Pattern Recognition 76 (2018) 228–241

Contents lists available at ScienceDirect

Pattern Recognition

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

# Structural property-aware multilayer network embedding for latent factor analysis

Jie Lu[a,*], Junyu Xuan[a,b], Guangquan Zhang[a], Xiangfeng Luo[b]

[a] *Centre for Artificial Intelligence, Faculty of Engineering and Information Technology, University of Technology Sydney, PO Box 123, Broadway, NSW 2007, Sydney, Australia*

[b] *School of Computer Engineering and Science, Shanghai University, 99 Shangda Road, Shanghai, China*

ARTICLE INFO

*Article history:*
Received 2 December 2016
Revised 28 September 2017
Accepted 4 November 2017
Available online 7 November 2017

*Keywords:*
Multilayer network
Network embedding
Nonnegative matrix factorization

ABSTRACT

Multilayer network is a structure commonly used to describe and model the complex interaction between sets of entities/nodes. A three-layer example is the author-paper-word structure in which authors are linked by co-author relation, papers are linked by citation relation, and words are linked by semantic relation. Network embedding, which aims to project the nodes in the network into a relatively low-dimensional space for latent factor analysis, has recently emerged as an effective method for a variety of network-based tasks, such as collaborative filtering and link prediction. However, existing studies of network embedding both focus on the single-layer network and overlook the structural properties of the network, e.g., the degree distribution and communities, which are significant for node characterization, such as the preferences of users in a social network. In this paper, we propose four multilayer network embedding algorithms based on Nonnegative Matrix Factorization (NMF) with consideration given to four structural properties: whole network (NNMF), community (CNMF), degree distribution (DNMF), and max spanning tree (TNMF). Experiments on synthetic data show that the proposed algorithms are able to preserve the desired structural properties as designed. Experiments on real-world data show that multilayer network embedding improves the accuracy of document clustering and recommendation, and the four embedding algorithms corresponding to the four structural properties demonstrate the differences in performance on these two tasks. These results can be directly used in document clustering and recommendation systems.



## 1. Introduction

Multilayer network [1] is a structure commonly used to describe and model the complex interaction between sets of entities/nodes. The structure has attracted the attention of researchers from many areas, such as computer scientists, sociologists, physicists, and biologists, due to its pervasiveness. As a result, research into multilayer network has become a multidisciplinary area of study. To date, many types of multilayer network with a variety of structures and names have been developed in the literature [1–3]. Multilayer network, as defined in this study, is composed of several homogeneous networks in multiple layers and the nodes in different layers may have external links across the layers, as illustrated in Fig. 1. One example, in the text mining area, is the author-paper-keyword structure shown in Fig. 2. This structure is a multilayer network because it is composed of three layered networks (i.e., author social network, paper citation network and keyword co-occurrence network) and the nodes across networks are also linked (i.e., an author and a paper are linked if this author writes this paper, and a paper and a word are linked if this paper contains this word). In recommender systems, a multilayer network is composed of tag-user-movie mapping relations with tag similarity network, user social network and movie similarity network, as shown in Fig. 3. Multilayer network would also be a good choice for big data modelling because there are complex interactions between multiple sources or attributes due to the Variety property of big data [4]. Hence, it is crucial and urgent to develop more effective analytic tools for multilayer network to obtain better understanding and improving behaviour prediction of its underlying complex systems.

Network embedding has recently emerged as an effective method for a variety of network-based tasks, such as collaborative filtering and link prediction. Its basic idea is to project the nodes in the network into a relatively low-dimensional space and provide each node with a new vector-based representation. It is

* Corresponding author.
*E-mail addresses:* jie.lu@uts.edu.au, jielu@it.uts.edu.au (J. Lu), junyu.xuan@uts.edu.au (J. Xuan), guangquan.zhang@uts.edu.au (G. Zhang), luoxf@shu.edu.cn (X. Luo).

https://doi.org/10.1016/j.patcog.2017.11.004
0031-3203/

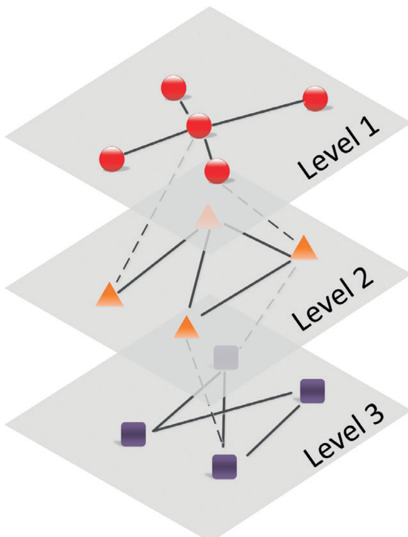


**Fig. 1.** Multilayer network.

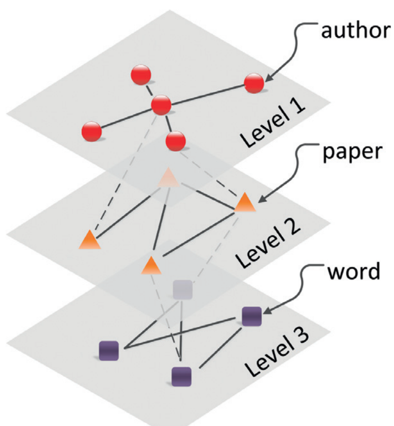


**Fig. 2.** An instance in text mining.

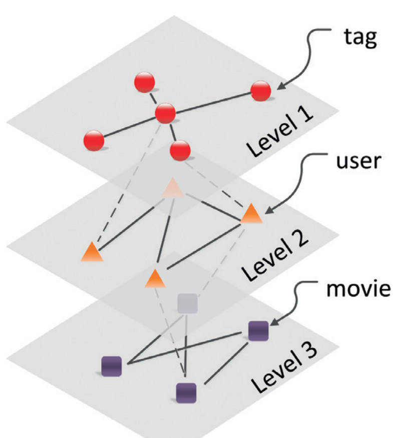


**Fig. 3.** An instance in recommender systems.

commonly believed that this new representation is not only more concise but also more innate, because it is expected to preserve the important characteristics of the network. One simple example is that each person is given a vector-based representation after a social network has been embedded, and two persons will be recommended as friends based on the similarity between their new vector-based representations. A similar example is that documents could be more accurately clustered using the new vector-based representation instead of the original word-based representation. A number of methods have been proposed in the literature for network embedding, including matrix factorization [5,6], random walk [7–9], deep learning [10,11], and so on. However, all these state-of-the-art methods are designed for a single-layer network and, importantly, do not consider the structural properties, i.e., community, degree distribution, and max spanning tree, during network embedding. Structural properties are the significant statistical properties of complex networks, and sometimes they have a greater ability to characterize the nature of a node than the whole network. Unfortunately, these structural properties are overlooked by existing network embedding methods.

In this paper, we propose four multilayer network embedding algorithms based on Nonnegative Matrix Factorization (NMF) [12] with considerations given to four different structural properties: whole network (NNMF), community (CNMF), degree distribution (DNMF) and max spanning tree (TNMF). Four objective functions are carefully designed to preserve the desired structural properties along with the multilayer network embedding. To optimize each objective function, the corresponding update rules are introduced. Experiments on synthetic data show that the designed algorithms have the ability to preserve the desired structural properties. To show the usefulness of these algorithms, two real-world tasks, document clustering and recommendation, are carried out. The results show that the proposed algorithms perform better than traditional NMF and other algorithms in achieving clustering and recommendation accuracy. A natural problem is to determine the difference between these structural properties in their impact on the real-world tasks, e.g., recommendation or clustering performance. To evaluate these differences, we conduct extensive experiments to compare the performance of each on two real-world tasks. As the experimental results show, we can achieve better performance by preserving the structural properties. We also compare the performance when different structural properties are retained in these tasks.

This paper makes the following contributions:

1. Four structural properties-aware multilayer network embedding algorithms based on nonnegative matrix factorization are proposed with consideration given to four significant structural properties.
2. Extensive experiments are conducted to show the ability of the proposed multilayer network embedding algorithms on the preservation of structural properties, and to compare the performances of the four designed algorithms on two real-world tasks.

The rest of this paper is organized as follows. Section 2 reviews related work. The problem is formally defined in Section 3. Our algorithms for multilayer network embedding are proposed in Section 4. Experiments on synthetic data and real-world data are conducted in Section 5. Lastly, Section 6 concludes the study and discusses future work.

## 2. Related work

Since our motivation is to use complex network structural properties as constraints for nonnegative matrix factorization, this section is composed of two parts: (1) we will discuss elementary introductions to, and research on, the structural properties of a complex network, and (2) we will discuss recent works on network-related factorization models or algorithms.