Link Prediction across Aligned Networks with Sparse and Low Rank Matrix Estimation

Jiawei Zhang[⋆], Jianhui Chen[†], Shi Zhi[‡], Yi Chang[♯], Philip S. Yu[⋆,¶], Jiawei Han[‡]
[⋆]University of Illinois at Chicago, Chicago, IL, USA
[†]Yahoo! Research, Sunnyvale, CA, USA
[‡]University of Illinois at Urbana-Champaign, IL, USA
[¶]Institute for Data Science, Tsinghua University, Beijing, China
[♯]Search Technology Lab, Huawei Research America, CA, USA
{jzhan9, psyu}@uic.edu, jianhui@yahoo-inc.com, yi.chang@huawei.com, {shizhi2, hanj}@illinois.edu

Abstract*—Users' addiction to online social networks is discovered to be highly correlated with their social connections in the networks. Dense social connections can effectively help online social networks retain their active users and improve the social network services. Therefore, it is of great importance to make a good prediction of the social links among users. Meanwhile, to enjoy more social network services, users nowadays are usually involved in multiple online social networks simultaneously. Formally, the social networks which share a number of common users are defined as the "aligned networks". With the information transferred from multiple aligned social networks, we can gain a more comprehensive knowledge about the social preferences of users in the pre-specified target network, which will benefit the social link prediction task greatly. However, when transferring the knowledge from other aligned source networks to the target network, there usually exists a shift in information distribution between different networks, namely domain difference. In this paper, we study the social link prediction problem of the target network, which is aligned with multiple social networks concurrently. To accommodate the domain difference issue, we project the features extracted for links from different aligned networks into a shared lower-dimensional feature space. Moreover, users in social networks usually tend to form communities and would only connect to a small number of users. Thus, the target network structure has both the low-rank and sparse properties. We propose a novel optimization framework, SLAMPRED, to combine both these two properties aforementioned of the target network and the information of multiple aligned networks with nice domain adaptations. Since the objective function is a linear combination of convex and concave functions involving non-differentiable regularizers, we propose a novel optimization method to iteratively solve it. Extensive experiments have been done on real-world aligned social networks, and the experimental results demonstrate the effectiveness of the proposed model.

I. INTRODUCTION

Online social networks (OSNs) have achieved a tremendous success in recent years, and a large number of online social networks have appeared and provided services for the public. In all kinds of OSNs, the social connections among users play an important role in steering their online social activities and network service usages. As proposed in [10], users' social network addiction and usage frequency are highly correlated to their friendship in the networks, and users who have more friends in the OSNs generally tend to use social network services more frequently. Thus, inferring the potential social connections among users is very crucial for the development of OSNs.

Formally, given a screen-shot of an OSN, the social link prediction problem aims at inferring the potential social connections that will be formed among users in the near future [11]. The link prediction problem has been studied for many years, and dozens of papers have been published so far. Depending on the supervision information used to train the models, existing link prediction methods can be generally divided into three main categories: (1) traditional unsupervised link predictors [11], [31], which infer the potential social connections among users merely by calculating their closeness scores; (2) supervised link prediction methods [5], [29], [28], [30], [27], which treat the existing and non-existing links as positive and negative instances respectively and build classification/regression models to infer the labels/scores of the potential links; and (3) semi-supervised link prediction models [37], [33], in which the existing and non-existing links are treated as positive and unlabeled instances respectively, and the link prediction problem is further transformed into a PU (Positive and Unlabeled instance) learning problem.

In recent years, the social network studies have developed into a new dimension. To enjoy different kinds of social network services, users nowadays are usually involved in multiple OSNs at the same time [28], and these shared users can act like anchors aligning these different OSNs together. According to the existing works [37], the common users shared by different networks are named as the *anchor users*, while the OSNs aligned by the anchor users are formally defined as the *aligned social networks*. The modeling of the multiple *aligned social networks* provides the opportunity for researchers to study *users' social behaviors* from a global perspective. With the abundant information about the users from multiple sites, we can gain a more comprehensive understanding about users' social activity patterns, which can greatly help the link prediction task of social networks.

Problem Studied: In this paper, we will study the link prediction problem for the target network, which is aligned with multiple source networks concurrently. Formally, the problem is named as the "Social Link Transfer" (SLT) problem. An

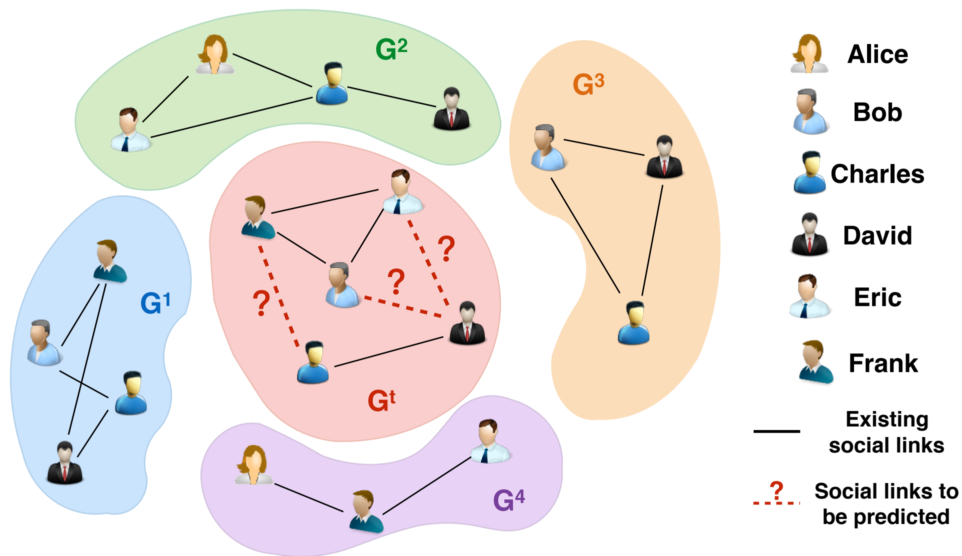


Fig. 1. An example of the SLT problem (Regions of different background colors: different social networks, Network G^t: the target network, Solid black line: existing social connections, Red dashed lines: links to be predicted).

example to illustrate the SLT problem is provided in Figure 1. In Figure 1, network G^t is the target network, and $G^1, \cdots, G^4$ are the other aligned source networks, which share a number of common users with G^t. With the information across networks $\{G^t, G^1, \cdots, G^4\}$, the objective of SLT is to infer potential social links (i.e., the red dashed lines) to be formed in the target network G^t.

The SLT problem studied in this paper is based on the same setting as those in [28], [37], but we propose a new model to address the problem. We summarize the differences of this work from these existing works as follows. Firstly, the link prediction model proposed in this paper is based on the matrix estimation, which is totally different from the classification based models proposed in [28], [37] and will not suffer from the class imbalance problem. Secondly, considering the connections among users in the networks are usually very sparse and users tend to form densely connected local communities, a sparse regularizer and a low-rank regularizer are incorporated in the objective function. Thirdly, these existing works [28], [37] transfer information across different networks without considering the domain differences. Meanwhile, based on the known anchor and social link information, our model overcomes the domain difference problem by mapping the feature vectors extracted for links from the aligned networks to a shared lower-dimensional latent feature space instead.

The SLT studied in this paper is very hard to solve due to the following challenges:

- *Heterogeneity of Networks*: The networks studied in this paper are all heterogeneous, involving both structure and various types of attribute information. Incorporating all these different categories of information in effective link prediction can be a great challenge.
- *Multiple Aligned Networks*: Inferring links in the target with information from multiple aligned networks simultaneously would suffer from the domain difference greatly, which makes the problem more challenging.
- *Sparse and Low-Rank Property*: The social networks are usually of a sparse and low-rank structure [38], [15], where the formed social links usually only occupy a small proportion of all potential social links and users tend to densely connect to local communities. Preserving and utilizing such properties in link prediction render the problem tougher.
- *Objective Function*: The objective function is a linear combination of convex and concave functions involving non-differentiable regularizers. How to effectively resolve the objective function is technically challenging .

To overcome these challenges, a novel link prediction model named SLAMPRED (Sparse Low-rAnk Matrix estimation based Prediction) is proposed in this paper. SLAMPRED formulates the link prediction problem as a sparse and low-rank matrix estimation problem. Heterogeneous information is used to calculate the similarity among users, and similar users tend be linked. With the existing anchor and social link information, SLAMPRED proposes to map the feature vectors of the social links extracted from the target and other aligned source networks to a common low-dimensional latent feature space. Two regularizers are introduced in the objective function of SLAMPRED to preserve the sparse and low-rank properties. Furthermore, SLAMPRED solves the objective function with the iterative CCCP (convex concave procedure), and in each iteration the involved non-differentiable sparsity and low-rank regularizers are effectively handled by the proximal operators.

The remaining part of this paper is organized as follows. In Section II, we will introduce the definitions of several important concepts and the formulation of the SLT problem. More detailed information about the link prediction model, SLAMPRED, will be introduced in Section III. Extensive experiments will be done to evaluate the performance of SLAMPRED in Section IV. Finally, in Section V, we will talk about the related works and conclude this paper in Section VI.

II. Problem Formulation

In this section, we will introduce the definitions of some important concepts used in this paper, and then provide the formulation of the SLT problem.

A. Terminology Definition

The networks studied in this paper all contain heterogeneous information involving various types of nodes and complex links among the modes, which can be represented as the *heterogeneous information networks* formally.

Definition 1 (Heterogeneous Information Network): A *heterogeneous information network* can be defined as graph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ and $\mathcal{E}$ denotes the union of different types of nodes and edges respectively.

More specifically, for the networks studied in this paper, the node set $\mathcal{V}$ can be represented as $\mathcal{V} = \mathcal{U} \cup \mathcal{P} \cup \mathcal{W} \cup \mathcal{T} \cup \mathcal{L}$ involving the nodes about the users, posts, words, timestamps, and location checkins. Meanwhile, the edge set $\mathcal{E}$ can be represented as $\mathcal{E} = \mathcal{E}_u \cup \mathcal{E}_p \cup \mathcal{E}_w \cup \mathcal{E}_t \cup \mathcal{E}_l$, which contains the links among users, between users and posts, as well as those between posts and words, timestamps and locations respectively. Furthermore, as mentioned in Section I, the networks studied in this paper may share some common users, which can be represented as the *multiple aligned networks* formally.

Definition 2 (Multiple Aligned Networks): To differentiate the target network from the other external aligned source networks, we indicate the target network specifically and define the multiple aligned networks as $\mathcal{G} = (\{G^t, G^1, G^2, \cdots, G^K\}, \{\mathcal{A}^{t,1}, \mathcal{A}^{t,2}, \cdots, \mathcal{A}^{K-1,K}\})$, where G^t is the target network, $G^1, G^2, \cdots, G^K$ are the K aligned source networks, and $\mathcal{A}^{t,1}, \mathcal{A}^{t,2}, \cdots, \mathcal{A}^{K-1,K}$ denote the sets of undirected *anchor links* among these networks respectively.

In this paper, we will follow the definitions about the *anchor users* and *anchor links* concepts proposed in [37], which are not introduced here due to the limited space.

B. Problem Definition

Definition 3 (The SLT Problem): Given the *multiple aligned online social networks* $\mathcal{G} = (\{G^t, G^1, G^2, \cdots, G^K\}, \{\mathcal{A}^{t,1}, \mathcal{A}^{t,2}, \cdots, \mathcal{A}^{K-1,K}\})$, the SLT problem studied in this paper aims at inferring the potential social connections among users in the target network G^t with information across all these networks. Formally, based on information available in $\mathcal{G}$, the objective of SLT is to build a social link prediction function $S : \mathcal{U}^t \times \mathcal{U}^t \setminus \mathcal{E}_u^t \to [0, 1]$ to infer the confidence scores of all the potential social connections among the users in the target network G^t, where $\mathcal{U}^t$ and $\mathcal{E}_u^t$ represent the existing users and social links in G^t respectively.

III. Proposed Method

In this section, we will first introduce the link prediction model built with the observed network connection information and other heterogeneous attribute information available in the target network. After that, we will talk about the target network link prediction problem with information across multiple aligned networks, where the features extracted from different networks are projected to a lower-dimensional feature space to accommodate the domain differences. Finally, we will introduce the joint optimization objective function, which can be resolved by the proximal operator based iterative CCCP algorithm effectively.

A. Notation

At the beginning of this section, we will first set some notations used in this paper. Throughout this paper, we will use lower case letters (e.g., x) to denote scalars, lower case bold letters (e.g., $\mathbf{x}$) to denote column vectors, upper case letters (e.g., X) to denote elements of matrices, upper case calligraphic letters (e.g., $\mathcal{X}$) to denote sets, and bold-face upper case letters (e.g., $\mathbf{X}$) to denote matrices and high-order tensors. In the sequel, the projection of a matrix $\mathbf{X}$ onto domain $\mathcal{S}$ is denoted by $P_{\mathcal{S}}(\mathbf{X})$. The matrix $(\mathbf{X})_+$ denotes the component-wise positive part of the matrix $\mathbf{X}$, and sgn$(\mathbf{X})$ is the sign matrix associated to $\mathbf{X}$ with the convention sgn$(0) = 0$. For a matrix $\mathbf{X}$, we denote $\mathbf{X}(i, :)$ (and $\mathbf{X}(:, j)$) as the i_{th} row (and the j_{th} column) of $\mathbf{X}$; while for a 3-way tensor $\mathbf{X}$, we denote $\mathbf{X}(k, :, :)$ as the k_{th} 2-way tensor slice (i.e., matrix) along the 1_{st} dimension and so forth. We use superscript $^\top$ to denote the transpose of a vector or a matrix. The notations $\|\cdot\|$, $\|\cdot\|_F$, $\|\cdot\|_p$ and $\|\cdot\|_*$ define the Euclidean norm, Frobenius norm, entry-wise l_p norm and trace norm respectively. More specifically, given a matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$, the entry-wise l_p norm of $\mathbf{X}$ can be represented as $\|\mathbf{X}\|_p = (\sum_{i,j} |X_{i,j}|^p)^{1/p}$; while the trace norm of $\mathbf{X}$ can be represented as $\|\mathbf{X}\|_* = \sum_{i=1}^{\text{rank}(\mathbf{X})} \sigma_i$, where rank$(\mathbf{X})$ denotes the rank and σ_i is the i_{th} singular value of matrix $\mathbf{X}$. In addition, we use Tr$(\mathbf{X})$ to denote the *trace* of matrix $\mathbf{X}$.

B. Intra-Network Link Prediction

Users' diverse online social activities may generate heterogeneous information in the online social networks, which include both the network structure information as well as different categories of attribute information about the users. In this subsection, we will introduce the link prediction method with the heterogeneous information available in the target network.

1) Intra-Network Link Prediction with Link Information: Give the target network G^t involving users $\mathcal{U}^t$, we can represent the observed social connection among the users with the binary social adjacency matrix $\mathbf{A}^t \in \{0, 1\}^{|\mathcal{U}^t| \times |\mathcal{U}^t|}$, where entry $A^t(i, j) = 1$ iff the corresponding social link (u_i^t, u_j^t) exists between users u_i^t and u_j^t in G^t. In the SLT problem, our objective is to infer the potential unobserved social links for the target network, which can be achieved by finding a sparse and low-rank predictor matrix $\mathbf{S} \in \mathcal{S}$ from some convex admissible set $\mathcal{S} \subset \mathbb{R}^{|\mathcal{U}^t| \times |\mathcal{U}^t|}$. Meanwhile, the inconsistency between the inferred matrix $\mathbf{S}$ and the observed social adjacency matrix $\mathbf{A}^t$ can be represented as the loss function $l(\mathbf{S}, \mathbf{A}^t)$. The optimal social link predictor for the target network can be achieved by minimizing the loss term, i.e.,

$$\arg\min_{\mathbf{S} \in \mathcal{S}} l(\mathbf{S}, \mathbf{A}^t).$$

The loss function $l(\mathbf{S}, \mathbf{A}^t)$ can be defined in many different ways, and, in this paper, we propose to approximate the *loss function* by counting the loss introduced by the existing social links in $\mathcal{E}_u^t$, i.e.,

$$l(\mathbf{S}, \mathbf{A}^t) = \frac{1}{|\mathcal{E}_u^t|} \sum_{(u_i^t, u_j^t) \in \mathcal{E}_u^t} \mathbb{1}\Big(\big(A^t(i,j) - \frac{1}{2}\big) \cdot S(i,j) \leq 0\Big).$$

2) Intra-Network Link Prediction with Heterogeneous Attribute Information: Besides the connection information, there also exists a large amount of attribute information available in the target network, e.g., *location checkin records*, *online social activity temporal patterns*, and *text usage patterns*, etc. Based on the attribute information, a set of features can be extracted for all the potential user pairs to denote their closeness, which are called the *intimacy features* formally. For instance, given user pair (u_i^t, u_j^t) in the target network, we can represent its *intimacy features* as vector $\mathbf{x}_{i,j}^t \in \mathbb{R}^{d^t}$ (d^t denotes the extracted intimacy feature number). According to the existing works [5], [28], different intimacy features can be extracted from the attribute information, and we will briefly introduce the extracted features in Section IV later.

More generally, we can represent the feature vectors extracted for user pairs as a 3-way tensor $\mathbf{X}^t \in \mathbb{R}^{d^t \times |\mathcal{U}^t| \times |\mathcal{U}^t|}$, where slice $\mathbf{X}^t(k, :, :)$ denote all the k_{th} intimacy features among all the user pairs. In online social networks, *homophily* principle [12] has been observed to widely structure the users' online social connections, and users who are close to each other are more likely to be friends. Based on such an intuition, we can infer the potential social connection matrix $\mathbf{S}$ by maximizing the overall intimacy scores of the inferred new social connections, i.e.,

$$\arg\max_{\mathbf{S} \in \mathcal{S}} int(\mathbf{S}, \mathbf{X}^t).$$

In this paper, we propose to define the intimacy score term $int(\mathbf{S}, \mathbf{X}^t)$ by enumerating and summing the *intimacy scores* of the inferred social connections, i.e.,

$$int(\mathbf{S}, \mathbf{X}^t) = \sum_{k=1}^{d^t} \left\| \mathbf{S} \circ \mathbf{X}^t(k, :, :) \right\|_1,$$

where operator $\circ$ denotes the Hadamard product (i.e., entry-wise product) of matrices.

3) Joint Optimization Function for Intra-Network Link Prediction: By considering the link and attribute information in the target network at the same time, we can represent the joint optimization for link prediction in the target network to be

$$\arg\min_{\mathbf{S} \in \mathcal{S}} l(\mathbf{S}, \mathbf{A}^t) - \alpha^t \cdot int(\mathbf{S}, \mathbf{X}^t) + \gamma \cdot \|\mathbf{S}\|_1 + \tau \cdot \|\mathbf{S}\|_*.$$

Considering that the social connections in online social networks are usually very sparse and of low-rank, the regularizers $\|\mathbf{S}\|_1$ and $\|\mathbf{S}\|_*$ are added to preserve the *sparse* and *low rank* properties of the inferred predictor matrix $\mathbf{S}$. Parameters α^t, γ, τ denote the importance scalars of different terms in the objective function.

C. Inter-Network Link Prediction

Besides the information available in the target network, a large amount of information about the users' social activities is available in other external source networks as well, which can be transferred to the target network to help improve the link prediction results, especially when the target network suffers from information sparsity problem. To be general, we can represent the *intimacy* features extracted for user pairs in source network G^i ($i \in \{1, 2, \cdots, K\}$) as a 3-way tensor $\mathbf{X}^i \in \mathbb{R}^{d^i \times |\mathcal{U}^i| \times |\mathcal{U}^i|}$, where $\mathcal{U}^i$ denotes the user set in G^i and d^i is the extracted feature number.

Meanwhile, different online social networks are constructed for different purposes, information from which may follow totally different distributions actually. To adopt the information domains of these different aligned networks, in this paper, we propose to project the extracted feature vectors from different networks (both G^t and aligned source networks $G^1, \cdots, G^K$) to a common lower-dimensional feature space instead. Given the $K+1$ partially aligned social networks, we formulate the information domain adaption problem as a mapping function inference problem instead. Our objective is to construct $K+1$ mapping functions, $f^t : \mathbb{R}^{d^t} \rightarrow \mathbb{R}^c, \cdots, f^K : \mathbb{R}^{d^K} \rightarrow \mathbb{R}^c$ to map the $K+1$ input features to a new c-dimensional latent space, where certain properties about the networks are still preserved.

In this paper, we propose achieve the objective by utilizing the existing anchor links and social links across the networks. As shown in Figure 2, the links in different social networks are first categorized into different sets: (1) social links aligned by anchor links (i.e., the *aligned social links* to be introduced later), (2) *similar social links* (i.e., connected user pairs or unconnected user pairs), and (3) *dissimilar social links* (i.e., the connected user pairs vs. the unconnected ones). Based on the categorization information about the links, in the link embedding process, we aim at placing aligned social links and similar social links closely in the common latent feature space, while placing the dissimilar ones far away from each other in the feature space. More information about these concepts and the embedding process will be introduced in the following parts in great details.

1) Anchor Link based Feature Space Projection: Before introducing the anchor link based feature space projection method, we first introduce the concept of *aligned social link* in this paper.

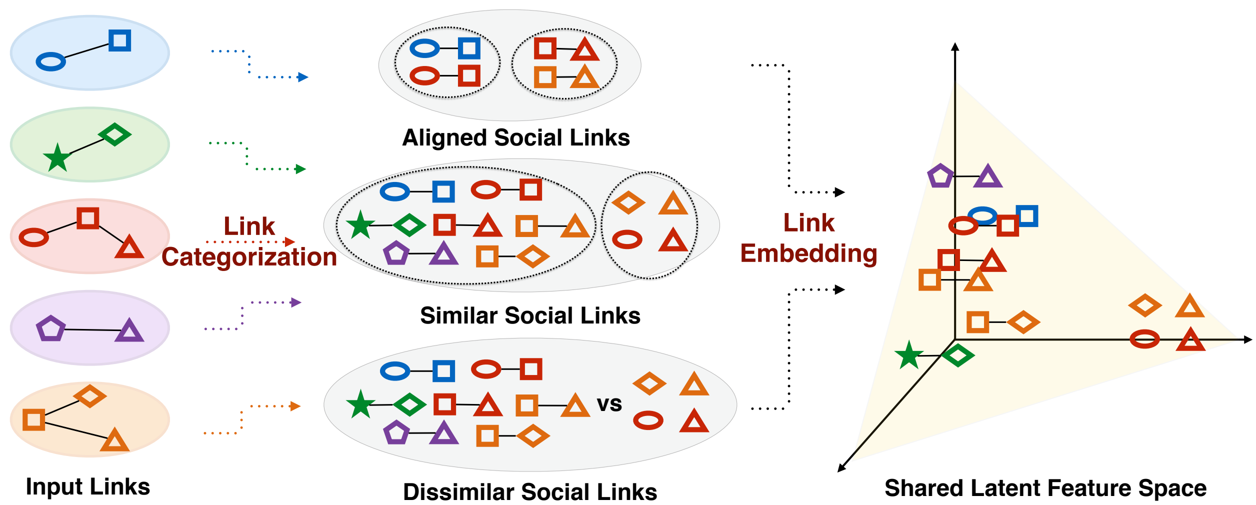


Fig. 2. An example of method SLAMPRED on social link embedding.

Definition 4 (Aligned Social Link): Given two social links (u_i^t, u_j^t) and (u_m^k, u_n^k) in networks G^t and G^k respectively, if u_i^t, u_m^k and u_j^t, u_n^k are both aligned by the anchor links (i.e., $(u_i^t, u_m^k) \in \mathcal{A}^{t,k}$ and $(u_j^t, u_n^k) \in \mathcal{A}^{t,k}$), then (u_i^t, u_j^t) and (u_m^k, u_n^k) are called the *aligned social links*.

Let sets $\mathcal{L}^t$ and $\mathcal{L}^k$ denote all the potential social links in networks G^t and G^k respectively, where $\mathcal{L}^t = \mathcal{U}^t \times \mathcal{U}^t \setminus \{(u,u)\}_{u \in \mathcal{U}^t}$ and $\mathcal{L}^k = \mathcal{U}^k \times \mathcal{U}^k \setminus \{(u,u)\}_{u \in \mathcal{U}^k}$. Based on the anchor links between networks G^t and G^k (i.e., $\mathcal{A}^{t,k}$), we can denote all the aligned social links with the *aligned social link indicator matrix* $\mathbf{W}_A^{t,k} \in \{0,1\}^{|\mathcal{L}^t| \times |\mathcal{L}^k|}$, where entry $W_A^{t,k}(i,j) = 1$ iff the corresponding social links $l_i^t \in \mathcal{L}^t$ and $l_j^k \in \mathcal{L}^k$ are *aligned social links*.

Generally, the *aligned social links* are actually connecting the accounts of the same users, and the feature vectors extracted for them from different networks should be mapped to close areas in the lower-dimensional latent feature space. Based on such an intuition, we can define the inconsistency introduced in projecting the features for aligned social links between networks G^t and other external source networks as term $Cost_A$:

$$Cost_A = \mu \sum_{m=t}^{K} \sum_{n=t}^{K} \sum_{i=1}^{|\mathcal{L}^m|} \sum_{j=1}^{|\mathcal{L}^n|} \left\| f^m(\mathbf{x}^m_{l_i^m}) - f^n(\mathbf{x}^n_{l_j^n}) \right\|^2 W_A^{m,n}(i,j),$$

where notation $\sum_{m=t}^{K}$ denotes the enumeration of all the networks in $\{G^t, G^1, \cdots, G^K\}$, and μ is the scalar.

Minimizing of the cost term will encourage the features extracted for social links corresponding to the aligned social links being mapped to similar locations in the latent feature space. Furthermore, for all the pairwise networks, we can group all the *aligned social link indicator matrices* together as the big *joint aligned social link indicator matrix* $\mathbf{W}_A \in \{0,1\}^{|\mathcal{L}| \times |\mathcal{L}|}$, where $\mathcal{L} = \mathcal{L}^t \cup \mathcal{L}^1 \cup \cdots \cup \mathcal{L}^K$. Formally, matrix $\mathbf{W}_A$ can be represented as

$$\mathbf{W}_A = \begin{bmatrix} \mathbf{W}_A^{t,t} & \mathbf{W}_A^{t,1} & \cdots & \mathbf{W}_A^{t,K} \\ \mathbf{W}_A^{1,t} & \mathbf{W}_A^{1,1} & \cdots & \mathbf{W}_A^{1,K} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{W}_A^{K,t} & \mathbf{W}_A^{K,1} & \cdots & \mathbf{W}_A^{K,K} \end{bmatrix}.$$

In addition, we can represent its Laplacian matrix as $\mathbf{L}_A = \mathbf{D}_A - \mathbf{W}_A$, where matrix $\mathbf{D}_A$ denotes the diagonal row sum matrix of $\mathbf{W}_A$ with entries $D_A(i,i) = \sum_j W_A(i,j)$ on the diagonal. Matrix $\mathbf{L}_A$ will be used in the projection function inference to be introduced in the following parts.

2) Existing Social Link based Feature Space Projection: Besides the anchor link information, we also propose to utilize the existing social connections among the users to help infer the *feature mapping functions*. Before introducing the detailed method, we propose to define the concept of *link existence label* $y(\cdot)$ first as follows:

Definition 5 (Link Existence Label): Given a link $l_i^k \in \mathcal{L}^k$ in network $G^k, k \in \{t, 1, 2, \cdots, K\}$, if link l_i^k exists in the network then its corresponding *link existence label* $y(l_i^k) = 1$, otherwise $y(l_i^k) = 0$.

Since our ultimate goal is to infer the potential feature vector mappings to the latent feature space to transfer information for the link prediction tasks, the social link existence information will plan a very important role in identifying the potential feature space mappings. Based on the known social connections in a pair of aligned networks G^t and G^k ($k \in \{1, 2, \cdots, K\}$), we can construct the *similar link existence label indicator matrix* $\mathbf{W}_S^{t,k} \in \{0,1\}^{|\mathcal{L}^t| \times |\mathcal{L}^k|}$ and *dissimilar link existence label indicator matrix* $\mathbf{W}_D^{t,k} \in \{0,1\}^{|\mathcal{L}^t| \times |\mathcal{L}^k|}$ between networks G^t and G^K. For any link instances $l_i^t \in \mathcal{L}^t$ and $l_j^k \in \mathcal{L}^k$, if l_i^t and l_j^s share the same *link existence label*, we will assign the corresponding entry in $\mathbf{W}_S^{t,k}$ with value 1 (and the corresponding entry in $\mathbf{W}_D^{t,k}$ with value 0); otherwise, we will assign the corresponding entry in $\mathbf{W}_S^{t,k}$ with value 0 (and the corresponding entry in $\mathbf{W}_D^{t,k}$ with value 1). Therefore, matrices $\mathbf{W}_S^{t,k}$ and $\mathbf{W}_D^{t,k}$ store all the link existence information in the networks G^t and G^k.

As pointed out in [21], the instances which share common labels tend to be projected together in the latent feature space, while those having different labels will be projected to be apart from each other instead. Based on such an intuition, we propose to define terms $Cost_S$ and $Cost_D$ to denote the mapping costs introduced by the *link existence label* information (for the links having *similar* and *different* labels) respectively:

$$Cost_S = \sum_{m=t}^{K} \sum_{n=t}^{K} \sum_{i=1}^{|\mathcal{L}^m|} \sum_{j=1}^{|\mathcal{L}^n|} \left\| f^m(\mathbf{x}^m_{l^m_i}) - f^n(\mathbf{x}^n_{l^n_j}) \right\|^2 W^{m,n}_S(i,j),$$

$$Cost_D = \sum_{m=t}^{K} \sum_{n=t}^{K} \sum_{i=1}^{|\mathcal{L}^m|} \sum_{j=1}^{|\mathcal{L}^n|} \left\| f^m(\mathbf{x}^m_{l^m_i}) - f^n(\mathbf{x}^n_{l^n_j}) \right\|^2 W^{m,n}_D(i,j).$$

If link instances l^t_i and l^k_j in networks G^t and G^k share the same *link existence label* (i.e., $W^{t,k}_S(i,j) = 1$), but their embeddings are far away from each other, then $Cost_S$ will be larger. Meanwhile, if link instances l^t_i and l^k_j have different *link existence labels* (i.e., $W^{t,k}_D(i,j) = 1$), and their embeddings are close to each other, the introduced $Cost_D$ will be small. Therefore, minimizing $Cost_S$ and maximizing $Cost_D$ simultaneously will encourage the link instances of the same label to be projected to similar areas, while those of different labels to be projected separately instead.

What's more, in a similar way, we can also group all the network pairwise *similar link existence label indicator matrices* and *dissimilar link existence label indicator matrices* together in the same order as matrix $\mathbf{W}_A$, which can be represented as $\mathbf{W}_S$ and $\mathbf{W}_D$. Their corresponding Laplacian matrices can be denoted as $\mathbf{L}_S$ and $\mathbf{L}_D$ respectively.

3) Joint Mapping Function Inference: We may want to ensure the mapping functions can achieve the above three objectives at the same time, which can be achieved by minimizing the overall cost function

$$\min Cost(f^t, f^1, f^2, \cdots, f^K) = \frac{Cost_A + Cost_S}{Cost_D}.$$

The projection mappings can be of different forms, and we will take the linear mapping as an example in this paper. In other words, the mappings f^t, f^1, f^2, $\cdots$, f^K can be represented as $K+1$ matrices $\mathbf{F}^t \in \mathbb{R}^{d^t \times c}$, $\mathbf{F}^1 \in \mathbb{R}^{d^1 \times c}$, $\cdots$, $\mathbf{F}^K \in \mathbb{R}^{d^K \times c}$ respectively, where $d^t, d^1, \cdots, d^K$ denote the length of features from networks $G^t, G^1, \cdots, G^K$ and c is the dimension of the projected feature space.

Formally, given all the feature vectors extracted for potential user pairs in the networks G^t, G^1, $\cdots$, G^K, we can group them together and represent it as matrix

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}^t & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}^1 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{Z}^K \end{bmatrix},$$

where submatrix $\mathbf{Z}^k = (\mathbf{z}^k_1, \mathbf{z}^k_1, \cdots, \mathbf{z}^k_{|\mathcal{L}^k| \times |\mathcal{L}^k|})$ and vector $\mathbf{z}^k_i \in \mathbb{R}^{d^k \times 1}$ represents the feature vector extracted for the i_{th} social link in network G^k. Furthermore, we can group all the projection function together and represent it as a $(d^t + d^1 + \cdots + d^K) \times c$ dimensional matrix

$$\mathbf{F} = \left((\mathbf{F}^t)^\top, (\mathbf{F}^1)^\top, \cdots, (\mathbf{F}^K)^\top \right)^\top,$$

which can be effectively inferred with the following Theorem 1.

Theorem 1. The projection functions that minimize the overall cost function are given by the eigenvectors corresponding to the smallest non-zero eigenvalues of the generalized eigenvalue decomposition

$$\mathbf{Z}(\mu \mathbf{L}_A + \mathbf{L}_S)\mathbf{Z}^\top \mathbf{x} = \lambda \mathbf{Z} \mathbf{L}_D \mathbf{Z}^\top \mathbf{x}.$$

Proof. Depending on the specific value of c, we propose to prove the theorem by considering two cases:
Case 1: if $c > 1$, with the above defined matrices, we can rewrite the introduced cost terms $Cost_A$, $Cost_S$ and $Cost_D$ in the linear algebra representation:

$$Cost_A = \text{Tr}(\mathbf{F}^\top \mathbf{Z} \mu \mathbf{L}_A \mathbf{Z}^\top \mathbf{F}),$$
$$Cost_S = \text{Tr}(\mathbf{F}^\top \mathbf{Z} \mathbf{L}_S \mathbf{Z}^\top \mathbf{F}),$$
$$Cost_D = \text{Tr}(\mathbf{F}^\top \mathbf{Z} \mathbf{L}_D \mathbf{Z}^\top \mathbf{F}).$$

Furthermore, the objective function can be represented as

$$\arg\min_{\mathbf{F}} \frac{\text{Tr}(\mathbf{F}^\top \mathbf{Z}(\mu \mathbf{L}_A + \mathbf{L}_S)\mathbf{Z}^\top \mathbf{F})}{\text{Tr}(\mathbf{F}^\top \mathbf{Z} \mathbf{L}_D \mathbf{Z}^\top \mathbf{F})}.$$

According to [21], [22], the matrix $\mathbf{F}$ which can minimize the objective function are actually the c eigenvectors corresponding to the c smallest non-zero eigenvalues of the following generalized eigenvalue decomposition function:

$$\mathbf{Z}(\mu \mathbf{L}_A + \mathbf{L}_S)\mathbf{Z}^\top \mathbf{x} = \lambda \mathbf{Z} \mathbf{L}_D \mathbf{Z}^\top \mathbf{x}.$$

Case 2: if $c = 1$, then matrix $\mathbf{F}$ to be inferred is actually a vector and the cost terms can be simply represented as

$$Cost_A = \mathbf{F}^\top \mathbf{Z} \mu \mathbf{L}_A \mathbf{Z}^\top \mathbf{F},$$
$$Cost_S = \mathbf{F}^\top \mathbf{Z} \mathbf{L}_S \mathbf{Z}^\top \mathbf{F},$$
$$Cost_D = \mathbf{F}^\top \mathbf{Z} \mathbf{L}_D \mathbf{Z}^\top \mathbf{F}.$$

The optimization objective function can be rewritten with the new cost representations as

$$\arg\min_{\mathbf{F}} \frac{\mathbf{F}^\top \mathbf{Z}(\mu \mathbf{L}_A + \mathbf{L}_S)\mathbf{Z}^\top \mathbf{F}}{\mathbf{F}^\top \mathbf{Z} \mathbf{L}_D \mathbf{Z}^\top \mathbf{F}},$$

which is actually the *Rayleigh quotient* of $(\mu \mathbf{L}_A + \mathbf{L}_S)$ relative to $\mathbf{L}_D$. According to the existing books on linear algebra and related works [19], [16], the optimal solution to the objective function can be represented as the eigenvectors corresponding to the c small non-zero eigenvalues of the generalized eigenvalue problem:

$$\mathbf{Z}(\mu \mathbf{L}_A + \mathbf{L}_S)\mathbf{Z}^\top \mathbf{x} = \lambda \mathbf{Z} \mathbf{L}_D \mathbf{Z}^\top \mathbf{x}.$$

□

Therefore, we can formally represent the feature tensors of network G^k (including both the target and aligned source networks) after the domain adaption as $\hat{\mathbf{X}}^k \in \mathbb{R}^{|\mathcal{U}^k| \times |\mathcal{U}^k| \times c}$ ($\forall k \in \{t, 1, 2, \cdots, K\}$), where feature vector

$$\hat{\mathbf{X}}^k(i,j,:) = (\mathbf{F}^k)^\top \mathbf{X}^k(i,j,:).$$

4) Inter-Network Link Prediction Objective Function: With the information from the external source networks, we can obtain more knowledge about the users and their social patterns. Based on the adapted feature tensors $\hat{\mathbf{X}}^1, \cdots, \hat{\mathbf{X}}^K$, we can represent the intimacy scores of the potential social links as

$$int(\mathbf{S}, \hat{\mathbf{X}}^1, \cdots, \hat{\mathbf{X}}^K) = \sum_{k=1}^{K} \alpha^i \cdot int(\mathbf{S}, \hat{\mathbf{X}}^k)$$

where term $int(\mathbf{S}, \hat{\mathbf{X}}^k) = \left\| \mathbf{S} \circ \hat{\mathbf{X}}^k \right\|_1$, and users in $\hat{\mathbf{X}}^k$ are organized in the same order as $\mathbf{X}^t$. Parameters α^i denotes the importance of the information transferred from the source network G^i. Furthermore, by adding the intimacy terms about the source networks into the objective function, we can rewrite it as follows:

$$\begin{aligned}\arg\min_{\mathbf{S}\in\mathcal{S}} \quad & l(\mathbf{S}, \mathbf{A}^t) - \alpha^t \cdot int(\mathbf{S}, \hat{\mathbf{X}}^t) - \sum_{k=1}^{K} \alpha^i \cdot int(\mathbf{S}, \hat{\mathbf{X}}^k)) \\ & + \gamma \cdot \|\mathbf{S}\|_1 + \tau \cdot \|\mathbf{S}\|_*\end{aligned}$$

D. Proximal Operator based CCCP Algorithm

By studying the objective function, we observe that the intimacy terms are convex while the empirical loss term $l(\mathbf{S}, \mathbf{A}^t)$ is non-convex. We propose to approximate it with other classical loss functions (e.g., the hinge loss and the Frobenius norm) instead, and the convex squared Frobenius norm loss function is used in this paper (i.e., $l(\mathbf{S}, \mathbf{A}^t) = \|\mathbf{S} - \mathbf{A}^t\|_F^2$). Therefore, the above objective function can be represented as a convex loss term minus another convex term together with two convex non-differentiable regularizers, which actually renders the objective function non-trivial. According to the existing works [23], [17], this kind of objective function can be addressed with the concave-convex procedure (CCCP). CCCP is a majorization-minimization algorithm that solves the difference of convex functions problems as a sequence of convex problems. Meanwhile, the regularization terms can be effectively handled with the proximal operators in each iteration of the CCCP process.

1) CCCP Algorithm: Formally, we can decompose the objective function into two convex functions:

$$\begin{aligned} u(\mathbf{S}) &= l(\mathbf{S}, \mathbf{A}^t) + \gamma \cdot \|\mathbf{S}\|_1 + \tau \cdot \|\mathbf{S}\|_*, \\ v(\mathbf{S}) &= \alpha^t \cdot int(\mathbf{S}, \hat{\mathbf{X}}^t) + \sum_{k=1}^{K} \alpha^i \cdot int(\mathbf{S}, \hat{\mathbf{X}}^k), \end{aligned}$$

With $u(\mathbf{S})$ and $v(\mathbf{S})$, we can rewrite the objective function to be

$$\arg\min_{\mathbf{S}\in\mathcal{S}} u(\mathbf{S}) - v(\mathbf{S}).$$

The CCCP algorithm can address the objective function with an iterative procedure that solves the following sequence of convex problems:

$$\mathbf{S}^{(h+1)} = \arg\min_{\mathbf{S}\in\mathcal{S}} u(\mathbf{S}) - \mathbf{S}^\top \nabla v(\mathbf{S}^{(h)})$$

It is easy to show that function $v(\mathbf{S})$ differentiable, and the derivative of function $v(\mathbf{S})$ is actually a constant term

$$\nabla v(\mathbf{S}) = \sum_{k=t}^{K} \alpha^i \sum_{i=1}^{c} \hat{\mathbf{X}}^k(i, :, :).$$

By relying on the Zangwill's global convergence theory [24] of iterative algorithms, it is theoretically proven in [17] that as such a procedure continues, the generated sequence of the variables $\{\mathbf{S}^{(h)}\}_{h=0}^{\infty}$ will converge to some stationary points $\mathbf{S}_*$ in the inference space $\mathcal{S}$.

2) Proximal Operators: Meanwhile, in each iteration of the CCCP updating process, objective function is not easy to address due to the non-differentiable regularizers. Some works have been done to deal with the objective function involving non-smooth functions. The Forward-Backward splitting method proposed in [2] can handle such a kind of optimization function with one single non-smooth regularizer based on the introduced proximal operators. More specifically, as introduced in [2], we can represent the proximal operators for the trace norm and L_1 norm as follows

$$\begin{aligned} &\text{prox}_{\tau\|\cdot\|_*}(\mathbf{S}) = \mathbf{U}\text{diag}((\sigma_i - \tau)_+)_i \mathbf{V}^\top, \\ &\text{prox}_{\gamma\|\cdot\|_1}(\mathbf{S}) = \text{sgn}(\mathbf{S}) \circ (|\mathbf{S}| - \gamma)_+, \end{aligned}$$

where $\mathbf{S} = \mathbf{U}\text{diag}(\sigma_i)_i \mathbf{V}^\top$ denotes the singular decomposition of matrix $\mathbf{S}$, and $\text{diag}(\sigma_i)_i$ represents the diagonal matrix with values σ_i on the diagonal.

Recently, some works have proposed the generalized Forward-Backward algorithm to tackle the case with $q (q \geq 2)$ non-differentiable convex regularizers [13]. These methods alternate the gradient step and the proximal steps to update the variables. For instance, given the above objective function in iteration h of the CCCP, we can represent the alternative updating equations in step k to address the objective function as follows:

$$\begin{cases} \mathbf{S}^{(k)} &= \mathbf{S}^{(k-1)} - \theta \cdot \nabla_{\mathbf{S}} \left(l(\mathbf{S}, \mathbf{A}) - \mathbf{S}^\top \nabla v(\mathbf{S}^{(h)}) \right), \\ \mathbf{S}^{(k)} &= \text{prox}_{\theta\tau\|\cdot\|_*}(\mathbf{S}^{(k)}), \\ \mathbf{S}^{(k)} &= \text{prox}_{\theta\gamma\|\cdot\|_1}(\mathbf{S}^{(k)}), \end{cases}$$

where the parameter θ denotes the learning rate and it is assigned with a very small value to ensure the converge of the above functions [15]. We will also give the convergence analysis about the model in the experiment section.

The pseudo-code of the Proximal Operators based CCCP algorithm is available in Algorithm 1.

IV. EXPERIMENTS

To test the effectiveness of the proposed model, we have conducted extensive experiments on real-world aligned networks. In this section, we will first introduce the datasets and the detailed experiment settings. After that we will show the experimental results together with the analysis about the parameters and convergence.

Algorithm 1 Proximal Operator Based CCCP Algorithm

Input: social adjacency matrix $\mathbf{A}$
projected feature tensors $\hat{\mathbf{X}}^t$, $\hat{\mathbf{X}}^1$, $\cdots$, $\hat{\mathbf{X}}^K$
Output: link predictor matrix $\mathbf{S}$

Initialize matrix $\mathbf{S}_{cccp} = \mathbf{A}$
Initialize CCCP convergence CCCP-tag = False
while CCCP-tag == False **do**
    Initialize Proximal convergence Proximal-tag = False
    Solve optimization function $\min_{\mathbf{S} \in \mathcal{S}} u(\mathbf{S}) - \mathbf{S}^\top \nabla v(\mathbf{S}_{cccp})$
    Initialize $\mathbf{S}_{po} = \mathbf{S}_{cccp}$
    while Proximal-tag == False **do**
        $\mathbf{S}_{po} = \mathbf{S}_{po} - \theta \nabla_{\mathbf{S}} \left(l(\mathbf{S}_{po}, \mathbf{A}) - \mathbf{S}_{po}^\top \nabla v(\mathbf{S}_{cccp}) \right)$
        $\mathbf{S}_{po} = \text{prox}_{\theta\tau\|\cdot\|_*}(\mathbf{S}_{po})$
        $\mathbf{S}_{po} = \text{prox}_{\theta\gamma\|\cdot\|_1}(\mathbf{S}_{po})$
        if $\mathbf{S}_{po}$ converges **then**
            Proximal-tag = True
            $\mathbf{S}_{cccp} = \mathbf{S}_{po}$
        end if
    end while
    if $\mathbf{S}_{cccp}$ converges **then**
        CCCP-tag = True
    end if
end while
Return $\mathbf{S}_{cccp}$

A. Dataset Description

The data used in the experiments include two *aligned social networks* Foursquare and Twitter simultaneously.

- *Foursquare*: The Foursquare network is a famous *location based social network*(LBSN), which provides users with various kinds of location-related services. Generally, users in Foursquare can perform various categories of social activities, like make friends with other users, write posts/reviews, check in at locations, etc.
- *Twitter*: The Twitter network is a famous *micro-blogging sites* that allows users to write, read, like, and share posts with their friends online. Formally, the short posts in Twitter is called the *tweets*, which can involve 140 characters (at most), images, hashtags, timestamps, location checkins, etc.

The basic statistical information about the Foursquare and Twitter datasets is available in Table I. We have crawled about $5,223$ Twitter users, and all the $164,920$ social follow connections among them. These crawled Twitter users have posted $9,490,707$ tweets, among which $615,515$ pieces of tweets are attached with geo-spatial location checkins. Among the $5,223$ Twitter users, $3,388$ of them are aligned by anchor links with users in Foursquare. From Foursquare, we have crawled $5,392$ users together with $76,972$ friendship links among them. These Foursquare users have written $48,756$ posts which are all attached with location checkins.

B. Experimental Setting

In this section, we will introduce the experimental settings, which include the detailed experiment setups, the comparison methods and the metrics used to evaluate the performance of the methods.

1) Experiment Setup: In the experiments, we use Twitter as the target network and use Foursquare as the source network.

TABLE I
PROPERTIES OF THE HETEROGENEOUS NETWORKS

		network	
	property	**Twitter**	**Foursquare**
# node	user	5,223	5,392
	tweet/tip	9,490,707	48,756
	location	297,182	38,921
# link	friend/follow	164,920	76,972
	write	9,490,707	48,756
	locate	615,515	48,756

Therefore, the two aligned networks used in the experiments can be represented as $\mathcal{G} = (\{G^t, G^s\}, \{A^{t,s}\})$. From the target network G^t, we extract the existing links, which are partitioned into 5 folds where 4 folds are used as the training set and 1 fold is used as the test set. We hide the test set as the ground truth, and try to build the model based on the training set to identify these hidden links from the network.

For our model SLAMPRED, we construct the social adjacency matrix based on the training set in the target network, and we use the same set of features introduced in [28] to denote the intimacy scores among the users in the target network and aligned source network respectively. After adapting the domain differences, we define and resolve the objective function for the inter-network link prediction. Based on the inferred link prediction matrix $\mathbf{S}$, we can obtain confidence scores of the corresponding links in the test set, which will be outputted as the inference results of the SLAMPRED model. In model SLAMPRED, the parameters $\mu = 1.0$ (weight of the anchor link based cost term), learning rate $\theta = 0.001$, weight of the regularization terms $\tau = 1.0$ and $\gamma = 1.0$. Analysis about weights of the attribute intimacy terms (i.e., α_t and α_s) is available in Section IV-D1.

2) Comparison Methods: Depending on the specific learning setting, the comparison methods used in the experiments can be divided into 4 main categories, and each category of methods also have several different variants. More detailed information about the comparison methods are listed as follows:

Sparse Low-Rank Matrix Estimation based Methods

- SLAMPRED: Method SLAMPRED is the link prediction method proposed in this paper. SLAMPRED is built based on sparse and low-rank matrix estimation model with both the social structure and other attribute information available in both target and other aligned source networks.
- SLAMPRED-T (SLAMPRED-Target): Method SLAMPRED-T is a variant of SLAMPRED, which is built based on the social structure information and attribute information in the target network only.
- SLAMPRED-H (SLAMPRED-Homogeneous): Similar to method SLAMPRED-T, the method SLAMPRED-H is built merely based on the social structure information in the target network without using other information.

PU Classification based Link Prediction

- PL [37]: Method PL is the PU link prediction method from the existing work [37] with the features extracted based on meta paths from both the target and the source networks simultaneously (Different from the method in [37], no feature selection is used to resolve the domain differences here).

- PL-T: Method PL-T is a variant of PL, and it is also a PU link prediction model built based on the features extracted from the target network only.
- PL-S: Different from PL-T, the PU link prediction method PL-S (a variant of PL) is built merely with the features extracted from the source network.

Supervised Classification based Link Prediction

- SCAN [28]: Supervised classification based link prediction method SCAN [28] is built based on the features extracted from both the target and other aligned source networks. SCAN labels the existing and non-existing links as the positive and negative instances, which are used to build the classifiers.
- SCAN-T: Method SCAN-T is a variant of SCAN, which utilizes the features extracted from the target network only.
- SCAN-S: Different from SCAN-T, the classification based link prediction method SCAN-S is built merely with the features from other aligned source networks.

Unsupervised Link Prediction

For completeness, we also compare the above link prediction methods with many other unsupervised link predictors, which include PA, CN and JC. Given the user paris (u, v) (with neighbor set $\Gamma(u)$ and $\Gamma(v)$ respectively), these methods calculate the following scores as the output.

- PA (Preferential Attachment Index) [1]: The PA predictor calculates the product of the u's and v's degrees as the confidence score, i.e., $|\Gamma(u)| \cdot |\Gamma(v)|$.
- CN (Common Neighbor) [6]: Link predictor CN counts the number of shared neighbors by u and v as the confidence score of user pair (u, v), i.e., $|\Gamma(u) \cap \Gamma(v)|$.
- JC (Jaccard's Coefficient) [6]: JC calculates the Jaccard's Coefficient of the neighborhood information as the confidence score for potential user pair (u, v), i.e., $\frac{|\Gamma(u) \cap \Gamma(v)|}{|\Gamma(u) \cup \Gamma(v)|}$.

3) Evaluation Metrics: The comparison link prediction methods can output the confidence scores of potential social links among users in the target network, whose performances are evaluated with the well-used evaluation metrics, like AUC and Precision@100, in the experiments.

C. Experimental Result

The experimental results obtained by the different comparison methods evaluated by AUC and Precision@100 are shown in Table II. In the table, the anchor link sampling rate denotes the proportion of anchor links connection the Twitter and Foursquare networks, where the ratio 0 denotes these two networks are unaligned and ratio 1.0 represents the networks are fully aligned. Generally, from the result, we can observe that as the anchor link sampling ratio increases, the performance achieved by the methods utilizing information from the source networks (i.e., SLAMPRED, PL, PL-S, SCAN, SCAN-S) will change accordingly, while the performance of the methods merely using information in the target network will stay the same.

Compared with SLAMPRED-T and SLAMPRED-H, with the heterogeneous information from both the target network and other aligned source networks, SLAMPRED can outperform these methods with one single kind of information (i.e., social structure information) or one single information source (i.e., the target network). It supports the motivation of using information from multiple aligned social networks to improve the link prediction results.

By comparing SLAMPRED with PL, SCAN as well as the unsupervised link predictors JC, CN and PA, method SLAMPRED proposed in this paper can overcome these comparison methods with great advantages. For instance, at anchor ratio 1.0, the AUC score obtained by SLAMPRED is 0.941, which is about 13% larger than that gained by PL; and over 46% greater than that achieved by SCAN, JC, CN and PA. Meanwhile, the Precision@100 obtained by SLAMPRED is almost the twice even three times larger than that obtained by PL, SCAN, JC, CN and PA. The comparison results demonstrate the effectiveness of SLAMPRED in predicting links across networks. With the sparse and low-rank regularization, SLAMPRED can also overcome the class imbalance problem encountered by the classification based models, like PL and SCAN.

For method SLAMPRED, which fuses information from multiple aligned networks in the link inference with domain accommodation, adding more available anchor links will improve the performance of SLAMPRED steadily and outperform the other baseline methods consistently. Meanwhile, for the methods without domain adaptions, e.g., PL and SCAN, adding more anchor links may not necessary leads to better performance. For instance, by comparing the AUC score achieved by PL and PL-T at anchor ratios 0.0 and 0.1, we observe that with more anchor links, the performance of PL will degrade a lot which is even worse than that of PL-T (using information merely in the target network) at anchor ratio 0.1. Similar results can be observed for methods SCAN, SCAN-T and SCAN-S. It can demonstrate the effectiveness of the feature projection step used in SLAMPRED in accommodating the domain difference.

D. Experimental Analysis

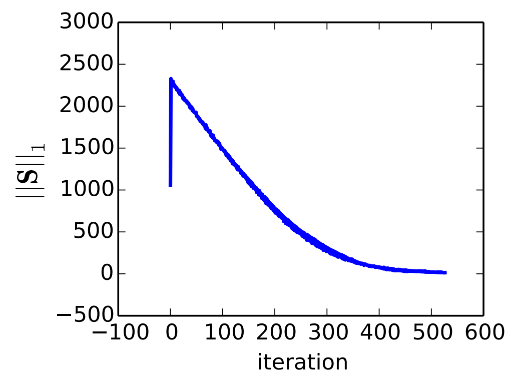
In this part, we will show the convergence analysis and parameter analysis about the proposed link prediction model.

1) Convergence Analysis: In building the model, we propose to apply the iterative CCCP to resolve the objective function, which calculates a series of the inferred social link prediction matrix $\mathbf{S}$ until convergence. To show that such a procedure will converge, we give the L_1 norm of the variable $\mathbf{S}$ in each iteration (i.e., $\left\|\mathbf{S}^h\right\|_1$) as well as the L_1 norm of the matrix changes (i.e., $\left\|\mathbf{S}^h - \mathbf{S}^{h-1}\right\|_1$) in Figure 3. From the plots, we can observe that the variable matrix $\mathbf{S}$ will converge in about 300 rounds of the iteration, where the changes of both the variable $\mathbf{S}^h$ itself as well as the updates of the variable

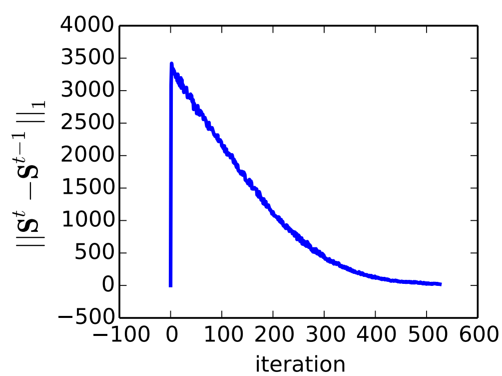
TABLE II

PERFORMANCE COMPARISON OF DIFFERENT METHODS FOR INFERRING SOCIAL LINKS FOR TWITTER WITH DIFFERENT ANCHOR LINK SAMPLING RATIOS.

		Different Ratios of Anchor Link between the Aligned Networks.										
measure	methods	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
AUC	SLAMPRED	**0.828±0.009**	**0.898±0.019**	**0.9±0.019**	**0.907±0.019**	**0.911±0.019**	**0.918±0.019**	**0.921±0.019**	**0.928±0.019**	**0.929±0.019**	**0.937±0.019**	**0.941±0.019**
	SLAMPRED-T	0.828±0.009	0.828±0.009	0.828±0.009	0.828±0.009	0.828±0.009	0.828±0.009	0.828±0.009	0.828±0.009	0.828±0.009	0.828±0.009	0.828±0.009
	SLAMPRED-H	0.776±0.03	0.776±0.03	0.776±0.03	0.776±0.03	0.776±0.03	0.776±0.03	0.776±0.03	0.776±0.03	0.776±0.03	0.776±0.03	0.776±0.03
	PL	0.706±0.018	0.637±0.115	0.687±0.068	0.699±0.103	0.795±0.013	0.779±0.022	0.796±0.023	0.795±0.013	0.819±0.024	0.817±0.011	0.834±0.015
	PL-T	0.706±0.018	0.706±0.018	0.706±0.018	0.706±0.018	0.706±0.018	0.706±0.018	0.706±0.018	0.706±0.018	0.706±0.018	0.706±0.018	0.706±0.018
	PL-S	0.5±0.0	0.48±0.023	0.462±0.047	0.649±0.075	0.724±0.021	0.768±0.012	0.776±0.021	0.79±0.022	0.802±0.015	0.813±0.008	0.843±0.005
	SCAN	0.730±0.009	0.730±0.009	0.738±0.005	0.725±0.01	0.73±0.01	0.719±0.013	0.717±0.017	0.725±0.014	0.722±0.009	0.673±0.02	0.643±0.024
	SCAN-T	0.730±0.009	0.730±0.009	0.730±0.009	0.730±0.009	0.730±0.009	0.730±0.009	0.730±0.009	0.730±0.009	0.730±0.009	0.730±0.009	0.730±0.009
	SCAN-S	0.5±0.0	0.529±0.038	0.628±0.018	0.649±0.015	0.69±0.003	0.697±0.012	0.693±0.011	0.661±0.039	0.674±0.038	0.586±0.016	0.565±0.026
	JC	0.624±0.014	0.624±0.014	0.624±0.014	0.624±0.014	0.624±0.014	0.624±0.014	0.624±0.014	0.624±0.014	0.624±0.014	0.624±0.014	0.624±0.014
	CN	0.631±0.017	0.631±0.017	0.631±0.017	0.631±0.017	0.631±0.017	0.631±0.017	0.631±0.017	0.631±0.017	0.631±0.017	0.631±0.017	0.631±0.017
	PA	0.557±0.02	0.557±0.02	0.557±0.02	0.557±0.02	0.557±0.02	0.557±0.02	0.557±0.02	0.557±0.02	0.557±0.02	0.557±0.02	0.557±0.02
Precision@100	SLAMPRED	**0.35±0.042**	**0.41±0.035**	**0.42±0.035**	**0.43±0.035**	**0.42±0.035**	**0.46±0.035**	**0.42±0.035**	**0.44±0.035**	**0.47±0.035**	**0.48±0.035**	**0.48±0.035**
	SLAMPRED-T	0.35±0.042	0.35±0.042	0.35±0.042	0.35±0.042	0.35±0.042	0.35±0.042	0.35±0.042	0.35±0.042	0.35±0.042	0.35±0.042	0.35±0.042
	SLAMPRED-H	0.28±0.012	0.28±0.012	0.28±0.012	0.28±0.012	0.28±0.012	0.28±0.012	0.28±0.012	0.28±0.012	0.28±0.012	0.28±0.012	0.28±0.012
	PL	0.20±0.008	0.17±0.029	0.27±0.029	0.18±0.038	0.3±0.035	0.19±0.021	0.2±0.025	0.19±0.008	0.26±0.015	0.28±0.029	0.23±0.012
	PL-T	0.20±0.008	0.20±0.03	0.20±0.017	0.20±0.03	0.20±0.02	0.20±0.031	0.0±0.025	0.20±0.026	0.20±0.015	0.20±0.028	0.20±0.015
	PL-S	0.0±0.0	0.01±0.008	0.0±0.004	0.05±0.019	0.11±0.012	0.08±0.024	0.11±0.019	0.08±0.024	0.12±0.015	0.12±0.017	0.15±0.033
	SCAN	0.27±0.055	0.23±0.032	0.24±0.026	0.25±0.05	0.24±0.058	0.23±0.023	0.24±0.052	0.26±0.028	0.26±0.035	0.27±0.027	0.26±0.029
	SCAN-T	0.27±0.055	0.27±0.041	0.27±0.055	0.27±0.028	0.27±0.033	0.27±0.026	0.27±0.024	0.27±0.023	0.27±0.021	0.27±0.043	0.27±0.023
	SCAN-S	0.0±0.0	0.1±0.044	0.18±0.032	0.21±0.033	0.17±0.036	0.21±0.032	0.24±0.036	0.22±0.029	0.23±0.024	0.24±0.029	0.26±0.033
	JC	0.12±0.029	0.12±0.029	0.12±0.029	0.12±0.029	0.12±0.029	0.12±0.029	0.12±0.029	0.12±0.029	0.12±0.029	0.12±0.029	0.12±0.029
	CN	0.12±0.014	0.12±0.014	0.12±0.014	0.12±0.014	0.12±0.014	0.12±0.014	0.12±0.014	0.12±0.014	0.12±0.014	0.12±0.014	0.12±0.014
	PA	0.07±0.033	0.07±0.033	0.07±0.033	0.07±0.033	0.07±0.033	0.07±0.033	0.07±0.033	0.07±0.033	0.07±0.033	0.07±0.033	0.07±0.033



(a) $\|\mathbf{S}^h\|_1$



(b) $\|\mathbf{S}^h - \mathbf{S}^{h-1}\|_1$

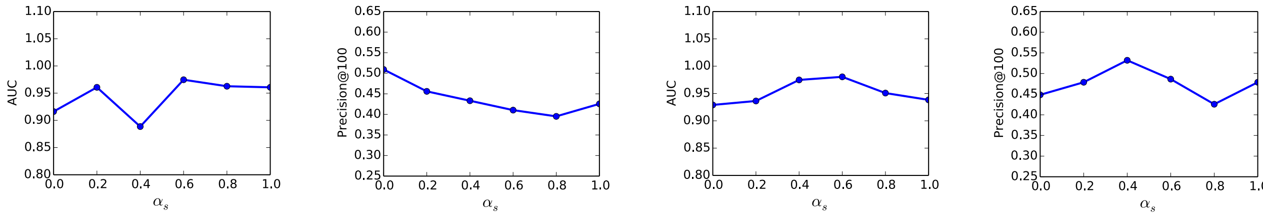
Fig. 3. Convergence Analysis of the iterative CCCP.

(i.e., $\mathbf{S}^h - \mathbf{S}^{h-1}$) will approach to the stationary states within a number of iterations respectively.

2) Parameter Analysis: The effects of the weight parameters α_t and α_s (i.e., the weights of the intimacy terms from G^t and G^s) on the performance of SLAMPRED will be studied in this part. In Figure 4, we fix parameter α_t but change parameter α_s with different values in $\{0.0, 0.2, \cdots, 1.0\}$. By assigning α_t with value 0.0 (i.e., intimacy term of G^t is not used), increasing the weight of the intimacy term about the aligned source network will slightly degrade the performance of SLAMPRED as shown in plots 4(a) and 4(b). The possible reason can be that the accommodated attribute information from the source network may still have some differences from the distribution of links in the target network, and assigning it with high weights may not necessary improve the performance of SLAMPRED.

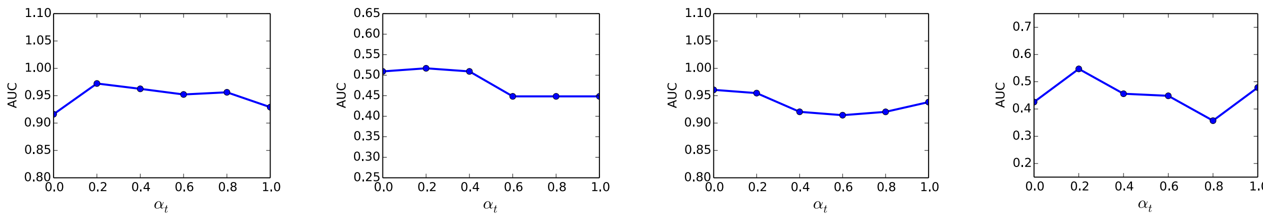
Meanwhile, as shown in plots 4(c) and 4(d), by fixing parameter α_t with value 1.0, increasing the value of α_s will improve the performance of SLAMPRED first and then degrade its performance. The reason can be that given the full consideration of attribute information in the target network, adding some complementary information from the source network will improve the inference performance, but assigning it with a too large weight will make the model overfit the attribute information from the source network instead.

For completeness, we also propose to fix parameter α_s but assign parameter α_t with values in $\{0.0, 0.2, \cdots, 1.0\}$. The results are shown in Figure 5. By discarding the attribute information from the source network (i.e., fixing α_s with value 0.0) or treating it as an important information source (i.e., fixing α_s with value 1.0), increasing the weight of the attribute information term about the target network will improve its performance first and then worsen its performance. The potential explanation for such an observation can be that assigning the attribute information terms with larger weights will make the model overfit the attribute information in the target network.



(a) AUC ($\alpha_t = 0.0$) (b) Prec@100 ($\alpha_t = 0.0$) (c) AUC ($\alpha_t = 1.0$) (d) Prec@100 ($\alpha_t = 0.0$)

Fig. 4. Parameter Analysis of α_s with Fixed α_t.



(a) AUC ($\alpha_s = 0.0$) (b) Prec@100 ($\alpha_s = 0.0$) (c) AUC ($\alpha_s = 1.0$) (d) Prec@100 ($\alpha_s = 0.0$)

Fig. 5. Parameter Analysis of α_t with Fixed α_s.

E. Experimental Discovery Summary

According to the above experimental results and analysis, we summarize the experimental discoveries as follows:

- *Anchor Link*: Introducing more anchor links between the target and other aligned source networks can help transfer more information from the source networks and improve the performance of link prediction model SLAMPRED.
- *Regularization*: The sparse and low-rank regularization terms works well in improving the performance of SLAMPRED, and can help overcome the class imbalance problem effectively.
- *Domain Adaption*: The domain adaption method proposed in this paper works well in accommodate the information distributions between the source and target domains.
- *Convergence of CCCP*: The CCCP can help identify solution to the objective function, and the updating process can converge within a reasonable number of iterations.
- *Parameter Selection*: Incorporating the attribute information from the target and source networks helps build better model. However, the weights of the attribute information from the target and source networks need to be selected carefully, and too large weights will make the model overfit the attribute information.

V. RELATED WORK

Link prediction and recommendation first proposed in[11] has become a very important problem in online social networks, which provides social network researchers with the opportunity to study both the network properties from the individuals social connection perspective. Traditional unsupervised link predictor proposed in [11] mainly calculate the closeness scores among users, and assume that close users tend to be friends in the network. Hasan et al. [5] is the first to study the link prediction problem as a supervised learning problem, where the existing and non-existing social links are treated as the positive and negative instances respectively. In [5], the authors propose to build supervised learning models to classify the link instances to do the prediction. Today, many social networks are heterogeneous and to conduct the link prediction in these networks, Sun et al. [18] propose a meta path-based prediction model to predict co-author relationship in the heterogeneous bibliographic network.

Most existing works solve link prediction problem with a single source of information. Nowadays, the researchers have pushed the problem boundary further forward by proposing the link prediction across multiple domains. Tang et al. [20] focus on inferring the particular type of links over multiple heterogeneous networks and develop a framework for classifying the type of social ties. To deal with the differences in information distributions of multiple networks, Qi et al. [4] propose to use biased cross-network sampling to do link prediction across networks. Meanwhile, some works have also been done on predicting multiple kinds of links simultaneously. I. Konstas et al. [9] propose to recommend multiple kinds of links with collaborative filtering methods. F. Fouss et al. [3] propose to use a traditional model, random walk, to predict multiple kinds of links simultaneously in networks.

Since Zhang et al. [8], [28] propose the concept of "*aligned social networks*", "*anchor links*", "*anchor users*", the social network studies across *multiple aligned social networks* have become a hot research area in recent years. Dozens of papers have been published around various problems about the multiple aligned networks, including *network alignment* [8], [35], [36], *link prediction* [29], [28], [37], [33], [27], *community detection* [32], [34], [7] and *information diffusion* [25], [26] etc. The link prediction models introduced in [29], [28], [37], [33] propose to combine the information from different sites by simply merging the extracted feature vectors together without considerations about the domain differences at all, which are totally different from the model introduced in this paper. The recent paper [27] aims at unifying the link prediction problems subject to different cardinality constraints, like *one-to-one*, *one-to-many* and *many-to-many*, and introduce a general scalable link prediction framework to solve the problem.

The link prediction model proposed in this paper is based on the sparse and low-rank matrix estimation. Richard et al. [15] introduce the penalized matrix estimation procedure aiming at solutions that are sparse and low-rank simultaneously. By incorporating the regularization terms about the sparsity and rank of the matrix to be inferred, the authors formulate the matrix estimation problem into a joint optimization problem. The model proposed in [15] can be applied in various types of application problems, which include *covariance matrix estimation*, *graph denoising*, and *link prediction*. By assuming the feature vectors to be autoregressive, Richard et al. [14] propose to study the link prediction problem in time-evolving graphs. The authors propose to formulate and address the problem as a sparse and low-rank matrix estimation problem, and provide the theoretical analysis about the introduced error bounds. Zhi et al. [38] study the link inference problem in the link-inconsistent case based on sparse and low-rank matrix estimation. Given a network and a small number of labeled nodes, they aim at learning a consistent network with more consistent and fewer inconsistent links than the original network.

VI. Conclusion

In this paper, we have studied the link prediction problem with multiple aligned networks by estimating a low-rank and sparse adjacency matrix. To address the domain difference issue, we project the feature vector of the users in multiple networks into a common low-dimensional space. Besides minimizing the loss function and maximizing the intimacy terms, two extra regularization terms are added to the objective function to guarantee the sparsity and low-rank properties. The objective function can be effectively resolved by the CCCP iteratively, and the non-differentiable regularization terms in the objective function are handled with the proximal operators. Based on two real-world partially aligned social networks (Foursquare and Twitter), extensive experiments have been done in this paper, and the experimental results have already demonstrated the effectiveness of the proposed model.

VII. Acknowledgement

This work is supported in part by NSF through grants IIS-1526499, and CNS-1626432, and NSFC 61672313. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X GPU used for this research.

References

[1] A.-L. Barabasi, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek. Evolution of the social network of scientific collaboration. In *Physica A*, pages 590–614, 2002.

[2] P. L. Combettes and V. Wajs. Signal Recovery by Proximal Forward-Backward Splitting. *Multiscale Modeling & Simulation*, 2005.

[3] F. Fouss, A. Pirotte, J. Renders, and M. Saerens. Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. *TKDE*, 2007.

[4] C. Aggarwal G. Qi and T. Huang. Link prediction across networks by biased cross-network sampling. In *ICDE*, 2013.

[5] M. Hasan, V. Chaoji, S. Salem, and M. Zaki. Link prediction using supervised learning. In *SDM*, 2006.

[6] M. Hasan and M. Zaki. A survey of link prediction in social networks. In C. Aggarwal, editor, *Social Network Data Analytics*. 2011.

[7] S. Jin, J. Zhang, P. Yu, S. Yang, and A. Li. Synergistic partitioning in multiple large scale social networks. In *IEEE BigData*, 2014.

[8] X. Kong, J. Zhang, and P. Yu. Inferring anchor links across multiple heterogeneous social networks. In *CIKM*, 2013.

[9] I. Konstas, V. Stathopoulos, and J. M. Jose. On social networks and collaborative recommendation. In *SIGIR*, 2009.

[10] H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In *WWW*, 2010.

[11] D. Liben-Nowell and J. Kleinberg. The link prediction problem for social networks. In *CIKM*, 2003.

[12] M. McPherson, L. Smith-Lovin, and J. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 2001.

[13] H. Raguet, J. Fadili, and G. Peyré. A generalized forward-backward splitting. *SIAM Journal on Imaging Sciences*, 2013.

[14] E. Richard, S. Gaïffas, and N. Vayatis. Link prediction in graphs with autoregressive features. *Journal of Machine Learning Research*, 2014.

[15] E. Richard, P. Savalle, and N. Vayatis. Estimation of simultaneously sparse and low rank matrices. In *ICML*, 2012.

[16] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000.

[17] B. Sriperumbudur and G. Lanckriet. On the convergence of concave-convex procedure. In *NIPS*, 2009.

[18] Y. Sun, R. Barber, M. Gupta, C. Aggarwal, and J. Han. Co-author relationship prediction in heterogeneous bibliographic networks. In *ASONAM*, 2011.

[19] Y. Sun and A. Damle. Matrix analysis notes. http://web.stanford.edu/~damle/refresher/notes/matrixanalysis.pdf. [Online; accessed 25-May-2016].

[20] J. Tang, T. Lou, and J. Kleinberg. Inferring social ties across heterogenous networks. In *WSDM*, 2012.

[21] C. Wang and S. Mahadevan. Heterogeneous domain adaptation using manifold alignment. In *IJCAI*, 2011.

[22] S. Wilks. *Mathematical Statistics*. Wiley, 1963.

[23] A. Yuille and A. Rangarajan. The concave-convex procedure. *Neural Computation*, 2003.

[24] W. Zangwill. *Nonlinear Programming*. Prentice-Hall, 1969.

[25] Q. Zhan, J. Zhang, S. Wang, P. Yu, and J. Xie. Influence maximization across partially aligned heterogenous social networks. In *PAKDD*. 2015.

[26] Q. Zhan, J. Zhang, P. Yu, and J. Xie. Discover tipping users for cross network influencing. In *IRI*, 2016.

[27] J. Zhang, J. Chen, J. Zhu, Y. Chang, and P. Yu. Link prediction with cardinality constraints. In *WSDM*, 2017.

[28] J. Zhang, X. Kong, and P. Yu. Predicting social links for new users across aligned heterogeneous social networks. In *ICDM*, 2013.

[29] J. Zhang, X. Kong, and P. Yu. Transfer heterogeneous links across location-based social networks. In *WSDM*, 2014.

[30] J. Zhang, Y. Lv, and P. Yu. Enterprise social link recommendation. In *CIKM*, 2015.

[31] J. Zhang and P. Yu. Link prediction across heterogeneous social networks: A survey. 2014.

[32] J. Zhang and P. Yu. Community detection for emerging networks. In *SDM*, 2015.

[33] J. Zhang and P. Yu. Integrated anchor and social link predictions across partially aligned social networks. In *IJCAI*, 2015.

[34] J. Zhang and P. Yu. Mcd: Mutual clustering across multiple social networks. In *IEEE BigData Congress*, 2015.

[35] J. Zhang and P. Yu. Multiple anonymized social networks alignment. In *ICDM*, 2015.

[36] J. Zhang and P. Yu. Pct: Partial co-alignment of social networks. In *WWW*, 2016.

[37] J. Zhang, P. Yu, and Z. Zhou. Meta-path based multi-network collective link prediction. In *KDD*, 2014.

[38] S. Zhi, J. Han, and Q. Gu. Robust classification of information networks by consistent graph learning. In *ECML/PKDD*, 2015.