PHYSA: 17429 Model 3Gsc pp. 1–11 (col. fig: NIL)

ARTICLE IN PRESS

Physica A xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Physica A

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

Predicting the evolution of complex networks via similarity dynamics

Tao Wu[a,b,c,*], Leiting Chen[a,b,c], Linfeng Zhong[d,e], Xingping Xian[f]

[a] Department of Computer Science and Engineering, University of Electronic Science and Technology of China, China
[b] Institute of Electronic and Information Engineering in Dongguan, University of Electronic Science and Technology of China, China
[c] Digital Media Technology Key Laboratory of Sichuan Province, China
[d] Web Science Center, University of Electronic Science and Technology of China, China
[e] Big Data Research Center, University of Electronic Science and Technology of China, China
[f] Department of Computer Science and Technology, Chengdu Neusoft University, China

HIGHLIGHTS

- Introduce a general solution for dynamic networks' evolution prediction.
- Propose an effective and robust link prediction index.
- Envision networks as dynamic systems and model similarity dynamics.
- Propose a position drift model to infer future network structure.

ARTICLE INFO

Article history:
Received 18 April 2016
Received in revised form 23 June 2016
Available online xxxx

Keywords:
Link prediction
Evolutionary dynamics
Spatial–temporal position drift
Network evolution

ABSTRACT

Almost all real-world networks are subject to constant evolution, and plenty of them have been investigated empirically to uncover the underlying evolution mechanism. However, the evolution prediction of dynamic networks still remains a challenging problem. The crux of this matter is to estimate the future network links of dynamic networks. This paper studies the evolution prediction of dynamic networks with link prediction paradigm. To estimate the likelihood of the existence of links more accurate, an effective and robust similarity index is presented by exploiting network structure adaptively. Moreover, most of the existing link prediction methods do not make a clear distinction between future links and missing links. In order to predict the future links, the networks are regarded as dynamic systems in this paper, and a similarity updating method, spatial–temporal position drift model, is developed to simulate the evolutionary dynamics of node similarity. Then the updated similarities are used as input information for the future links' likelihood estimation. Extensive experiments on real-world networks suggest that the proposed similarity index performs better than baseline methods and the position drift model performs well for evolution prediction in real-world evolving networks.

* Corresponding author at: Department of Computer Science and Engineering, University of Electronic Science and Technology of China, China.
E-mail address: wutaoadeny@gmail.com (T. Wu).

http://dx.doi.org/10.1016/j.physa.2016.08.013
0378-4371/

Please cite this article in press as: T. Wu, et al., Predicting the evolution of complex networks via similarity dynamics, Physica A (2016), http://dx.doi.org/10.1016/j.physa.2016.08.013

1. Introduction

Complex network theory, a marriage of ideas and methods from statistical physics and graph theory, provides an ideal tool for studying complex systems and leads to major advances in our understanding of metabolic networks, urban road networks, social and communication networks [1]. In the last decade, many works have been done in the research community of complex networks, including structure analysis [2–5], spreading modeling [6,7], link prediction [8,9], similarity measure [10,11], network reconstruction [12,13] and information filtering [14,15]. In particular, the evolution prediction of network structure is a critical topic in complex network research. Understanding the underlying mechanisms and predicting the evolution of dynamic complex networks is fundamental to many applications, including suppressing virus propagation, controlling rumor diffusion and protecting ecological network system. These problems are all equivalent to asking how about the future structure of the network systems. Since sensing future network structure can guide individuals' behavior in the exploration of complex systems, the question which we will address here is how to predict the evolution of complex networks. According to Ref. [16], every effective link prediction algorithm corresponds to one or more mechanisms of network organization and evolution. Hence, we will in fact answer the above question with link prediction paradigm.

Link prediction problem has received extensive research in complex network studies. Specifically, D. Liben-Nowell and J. Kleinberg [17] argue that the link prediction problem asks to what extent can the evolution of a complex network be modeled using features intrinsic to the network itself? They summarized many similarity indices based on network structure and found that there is indeed useful information contained in the network topology alone by comparing them with random predictors. Commonly, the topology-based methods can be divided into three classes. The first class is similarity-based methods, which assumes that the links between more similar nodes are of higher existing likelihood, and the similarity of the two endpoints can be transferred through the links. The similarity-based methods can be subdivided into neighbor-based methods and distance-based methods. Neighbor-based methods are based on the idea that two nodes are more likely to generate a link if they have more common neighbors, such as common neighbors index (CN), Adamic–Adar index (AA) [18], resource allocation index (RA) [19], Leicht–Holme–Newman index (LHN) [20]. Distance-based methods suppose that link probability is determined by the distance or the number of the shortest path between nodes, such as local path index (LP) [19], Katz index [21], Leicht–Holme–Newman index (LHN-II) [20]. The second class of link prediction methods is maximum likelihood estimation methods. Two popular methods of this type are hierarchical structure model (HSM) [22] and stochastic block model (SBM) [23]. The third class is machine learning based methods. The main methods of this type are supervised learning method [24] and negative matrix factorization (NMF) method [25]. Owning to its simplicity, the study on similarity-based algorithms is the mainstream issue.

In most of the existing link prediction works, a typical evaluation method is to calculate an algorithm's accuracy by reproducing the known links that have been removed from test set. Moreover, most of the works do not make a clear distinction between future links and missing links, except Ref. [26] offers an evidence that missing links are more likely to be the links connecting low-degree nodes. Thus an accurate prediction of network links is not necessarily a useful one. Consequently, predicting future links without losing generality in dynamic networks is still one of the most challenging tasks.

Up to now, many dynamic networks have been investigated empirically to uncover their evolution mechanisms. From the perspective of group evolution [27,28], it has already been pointed out that many new group members are the neighbors or second level neighbors of the current group members, and the neighbor nodes are more likely to join the cohesive groups than the unstructured ones. In other words, the neighbor nodes prefer to connect with the nodes in groups with higher clustering level. Ref. [29] presents a detailed study of network evolution by analyzing four large online social networks with full temporal information. The study shows that most new edges span very short distances, typically closing triangles. Furthermore, the result of Ref. [30] shows that the auto-correlation function of the successive states of evolving communities is continuous, which indicates that the states of networks are associated with the states at the last time point. Thus the timestamps of network interactions have potential to influence network evolution. Now the big question is obviously, how to utilize the spatial and temporal factors of networks to predict future links?

The essence of link prediction is to estimate the topology similarity of node pairs based on the observed network structure, i.e. the similarity relations between nodes. In order to estimate the topology similarity in networks with different structure properties, the paper firstly proposes a robust and structure-dependent index. Moreover, the crux of dynamic networks' evolution prediction is the future links prediction. According to the essence of the link prediction problem, the future similarity relations are the basis of future links prediction. To infer the future similarity relations, we envision evolving networks as dynamic systems and investigate the similarity dynamics based on the current network condition, in which node's influence is defined based on spatial and temporal factors and node's network position is defined as the similarity relations between the node and their neighbors. Then a spatial–temporal position drift model is proposed to update node's network position iteratively according to the node influence. The variation of the similarities of node pairs reflects the underlying evolution trend of current network, and the iterative updating of nodes' network position would lead to a drifted network structure. Finally, according to the experimental results in real-world networks, we find that the structure-dependent index is effective and robust for link prediction and the spatial–temporal position drift model performs well in the prediction of network evolution.

The rest of the paper is organized as follows. Section 2 introduces some indices as baselines. Section 3 presents the structure-dependent index and the spatial–temporal position drift model. Section 4 gives the experimental results. Discussion and conclusion are drawn in Section 5.

Please cite this article in press as: T. Wu, et al., Predicting the evolution of complex networks via similarity dynamics, Physica A (2016), http://dx.doi.org/10.1016/j.physa.2016.08.013