User Alignment with Jumping Seed Alignment Information Propagation

Xiang Li[1,2,3], Yijun Su[2,3], Neng Gao[1,3], Ji Xiang[3] *, Yuewu Wang[1,2,3]

[1]*State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences*
[2]*School of Cyber Security, University of Chinese Academy of Sciences*
[3]*Institute of Information Engineering, Chinese Academy of Sciences*
Beijing, China
{lixiang9015, suyijun, gaoneng, xiangji, wangyuewu} @iie.ac.cn

Abstract—User Alignment is to find users belonging to a same real person on different social networks and has become a fundamental task for many sequent applications such as cross-network recommendation systems. When matching users in multiple social networks, existing approaches always know some correctly matched users, which can be called seeds. Then, existing methods strongly depend on the neighboring users of each user to propagate alignment information from seeds and align probable matching users implicitly. However, the completeness and validity of original alignment information among seeds cannot be fully preserved when learning and aligning multiple user spaces. In this paper, we propose a unified framework named Jumping Seed Alignment Information Propagation (JSAIP) to flexibly leverage, for each user, complete and correct alignment information from seeds. Specifically, JSAIP learns a reasonable user space for each social network by preserving enough original network and label information. Then, JSAIP ensures the correct alignment among seeds and shared labels to reduce the diversity between different user spaces. Finally, JSAIP constructs jumping links from seeds to each user in each social network and ultilizes original seed alignment information to enhance or rectify the alignment information propagated from neighbors. Experiments on real world datasets demonstrate the effectiveness of our proposed JSAIP method compared to several state-of-the-art methods.

Index Terms—social network, user alignment, network integration

I. Introduction

Nowadays, large amounts of social networks have grown quickly. People have many choices to use different social network for different personal demands. Massive user data promotes the development of sequent cross-network applications such as recommendation [1]–[4], link prediction [5]–[7] and topic analysis [8]. For example, cold-start problem is one of the main challenges for recommendation system. When the users in different social networks can be aligned, for new users in the second social network, their explicit user information in the first social network can provide enough useful information and greatly alleviate the cold-start problem in the second social network. Therefore, user alignment is a fundamental and meaningful task for common data mining fields.

User Alignment is to align, for a real person, different user accounts in different social networks and has attracted much
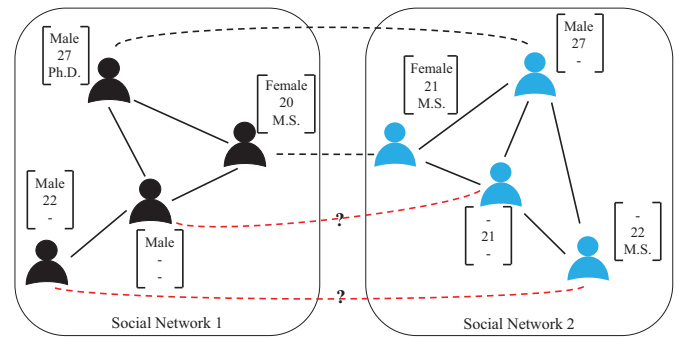
* Corresponding author.

Fig. 1. An example of user alignment task. The black dash line means the connected two users belongs to a same real person. Users linked by red dash line are the user pair to be identified.

attention. As shown in Fig. 1, besides label information and network information, researchers also know some correctly matched users (i.e., seeds) by some ways. Based on known above information, the main aim is to identify whether two users belong to the same real person. Early work focus on feature engineering to construct useful features based on user profiles or local structure and use seeds to train the prediction model to align probable matched users [9]–[13].

For better practical application, existing methods map all users in the same social network into a same user space by preserving useful original user profiles or local structure [14]–[16]. Moreover, these user space methods integrate the alignment information from seeds with the process of learning user spaces in an iterative manner to reduce the diversity between multiple user spaces. For each iteration, user space methods firstly learn the user space for each social network from user profiles and local structure under the guidance of probable aligned users in previous iteration. Then, these methods align the most probable matched users implicitly by learned user representations in the first step. Finally, above two steps iteratively run until end. Briefly, existing user space methods try to propagate the seed alignment information from seed users to other users in order to fuse it with preserved label and structure information. Therefore, making the seed alignment information complete and valid is a crucial challenge for user space methods.

However, when the number of seeds is very small compared to the total users in all social networks, existing work cannot guarantee the alignment information from seeds can be completely and effectively propagated to each user. Especially, for users far away from seeds, only label and structure information can be utilized for alignment. Besides, when some label information of some users are missing, the diversity between different user spaces cannot be effectively reduced and the final alignment performance will sharply drop.

In this paper, we propose a unified framework called **J**umping **S**eed **A**lignment **I**nformation **P**ropagation (JSAIP) method to flexibly leverage, for each user, complete and correct alignment information from seeds. Firstly, when learning user spaces, JSAIP preserves partial network information and label information. Besides, considering missing labels, JSAIP implicitly predicts the missing labels. Then, to reduce the diversity between different user spaces, JSAIP makes full use of seeds to align multiple user spaces. Finally, the key of JSAIP is to construct the jumping links for seeds and other users to directly propagate seed alignment information to enhance or rectify the neigboring alignment information and preserved original information.

Our contributions can be summarized as follows:

- We detaily analyze the effect of different information and the crucial role of seed alignment information from the local and global view. Then, we propose the challenge of effective seed alignment information propagation, which is very important but not valued by existing methods.
- We propose the JSAIP method to solve the challenge of seed alignment information propagation. The key of JSAIP is to build jumping links to connect each user with seed users. Besides, we formally represent our method as an optimization problem and design a fast and efficient optimization algorithm to solve the problem.
- We conduct extensive experiments on real-world datasets to evaluate the performance on aligning users across multiple social networks. The results demonstrate its superiority over the state-of-the-art methods, with signifcant improvement on several datasets.

The rest of this paper is organized as follows: We review related work in Section II. Section III presents specifics of our proposed JSAIP approach and the proposed optimization method is showed in Section IV. Experimental evaluation and comparison are shown in Section V. At last, Section VI concludes the paper with a brief discussion.

II. Related Work

In this section, we review the main works on user alignment. Firstly, we briefly introduce traditional methods. Then, we discuss the progress of user space methods.

Traditional methods have paid much attention to extract useful featuers from user profiles and content. The first work on user alignment uses usernames [9]–[11]. They study the behavior patterns during selecting usernames and construct totally more than four handreds features. Then, based on seed users, some common classifiers were used to predict matched users. Besides, spatio-temporal information has been deeply studied for extracting useful features [12], [13] such as spatio distribution and temporal distribution. For content information, topic distribution has been demonstrated the effect [17].

Considering the cost of artificial features, user space methods have attracted much attention. Different user space methods have been proposed for different types of information. With the rapid development of network representation learning [18]–[20], methods based on network information have made great progress. For example, PALE [14] preserves local structure in user space for each social network independently. Then, PALE learns the linear/non-linear mapping by using seed users as training data. IONE [15] considers the directed followee/follower relationship and learns three representations for each user. Seed users were used in IONE only for aligning the vectors of seed users. DeepLink [16] introduces the deep neural network based on the learned users' representations by random walk. To integrate network information and label information, MAH [21] firstly captures the high-order structure formed by using neighbor users and labels to construct hypergraph. Based on MAH, UMAH emphasizes the effect of shared labels among different social networks and automatically learns the weights of different types of labels [22]. Besides, a lot of methods also utilize the local structure and label information. For example, MASTER [23] firstly computes the similarity matrices and learns the user space by matrix factorization with kernel tricks. MEgo2Vec [24] uses deep learning techniques to capture user's profiles and ego network in the user space. Recently, OURLACER [25] is the first work to model the missing labels by implicitly predicting missing information.

When learning the user space for each social network, an inevitable challenge is how to reduce the diversity between vector spaces by seed users. Existing user space methods mainly utilize the consistence between seed users [14]–[16], [21]–[25], which means the alignment information cannot be propagated or only propagated to neighboring users. However, compared the large amount of users in each social network, the alignment information propagated from neighboring users cannot take effect for users far away from seeds. Besides, the error occurring in propagated alignment information has great impact on sequent propagation and alignment.

In this paper, we propose JSAIP method to solve the propagation of seed alignment information. Similar to OURLACER, JSAIP is able to implicitly complete missing label information. Furthermore, JSAIP can propagate correct alignment information for each user in each social network by building jumping links, which is the first work on deeply investigating the effect of propagatiing alignment information.

III. Proposed Method

In this section, we firstly introduce some basic notations and notions. Then, we present the JSAIP method from three modules: preserving original information, aliging multiple user spaces and building jumping links. Finally, we give the optimization objective of our proposed JSAIP method.

For the i-th social network, we represent it as $G_i = (V_i, A_i, L_i)$, where V_i is the set of users in the i-th social network and A_i is the adjacency matrix constructed by edges in the i-th social network. Besides, $L_i \in R^{n_i \times d_i}$ denotes the user labels and n_i is the total number of all users in the i-th social network. For seed users, we extract the subset of the edges connected to seed users and represent the adjacency matrix as $H_i \in \{0, 1\}^{s \times n_i}$, where s is the number of seed users. For each row in H_i, we set value of the j-th column to one if the j-th user in the i-th social network is connected to this seed user. The aim of user alignment using user space is to learn the vector representation matrix $U_i \in R^{n_i \times d}$ for the i-th social network. Due to the emergence of label information, we also learn the vector representation $Q_i \in R^{d_i \times d}$ for labels in the i-th social network.

A. Preserving Original Information

In real life, people always use multiple social networks at the same time. Though people use different social network for different purposes, for a real person, his behaviors in different social network are similar. For example, the social circle in one social network will be partially same as it in other social network. Hence, we should preserve the original network information in the social network. Formally, we expect the learned vector representations can reconstruct the original network, which means the vectors of users connected should be similar. From the global view, we can express it in a form of matrix:

$$\min_{U_i} \quad \frac{\alpha_i}{2}||A_i - U_i U_i^T||_F^2 + \frac{\beta}{2}||U_i||_F^2 , \tag{1}$$

where α_i is the parameter to control the amount of preserved network information and β is to control the complexity of U_i. $||\cdot||_F$ stands for Frobenius norm.

Considering the label information, we also need to preserve the label information. A basic idea is to reconstruct the existing labels. However, as shown in Fig. 1, we often cannot get the full label information because of privacy demand. If we still reconstruct original label information exactly, the missing labels will be mistakenly recovered because the value 0 in labe matrix L_i is seen as no labels rather than missing labels. Therefore, in this paper, we use collective matrix factorization to implicitly predict missing labels:

$$\min_{U_i, Q_i} \quad \frac{\alpha_i}{2}||A_i - U_i U_i^T||_F^2 + \frac{\beta}{2}(||U_i||_F^2 + ||Q_i||_F^2)$$
$$+ \frac{1}{2}||I_i^L \odot (L_i - U_i Q_i^T)||_F^2 , \tag{2}$$

where $\odot$ is the Hadamard (element-wise) product and I_i^L is an indicator matrix. $I_i^L(p, q) = 0$ if $L_i(p, q)$ is missing, and otherwise $I_i^L(p, q) = 1$. Considering the value in A_i and L_i,

we limit the value in U_i and Q_i to range from 0 to 1. Then, the optimization problem can be formulated as

$$\min_{U_i, Q_i} \quad \frac{\alpha_i}{2}||A_i - U_i U_i^T||_F^2 + \frac{\beta}{2}(||U_i||_F^2 + ||Q_i||_F^2)$$
$$+ \frac{1}{2}||I_i^L \odot (L_i - U_i Q_i^T)||_F^2 \tag{3}$$
$$s.t. \quad 0 \leq U_i \leq 1 ,$$
$$0 \leq Q_i \leq 1 .$$

Through above optimization problem, we can learn the user space preserving enough original information for each social network.
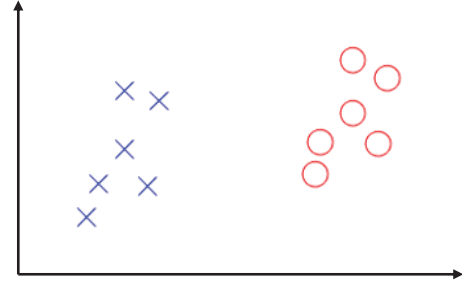
B. Aligning Multiple User Spaces

Fig. 2. An example of learned user spaces. Independent learning may lead to above non-overlapping situation.

For the problem of user alignment, if we independently learn the user space for each social network, we might get the result as shown in Fig. 2. Therefore, we should make full use of owning information to make the diversity between different social networks as small as possible.

To make multiple user spaces correctly overlapped, we firstly make seed users play the role of anchor, which can be written as

$$\min_{U_i, Q_i} \quad \sum_i (\frac{\alpha_i}{2}||A_i - U_i U_i^T||_F^2 + \frac{\beta}{2}(||U_i||_F^2 + ||Q_i||_F^2)$$
$$+ \frac{1}{2}||I_i^L \odot (L_i - U_i Q_i^T)||_F^2)$$
$$+ \frac{\gamma}{2}||T_1 U_1 - T_2 U_2||_F^2 \tag{4}$$
$$s.t. \quad 0 \leq U_i \leq 1 ,$$
$$0 \leq Q_i \leq 1 ,$$

where $T_i \in \{0, 1\}^{s \times n_i}$ is used to extract seed user representations in the i-th social network.

Besides seed users, different social networks may have same labels such as age and sex. These shared labels in multiple social networks can also make different user spaces more
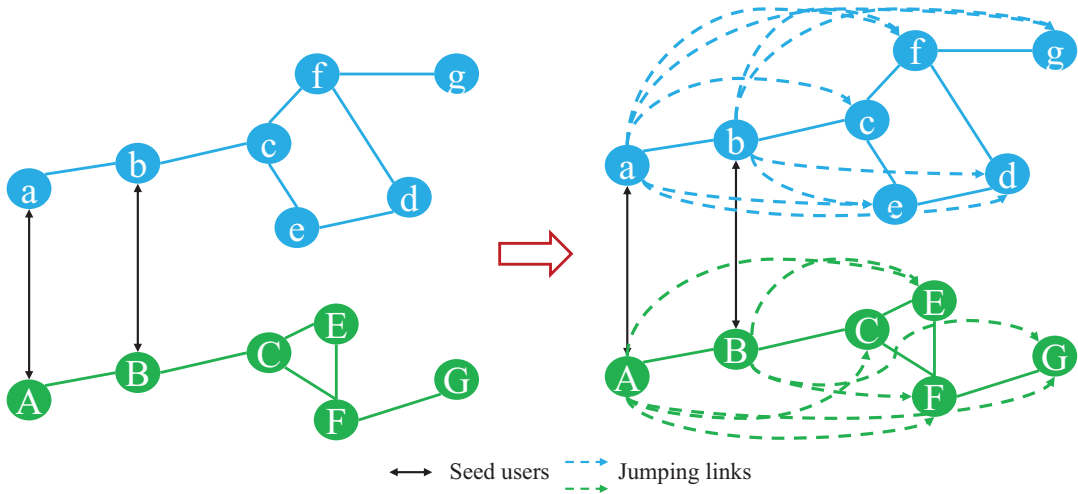
Fig. 3. An example of building jumping links. The left is original social networks and the right is the recovered social networks with jumping links.

overlapped. Then, the optimization problem can be rewritten as

$$\min_{U_i, Q_i} \quad \sum_i (\frac{\alpha_i}{2}||A_i - U_i U_i^T||_F^2 + \frac{\beta}{2}(||U_i||_F^2 + ||Q_i||_F^2)$$
$$+ \frac{1}{2}||I_i^L \odot (L_i - U_i Q_i^T)||_F^2)$$
$$+ \frac{\gamma}{2}(\sum_{i,j,i\neq j} ||T_i U_i - T_j U_j||_F^2 + ||M_i Q_i - M_j Q_j||_F^2)$$
$$s.t. \quad 0 \leq U_i \leq 1 ,$$
$$0 \leq Q_i \leq 1 ,$$

(5)

where M_i is to used to extract shared label representations in the i-th social network.

C. Building Jumping Links

To learn and align multiple user spaces, existing methods usually use iterative optimization methods to solve the optimization problem. Then, the alignment information can be propagated from user to user by iteration. In essence, alignment information means the correct overlapping information between users. If each pair of true matched users can be overlapped correctly in the user space in each iteration, the whole user spaces will be totally overlapped after enough iterations.

However, the alignment information propagated from neighboring users may be not complete and correct. For wrong alignment information, it will destroy the overlapping situation. For user far away from seeds, incomplete alignment information cannot take effect.

Therefore, to address the challenge of seed alignment information propagation, we build the jumping links as shown in Fig. 3. Through jumping links, seed alignment information can be directly propagated to each user. Then, each user can use complete and valid alignment information to enhance or rectify the alignment information from neighoring users. Formally, we

firstly need to preserve existing seed alignment information by following optimization objective

$$\min_{U_i, P_i} \quad ||H_i - U_i P_i^T||_F^2 ,$$

(6)

where $P_i \in R^{s \times d}$ represents the vector representations of seed users in the i-th social network. Objective (6) can preserve original alignment information for the neighboring users of seed users. To construct the jumpling links, we use collective matrix factorization to predict the links between seed users and other users. Then, the optimization objective can be reformulated as

$$\min_{U_i, P_i} \quad ||H_i \odot (H_i - U_i P_i^T)||_F^2 .$$

(7)

D. Final Optimization Objective

As shown in previous content, we have presented three important modules of proposed JSAIP method. The first module is to preserve the local structure and label information for each social network. Besides, we consider the missing labels and implicitly predict the missing labels. The second module is to align different user spaces by aligning seed users and shared labels in different user spaces. The third module is to build jumping links to propagate complete and valid alignment information to each user and enhance the alignment performance. By integrating the optimization objectives of above

three modules, we can get following optimization problem:

$$\min_{U_i,Q_i} \quad \sum_i (\frac{\alpha_i}{2}||A_i - U_iU_i^T||_F^2 + \frac{\beta}{2}(||U_i||_F^2 + ||Q_i||_F^2)$$
$$+ \frac{1}{2}||I_i^L \odot (L_i - U_iQ_i^T)||_F^2$$
$$+ \frac{1}{2}||H_i \odot (H_i - U_iP_i^T)||_F^2)$$
$$+ \frac{\gamma}{2}(\sum_{i,j,i\neq j} ||T_iU_i - T_jU_j||_F^2 + ||M_iQ_i - M_jQ_j||_F^2)$$
$$s.t. \quad 0 \leq U_i \leq 1 ,$$
$$0 \leq Q_i \leq 1 .$$
$$(8)$$

We can find this optimization objective cannot restrain P_i. Therefore, we revise the objective of the second module and the final optimization problem can be written as

$$\min_{U_i,Q_i} \quad \sum_i [\frac{\alpha_i}{2}||A_i - U_iU_i^T||_F^2 + \frac{1}{2}||I_i^L \odot (L_i - U_iQ_i^T)||_F^2$$
$$+ \frac{\beta}{2}(||U_i||_F^2 + ||Q_i||_F^2)$$
$$+ \frac{1}{2}||H_i \odot (H_i - U_iP_i^T)||_F^2]$$
$$+ \frac{\gamma}{2}(\sum_i ||T_iU_i - P_i||_F^2$$
$$+ \sum_{i,j,i\neq j} ||P_i - P_j||_F^2 + ||M_iQ_i - M_jQ_j||_F^2)$$
$$s.t. \quad 0 \leq U_i \leq 1 ,$$
$$0 \leq Q_i \leq 1 .$$
$$(9)$$

IV. OPTIMIZATION

In this section, we present the optimization algorithm to solve optimization problem (9). It is hard to get the optimal solution due to the nonconvexity of optimization objective (9). Therefore, we apply two common optimization methods to optimize (9). The first is alternative optimization, which means we optimize U_i, Q_i and P_i in an alternative way. The second is stochastic gradient method with multiplicative updating rules, which can ensure the nonnegativity of U_i, Q_i and P_i. Our final algorithm is as follows.

a) Optimize U_i: For each social network, U_i is the key variable. The gradient related to U_i is

$$\frac{\partial O}{\partial U_i} = \alpha_i(U_iU_i^T - A_i)U_i + I_i^L \odot (U_iQ_i^T - L_i)Q_i$$
$$+ \beta U_i + H_i \odot (U_iP_i^T - H_i)P_i + \gamma T_i^T(T_iU_i - P_i)$$
$$(11)$$

Then, the updating rule is shown in (10).

b) Optimize Q_i: The partial derivate of objective (9) related to Q_i is

$$\frac{\partial O}{\partial Q_i} = (I_i^L)^T \odot (Q_iU_i^T - L_i^T)U_i + \beta Q_i$$
$$+ \gamma \sum_{j,j\neq i} M_i^T(M_iQ_i - M_jQ_j)$$
$$(12)$$

Then, the updating rule is shown in (10).

c) Optimize P_i: The partial derivate of objective (9) related to P_i is

$$\frac{\partial O}{\partial P_i} = H_i^T \odot (P_iU_i^T - H_i^T)U_i$$
$$+ \gamma(P_i - T_iU_i) + \gamma \sum_{j,j\neq i} (P_i - P_j)$$
$$(13)$$

The updating rule is shown in (10).

At the beginning of optimization algorithm, we should initialize U_i, Q_i and P_i with random nonnegative values. Then, we iteratively optimize U_i, Q_i and P_i by updating rules (10). Noting that the the updating rules cannot make the value smaller than 1. Hence, we utilize the projection technique [26], [27] to project elements greater than 1 in U_i, Q_i and P_i to 1 after each update process. When the change of the value of objective (9) is smaller than predefined threshold such as 10^{-4}, we stop the iteration and use result U_i to predict probable matched users.

V. EXPERIMENT

In this section, we evaluate the performance compared to several state-of-the-art methods and study the performance of our proposed JSAIP method.

A. Compared Methods

The main compared methods used in experiments include:

- Global Method (GM) [11]: By constructing spectral embedding of each user, this algorithm learns a linear transformation by seed users and this method can be seen as a basic version of PALE [14].
- MAH [21]: By constructing hypergraphs based on labels and edges, each user has a vector representation and seed users in different social networks own totally same vector representations.
- UMAH [22]: Based on MAH, this method automatically learns the weights of different types of shared labels.
- OURLACER [25]: This method can learn the vector repesentation for each user under user and label consistence contraints.
- JSAIP: Our proposed JSAIP method considers the important effect of seed alignment information and provides useful paths for propagation.

B. Datasets

We use two real-world datasets to evaluate the performance: (1) Twitter *vs.* BlogCatalog: This dataset is provided by [22] and contains 2710 aligned users in both networks. For each user, this dataset has friendship and location information. For location information, 6.38% users do not reveal their location information in both networks and 31.03% only publish location information in one network. In the remaining users (62.59%), only 14.39% users input exactly the same location information in the two networks. (2) DBLP 2015 *vs.* 2016: This dataset is provided by [25]. This dataset contains two co-author networks. These two networks are the part of the DBLP 2015 and 2016. Besides, the conferences/journals published

$$U_i = U_i \odot \sqrt{\frac{\alpha_i A_i U_i + I_i^L \odot L_i Q_i + H_i P_i + \gamma T_i^T P_i}{\alpha_i U_i U_i^T U_i + (I_i^L \odot U_i Q_i^T) Q_i + \beta U_i + (H_i \odot U_i P_i^T) P_i + \gamma T_i^T T_i U_i}}$$

$$Q_i = Q_i \odot \sqrt{\frac{L_i^T U_i + \gamma \sum_{j,j \neq i} M_i^T M_j Q_j}{(I_i^L)^T \odot Q_i U_i^T U_i + \beta Q_i + \gamma \sum_{j,j \neq i} M_i^T M_i Q_i}} \qquad (10)$$

$$P_i = P_i \odot \sqrt{\frac{H_i^T U_i + \gamma T_i U_i + \gamma \sum_{j,j \neq i} P_j}{(H_i^T \odot P_i U_i^T) U_i + 2\gamma P_i}}$$

at least once in one year are used as the labels of users in that year. Finally, there are 2845 users in 2015, 3234 users in 2016 and 2169 aligned users between two networks. For label information, user in 2015 and 2016 respectively owns 882 and 1005 unique labels. Except unique labels, the number of shared labels is 945.

C. Experimental Settings

1) Performance Metric: To evaluate the performance of comparison methods, we utilize *Accuracy* and *Hit Precision@k* to evaluate the exact prediction and top-k prediction [22]. Specially, *Hit Precision@k* allocates the different weights for different rank k:

$$h(x) = \frac{k - (hit(x) - 1)}{k} \ ,$$

where $hit(x)$ is the position of correct matched user in the returned top-k candidate users. Then, *Hit Precision@k* can be computed on N test users by $\frac{\sum_i^N h(x_i)}{N}$.

2) Experiment Setups: UMAH and OURLACER have provided their source codes. For MAH, we implement it by matlab according to original paper and the implement of UMAH. We also implement GM by matlab. According to the settings in UMAH, we set 30% matched users as seeds and other users as test data. Considering existing study on the effect of dimension, we set the dimension of user representation to a big value such as 500, which can achieve best performance for most compared methods. When setting the parameters of our method, we set γ to a bigger value such as 10 to make loss of the diversity between different user spaces as small as possible. For parameter α_i, we set it to a same value if two datasets have similar number of users. For parameters of other compared methods, we set them to reasonable values according to original papers. Considering the effect of different order of networks for alignment, we use mean value of above metrics. For example, for users in the first network, we predict the probable aligned users in the second network and compute the metrics. Then, we predict probable aligned users in the first network for users in the second network. Finally, we report the average metrics in the experiment results. The training data and testing data are totally same for all methods.

D. Experiment Results

1) Overall Prediction Performance: We evaluate the overall prediction performance for different methods. As shown in Table I, our proposed JSAIP always behaves best than other methods. Among all methods, only GM have not used label information and achieve worst performance, which shows label information has great effect. Among methods using label information, only MAH doesn't consider the shared lables and also behaves worse than UMAH, OURLACER and JSAIP, which demonstrate the potential good effect of shared labels for reducing the diversity of different user spaces. Furthermore, by comparing UMAH, OURLACER and JSAIP, we can find modeling the shared labels in two networks simultaneously is better than modeling labels separately. Finally, JSAIP is still much better than OURLACER, which means the great effects of constructing jumping links to directly propagate original seed alignment information.

TABLE I
OVERALL PREDICTION PERFORMANCE ON TWO DATASETS

Metric	Method	DBLP 15-16	Twitter-BlogCatalog
Accuracy	GM	2.93	3.54
	MAH	5.40	9.39
	UMAH	6.45	47.27
	OURLACER	11.59	52.63
	JSAIP	**13.93**	**55.93**
Hit Prec@5	GM	5.40	6.72
	MAH	12.06	17.40
	UMAH	11.55	52.13
	OURLACER	19.61	55.88
	JSAIP	**23.07**	**59.78**

2) Visualization of User Spaces: The visualization of user spaces in OURLACER [25] has shown the best overlap than other methods. Therefore, we visualize the user spaces learned by OURLACER and our proposed JSAIP method. We plot all matched users in different social networks and use t-sne method to reduce the dimension. As shown in Fig 5, both OURLACER and JSAIP can achieve good overlap, which shows the importance of local structure information and label information with seed alignment. Besides, both OURLACER and JSAIP tend to learn clustered representations due to the use of collective matrix factorization. However, compared with OURLACER, the user space learned by JSAIP has more fine-grained clusters, which means, for a same user, JSAIP method can provide more information. Hence, the jumping links are demonstrated to be able to provide more alignment information and important to promote the ratio of overlap. Briefly, JSAIP method can learn better aligned user spaces

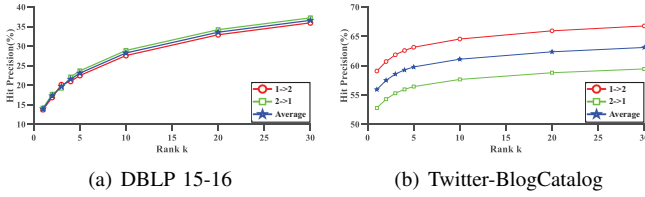and user representations than OURLACER due to jumping links.

Fig. 4. The performance of our proposed JSAIP method with different Rank k on datasets.

3) Effect of Different Rank k: During evaluation, we set the rank k to 5. However, in practical applications, we may have different setting on the rank k. Besides, different order for evaluation can lead to different performance. We use '1→2' to represent that we predict the aligned users for user in the first social network. Similarly, '2→1' means we predict the aligned users for user in the second social network. 'Average' refers to the average result of '1→2' and '2→1'. As shown in Fig. 4, we vary the rank from 1 to 30 and report the performance metric *Hit Precision@k*. We can find that the performances are approximately equal on dataset DBLP 15-16 while behaves discrepant on dataset Twitter-BlogCatalog. In general, JSAIP can achieve better performance with the rise of rank k, which shows the effectiveness of our proposed JSAIP method.

4) Effect of Dimension d: Dimension of user space is a significant parameter for practical applications. We study the influence of different dimension for our proposed JSAIP method. As shown in Fig. 7, when the dimension is smaller than 100, the performance metrics *Accuracy* and *Hit Precision@5* rise quickly. Afterwards, we can find the performance rises slowly. Generally, bigger dimension means more running time and complexity. Our proposed JSAIP method can achieve enough performance even the dimension is only 100, which shows good practicality of JSAIP method.

5) Empirical Analysis of Convergence: In the optimization problem of JSAIP method, we use stochastic gradient method, which cannot guarantee the solution is globally optimal. Therefore, we investigate the convergence empirically. We record the performance metrics *Accuracy* and *Hit Precision@5* after each iteration. Besides, we also record the absolute value of the change of final loss. As shown in Fig. 6, we plot the $|\Delta\text{loss}|$ and performance after each iteration. We can find the loss decreases quickly. After around 150 iterations, the final loss keeps stable and the change tends to be zero, which show the JSAIP method can run quickly and be easy to be used in practical scenario. From the view of performance metrics, the performances increases with increasing iterations and at the top after 400 iterations on dataset DBLP 15-16. For dataset Twitter-BlogCatalog, the performances rise sharply and achieve the top only after about 30 iterations. Briefly, our proposed JSAIP method can achieve good performance only after a few iterations. Besides, our optimization algorithm can converge fast and the performance can rises continuously,

which demonstrates the high efficiency of our proposed JSAIP method.

VI. Conclusion

Existing user space methods on user alignment have paid much attention on extracting more useful features automatically while ignore the fundamental effect of alignment information propagation. The complete and valid propagation of alignment information is a great challenge for user alignment. In this paper, we propose a unified framework called **J**umping **S**eed **A**lignment **I**nformation **P**ropagation (JSAIP) method to flexibly leverage, for each user, complete and correct alignment information from seeds. JSAIP method constructs jumping links from seeds to directly propagate seed alignment information to each user and contains three modules: preserving original information, aliging multiple user spaces and building jumping links. We formalize an optimization problem and design an optimization algorithm to effectively solve the optimization problem. Experiments show the superiority of our proposed JSAIP method compared to several state-of-the-art methods. Future work is to find more powerful seed users based on JSAIP method and deeply study the impact of jumping links.

References

[1] M. Yan, J. Sang, and C. Xu, "Mining cross-network association for youtube video promotion," in *Proceedings of the 22nd ACM International Conference on Multimedia*, 2014, pp. 557–566.

[2] M. Yan, J. Sang, C. Xu, and M. S. Hossain, "A unified video recommendation by cross-network user modeling," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 12, pp. 53:1–53:24, 2016.

[3] G. Hu, Y. Zhang, and Q. Yang, "Conet: Collaborative cross networks for cross-domain recommendation," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 2018, pp. 667–676.

[4] D. Perera and R. Zimmermann, "LSTM networks for online cross-network recommendations," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018, pp. 3825–3833.

[5] J. Zhang, X. Kong, and P. S. Yu, "Predicting social links for new users across aligned heterogeneous social networks," in *Proceedings of the 13th IEEE International Conference on Data Mining*, 2013, pp. 1289–1294.

[6] J. Zhang, J. Chen, S. Zhi, Y. Chang, P. S. Yu, and J. Han, "Link prediction across aligned networks with sparse and low rank matrix estimation," in *Proceedings of the 33rd IEEE International Conference on Data Engineering*, 2017, pp. 971–982.

[7] X. Cao, H. Chen, X. Wang, W. Zhang, and Y. Yu, "Neural link prediction over aligned networks," in *Proceedings of the 32th AAAI Conference on Artificial Intelligence*, 2018, pp. 249–256.

[8] R. K.-W. Lee, T.-A. Hoang, and E.-P. Lim, "On analyzing user topic-specific platform preferences across multiple social media sites," in *Proceedings of the 26th International Conference on World Wide Web*, 2017, pp. 1351–1359.

[9] R. Zafarani and H. Liu, "Connecting corresponding identities across communities," in *Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media*, 2009.

[10] R. Zafarani, "Connecting users across social media sites: A behavioral-modeling approach," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2013, pp. 41–49.
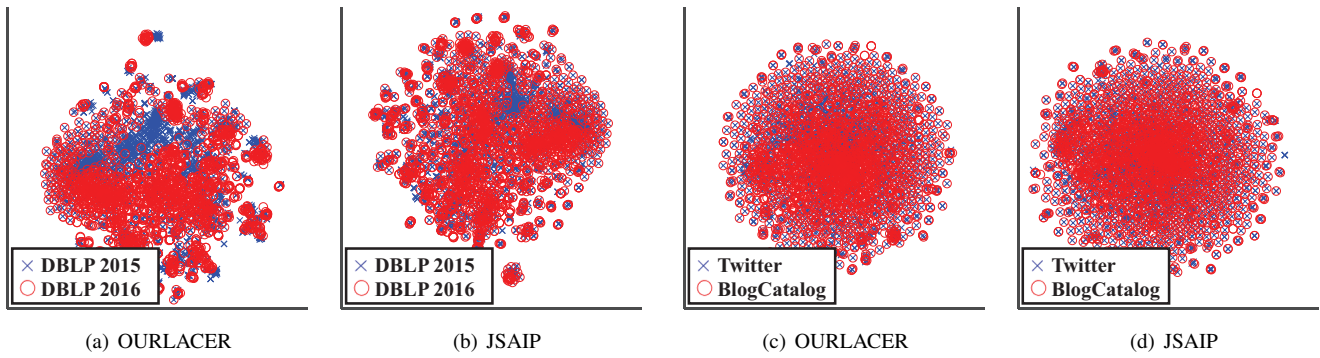
(a) OURLACER   (b) JSAIP   (c) OURLACER   (d) JSAIP

Fig. 5. Visualization of user spaces learned by OURLACER and JSAIP.

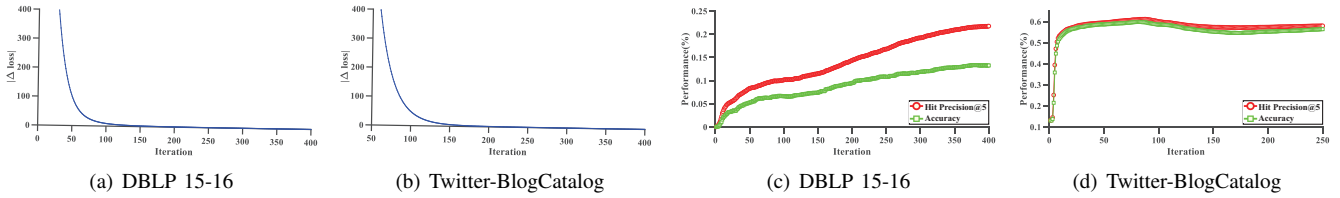(a) DBLP 15-16   (b) Twitter-BlogCatalog   (c) DBLP 15-16   (d) Twitter-BlogCatalog

Fig. 6. The $|\Delta \text{loss}|$ and performance after each iteration for JSAIP method on datasets DBLP 15-16 and Twitter-BlogCatalog.
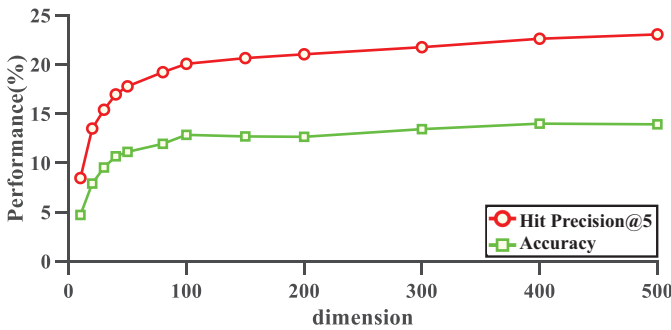

Fig. 7. The performance of our proposed JSAIP method with different dimension on dataset DBLP 15-16.

[11] R. Zafarani, L. Tang, and H. Liu, "User identification across social media," *ACM Trans. Knowl. Discov. Data*, vol. 10, no. 2, pp. 16:1–16:30, 2015.

[12] W. Chen, H. Yin, W. Wang, L. Zhao, W. Hua, and X. Zhou, "Exploiting spatio-temporal user behaviors for user linkage," in *Proceedings of the 26th ACM on Conference on Information and Knowledge Management*, 2017, pp. 517–526.

[13] W. Chen, H. Yin, W. Wang, L. Zhao, and X. Zhou, "Effective and efficient user account linkage across location based social networks," in *Proceedings of the 34th IEEE International Conference on Data Engineering*, 2018, pp. 1085–1096.

[14] T. Man, H. Shen, S. Liu, X. Jin, and X. Cheng, "Predict anchor links across social networks via an embedding approach," in *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, 2016, pp. 1823–1829.

[15] L. Liu, W. K. Cheung, X. Li, and L. Liao, "Aligning users across social networks using network embedding," in *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, 2016, pp. 1774–1780.

[16] F. Zhou, L. Liu, K. Zhang, G. Trajcevski, J. Wu, and T. Zhong, "Deeplink: A deep learning approach for user identity linkage," in *Proceedings of the 37th IEEE Conference on Computer Communications*, 2018, pp. 1313–1321.

[17] Y. Nie, Y. Jia, S. Li, X. Zhu, A. Li, and B. Zhou, "Identifying users across social networks based on dynamic core interests," *Neurocomputing*, vol. 210, pp. 107–115, 2016.

[18] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: online learning of social representations," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2014, pp. 701–710.

[19] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, "LINE: large-scale information network embedding," in *Proceedings of the 24th International Conference on World Wide Web*, 2015, pp. 1067–1077.

[20] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 855–864.

[21] S. Tan, Z. Guan, D. Cai, X. Qin, J. Bu, and C. Chen, "Mapping users across networks by manifold alignment on hypergraph," in *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, 2014.

[22] W. Zhao, S. Tan, Z. Guan, B. Zhang, M. Gong, Z. Cao, and Q. Wang, "Learning to map social network users by unified manifold alignment on hypergraph," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, pp. 5834–5846, 2018.

[23] S. Su, L. Sun, Z. Zhang, G. Li, and J. Qu, "MASTER: across multiple social networks, integrate attribute and structure embedding for reconciliation," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018, pp. 3863–3869.

[24] J. Zhang, B. Chen, X. Wang, H. Chen, C. Li, F. Jin, G. Song, and Y. Zhang, "Mego2vec: Embedding matched ego networks for user alignment across social networks," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 2018, pp. 327–336.

[25] X. Li, Y. Su, N. Gao, W. Tang, J. Xiang, and Y. Wang, "Aligning users across social networks by joint user and label consistence representation," in *Proceedings of the 26th International Conference on Neural Information Processing*, 2019, pp. 656–668.

[26] D. Koutra, H. Tong, and D. Lubensky, "BIG-ALIGN: fast bipartite graph alignment," in *Proceedings of the 13th IEEE International Conference on Data Mining*, 2013, pp. 389–398.

[27] J. Zhang and P. S. Yu, "Multiple anonymized social networks alignment," in *Proceedings of the 15th IEEE International Conference on Data Mining*, 2015, pp. 599–608.