Physica A 450 (2016) 609–623

Contents lists available at ScienceDirect

Physica A

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

Link prediction based on hyperbolic mapping with community structure for complex networks

Zuxi Wang, Yao Wu, Qingguang Li *, Fengdong Jin, Wei Xiong

National Key Laboratory of Science and Technology on Multi-spectral Information Processing, Key Laboratory of Ministry of Education for Image Processing and Intelligence Control, School of Automation, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan 430074, PR China

HIGHLIGHTS

- A novel link prediction method based on hyperbolic geometry framework.
- Use community structure as the heuristic information to map the network.
- Proposed method outperforms many existing popular prediction methods.
- The time complexity of the proposed algorithm is linear.

ARTICLE INFO

Article history:
Received 3 March 2015
Received in revised form 29 November 2015
Available online 14 January 2016

Keywords:
Link prediction
Community structure
Hierarchical organization
Hyperbolic geometry
Complex networks

ABSTRACT

Link prediction is becoming a concerned topic in the complex network field in recent years. However, the existing link prediction methods are unsatisfactory for processing topological information and have high time complexity. This paper presents a novel method of Link Prediction with Community Structure (*LPCS*) based on hyperbolic mapping. Different from the existing link prediction methods, to utilize global structure information of the network, *LPCS* deals with the network from an overall perspective. *LPCS* takes full advantage of the community structure and its hierarchical organization to map networks into hyperbolic space, and obtains the hyperbolic coordinates which depict the global structure information of the network, then uses hyperbolic distance to describe the similarity between the nodes, finally predicts missing links according to the degree of the similarity between unconnected node pairs. The combination of the hyperbolic geometry framework and the community structure makes *LPCS* perform well in predicting missing links, and the time complexity of *LPCS* is linear, which makes *LPCS* can be applied to handle large scale networks in acceptable time. *LPCS* outperforms many state-of-the-art link prediction methods in the networks obeying power-law degree distribution.

1. Introduction

The well-being of various complex systems greatly affects our life. Many complex systems can be well described by networks, including social networks, technical networks, biological networks, etc. Great efforts have been made to understand the evolution of networks, the relationships between the structures and the functions, and the network characteristics [1–4]. Recently, link prediction [5,6] in complex networks, which aims at estimating the likelihood of the emergence of a link not existing originally between two nodes based on the known structure information of the network,

* Corresponding author.
E-mail address: qg_li@hust.edu.cn (Q. Li).

http://dx.doi.org/10.1016/j.physa.2016.01.010
0378-4371/

has attracted the increasing attention of researchers in the fields of physical and computer science. The study of link prediction, which is closely connected with the understanding of the structure and evolution of complex networks, has important theoretical significance. The excellent link prediction models help to understand the structure of networks [7]. Link prediction is expected to provide a fair and uniform comparison platform for numerous evolution methods to push forward the study of network evolution [8,9]. Moreover, link prediction has huge practical application value, such as guiding the protein interaction experiments [10,11], solving the label classification problem [12], building recommendation system for on-line social networks [13], etc.

Link prediction can detect three kinds of links: (1) missing links in sampling networks, such as World Wide Web; (2) spurious links [14] in the known networks, such as the protein–protein interaction networks; (3) future links that may emerge in the evolving networks, such as on-line social networks. It is significant to detect out these links. The detected missing and spurious links can be used to complete, correct and even reconstitute the sampling networks. And the detected future links help us to study the evolution of networks and deal with evolving networks.

So far, many link prediction algorithms based on the structure information of the networks have been presented. Most of them belong to the similarity-based prediction algorithms which depend a lot on the proximity called similarity between two nodes. These methods are proposed under an important assumption that the likelihood of the existence of a link between two nodes is proportional to the similarity between them [15]. The Common Neighbors (*CN*) [16] was a classic similarity-based link prediction index, which defined the node similarity as the number of common neighbors between the two nodes. Based on common neighbors, some indices were proposed by considering the effects caused by different nodes, such as the Adamic–Adar (*AA*) index [17] that weakened the contribution of large-degree nodes and the Resource Allocation (*RA*) index [18]. There are also many similarity-based indices depending on path information and random walk [19–21]. Another type of link prediction methods is based on maximum likelihood analysis [14,22,23], which includes Hierarchical Structure Model (*HSM*) and Stochastic Block Model (*SBM*) [14]. *HSM* took advantage of hierarchical structure information to predict the missing links in the partly known networks with high precision. *SBM* could not only predict out the missing links but also correct the spurious links, such as the spurious links in the protein–protein interaction network. In generally, the accuracy of the methods based on maximum likelihood analysis is higher than that of similarity-based methods for most networks [9]. There are some other link prediction methods that are proposed according to different ideas, such as the probability model [24], machine learning [25], etc. In addition to undirected and unweighted networks, researchers also studied the link prediction in weighted networks [26–28], directed networks [29–31] and bipartite networks [9].

Some recent works offered their studies on link prediction from new perspectives. Zhou et al. [32] studied the predictability of the networks and proposed the structure consistence index to quantify the extent that the network could be predicted, and further presented the structure perturbation method to predict the missing and noise links in the network. Soundarajan and Hopcroft [33] found that the community structure [34] could be used to improve the precision of the link prediction indices based on similarity, and the improvement was obvious for some indices, especially the *RA* index. Liu et al. [35] studied the relationship between the community structure and link prediction. The study of hyperbolic geometry underlying complex network was also found to be helpful in the study of link prediction. Based on hyperbolic geometry of networks, Papadopoulos et al. [36] proposed *HyperMap* method to map target networks into hyperbolic space, then used the coordinates of nodes to predict missing links in the network and obtained highly accurate result.

However, there are still some problems existing among the existing link prediction algorithms. One is that the accuracy of the similarity-based prediction algorithms mainly depends on the matching degree between the definition of the node similarity and the structure characteristics of the target networks [9]. These methods are always hard to deal with the sparse networks without abundant structure information. The other problem is how to handle the huge size of real systems in acceptable time [6,21], because the time complexity of the major algorithms with high accuracy is usually high. Generally, the methods which use global structure information can obtain higher prediction accuracy. However, the computation with global information increases the time complexity vastly, for example, the time complexity of *HyperMap* is $O(n^3)$, where n is the total number of nodes in the network. Therefore, it is a challenge to take advantage of more structure information of the networks to design a fast and accurate link prediction algorithm for sparse and huge networks.

This paper presents a simple and fast method for Link Prediction with Community Structure (*LPCS*) based on hyperbolic mapping. Different from the existing link prediction methods, to utilize the global structure information of network, *LPCS* attempts to deal with the network from an overall perspective. *LPCS* considers the combination of the global information, the meso-level structure known as community structure and the hyperbolic geometry of complex network. It firstly takes advantage of community information to map a target network into hyperbolic space, and then predicts missing links in the network according to the hyperbolic distance between two nodes. Meanwhile, *LPCS* improves the prediction accuracy by further utilizing the hierarchical organization of community structure. It is an innovation that *LPCS* uses community information as the heuristic information to map the network, which greatly decreases the time complexity of mapping. The way that deals with the network from an overall perspective may provide a new viewpoint for the study of link prediction.

The rest of the paper is organized as follows. Section 2 presents the *LPCS* method. Section 3 validates the mapping accuracy and link prediction performance of the *LPCS* method on generated networks and real networks. In Section 4 we analyze and discuss the experimental results, and in Section 5 we conclude the paper.