electronics

MDPI

Article

Cross-Social-Network User Identification Based on Bidirectional GCN and MNF-UI Models

Song Huang [1], Huiyu Xiang [1,*], Chongjie Leng [1] and Feng Xiao [2]

[1] School of Computer and Artificial Intelligence, Beijing Technology and Business University, Beijing 100048, China; 2230601007@st.btbu.edu.cn (S.H.); lengchongjie@163.com (C.L.)
[2] Cyberspace Security Department, Beijing Electronic Science & Technology Institute, Beijing 102627, China; xiaofrank777@163.com
* Correspondence: xianghy@th.btbu.edu.cn

Abstract: Due to the distinct functionalities of various social network platforms, users often register accounts on different platforms, posing significant challenges for unified user management. However, current multi-social-network user identification algorithms heavily rely on user attributes and cannot perform user identification across multiple social networks. To address these issues, this paper proposes two identity recognition models. The first model is a cross-social-network user identification model based on bidirectional GCN. It calculates user intimacy using the Jaccard similarity coefficient and constructs an adjacency matrix to accurately represent user relationships in the social network. It then extracts cross-social-network user information to accomplish user identification tasks. The second model is the multi-network feature user identification (MNF-UI) model, which introduces the concept of network feature vectors. It effectively maps the structural features of different social networks and performs user identification based on the common features of seed nodes in the cross-network environment. Experimental results demonstrate that the bidirectional GCN model significantly outperforms baseline algorithms in cross-social-network user identification tasks. The MNF-UI (multi-network feature user identification) model can operate in situations with two or more networks with inconsistent structures, resulting in improved identification accuracy. These two user identification algorithms provide technical and theoretical support for in-depth research on social network information integration and network security maintenance.

Keywords: user identification; cross-social network; bidirectional GCN; MNF-UI; network embeddedness

check for updates

Citation: Huang, S.; Xiang, H.; Leng, C.; Xiao, F. Cross-Social-Network User Identification Based on Bidirectional GCN and MNF-UI Models. *Electronics* **2024**, *13*, 2351. https://doi.org/10.3390/electronics13122351

Academic Editor: Christos J. Bouras

Received: 23 April 2024
Revised: 27 May 2024
Accepted: 31 May 2024
Published: 15 June 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In the era of rapid mobile internet development, social networks have become an integral part of daily life, increasingly recognized for their importance [1]. Various social network platforms vie for user registration by leveraging their unique core functionalities, allowing users to experience these specific features. However, this trend poses significant challenges to the management of user accounts uniformly. In response, multi-social-network user identification technology has emerged, aiming to identify accounts of the same individuals across different social network platforms, thus enabling cross-platform data integration [2]. This technology has profound impacts on user recommendation systems, network security, and public opinion monitoring. Currently, the field of social network user identification employs various methods, mainly categorized into three types: based on user attributes, based on user relationships (i.e., the topology of social networks), and based on user behavior information [3].

Firstly, methods based on user attributes rely on personal information such as age, gender, occupation, and other static information for identification. These methods are popular for their simplicity and ease of data acquisition. For instance, Zeng W et al. proposed a method for user identification across online social networks by integrating

multi-dimensional information such as personal profiles, social relationships, and online behaviors, achieving efficient cross-platform user identity recognition [4]. Çoban Ö explored a username-based method for identifying users across online social networks, using usernames to match accounts belonging to the same user, showing that the learning function constructed by binary classification outperformed similarity methods, reaching an optimal F-score of 0.921 without feature selection and expansion [5]. Li Y investigated a social network user identification method based on usernames and display names (UISN-UD), demonstrating the possibility of matching user accounts by comparing display names or usernames with limited online data, providing practical support for building better user profiles [6]. Solanki P and Harwood A discussed a method for cross-platform user identification using personal profiles and posting patterns on social network sites. Their study developed an algorithm by analyzing personal information such as interests and professions to match and identify the same user on different social network platforms [7]. Secondly, methods based on user relationships utilize the web of relationships in social networks, i.e., the interactions and social circles forming the topology. These methods reflect users' social activities and preferences well, with easier and more authentic network structure acquisition. For example, Qu Y and his team proposed a friendship-based identification method (FBI) by analyzing and learning the friendship link patterns of users on different social networks [8]. Kojima K et al. explored a method based on the similarity of friends' location distributions across online social networks, assuming that users' social circles have certain similarities even on different networks. By comparing and analyzing the location distribution of users' friends, the study proposed an innovative identification strategy that effectively identifies the same user across platforms [9]. Ahmad W and Ali R introduced a method for user identification across multiple online social networks, relying on cross-link attributes and network relationships, developing an algorithm to identify and link the same user on different platforms [10]. Qu Y et al. introduced a method for cross-social network user identification by utilizing users' friendship networks. This approach matches and identifies user identities based on their social relationships across different platforms, demonstrating the potential of leveraging structural information of social networks in user identification [11]. Lastly, methods based on user behavior information focus on the behavior patterns on social platforms, such as topics of posts and activity times. These methods capture dynamic characteristics of users, providing a richer dimension for user identification. Xing L et al. focused on identifying users across platforms by their behavior habits in social networks, based on the key observation that despite different activities, users' behavior habits and patterns have certain consistencies. By analyzing and comparing users' behavior habits, such as posting frequency, active time slots, and interaction preferences, the research team developed a novel identification algorithm [12]. Solanki P and Harwood A explored a method using user profiles and posting patterns for cross-social-network site user identification, proposing an innovative algorithm that considers both personal profile information and posting behavior patterns on different social networks [7]. Lei T and colleagues proposed a user identification method based on machine learning. This method effectively improves the accuracy of identification by analyzing users' behaviors and social relationships across different platforms. This work not only offers a new technical approach for cross-social- network user identification, but also opens new possibilities for research and applications in related fields [13]. In recent years, graph neural networks (GNNs) have made significant progress in processing graph-structured data, finding wide applications in social network analysis, recommendation systems, bioinformatics, and other fields. However, existing GNN methods still face numerous challenges when dealing with complex graph structures and large-scale data. To improve the performance and efficiency of GNNs, researchers have proposed a series of enhancement methods, including APGVAE (adaptive disentangled representation learning with the graph-based structure information) and injective aggregation techniques. APGVAE is a novel adaptive graph variational autoencoder method that aims to improve the performance of GNNs by incorporating graph structure information for disentangled

representation learning [14]. This method not only learns more interpretable representations, but also enhances representation learning by adaptively integrating graph structure information. However, the complexity and computational resource requirements of the APGVAE method present challenges when processing large-scale graph data. Another study proposed improving GNN performance and efficiency through injective aggregation techniques [15]. Injective aggregation ensures that the aggregation function uniquely maps different input sets, thereby better capturing the complex dependencies in graph data. This approach has shown significant performance improvements in practical applications, but it also faces challenges related to the increased implementation complexity and computational resource demands. Despite achievements in addressing user identification issues, challenges remain due to specific reasons, summarized as follows:

1. Missing user attributes. The absence or inaccuracy of user attribute information primarily stems from three aspects. First, users often only provide the most basic mandatory information when filling in personal details, due to concerns over privacy protection or to avoid the hassle. Secondly, as privacy awareness increases, more and more users opt to set their personal information to limited visibility, such as visible only to friends or completely confidential, making direct access to user information more challenging. Lastly, some users may fill in incorrect or false personal details due to reluctance to disclose true information. Therefore, traditional methods based on user attribute matching significantly lose effectiveness under these circumstances, urgently necessitating the exploration of more effective user identification techniques.

2. Heterogeneity of data structures. Different social network platforms, due to their unique functional requirements, lead to diverse rules that users follow when submitting information at the time of registration. This results in a variety of data formats. For example, Facebook allows the collection of diverse information including usernames, nicknames, addresses, interests, etc., while Twitter mainly limits data collection to basic information. Meanwhile, the representation of user relationships across social networks also varies, such as the bilateral friend relationships in WeChat versus the unilateral following mechanism in Weibo. These differences increase the complexity of cross-platform user identification.

3. Incompleteness of user matching. Usually, users on social networking platforms do not correspond completely. That is, some users may be active on one platform, but not exist on others. Therefore, effective user identification requires finding corresponding user accounts across multiple social networks, a process that considers many factors aiming to ensure high accuracy and precision, constituting a challenge.

In response to the limitations observed in current user identification methods across social networks, which predominantly rely on user attributes, we conducted a comprehensive analysis of data from multiple social networking platforms. This analysis particularly focused on friend relationship data in social networks to devise an innovative user identification algorithm. Our paper introduces a cross-social-network user identification model based on bidirectional graph convolutional networks (GCNs), addressing the deficiencies of traditional attribute-dependent methods. This model represents a significant advance in neural network applications by leveraging node classification technology within GCNs to identify the same user across multiple social networks. Key contributions include:

1. Innovative model design: we propose a cross-social-network user identification model that utilizes bidirectional GCNs. This model innovates by initializing user node feature vectors, which allow each node in the network to aggregate information from neighboring nodes. This aggregation process generates a new, enhanced node representation, thereby significantly boosting the model's performance through iterative training.

2. Application to multi-social-network scenarios: our model addresses the challenge of user identification in environments where multiple social networks are involved, and there is a scarcity of user attribute information. Unlike traditional methods, our model, termed MNF-UI, is designed to operate effectively regardless of the number of networks and the variance in their topology structures. This adaptability enhances the applicability and accuracy of the model, setting a new standard in the field.

This paper aims to fill the gap in existing methods by providing a robust solution to user identification across multiple social network platforms without relying on direct user attribute data, using advanced computational techniques in neural networks. Through our proposed models—particularly the MNF-UI—we demonstrate the feasibility of cross-social-network user identification, significantly advancing the field's current methodologies. The technical roadmap of this paper, as shown in Figure 1, and the subsequent work of this paper are as follows: Section 2 details the cross-social-network user identification model based on bidirectional GCN, describing its design and implementation and validating the effectiveness of the proposed model through a series of experiments. Section 3 focuses on the multi-social-network user identification model based on MNF-UI, outlining the overall design of the model framework, delving into the specific implementation of the multi-social-network user identification model based on MNF-UI, and verifying the model's performance through experiments. Section 4 concludes the paper, aiming to summarize the work presented.

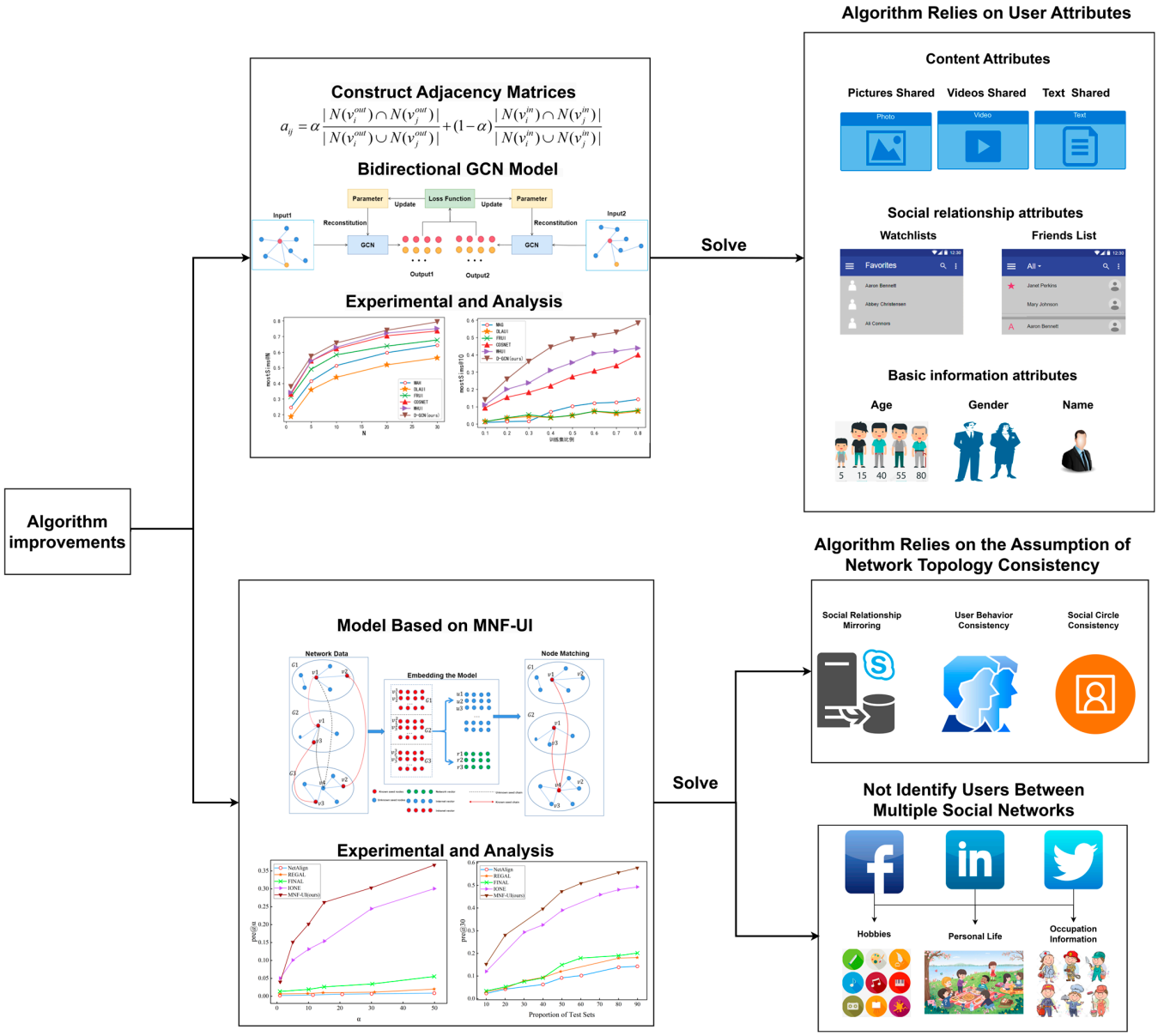


Figure 1. Technical roadmap.

2. Cross-Social-Network User Identification Model Based on Bidirectional GCN

In social networks, each node possesses not only basic attribute information but also embeds complex social connections, such as "friendship relations" of varying scales that construct a sophisticated social network system [16,17]. Against this backdrop, accurately identifying the same users in the real world has become a hot research topic in the field of social network analysis. Traditional identification methods mainly rely on user attribute information, network structural information, or a comprehensive application of both. However, with the increasing awareness of user privacy protection, obtaining users' attribute information has become increasingly difficult, directly leading to a decline in the accuracy of traditional methods for user identification. Methods that rely solely on network structure lack in-depth analysis of individual user characteristics, thereby affecting the precision of identification. To address this issue, this paper develops a bidirectional graph convolutional network (GCN) model that ingeniously circumvents the negative impact caused by missing user attribute information through random initialization of user attributes. Compared to traditional methods that depend solely on user social relationships, our model achieves a significant improvement in identification accuracy. This innovation not only makes up for the deficiencies of existing methods, but also offers a new perspective and technical path for user identification research in social networks.

2.1. Problem Definition and Model Framework

(1) Problem definition

Graphs are the optimal way to represent network information, where users and their relationships within social networks can be represented by a graph $G = (V, E)$, with V representing the set of all user nodes in the graph and E representing the set of all user relationship links within the graph. The following definitions are necessary for the identification of users across multiple social networks:

Definition 1. *User identification can be defined as the process of identifying accounts belonging to the same real-world user across two distinct social networks $G^X(V^X, E^X)$ and $G^Y(V^Y, E^Y)$. This involves identifying user pairs (v_i^X, v_j^Y) representing the same real-world user, where $v_i^X \in V^X$ and $v_j^Y \in V^Y$.*

Definition 2. *Seed nodes are defined as users who have already been identified across different social networks as the same user. They are represented by (s_i^X, s_i^Y), where $s_i^X \in V^X$ and $s_i^Y \in V^Y$ correspond to the same user in the real world, with i being the index of the seed node. The set of seed nodes is called the seed set, represented by $S = \{(s_1^X, s_1^Y), (s_2^X, s_2^Y), \cdots, (s_n^X, s_n^Y)\}$.*

The network topology can be represented by an adjacency matrix $A \in R^{N*N}$, and the features of the nodes can be represented by a feature matrix $H \in R^{N*F}$, where R is the set of real numbers, N is the number of nodes, and F is the number of features per node. As illustrated in Figure 2, the problem to be solved in the identification of users across multiple social networks is to determine whether a user v_i^X in social network G^X and a user v_j^Y in social network G^Y are the same user.

(2) Model framework

Graph convolutional networks (GCN) are a neural network architecture designed specifically for graph data. The core idea is to use the edge and node information within a graph for comprehensive aggregation, thus producing updated node representations [18]. A key aspect of this architecture is that, through the learning changes of the convolutional factors (see Equation (1)) between layers, it achieves the aggregation of surrounding node information and generates new node representations, where H represents the feature matrix, A is the adjacency matrix, and W is the weight matrix. Given that the adjacency matrix itself does not include self-loops, making it incapable of integrating self-information, self-loops are typically added during the node self-update process to incorporate the node's

own information, thereby achieving a comprehensive update of a node's features and its neighbors'. In social networks, the degree of each node varies and, without normalization, nodes with higher degrees may have disproportionately large feature values which could affect feature extraction and model convergence. To address this issue, this study optimizes for the limitation of not considering self-information using the Laplacian matrix and normalizes the feature vectors by performing a square root inverse operation on the degree matrix of the adjacency matrix, as shown in Equation (2), where $\hat{A} = A + I$, A is the network adjacency matrix, and I is the identity matrix, $\hat{D}$ is the degree matrix of $\hat{A}$. Based on this, the activation function σ is where the GCN model employs the ReLU function, as described in Equation (3). The basic implementation process of GCN is illustrated in Figure 3, where graph data are processed and transformed into the graph's adjacency matrix, degree matrix, and node feature matrix, among other forms, to be input into the GCN model. After transformation by the convolutional factors, an updated feature matrix is output.

$$H^{(l+1)} = \sigma(AH^{(l)}W^{(l)}) \tag{1}$$

$$H^{(l+1)} = \sigma(\hat{D}^{-\frac{1}{2}}(\hat{D} - \hat{A})\hat{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}) \tag{2}$$

$$relu(x) = \max(0, x) \tag{3}$$

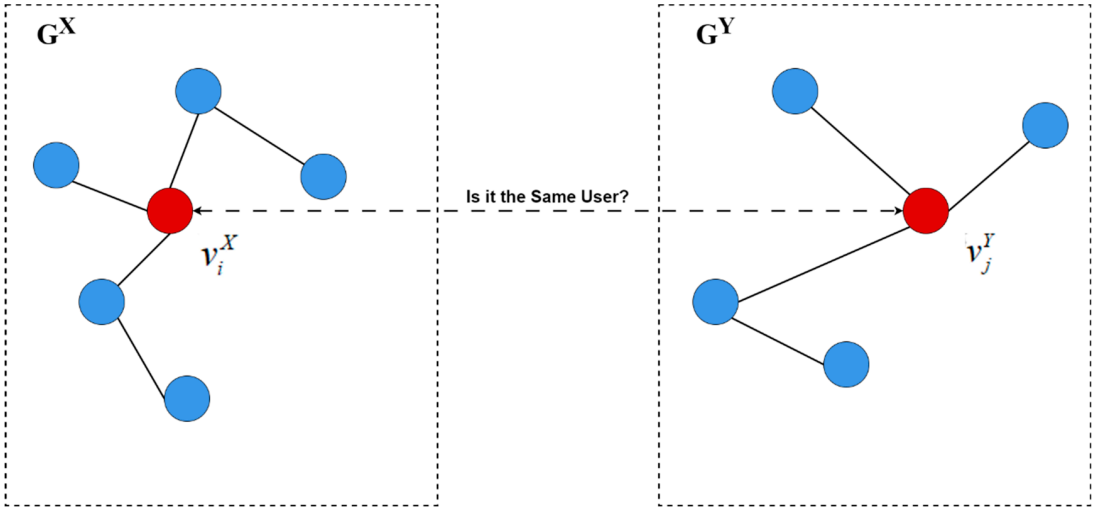


Figure 2. Identifying problems across social network users.

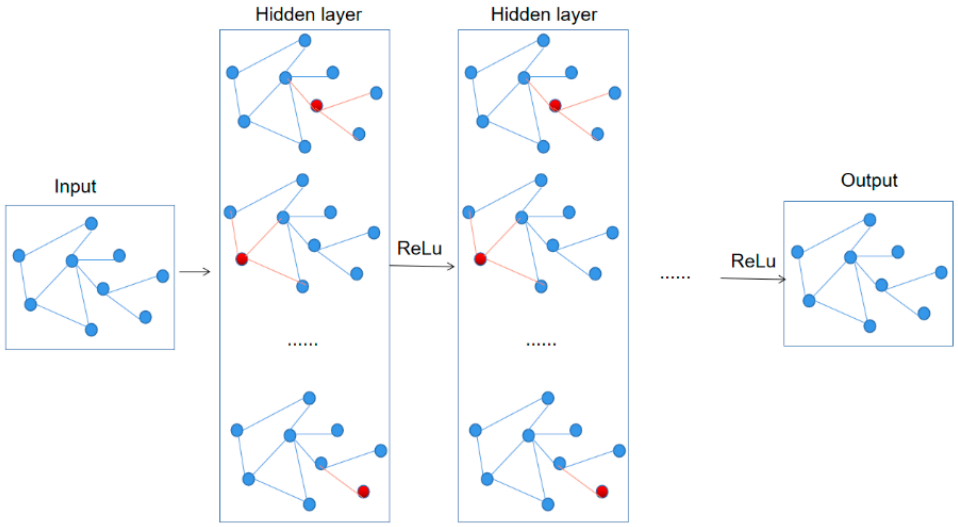


Figure 3. Basic implementation of GCN.

2.2. Model Building

In the field of network science, the adjacency matrix is widely used to describe the topological structure of a network. In this matrix, the value of elements indicates the connectivity between nodes in the graph, with a conventional practice of setting the value between connected nodes to 1 and between unconnected nodes to 0. However, in the context of social networks, the intimacy between users is not uniformly the same. For example, in the friendship relations formed by user node v_1 with user nodes v_2 and v_3, the intimacy between v_1 and v_2 is significantly higher than that between v_1 and v_3. This phenomenon manifests in social networks as the co-occurrence probability between user node v_1 and v_2 being higher than that between v_1 and v_3. Clearly, representing the degree of intimacy between nodes merely with binary values of 0 and 1 would result in the loss of a considerable amount of critical information. To address this issue, this study adopts the Jaccard coefficient to quantify the intimacy between users. The calculation formula for element a_{ij} in the adjacency matrix is shown in Equation (4), where $N(v)$ represents the set of neighbor nodes of user v. Given that most user relationships in social networks are unidirectional, forming directed graphs, such as the "follow" and "be followed" relationships on Weibo. Among all social activities of users, active following and passive being followed reflect different user behaviors, which reveal users' preferences and social circles. Based on this, this study proposes a new method to measure the intimacy between users using the Jaccard coefficient, as detailed in Equation (5). Here, $(v_i, v_j) \in E$, $N(v^{out})$) and $N(v^{in})$ represent the set of neighbor nodes of user's out-links and in-links, respectively. Furthermore, in social networks, there exists a special mutual following relationship, i.e., $(v_i, v_j) \in E$ and $(v_j, v_i) \in E$, for which the adjacency matrices of the two graphs can be calculated separately using the aforementioned method.

$$a_{ij} = \frac{\left|N(v_i) \cap N(v_j)\right|}{\left|N(v_i) \cup N(v_j)\right|} \tag{4}$$

$$a_{ij} = \alpha \frac{\left|N(v_i^{out}) \cap N(v_j^{out})\right|}{\left|N(v_i^{out}) \cup N(v_j^{out})\right|} + (1-\alpha) \frac{\left|N(v_i^{in}) \cap N(v_j^{in})\right|}{\left|N(v_i^{in}) \cup N(v_j^{in})\right|} \tag{5}$$

The core logic of identifying users across social networks is based on the assumption that, if two nodes belonging to different social networks exhibit similar attribute characteristics and close topological structures, then, these two accounts are very likely to belong to the same real-world user [19]. Given the difficulty in directly obtaining node attributes, studies commonly generate users' feature vectors through random initialization methods and, then, based on the connections between users, calculate structural similarity to identify the same user across different social networks [20,21]. Based on this, this study proposes a cross-social-network user identification model based on bidirectional graph convolutional networks (GCN).

The architecture of the model is shown in Figure 4, where Input1 and Input2 represent the comprehensive input information generated by the adjacency matrix construction method and node feature matrix initialization technique mentioned above, respectively. After processing by the graph convolutional network (GCN), the model outputs the node feature matrix. Subsequently, the model optimizes the weight matrix W by minimizing the loss function based on the average distance between corresponding seed nodes in the feature matrices of two social networks (see Equation (6)) and iteratively updates the GCN until the loss function converges. Finally, the output feature matrix is used for further analysis. In this process, n represents the number of seed nodes, d represents the feature dimension of each seed node, v_i^1 and v_i^2 represent the i-th seed node in social networks G^1

and G^2, respectively, and H^1, H^2 are the feature matrices output by the GCN for G^1 and G^2, respectively.

$$f(n, d, v_i^1, v_i^2) = \frac{\sum_{i,j} (H_{i,j}^1 - H_{i,j}^2)}{n} \tag{6}$$

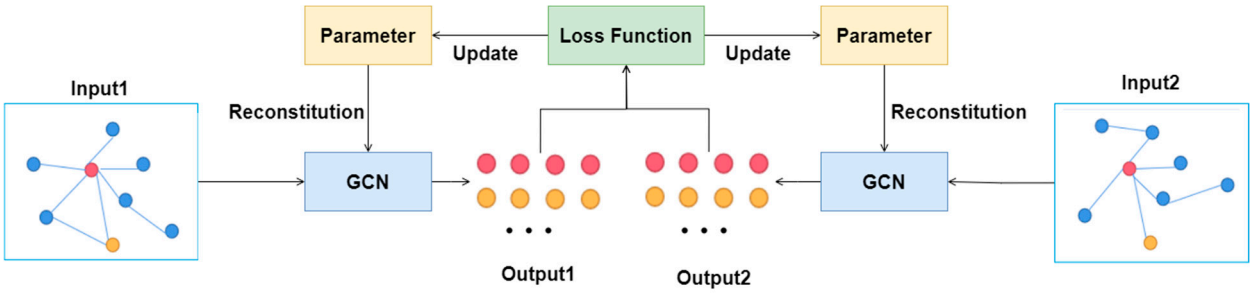


Figure 4. Bidirectional GCN multi-social-network user model.

By applying a finely trained graph convolutional network (GCN) model, this study successfully learned the vector representations of user nodes. For any given two user nodes $v_i^X \in V^X$ and $v_j^Y \in V^Y$, the distance between them is calculated using Equation (7). This research is based on a core assumption: the spatial distance between pairs of seed nodes should be as small as possible, while the spatial distance between pairs of non-seed nodes should be as large as possible. To achieve this goal, the model adopts an optimization strategy, the specifics of which are described in Equation (8). Here, C^{-1} represents the set of negative examples, γ is a parameter used to adjust the distance between positive and negative examples, and $[x]_+ = \max(0, x)$ defines a non-negative activation function. For the objective function proposed, this study employs the stochastic gradient descent (SGD) method for optimization, with the gradient update rule detailed in Equation (9).

$$d(v_i, v_j) = ||v_i - v_j|| \tag{7}$$

$$J = \sum_{(c_i, c_j \in C)} \sum_{(c_i^-, c_j^- \in C^-)} [d(c_i, c_j) + \gamma - d(c_i^-, c_j^-)]_+ \tag{8}$$

$$\theta = \theta - \eta \cdot \nabla_\theta J(\theta; x^{(i)}; y^{(i)}) \tag{9}$$

2.3. Experiments and Result Analysis

(1) Experimental datasets

The models in this chapter were tested on real-world datasets, including the Stanford Twitter Dataset, LiveJournal–MySpace, Facebook Social Circles Data, and LinkedIn API, validating the effectiveness of the model. Detailed information about these four social networks is presented in Table 1.

Table 1. Information about the dataset used in the experiment.

Social Networks	Number of Users	Number of Relationships	Number of Seed Nodes
Twitter	4572	154,921	1703
MySpace	4716	93,725	
Facebook	4039	88,234	1097
LinkedIn	3718	63,528	

(2) Evaluation criteria and experiments

In this project, *mostSims@N* was selected as the primary evaluation metric for assessing model performance, as defined in Equation (10). Here, *RightCount@N* represents the

number of correct matches among the top N most similar users for each user in the test set, with the correctness of a match indicated by the subscript. *TestAnchors* represents the total number of correct seed node pairs in the test set. An improvement in the *mostSims@N* value directly reflects better identification results. For this project's model, the parameter settings were as follows: the number of negative samples was set to five to maintain a balance between positive and negative samples, the gap parameter $\gamma = 1$, the dimension of the user node feature vector was 100, and the user closeness coefficient $\alpha = 0.9$. This parameter configuration was designed to optimize model performance and ensure maximum precision.

$$mostSims@N = \frac{RightCount@N_{XY} + RightCount@N_{YX}}{TestAnchors \times 2} \tag{10}$$

(3) Analysis of experimental results

Table 2 shows the benchmark algorithm selected in this paper.

Table 2. Benchmark algorithms.

Algorithm	Description
MAH [22]	Based on the traditional network topology, the semi-supervised learning model realizes user identification across two social networks through a small number of seed nodes, and efficiently uses a small amount of annotated data.
DLAUI [23]	Through the in-depth modeling of friendship and the introduction of manifold alignment framework, user information is mapped to low-dimensional space, which significantly reduces the computational complexity and simplifies the calculation process of user identification.
FRUI [24]	The user similarity is measured by using the information between the known nodes and the nodes to be matched, and the user identification is carried out through the node similarity, which simplifies the identification process.
COSNET [25]	Starting from the research of energy model, the energy level difference is used to distinguish different structure matching modes, and the energy model is optimized by the subgradient algorithm, and the lowest energy is the optimal matching result.
WHUI [9]	By comprehensively considering the network friendship and other information shared by users, such as interest group affiliation, etc., the similarity is calculated by using weighted supergraphs, and the threshold pruning is set to improve the accuracy and efficiency of recognition.

In experiments conducted on the Twitter–MySpace dataset, multiple evaluation metrics (*mostSims@N*) were used to measure the performance of the proposed model. As shown in Table 3, the model proposed in this study significantly outperformed several existing algorithms on all evaluated metrics, including MAH, DLAUI, FRUI, COSNET, and WHUI, especially noting that the performance of the COSNET and WHUI algorithms is close to that of the model proposed in this research. This is further illustrated in a detailed comparison in Figure 5, where the x-axis represents the top N most similar nodes selected from the test set, and the y-axis represents the probability of correctly matching nodes between two networks among these top N nodes. Figure 5 reveals an important phenomenon: as N increased from 0 to 5, there was a noticeable improvement in the model's identification performance. However, as N continued to increase, the rate of performance improvement gradually diminished. This phenomenon mainly resulted from sorting the similarity of seed nodes awaiting matching, with seed nodes ranked higher being more easily matched correctly. This analysis not only highlights the effectiveness of the model in processing social network structural information, but also provides potential directions for further optimizing model performance in future research.

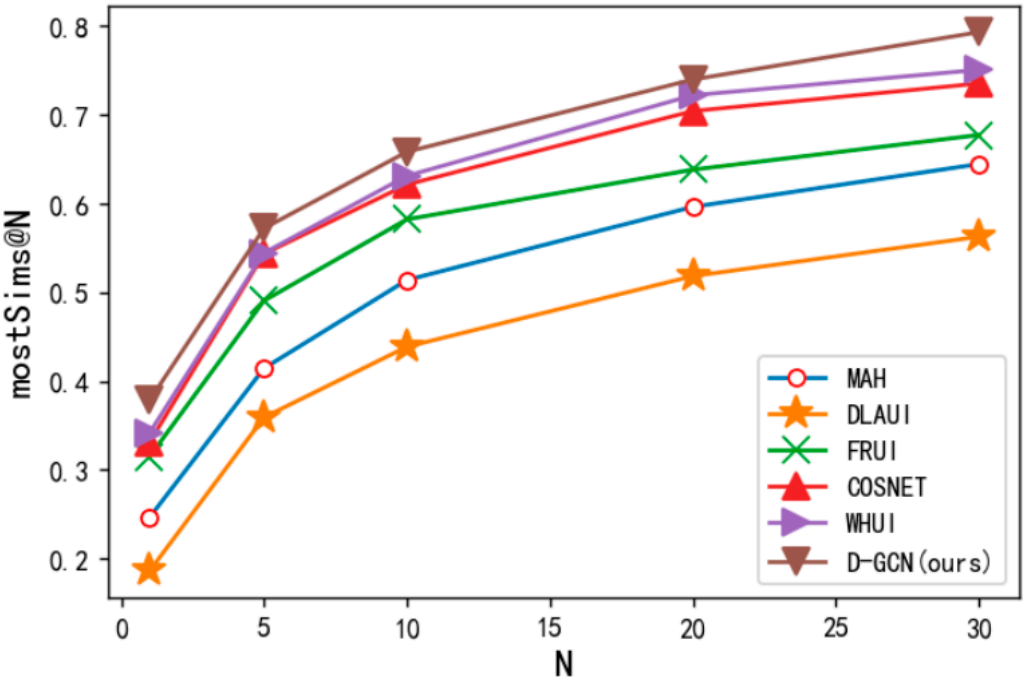


Figure 5. Comparison of the optimal *mostSims@N* of each algorithm.

Table 3. Different recognition effects are set for each algorithm on the Twitter–MySpace dataset.

Algorithm	*mostSims*@1	*mostSims*@5	*mostSims*@10	*mostSims*@20	*mostSims*@30
MAH	0.2468	0.4145	0.5139	0.5962	0.6444
DLAUI	0.1865	0.3593	0.4388	0.5183	0.5626
FRUI	0.3152	0.4906	0.5822	0.6381	0.6771
COSNET	0.3314	0.5433	0.6213	0.7041	0.7351
WHUI	0.3419	0.5444	0.6311	0.7219	0.7505
D-GCN(ours)	0.3787	0.5721	0.6582	0.7396	0.7929

This study demonstrates that our model can achieve significant identification performance solely through the structure of social networks. This effectiveness is primarily attributed to the differences between our model and algorithms such as N-GCN, MAH, and DLAUI. The latter rely solely on the superficial information of friendships, without delving into the underlying aspects of these relationships, such as whether users follow the same influencers or share the same idol preferences, leading to their limited identification accuracy. In contrast, the FRUI algorithm shows advancement in calculating node similarity by considering successfully identified pairs of seed nodes. Its advantages are mainly in two aspects: first, the FRUI model believes that the joint representation of a node and its neighboring nodes can provide more comprehensive information for user identification; second, the model can calculate the network similarity between nodes in different networks, effectively reducing the impact of different network structures. However, the FRUI algorithm does not fully consider the heterogeneity between nodes, making its identification accuracy inferior to the WHUI algorithm. The WHUI algorithm, by introducing a hypergraph structure that considers additional information beyond friendships, more accurately reflects the actual situation. However, due to its fixed weight settings, which require manual configuration and lack the ability to adaptively adjust according to actual conditions, its identification performance is somewhat lacking compared to the N-GCN algorithm.

In this paper, the N-GCN algorithm, catering to the needs of real-world application scenarios, not only considers friendships, but also delves into the intimacy between friends by reconstructing the adjacency matrix through the Jaccard coefficient, revealing deeper information behind friendships. Thanks to the natural advantage of GCN in processing graph data, namely representing nodes with new feature vectors after learning information from neighboring nodes, this model can integrate more factors and align more closely with actual situations, thus improving the accuracy of the experiments. As shown in Table 3, with the increase in N in the evaluation metric *mostSims*@N, the identification performance of each model has been improved.

After extensive benchmark model comparisons, this study further evaluated the performance of the $D-GCN$ model and its various versions on the Facebook–LinkedIn dataset, specifically including $D-GCN_{nor}$, $D-GCN_{in_Jaccard}$, $D-GCN_{out_Jaccard}$, and $D-GCN$. These versions represent the use of traditional adjacency matrices, out-degree neighbor Jaccard coefficient matrices, in-degree neighbor Jaccard coefficient matrices, and the adjacency matrix construction method proposed in this research, respectively. The research results (see Table 4) show that the closeness between users significantly affected the performance of cross-social-network user identification. Specifically, the D-GCN model proposed in this study significantly improved the accuracy of user identification by effectively utilizing the close relationships between users to construct and learn adjacency matrices.

Table 4. Performance statistics of the Facebook–LinkedIn dataset for each variant.

Algorithm	*mostSims*@1	*mostSims*@15	*mostSims*@30
$D-GCN_{nor}$	0.3136	0.6331	0.7692
$D-GCN_{in_Jaccard}$	0.3432	0.6627	0.7751
$D-GCN_{out_Jaccard}$	0.3609	0.6746	0.7870
$D-GCN$	0.3719	0.6937	0.7916

To explore the relationship between model performance and training sample size, this study conducted a comparative analysis of the *mostSims*@10 metric of each algorithm under different training set sizes. By gradually increasing the proportion of the training set, the study meticulously examined the precision performance of different models. The specific experimental results are displayed in Figure 6, where the brown downward triangle curve represents the $D-GCN$ model, the purple rightward triangle curve represents the WHUI model, the red upward triangle curve represents the COSNET model, the green star-shaped line represents the FRUI model, the yellow star-shaped line represents the DLAUI model, and the white circular curve represents the MAH model. The x-axis represents the proportion of the training set, and the y-axis represents the *mostSims*@10 metric at different training set proportions. In this experiment, the training set proportion of the Facebook–LinkedIn dataset was gradually increased from 10% to 80%, while keeping the testing set proportion at 10%. The experimental results indicate that, regardless of the size of the training set, as long as friendship relations were considered, our model significantly outperformed other models. This advantage stems from our model not only incorporating additional information, especially in the reconstruction of adjacency matrices and application of the GCN model, but also presenting, compared to other models, weight adjustments that are automatically optimized based on outcomes, thereby achieving higher identification accuracy.

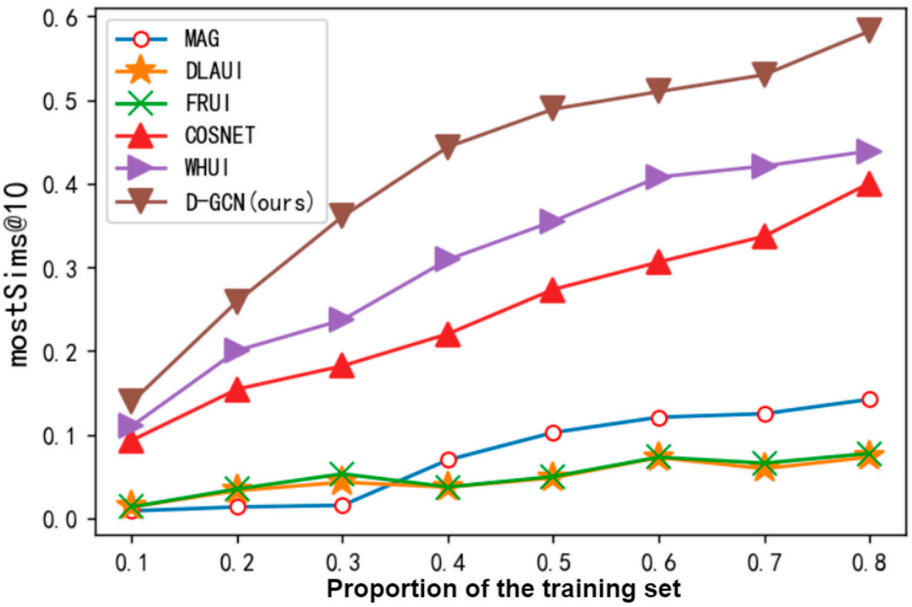


Figure 6. Comparison of the optimal *mostSims@10* at different training set ratios.

3. Multi-Social-Network User Identification Model Based on MNF-UI

With the rapid growth of information diversity, various social networking media such as Weibo, Kuaishou, and Zhihu have emerged. Currently, users commonly participate in multiple different types of social network platforms, revealing two pieces of information: on the one hand, each social network can map out the unique topological structure of the relationships formed by its users; on the other hand, there are the same users across different social networks, thereby having the potential to interconnect various networks through user relationships. Although multiple related networks may share the same user group, they are often isolated in different social network platforms without clear interconnecting links. Facing this challenge, our goal is to utilize multi-network fusion technology to explore and reveal the unknown links between these different social networks, thereby achieving the identification work of the same user across social networks.

3.1. Problem Definition and Model Framework

(1) Problem definition

In the following definitions, we assume that the edges are unweighted and that all edges are directed, as an undirected edge can be converted into two directed edges.

Definition 3. *(Multi-network matching): define a set of networks* $G = ((G^1, G^2, \cdots, G^N), (A^{(1,2)}, A^{(1,3)}, \cdots, A^{(N-1,N)}))$ *as matching networks, where* $G_i, i \in \{1, 2, \cdots, N\}$ *represents the i-th network in G, N is the number of related networks.* $A^{(i,j)}, i, j \in \{1, 2, \cdots, N\}$ *represents the set of seed linkages between* G^i *and* G^j*. We define each network as* $G^i = (V^i, E^i)$*, where* V^i *is the set of nodes, and* E^i *is the set of edges.*

Definition 4. *(Seed linkage): for two networks, we define a seed linkage* $(v_k^i, v_k^j) \in A^{(i,j)}$*, where* v_k^i, v_k^j *are the seed nodes in the two social networks, respectively. If* $(v_k^i, v_k^j) \in A^{(i,j)}$ *and* $(v_k^i, v_k^h) \in A^{(i,h)}$*, then* $(v_k^j, v_k^h) \in A^{(j,h)}$*. For ease of notation, nodes with the same subscript in different networks are referred to as known seed nodes.*

Definition 5. *(Multi-network matching problem): given a set of partially known seed linkages between multiple social networks, the problem is to discover unknown or potential seed linkages. It is noteworthy that the social networks are partially matched, meaning that not all nodes have corresponding nodes in other social networks, and the seed nodes follow a one-to-one matching constraint.*

(2) Model framework

As shown in Figure 7, the overall structural framework of the MNF-UI model mainly consists of three core components: network data representation, multi-network embedding, and node matching tasks. First, for the network data representation part, the MNF-UI model views social networks as a topological structure graph, where nodes symbolize users on social networks and edges represent social relationships between users, such as friendship links on Twitter or follow relationships on Weibo. In this context, different accounts of the same real-world individual are defined as seed node pairs. Specifically, identified seed node pairs are collectively referred to as the seed node pair set. Subsequently, the design of the multi-network embedding part aims to map two different social networks into a unified, low-dimensional vector space through the seed node pair set. In this space, since the seed node pairs represent already matched node pairs, their relative distances in the vector space will be closer, while those unmatched node pairs will be further apart. Our goal was to minimize the distance between seed node pairs as much as possible in this embedding space, while maximizing the distance between non-seed node pairs. By stripping away the interference of the network environment, we can obtain the internet vector, which can effectively be used for node matching like seed nodes. Finally, the node matching part calculates the relative distance between two nodes using cosine similarity and identifies seed node pairs by setting a certain threshold. Those seed node distances greater than the threshold will be ranked and selected to find the matching seed node pairs with the smallest distance, while those below the threshold are considered unsuccessfully matched seed nodes and are excluded.

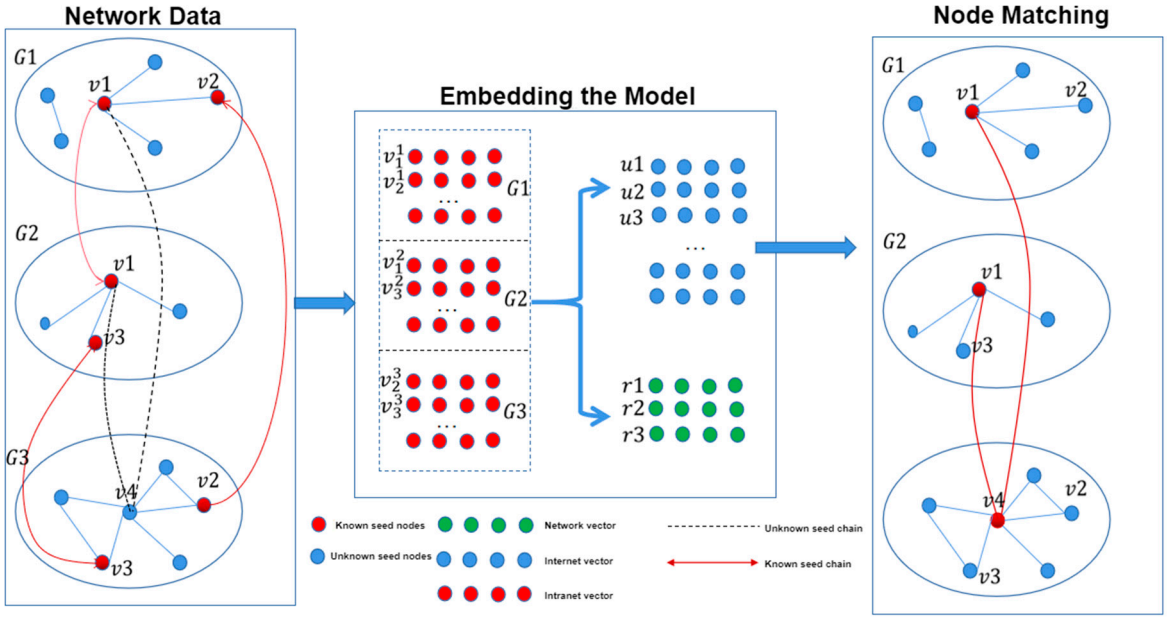


Figure 7. Overall framework of the MNF-UI model.

3.2. Model Building

In different network structures, seed nodes exhibit varying patterns of interaction behavior. As shown in Figure 7, seed node v_1 presents different connection patterns in networks G^1 and G^2. As the embodiment of the same entity in the real world, seed nodes maintain some common characteristics across networks. For example, if it is known that node v^4 in network G^3 corresponds to node v^1 in networks G^1 and G^2, we observe that seed node v^1 tends to interact with nodes v^2 or v^3 within each network. Based on this, we infer that, in selected networks, seed nodes can not only display structural characteristics similar to their corresponding nodes, but also exhibit different connection patterns due to different network environments. These common characteristics between seed nodes are crucial for user identification. Therefore, this study introduces the internet vector to capture the

common characteristics between seed nodes. Through the training process, we expect an unknown seed node's internet vector to approximate the vectors of its corresponding nodes across all network environments. However, due to the lack of direct connections between unknown seed nodes, training to obtain the corresponding vectors poses a significant challenge, such as the lack of a direct connection between node v_1 in v_4 and node v_4 in G^3. Thus, it is necessary to learn the internet vector indirectly. Through network fusion methods, it is intuitive to extract the structural characteristics of nodes, referred to as the intranet vector in our MNF-UI model. Due to differences in network environments, this vector simultaneously contains the commonality between seed nodes and the specific connection characteristics within their networks. Only by eliminating the influence of network environments can we apply the internet vector to the problem of multi-social network user identification. For this purpose, we constructed Formula (11) to establish the relationship among the intranet vector, internet vector, and network vector:

$$v_i^k = u_i + r^k \tag{11}$$

In network G^k, the intranet vector (v_i^k) of node v_i is easy to learn, while the internet vector (u_i) of node v_i shares this vector with its known corresponding seed nodes; r^k represents the network vector formed by extracting the structural characteristics of G^k, reflecting the global differences between networks. Therefore, by training based on the composite intranet vector, we can indirectly learn the internet vector of seed nodes. Taking Figure 7 as an example for further elaboration, by observing seed node v_1, one can find commonalities between the two networks, which can be represented through the internet vector. Meanwhile, due to G^1 and G^2 being different networks, v_1's local links also differ due to network structural differences. Based on Formula (11), we can define the intranet vector for node v_1 in G^1 and G^2, as shown in Formula (12).

$$\begin{aligned} v_1^1 &= u_1 + r^1 \\ v_1^2 &= u_1 + r^2 \end{aligned} \tag{12}$$

Firstly, by jointly training the intranet vector, the shared u_1 can store complementary information between two networks. For instance, v_1 might establish connections with both v_2 and v_3. Secondly, considering the information conveyed in u_1 is consistent, and r^1 and r^2 can reveal the global differences between the two networks, the intranet vector of node v_4 in network G^3 can be expressed as Formula (13). By learning the network structure information in G^3, v_4^3 might include features similar to v_1^1 or v_1^2, for example, v_4 might also tend to interact with nodes v_2 and v_3. Since there are other known seed linkages between G^3 and other networks, r^3 can reflect the specific structural characteristics of G^3. Therefore, by eliminating the influence of network structural differences, $u_4 = v_4^3 - r^3$, the internet vector u_4 can exhibit features more similar to u_1, thereby inferring that v_4 and v_1 form a pair of seed nodes. On this basis, this paper introduces the transformation matrix W, to align u across different dimensions. Equation (11) is thus rewritten as Formula (14), where U^{d_1} and R^{d_2} represent different vector spaces of dimensions d_1 and d_2, respectively. W, u_i, and r^k are parameters to be learned in the MNF-UI model. From the above discussion, theoretically, the more networks involved, the higher the accuracy of node matching. Therefore, we propose a total objective function for joint training across all networks, as shown in Formula (15), where J^k is the objective function for each network G^k, capable of preserving the structural information of each node in G^k. For every directed edge (v_i^k, v_j^k) in network G^k, we define the conditional probability of generating v_j^k from v_i^k as Formula (16). The objective function for each network G^k can be defined as Formula (17).

$$v_4^3 = u_4 + r^3 \tag{13}$$

$$v_i^k = Wu_i + r^k, u_i \in U^{d_1}, v_i^k \in R^{d_2} \tag{14}$$

$$J = \sum_{k} J^k \tag{15}$$

$$p(v_j^k \Big| v_i^k) = \frac{\exp((Wu_i + r^k) \cdot (Wu_j + r^k))}{\sum_{v_z^k \in v^k} \exp((Wu_i + r^k) \cdot (Wu_z + r^k))} \tag{16}$$

$$J^k = \sum_{(v_j^k, v_i^k) \in E^k} p(v_j^k \Big| v_i^k) \tag{17}$$

During the training process, the known seed nodes will play their utmost role, passing structural information through the shared internet vector u, ensuring this vector captures the commonality between seed nodes. Meanwhile, with Formula (16), the intranet vector v of each seed node can maintain its local structural features in different networks, thus allowing the network vector r^k to extract and identify the global structural features of the network itself. For each unknown seed node v_i^k, although we cannot establish a direct connection with the corresponding network, the composite vector, i.e., the intranet vector, can identify some similar features in other networks. By eliminating the influence of network structure r^k, the internet vector will exhibit features similar to those of seed nodes, something which is highly effective for node matching.

After training, we will be able to obtain the internet vector for each node, representing the common information between different seed nodes. In the problem of multi-social network user identification, a basic approach is to calculate the similarity between nodes. We use cosine similarity to measure the distance between two nodes and set a threshold. Node pairs with a similarity above this threshold are sorted by similarity, and the pair with the highest similarity is selected as the successfully identified seed node pair. Nodes below the threshold are discarded, indicating that the corresponding seed node pair was not identified in this match.

3.3. Experiments and Analysis of Results

We conducted a series of experiments aimed at validating the effectiveness of our proposed multi-network fusion for user identification (MNF-UI) model. By comparing our model with several leading methods in the field of multi-social-network user identification, we demonstrated the advantages and application potential of our approach.

(1) Experimental datasets

To validate the effectiveness of the MNF-UI model, this study selected widely recognized public datasets in the field of multi-social-network user identification for experimentation. Specifically, the first dataset chosen was based on the data from the 2013 World Athletics Championships which divided the Twitter dataset into three independent networks: Twitter-RT, Twitter-MT, and Twitter-RE through behaviors such as retweeting, mentioning, and replying. This dataset contains 88,804 nodes, 210,250 edges, and 55,362 seed links. The second dataset encompasses four real social network platforms: Twitter, MySpace, Facebook, and LinkedIn. Detailed information about this dataset is provided in Section 2.3 and its summary information is also presented in Table 5. The selection of these two datasets aimed to comprehensively assess the MNF-UI model's ability to identify users in different social network environments and the model's adaptability and effectiveness in processing different types of network structural information.

Table 5. Twitter–MySpace–Facebook–LinkedIn dataset information.

Social Networks	Number of Users	Number of Relationships	Number of Seed Nodes
Twitter	4572	154,921	653
MySpace	4716	93,725	
Facebook	4039	88,234	
LinkedIn	3718	63,528	

(2) Evaluation indexes and experimental settings

To validate the effectiveness of our proposed method in solving the problem of multi-social-network user identification, this study adopted an approach that solely utilizes network structural information and compared it with four different baseline methods, as detailed in Table 6.

Table 6. Four baseline methods.

Name of Method	Algorithm Description	Type
NetAlign [26]	The messaging algorithm is used to identify social network users in unsupervised mode and explore the potential correspondence between network structures.	Unsupervised
REGAL [27]	The similarity-based graph vector representation is extracted, and the same user across multiple social networks is identified by comparing these learned embedding vectors.	Unsupervised
FINAL [28]	From the perspective of optimization, according to the principle of network consistency, supplemented by attribute information, the same user across multiple social networks is accurately identified through a series of attribute recognition algorithms.	Unsupervised
IONE [29]	Node vectors, including the input vectors and output vectors of nodes, are used to jointly define the representation of user nodes. This representation strategy effectively maintains the mapping of the relationship between the concerned and the followed between users in the network in the embedded space.	Semi-oversight

To further explore the performance differences between the MNF-UI model and existing methods in the field of multi-social network user identification, this study conducted experimental validations using two different datasets. The first dataset was based on the 2013 World Athletics Championships and divided Twitter data into three independent networks through actions such as retweeting, mentioning, and replying. The second dataset spanned Twitter, MySpace, Facebook, and LinkedIn, providing a diversified social network environment. To ensure the fairness of the experiments, all embedding-based methods adopted the same node dimension setting, i.e., $d = 100$. In the MNF-UI model, we specifically set $d_1 = 100$ and $d_2 = 100$ and, to accelerate the training process, the number of negative samples was set to 1. In the multi-social-network user identification task, we utilized the internet vector as a key tool. For experimental design, we randomly selected 50% of the seed links as the test set, with the remainder used as the training set. Moreover, we compared the identification accuracy of the MNF-UI model with several other models.

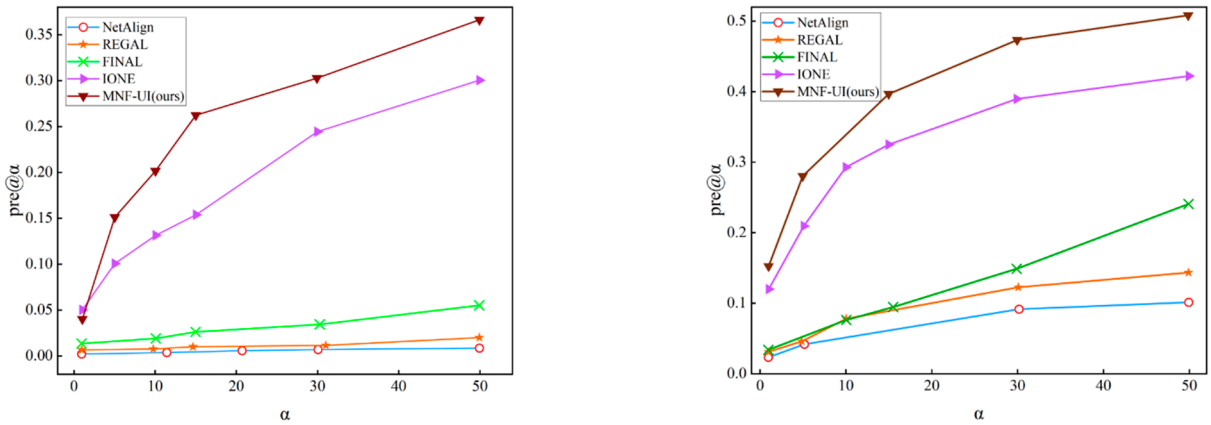
In this study, we adopted the widely recognized evaluation standard in the field of multi-social-network user identification, i.e., $pre@\alpha$, as the key indicator to assess the performance of our model. The design goal of $pre@\alpha$ is to quantify the accuracy of seed node identification between two networks G^i and G^j. The calculation formula for this evaluation metric is as described in Formula (18), where $UnMappedAnchors^{(i,j)}$ represents the number of seed node links that were not successfully matched, including both correct and incorrect matches, $RightNodes^{(i,j)}@\alpha$ refers to the number of seed node links that were correctly matched. Further, the specific expression of this evaluation metric can be referred to in Formula (19), where N represents the total number of networks:

$$pre^{(i,j)}@\alpha = \frac{\left|\text{Right}Nodes^{(i,j)}@\alpha\right|}{\left|UnMappedAnchors^{(i,j)}\right|} \tag{18}$$

$$pre@\alpha = \frac{1}{N(N-1)}\sum_{i}\sum_{j\neq i} pre^{(i,j)}@\alpha \tag{19}$$

(3) Analysis of experimental results

First, we set the proportion of the training set to 50% and adjusted the value of parameter α. Based on this, we conducted a comprehensive comparison of several algorithms on two different datasets. As shown in Figure 8, the x-axis represents the number of top α similar nodes selected from the test set, while the y-axis reflects the probability of correctly matching nodes between different networks among the top α nodes. By gradually increasing the value of α in intervals of 10, it can be clearly observed in Figure 8a,b that both MNF-UI and IONE demonstrated superior performance in relation to the two datasets. In contrast, unsupervised learning methods such as NetAlign, REGAL, and FINAL performed poorly relative to all datasets examined. The fundamental reason for this phenomenon is that the structure of each social network does not strictly adhere to the topological structure consistency principle assumed by these algorithms, leading to their identification effects being far inferior than that of the model we proposed.



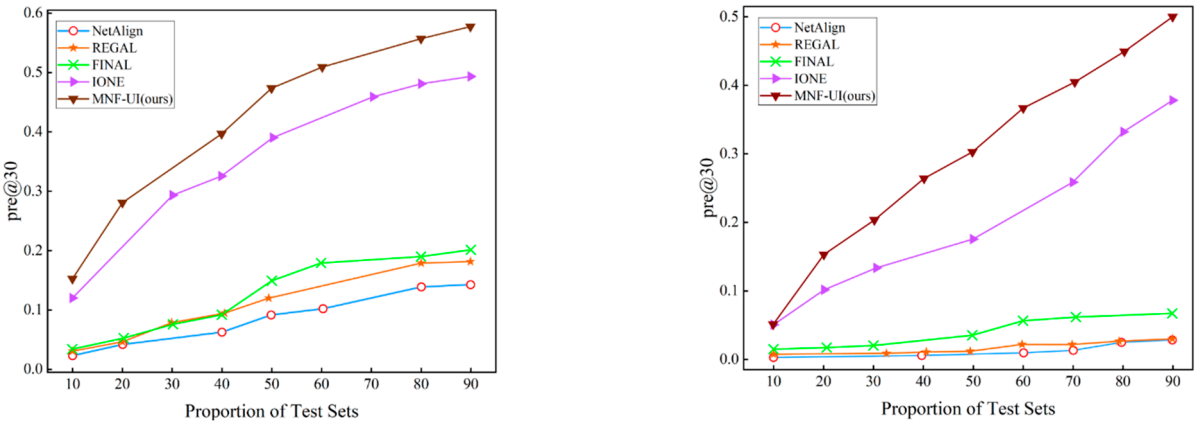
(**a**) The first dataset (**b**) The second dataset

Figure 8. Change the experimental results of α.

From the analysis of Figure 8a,b, it is evident that the model proposed in this study significantly outperformed the other benchmark models with respect to the vast majority of training sets. Notably, even under conditions of extremely limited training data, our model still demonstrated certain effectiveness. This experimental result proves that our method can still function even in situations where seed link information is scarce, something which is very common in practical application scenarios. Although the network embedding-based IONE method can achieve good identification results on various datasets, it only considers pairwise learning of two networks, thus ignoring the complementary information between many networks. At the same time, this method does not fully consider the characteristics of the network structure, leading to ineffective matching of seed nodes where structural local differences exist. Therefore, the recognition effect of the IONE method is limited in scenarios with small structural differences between networks. For example, in the Twitter dataset, because the dataset is divided into retweet, mention, and reply sub-networks based on the same event, these three networks show a similar global structure, and the distribution of seed nodes is relatively uniform, thus IONE performed well on this dataset. Especially when the training set ratio was 50% and $\alpha = 1$, the recognition effect of IONE was better than our model. However, as the training set ratio increased, our model gradually demonstrated a recognition performance that surpassed that of other models by integrating the collective learning ability of all networks.

Setting the parameter α to 30 and adjusting the proportion of the training set, we compared the performance of various algorithms using two different datasets. The results are shown in Figure 9. The x-axis represents the selected test set proportion, while the y-axis indicates the probability of successfully matching the top 30 nodes with the target node in the test set. The proportion of the test set starts at 0.1 and increases in increments

of 0.1 to 0.9 to evaluate the two datasets. The results show that MNF-UI and IONE performed well under different test set proportions, while methods such as NetAlign, REGAL, and FINAL exhibited a relatively mediocre performance across all test proportions. Part of the reason for this phenomenon is that these social network structures do not follow the assumed topological structure consistency, leading to significant local structural differences of the same users in different networks. Therefore, even when the training set proportion reached 0.9, these methods still failed to achieve significant effects. Although IONE performed better than the other three methods under different test proportions, since it only involves pairwise learning among multiple networks without fully exploiting the potential complementary information among multiple networks, its effectiveness was still inferior to that of our proposed model.



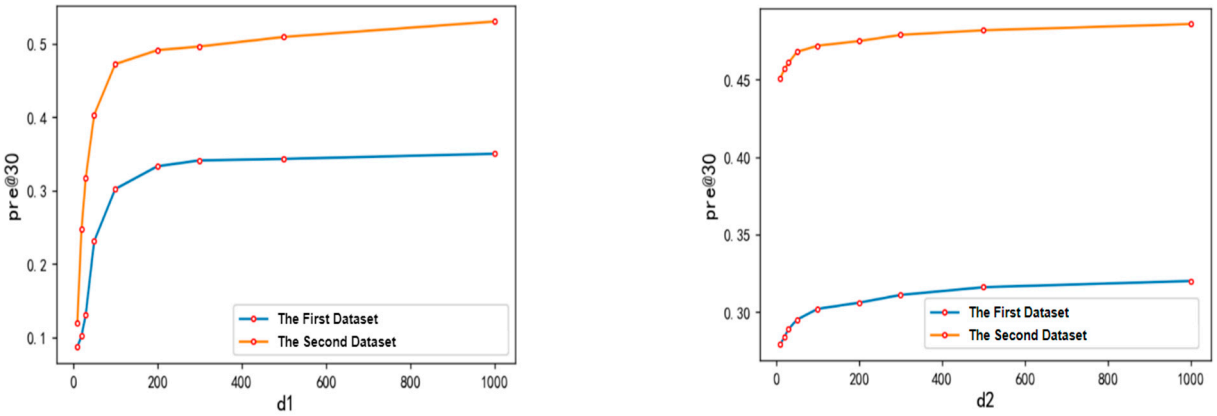
(**a**) The first dataset (**b**) The second dataset

Figure 9. Experimental results adjusted for the scale of the test set.

On the Twitter–MySpace–Facebook–LinkedIn dataset, the IONE method did not achieve better recognition results due to significant structural differences among these four social networks. IONE is limited to pairwise learning of matching information between two networks and lacks the ability to handle structural differences among multiple networks. In contrast, our proposed model, through joint learning of four networks, can not only effectively integrate cross-network information, but also mitigate the impact of network structural differences to a certain extent. This characteristic significantly enhances the model's efficiency and accuracy in user identification across multi-network environments, thereby achieving superior recognition results relative to this dataset.

To explore the specific impact of dimension parameters on experimental effects, we conducted detailed experimental research on the dimension parameters d_1 and d_2 for two key vectors—internet vector and intranet vector. Initially, we set both d1 and d2 to 100 as the baseline parameters. During the experiments, we tested the impact of parameter changes on the experimental results by fixing one parameter while changing the other. The experiments were conducted on the Twitter–MySpace–Facebook–LinkedIn dataset, with the training set proportion set to 50% and parameter α set to 30. According to the experimental results (as shown in Figure 10a,b), the following phenomena can be observed: as the dimension of d_1 increased, the model's performance significantly improved. This indicates that increasing the dimension of the internet vector effectively enhanced the model's ability to capture network structural information, thereby improving the accuracy of user identification. However, when d_2 exceeded a certain threshold, performance tended to stabilize, and this threshold appears to be related to the scale of the network. Specifically, on the Twitter–MySpace–Facebook–LinkedIn dataset, performance gains tended to plateau when d_1 exceeded 200. Meanwhile, the impact of the d_2 dimension on model performance was relatively limited. This finding suggests that adopting a smaller value strategy for d_2 can not only save computational resources, but also optimize model performance to some extent, under the premise of ensuring model efficiency. Therefore, selecting appropriate

values for d_1 and d_2 is of great significance for resource optimization and performance enhancement. For instance, setting d_1 to 300 and d_2 to 500 can ensure model performance while being more resource-efficient.



(**a**) experimental result d1 (**b**) experimental result d2

Figure 10. Effect of dimensions on the recognition effect of MNF-UI model.

4. Conclusions

With the rapid development of internet technology, various social networking platforms with different functionalities have emerged one after the other. While people enjoy the convenience brought by these platforms, it also exerts certain impacts on network security. Therefore, the problem of social network user identification holds significant research importance. It can integrate user information across multiple social networks, which is crucial for user recommendation, network security, and public opinion supervision. This paper first introduces the background and significance of multi-social-network user identification, analyzes the current research status, clarifies the challenges faced by multi-social-network user identification, and points out the problems existing in current research methods and the directions for improvement. To address the issues in multi-social-network user identification, this paper proposes two methods. The main research contents are as follows:

1. This paper applies graph convolutional networks to the task of cross-social-network user identification and proposes a cross-social-network user identification model based on bidirectional GCN. First, we propose a user intimacy measurement method based on the Jaccard coefficient which reflects the degree of intimacy between social network users in the adjacency matrix. Secondly, we reconstruct the adjacency matrix of social networks. Finally, we randomly initialize user features to address the problem of unobtainable user attributes. Experimental results show that our model significantly outperforms the baseline models.

2. In response to the current methods' over-reliance on the assumption of network topological structure consistency and the inability to identify users across multiple social networks, we propose a multi-social network user identification model based on MNF-UI. Our model defines a network feature vector representing the structural features of each social network which can reflect the global structural differences between social networks. It represents each seed node's vector as the sum of the cross-network common features inherent in the seed node and the network features of the network where the seed node is located. Then, it uses the cross-network common features inherent in the seed node for user identification, successfully solving the problems existing in current methods. Moreover, our model achieved better recognition effects on two datasets compared to baseline models.

In summary, based on existing user identification methods, this paper innovatively combines the advantages of graph convolutional networks, network embedding, and data mining to conduct a deeper study on the algorithm for multi-social network user identification. Finally, it proposes two algorithms that can be used for multi-social-network

user identification, providing technical and theoretical support for in-depth research on social network information fusion and network security maintenance.

Author Contributions: Software, C.L.; Validation, F.X.; Writing—original draft, S.H.; Writing—review & editing, H.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Van Dülmen, C.; Klärner, A. Places that bond and bind: On the interplay of space, places, and social networks. *Soc. Incl.* **2022**, *10*, 248–261. [CrossRef]
2. Xing, L.; Deng, K.; Wu, H.; Xie, P.; Zhao, H.V.; Gao, F. A survey of across social networks user identification. *IEEE Access* **2019**, *7*, 137472–137488. [CrossRef]
3. Zahra, K.; Azam, F.; Butt, W.H.; Ilyas, F. User identification on social networks through text mining techniques: A systematic literature review. In *Information Science and Applications 2018*; Springer Nature: Singapore, 2019; pp. 485–498.
4. Zeng, W.; Tang, R.; Wang, H.; Chen, X.; Wang, W. User identification based on integrating multiple user information across online social networks. *Secur. Commun. Netw.* **2021**, *2021*, 5533417. [CrossRef]
5. Çoban, Ö.; Ali, İ.; Ozels, A. Your username can give you away: Matching turkish OSN users with usernames. *Int. J. Inf. Secur. Sci.* **2021**, *10*, 1–15.
6. Li, Y.; Peng, Y.; Zhang, Z.; Yin, H.; Xu, Q. Matching user accounts across social networks based on username and display name. *World Wide Web* **2019**, *22*, 1075–1097. [CrossRef]
7. Solanki, P.; Harwood, A. User identification across social networking sites using user profiles and posting patterns. In Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 18–22 July 2021; IEEE: New York, NY, USA, 2021; pp. 1–8.
8. Qu, Y.; Yu, S.; Zhou, W.; Niu, J. FBI: Friendship learning-based user identification in multiple social networks. In Proceedings of the 2018 IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, United Arab Emirates, 9–13 December 2018; IEEE: New York, NY, USA, 2018; pp. 1–6.
9. Kojima, K.; Ikeda, K.; Masahiro, T. Short Paper: User Identification across Online Social Networks Based on Similarities among Distributions of Friends' Locations. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; IEEE: New York, NY, USA, 2019; pp. 4085–4088.
10. Ahmad, W.; Ali, R. User identification across multiple online social networks using cross link attribute and network relationship. *J. Interdiscip. Math.* **2020**, *23*, 205–214. [CrossRef]
11. Qu, Y.; Xing, L.; Ma, H.; Wu, H.; Zhang, K.; Deng, K. Exploiting user friendship networks for user identification across social networks. *Symmetry* **2022**, *14*, 110. [CrossRef]
12. Xing, L.; Deng, K.; Wu, H.; Xie, P.; Gao, J. Behavioral habits-based user identification across social networks. *Symmetry* **2019**, *11*, 1134. [CrossRef]
13. Lei, T.; Ji, L.; Liu, S. Investigation of Cross-Social Network User Identification. In Proceedings of the 2021 International Conference on Advanced Computing and Endogenous Security, Nanjing, China, 21–22 April 2022; IEEE: New York, NY, USA, 2022; pp. 1–7.
14. Ke, Q.; Jing, X.; Woźniak, M.; Xu, S.; Liang, Y.; Zheng, J. APGVAE: Adaptive disentangled representation learning with the graph-based structure information. *Inf. Sci.* **2024**, *657*, 119903. [CrossRef]
15. Dong, W.; Wu, J.; Zhang, X.; Bai, Z.; Wang, P.; Woźniak, M. Improving performance and efficiency of Graph Neural Networks by injective aggregation. *Knowl. Based Syst.* **2022**, *254*, 109616. [CrossRef]
16. Lavanya, R.; Saksena, A.; Singh, A. Effective networking on social media platforms for building connections and expanding E-commerce business by analyzing social networks and user's nature and reliability. In *Artificial Intelligence Techniques for Advanced Computing Applications: Proceedings of ICACT 2020*; Springer: Singapore, 2021; pp. 503–514.
17. Yang, D.; Qu, B.; Yang, J.; Cudré-Mauroux, P. Lbsn2vec++: Heterogeneous hypergraph embedding for location-based social networks. *IEEE Trans. Knowl. Data Eng.* **2020**, *34*, 1843–1855. [CrossRef]
18. Pei, H.; Wei, B.; Chang KC, C.; Lei, Y.; Yang, B. Geom-gcn: Geometric graph convolutional networks. *arXiv* **2020**, arXiv:2002.05287.
19. Lei, T.; Ji, L.; Wang, G.; Liu, S.; Wu, L.; Pan, F. Transformer-Based User Alignment Model across Social Networks. *Electronics* **2023**, *12*, 1686. [CrossRef]
20. Ma, T.; Guo, L.; Wang, X.; Qian, Y.; Tian, Y.; Al-Nabhan, N. Friend closeness based user matching cross social networks. *Math. Biosci. Eng.* **2021**, *18*, 4264–4292. [CrossRef] [PubMed]
21. Ding, X.; Zhang, H.; Ma, C.; Zhang, X.; Zhong, K. User identification across multiple social networks based on naive Bayes model. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *35*, 4274–4285. [CrossRef] [PubMed]
22. Qu, Y.; Ma, H.; Wu, H.; Zhang, K.; Deng, K. A Multiple Salient Features-Based User Identification across Social Media. *Entropy* **2022**, *24*, 495. [CrossRef] [PubMed]

23. Zhou, F.; Liu, L.; Zhang, K.; Trajcevski, G.; Wu, J.; Zhong, T. Deeplink: A deep learning approach for user identity linkage. In Proceedings of the IEEE INFOCOM 2018-IEEE Conference on Computer Communications, Honolulu, HI, USA, 16–19 April 2018; IEEE: New York, NY, USA, 2018; pp. 1313–1321.
24. Shu, K.; Wang, S.; Tang, J.; Zafarani, R.; Liu, H. User identity linkage across online social networks: A review. *Acm Sigkdd Explor. Newsl.* **2017**, *18*, 5–17. [CrossRef]
25. Zhang, Y.; Tang, J.; Yang, Z.; Pei, J.; Yu, P.S. Cosnet: Connecting heterogeneous social networks with local and global consistency. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, Australia, 10–13 August 2015; ACM: New York, NY, USA, 2015; pp. 1485–1494.
26. Huang, X.; Chen, D.; Ren, T.; Wang, D. A survey of community detection methods in multilayer networks. *Data Min. Knowl. Discov.* **2021**, *35*, 1–45. [CrossRef]
27. Heimann, M.; Shen, H.; Safavi, T.; Koutra, D. Regal: Representation learning-based graph alignment. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Torino, Italy, 22–26 October 2018; ACM: New York, NY, USA, 2018; pp. 117–126.
28. Bhowmick, S.; Bell, P.; Taufer, M. A Survey of Graph Comparison Methods with Applications to Nondeterminism in High-Performance Computing. *Int. J. High Perform. Comput. Appl.* **2023**, *37*, 306–327. [CrossRef]
29. Liu, L.; Cheung, W.K.; Li, X.; Liao, L. Aligning Users across Social Networks Using Network Embedding. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-16), New York, NY, USA, 9–15 July 2016; IJCAI: New York, NY, USA, 2016; pp. 1774–1780.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.