



Link Prediction in Social Networks using improving Adamic-Adar and Jaccard methods

Hadis Bashiri¹, Gholamhossein Dastghaibfard²

¹MSC Student, Shiraz University, hadis.bashiri1@gmail.com

² Associate Professor, Shiraz University, dstghaib@shirazu.ac.ir

Abstract

In recent decade, social networks are an important issue and those are a part of people life. The number of connected users are increasing daily. Active users in social networks are more than one hundred million. People in social networks interact together and share their life. Some example of social network are Facebook, google+, Instagram and etc. link prediction is an important and challenging problem in social networks. Social networks are dynamic and it give the importance of link prediction. Dynamic means that it is maybe members and relations not exist in next moment. On the other hand maybe in the next time will create new member or new relation. So there is a need to predict not exists links at this moment that maybe will create in the next moment. We need some information to link prediction, that we can gain those from the network graph. Forasmuch as Adamic-Adar have good precision on the different datasets, in this paper, we improving Adamic-Adar Method using Jaccard index. To evaluate proposed method, we compared it with several link prediction method. The proposed method was successful that Increasing the Adamic-Adar precision of Facebook's dataset 0.2%, and the rate of improvement on the hamster dataset is 23%.

Keywords: Link Prediction, Adamic-Adar, Jaccard.

Introduction

In true life, the people are not independent of each Other, they are interacted and affected with other. If we only pay attention to individual characteristics and neglect relationships between them, this is bound to affect the efficiency and comprehensiveness of analysis. Social Networks have achieved great popularity in recent years. These days human beings communicate together through social networks more than past.

A graph G is defined as $G = (V, E)$, where V is a set of nodes or vertices and E is a set of edges or links between them. We use undirected graph in this type if exist a link between u and v also this link exists between v and u .

Social networks are too similar to a graph that include nodes and edges. That can be directed like Google+ and can be undirected like Facebook.

Links are very important in social network analysis, since without these links, network lost it nature and become so meaningless also by analyzing the links we can earn much information from links.

So, studying social networks have been significant for todays life. For anyone who administers a social network, it is an important matter to correctly predict the relationship between users in the network, by using this can create a better environment for people and customers. Also for other people is important to find a person that is like itself or is an old friend in the real world. Therefore link prediction is an important issue in the social network. Link prediction is not only to use in social network field, it also can be used in other fields. It can be used to security issues i.e. to recognize and composition two guilty and terrorist network [1]. It can be used to bioinformatics, i.e. to recognize communications between proteins[2]. It can be used to electronic commerce. It can be used to recommendation system [3] i.e. predicting a product for a person, because a purchase always not from experience, sometimes moreover appearance and experience there are other things too, i.e. we can use friends of people to predict a product for them.

Link prediction definition: we have a snapshot of a social network in time t at form of a graph and we want to accurately predict the edges that will be added to the network in time $t+1$. Figure 1 show two small network in time t and $t+1$.

As has shown in the figure 1, in time t dataset have 5 nodes and in time $t+1$ the number of nodes has increased, but the number of edge maybe increase or decrease. In our picture two edges added to graph. In time t between 1-4 , 1-5 , 4-5 , 2-5 and etc. an edges rely on the method that used can be added. For example two edges added to graph at time $t+1$.

there are many challenges in link prediction according to [1]: (1) in the heterogeneous network, it is not clear how is the best way to combine the topology and information, (2) total imbalance link prediction datasets, that is, the number of known