MASTER: across Multiple social networks, integrate Attribute and STructure Embedding for Reconciliation

Sen Su, Li Sun, Zhongbao Zhang*, **Gen Li** and **Jielun Qu**
State Key Laboratory of Networking and Switching Technology,
Beijing University of Posts ands Telecommunications, Beijing 100876, China
{susen,l.sun,zhongbaozb,ligen,qujl}@bupt.edu.cn,

Abstract

Recently, reconciling social networks receives significant attention. Most of the existing studies have limitations in the following three aspects: multiplicity, comprehensiveness and robustness. To address these three limitations, we rethink this problem and propose the MASTER framework, *i.e.*, across Multiple social networks, integrate Attribute and STructure Embedding for Reconciliation. In this framework, we first design a novel Constrained Dual Embedding model by simultaneously embedding and reconciling multiple social networks to formulate our problem into a unified optimization. To address this optimization, we then design an effective algorithm called NS-Alternating. We also prove that this algorithm converges to KKT points. Through extensive experiments on real-world datasets, we demonstrate that MASTER outperforms the state-of-the-art approaches.

1 Introduction

Nowadays, social network is becoming increasingly important in people's lives. People often have several social network accounts, *e.g.*, Twitter for news, Facebook for friends and LinkedIn for jobs. However, these accounts are often independent from each other. It arises the problem of identifying the corresponding accounts belonging to the same individual, which is termed as reconciling social networks. Reconciling social networks can support a wide range of applications, *e.g.*, network fusion [Zhang and Yu, 2016], link prediction [Zhang *et al.*, 2017b] and cross-domain recommendation [Man *et al.*, 2017].

This problem still remains open as most of the existing methods have several limitations as follows:

- **Multiplicity**: In real world, people usually have several social network accounts. However, most of the existing methods [Kong *et al.*, 2013; Liu *et al.*, 2014; Korula and Lattanzi, 2014; Liu *et al.*, 2016; Man *et al.*, 2016] focus on reconciling only two social networks and cannot pairwise reconcile multiple social networks due to the global inconsistency.

*Corresponding Author

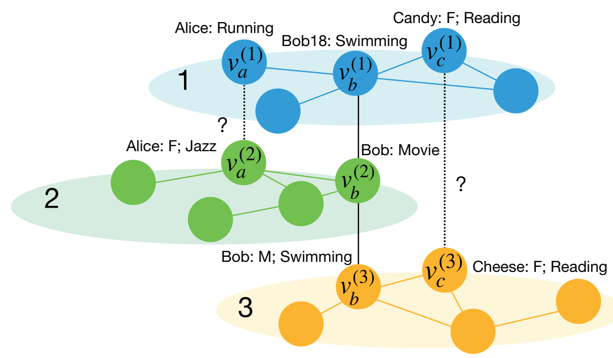


Figure 1: This figure demonstrates the problem of reconciling across multiple social networks. Graphs of different colors denote different social networks. Nodes (social accounts) are associated with attributes (*e.g.*, screen name, gender and hobby). Black lines denote the correspondences among the nodes. The full lines are known in advance while the dotted ones (with marks) are to be identified.

- **Comprehensiveness**: Social networks usually have two categories of spaces, *i.e.*, attribute space and structure space. However, most of existing methods [Korula and Lattanzi, 2014; Liu *et al.*, 2016; Man *et al.*, 2016; Mu *et al.*, 2016] do not comprehensively exploit the information of both spaces to reconcile social networks.
- **Robustness**: The social network is noisy and most of the existing methods [Zafarani and Liu, 2013; Kong *et al.*, 2013; Liu *et al.*, 2014; Zhang *et al.*, 2015] struggle in defining effective features sensitive to data. Therefore, they are still away from robustly reconciling networks.

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

The answer is **YES**! In this paper, for the first time, we propose the **MASTER** framework, *i.e.*, across **M**ultiple social networks, integrate **A**ttribute and **ST**ructure **E**mbedding for **R**econciliation. In this framework, there are two main challenges: **(1) Modeling**: 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 is to be addressed in the embedding. **(2) Optimizing**: The network embedding problem is often formulated as an optimization problem. In our framework, the optimization problem behind tends to be non-convex and hence much tougher.

To address the issue **(1)**, we design a novel **C**onstrained **D**ual **E**mbedding (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 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 and we finally give the formulation of the unified optimization.

To address the issue **(2)**, we design an effective NS-Alternating algorithm to approach the optima of the high-order matrix optimization. Specifically, we first reformulate the optimization problem, inspired by a recent advance in computational mathematics. We then alternately solve the representation matrices subproblem and the kernel matrices subproblem of the reformulated problem 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.

We validate MASTER through extensive experiments on real-world datasets and show that MASTER outperforms several state-of-the-art methods.

To summarize, we make the following contributions:

- To the best of our knowledge, our proposed MASTER is the first attempt to robustly reconcile multiple social networks comprehensively exploiting attribute and structure information via an embedding approach.
- We design a novel model (CDE) to formulate the problem of reconciliation into a unified optimization.
- We design an effective NS-Alternating algorithm to address the optimization and prove that it converges to KKT points.
- We conducted extensive experiments on real-world datasets and the experiment results demonstrate the superiority of our approach.

2 Problem Definition

In this paper, we consider a set of M social networks $\{S^{(m)}\}$. A social network $S^{(m)}$ of $N^{(m)}$ 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 $v_i^{(m)}$. Part of the 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 $\phi^{(1)}(v_{(\cdot)}^{(1)}) = ... = \phi^{(m)}(v_{(\cdot)}^{(m)}) = ... = \phi^{(M)}(v_{(\cdot)}^{(M)})$ to identify correspondences of shared users.*

To address this problem, we propose the MASTER framework. In MASTER, we design a novel model, **C**onstrained **D**ual **E**mbedding (CDE), to formulate the problem into a unified optimization (Section 3). To address the optimization, we design an effective NS-Alternating algorithm (Section 4).

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 $\boldsymbol{h}_i^{(m)}$ is the d-dimensional vector of $v_i^{(m)}$ in the latent space, capturing the observations of both structure 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 [Tang *et al.*, 2015], 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 of $\mathbf{G}_i^{(m)}$ and $\mathbf{G}_j^{(m)}$, where $\mathbf{G}_i^{(m)}$ is the i^{th} row of $\mathbf{G}^{(m)}$.*

The inner product similarity is into account 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 η 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 $\boldsymbol{h}_i^{(m)}$. Assume that $\boldsymbol{h}_i^{(m)}$ is projected onto structure space and attribute space via different projection $\varphi(.)$. We introduce kernel technique to bridge the inner product $\langle \cdot, \cdot \rangle$ of the $\varphi(\boldsymbol{h}_i^{(m)})$ with that of $\boldsymbol{h}_i^{(m)}$, *i.e.*, $\langle \varphi(\boldsymbol{h}_i^{(m)}), \varphi(\boldsymbol{h}_j^{(m)}) \rangle = \boldsymbol{h}_i^{(m)} \mathbf{K}_\varphi^{(m)} \boldsymbol{h}_j^{(m)}$, 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 the optimization of a collaborative matrix factorization as below:

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

where $||\cdot||_F$ is Frobenius norm, $\mathbb{S}_+^d$ denotes semidefinite cone, and α and β are positive parameters weighting the observed similarities in structure space and attribute space respectively.

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, $\boldsymbol{h}_b^{(1)}$ will be closer to $\boldsymbol{h}_a^{(1)}$ than $\boldsymbol{h}_c^{(1)}$, and $\boldsymbol{h}_a^{(1)}$ and $\boldsymbol{h}_c^{(1)}$ 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) the $\boldsymbol{h}_l^{(\cdot)}$ 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 $\boldsymbol{h}_{(\cdot)}^{(m)}$.

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 $\boldsymbol{h}_{(\cdot)}^{(m)}$ according to the correspondences in $L(m,n)$. We obtain the constraints as follows:

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

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 rotating matrices $\tilde{\mathbf{D}}_p$ and $\tilde{\mathbf{D}}_q$. If M is odd,

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

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 $\tilde{}$ denote the block diagonal matrix, *i.e.*, $\tilde{\mathbf{X}} = diag(\{\mathbf{X}^{(m)}\})$. We obtain the equivalent equation pair:

$$L_p(\tilde{\mathbf{H}}) = ||\tilde{\mathbf{E}}\tilde{\mathbf{H}} - \mathcal{R}_p(\tilde{\mathbf{E}}\tilde{\mathbf{H}})||_F^2 = 0 \tag{3}$$

$$L_q(\tilde{\mathbf{H}}) = ||\tilde{\mathbf{E}}\tilde{\mathbf{H}} - \mathcal{R}_q(\tilde{\mathbf{E}}\tilde{\mathbf{H}})||_F^2 = 0 \tag{4}$$

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, which is equivalent to combining M objectives of uni-embedding, as follows:

$$\frac{\alpha}{2} ||\tilde{\mathbf{M}} - \tilde{\mathbf{H}}\tilde{\mathbf{B}}\tilde{\mathbf{H}}^T||_F^2 + \frac{\beta}{2} ||\tilde{\mathbf{W}} - \tilde{\mathbf{H}}\tilde{\mathbf{C}}\tilde{\mathbf{H}}^T||_F^2. \tag{5}$$

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

Finally, we remove the constraint by adding penalty with a coefficient γ and obtain the unified optimization objective:

$$\begin{aligned} \min_{\tilde{\mathbf{H}},\tilde{\mathbf{B}},\tilde{\mathbf{C}}} \quad & \frac{\alpha}{2} \left\| \tilde{\mathbf{M}} - \tilde{\mathbf{H}}\tilde{\mathbf{B}}\tilde{\mathbf{H}}^T \right\|_F^2 + \frac{\beta}{2} \left\| \tilde{\mathbf{W}} - \tilde{\mathbf{H}}\tilde{\mathbf{C}}\tilde{\mathbf{H}}^T \right\|_F^2 \\ & + \frac{\gamma}{2} [L_p(\tilde{\mathbf{H}}) + L_q(\tilde{\mathbf{H}})] \\ s.t. \quad & \tilde{\mathbf{B}}, \tilde{\mathbf{C}} \in \mathbb{S}_+^{Md} \end{aligned} \tag{6}$$

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 $\boldsymbol{h}_b^{(1)} = \boldsymbol{h}_b^{(2)} = \boldsymbol{h}_b^{(3)}$ 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.

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 [Lu *et al.*, 2017], we design an effective NS-Alternating algorithm. In this algorithm, we first reformulate the problem (6) and alternately solve the subproblems of the reformulation.

4.1 Problem Reformulation

The high-order objective (6) 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 formulate the problem as follows:

$$\begin{aligned} \min_{\tilde{\mathbf{H}},\tilde{\mathbf{B}},\tilde{\mathbf{C}},\mathbf{V}} \quad & \mathcal{J}(\tilde{\mathbf{H}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}, \mathbf{V}) \\ & = \frac{\alpha}{2} \left\| \tilde{\mathbf{M}} - \tilde{\mathbf{H}}\tilde{\mathbf{B}}\mathbf{V}^T \right\|_F^2 + \frac{\beta}{2} \left\| \tilde{\mathbf{W}} - \tilde{\mathbf{H}}\tilde{\mathbf{C}}\mathbf{V}^T \right\|_F^2 \\ & + \frac{\gamma}{2} (L_p(\tilde{\mathbf{H}}) + L_q(\tilde{\mathbf{H}})) \\ s.t. \quad & \tilde{\mathbf{B}}, \tilde{\mathbf{C}} \in \mathbb{S}_+^{Md}, \tilde{\mathbf{H}} = \mathbf{V}, ||\mathbf{V}_i||_2^2 < \tau, \forall i \end{aligned} \tag{7}$$

According to the study [Vandaele *et al.*, 2016], problem (6) is *equivalent* to problem (7) in the sense of KKT points if $\tau = \sqrt{C}$ is sufficiently large ($C = \max\{||\tilde{\mathbf{M}}||_F^2, ||\tilde{\mathbf{W}}||_F^2\}$ in our algorithm). That is, the KKT points of problem (6) and problem (7) have a one-to-one correspondence.

We alternately solve $(\tilde{\mathbf{H}}, \mathbf{V})$ and $(\tilde{\mathbf{B}}, \tilde{\mathbf{C}})$ of problem (7), 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:

$$\mathbf{V}^{(t+1)} = \arg\min_{||\mathbf{V}_i||_2^2<\tau,\forall i} \mathcal{L}(\tilde{\mathbf{H}}^{(t)}, \mathbf{V}; \mathbf{\Lambda}^{(t)}) + \frac{\xi^{(t)}}{2}||\mathbf{V} - \mathbf{V}^{(t)}||_F^2 \tag{8}$$

$$\tilde{\mathbf{H}}^{(t+1)} = \arg\min \mathcal{L}(\tilde{\mathbf{H}}, \mathbf{V}^{(t+1)}; \mathbf{\Lambda}^{(t)}) \tag{9}$$

$$\mathbf{\Lambda}^{(t+1)} = \mathbf{\Lambda}^{(t)} + \rho(\mathbf{V}^{(t+1)} - \tilde{\mathbf{H}}^{(t+1)}) \tag{10}$$

$$\xi^{(t+1)} = \tfrac{6}{\rho} \cdot \mathcal{J}(\tilde{\mathbf{H}}^{(t+1)}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}, \mathbf{V}^{(t+1)}) \tag{11}$$

$$\mathcal{L}(\tilde{\mathbf{H}}, \mathbf{V}; \mathbf{\Lambda}) = \mathcal{J}(\tilde{\mathbf{H}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}, \mathbf{V}) + \tfrac{\rho}{2}||\mathbf{V} - \tilde{\mathbf{H}} + \Lambda/\rho||_F^2 \tag{12}$$

$\mathcal{L}(\tilde{\mathbf{H}}, \mathbf{V}; \mathbf{\Lambda})$ 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 [Lu *et al.*, 2017]. The optimization *w.r.t.* $\mathbf{V}$ can be decomposed into k separable problems, each of which can be solved using gradient projection:

$$\mathbf{V}_{[i]}^{(r+1)} = proj_{\mathbf{V}}(\mathbf{V}_{[i]}^{(r)} - \lambda(\mathbf{A}_V^{(t)}\mathbf{V}_{[i]}^{(r)} - \mathbf{B}_{V}{}_{[i]}^{(t)})) \tag{13}$$

$$\mathbf{A}_V^{(t)} = \alpha\tilde{\mathbf{B}}\tilde{\mathbf{H}}^T\tilde{\mathbf{H}}\tilde{\mathbf{B}} + \beta\tilde{\mathbf{C}}^T\tilde{\mathbf{H}}^T\tilde{\mathbf{H}}\tilde{\mathbf{C}} + (\xi^{(t)} + \rho)\mathbf{I} \tag{14}$$

$$\mathbf{B}_V^{(t)} = \alpha\tilde{\mathbf{M}}\tilde{\mathbf{H}}\tilde{\mathbf{B}} + \beta\tilde{\mathbf{S}}^T\tilde{\mathbf{H}}\tilde{\mathbf{C}} + \xi^{(t)}\mathbf{V}^{(t)} + \rho\tilde{\mathbf{H}} - \mathbf{\Lambda} \tag{15}$$

$$proj_{\mathbf{V}}(\boldsymbol{w}) = \sqrt{\tau}\boldsymbol{w}/\max\{\sqrt{\tau}, ||\boldsymbol{w}||_2\},\ \forall \boldsymbol{w} \in \mathbb{R}^n \tag{16}$$

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

$$\begin{aligned}\nabla_{\tilde{\mathbf{H}}}\mathcal{L} = &\ \alpha(\tilde{\mathbf{H}}\tilde{\mathbf{B}}{\mathbf{V}^{(t+1)}}^T - \tilde{\mathbf{M}})\mathbf{V}^{(t+1)}\tilde{\mathbf{B}}^T \\ &+ \beta(\tilde{\mathbf{H}}\tilde{\mathbf{C}}{\mathbf{V}^{(t+1)}}^T - \tilde{\mathbf{S}})\mathbf{V}^{(t+1)}\tilde{\mathbf{C}}^T \\ &+ \gamma\tilde{\mathbf{E}}[\mathcal{R}_p(\tilde{\mathbf{E}}\tilde{\mathbf{H}}) + \mathcal{R}_q(\tilde{\mathbf{E}}\tilde{\mathbf{H}})] + 2\gamma\tilde{\mathbf{E}}^T\tilde{\mathbf{E}}\tilde{\mathbf{H}} \\ &+ \rho(\tilde{\mathbf{H}} - \mathbf{V}^{(t+1)} - \mathbf{\Lambda}^{(t)}/\rho),\end{aligned} \tag{17}$$

as $\tilde{\mathbf{D}}_p = \tilde{\mathbf{D}}_p^{-1}$ and $\tilde{\mathbf{D}}_q = \tilde{\mathbf{D}}_q^{-1}$.

4.3 Kernel Matrix Subproblem

Utilizing $||\mathbf{X}||_F^2 = tr(\mathbf{X}^T\mathbf{X})$, we reformulate the optimization *w.r.t.* $\mathbf{B}$ into an inner-product form ($\tilde{\mathbf{C}}$-subproblem is the same as $\tilde{\mathbf{B}}$-subproblem, and omitted due to the space limit):

$$\begin{aligned}\min_{\tilde{\mathbf{B}}}\quad & tr(\tilde{\mathbf{H}}\tilde{\mathbf{B}}\mathbf{V}^T\mathbf{V}\tilde{\mathbf{B}}\tilde{\mathbf{H}}^T) - 2tr(\mathbf{M}\tilde{\mathbf{H}}\tilde{\mathbf{B}}\tilde{\mathbf{V}}^T) \\ & = \langle \mathcal{Q}(\tilde{\mathbf{B}}), \tilde{\mathbf{B}}\rangle - 2\langle \mathbf{A}, \tilde{\mathbf{B}}\rangle, \quad \tilde{\mathbf{B}} \in \mathbb{S}_+^{Md},\end{aligned} \tag{18}$$

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 constraint $\tilde{\mathbf{H}} = \mathbf{V}$ holds when representation matrix subproblem converges. Let $\tilde{\mathbf{H}}^T\tilde{\mathbf{H}} = \mathbf{P}$ and we further analyze $\mathcal{Q}$. For arbitrary $\mathbf{X}, \mathbf{Y}$, the following equations hold:

$$\langle \mathcal{Q}(\mathbf{X}), \mathbf{X}\rangle = tr(\mathbf{PXPX}) = ||\mathbf{PX}||_F^2 \geq 0, \tag{19}$$

$$\langle \mathcal{Q}(\mathbf{X}), \mathbf{Y}\rangle = tr(\mathbf{PXPY}) = tr(\mathbf{XPYP}) = \langle \mathcal{Q}(\mathbf{Y}), \mathbf{X}\rangle. \tag{20}$$

That is, $\mathcal{Q}$ is semidefinite (Eq. 19) and self-adjoint (Eq. 20). According to the study [Toh, 2008], we conclude that:

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

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

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 $\quad(\tilde{\mathbf{H}}, \mathbf{V})^{(n+1)} = \arg\min l(\tilde{\mathbf{H}}, \mathbf{V}, \tilde{\mathbf{B}}^{(n+1)}, \tilde{\mathbf{C}}^{(n+1)})$;
5 $\quad(\tilde{\mathbf{B}}, \tilde{\mathbf{C}})^{(n+1)} = \arg\min_{\tilde{\mathbf{B}}, \tilde{\mathbf{C}} \in \mathbb{S}_+^{Md}} l(\tilde{\mathbf{H}}^{(n)}, \mathbf{V}^{(n)}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}})$;
6 $\quad n = n + 1$;
7 **end**
8 **return** $\{\mathbf{H}^{(\cdot)}\}$ from $\tilde{\mathbf{H}}$;

4.4 Convergence and Complexity Analysis

Convergence Analysis: Utilizing the convergence properties [Hong *et al.*, 2016], we can conclude that:

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

$$\rho_1 = 6N\tau\left(||\tilde{\mathbf{B}}||_F^4 + ||\tilde{\mathbf{C}}||_F^4\right) / \left(||\tilde{\mathbf{B}}||_F^2 + ||\tilde{\mathbf{C}}||_F^2\right)$$

$$\rho_2 + 2||\tilde{\mathbf{E}}||_F^2 = \tfrac{6}{\rho_2}\left(16N + N\tau\left(||\tilde{\mathbf{B}}||_F^2 + ||\tilde{\mathbf{C}}||_F^2\right)\right)^2$$

$$\rho_3 = ||\tilde{\mathbf{B}}||_F^2 + ||\tilde{\mathbf{C}}||_F^2 + ||\mathcal{R}_p(\tilde{\mathbf{E}}) + \mathcal{R}_q(\tilde{\mathbf{E}})||_F^2$$

We can claim that:

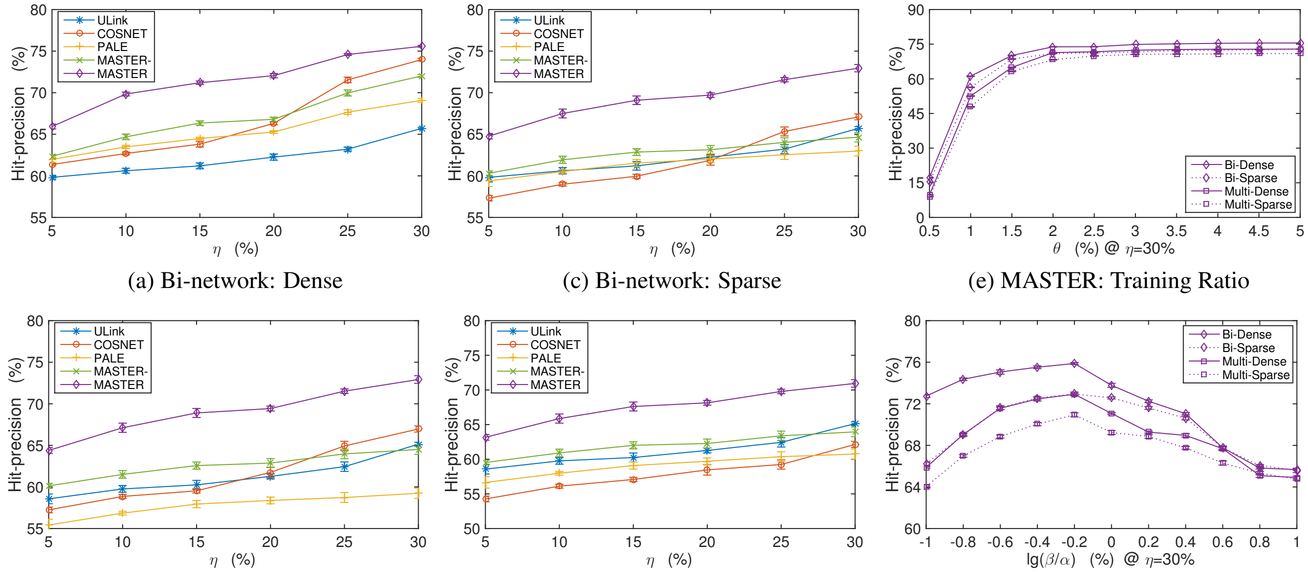
- *The equality constraint on the auxiliary matrix is satisfied in the limit,* i.e., $\lim_{t\to\infty}||\tilde{\mathbf{H}}^{(t)} - \mathbf{V}^{(t)}||_F^2 = 0$.
- *The sequence* $\{\tilde{\mathbf{H}}^{(t)}, \mathbf{V}^{(t)}, \mathbf{\Lambda}^{(t)}\}$ *generated by the NS-Alternating algorithm is bounded, and every limit point of the sequence is a KKT point of problem (6).*

The detailed proof can be found at the website.[1]

Computational Complexity: The outer loop of Algo 1 (line 3-7) achieves the satisfactory accuracy in a few iterations. 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. Therefore, the computational complexity depends on matrix operations in updating rules. The inversions in *QSDP* is $O(d^3)$ and matrix multiplication is $O(N_{max}^2 d)$ where $N_{max} = \max\{N^{(m)}\}$ and $d \ll \min\{N^{(m)}\}$. Moreover, there are quantities of optimized libraries (*e.g.*, OpenBLAS) to speed up the most expensive multiplication operations.

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

[1] http://www.zhongbaozhang.com/publications



(a) Bi-network: Dense
(c) Bi-network: Sparse
(e) MASTER: Training Ratio
(b) Multi-network: Dense
(d) Multi-network: Sparse
(f) MASTER: Contribution Analysis

Figure 2: Experimental result on Twitter-Foursquare dataset

5 Experiment

5.1 Experimental Setup

Datasets: We use the Twitter-Foursquare dataset [Kong *et al.*, 2013]. 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 (η), 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 η, called dense pairs. Moreover, in order to evaluate the robustness, we generated a sparse pair for each η-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 η.

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

Competitive Methods: To evaluate the performance of the proposed MASTER, we compared it with several state-of-the-art methods listed as follows:

- **ULink** [Mu *et al.*, 2016]: This method links user identities by modeling users' attributes in the latent user space. We implemented the CCP version of this method.
- **PALE** [Man *et al.*, 2016]: This method performs reconciliation in an embedding-matching framework. We implemented this method with the matching of MLP.
- **COSNET** [Zhang *et al.*, 2015]: This method considers the local and global consistency in reconciliation.
- **MASTER-**: We implemented a degraded version of MASTER ignoring the attribute space to emphasize the importance of comprehensiveness.

In all the experiments, the dimension of the representation vector in MASTER and PALE is set to be 100.

5.2 Experimental Results

We repeated each experiment for 10 times and both the mean and 95% confidence interval are reported. The experimental results are summarized as follows:

Bi-network Case: We evaluated competitive methods on dense and sparse pairs with different $\eta = [5\%, 10\%, ..., 30\%]$. 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: (i) In MASTER, both observations of attribute and structure space are comprehensively exploited. (ii) The embeddings, capturing the intrinsic relation between users, facilitate the robust reconciliation. MASTER performs better than MASTER- consistently, demonstrating the necessity of comprehensiveness.

Model	Multiplicity	Attributes	Structure	Robustness
MNA [Kong *et al.*, 2013]		√	√	
ULink [Mu *et al.*, 2016]	√	√		
COSNET [Zhang *et al.*, 2015]	√	√	√	
IONE [Liu *et al.*, 2016]			√	√
PALE [Man *et al.*, 2016]			√	√
MOBIUS [Zafarani and Liu, 2013]		√		
HYDRA [Liu *et al.*, 2014]		√	√	
User-Matching [Korula and Lattanzi, 2014]			√	
MASTER (our model)	√	√	√	√

Table 1: A summary of most of typical reconciliation models

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 $\eta = [5\%, 10\%, ..., 30\%]$, 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. Although 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 our solution, we conducted experiments the sparse pairs and sparse groups, reported in Fig. 2(c) and 2(d) respectively. The performances of COSNET become noticeably worser in sparse pairs (groups) as less information can be leveraged. The robustness of ULink is not evaluated, since the difference among them is in the structure space neglected by ULink. However, the embedding based approaches, *i.e.*, MASTER, MASTER- and PALE, lose much less hit-precision in sparse pairs (groups) and MASTER achieves the highest robustness.

The impacts of parameters: To further evaluate MASTER, we conducted experiments to evaluate the effect of training ratio θ and the contribution of each space. Regarding the impact of θ, we fix $\eta = 30\%$, $\beta/\alpha = 2/3$, and vary the value of θ to $[0.5\%, 1\%, 1.5\%, ..., 5\%]$. We report the corresponding result 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 θ 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 $[-1, -0.8, -0.6, ..., 1]$. We report the corresponding result 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 provide truthful personal information (*e.g.*, location, birthday), which results in lots of noise in the attribute space.

6 Related Work

The MASTER reconciles multiple social networks in an embedding approach. We briefly summarize the related work in the problem of reconciliation and network embedding:
Reconciliation: The problem is generally regarded as a (semi-) supervised task as the labels can be observed [Shu *et al.*, 2017]. Most of the existing models focus on reconciling only two social networks. A few models reconcile multiple networks, however, ULink [Mu *et al.*, 2016] suffers from the noise in attribute space while COSNET [Zhang *et al.*, 2015] and UniRank [Zhang *et al.*, 2017a] struggle in defining local consistency. Most of the typical models are summarized in Table 1. The difference between our model and the others lies in that we, for the first time, address all these limitations.
Network embedding: Network embedding can be addressed through several techniques: (1) matrix factorization, *e.g.*, TADW [Yang *et al.*, 2015], HOPE [Ou *et al.*, 2016], LANE [Huang *et al.*, 2017], M-NMF [Wang *et al.*, 2017b]; (2) deep neural network, *e.g.*, SDAE [Cao *et al.*, 2016], SiNE [Wang *et al.*, 2017a]; (3) random walk, *e.g.*, DeepWalk [Perozzi *et al.*, 2014], Node2Vec [Grover and Leskovec, 2016]. Moreover, both DeepWalk and LINE [Tang *et al.*, 2015] are proved to be equivalent to matrix factorization recently [Qiu *et al.*, 2018]. However, different from all these models, our model is tailored for reconciling multiple social networks.

7 Conclusion

To address the problem of robustly reconciling multiple social networks, we, for the first time, propose the MASTER framework. In MASTER, we design the CDE model to formulate the reconciliation problem into a unified optimization where we embed and reconcile multiple social networks in the joint latent space constructed by uni- and joint-embedding. We design an effective NS-Alternating algorithm to solve the nonconvex optimization of CDE and further prove the KKT convergence of the algorithm. We conducted extensive experiments on real-world datasets and demonstrate that MASTER outperforms several state-of-the-art methods.

Acknowledgments

This work was supported in part by the following funding agencies of China: National Key Research and Development Program under Grant 2016QY01W0200 and National Natural Science Foundation under Grant 61602050 and U1534201.

References

[Cao *et al.*, 2016] Shaosheng Cao, Wei Lu, and Qiongkai Xu. Deep neural networks for learning graph representations. In *Proceedings of the AAAI*, pages 1145–1152, 2016.

[Grover and Leskovec, 2016] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the SIGKDD*, pages 855–864, 2016.

[Hong *et al.*, 2016] Mingyi Hong, Zhi-Quan Luo, and Meisam 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.

[Huang *et al.*, 2017] Xiao Huang, Jundong Li, and Xia Hu. Label informed attributed network embedding. In *Proceedings of the WSDM*, pages 731–739, 2017.

[Kong *et al.*, 2013] Xiangnan Kong, Jiawei Zhang, and Philip S Yu. Inferring anchor links across multiple heterogeneous social networks. In *Proceedings of the CIKM*, pages 179–188, 2013.

[Korula and Lattanzi, 2014] Nitish Korula and Silvio Lattanzi. An efficient reconciliation algorithm for social networks. *Proceedings of the VLDB*, 7(5):377–388, 2014.

[Liu *et al.*, 2014] Siyuan Liu, Shuhui Wang, Feida Zhu, Jinbo Zhang, and Ramayya Krishnan. Hydra: Large-scale social identity linkage via heterogeneous behavior modeling. In *Proceedings of the SIGMOD*, pages 51–62, 2014.

[Liu *et al.*, 2016] Li Liu, William K Cheung, Xin Li, and Lejian Liao. Aligning users across social networks using network embedding. In *Proceedings of the IJCAI*, pages 1774–1780, 2016.

[Lu *et al.*, 2017] Songtao Lu, Mingyi Hong, and Zhengdao Wang. A nonconvex splitting method for symmetric nonnegative matrix factorization: Convergence analysis and optimality. *IEEE Trans. on Signal Processing*, 2017.

[Man *et al.*, 2016] Tong Man, Huawei Shen, Shenghua Liu, Xiaolong Jin, and Xueqi Cheng. Predict anchor links across social networks via an embedding approach. In *IJCAI*, pages 1823–1829, 2016.

[Man *et al.*, 2017] Tong Man, Huawei Shen, Xiaolong Jin, and Xueqi Cheng. Cross-domain recommendation: An embedding and mapping approach. In *Proceedings of the IJCAI*, pages 2464–2470, 2017.

[Mu *et al.*, 2016] Xin Mu, Feida Zhu, Ee-Peng Lim, Jing Xiao, Jianzong Wang, and Zhi-Hua Zhou. User identity linkage by latent user space modelling. In *Proceedings of the SIGKDD*, pages 1775–1784, 2016.

[Ou *et al.*, 2016] Mingdong Ou, Peng Cui, Jian Pei, Ziwei Zhang, and Wenwu Zhu. Asymmetric transitivity preserving graph embedding. In *Proceedings of the SIGKDD*, pages 1105–1114, 2016.

[Perozzi *et al.*, 2014] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the SIGKDD*, pages 701–710, 2014.

[Qiu *et al.*, 2018] Jiezhong Qiu, Yuxiao Dong, Hao Ma, Jian Li, Kuansan Wang, and Jie Tang. Network embedding as matrix factorization: Unifying deepwalk, line, pte, and node2vec. In *Proceedings of the WSDM*, 2018.

[Shu *et al.*, 2017] Kai Shu, Suhang Wang, Jiliang Tang, Reza Zafarani, and Huan Liu. User identity linkage across online social networks: A review. *ACM SIGKDD Explorations Newsletter*, 18(2):5–17, 2017.

[Tang *et al.*, 2015] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. Line: Large-scale information network embedding. In *Proceedings of the WWW Conference*, pages 1067–1077, 2015.

[Toh, 2008] Kim-Chuan Toh. An inexact primal–dual path following algorithm for convex quadratic sdp. *Mathematical programming*, 112(1):221–254, 2008.

[Vandaele *et al.*, 2016] Arnaud Vandaele, Nicolas Gillis, Qi Lei, Kai Zhong, and Inderjit Dhillon. Efficient and non-convex coordinate descent for symmetric nonnegative matrix factorization. *IEEE Trans. on Signal Processing*, 64(21):5571–5584, 2016.

[Wang *et al.*, 2017a] Suhang Wang, Jiliang Tang, Charu Aggarwal, Yi Chang, and Huan Liu. Signed network embedding in social media. In *Proceedings of the SDM*, pages 327–335, 2017.

[Wang *et al.*, 2017b] Xiao Wang, Peng Cui, Jing Wang, Jian Pei, Wenwu Zhu, and Shiqiang Yang. Community preserving network embedding. In *Proceedings of the AAAI*, pages 203–209, 2017.

[Yang *et al.*, 2015] Cheng Yang, Zhiyuan Liu, Deli Zhao, Maosong Sun, and Edward Y Chang. Network representation learning with rich text information. In *Proceedings of the IJCAI*, pages 2111–2117, 2015.

[Zafarani and Liu, 2013] Reza Zafarani and Huan Liu. Connecting users across social media sites: a behavioral-modeling approach. In *Proceedings of the SIGKDD*, pages 41–49, 2013.

[Zhang and Yu, 2016] Jiawei Zhang and Philip S Yu. Pct: partial co-alignment of social networks. In *Proceedings of the WWW Conference*, pages 749–759, 2016.

[Zhang *et al.*, 2015] Yutao Zhang, Jie Tang, Zhilin Yang, Jian Pei, and Philip S Yu. Cosnet: Connecting heterogeneous social networks with local and global consistency. In *Proceedings of the SIGKDD*, pages 1485–1494, 2015.

[Zhang *et al.*, 2017a] Zhongbao Zhang, Qihang Gu, Tong Yue, and Sen Su. Identifying the same person across two similar social networks in a unified way: Globally and locally. *Information Sciences*, 395:53–67, 2017.

[Zhang *et al.*, 2017b] Zhongbao Zhang, Jian Wen, Li Sun, Qiaoyu Deng, Sen Su, and Pengyan Yao. Efficient incremental dynamic link prediction algorithms in social network. *Knowledge-Based Systems*, 132:226–235, 2017.