Engineering Applications of Artificial Intelligence 70 (2018) 16–24

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

# Link prediction in weighted social networks using learning automata

Behnaz Moradabadi, Mohammad Reza Meybodi *

*Department of Computer Engineering, Amirkabir University of Technology, Tehran, Iran*

A R T I C L E I N F O

*Keywords:*
Social network
Link prediction
Weighted network
Learning automata

A B S T R A C T

Link prediction is an important task in Social Network Analysis. The present paper addresses predicting the emergence of future relationships among nodes in a social network. Our study focuses on a strategy of learning automata for link prediction in weighted social networks. In this paper, we try to estimate the weight of each test link directly from the weights information in the network. To do so, we take advantage of using learning automata, intelligent tools that try to learn the optimal action based on reinforcement signals. In the method proposed here, there exist one learning automata for each test link that must be predicted and each learning automata tries to learn the true weight of the corresponding link based on the weight of links in the current network. All learning automata iteratively select their action as the weight of corresponding links. The set of learning automata actions will then be used to calculate the weight of training links and each learning automata will be rewarded or punished according to its influence upon the true weight estimating of the training set. A final prediction is then performed based on the estimated weights. Our preliminary link prediction experiments with co-authorship and email networks have provided satisfactory results when weights are considered.



## 1. Introduction

The advancement of the internet has provided better chances of collaboration and interaction among people and organizations. The advancement has paved the way for the emergence of social networks over the internet which is nowadays very popular. A social network can be formally shown as a graph, where the vertices represent people or organizations, and the connecting edges indicate social connections. Social Network Analysis (SNA) is a vast area of research dealing with techniques and strategies for the study of social networks (Liben-Nowell and Kleinberg, 2007). The analysis and knowledge of networks widely employed to understand the behavior of a community (Liben-Nowell and Kleinberg, 2007; Al Hasan and Zaki, 2011). SNA gives us opportunities and benefits in different areas like marketing, economics, health, sociology and safety (Al Hasan and Zaki, 2011). Link prediction is one of the main tasks undertaken by SNA. The task is concerned with the problem of predicting the prospective existence of relationships among nodes in a network, based on patterns observed in the existing nodes and relations. Link prediction can help us make out the mechanisms that trigger the evolution in a social network and it can be applied to many application areas. For instance, in the area of Internet and web science, it can be used in tasks such as automatic web hyper-link creation (Adafre and de Rijke, 2005) as well as web site hyper-link prediction (Zhu et al., 2002). In e-commerce, one of the most prevalent usages of link prediction is to build recommendation systems (Li and Chen, 2009; Huang et al., 2005). It also finds various applications in other scientific fields. For example in bibliography and library science, it can be tapped for de-duplication (Malin et al., 2005) as well as record linkage (Elmagarmid et al., 2007). In bioinformatics, nevertheless, it has been used in protein–protein interaction (PPI) prediction (Freschi, 2009). In security-related areas, it can be applied to identify hidden groups of terrorists and criminals (Al Hasan and Zaki, 2011). The literature shows various strategies and approaches to treating this problem (Xiang, 2008; Wang et al., 2007). In general, the most widely used techniques are based on one of three approaches, namely: structural measures or patterns in the network; the similarity between nodes (content and/or semantics of the nodes); probabilistic models. These approaches will be briefly explained in Section 2.

Weighted networks are a kind of social network in which each link has a weight that indicates the strength of the corresponding link (Liben-Nowell and Kleinberg, 2007). Link prediction in such networks is required to adapt the current methods such as adopting the similarity metric based link prediction to consider weights in the network. But in this area, there are some researches that show the strong links are important in link prediction (Murata and Moriyasu, 2007). On the other hand, there are studies that show weak links are important in

* Corresponding author.
*E-mail addresses:* moradabadi@aut.ac.ir (B. Moradabadi), mmeybodi@aut.ac.ir (M.R. Meybodi).

https://doi.org/10.1016/j.engappai.2017.12.006
Received 1 April 2016; Received in revised form 17 November 2017; Accepted 6 December 2017
0952-1976/

the link prediction (Lü and Zhou, 2009). So, in this research, we will try to estimate the weight of each test link directly from the links weight information in the network. To do so, we use learning automata, intelligent tools that try to learn the optimal action based on the reinforcement signal.

A learning automata (LA) is an adaptive decision-making unit that tries to learn the optimal action from a set of allowable actions by interacting with a random environment (Thathachar and Sastry, 2011). Within each step, an LA selects an action from its action-set. Action selection in the learning automata is based on a probability distribution over the action set. The selected action is applied to the environment, after which a reinforcement signal is produced by the environment. The learning automata update the probability distribution of its actions according to both reinforcement signal and a learning algorithm. Then it will choose an action once more. These steps are repeated until the automata converge to some action.

In the proposed method there exist one learning automata for each link that must be predicted, and each LA tries to learn the true weight of the corresponding link according to the current network's links weight information. We also partition the network links in two sets: the training set that we use for training LAs, and the test set that must be predicted. In each iteration of the proposed algorithm, each LA chooses a weight as its action. After choosing actions, we will have a weighed network of the test links. Now, we define some metrics to calculate the weight of the training set using these new weights. After calculating the weight of the training set from the weights of the test set, we generate a reinforcement signal for each LA based on its influence on the true weight estimation of the training set, and each LA updates its action probability distribution according to its reinforcement signal. After estimating the weight of test links, we sort them by their weights and predict the existence/absence of each link based on its weight. Our experiments demonstrate that link prediction in the proposed method out-performs other link prediction methods.

Section 2 briefly discusses the link prediction problem and related works in the weighted network. Section 3 will then present the used weighted similarity metrics for calculating the weight of links from the weights information in the network. Section 4 reviews learning automata briefly. After that, Section 5 leads through the proposed algorithm and procedures for weighted link prediction problem based on learning automata. Section 6 brings the experiments and obtained results to a conclusion.

## 2. Link prediction

A classic definition of the link prediction problem is expressed by: "Given a snapshot of a social network at time $t$, we seek to accurately predict the edges that will be added to the network during the interval from time $t$ to a given future time $t + 1$" (Liben-Nowell and Kleinberg, 2007). The most widespread approach to the problem is to explore the topological/structural patterns from the social network of interest (Huang, 2010; Murata and Moriyasu, 2008). Different metrics to describe node pairs have already been adopted in previous works, including for example the number of common neighbors, the path distance between the two nodes, Jaccard's coefficient, and the Adamic–Adar coefficient, among others Al Hasan et al. (2006), Liben-Nowell and Kleinberg (2007) and Murata and Moriyasu (2008). Such metrics explore structural patterns of the network and commonly provide a degree of proximity/similarity between the nodes. The used metrics can be either local (limited to the direct neighbors of nodes) or global (covering the entire network). As previously mentioned, the starting point in these techniques is to extract the values/scores of different metrics that represent the proximity of pairs of nodes. Then, the obtained data are processed to build a model which can predict the hidden links or links that will appear in the future.

The node-wise similarity based approach searches appropriate measures of similarity between two nodes according to the content and/or semantics they present (Al Hasan et al., 2006). Each node on the network can be represented as a vector of features. The more similar two nodes are in terms of their particular attributes, the more likely they are to relate. Cosine coefficient, mutual information, and Dice coefficient are examples of techniques used in this approach.

The approaches that are based on probabilistic models aim to learn the best probabilistic model that abstracts the network information. The basic idea here is to create the model through a set of parameters $\theta$, given the observed social network $G = (V, E)$. The existence of the link between the pair of nodes $x$ and $y$ is estimated by the conditional probability $P(e(x, y)|\theta)$. This approach examines the elements of the network with the help of relational data models and it enables us to encapsulate relevant information from nodes relationships. Relational Markov Networks and Relational Bayesian Networks are two examples of models dealt with in this approach.

With regard to link prediction considering structural patterns, pairs of non-connected nodes are at first ranked according to a chosen metric (for instance, the number of common neighbors) (Huang, 2010; Murata and Moriyasu, 2008). Then, the top $L$ ranked pairs are assigned as predicted links. To put it another way, it is always assumed that links that have the highest scores are most likely to occur.

We should underline here that earlier works in supervised link prediction consider metrics computed for unweighted social networks. In this type of network, the strength of relationships is not taken into account (only their existence is considered) (Al Hasan et al., 2006). In the rest of this section, we go on to review related link prediction methods in weighted social networks:

In De Sá and Prudêncio (2011) a supervised machine learning strategy for link prediction in the weighted network is proposed. The method uses link weights that express the "strength" of relationships. Here, the results of supervised prediction on a co-authorship network revealed satisfactory results when weights were taken into account.

Murata and Moriyasu (2007) indicates that link prediction based on graph proximity measures fits open and dynamic online social networks. It proffers new weighted graph proximity measures for link prediction of social networks. The method relies on an assumption that proximities between nodes would be better estimated by using both graph proximity measures and the weights of existing links in a social network. By taking into consideration the weights of links, link prediction performance is improved via previous proximity measures.

Paper (Wind and Morup, 2012) has studied the effect of using weight information when recovering missing edges in a network following the framework of Wang et al. (2007). The researchers have observed that the application of a Poisson-based model on a binary network does not hamper the structure modeling. Using Poisson-based models for weighted networks, and for binary versions of the same networks, they observed that weight information did not improve link prediction. They further witnessed that complex and flexible models in general performed better than simpler models regardless of the available information (i.e., a fraction of edges treated as missing). When predicting the weights of the missing edges, the researchers in the said study saw that complex model overfits to the edges, resulting in a poor recovery of the true edge-weight. Also, there are some relevant works about the prediction in the weighted social networks: In Dong et al. (2013) the dynamics properties of mobile calling patterns and some social characteristics are studied based on a large mobile call duration network where the weights of the links are call durations. They found that the stronger ties have lower call duration; the average call duration get shorter when the end point of call have more common neighbors; the opinion leaders have shorter call duration and the social balance tends to shorter call duration. Based on these facts they proposed a probabilistic model to predict the call duration and they compared their methods with some based methods. In Gupte and Eliassi-Rad (2012) an obvious method to infer the tie strength between the users using bipartite event and people network is proposed. They modeled the characterizations of functions that could serve as a measure of tie strength. They showed that for