Physica A 495 (2018) 324–335

Contents lists available at ScienceDirect

Physica A

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

Attributed community mining using joint general non-negative matrix factorization with graph Laplacian

Zigang Chen [a,b], Lixiang Li [a,*], Haipeng Peng [a], Yuhong Liu [c], Yixian Yang [a]

[a] Information Security Center, State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, P.O. Box 145, Beijing 100876, China
[b] College of Computer and Information Engineering, Nanyang Institute of Technology, Nanyang 473004, China
[c] Department of Computer Engineering, Santa Clara University, Santa Clara, CA 95053, USA

ARTICLE INFO

Article history:
Received 19 October 2017
Available online 22 December 2017

Keywords:
Community mining
Complex social networks
Graph clustering
General non-negative matrix factorization (GNMF)

ABSTRACT

Community mining for complex social networks with link and attribute information plays an important role according to different application needs. In this paper, based on our proposed general non-negative matrix factorization (GNMF) algorithm without dimension matching constraints in our previous work, we propose the joint GNMF with graph Laplacian (LJGNMF) to implement community mining of complex social networks with link and attribute information according to different application needs. Theoretical derivation result shows that the proposed LJGNMF is fully compatible with previous methods of integrating traditional NMF and symmetric NMF. In addition, experimental results show that the proposed LJGNMF can meet the needs of different community minings by adjusting its parameters, and the effect is better than traditional NMF in the community vertices attributes entropy.

1. Introduction

Social networks are expanding quickly with the emergence and rapid proliferation of social applications and media. Community mining for complex social networks plays an important role in the studies of social network topologies and their functional characteristics, as well as online social networks' analysis and forecasting [1,2]. A complex social network contains not only link information, but also attribute information, where the attributes describe the features of a vertex (person), and the topological structure represents the relationships among a group of vertices. Therefore, it is also called attributed graph [3].

For instance, Lada Adamic and Natalie Glance compiled the political blogosphere Feb. 2005 data [4]. In the dataset, there is not only link information between vertices, but also a "value" attribute with "liberal" or "conservative" to indicate the political leaning of each vertex. Meanwhile, each vertex in the dataset has one or more of these seven "source" attributes with "Blogarama", "BlogCatalog", "LeftyDirectory", "eTalkingHead", "CampaignLine", "LabeledManually", and "BlogPulse", and these attributes can describe the source features of the vertex. Normally there are three respects to study the community mining, such as only vertex link information, only the attribute information, and both the link and attribute information of vertices according to different application needs.

* Corresponding author.
E-mail address: li_lixiang2006@163.com (L. Li).

https://doi.org/10.1016/j.physa.2017.12.038

A complex network that contains both link information and attribute information is denoted as $G = (V, E, A)$, where $V = \{v_1, v_2, \ldots, v_t\}$ is the set of vertices, and $|V| = t$ is the number of vertices. E is the set of edges; and $A = \{a_1, a_2, \ldots, a_r\}$ used to represent the properties of every vertex, is the set of attributes associated with vertices in V. The purpose of community mining is to find $\{V_1, V_2, \ldots, V_p\}$, where p is the total number of communities in the community mining, and $V_i \in \{V_1, V_2, \ldots, V_p\}$ is a community.

Matrix factorization has recently been adopted by many researchers to perform community mining. For example, non-negative matrix factorization (NMF) and symmetric non-negative matrix factorization adopted by [5–10] are shown to be able to effectively extract the community structure of complex social networks.

However, most of the matrix factorization based community mining approaches consider only the link information or the attribute information of social networks, which often cannot achieve the desired community mining results. Then community mining methods integrating link and attribute information have been studied. Representative works are the SACluster approach proposed by Cheng et al. [11] and the BAGC approach proposed by Xu et al. [3] However, such methods often need to design different models to process link information and attribute differently. Then the community mining process is divided into multiple stages, which is too complex to be implemented for the practical use.

Link and attribute information based on NMF (LANMF) approach proposed by He et al. [12] can integrate both link and attribute informations to mine community in complex networks. Furthermore, experimental results showed that the quality of community mining using LANMF was better than SACluster approach and BAGC approach, and it can mine community directly and effectively.

However, the approach [12,13] similar to LANMF fixed the ratio of link information and attribute information, which cannot adjust the proportional parameters according to different application needs.

Our work in this paper is to propose a novel approach that considers three cases, only the link information, only the attribute information, and both of them to mine communities according to different application needs. Based on the general NMF (GNMF) proposed in our previous work, we aim to propose a novel GNMF attached with graph Laplacian to improve the quality of clustering when mining attributed communities.

Our works are as follows:

(1) Based on the GNMF proposed in our previous work, we propose a joint GNMF with graph Laplacian (LJGNMF) to mine the communities in social networks with the link and attribute information of vertices according to different application needs. Theoretical derivation result shows that the proposed LJGNMF is fully compatible with previous methods of integrating traditional NMF and symmetric NMF.
(2) We conduct community mining tasks on three social networks to show the performance evaluations of the proposed LJGNMF.

The remainder of this paper is organized as follows. Section 2 introduces the traditional NMF and the GNMF proposed in our previous work. Section 3 presents the proposed JGNMF and LJGNMF methods, and their algorithms analysis for attributed graphs. Community mining tasks on three social networks are used for the performance evaluations of the proposed LJGNMF algorithm, and the details are shown in Section 4. Section 5 concludes the paper.

2. Preliminaries

2.1. Non-negative matrix factorization

Non-negative Matrix Factorization (NMF) [14] is able to learn parts of faces and semantic features of the text. NMF factorizes an input non-negative matrix into two non-negative matrices of lower rank.

Given a matrix $X \in R_+^{s\times t}$, NMF is trying to find two non-negative matrices $W \in R_+^{m\times n}$ and $H \in R_+^{p\times q}$, so that

$$X_+^{s\times t} \approx W_+^{m\times n} H_+^{p\times q}, \tag{1}$$

where $s = m$, $n = p$, and $t = q$.

In general, the factorized matrices W and H are known as the basis matrix and the coefficient matrix, respectively.

In order to obtain these two non-negative matrices in Eq. (1), we can quantify the quality of the approximation by using a cost function with Euclidean distance, and the problem turns to minimize the following objective function.

$$\begin{aligned} F_{NMF}(W, H) &= \frac{1}{2}\|X - WH\|_F^2 \\ &= \frac{1}{2}\sum_{i=1}^{s}\sum_{j=1}^{t}[X_{ij} - (WH)_{ij}]^2, \end{aligned} \tag{2}$$

where $\|\cdot\|$ is the matrix Frobenius norm.