Chaos, Solitons and Fractals 78 (2015) 140–147

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

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

Bi-scale link prediction on networks☆

Enming Dong*, Jianping Li, Zheng Xie, Ning Wu

Department of Mathematics and System Science, College of Science, National University of Defense Technology, Changsha 410073, China

ARTICLE INFO

Article history:
Received 13 January 2015
Accepted 12 July 2015
Available online 1 September 2015

Keywords:
Complex networks
Link prediction
Convex nonnegative matrix factorization
Resource allocation

ABSTRACT

Link prediction is important for inferring interactions among members in incomplete networks. For a given snapshot of network by sparse sampling, most link prediction methods only consider one scale information, like global or local information, and it is hard to combine them together. A probabilistic model is established to give a theoretical guarantee of the information combinations. Meanwhile a bi-scale method is proposed to combine the information of microscale (neighbors) and mesoscale (communities) in the observed networks. Experiments on several social networks demonstrate that the approach always outperforms local information based methods, and it is faster than the global methods with competitive results.

1. Introduction

As a fundamental problem in the network research works, link prediction attempts to estimate the likelihood of links between nodes, based on the observed links and the property of nodes. A large number of link prediction methods have already been proposed in recent years. Many of them assume that two nodes are more likely to be connected if they are similar. The similarities of nodes can be measured by the different scales information in networks [1].

Considering the globle information(macroscale information), the Katz method uses the ensemble of all paths [2] of the network topology. And Hierarchical random graph(HRG) [3] is based on maximum likelihood estimation of the macroscale structure–latent hierarchical tree in network. Although those global similarity approaches make good performance, they carry high computational complexity. Considering local information (microscale information), a class of predicting methods are proposed based on the idea that two nodes are more likely to form a link if they share more neighbors [5], such as Common Neighbors(CN) [6], Resource Allocation(RA) [17], and Adamic-Adar. If the observed data are sampled sufficiently, those methods are fast and effective for predicting links in social networks [6]. Mesoscale information such as clusterings and communities is a significant feature of the networks. Methods based on clusterings like Stochastic Block Model(BM) [4] and approaches which draw attention to community structure [7,8] can improve the prediction accuracy in a certain range.

Link prediction essentially can be regarded as a completion of adjacency matrix of the network. Low rank approximation is a type of such techniques that generates some of the structures in the adjacent matrix of network with a highly simplified representation. Prediction based on the reconstructed matrix instead of the original one can get better results [9]. Dealing with large networks, Cluster Low Rank Approximation(CLRA) [10,11] using the clusterings of nodes can give a quick low rank approximation. Combining CLRA and hierarchical clustering, a multi-scale predicting method is proposed to handle massive networks in the low rank matrix [12]. However, it is hard for low rank approximation to reveal the local structure when the network structure is complex. Chen and Zhang [13] treats the link prediction as a problem of matrix denoising, and propose a marginalized denoising model. In our recent paper, we propose a convex

☆ This work is partially supported by NSFC (No. 11001237) NUDT Preparing Research Project (JC110204)

* Corresponding author. Tel.: +4407704348784.
E-mail address: dongenming@gmail.com (E. Dong).

http://dx.doi.org/10.1016/j.chaos.2015.07.014
0960-0779/

nonnegative matrix factorization method, however, it only consider the information of clusterings [14].

Methods mentioned above based on single scale information are efficient, but still have limitations. For example, if the observed network is sparse, the number of common neighbors for arbitrary pairs of nodes tends to be zero. So it is hard to find the missing links by those approaches. Prediction using communities or mesoscale information overcome this drawbacks to some extent [3]. However, they are uncompetitive with the microscale ones when dealing with the dense networks. One reason is the resolution limitation of communities detection methods. For example, the Multilevel algorithm for modularity clustering [15] fails to resolve small communities. The similarities of nodes based solely on the network structure cannot be expressed by big communities sufficiently, but they can be revealed by common neighbors. In all, networks are likely to be driven by multiple mechanisms, and combinations of methods in different scales has already shown better efficiencies [12,16]. Since information of different scales can all contribute to the prediction results, it is a natural idea to consider combining them together, however if we simply combine two methods together, we may not always get a better result, for the errors of the two methods may be strengthened.

Based on statistic inference, we establish a probabilistic model to give a theoretical guarantee of a bi-scale combination. And a bi-scale model(BSM) is proposed to inherit the merits of both single scale approaches. In microscale view, we prove that the model can be treated exactly as a probabilistic explanation for the local method RA. In mesoscale view, we show that the model can be mapped directly onto calculating matrix factorization for the blocks of the networks' adjacency matrices, which allows us to apply any non-negative factorizations. In this paper, we use the Convex Nonnegative Matrix Factorization(CNMF) [18], where convex means a convex constraint on the results of the factorization. Experiments on several networks show that BSM outperforms single scale methods. Furthermore, we find that BSM is competitive with some classical methods, such as BM and HRG, with a less computational complexity.

2. Model

In this section, we propose our Bi-scale method and use a statistic inference model to give a theoretical guarantee. The statistic inference model we concerned is an underlying probability model, either because the observed network is the result of a stochastic process, or because the sampling is uncertain. Commonly the observed network of N nodes can be represented by an $N \times N$ adjacency matrix A^o. In our model We re-scale A^o_{ij} by $A^o_{ij} \leftarrow A^o_{ij}/\sum_{ij} A^o_{ij}$. With this stochastic normalization, $\sum_{ij} A^o_{ij} = 1$. The entity A^o_{ij} can be thought as a joint occurrence probability $P(X = i, X = j)$, where X is a variable for the index of nodes.

Nodes in real data are often organized into clusters and the probability of a link between two nodes depends on the groups containing them. However these clusters memberships are still unknown to us. In the language of statistical inference, they are latent variable. Assuming each cluster is a combination of objects, the joint occurrence probability can be factorized by Bayes' formula as follows:

$$P(X = i, X = j) = \sum_{l=1}^{L} P(X = i, X = j|C = l)P(C = l) = \sum_{l=1}^{L} P(X = i, C = l)P(X = j|C = l), \quad (1)$$

where C is the variable for the index of clusters. Here, we assume that the random variables X are conditional independent for a given C.

Under the preparation above, the bi-scale model can be divided into two parts using microscale and mesoscale information respectively. Scale can be determined by the sizes of latent clusters. For microscale information, we can treat each node as a cluster. And if clusters contain more than one nodes, they can release mesoscale information. For each scale, the proximate measure can be computed as a matrix. Each component of the matrix is a result of joint occurrence probability $P(X = i, X = j)$. This gives two proximity measures for each link, and they can be combined together to make final predictions.

In order to consider the microscale part, we assume that each node is a cluster, so there are N-clusters. Eq. (1) can be written as

$$P(X = i, X = j) = \sum_{l=1}^{N} \frac{P(X = i, C = l)P(X = j, C = l)}{P(C = l)} = \sum_{l=1}^{N} \frac{P(X = i, C = l)P(X = j, C = l)}{\sum_{k=1}^{N} P(X = K, C = l)}. \quad (2)$$

Suppose $P(X = i, C = l) = A^o_{il}$, we can express Eq. (2) as

$$P(X = i, X = j) = \sum_{l=1}^{N} \frac{A^o_{il}A^o_{jl}}{\sum_{k=1}^{K} A^o_{lk}}. \quad (3)$$

Obviously, this measure is symmetrical. and a pair of nodes can transfer some resource to each other, during which their common neighbors can be treated as transmission media. Each medium has a unit of the resource and distributes it to its neighbors averagely. This is exactly the RA index of re-scale matrix A^o.

Latent clusters contain more than one nodes in the mesoscale part. For the reason of interpretability, we treat the adjacent matrix A^o as an object-feature matrices for objects $\{x_i, i = 1, \ldots, N\}$ and features $\{y_j, j = 1, \ldots, N\}$. In order to reduce computability and get better results, we take the strategy of blockwise approximation. Denote $\{\bar{A}^o_k, k = 1, \ldots, K\}$ as the K blocks in A^o finding by fast modularity clustering algorithm [19]. For each block $\bar{A}^o_k$, we extend it to a $N \times N$ matrix by using 0 to complete the matrix and denote it by A^o_k. Then the joint occurrence probability of an object and a feature can be factorized as

$$P(X = i, Y = j) = \sum_{k=1}^{K} P(X = i|Y = j, A = k)P(A = k|Y = j) \times P(Y = j), \quad (4)$$

where X is a variable for the index of objects, Y is a variable for the index of features, A is a variable for the index