IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 18, 2019, accepted October 5, 2019, date of publication October 11, 2019, date of current version October 30, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2946744

Multi-Mode Social Network Clustering via Non-Negative Tri-Matrix Factorization With Cluster Indicator Similarity Regularization

LI NI[1,2], PENG MANMAN[1], AND WU QIANG[1]
[1]College of Information and Engineering, Hunan University, Changsha 41008, China
[2]College of Information and Electronic Engineering, Hunan City University, Yiyang 413000, China

Corresponding author: Peng Manman (pengmanman@hnu.edu.cn)

The research was partially funded by the Key Program of National Natural Science Foundation of China (Grant Nos. 61133005, 61432005), the National Natural Science Foundation of China (Grant Nos. 61370095, 61472124, 61572175), International Science & Technology Cooperation Program of China (2015DFA11240). The work was supported by Research Project of the Education Department of Hunan Province (Grant No. 14C0210).

ABSTRACT Community discovery algorithms are important aspects of network science, especially as social network structures become more complex. Multi-mode social networks have recently become a challenging and popular topic in this field. At present, inner-mode relationship is mainly considered in community discovery algorithms for social networks. Thus, the effect of the these methods is not well in clustering as the intra-mode relationship is not considered in the clustering methods. In this paper, we propose a flexible and robust clustering framework, MRTA (the Multi-Similarity Regular Tri-Factorization Algorithm), based on non-negative tri-matrix factorization. MRTA has several advantages over the existing methods. First, it achieves more consistent clustering results based on cluster indicator of inner-mode and intra-mode relationships of multi-mode networks. Second, it can simultaneously cluster multiple modes, which is impossible for single-mode clustering algorithms. Finally, it provides a multi-mode clustering solution that is more robust to noise. We perform an efficient iterative update algorithm, and theoretically prove its accuracy. Extensive experimental results of a variety of real and synthetic networks demonstrate the effectiveness of our approach.

INDEX TERMS Cluster indicator, multi-mode social network, non-negative tri-matrix factorization.

I. INTRODUCTION

Ever since Watts proposed the small world phenomenon, multi-mode and heterogeneous networks have been a subject of keen interest for network science researchers [1], [4], [5], [9], [10]. Algorithms for discovering social network communities represent a classic yet challenging problem [21] for this field of study. Some researchers have proposed that complex networks have a community structure [22], [23], i.e., they consist of multiple tightly knitted communities of varying size. This property allows complex networks to be examined from an entirely new perspective: the communities that make up the complex network. For example, in the Weibo social network, a community would refer to users with the same interests and hobbies (e.g., music and movies). In a research cooperation network, a community would consist of researchers with the same research interests or field of study. Several highly effective community discovery algorithms are already available in the literature, like k-means clustering [15], [24], spectral clustering [13], [25], [26], non-negative matrix factorization (NMF) [6], [7], [12], [17], [27], and non-negative matrix tri-factorization (NMTF) [19]–[21], [28], [29].

In complex social networks, network may be collected from multiple modes(objects) [2]. For example, Twitter network, as shown in Figure 1, the network includes two different modes: user mode and Twitter mode, and corresponding relationships (interact, post, involve). In many applications, the relational topic model (RTM) algorithm [30], [31] are proposed in document network, which used the relationships in document clustering. Ou-Yang *et al.* [18] proposed a multi-network clustering method for multiple protein heterogeneous networks. NMF-based combinatorial clustering algorithms [3], [8], [11], [20] had also been proposed in

The associate editor coordinating the review of this manuscript and approving it for publication was Maode Ma.

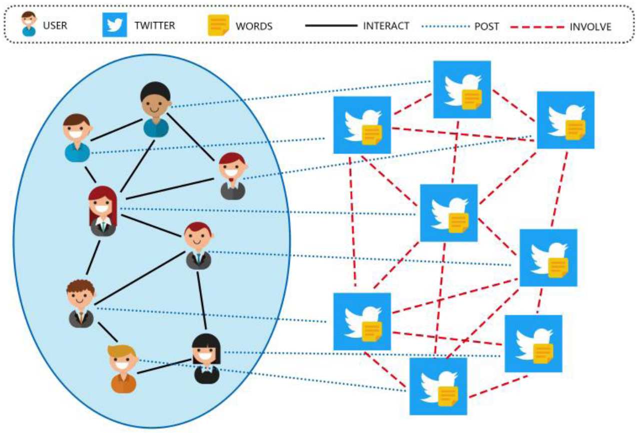


FIGURE 1. Twitter network instance.

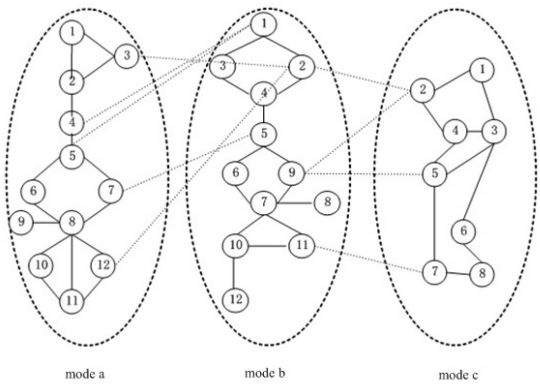


FIGURE 2. An example of relationship graph of a three-mode network: mode a, mode b, and mode c.

multi-mode network. Multi-mode network can fuse more modes and their interaction relationships, the correctness of clustering had been improved for existing clustering algorithms based on multi-mode networks.

Despite the success of previous approaches in network clustering, they still suffer from two limitations. First, they usually assume that the nodes in different modes are for the same type of nodes. First, the intra-mode nodes relationships are one-to-one correspondence. But this assumption may not be inconsistent with the nature of multi-mode networks. Second, existing approaches tend to focus inner-mode relationships and ignore intra-mode relationships that may affect their clustering results. For example, in the Twitter network, a pair of users may belong to the same community if they publish similar content on Twitter, even if they rarely interact with each other.

For social networks, relationship graphs can be constructed between nodes of the inner-mode and between nodes with intra-mode. Fig. 2 shows a example mode of a three-mode network representing a social network. It shows 3 clusters in mode a, 3 clusters in mode b and 2 clusters in mode c. It is summarized in Table 1. This model graph describes three key characteristics in multi-mode network graphs that are relevant to cluster indicators:

1) Nodes of the inner-mode will form clustering relationships according to topological similarity, and the nodes in such a relationship are typically unrelated to other modes (e.g., Nodes 1 and 2 in mode a, Nodes 7 and 8 in mode b, and Nodes 6 and 8 in mode c).
2) If the nodes of two modes are connected, a node will have two types of connections: with nodes of the inner-mode and with nodes from other modes. In this case, the clustering relationship between nodes of the inner-mode is affected by inter-mode connections (e.g., in mode a, Node 4 is connected to Nodes 2 and 5 as well as Node 1 in mode b). If the topological similarity with Node 4 is only considered for nodes of the inner-mode, ascertaining the clustering of Node 4 is difficult. In this case, the relationship between Node 4 and the nodes of other modes may play an important role in determining its clustering. Because Node 4 of Mode a is connected to Node 1 of Mode b and Node 1 of Mode b is also connected to Node 5 of mode a, Node 4 should be clustered with Node 5 in mode a.
3) Different modes may not have the same number of nodes or clusters. For example, mode b has 12 nodes, whereas mode c only has eight nodes. Mode b also has three clusters while mode c has two clusters.

TABLE 1. Cluster pairs in Fig 2.

mode	1st cluster pair	2st cluster pair	3st cluster pair
mode a	$\{1,2,3\}$	$\{4,5,6,7\}$	$\{8,9,10,11,12\}$
mode b	$\{1,2,3,4,5\}$	$\{6,7,8,9\}$	$\{10,11,12\}$
mode c	$\{1,2,3,4,5\}$	$\{6,7,8\}$	

Based on these advances, we propose a generalized clustering model for multi-mode networks that is based on the aforementioned clustering-relevant characteristics of multi-mode network graphs and integrates the characteristics of social media content with social relationships between different types of entities. The model algorithm is based on the matrix decomposition of social network relationships and social media contents and incorporates clustering similarity relationships for modes of the same type and of different types. The algorithm also sparsely regularizes the correlation matrix to enhance its robustness against noise. The proposed model is a community discovery algorithm called the Multi-Similarity Regular Tri-factorization Algorithm (MRTA). The contributions to this paper are summarized below:

1 We propose a community discovery method based on regularized cluster indicator relationships. We not only consider the inner-mode relationships of clustering, but also the intra-mode relationships of clustering. And, our method also accounts for differences in the size of the modes and the number of clusters of each mode. Thus, our model can be adapted to a variety of applications.
2 Because intra-mode relationships are intrinsically noisy, this may decrease clustering effectiveness, the L1-norm regularization is used to reduce the impact on inconsistent correlations between the clustering process. This ensures that the algorithm is robust against noise.

3 The proposed model is capable of simultaneously clustering multiple modes and simultaneously enhancing cluster associations between different modes.

4 The accuracy, robustness and convergence of the proposed model were theoretical analysed. In addition, it was experimentally assessed with simulated data and several real multi-mode datasets. The effectiveness of our method was thus validated theoretically and experimentally.

The rest of this paper is organized as follows: Section II reviews related studies. Section III shows related models and preliminaries, presents the verification of the MRTA algorithm,. Section IV describes experimental results from this study. Section V concludes this study.

II. RELATED WORKS

When spectral clustering is performed with a conventional matrix factorization clustering algorithm, the similarity matrix A is usually assumed to be positive. In practical applications, however, the similarity matrix is actually non-negative. The generalized NMF proposed by Ding *et al.* [15] is expressed as follows:

$$\min_{C,R} J = ||A - CRC^T||_F^2, \tag{1}$$

where $A \in R_+^{n\times n}$, $H \in R_+^{n\times k}$, and $R \in R_+^{k\times k}$; A is a non-negative and non-symmetric positive definite (s.p.d.) similarity matrix; and C is the cluster indicator matrix. According to Cholesky factorization, $A \cong CRC^T$. This has been proven to be the best way to express matrix decomposition in clustering algorithms, and we refer to this expression as weighted NMF. Although the objective function is a convex function of any C or R variable, the objective function of both C and R would is a non-convex function. Based on the Karush-Kuhn-Tucker (KKT) optimality condition, Shang *et al.* [7] and Meng *et al.* [8] proposed a multiplicative iterative algorithm for objective functions in the form of Eq. (1):

$$\begin{aligned} R_{ik} &\leftarrow R_{ik}\frac{(C^TAC)_{ik}}{(C^TCRC)_{ik}}, \\ C_{ik} &\leftarrow C_{ik}\frac{(ACR)_{ik}}{(CRC^TCR)_{ik}}. \end{aligned} \tag{2}$$

Most studies have focused on matrix factorization for single-mode networks and have generally emphasized the importance of non-negativity and orthogonality for clustering operations. However, modern information networks are usually multi-mode networks that contain a variety of subjects and correlations. The work of Ding *et al.* [15] on non-negative matrix factorization has been extended to multi-mode networks via the tri-factorization method (i.e.,NMTF) for relational graphs [20], [21], [25], [28], [29]. NMTF-based clustering adds orthogonality and non-negativity constraints to matrix factorization to improve clustering accuracy. The NMTF objective function is

$$\begin{aligned} &\min_{F,S,G} J = ||S - FRG^T||_F^2, \\ &s.t.\ F \geq 0, \quad G \geq 0,\ S \geq 0, F^TF = I_k,\ G^TG = I_k. \end{aligned} \tag{3}$$

Likewise, the multiplicative iteration algorithm for G,F, and R is:

$$\begin{aligned} G_{ik} &\leftarrow G_{ik}\sqrt{\frac{(R^TFS)_{ik}}{(GG^TR^TFS)_{ik}}}, \\ F_{ik} &\leftarrow F_{ik}\sqrt{\frac{(RGS^T)_{ik}}{(F^TFSG^TG)_{ik}}}, \\ R_{ik} &\leftarrow R_{ik}\sqrt{\frac{(F^TSG)_{ik}}{(F^TFRG^TG)_{ik}}}. \end{aligned} \tag{4}$$

In the paper, we focus on the study of NMTF

III. NON-NEGATIVE MATRIX FACTORIZATION BASED ON CLUSTER INDICATOR SIMILARITY REGULARIZATION

Suppose that there exists a multi-mode network graph with m modes $\{D_1, D_2, \ldots, D_m\}$ and the inner-mode relationships of each mode may be expressed as a matrix $A_p(1 \leq p \leq m)$ with $S_{pq}(1 \leq p, q \leq m)$ being the mode correlation matrix. The definitions of each symbol are given in Table 1.

In this section, in subsection A, we first define the relations between the network similarity matrix of cluster indicators. Then, in subsection B, we present the proposed MRTA. Finally, in subsection C, we propose an alternating iterative update algorithm for C and R, prove its convergence, and analyzed the time complexity of MRTA.

A. MATRIX GRAPH OF CLUSTER INDICATOR SIMILARITY

The relationship between the network similarity matrix and cluster indicators can be categorized into two cases:

1 In nodes of the inner-mode, a greater similarity between the i-th and j-th nodes of mode q ($(x_q)_i$ and $(x_q)_j$, respectively), implies a greater similarity between their cluster indicators $(C_q)_i$ and $(C_q)_j$.

2 In multi-mode nodes, a stronger association between the i-th node of mode p $(x_p)_i$ and the j-th node of mode q $(x_q)_j$ implies greater similarity between their cluster indicators ($(C_p)_i$ and $(C_q)_j$, respectively).

There are many ways to express the above conditions. In Case 1, the loss function may be expressed as

$$W_1 = \frac{1}{2}\sum_{i,j=1}^{n_q} ||(C_q)_i - (C_q)_j||_F^2 (A_q)_{ij}, \tag{5}$$

where

$$(A_q)_{ij} = \begin{cases} 1 & (x_q)_j \in N((x_q)_i) \\ 0 & otherwise \end{cases} \quad i,j = 1,2,\ldots,n_q, \tag{6}$$

here, $N((x_q)_j)$ is the set of nodes that are adjacent to $(x_q)_i$. Equation (5) may be expressed in matrix form as follows:

$$\begin{aligned} W_1 &= \frac{1}{2}\sum_{i,j=1}^{n_q} ||(C_q)_i - (C_q)_j||_F^2 (A_q)_{ij} \\ &= \sum_{i=1}^{n_q}(C_q)_i((C_q)_i)^T(D_q)_{ii} - \sum_{i,j=1}^{n_q}(C_q)_j((C_q)_j)^T(A_q)_{ij} \end{aligned}$$

$$= tr(C_q^T D_q C_q) - tr(C_q^T A_q C_q)$$
$$= tr(C_q^T L_q C_q), \quad (7)$$

where the Laplacian matrix of mode p is given by $L_q = D_q - A_q$. D_q is the angle matrix (i.e.,$(D_q)_{ii} = \sum_j (A_q)_{ij}$. $(C_q)_i$ is the cluster indicator vector of node $(x_q)_i$.

In Case 2, the loss function cannot be expressed with functions in the form of Equations (5) − (7) because the existence of inter-mode relations clearly affects the cluster indicators of each mode. Because the number of clusters in each mode can be different, the number of nodes in each mode can also vary. In this paper, the similarity between the cluster indicators of $(x_q)_i$ and $(x_q)_j$ ($(C_q)_i$ and $(C_q)_j$, respectively) is expressed by the cosine similarity. Therefore, the cluster indicator similarity matrix of the nodes in mode q is expressed as $CS_q = C_q C_q^T$. The cluster mapping index matrix of the nodes of mode p in mode q is defined as $\widetilde{C_q} = S_{pq} C_p$. Because this is also expressed in terms of cosine similarity, $\widetilde{CS_q} = \widetilde{C_q}\widetilde{C_q}^T$. Based on our definition of the cluster indicator similarity, the loss function is defined as

$$W_2 = ||CS_q - \widetilde{CS_q}||_F^2$$
$$= \sum_{p,q} ||S_{pq} C_p (S_{pq} C_p)^T - C_q C_q^T||_F^2, \quad (8)$$

in the equation, the similarity association matrix(S_{pq}) between modes q and p is defined as:

$$(S_{pq})_{ij} = \begin{cases} 1 & (x_p)_j \in N_{pq}((x_q)_i) \\ 0 & otherwise \end{cases}$$
$$i = 1, 2, \ldots, \quad n_q, j = 1, 2, \ldots n_p, \quad (9)$$

where $N_{pq}((x_q)_i)$ is the set of adjacent nodes of node i of mode q ($(x_q)_i$) in mode p.

B. MRTA MODEL

Based on the two similarity matrix graphs above, we propose a non-negative matrix factorization model for a network with m modes that is based on similarity graph regularization:

$$J_{MRTA}$$
$$= \sum_{p,q} ||S_{pq}^T - C_p R_{pq} C_q^T||_F^2 + a_p \sum_p tr(C_p^T L_p C_p)$$
$$+ b_{pq} \sum_{p,q} ||S_{pq} C_p (S_{pq} C_p)^T - C_q C_q^T||_F^2$$
$$+ c_{pq} \sum_{p,q} ||R_{pq}||_1$$
$$s.t. \; R_{pq} \geq 0, C_p \geq 0, C_p C_p^T = I_p, S_{pq} \in R^{n_q \times n_p},$$
$$C_p \in R^{n_p \times k_p}, R_{pq} \in R^{k_p \times k_q}, p, q = 1, \ldots, m, \quad (10)$$

where a_p, b_{pq}, and c_{pq} are regularization parameters. a_p and b_{pq} use cluster indicator similarity matrix graphs to balance the first-order loss function, where as c_{pq} reduces the noise of the association graph S_{pq}. Because the correlation matrix or graph is intrinsically noisy, sparse L1-norm regularization is used to reduce the noise. The optimization function in Eq. (10) is obtained by combining these parts.

C. ITERATIVE UPDATE ALGORITHM

Because MRTA is a non-convex function of C_p, R_{pq}, and C_q, directly solving for the global optimum of this model is not realistically possible. Therefore, we propose an alternating iterative optimization algorithm where one variable is fixed while the convex optimization problem of the other variable is minimized. This approach ultimately produces a stable solution or local extremum.

1) UPDATING THE C_π VARIABLE

We must now solve for Eq. (10). According to the KKT optimality condition, the Lagrangian objective function of $C_\pi(J(C_\pi))$ is

$$J(C_\pi) = \sum_{p\pi \in A} ||S_{p\pi}^T - C_p R_{p\pi} C_\pi^T||_F^2$$
$$+ \sum_{\pi q \in A} ||S_{\pi q}^T - C_\pi R_{\pi q} C_q^T||_F^2$$
$$+ \sum_{p\pi \in A} b_{p\pi} ||S_{p\pi} C_p (S_{p\pi} C_p)^T - C_\pi C_\pi^T||_F^2$$
$$+ \sum_{\pi q \in A} b_{\pi q} ||S_{\pi q} C_\pi (S_{\pi q} C_\pi)^T - C_q C_q^T||_F^2$$
$$+ a_\pi tr(C_\pi^T L_\pi C_\pi) - tr(\Lambda_\pi C_\pi^T)$$
$$+ tr(\Gamma_\pi (C_\pi^T C_\pi - I_{k_\pi})). \quad (11)$$

To generalize this expression, we calculated the update rules of mode. In particular, Λ_π and Γ_π are matrices where every term is a Lagrangian operator. Taking the derivative of $L(C_\pi)$ with respect to C_π yields

$$\nabla_{C_\pi} = -2 \sum_{p\pi \in A} S_{p\pi} C_p R_{p\pi} + 2 \sum_{p\pi \in A} C_\pi R_{p\pi}^T C_p^T C_p R_{p\pi}$$
$$-2 \sum_{\pi q \in A} S_{\pi q}^T C_q R_{\pi q}^T + 2 \sum_{\pi q \in A} C_\pi R_{\pi q} C_q^T C_q R_{\pi q}^T$$
$$+2 a_\pi L_\pi C_\pi + 4 \sum_{\pi q \in A} b_{\pi q} S_{\pi q}^T S_{\pi q} C_\pi C_\pi^T S_{\pi q}^T S_{\pi q} C_\pi$$
$$+2 C_\pi \Gamma_\pi. \quad (12)$$

Because of the KKT optimality condition, $C_\pi \geq 0$. Hence, $\Lambda_\pi \circ C_\pi = 0$ and $\nabla_{C_\pi} = 0$. This yields

$$C_\pi \leftarrow C_\pi \circ \sqrt{\frac{F(C_\pi)}{M(C_\pi)}}, \quad (13)$$

where

$$F(C_\pi)$$
$$= 2 \sum_{p\pi \in A} S_{p\pi} C_p R_{p\pi} + 2 \sum_{\pi q \in A} S_{\pi q}^T C_q R_{\pi q}^T$$
$$+2 a_\pi L_\pi^- C_\pi + 4 \sum_{p\pi \in A} b_{p\pi} S_{p\pi} C_p C_p^T S_{p\pi}^T C_\pi$$
$$+4 \sum_{\pi q \in A} b_{\pi q} S_{\pi q}^T C_q C_q^T S_{\pi q} C_\pi, \quad (14)$$

$$
\begin{aligned}
&M(C_\pi)\\
&\quad = C_\pi C_\pi^T[2\sum_{p\pi\in A} S_{p\pi}C_pR_{p\pi} + 2\sum_{\pi q\in A} S_\pi q^T C_q R_{\pi q}^T\\
&\qquad +4\sum_{p\pi\in A} b_{p\pi}S_{p\pi}C_pC_p^TS_{p\pi}^TC_\pi\\
&\qquad +4\sum_{\pi q\in A} b_{\pi q}S_{\pi q}^TC_qC_q^TS_{\pi q}C_\pi]\\
&\qquad +4\sum_{\pi q\in A} b_{\pi q}S_{\pi q}^TS_{\pi q}C_\pi C_\pi^TS_{\pi q}^TS_{\pi q}C_\pi + 2a_\pi L_\pi^- C_\pi\\
&\qquad -4C_\pi C_\pi^T\sum_{\pi q\in A} b_{\pi q}S_{\pi q}^TS_{\pi q}C_\pi C_\pi^TS_{\pi q}^TS_{\pi q}C_\pi, \qquad (15)
\end{aligned}
$$

here,L_π^- is the negative part of L_π, i.e., $L_\pi^- = \frac{|L_\pi|-L_\pi}{2}$. The symbols $\circ$, $\frac{(\cdot)}{(\cdot)}$, and $\sqrt{\cdot}$ symbols used in this equation are all operation symbols based on matrix terms.

2) UPDATING THE R_{PQ} VARIABLE

Because of the intrinsic properties of matrices, $tr(AB) = tr(BA), tr(A) = tr(A^T)$ and $||S_{pq}^T - C_pR_{pq}C_q^T||_F^2 = tr((S_{pq}^T - C_pR_{pq}C_q^T)(S_{pq}^T - C_pR_{pq}C_q^T)^T)$. Therefore, the Lagrangian objective function of $R_{pq}(J(R_{pq}))$ may be expressed as:

$$
\begin{aligned}
J(R_{pq}) &= tr(S_{pq}S_{pq}^T) - 2tr(C_p^TS_{pq}^TC_qR_{pq}^T)\\
&\quad + tr(C_pR_{pq}C_q^TC_qR_{pq}^TC_p^T) + c_{pq}R_{pq} - tr(\Lambda_{pq}R_{pq}^T), \qquad (16)
\end{aligned}
$$

where Λ_{pq} is a symmetric matrix in which every term is a Lagrangian operator. Because $R_{pq} \geq 0$, the Lagrangian of $||R_{pq}||_1$ is $c_{pq}R_{pq}$. Taking the derivative of $J(R_{pq})$ with respect to R_{pq} yields

$$
\nabla_{R_{pq}} = -2C_p^TR_{pq}C_q + 2C_p^TC_pR_{pq}C_q^TC_q + c_{pq} - \Lambda_{pq}. \qquad (17)
$$

Because KKT optimality requires $R_{pq} \geq 0$, then $\nabla_{R_{pq}} = 0$ and $\Lambda_{pq}R_{pq} = 0$. Therefore,

$$
R_{pq} \leftarrow R_{pq} \circ \sqrt{\frac{C_p^TS_{pq}^TC_q}{C_p^TC_pR_{pq}C_q^TC_q + \frac{1}{2}c_{pq}}}. \qquad (18)
$$

In summary, MRTA may be described as follows:

3) CONVERGENCE ANALYSIS

To ensure that the alternating update algorithm in Section C. Causes the objective function to decrease monotonically, an auxiliary function in an EM-like algorithm is used to verify the convergence of the objective function.

Definition 1: $G(u, u')$ is an auxiliary function of $F(u)$ if $G(u, u') \geq F(u)$, *and* $G(u, u) = F(u)$ is an auxiliary function of F(u).

Theorem 1: If G is an auxiliary function of F, then F decreases monotonically with an updating equation in the following form:

$$
u^{(t+1)} = \arg\min_u G(u, u^{(t)}). \qquad (19)
$$

Algorithm 1 MRTA for Clustering

Require: m-mode graph, where the similarity matrix of each mode is $A_p, p = 1, 2, \ldots, m$, the mode correlation matrix is $S_{pq}, p, q = 1, 2, \ldots, m$, and the parameters are $a_p, b_{pq}, c_{pq}, p, q = 1, 2, \ldots, m.$
Ensure: Clustering of each mode $C_p, p = 1, 2, \ldots, m$

```
Normalize all similarity and correlation matrices
for p=1 to m do
  randomly initialize C_p as a value in (0, 1]
end for
for p=1 to m do
  for q=1 to m do
    randomly initialize the cluster association index matrix R_pq
  end for
end for
repeat
  for p=1 to m do
    update C_p according to Equation (13)
  end for
  for p=1 to m do
    for q=1 to m do
      update R_pq according to Equation (18)
    end for
  end for
until convergence
```

Proof 1: Because $G(u^{(t+1)}, u^{(t)}) \leq G(u^{(t)}, u^{(t)})$, Definition 1 then gives $F(u^{(t+1)}) \leq G(u^{(t+1)}, u^{(t)}) \leq G(u^{(t)}, u^{(t)}) = F(u^{(t)})$.
Below, we show that the equation for updating R_{pq} also has an auxiliary function.

Theorem 2: The following is the auxiliary function of $J(R_{pq})$:

$$
\begin{aligned}
&G(R_{pq}, R_{pq}^{(t)})\\
&\quad = -2\sum_{i,k}(C_p^TS_{pq}^TC_q)_{ik}(R_{pq}^{(t)})_{ik}\left(1 + \frac{log(R_{pq})_{ik}}{(R_{pq}^{(t)})_ik}\right)\\
&\qquad + \sum_{i,k}\frac{(C_p^TC_pR_{pq}^{(t)}C_q^TC_q)_{ik}(R_{pq})_{ik}^2}{(R_{pq}^{(t)})_{ik}}\\
&\qquad + c_{pq}\sum_{i,k}\frac{(R_{pq})_{ik}^2 + (R_{pq}^{(t)})_{ik}^2}{2(R_{pq}^{(t)})_{ik}}\\
&\qquad - \sum_{i,k}(\Lambda_{pq})_{ik}(R_{pq}^{(t)})_{ik}\left(1 + \frac{log(R_{pq})_{ik}}{(R_{pq}^{(t)})_{ik}}\right). \qquad (20)
\end{aligned}
$$

Proof 2: The function $F(R_{pq})$ of $J(R_{pq})$ that depends only on R_{pq} is known to be expressed as follows:

$$
\begin{aligned}
F(R_{pq}) &= -2tr(C_p^TS_{pq}^TC_qR_{pq}^T) + tr(C_pR_{pq}C_q^TC_qR_{pq}^TC_p^T)\\
&\quad + c_{pq}R_{pq} - tr(\Lambda_{pq}R_{pq}^T).
\end{aligned}
$$

The first term of $F(R_{pq})$ is then

$$-2tr(C_p^T S_{pq}^T C_q R_{pq}^T) = -2\sum_{i,k}(C_p^T S_{pq}^T C_q)_{ik}(R_{pq})_{ik}.$$

Because $Z \geq 1 + log(Z)$ is known, it follows that $Z \geq 0$. If $Z = \frac{(R_{pq})_{ik}}{(R_{pq}^{(t)})_{ik}}$, then

$$(R_{pq}^{(t)})_{ik}\left(1 + log\frac{(R_{pq})_{ik}}{(R_{pq}^{(t)})_{ik}}\right) \leq (R_{pq})_{ik}.$$

Hence, the first term of $G(R_{pq}, R_{pq}^{(t)})$ must be greater than or equal to the first term of $F(R_{pq})$:

$$-2\sum_{i,k}(C_p^T S_{pq}^T C_q)_{ik}(R_{pq}^{(t)})_{ik}(1 + log\frac{(R_{pq})_{ik}}{(R_{pq}^{(t)})_{ik}}) \geq -2tr(C_p^T S_{pq}^T C_q R_{pq}^T).$$

The second term of $F(R_{pq})$ is $tr(C_p R_{pq} C_q^T C_q R_{pq}^T C_p^T) = tr(R_{pq}^T C_p^T C_p R_{pq} C_q^T C_q)$. According to Theorem 6 in [29]: The second term of $G(R_{pq}, R_{pq}^{(t)})$ is then greater than or equal to the second term of $F(R_{pq})$:

$$\sum_{i,k}\frac{(C_p^T C_p R_{pq}^{(t)} C_q^T C_q)_{ik}(R_{pq})_{ik}^2}{(R_{pq}^{(t)})_{ik}} \geq tr(C_p R_{pq} C_q^T C_q R_{pq}^T C_p^T).$$

Given the inequality where $X^2 + Y^2 \geq 2XY$, $X = (R_{pq}^{(t)})_{ik}, Y = (R_{pq})_{ik}$. The third term of $G(R_{pq}, R_{pq}^{(t)})$ is greater than or equal to the third term of $F(R_{pq})$:

$$c_{pq}\sum_{i,k}\frac{(R_{pq})_{ik}^2 + (R_{pq}^{(t)})_{ik}^2}{2(R_{pq}^{(t)})_{ik}} \geq c_{pq}(R_{pq})_{ik}.$$

Like the proof for the first term, the fourth term of $G(R_{pq}, R_{pq}^{(t)})$ is greater than or equal to the fourth term of $F(R_{pq})$:

$$-\sum_{i,k}(\Lambda_{pq})_{ik}(R_{pq}^{(t)})_{ik}(1 + \frac{log(R_{pq})_{ik}}{(R_{pq}^{(t)})_{ik}}) \geq -tr(\Lambda_{pq} R_{pq}^T).$$

Ergo, $G(R_{pq}, R_{pq}^{(t)}) \geq F(R_{pq})$. Furthermore, when $R_{pq}^{(t)} = R_{pq}$, $G(R_{pq}, R_{pq}) = F(R_{pq})$.

The updating process based on $G(R_{pq}, R_{pq}^{(t)})$ is shown below:

$$\frac{\partial G(R_{pq}, R_{pq}^{(t)})}{\partial (R_{pq})_{ik}} = -2(C_p^T S_{pq}^T C_q)_{ik}\frac{(R_{pq}^{(t)})_{ik}}{(R_{pq})_{ik}} + 2\frac{(C_p^T C_p R_{pq}^{(t)} C_q^T C_q)_{ik}(R_{pq})_{ik}}{(R_{pq}^{(t)})_{ik}} + c_{pq}\frac{(R_{pq})_{ik}}{(R_{pq}^{(t)})_{ik}} - (\Lambda_{pq})_{ik}\frac{(R_{pq}^{(t)})_{ik}}{(R_{pq})_{ik}}.$$

Because $\frac{\partial G(R_{pq}, R_{pq}^{(t)})}{\partial (R_{pq})_{ik}} = 0$, then

$$(\Lambda_{pq})_{ik} = -2(C_p^T S_{pq}^T C_q)_{ik} + 2\frac{(C_p^T C_p R_{pq}^{(t)} C_q^T C_q)_{ik}(R_{pq})_{ik}^2}{(R_{pq}^{(t)})_{ik}^2} + c_{pq}\frac{(R_{pq})_{ik}^2}{(R_{pq}^{(t)})_{ik}^2}.$$

Given that R_{pq} is non-negative,

$$(R_{pq}^{(t+1)})_{ik} \leftarrow (R_{pq}^{(t)})_{ik}\sqrt{\frac{(C_p^T S_{pq}^T C_q)_{ik}}{(C_p^T C_p R_{pq}^{(t)} C_q^T C_q)_{ik} + \frac{1}{2}c_{pq}}}.$$

Because $G(R_{pq}, R_{pq}^{(t)})$ is the auxiliary function of $J(R_{pq})$, the updating equation causes $J(R_{pq})$ to decrease monotonically.

Theorem 3: The following is the auxiliary function of $J(C_\pi)$:

$$\begin{aligned}
&G(C_\pi, C_\pi^{(t)})\\
&= -2\sum_{p\pi\in A}\sum_{i,k}(S_{p\pi} C_p R_{p\pi})_{ik}(C_\pi^{(t)})_{ik}(1 + log\frac{(C_\pi)_{ik}}{(C_\pi^{(t)})_{ik}})\\
&\quad + \sum_{p\pi\in A}\sum_{i,k}\frac{(C_\pi^{(t)} R_{p\pi}^T C_p^T C_p R_{p\pi})_{ik}(C_\pi)_{ik}^2}{(C_\pi^{(t)})_{ik}}\\
&\quad -2\sum_{\pi q\in A}\sum_{i,k}(S_{\pi q}^T C_q R_{\pi q}^T)_{ik}(C_\pi^{(t)})_{ik}(1 + log\frac{(C_\pi)_{ik}}{(C_\pi^{(t)})_{ik}})\\
&\quad + \sum_{\pi q\in A}\sum_{i,k}\frac{(C_\pi^{(t)} R_{\pi q} C_q^T C_q R_{\pi q}^T)_{ik}(C_\pi)_{ik}^2}{(C_\pi^{(t)})_{ik}}\\
&\quad + a_\pi\sum_{i,k}\frac{((L_\pi^+ C_\pi^{(t)})_{ik}(C_\pi)_{ik}^2)}{(C_\pi^{(t)})_{ik}}\\
&\quad - a_\pi\sum_{i,k}(L_\pi^-)_{ik}(C_\pi^{(t)}(C_\pi^{(t)})^T)_{ik}(1 + log\frac{(C_\pi C_\pi^T)_{ik}}{(C_\pi^{(t)}(C_\pi^{(t)})^T)_{ik}})\\
&\quad -2\sum_{p\pi\in A} b_{p\pi}\sum_{i,k}(S_{p\pi} C_p C_p^T S_{p\pi}^T)_{ik}(C_\pi^{(t)}(C_\pi^{(t)})^T)_{ik}\\
&(1 + log\frac{(C_\pi C_\pi^T)_{ik}}{(C_\pi^{(t)}(C_\pi^{(t)})^T)_{ik}})\\
&\quad + \sum_{p\pi\in A} b_{p\pi}\sum_{i,k}\frac{(C_\pi^{(t)})_{ik}(C_\pi)_{ik}^2}{(C_\pi^{(t)})_{ik}}\\
&\quad -2\sum_{\pi q\in A} b_{\pi q}\sum_{i,k}(S_{\pi q}^T C_q C_q^T S_{\pi q})_{ik}(C_\pi^{(t)}(C_\pi^{(t)})^T)_{ik}\\
&(1 + log\frac{(C_\pi C_\pi^T)_{ik}}{(C_\pi^{(t)}(C_\pi^{(t)})^T)_{ik}})\\
&\quad + \sum_{\pi q\in A} b_{\pi q}\sum_{i,k}\frac{(S_{\pi q}^T S_{\pi q} C_\pi^{(t)}(C_\pi^{(t)})^T S_{\pi q}^T S_{\pi q})_{ik}(C_\pi C_\pi^T)_{ik}^2}{(C_\pi^{(t)}(C_\pi^{(t)})^T)_{ik}}\\
&\quad - \sum_{i,k}(\Lambda_\pi)_{ik}(C_\pi^{(t)})_{ik}(1 + log\frac{(C_\pi)_{ik}}{(C_\pi^{(t)})_{ik}})\\
&\quad + \sum_{i,k}\frac{(C_\pi^{(t)}\Gamma_\pi)_{ik}(C_\pi)_{ik}^2}{(C_\pi^{(t)})_{ik}}.
\end{aligned} \tag{21}$$

Proof 3: Let

$$J(C_\pi) = \sum_{p\pi\in A}||S_{p\pi}^T - C_p R_{p\pi} C_\pi^T||_F^2 + \sum_{\pi q\in A}||S_{\pi q}^T - C_\pi R_{\pi q} C_q^T||_F^2$$

$$+a_\pi tr(C_\pi^T L_\pi C_\pi) + \sum_{p\pi\in A} b_{p\pi}||S_{p\pi}C_p(S_{p\pi}C_p)^T - C_\pi C_\pi^T||_F^2 + \sum_{\pi q\in A} b_{\pi q}||S_{\pi q}C_\pi(S_{\pi q}C_\pi)^T - C_q C_q^T||_F^2 - tr(\Lambda_\pi C_\pi^T) + tr(\Lambda_\pi(C_\pi^T C_\pi - I_{k_\pi}).$$

If only the part of the expression that is relevant to C_π is considered, then

$$\begin{aligned} F(C_\pi) &= \sum_{p\pi\in A}(-2tr(C_p^T S_{p\pi}^T C_\pi R_{p\pi}^T) + tr(C_p R_{p\pi} C_\pi^T C_\pi R_{p\pi}^T C_p^T)) \\ &+ \sum_{\pi q\in A}(-2tr(C_\pi^T S_{\pi q}^T C_q R_{\pi q}^T) + tr(C_\pi R_{\pi q} C_q^T C_q R_{\pi q}^T C_\pi^T)) \\ &+ a_\pi tr(C_\pi^T L_\pi^+ C_\pi) - a_\pi tr(C_\pi^T L_\pi^- C_\pi) \\ &+ \sum_{p\pi\in A} b_{p\pi}(-2tr(S_{p\pi}C_p C_p^T S_{p\pi}^T C_\pi C_\pi^T) + tr(C_\pi C_\pi^T)) \\ &+ \sum_{\pi q\in A} b_{\pi q} tr(S_{\pi q}C_\pi C_\pi^T S_{\pi q}^T S_{\pi q} C_\pi C_\pi^T S_{\pi q}^T \\ &- 2S_{\pi q}C_\pi C_\pi^T S_{\pi q}^T C_q C_q^T) - tr(\Lambda_\pi C_\pi^T) \\ &+ tr(_\pi(C_\pi^T C_\pi - I_{k_\pi}). \end{aligned}$$

Therefore, according to the properties of traces $tr(ABCD) = tr(BCDA) = tr(CDAB) = tr(DABC)$, $tr(A^T B) = tr(AB^T) = tr(B^T A) = tr(BA^T) = \sum_{i,k} A_{ik}B_{ik}$ We have

$$tr(C_p^T S_{p\pi}^T C_\pi R_{p\pi}^T) = tr(R_{p\pi}^T S_{p\pi}^T C_p^T C_\pi) = \sum_{i,k}(C_p S_{p\pi} R_{p\pi})_{ik}(C_\pi)_{ik}.$$

because $(C_\pi^{(t)})_{ik}(1 + log\frac{(C_\pi)_{ik}}{(C_\pi^{(t)})_{ik}}) \leq (C_\pi)_{ik}$ Therefore,

$$\sum_{i,k}(C_p S_{p\pi} R_{p\pi})_{ik}(C_\pi^{(t)})_{ik}(1+log\frac{(C_\pi)_{ik}}{(C_\pi^{(t)})_{ik}}) \leq tr(C_p^T S_{p\pi}^T C_\pi R_{p\pi}^T).$$

Similarly,

$$-2tr(S_{\pi q}^T C_q R_{\pi q}^T C_\pi^T) = -2\sum_{i,k}(S_{\pi q}^T C_q R_{\pi q}^T)_{ik}(C_\pi)_{ik} \leq -2\sum_{i,k}(S_{\pi q}^T C_q R_{\pi q}^T)_{ik}(C_\pi^{(t)})_{ik}(1 + log\frac{(C_\pi)_{ik}}{(C_\pi^{(t)})_{ik}}).$$

According to Theorem 6 in [29],

$$tr(C_\pi R_{\pi q} C_q^T C_q R_{\pi q}^T C_\pi^T) = tr(C_\pi^T C_\pi R_{\pi q} C_q^T C_q R_{\pi q}^T) \leq \sum_{i,k}\frac{(C_\pi^{(t)} R_{\pi q} C_q^T C_q R_{\pi q}^T)_{ik}(C_\pi)_{ik}^2}{(C_\pi^{(t)})_{ik}}.$$

No further elaboration is provided since the proof is similar to the one given above. The auxiliary function described above is thus obtained.

The convergence of $F(C_\pi)$ when updated by $C_\pi \leftarrow C_\pi \circ \sqrt{\frac{F(C_\pi}{M(C_\pi)}}$ is similar to the updating process based on $G(R_{pq}, R_{pq}^{(t)})$. Because $G(C_\pi, C_\pi^{(t)}) \geq F(C_\pi), G(C_\pi, C_\pi) = F(C_\pi)$, and $\frac{\partial G(C_\pi, C_\pi^{(t)})}{\partial (C_\pi)_{ik}} = 0$ are established, it can be readily demonstrated that $C_\pi \leftarrow C_\pi \circ \sqrt{\frac{F(C_\pi}{M(C_\pi)}}$ under the KKT optimality condition. However, we have chosen to omit this detail since the proof is similar to the one above and to limit the length of this paper.

4) COMPLEXITY ANALYSIS

The time complexity of MRTA is examined in this section. First, the updating of R_{pq} and C_π involves matrix multiplications. The time complexity of C_π is $O(tm(\hat{n}^2\hat{k} + \hat{n}\hat{k}^2 + \hat{n}^3))$, where m is the number of modes in the network, $\hat{n}$ is the maximum value of n_p, $\hat{k}$ is the maximum value of k_p, and t is the number of iterations needed for convergence. For example, Eq.(13) shows that $\hat{n}^3$ can be obtained from the matrix multiplication of $S_{\pi q}^T S_{\pi q}$. The time complexity of R_{pq} is $O(tm(\hat{n}^2\hat{k} + \hat{n}\hat{k}^2 + \hat{k}^3))$. Similarly, Eq. (18) shows that $\hat{n}^2\hat{k}$ can be calculated from the matrix multiplication of $C_p^T C_p$. Therefore, the total time complexity of MRTA is $O(tm(\hat{n}^2\hat{k} + \hat{n}\hat{k}^2 + \hat{n}^3))$.

IV. EXPERIMENTAL VALIDATION

A. DATA SETS

To evaluate the efficiency of our proposed MRTA, we test them on three social network datasets were used in this experiment: Politics-UK, Politics-IE and DBLP in the following experiments. We give more descriptions for these three social datasets:

The Politics-UK dataset contained the Twitter content and social network relationships of 419 Members of Parliament (MPs) in the United Kingdom from five different parties. The Politics-IE dataset contained the Twitter content and social network relationships of 348 MPs in Ireland from seven different parties. The association relationships and data contents of the Politics-UK and Politics-IE datasets consisted of the follower relationships and tweets of the MPs.

The DBLP bibliography contained 8293 papers published by 6604 authors in 16 top-level conferences across four fields of study. The details are presented in Table 2. We focused on the author and conference modes of the DBLP dataset. The relationships between the author and conference modes are shown in Fig. 1.In the DBLP dataset, the similarity matrix A_1 of the G_1 data mode (conferences) was the cosine between the eigenvectors of two conferences, i.e., $A_1 = cos(F_1, F_2)$. The G_2 data mode (authors) represented collaborations between authors; in other words, each $A_2^{(i,j)}$ term in the similarity matrix A_2 of G_2 represents the number of collaborations between the i-th and j-th authors. The author-conference correlation matrix reflects the number of articles published by an author in a conference, while the paper-conference correlation matrix reflects the publication of a paper in a conference. Finally, the author- paper correlation matrix reflects the writing of a paper by an author.

TABLE 2. Description of various symbols.

Symbols	Explanation
m	Number of modes
S	Relationship matrix sets between mode and mode $S_{pq}, p, q = 1, 2, ..., m$
p, q, π	Digital label of mode
n_p	Number of nodes of p-mode network
k_p	Number of clustering of p-mode network
$(x_p)_i$	The i-th node of p-mode network
L_p	Laplacian matrix of p-mode network
C_p	Cluster indicator matrix of p-mode network
A_p	Inner-mode relationship matrix of p-mode network
CS_p	Clustering similarity matrix of p-mode network
S_{pq}	Intra-mode relationship matrix between p-mode network and q-mode network
R_{pq}	Cluster correlation matrix between p-mode network and q-mode network

TABLE 3. Brief Statistics of Datasets.

	Politics-UK	Politics-IE	DBLP
Number of users	419	348	6604
Number of messages	72314	45805	8238
Number of social relations	19868	16856	17029

B. PARAMETER SELECTION AND EVALUATION METRICS

To facilitate the assessment of our algorithm, all of the algorithms was tuned for leave-one-out cross validation. The proposed MRTA had three parameters: the same-mode similarity regularization parameter (a_p), two-mode similarity regularization parameter (b_{pq}), and robustness parameter (c_{pq}). The same values were assigned to these parameters in all of the following experiments. The number of clusters was equal to the number of types of data, and the performance of each algorithm was evaluated according to the clustering accuracy (ACC) and normalized mutual information (NMI) indices. The evaluating indices were defined as follows:

Definition 2: Clustering accuracy (ACC):

$$ACC = \frac{\sum_{i=1}^{n} \delta(\tilde{C}_i, map(\tilde{C}_i))}{n}. \tag{22}$$

$\tilde{C}_i$ and C_i correspond to the cluster indicator vector of x_i and the cluster indicator vector calculated by the clustering algorithm, respectively. $\delta(x, y)$ is a delta function that is 1 if x and y are identical and 0 otherwise. The map function is the optimal mapping function. A higher ACC value indicates that the clustering algorithm is more effective.

Definition 3: Normalized mutual information (NMI):

$$NMI = \frac{MI(\tilde{C}, C)}{max(H(\tilde{C}), H(C))}. \tag{23}$$

$MI(\tilde{C}, C)$ is the mutual information of $\tilde{C}$ and C, while $H(\tilde{C})$ and $H(C)$ are the information entropies of $\tilde{C}$ and C, respectively. A higher NMI indicates that the clustering algorithm is better quality.

C. COMPARATIVE ALGORITHMS

To demonstrate the effectiveness of our algorithm, we compared our algorithm with Spectral clustering (SC) [32], the Non-symmetric matrix factorization(SNMF) [6], the Non-negative tri-factorization NMTF [20] and The relational topic model (RTM) [30]. The comparative algorithms are all listed below. MRTA was compared to the following algorithms:

(1) SC is a matrix decomposition algorithm based on the graph similarity matrix A that calculates the k the largest singular eigenvectors of this matrix. Graph clustering is then performed by applying the k-means clustering algorithm to these eigenvectors.
(2) SNMF is a binary matrix factorization algorithm based on the graph similarity matrix A that uses an optimization algorithm to obtain the values of the cluster indicators.
(3) NMTF is a matrix tri-factorization algorithm based on the correlation matrix S that uses an optimization algorithm to obtain the values of the cluster indicators.
(4) RTM models inter-mode links S as a random binary number. RTM is a clustering matrix factorization algorithm that uses the graph similarity matrix A as a reference and inter-mode links S as regularization terms.

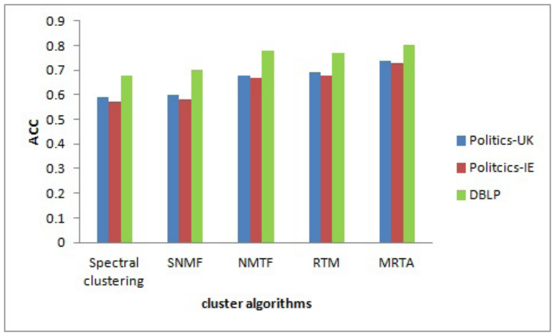


FIGURE 3. ACC with the Politics-UK, Politics-IE, and DBLP datasets.

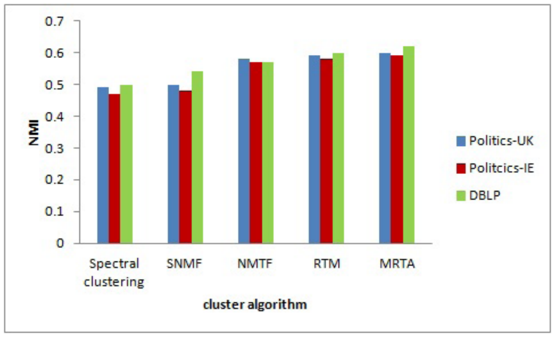
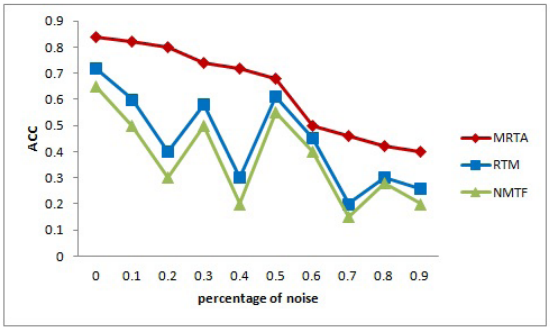


FIGURE 4. NMI with the Politics-UK, Politics-IE, and DBLP datasets.

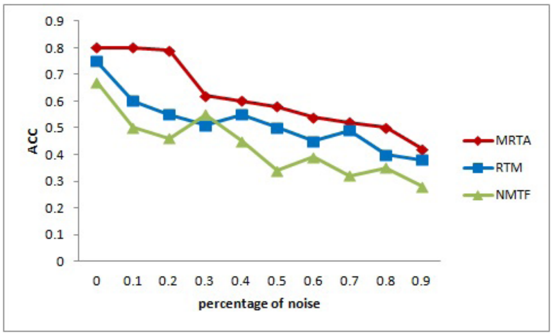
D. PERFORMANCE ANALYSIS

The effectiveness of MRTA was assessed via comparisons with conceptually similar matrix factorization algorithms. Spectral clustering and SNMF are matrix factorization algorithms meant for single-mode data, whereas NMTF, RTM, and MRTA are matrix factorization algorithms meant for multi-mode data. The proposed MRTA is based on matrix factorization but also considers the data content and the relationship between the mode correlation matrix and cluster indicator matrix.

In the following experiment, we compiled the clustering results obtained by the authors of each method. The clustering



(a)



(b)

FIGURE 5. **ACC as a function of the percentage increase of noise in R_{12} on with the simulated network: (a) G_1 and (b) G_2.**

TABLE 4. **Number of Iterations to Converge.**

dataset	size of network G_1	size of network G_2	number of iterations
synthetic	12	15	6
politics-IE	348	512	40
politics-Uk	419	823	48
DBLP	80	6604	274

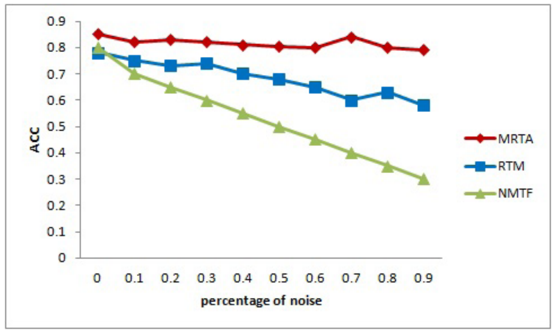
quality of each clustering algorithm was then compared via ACC and NMI.

The averaged clustering results obtained by MRTA and the four above algorithms after 30 trials are shown in Figs. 3 and 4. For multi-mode social networks, matrix factorization methods that considered correlated relations showed markedly better clustering effectiveness than matrix factorization methods that only considered the contents of the social network data. This is because social relationships reflect a users interests more accurately than social network data because a user can publish a wide variety of data in social networks. MRTA had the best clustering effectiveness of the five algorithms because it also incorporates a correlation matrix as well as the relationship between the data content and cluster indicators.

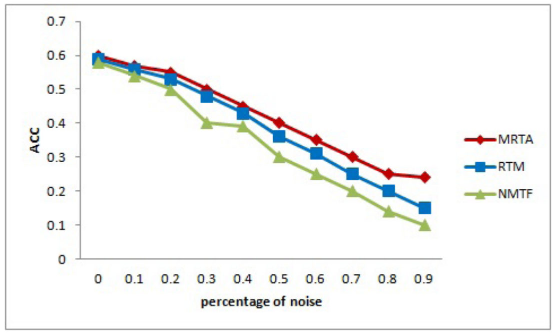
E. ROBUSTNESS ANALYSIS

Two simulated data networks (G_1 and G_2) were constructed to compare the robustness of MRTA to that of RTM and NMTF. Only multi-mode clustering algorithms were included in this comparison because the robustness analysis was conducted by adding noise to the correlation data.

Fig. 5 shows that MRTA produced better clustering results than NMTF and RTM with varying noise levels.

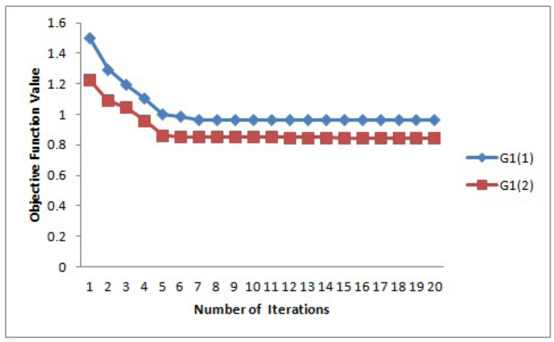


(a)

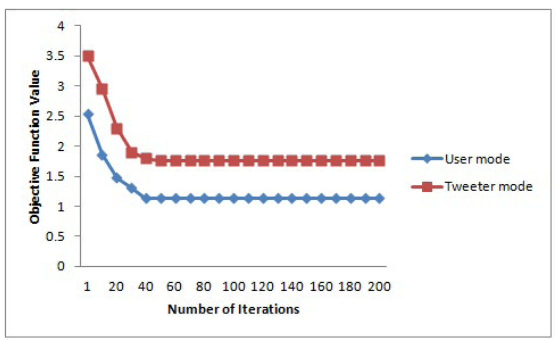


(b)

FIGURE 6. **ACC as a function of the percentage increase of noise in R_{12} on with the DBLP dataset: (a) conference network G_1 and (b) author network G_2.**



(a)



(b)

FIGURE 7. **Convergence curves of the MRTA with the simulated and Politics-IE datasets: (a) simulated network G_1, (b) the Politics-IE dataset.**

Fig. 5(a) shows that MRTA had a gradually decreasing ACC with increasing noise, whereas RTM and NMTF exhibited substantial fluctuations in ACC with increasing noise. Thus, MRTA is more robust than other currently available multi-mode clustering algorithms.

To further assess the robustness of MRTA, it was compared with the above algorithms when noise was added in varying proportions to the author-conference correlation matrix of the DBLP dataset (a real dataset). The details of this dataset are given in Table 2. Fig. 6 shows that the clustering of the conference mode was generally immune to noise, whereas the clustering of the author mode was sensitive to noise. This was mainly because similar authors are more likely to publish articles in similar conferences. MRTA still outperformed all other algorithms in terms of robustness.

F. CONVERGENCE ANALYSIS

The convergence analysis was conducted by plotting the convergence of MRTA when clustering the simulated dataset and Politics-IE dataset. Fig. 7 illustrates the convergence of the simulated dataset and Politics-IE dataset with a varying number of iterations. TABLE 4 describes the number of iterations to converge. The number of iterations required for convergence increased with the network size. Nonetheless, the objective function always converged to a fixed value as the number of iterations increased.

V. CONCLUSION

We propose MRTA as a novel method for clustering multi-mode social networks. MRTA is a clustering algorithm that is well-suited to the characteristics of modern social networks. Because of the non-negativity and orthogonality of cluster indicators in matrix tri-factorization clustering algorithms, MRTA is also constrained by similarity relationships between cluster indicators of the inner-mode or different modes. These constraints are based on the relationship between cluster indicators and the similarity graph of data belonging to the inner-mode as well as the relationship between cluster indicators and the mode correlation graph. This approach provides a more accurate reflection of the cluster indicator relationships of multi-mode social networks under conventional cluster indicator constraints.

The proposed MRTA is based on two clustering relationship graphs: single-mode cluster indicator relationship graphs and multi-mode cluster indicator relationship graphs. These graphs are based on the three features of social network data proposed in this work. MRTA also includes measures to improve its robustness with respect to multi-mode correlation relationships. In reality, network data tend to contain a large number of uncertainties (i.e., noise), which interferes with clustering operations. Therefore, the resistance of the clustering algorithms performance to noise needs to be strengthened to improve its robustness.

An alternating iterative update rule was derived, and the accuracy and convergence of the iterative algorithm were validated via auxiliary functions. MRTA was experimentally validated via performance, robustness, and convergence analyses with simulated and real social network datasets. The numerous experimental analyses showed that MRTA outperformed currently available clustering algorithms in terms of the clustering quality and robustness.

REFERENCES

[1] D. Huang, C.-D. Wang, J. S. Wu, J.-H. Lai, and C. K. Kwoh, "Ultra-scalable spectral clustering and ensemble clustering," *IEEE Trans. Knowl. Data Eng.*, to be published.

[2] C. Shi, Y. Sun, and P. S. Yu, "Current development and trend of mining heterogeneous information networks," *Commun. Comput.*, vol. 13, no. 11, pp. 36–42, 2017.

[3] H. Yu, J. Wang, Y. Bai, W. Yang, and G.-S. Xia, "Analysis of large-scale UAV images using a multi-scale hierarchical representation," *Geo-Spatial Inf. Sci.*, vol. 21, no. 3, pp. 33–44, 2018.

[4] J. Ni, W. Cheng, W. Fan, and X. Zhang, "Self-grouping multi-network clustering," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2017, pp. 1119–1124.

[5] J. Ni, W. Cheng, W. Fan, and X. Zhang, "ComClus: A self-grouping framework for multi-network clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 3, pp. 435–448, Mar. 2018.

[6] D. Kuang, S. Yun, and H. Park, "SymNMF: Nonnegative low-rank approximation of a similarity matrix for graph clustering," *J. Global Optim.*, vol. 62, no. 3, pp. 545–574, 2015.

[7] F. Shang, L. C. Jiao, and F. Wang, "Graph dual regularization non-negative matrix factorization for co-clustering," *Pattern Recognit.*, vol. 45, no. 6, pp. 2237–2250, 2012.

[8] Y. Meng, R. Shang, L. Jiao, W. Zhang, and S. Yang, "Dual-graph regularized non-negative matrix factorization with sparse and orthogonal constraints," *Eng. Appl. Artif. Intell.*, vol. 69, pp. 24–35, Mar. 2018.

[9] P. Luo, J. Peng, Z. Guan, and J. Fan, "Dual regularized multi-view non-negative matrix factorization for clustering," *Neurocomputing*, vol. 294, pp. 1–11, Jun. 2018.

[10] P. Luo, J. Peng, and J. Fan, "Dual hybrid manifold regularized non-negative matrix factorization with discriminability for image clustering," in *Proc. Int. Conf. Frontiers Adv. Data Sci.*, Oct. 2017, pp. 7–11.

[11] X. Zhang, Z. Wang, L. Zong, and H. Yu, "Multi-view clustering via graph regularized symmetric nonnegative matrix factorization," in *Proc. IEEE Int. Conf. Cloud Comput. Big Data Anal. (ICCCBDA)*, Jul. 2016, pp. 109–114.

[12] Y. Ma, X. Hu, T. He, and X. Jiang, "Multi-view clustering microbiome data by joint symmetric nonnegative matrix factorization with Laplacian regularization," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2016, pp. 625–630.

[13] A. Kumar and H. Daumé, III, "A co-training approach for multi-view spectral clustering," in *Proc. Int. Conf. Mach. Learn.*, 2011, pp. 393–400.

[14] Q. Gu and J. Zhou, "Co-clustering on manifolds," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Paris, France, vol. 1, Jun./Jul. 2009, pp. 359–368.

[15] C. Ding, X. He, H. Simon, and R. Jin, "On the equivalence of nonnegative matrix factorization and K-means—Spectral clustering," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2005. doi: 10.1137/1.9781611972757.70.

[16] S. Bickel and T. Scheffer, "Multi-view clustering," in *Proc. IEEE Int. Conf. Data Mining*, Nov. 2004, pp. 1–8.

[17] D. Kong, C. Ding, and H. Huang, "Robust nonnegative matrix factorization using L21-norm," in *Proc. ACM Int. Conf. Inf. Knowl. Manage.*, 2011, pp. 673–682.

[18] L. Ou-Yang, H. Yan, and X.-F. Zhang, "A multi-network clustering method for detecting protein complexes from multiple heterogeneous networks," *BMC Bioinf.*, vol. 18, no. S13, 2017, Art. no. 463.

[19] W. Cheng, Z. Guo, X. Zhang, and W. Wang, "CGC: A flexible and robust approach to integrating co-regularized multi-domain graph for clustering," *ACM Trans. Knowl. Discovery Data*, vol. 10, no. 4, 2016, Art. no. 46.

[20] N. Del Buono and G. Pio, "Non-negative matrix tri-factorization for co-clustering: An analysis of the block matrix," *Inf. Sci.*, vol. 301, pp. 13–26, Apr. 2015.

[21] L. Zhu, A. Galstyan, J. Cheng, and K. Lerman, "Tripartite graph clustering for dynamic sentiment analysis on social media," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2014, pp. 1531–1542.

[22] N. M. Arqué and D. F. Nettleton, "Analysis of on-line social networks represented as graphs—Extraction of an approximation of community structure using sampling," in *Proc. Int. Conf. Modeling Decis. Artif. Intell.* Berlin, Germany: Springer-Verlag, 2012, pp. 149–160.

[23] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 12, pp. 7821–7826, 2002.

[24] C. Bauckhage, "k-means clustering is matrix factorization," 2015, *arXiv:1512.07548*. [Online]. Available: https://arxiv.org/abs/1512.07548

[25] B. Long, Z. Zhang, X. Wú, and P. S. Yu, "Spectral clustering for multi-type relational data," in *Proc. 23rd Int. Conf. Mach. Learn.*, 2006, pp. 585–592.

[26] U. Shaham, K. Stanton, H. Li, B. Nadler, R. Basri, and Y. Kluger, "SpectralNet: Spectral clustering using deep neural networks," 2018, *arXiv:1801.01587*. [Online]. Available: https://arxiv.org/abs/1801.01587

[27] 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, Oct. 1999.

[28] H. Wang, F. Nie, H. Huang, and F. Makedon, "Fast nonnegative matrix tri-factorization for large-scale data co-clustering," in *Proc. 22nd Int. Joint Conf. Artif. Intell.*, Catalonia, Spain, Jul. 2011, pp. 1–6.

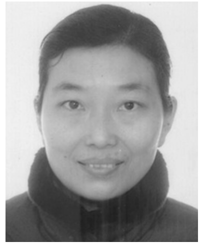
[29] C. Ding, T. Li, W. Peng, and H. Park, "Orthogonal nonnegative matrix t-factorizations for clustering," in *Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2006, pp. 126–135.

[30] A. Zhang, J. Zhu, and B. Zhang, "Sparse relational topic models for document networks," in *Machine Learning and Knowledge Discovery in Databases*. Berlin, Germany: Springer, 2013.

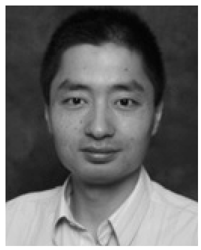
[31] J. Chang and D. M. Blei, "Relational topic models for document networks," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2009, pp. 81–88.

[32] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

LI NI received the bachelor's degree in computer science from the University of National Defense Technology, Changsha, China, in 2002, and the master's degree in computer software and theory from Xiangtan University, in 2009. She is currently pursuing the Ph.D. degree with Hunan University, Changsha, China. Her research interests include machine learning, data mining, and social community detection.



PENG MANMAN received the bachelor's degree from the Department of Computer Science, Hunan University, in 1985, the master's degree in computer science from Hunan University, in 1988, and the Ph.D. degree from the College of Computer and Communication, Hunan University, in 2006. After graduation, she was involved in Postgraduate Teaching. She is currently a Professor and a Doctoral Tutor. Her research interests include data mining, computer architecture, and big data computing.



WU QIANG received the master's degree in computer science from Tsinghua University, in 2002, and the Ph.D. degree from the Imperial College of Technology, in 2009. He is currently an Associate Professor and a master's degree Supervisor. His research interests include data mining, computer architecture, big data computing, and machine learning.