



# Graph regularized nonnegative matrix factorization for temporal link prediction in dynamic networks

Xiaoke Ma<sup>a,\*</sup>, Penggang Sun<sup>a,\*</sup>, Yu Wang<sup>b</sup>

<sup>a</sup> School of Computer Science and Technology, Xidian University, No. 2 South Taibai Road, Xi'an, Shaanxi, China

<sup>b</sup> School of Economics and Management, Xidian University, No. 2 South Taibai Road, Xi'an, Shaanxi, China

## HIGHLIGHTS

- Two graph regularized frameworks for temporal link prediction are proposed.
- Graph regularized algorithms have been developed.
- The proposed algorithms outperform state-of-the-art methods without increasing the time complexity.

## ARTICLE INFO

### Article history:

Received 28 August 2017

Received in revised form 30 October 2017

Available online 26 December 2017

### Keywords:

Temporal link prediction

Dynamic networks

Graph regularization

Nonnegative matrix factorization

## ABSTRACT

Many networks derived from society and nature are temporal and incomplete. The temporal link prediction problem in networks is to predict links at time  $T + 1$  based on a given temporal network from time 1 to  $T$ , which is essential to important applications. The current algorithms either predict the temporal links by collapsing the dynamic networks or collapsing features derived from each network, which are criticized for ignoring the connection among slices. To overcome the issue, we propose a novel graph regularized nonnegative matrix factorization algorithm (GrNMF) for the temporal link prediction problem without collapsing the dynamic networks. To obtain the feature for each network from 1 to  $t$ , GrNMF factorizes the matrix associated with networks by setting the rest networks as regularization, which provides a better way to characterize the topological information of temporal links. Then, the GrNMF algorithm collapses the feature matrices to predict temporal links. Compared with state-of-the-art methods, the proposed algorithm exhibits significantly improved accuracy by avoiding the collapse of temporal networks. Experimental results of a number of artificial and real temporal networks illustrate that the proposed method is not only more accurate but also more robust than state-of-the-art approaches.

© 2017 Elsevier B.V. All rights reserved.

## 1. Introduction

Network (also called graph) effectively models and analyzes many complex systems, where each vertex represents an entity, such as a gene, a web user, or a terminal in Internet, and each link denotes an interaction between a pair of vertices. There are various networks such as social networks [1,2], web networks [3] and biological networks [4,5]. Network analysis is a hot topic in modern science with an immediate purpose to discover graph patterns by elucidating the structure–function

\* Corresponding authors.

E-mail addresses: [xkma@xidian.edu.cn](mailto:xkma@xidian.edu.cn) (X. Ma), [psun@mail.xidian.edu.cn](mailto:psun@mail.xidian.edu.cn) (P. Sun).

relationship of the overall systems. For example, a community in social networks corresponds to a group of individuals with the same or similar backgrounds [2].

Thus, the analysis of networks has been extensively studied to extract the graph patterns, which sheds light on the understanding of mechanisms of networks. For instance, Desai et al. [6] constructed the topical hierarchies by analyzing the heterogeneous information networks; Ma et al. [7] discovered the evolving clusters in dynamics networks using the semi-supervised evolutionary nonnegative matrix factorization algorithm (sE-NMF). However, many networks in real world are incomplete because of the limitations of our knowledge, which is a huge obstacle to the application of network analysis. For example, nearly 99% of protein interactions within human remain unknown [8]. Accordingly, link prediction plays a critical role in network analysis because it facilitates the construction and analysis of the overall systems [9–11].

Therefore, great efforts have been devoted to the link prediction since it is of theoretical interest and practical significance for network analysis. The most intuitive strategy is to design experiments to validate the existence of links. However, it is unpractical due to the limitations of finance and technology. For instance, it is costly and time-consuming to validate the protein–protein interactions using the biological experiments [12]. To this end, the computational methods for link prediction based on the known links provide an alternative for the experimental strategy. And, many effective and efficient algorithms have been developed [13–18]. For example, Zhou et al. [13] exploited the local topological information of networks to predict the missing links in order to avoid computing the global information; Lü et al. [14] addressed the link prediction problem in weighted networks using the so-called weak tie effect, indicating that the social phenomenon in networks is also valuable for link prediction; Kunegis et al. [15] designed the spectral clustering algorithm for the link prediction problem. Zhou et al. [17] developed the robust principle component analysis approach for the link prediction and predicted the missing interaction between drugs and targets using sparse learning [18]. More information can be referred to [19–21].

However, these algorithms are solely applicable for static networks. Actually, the social networks are not only incomplete, but also temporal (dynamic) [22], implying that the network structure evolves at various time steps or conditions. For example, in social networks, links often are dynamic changing in response to the behavior of individual's social partners [23]; In disease networks, cancer cell immigration leads to the cancer metastasis, which is critical for the cancer diagnosis and therapy [24]. Compared to static networks, the dynamic networks are more precise in characterizing the complex systems. Thus, the analysis of temporal networks has received considerable attention because the evolutionary patterns provide novel insights into the function–structure relation of the overall system [25].

Predicting links in temporal networks is promising since it is critical for the analysis of dynamic networks, which is the so called *temporal link prediction problem*. Specifically, it predicts links in the network at time  $T + 1$  based on the given temporal networks from time 1 to  $T$ , whereas the *missing link prediction problem* aims at obtaining the missing links in static networks. However, it is highly non-trivial to design effective and efficient algorithms for the prediction of temporal link prediction due to many reasons. The first reason is that it is difficult to extract the features because they must characterize the evolving patterns of dynamic networks. Secondly, the temporal networks pose a great challenge for designing efficient computational algorithms.

Even though it is difficult, great efforts have been devoted to the temporal link prediction problem [26–30]. The current algorithms collapse either the dynamic networks or the features to predict temporal links. For example, Sharan et al. [26] collapsed the dynamic networks to predict temporal links by summing the matrices associated with networks, thereby saving running time by sacrificing accuracy. To fully utilize topology structure, Katz [31] predicted temporal links by counting the number of paths. Matrix decomposition-based algorithms [27,28], such as singular value decomposition (SVD) and eigen-decomposition (ED), have been developed to predict temporal links using low-rank approximation. These algorithms initially collapse temporal networks and then predict temporal links based on the collapsed network, which ignore the critical information hidden in dynamic networks, thereby affecting the performance of algorithms. To avoid collapsing the temporal networks, Acar et al. [27] provided the tensor decomposition (TD) method for the temporal link prediction problem, which dramatically improved the accuracy of algorithms.

Although many algorithms have been successfully devoted to the temporal link prediction problem, they ignore the connection among the dynamic networks, resulting in undesirable performance. Actually, incorporating the evolving information into feature extraction is critical for the temporal network analysis [22,32]. Therefore, extracting the features associated with the dynamic networks is critical for the temporal link prediction. Fortunately, the regularization provides a novel strategy to overcome this issue, where the intrinsic geometrical structure of the data space is integrated. For example, Cai et al. [33] provided the regularized nonnegative matrix factorization (RNMF) for data representation, which significantly improves the performance of NMF. The regularization based algorithms have been widely applied to many fields, such as Alzheimer's disease diagnosis [34] and image process [35].

However, as far as we know, no attempt has been devoted to the temporal link prediction problem using the regularization strategy. Thus, we explore the possibility of improving the performance of algorithms using the graph regularization. The GrNMF algorithm is proposed without collapsing the dynamic networks, which factorizes the network at time  $T$  by setting networks from 1 to  $T - 1$  as a regularizer. It provides a better way to characterize the topological information of temporal links since the connection of networks is incorporated into the objective function.

It is worthwhile to highlight several aspects of the proposed approach here:

- To avoid collapsing the dynamic networks or features, we propose the GrNMF algorithm for the temporal link prediction problem, where the dynamic network is transformed as a regularizer. To further quantify the importance of each slice in the temporal networks, we weight each slice that is incorporated into the objective function of GrNMF.

Download English Version:

<https://daneshyari.com/en/article/7376099>

Download Persian Version:

<https://daneshyari.com/article/7376099>

[Daneshyari.com](https://daneshyari.com)