challenges in link prediction especially faced by supervised classification based models. Section 6 presents details of our work in the field of link prediction. Section 7 concludes the chapter.

2. PROBLEM DESCRIPTION AND NOTATIONS

The problem of prediction of new links (or simply called link prediction problem) refers to a question of inferring the formation of links at a future time, by studying the history of appearance or disappearance of links in a network over a period of time. In topology based link prediction approaches, only structural properties of the network are used to implement learning methods and to find a model that will be used to predict links.

For prediction of new links at a certain point of time t_{n+1} having network information till time t_n, the network can be presented as a sequence of graphs representing different snapshots of the network at different points of time $\langle t_0,\ t_1,\ \ldots,\ t_n \rangle$. Suppose the temporal sequence of graphs $G = \langle G_0, G_1, ..., G_n \rangle$ each having their own sets of nodes and edges. The network can also be represented as a graph $G = \langle V, E \rangle$ such that $V = \bigcup_{i=0}^{n} V_i$ and $E = \bigcup_{i=0}^{n} E_i$. The goal of a link prediction approach is to find the likelihood