# Adversarial Learning for Weakly-Supervised Social Network Alignment

**Chaozhuo Li,[1] Senzhang Wang,[2] Yukun Wang,[3] Philip Yu,[4,5] Yanbo Liang,[6] Yun Liu,[1] Zhoujun Li[1]***
[1]State Key Lab of Software Development Environment, Beihang University; [2] College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics; [3]Department of Electrical Computer Engineering, National University of Singapore; [4]Tsinghua University; [5]Computer Science Department, University of Illinois at Chicago; [6]Hortonworks,USA
{lichaozhuo,gz_liuyun,lizj}@buaa.edu.cn, szwang@nuaa.edu.cn, wykun06@gmail.com, psyu@uic.edu, yliang@apache.org

## Abstract

Nowadays, it is common for one natural person to join multiple social networks to enjoy different kinds of services. Linking identical users across multiple social networks, also known as social network alignment, is an important problem of great research challenges. Existing methods usually link social identities on the pairwise sample level, which may lead to undesirable performance when the number of available annotations is limited. Motivated by the isomorphism information, in this paper we consider all the identities in a social network as a whole and perform social network alignment from the distribution level. The insight is that we aim to learn a projection function to not only minimize the distance between the distributions of user identities in two social networks, but also incorporate the available annotations as the learning guidance. We propose three models SNNA$_u$, SNNA$_b$ and SNNA$_o$ to learn the projection function under the weakly-supervised adversarial learning framework. Empirically, we evaluate the proposed models over multiple datasets, and the results demonstrate the superiority of our proposals.

## Introduction

Recently social networks are becoming increasingly popular, and users can register in multiple platforms simultaneously to enjoy different types of services. In each social platform, a user can create an identity to represent his/her unique personal figure. Aligning identities of the same natural person across multiple social platforms, which is refereed to *Social Network Alignment*, has attracted increasing attention considering its tremendous practical value. The successful network alignment benefits many applications, such as friend recommendation (Shu et al. 2017), information diffusing prediction (Zafarani and Liu 2014; Wang et al. 2014; Zhan et al. 2015) and network dynamics analysis (Wang et al. 2015; Zafarani and Liu 2016).

Most existing methods are supervised, which need a large number of manually labeled samples to train a classifier to separate matched identity pairs from the non-matched ones (Motoyama and Varghese 2009; Vosecky, Hong, and Shen 2009; Iofciu et al. 2011; Perito et al. 2011; Peled et al. 2013; Zhang et al. 2014; Mu et al. 2016; Man et al. 2016; Nie et al. 2016). Considering the high cost of obtaining the labeled instances, several semi-supervised approaches are proposed to incorporate the unlabeled instances to provide complementary information (Tan et al. 2014; Korula and Lattanzi 2014; Zafarani, Tang, and Liu 2015; Zhang et al. 2015; Liu et al. 2016; Zhong et al. 2018). Semi-supervised methods can utilize the unlabeled data to help capture the shape of the underlying data distribution, which are more promising to perform social network alignment in practice.

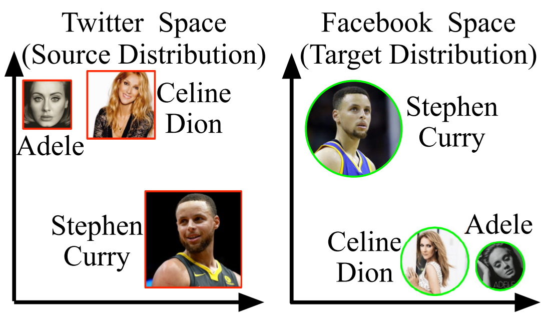
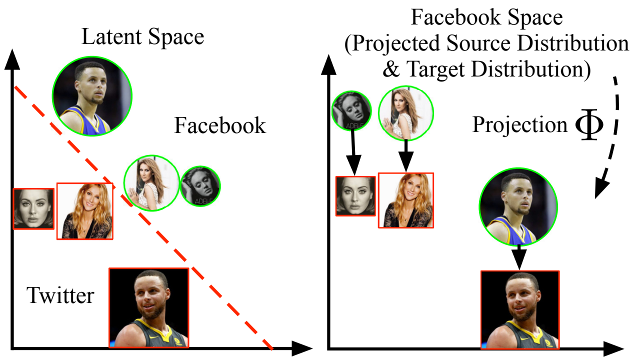



Figure 1: Original feature spaces of Twitter and Facebook.

Existing semi-supervised methods usually perform network alignment on the pairwise sample level. They first embed identities from different social networks into a common latent space, in which the similar identities from the same network should be closely distributed while the annotated identity pairs across social networks also should be grouped together. Then the distance between identities is viewed as the indicator of network alignment. A major limitation of such methods is that they still need plenty of annotations to ensure the performance. As shown in Figure 1, Celine Dion and Adele are popular singers and Stephen Curry is a NBA player. Celine Dion and Adele are closer in both platforms (Twitter and Facebook) considering their shared interests. Assume Celine Dion is selected as the annotated identity pair. Traditional sample level semi-supervised methods will generate the latent space as shown in Figure 2a. The identities of Celine Dion distributed closely while the latent space preserves the original identity similarities inside the single network. However, for Stephen Curry in Twitter, his nearest Facebook neighbor is Adele rather than his identity in Facebook, leading to a mistake match.

*Zhoujun Li is the corresponding author of this paper.




(a) Sample-level latent space. (b) Distribution-level space.

Figure 2: The illustration of traditional sample-level space and the distribution-level space.

Different of previous semi-supervised works, we view all the identities in a social network as a whole and perform identity alignment in the distribution level. As shown in Figure 1, the distributions of identities in the two original feature spaces present similar shape due to the shared interests, which is refereed to the isomorphism across social networks (McKay and others 1981; Li et al. 2018). As shown in Figure 2b, if we can convert the identity distribution in Twitter space by a set of operations $\Phi$ (e.g., transposing) to minimize the distance between it and the identity distribution in Facebook, the identities of same natural person will be grouped close to each other. Inspired by the isomorphism, we transform the social network alignment problem to the learning of the operation $\Phi$ to minimize the distance between two distributions. Following the previous work (Mu et al. 2016), we refer the operation $\Phi$ to the projection function. By introducing the isomorphism from the global perspective, the requirement of sample level supervisions is further reduced.

The motivation requires a metric of distribution distance, for which we introduce the wasserstein distance (WD). Compared with other metrics such as KL divergence, WD is symmetric and able to measure the distance between two distributions even if they have no overlap (Arjovsky, Chintala, and Bottou 2017). We view each identity distribution as a set of weighted points, and the WD measures the minimum cost of transporting one set of points into the other. However, the WD minimization is performed on the distribution level in the unsupervised manner, while the labeled identity pairs preserve the guidance information in the sample level. Considering the totally different purposes and scenarios, it is challenging to utilize the available annotations as indicators to guide the distribution minimization problem.

In this paper, we introduce an adversarial learning framework named SNNA to solve the weakly-supervised identity alignment problem. The discriminator is designed to estimate the WD between the projected source distribution and the target distribution, while the projection function is viewed as the generator to minimize the approximated WD. Through the competition between the generator and the discriminator, the approximated WD can be minimized while the projection function can be learned to find a neighborhood of a good optimum. We also design another objective function to incorporate available annotations, which guides the projected point of a source identity close to its corresponding target identity. The two objective functions will be jointly trained under a unified framework. Specifically, we propose three variants of SNNA model including a unidirectional model SNNA$_u$, a bidirectional model SNNA$_b$ and a model SNNA$_o$ to introduce more stricter orthogonal restriction. The experimental results demonstrate our proposals significantly outperform the baseline methods.

We summarize our main contributions as follows.

- We study the novel problem of weakly-supervised social network alignment from a new perspective. The distribution closeness is introduced to provide the complementary information.
- We design three adversarial learning based models to minimize the distribution distance and incorporate the available annotations simultaneously.
- Extensively, we evaluate the proposals on five groups of datasets. Experimental results show the superior performance of the proposed models.

## Preliminaries and Problem Definition

### Wasserstein Distance

WD measures the closeness between two distributions by estimating the minimum amount of works to change one distribution into the other. WD can be formally defined as follows (Arjovsky, Chintala, and Bottou 2017; Zhang et al. 2017b):

$$\mathrm{W}(\mathbb{P}^I, \mathbb{P}^J) = \inf_{\gamma \in \Gamma(\mathbb{P}^I, \mathbb{P}^J)} \mathbb{E}_{(x,y)\sim\gamma}[d(x, y)]. \tag{1}$$

In this task $\mathbb{P}^I$ and $\mathbb{P}^J$ are two discrete probability distributions in the form of $\mathbb{P} = \sum_i p_i \delta_{x_i}$, in which $x_i$ is a sample in the distribution $\mathbb{P}$, $p_i$ is its corresponding probability and $\delta_{x_i}$ is the Dirac delta function (Chakraborty 2008). $\Gamma(\mathbb{P}^I, \mathbb{P}^J)$ represents the joint probability distribution $\gamma(x, y)$ with marginals $\mathbb{P}^I$ and $\mathbb{P}^J$. Function $d$ measures the ground distance (e.g., Euclidean distance) between two samples. WD aims to find the desirable joint distribution $\Gamma$ to reach the expectation infimum.

### Problem Definition

We denote a social network as $N = \{V, W, P\}$, where $V = \{v_1, v_2, \cdots, v_n\}$ is the user set containing $n$ users. Each user $v_i$ is represented by a $d$-dimensional feature vector $w_i$, which forms the feature matrix $W \in \mathbb{R}^{d\times n}$. $P \in \mathbb{R}^{1\times n}$ contains the topology influence of the social users, such as the count of the in-degrees or out-degrees. We formally define the studied problem as follows:

**Definition 1 Weakly-supervised Social Network Alignment.** Given two partially aligned social networks $O = \{V_O, W_O, P_O\}$, $E = \{V_E, W_E, P_E\}$ and a few available matched identity pairs $M = \{(v_o, v_e)|v_o \in V_O, v_e \in V_E\}$, we aim to find all the other matched identity pairs $Y = \{(v_o, v_e)|v_o \in V_O, v_e \in V_E, (v_o, v_e) \notin M\}$, in which $v_o$ and $v_e$ belong to the same natural person.

We assume the dimension of feature vectors in both networks is $d$, which can be easily satisfied by the popular network embedding models (Li et al. 2017a; Li et al. 2017b; Yang et al. 2017). Here we aim to learn a desirable projection function to match identities. Thus the studied problem can be further clarified as follows:

**Definition 2 Projection Function Learning for Social Alignment.** Given the source distribution $\mathbb{P}^O$, the target distribution $\mathbb{P}^E$ and the annotation set $M = \{(v_o, v_e)|v_o \in V_O, v_e \in V_E\}$, we aim to learn a projection function $\Phi$ which satisfies: 1) $\Phi$ should minimize the wasserstein distance between the projected source distribution $\mathbb{P}^{\Phi(O)}$ and the target distribution $\mathbb{P}^E$; and 2) for a matched identity pair $(v_o, v_e)$ in $M$, $\Phi$ should minimize the distance between the projected source point $\Phi(v_o)$ and the target point $v_e$.

After the training process, given a source identity $v_o$, his/her matched candidates can be selected according to the ground distance $d(\Phi(v_o), v_e)$ with the identity $v_e$ in target social network. A smaller ground distance means the two identities has a larger chance to be the same natural person.

## Adversarial Learning Framework

Following previous works (Mu et al. 2016; Man et al. 2016; Li et al. 2018), we choose the linear transformation as the projection function $\Phi$. Given a source node $v_o$ with its feature vector $w_o$, its projected point is defined as: $\Phi(w_o) = G \times w_o$, where $G \in \mathbb{R}^{d\times d}$ is the transformation matrix. The studied problem can be understood as the learning of the matrix $G$. We also tried non-linear projection functions using neural networks but they do not work well. This may be because the non-linear projection seriously alters the input distribution and further destroys the isomorphism. Zhang et al. (Zhang et al. 2017a; Zhang et al. 2017b) introduce GAN model to perform bilingual lexicon induction task. Inspired by this work, we design the following SNNA models, which can not only minimize the WD but also incorporate the annotations.

**Unidirectional Projection Model SNNA$_u$** Firstly we introduce the unidirectional projection model SNNA$_u$, which only projects the source distribution to the target social space in one-way. Figure 3 shows the framework of SNNA$_u$. The generator $G$ can be considered as the projection function $\Phi$ while the discriminator $D$ is designed to estimate the WD between the projected source distribution $\mathbb{P}^{G(O)}$ and the target distribution $\mathbb{P}^E$. The objective of SNNA$_u$ can be formally defined as follows:

$$\min_G \mathrm{W}(\mathbb{P}^E, \mathbb{P}^{G(O)}) = \inf_{\gamma\in\Gamma(\mathbb{P}^E,\mathbb{P}^{G(O)})} \mathbb{E}_{(w_e, Gw_o)\sim\gamma}[d(w_e, Gw_o)]$$

where $w_o$ is the feature vector of the source identity $v_o$ sampled from the source distribution $\mathbb{P}^O$ according to its topology influence $p_o$. $w_e$ is sampled from the target distribution in a similar way. It is intractable to traverse all the possible joint distributions to compute the expectation infimum $\inf_{\gamma\in\Gamma(\mathbb{P}^E,\mathbb{P}^{G(O)})}$ (Zhang et al. 2017b). Vallani et al. (Villani 2008) proposed a simple version of the WD minimization objective based on the Kantorovich-Rubinstein duality when the ground distance $d$ is defined as the Euclidean distance:

$$\mathrm{W} = \frac{1}{K} \sup_{\|f\|_L \le K} \mathbb{E}_{w_e\sim\mathbb{P}^E} f(w_e) - \mathbb{E}_{Gw_o\sim\mathbb{P}^{G(O)}} f(Gw_o)$$

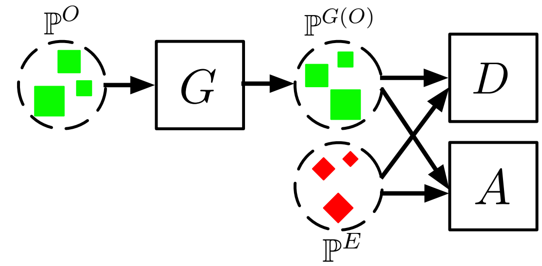



Figure 3: An illustration of the unidirectional projection model SNNA$_u$.

Function $f$ is required to be K-Lipschitz continuous, which means $|f(x_1) - f(x_2)| \leqslant K|x_1 - x_2|$ for all $x_1, x_2 \in \mathbb{R}$ and $K \geqslant 0$ is the Lipschitz constant. This objective aims to locate the supremum over all the possible K-Lipschitz functions. Feed forward neural networks own powerful approximation capabilities (Hornik 1991). Hence we select a multilayer feed forward network to find a desirable function $f$, which is defined as the discriminator $D$ in Figure 3. The objective function of the discriminator is to learn a desirable function $f$ to estimate the WD between $\mathbb{P}^E$ and $\mathbb{P}^{G(O)}$:

$$\max_{\theta:\|f_\theta\|_L\le K} L_D = \mathbb{E}_{w_e\sim\mathbb{P}^E}[f_\theta(w_e)] - \mathbb{E}_{Gw_o\sim\mathbb{P}^{G(O)}}[f_\theta(Gw_o)] \tag{2}$$

in which $\theta$ is the parameter set in the multi-layer neural network used in the discriminator. We introduce the clipping trick (Arjovsky, Chintala, and Bottou 2017) to satisfy the K-Lipschitz restriction, which clamps the weights $\theta$ to a small window $[-c, c]$ after every gradient updating.

The generator $G$ is designed to minimize the estimated WD. In Formula (2), $G$ only exists in the second term and thus we aim to learn a desirable generator by minimizing the following objective:

$$\min_{G\in\mathbb{R}^{d\times d}} L_G = -\mathbb{E}_{Gw_o\sim\mathbb{P}^{G(O)}}[f_\theta(Gw_o)] \tag{3}$$

With the decreasing of the generator loss, the WD estimated by the discriminator will be gradually reduced, leading to the identities belonging to the same person grouped together in the target space.

Meanwhile, we also incorporate a few annotations to guide the learning process of the projection function, which is shown as the component $A$ in Figure 3. Assuming in a training batch, we have a set of source identities and their matched target identities denoted as $M_t \subset M$. For the matched identity pair $(v_o, v_e)$ in the labeled set $M_t$, we aim to minimize the distance between the projected source node $G(v_o)$ and the target node $v_e$:

$$\min_{G\in\mathbb{R}^{d\times d}} L_C = \frac{\lambda_c}{|M_t|} \sum_{(v_o,v_e)\in M_t} d(Gw_o, w_e) \tag{4}$$

where $w$ is the feature vector of the corresponding identity. This objective incorporates the available annotations to facilitate the learning of the projection function. $\lambda_c$ is a hyperparameter to control the weight of loss $L_C$.

Here we briefly introduce the training steps of SNNA$_u$ in Algorithm 1. Line 2 to 12 represents a training iteration.

**Algorithm 1** Training process of SNNA$_u$

**Require:** the learning rate $\alpha$, the clipping weight $c$, the minimal training batch size $m$, the number of discriminator training in each loop $n_d$ and the annotation guided weight $\lambda_c$.
**Require:** the initial generator parameters $G_0$, the initial discriminator parameters $\theta_0$.
1: **while** $G$ has not converged **do**
2:   **for** $i = 0 \to n_d$ **do**
3:     Sample a batch from source distribution: $\{w_o^{(i)}\}_{i=1}^m$
4:     Sample a batch from target distribution: $\{w_e^{(i)}\}_{i=1}^m$
5:     $g_\theta \leftarrow \nabla_\theta[\frac{1}{m}\sum_{i=1}^m f_\theta(w_e^{(i)}) - \frac{1}{m}\sum_{i=1}^m f_\theta(Gw_o^{(i)})]$
6:     $\theta \leftarrow \theta + \alpha\cdot$ RMSProp$(\theta, g_\theta)$
7:     $\theta \leftarrow$ clip$(\theta, -c, c)$
8:   **end for**
9:   Sample a batch from source distribution: $\{w_o^{(i)}\}_{i=1}^m \sim \mathbb{P}^O$
10:   $g_G \leftarrow \nabla_G(\frac{1}{m}\sum_{i=1}^m(-f_\theta(Gw_o^{(i)}))$
11:   $g'_G \leftarrow \nabla_G \lambda_c \cdot \frac{1}{|M_t|}\sum_{(v_o,v_e)\in M_t} d(Gw_o, w_e)$
12:   $G \leftarrow G - \alpha\cdot$ RMSProp$(G, g_G + g'_G)$
13: **end while**

Firstly, we train the discriminator $n_d$ times from line 2 to 8, which is designed to avoid the collapsed GAN risk (Arjovsky, Chintala, and Bottou 2017). Then the generator is updated by minimizing the weighted combination of objectives (3) and (4) as shown from line 9 to 13, which means the learned generator not only minimizes the estimated WD but also fits the available annotations.

**Bidirectional Projection Model SNNA$_b$** The unidirectional model SNNA$_u$ has no constraint on the projection matrix $G$. Mu et al. (Mu et al. 2016) has proven the orthogonal projection contributes to better aligning user identities. An orthogonal projection is theoretically appealing for its numerical stability (Smith et al. 2017; Li et al. 2018). With the orthogonal projection, the projected distribution is the reflection of the original distribution in a plane by rotating and scaling, which will preserve the interior characteristics of the original distribution. Hence we add the orthogonal constraint on the learning of projection matrix.

Introducing the traditional orthogonal constraint into adversarial learning is cumbersome as their optimizations are intractable (Smith et al. 2017; Zhang et al. 2017a). We further design a bidirectional projection model SNNA$_b$ to latently introduce the orthogonal constraint. As shown in Figure 4, SNNA$_b$ performs the projection in both directions. If the projection function $G$ minimizes the WD between distributions $\mathbb{P}^{G(O)}$ and $\mathbb{P}^E$, its transpose version $G^\intercal$ should be able to minimize the WD between distributions $\mathbb{P}^{G(E)}$ and $\mathbb{P}^O$. Considering the input networks are partially aligned while distributions $\mathbb{P}^E$ and $\mathbb{P}^O$ are different, the learned projection matrix can only be self-consistent, which can be close to the orthogonal matrix but not exactly orthogonal.

SNNA$_b$ model can be easily implemented by two SNNA$_u$ models with a shared projection matrix $G$. The first SNNA$_u$ model utilizes the projection function $G$ as the generator and the discriminator $D_e$ to estimate the WD between the distributions $\mathbb{P}^{G(O)}$ and $\mathbb{P}^E$. The second SNNA$_u$ model utilizes the generator $G^\intercal$ as the projection and the discriminator $D_o$

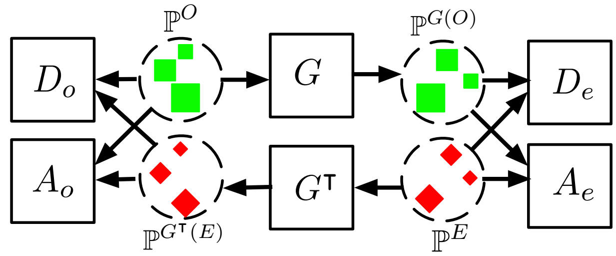



Figure 4: An illustration of the bidirectional projection model SNNA$_b$.

to estimate the WD between the distributions $\mathbb{P}^{G^\intercal(E)}$ and $\mathbb{P}^O$. The optimization of the two SNNA$_u$ models are processed iteratively. After the training finished, we still use the learned projection function $G$ to select target candidates for the source identities in network $O$.

**Orthogonal Projection Model SNNA$_o$** Different from the self-consistent assumption used in SNNA$_b$, we further introduce a stricter orthogonal constraint. If $G$ is an orthogonal matrix, the source distribution should be easily recovered from its projected version with the transpose matrix: $G^\intercal G w_o = w_o$, which ensures the social identities and the natural people can be transformed in bi-direction. As the reconstructed distribution has potential to be same to the original one, the learned projection matrix is more closer to the orthogonal matrix than the SNNA$_b$ mode. Hence, we design a reconstruction component to integrate the stricter orthogonal constraint into the adversarial training model.

As shown in Figure 5, we aim to reconstruct the original source distribution $\mathbb{R}^O$ from the projected distribution $\mathbb{R}^{G(O)}$ with the transpose matrix $G^\intercal$. The reconstructed source distribution is defined as $\mathbb{R}^{O'}$, and we introduce the following objective function to minimize the difference between the original distribution and the reconstructed one:

$$\min_{G\in\mathbb{R}^{d\times d}} L_R = \lambda_r \mathbb{E}_{w_o\sim\mathbb{P}^O}[d(w_o, G^\intercal G w_o)] \qquad (5)$$

where $\lambda_r$ is a hyper-parameter to control the weight of reconstruction errors. With the minimization of the loss $L_R$, the learned matrix $G$ will be more orthogonal. The training process of SNNA$_o$ can be easily expanded from SNNA$_u$ by adding the new defined reconstruction loss into line 12 of the Algorithm 1. The projection matrix will be learned to be orthogonal, fitting the available annotations and minimizing the WD between the projected source distribution and the target distribution. The loss function of generator in SNNA$_o$ is the weighted sum of the distance minimization loss $L_G$, annotation guided loss $L_C$ and the reconstruction loss $L_R$.

## Experiments

**Datasets** We use two pairs of social network datasets and three pairs of academic co-author datasets for evaluation. The datasets are crawled and formatted by our corporation coauthor. Table 1 shows the detailed statistics.

- **Twitter-Flickr**: Twitter and Flickr are two popular social networks, and it is difficult to obtain the matched identities across these two platforms. Fortunately, social users

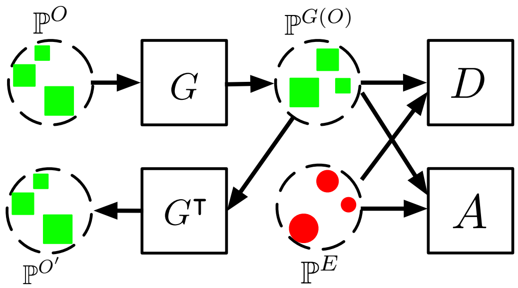



Figure 5: An illustration of the orthogonal projection model SNNA$_o$.

can link and present their identities of different social platforms in the about.me website. Based on the matched pairs collected from about.me, we can crawl the social data of the same natural person from Twitter and Flickr. After removing users with scarce attributes, we finally obtain 3259 Twitter identities and 4308 Flickr identities, in which 2773 pairs are matched as the ground truth.

- **Weibo-Douban**: Sina Weibo is one of the most influential social platforms in China, and Douban is a social networking service website allowing registered users to record information and create content related to films, books and musics. Douban users can present their Weibo accounts in the homepage, and hence we can collect the matched identities as the ground truth based on this information. Besides we also randomly select a set of unmatched social identities from both platforms to form the partially aligned dataset. Finally we obtain 4119 Weibo identities and 4554 Douban identities, in which 3235 pairs are matched as the ground truth.
- **DBLP**: DBLP (*http://dblp.uni-trier.de/*) is a computer science bibliography website, and its dataset is publicly available[1]. We select the published papers along with the authors in three years (2015, 2016 and 2017) to form three co-author networks. For each year, we select *Yoshua Bengio* as the center node, and then construct a coauthor subnetwork by locating the coauthors who can be reached within three steps from the center node. The published papers of one author in this year are considered as his/her attributes. We aim to match two nodes in the different coauthor networks, and the author identities from the DBLP dataset are considered as the ground truth.

**Data Preprocessing** For the Twitter users, their feature space is constructed based on the published tweets and the friend following relations. For the tweet text, we first process the crawled tweets using NTLK[2] stemmer and removing stop or rare words. After that, the tweets published by a single user are collected as a whole and then represented by a $tf$-$idf$ feature vector. We further utilize an attribute preserving network embedding model TADW (Yang et al. 2015) to encapsulate the network topology information and the text attributes into the low-dimensional latent space. The feature

[1] http://dblp.uni-trier.de/xml/

[2] https://www.nltk.org/

Table 1: Statistics of the datasets. The numbers in the brackets are the counts of nodes. $\#_M$ is the number of matched identity pairs across networks.

| Dataset | Source Network | Target Network | $\#_M$ |
|---|---|---|---|
| Twi.-Fli. | Twitter (3,259) | Flickr (4,308) | 2,773 |
| Wei.-Dou. | Weibo (4,119) | Douban (4,554) | 3,235 |
| DBLP15-16 | DBLP15 (3,881) | DBLP16 (5,989) | 1,852 |
| DBLP16-17 | DBLP16 (5,989) | DBLP17 (7,073) | 2,570 |
| DBLP15-17 | DBLP15 (3,881) | DBLP17 (7,073) | 1,492 |

spaces of other datasets are also constructed using TADW, but differ in the user attributes (the published picture tags and joined groups for Flickr users, the microblog text and the hashtags for Sina Weibo users, and the interest tags and joined groups for Douban users). For the social networks, the normalized count of the followers is viewed as the topology weight $p_i$ of the user $v_i$, which will be utilized to sample the training samples from the identity distributions.

For the DBLP datset, we first construct three co-author subnetworks according to the publications in 2015, 2016 and 2017. For each node (author) in the co-author subnetworks, we collect the published papers of this author in the corresponding years, and view the titles and abstracts of the published papers as the node attributes. The text attributes are formatted into the $tf$-$idf$ vectors, and then are embedded into the latent feature vectors with the co-author relationships by TADW. For the co-author networks, we utilize the degree count of a node as its sampling weight. Note that, the feature spaces of different networks are learned independently, which ensures the generality of our proposals.

**Baseline Methods** We compare our models with the following state-of-the-art baseline methods, including both semi-supervised and supervised models.

- **MAH** (Tan et al. 2014) is a semi-supervised model that utilizes social structures to improve the linkage performance by a subspace learning algorithm.
- **COSNET** (Zhang et al. 2015) is an energy-based model considering both local and global consistency among multiple networks. An efficient subgradient algorithm is developed to train the model.
- **IONE** (Liu et al. 2016) is a unified optimization framework to jointly train the the network embedding objective for capturing the identity similarities, and the user alignment objective for linking identities across the networks.
- **CoLink** (Zhong et al. 2018) is a weakly-supervised model which employs a co-training algorithm to manipulate two independent components: the attribute-based model and the relationship-based model.
- **ULink** (Mu et al. 2016) is a supervised model to link identities by latent user space modeling.

**Parameter Setup** For our proposals, the dimension $d$ of the latent feature space is set to 100. The discriminator $D$ in all SNNA models is a multi-layer perceptron network with only one hidden layer, as a too powerful discriminator may lead

Table 2: Comparison with the baseline methods ($Hit\text{-}Precision$ score).

| | Twitter-Flickr | | | Weibo-Douban | | | DBLP15-16 | | | DBLP16-17 | | | DBLP15-17 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| k | k=3 | k=5 | k=10 | k=3 | k=5 | k=10 | k=3 | k=5 | k=10 | k=3 | k=5 | k=10 | k=3 | k=5 | k=10 |
| MAH | 0.132 | 0.153 | 0.192 | 0.125 | 0.142 | 0.191 | 0.277 | 0.309 | 0.354 | 0.275 | 0.305 | 0.356 | 0.267 | 0.311 | 0.363 |
| COSNET | 0.144 | 0.187 | 0.236 | 0.132 | 0.161 | 0.194 | 0.292 | 0.330 | 0.373 | 0.288 | 0.332 | 0.386 | 0.289 | 0.338 | 0.375 |
| IONE | 0.161 | 0.196 | 0.242 | 0.150 | 0.189 | 0.232 | 0.302 | 0.347 | 0.397 | 0.308 | 0.345 | 0.396 | 0.310 | 0.352 | 0.377 |
| CoLink | 0.193 | 0.225 | 0.267 | 0.171 | 0.193 | 0.244 | 0.322 | 0.379 | 0.414 | 0.310 | 0.345 | 0.400 | 0.317 | 0.366 | 0.395 |
| ULink | 0.141 | 0.162 | 0.199 | 0.113 | 0.142 | 0.198 | 0.283 | 0.318 | 0.359 | 0.304 | 0.317 | 0.375 | 0.278 | 0.325 | 0.366 |
| SNNA$_u$ | 0.228 | 0.244 | 0.295 | 0.215 | 0.246 | 0.282 | 0.342 | 0.388 | 0.437 | 0.323 | 0.353 | 0.427 | 0.331 | 0.376 | 0.423 |
| SNNA$_b$ | 0.235 | 0.252 | 0.304 | 0.237 | 0.252 | 0.298 | 0.353 | 0.394 | 0.441 | 0.332 | 0.379 | 0.439 | 0.344 | 0.382 | 0.437 |
| SNNA$_o$ | **0.263** | **0.283** | **0.321** | **0.251** | **0.282** | **0.311** | **0.383** | **0.420** | **0.461** | **0.350** | **0.399** | **0.457** | **0.373** | **0.417** | **0.469** |

to the corruption of the GAN training and make the generator lose the adversarial ability (Arjovsky, Chintala, and Bottou 2017). For the generator $G$, its projection matrix is randomly initialized as an orthogonal matrix. The size of minimal training batch is 256, and the learning rate $\alpha$ is set to $0.0001$. As mentioned in Algorithm 1, the discriminator will be trained $n_d$ times in each training iteration and $n_d$ is set to 5. The clipping weight $c$ is 0.01, the annotation weight $\lambda_c$ is set to 0.2 and the reconstruction weight $\lambda_r$ is set to 0.3. The baselines are implemented according to the original papers. For the CoLink model, we utilize SVM classifier as the attribute-based model. The ULink model is trained by the constrained concave convex procedure optimization.

**Evaluation Metric**. Following the previous work (Mu et al. 2016), we select $Hit\text{-}Precision$ as the evaluation metric:

$$h(x) = \frac{k - (hit(x) - 1)}{k} \tag{6}$$

where $hit(x)$ is the rank position of the matched target user in the returned top-k candidate target identities. The top candidates are selected according to the ground distances between the projected source identity and the target identities. The $Hit\text{-}Precision$ is calculated by the average on the scores of the matched identity pairs: $\frac{\sum_{i=0}^{i=m} h(x_i)}{m}$, in which $m$ is the number of source identities in the the matched pairs.

**Experimental Results** For each dataset, we randomly select $T_{tr}$ portion of matched identity pairs as the training data, and $N_{te}$ matched identity pairs are randomly selected as the test set. Here we fix the training ration $T_{tr}$ as 10%, and the size of test set $N_{te}$ is set to 500. We compare the proposed models with the baselines, and report the $Hit\text{-}Precision$ scores with different settings of $k$. We repeat this process three times and report the average scores.

Table 2 shows the experimental results. One can see that all the methods perform better on the DBLP datasets than on the social networks. This is probably because the co-author networks are more denser and the user attributes are formatted and clean. COSNET outperforms MAH method as it introduces both the global and local topology similarities. As a supervised model, ULink achieves an undesirable performance in the weakly-supervised learning setting as it needs a large portion of annotations (e.g., 80% for the original work) to achieve a desirable performance. CoLink achieves the best performance among the baselines, because it carefully designs an objective function to incorporate the attributes, while the attributed-based model and the relationship based model can facilitate each other.

One can also see that out proposals all outperform the baseline methods on both datasets with different settings. This is because the distribution closeness information is introduced as the complementary. The unidirectional model SNNA$_u$ beats the best baseline (CoLink) by around 3%. By introducing the self-consistent constraint, the performance of SNNA$_b$ is further improved, which proves a orthogonal projection matrix contributes to better aligning identities. SNNA$_o$ achieves the best performance among all the methods, which beats the best baseline (CoLink) by 7%. With the stricter orthogonal constraint, SNNA$_o$ further improves the performance by 3% compared with SNNA$_b$.

**Learning Behavior of the SNNA$_o$** The adversarial learning is also famous for its instable training behavior. Hence we present the training trajectory of SNNA$_o$ model on DBLP15-16 dataset with $k$=5. After each 1,0000 training batches, we will save a check point model, and finally we can obtain 100 check points models. For each checkpoint model, we record the output value from its discriminator as the approximated WD value, and its $Hit\text{-}Precision$ score on the social network alignment task. Note that we rescale the approximated Wasserstein distance to the range of 0 to 10. As shown in Figure 6, one can see that with the increasing of training batches, the WD decreases while the $Hit\text{-}Precision$ score increases. The results demonstrate that: 1) the proposed model can effectively reduce the WD in the dynamic distribution scenario; and 2) a smaller WD leading to a better network alignment performance. Therefore, we can save the check point model with the lowest estimated Wasserstein distance as the final model.

**Parameter Sensitivity Study.** Finally we analyze the parameter sensitivity of the proposed SNNA$_o$ model. We first analyze the effect of the training ratio $T_{tr}$ on the model performance over the Twitter-Flickr dataset. We fix $k$=5 and show the performance of SNNA$_o$ with different settings of $T_{tr}$. The best baseline CoLink is also used for comparison. From Figure 7a, one can see that with the increase of $T_{tr}$, the performance of both methods increases. SNNA$_o$ consistently outperforms the CoLink, while the performance gap between them tends to be smaller. This is because annotations can remedy the limitation of CoLink ignoring the distribution closeness information. Next we study the effect of the reconstruction weight $\lambda_r$ on the model

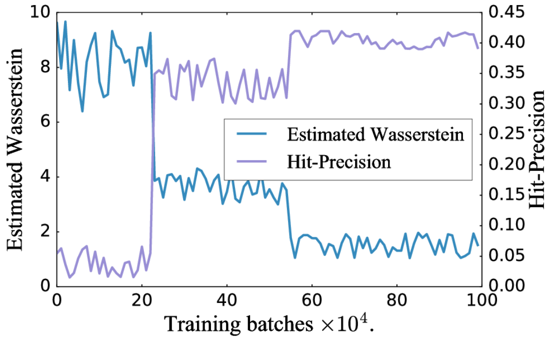
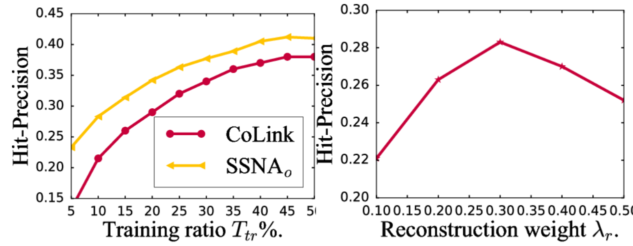



Figure 6: Training trajectory of SNNA$_o$.




(a) Sensitivity to $T_r$. (b) Sensitivity to $\lambda_r$.

Figure 7: Core parameter sensitivity analysis.

performance. When we increase the value of $\lambda_r$, the $Hit\text{-}Precision$ score first increases and then decreases, which demonstrates that incorporating appropriate orthogonal constraint contributes to improving the performance. A larger $\lambda_r$ will make the training process focus more on the reconstruction task, which may interrupt the optimization to reach the minimal wasserstein distance.

## Related Work

Existing social network alignment methods can be roughly categorized into supervised, semi-supervised and unsupervised methods. Most existing related works are supervised, which aim to train a binary classifier to separate the matched user identity pairs from the unmatched ones (Vosecky, Hong, and Shen 2009; Motoyama and Varghese 2009; Iofciu et al. 2011; Perito et al. 2011; Peled et al. 2013; Zhang et al. 2014; Mu et al. 2016; Man et al. 2016; Nie et al. 2016). Man et al. (Man et al. 2016) proposed a supervised social network alignment model by linking identities in the latent low dimensional space. The user identities from both networks are mapped into a latent space, and a projection function is learned to link identities belonging to the same natural person. ULink (Mu et al. 2016) is also an embedding base supervised approach. ULink first mapped the user identities in multiple networks into a latent space, and then minimized the distance between the user identities of the same person and maximize the distance between user identities belonging to different people. Considering it is non-trivial and time consuming to achieve enough annotations to fully train a supervised mode, some unsupervised methods are proposed, which mainly rely on the strong discriminative features to link user identities (Labitzke, Taranu, and Hartenstein 2011; Liu et al. 2013; Lacoste-Julien et al. 2013; Riederer et al. 2016). UUIL (Li et al. 2018) is the basis of this work, but UUIL focus on the unsupervised learning, while SNNA aims to incorporate few annotations to improve the alignment performance. Recently several semi-supervised methods are proposed to incorporate the unlabeled data to capture the inner data distribution (Zafarani, Tang, and Liu 2015; Korula and Lattanzi 2014; Zhang et al. 2015; Tan et al. 2014; Liu et al. 2016; Zhang, Saha, and Al Hasan ). Korula et al. (Korula and Lattanzi 2014) introduced the label propagation, a popular semi-supervised model, to perform UIL task according to neighborhood-based network features. CosNet (Zhang et al. 2015) was an energy-based model to link user identities by considering both local and global consistency. Existing semi-supervised methods usually link user identities from the pairwise sample level, which cannot achieve desirable performance with very limited annotations. Hence in this paper, we aim to perform social network alignment from the distribution level. The studied problem is converted to the learning of a distribution projection function, which can be solved under an adversarial training framework.

## Conclusion

In this paper, we studied the novel problem of weakly-supervised social network alignment. The insight is that we perform the social network alignment from the identity distribution level, which contributes to reduce the number of needed annotations. The studied problem is converted into the learning of a desirable projection function, which can not only minimize the wasserstein distance between the identity distributions from two social networks, but also group the available matched identities together in the projected space. Furthermore, we proposed three models SNNA$_u$, SNNA$_b$ and SNNA$_o$ with different levels of orthogonal constraints. We evaluated our proposals on multiple datasets ,and the experimental results proven the superiority of SNNA models.

## Acknowledgments

This work was supported the Natural Science Foundation of China (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 by NSF through grants IIS-1526499, IIS-1763325, and CNS-1626432, and NSFC 61672313. We appreciate the helps from Meng Zhang (Tsinghua University).

## References

Arjovsky, M.; Chintala, S.; and Bottou, L. 2017. Wasserstein gan. *arXiv* 1–32.

Chakraborty, S. 2008. Some applications of dirac's delta function in statistics for more than one random variable. *Applications and applied mathematics* 42–54.

Hornik, K. 1991. Approximation capabilities of multilayer feedforward networks. *Neural networks* 251–257.

Iofciu, T.; Fankhauser, P.; Abel, F.; and Bischoff, K. 2011. Identifying users across social tagging systems. In *ICWSM*.

Korula, N., and Lattanzi, S. 2014. An efficient reconciliation algorithm for social networks. *VLDB* 377–388.

Labitzke, S.; Taranu, I.; and Hartenstein, H. 2011. What your friends tell others about you: Low cost linkability of social network profiles. In *Social Network Mining and Analysis*, 1065–1070.

Lacoste-Julien, S.; Palla, K.; Davies, A.; Kasneci, G.; Graepel, T.; and Ghahramani, Z. 2013. Sigma: Simple greedy matching for aligning large knowledge bases. In *KDD*, 572–580. ACM.

Li, C.; Li, Z.; Wang, S.; Yang, Y.; Zhang, X.; and Zhou, J. 2017a. Semi-supervised network embedding. In *DASFAA*, 131–147. Springer.

Li, C.; Wang, S.; Yang, D.; Li, Z.; Yang, Y.; Zhang, X.; and Zhou, J. 2017b. Ppne: property preserving network embedding. In *DASFAA*, 163–179. Springer.

Li, C.; Wang, S.; Yu, P. S.; Zheng, L.; Zhang, X.; Li, Z.; and Liang, Y. 2018. Distribution distance minimization for unsupervised user identity linkage. In *CIKM*, 447–456. ACM.

Liu, J.; Zhang, F.; Song, X.; Song, Y.-I.; Lin, C.-Y.; and Hon, H.-W. 2013. What's in a name?: an unsupervised approach to link users across communities. In *WSDM*, 495–504. ACM.

Liu, L.; Cheung, W. K.; Li, X.; and Liao, L. 2016. Aligning users across social networks using network embedding. In *IJCAI*, 1774–1780.

Man, T.; Shen, H.; Liu, S.; Jin, X.; and Cheng, X. 2016. Predict anchor links across social networks via an embedding approach. In *IJCAI*, 1823–1829.

McKay, B. D., et al. 1981. Practical graph isomorphism.

Motoyama, M., and Varghese, G. 2009. I seek you: searching and matching individuals in social networks. In *WSDM*, 67–75. ACM.

Mu, X.; Zhu, F.; Lim, E.-P.; Xiao, J.; Wang, J.; and Zhou, Z.-H. 2016. User identity linkage by latent user space modelling. In *KDD*, 1775–1784. ACM.

Nie, Y.; Jia, Y.; Li, S.; Zhu, X.; Li, A.; and Zhou, B. 2016. Identifying users across social networks based on dynamic core interests. *Neurocomputing* 107–115.

Peled, O.; Fire, M.; Rokach, L.; and Elovici, Y. 2013. Entity matching in online social networks. In *SocialCom*, 339–344. IEEE.

Perito, D.; Castelluccia, C.; Kaafar, M. A.; and Manils, P. 2011. How unique and traceable are usernames? In *International Symposium on Privacy Enhancing Technologies Symposium*, 1–17. Springer.

Riederer, C.; Kim, Y.; Chaintreau, A.; Korula, N.; and Lattanzi, S. 2016. Linking users across domains with location data: Theory and validation. In *WWW*, 707–719.

Shu, K.; Wang, S.; Tang, J.; Zafarani, R.; and Liu, H. 2017. User identity linkage across online social networks: A review. *ACM SIGKDD Explorations Newsletter* 5–17.

Smith, S. L.; Turban, D. H.; Hamblin, S.; and Hammerla, N. Y. 2017. Offline bilingual word vectors, orthogonal transformations and the inverted softmax. *arXiv*.

Tan, S.; Guan, Z.; Cai, D.; Qin, X.; Bu, J.; and Chen, C. 2014. Mapping users across networks by manifold alignment on hypergraph. In *AAAI*, 159–165.

Villani, C. 2008. *Optimal transport: old and new*, volume 338. Springer Science & Business Media.

Vosecky, J.; Hong, D.; and Shen, V. Y. 2009. User identification across multiple social networks. In *Networked Digital Technologies*, 360–365. IEEE.

Wang, S.; Hu, X.; Yu, P. S.; and Li, Z. 2014. Mmrate: inferring multi-aspect diffusion networks with multi-pattern cascades. In *KDD*, 1246–1255. ACM.

Wang, S.; Yan, Z.; Hu, X.; Philip, S. Y.; and Li, Z. 2015. Burst time prediction in cascades. In *AAAI*, 325–331.

Yang, C.; Liu, Z.; Zhao, D.; Sun, M.; and Chang, E. Y. 2015. Network representation learning with rich text information. In *IJCAI*, 2111–2117.

Yang, D.; Wang, S.; Li, C.; Zhang, X.; and Li, Z. 2017. From properties to links: deep network embedding on incomplete graphs. In *CIKM*, 367–376. ACM.

Zafarani, R., and Liu, H. 2014. Users joining multiple sites: distributions and patterns. In *ICWSM*.

Zafarani, R., and Liu, H. 2016. Users joining multiple sites: Friendship and popularity variations across sites. *Information Fusion* 83–89.

Zafarani, R.; Tang, L.; and Liu, H. 2015. User identification across social media. *TKDD* 16.

Zhan, Q.; Zhang, J.; Wang, S.; Philip, S. Y.; and Xie, J. 2015. Influence maximization across partially aligned heterogenous social networks. In *PAKDD*, 58–69. Springer.

Zhang, H.; Kan, M.-Y.; Liu, Y.; and Ma, S. 2014. Online social network profile linkage. In *Asia Information Retrieval Symposium*, 197–208. Springer.

Zhang, Y.; Tang, J.; Yang, Z.; Pei, J.; and Yu, P. S. 2015. Cosnet: Connecting heterogeneous social networks with local and global consistency. In *KDD*, 1485–1494. ACM.

Zhang, M.; Liu, Y.; Luan, H.; and Sun, M. 2017a. Adversarial training for unsupervised bilingual lexicon induction. In *ACL*, volume 1, 1959–1970.

Zhang, M.; Liu, Y.; Luan, H.; and Sun, M. 2017b. Earth mover's distance minimization for unsupervised bilingual lexicon induction. In *EMNLP*, 1934–1945.

Zhang, B.; Saha, T. K.; and Al Hasan, M. Name disambiguation from link data in a collaboration graph. In *ASONAM*, 81–84.

Zhong, Z.; Cao, Y.; Guo, M.; and Nie, Z. 2018. Colink: An unsupervised framework for user identity linkage. In *AAAI*, 3379–3385.