

# Reconciling Multiple Social Networks Effectively and Efficiently: An Embedding Approach

Zhongbao Zhang, Li Sun, Sen Su, Jielun Qu, Gen Li

**Abstract**—Recently, reconciling social networks, identifying the accounts belonging to the same individual across social networks, receives significant attention from both academic and industry. Most of the existing studies have limitations in the following three aspects: multiplicity, comprehensiveness and robustness. To address these limitations, we rethink this problem and, for the first time, robustly and comprehensively reconcile multiple social networks. In this paper, we propose two frameworks, MASTER and MASTER+, i.e., across Multiple social networks, integrate Attribute and STructure Embedding for Reconciliation. In MASTER, we first design a novel Constrained Dual Embedding model, simultaneously embedding and reconciling multiple social networks, to formulate this problem into a unified optimization. To address this optimization, we then design an effective NS-Alternating algorithm and prove it converges to KKT points. To further speed up MASTER, we propose a scalable framework, namely MASTER+. The core idea is to group accounts into clusters and then perform MASTER in each cluster in parallel. Specifically, we design an efficient Augmented Pre-Embedding model and Balance-aware Fuzzy Clustering algorithm for the high efficiency and the high accuracy. Extensive experiments demonstrate that both MASTER and MASTER+ outperform the state-of-the-art approaches. Moreover, MASTER+ inherits the effectiveness of MASTER and enjoys higher efficiency.

**Index Terms**—social network reconciliation, network alignment, network embedding, matrix factorization, semidefinite programming

## 1 INTRODUCTION

NOWADAYS, social network is becoming increasingly important in people's lives. People prefer to join in multiple social networks to enjoy more social network services, e.g., Twitter for news, Facebook for friends and LinkedIn for jobs. However, these accounts are often independent from each other, that is, the correspondence information of the accounts across these social networks is missing. How to identify the accounts belonging to the same individual poses the problem of *reconciling social networks*. This problem illustrated in Fig. 1 routinely finds itself in a wide spectrum of applications, ranging from network fusion [32], link prediction [35], information diffusion [31] to cross-domain recommendation [14]. On the whole, reconciling social networks paves the way for the social network analysis.

The problem of reconciling social networks still largely remains open as most of the existing approaches have

- Some preliminary results of this paper were published in IJCAI 2018 [21]. This journal version significantly strengthens the previous conference version by adding new technical contents. We highlight the following changes: (1) We added a new section (Section 5) titled *Convergence Analysis* to give the lemmas and theorem on convergence. The key derivation of the lemmas are included in this section; (2) We added another new section (Section 6) titled *MASTER+*. In this section, we propose the scalable *MASTER+* framework, which greatly enhances the efficiency of *MASTER* while preserving its effectiveness; (3) We conducted more experiments and added the experiment result in Fig. 2, Table 2 and 3 in Section 7.3; (4) We added more technical details in Section 4; (5) We added several reference papers in the related work (Section 8); (6) We significantly improved the presentation of this paper by rewriting several sections of this paper, e.g., Section 1, 3 and 9.
- Zhongbao Zhang, Li Sun, Su Sen and Gen Li are with State Key Laboratory of Networking and Switching Tech., Beijing University of Posts and Telecomm., China. Jielun Qu is with New York University, USA. Corresponding author: Sen Su.  
E-mail: {zhongbaobz, l.sun, susen, ligen}@bupt.edu.cn; jq653@nyu.edu.

Manuscript received XX,XX 2018.

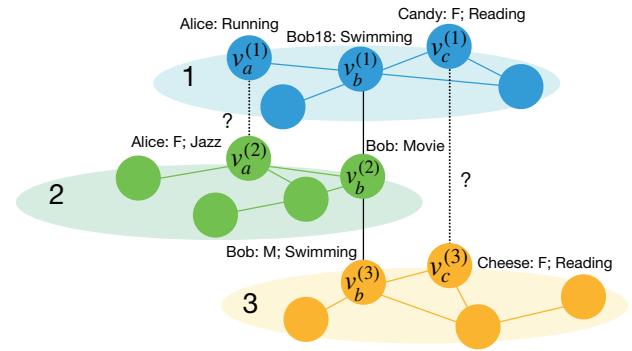


Fig. 1: This figure gives a toy example to demonstrate the task of reconciling social networks. Graphs of different colors denote different social networks. Nodes (accounts) are associated with attributes including profiles and contents (e.g., screen name and favored topics). Black lines denote the correspondences among the nodes. The accounts linked via full lines are the few common users known in advance, and the task of reconciling social networks is to identify whether or not the accounts, linked via dotted lines with mark, belong to the same individual.

limitations in following the three aspects:

- **Multiplicity:** In real world, people usually have several social network accounts. However, most of the existing methods [9], [10], [11], [12], [15] focus on reconciling only two social networks and cannot pairwise reconcile multiple social networks due to the global inconsistency.
- **Comprehensiveness:** Social networks usually have two categories of spaces, i.e., attribute space and structure space. However, most of existing methods [10], [11], [15], [16] cannot comprehensively explore these two spaces for reconciling multiple social net-

works.

- **Robustness:** The social network is noisy and most of the existing methods [9], [12], [30], [33] struggle in defining effective features sensitive to data. Therefore, they are still far away from effectively reconciling networks due to the lack of robustness.

These limitations motivate us to rethink: *Can we robustly and comprehensively reconcile multiple social networks?*

The answer is YES! In this paper, we, for the first time, propose to robustly and comprehensively reconcile multiple social networks. Nevertheless, it faces three challenging issues: (1) **Model:** To the best of our knowledge, there is no embedding model reconciling multiple social networks in the literatures. Both spaces of multiple networks should be integrated and, moreover, the problem of global inconsistency ought to be addressed inherently in the proposed model. (2) **Optimization:** The network embedding problem is often explicitly or implicitly formulated as an optimization problem. In this paper, the optimization problem behind tends to be non-convex and hence much tougher. (3) **Scalability:** The real-world social networks are growing larger and larger. Therefore, how to scale up the proposed algorithm is another important issue and, moreover, the accuracy ought to be ensured in this process.

To address the issue (1) and (2), we propose a novel semi-supervised embedding framework, MASTER [21], i.e., across Multiple social networks, integrate Attribute and STructure Embedding for Reconciliation.

In the MASTER framework, we design a novel Constrained Dual Embedding (CDE) model to formulate the reconciling social network problem. The core idea of CDE is to simultaneously embed and reconcile multiple social networks in the joint latent space via *uni-* and *joint-embedding*. For uni-embedding, we perform collaborative matrix factorization to independently embed each network into a latent space, which collaboratively captures the observations of both attribute space and structure space. For joint-embedding, we align these embedded social networks at the known correspondences to construct the joint latent space for consistent reconciliation. With some algebraic reformulation, we give the unified optimization for robustly reconciling multiple social networks.

To address the optimization of CDE, we design an effective NS-Alternating algorithm to approach the optima of the non-convex matrix optimization. Specifically, inspired by a recent advance in computational mathematics, we reformulate the optimization problem with an auxiliary variable. In this reformulated problem, we alternately solve the representation matrices subproblem and the kernel matrices subproblem via first-order method and semidefinite programming, respectively. Moreover, we make analysis of the convergence property in depth and give the sufficient condition of Karush-Kuhn-Tucker (KKT) convergence.

Finally, to address the issue (3), we propose a scalable framework based on MASTER, namely MASTER+. The core idea is to group the accounts, possibly belonging to the same individuals, into clusters and perform MASTER in each cluster in parallel. Regarding the clustering, the key challenge lies in *how to enhance the efficiency while ensuring the accuracy*. To this end, specifically, we first design an

Notation	Definition
$S^{(m)}$	$m^{th}$ social network ( $m \in [1, M], m \in Z$ )
$N^{(m)}$	number of the users in $S^{(m)}$
$\mathbf{G}^{(m)}$	adjacency matrix of $S^{(m)}$
$\mathbf{A}^{(m)}$	attribute matrix of $S^{(m)}$
$\hat{\mathcal{L}}$	known correspondences set
$\mathbf{H}^{(m)}$	representation matrix of $S^{(m)}$
$d$	dimension of the embeddings
$\alpha, \beta, \gamma$	parameters of CDE model
$\mathbf{G}_P$	adjacency matrix of the augmented graph
$\mathbf{A}_P$	attribute matrix of the augmented graph
$\mathbf{H}_P$	representation matrix of the augmented graph
$\mathbf{U}$	membership matrix
$m, \zeta$	parameter of BFC algorithm

TABLE 1: Main symbols and definitions in the paper efficient Augmented Pre-Embedding (APE) model, employing symmetric matrix factorization, to embed the accounts of multiple networks into a vector space. We then design a novel Balance-aware Fuzzy Clustering (BFC) algorithm for soft clustering in the embedded space. We introduce the balance-aware regularizer to the matrix form of fuzzy C-Means, formulating a matrix optimization. This APE-BFC process of balance-aware and soft clustering inherently helps MASTER+ achieve the dual guarantee of high efficiency and high accuracy.

We validate these two proposed frameworks through extensive experiments on real-world datasets. Both MASTER and MASTER+ significantly outperform several state-of-the-art methods. Moreover, MASTER+ inherits the effectiveness of MASTER and enjoys higher efficiency.

To summarize, the main contributions are listed below:

- To the best of our knowledge, this work is the first attempt to robustly reconcile multiple social networks, comprehensively exploiting attribute and structure information via an embedding approach.
- We first design the MASTER framework, where a novel CDE model is proposed to formulate this problem into a unified optimization and NS-Alternating algorithm is proposed to address the optimization. Moreover, we prove the (KKT) convergence of the algorithm.
- We further design the scalable MASTER+ framework to speed up MASTER, where the efficient APE-BFC process of balance-aware and soft clustering is proposed to achieve the dual guarantee of high efficiency and high accuracy.
- We conducted extensive experiments on the real-world datasets and the experiment results demonstrate the superiority of our proposed frameworks.

## 2 PROBLEM DEFINITION

In this paper, we consider a set of  $M$  social networks  $\{S^{(m)}\}$ . A social network  $S^{(m)}$  with  $N^{(m)}$  users is denoted as  $(\mathbf{G}^{(m)}, \mathbf{A}^{(m)})$ . The adjacency matrix  $\mathbf{G}^{(m)} \in \mathbb{R}^{N^{(m)} \times N^{(m)}}$  represents the structure space, where binary  $\mathbf{G}_{ij}^{(m)}$  indicates whether or not a social connection exists between user account  $v_i^{(m)}$  and  $v_j^{(m)}$ .  $\mathbf{G}^{(m)}$  is symmetric as the network is considered to be undirected.  $\mathbf{A}^{(m)} \in \mathbb{R}^{N^{(m)} \times l}$  represents the attribute space and its  $i^{th}$  row  $\mathbf{a}_i^{(m)}$  denotes

the  $l$ -dimensional attribute vector associated with  $v_i^{(m)}$ . The attributes vector includes the features extracted from both the profile information (e.g., gender, location) and the user generated contents (e.g., interests, frequent words). Some user account correspondences can be obtained from user profiles or some third-party platforms. Such information is represented in a label set  $\hat{L} = \{L(m, n)\}$ , where  $L(m, n)$  is the set of known account pairs between  $S^{(m)}$  and  $S^{(n)}$  of the same individual.

Without loss of generality, we assume that social networks are partially overlapped. We formally define the problem of reconciling multiple social networks as follows:

**Problem Definition.** Given the set  $\{S^{(m)}\}$  with labels  $\hat{L}$ , the problem of reconciling multiple social networks is to find a  $\phi^{(m)}$ , mapping the user account to its owner, for each  $S^{(m)}$  so that  $\phi^{(1)}(v_{(\cdot)}^{(1)}) = \dots = \phi^{(m)}(v_{(\cdot)}^{(m)}) = \dots = \phi^{(M)}(v_{(\cdot)}^{(M)})$  if and only if the accounts  $\{v_{(\cdot)}^{(m)}\}$  are of the same individual.

To address this problem, we first propose the MASTER framework. In MASTER, we design a novel model, Constrained Dual Embedding (CDE), to formulate the problem into a unified optimization (Sec. 3). To address the optimization, we design an effective NS-Alternating algorithm (Sec. 4) and give its convergence analysis (Sec. 5). We then propose a parallel framework based on MASTER, namely MASTER+ (Sec. 6).

### 3 MODELING: CONSTRAINED DUAL EMBEDDING

In CDE, we independently embed each social network via *uni-embedding* and simultaneously reconcile these embedded networks via *joint-embedding*.

#### 3.1 Uni-embedding

The goal of uni-embedding is, for each social network  $S^{(m)}$ , to obtain the representation matrix  $\mathbf{H}^{(m)} \in \mathbb{R}^{N^{(m)} \times d}$  ( $d \ll \min\{N^{(m)}\}$ ), whose  $i^{th}$  row  $\mathbf{h}_i^{(m)}$  is the  $d$ -dimensional vector of  $v_i^{(m)}$  in the latent space, capturing the observations of both structure space and attribute space.

To achieve this goal, first, we construct the similarity matrix  $\mathbf{M}^{(m)}$  of structure space. Note that  $\mathbf{G}_{ij}^{(m)}$  encodes the first-order proximity, defined in the study [22], which is measured by whether or not  $v_i^{(m)}$  and  $v_j^{(m)}$  have a direct connection. Obviously, it is necessary to preserve the first-order proximity as it depicts the original structure of the social network. However, the observed edges are usually sparse in the network. For two user accounts with no direct connection, an alternative way to imply the proximity is to measure their neighbors. Intuitively, the more similar their neighbors are, the higher proximity they share. Therefore, we formally define the second-order proximity as follows:

**Definition (second-order proximity).** Given the adjacency matrix (or first-order proximity matrix)  $\mathbf{G}^{(m)}$ , the second-order proximity  $\bar{\mathbf{G}}_{ij}^{(m)}$  between  $v_i^{(m)}$  and  $v_j^{(m)}$  is the similarity between  $\mathbf{G}_i^{(m)}$  and  $\mathbf{G}_j^{(m)}$ , the  $i^{th}$  row and  $j^{th}$  row of  $\mathbf{G}^{(m)}$ .

The inner product similarity is utilized in this paper, i.e.,  $\bar{\mathbf{G}}^{(m)} = \mathbf{G}^{(m)2}$  as  $\mathbf{G}^{(m)}$  is symmetric. To incorporate the

first- and second-order proximity,  $\mathbf{M}^{(m)} = \mathbf{G}^{(m)} + \eta \bar{\mathbf{G}}^{(m)}$  where  $\eta$  is a non-negative weight.

Second, we derive the similarity matrix  $\mathbf{W}^{(m)}$  of attribute space by computing the pairwise inner product of the attribute vector, i.e.,  $\mathbf{W}^{(m)} = \mathbf{A}^{(m)} \mathbf{A}^{(m)T}$ .

Finally, we approximate the pairwise similarity in each space by the inner product of  $\mathbf{h}_i^{(m)}$ . Assume that  $\mathbf{h}_i^{(m)}$  is projected onto structure space and attribute space via different projections  $\varphi(\cdot)$ . We introduce kernel technique to bridge the inner product  $\langle \cdot, \cdot \rangle$  of  $\varphi(\mathbf{h}_i^{(m)})$  with that of  $\mathbf{h}_i^{(m)}$ , i.e.,  $\langle \varphi(\mathbf{h}_i^{(m)}), \varphi(\mathbf{h}_j^{(m)}) \rangle = \mathbf{h}_i^{(m)} \mathbf{K}_\varphi^{(m)} \mathbf{h}_j^{(m)}$ , where  $\mathbf{K}_\varphi^{(m)}$  is semidefinite. Let  $\mathbf{B}^{(m)}$  and  $\mathbf{C}^{(m)}$  denote  $\mathbf{K}_\varphi^{(m)}$  of the projection onto structure space and attribute space, respectively.  $\mathbf{H}^{(m)}$  can be learned by optimization of a collaborative matrix factorization as below:

$$\begin{aligned} \min_{\mathbf{H}^{(m)}, \mathbf{B}^{(m)}, \mathbf{C}^{(m)}} & \frac{\alpha}{2} d\left(\mathbf{M}^{(m)}, \mathbf{H}^{(m)} \mathbf{B}^{(m)} \mathbf{H}^{(m)T}\right) \\ & + \frac{\beta}{2} d\left(\mathbf{W}^{(m)}, \mathbf{H}^{(m)} \mathbf{C}^{(m)} \mathbf{H}^{(m)T}\right) \\ \text{s. t. } & \mathbf{B}^{(m)}, \mathbf{C}^{(m)} \in \mathbb{S}_+^d \end{aligned} \quad (1)$$

where  $d(\cdot, \cdot)$  denotes the distance metric,  $\mathbb{S}_+^d$  denotes semidefinite cone, and  $\alpha$  and  $\beta$  are positive parameters weighting the observed similarities in structure space and attribute space, respectively. Note that, we use Euclidean distance, i.e.,  $d(\mathbf{X}, \mathbf{Y}) = \|\mathbf{X} - \mathbf{Y}\|_F^2$ .

Take Fig. 1 for example. For each social network, e.g.,  $S^{(1)}$ , we first calculate the similarity matrices  $\mathbf{M}^{(1)}$  and  $\mathbf{W}^{(1)}$ . Then, we perform the collaborative matrix factorization in optimization (1) to embed  $S^{(1)}$  into a latent space, represented by  $\mathbf{H}^{(1)}$ . In the latent space,  $\mathbf{h}_b^{(1)}$  will be closer to  $\mathbf{h}_a^{(1)}$  than  $\mathbf{h}_c^{(1)}$ , and  $\mathbf{h}_a^{(1)}$  and  $\mathbf{h}_c^{(1)}$  are far away.

#### 3.2 Joint-embedding

Based on uni-embedding, joint-embedding aims to construct the joint latent space by aligning latent spaces of  $S^{(m)}$  at the known correspondences  $\hat{L}$  so that: (1)  $\mathbf{h}_i^{(\cdot)}$  of the correspondences coincides in the joint latent space and (2) the proximity of both structure and attribute spaces within the individual networks is captured in  $\mathbf{h}_i^{(\cdot)}$ .

To achieve the first goal, we leverage natural constraints to encode the correspondences and thus force alignment. To give its matrix form, we introduce an elementary matrix  $\mathbf{E}^{(m)}$  for each social network  $S^{(m)}$ . Each row of  $\mathbf{E}^{(m)}$  has only one non-zero element (i.e., 1), to select a  $\mathbf{h}_i^{(\cdot)}$  according to the correspondences in  $L(m, n)$ . We obtain the constraints as follows:

$$\forall L(m, n) \in \hat{L} : \mathbf{E}^{(m)} \mathbf{H}^{(m)} = \mathbf{E}^{(n)} \mathbf{H}^{(n)}, \quad (2)$$

where  $\mathbf{E}^{(m)} \in \mathbb{R}^{|L(m, n)| \times N^{(m)}}$  and  $\mathbf{E}^{(n)} \in \mathbb{R}^{|L(m, n)| \times N^{(n)}}$ . This is an equation system of  $|\hat{L}| = M \cdot (M - 1)$  equations.

We further formulate the equation system into a unified equation pair, despite of the number of networks  $M$ . First, we define two rotating matrices,  $\tilde{\mathbf{D}}_p$  and  $\tilde{\mathbf{D}}_q$ . If  $M$  is odd, we have

$$\tilde{\mathbf{D}}_p = \begin{bmatrix} \mathbf{D} & & & \\ & \mathbf{D} & & \\ & & \ddots & \\ & & & \mathbf{I} \end{bmatrix}, \quad \tilde{\mathbf{D}}_q = \begin{bmatrix} \mathbf{I} & & & \\ & \mathbf{D} & & \\ & & \ddots & \\ & & & \mathbf{D} \end{bmatrix},$$

otherwise,

$$\tilde{\mathbf{D}}_p = \begin{bmatrix} \mathbf{D} & & \\ & \mathbf{D} & \\ & & \ddots \\ & & & \mathbf{D} \end{bmatrix}, \quad \tilde{\mathbf{D}}_q = \begin{bmatrix} \mathbf{I} & & \\ & \mathbf{D} & \\ & & \ddots \\ & & & \mathbf{I} \end{bmatrix},$$

where  $\mathbf{D} = \begin{bmatrix} & \mathbf{I} \\ \mathbf{I} & \end{bmatrix}$  and  $\mathbf{I}$  is the identity matrix. Second, we let  $\mathcal{R}_p(\mathbf{X}) = \tilde{\mathbf{D}}_p \mathbf{X} \tilde{\mathbf{D}}_p$  and  $\mathcal{R}_q(\mathbf{X}) = \tilde{\mathbf{D}}_q \mathbf{X} \tilde{\mathbf{D}}_q$  for rotating operation. Let  $\tilde{\mathbf{X}}$  denote the block diagonal matrix, i.e.,  $\tilde{\mathbf{X}} = \text{diag}(\{\mathbf{X}^{(m)}\})$ . We obtain the equivalent equation pair  $\{L_p(\tilde{\mathbf{H}}) = 0, L_q(\tilde{\mathbf{H}}) = 0\}$ , where

$$L_{(\cdot)}(\tilde{\mathbf{H}}) = d(\tilde{\mathbf{E}}\tilde{\mathbf{H}}, \mathcal{R}_{(\cdot)}(\tilde{\mathbf{E}}\tilde{\mathbf{H}})). \quad (3)$$

To achieve the second goal, similarly, we incorporate uni-embedding preserving the proximity within each network. Utilizing  $\sum_i \|\mathbf{X}\|_F^2 = \|\tilde{\mathbf{X}}\|_F^2$ , we obtain the unified objective on the semidefinite cone, which is equivalent to combining  $M$  objectives of uni-embedding, as follows:

$$\frac{\alpha}{2} \|\tilde{\mathbf{M}} - \tilde{\mathbf{H}}\tilde{\mathbf{B}}\tilde{\mathbf{H}}^T\|_F^2 + \frac{\beta}{2} \|\tilde{\mathbf{W}} - \tilde{\mathbf{H}}\tilde{\mathbf{C}}\tilde{\mathbf{H}}^T\|_F^2. \quad (4)$$

Note that,  $\tilde{\mathbf{B}}$  and  $\tilde{\mathbf{C}}$  inherit the semidefiniteness while  $\tilde{\mathbf{M}}$  and  $\tilde{\mathbf{W}}$  remain to be symmetric.

Finally, we remove the constraint by adding penalty with a coefficient  $\gamma$ , which is sufficiently large, and obtain the unified optimization objective:

$$\begin{aligned} \min_{\tilde{\mathbf{H}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}} \quad & \frac{\alpha}{2} \|\tilde{\mathbf{M}} - \tilde{\mathbf{H}}\tilde{\mathbf{B}}\tilde{\mathbf{H}}^T\|_F^2 + \frac{\beta}{2} \|\tilde{\mathbf{W}} - \tilde{\mathbf{H}}\tilde{\mathbf{C}}\tilde{\mathbf{H}}^T\|_F^2 \\ & + \frac{\gamma}{2} (L_p(\tilde{\mathbf{H}}) + L_q(\tilde{\mathbf{H}})) \end{aligned} \quad (5)$$

s.t.  $\tilde{\mathbf{B}}, \tilde{\mathbf{C}} \in \mathbb{S}_+^{Md}$

Recall the example in Fig. 1. In this example, since the correspondence  $\{v_b^{(1)}, v_b^{(2)}, v_b^{(3)}\}$  is known in advance, we will force  $\mathbf{h}_b^{(1)} = \mathbf{h}_b^{(2)} = \mathbf{h}_b^{(3)}$  to align the embedded space of  $S^{(1)}$ ,  $S^{(2)}$  and  $S^{(3)}$ . Those who are close in the joint latent space from different  $S^{(\cdot)}$  are regarded as good candidates for reconciliation.

The benefits of CDE model are two-folded: (1) both spaces are comprehensively exploited and (2) the problem of reconciliation is formulated in a unified approach for effective reconciliation, regardless of the number of networks.

## 4 OPTIMIZATION: NS-ALTERNATING

To address the optimization problem of the CDE model, inspired by Non-convex Spiting framework [13], we design an effective NS-Alternating algorithm. In this algorithm, we first reformulate the problem (5) into an equivalent form and alternately solve the subproblems of this equivalent form.

### 4.1 Problem Reformulation

The high-order objective (5) is not jointly convex over  $\tilde{\mathbf{H}}$ ,  $\tilde{\mathbf{B}}$  and  $\tilde{\mathbf{C}}$ . Therefore, we reduce the order by introducing an auxiliary matrix  $\mathbf{V} = \tilde{\mathbf{H}}$ , and reformulate the problem as follows:

$$\begin{aligned} \min_{\tilde{\mathbf{H}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}, \mathbf{V}} \quad & \mathcal{J}(\tilde{\mathbf{H}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}, \mathbf{V}) \\ = & \frac{\alpha}{2} \|\tilde{\mathbf{M}} - \tilde{\mathbf{H}}\tilde{\mathbf{B}}\mathbf{V}^T\|_F^2 + \frac{\beta}{2} \|\tilde{\mathbf{W}} - \tilde{\mathbf{H}}\tilde{\mathbf{C}}\mathbf{V}^T\|_F^2 \\ & + \frac{\gamma}{2} (L_p(\tilde{\mathbf{H}}) + L_q(\tilde{\mathbf{H}})) \end{aligned} \quad (6)$$

s.t.  $\tilde{\mathbf{B}}, \tilde{\mathbf{C}} \in \mathbb{S}_+^{Md}, \tilde{\mathbf{H}} = \mathbf{V}, \|\mathbf{V}_i\|_2^2 < \tau, \forall i$

According to the study [24], problem (5) is *equivalent* to problem (6) in the sense of KKT points if  $\tau = \sqrt{C}$  is sufficiently large. That is, the KKT points of problem (5) and problem (6) have a one-to-one correspondence. We set  $C = \max\{\|\tilde{\mathbf{M}}\|_F^2, \|\tilde{\mathbf{W}}\|_F^2\}$  in our algorithm.

We alternately solve  $(\tilde{\mathbf{H}}, \mathbf{V})$  and  $(\tilde{\mathbf{B}}, \tilde{\mathbf{C}})$  of problem (6), referred to as representation matrices subproblem and kernel matrices subproblem, respectively.

### 4.2 Representation Matrix Subproblem

Fixing kernel matrices, the updating rules are given below:

$$\mathbf{V}^{(t+1)} = \arg \min_{\|\mathbf{V}_i\|_2^2 < \tau, \forall i} \mathcal{L}(\tilde{\mathbf{H}}^{(t)}, \mathbf{V}; \Lambda^{(t)}) + \frac{\xi^{(t)}}{2} \|\mathbf{V} - \mathbf{V}^{(t)}\|_F^2 \quad (7)$$

$$\tilde{\mathbf{H}}^{(t+1)} = \arg \min \mathcal{L}(\tilde{\mathbf{H}}, \mathbf{V}^{(t+1)}; \Lambda^{(t)}) \quad (8)$$

$$\Lambda^{(t+1)} = \Lambda^{(t)} + \rho(\mathbf{V}^{(t+1)} - \tilde{\mathbf{H}}^{(t+1)}) \quad (9)$$

$$\xi^{(t+1)} = \frac{6}{\rho} \cdot \mathcal{J}(\tilde{\mathbf{H}}^{(t+1)}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}, \mathbf{V}^{(t+1)}) \quad (10)$$

$$\mathcal{L}(\tilde{\mathbf{H}}, \mathbf{V}; \Lambda) = \mathcal{J}(\tilde{\mathbf{H}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}, \mathbf{V}) + \frac{\rho}{2} \|\mathbf{V} - \tilde{\mathbf{H}} + \Lambda/\rho\|_F^2 \quad (11)$$

$\mathcal{L}(\tilde{\mathbf{H}}, \mathbf{V}; \Lambda)$  is the augmented Lagrangian. Note that, proximal term  $\|\mathbf{V} - \mathbf{V}^{(t)}\|_F^2$  and penalty parameter  $\xi^{(t)}$  are added, according to the study [13], to smooth this optimizing. The optimization w.r.t.  $\mathbf{V}$  can be decomposed into  $k$  separable problems, each of which can be solved using gradient projection. The column-wise updating rule of  $\mathbf{V}$  is given as follows:

$$\mathbf{V}_{[i]}^{(r+1)} = \text{proj}_{\mathbf{V}}(\mathbf{V}_{[i]}^{(r)} - \lambda(\mathbf{A}_V^{(t)} \mathbf{V}_{[i]}^{(r)} - \mathbf{B}_V^{(t)})) \quad (12)$$

$$\mathbf{A}_V^{(t)} = \alpha \tilde{\mathbf{B}} \tilde{\mathbf{H}}^T \tilde{\mathbf{H}} \tilde{\mathbf{B}} + \beta \tilde{\mathbf{C}}^T \tilde{\mathbf{H}}^T \tilde{\mathbf{H}} \tilde{\mathbf{C}} + (\xi^{(t)} + \rho) \mathbf{I} \quad (13)$$

$$\mathbf{B}_V^{(t)} = \alpha \tilde{\mathbf{M}} \tilde{\mathbf{H}} \tilde{\mathbf{B}} + \beta \tilde{\mathbf{S}}^T \tilde{\mathbf{H}} \tilde{\mathbf{C}} + \xi^{(t)} \mathbf{V}^{(t)} + \rho \tilde{\mathbf{H}} - \Lambda \quad (14)$$

$$\text{proj}_{\mathbf{V}}(\mathbf{w}) = \sqrt{\tau} \mathbf{w} / \max\{\sqrt{\tau}, \|\mathbf{w}\|_2\}, \forall \mathbf{w} \in \mathbb{R}^n \quad (15)$$

where  $r$  denotes the inner-iteration number,  $\lambda$  denotes the step size and  $\mathbf{V}_{[i]}$  denotes the  $i^{th}$  column of matrix  $\mathbf{V}$ . For a given vector  $\mathbf{w}$ ,  $\text{proj}_{\mathbf{V}}(\cdot)$  projects it onto the feasible set of  $\mathbf{V}_{[i]}$ .  $\tilde{\mathbf{H}}$  can be solved via 1<sup>st</sup>-order method, whose gradient is:

$$\begin{aligned} \nabla_{\tilde{\mathbf{H}}} \mathcal{L} &= \nabla_{\tilde{\mathbf{H}}} \mathcal{J} + \frac{\rho}{2} \nabla_{\tilde{\mathbf{H}}} \|\mathbf{V} - \tilde{\mathbf{H}} + \Lambda/\rho\|_F^2 \\ &= \alpha (\tilde{\mathbf{H}} \tilde{\mathbf{B}} \mathbf{V}^{(t+1)} - \tilde{\mathbf{M}}) \mathbf{V}^{(t+1)} \tilde{\mathbf{B}}^T \\ &\quad + \beta (\tilde{\mathbf{H}} \tilde{\mathbf{C}} \mathbf{V}^{(t+1)} - \tilde{\mathbf{S}}) \mathbf{V}^{(t+1)} \tilde{\mathbf{C}}^T \\ &\quad + \gamma \tilde{\mathbf{E}} [\mathcal{R}_p(\tilde{\mathbf{E}} \tilde{\mathbf{H}}) + \mathcal{R}_q(\tilde{\mathbf{E}} \tilde{\mathbf{H}})] + 2\gamma \tilde{\mathbf{E}}^T \tilde{\mathbf{E}} \tilde{\mathbf{H}} \\ &\quad + \rho (\tilde{\mathbf{H}} - \mathbf{V}^{(t+1)} - \Lambda/\rho), \end{aligned} \quad (16)$$

as  $\tilde{\mathbf{D}}_{(\cdot)}$  is self-inversed, i.e.,  $\tilde{\mathbf{D}}_p = \tilde{\mathbf{D}}_p^{-1}$  and  $\tilde{\mathbf{D}}_q = \tilde{\mathbf{D}}_q^{-1}$ .

### 4.3 Kernel Matrix Subproblem

Since the structure of  $\tilde{\mathbf{B}}$ -subproblem is the same with that of  $\tilde{\mathbf{C}}$ -subproblem, we only discuss the optimization of  $\tilde{\mathbf{B}}$  for the limit of space. Utilizing  $\|\mathbf{X}\|_F^2 = \text{tr}(\mathbf{X}^T \mathbf{X})$ , we reformulate the optimization w.r.t.  $\mathbf{B}$  into an inner-product form:

$$\begin{aligned} \min_{\tilde{\mathbf{B}}} \quad & \text{tr}(\tilde{\mathbf{H}} \tilde{\mathbf{B}} \mathbf{V}^T \mathbf{V} \tilde{\mathbf{B}} \tilde{\mathbf{H}}^T) - 2\text{tr}(\mathbf{M} \tilde{\mathbf{H}} \tilde{\mathbf{B}} \tilde{\mathbf{V}}^T) \\ & = \langle \mathcal{Q}(\tilde{\mathbf{B}}), \tilde{\mathbf{B}} \rangle - 2\langle \mathbf{A}, \tilde{\mathbf{B}} \rangle, \quad \tilde{\mathbf{B}} \in \mathbb{S}_+^{Md}, \end{aligned} \quad (17)$$

where  $\mathcal{Q}(\tilde{\mathbf{B}}) = \mathbf{V}^T \mathbf{V} \tilde{\mathbf{B}} \tilde{\mathbf{H}}^T \tilde{\mathbf{H}}$  and  $\mathbf{A} = \mathbf{V}^T \tilde{\mathbf{M}} \tilde{\mathbf{H}}$ . The equality  $\tilde{\mathbf{H}} = \mathbf{V}$  holds when representation matrix subproblem converges.

**Lemma 1.**  $\mathcal{Q}$  in optimization (17) is a self-adjoint positive semidefinite linear operator.

*Proof.* A linear operator  $\mathcal{Q} : \mathbb{S}^n \rightarrow \mathbb{S}^n$ , where  $\mathbb{S}^n$  is symmetric cone, mapping an inner-product space  $\mathbb{H}$  to itself is said to be self-adjoint positive semidefinite if and only if (1)  $\langle \mathcal{Q}(\mathbf{X}), \mathbf{Y} \rangle = \langle \mathcal{Q}(\mathbf{Y}), \mathbf{X} \rangle$  (self-adjoint) and (2)  $\langle \mathcal{Q}(\mathbf{X}), \mathbf{Y} \rangle \geq 0$  (positive semidefinite) for all  $\mathbf{X}, \mathbf{Y} \in \mathbb{H}$  [8].

Let  $\tilde{\mathbf{H}}^T \tilde{\mathbf{H}} = \mathbf{P} \in \mathbb{S}_+^{Md}$  and  $\mathcal{Q}(\mathbf{X}) = \mathbf{P} \mathbf{X} \mathbf{P}$ . Utilizing  $\langle \mathbf{X}, \mathbf{Y} \rangle = \text{tr}(\mathbf{X} \mathbf{Y})$ , we first show that  $\mathcal{Q}$  is self-adjoint:

$$\langle \mathcal{Q}(\mathbf{X}), \mathbf{Y} \rangle = \text{tr}(\mathbf{P} \mathbf{X} \mathbf{P} \mathbf{Y}) = \text{tr}(\mathbf{X} \mathbf{P} \mathbf{Y} \mathbf{P}) = \langle \mathcal{Q}(\mathbf{Y}), \mathbf{X} \rangle, \quad (18)$$

and then demonstrate that  $\mathcal{Q}$  is positive semidefinite:

$$\langle \mathcal{Q}(\mathbf{X}), \mathbf{X} \rangle = \text{tr}(\mathbf{P} \mathbf{X} \mathbf{P} \mathbf{X}) = \|\mathbf{P} \mathbf{X}\|_F^2 \geq 0. \quad (19)$$

Knowing  $\mathcal{Q}$  is self-adjoint positive semidefinite, it is easy to check that  $\tilde{\mathbf{B}}$ -subproblem is equivalent to the QSDP problem [23]. Finally, we obtain the theorem as follows:

**Theorem 1.**  $\tilde{\mathbf{B}}$ -subproblem (as well as  $\tilde{\mathbf{C}}$ -subproblem) is a convex Quadratic Semi-Definite Programming (QSDP) problem with  $\mathcal{Q}$  of  $\mathbf{P} \mathbf{X} \mathbf{P}$  form and has the solution of the global optima with quadratic convergence rate.

---

### Algorithm 1: NS-Alternating

---

**Input:** observed  $\{\mathbf{G}^{(\cdot)}, \mathbf{A}^{(\cdot)}\}$  of  $S^{(\cdot)}$  and  $\hat{L}$   
**Output:**  $\mathbf{H}^{(\cdot)}$  for each  $S^{(\cdot)}$  of the joint latent space  
1 Compute  $\tilde{\mathbf{M}}, \tilde{\mathbf{W}}, \{\tilde{\mathbf{E}}^{(\cdot,\cdot)}\}$ ;  
2 Initialize  $\tilde{\mathbf{H}}^{(0)}, \mathbf{V}^{(0)} = \tilde{\mathbf{H}}^{(0)}, \tilde{\mathbf{B}}^{(0)}, \tilde{\mathbf{C}}^{(0)}, n = 0$ ;  
3 **while** not converge **do**  
4     // representation matrix subproblem  
5      $(\tilde{\mathbf{H}}, \mathbf{V})^{(n+1)} = \arg \min l(\tilde{\mathbf{H}}, \mathbf{V}, \tilde{\mathbf{B}}^{(n+1)}, \tilde{\mathbf{C}}^{(n+1)})$ ;  
6     // kernel matrix subproblem  
7      $(\tilde{\mathbf{B}}, \tilde{\mathbf{C}})^{(n+1)} = \arg \min_{\tilde{\mathbf{B}}, \tilde{\mathbf{C}} \in \mathbb{S}_+^{Md}} l(\tilde{\mathbf{H}}^{(n)}, \mathbf{V}^{(n)}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}})$ ;  
8      $n = n + 1$ ;  
9 **end**  
10 **return**  $\mathbf{H}^{(\cdot)}$  from  $\tilde{\mathbf{H}}$ ;

---

We summarize the overall process of NS-Alternating in Algo. 1, where line 4 and line 5 refer to representation matrix subproblem and kernel matrix subproblem, respectively.

Recall the toy example in Fig. 1. Optimizing via the Algo. 1, we obtain  $\mathbf{h}_{(\cdot)}^{(1)}, \mathbf{h}_{(\cdot)}^{(2)}$  and  $\mathbf{h}_{(\cdot)}^{(3)}$  and then,  $\forall i, j \in \{1, 2, 3\}$  ( $i \neq j$ ), we calculate  $\|\mathbf{h}_{(\cdot)}^{(i)} - \mathbf{h}_{(\cdot)}^{(j)}\|_F^2$  to identify the candidates for correspondence.

### 4.4 Computational Complexity

The outer loop of Algo. 1 (line 3-7) achieves the satisfactory accuracy in a few iterations. The representation matrix subproblem (line 4) converges in a few iterations and the kernel matrix subproblem (line 5). QSDP with  $\mathcal{Q}$  of  $\mathbf{P} \mathbf{X} \mathbf{P}$  form is well designed to be quadratic convergence. Note that,  $\tilde{\mathbf{B}}$  and  $\tilde{\mathbf{C}}$  of convex QSDP, blocks  $\mathbf{H}^{(\cdot)}$  of  $\tilde{\mathbf{H}}$  and columns of  $\mathbf{V}$  can be computed in parallel. The computational complexity depends on matrix inversions and multiplication in updating rules. The complexity of matrix inversions is  $O(d^3)$  and the most expensive multiplication term is  $O(N_{max} d^2 + N_{max}^2 d + E_{max} d)$ , where  $N_{max} = \max\{N^{(m)}\}$ ,  $E_{max} = \max\{E^{(m)}\}$  and  $E^{(m)}$  denotes the number of edges in the social network  $S^{(m)}$ . We have  $E_{max} < N_{max}^2$  owing to the sparsity of social networks. With the times of multiplications and inversions set to small constants and  $d \ll N_{max}$ , it is easy to figure out that the overall computational complexity of the proposed MASTER is in the order of  $O(N_{max}^2)$ . The computational complexity MASTER is of the same order with that of the state-of-the-art methods, such as PALE [15], ULink [16] and IONE [11], and lower than that of the methods leveraging the deep neural networks. Note that, there are quantities of optimized libraries (e.g., OpenBLAS) to speed up the most expensive multiplications.

The benefit of our proposed NS-Alternating algorithm lies in that, besides monotonously non-increasing the objective, it will converge to a KKT point of the original optimization of CDE in a modest computational complexity with solid theoretical guarantees.

## 5 CONVERGENCE ANALYSIS

In this section, we first present Lemma 2-4 and then present the convergence theorem of the NS-Alternating algorithm.

**Lemma 2.** We have a bounded successive difference of the multipliers, that is,

$$\begin{aligned} \|\Lambda^{(t+1)} - \Lambda^{(t)}\|_F^2 \leq & 3c_1 \cdot \|\tilde{\mathbf{H}}^{(t+1)} - \tilde{\mathbf{H}}^{(t)}\|_F^2 \\ & + 3c_2 \cdot \|\mathbf{V}^{(t+1)} - \mathbf{V}^{(t)}\|_F^2 \\ & + 3c_3 \cdot \|\tilde{\mathbf{H}}^{(t+1)} (\mathbf{V}^{(t+1)} - \mathbf{V}^{(t)})\|_F^2, \end{aligned}$$

where  $c_i, i = \{1, 2, 3\}$ , are positive scalars:

$$\begin{aligned} c_1 &= \left(16N + \tau N \left(\|\tilde{\mathbf{B}}\|_F^2 + \|\tilde{\mathbf{C}}\|_F^2\right)\right)^2 \\ c_2 &= \|\tilde{\mathbf{B}}\|_F^2 \cdot \|\tilde{\mathbf{H}}^{(t)} (\mathbf{V}^{(t)} \tilde{\mathbf{B}})^T - \tilde{\mathbf{M}}\|_F^2 \\ & + \|\tilde{\mathbf{C}}\|_F^2 \cdot \|\tilde{\mathbf{H}}^{(t)} (\mathbf{V}^{(t)} \tilde{\mathbf{C}})^T - \tilde{\mathbf{W}}\|_F^2 \\ c_3 &= N\tau \cdot \left(\|\tilde{\mathbf{B}}\|_F^4 \cdot \|\tilde{\mathbf{C}}\|_F^4\right) \end{aligned} \quad (20)$$

*Proof.* We give the idea to prove this lemma here<sup>1</sup>. Using the optimal condition of  $\Lambda$  together with its updating rule, we obtain:

$$\begin{aligned} \Lambda^{(t+1)} &= (\tilde{\mathbf{H}}^{(t+1)} \tilde{\mathbf{B}} \mathbf{V}^{(t+1)T} - \tilde{\mathbf{M}}) \cdot \mathbf{V}^{(t+1)} \tilde{\mathbf{B}}^T \\ & + (\tilde{\mathbf{H}}^{(t+1)} \tilde{\mathbf{C}} \mathbf{V}^{(t+1)T} - \tilde{\mathbf{W}}) \cdot \mathbf{V}^{(t+1)} \tilde{\mathbf{C}}^T \\ & + 2 \left(\mathbf{E}_1^T \mathbf{E}_1 - \mathbf{E}_1^T \mathbf{E}_2 - \mathbf{E}_2^T \mathbf{E}_1 + \mathbf{E}_2^T \mathbf{E}_2\right) \tilde{\mathbf{H}}^{(t+1)} \end{aligned} \quad (21)$$

1. For the ease of derivation, we introduce  $\mathbf{E}_1$  and  $\mathbf{E}_2$  to encode the transformation of the selecting (as well as the rotating via  $\mathcal{R}_{(\cdot)}$ ) of the representation matrix  $\tilde{\mathbf{H}}$ .

Then, the successive difference of  $\Lambda$  is given below:

$$\begin{aligned} & \Lambda^{(t+1)} - \Lambda^{(t)} \\ &= \left( \tilde{\mathbf{H}}^{(t)} \left( \mathbf{V}^{(t)} \tilde{\mathbf{B}}^T \right)^T - \tilde{\mathbf{M}} \right) \left( \mathbf{V}^{(t+1)} \tilde{\mathbf{B}}^T - \mathbf{V}^{(t)} \tilde{\mathbf{B}}^T \right) \\ &\quad - \left( \tilde{\mathbf{H}}^{(t)} \left( \mathbf{V}^{(t)} \tilde{\mathbf{C}}^T \right)^T - \tilde{\mathbf{W}} \right) \left( \mathbf{V}^{(t+1)} \tilde{\mathbf{C}}^T - \mathbf{V}^{(t)} \tilde{\mathbf{C}}^T \right) \\ &\quad + \left[ \tilde{\mathbf{H}}^{(t+1)} - \tilde{\mathbf{H}}^{(t)} \right] \left[ \left( \mathbf{V}^{(t+1)} \tilde{\mathbf{B}}^T \right)^T \left( \mathbf{V}^{(t+1)} \tilde{\mathbf{B}}^T \right) \right] \\ &\quad + \left[ \tilde{\mathbf{H}}^{(t+1)} - \tilde{\mathbf{H}}^{(t)} \right] \left[ \left( \mathbf{V}^{(t+1)} \tilde{\mathbf{C}}^T \right)^T \left( \mathbf{V}^{(t+1)} \tilde{\mathbf{C}}^T \right) \right] \\ &\quad + 2 \left[ \tilde{\mathbf{H}}^{(t+1)} - \tilde{\mathbf{H}}^{(t)} \right] \left( \mathbf{E}_1^T \mathbf{E}_1 - \mathbf{E}_1^T \mathbf{E}_2 - \mathbf{E}_2^T \mathbf{E}_1 + \mathbf{E}_2^T \mathbf{E}_2 \right) \\ &\quad + \tilde{\mathbf{H}}^{(t)} \left[ \left( \mathbf{V}^{(t+1)} \tilde{\mathbf{B}}^T - \mathbf{V}^{(t)} \tilde{\mathbf{B}}^T \right)^T \left( \mathbf{V}^{(t+1)} \tilde{\mathbf{B}}^T \right) \right] \\ &\quad + \tilde{\mathbf{H}}^{(t)} \left[ \left( \mathbf{V}^{(t+1)} \tilde{\mathbf{C}}^T - \mathbf{V}^{(t)} \tilde{\mathbf{C}}^T \right)^T \left( \mathbf{V}^{(t+1)} \tilde{\mathbf{C}}^T \right) \right] \end{aligned} \quad (22)$$

Using triangle inequality, we obtain:

$$\begin{aligned} & \|\Lambda^{(t+1)} - \Lambda^{(t)}\|_F \\ &\leq \|\tilde{\mathbf{H}}^{(t+1)} - \tilde{\mathbf{H}}^{(t)}\|_F \left( \mathbf{V}^{(t+1)} \tilde{\mathbf{B}}^T \right)^T \left( \mathbf{V}^{(t+1)} \tilde{\mathbf{B}}^T \right) \\ &\quad + 2 \left( \mathbf{E}_1^T \mathbf{E}_1 - \mathbf{E}_1^T \mathbf{E}_2 - \mathbf{E}_2^T \mathbf{E}_1 + \mathbf{E}_2^T \mathbf{E}_2 \right) \\ &\quad + \left( \mathbf{V}^{(t+1)} \tilde{\mathbf{C}}^T \right)^T \left( \mathbf{V}^{(t+1)} \tilde{\mathbf{C}}^T \right) \|_F \\ &\quad + \|\tilde{\mathbf{H}}^{(t)} \left( \mathbf{V}^{(t)} \tilde{\mathbf{B}}^T \right)^T - \tilde{\mathbf{M}}\|_F \|\mathbf{V}^{(t+1)} \tilde{\mathbf{B}}^T - \mathbf{V}^{(t)} \tilde{\mathbf{B}}^T\|_F \\ &\quad + \|\tilde{\mathbf{H}}^{(t)} \left( \mathbf{V}^{(t)} \tilde{\mathbf{C}}^T \right)^T - \tilde{\mathbf{W}}\|_F \|\mathbf{V}^{(t+1)} \tilde{\mathbf{C}}^T - \mathbf{V}^{(t)} \tilde{\mathbf{C}}^T\|_F \\ &\quad + \|\tilde{\mathbf{H}}^{(t)} \left( \mathbf{V}^{(t+1)} \tilde{\mathbf{B}}^T - \mathbf{V}^{(t)} \tilde{\mathbf{B}}^T \right)^T\|_F \|\mathbf{V}^{(t+1)} \tilde{\mathbf{B}}^T\|_F \\ &\quad + \|\tilde{\mathbf{H}}^{(t)} \left( \mathbf{V}^{(t+1)} \tilde{\mathbf{C}}^T - \mathbf{V}^{(t)} \tilde{\mathbf{C}}^T \right)^T\|_F \|\mathbf{V}^{(t+1)} \tilde{\mathbf{C}}^T\|_F \end{aligned} \quad (23)$$

Given  $\|\mathbf{V}\|_F \leq \sqrt{N\tau}$ , it is easy to check Lemma 2 holds.  $\square$

**Lemma 3.** If the equations below are satisfied,

$$\begin{aligned} \rho &> \max\{\rho_1, \rho_2\} \\ \rho_1 &= 6N\tau \left( \|\tilde{\mathbf{B}}\|_F^4 + \|\tilde{\mathbf{C}}\|_F^4 \right) / \left( \|\tilde{\mathbf{B}}\|_F^2 + \|\tilde{\mathbf{C}}\|_F^2 \right) \\ \rho_2 &= -2\|\mathbf{E}\|_F^2 + \frac{6}{\rho_2} \left( 16N + N\tau \left( \|\tilde{\mathbf{B}}\|_F^2 + \|\tilde{\mathbf{C}}\|_F^2 \right) \right)^2 \\ \beta^{(t)} &> -\rho + \frac{6}{\rho} \|\tilde{\mathbf{B}}\|_F^2 \cdot \|\tilde{\mathbf{H}}^{(t)} \left( \mathbf{V}^{(t)} \tilde{\mathbf{B}}^T \right)^T - \tilde{\mathbf{M}}\|_F^2 \\ &\quad + \frac{6}{\rho} \|\tilde{\mathbf{C}}\|_F^2 \cdot \|\tilde{\mathbf{H}}^{(t)} \left( \mathbf{V}^{(t)} \tilde{\mathbf{C}}^T \right)^T - \tilde{\mathbf{W}}\|_F^2 \end{aligned}$$

we have positive scalars  $c_i, i = \{1, 2, 3, 4\}$ , so that:

$$\begin{aligned} & \mathcal{L} \left( \tilde{\mathbf{H}}^{(t+1)}, \mathbf{V}^{(t+1)}, \Lambda^{(t+1)} \right) - \mathcal{L} \left( \tilde{\mathbf{H}}^{(t)}, \mathbf{V}^{(t)}, \Lambda^{(t)} \right) \\ &< -c_1 \|\tilde{\mathbf{H}}^{(t+1)} - \tilde{\mathbf{H}}^{(t)}\|_F^2 - c_2 \|\left( \tilde{\mathbf{H}}^{(t+1)} - \tilde{\mathbf{H}}^{(t)} \right) \mathbf{V}^{(t+1)}\|_F^2 \\ &\quad - c_3 \|\mathbf{V}^{(t+1)} - \mathbf{V}^{(t)}\|_F^2 - c_4 \|\tilde{\mathbf{H}}^{(t)} \left( \mathbf{V}^{(t+1)} - \mathbf{V}^{(t)} \right)\|_F^2 \end{aligned}$$

*Proof.* We give the idea to prove this lemma as follows:  
First, we let

$$\begin{aligned} A &\triangleq \mathcal{L} \left( \tilde{\mathbf{H}}^{(t)}, \mathbf{V}^{(t+1)}, \Lambda^{(t)} \right) - \mathcal{L} \left( \tilde{\mathbf{H}}^{(t)}, \mathbf{V}^{(t)}, \Lambda^{(t)} \right) \\ B &\triangleq \mathcal{L} \left( \tilde{\mathbf{H}}^{(t+1)}, \mathbf{V}^{(t+1)}, \Lambda^{(t)} \right) - \mathcal{L} \left( \tilde{\mathbf{H}}^{(t)}, \mathbf{V}^{(t+1)}, \Lambda^{(t)} \right) \\ C &\triangleq \mathcal{L} \left( \tilde{\mathbf{H}}^{(t+1)}, \mathbf{V}^{(t+1)}, \Lambda^{(t+1)} \right) - \mathcal{L} \left( \tilde{\mathbf{H}}^{(t+1)}, \mathbf{V}^{(t+1)}, \Lambda^{(t)} \right) \\ \hat{A} &= \hat{\mathcal{L}} \left( \tilde{\mathbf{H}}^{(t)}, \mathbf{V}^{(t+1)}, \Lambda^{(t)} \right) - \mathcal{L} \left( \tilde{\mathbf{H}}^{(t)}, \mathbf{V}^{(t)}, \Lambda^{(t)} \right) \end{aligned} \quad (24)$$

where

$$\begin{aligned} & \hat{\mathcal{L}} \left( \tilde{\mathbf{H}}^{(t)}, \mathbf{V}, \Lambda^{(t)} \right) \\ &\triangleq \frac{1}{2} \left\| \tilde{\mathbf{H}}^{(t)} \tilde{\mathbf{B}} \mathbf{V}^T - \tilde{\mathbf{M}} \right\|_F^2 + \frac{1}{2} \left\| \tilde{\mathbf{H}}^{(t)} \tilde{\mathbf{C}} \mathbf{V}^T - \tilde{\mathbf{W}} \right\|_F^2 \\ &\quad + \frac{\rho}{2} \left\| \tilde{\mathbf{H}}^{(t)} - \mathbf{V} + \Lambda^{(t)} / \rho \right\|_F^2 + \left\| \mathbf{E}_1 \tilde{\mathbf{H}} - \mathbf{E}_2 \tilde{\mathbf{H}} \right\|_F^2 \\ &\quad + \frac{\beta^{(t)}}{2} \left\| \mathbf{V} - \mathbf{V}^{(t)} \right\|_F^2 \end{aligned} \quad (25)$$

is the upper bound of  $\mathcal{L} \left( \tilde{\mathbf{H}}^{(t)}, \mathbf{V}, \Lambda^{(t)} \right)$ .

Then, we have

$$\begin{aligned} & \mathcal{L} \left( \tilde{\mathbf{H}}^{(t+1)}, \mathbf{V}^{(t+1)}, \Lambda^{(t+1)} \right) - \mathcal{L} \left( \tilde{\mathbf{H}}^{(t)}, \mathbf{V}^{(t)}, \Lambda^{(t)} \right) \\ &= A + B + C \leq \hat{A} + B + C \end{aligned} \quad (26)$$

Specifically,

$$\begin{aligned} \hat{A} &= \frac{1}{2} \left\| \tilde{\mathbf{H}}^{(t)} \tilde{\mathbf{B}} \mathbf{V}^{(t+1)} - \tilde{\mathbf{M}} \right\|_F^2 - \frac{1}{2} \left\| \tilde{\mathbf{H}}^{(t)} \tilde{\mathbf{B}} \mathbf{V}^{(t)} - \tilde{\mathbf{M}} \right\|_F^2 \\ &\quad + \frac{1}{2} \left\| \tilde{\mathbf{H}}^{(t)} \tilde{\mathbf{C}} \mathbf{V}^{(t+1)} - \tilde{\mathbf{W}} \right\|_F^2 - \frac{1}{2} \left\| \tilde{\mathbf{H}}^{(t)} \tilde{\mathbf{C}} \mathbf{V}^{(t)} - \tilde{\mathbf{W}} \right\|_F^2 \\ &\quad + \frac{\rho}{2} \left\| \tilde{\mathbf{H}}^{(t)} - \mathbf{V}^{(t+1)} + \frac{\Lambda^{(t)}}{\rho} \right\|_F^2 - \frac{\rho}{2} \left\| \tilde{\mathbf{H}}^{(t)} - \mathbf{V}^{(t)} + \frac{\Lambda^{(t)}}{\rho} \right\|_F^2 \\ &\quad + \frac{\beta^{(t)}}{2} \left\| \mathbf{V}^{(t+1)} - \mathbf{V}^{(t)} \right\|_F^2 \\ &\stackrel{(a)}{\leq} \left\langle \mathbf{V}^{(t+1)} \tilde{\mathbf{B}}^T \tilde{\mathbf{H}}^{(t)} \tilde{\mathbf{H}}^{(t)} \tilde{\mathbf{B}} - \tilde{\mathbf{B}}^T \tilde{\mathbf{H}}^{(t)} \tilde{\mathbf{M}}, \mathbf{V}^{(t+1)} - \mathbf{V}^{(t+1)} \right\rangle \\ &\quad + \left\langle \mathbf{V}^{(t+1)} \tilde{\mathbf{C}}^T \tilde{\mathbf{H}}^{(t)} \tilde{\mathbf{H}}^{(t)} \tilde{\mathbf{C}} - \tilde{\mathbf{C}}^T \tilde{\mathbf{H}}^{(t)} \tilde{\mathbf{W}}, \mathbf{V}^{(t+1)} - \mathbf{V}^{(t+1)} \right\rangle \\ &\quad - \frac{1}{2} \left\| \tilde{\mathbf{B}} \right\|_F^2 \cdot \left\| \tilde{\mathbf{H}}^{(t)} \cdot \left( \mathbf{V}^{(t+1)} - \mathbf{V}^{(t)} \right) \right\|_F^2 \\ &\quad - \frac{1}{2} \left\| \tilde{\mathbf{C}} \right\|_F^2 \cdot \left\| \tilde{\mathbf{H}}^{(t)} \cdot \left( \mathbf{V}^{(t+1)} - \mathbf{V}^{(t)} \right) \right\|_F^2 \\ &\quad + \rho \left\langle \tilde{\mathbf{H}}^{(t)} - \mathbf{V}^{(t+1)} + \frac{\Lambda^{(t)}}{\rho}, \mathbf{V}^{(t+1)} - \mathbf{V}^{(t)} \right\rangle \\ &\quad - \frac{\rho}{2} \left\| \mathbf{V}^{(t+1)} - \mathbf{V}^{(t)} \right\|_F^2 + \frac{\beta^{(t)}}{2} \left\| \mathbf{V}^{(t+1)} - \mathbf{V}^{(t)} \right\|_F^2 \\ &\stackrel{(b)}{\leq} -\frac{1}{2} \left\| \tilde{\mathbf{B}} \right\|_F^2 \cdot \left\| \tilde{\mathbf{H}}^{(t)} \cdot \left( \mathbf{V}^{(t+1)} - \mathbf{V}^{(t)} \right) \right\|_F^2 \\ &\quad - \frac{1}{2} \left\| \tilde{\mathbf{C}} \right\|_F^2 \cdot \left\| \tilde{\mathbf{H}}^{(t)} \cdot \left( \mathbf{V}^{(t+1)} - \mathbf{V}^{(t)} \right) \right\|_F^2 \\ &\quad - \frac{\rho}{2} \left\| \mathbf{V}^{(t+1)} - \mathbf{V}^{(t)} \right\|_F^2 + \frac{\beta^{(t)}}{2} \left\| \mathbf{V}^{(t+1)} - \mathbf{V}^{(t)} \right\|_F^2 \end{aligned} \quad (27)$$

where (a) is the Taylor expansion and (b) is the optimal condition.

Similarly, we have

$$\begin{aligned} B &\leq -\frac{1}{2} \left( \left\| \tilde{\mathbf{B}} \right\|_F^2 + \left\| \tilde{\mathbf{C}} \right\|_F^2 \right) \left\| \left( \tilde{\mathbf{H}}^{(t+1)} - \tilde{\mathbf{H}}^{(t)} \right) \mathbf{V}^{(t+1)} \right\|_F^2 \\ &\quad - \left( \frac{\rho}{2} + \left\| \mathbf{E}_1 - \mathbf{E}_2 \right\|_F^2 \right) \left\| \tilde{\mathbf{H}}^{(t+1)} - \tilde{\mathbf{H}}^{(t)} \right\|_F^2 \end{aligned} \quad (28)$$

and

$$\begin{aligned} C &= \frac{\rho}{2} \left\| \mathbf{V}^{(t+1)} - \tilde{\mathbf{H}}^{(t+1)} + \frac{\Lambda^{(t+1)}}{\rho} \right\|_F^2 - \frac{\rho}{2} \left\| \mathbf{V}^{(t+1)} - \tilde{\mathbf{H}}^{(t+1)} + \frac{\Lambda^{(t)}}{\rho} \right\|_F^2 \\ &\stackrel{(9)}{=} \frac{1}{\rho} \left\| \Lambda^{(t+1)} - \Lambda^{(t)} \right\|_F^2 \end{aligned} \quad (29)$$

Then, we need to incorporate the result of Lemma 2, i.e., equation (21), into  $C$ .<sup>2</sup>

<sup>2</sup> We prefer to use  $\mathcal{L} \left( \tilde{\mathbf{H}}, \mathbf{V}, \Lambda \right)$  in the proof as  $\Lambda$  is of no difference with  $\tilde{\mathbf{H}}$  or  $\mathbf{V}$  as a matrix variable in the derivation.

Finally, we obtain the successive difference as follows:

$$\begin{aligned}
 & \mathcal{L}(\tilde{\mathbf{H}}^{(t+1)}, \mathbf{V}^{(t+1)}, \boldsymbol{\Lambda}^{(t+1)}) - \mathcal{L}(\tilde{\mathbf{H}}^{(t)}, \mathbf{V}^{(t)}, \boldsymbol{\Lambda}^{(t)}) \\
 & \leq \hat{A} + B + C \\
 & \leq -\frac{1}{2} \left( \|\tilde{\mathbf{B}}\|_F^2 + \|\tilde{\mathbf{C}}\|_F^2 \right) \|\tilde{\mathbf{H}}^{(t)} (\mathbf{V}^{(t+1)} - \mathbf{V}^{(t)})\|_F^2 \\
 & \quad - \frac{\rho}{2} \|\mathbf{V}^{(t+1)} - \mathbf{V}^{(t)}\|_F^2 + \frac{\beta^{(t)}}{2} \|\mathbf{V}^{(t+1)} - \mathbf{V}^{(t)}\|_F^2 \\
 & \quad - \frac{1}{2} \left( \|\tilde{\mathbf{B}}\|_F^2 + \|\tilde{\mathbf{C}}\|_F^2 \right) \|(\tilde{\mathbf{H}}^{(t+1)} - \tilde{\mathbf{H}}^{(t)}) \mathbf{V}^{(t+1)T}\|_F^2 \\
 & \quad - \left( \frac{\rho}{2} + \|\mathbf{E}_1 - \mathbf{E}_2\|_F^2 \right) \|\tilde{\mathbf{H}}^{(t+1)} - \tilde{\mathbf{H}}^{(t)}\|_F^2 \\
 & \quad + \frac{3}{\rho} \left( 16N + N\tau \left( \|\tilde{\mathbf{B}}\|_F^2 + \|\tilde{\mathbf{C}}\|_F^2 \right) \right)^2 \|\tilde{\mathbf{H}}^{(t+1)} - \tilde{\mathbf{H}}^{(t)}\|_F^2 \\
 & \quad + \frac{3}{\rho} \|\tilde{\mathbf{B}}\|_F^2 \cdot \|\tilde{\mathbf{H}}^{(t)} \cdot (\mathbf{V}^{(t)} \cdot \tilde{\mathbf{B}}^T)^T - \tilde{\mathbf{M}}\|_F^2 \cdot \|\mathbf{V}^{(t+1)} - \mathbf{V}^{(t)}\|_F^2 \\
 & \quad + \frac{3}{\rho} \|\tilde{\mathbf{C}}\|_F^2 \cdot \|\tilde{\mathbf{H}}^{(t)} \cdot (\mathbf{V}^{(t)} \cdot \tilde{\mathbf{C}}^T)^T - \tilde{\mathbf{W}}\|_F^2 \cdot \|\mathbf{V}^{(t+1)} - \mathbf{V}^{(t)}\|_F^2 \\
 & \quad + \frac{3}{\rho} N\tau \left( \|\tilde{\mathbf{B}}\|_F^4 + \|\tilde{\mathbf{C}}\|_F^4 \right) \|\tilde{\mathbf{H}}^{(t)} (\mathbf{V}^{(t+1)} - \mathbf{V}^{(t)})\|_F^2 \\
 & < -c_1 \|\tilde{\mathbf{H}}^{(t+1)} - \tilde{\mathbf{H}}^{(t)}\|_F^2 - c_2 \|(\tilde{\mathbf{H}}^{(t+1)} - \tilde{\mathbf{H}}^{(t)}) \mathbf{V}^{(t+1)T}\|_F^2 \\
 & \quad - c_3 \|\mathbf{V}^{(t+1)} - \mathbf{V}^{(t)}\|_F^2 - c_4 \|\tilde{\mathbf{H}}^{(t)} (\mathbf{V}^{(t+1)} - \mathbf{V}^{(t)})\|_F^2
 \end{aligned} \tag{30}$$

Let  $c_i > 0, i = \{1, 2, 3, 4\}$ , and we have Lemma 3.  $\square$

**Lemma 4.** If the equation below is satisfied,

$$\rho \geq \|\tilde{\mathbf{B}}\|_F^2 + \|\tilde{\mathbf{C}}\|_F^2 + 2\|\mathcal{R}_p(\mathbf{E}) + \mathcal{R}_q(\mathbf{E})\|_F^2$$

we have a lower bound of 0, that is,

$$\mathcal{L}(\tilde{\mathbf{H}}^{(t+1)}, \mathbf{V}^{(t+1)}; \boldsymbol{\Lambda}^{(t+1)}) \geq 0$$

*Proof.* We give the idea to prove this lemma as follows:

$$\begin{aligned}
 & \mathcal{L}(\mathbf{H}^{(t+1)}, \mathbf{V}^{(t+1)}, \boldsymbol{\Lambda}^{(t+1)}) \\
 & = \frac{1}{2} \|\tilde{\mathbf{H}}^{(t+1)} \tilde{\mathbf{B}} \mathbf{V}^{(t+1)T} - \tilde{\mathbf{M}}\|_F^2 + \frac{1}{2} \|\tilde{\mathbf{H}}^{(t+1)} \tilde{\mathbf{C}} \mathbf{V}^{(t+1)T} - \tilde{\mathbf{W}}\|_F^2 \\
 & \quad + \|\mathbf{E}_1 \tilde{\mathbf{H}}^{(t+1)} - \mathbf{E}_2 \tilde{\mathbf{H}}^{(t+1)}\|_F^2 + \frac{\rho}{2} \left\| \mathbf{V}^{(t+1)} - \tilde{\mathbf{H}}^{(t+1)} + \frac{\boldsymbol{\Lambda}^{(t+1)}}{\rho} \right\|_F^2 \\
 & \stackrel{(21)}{=} \frac{1}{2} \|\tilde{\mathbf{H}}^{(t+1)} \tilde{\mathbf{B}} \mathbf{V}^{(t+1)T} - \tilde{\mathbf{M}}\|_F^2 + \frac{1}{2} \|\tilde{\mathbf{H}}^{(t+1)} \tilde{\mathbf{C}} \mathbf{V}^{(t+1)T} - \tilde{\mathbf{W}}\|_F^2 \\
 & \quad + \|\mathbf{E}_1 \tilde{\mathbf{H}}^{(t+1)} - \mathbf{E}_2 \tilde{\mathbf{H}}^{(t+1)}\|_F^2 + \frac{\rho}{2} \|\tilde{\mathbf{H}}^{(t+1)} - \mathbf{V}^{(t+1)}\|_F^2 \\
 & \quad + \langle \tilde{\mathbf{H}}^{(t+1)} - \mathbf{V}^{(t+1)}, (\tilde{\mathbf{H}}^{(t+1)} \tilde{\mathbf{B}} \mathbf{V}^{(t+1)T} - \tilde{\mathbf{M}}) \mathbf{V}^{(t+1)T} \rangle + \\
 & \quad + \langle \tilde{\mathbf{H}}^{(t+1)} - \mathbf{V}^{(t+1)}, (\tilde{\mathbf{H}}^{(t+1)} \tilde{\mathbf{C}} \mathbf{V}^{(t+1)T} - \tilde{\mathbf{W}}) \mathbf{V}^{(t+1)T} \rangle + \\
 & \quad + 2 \langle \tilde{\mathbf{H}}^{(t+1)} - \mathbf{V}^{(t+1)}, (\mathbf{E}_1^T \mathbf{E}_1 - \mathbf{E}_1^T \mathbf{E}_2 - \mathbf{E}_2^T \mathbf{E}_1 + \mathbf{E}_2^T \mathbf{E}_2) \tilde{\mathbf{H}}^{(t+1)} \rangle
 \end{aligned} \tag{31}$$

First, we note that,

$$\begin{aligned}
 0 & \leq \|\tilde{\mathbf{H}}^{(t+1)} - \mathbf{V}^{(t+1)}\|_F^2 + \|\tilde{\mathbf{H}}^{(t+1)} \tilde{\mathbf{B}} \mathbf{V}^{(t+1)T} - \tilde{\mathbf{M}}\|_F^2 \\
 & = \|\tilde{\mathbf{H}}^{(t+1)} - \mathbf{V}^{(t+1)}\|_F^2 + \|\tilde{\mathbf{H}}^{(t+1)} \tilde{\mathbf{B}} \mathbf{V}^{(t+1)T} - \tilde{\mathbf{M}}\|_F^2 \\
 & \quad + 2 \langle \tilde{\mathbf{H}}^{(t+1)} - \mathbf{V}^{(t+1)}, \tilde{\mathbf{B}} \mathbf{V}^{(t+1)T}, \tilde{\mathbf{H}}^{(t+1)} \tilde{\mathbf{B}} \mathbf{V}^{(t+1)T} - \tilde{\mathbf{M}} \rangle \\
 & = \|\tilde{\mathbf{H}}^{(t+1)} - \mathbf{V}^{(t+1)}\|_F^2 + \|\tilde{\mathbf{H}}^{(t+1)} \tilde{\mathbf{B}} \mathbf{V}^{(t+1)T} - \tilde{\mathbf{M}}\|_F^2 \\
 & \quad + 2 \langle \tilde{\mathbf{H}}^{(t+1)} - \mathbf{V}^{(t+1)}, (\tilde{\mathbf{H}}^{(t+1)} \tilde{\mathbf{B}} \mathbf{V}^{(t+1)T} - \tilde{\mathbf{M}}) \mathbf{V}^{(t+1)T} \rangle,
 \end{aligned} \tag{32}$$

where we have

$$\langle (\mathbf{X} - \mathbf{Y}) \mathbf{Y}^T, \mathbf{XY} - \mathbf{Z} \rangle = \langle \mathbf{X} - \mathbf{Y}, (\mathbf{XY} - \mathbf{Z}) \mathbf{Y} \rangle. \tag{33}$$

That is,

$$\begin{aligned}
 & \|\tilde{\mathbf{H}}^{(t+1)} \tilde{\mathbf{B}} \mathbf{V}^{(t+1)T} - \tilde{\mathbf{M}}\|_F^2 + \\
 & \quad + 2 \langle (\tilde{\mathbf{H}}^{(t+1)} - \mathbf{V}^{(t+1)}), (\tilde{\mathbf{H}}^{(t+1)} \tilde{\mathbf{B}} \mathbf{V}^{(t+1)T} - \tilde{\mathbf{M}}) \mathbf{V}^{(t+1)T} \rangle \\
 & \leq -\|(\tilde{\mathbf{H}}^{(t+1)} - \mathbf{V}^{(t+1)}) \tilde{\mathbf{B}} \mathbf{V}^{(t+1)T}\|_F^2 \\
 & \leq -\|\tilde{\mathbf{B}}\|_F^2 N\tau \cdot \|\tilde{\mathbf{H}}^{(t+1)} - \mathbf{V}^{(t+1)}\|_F^2.
 \end{aligned} \tag{34}$$

Therefore, we obtain

$$\begin{aligned}
 & \frac{1}{2} \|\tilde{\mathbf{H}}^{(t+1)} \tilde{\mathbf{B}} \mathbf{V}^{(t+1)T} - \tilde{\mathbf{M}}\|_F^2 \\
 & \quad + \langle (\tilde{\mathbf{H}}^{(t+1)} - \mathbf{V}^{(t+1)}), (\tilde{\mathbf{H}}^{(t+1)} \tilde{\mathbf{B}} \mathbf{V}^{(t+1)T} - \tilde{\mathbf{M}}) \mathbf{V}^{(t+1)T} \rangle \\
 & \leq -\frac{1}{2} \|\tilde{\mathbf{B}}\|_F^2 N\tau \cdot \|\tilde{\mathbf{H}}^{(t+1)} - \mathbf{V}^{(t+1)}\|_F^2.
 \end{aligned} \tag{35}$$

Similarly, we can obtain following inequalities:

$$\begin{aligned}
 & \frac{1}{2} \|\tilde{\mathbf{H}}^{(t+1)} \tilde{\mathbf{C}} \mathbf{V}^{(t+1)T} - \tilde{\mathbf{W}}\|_F^2 \\
 & \quad + \langle (\tilde{\mathbf{H}}^{(t+1)} - \mathbf{V}^{(t+1)}), (\tilde{\mathbf{H}}^{(t+1)} \tilde{\mathbf{C}} \mathbf{V}^{(t+1)T} - \tilde{\mathbf{W}}) \mathbf{V}^{(t+1)T} \rangle \\
 & \leq -\frac{1}{2} \|\tilde{\mathbf{C}}\|_F^2 N\tau \cdot \|\tilde{\mathbf{H}}^{(t+1)} - \mathbf{V}^{(t+1)}\|_F^2
 \end{aligned} \tag{36}$$

and

$$\begin{aligned}
 & 2 \langle \tilde{\mathbf{H}}^{(t+1)} - \mathbf{V}^{(t+1)}, (\mathbf{E}_1^T \mathbf{E}_1 - \mathbf{E}_1^T \mathbf{E}_2 - \mathbf{E}_2^T \mathbf{E}_1 + \mathbf{E}_2^T \mathbf{E}_2) \tilde{\mathbf{H}}^{(t+1)} \rangle \\
 & \quad + \|\mathbf{E}_1 \tilde{\mathbf{H}}^{(t+1)} - \mathbf{E}_2 \tilde{\mathbf{H}}^{(t+1)}\|_F^2 \\
 & \geq -\|\mathbf{E}_1 \tilde{\mathbf{H}}^{(t+1)} - \mathbf{E}_2 \tilde{\mathbf{H}}^{(t+1)}\|_F^2 \cdot \|\tilde{\mathbf{H}}^{(t+1)} - \mathbf{V}^{(t+1)}\|_F^2
 \end{aligned} \tag{37}$$

Finally, we can formulate:

$$\mathcal{L}(\tilde{\mathbf{H}}^{(t+1)}, \mathbf{V}^{(t+1)}, \boldsymbol{\Lambda}^{(t+1)}) \geq c \cdot \|\tilde{\mathbf{H}}^{(t+1)} - \mathbf{V}^{(t+1)}\|_F^2 \tag{38}$$

Let  $c > 0$  and we have Lemma 4.  $\square$

With Lemma 2-4, we can claim two propositions: (1) the size of the successive difference of the multipliers is bounded by that of the successive difference of the primal variables, and (2) the augmented Lagrangian is decreasing and lower bounded. These propositions hold in the updating rules (8)-(11). According to the convergence properties [6], we finally conclude the sufficient condition of KKT-convergence as follows:

**Theorem 2.** With given  $\tilde{\mathbf{B}}$  and  $\tilde{\mathbf{C}}$ , if  $\rho > \max\{\rho_1, \rho_2, \rho_3\}$ :

$$\begin{aligned}
 \rho_1 & = 6N\tau \left( \|\tilde{\mathbf{B}}\|_F^4 + \|\tilde{\mathbf{C}}\|_F^4 \right) / \left( \|\tilde{\mathbf{B}}\|_F^2 + \|\tilde{\mathbf{C}}\|_F^2 \right) \\
 \rho_2 & = 2\|\tilde{\mathbf{B}}\|_F^2 = \frac{6}{\rho_2} \left( 16N + N\tau \left( \|\tilde{\mathbf{B}}\|_F^2 + \|\tilde{\mathbf{C}}\|_F^2 \right) \right)^2 \\
 \rho_3 & = \|\tilde{\mathbf{B}}\|_F^2 + \|\tilde{\mathbf{C}}\|_F^2 + \|\mathcal{R}_p(\tilde{\mathbf{E}}) + \mathcal{R}_q(\tilde{\mathbf{E}})\|_F^2
 \end{aligned}$$

We can claim that:

- The equality constraint on the auxiliary matrix is satisfied in the limit, i.e.,  $\lim_{t \rightarrow \infty} \|\tilde{\mathbf{H}}^{(t)} - \mathbf{V}^{(t)}\|_F^2 = 0$ .
- The sequence  $\{\tilde{\mathbf{H}}^{(t)}, \mathbf{V}^{(t)}, \boldsymbol{\Lambda}^{(t)}\}$  generated by the NS-Alternating algorithm is bounded, and every limit point of the sequence is a KKT point of problem (5).

That is, if the value of  $\rho$  satisfies the criteria in Theorem 2, every limit point generated by the NS-Alternating algorithm is a KKT point of the optimization proposed by CDE under the global optimal kernel matrices of  $QSDP$ . Please see the web link<sup>3</sup> for full version of the proofs.

## 6 MASTER+

### 6.1 Motivation

To further speed up the MASTER framework, we propose a *scalable* framework MASTER+, illustrated in Algo. 2. Its core idea is that we first group accounts, probably belonging to the same users, into clusters and then simultaneously perform the MASTER framework in each cluster in parallel.

Regarding the clustering, the key challenge lies in the *dual guarantee of high efficiency and high accuracy*. The high efficiency requires the clusters tend to be balanced as the computing scale is determined by the largest one. The high accuracy requires the clusters tend to be overlapped, *i.e.*, soft clustering, so that each account in the cluster, especially for the marginal account, is capable to identify its correspondent account. Note that, this process itself cannot be computationally expensive.

To address these issues, specifically, to facilitate the clustering in the parallelization, we first design an efficient Augmented Pre-Embedding (APE) model encoding multiple overlapped networks into a vector space, then design an efficient Balanced-aware Fuzzy Clustering (BFC) algorithm to perform soft clustering in the embedded vector space. The intrinsic benefit of this APE-BFC process lies in that the dual guarantee is achieved via balance-aware and soft clustering.

---

#### Algorithm 2: MASTER+

```

Input: observed  $\{\mathbf{G}^{(\cdot)}, \mathbf{A}^{(\cdot)}\}$  of  $S^{(\cdot)}$  and  $\tilde{L}$ 
Output:  $\mathbf{H}^{(\cdot)}$  for each  $S^{(\cdot)}$  of the joint latent space
// perform APE sub-process
1 Compute  $\mathbf{G}_P$  and  $\mathbf{W}_P$  for the augmented graph;
2 Solve the optimization (39) using SGD;
// perform BFC sub-process
3 while not converge do
4   | Compute the updating rules (46) and (47);
5 end
// perform MASTER in parallel
6 for each cluster do
7   | Call NS-Alternating( $\{S_R^{(\cdot)}\}, \hat{L}$ ) (Algo. 1);
8 end
9 return  $\{\mathbf{H}^{(\cdot)}\}$ ;

```

---

### 6.2 Augmented Pre-embedding

The goal of the Augmented Pre-Embedding (APE) is, for all the accounts of  $M$  overlapped networks, to obtain the pre-embedding matrix  $\mathbf{H}^{(P)} \in \mathbb{R}^{N_{sum} \times d}$ . The similarity of the accounts across networks and the proximity within individual network are captured in the pre-embeddings. This embedding transforms the raw input for clustering to facilitate parallelization in MASTER+.

To achieve this goal, we first construct an augmented graph, denoted as  $(\mathbf{G}_P, \mathbf{A}_P)$ .  $M$  social networks are integrated in the augmented graph by integrating the accounts

3. <https://www.zhongbaozhang.com/publications>

with known correspondence into one node. All edges of the integrated node within each network are inherited and the attributes are integrated. The adjacency matrix  $\mathbf{G}_P \in \mathbb{R}^{N_{sum} \times N_{sum}}$  represents the structure space, where  $N_{sum} = \sum_{i=1}^M N^{(m)} - |\hat{L}|$ .  $\mathbf{A}_P \in \mathbb{R}^{N_{sum} \times l_{sum}}$  represents the attribute space and each row denotes the  $l_{sum}$ -dimensional attribute vector integrated from  $\{\mathbf{a}_{(\cdot)}^{(i)}\}$ .

Second, we construct the similarity matrix. The first-order proximity of the structure space, encoded in the adjacency matrix  $\mathbf{G}_P$ , and the pairwise similarity of the attribute, encoded in the inner product of attribute matrix ( $\mathbf{A}_P^T \mathbf{A}_P = \mathbf{W}_P$ ), are incorporated, *i.e.*,  $\mathbf{G}_P + \mathbf{W}_P$ .

Finally, we approximate the pairwise similarity within the augmented graph by the inner product of the  $\mathbf{h}$ . To learn the pre-embeddings, we perform the symmetric matrix factorization via the following optimization:

$$\min d(\mathbf{G}_P + \mathbf{W}_P, \mathbf{H}^{(P)} \mathbf{H}^{(P)T}). \quad (39)$$

Euclidean distance is employed in this paper, *i.e.*,  $d(\mathbf{X}, \mathbf{Y}) = \|\mathbf{X} - \mathbf{Y}\|_F^2$ . The pre-embeddings  $\mathbf{h}$  can be learned via 1<sup>st</sup>-order method. The gradient of  $d$  w.r.t.  $\mathbf{h}_i^{(P)}$  is given below:

$$\nabla_{\mathbf{h}_i^{(P)}} d = - \sum_{j \in \mathcal{N}(i)} ([\mathbf{G}_P + \mathbf{W}_P]_{ij} - \langle \mathbf{h}_i^{(P)}, \mathbf{h}_j^{(P)} \rangle) \mathbf{h}_j, \quad (40)$$

where  $\mathcal{N}(i) = \{j | [\mathbf{G}_P]_{ij} = 1\}$  denotes the neighbors set of the account  $i$  in the augmented graph. With the gradient, Stochastic Gradient Descent (SGD) is readily to be employed for solving this non-convex optimization.

### 6.3 Balance-aware Fuzzy Clustering

Inspired by the Fuzzy C-Means (FCM), we propose a novel Balance-aware Fuzzy Clustering (BFC) algorithm to group  $N_{sum}$  pre-embeddings into  $K$  overlapped clusters, where  $K$  is a pre-defined positive integer. We first formulate the FCM in the matrix form as follows:

$$\begin{aligned} \min_{\mathbf{U}, \{\mathbf{c}\}} & \|\mathbf{U}_m(\mathbf{U}, m) \odot \mathbf{D}(\mathbf{h}^{(P)}, \mathbf{c})\|_F^2 \quad (m > 1) \\ \text{s. t. } & \|\mathbf{U}_{[j]}\|_{p,p}=1, \forall j, \end{aligned} \quad (41)$$

where

$$[\mathbf{U}_m(\mathbf{U}, m)]_{ij} = u_{ij}^m \quad (42)$$

$$[\mathbf{D}(\mathbf{h}^{(P)}, \mathbf{c})]_{ij} = \|\mathbf{h}_j^{(P)} - \mathbf{c}_i\|_F^2. \quad (43)$$

In this optimization,  $\mathbf{U} \in \mathbb{R}^{K \times N_{sum}}$  is the membership matrix. The element of  $\mathbf{U}$  ( $u_{ij}$ ) indicates the probability of embedding  $\mathbf{h}_j^{(P)}$  assigned to the cluster whose mean is  $\mathbf{c}_i$ .  $m$  is a parameter controlling the fuzziness ( $m > 1$ ).  $\mathbf{U}_{[j]}$  denotes the  $j^{th}$  column of  $\mathbf{U}$  and  $\|\mathbf{x}\|_p$  denotes the  $p$ -norm of a  $n$ -dimensional vector  $\mathbf{x}$ , which is defined as  $\|\mathbf{x}\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$ .

We then introduce the balance-aware regularizer  $\|\mathbf{U}\|_p$ . Similar to  $\|\mathbf{x}\|_p$ ,  $\|\mathbf{X}\|_p$  denotes the  $p$ -norm of a matrix  $\mathbf{X}$ , which is defined as  $\|\mathbf{X}\|_p = (\sum_{i,j} |\mathbf{X}_{ij}|^p)^{1/p}$ . The balance-aware regularizer will enlarge the probability of the pre-embeddings (points) away from the means assigned to small clusters (*i.e.*, the clusters with fewer members), where the Euclidean distance is employed.

The overall objective of our proposed BFC is given as follows:

$$\begin{aligned} \min_{\mathbf{U}, \{\mathbf{c}\}} & \|\mathbf{U}_m(\mathbf{U}, m) \odot \mathbf{D}(\mathbf{h}^{(P)}, \mathbf{c})\|_F^2 + \zeta \|\mathbf{U}\|_{p|p=m}^m \quad (m > 1) \\ \text{s. t. } & \|\mathbf{U}_{[j]}\|_{p|p=1} = 1, \forall j, \end{aligned} \quad (44)$$

where  $\zeta$  is a positive parameter weighting the importance of the balance-awareness.

To address the optimization (45) of BFC, we introduce Lagrangian multipliers  $\lambda$  to construct the augmented objective:

$$\mathcal{L}_B = \sum_{i=1}^K \sum_{j=1}^{N_{sum}} u_{ij}^m (\|\mathbf{h}_j^{(P)} - \mathbf{c}_i\|^2 + \zeta) + \sum_{j=1}^{N_{sum}} \lambda_j \sum_{i=1}^K (u_{ij}^m - 1). \quad (45)$$

Set the partial differentials  $\frac{\partial}{\partial u_{ij}} \mathcal{L}_B$ ,  $\frac{\partial}{\partial \mathbf{c}_i} \mathcal{L}_B$  and  $\frac{\partial}{\partial \lambda_j} \mathcal{L}_B$  to be 0, respectively. Note that, the multipliers  $\lambda$  are eliminated in solving the equation system. Finally, we obtain the updating rules as follows:

$$\mathbf{c}_i = \left( \sum_{j=1}^{N_{sum}} u_{ij}^m \mathbf{h}_j^{(P)} \right) \left( \sum_{i=1}^K u_{ij}^m \right)^{-1}, \quad (46)$$

$$u_{ij} = \sum_{k=1}^K \left( \frac{\|\mathbf{h}_j^{(P)} - \mathbf{c}_i\|^2 + \zeta}{\|\mathbf{h}_j^{(P)} - \mathbf{c}_k\|^2 + \zeta} \right)^{-\frac{1}{m-1}}. \quad (47)$$

The procedure of MASTER+ is summarized in Algo. 2. The input of MASTER (line 7) is the set  $\hat{L}$  and reduced networks  $\{S_R^{(\cdot)}\}$ . For each cluster,  $S_R^{(m)}$  is formed by the accounts of  $S^{(m)}$  in the generated clusters.

The benefit of this APE-BFC process, regarding the clustering, is two-folded: (1) The high efficiency is enhanced via the balanced-awareness. (2) The high accuracy is ensured via the fuzziness.

#### 6.4 Computational Complexity

We first analyze the APE sub-process (line 2 in Algorithm 2). In each iteration for updating pre-embeddings, the computational complexity is  $O(E_{sum})$ , where  $E_{sum} = \|\mathbf{G}_P\|_0$  and  $\|\mathbf{G}_P\|_0$ , 0-norm of  $\mathbf{G}_P$ , is the number of non-zero elements of  $\mathbf{G}_P$ . We proceed to analyze the BFC sub-process (line 3-5 in Algorithm 2). In each iteration, the computational complexity of BFC algorithm is  $O(KdN_{sum})$  and  $d$  is the dimension of the pre-embedding, which is the same as that of CDE in our paper. With given  $K$  and  $d$ , BFC enjoys high scalability as its computational complexity is proportional to the number of data points (users). Note that, both APE and BFC will converge in a few iterations, and the computational complexity of both sub-processes is less than that of processing MASTER in parallel.

Finally, we can conclude that the overall computation complexity of MASTER+ is in the order of  $O(N_{Rmax}^2)$ , where  $N_{Rmax} = \max\{N_R^{(m)}\}$  is the number of accounts in the largest reduced network. Both MASTER and MASTER+ are of square computation complexity regarding the number of accounts. As  $N_{Rmax}$  is far less than  $N_{max}$ , the computation complexity of MASTER+ is far less than that of MASTER. It is obvious that, compared to MASTER, MASTER+ enjoys much higher efficiency.

The benefit of MASTER+ framework lies in that it is more efficient while preserving the accuracy of MASTER via the parallelization of the efficient APE-BFC process.

## 7 EXPERIMENT

### 7.1 Experimental Setup

**Datasets:** We use the Twitter-Foursquare dataset [9]. Since Foursquare can be registered by Twitter account, we regard this part of data as ground truth. Twitter dataset consists of 5,220 users and 164,917 connections while Foursquare dataset consists of 5,315 users and 76,972 connections. There are 1,610 shared users. We evaluated the performance of competing methods in the following two cases:

- **Bi-network case:** We generated a series of network pairs with different overlap rates ( $\eta$ ), measured by  $\frac{2N_s}{N_T+N_F}$ , where  $N_s$ ,  $N_T$  and  $N_F$  denote the number of shared users, Twitter users and Foursquare users respectively. Specifically, for each network pair, we sample users from this dataset according to  $\eta$ , called dense pairs. Moreover, in order to evaluate the robustness, we generated a sparse pair for each  $\eta$ -overlap dense pair by randomly removing 30% connections.
- **Multi-network case:** We generated two networks from Twitter by inheriting all the users and randomly sampling 70% of the connections and attributes (e.g., profiles, generated contents), and also generated two networks from Foursquare via the same process. Then, we simulated a series of four-network dense groups and corresponding sparse groups with different values of  $\eta$ .

**Performance Metric:** We evaluated all the competitive methods by the *hit-precision* of the candidate lists, which is measured by  $\frac{1}{M} \sum_m \mathbb{E}_i \left[ \frac{(K+1)-hit_m(v_i^{(\cdot)})}{K} \right]$ , where  $\mathbb{E}[\cdot]$  denotes the expectation and  $K = 5$ . For instance, for social network  $S^{(m)}$ , we obtain a top-K candidates list  $\{v_6^{(m)}, v_9^{(m)}, \dots, v_1^{(m)}\}$  for  $v_5^{(1)}$ . If  $v_9^{(m)}$  hits the ground truth,  $hit_m(v_5^{(1)}) = 2$ ;  $hit_m(v_5^{(1)}) = 6$  for not hitting.

**Competitive Methods:** To evaluate the performance of the proposed MASTER and MASTER+ frameworks, we compared them with several state-of-the-art methods.

- **ULink** [16]: This method links user identities by modeling users' attributes in the latent user space. In this paper, we implemented the ULink of the concave-convex procedure.
- **PALE** [15]: This method reconciles social networks via an embedding-matching framework. In this paper, we implemented the PALE whose matching component is a Multi-Layer Perception.
- **COSNET** [33]: This method considers the local as well as global consistency in reconciling multiple social networks. The optimization of the energy-based model is formulated and addressed by a proposed sub-gradient algorithm in the study.
- **NR-GL** [34]: This method expands the seed set, the set of known common users, with a proposed Uni-Rank metric in account of local and global features.
- **CoLink** [36]: This method employs a novel co-training algorithm manipulating the attribute-based model and the relationship-based model, and makes two models reinforce each other iteratively with an unsupervised fashion.

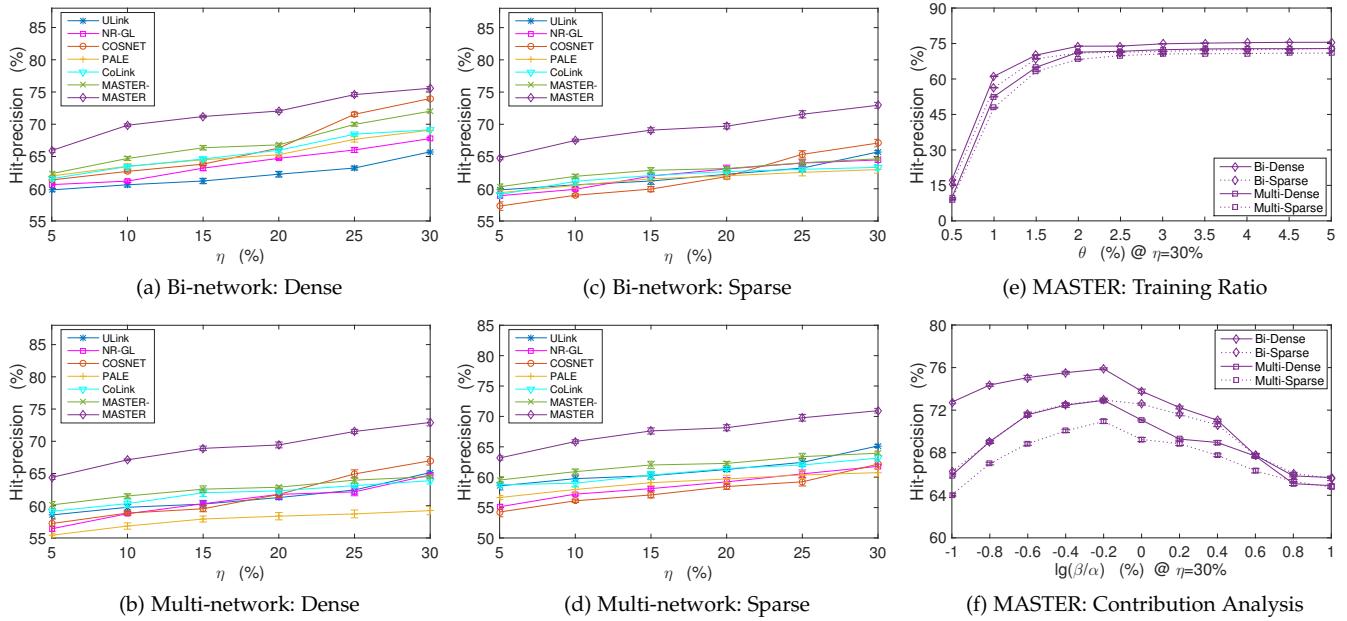


Fig. 2: Experimental results on Twitter-Foursquare dataset

- **MASTER-**: We implemented a degraded version of MASTER ignoring the attribute space (*i.e.*,  $\beta$  is set to 0) to study the importance of comprehensiveness.

In all experiments, the dimension of the representation vector is set to 100. The penalty coefficient  $\gamma$  of MASTER is set to be  $10^6$ . We repeated each experiment for 10 times and reported the mean with 95% confidence interval.

## 7.2 Evaluation on MASTER

**Bi-network Case:** We evaluated competitive methods on dense and sparse pairs with different  $\eta = [5\%, 10\%, \dots, 30\%]$ . Experimental results are reported in Fig. 2 (a). In all the experiments, MASTER achieves the highest hit-precision. MASTER has an improvement of 4.92%, 6.21% and 9.40% in average compared to COSNET, PALE and ULink respectively. It is expected. The reasons are two-folded: (1) In MASTER, both observations of attribute and structure space are comprehensively exploited. (2) The embeddings, capturing the intrinsic relation among users, facilitate the robust reconciliation. MASTER performs better than MASTER- consistently, demonstrating the necessity of comprehensiveness. COSNET struggles in defining local consistency and ULink performs the worst due to that it only considers the attribute space with lots of noise.

**Multi-network Case:** We conducted experiments on the four-network groups with different  $\eta = [5\%, 10\%, \dots, 30\%]$ , reported in Fig. 2 (b). It is evident that MASTER remains the best. This is expected owing to following reasons: MASTER consistently and robustly reconciles multiple social networks in the joint latent space where the information of both space is comprehensively captured and global inconsistency is naturally eliminated. In contrast, PALE presents low hit-precision due to global inconsistency. MASTER-, COSNET and ULink address the problem of global inconsistency. However, MASTER- ignores the information in attribute

space, COSNET is limited to the sparse information in this case and ULink still suffers from the noise in attribute space.

**Robustness:** To further evaluate the robustness of MASTER, we conducted experiments in the sparse case, reported in Fig. 2 (c) and (d). As shown in Fig. 2 (a), (b), (c) and (d), the proposed MASTER consistently outperforms its competitors under various settings of the real-world dataset regarding sparsity and multiplicity, validating the robustness. Moreover, zooming in the performance of MASTER in Fig. 2 (a) and (b), the performance of MASTER in dense network is competitive to that in sparse network, which gives detailed evidence on the robustness of MASTER. The underlying reason of robustness of the proposed MASTER lies in two-folds: (1) The proposed CDE model the with powerful learning ability recursively represents an account via the pairwise proximity regarding the structure and attributes. (2) The proposed NS-Alternating algorithm effectively learns the embeddings of the CDE model with solid theoretical guarantees. Thus, MASTER can effectively identify the intrinsic regularity of an account even in the partial observed dataset, *i.e.*, MASTER presents better robustness compared to its competitors.

**The impacts of parameters:** To further evaluate MASTER, we conducted experiments to evaluate the effect of training ratio  $\theta$  and the contribution of each space. Regarding the impact of  $\theta$ , we fix  $\eta = 30\%$ ,  $\beta/\alpha = 2/3$ , and vary the value of  $\theta$  to  $[0.5\%, 1\%, 1.5\%, \dots, 5\%]$ . We report the corresponding results in Fig. 2 (e). From this figure, we observe that the hit-precision of MASTER raises quickly as the training ratio increases from 0.5% to 2%, and slows down when  $\theta$  exceeds 2%. That is to say, MASTER can achieve good performance with relatively less label information. Regarding the impact of the contribution of each space, we set  $\eta = 30\%$ ,  $\theta = 5\%$ , and vary the value of  $lg(\beta/\alpha)$  to  $[-1, -0.8, -0.6, \dots, 1]$ . We report the corresponding results in Fig. 2 (f). From this figure, it can be inferred that the structure space has higher contribution than that in attribute space. A reasonable interpretation behind this is that people are not willing to pro-

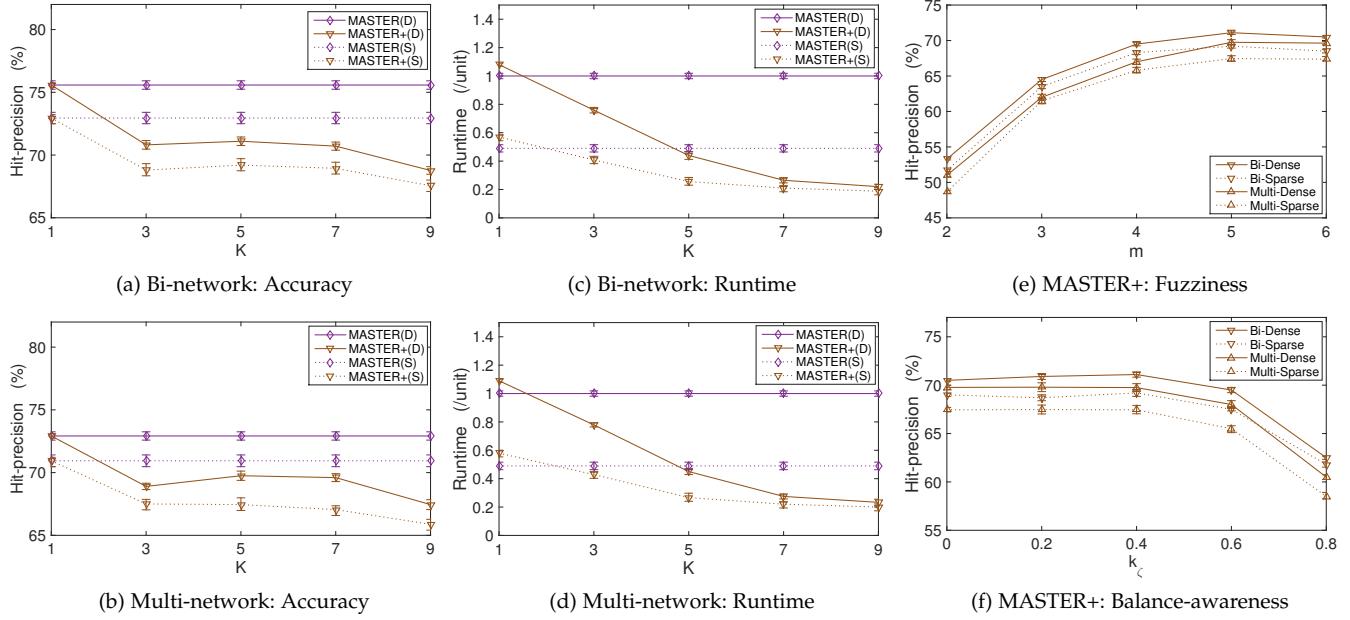


Fig. 3: Experimental results of MASTER+ on Twitter-Foursquare dataset

vide truthful personal information (e.g., location, birthday), which results in lots of noise in the attribute space.

### 7.3 Evaluation on MASTER+

**Accuracy:** First, regarding bi-network case, we evaluated MASTER+ on dense and sparse pairs with overlap rate  $\eta = 30\%$ . We varied the value of  $K$  to  $[1, 3, 5, \dots, 9]$ . The experimental results are reported in Fig. 3 (a). The hit-precision of MASTER+ is the same as that of MASTER when  $K = 1$ , as MASTER+ in fact degrades into MASTER in this case. The hit-precision remains smooth and shrinks when  $K$  is larger than 7. On average, the hit-precision of MASTER+ achieves 94.46% and 95.25% of that of MASTER on dense and sparse pairs, respectively. This is expected and the reasons are two-folded: (1) The simple yet effective pre-embedding captures the intra- and inter-network similarity of the social accounts so that the account embeddings of the same user are proximal in the embedded space. (2) The novel BFC algorithm generates the fuzzy clusters of enough search space via the soft clustering regularized by the balance-aware regularizer. However, as the number of the clusters becomes large enough (e.g.,  $K > 7$  on this Twitter-Foursquare dataset), the limited space of the cluster cannot cover the few account embeddings of the same user away from each other. Hence, the accuracy of MASTER+, in general, tends to be lower than that of MASTER in an acceptable extend.

Second, regarding the multi-network case, we evaluated MASTER+ on dense and sparse groups with overlap rate  $\eta = 30\%$ . We used the same value set of  $K$  as that of the bi-network case. The experimental results are reported in Fig. 3 (b). As shown in this figure, MASTER+ presents similar performance in the multi-network case to that of the bi-network case. Recall that the best-performance competitive methods in the dense and sparse scenario are COSNET with the hit-precision of 66.97% and ULink with the hit-precision

of 65.14%, respectively. The hit-precision of MASTER+ remains higher than that of the competitive methods of best-performance in both scenarios, verifying the effectiveness of the proposed MASTER+.

On average, the hit-precision of MASTER+ achieves 95.62% and 95.52% of that of MASTER on dense and sparse groups, respectively. The performance of MASTER+ in sparse data scenario is comparable to or even better than that of dense data scenario in both bi-network and multi-network case. Therefore, we can conclude that the proposed MASTER+ not only remains high effectiveness but also inherits high robustness of MASTER.

**Efficiency:** We proceeded to study the runtime of MASTER and MASTER+. It shares the same settings with last subsection (Accuracy). The experimental results are reported in Fig. 3 (c) and (d), where the runtime of MASTER is regarded as one time unit. Note that, MASTER+ ( $K = 1$ ) degrades into MASTER with the APE-FBC process. We can infer that the runtime of the APE-FBC process is  $1.08 - 1 = 0.08$  time unit, which is *trivial* compared to that of MASTER.

Recall the computational complexity in Section 4.4 and 5.4. It is easy to check that the ratio of the runtime between MASTER+ and MASTER is  $O(T_3 E_{sum} + T_4 K d N_{sum} + T_1 d N_{Rmax}^2 + T_2 d^3) / O(T_1 d N_{max}^2 + T_2 d^3)$ . This ratio is roughly asymptotic to  $(\frac{N_{Rmax}}{N_{max}})^2$  as the values of other terms (on numerator and denominator) are far less than that of the term notating the expense of matrix multiplications. As shown in Fig. 3 (c) and (d), compared to the runtime of MASTER, the runtime of MASTER+ is in accordance with the theoretical analysis and presents evident decrease in a near inverse-square pattern. This result convincingly signifies that MASTER+ owns much higher efficiency. This is expected. The reasons lies in that novel BFC algorithm (1) generates (overlapped) clusters with enough search space with the property of fuzziness and (2) avoids any reduced network growing too large with the balance-awareness

$m$	B-D	B-S	M-D	M-S
2	(37.01 $\pm$ 0.11)%	(22.61 $\pm$ 0.07)%	(38.03 $\pm$ 0.05)%	(23.55 $\pm$ 0.11)%
3	(39.16 $\pm$ 0.03)%	(23.59 $\pm$ 0.15)%	(40.15 $\pm$ 0.05)%	(24.63 $\pm$ 0.07)%
4	(42.95 $\pm$ 0.03)%	(24.58 $\pm$ 0.06)%	(43.85 $\pm$ 0.02)%	(25.58 $\pm$ 0.10)%
5	(44.05 $\pm$ 0.07)%	(25.60 $\pm$ 0.13)%	(45.02 $\pm$ 0.09)%	(26.71 $\pm$ 0.11)%
6	(45.96 $\pm$ 0.05)%	(26.71 $\pm$ 0.12)%	(46.95 $\pm$ 0.07)%	(27.49 $\pm$ 0.15)%

TABLE 2: MASTER+: The impact of parameter  $m$  (fuzziness) on runtime. B-D, B-S, M-D and M-S in the table represents the bi-network case of dense scenario, the bi-network case of sparse scenario, the multi-network case of dense scenario and the multi-network case of sparse scenario, respectively.

$k_\zeta$	B-D	B-S	M-D	M-S
0	(46.05 $\pm$ 0.07)%	(25.84 $\pm$ 0.07)%	(46.59 $\pm$ 0.03)%	(27.52 $\pm$ 0.05)%
0.2	(45.12 $\pm$ 0.15)%	(25.70 $\pm$ 0.10)%	(46.03 $\pm$ 0.11)%	(26.87 $\pm$ 0.09)%
0.4	(44.05 $\pm$ 0.07)%	(25.60 $\pm$ 0.13)%	(45.02 $\pm$ 0.09)%	(26.71 $\pm$ 0.11)%
0.6	(43.75 $\pm$ 0.03)%	(26.26 $\pm$ 0.09)%	(44.86 $\pm$ 0.02)%	(27.11 $\pm$ 0.15)%
0.8	(43.88 $\pm$ 0.03)%	(25.33 $\pm$ 0.10)%	(44.78 $\pm$ 0.11)%	(26.91 $\pm$ 0.07)%

TABLE 3: MASTER+: The impact of parameter  $\zeta$  (balance-awareness) on runtime. The meaning of B-D, B-S, M-D and M-S is same as that of the Table. 2.

regularizer. Therefore, we can further conclude that the proposed MASTER+ will largely speed up MASTER while preserving the high effectiveness and robustness.

**The impacts of parameters:** To further evaluate MASTER+, we proceeded to analyze the impacts of the parameter  $m$ , controlling the fuzziness of the clustering, and the parameter  $\zeta$ , controlling the balance-awareness.

Regarding the fuzziness, we conducted experiments in both bi-network case and multi-network case with the fixed overlap rate  $\eta = 30\%$  and the number of clusters  $K = 5$ . We let  $\zeta = \frac{k_\zeta}{K^2} \mathbb{E}_{\mathbf{h}^{(P)}} [\sum_{i=1}^K d(\mathbf{c}_i^{(0)}, \mathbf{h}^{(P)})]$  and fixed the value of  $k_\zeta$  to 0.4, where  $\mathbf{c}_i^{(0)}$  denotes the initial cluster mean,  $\mathbf{h}^{(P)}$  denotes the pre-embeddings and  $d(\cdot, \cdot)$  is the distance metric used in clustering. We varied the value of  $m$  to  $[1, 2, \dots, 5]$ . We study the impacts of  $m$  on accuracy and runtime, whose experimental results are reported in Fig. 3 (e) and Table. 2, respectively. Similar to the evaluation of efficiency, the runtime of MASTER in bi-network case and multi-network case of the dense scenario is regarded as the time unit for bi-network case and multi-network case, respectively. As shown in Fig. 3 (e), the runtime becomes longer as the value of  $m$  increases. The reason lies in that the largest reduced network tends to be larger when the value of  $m$  becomes larger. As shown in Table. 2, we observe that the hit-precision of MASTER raises as the value of  $m$  increases from 2 to 4 and slows down when  $m$  exceeds 4. We can infer that the search space within the clusters are large enough when  $m \geq 4$ . We set  $m = 5$  when reporting the accuracy and runtime of MASTER+. More importantly, the results show that fuzziness is necessary to ensure the accuracy.

Regarding the balance-awareness, we also conducted experiments in both bi-network case and multi-network case. In this case, we first set  $\eta = 30\%$ ,  $K = 5$  and  $m = 5$ . Then, we varied the value of  $\zeta$  by varying the value of  $k_\zeta$  to  $[0, 0.2, 0.4, \dots, 0.8]$ . The proposed BFC degrades into FCM when  $k_\zeta = 0$ . We study the impacts of  $\zeta$  on the accuracy and runtime, whose experimental results are reported in Fig. 3 (f) and Table. 3, respectively. As shown in Fig. 3 (f), we observe that the runtime of MASTER decreases as the value of  $k_\zeta$  increases from 0 to 0.4 and slows down

when  $k_\zeta$  exceeds 0.4. This is expected and the reason lies in that BFC avoids a few reduced network growing too large as the balance-awareness is enhanced. As shown in Table. 3, we observe that the hit-precision of MASTER remains smooth and shrinks when  $k_\zeta$  exceeds 0.4. The reason is that the balance-awareness regularizer with too much weight will disturb the clustering of BFC. We set  $k_\zeta = 0.4$  when reporting the accuracy and runtime of MASTER+. We can conclude that the balance-awareness of reasonable extend will effectively decrease the runtime.

## 8 RELATED WORK

There are three lines of research related to the proposed frameworks: reconciling social networks, network embedding and clustering. We briefly summarize these related works as follows:

**Reconciling social networks:** This problem, also known as anchor link prediction or user identity linkage, is to identify the accounts across different social networks belonging to the same individual. This problem is first addressed by the study [29] via the similar pattern in user name. Most of the existing models [10], [11], [12], [15], [34], [36] focus on reconciling only two social networks. Some studies (e.g., [9], [12], [34]) incorporate heterogeneous attributes, e.g., profile, topical interest, for reconciliation, while others (e.g., [10], [11], [15]) focus on the structure information. A few models reconcile multiple networks. COSNET is prior work reconciling multiple social networks, and ULink gives an incremental solution. However, ULink [16] suffers from noises in attribute space while COSNET [33] struggles in defining local consistency. The study [15] argues that existing methods lack robustness, and proposes an embedding-matching framework, PALE. However, PALE is originally designed for reconciling two networks, and incurs global inconsistency in reconciling multiple social networks. Recently, CoLink [36], a novel unsupervised model, enhances the reconciliation by performing its attribute sub-model and structure sub-model iteratively.

Model	Multiplicity	Attributes	Structure	Robustness
MNA [9]		✓	✓	
ULink [16]	✓	✓		
COSNET [33]	✓	✓	✓	
IONE [11]			✓	
PALE [15]			✓	✓
MOBIUS [30]		✓		
HYDRA [12]		✓	✓	
User-Matching [10]			✓	
CoLink [36]		✓	✓	
MASTER/MASTER+	✓	✓	✓	✓

TABLE 4: A summary of typical reconciliation models

This is the first attempt to model the attribute alignment problem as machine translation. Most of the typical models are summarized in Table 4, regarding multiplicity, robustness, and whether or not attributed or structure information can be incorporated. The study [20] gives a comprehensive survey on this problem. The essential difference between our frameworks and the others lies in that we, for the first time, robustly reconcile multiple social networks, addressing all these limitations, i.e., robustness, multiplicity and comprehensiveness.

**Network embedding:** Network embedding aims to learn a low-dimensional vector representation for each node in the network, obtaining an isomorphism with respect to some required properties, e.g., proximity, community and other structural regularity. Network embedding can be addressed through several techniques: (1) matrix factorization, learning the low-rank space of original representation matrix, e.g., TADW [28], HOPE [17], LANE [7], M-NMF [26]; (2) deep neural network, learning the mapping function between two spaces, e.g., SDAE [3], SiNE [25]; (3) random walk, generating random path over a network to capture structure information, e.g., DeepWalk [18], Node2Vec [5]. Moreover, both DeepWalk and LINE [22] are proved to be equivalent to matrix factorization recently [19], which motivates us to employ such framework. However, different from all these models, the proposed CDE model, for the first time, jointly embedding multiple partially overlapped networks, is tailored for the problem of reconciling multiple social networks.

**Clustering:** Clustering is a typical task in machine learning and data mining. It is categorized into soft clustering (e.g., fuzzy C-Means [4], topic model [2]) and hard clustering (e.g., K-Means) by whether or not a data point can be assigned to more than one cluster. The proposed BFC algorithm is in line with soft clustering. To the best of our knowledge, the BFC algorithm is a prior work considering the balance across the clusters, where a novel balance-awareness regularizer is introduced in the matrix optimization. Please refer to the study [1], [27] for more about clustering methods. Distinguishing from all these clustering methods, the proposed BFC algorithm in MASTER+ framework is designed for performing clustering on the partially aligned social networks, which aims at enhancing the scalability of reconciling multiple social networks.

## 9 CONCLUSION

In this paper, we propose the MASTER and MASTER+ frameworks, for the first time, robustly reconciling multiple social networks. In MASTER, we design the CDE model to embed and reconcile multiple networks in the joint latent space constructed via uni- and joint-embedding. We formulate the CDE model into a unified optimization on the semidefinite cone. We then design an effective NS-Alternating algorithm to solve this non-convex optimization of CDE and give the sufficient condition of its KKT convergence. To further speed up MASTER, we propose a scalable MASTER+ framework of dual guarantee (i.e., high efficiency and high accuracy), which performs MASTER in each cluster in parallel. Regarding the clustering, we first design the efficient APE model to transform the multiple overlapped networks into a vector space, then design a novel BFC algorithm to perform the soft clustering in the embedded vector space. We conducted extensive experiments on real-world datasets and demonstrate that MASTER and MASTER+ outperform several state-of-the-art methods. Moreover, MASTER+ inherits the effectiveness of MASTER and enjoys higher efficiency.

## ACKNOWLEDGMENTS

This work was supported in part by the following funding agencies of China: National Key Research and Development Program under Grant 2018YFB1003804 and National Natural Science Foundation under Grant 61602050 and U1534201.

## REFERENCES

- [1] C. C. Aggarwal and C. K. Reddy. *Data clustering: algorithms and applications*. CRC press, 2013.
- [2] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.
- [3] S. Cao, W. Lu, and Q. Xu. Deep neural networks for learning graph representations. In *Proceedings of the AAAI*, pages 1145–1152, 2016.
- [4] J. C. Dunn. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3(3):32–57, 1973.
- [5] A. Grover and J. Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the SIGKDD*, pages 855–864, 2016.
- [6] M. Hong, Z.-Q. Luo, and M. Razaviyayn. Convergence analysis of alternating direction method of multipliers for a family of non-convex problems. *SIAM Journal on Optimization*, 26(1):337–364, 2016.
- [7] X. Huang, J. Li, and X. Hu. Label informed attributed network embedding. In *Proceedings of the WSDM*, pages 731–739, 2017.
- [8] V. Hutson, J. Pym, and M. Cloud. *Applications of functional analysis and operator theory*, volume 200. Elsevier, 2005.

- [9] X. Kong, J. Zhang, and P. S. Yu. Inferring anchor links across multiple heterogeneous social networks. In *Proceedings of the CIKM*, pages 179–188, 2013.
- [10] N. Korula and S. Lattanzi. An efficient reconciliation algorithm for social networks. *Proceedings of the VLDB*, 7(5):377–388, 2014.
- [11] L. Liu, W. K. Cheung, X. Li, and L. Liao. Aligning users across social networks using network embedding. In *Proceedings of the IJCAI*, pages 1774–1780, 2016.
- [12] S. Liu, S. Wang, F. Zhu, J. Zhang, and R. Krishnan. Hydra: Large-scale social identity linkage via heterogeneous behavior modeling. In *Proceedings of the SIGMOD*, pages 51–62, 2014.
- [13] S. Lu, M. Hong, and Z. Wang. A nonconvex splitting method for symmetric nonnegative matrix factorization: Convergence analysis and optimality. *IEEE Trans. on Signal Processing*, 2017.
- [14] T. Man, H. Shen, X. Jin, and X. Cheng. Cross-domain recommendation: An embedding and mapping approach. In *Proceedings of the IJCAI*, pages 2464–2470, 2017.
- [15] T. Man, H. Shen, S. Liu, X. Jin, and X. Cheng. Predict anchor links across social networks via an embedding approach. In *IJCAI*, pages 1823–1829, 2016.
- [16] X. Mu, F. Zhu, E.-P. Lim, J. Xiao, J. Wang, and Z.-H. Zhou. User identity linkage by latent user space modelling. In *Proceedings of the SIGKDD*, pages 1775–1784, 2016.
- [17] M. Ou, P. Cui, J. Pei, Z. Zhang, and W. Zhu. Asymmetric transitivity preserving graph embedding. In *Proceedings of the SIGKDD*, pages 1105–1114, 2016.
- [18] B. Perozzi, R. Al-Rfou, and S. Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the SIGKDD*, pages 701–710, 2014.
- [19] J. Qiu, Y. Dong, H. Ma, J. Li, K. Wang, and J. Tang. Network embedding as matrix factorization: Unifying deepwalk, line, pte, and node2vec. In *Proceedings of the WSDM*, 2018.
- [20] K. Shu, S. Wang, J. Tang, R. Zafarani, and H. Liu. User identity linkage across online social networks: A review. *ACM SIGKDD Explorations Newsletter*, 18(2):5–17, 2017.
- [21] S. Su, L. Sun, Z. Zhang, G. Li, and J. Qu. Master: across multiple social networks, integrate attribute and structure embedding for reconciliation. In *Proceedings of the IJCAI*, 2018.
- [22] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei. Line: Large-scale information network embedding. In *Proceedings of the WWW Conference*, pages 1067–1077, 2015.
- [23] K.-C. Toh. An inexact primal-dual path following algorithm for convex quadratic sdp. *Mathematical programming*, 112(1):221–254, 2008.
- [24] A. Vandaele, N. Gillis, Q. Lei, K. Zhong, and I. Dhillon. Efficient and non-convex coordinate descent for symmetric nonnegative matrix factorization. *IEEE Trans. on Signal Processing*, 64(21):5571–5584, 2016.
- [25] S. Wang, J. Tang, C. Aggarwal, Y. Chang, and H. Liu. Signed network embedding in social media. In *Proceedings of the SDM*, pages 327–335, 2017.
- [26] X. Wang, P. Cui, J. Wang, J. Pei, W. Zhu, and S. Yang. Community preserving network embedding. In *Proceedings of the AAAI*, pages 203–209, 2017.
- [27] R. Xu and D. Wunsch. Survey of clustering algorithms. *IEEE Transactions on neural networks*, 16(3):645–678, 2005.
- [28] C. Yang, Z. Liu, D. Zhao, M. Sun, and E. Y. Chang. Network representation learning with rich text information. In *Proceedings of the IJCAI*, pages 2111–2117, 2015.
- [29] R. Zafarani and H. Liu. Connecting corresponding identities across communities. In *Proceedings of the ICWSM*, volume 9, pages 354–357, 2009.
- [30] R. Zafarani and H. Liu. Connecting users across social media sites: a behavioral-modeling approach. In *Proceedings of the SIGKDD*, pages 41–49, 2013.
- [31] Q. Zhan, J. Zhang, S. Wang, P. S. Yu, and J. Xie. Influence maximization across partially aligned heterogeneous social networks. In *Proceedings of the PAKDD*, pages 58–69, 2015.
- [32] J. Zhang and P. S. Yu. Pct: partial co-alignment of social networks. In *Proceedings of the WWW Conference*, pages 749–759, 2016.
- [33] Y. Zhang, J. Tang, Z. Yang, J. Pei, and P. S. Yu. Cosnet: Connecting heterogeneous social networks with local and global consistency. In *Proceedings of the SIGKDD*, pages 1485–1494, 2015.
- [34] Z. Zhang, Q. Gu, T. Yue, and S. Su. Identifying the same person across two similar social networks in a unified way: Globally and locally. *Information Sciences*, 395:53–67, 2017.
- [35] Z. Zhang, J. Wen, L. Sun, Q. Deng, S. Su, and P. Yao. Efficient incremental dynamic link prediction algorithms in social network. *Knowledge-Based Systems*, 132:226–235, 2017.
- [36] Z. Zhong, Y. Cao, M. Guo, and Z. Nie. Colink: An unsupervised framework for user identity linkage. In *AAAI Conference on Artificial Intelligence*, 2018.



**Zhongbao Zhang** received the Ph.D. Degree in Computer Science from Beijing University of Posts and Telecommunications, China, in 2014. He is an associate professor in Beijing University of Posts and Telecommunications now. His research interests include network visualization, social network and big data.



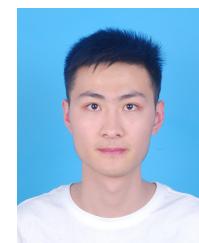
**Li Sun** is a 2nd year Ph.D. Candidate in Computer Science, and received the B.E. Degree in Internet of Things in 2016, both from Beijing University of Posts and Telecommunications. He won the National Scholarships in 2014 and 2015, and won the Outstanding Graduate of Beijing in 2016. His research areas are machine learning and data mining, especially social network analysis and network embedding.



**Sen Su** received the Ph.D. Degree in Computer Science from the University of Electronic Science and Technology, China in 1998. He is currently a Professor of Beijing University of Posts and Telecommunications. His research interests include data privacy and data mining.



**Jielun Qu** received the B.E. Degree in Telecommunication Engineering with Management, Beijing University of Posts and Telecommunications. He is with Tandon School of Engineering, New York University for M.S. degree from Aug. 2018. His main research interest is data mining.



**Gen Li** received the B.S. Degree in Applied Physics in 2017 and is working towards the master degree in Computer Science, both from Beijing University of Posts and Telecommunications. His current research interests include social network analysis and personal credit assessment.