Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)

Finding Cut from the Same Cloth: Cross Network Link Recommendation via Joint Matrix Factorization

Arun Reddy Nelakurthi, Jingrui He
Arizona State University, Tempe, AZ, USA
{anelakur,jingrui.he}@asu.edu

Abstract

With the emergence of online forums associated with major diseases, such as diabetes mellitus, many patients are increasingly dependent on such disease-specific social networks to gain access to additional resources. Among these patients, it is common for them to stick to one disease-specific social network, although their desired resources might be spread over multiple social networks, such as patients with similar questions and concerns. Motivated by this application, in this paper, we focus on cross network link recommendation, which aims to identify similar users across multiple heterogeneous social networks. The problem setting is different from existing work on cross network link prediction, which either tries to link accounts of the same user from different social networks, or aims to match users with complementary expertise or interest.

To approach the problem of cross network link recommendation, we propose to jointly decompose the user-keyword matrices from multiple social networks, while requiring them to share the same topics and user group-topic association matrices. This constraint comes from the fact that social networks dedicated to the same disease tend to share the same topics as well as the interests of users groups in certain topics. Based on this intuition, we construct a generic optimization framework, provide four instantiations and an iterative optimization algorithm with performance analysis. In the experiments, we demonstrate the superiority of the proposed algorithm over state-of-the-art techniques on various real-world data sets.

1 Introduction

Nowadays, online social networks has become an important portal for patients with major diseases, such as diabetes mellitus, to connect with physicians as well as other patients. Compared with the generic social networks such as Twitter and Facebook, the disease-specific social networks (e.g., TuDiabetes (2016) and DiabetesSisters (2016)) have a greater concentration of patients with similar conditions, and the patients expect to obtain additional resources from these social networks. However, when it comes to using these social networks, it is often the case that a patient would stick to a single social network, and rarely look at the other social networks, thus limiting their access to the online resources, especially the patients with similar questions and concerns from the other social networks. Motivated by this application, in this paper, we focus on cross network link recommendation, which aims to identify similar actors across multiple heterogeneous social networks. In this way, we will be able to form support groups consisting of patients from multiple disease-specific networks, all sharing the same questions and concerns.

The problem setting studied in this paper is similar and yet significantly different from existing work on cross network link prediction. In particular, existing work either links different accounts belonging to the same user across multiple social networks (Zhang et al. 2015), or links users with complementary expertise or interest (Tang et al. 2012). In contrast, we aim to find *similar* users using different social networks, which enables them to exchange important information regarding their shared questions or concerns.

Based on the observation that different disease-specific social networks tend to share the same topics as well as the interests of user groups in certain topics, we propose to jointly decompose the user-keyword matrices from these social networks, while requiring them to share the same topics and user group-topic association matrices. To be specific, we form a generic optimization framework, and instantiate it with variations of the constraints. Then we propose an iterative optimization algorithm and analyze its performance from multiple perspectives. Finally, we test the performance of this algorithm on various real-world data sets, which outperforms state-of-the-art techniques.

The rest of the paper is organized as follows, Section 2 discusses the related work in the field of link prediction and non-negative matrix factorization. Section 3 formalizes the problem of cross network link prediction and describes the proposed approach as well as the optimization algorithm. In Section 4, we evaluate the performance of our proposed algorithm and discuss the results on different data sets. Finally, we conclude the paper in Section 5.

2 Related Work

In this section, we briefly review the related work on link prediction and non-negative matrix factorization.

Link prediction is a widely studied problem in the field of Social Network Analysis (Liben-Nowell and Kleinberg 2007; Al Hasan and Zaki 2011; Wang et al. 2014). Link prediction can be broadly classified into two types: (1)

Classical link prediction which aims at predicting the missing links in a given social network (Al Hasan et al. 2006; Fortunato 2010); (2) Cross network link prediction that recommends the links across two or more social networks. Tang et al. (2012) modeled users as feature vector with in-domain and cross-domain topic distributions, and used it to learn associations between users across source and target domains. Kong, Zhang, and Yu (2013) suggested a multi-network anchoring algorithm to discover the correspondence between accounts of the same user in multiple networks. Zhang et al. (2015) proposed an energy-based framework COSNET for cross network link prediction in heterogeneous networks. Our problem differs with previous cross network link prediction problems, as we recommend links between similar actors across social networks.

Non-negative matrix factorization (NMF) is widely used for co-clustering problems. Li and Ding (2006) demonstrated a NMF framework for document-word co-clustering. Cai et al. (2011) improved Li and Ding (2006) framework by adding a graph regularizer which captures geometric information embedded in the data. Gu, Ding, and Han (2011) proposed an orthogonal framework to fix scaling problem in Cai et al. (2011). Wang, Nie, and Huang (2015) proposed a NMF based Dual Knowledge Transfer approach for cross-language Web page classification. Our approach differs from previous works as we jointly factor user-keyword matrices from multiple social networks to learn latent features on the combined set of keywords from all the social networks and users from each social network. Chakraborty and Sycara (2015) proposed a constrained NMF framework for community detection in social networks which is closely related to our work. Our problem is different from the community detection problem, which finds communities of closely related actors inside a single social network, whereas we find closely related actors across multiple social networks.

3 Cross Network Link Recommendation

In this section, we formally introduce the cross network link recommendation problem, followed by the proposed generic optimization framework and its instantiations. Then we present the iterative optimization algorithm as well as its performance analysis.

3.1 Notation and Problem Definition

Suppose that we have K disease-specific social networks: $\mathcal{G}_k = \langle V_k^U, E_k^U \rangle, k = 1, \ldots, K$, where V_k^U is the set of user nodes $|V_k^U| = m_k$ and $E_k^U \subseteq V_k^U \times V_k^U$ is the set of edges representing the connection between users in the same social network. Self-connections and multiple links between two user nodes are not allowed. Let $\mathbf{A}_k \subset \{0,1\}^{m_k \times m_k}$ denote the user-user adjacency matrix for the k^{th} social network $k = 1, \ldots, K$, where the edge weight is set to 1 if there is a connection between two users. Notice that we focus on the more challenging case where: (1) there are no shared user nodes across the social networks, i.e., $V_i^U \cap V_j^U = \emptyset, i \neq j \; \forall \, i, j = 1, \ldots, K$, and (2) there are no cross network links available between the users in different social networks. The goal of cross network link recommendation is to identify similar actors across multiple social networks. This is different from existing work on cross network link prediction which focuses on linking different accounts of the same user, or finding users with complementary expertise or interest.

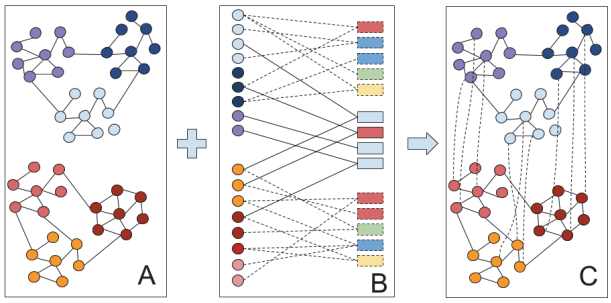

Figure 1: Cross network link prediction problem: A) Two social networks with user nodes represented by circles and user-user associations represented by edges joining two nodes. Different colors represent different user groups. B) User-keyword bipartite graph, circles represent users from different social networks, squares represent keywords from vocabulary space for different social networks. Dotted lines link the users to unique keywords in a social network and solid lines link users to shared keywords. C) Dotted lines represent the recommended links between similar actors across social networks.

Let $\mathcal{G}_k^{'} = \langle V_k^U, V_k^W, E_k^{UW} \rangle$ denote the undirected user-keyword bipartite graph for the k^{th} social network, where V_k^W is the set of keyword nodes $|V_k^W| = n_k$ and $E_k^{UW} \subseteq V_k^U \times V_k^W$ is the set of edges connecting the user nodes and the keyword nodes. Let $\mathbf{X}_k \subset \mathbb{R}^{m_k \times n_k}$ be the user-keyword adjacency matrix constructed from the bipartite graph $\mathcal{G}_k^{'}$, $k = 1, \ldots, K$. Let d be the size of the vocabulary for all the social networks combined, i.e., $|V_1^W \cup V_2^W \cup .. \cup V_K^W| = d$.

Figure 1 illustrates the cross network link recommendation problem with two social networks $K = 2$. Figure 1(A) shows the user-user connection graphs $\mathcal{G}_1$ and $\mathcal{G}_2$. Figure 1(B) represents the user-keyword bipartite graphs $\mathcal{G}_1^{'}$ and $\mathcal{G}_2^{'}$. Figure 1(C) represents the problem of cross network link recommendation that recommends links between user nodes from different social networks $\mathcal{G}_1$ and $\mathcal{G}_2$.

Problem. *Cross network link prediction across multiple social networks.*

Input: The input to the problem is a set of user-user adjacency matrices $\{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_K\}$ constructed from user relationship graphs $\mathcal{G}_k, k = 1, \ldots, K$ and a set of user-keyword adjacency matrices $\{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_K\}$ constructed from user-keyword bipartite graphs $\mathcal{G}_k^{'}, k = 1, \ldots, K$.

Output: A set of cross network links $E^U \subseteq V_i^U \times V_j^U$ connecting similar user nodes V_i^U from the social network $\mathcal{G}_i$ to user nodes V_j^U from the social network $\mathcal{G}_j$, where $i \neq j$ and $i, j = 1, \ldots, K$.

3.2 Matrix Factorization for Cross Network Link Recommendation

In order to identify the similar actors across multiple disease-specific social networks, we propose to perform co-clustering on user-keyword graphs to learn the representation of users and keywords in a latent feature space, and then recommend the links between similar actors across the networks through the respective user latent features learned from each network. To be specific, we propose a constrained non-negative matrix tri-factorization (NMTF) approach with a graph regularizer obtained from the user-user adjacency matrices.

We begin by considering existing NMTF approaches and later introduce our approach for link recommendation. NMTF as shown in eq (1) involves decomposing a matrix $\mathbf{X} \subset \mathbb{R}^{m \times n}$, into three non-negative latent factor matrices $\mathbf{F} \subset \mathbb{R}_+^{m \times p}$, $\mathbf{S} \subset \mathbb{R}_+^{p \times o}$ and $\mathbf{G} \subset \mathbb{R}_+^{n \times o}$ that can best approximate $\mathbf{X}$. For example, in the context of social network analysis, given the user-keyword matrix for a social network, NMTF co-clusters users and keywords into p user groups and o keyword groups.

$$\mathbf{X} = \mathbf{F}\mathbf{S}\mathbf{G}^T \tag{1}$$

Cai et al. (2011) proposed a co-clustering method called Graph based non-negative matrix factorization (GNMF) that adds a graph regularizer to NMF imposing manifold assumptions. The factors for multiple social networks can be computed individually through K subproblems as follows:

$$\begin{aligned} \min \quad & \left\|\mathbf{X}_k - \mathbf{F}_k\mathbf{G}_k^T\right\|_F^2 + \alpha_k \mathbf{tr}\Big(\mathbf{F}_k^T\mathbf{L}_k\mathbf{F}_k\Big) \\ \text{s.t.} \quad & \mathbf{F}_k \geq 0,\ \mathbf{G}_k \geq 0,\ k = 1, \ldots, K \end{aligned} \tag{2}$$

where $\mathbf{tr}(.)$ is the trace of the matrix, $\mathbf{L}_k = \mathbf{D}_k - \mathbf{A}_k$ is the graph Laplacian of user-user adjacency matrix $\mathbf{A}_k$, $\mathbf{D}_k = \sum_j \mathbf{A}_k^{ij}$ is the degree matrix, α_k is the regularization parameter on the user groups and $||.||_F^2$ is the Frobenius norm. The first term in the objective function minimizes the reconstruction error and the second term is a manifold regularizer on user-user relations which incorporates the geometric information of the data.If two users are closely connected to each other, they belong to the same group.

Gu, Ding, and Han (2011) and Huang et al. (2014) showed that when regularization parameter α_k is set to a large value GNMF ends up in a trivial solution, associating all the users to one group. Also GNMF is prone to scale transfer problems, when the parameters in the objective function multiplied by any scalar ($\gamma > 1$) results in a solution which is different from the optimal solution. To fix these two issues, Gu, Ding, and Han (2011) proposed a graph based NMTF approach (IGNMTF), with three factors and orthogonal constraints to allow more degrees of freedom between user and keyword latent factors. Huang et al. (2014) added orthogonal constraints to eq (2) to fix scale transfer problems. Similar as before, we have the following K subproblems:

$$\begin{aligned} \min \quad & \left\|\mathbf{X}_k - \mathbf{F}_k\mathbf{S}_k\mathbf{G}_k^T\right\|_F^2 - \alpha_k \mathbf{tr}\Big(\mathbf{F}_k^T\mathbf{A}_k\mathbf{F}_k\Big) \\ & - \Big(\mathbf{G}_k^T\mathbf{A}_k^{'}\mathbf{G}_k\Big) \\ \text{s.t.} \quad & \mathbf{F}_k \geq 0,\ \mathbf{S}_k \geq 0,\ \mathbf{G}_k \geq 0,\ k = 1, \ldots, K \\ & \mathbf{F}_k^T\mathbf{D}_k\mathbf{F}_k = \mathbf{I}, \mathbf{G}_k^T\mathbf{D}_k^{'}\mathbf{G}_k = \mathbf{I} \end{aligned} \tag{3}$$

where $\mathbf{A}_k^{'}$ is the keyword-keyword adjacency matrix, $\mathbf{D}_k^{'} = \sum_j \mathbf{A}_k^{'ij}$ is the degree matrix, $\mathbf{I}$ is the identity matrix of the appropriate size. The main difference between GNMF eq (2) and IGNMTF eq (3) is the orthogonal constraints, which fix both the scale transfer and trivial solution problems. Without the constraints the optimization problem in eq (2) can be seen as a special case of eq (3) by absorbing $\mathbf{S}_k$ into $\mathbf{F}_k$. Also, as shown in Nie et al. (2010) when orthonormal and non-negative constraints of $\mathbf{F}_k$ and $\mathbf{G}_k$ are simultaneously satisfied, then it can be proved that in each row of $\mathbf{F}_k$ and $\mathbf{G}_k$, only one element could be positive and others are zeros, which can be directly used to assign cluster labels to data points.

3.3 Proposed Framework

As shown in the last subsection, existing work on NMTF is designed for a single social network, and cannot be readily applied to model multiple social networks and identify similar actors. Notice that disease-specific social networks often share the same set of topics. For example, for diabetes-specific social networks, the set of topics usually include Type I diabetes, Type II diabetes, gestational diabetes, diet and exercise, etc. Furthermore, the users of these social networks tend to form the same groups with interest in certain topics. For example, on both TuDiabetes and DiabetesSisters, there are user groups associated with Type I diabetes, Type II diabetes and gestational diabetes. Based on this observation, in this subsection, we present our proposed optimization framework named CrossNet, which jointly decomposes the user-keyword matrices from multiple social networks, while requiring them to share the same topics as well as user group-topic association matrices.

$$\begin{aligned} \min \quad & \sum_{k=1}^{K} \Big\{ \left\|\mathbf{X}_k - \mathbf{F}_k\mathbf{S}\mathbf{G}^T\right\|_F^2 + \alpha_k tr\Big(\mathbf{F}_k^T\mathbf{L}_k^s\mathbf{F}_k\Big)\Big\} \\ \text{s.t.} \quad & N_F(\mathbf{F}_k), N_G(\mathbf{G}), N_S(\mathbf{S}) \\ & O_F(\mathbf{F}_k), O_G(\mathbf{G}), k = 1, \ldots, K \end{aligned} \tag{4}$$

where $\mathbf{L}_k^s = \mathbf{I} - \mathbf{D}_k^{-\frac{1}{2}}\mathbf{A}_k\mathbf{D}_k^{-\frac{1}{2}}$ is the symmetric normalized Laplacian of the user-user adjacency matrix $\mathbf{A}_k$, $N_F(\cdot)$, $N_G(\cdot)$, and $N_S(\cdot)$ denote the non-negative constraint on a certain matrix, $O_F(\cdot)$ and $O_G(\cdot)$ denote the orthogonal constraint on the input matrix. Notice that we use the symmetric normalized Laplacian as it provides more robust results as compared to the one used in eq (2).

Compared with eq (2) and eq (3), the major difference is that we couple the K subproblems by requiring them to share the same matrices $\mathbf{S}$ and $\mathbf{G}$. This is because multiple disease-specific social networks tend to share the same topics ($\mathbf{G}$) as well as the user group-topic matrix $\mathbf{S}$. Depending

on the specific form of the non-negative constraint $N(\cdot)$ and the orthogonal constraint $O(\cdot)$, CrossNet can be instantiated in four different ways as follows.
CrossNet-I:

$$\mathbf{F}_k \geq 0, \mathbf{G} \geq 0$$
$$\mathbf{F}_k^T\mathbf{F}_k = \mathbf{I}_F, \sum_j \mathbf{G}_{i,j} = 1, k = 1, \ldots, K. \quad (5)$$

CrossNet-II:

$$\mathbf{F}_k \geq 0, \mathbf{S} \geq 0, \mathbf{G} \geq 0$$
$$\mathbf{F}_k^T\mathbf{F}_k = \mathbf{I}_F, \sum_j \mathbf{G}_{i,j} = 1, k = 1, \ldots, K. \quad (6)$$

CrossNet-III:

$$\mathbf{F}_k \geq 0, \mathbf{G} \geq 0$$
$$\mathbf{F}_k^T\mathbf{D}_F\mathbf{F}_k = \mathbf{I}_F, \sum_j \mathbf{G}_{i,j} = 1, k = 1, \ldots, K. \quad (7)$$

CrossNet-IV:

$$\mathbf{F}_k \geq 0, \mathbf{S} \geq 0, \mathbf{G} \geq 0$$
$$\mathbf{F}_k^T\mathbf{D}_F\mathbf{F}_k = \mathbf{I}_F, \sum_j \mathbf{G}_{i,j} = 1, k = 1, \ldots, K. \quad (8)$$

Notice that in all four instantiations, the orthogonal constraint on $\mathbf{G}$ is designed in such a way that its row sums are equal to 1. In this way, we allow the keywords to be part of multiple keyword groups (topics) instead of a single one.

3.4 Optimization Algorithm

In this subsection we provide the optimization algorithm for CrossNet with the constraint instantiation in eq (8). The algorithm for the other instantiations can be designed in a similar way. The objective function in eq (4) that we minimize is the following sum of squared residuals:

$$f = \sum_{k=1}^{K} \Big\{ \mathbf{tr}\Big(\mathbf{X}_k^T\mathbf{X}_k - 2\mathbf{G}^T\mathbf{X}_k^T\mathbf{F}_k\mathbf{S} + \mathbf{F}_k^T\mathbf{F}_k\mathbf{S}\mathbf{G}^T\mathbf{G}\mathbf{S}^T\Big) + \alpha_k \mathbf{tr}\Big(\mathbf{F}_k^T\mathbf{L}_k^s\mathbf{F}_k\Big)\Big\}$$

Following the standard theory of constrained optimization, we introduce the following Lagrangian function where Lagrange multiplier $\mathbf{\Lambda}_k$ enforce the constraints $\mathbf{F}_k^T\mathbf{D}_k\mathbf{F}_k = \mathbf{I}$ in eq (8).

$$\mathcal{L} = \sum_{k=1}^{K} \Big\{ \mathbf{tr}\Big(\mathbf{X}_k^T\mathbf{X}_k - 2V^T\mathbf{X}_k^T\mathbf{F}_k S + \mathbf{F}_k^T\mathbf{F}_k\mathbf{S}\mathbf{G}^T\mathbf{G}\mathbf{S}^T\Big) + \alpha_k \mathbf{tr}\Big(\mathbf{F}_k^T\mathbf{L}_k^s\mathbf{F}_k\Big) + \mathbf{\Lambda}_k\Big(\mathbf{I} - \mathbf{F}_k^T\mathbf{D}_k\mathbf{F}_k\Big)\Big\} \quad (9)$$

Computing $\mathbf{F}_k$: Fixing $\mathbf{S}$ and $\mathbf{G}$, the gradient $\nabla\mathcal{L}(\mathbf{F}_k)$ is

$$\nabla\mathcal{L}(\mathbf{F}_k) = 2(\mathbf{F}_k\mathbf{S}\mathbf{G}^T\mathbf{G}\mathbf{S}^T + \alpha_k\mathbf{L}_k^s\mathbf{F}_k - \mathbf{X}_k\mathbf{G}\mathbf{S}^T - \mathbf{D}_k\mathbf{F}_k\mathbf{\Lambda}_k)$$

By the KKT complementary slackness we have $\nabla\mathcal{L}(\mathbf{F}_k)^{ij}\mathbf{F}_k^{ij} = 0$, so

$$(\mathbf{F}_k\mathbf{S}\mathbf{G}^T\mathbf{G}\mathbf{S}^T + \alpha_k\mathbf{L}_k^s\mathbf{F}_k - \mathbf{X}_k\mathbf{G}\mathbf{S}^T - \mathbf{D}_k\mathbf{F}_k\mathbf{\Lambda}_k)^{ij}\mathbf{F}_k^{ij} = 0$$

The Lagrangian multiplier $\mathbf{\Lambda}^k$ is calculated as given in the (Ding et al. 2006) by summing up across i index. That gives

$$\mathbf{\Lambda}_k = \mathbf{F}_k^T\mathbf{X}_k\mathbf{G}\mathbf{S}^T - \mathbf{S}\mathbf{G}^T\mathbf{G}\mathbf{S}^T - \alpha_k\mathbf{F}_k^T\mathbf{L}_k^s\mathbf{F}_k$$

As $\mathbf{\Lambda}_k$ has negative components, it can be expressed as a difference of two non-negative components $\mathbf{\Lambda}_k = \mathbf{\Lambda}_k^+ - \mathbf{\Lambda}_k^-$, where $\mathbf{\Lambda}_k^+ = \frac{|\mathbf{\Lambda}_k| + \mathbf{\Lambda}_k}{2}$ and $\mathbf{\Lambda}_k^- = \frac{|\mathbf{\Lambda}_k| - \mathbf{\Lambda}_k}{2}$. Substituting the non-negative components in the equation (3.4) we get

$$(\mathbf{F}_k\mathbf{S}\mathbf{G}^T\mathbf{G}\mathbf{S}^T + \alpha_k\mathbf{L}_k^s\mathbf{F}_k - \mathbf{X}_k\mathbf{G}\mathbf{S}^T - \mathbf{D}_k\mathbf{F}_k\mathbf{\Lambda}_k^+ + \mathbf{D}_k\mathbf{F}_k\mathbf{\Lambda}_k^-)^{ij}\mathbf{F}_k^{ij} = 0$$

As the constraint, $\mathbf{I} - \mathbf{F}_k^T\mathbf{D}_k\mathbf{F}_k$ is symmetric, As suggested in (Gu, Ding, and Han 2011) we have $\mathbf{tr}(\mathbf{\Lambda}_k(\mathbf{I} - \mathbf{F}_k^T\mathbf{D}_k\mathbf{F}_k)) = tr((\mathbf{I} - \mathbf{F}_k^T\mathbf{D}_k\mathbf{F}_k)\mathbf{\Lambda}_k^T)$. Therefore only symmetric part of $\mathbf{\Lambda}_k$ contributes to $\mathcal{L}$. So $\mathbf{\Lambda}_k$ should be symmetric, we use $\mathbf{\Lambda}_k^{'} = \frac{\mathbf{\Lambda}_k + \mathbf{\Lambda}_k^T}{2}$ instead of $\mathbf{\Lambda}_k$. This leads to the following update rule for calculating $\mathbf{F}_k$:

$$\mathbf{F}_k^{ij} \Leftarrow \mathbf{F}_k^{ij} \sqrt{\frac{\Big\{\mathbf{X}_k\mathbf{G}\mathbf{S}^T + \mathbf{D}_k\mathbf{F}_k\mathbf{\Lambda}_k^{'+}\Big\}^{ij}}{\Big\{\mathbf{F}_k\mathbf{S}\mathbf{G}^T\mathbf{G}\mathbf{S}^T + \alpha_k\mathbf{L}_k^s\mathbf{F}_k + \mathbf{D}_k\mathbf{F}_k\mathbf{\Lambda}_k^{'-}\Big\}^{ij}}} \quad (10)$$

Computing $\mathbf{G}$: Fixing $\mathbf{S}$ and $\mathbf{F}_k$, setting $\nabla\mathcal{L}(\mathbf{G}) = 0$ and following the similar steps in computing $\mathbf{F}_k$ we get the following update rule for $\mathbf{G}$:

$$\mathbf{G}^{ij} \Leftarrow \mathbf{G}^{ij} \sqrt{\frac{\Big\{\sum_{t=1}^{T}\mathbf{X}_k^T\mathbf{F}_k\mathbf{S}\Big\}^{ij}}{\Big\{\sum_{t=1}^{T}\mathbf{S}^T\mathbf{F}_k^T\mathbf{F}_k\mathbf{S}\mathbf{G}\Big\}^{ij}}} \quad (11)$$

The orthogonal constraint $\sum_j \mathbf{G}_{i,j} = 1$ on $\mathbf{G}$ is enforced by row normalizing the $\mathbf{G}$ factor after every iteration.
Computing $\mathbf{S}$: Fixing $\mathbf{G}$ and $\mathbf{F}_k$, setting $\nabla\mathcal{L}(\mathbf{S}) = 0$ and following the similar steps in computing $\mathbf{F}_k$ we get the following update rule for $\mathbf{S}$:

$$\mathbf{S}^{ij} \Leftarrow \mathbf{S}^{ij} \sqrt{\frac{\sum_{k=1}^{K}\Big\{\mathbf{F}_k^T\mathbf{X}_k\mathbf{G}\Big\}^{ij}}{\sum_{k=1}^{K}\Big\{\mathbf{F}_k^T\mathbf{F}_k\mathbf{S}\mathbf{G}^T\mathbf{G}\Big\}^{ij}}} \quad (12)$$

Theorem 1. *The objective function in eq* (5) *is lower-bounded, and monotonically decreasing (non-increasing) with the update rules eq* (10)*, eq* (11) *and eq* (12)*. Hence CrossNet converges.*

Proof Sketch. First of all, it is easy to see that the objective function in eq (5) is lower-bounded. Second, it consists of two terms, and it suffices to show that each of these terms is monotonically decreasing. As the second term depends on $\mathbf{U}$ only, the update functions are similar between CrossNet and general NMTF. Following the steps in (Ding et al. 2006; Gu, Ding, and Han 2011), it can be shown that the first term is monotonically decreasing under the update rules. For the second term, by introducing an auxiliary function as in (Cai et al. 2011), it can be shown that the second term is also

Algorithm 1: CrossNet Algorithm

Input: A set of user-user adjacency matrices $\{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_K\}$ constructed from user relationship graphs $\mathcal{G}_k, k = 1, \ldots, K$ and a set of user-keyword adjacency matrices $\{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_K\}$ constructed from user-keyword bipartite graphs $\mathcal{G}_k^{'}, k = 1, \ldots, K$. The regularization parameter α^k. Number of iterations t.

Output: The user latent factors $\mathbf{F}_k$ for all the disease-specific social networks $k = 1, \ldots, K$.

Initialize the factor matrices $\mathbf{F}_k$ and $\mathbf{G}$ using k-means.
for $i \leftarrow 1$ ***to*** t **do**
  Update $\mathbf{S}$ using eq (12)
  Update $\mathbf{G}$ using eq (11)
  Update $\mathbf{F}_k$ using eq (10) $\forall\ k = 1, \ldots, K$
end
Return user latent factors $\mathbf{F}_k$.

monotonically decreasing. Putting everything together, the update rules converge to the local optimal solution. Hence CrossNet converges. Details omitted due to space limit.■

With the update rules eq (10), eq (11) and eq (12) the optimization algorithm for link prediction problem is presented in the Algorithm 1.

3.5 Link Recommendation

Using NMTF we represent the users in a latent feature space shared across all the networks. For link prediction we leverage the learned shared user space along with user associations in each social network. We combine user-user associations and user-user latent features space as a graph. We use neighborhood formation using random walk with restarts (RWR) (Sun et al. 2005) to learn the cross network user-user relations. As the social networks are dynamic in nature (users join and leave over time), our approach is more robust and works for new users as we can leverage user-user associations to predict links between cross network users.

3.6 Complexity Analysis

The user-keyword matrix $\mathbf{X} \subset \mathbb{R}^{m \times n}$ is typically very sparse . Using NMTF, $\mathbf{X}$ is factorized into three latent factors as shown in eq (1). Updating $\mathbf{F}_k$, $\mathbf{S}$ and $\mathbf{G}$ using a multiplicative update algorithm takes $O(k^2(m+n))$ in each iteration for computation. And other $O(zk)$ cost for component wise addition where $z << mn$ is the number of non-zero elements in $\mathbf{X}$. Using the multiplicative algorithms for sparse computation, the efficiency of our algorithm can be improved tremendously. As the value of k is very small (usually < 100), we can consider that the algorithm is linear per computation. Empirically we found that number of iterations it takes to converge is $t < 100$. So the total cost of complexity is $O(tk^2(m+n) + tkz)$ which is still linear. So computationally, CrossNet scales to large data sets.

arXiv	# papers	# nodes	# edges
Artificial Intelligence (cs.AI)	6972	10272	31266
Computer Vision (cs.CV)	5321	10156	19284
Databases (cs.DB)	2070	4297	6492
Machine Learning (cs.LG)	7321	11103	39349
Software (cs.SE)	2753	5514	18462
diabetes	**# posts**	**# nodes**	**# edges**
Diabetes Sisters	2643	750	4118
TuDiabetes	3742	1032	6323

Table 1: Statistics of arXiv and diabetes-specific social network data sets.

4 Experimental Results

In this section we compare CrossNet with other state-of-the-art approaches on an academic publications data set. We also demonstrate the effectiveness of CrossNet through a case study on a diabetes-specific social network data set.

4.1 Data Sets

The first data set is from the online repository of electronic preprints arXiv (2016), which contains scientific papers related to artificial intelligence (cs.AI), computer vision (cs.CV), databases (cs.DB), machine learning (cs.LG) and software (cs.SE) categories in the field of computer science. Each category represents a social network with user-user associations based on the co-authorship information. Keywords are extracted from the abstract of each scientific paper. For each author (user), we combine all the abstracts from the papers authored or co-authored by the author. The ground truth for this data set is computed from the existing cross network links (authors common to different networks). The neighborhood formation algorithm based on RWR is used to estimate the cross network link associations.

We also demonstrate the applicability of CrossNet to a real world setting through a case study on diabetes-specific social networks. The user posts from two diabetes-specific social networks – TuDiabetes (2016) and DiabetesSisters (2016) are crawled. The user-user associations in the forums are missing, so we considered the users who post in any given thread as related, i.e., there exists an edge between the users responding to the same thread. Keywords are extracted from the posts. Several pre-processing steps were taken before the experiments, including stemming, stop word removal, etc. Each user is represented as a binary feature vector with bag of words with n-grams $n = \{1, 2, 3\}$. Table 1 shows the data set statistics.

4.2 Experiment Setup

We compare the proposed CrossNet approaches with other state-of-the-art approaches including: (1) GNMF (Cai et al. 2011); (2) IGNMTF (Gu, Ding, and Han 2011); (3) CoupledLP (Dong et al. 2015) modified for cross network links; and (4) COSNET (Zhang et al. 2015).

We have used Precision at K (P@K) as an evaluation metric to compare the performance of different algorithms. It

	DB - SE		DB - LG		LG - SE		AI - LG		AI - CV		CV - LG	
	P@10	P@20	P@10	P@20	P@10	P@20	P@10	P@20	P@10	P@20	P@10	P@20
GNMF	15.48	12.84	12.88	12.77	8.64	8.77	18.72	17.47	15.03	14.9	11.73	11.9
IGNMTF	26.28	18.36	18.92	16.3	24.25	18.96	33.23	30.26	22.08	19.02	32.9	25.73
CoupledLP	23.4	20.28	31.58	21.77	24.64	24.12	37.86	40.87	36.84	25.4	33.43	32.73
COSNET	31.68	29.88	35.8	26.56	31.99	28.64	**47.58**	**45.86**	41.76	30.99	43.4	38.85
CrossNet - I	35.28	30.48	36.71	32.95	33.8	31.48	44.62	42.73	42.83	38.44	45.85	42.7
CrossNet - II	**36.12**	30.59	36.94	33.29	34.06	31.61	44.77	42.32	43.09	38.84	46.2	42.88
CrossNet - III	35.28	30.36	36.59	33.17	33.93	31.48	44.77	42.54	42.69	38.7	46.03	42.7
CrossNet - IV	35.41	**30.63**	**37.05**	**33.63**	**34.19**	**31.73**	45.08	42.81	**43.23**	**39.24**	**46.38**	**43.05**

Table 2: arXiv results

topic-1 healthy eating	topic-2 insurance	topic-3 exercise	topic-4 products	topic-5 diet	topic-6 diagnosis	topic-7 research
food[12]	medical insurance[12]	running[12]	pump[12]	insulin[12]	diagnosed[12]	patients study[1]
healthy eating[12]	cost information[12]	ginger[2]	cgm[12]	dose[12]	diabetes[12]	levels[12]
carbs[12]	money[12]	training[1]	minimed[12]	carbs[12]	family doctor[12]	doctor[12]
protein[2]	insulin supplies[12]	yoga[12]	infusion pumps[1]	low carb[12]	hospital[12]	ADA [1]
veggies[1]	strips[2]	gym[12]	insulin use[12]	high day[2]	symptoms[12]	people[12]
bread[12]	companies[12]	workout[12]	omnipod[12]	bg[12]	months[12]	clinical treatment[1]
diet[12]	doctors[12]	muscle[2]	pumping set[12]	basal hours[12]	told diabetic[12]	disease research[2]

Table 3: Diabetes keyword groups (top 7). [1] represents keywords from Diabetes Sisters, [2] from TuDiabetes and [12] from both.

computes the percentage of the relevant links among the top-K links predicted by the algorithm. For evaluation we compute P@10 and P@20 for all the algorithms and data set combinations. Here relevant links refer to the links between similar actors across the networks.

Regarding the parameters, we use grid-search to set regularization parameters $\alpha_1 = \alpha_2 = 0.01$ for CrossNet, the number of user groups and keyword groups $o = p = 40$ and iterations $t = 100$. From the results in Table 2 CrossNet outperforms all other approaches. Jointly factorizing keywords across all the networks through $\mathbf{G}$ resulted in significant improvement over GNMF and IGNMF approaches. CrossNet outperformed modified CoupledLP as it uses both the user-user associations and user-keyword bipartite graphs unlike CoupledLP that relies on user-user network structure only. COSNET performs closely as it leverages both the user-user and user-keyword graphs, but it identifies the distinct user-user links across networks to the similar ones. Among the four constraint instantiations, setting $\mathbf{S} \geq 0$ and orthogonal constraint with degree matrix led to a better performance.

4.3 Case Study

We also conduct a case study on diabetes-specific social networks. Notice that CrossNet has two steps: (1) jointly decomposing the user-keyword matrices from each network into respective user factors and a combined keyword latent factor matrix; (2) using RWR on user-user associations and user factor matrices for each network to recommend links between similar actors across different networks. Table 3 shows the keyword latent factors from all the networks combined ($K = 2, p = 7$). It can be observed that our joint factorization approach clustered similar keywords from different networks into one group. The following is an example of two posts generated by two users from different social networks, between whom CrossNet recommends a link.

***User A**: I have been diagnosed with Type 1 for about 5 years. I had my blood glucose with an A1C over 9. I am worried!*
***User B**: I am a 22 year old female recently diagnosed type 1 diabetic. I found out that my blood glucose was over 400. I came here looking for support.*

As we can see, both users are concerned about their blood glucose level and have been diagnosed with Type I diabetes.

5 Conclusion

In this paper, motivated by the use of disease-specific social networks, we studied the problem of cross network link recommendation, where we aim to identify similar patients across multiple heterogeneous networks, such that they can form support groups to exchange information and resources. This is different from existing work on cross network link prediction where the goal is to link accounts belonging to the same user from different social networks or to find users with complementary expertise or interests. To address this problem, we propose an optimization framework named CrossNet with four instantiations, which can be solved using an iterative algorithm. The performance of the proposed algorithm is evaluated both analytically in terms of convergence and computational complexity, and empirically on various real data sets.

Acknowledgements

This work is supported by the NSF research grant IIS-1552654, ONR research grant N00014-15-1-2821, and an IBM Faculty Award. The views and conclusions are those of the authors and should not be interpreted as representing the official policies of the funding agencies or the government.

References

Al Hasan, M., and Zaki, M. J. 2011. A survey of link prediction in social networks. In *Social Network Data Analytics*. Springer US. 243–275.

Al Hasan, M.; Chaoji, V.; Salem, S.; and Zaki, M. 2006. Link prediction using supervised learning. In *SDM06: workshop on link analysis, counter-terrorism and security*.

arXiv. 2016. arxiv.org e-print archive. [Online; accessed 29-November-2016].

Cai, D.; He, X.; Han, J.; and Huang, T. S. 2011. Graph regularized nonnegative matrix factorization for data representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 33(8):1548–1560.

Chakraborty, Y. P. N., and Sycara, K. 2015. Nonnegative matrix tri-factorization with graph regularization for community detection in social networks.

DiabetesSisters. 2016. Diabetessisters. [Online; accessed 29-November-2016].

Ding, C.; Li, T.; Peng, W.; and Park, H. 2006. Orthogonal nonnegative matrix t-factorizations for clustering. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '06, 126–135. New York, NY, USA: ACM.

Dong, Y.; Zhang, J.; Tang, J.; Chawla, N. V.; and Wang, B. 2015. CoupledLP: Link prediction in coupled networks. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, 199–208. New York, NY, USA: ACM.

Fortunato, S. 2010. Community detection in graphs. *Physics reports* 486(3):75–174.

Gu, Q.; Ding, C.; and Han, J. 2011. On trivial solution and scale transfer problems in graph regularized nmf. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, volume 22, 1288. people.virginia.edu.

Huang, J.; Nie, F.; Huang, H.; and Ding, C. 2014. Robust manifold nonnegative matrix factorization. *ACM Trans. Knowl. Discov. Data* 8(3):11:1–11:21.

Kong, X.; Zhang, J.; and Yu, P. S. 2013. Inferring anchor links across multiple heterogeneous social networks. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, 179–188. ACM.

Li, T., and Ding, C. 2006. The relationships among various nonnegative matrix factorization methods for clustering. In *Data Mining, 2006. ICDM'06. Sixth International Conference on*, 362–371. IEEE.

Liben-Nowell, D., and Kleinberg, J. 2007. The link-prediction problem for social networks. *J. Am. Soc. Inf. Sci.* 58(7):1019–1031.

Nie, F.; Ding, C.; Luo, D.; and Huang, H. 2010. Improved minmax cut graph clustering with nonnegative relaxation. In *Proceedings of the 2010 European Conference on Machine Learning and Knowledge Discovery in Databases: Part II*, ECML PKDD'10, 451–466. Berlin, Heidelberg: Springer-Verlag.

Sun, J.; Qu, H.; Chakrabarti, D.; and Faloutsos, C. 2005. Neighborhood formation and anomaly detection in bipartite graphs. In *Fifth IEEE International Conference on Data Mining (ICDM'05)*, 8 pp.–.

Tang, J.; Wu, S.; Sun, J.; and Su, H. 2012. Cross-domain collaboration recommendation. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, 1285–1293. New York, NY, USA: ACM.

TuDiabetes. 2016. A community of people touched by diabetes, a program of the diabetes hands foundation. [Online; accessed 29-November-2016].

Wang, P.; Xu, B.; Wu, Y.; and Zhou, X. 2014. Link prediction in social networks: the State-of-the-Art.

Wang, H.; Nie, F.; and Huang, H. 2015. Large-scale cross-language web page classification via dual knowledge transfer using fast nonnegative matrix trifactorization. *ACM Trans. Knowl. Discov. Data* 10(1):1:1–1:29.

Zhang, Y.; Tang, J.; Yang, Z.; Pei, J.; and Yu, P. S. 2015. COSNET: Connecting heterogeneous social networks with local and global consistency. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, 1485–1494. New York, NY, USA: ACM.