

# Adversarial-Enhanced Hybrid Graph Network for User Identity Linkage

Xiaolin Chen<sup>†</sup>, Xuemeng Song<sup>†\*</sup>, Guozhen Peng<sup>†</sup>, Shanshan Feng<sup>§</sup>, Liqiang Nie<sup>†\*</sup>  
<sup>†</sup>Shandong University, Shandong, China, <sup>§</sup>Harbin Institute of Technology, Shenzhen, China  
 {cxlicd, sxmustc}@gmail.com, {guozhen.sdu, victor\_fengss}@foxmail.com, nieliqiang@gmail.com

## ABSTRACT

In this work, we investigate the user identity linkage task across different social media platforms based on heterogeneous multi-modal posts and social connections. This task is non-trivial due to the following two challenges. 1) As each user involves both intra multi-modal posts and inter social connections, how to accurately fulfil the user representation learning from both intra and inter perspectives constitutes the main challenge. And 2) even representations distributed on different platforms of the same identity tend to be distinct (i.e., the semantic gap problem) owing to discrepant data distribution of different platforms. Hence, how to alleviate the semantic gap problem poses another tough challenge. To this end, we propose a novel adversarial-enhanced hybrid graph network (AHG-Net), consisting of three key components: *user representation extraction*, *hybrid user representation learning*, and *adversarial learning*. Specifically, AHG-Net first employs advanced deep learning techniques to extract the user's intermediate representations from his/her heterogeneous multi-modal posts and social connections. Then AHG-Net unifies the intra-user representation learning and inter-user representation learning with a hybrid graph network. Finally, AHG-Net adopts adversarial learning to encourage the learned user presentations of the same identity to be similar using a semantic discriminator. Towards evaluation, we create a multi-modal user identity linkage dataset by augmenting an existing dataset with 62,021 images collected from Twitter and Foursquare. Extensive experiments validate the superiority of the proposed network. Meanwhile, we release the dataset, codes, and parameters to facilitate the research community.

## CCS CONCEPTS

• **Information systems** → **Retrieval tasks and goals.**

## KEYWORDS

User Identity Linkage; Graph Neural Network; Adversarial Learning.

\*Corresponding authors: Xuemeng Song and Liqiang Nie.

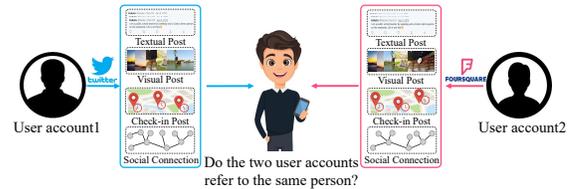
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components 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](mailto:permissions@acm.org).

SIGIR '21, July 11–15, 2021, Virtual Event, Canada

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-8037-9/21/07...\$15.00

<https://doi.org/10.1145/3404835.3462946>



**Figure 1: Illustration of the user identity linkage task based on the user's multi-modal posts and social connections. We take platforms of Twitter and Foursquare as an example.**

## ACM Reference Format:

Xiaolin Chen, Xuemeng Song, Guozhen Peng, Shanshan Feng, Liqiang Nie. 2021. Adversarial-Enhanced Hybrid Graph Network for User Identity Linkage. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21)*, July 11–15, 2021, Virtual Event, Canada. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3404835.3462946>

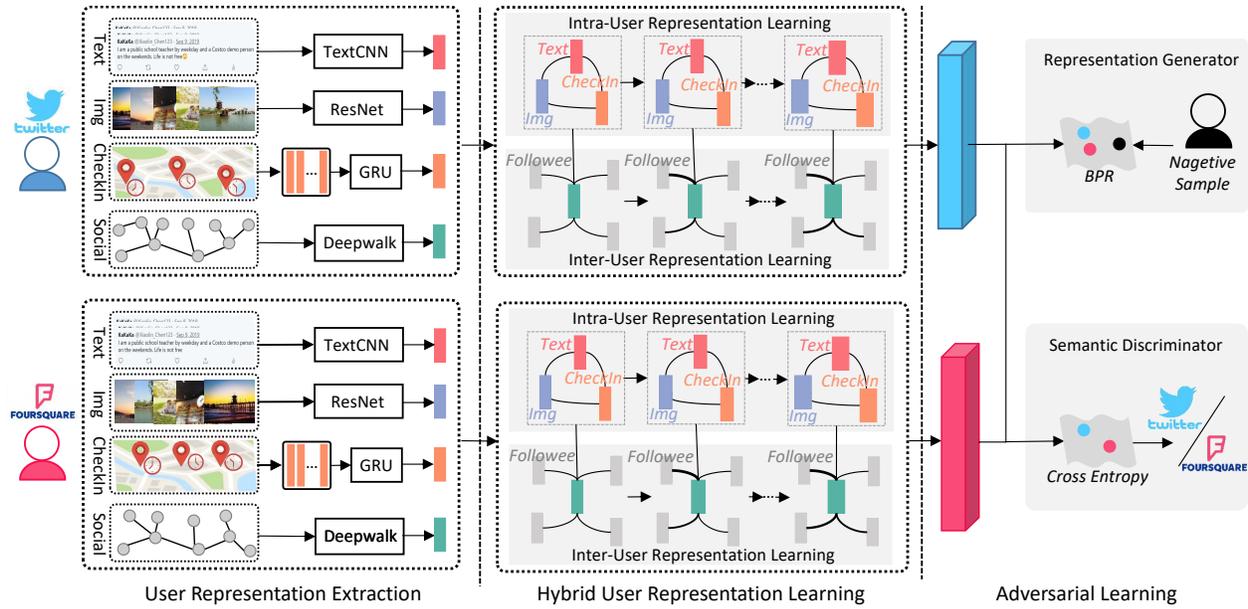
## 1 INTRODUCTION

In recent years, owing to diverse service spotlights of different social platforms, ranging from self-promoting to photo sharing, people tend to embrace multiple social platforms (e.g., Twitter, Foursquare, and Instagram) concurrently. According to the report of Pew Research Center<sup>1</sup>, roughly 73% of Internet users utilize more than one social media platform simultaneously, where 90% of Twitter users use Facebook, and 47% Facebook users engage in Instagram. Therefore, a surge of researches are dedicated to exploiting tasks across social media platforms, such as the cross platform recommendation [41] and the information diffusing prediction [44]. As a vital prerequisite of the above studies, the user identity linkage task, which aims to link individual's accounts on different social media platforms, is increasingly indispensable.

In fact, various methods have been presented for the user identity linkage task [18, 24, 25, 43], where user's textual posts, check-in posts, social connections, and the combinations of these contents, have been well explored. Despite the prominent success achieved by these methods, they generally overlook the fact that the visual modality also merits our attention in the context of user identity linkage, since users tend to share similar visual posts on different social platforms. In light of this, as shown in Figure 1, we aim to solve the user identity linkage across different social platforms by simultaneously exploring the user's multi-modal posts, including textual, visual, check-in posts, as well as social connections.

However, comprehensively fulfilling the user identity linkage task with both heterogeneous multi-modal posts and unstructured social connections of the user is non-trivial due to the following

<sup>1</sup><https://tinyurl.com/t9uqq3j>.



**Figure 2: Illustration of the proposed network for user identity linkage. AHG-Net contains three key components: user representation extraction, hybrid user representation learning and adversarial learning. “Text”, “Img”, “CheckIn” and “Social” refer to the textual post, visual post, check-in post, and social connection, respectively.**

two challenges. 1) In a sense, each user can be characterized from both his/her intra multi-modal posts and inter social connections. Regarding the user’s intra multi-modal post fusion, although different modalities, i.e., textual, visual, and check-in, may reflect different views of the user, they share certain intrinsic semantic relations. Intuitively, users tend to share delicious food images if they frequently check in restaurant landmarks, and express positive feelings with corresponding images. Meanwhile, pertaining to the inter-user correlation modeling, different followees may have distinct influences concerning the user characterization due to their different levels of intimate degree. Therefore, how to properly fuse the user representation learning from the above two perspectives constitutes the main challenge for us. 2) In fact, even for the same user identity, his/her representations on diverse social media platforms tend to be dissimilar (i.e., the semantic gap problem), owing to the discrepant data distribution of social media platforms. Therefore, existing studies that tackle the user identity linkage task by directly seeking a latent user representation space, where the user representations on different social platforms belonging to the same user identity should be close, can yield suboptimal performance. Hence, how to effectively alleviate the aforementioned semantic gap problem poses another tough challenge.

To address the aforementioned challenges, we propose a novel Adversarial-enhanced Hybrid Graph Network for user identity linkage, AHG-Net for short. As illustrated in Figure 2, the proposed AHG-Net contains three key components: *user representation extraction*, *hybrid user representation learning*, and *adversarial learning*. In particular, we first extract the user representation from the heterogeneous multi-modal posts as well as social connections via advanced deep learning techniques. Thereafter, we employ a

hybrid graph neural network to unify the intra-user representation learning and inter-user representation learning, where the semantic correlation among modalities and the adaptive influences among users are jointly modeled. In addition, we exploit the adversarial learning and deploy the semantic discriminator to distinguish the platform source of each representation, which promotes the learned representations from different platforms to be undistinguishable and thus facilitates the user identity linkage. Moreover, towards our model evaluation, we further build a new multi-modal dataset by augmenting the existing public dataset [31] with 62,021 visual posts from Twitter and Foursquare. Extensive experiments on this dataset have fully validated the effectiveness of our proposed AHG-Net.

Our main contributions can be summarized as follows:

- To the best of our knowledge, we are among the first to tackle the user identity linkage task by taking into account both heterogeneous multi-modal posts (including the visual cue) as well as social connections of users.
- We propose an adversarial-enhanced hybrid graph network for user identity linkage, which unifies the intra-user representation learning and inter-user representation learning in a hybrid graph network, and alleviates the semantic gap problem caused by distinct data distribution of social media platforms with the adversarial learning.
- We create a new multi-modal dataset based on an existing dataset, where 62,021 visual posts are additionally collected from Twitter and Foursquare. Extensive experiments on the real-world dataset demonstrate the superiority of our proposed model. As a byproduct, we have released the dataset, codes, and involved parameters to facilitate the research community<sup>2</sup>.

<sup>2</sup><https://anonymous819.wixsite.com/ahg-net>.

## 2 RELATED WORK

### 2.1 User Identity Linkage

According to the review [33], existing efforts on user identity linkage mainly aim to tackle the problem by exploring diverse types of user information (e.g., profile, social connection, and content) in social media platforms. In a sense, existing studies can be roughly divided into the following two categories.

The first category focuses on exploiting *a single type of user information* for user identity linkage. The most straightforward solution is to utilize the user’s profile information [25, 43], such as usernames, genders and birthdays. For example, Mu et al. [25] implemented user identity linkage by utilizing linear regression to project user accounts of different platforms into the latent user space based on their’s profiles. Nevertheless, for the privacy concern, a user profile may be deliberately counterfeited and contains some inconsistent values, which makes the basic user profile fragile and unreliable. Beyond this, several studies have resorted to the user’s social connections [18, 24], which are more reliable than the user profile. For example, Man et al. [24] tackled the user identity linkage task by network embedding, where observed anchor links are used as supervision to capture the structural regularity. Moreover, different from the above studies, several efforts have been made to explore the user content [2, 10, 45]. For instance, Rong et al. [45] explored users’ writing style features to solve the authorship identification problem. Although these efforts have achieved compelling success, they only consider one specific type of user information, which cannot comprehensively characterize the user.

Beyond methods of the first category, the second category aims to collectively explore *multiple types of user information* to boost the performance of user identity linkage [12, 17, 20, 22, 31, 38]. For example, Liu et al. [22] designed a heterogeneous behavior model to measure the user behavior similarity based on all above types of user information (i.e., profile, social connection and content). Similarly, due to the concern on the ambiguous and unreliable user profiles, recent studies of this category mainly focus on the user’s social connection and content information. To be more specific, Ren [31] introduced a network alignment model, which predefines a set of meta diagrams to extract user features from his/her social connections and content information. Although these studies obtain remarkable performance, they treat the different modality content independently and learn the user representation from different cues separately. Beyond that, in this work, we aim to explore the underlying relations residing in users’ heterogeneous multi-modal posts as well as social connections, and alleviate the semantic gap problem among different social media platforms.

### 2.2 Generative Adversarial Networks

In recent years, GANs has attracted increasing research attention and achieved prominent success in diverse tasks, ranging from the sequential recommendation [30] to cross-modality search [36, 42]. For example, Yang et al. [42] utilized GANs to enhance the visual understanding in fashion search by directly synthesizing the target item image. In addition, Ren et al. [30] presented a multi-factor generative adversarial network to explicitly model the effect of context information for the sequence recommendation task. In the

domain of user identity linkage, Li et al. [19] proposed an adversarial learning based framework to solve the semi-supervised user identity linkage task under the multi-platform setting. To be specific, they employed an auto-encoder to map feature vectors in one platform into another social platform. Then, for each platform, they set a discriminator to distinguish whether the given feature vector is original or mapped. Different from existing work, we target at bridging the aforementioned semantic gap between different social media platforms by deploying a semantic discriminator to distinguish the platform source of each user representation, in order to facilitate the conduction of user identity linkage.

## 3 MODEL

### 3.1 Problem Formulation

In this work, we aim to investigate the task of user identity linkage across different social media platforms by answering the question “*whether the given pair of user accounts on different social media platforms refer to the same user identity based on their heterogeneous multi-modal posts and social connections*”. Without loss of generality, we particularly focus on linking user identities between two platforms (i.e.,  $O_1$  and  $O_2$ ), while the cases of multiple social media platforms can be easily extended. Suppose we have a set of user accounts  $\mathcal{U}_1 = \{u_1^1, u_1^2, \dots, u_1^{N_1}\}$  on the platform  $O_1$  and a set of user accounts  $\mathcal{U}_2 = \{u_2^1, u_2^2, \dots, u_2^{N_2}\}$  on  $O_2$ , where  $N_1$  and  $N_2$  denote the number of user accounts on platforms  $O_1$  and  $O_2$ . Each user account  $u_1^i$  on  $O_1$  has a set of social connections  $\mathcal{S}_1^i$  and a set of social posts  $\mathcal{X}_1^i$ , which can be further grouped into three subsets according to their modalities: the set of textual posts  $\mathcal{C}_1^i$ , the set of visual posts  $\mathcal{V}_1^i$ , and the set of check-in posts  $\mathcal{T}_1^i$ , respectively. Notably, the check-in posts reveal the user’s historical trajectories, which are derived from the user’s social posts that contain location tags. Analogously, for each user account  $u_2^j$  on  $O_2$ , we also have his/her social connection set  $\mathcal{S}_2^j$ , textual post set  $\mathcal{C}_2^j$ , visual post set  $\mathcal{V}_2^j$ , and check-in post set  $\mathcal{T}_2^j$ . Let  $y_i^j = 1$ , if  $u_1^i$  and  $u_2^j$  refer to the same user identity in the real world, and  $y_i^j = 0$  otherwise.

In a sense, we aim to devise a novel model  $\mathcal{F}$  which can accurately predict whether the given two user accounts on different social platforms refer to the same user identity as follows,

$$\mathcal{F}(u_1^i, u_2^j | \Theta_{\mathcal{F}}) \rightarrow y_i^j, \quad (1)$$

where  $\Theta_{\mathcal{F}}$  represents the model parameters.

### 3.2 User Representation Extraction

First, we introduce how to extract the user representation from the heterogeneous multi-modal posts and the unstructured social connections. For the ease of illustration, we temporally omit both the superscript and the subscript, since the user representation extraction for any user on any social media platform can be derived in the same manner.

**Textual Representation.** To obtain the underlying semantic information delivered by the user’s textual posts and characterize the user from the textual modality, we leverage the textual convolutional neural network (TextCNN) [16], which has achieved compelling success in the representation learning task [3, 37]. In particular, given the set of textual posts  $C = \{c_1, c_2, \dots, c_n\}$

consisting of  $n$  posts, we first embed each post  $c_p, p = \{1, 2, \dots, n\}$  into the vector  $\mathbf{e}_p \in \mathbb{R}^{D_e}$  with the help of the pre-trained Bidirectional Encoder Representations from Transformer (BERT) [6, 14].  $D_e$  is the dimension of the textual post embedding generated by BERT. Thereafter, we stack the textual post embeddings chronologically, namely, according to their post time, and obtain the textual post embedding matrix  $\mathbf{C} \in \mathbb{R}^{D_e \times n}$ . Then we employ a one-layer CNN with  $K$  filters  $\mathcal{W}_f = \{\mathbf{W}_{f_1}, \mathbf{W}_{f_2}, \dots, \mathbf{W}_{f_K}\}$ , where  $\mathbf{W}_{f_k} \in \mathbb{R}^{h \times D_e}$  is the  $k$ -th filter working on summarizing a window of  $h$  posts into a new feature to learn the user’s textual representation. One advantage of TextCNN is that it incorporates the local post relevance, which benefits the textual context capturing. Ultimately, we project the intermediate representation into the latent user representation space to obtain the final textual representation with the average pooling operation and a fully connected layer as follows,

$$\begin{cases} \tilde{\mathbf{c}} = \text{avg}[\rho(\mathbf{W}_{f_1}, \mathbf{C}), \rho(\mathbf{W}_{f_2}, \mathbf{C}), \dots, \rho(\mathbf{W}_{f_K}, \mathbf{C})], \\ \mathbf{c} = \xi(\mathbf{W}_c \tilde{\mathbf{c}} + \mathbf{b}_c), \end{cases} \quad (2)$$

where  $\text{avg}[\cdot]$  represents the average pooling operation,  $\rho(\cdot)$  refers to the convolutional operation and  $\xi(\cdot)$  denotes the LeakyRelu activation function.  $\mathbf{W}_c$  and  $\mathbf{b}_c$  are the weight matrix and the bias vector, respectively.

**Visual Representation.** Regarding the visual posts of a user, we encode them with the pre-trained Residual Network (ResNet) [13], which has shown the superior performance in the computer vision task [39]. Given the set of visual posts  $\mathcal{V} = \{v_1, v_2, \dots, v_m\}$ , we first employ ResNet to encode each visual post  $v_p$  and then utilize the average pooling and a fully connected layer to get the user’s visual representation  $\mathbf{v} \in \mathbb{R}^D$  as follows,

$$\begin{cases} \tilde{\mathbf{v}} = \text{avg}[\text{Res}(v_1|\Theta_R), \text{Res}(v_2|\Theta_R), \dots, \text{Res}(v_m|\Theta_R)], \\ \mathbf{v} = \xi(\mathbf{W}_r \tilde{\mathbf{v}} + \mathbf{b}_r), \end{cases} \quad (3)$$

where  $\mathbf{W}_r$  and  $\mathbf{b}_r$  are the weight matrix and bias vector of the fully connected layer, respectively.  $\Theta_R$  refers to the network parameters of ResNet and  $\xi(\cdot)$  denotes the LeakyRelu activation function.

**Check-In Representation.** Intuitively, check-in posts may reflect the user’s life circles to some extent, which is of crucial importance for the user identity linkage. To yield the check-in representation of the user, one naive way is to introduce a location vocabulary, and represent each user with a one-hot representation, where 1 indicates that the user once posted at the corresponding location. However, this method neglects the temporal factor in the check-in pattern. For example, given a user account  $u_1$  on the platform  $O_1$ , who frequently publishes posts at Beijing in June 2020 and New York in August 2020, and another user account  $u_2$  on the platform  $O_2$  who always posts at New York in June 2020 and Beijing in August 2020, they tend to be classified as the same user identity based on the simple spatial distribution. However, on the basis of their check-in patterns, user accounts  $u_1$  and  $u_2$  are less possible to be the same user identity as they have not been at the same place for the same period [31]. Accordingly, we propose to utilize the temporal-spatial distribution to comprehensively characterize the user from the check-in perspective.

Let  $\mathcal{T} = \{(t_g, q_g)\}_{g=1}^k$  denote the set of check-in posts.  $t_g$  and  $q_g$  represent the time slot and the location of  $g$ -th check-in post.  $k$  is

the total number of check-in posts of the user. We first construct a vocabulary of  $M$  time slots and a vocabulary of  $K$  locations that can be shared by both social media platforms, i.e.,  $O_1$  and  $O_2$ . Then based on a user’s check-in posts, we can derive a spatial-temporal co-occurrence matrix  $\mathbf{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_M\}^T \in \mathbb{R}^{M \times K}$ , where  $\mathbf{a}_m = (a_m^1, a_m^2, \dots, a_m^K)$  denotes the one-hot spatial distribution vector for the  $m$ -th time slot. To be specific,  $a_m^k = 1$  if the user appears at the  $k$ -th location in the  $m$ -th time slot, and  $a_m^k = 0$  otherwise. Since the spatial distribution has clear sequential relationships, we adopt the gated recurrent units (GRU) [5] to derive the final check-in representation, considering its superior performance in various sequence modeling tasks [10, 15]. The final check-in representation  $\mathbf{t}$  can be obtained as follows,

$$\begin{cases} \tilde{\mathbf{t}} = \text{GRU}(\mathbf{a}_m | \Theta_G), m \in \{1, 2, \dots, M\}, \\ \mathbf{t} = \xi(\mathbf{W}_t \tilde{\mathbf{t}} + \mathbf{b}_t), \end{cases} \quad (4)$$

where  $\mathbf{W}_t$  and  $\mathbf{b}_t$  are the weight matrix and bias vector of a fully connected layer, respectively.  $\xi(\cdot)$  is the LeakyRelu activation function and  $\Theta_G$  refers to the network parameters of GRU.

**Social Representation.** Undoubtedly, one essential function of social platforms is to connect users. Typically, users on social media platforms can follow each other due to various reasons, ranging from personal interest to making friends in real life. Therefore, one’s followee connections to some extent reflect the topics that the user is interested in or the user’s friend relationships. Hence, in this work, we focus on the followee connections to characterize each user. In particular, we employ DeepWalk [26] to learn the intermediate representation of each user on the social platform. Thereafter, a fully-connected layer with the LeakyRelu activation function is used to transform the intermediate user representation to the latent user representation space. Formally, let  $\mathbf{s}$  be the latent social representation of a given user.

### 3.3 Hybrid Graph-based User Representation Learning

Having obtained heterogeneous multi-modal post representations and social connection representation of each user, we should proceed to how to seamlessly integrate these representations towards the accurate user representation learning. In fact, each user’s representation can be accounted by both his/her social connections’ representations and his/her own multi-modal post representations. Meanwhile, although the multi-modal post representations convey different cues with regard to the user characterization, essentially they reflect the same user and hence share certain semantic relations. Intuitively, users that like to post at restaurants are more likely to share food photos, while those who prefer to post scene pictures tend to check in travel spots. Therefore, in order to model the user-user interaction and modal-modal semantic relation, we resort to graph neural networks, which have shown to be effective in relation reasoning [3, 8, 9, 40]. Intuitively, we propose a hybrid user representation learning scheme consisting of two key components: *intra-user representation learning* and *inter-user representation learning*.

**3.3.1 Intra-User Representation Learning.** In order to model the underlying semantic relation among different modality cues (i.e., textual, visual, and check-in), we first devise an undirected graph

$\mathcal{G} = (\mathcal{M}, \mathcal{E})$ , where the set of nodes  $\mathcal{M} = \{\mathbf{m}_i\}_{i=1}^Q$  correspond to the  $Q$  modality post representations, i.e.,  $\mathbf{c}$ ,  $\mathbf{v}$ , and  $\mathbf{t}$ , respectively, while the set of edges  $\mathcal{E} = \{(m_i, m_j) | i, j \in [1, \dots, Q]\}$  indicate the semantic relations between different modalities.

According to GCN [3], we can update one modality representation based on its correlated modality representation. Towards this, we define the semantic similarity-based adjacent matrix  $\mathbf{A} \in \mathbb{R}^{Q \times Q}$ , whose  $(i, j)$ -th entry can be derived as follows,

$$\mathbf{A}_{i,j} = \begin{cases} \cos(\mathbf{m}_i, \mathbf{m}_j), & \text{if } i \neq j, \\ 1, & \text{otherwise,} \end{cases} \quad (5)$$

where  $\cos(\mathbf{m}_i, \mathbf{m}_j)$  is the cosine similarity between the  $i$ -th modality representation and  $j$ -th one. Then given the semantic adjacent matrix  $\mathbf{A}$ , each GCN layer can be formulated as follows,

$$\mathbf{H}^{(l+1)} = g(\mathbf{A}\mathbf{H}^{(l)}\mathbf{W}^{(l)}), l \in \{0, 1, \dots, L-1\}, \quad (6)$$

where  $\mathbf{H}^{(l)} = \{\mathbf{h}_1^{(l)}, \mathbf{h}_2^{(l)}, \dots, \mathbf{h}_Q^{(l)}\}^T \in \mathbb{R}^{Q \times d_l}$ , and  $\mathbf{h}_n^{(l)}$  is the latent embedding of the  $n$ -th modality at the  $l$ -th layer. In particular,  $\mathbf{H}^{(0)} = [\mathbf{c}, \mathbf{v}, \mathbf{t}]$  is the initial embedding matrix.  $g(\cdot)$  is a non-linear operation, where we employ the LeakyRelu.  $L$  indicates the total number of GCN layers, and  $\mathbf{W}^{(l)} \in \mathbb{R}^{d_l \times d_{(l+1)}}$  is the to-be-learned transformation matrix for the  $l$ -th layer.  $d_l$  and  $d_{(l+1)}$  represent the embedding dimensions of the  $l$ -th and  $(l+1)$ -th layers, respectively. Ultimately, we treat the output of the  $L$ -th layer as the final multi-modal representation of the user, i.e.,  $\mathbf{H}^{(L)} = \{\mathbf{h}_1^{(L)}, \mathbf{h}_2^{(L)}, \dots, \mathbf{h}_Q^{(L)}\}^T$ .

**3.3.2 Inter-User Representation Learning.** Having obtained the user representations of different modalities with the intra-user representation learning, we can move forward to model the user-user relations. As aforementioned, each user's representation can be reflected by the social connections, i.e., followees. One naive way is to equally fuse all the user representations of one's social connections. However, this method overlooks that different connections tend to have distinguished influences on the user characterization due to different levels of intimate degree, thereby easily resulting in suboptimal representations. Towards this end, to distinguish the informative connections in representing the user, we propose to leverage graph attention network (GAT) [35], which has achieved conspicuous performance [4, 23].

In particular, given a user  $u$ , let  $\mathcal{S} = \{\mathbf{s}_{c_1}, \mathbf{s}_{c_2}, \dots, \mathbf{s}_{c_S}\}$  denote the set of representations of his/her social connections, all of which are derived by the aforementioned social representation learning.  $c_S$  is the total number of social connections of the user  $u$ . Then to retain the user's own features during the representation learning, we extend  $\mathcal{S}$  with the user's own representation, i.e.,  $\bar{\mathcal{S}} = \{\mathbf{s}_{c_0}, \mathbf{s}_{c_1}, \mathbf{s}_{c_2}, \dots, \mathbf{s}_{c_S}\}$ , where  $\mathbf{s}_{c_0} = \mathbf{s} + \sum_{n=1}^Q \mathbf{h}_n^{(L)}$  refers to the user  $u$ 's own representation. Thereafter, we can assign the confidences for different social connections as follows,

$$\alpha_{c_g} = \frac{\exp(\xi(\mathbf{a}_1^T(\mathbf{W}_1 \mathbf{s}_{c_0} \oplus \mathbf{W}_1 \mathbf{s}_{c_g})))}{\sum_{g=0}^S \exp(\xi(\mathbf{a}_1^T(\mathbf{W}_1 \mathbf{s}_{c_0} \oplus \mathbf{W}_1 \mathbf{s}_{c_g})))}, g \in \{0, 1, 2, \dots, S\}, \quad (7)$$

where  $\alpha_{c_g}$  denotes the confidence for the  $g$ -th social connection,  $\oplus$  represents the concentration operation [28], and  $\mathbf{W}_1$  is the shared global weight matrix of users on  $O_1$ . In a sense, the vector  $\mathbf{a}_1$  can be treated as the latent representation of the query "which

social connection of the given user conveys more vital cues to the user characterization". Accordingly, we can reach the final user representation  $\mathbf{u}$  as follows,

$$\mathbf{u} = \sum_{g=0}^S \alpha_{c_g} \mathbf{W}_1 \mathbf{s}_{c_g}. \quad (8)$$

Moreover, to boost the performance towards the user representation learning, we employ the multi-head graph attention mechanism [34], where we incorporate  $R$  independent single attention modules concurrently to stabilize the learning process of graph attention mechanism, and concentrate their output as the final user representation. Accordingly,  $\mathbf{u}$  can be re-written as,

$$\mathbf{u} = \sum_{g=0}^S \alpha_{c_g}^1 \mathbf{W}_1^1 \mathbf{s}_{c_g} \oplus \dots \oplus \sum_{g=0}^S \alpha_{c_g}^r \mathbf{W}_1^r \mathbf{s}_{c_g} \oplus \dots \oplus \sum_{g=0}^S \alpha_{c_g}^R \mathbf{W}_1^R \mathbf{s}_{c_g}, \quad (9)$$

where  $\alpha_{c_g}^r$  is the confidence derived from the  $r$ -th attention module, and  $\mathbf{W}_1^r$  is the corresponding weight matrix.

### 3.4 Adversarial Learning

As a matter of fact, even the representations distributed on different social media of the same user identity tend to suffer from the semantic gap problem due to their distinct data distribution [10], which may impede the conduction of the user identity linkage task. Impelled by the massive success achieved by adversarial learning in diverse representation learning tasks [7, 30, 42], we explore adversarial learning to address the semantic gap problem and enhance the user representation learning in the context of the user identity linkage task.

**3.4.1 Representation Generator.** On one hand, we treat the above hybrid graph network as a user representation generator. Essentially, the generator works on learning the latent user representation with the assumption that user accounts on different social media platforms that refer to the same user identity tend to be more similar than those of different user identities. In particular, we measure the similarity score between user accounts on two platforms as follows,

$$q_i^j = (\mathbf{u}_1^i)^T \mathbf{u}_2^j, \quad (10)$$

where  $\mathbf{u}_1^i$  and  $\mathbf{u}_2^j$  refer to the user representation of the user account  $u_1^i$  on  $O_1$  and  $u_2^j$  on  $O_2$ , both of which can be derived according to Eqn. (9). Then following the Bayesian Personalized Ranking (BPR) [32] framework, which has demonstrated superiority in the classification task [1], we have the objective function for the generator as follows,

$$\mathcal{L}_{G_{bpr}} = \sum_{i=1}^{N^+} \mathcal{L}_{bpr}(q_i^+, q_i^-) = \sum_{i=1}^{N^+} -\ln(\sigma(q_i^+ - q_i^-)), \quad (11)$$

where  $q_i^+$  refers to the similarity score between the positive user pair  $(u_1^i, u_2^{i+})$ , while  $q_i^-$  is between the negative user pair  $(u_1^i, u_2^{i-})$ .  $N^+$  is the total number of the positive user pairs, and  $\sigma(\cdot)$  is the sigmoid activation function.

**Table 1: Statistics of TWFQ.**

| Category           | Twitter   | Foursquare | Total Number |
|--------------------|-----------|------------|--------------|
| #User              | 3,463     | 3,833      | 7,296        |
| #Textual Post      | 2,197,830 | 18,442     | 2,216,272    |
| #Visual Post       | 33,454    | 28,567     | 62,021       |
| #Check-in Post     | 198,105   | 18,417     | 216,522      |
| #Social Connection | 53,137    | 24,780     | 77,917       |

**3.4.2 Semantic Discriminator.** On the other hand, we introduce a platform semantic discriminator  $D_p$  that aims to distinguish the user representations derived from different platforms of the same identity (e.g.,  $u_1^i$  or  $u_2^{i+}$ ). This can be cast as a binary classification task, i.e., determining the platform source, i.e.,  $O_1$  or  $O_2$ , of each user representation. Specifically, taking the positive pair  $(u_1^i, u_2^{i+})$  as an example, we feed the corresponding representation  $(u_1^i, u_2^{i+})$  into the Multiple Layer Perceptron (MLP) and then adopt the cross-entropy [21] loss as follows,

$$\mathcal{L}_{D_s} = -\frac{1}{N^+} \sum_{i=1}^{N^+} \mathbf{m}_i (\log D_s(u_1^i | \Theta_{D_s}) + \log(1 - D_s(u_2^{i+} | \Theta_{D_s}))), \quad (12)$$

where  $\mathbf{m}_i$  is the platform label of each user representation, defined as one-hot vector.  $D_s(\cdot | \Theta_{D_s})$  is the predicted platform probability vector of the given user representation.  $\Theta_{D_s}$  denotes the parameters of the semantic discriminator.

**3.4.3 Joint Optimization.** Finally, we optimize the generator and discriminator as a minimax game [11] since the optimization goals of the two components are opposite. To be specific, we parameterize the final objective function for user identity linkage as follows,

$$\begin{cases} \Phi^* = \operatorname{argmax}_D(\mathcal{L}_{D_s}), \\ \Theta^* = \operatorname{argmin}_G(\mathcal{L}_{G_{bpr}}), \end{cases} \quad (13)$$

where  $\Phi^*$  refers to the generator parameters, and  $\Theta^*$  indicates the parameters of the semantic discriminator.

## 4 EXPERIMENT

In this section, we first introduce the dataset as well as experiment setting, and then detail the experiments by answering the following research questions:

- **RQ1:** Does our AHG-Net surpass state-of-the-art methods?
- **RQ2:** How do the hybrid user representation learning and adversarial learning affect the AHG-Net?
- **RQ3:** How do different modalities influence the AHG-Net?
- **RQ4:** How does AHG-Net perform with missing data?

### 4.1 Dataset

Since this is the first research attempting to explore the user’s multi-modal social posts and social connection simultaneously, there is no publicly available dataset that exactly supports our goal. In light of this, we created our own dataset, named TWFQ, on the basis of one existing dataset introduced by Ren et al. [31], which consists of 3,282 positive user pairs with their corresponding social posts and social connections from Twitter and Foursquare. Specifically, we first only retained users with complete modalities (i.e., textual, check-in, and social) and obtained 1,071 positive user pairs. Thereafter, we crawled these users’ visual

**Table 2: Performance comparison among different methods.**

| Models         | Accuracy      | Precision     | Recall        | F1-score      |
|----------------|---------------|---------------|---------------|---------------|
| DPLink         | 58.86%        | 58.68%        | 77.97%        | 66.96%        |
| MV_URL         | 62.62%        | 67.31%        | 60.34%        | 63.64%        |
| DLHD           | 63.55%        | 61.45%        | 87.93%        | 72.34%        |
| ACTIVER_V      | 75.23%        | 75.47%        | 74.77%        | 75.12%        |
| MSUIL_V        | 77.57%        | 77.42%        | 82.76%        | 80.00%        |
| MNA_V          | 78.97%        | 78.18%        | 80.37%        | 79.26%        |
| ACTIVER        | 89.72%        | 91.23%        | 89.66%        | 90.43%        |
| <b>AHG-Net</b> | <b>91.59%</b> | <b>92.98%</b> | <b>91.38%</b> | <b>92.17%</b> |

posts (i.e., images) according to the user IDs provided by the original dataset. In particular, we employed Twitter API<sup>3</sup> and Foursquare API<sup>4</sup> to collect user’s most recent 50 images from Twitter and Foursquare, respectively. Ultimately, we obtained 33,454 visual posts from Twitter, while 28,567 visual posts from Foursquare. Table 1 shows the statistics of our dataset.

### 4.2 Experiment Setting

Firstly, we divided the positive user account pairs into three chunks: 80% for training, 10% for validation, and 10% for testing. Then for each positive pair  $(u_1^i, u_2^{i+})$ , we randomly sampled a negative user  $u_2^{i-}$  on  $O_2$  to constitute a training triplet. We selected the following common evaluation metrics: Accuracy, Precision, Recall, and F1-score, to evaluate our proposed scheme. For optimization, we utilized the Adaptive Moment Estimation (Adam) Optimizer, and adopted the grid search strategy to get the optimal values for hyper-parameters. The learning rate, the number of filters (i.e.,  $K$ ), and the batch size were searched in ranges of [0.00001, 0.0001, 0.001, 0.01, 0.1], [64, 128, 512, 1024], and [16, 32, 64, 128], respectively. Moreover, we fine-tuned the proposed AHG-Net based on the training and validation dataset with 200 epochs, and reported the model performance on the testing dataset. All the experiments were conducted on a server equipped with 32 Cores Intel (R) Xeon (R) CPU E5-2620 v4 2.10GHz and four TITAN Xp GPUs.

### 4.3 On Model Comparison (RQ1)

To verify the effectiveness of our AHG-Net, we chose the following state-of-the-art methods on user identity linkage as baselines.

- **DPLink** [10] resorts to the user’s check-in posts to tackle the user identity linkage task, where the spatial one-hot representation and corresponding temporal one-hot representation was concatenated as the check-in post representation, and the recurrent network was used to derive the user representation from the check-in post sequence.
- **MV\_URL** [38] captures the multi-view representation of the user by constructing heterogeneous graphs (e.g., User-User graph and User-Word graph) based on the user’s textual posts and social connections.
- **DLHD** [12] first measures the modality similarities, i.e., textual and social similarity, between two user accounts on different platforms with MLPs, and then aggregates them with a MLP to derive the final user similarity score.

<sup>3</sup><https://developer.twitter.com/en/docs/twitter-api>.

<sup>4</sup><https://developer.foursquare.com/>.

|        | $u_1$  |  |  | $u_2$  |  |  |
|--------|--|--|--|--|--|--|
| Text   | Beer Sunday specials happy hour at 5 pm at an Irish pub with some flair & great tasting wings. Try Moriarty's Pub. |  |  | Be sure to try Chef Beck's signature sandwich, the Train Wreck. Fantastic. Any of the food is good.                |  |  |
| Image  |  |  |  |  |  |  |
| Check  |  |  |  |  |  |  |
| Social | ABC, dfsppgh, JackWagner54, keewood, phitocomedy, Roose_TTB, ScottBarnes, sparkey215, UnleashedPetSpa, ambde       |  |  | ambde, denims, drinkphilly, jackwagner54, phitocomedy, scottbarnes, sparkey215, unleashedpetspa, beejer, roose_ttb |  |  |

Figure 3: Illustration of the positive user pair that is correctly judged by AHG-Net. “Text”, “Image”, “Check”, and “Social” refer to textual posts, visual posts, check-in posts, and social connection, respectively.

- MSUIL\_V [19] employs an encoder and decoder to project the representation from one platform into another platform, where the user’s textual, visual, check-in and social posts are simultaneously explored. Thereafter, for each platform, it further sets a discriminator to distinguish whether the given vector is original or projected.
- MNA\_V, derived from [17], utilizes five hand-crafted heterogeneous features, including textual, visual, temporal, spatial and social features to characterize each user, and employs SVM to fulfil the user identity linkage.
- ACTIVER [31] treats the user pair as a whole and learns the pair representation based on a series of pre-defined meta diagrams, where the users’ textual, temporal, spatial, and social modalities are jointly utilized. Ultimately, the user similarity is accessed by the learned pair representation.
- ACTIVER\_V is an extension of ACTIVER, where visual posts are also added into the pre-defined meta diagrams via histogram features [27] following [31].

Table 2 shows the performance comparison among different methods with respect to different evaluation metrics. From this table, we can draw the following observations: 1) AHG-Net consistently surpasses all the baselines, exhibiting the effectiveness of the proposed network. This may be attributed to the fact that AHG-Net is able to capture the hybrid user representation from both intra-user and inter-user perspectives and alleviate the semantic gap problem between different social platforms. 2) AHG-Net, ACTIVER, MNA\_V, MSUIL\_V and ACTIVER\_V outperform DLHD and MV\_URL, which indicates the advantage of simultaneously incorporating heterogeneous multi-modal posts in the user characterization. 3) DPLink gets the worst performance compared to other methods. This may be due to the fact that DPLink focuses on the single type of user information, i.e., the check-in posts, while overlooks other multi-modal posts and social connections, which limits its representation capability for the user. 4) AHG-Net exceeds MNA\_V, MSUIL\_V and ACTIVER\_V, which confirms the benefit of exploiting complicated relations hidden in heterogeneous posts and social connections. And 5) we found that ACTIVER shows superiority over ACTIVER\_V, which reflects that simply integrating the visual posts may bring noise to the user representation and thus hurt the performance.

Table 3: Performance comparison between AHG-Net and its derivatives.

| Models             | Accuracy      | Precision     | Recall        | F1-score      |
|--------------------|---------------|---------------|---------------|---------------|
| AHG-Net-w/o-inter  | 59.81%        | 65.31%        | 55.17%        | 59.81%        |
| AHG-Net-w/o-interG | 90.65%        | 92.86%        | 89.66%        | 91.23%        |
| AHG-Net-w/o-intra  | 89.72%        | 91.23%        | 89.66%        | 90.43%        |
| AHG-Net-w/o-intraG | 88.79%        | 91.07%        | 87.93%        | 89.47%        |
| AHG-Net-w/o-A      | 88.79%        | 92.59%        | 86.21%        | 89.29%        |
| <b>AHG-Net</b>     | <b>91.59%</b> | <b>92.98%</b> | <b>91.38%</b> | <b>92.17%</b> |

To intuitively show the effectiveness of our AHG-Net, we sampled a positive testing user account pair that our model correctly classified as the same user identity. Due to the space limit, we only show the meaningful posts and social connections in Figure 3. As can be seen, the same user identity does share similar multi-modal cues and social connections on different social platforms.

#### 4.4 On Ablation Study (RQ2)

To get a thorough understanding of our proposed model, we compared AHG-Net with the following five derivations. 1) **AHG-Net-w/o-inter**. We disabled the inter-user representation learning. 2) **AHG-Net-w/o-interG**. We kept the inter-user representation learning, but removed its attention mechanism in the information aggregation by allocating the same confidence for all connections of the user. 3) **AHG-Net-w/o-intra**. We discarded the intra-user representation learning. 4) **AHG-Net-w/o-intraG**. We retained the intra-user representation learning, but replaced the GCN with an average pooling over the user’s multi-modal representations (i.e., c, v, t). And 5) **AHG-Net-w/o-A**. We removed the semantic discriminator from AHG-Net.

Table 3 illustrates the performance comparison between our AHG-Net and its derivatives. Firstly, as can be seen, AHG-Net outperforms both AHG-Net-w/o-inter and AHG-Net-w/o-intra, demonstrating that removing either the inter-user or the intra-user representation learning will hurt the performance of AHG-Net to some extent. The rationale behind is that the intra-user representation learning can capture the semantic relations among users’ heterogeneous multi-modal posts, while the inter-user representation learning is able to distinguish the informative followees, both benefiting the final user representation learning. Secondly, we found that AHG-Net-w/o-intra performs better than

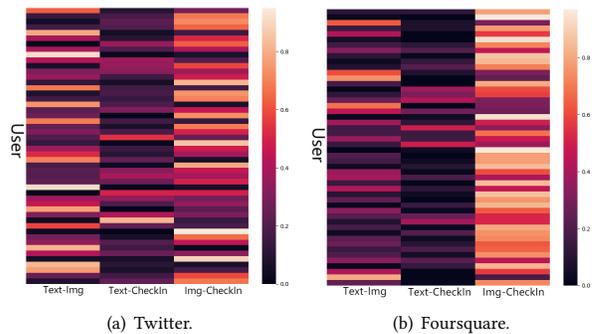


Figure 4: Visualization of the semantic adjacent matrix in the intra-user representation learning.

|              | User  | Followee1 $\alpha_{c_1}^1: 0.04$   | Followee2 $\alpha_{c_2}^1: 0.06$   |
|--------------|---|--|--|
| Textual Post | Big Mixerz Sound Lab will have a hybrid analog/digital recording system with 40 analog tracks available at mixdown...WOWWWW!!!!!! | New Jersey has resolved yearslong litigation in connection to its water crisis, in which city drinking water was polluted with illegally high levels of lead.            | Sometimes all you need is a good sunset and a piano especially with how this week has been See ya'll tonight.                        |
|              | Luck" is when preparation meets opportunity.SO PREPARE YOURSELF!! #BIGMIXERZ!!!   | A tornado strikes Florida, damaging Tallahassee International Airport.   | Y'all asked for it!! The love 4 Wasted Energy is! Cooked up some magic on the remix 🌟🌟   |
|              | 10 songs on industry beats mixed and mastered \$250 Or 10 songs on original beats mixed and mastered for \$500 Hit me up.!!!!     | A tornado described as "large and extremely dangerous" by the National Weather Service in Birmingham, Alabama, ripped through a hotel.                                   | Big things coming in 2021 for all of us! No holding back from your greatness!! Ya'll ready???Drop some🌟🌟 in the comments if you are! |
| Visual Post  |    |   |   |
|              |    | <br> |   |

Figure 5: Illustration of different confidences over different followees.

AHG-Net-w/o-inter, indicating that the inter-user representation learning contributes more to the user characterization, as compared with the intra-user representation learning. This suggests that social connections, especially the followee relations, are more reliable than one’s social posts in terms of the user representation. Thirdly, it is surprising that AHG-Net-w/o-intra somehow outperforms AHG-Net-w/o-intraG. One possible explanation is that directly combining the representations of multi-modal posts can fuse noisy user representation from the least reliable modality, where the user’s data may be insufficient to support the user representation learning. This also confirms that it is essential to capture the semantic relations among users’ heterogeneous multi-modal posts and achieve the optimal user representation. Fourthly, we observed that AHG-Net surpasses AHG-Net-w/o-interG, which implies the advantage of adaptively integrating the social connection representations to learn the user representation. Last but not least, AHG-Net shows superiority over AHG-Net-w/o-A, demonstrating the crucial importance of the semantic discriminator towards the user identity linkage.

To obtain deeper insights on the intra-user representation learning, we randomly sampled 50 testing users on each social media platform, and exhibited their learned semantic adjacent matrices in Eqn. (5), as shown in Figure 4. The lighter the color

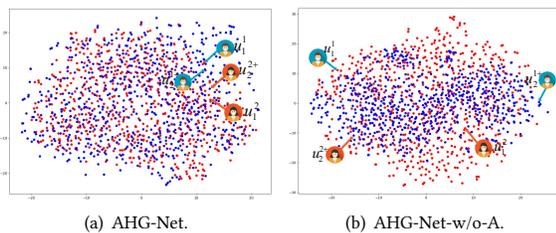


Figure 6: Visualization of the user representation distribution on Twitter and Foursquare by AHG-Net and AHG-Net-w/o-A. The red points represent the user representations on Twitter, and the blue ones refer to that on Foursquare.  $(u_1^1, u_2^1)$  and  $(u_1^2, u_2^2)$  are positive user pairs.

is, the higher the semantic similarity between the two modalities. Each line corresponds to one user’s semantic similarities among different modalities. As can be seen from Figure 4, there do exist distinguished semantic similarities among different modalities. In particular, we observed that the semantic similarity between textual and check-in modalities is consistently the weakest one, while that between the visual and the check-in modalities is consistently the most prominent one on both platforms. This may be due to the fact that images are more likely to be posted with check-in tags and textual descriptions.

To intuitively show the effectiveness of our inter-user representation learning, we also performed the case study on the followee confidence assignment with a testing user and his/her two followees, as shown in Figure 5. Due to the limited space, we only provided several textual and visual posts of the user. As we can see, towards the user representation learning, AHG-Net assigned the higher confidence to the followee2, as compared to the followee1. Checking the user’s historical posts, we learned that the given user is a music producer and always shares his music life. Meanwhile, we found that the followee1 likes to broadcast news, while the followee2 is a singer who often posts her daily life and music production. In light of this, the confidence assignment of our model regarding these two followees for the given user is reasonable. This suggests that our model is able to learn the latent similarity between users.

To intuitively reflect the effectiveness of the semantic discriminator, we visualized the learned user representations on both platforms by AHG-Net and AHG-Net-w/o-A with the help of tSNE [29] in Figure 6. The red points denote the user representations on Twitter, and the blue ones refer to that on Foursquare.  $(u_1^1, u_2^1)$  and  $(u_1^2, u_2^2)$  are positive user pairs. As we can see, user representations on different platforms achieved by AHG-Net are more uniformly distributed and less distinguishable from each other, as compared to that obtained by AHG-Net-w/o-A. On one hand, this confirms that the semantic gap problem does exist and the adversarial learning can well fix it. On the other hand, this implies that with adversarial learning, AHG-Net is able to push

**Table 4: Performance of AHG-Net with different modality configurations.**

| Models             | Accuracy      | Precision     | Recall        | F1-score      |
|--------------------|---------------|---------------|---------------|---------------|
| AHG-Net-w/o-Text   | 86.92%        | 89.29%        | 86.21%        | 87.72%        |
| AHG-Net-w/o-Image  | 90.65%        | 92.86%        | 89.66%        | 91.23%        |
| AHG-Net-w/o-Check  | 88.79%        | 91.07%        | 87.93%        | 89.47%        |
| AHG-Net-w/o-Social | 59.81%        | 65.31%        | 55.17%        | 59.81%        |
| <b>AHG-Net</b>     | <b>91.59%</b> | <b>92.98%</b> | <b>91.38%</b> | <b>92.17%</b> |

the representations of the same user identity on different social platforms to be similar, such as  $(u_1^1, u_2^1)$  and  $(u_1^2, u_2^2)$ .

#### 4.5 On Modality (RQ3)

To explore the roles of different modalities about characterizing the user, we conducted the comparative experiment with the following derivatives: **AHG-Net-w/o-Text**, **AHG-Net-w/o-Image**, **AHG-Net-w/o-Check**, **AHG-Net-w/o-Social**, where textual posts, visual posts, check-in posts and social connections are removed, respectively. Notably, since discarding the users' social connection inevitably disables the inter-user representation learning, AHG-Net-w/o-Social is essentially the same as AHG-Net-w/o-Inter, which has been mentioned in the ablation study.

Table 4 summarizes the performance of AHG-Net with different modality configurations. From Table 4, we had the following observations. Firstly, removing any type of posts or the social connections would inevitably hurt the performance of AHG-Net, which validates the necessity of incorporating both heterogeneous multi-modal posts and social connections in the context of user identity linkage. In a sense, this illustrates that different modality posts can convey distinguished cues and complement each other towards the user characterization. Secondly, we observed that AHG-Net-w/o-Social obtains the worst performance, which suggests the dominant role of the social connection in the user identity linkage. Thirdly, AHG-Net-w/o-Image and AHG-Net-w/o-Check show superiority over AHG-Net-w/o-Text, indicating the effectiveness of the image modality and check-in modality are limited as compared to the textual modality. The possible reason may be two folds. 1) People tend to post similar textual posts on different social platforms, as compared to visual posts and check-in posts. In fact, people usually prefer to post check-ins and images on Foursquare rather than Twitter. And 2) the textual posts are more straightforward to characterize the user, while the image contents and check-in posts are relatively implicit towards the user identity linkage.

#### 4.6 On Missing Data (RQ4)

Due to the concern that not every user is attached to the completed modality information, we further investigated the effectiveness of our model with incomplete dataset, that is, some modalities can be

**Table 5: Statistics of TWFQ-M.**

| Category           | Twitter   | Foursquare | Total Number |
|--------------------|-----------|------------|--------------|
| #User              | 4,014     | 4,505      | 8,519        |
| #Textual Post      | 4,097,233 | 31,702     | 4,128,935    |
| #Visual Post       | 52,359    | 42,495     | 94,854       |
| #Check-in Post     | 292,254   | 25,799     | 318,053      |
| #Social Connection | 94,288    | 44,723     | 139,011      |

**Table 6: Performance comparison among different methods in the condition of missing data.**

| Models           | Accuracy      | Precision     | Recall        | F1-score      |
|------------------|---------------|---------------|---------------|---------------|
| DPLink-M         | -             | -             | -             | -             |
| MV_URL-M         | 56.68%        | 62.86%        | 54.55%        | 58.41%        |
| DLHD-M           | 57.14%        | 60.94%        | 64.46%        | 62.65%        |
| MNA_V-M          | 70.51%        | 79.47%        | 55.30%        | 65.22%        |
| ACTIVER_V-M      | 76.50%        | 80.10%        | 70.51%        | 75.00%        |
| MSUIL_V-M        | 82.03%        | 85.96%        | 80.99%        | 83.40%        |
| ACTIVER-M        | 83.18%        | 83.33%        | 82.95%        | 83.14%        |
| <b>AHG-Net-M</b> | <b>88.02%</b> | <b>90.60%</b> | <b>87.60%</b> | <b>89.08%</b> |

missing for some users. In particular, to guarantee the network integrity, we retained those users who have at least the social connections as well as one type of multi-modal post. In this way, we obtained another dataset, named TWFQ-M, with 2,175 positive user pairs. Table 5 refers to the statistics of TWFQ-M. To deal with TWFQ-M, we adapted the proposed AHG-Net into AHG-Net-M by masking the corresponding network structures in the user representation extraction, and utilizing the user representation learning zero padding.

Table 6 shows the performance comparison between AHG-Net-M and baselines. Notably, DPLink-M is not adopted for comparison, since it only utilizes the check-in posts, which cannot support user identity linkage with missing data. As can be seen, our AHG-Net-M outperforms all baselines, validating the effectiveness of AHG-Net-M in the cases where users' data is incomplete.

## 5 CONCLUSION

In this paper, we investigate the user identity linkage task based on heterogeneous multi-modal posts as well as social connections. In particular, we propose a novel adversarial-enhanced hybrid graph network for user identity linkage, named AHG-Net, which consists of three pivotal components: *user representation extraction*, *hybrid user representation learning*, and *adversarial learning*. To promote the evaluation, we build a multi-modal dataset by augmenting the existing public dataset with 62,021 visual posts. Extensive experiments on this dataset validate the effectiveness of the proposed AHG-Net. Interestingly, we observe that the intra-user representation learning and the inter-user representation learning are conducive to the user characterization, and they can further complement each other. Besides, the semantic gap problem caused by different data distributions of social media platforms does exist in the context of user identity linkage and should be taken into account. In addition, incorporating both heterogeneous multi-modal posts (i.e., textual, visual, and check-in) and social connection in user characterization can facilitate the conduction of user identity linkage. We have released the dataset, codes, and parameters to facilitate other researchers. Currently, we center on the supervised user identity linkage, which needs a lot of annotations. In the future, we plan to explore the unsupervised user identity linkage.

## ACKNOWLEDGMENTS

This work is supported by the National Key Research and Development Project of New Generation Artificial Intelligence, No.:2018AAA0102502.

## REFERENCES

- [1] Da Cao, Liqiang Nie, Xiangnan He, Xiaochi Wei, Shunzhi Zhu, and Tat-Seng Chua. 2017. Embedding Factorization Models for Jointly Recommending Items and User Generated Lists. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 585–594.
- [2] Xiaolin Chen, Xuemeng Song, Siwei Cui, Tian Gan, and Liqiang Nie. 2020. User Identity Linkage across Social Media via Attentive Time-aware User Modeling. *IEEE Transactions on Multimedia* PP, 99 (2020), 1–1.
- [3] Xiaolin Chen, Xuemeng Song, Ruiyang Ren, Lei Zhu, Zhiyong Cheng, and Liqiang Nie. 2020. Fine-Grained Privacy Detection with Graph-Regularized Hierarchical Attentive Representation Learning. *ACM Transactions on Information Systems* 38, 4 (2020), 37:1–37:26.
- [4] Dawei Cheng, Yi Tu, Zhen-Wei Ma, Zhibin Niu, and Liqing Zhang. 2019. Risk Assessment for Networked-guarantee Loans Using High-order Graph Attention Representation. In *Proceedings of the International Joint Conference on Artificial Intelligence*. ijcai.org, 5822–5828.
- [5] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *CoRR* abs/1412.3555 (2014).
- [6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, 4171–4186.
- [7] Fuli Feng, Xiangnan He, Jie Tang, and Tat-Seng Chua. 2019. Graph Adversarial Training: Dynamically Regularizing Based on Graph Structure. *IEEE Transactions on Knowledge and Data Engineering* (2019).
- [8] Fuli Feng, Xiangnan He, Xiang Wang, Cheng Luo, Yiqun Liu, and Tat-Seng Chua. 2019. Temporal Relational Ranking for Stock Prediction. *ACM Transactions on Information Systems* 37, 2 (2019), 1–30.
- [9] Fuli Feng, Weiran Huang, Xiangnan He, Xin Xin, Qifan Wang, and Tat-Seng Chua. 2021. Should Graph Convolution Trust Neighbors? A Simple Causal Inference Method. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. Association for Computing Machinery.
- [10] Jie Feng, Mingyang Zhang, Huandong Wang, Zeyu Yang, Chao Zhang, Yong Li, and Depeng Jin. 2019. DPLink: User Identity Linkage via Deep Neural Network From Heterogeneous Mobility Data. In *The World Wide Web Conference*. 459–469.
- [11] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. 2014. Generative Adversarial Nets. In *Annual Conference Neural Information Processing Systems*. 2672–2680.
- [12] Asmelash Teka Hadgu and Jayanth Kumar Reddy Gundam. 2019. User Identity Linking Across Social Networks by Jointly Modeling Heterogeneous Data with Deep Learning. In *Proceedings of the ACM Conference on Hypertext and Social Media*. ACM, 293–294.
- [13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 770–778.
- [14] Yupeng Hu, Meng Liu, Xiaobin Su, Zan Gao, and Liqiang Nie. 2021. Video Moment Localization via Deep Cross-modal Hashing. *IEEE Transactions on Image Processing* 30 (2021), 4667–4677.
- [15] Yupeng Hu, Peng Zhan, Yang Xu, Jia Zhao, Yujun Li, and Xueqing Li. 2021. Temporal Representation Learning for Time Series Classification. *Neural Computing and Applications* 33, 8 (2021), 3169–3182.
- [16] Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. ACL, 1746–1751.
- [17] Xiangnan Kong, Jiawei Zhang, and Philip S. Yu. 2013. Inferring Anchor Links across Multiple Heterogeneous Social Networks. In *ACM International Conference on Information and Knowledge Management*. ACM, 179–188.
- [18] Nitish Korula and Silvio Lattanzi. 2014. An Efficient Reconciliation Algorithm for Social Networks. *Proceedings of the VLDB Endowment* 7, 5 (2014), 377–388.
- [19] Chaozhuo Li, Senzhang Wang, Hao Wang, Yanbo Liang, Philip S. Yu, Zhoujun Li, and Wei Wang. 2019. Partially Shared Adversarial Learning For Semi-supervised Multi-platform User Identity Linkage. In *Proceedings of the ACM International Conference on Information and Knowledge Management*. ACM, 249–258.
- [20] 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 *Proceedings of the ACM International Conference on Information and Knowledge Management*. ACM, 447–456.
- [21] Chun Hung Li and C. K. Lee. 1993. Minimum Cross Entropy Thresholding. *Pattern Recognition* 26, 4 (1993), 617–625.
- [22] Siyuan Liu, Shuhui Wang, Feida Zhu, Jinbo Zhang, and Ramayya Krishnan. 2014. HYDRA: Large-Scale Social Identity Linkage via Heterogeneous Behavior Modeling. In *International Conference on Management of Data*. ACM, 51–62.
- [23] Lihua Lu, Yao Lu, Ruizhe Yu, Huijun Di, Lin Zhang, and Shunzhou Wang. 2020. GAIM: Graph Attention Interaction Model for Collective Activity Recognition. *IEEE Transactions on Multimedia* 22, 2 (2020), 524–539.
- [24] Tong Man, Huawei Shen, Shenghua Liu, Xiaolong Jin, and Xueqi Cheng. 2016. Predict Anchor Links across Social Networks via an Embedding Approach. In *Proceedings of the International Joint Conference on Artificial Intelligence*. IJCAI AAAI Press, 1823–1829.
- [25] 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 *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1775–1784.
- [26] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. DeepWalk: Online Learning of Social Representations. In *The ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 701–710.
- [27] Dung Phan, In Seop Na, and Soo-Hyung Kim. 2014. Local Features and Histogram Based Planar Object Recognition. In *Proceedings of the International Conference on Ubiquitous Information Management and Communication*. ACM, 49:1–49:4.
- [28] Jiezhong Qiu, Jian Tang, Hao Ma, Yuxiao Dong, Kuansan Wang, and Jie Tang. 2018. DeepInf: Social Influence Prediction with Deep Learning. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery Data Mining*. ACM.
- [29] Paulo E. Rauber, Alexandre X. Falcao, and Alexandru C. Telea. 2016. Visualizing Time-Dependent Data Using Dynamic t-SNE. In *Eurographics Conference on Visualization*. Eurographics Association, 73–77.
- [30] Ruiyang Ren, Zhaoyang Liu, Yaliang Li, Wayne Xin Zhao, Hui Wang, Bolin Ding, and Ji-Rong Wen. 2020. Sequential Recommendation with Self-Attentive Multi-Adversarial Network. In *Proceedings of the International ACM SIGIR conference on research and development in Information Retrieval*. ACM, 89–98.
- [31] Yuxiang Ren, Charu Aggarwal, and Jiawei Zhang. 2019. Activer: Meta Diagram based Active Learning in Social Networks Alignment. *IEEE Transactions on Knowledge and Data Engineering* (2019).
- [32] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*. AUAI Press, 452–461.
- [33] Kai Shu, Shuhang Wang, Jiliang Tang, Reza Zafarani, and Huan Liu. 2016. User Identity Linkage across Online Social Networks: A Review. *SIGKDD Explorations* 18, 2 (2016), 5–17.
- [34] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All You Need. In *Annual Conference on Neural Information Processing Systems*. 5998–6008.
- [35] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2018. Graph Attention Networks. In *International Conference on Learning Representations*.
- [36] Bokun Wang, Yang Yang, Xing Xu, Alan Hanjalic, and Heng Tao Shen. 2017. Adversarial Cross-Modal Retrieval. In *Proceedings of the ACM on Multimedia Conference*. ACM, 154–162.
- [37] Pengfei Wang, Yu Fan, Shuzi Niu, Ze Yang, Yongfeng Zhang, and Jiafeng Guo. 2019. Hierarchical Matching Network for Crime Classification. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 325–334.
- [38] Weiqing Wang, Hongzhi Yin, Xingzhong Du, Wen Hua, Yongjun Li, and Quoc Viet Hung Nguyen. 2019. Online User Representation Learning Across Heterogeneous Social Networks. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 545–554.
- [39] Yinwei Wei, Xiang Wang, Weiwei Guan, Liqiang Nie, Zhouchen Lin, and Baoquan Chen. 2019. Neural multimodal cooperative learning toward micro-video understanding. *IEEE Transactions on Image Processing* 29 (2019), 1–14.
- [40] Yinwei Wei, Xiang Wang, Liqiang Nie, Xiangnan He, Richang Hong, and Tat-Seng Chua. 2019. MMGCN: Multi-modal Graph Convolution Network for Personalized Recommendation of Micro-video. In *Proceedings of the ACM International Conference on Multimedia*. 1437–1445.
- [41] Ming Yan, Jitao Sang, and Changsheng Xu. 2014. Mining Cross-network Association for YouTube Video Promotion. In *Proceedings of the ACM International Conference on Multimedia*. ACM, 557–566.
- [42] Xin Yang, Xuemeng Song, Xianjing Han, Haokun Wen, Jie Nie, and Liqiang Nie. 2020. Generative Attribute Manipulation Scheme for Flexible Fashion Search. In *Proceedings of the International ACM SIGIR conference on research and development in Information Retrieval*. ACM, 941–950.
- [43] Reza Zafarani and Huan Liu. 2009. Connecting Corresponding Identities across Communities. In *Proceedings of the Third International Conference on Weblogs and Social Media*. The AAAI Press.
- [44] Qianyi Zhan, Jiawei Zhang, Senzhang Wang, Philip S. Yu, and Junyuan Xie. 2015. Influence Maximization Across Partially Aligned Heterogeneous Social Networks. In *Advances in Knowledge Discovery and Data Mining*. Springer, 58–69.
- [45] Rong Zheng, Jiexun Li, Hsinchun Chen, and Zan Huang. 2006. A Framework for Authorship Identification of Online Messages: Writing-Style Features and Classification Techniques. *Journal of the American society for information science and technology* 57, 3 (2006), 378–393.