

Temporal link prediction algorithm based on local random walk

Yuanxiao Fan¹, Pei-ai zhang²

¹*School of Information and Science Technology, Jinan University,
Guangzhou, 510632, China, E-mail: 1329791691@qq.com.*

²*School of Information and Science Technology, Jinan University,
Guangzhou, 510632, China, E-mail: tzhangqpa@jnu.edu.cn.*

(Received July 15, 2017, accepted September 25, 2017)

Abstract. Link prediction is an important part of complex network research. Traditional static link prediction algorithm ignores that nodes and links in network are added and removed over time. But temporal link prediction can use the information of historical network to make better prediction. Based on local random walk, this paper proposes a time-series random walk algorithm. Given link data for times 1 through T, then we predict the links at time T+1. The algorithm first computes the Markov probability transfer matrix at each time, then combines them into a transformation matrix, and applies the local random walk algorithm to obtain the final prediction result. The experimental results on real networks show that our algorithm demonstrates better than other algorithms.

Keywords: Strong and Weak solutions, temporal link prediction, Markov probability transfer matrix, local random walk.

1. Introduction

Many social, biological, and information systems can be well described by networks, where nodes represent individuals, biological elements, computers, web users, and so on, and links denote the relations or interactions between nodes. The study of complex networks has therefore become a common focus of many branches of science. Link prediction [1] aims at estimating the likelihood of the existence of links between nodes. The prediction of existent yet unknown links is similar to the data mining process, while the future links relates to the network evolution.

Link prediction has huge theoretical and practical value. It has attracted a great deal of interest of many scientists from different fields. In many biological networks, such as food webs, protein-protein interaction networks and metabolic networks, whether a link between two nodes exists must be demonstrated by a large number of laboratorial experiments, which are usually very costly. Instead of blindly checking all possible interactions, to predict based on known interactions and focus on those links most likely to exist can sharply reduce the experimental costs if the predictions are accurate enough [2,3]. In rapid development of social networks and e-commerce platforms, link prediction is widely used in personalized recommendation system. Specifically, in online social network, based on some information of the current network, link prediction can recommend friends to users by predicting their possible relationship [4]. In e-commerce platform, link prediction can recommend goods or information to users. It can not only improve the users' experience, but also solve the problem of data sparseness in commodity recommended system [5]. In addition, in scientific collaboration network, link prediction can be used to identify the possibility of cooperation [6]; In the aeronautical network, link prediction can provide transportation management strategy for airlines [7].

In recent years, the research of link prediction algorithms have attracted much attention. Document [8] is a summary literature of link prediction in complex networks, and link prediction algorithms are classified into three categories: (1) similarity-based algorithm; (2) the maximum likelihood algorithm; (3) probabilistic model for link prediction. This kind of classification reflects different modeling ideas of link prediction problem. The similarity-based algorithm is the most commonly used algorithm, which is characterized by simple algorithm, low computational complexity and widely apply. In this approach, some of the basic properties of nodes can be used to define their similarities, such as common features or topologies between nodes [9].

Most of the existing link prediction algorithms are still based on static network, that is, according to the static network at a certain time, the possibility of the link between nodes at next time is predicted, which is called static link prediction method [10]. In fact, these methods ignore the temporal characteristics of

network evolution, which is unreasonable in many specific application scenarios. Huang et al. [10] introduced temporal link prediction method, taking temporal evolutions of link occurrences into consideration to predict link occurrence probabilities at a particular time. Bliss et al. [11] proposed an evolutionary algorithm to integrate the topological features and node attribute characteristics in the network to improve the link prediction accuracy. Dunlavy et al. [12] considered bipartite graphs that evolved over time and considered to use matrix and tensor-based methods to predict future links. In addition, a weight-based approach is proposed to integrate multiyear data into a single matrix, and future links can be predicted by using a truncated singular value. Behnaz Moradabadi et al. [13] proposed a new time series link prediction based on learning automata, for each link that must be predicted there is one learning automaton and each learning automaton tried to predict the existence or non-existence of the corresponding link. Deng et al. [14] proposed the concept of a temporal link prediction method based on the prediction error correction. Ahmed et al. [15] proposed a method for link prediction in temporal uncertain networks, which formalized the predicting problem by designing a random walk in temporal uncertain networks to obtain more accurate results.

This paper is organized as follows: Section 2 describes the temporal link prediction problem. Section 3 provides some concepts, presents the local random walk method, and then gives a temporal link prediction algorithm. Section 4 presents and analyzes the experimental results for link prediction algorithm on real datasets. Finally, Section 5 offers conclusions and possibilities for future work.

2. Problem description

Given a network $G = \{G_1, G_2, \dots, G_t, \dots\}$ represents an evolution of network (as shown in Figure 1). A temporal network can be described by snapshots $G_t = (V_t, E_t, A_t)$ for $t = t_0, t_0 + 1, \dots, t_0 + T - 1$, where T is the window size, V_t is the node set of G_t , E_t is the edge set of G_t , and A_t is the adjacency matrix. The temporal link prediction problem is to predict the probability of occurrence of edges between two nodes in the network at time $t_0 + T$. At time t , a link prediction algorithm is given to predict the probability s_{xy} that any nodes pair (v_x, v_y) will generate a new link at the next moment, that is, a representation of a s_{xy} corresponds to one link prediction method, temporal link prediction method (TLPM) can be represented by a mapping:

$$\text{TLPM: } G' \rightarrow S$$

Where $G' \subset G$ is historical network topology information that has been observed, and contains multiple continuous network topological graphs; S represents the link probability matrix, which is a guess about network topology in the future. Lv Linyuan and Zhou Tao [8] summed up different link prediction algorithms, and this paper selects common neighbors (CN) index, Adamic-Adar (AA) index, priority link (PA) index, Katz index as a reference.

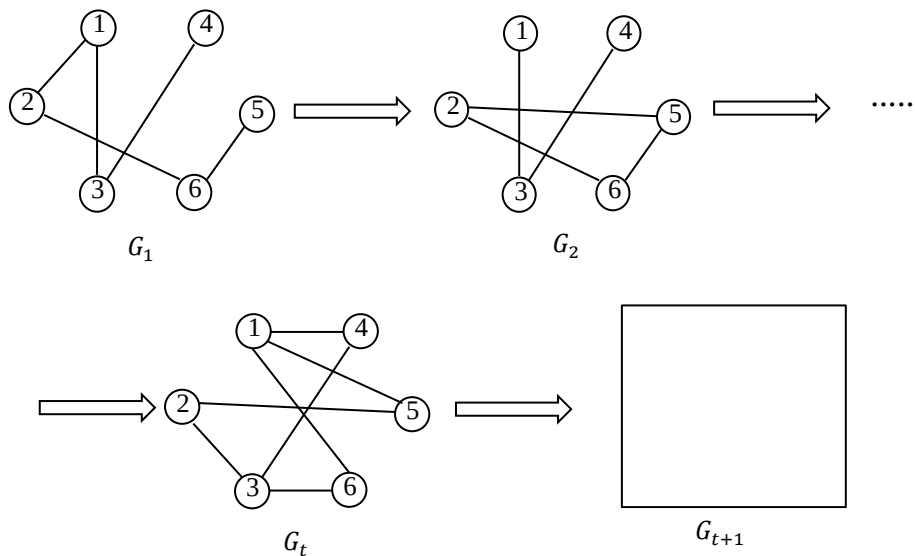


Fig. 1: temporal link prediction diagram of undirected simple network