

Distribution Distance Minimization for Unsupervised User Identity Linkage

Chaozhuo Li
Beihang University
lichaozhuo@buaa.edu.cn

Senzhang Wang
Nanjing University of
Aeronautics and
Astronautics
szwang@nuaa.edu.cn

Philip S. Yu
Fudan University
University of Illinois at
Chicago
psyu@uic.edu

Lei Zheng
University of Illinois at
Chicago
lzheng21@uic.edu

Xiaoming Zhang
Beihang University
yolixs@buaa.edu.cn

Zhoujun Li*
Beihang University
lizj@buaa.edu.cn

Yanbo Liang
Hortonworks
yliang@apache.org

ABSTRACT

Nowadays, it is common for one natural person to join multiple social networks to enjoy different services. Linking identical users across different social networks, also known as the User Identity Linkage (UIL), is an important problem of great research challenges and practical value. Most existing UIL models are supervised or semi-supervised and a considerable number of manually matched user identity pairs are required, which is costly in terms of labor and time. In addition, existing methods generally rely heavily on some discriminative common user attributes, and thus are hard to be generalized. Motivated by the isomorphism across social networks, in this paper we consider all the users in a social network as a whole and perform UIL from the user space distribution level. The insight is that we convert the unsupervised UIL problem to the learning of a projection function to minimize the distance between the distributions of user identities in two social networks. We propose to use the earth mover's distance (EMD) as the measure of distribution closeness, and propose two models $UUIL_{gan}$ and $UUIL_{omt}$ to efficiently learn the distribution projection function. Empirically, we evaluate the proposed models over multiple social network datasets, and the results demonstrate that our proposal significantly outperforms state-of-the-art methods.

KEYWORDS

User Identity Linkage, Earth Mover's Distance, Generative Adversarial Network

ACM Reference Format:

Chaozhuo Li, Senzhang Wang, Philip S. Yu, Lei Zheng, Xiaoming Zhang, Zhoujun Li, and Yanbo Liang. 2018. Distribution Distance Minimization for Unsupervised User Identity Linkage. In *The 27th ACM International Conference on Information and Knowledge Management (CIKM '18)*, October

*Zhoujun Li is the corresponding author of this paper.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CIKM '18, October 22–26, 2018, Torino, Italy

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-6014-2/18/10...\$15.00

<https://doi.org/10.1145/3269206.3271675>

22–26, 2018, Torino, Italy. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3269206.3271675>

1 INTRODUCTION

Social media sites become increasingly important in our daily life due to their rising popularity and diversity. It is common that one natural person participates in multiple social platforms to enjoy different types of services. One natural person may create a social identity in each social platform to represent his or her unique public figure, which can be reflected by the social profiles, published posts or the following relationships. The problem of *User Identity Linkage* (UIL), which aims to link the identities of the same natural person across different social platforms, has recently attracted an increasing amount of research attention considering its significant research challenges and the immense practical value. The successful UIL contributes to facilitating various applications including friend recommendation [32], information diffusing prediction [41] and network dynamics analysis [42].

Most existing UIL methods are supervised, which aim to learn a binary classifier to separate matched user identity pairs from non-matched user identity pairs [10, 20, 22–24, 26, 27, 36, 46]. Due to the high cost to obtain the labeled instances, several semi-supervised approaches are proposed to utilize both labeled and unlabeled user identities, in which the unlabeled identity pairs are predicted during the learning process [11, 18, 34, 43, 50, 52]. Recently, several works also tried to perform UIL in an unsupervised manner [12, 13, 17, 29]. Lacoste-Julien et al. [13] proposed a greedy approach to align the attributes of users according to the heuristic string similarities, which is sensitive to user profiles and thus leads to poor generality. CoLink [51] is the most recent weakly-supervised method, and it still needs some labeled seeds to train the co-training algorithm.

Existing unsupervised or weakly-supervised UIL methods rely heavily on some highly discriminative user attributes and suffer from poor generality. The user identities are compared and linked based on some common attributes such as user names and locations. If the attributes of a user u_1 in a social network are very similar to a user v_1 in another social network, u_1 and v_1 tend to be a same person. However, the user attributes in different networks are usually defined and formatted very differently. For example, Twitter users tend to publish tweets while the Instagram users usually share the images. It is difficult to measure the similarity

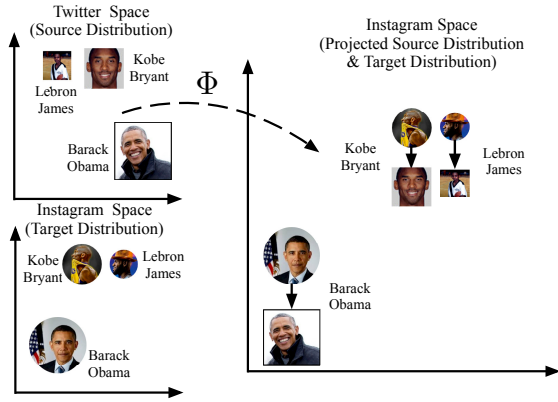


Figure 1: An illustration of the distribution distance minimization in two social networks. We expect to find a projection Φ to minimize the distance between the target distribution in Instagram space and the projected source distribution from Twitter space for unsupervised UIL.

between a tweet and an image. When the common attributes are insufficient, existing methods may not work very well.

In this paper we study unsupervised UIL from a new perspective. Instead of matching the user identities one by one, we consider all the users in a social network as a whole and perform UIL from the user distribution level. The idea is that two independent feature spaces can be constructed based on the text similarities between the Twitter users and the image similarities between the Instagram users. From the global perspective, given a set of natural persons, the distributions of their social identities in different social feature spaces present similar structure, which is referred to the isomorphism across social networks [21]. For example, as shown in Figure 1, Kobe Bryant and Lebron James are both NBA stars and Barack Obama is a politician. Due to the shared interests, Kobe Bryant and Lebron James are more closer in both platforms than Kobe Bryant and Barack Obama, and thus the distributions of them in both feature spaces present similar shape. If we can transform their distribution in Twitter by a set of operations Φ (e.g., shifting, rotation and extension) to minimize the difference between it and the distribution in Instagram, the user identities of the same person will be gathered together as shown in the right part of Figure 1. Inspired by the isomorphism, we convert the UIL problem to the learning of the operations to minimize the distance between the two distributions, which does not need any cross-network indicators.

The operation Φ in Figure 1 is defined as the projection function by [23]. It assumes that the observed data of the social identities are the projection of the natural people constrained by the features and structures provided by the particular platform. Here we can view users in a social network (target network) as the natural people and their identities in another network (source network) as their projections. Although this assumption has been adopted by several recent works [18, 20, 23], they are all supervised methods and annotations are required to learn the projection function by minimizing the distances between the projected source users and the corresponding target users. Different from them, in this paper

the minimization is performed at the distribution level, and hence sample level supervision is not required. This objective frees the model from focusing on local similarity in the sample level, but rather on examining the intrinsic structure of distributions.

This intuition naturally calls for a measure of distribution closeness, for which we introduce the earth mover’s distance (EMD). EMD has been proven as an effective distance measurement of distributions in information retrieval [30, 31], computer vision [25, 28] and natural language processing domains [48, 49]. EMD holistically measures the closeness between two set of weighted points by computing the minimal cost of transporting one set of points to the other. In the UIL task, we view social identities as the weighted points and aim to learn a projection Φ to minimize EMD between the projected source distribution and the target distribution.

However, it is very challenging to effectively solve the dynamic EMD minimization problem under the UIL scenario. The traditional solutions usually focus on the EMD minimization between two fixed distributions. The two input distributions are considered as two independent node sets in a bipartite graph. The weights of edges are the ground distances between the connected nodes in different sets. Thus EMD minimization problem can be considered as the minimum cost flow problem, which can be solved by the hungarian algorithm [14]. However, hungarian algorithm can only be applied on the static bipartite graph. In this task, we aim to learn a projection function by minimizing the EMD between two node distributions, while the projected distribution changes with the learning of the projection function. Therefore, the bipartite graph is dynamic rather than static and hungarian algorithm cannot apply. The optimal solution of dynamic EMD minimization is still unclear as Ding et al. [6] has proven its exact solution is NP-hard.

In this paper, we study unsupervised user identity linkage (UUIL) through learning a projection function to efficiently minimize the EMD between the distributions of user identities in two social networks. Specifically, we propose two optimization algorithms UUIL_{gan} and UUIL_{omt} to learn the projection function. UUIL_{gan} utilizes the Wasserstein Generative Adversarial Net [1] to minimize the Wasserstein distance between two distributions, which generalizes EMD to allow continuous distributions. UUIL_{omt} is an orthogonal matrix factorization based method to guarantee the minimization procedure converge to a local minimum. We extensively evaluate their performance on two pairs of social network datasets and three pairs of co-author datasets. The experimental results demonstrate UUIL_{omt} can locate a good neighbor of the optimum and provide a reasonable initialization for the UUIL_{gan} model. The combination of the two models outperforms unsupervised baselines by a large margin and achieve comparable performance with semi-supervised methods.

We summarize our main contributions as follows.

- We study the novel problem of UUIL. Different from existing sample-level methods, we are the first to perform UIL through minimizing the distance between user distributions.
- Earth Mover’s Distance is introduced as the measure of distribution closeness. We propose two optimization methods to minimize the dynamic EMD: a generative adversarial network based model UUIL_{gan} and an orthogonal matrix transformation based model UUIL_{omt}.

- Extensively, we evaluate the proposed models on two pairs of social network datasets and three pairs of paper co-author network datasets. The results show the superior performance of the proposed models by comprehensive comparison with state-of-the-art baseline methods.

The rest of this paper is organized as follows. Section 2 briefly introduces the mathematical background of EMD and formally defines the studied problem. Section 3 introduces the proposed models. Section 4 presents the experimental results. Section 5 summarizes the related works. Finally, we conclude this work in Section 6.

2 PRELIMINARIES AND PROBLEM DEFINITION

2.1 Earth Mover's Distance

EMD defines a distance between two probability distributions, which estimates the minimum amount of work required to change one distribution into the other. A discrete probability distribution can be depicted by a sum of Dirac delta functions: $\mathbb{P} = \sum_i p_i \delta_{x_i}$, in which x_i is a sample in the distribution \mathbb{P} , p_i is its corresponding probability and δ_{x_i} is the Dirac delta function [2, 48]. For a pair of discrete distributions $\mathbb{P}^I = \sum_i u_i \delta_{x_i}$ and $\mathbb{P}^J = \sum_j v_j \delta_{y_j}$, the EMD between them is defined as:

$$\text{EMD}(\mathbb{P}^I, \mathbb{P}^J) = \min_{E \in \mathcal{U}(\mathbb{P}^I, \mathbb{P}^J)} \sum_i \sum_j E_{ij} d(x_i, y_j) \quad (1)$$

where $d(x_i, y_j)$ measures the ground distance between x_i and y_j . $\mathcal{U}(u, v)$ is the earth transport polytope records the transportation strategy between distributions:

$$\mathcal{U}(\mathbb{P}^I, \mathbb{P}^J) = \{E | E_{ij} \geq 0, \sum_j E_{ij} = u_i, \sum_i E_{ij} = v_j, \forall i, j\} \quad (2)$$

In our UIL task, we view the projected source distribution as the piles of earth, and the target distribution as holes to be filed. The EMD estimates the minimum cost of turning the earth into the holes, which is calculated by the amount of dirt moved times the moving distance. E_{ij} represents the amount of the dirt moved from the pile y_j to the hold x_i . Larger E_{ij} means users x_i and y_j has a larger chance to be a same person.

In addition, Wasserstein distance is the extension of EMD on the continuous distributions [5, 48], which is defined as:

$$W(\mathbb{P}^I, \mathbb{P}^J) = \inf_{\gamma \in \Gamma(\mathbb{P}^I, \mathbb{P}^J)} \int d(x, y) d\gamma(x, y), \quad (3)$$

where $\Gamma(\mathbb{P}^I, \mathbb{P}^J)$ denotes the a joint probability distribution $\gamma(x, y)$ with marginals \mathbb{P}^I and \mathbb{P}^J . Formula (3) aims to find the joint distribution to reach the expectation infimum of the ground distances. Considering the sampling process $(x, y) \sim \gamma$ is based on their weights, formula (3) is equivalent to the objective of EMD [5].

2.2 Problem Definition

The social network is defined as $N = \{V, W, P\}$, in which $V = \{v_1, v_2, \dots, v_n\}$ is the set of users in this network. $W \in \mathbb{R}^{d \times n}$ represents the feature matrix in which $w_i \in \mathbb{R}^{d \times 1}$ is the feature vector of identity v_i with the dimension of d . $P \in \mathbb{R}^{1 \times n}$ represents the topology importance of the users (e.g. the count of followers). We formally define the studied problem as follows.

Definition 2.1. Unsupervised User Identity Linkage. Given two input social networks $S = \{V_S, W_S, P_S\}$ and $T = \{V_T, W_T, P_T\}$, we aim to locate a set of user identity pairs $U = \{(v_s, v_t) | v_s \in V_S, v_t \in V_T\}$ without any annotated pairs as supervision, in which v_s and v_t are the identities of the same natural person in the two networks.

Here we view the users of an input social network N as a distribution $\mathbb{P}^N = \sum_n p_n \delta_{w_n}$, in which w_n defines the position of the user v_n in this social space and p_n is the his topology importance. Then we can define the problem in the distribution level as follows.

Definition 2.2. Earth Mover's Distance Minimization for Unsupervised User Identity Linkage. Given the source distribution \mathbb{P}^S and the target distribution \mathbb{P}^T , we aim to learn a projection Φ to achieve the minimal earth mover's distance between the projected source distribution $\mathbb{P}^{\Phi(S)}$ and the target distribution \mathbb{P}^T .

After the projection process, the ground distance $d(\Phi(u_s), u_t)$ between the projected point of the source user u_s and the target user u_t , or the transportation probability E_{ij} can be viewed as the indicator of UIL. Larger E_{ij} or a smaller distance c represents two users has a larger chance to be a same person. We assume the input two social networks share the same feature dimension d , which can be easily satisfied by popular node embedding methods [15, 16, 40, 44]. Following previous works [23, 47], we consider the two input networks are partially aligned with different number of identities $|V_T| \neq |V_S|$.

3 UNSUPERVISED USER IDENTITY LINKAGE

In this section we will introduce the details of the proposed unsupervised user identity linkage models UUIL_{gan} and UUIL_{omt} .

3.1 Projection Function

The approximate isomorphism across networks inspires researchers to use a simple form of projection. Mu et al. [23] utilizes the linear transformation as the projection. Similar strategy has been adopted by [20] and achieves desirable performance. Following the previous works, we select the linear transformation as the projection, which is defined as: $\Phi(w_s) = G \times w_s$, in which $G \in \mathbb{R}^{d \times d}$ is the transformation matrix to be learned. We also tried the non-linear transformations (multi-layer perceptron with non-linear activations) but their performances are unsatisfied. This may be because the non-linear functions seriously change the distributions and destroy the isomorphism. In addition, we restrict that the transformation matrix G has to be an orthogonal matrix. Geometrically, converting a distribution by an orthogonal matrix reflects the distribution in some plane and rotates it, which will preserve the intrinsic structure of this distribution. An orthogonal projection is also theoretically appealing for its numerical stability [33, 48].

3.2 Generative Adversarial Training for Unsupervised User Identity Linkage

In this subsection we introduce the details of the proposed UUIL_{gan} , which is a generative adversarial training based model. GAN is currently a state-of-the-art generative model [7], which performs an adversarial game between the generator and the discriminator. This inspires us to view the projection as the generator, and the discriminator is designed to distinguish the projected source distribution

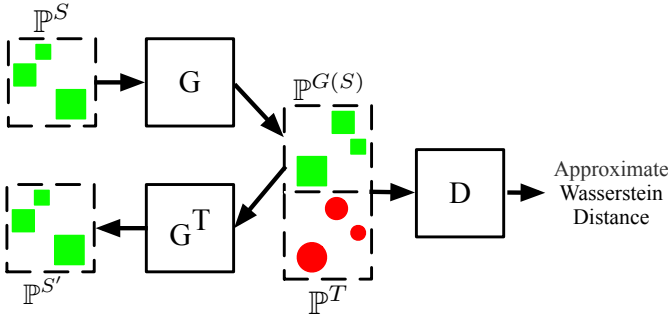


Figure 2: An illustration of the proposed UUIL-gan model, which is a generative adversarial network.

and the target distribution. However, the origin GAN is very difficult to train and unpredictable. Recently an novel variation WGAN [1] is proposed to perform more stable training, which aims to minimize the Wasserstein distance between the actual distribution and the generated one. Zhang et al. [48] introduces the WGAN to perform bilingual lexicon induction task. Inspired this work, we introduce WGAN as the foundation of the UUIL-gan model.

Figure 2 represents the framework of the proposed UUIL-gan model. The generator G projects the distribution of the source platform \mathbb{P}^S into the projected source distribution $\mathbb{P}^{G(S)}$, and the discriminator D estimates the Wasserstein distance between the target distribution \mathbb{P}^T and $\mathbb{P}^{G(S)}$. The objective of UUIL-gan is to learn the projection matrix used in the generator to minimize the Wasserstein distance estimated by the discriminator:

$$\min_{G \in \mathbb{R}^{d \times d}} W(\mathbb{P}^T, \mathbb{P}^{G(S)}) = \inf_{\gamma \in \Gamma(\mathbb{P}^T, \mathbb{P}^{G(S)})} \mathbb{E}_{(w_t, Gw_s) \sim \gamma} [d(w_t, Gw_s)], \quad (4)$$

in which the symbols are the same meanings to the ones in formula (3). This objective is not directly solvable because it is cumbersome to enumerate all the possible joint distributions in $\Gamma(\mathbb{P}^T, \mathbb{P}^{G(S)})$. If the ground distance function d is set to L2 Euclidean distance, Vallani et al. [35] proves the Wasserstein distance has a simple format based on the Kantorovich-Rubinstein duality:

$$W(\mathbb{P}^T, \mathbb{P}^{G(S)}) = \frac{1}{K} \sup_{\|f\|_L \leq K} \mathbb{E}_{w_t \sim \mathbb{P}^T} [f(w_t)] - \mathbb{E}_{Gw_s \sim \mathbb{P}^{G(S)}} [f(Gw_s)] \quad (5)$$

in which f is the K-Lipschitz continuous function and the Wasserstein distance can be understood as the proportion to the supremum over all K-Lipschitz functions. Considering the strong approximate ability of the neural networks [9], here we utilize a feed-forward neural network as the approximation of function f , which is the discriminator D shown in Figure 2. Hence, the objective of D is to learn the parameters of the neural network to find the supremum over K-Lipschitz functions:

$$\max_{\theta: \|f_\theta\|_L \leq K} L_D = \mathbb{E}_{w_t \sim \mathbb{P}^T} [f_\theta(w_t)] - \mathbb{E}_{Gw_s \sim \mathbb{P}^{G(S)}} [f_\theta(Gw_s)] \quad (6)$$

where θ is the parameter set of the discriminator. Under the limitation of K-Lipschitz continuity, if we can learn a desired discriminator D to maximize objective (6), its calculation result is the estimated Wasserstein distance up to a multiplicative constant K . To maintain the K-Lipschitz continuity of f_θ , Arjovsky et al. [1] proposed a

Algorithm 1 Training process of UUIL-gan

Require: the learning rate α , the clipping parameter c , the batch size m , the number of discriminator updates in each iteration n_d and the reconstruction weight λ .

Require: the initial generator parameters G_0 , the initial discriminator parameters θ_0 .

```

1: while  $G$  has not converged do
2:   for  $i = 0 \rightarrow n_d$  do
3:     Sample a batch from source distribution:  $\{w_s^{(i)}\}_{i=1}^m \sim \mathbb{P}^S$ 
4:     Sample a batch from target distribution:  $\{w_t^{(i)}\}_{i=1}^m \sim \mathbb{P}^T$ 
5:      $g_\theta \leftarrow \nabla_\theta [\frac{1}{m} \sum_{i=1}^m f_\theta(w_t^{(i)}) - \frac{1}{m} \sum_{i=1}^m f_\theta(Gw_s^{(i)})]$ 
6:      $\theta \leftarrow \theta + \alpha \cdot \text{RMSProp}(\theta, g_\theta)$ 
7:      $\theta \leftarrow \text{clip}(\theta, -c, c)$ 
8:   end for
9:   Sample a batch from source distribution:  $\{w_s^{(i)}\}_{i=1}^m \sim \mathbb{P}^S$ 
10:   $g_G \leftarrow \nabla_G [\frac{1}{m} \sum_{i=1}^m [-f_\theta(Gw_s^{(i)}) + \lambda \cdot d(w_s^{(i)}, G^T Gw_s^{(i)})]]$ 
11:   $G \leftarrow G - \alpha \cdot \text{RMSProp}(G, g_G)$ 
12: end while

```

simple but effective trick: clamp the weights θ to a small window after every gradient updating.

The target of discriminator D is to estimate the Wasserstein distance between distributions \mathbb{P}^T and $\mathbb{P}^{G(S)}$, while the generator G aims to minimize the estimated distance. The first term of formula (6) has nothing to do with the generator G . Hence the objective function of G can be defined as:

$$\min_{G \in \mathbb{R}^{d \times d}} L'_G = -\mathbb{E}_{Gw_s \sim \mathbb{P}^{G(S)}} [f_w(Gw_s)] \quad (7)$$

As the generator loss (7) decreases in the training process, the Wasserstein distance becomes smaller and the user identities of the same person will move closer.

As discussed in subsection 4.1, we hope the learned projection matrix should be orthogonal considering its superiority. However, directly introducing traditional optimization methods under the orthogonal constraint in GAN is intractable [33]. Hence another component is proposed to latently process the orthogonal constraint as shown in the bottom left part of figure 2. We assume the transposed projection matrix G^T should be able to reconstruct the source distribution \mathbb{P}^S from its projected version $\mathbb{P}^{G(S)}$, which can be understood as the minimization of the distance between the original distribution \mathbb{P}^S and the reconstructed one $\mathbb{P}^{S'}$.

$$\min_{G \in \mathbb{R}^{d \times d}} L_R = \mathbb{E}_{w_s \sim \mathbb{P}^S} [d(w_s, G^T Gw_s)] \quad (8)$$

Through this way we can force the learned matrix to be orthogonal. Hence, the entire objective of the generator is the weighted sum of L'_G and L_R as follows.

$$\min_{G \in \mathbb{R}^{d \times d}} L_G = L'_G + \lambda \cdot L_R \quad (9)$$

where λ is a hyper parameter that balances the two terms and larger λ should enforce a stricter orthogonal constraint.

Algorithm 1 represents the training procedure of the UUIL-gan model. In each iteration from line 2 to 11, we train the discriminator n_d times but only train generator once, which is a popular strategy to prevent the collapsed GAN problem [1]. After sampling the

training batches from the two distributions, the parameter set θ of the discriminator are updated by the RMSProp optimizer following formula (6) in line 5 and 6. Line 7 shows the clip operation to ensure the K-Lipschitz continuity. Lines 9 to 11 show the updating steps of generator following the formula (9). Here we utilize the distance between the projected source users and the target users as the measurement. A smaller distance represents a higher probability belonging to the same person.

3.3 Orthogonal Matrix Transformation for Unsupervised User Identity Linkage

The UUIL_{gan} model aims to learn the projection by minimizing the duality of Wasserstein distance. Different from UUIL_{gan} focusing on the duality of the objective, we further design another model UUIL_{omt} to solve the minimization problem of the original EMD objective, which is defined as:

$$\min_{G \in \mathbb{R}^{d \times d}} \text{EMD}(\mathbb{P}^{G(S)}, \mathbb{P}^T). \quad (10)$$

After replacing EMD function with formula (1), we can get the following objective:

$$\min_{G \in \mathbb{R}^{d \times d}} \min_{E \in \mathcal{U}(p^S, p^T)} \sum_{s=1}^{|V^S|} \sum_{t=1}^{|V^T|} E_{st} d(Gw_s, w_t) \quad (11)$$

$$\text{st} : G^T G = I.$$

To minimize the EMD between distributions $\mathbb{P}^{G(S)}$ and \mathbb{P}^T , we need to locate two matrices: the earth transport matrix E to measure the EMD, and the projection matrix G to minimize the calculated distance. Ding et al. [6] has proven the exact solution to this dynamic EMD minimization problem is NP-hard. However, Cohen et al. [3] demonstrate this objective can converge to a local minimum by an gradual convergence framework.

The projection matrix $G^{(0)}$ is initialized as a randomly generated orthogonal matrix. In the k -th training iteration, we first fix the matrix G to calculate the transport matrix E by minimizing the following objective:

$$E^{(k)} = \min_{E^{(k)} \in \mathcal{U}(p^S, p^T)} \sum_{s=1}^{|V^S|} \sum_{t=1}^{|V^T|} E_{st}^{(k)} d(G^{(k-1)}w_s, w_t). \quad (12)$$

The EMD between the two distributions is the sum of the right part of formula (12). Then based on the learned matrix $E^{(k)}$, we aim to find the desired $G^{(k)}$ to minimize the calculated EMD:

$$G^{(k)} = \min_{G^{(k)} \in \mathbb{R}^{d \times d}} \sum_{s=1}^{|V^S|} \sum_{t=1}^{|V^T|} E_{st}^{(k)} d(G^{(k)}w_s, w_t) \quad (13)$$

$$\text{st} : (G^{(k)})^T G^{(k)} = I$$

These two objectives are optimized alternately and repeatedly in the model training iterations.

The objective (12) aims to find the transport matrix E which can minimize EMD between two distributions. This is the traditional EMD optimization problem and a set of methods can solve it. Here we choose a fast and scalable model [4]. The objective (13) aims to learn the projection matrix G given the fixed matrix E . It seems difficult to find the solution to this objective, especially with the orthogonal constraint. However, if we select the squared Euclidean

distance L_2^2 as the ground distance function d , the objective can be viewed as an extension of the Orthogonal Procrustes problem [8, 48], which has a explicit solution by singular value decomposition:

$$G^{(k)} = UV^T, \quad (14)$$

in which matrices U and V can be learned by the SVD of the following matrix:

$$\sum_{s=1}^{|V^S|} \sum_{t=1}^{|V^T|} E_{st}^{(k)} w_s w_t^T = U \Sigma V^T. \quad (15)$$

This calculation of SVD is effective due to the generated matrix in the left part of formula (15) is a $d \times d$ matrix. When this optimization converges to a local minimum, we can utilize the learned transport matrix E as the indicator of UIL. A larger value of E_{st} means the user v_t in network T and user v_s in network S has a larger chance to be a same person.

4 EXPERIMENTS

In this section, we will introduce the experimental part. First we will present the experiment setup, including the datasets, baseline methods and the parameter setup. Then we will present and discuss the performance of different UIL methods. Finally we show that EMD can be also used to measure the difference between two social networks, which is an additional interesting discovery.

4.1 Datasets

In this paper, we use two pairs of social network datasets and three pairs of academic co-author datasets for evaluation. The datasets are crawled and organized by our corporation coauthor. The details of the datasets are introduced as follows.

- **Twitter-Flickr:** Twitter (<https://twitter.com>) is one of the most popular social networking platforms where users can post tweets to share their opinions. Flickr (<https://www.flickr.com/>) is a popular image sharing platform where users can share pictures and join groups of interest. Although many Twitter and Flickr datasets are publicly available, it is difficult to have a set of linked users as the ground truth. Luckily, the third part website about.me (<https://about.me/>) provides a platform where people can link their identities in different social networks. From an open about.me dataset¹, we crawl the social data of the same natural person from Twitter and Flickr networks. After removing the canceled identities and the users with few attributes, we finally have 3259 Twitter identities and 4308 Flickr identities, in which 2773 pairs of identities are matched as the ground truth.
- **Weibo-Douban:** Sina Weibo (<http://weibo.com/>) is a popular Chinese social platform, which can be viewed as a hybrid of Twitter and Facebook. Douban (<https://www.douban.com/>) is a social service for people to share comments on movies, books and musics. Douban users can show their corresponding Weibo account in their home pages, and thus we can obtain the ground truth users through this information. After crawling the linkage information and user data from the two social networks, we obtain 3235 pairs of match identities.

¹<https://github.com/kite1988/aboutme>

Table 1: Statistics of the datasets. The numbers in the brackets are the counts of nodes in the networks.

Dataset	Source Network	Target Network	Matched Pairs
Twitter-Flickr	Twitter (3,259)	Flickr (4,308)	2,773
Weibo-Douban	Weibo (4,119)	Douban (4,554)	3,235
DBLP15-DBLP16	DBLP15 (3,881)	DBLP16 (5,989)	1,852
DBLP16-DBLP17	DBLP16 (5,989)	DBLP17 (7,073)	2,570
DBLP15-DBLP17	DBLP15 (3,881)	DBLP17 (7,073)	1,492

Furthermore, a set of unmatched identities are randomly selected to form a partially aligned network. Finally this dataset contains 4119 Weibo accounts and 4554 Douban accounts.

- **DBLP**: Besides the social network datasets, we also introduce several co-author networks to evaluate the proposed models. DBLP (<http://dblp.uni-trier.de/>) is the on-line reference for open bibliographic information on computer science papers, and the dataset is publicly available². We first split the whole dataset to three co-author networks in different years (2015, 2016 and 2017). In each network, we choose *Yoshua Bengio* as the center node, and then locate the co-authors that can be reached from the center node with no more than three hops to construct the co-author subnetwork. We aim to link the two nodes in two different subnetworks according to the co-author relationships and the published papers. The author identities in the DBLP are used as the ground truth.

Table 1 shows the detailed information of the datasets.

4.2 Data Preprocessing

For each dataset, we crawl the user information from the two networks to construct their features. For the Twitter users, we gather their published tweets and the following relationships. After processing the tweets using NTLK³ stemmer and removing some stop or rare words (occurrence less than 10 times), the tweets published by a user are represented as a tf-idf vector. In order to incorporate the topology information and reduce the feature dimension to speed up the training process, we use TADW [39], a popular network embedding model, to embed the following relationships and the text features into a low dimensional latent vector. Finally the vectors are normalized with L2-normalization. Flickr, Weibo and Douban datasets are processed in a similar way but only differs in the user attributes. We crawl the image tags and interest groups for Flickr users, interest tags and the hashtags for Weibo users, and the interests tags and the joined groups for Douban users.

For the DBLP dataset, we first split the dataset to three networks according to the years. For each author in a network, we identify all the published papers of the author in this year and extract the titles and abstracts of the papers as the attributes, which are represented by a tf-idf vector. We utilize TADW to embed the co-author relationships and attributes into the latent feature vectors.

Note that for the two networks of a network pair, their feature spaces are trained independently, which ensure the generality of the proposed models. For the social networks, we view the normalized

count of the followers as the weight p_i of the user v_i , which will be used as the guide to sampling the training samples in $UUIL_{gan}$ and calculate the transport matrix E in the $UUIL_{omt}$. For the co-author networks, we utilize the degree of a node as its weight.

4.3 Comparison Methods

Here we select the following state-of-the-art UIL methods to make a comparison. Note that the baselines contain both unsupervised, weakly supervised and supervised methods to extensively evaluate the proposed models.

- **Degree-Based Alignment (DBA)**: Nodes between two networks are matched in terms of their degrees, offering a trivial baseline for UIL. First we rank user identities in the input networks based on their degrees. Then given the i -th ranked identity in the source network, the matched candidates of the selected identity are the $i-k/2$ to $i+k/2$ ranked identities in the other network.
- **KNN**: We use the K-Nearest-Neighbor (KNN) as another unsupervised UIL baseline, which is a popular baseline adopted by related works. For a user in a network, the users in the other network are ranked according to the attribute similarities with the source user. We use cosine similarity as the measurement in KNN and select k nearest users in the target network as the matched candidates.
- **Alias-Disamb** [17]: This is also an unsupervised approach based on screen names. Firstly, the training data is automatically generated by evaluating the rareness of user names in both networks. The n-gram probability is utilized to estimate the rareness or commonness of a name. We label the candidate pairs with top 5% rarest names as positive examples, and candidate pairs with top 5% most common name as negative examples. Then we train a binary SVM classifier using the generated training pairs, in which the user features are the tf-idf attributes vectors as shown in the preprocessing step.
- **CoLink** [51]: This is a weakly supervised UIL model which employs a co-training algorithm to manipulate two independent components: the attribute-based model and the relationship-based models. These two models reinforce each other iteratively in an weak supervised way with a few selected seeds.
- **ULink** [23]: This is a supervised UIL model, which aims to minimize the distances between the identities of the same people, while maximizes the distances between the identities of different people in the latent space.
- **UUIL_{gan}**: This is the proposed GAN based model.
- **UUIL_{gan} + UUIL_{omt}**: Considering the performance of $UUIL_{omt}$ is seriously affected by the initialization, we design a combined model to overcome this drawback. We use the projection matrix G learned by the $UUIL_{gan}$ model as the initialized matrix $G^{(0)}$ of $UUIL_{omt}$. The combined model can be benefited by the merits of both methods.

4.4 Parameter Setup

For the proposed models, the dimension of the feature space d of the input networks is set to 100. For the $UUIL_{gan}$ model, the discriminator D is implemented as a feed-forward neural network

²<http://dblp.uni-trier.de/xml/>

³<https://www.nltk.org/>

Table 2: Comparison with unsupervised methods (*Hit-Precision score*).

	Twitter-Flickr	Weibo-Douban	DBLP15-16	DBLP16-17	DBLP15-17
DBA	0.166	0.125	0.273	0.278	0.230
KNN	-	0.279	0.356	0.353	0.321
Alias-Disamb	0.291	0.223	0.347	0.376	0.351
UUIL _{gan}	0.323	0.310	0.442	0.467	0.432
UUIL _{gan} + UUIL _{omt}	0.354	0.366	0.482	0.497	0.461

with one hidden layer of 256 neurons. Considering a too strong discriminator may cause the GAN corruption and make the generator lose the adversarial ability, we implement the discriminator D as a feed-forward neural network with only one hidden layer of 128 neurons. The projection matrix used in the generator G is initialized with a random orthogonal matrix. The batch size m is set to 256 and the learning rate α is set to 0.0001. The number of discriminator training times in a iteration n_d is set to 5, the clipping parameter c is set to 0.01 and the reconstruct weight λ is set to 0.3.

For the baseline methods, the parameter k in DBA and KNN is set to 20. The codes for Alias-Disamb, ULink and the CoLink methods are implemented based on the original papers. The SVM used in the Alias-Disamb is implemented by LibSVM⁴ tool with RBF kernel. The coefficient parameter C used in the ULink models is also tuned in the randomly selected validation dataset. For the CoLink model, we employ SVM as the attribute-based model and randomly select the training seeds from the matched identity pairs.

Evaluation Metrics. Following previous work [23], we select the *Hit-Precision* as the evaluation metric to compare the top-k candidates. Here we set $k = 20$ and evaluate all the methods by computing *Hit-Precision* for each source user as follows:

$$h(x) = \frac{k - (\text{hit}(x) - 1)}{k} \quad (16)$$

where $\text{hit}(x)$ represents the position of the correct linked user in the returned list of the top-k candidate target identities. Then the *Hit-Precision* is calculated by the average on the scores of the matched source identities: $\frac{\sum_{i=0}^m h(x_i)}{m}$, in which m is the number of the source identities in the matched identity pairs.

4.5 Experimental Results

4.5.1 Comparison with Unsupervised Baselines. We first compare our model with unsupervised baselines, and the results are shown in Table 2. All the methods perform better on the DBLP datasets than on the social networks. This is probably because the co-author network is more denser and the user attributes are formatted and clean. As a trivial method, DBA achieves the lowest *Hit-Precision* scores in all the datasets. DBA performs better on the DBLP datasets, because the co-author relationships are more consistent than the following relationships in social networks (e.g. an influential researcher has a larger chance to keep his influence in the next year). KNN performs better than DBA, showing the user attributes contribute more than the topology information. As the attributes of Twitter users (tweets) and Flickr users (tags and groups) are totally different, it is very difficult to obtain enough

common cross-network attributes and thus the promising feature similarity is hard to be obtained. Utilizing stronger discriminative feature (screen name), Alias-Disamb performs better than the previous unsupervised methods. Alias-Disamb performs better on Twitter-Flickr than on Weibo-Douban, which may be because Chinese users more like to use different aliases in different social networks.

The proposed UUIL_{gan} model outperforms all the baseline methods over all the datasets. By matching the two networks in the distribution level, UUIL_{gan} captures the global similarities, including both the topology and content information, to link the user identities. Without using the strong discriminative features, the UUIL_{gan} achieves much better performance than Alias-Disamb. Meanwhile, UUIL_{gan} is a general model, in which the two input feature spaces are learned independently. Hence UUIL_{gan} can be applied on networks with totally different user attributes like Twitter-Flickr. Based on the UUIL_{gan} model, we utilize the learned projection matrix as the initialization of UUIL_{omt} model, which further improve the performance by 3%. This enhancement proves the UUIL_{omt} can better explore the parameter space located by the UUIL_{gan} model to reach a desirable local minimal. By combining these two models, we can enjoy the merits of both approaches.

4.5.2 Comparison with Supervised Baselines. We next compare the proposed methods with the state-of-the-art supervised and weakly-supervised methods. We only show the results of the combined model of UUIL_{gan}+UUIL_{omt}, as it achieves the best performance. The results are shown in Figure 3. The x-axis in the figure is the number of training seeds, which are randomly selected from the ground truth. From Figure 3, one can see that the performance of the supervised method ULink and the weakly-supervised method CoLink increases with the increase of the available training seeds. CoLink consistently outperforms ULink. This is because CoLink is a weakly supervised model, which utilizes the topology and attribute similarities of the unlabeled samples as complementary, while ULink is a supervised model whose training process is totally guided by the labeled seeds. For all the four datasets, the proposed UUIL_{gan}+UUIL_{omt} significantly outperforms CoLink and ULink when very few annotations are available. One can see that UUIL_{gan}+UUIL_{omt} achieves comparable performance with or even outperforms the two baselines with 250 labeled seeds, which proves UUIL_{gan}+UUIL_{omt} can effectively reduce the requirement of annotations. For different datasets, the needed number of training seeds for the two baselines to achieve the comparable performance with UUIL_{gan}+UUIL_{omt} are different due to the various network characteristics. For example, it requires fewer labeled samples as

⁴<https://www.csie.ntu.edu.tw/~cjlin/libsvm/>

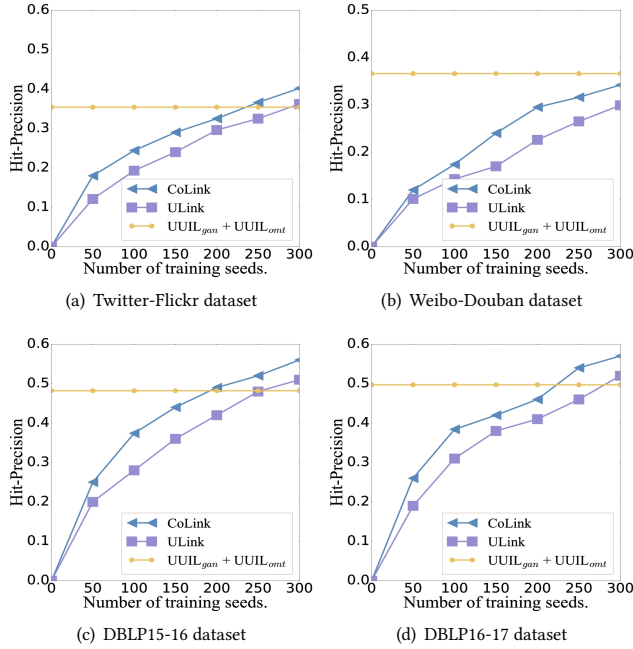


Figure 3: Comparison with the supervised methods.

seeds for the co-author networks than the social networks because the social media data is more noisy and sparse.

4.5.3 Learning Behavior of the UUIL_{GAN}. It is well known that the training process of GAN is instable. Hence we analyze the learning behavior of the UUIL_{gan} model. Figure 4 shows the training trajectory of the UUIL_{gan} on the Weibo-Douban dataset.

During the model training process, we save the parameters after every 1,0000 batches and finally we can obtain 100 checkpoint models. We utilize these models to estimate the Wasserstein distance in between the projected source distribution and the target distribution in these 100 checkpoints. Besides, we also record the *Hit-Precision* score of each checkpoint model. The Wasserstein distance and the *Hit-Precision* scores of the checkpoint models can indicate the training behavior of UUIL_{gan} as shown in Figure 4. The blue curve is the training trajectory of Wasserstein distance, while the purple one represents the *Hit-Precision* scores. Note that we rescale the Wasserstein distance to the range of 0 to 10.

From a wholeness perspective on Figure 4, one can see the Wasserstein distance decreases with the increase of the training batches, which proves the effectiveness of the proposed models in minimizing EMD. Besides, the correlation between the estimated Wasserstein distance and the *Hit-Precision* curves show that a smaller Wasserstein distance results in a better user identity linkage performance. Hence, the estimated Wasserstein distance can be utilized as a indicator of the identity linkage performance. Therefore, we can save the model with the lowest Wasserstein distance during training as the final model. In addition, one can see the curves in Figure 4 are not very smooth. Although UUIL_{gan} can achieve better performance and learn a desirable projection matrix, it suffers from unstable training problem. However, UUIL_{omt} model

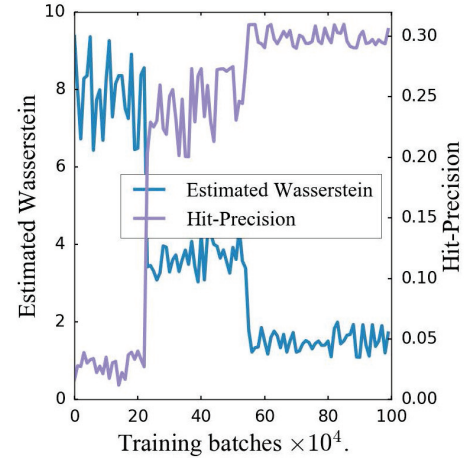


Figure 4: The training trajectory of the UUIL_{gan} model.

can provide stable training process, and thus a combination of the two can enjoy the advantages of both methods.

4.5.4 Parameter Sensitivity Study. Here we study the sensitivity of the proposed unsupervised user identity linkage models on some core parameters. The UUIL_{omt} model has no hyper-parameters, so we only study the effect of different parameter settings to UUIL_{gan}. We select the feature dimension d and the reconstruct weight λ as the core parameters, and perform model training with different settings of d and λ on the Weibo-Douban dataset. The feature dimension d represents the scale of input network information, while the parameter λ controls the weight of the reconstruction error in the final objective function.

Figure 5(a) shows the sensitivity of the model performance on different feature dimension d . We let d vary from 25 to 125 and λ vary from 0.1 to 0.5. One can see the performance of UUIL_{gan} first increases and then keeps stable with the increase of the dimension. It demonstrates that a higher dimensional feature space helps to better preserve the user information, which further contributes to improving the UIL performance. Considering the larger dimension will cost more computational resources and the time complexity is high, we select $d = 100$ to trade off between the effectiveness and efficiency. When we increase the reconstruct weight λ , the *Hit-Precision* score first increases and then decreases. It shows that properly introducing reconstruction loss contributes to improving the performance. A larger λ will introduce a stricter orthogonal constraint, which may interrupt the optimization procedure to reach the minimal Wasserstein distance.

4.5.5 EMD as a measure of network difference. As an additional interesting discovery, we find the calculated EMD can indicate the degree of difference between two networks. Here we take the DBLP dataset as an example. From an intuitive perspective, the publications in two consecutive years should be more similar than in two inconsecutive years, and thus the difference between DBLP15 and DBLP16 networks should be smaller than DBLP15 and DBLP17. After training UUIL_{omt}, we can obtain the transportation matrix E and the projection matrix G . Then we can calculate the EMD

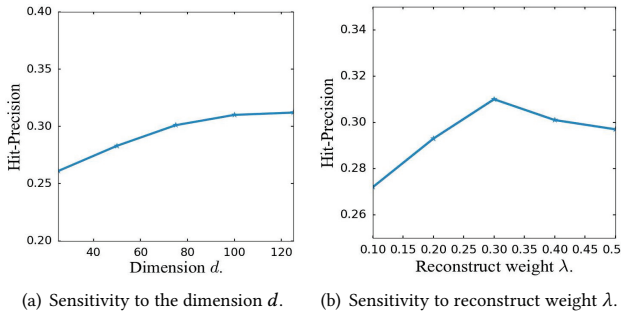


Figure 5: Parameter sensitivity study.

between two networks according to the formula (11). The results are shown in Table 3. The EMD between DBLP15 and DBLP17 is much larger than other datasets, which is consistent with the intuition. Furthermore, we select the average percentage of matched nodes in each pair of networks as another indicator. Networks with more matched nodes should be more similar, and vice versa. From the Table 3, one can see the matched nodes percentage shows similar variation trend to EMD among different datasets. The DBLP15-DBLP17 dataset has the lowest percentage of matched nodes and the highest EMD, both indicating the DBLP15 network is dissimilar to the DBLP17 network. The results of EMD are consistent with the intuitive sense and the straightforward indicator, hence it can be used to measure the difference between two networks.

5 RELATED WORKS

According to the portion of used annotations, existing UIL methods can be roughly categorized into supervised, semi-supervised and unsupervised methods. Most existing works are supervised, which view the UIL task as a typical binary classification problem to separate the matched user identity pairs from the unmatched ones [10, 20, 22–24, 26, 27, 36, 46]. Man et al. [20] proposed a supervised embedding based UIL framework, in which the topology structures were mapped into a low dimensional space, and then a projection method was introduced to perform linkage in the latent space. ULink [23] first projected the user identities in multiple networks into a latent user space, and then minimized the distance between the user identities of the same person and maximized the distance between user identities belonging to different people. Zhang et al. [46] proposed an online UIL model under a cost-sensitive setting. They extracted a set of extensive profiles and then a probabilistic classifier was applied. As it is difficult and extremely expensive to obtain the annotations across different networks, several semi-supervised methods are proposed to utilize the similarities between unlabeled samples along with a few annotations to link user identities [11, 18, 34, 43, 45, 50, 52]. Korula et al. [11] introduced the label propagation, a popular semi-supervised model, to perform UIL task according to neighborhood-based network features. CosNet [50] was an energy-based model to link user identities by considering both local and global consistency. COSNET first extracted distance-based profile features and neighborhood-based topology features and then used an aggregating algorithm to obtain local consistency. HYDRA [19] is a multi-objective semi-supervised framework by

Table 3: Calculated EMD on DBLP datasets.

Dataset	EMD	Average percentage of matched nodes
DBLP15-DBLP16	0.56	39.3%
DBLP16-DBLP17	0.52	39.6%
DBLP15-DBLP17	0.74	29.7%

jointly modeling heterogeneous behaviors and structure consistency by a propagation algorithm.

Compared to the UIL model with annotations, unsupervised user linkage problem is less explored [12, 13, 17, 29]. Liu et al. [17] proposed to first automatically generate a set of training samples according to the rareness of the user names in two social networks, and then use these samples to train a binary classification. Lacoste-Julien et al. [13] proposed a greedy approach to align the attributes of users according to the heuristic string similarities. POIS [29] used the trajectory-based attribute features to link user identities. POIS first calculated the affinity scores between user identities based on the timestamped location data, and then a maximum weighted matching scheme is utilized to find the most likely matching pairs. Zhong et al. [51] propose a co-training model CoLink to iteratively optimize the topology-based and the attribute-based objectives. Although CoLink was claimed as an unsupervised model, it actual is a weakly-supervised method as the few annotations are still needed to train the co-training algorithm. Existing unsupervised UIL models usually heavily rely on the strong discriminative features or the cross-network common attributes, which greatly limits its generality. Hence in this paper, we aim to propose a general unsupervised model to perform UIL in the distribution level, which contributes to solve a set of social network applications [37, 38].

6 CONCLUSION

In this paper, we studied the novel problem of unsupervised UIL. The insight is that we link user identities from the distribution level, hence no annotations are required. We transformed the unsupervised UIL problem to the learning of a projection function to minimize the EMD between the distributions of user identities in two social networks. Furthermore, we proposed two unsupervised UIL models. UIL_{gan} is a GAN based model to minimize the duality of Wasserstein distance, while UIL_{omt} is an orthogonal matrix transformation based methods to minimize the EMD between distributions. We evaluated the proposed models on several social network datasets and co-author network datasets. The experimental results demonstrated the combination of the two proposed models outperformed the state-of-the-art unsupervised baselines and presented comparative performance with supervised baselines. In addition, we found the calculated EMD can be also utilized to measure the difference between two networks.

ACKNOWLEDGMENTS

This work was supported in part by the Natural Science Foundation of China (Grand Nos. U1636211, 61672081, 61370126, 61602237), Beijing Advanced Innovation Center for Imaging Technology (No. BAICIT-2016001), National Key R&D Program of China (No. 2016QY04W0802) and Natural Science Foundation of Jiangsu Province (No. BK20171420)

. This work was also supported in part by NSF through grants IIS-1526499, IIS-1763325, and CNS-1626432, and NSFC 61672313. We appreciate the helps from Meng Zhang (Tsinghua University) and Feiran Huang (Beihang University).

REFERENCES

- [1] Martin Arjovsky, Soumith Chintala, and Léon Bottou. 2017. Wasserstein gan. *arXiv* (2017), 1–32.
- [2] Santanu Chakraborty. 2008. Some applications of Dirac's delta function in statistics for more than one random variable. *Applications and applied mathematics* (2008), 42–54.
- [3] Scott Cohen and L. Guibas. 1999. The earth mover's distance under transformation sets. In *ICCV. IEEE*, 1076–1083.
- [4] Marco Cuturi. 2013. Sinkhorn distances: Lightspeed computation of optimal transport. In *NIPS*. 2292–2300.
- [5] Marco Cuturi and Arnaud Doucet. 2014. Fast computation of Wasserstein barycenters. In *ICML*. 685–693.
- [6] Hu Ding and Jinhui Xu. 2013. FPTAS for minimizing earth mover's distance under rigid transformations. In *European Symposium on Algorithms*. Springer, 397–408.
- [7] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *NIPS*. 2672–2680.
- [8] John C Gower and Garnt B Dijkstra. 2004. *Procrustes problems*. Oxford University.
- [9] Kurt Hornik. 1991. Approximation capabilities of multilayer feedforward networks. *Neural networks* (1991), 251–257.
- [10] Tereza Iofciu, Peter Fankhauser, Fabian Abel, and Kerstin Bischoff. 2011. Identifying Users Across Social Tagging Systems. In *ICWSM*.
- [11] Nitish Korula and Silvio Lattanzi. 2014. An efficient reconciliation algorithm for social networks. *VLDB* (2014), 377–388.
- [12] Sebastian Labitzke, Irina Taranu, and Hannes Hartenstein. 2011. What your friends tell others about you: Low cost linkability of social network profiles. In *Social Network Mining and Analysis*. 1065–1070.
- [13] Simon Lacoste-Julien, Konstantina Palla, Alex Davies, Gjergji Kasneci, Thore Graepel, and Zoubin Ghahramani. 2013. Sigma: Simple greedy matching for aligning large knowledge bases. In *KDD. ACM*, 572–580.
- [14] Elizaveta Levina and Peter Bickel. 2001. The earth mover's distance is the mallows distance: Some insights from statistics. In *ICCV. IEEE*, 251–256.
- [15] Chaozhuo Li, Zhoujun Li, Senzhang Wang, Yang Yang, Xiaoming Zhang, and Jianshe Zhou. 2017. Semi-supervised network embedding. In *International Conference on Database Systems for Advanced Applications*. Springer, 131–147.
- [16] Chaozhuo Li, Senzhang Wang, Dejian Yang, Zhoujun Li, Yang Yang, Xiaoming Zhang, and Jianshe Zhou. 2017. PPNE: property preserving network embedding. In *International Conference on Database Systems for Advanced Applications*. Springer, 163–179.
- [17] Jing Liu, Fan Zhang, Xinying Song, Young-In Song, Chin-Yew Lin, and Hsiao-Wuen Hon. 2013. What's in a name?: An unsupervised approach to link users across communities. In *WSDM. ACM*, 495–504.
- [18] Li Liu, William K Cheung, Xin Li, and Lejian Liao. 2016. Aligning Users across Social Networks Using Network Embedding. In *IJCAI*. 1774–1780.
- [19] Siyuan Liu, Shuhui Wang, Feida Zhu, Jinbo Zhang, and Ramayya Krishnan. 2014. Hydra: Large-scale social identity linkage via heterogeneous behavior modeling. In *KDD. ACM*, 51–62.
- [20] Tong Man, HuaWei Shen, Shenghua Liu, Xiaolong Jin, and Xueqi Cheng. 2016. Predict Anchor Links across Social Networks via an Embedding Approach. In *IJCAI*. 1823–1829.
- [21] Brendan D McKay et al. 1981. Practical graph isomorphism. (1981).
- [22] Marti Motoyama and George Varghese. 2009. I seek you: searching and matching individuals in social networks. In *WSDM. ACM*, 67–75.
- [23] Xin Mu, Feida Zhu, Ee-Peng Lim, Jing Xiao, Jianzong Wang, and Zhi-Hua Zhou. 2016. User identity linkage by latent user space modelling. In *KDD. ACM*, 1775–1784.
- [24] Yuanping Nie, Yan Jia, Shudong Li, Xiang Zhu, Aiping Li, and Bin Zhou. 2016. Identifying users across social networks based on dynamic core interests. *Neurocomputing* (2016), 107–115.
- [25] Ofir Pele and Michael Werman. 2009. Fast and robust earth mover's distances. In *ICCV. IEEE*, 460–467.
- [26] Olga Peled, Michael Fire, Lior Rokach, and Yuval Elovici. 2013. Entity matching in online social networks. In *SocialCom. IEEE*, 339–344.
- [27] Daniele Perito, Claude Castelluccia, Mohamed Ali Kaafar, and Pere Manils. 2011. How unique and traceable are usernames?. In *International Symposium on Privacy Enhancing Technologies Symposium*. Springer, 1–17.
- [28] Zhou Ren, Junsong Yuan, and Zhengyou Zhang. 2011. Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera. In *MM. ACM*, 1093–1096.
- [29] Christopher Riederer, Yunsung Kim, Augustin Chaintreau, Nitish Korula, and Silvio Lattanzi. 2016. Linking users across domains with location data: Theory and validation. In *WWW*. 707–719.
- [30] Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. 1998. A metric for distributions with applications to image databases. In *ICCV. IEEE*, 59–66.
- [31] Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. 2000. The earth mover's distance as a metric for image retrieval. *IJCV* (2000), 99–121.
- [32] Kai Shu, Suhang Wang, Jiliang Tang, Reza Zafarani, and Huan Liu. 2017. User identity linkage across online social networks: A review. *ACM SIGKDD Explorations Newsletter* (2017), 5–17.
- [33] Samuel L Smith, David HP Turban, Steven Hamblin, and Nils Y Hammerla. 2017. Offline bilingual word vectors, orthogonal transformations and the inverted softmax. *arXiv* (2017).
- [34] Shulong Tan, Ziyu Guan, Deng Cai, Xuzhen Qin, Jiajun Bu, and Chun Chen. 2014. Mapping Users across Networks by Manifold Alignment on Hypergraph. In *AAAI*. 159–165.
- [35] Cédric Villani. 2008. *Optimal transport: old and new*. Vol. 338. Springer Science & Business Media.
- [36] Jan Vosecky, Dan Hong, and Vincent Y Shen. 2009. User identification across multiple social networks. In *Networked Digital Technologies. IEEE*, 360–365.
- [37] Senzhang Wang, Xia Hu, Philip S Yu, and Zhoujun Li. 2014. MMRate: inferring multi-aspect diffusion networks with multi-pattern cascades. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM*, 1246–1255.
- [38] Senzhang Wang, Zhao Yan, Xia Hu, S Yu Philip, and Zhoujun Li. 2015. Burst Time Prediction in Cascades. In *AAAI*. 325–331.
- [39] Cheng Yang, Zhiyuan Liu, Deli Zhao, Maosong Sun, and Edward Y Chang. 2015. Network Representation Learning with Rich Text Information. In *IJCAI*. 2111–2117.
- [40] Dejian Yang, Senzhang Wang, Chaozhuo Li, Xiaoming Zhang, and Zhoujun Li. 2017. From properties to links: deep network embedding on incomplete graphs. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. ACM*, 367–376.
- [41] Reza Zafarani and Huan Liu. 2014. Users joining multiple sites: distributions and patterns. In *ICWSM*.
- [42] Reza Zafarani and Huan Liu. 2016. Users joining multiple sites: Friendship and popularity variations across sites. *Information Fusion* (2016), 83–89.
- [43] Reza Zafarani, Lei Tang, and Huan Liu. 2015. User identification across social media. *TKDD* (2015), 16.
- [44] Baichuan Zhang and Mohammad Al Hasan. 2017. Name Disambiguation in Anonymized Graphs using Network Embedding. In *Proceedings of the 26th ACM International on Conference on Information and Knowledge Management*.
- [45] Baichuan Zhang, Tanay Kumar Saha, and Mohammad Al Hasan. [n. d.]. Name disambiguation from link data in a collaboration graph. In *IEEE/ACM 5th International Conference on Advances in Social Networks Analysis and Mining*. 81–84.
- [46] Haochen Zhang, Min-Yen Kan, Yiqun Liu, and Shaoping Ma. 2014. Online social network profile linkage. In *Asia Information Retrieval Symposium*. Springer, 197–208.
- [47] Jiawei Zhang, Philip S Yu, and Zhi-Hua Zhou. 2014. Meta-path based multi-network collective link prediction. In *KDD. ACM*, 1286–1295.
- [48] Meng Zhang, Yang Liu, Huanbo Luan, and Maosong Sun. 2017. Earth Mover's Distance Minimization for Unsupervised Bilingual Lexicon Induction. In *EMNLP*. 1934–1945.
- [49] Meng Zhang, Haoruo Peng, Yang Liu, Huan-Bo Luan, and Maosong Sun. 2017. Bilingual Lexicon Induction from Non-Parallel Data with Minimal Supervision. In *AAAI*. 3379–3385.
- [50] Yutao Zhang, Jie Tang, Zhilin Yang, Jian Pei, and Philip S Yu. 2015. Cosnet: Connecting heterogeneous social networks with local and global consistency. In *KDD. ACM*, 1485–1494.
- [51] Zexuan Zhong, Yong Cao, Mu Guo, and Zaiqing Nie. 2018. CoLink: An Unsupervised Framework for User Identity Linkage. In *AAAI*. 3379–3385.
- [52] Xiaoping Zhou, Xun Liang, Haiyan Zhang, and Yuefeng Ma. 2016. Cross-platform identification of anonymous identical users in multiple social media networks. *TKDE* (2016), 411–424.