IEICE TRANS. INF. & SYST., VOL.E100–D, NO.6 JUNE 2017

PAPER

Community Discovery on Multi-View Social Networks via Joint Regularized Nonnegative Matrix Triple Factorization

Liangliang ZHANG[†a)], Longqi YANG[†], Yong GONG[†], *Nonmembers*, Zhisong PAN[†], *Member*, Yanyan ZHANG[†], *and* Guyu HU[†b)], *Nonmembers*

SUMMARY In multi-view social networks field, a flexible Nonnegative Matrix Factorization (NMF) based framework is proposed which integrates multi-view relation data and feature data for community discovery. Benefit with a relaxed pairwise regularization and a novel orthogonal regularization, it outperforms the-state-of-art algorithms on five real-world datasets in terms of accuracy and NMI.

key words: *data mining, community discovery, social network, nonnegative matrix factorization*

1. Introduction

In recent years, research on community discovery has received considerable attention in the data mining field [1]–[3], particularly in the area of social media.

Finding a community in a social network corresponds to identifying a set of nodes such that they interact with each other more frequently than with those nodes outside the group.

Community discovery can facilitate other social computing tasks and is applied in many real-world applications. For instance, the grouping of customers with similar interests in social media leads to efficient recommendations that expose customers to a wide range of relevant items to enhance transaction success rates. Communities can also be used to compress an extremely large network, thereby resulting in a smaller network. In other words, problem solving is accomplished at the group level rather than at the node level. Similarly, an extremely large network can be visualized at different resolutions, offering an intuitive solution for network analysis and navigation.

However, community discovery in social networks with heterogeneous entities and interactions is still challenging.

Take Twitter as an example. A network in Twitter can encompass entities such as users, tweets and lists. Multiple interactions can exist between users: a user x may follow, retweet or mention another user y. Thus, a variety of interactions exist between the same set of users in a network. Each type of interaction between users forms a view of the network, i.e., relation view; thus, the relation views of Twitter could be named follow, retweet and mention. Moreover, the entities of Twitter such as lists and tweets form feature views, which may be named lists and tweets, respectively. These distinctive types of entities or interactions form multi-type, multi-view networks in social media. With a heterogeneous network, our goal is to determine the hidden communities through integrating multi-view relations and features.

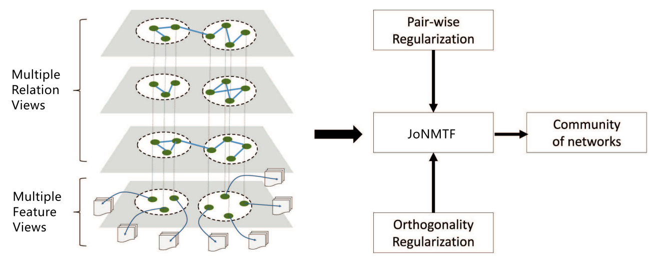


Fig. 1 JoNMTF framework

Nonnegative matrix factorization (NMF) has a solid theoretical basis, good interpretability and high computational efficiency, and it has recently been widely applied in community discovery [4]–[7].

In this work, we propose a framework named joint-regularized nonnegative matrix triple factorization (JoNMTF), which extends NMF to integrate two distinct types of data sources, as shown in Fig. 1. JoNMTF is discussed in Sect. 4.

The main contributions of this paper are as follows:

- Proposing the flexible JoNMTF framework that combines multi-type, multi-view data sources (i.e., relation views and feature views);
- Introducing two novel regularizations (i.e., pairwise regularization and orthogonal regularization) to extend NMF for multi-type, multi-view community discovery;
- Applying JoNMTF to five real-world datasets and demonstrating the effectiveness of these solutions for community discovery.

Section 2 surveys the works related to multi-view clustering. In Sect. 3, we formalize our research problem and study the problem in a preliminary study on multi-view methods. Section 4 outlines the proposed joint-regularized NMTF framework and discusses its correctness. In Sect. 5, we evaluate our proposed method. The paper is concluded in Sect. 6.

Manuscript received January 5, 2017.
Manuscript revised February 28, 2017.
Manuscript publicized March 21, 2017.

[†]The authors are with the College of Command Information System, PLA University of Science and Technology, Nanjing 210007, China.
a) E-mail: vermouthlove@hotmail.com
b) E-mail: guyuhu@189.com (Corresponding author)

DOI: 10.1587/transinf.2017EDP7004

2. Related Works

We first review the literature on the general problem of single-view NMF-based community discovery. We then review works on multi-view community discovery, which represent a collection of methods of which our specific proposal of JoNMTF is an instance.

The NMF-based methods for community discovery on social networks can be classified as single-view community discovery and multi-view community discovery according to the amount of views (i.e., networks).

2.1 Single-View NMF-Based Community Discovery

NMF can be traced back to the 1970s (notes from G. Golub) and was extensively studied by Paatero [8]. The work of Lee and Seung [9] brought considerable attention to NMF in the fields of machine learning and data mining. Most applications make use of the clustering aspect of NMF, which is inspired by the work of Lee and Seung.

In general, the input data for single-view NMF methods include relation data, feature data (i.e., content data), and so forth.

For feature data, normal NMF factorizes a nonnegative input feature data matrix $X \in \mathbb{R}_+^{p\times n}$, as shown in Eq. (1),

$$X \approx FG^T, \tag{1}$$

where $F \in \mathbb{R}_+^{p\times k}$ and $G \in \mathbb{R}_+^{n\times k}$.

Some preliminary work was conducted in 2006; Ding et al. [10] presented Non-negative Matrix Tri-Factorization (NMTF), as shown in Eq. (2),

$$\min_{F\geq 0, G\geq 0, S\geq 0} \|X - FSG^T\|, \tag{2}$$

where $\|\cdot\|$ denotes the squared sum of all elements in the matrix, $F \in \mathbb{R}_+^{p\times k}$, $S \in \mathbb{R}_+^{k\times l}$, and $G \in \mathbb{R}_+^{n\times l}$. To simultaneously cluster the rows and columns of the input data matrix, F provides row clusters, G provides column clusters, and S provides additional degrees of freedom such that the low-rank matrix representation remains accurate. More precisely, we solve Eq. (3)

$$\min_{F\geq 0, G\geq 0, S\geq 0} \|X - FSG^T\|, \quad s.t. F^TF = I, G^TG = I. \tag{3}$$

The value of $X - FSG^T$ with bi-orthogonal constraints is smaller than that without bi-orthogonal constraints. We typically set $k = l$. If the orthogonal constraint is only applied on G, namely, one-sided ONMF, then it makes no difference by setting $F \leftarrow FS$, and NMTF can be transformed to $X \approx FG^T$, which is a normal NMF 1. This form Eq. (3) gives a good framework for simultaneously clustering the rows and columns of X.

For relation data, when the input X is a symmetric matrix of pairwise similarities, i.e., $X_{ij} = X_{ji}$, it is a special case of NMTF. Ding et al. [10] proposed the symmetric NMTF (symNMTF) method to solve this problem by setting $X = X^T = W$ and $F = G = H$ as

$$\min_{H\geq 0, S\geq 0} J_{symNMTF} = \|W - HSH^T\|, \quad s.t. H^TH = I. \tag{4}$$

Although these NMF-based algorithms exhibit high interpretability, they are limited for community discovery in single-view networks. We will discuss some NMF-based methods for multi-view community discovery in the next section.

2.2 Multi-View NMF-Based Community Discovery

Single-view NMF focused on only one aspect of a network; it performs well when the data source is complete and accurate. In reality, the poor performance of single-view NMF is due to missing data, data errors and so on. Integrating multiple types of measurements for the same users allows us to gain a deeper understanding of the data and refine the clustering. It is intuitive to integrate multiple-view information to obtain a more stable and accurate partition.

Multi-view NMF-based community discovery methods can be grouped into three categories [6] - early, immediate and late integration - according to when the information from single views are integrated.

- **Early integration.** In these approaches, the individual views are directly integrated into a unified view before learning. Then, the classical clustering algorithm can be performed on the views. Greene and Cunningham [11] proposed a scheme to effectively solve the multi-view problem. Their proposed scheme unifies multi-view data into a single graph, and the classical clustering method k-nearest neighbor was applied to this graph. This scheme could handle both relation and feature data. The main weakness of this study is that information may be lost at the integration step. However, verifying the correctness and accuracy of the unified graph based only on clustering performance is difficult.
- **Late Integration.** In these approaches, the clustering algorithm is first conducted on the individual views. Then, a strategy is adopted to combine the individual results. Hindle et al. [12] generated the clustering of different views by treating the optimal clustering as hidden factors. PLSA was adopted to solve this problem.
- **Immediate integration.** In these approaches, multiple views are fused during the clustering process. Jing et al. [6] proposed a multi-view clustering algorithm called MultiNMF, which formulates a joint matrix factorization process with the constraint. MultiNMF pushes the clustering solution of each view toward a common consensus, which may be too strict in some fields. As the regularization parameter increases, the membership matrices become similar, which may be not a good approximation of real data. He et al. [7] proposed CoNMF, the pair-wise regularization which relaxed the regularization of MultiNMF. CoNMF is reported has better performance. Hidru and Goldenberg [13] proposed a graph-regularized multi-view

NMF-based method named EquiNMF. Due to the various different network types (e.g., undirected, directed and compound networks), Wang et al. proposed a series of NMF-based algorithms (SNMF, ANMF, and JNMF) for different networks [5]. JNMF is designed for compound networks; it combines user-user, user-movie and movie-movie information into one model.

3. Preliminaries

Before describing JoNMTF, we discuss some necessary preliminaries. We first present a formal problem statement for multi-type, multi-view community discovery and then introduce the evaluation criteria. We then conduct an initial study on a multi-view community discovery method - CoNMF [7]. Inspired by CoNMF and the advantages of symmetric NMTF, we propose a community discovery method called CoNMTF for multi-view relation data.

3.1 Problem Formulation

In this paper, we define a multi-type, multi-view social network as a collection of relation views and feature views. We investigate how relation data and feature data are best used to assist community discovery.

Input: For relation views, G consists of n_r view networks. Formally, we are given n_r view relation data denoted as $G = \{G^1, \ldots, G^{n_r}\} \in \mathbb{R}_+^{n\times n}$, where n denotes the total number of users. For view i, $G^i = (V^i, E^i)$, where V^i and E^i are the set of nodes and set of edges, respectively. A node may appear in one or multiple views.

For feature views, X consists of n_f data matrices for each view network. Define $N = n_r + n_f$. Let $X = \{X^{n_r+1}, \ldots, X^N\}$ be n_f views of a set of V_i data points such that each row of X_i represents a user and each column represents an attribute.

Output: $H \in \mathbb{R}_+^{n\times k}$ denotes the community membership matrix, i.e., indicator matrix. The i-th row of H not only indicates the degree of attribution of the nodes in all communities but also depicts the distribution structure and characteristics of nodes in k-dimensional space.

Thus, the target of community discovery in multi-view social networks is to integrate both relation views and feature views and then factorize into H.

3.2 Evaluation Metrics

Clustering accuracy (AC) and the normalized mutual information metric (NMI) are used to measure the community discovery performance [14].

Let l_i and α_i be the cluster label and the ground truth label, respectively. The AC is defined as follows:

$$AC = \frac{\sum_{i=1}^{n} \delta(\alpha_i, map(l_i))}{n}, \tag{5}$$

where n denotes the total number of nodes; $\delta(x, y)$ is the delta function, which equals one if $x = y$ and equals zero otherwise; and map(l_i) is the mapping function that maps each cluster label l_i to the equivalent label from the ground truth. The best mapping can be found by using the Kuhn-Munkres algorithm.

Moreover, given two sets of communities C and C', their mutual information metric $MI(C, C')$ is defined as follows:

$$MI(C, C') = \sum_{c_i \in C, c_j' \in C'} p(c_i, c_j') \odot \log_2 \frac{p(c_i, c_j')}{p(c_i)p(c_j')}, \tag{6}$$

where $p(c_i)$ and $p(c_j')$ denote the probabilities that a user belongs to community c_i or c_j', respectively. $p(c_i, c_j')$ denotes the joint probability that user belongs to c_i and c_j'.

For example, the indicator matrix is $H \in \mathbb{R}_+^{n\times k}$, where n is the number of users and k is the number of communities. Considering the hard (single-assignment) clustering problem, we take the most likely cluster in the soft assignment to yield a hard assignment.

For the i-th user, H_{ij} denotes the probability that the i-th user belongs to the j-th cluster. We define $A(i) = \arg\max_j H_{ij}$, which means that the i-th user is assigned to the j-th cluster.

We now discuss the calculation of $p(c_i)$, $p(c_j')$ and $p(c_i, c_j')$.

$$p(c_i) = \frac{1}{N} \sum_{x=1}^{N} \delta(c_i, (A(x))), \tag{7}$$

where $I_c(x)$ is the indicator function, which equals one if $x = c$ and equals zero otherwise. The ground truth of clustering is denoted as GT such that

$$p(c_j') = \frac{1}{N} \sum_{x=1}^{N} \delta(c_j', (GT(x))). \tag{8}$$

$$p(c_i, c_j') = \frac{1}{N} \sum_{x=1}^{N} \delta(c_i, (A(x))) \times \delta(c_j', (GT(x))). \tag{9}$$

$H(C)$ and $H(C')$ are the entropies of C and C', respectively. The normalized metric (NMI) is:

$$NMI(C, C') = \frac{MI(C, C')}{\max(H(C), H(C'))}, \tag{10}$$

where the $\odot$ in this matrix context denotes element-wise multiplication.

3.3 Co-Regularized Non-Negative Matrix Factorization

He et al. [7] proposed an approach named Co-regularized Non-negative Matrix Factorization (CoNMF), which is for multi-view community discovery, and it used a pairwise constraint to force each pair of indicator matrices H to be similar.

Given n_v view data denoted as $\{X^1, \ldots, X^{n_v}\}$, each view is factorized as $X^i \approx H^i V^i$, where H^i have the same dimensions $n \times k$, and V^i have dimensions $k \times m_i$, differing per view.

The objective function of CoNMF is Eq. (11)

$$J_{CoNMF} = \sum_{i=1}^{n_v} \theta_i \|X^i - H^i V^i\| + R_p \quad s.t. H^i \geq 0, V^i \geq 0, \tag{11}$$

where $R_p = \sum_{i=1}^{n_v} \sum_{j=1}^{n_v} \|H^i - H^j\|$ is the pairwise regularization that relaxed the regularization of MultiNMF [6].

3.4 Community Discovery for Multi-View Relation Data

Consider an undirected network, the relation data can be formulated as a symmetric matrix of pairwise user similarities, we denote it as $W \in \mathbb{R}_+^{n \times n}$.

Symmetric NMTF is reported to have good performance on symmetric data [15], [16]. It is formulated as $W \approx HSH^T s.t. H^T H = I$, where $W \in \mathbb{R}_+^{n \times n}$, $H \in \mathbb{R}_+^{n \times k}$, and $S \in \mathbb{R}_+^{k \times k}$.

NMTF has the following four advantages:

- S absorbs both positive and negative eigenvalues such that $W \approx HSH^T$ has a better approximation than $W \approx \tilde{H}\tilde{H}^T$.
- S in HSH^T provides extra degrees of freedom such that H is much closer to the form of cluster indicators. One key of $W \approx HSH^T$ is the orthogonal constraints on H, which lead to a more sparse H than that in $W \approx \tilde{H}\tilde{H}^T$.
- There is a special meaning on k-by-k matrix S. H are vigorous cluster indicators, and $H^T H = I$. Consider the derivative $\frac{\partial J_{symNMF}}{\partial S}$; we obtain $S = H^T W H$ or $S_{lk} = h_l^T X h_k = \frac{\sum_{i \in C_l} \sum_{j \in C_k} w_{ij}}{\sqrt{n_l n_k}}$. S represents the within cluster sum of weight ($l = k$) and between cluster sum of weights ($l \neq k$). The elements of the off diagonal of S are considerably smaller than the diagonal elements if the clusters are well separated.
- In social networks, the relation view data are often symmetric. Each node represents one user, and each link represents the relationship between users. In 2011, Wang et al. [5] first used NMTF for community discovery. Specifically, Wang has reported a widely discussed position that HSH^T has a clear physical interpretability compared to other clustering methods.

3.4.1 CoNMTF

Inspired by CoNMF [7], we proposed a multi-view community discovery solution (CoNMTF) by using symNMTF to factorize a relation view instead of standard NMF.

Formally, given n_r view relation data denoted as $\{W^1, \ldots, W^{n_r}\}$, each view is factorized as $W^i \approx H^i S^i H^{i^T}$, where H^i have the same dimensions $n \times k$. S^i are of dimensions $k \times k$. S^i denotes the relationship of communities. The multi-view relation data can be solved by optimizing the following objective function Eq. (12),

$$J_{CoNMTF} = \sum_{i=1}^{n_r} \theta_i \|W^i - H^i S^i H^{i^T}\| + R_p \quad s.t. H^i \geq 0, S^i \geq 0, \tag{12}$$

where θ_i is the weight of each view; R_p is a pairwise constraint, and it relaxes the MultiNMF's [6] constraints and forces the indicator matrices of each pair of views to complement each other during the factorization process.

CoNMTF and CoNMF are the basis for constructing the multi-type, multi-view community discovery method, and they will be conducted as baseline methods to compare with the proposed method during the experiments.

4. Joint-Regularized Nonnegative Matrix Tri-Factorization

Our solution is finding a principled method to combine multi-type, multi-views (i.e., relation views and feature views) adopting the NMF technique.

After briefly reviewing solutions on multi-view data (CoNMF and CoNMTF) in Sect. 3, we introduced our solution - the JoNMTF to integrate multi-type, multi-view data. In addition, we prove the correctness and derive the time complexity of our proposed method.

4.1 JoNMTF Framework

The idea behind multi-type, multi-view community discovery is that different types of views should admit the same underlying community of the data.

Thus, the indicator matrices H of relation views and feature views are similar. Symmetric NMTF and standard NMF are applied to factorize relation views and feature views, respectively. Thus, the objective function of the proposed JoNMTF can be formulated as follows,

$$J = \sum_{i=1}^{n_r} \theta_i \|G^i - H^i S^i H^{i^T}\| + \sum_{i=n_r+1}^{N} \theta_i \|X^i - H^i V^i\| + R \quad s.t. H^i \geq 0, S^i \geq 0, V^i \geq 0, \tag{13}$$

where θ_i balances the factorization of different views. R is a proposed multi-view joint regularization function for multi-type, multi-view social networks. The regularization function R is defined as the sum of a pairwise regularization function R_p and an orthogonal regularization function R_o. We will detail these in the following sections.

4.2 Pairwise Regularization

As He introduced in [7], pairwise constraints relax MultiNMF's constraints rather than imposing similarity constraints on each pair of views. We expect that the indicator matrices Hs learned from two views can complement each other during the factorization process. The corresponding co-regularization function R_p is intuitively defined as follows:

$$R_p = \sum_{i=1}^{N} \sum_{j=1}^{N} \|H^i - H^j\|. \tag{14}$$

4.3 Orthogonal Regularization

The orthogonality principle was first employed by Li et al. [17] to minimize the redundancy between different bases, and then Ding et al. [10] broached the concept of orthogonal NMF explicitly. Under the condition of nonnegativity, orthogonality will necessarily result in sparseness [18].

Recall the NMTF, which is discussed in the related works section; from the perspective of NMTF, $W \approx HSH^T$ is a more effective factorization than $W \approx \tilde{H}\tilde{H}^T$, and the orthogonal constraint on indicator matrix H is necessary to construct NMTF [15].

The orthogonality regularization can be formulated as follows:

$$R_o = \sum_{i=1}^{n_r} \|{H^i}^T H^i - I\|, \tag{15}$$

where I is the identity matrix.

4.4 Optimization

We propose two algorithms: one with an orthogonal constraint (JoNMTF-OP) and the other without an orthogonal constraint (JoNMTF-P).

First, we ignore the orthogonal constraint, only considering the pairwise constraint in our model. Let $R = R_p$; by substituting R into Eq. (13), we obtain the objective function of JoNMTF-P:

$$\begin{aligned} J_{JoNMTF-P} = &\sum_{i=1}^{n_r} \theta_i \|G^i - H^i S^i {H^i}^T\| \\ &+ \sum_{i=n_r+1}^{N} \theta_i \|X^i - H^i V^i\| \\ &+ \sum_{s=1}^{N} \sum_{t=1}^{N} \|H^s - H^t\| \\ &s.t. H \geq 0, S \geq 0, V \geq 0, \end{aligned} \tag{16}$$

Second, we consider both the pairwise constraint and orthogonal constraint in our model. Let $R = R_p + R_o$; by substituting R into Eq. (13), we obtain the objective function of JoNMTF-OP:

$$\begin{aligned} J_{JoNMTF-OP} = &\sum_{i=1}^{n_r} \theta_i \|G^i - H^i S^i {H^i}^T\| \\ &+ \sum_{i=n_r+1}^{N} \theta_i \|X^i - H^i V^i\| \\ &+ \sum_{s=1}^{N} \sum_{t=1}^{N} \|H^s - H^t\| + \sum_{k=1}^{n_r} \|{H^k}^T H^k - I\| \\ &s.t. H \geq 0, S \geq 0, V \geq 0, \end{aligned} \tag{17}$$

Algorithm 1 Joint-regularized Nonnegative Matrix TriFactorization (JoNMTF)

Require: Relation matrices $\{G^1, \ldots, G^{n_r}\}$, feature matrices $\{X^1, \ldots, X^{n_f}\}$, parameters $\{\theta_1, \ldots, \theta_N, \alpha\}$, and number of community K;
Ensure: Membership matrix H^*

```
Normalize each relation-based view G^i such that ||G^i_{s.}|| = 1;
Initialize matrices {H^i}, {H^j}, {S^i} and {V^j}.
while Eq. (17) not converges do
    for each i from 1 to n_r do
        Update H^i and S^i using Eq. (A·8)
    end for
    for each j from n_r + 1 to N do
        Update H^j and V^j using Eq. (A·9)
    end for
end while
H^* = (1/N) Σ_{i=1}^{N} H^i
Normalized the membership matrix H^* such that Σ_j H^*_{ij} = 1.
return H^*
```

The optimizations of Eq. (16) and Eq. (17) are similar; thus, we only discuss the process for Eq. (17).

After solving the optimization problem of Eq. (17), we derive the update rules. The details of the optimization are provided in the Appendix.

Since matrices G, X, H, S, and V are all nonnegative during the updating process, the final X, S and V will also be nonnegative. Therefore, we prove the correctness of our algorithm. The convergence of the proposed Algorithm 1 can be proven via an auxiliary function following [4].

After convergence, we applied a simple late integration strategy to the membership matrices $\{H^1, \ldots, H^N\}$ of each individual view. The obtained H^* is simply the scale partition matrix, whose i-th row corresponds to the community membership of the i-th user. H^* is normalized as $\sum_j H^*_{ij} = 1$ such that H^*_{ik} corresponds to the posterior probability that the i-th user belongs to the k-th community.

$$H^* = \frac{1}{N} \sum_{i=1}^{N} H^i. \tag{18}$$

4.5 Time Complexity Analysis

We discuss the computational complexity of the proposed JoNMTF algorithm in contrast to standard NMF using big O notation.

The cost for NMF's update rules in each iteration is $O(nmk)$.

Now, consider one iteration of the proposed method. For relation-based views, the cost for update rule H^i and S^i in each single view is $O(n^2k)$. For feature-based views, the cost for update rule H^j and V^j in each single view is $O(nmk)$.

Suppose that the algorithm converges after T iterations and that the computational cost of JoNMTF is $O(T(n_r n^2 k + n_f nmk))$. The total number of relation-based views n_r and the total number of feature-based views n_f are always small constants in the context of multi-view social networks. Therefore, the overall running time of JoNMTF is linear with respect to the number of users, communities and attributes. The computational complexity of the proposed method is equal to that of a standard NMF up to a constant term.

5. Experiments

Our evaluation focuses on evaluating JoNMTF for multi-type, multi-view clustering. We first benchmark the performance computed from single views, and then we contrast it against the performance on multi-view clustering. We also compare JoNMTF against other multi-view clustering techniques.

5.1 Dataset

We experiment with five real-world datasets from Twitter: football, olympics, politics of UK (politics-uk), politics of Ireland (politics-ie) and rugby.

Each dataset includes relation views and feature views. The relation views consist of follows, mentions and retweets. Follows, mentions and retweets are constructed by binary user profile vectors based on the user whom they follow, mention and retweet, respectively. The feature views include listmerged500, lists500 and tweets500. List500 means list content profiles, constructed from the concatenation of both the names and the descriptions of the 500 Twitter lists to which each user has most recently been assigned. Listmerged500 (listmgd500) merges the items from list500 according to the name. Tweets500 means user content profiles, constructed from the concatenation of the 500 most recently posted tweets for each user. Table 1 lists the details of the five Twitter datasets.

5.2 Baselines

In addition to the baseline NMF, we also compare with the following five algorithms.

1.NMF. Since the proposed method is an extension of NMF, it is reasonable to choose NMF as the baseline of the single-view clustering algorithm. The worst and best performances of NMF for single view are reported as WorstNMF and BestNMF, respectively.

2.UniNMF. UniNMF is a two-step algorithm. The first step is an early integration process; it combines all relation views and feature views into a unified graph using the method in [11], and then standard NMF is performed on this graph as a single-view community discovery problem.

3.MultiNMF [6]. MultiNMF solve a multi-view community discovery problem by optimizing following objective function,

$$J_{multiNMF} = \sum_{v=1}^{n_v} \|X^{(v)} - U^{(v)}(V^v)^T\|_F^2 + \sum_{v=1}^{n_v} \lambda_v \|V^{(v)} - V^*\|_F^2$$
$$s.t. 1 \leq k \leq K, \|U_{\cdot,k}^{(v)}\|_1 = 1 and U^{(v)}, V^{(v)}, V^* \geq 0. \quad (19)$$

We set the regularization parameters λ_v uniformly as 0.01 as suggested.

4.CoNMF [7]. CoNMF solves the problem of Eq. (11). The parameters are set to 1 as suggested by the authors.

5.CoNMTF. CoNMTF is discussed in Sect. 3, it solves the problem of Eq. (12). It is based on the non-negative matrix triple factorization. The parameters are set to 1.

The five baselines cover all three types of integration strategy. UniNMF is an early integration algorithm, whereas MultiNMF is an immediate integration algorithm. CoNMF, CoNMTF and JoNMTF are immediate and late integration algorithms, and the late integration strategy is applied to obtain the final clustering result. Table 2 presents the comparison of these algorithms.

5.3 Setup

In our JoNMTF settings, the regularization parameters are set to 1 for all views. For each algorithm, 10 test runs with different random initializations were conducted, and the average score is reported. In the following, we report the statistical significance (judged at the 5% level by a one-tailed two-sample t-test) where appropriate.

5.4 Result and Analysis

Tables 3 and 4 show the algorithm performance for all five datasets, the average performance along with the standard deviation are reported.

From the results in Table 3 and Table 4, we can see that the proposed JoNMTF-P and JoNMTF-OP methods consistently outperformed the other compared methods, sometimes very significantly, which demonstrate the advantage of our approaches in terms of clustering performance.

The results show that JoNMTF-P outperformed the second-best baseline algorithm in terms of accuracy/NMI (1.60%/1.60%,for the football dataset; 3.66%/2.77% for the

Table 1 Summary of five Twitter data sets.

Datasets	football	olympics	politics-ie	politics-uk	rugby
users	248	464	348	419	854
communities	20	28	7	5	15
follows	3819	10642	16856	27340	35757
mentions	3312	9615	6318	14788	33832
retweets	1350	3740	3019	7270	12472
lists500	7814	4942	1047	3614	5900
listmgd500	3601	3907	1051	2879	3785
tweets500	11806	18455	14377	19868	28903

Table 2 Summary of six algorithms

Algorithm	Data Type	Data Source	Integration
WorstNMF	single-view	relation	none
BestNMF	single-view	relation	none
UniNMF	multi-type multi-view	relation + feature	early
MultiNMF	multi-type multi-view	relation + feature	immediately
CoNMF	multi-type multi-view	relation + feature	immediately+late
CoNMTF	single type multi-view	relation	immediately+late
JoNMTF	multi-type multi-view	relation + feature	immediately+late

Table 3 Average of accuracy on five real world datasets (%)

Algorithm	football	olympics	politics-ie	politics-uk	rugby
WorstNMF	20.16 ± 0.03	15.95 ± 0.01	58.47 ± 0.09	57.76 ± 0.04	12.53 ± 0.07
BestNMF	25.00 ± 0.01	18.32 ± 0.02	64.73 ± 0.12	61.61 ± 0.11	17.21 ± 0.03
UniNMF	42.63 ± 3.43	33.27 ± 3.52	59.00 ± 0.15	56.30 ± 0.17	46.60 ± 3.23
MultiNMF	57.66 ± 0.04	65.23 ± 0.06	64.08 ± 0.03	68.39 ± 0.03	62.93 ± 0.04
CoNMF	79.63 ± 0.03	77.58 ± 0.03	55.35 ± 0.02	64.94 ± 0.03	59.95 ± 0.03
CoNMTF	86.69 ± 0.12	80.82 ± 0.20	76.15 ± 0.14	86.16 ± 0.17	57.61 ± 0.11
JoNMTF-P	**88.29** ± 0.21	**84.48** ± 0.19	**83.05** ± 0.13	**89.32** ± 0.12	**65.53** ± 0.32
JoNMTF-OP	**89.11** ± 0.32	**86.76** ± 0.13	**85.34** ± 0.21	**90.84** ± 0.32	**67.11** ± 0.39

Table 4 Average of NMI on five real world datasets (%)

Algorithm	football	olympics	politics-ie	politics-uk	rugby
WorstNMF	29.25 ± 0.33	27.48 ± 0.58	48.49 ± 0.87	53.64 ± 0.55	6.38 ± 0.21
BestNMF	33.91 ± 0.54	30.37 ± 0.63	68.53 ± 0.34	62.56 ± 0.29	9.23 ± 0.13
UniNMF	47.15 ± 2.54	44.89 ± 4.19	33.09 ± 0.05	37.98 ± 0.14	43.74 ± 4.31
MultiNMF	61.67 ± 0.06	51.40 ± 0.03	50.11 ± 0.04	58.06 ± 0.43	48.74 ± 0.02
CoNMF	82.72 ± 0.01	85.29 ± 0.01	49.66 ± 0.05	67.60 ± 0.04	62.32 ± 0.02
CoNMTF	89.81 ± 0.31	89.99 ± 0.43	80.46 ± 0.41	78.20 ± 0.65	61.10 ± 0.77
JoNMTF-P	**91.41** ± 0.20	**92.76** ± 0.31	**82.09** ± 0.25	**83.20** ± 0.34	**65.46** ± 0.35
JoNMTF-OP	**92.35** ± 0.17	**93.41** ± 0.26	**84.12** ± 0.28	**85.32** ± 0.13	**68.94** ± 0.59

Olympics dataset; 6.90%/1.63% for the politics-ie dataset; 3.16%/5.00% for the politics-uk dataset, and 2.60%/3.14% for the rugby dataset).

For the football dataset, JoNMTF-P outperformed the baseline methods with a margin (1.60%/1.60%). MultiNMF achieves (57.66% and 61.67%) while the performance of JoNMTF is approximately 90%. One of the possible reasons is that the relaxed constraint involved the complement information regarding each pair of views. The best and worst performances of NMF for individual view were approximately 25%/20%. JoNMTF-P was able to detect the true communities by finding the right direction during optimization process.

For the Olympics dataset, JoNMTF-P performed well (over 84%/92%). Notably, CoNMF achieved 77%/85% which outperformed MultiNMF (65%/51%).

For the politics-ie and politics-uk datasets, the performance of individual view was comparable with that of CoNMF (approximately 50% and 65%), while JoNMTF-P achieves a performance over 80%. These results support our claim that the proposed JoNMTF-P algorithm is useful when dealing with multiple types of data sources.

For the rugby dataset, JoNMTF-P performed the worst among all the five datasets, which is close to the best of baseline methods. Notably, the multi-view method still outperformed the standard NMF with a large margin over 30%.

Next, we consider the orthogonality of H, whose importance w.r.t. clustering is emphasized in Sect. 2 and Sect. 3. The difference between JoNMTF-P and JoNMTF-OP is that JoNMTF-OP with an orthogonal constraint and JoNMTF-P without orthogonal constraint. We compute the normalized orthogonality, $(H^T H)_{nm} = D^{-1/2}(H^T H)D^{-1/2}$, where $D = diag(H^T H)$. Thus the diagonal is normalized to 1, and derivation can be clearly judged. The off-diagonal elements in $(H^T H)_{nm}$ reflect the orthogonality of H. We compute the max and mean value of off-diagonal elements and results are given in Table 5. One can see that off-diagonal elements of JoNMTF-OP (with orthogonal constraint) is smaller than that of JoNMTF-P (without orthogonal constraint) on five datasets. With the orthogonal constraints on H, JoNMTF-OP outperformed JoNMTF-P on all five datasets.

Table 5 The mean and max value of off-diagonal elements in $(H^T H)_{nm}$

Dataset	JoNMTF-P		JoNMTF-OP	
	Mean	Max	Mean	Max
football	0.0356	0.3616	**0.0280**	**0.1473**
olympics	0.0337	0.6617	**0.0201**	**0.4447**
politics-ie	0.1356	0.5620	**0.0787**	**0.3958**
politics-uk	0.1362	0.5121	**0.0718**	**0.2336**
rugby	0.0398	0.1762	**0.0355**	**0.1652**

6. Conclusions

Community discovery in social networks with multi-type multi-view data is more difficult than that for single-view data. Data from different perspectives can be complementary, making multi-view data more adequate than information obtained from a single view. In this paper, we proposed an algorithm (JoNMTF) for combining multi-view relation data and feature data in social networks. A pairwise regularization was introduced in the algorithm to complement each pair of views. A novel orthogonal regularization was proposed for an better approximation and a more sparse solution. The correctness and time complexity of the algorithm were discussed. The algorithm was tested on five real world datasets, and the results were compared with those obtained using existing methods. The results demonstrated that our algorithm is an improvement over existing methods.

References

[1] J. Tang, C. Zhang, K. Cai, L. Zhang, and Z. Su, "Sampling representative users from large social networks," Twenty-Ninth AAAI

Conference on Artificial Intelligence, 2015.
[2] M. Zitnik and B. Zupan, "Data fusion by matrix factorization," IEEE Transactions on Pattern Analysis & Machine Intelligence, vol.36, no.1, pp.41–53, 2013.
[3] B. Li, G. Zhou, and A. Cichocki, "Two efficient algorithms for approximately orthogonal nonnegative matrix factorization," IEEE Signal Process. Lett., vol.22, no.7, pp.843–846, 2015.
[4] D.D. Lee, "Algorithms for non-negative matrix factorization," Advances in Neural Information Processing Systems, vol.13, no.6, pp.556–562, 2001.
[5] F. Wang, T. Li, X. Wang, S. Zhu, and C. Ding, "Community discovery using nonnegative matrix factorization," Data Mining & Knowledge Discovery, vol.22, no.3, pp.493–521, 2011.
[6] G. Jing, H. Jiawei, L. Jialu, and W. Chi, "Multi-view clustering via joint nonnegative matrix factorization," SIAM International Conference on Data Mining, 2013.
[7] X. He, M.-Y. Kan, P. Xie, and X. Chen, "Comment-based multi-view clustering of web 2.0 items," Proceedings of the 23rd international conference on World wide web, pp.771–782, 2014.
[8] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," Environmetrics, vol.5, no.2, pp.111–126, 1994.
[9] D.D. Lee and H.S. Seung, "Learning the parts of objects by non-negative matrix factorization," Nature, vol.401, no.6755, pp.788–791, 1999.
[10] C. Ding, T. Li, W. Peng, and H. Park, "Orthogonal nonnegative matrix t-factorizations for clustering," Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Philadelphia, Pa, Usa, pp.126–135, Aug. 2006.
[11] D. Greene and P. Cunningham, "Producing a unified graph representation from multiple social network views," Association for Computing Machinery, pp.118–121, 2013.
[12] A. Hindle, J. Shao, D. Lin, J. Lu, and R. Zhang, "Clustering web video search results based on integration of multiple features," World Wide Web-internet & Web Information Systems, vol.14, no.1, pp.53–73, 2011.
[13] D. Hidru and A. Goldenberg, "Equinmf: Graph regularized multiview nonnegative matrix factorization," Computer Science, 2014.
[14] W. Xu, X. Liu, and Y. Gong, "Document clustering based on non-negative matrix factorization," International ACM SIGIR Conference on Research and Development in Informaion Retrieval, pp.267–273, 2003.
[15] C. Ding, X. He, H.D. Simon, and R. Jin, "On the equivalence of nonnegative matrix factorization and k-means-spectral clustering," Lawrence Berkeley National Laboratory, 2008.
[16] D. Kuang, S. Yun, and H. Park, "Symnmf: nonnegative low-rank approximation of a similarity matrix for graph clustering," Journal of Global Optimization, vol.62, no.3, pp.545–574, 2015.
[17] S.Z. Li, X.W. Hou, H.J. Zhang, and Q.S. Cheng, "Learning spatially localized, parts-based representation," Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. CVPR 2001. vol.1, pp.I-207–I-212, 2001.
[18] Y.-X. Wang and Y.-J. Zhang, "Nonnegative matrix factorization: A comprehensive review," IEEE Transactions on Knowledge & Data Engineering, vol.25, no.6, pp.1336–1353, 2013.

Appendix: Optimization

We adopt alternating optimization to minimize the objective function Eq. (17). First, the value of S^i is fixed while minimizing J over H^i. Then, the value of H^i is fixed while minimizing J over S^i. After that, the value of V^j is fixed while while minimizing J over H^j. Then, the value of H^j is fixed while minimizing J over V^j. These steps are iteratively executed until the stop condition was satisfied. The stop condition was either convergence or exceeding the threshold of iteration.

Rewritten the objective function J from Eq. (17) as:

$$\begin{aligned} J = &\sum_{i=1}^{n_r} \theta_i tr(W^{i^T} W^i + H^i S^{i^T} H^{i^T} H^i S^i H^{i^T} \\ &\qquad - W H^i S^{i^T} H^{i^T} - W^T H^i S^i H^{i^T}) \\ &+ \sum_{j=n_r+1}^{N} \theta_j tr(X^{j^T} X^j - 2X^{j^T} H^j V^j + V^{j^T} H^{j^T} H^j V^j) \\ &+ \sum_{s=1}^{N} \sum_{t=1}^{N} tr(H^{s^T} H^s - 2H^{s^T} H^t + H^{t^T} H^t) \\ &+ \alpha \sum_{k=1}^{N} tr(H^{k^T} H^k H^{k^T} H^k - 2H^{k^T} H^k + I) \end{aligned} \tag{A· 1}$$

where $tr(\cdot)$ denotes the trace function. Let $\omega_1, \omega_2, \omega_3$ and ω_4 be the Lagrange multiplier matrix for the constraint, respectively. L be the Lagrange,

$$\begin{aligned} L = &J + \omega_1 tr(H^{i^T}) + \omega_2 tr(H^{i^T}) + \omega_3 tr(S^{i^T}) \\ &+ \omega_4 tr(V^{j^T}) \end{aligned} \tag{A· 2}$$

Then, we obtain the derivatives of L with respect to H^i, H^j, S^i and V^j.

$$\begin{aligned} \frac{\partial L}{\partial H^i} = &4\theta_i (H^i S^i H^{i^T} H^i S^i - W H^i S^i) + \sum_{j=1}^{N} 2(H^i - H^j) \\ &+ 4\alpha (H^i H^{i^T} H^i - H^i) + \omega_1 \end{aligned} \tag{A· 3}$$

$$\begin{aligned} \frac{\partial L}{\partial H^j} = &2\theta_j (-X^j V^{j^T} + H^j V^j V^{j^T}) + \sum_{i=1}^{N} 2(H^j - H^i) \\ &+ 4\alpha (H^j H^{j^T} H^j - H^j) + \omega_2 \end{aligned} \tag{A· 4}$$

$$\frac{\partial L}{\partial S^i} = 2\theta_i (H^{i^T} H^i S^i H^{i^T} H^i - H^{i^T} W^i H^i) + \omega_3 \tag{A· 5}$$

$$\frac{\partial L}{\partial V^j} = \theta_j (-2H^{j^T} X^j + 2H^{j^T} H^j V^j) + \omega_4 \tag{A· 6}$$

Let $\omega_1(s,t) H^i(s,t) = 0$, $\omega_2(s,t) H^j(s,t) = 0$, $\omega_3(s,t) S^i(s,t) = 0$ and $\omega_4(s,t) V^j(s,t) = 0$. Follow the Karush-Kuhn-Tucker (KKT) complementary slackness condition, we have

$$\begin{aligned} &\frac{\partial L}{\partial H^i} \odot H^i = 0, \frac{\partial L}{\partial H^j} \odot H^j = 0, \\ &\frac{\partial L}{\partial S^i} \odot S^i = 0, \frac{\partial L}{\partial V^j} \odot V^j = 0 \end{aligned} \tag{A· 7}$$

where the $\odot$ and the division symbol in this matrix context denote element-wise multiplication and division. For example, $(A \odot B)_{ij} = A_{ij} B_{ij}$. Same for element-wise division.

Solving the above equations, we derive the following update rules:

$$H^i \leftarrow H^i \odot \frac{\theta_i W H^i S^i + \alpha H^i + \sum_{j=1}^{N} H^j}{\theta_i H^i S^i H^{i^T} H^i S^i + \alpha H^i H^{i^T} H^i + \sum_{j=1}^{N} H^i} \tag{A$\cdot$8}$$

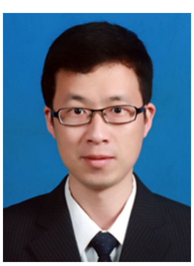
$$H^j \leftarrow H^j \odot \frac{\theta_j X^j V^{j^T} + \alpha H^j + \sum_{i=1}^{N} H^i}{\theta_j H^j V^j V^{j^T} + \alpha H^j H^{j^T} H^j + \sum_{i=1}^{N} H^j} \tag{A$\cdot$9}$$

$$S^i \leftarrow S^i \odot \frac{H^{i^T} W^i H^i}{H^{i^T} H^i S^i H^{i^T} H^i} \tag{A$\cdot$10}$$

$$V^j \leftarrow V^j \odot \frac{H^{j^T} X^j}{H^{j^T} H^j V^j} \tag{A$\cdot$11}$$

Liangliang Zhang received the B.S. and M.S. degrees in computer science and technology from PLA University of Science and Technology in 2011 and 2014, respectively. Since 2014, she is a Ph.D. student in PLA University of Science and Technology. Her research interests include complex network, machine learning, and pattern recognition.

Longqi Yang received the B.S. in computer science and technology from Tongji University and received M.S. degrees in computer network from PLA University of Science and Technology in 2010 and 2015, respectively. His research interests include large-scale computing, machine learning, and pattern recognition.



Yong Gong received B.S. and M.S. degrees from Air Force Engineering University in 2003 and 2006, respectively. He received a Ph.D. from PLA University of Science and Technology and now a postdoctoral researcher. His subjects are machine learning, deep learning, and complex networks.

Zhisong Pan received the B.S. and M.S. in computer science and technology from the PLA Information Engineering University, respectively. He received Ph.D. degree from Nanjing University of Aeronautics and Astronautics in 2003. He is now a Professor at PLA University of Science and Technology. His research interests include deep learning, machine learning, and pattern recognition.

Yanyan Zhang received the B.S. and M.S. in computer science and technology from Nanjing University of Aeronautics and Astronautics in 2009 and 2012, respectively. She is now Researcher/Lecturer at PLA University of Science and Technology. Her research interests include machine learning, and pattern recognition.

Guyu Hu received the B.S. in Zhejiang University. After that, he received M.S. and Ph.D. in communication and information system from PLA University of Science and Technology. He is now a Professor at PLA University of Science and Technology. In 2010, he was owned the outstanding contributions of young and middle-aged experts in Jiangsu province. His research interests include network operation, data analysis and big data.