

## Research Article

# Two-Stage User Identification Based on User Topology Dynamic Community Clustering

Jiajing Zhang,<sup>1</sup> Zhenhua Yuan ,<sup>2</sup> Neng Xu,<sup>2</sup> Jinlan Chen,<sup>3</sup> and Juxiang Wang<sup>1</sup>

<sup>1</sup>School of Mathematics & Physics, Anhui Jianzhu University, Hefei 230601, China

<sup>2</sup>School of Electronics and Information, Anhui Jianzhu University, Hefei 230601, China

<sup>3</sup>School of Mechanical and Electrical Engineering, Anhui Jianzhu University, Hefei 230601, China

Correspondence should be addressed to Zhenhua Yuan; yzhua777@163.com

Received 28 February 2021; Revised 17 March 2021; Accepted 29 March 2021; Published 17 April 2021

Academic Editor: M. Irfan Uddin

Copyright © 2021 Jiajing Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to solve the problem of node information loss during user matching in the existing user identification method of fixed community across the social network based on user topological relationship, Two-Stage User Identification Based on User Topology Dynamic Community Clustering (UIUTDC) algorithm is proposed. Firstly, we perform community clustering on different social networks, calculate the similarity between different network communities, and screen out community pairs with greater similarity. Secondly, two-way marriage matching is carried out for users between pairs of communities with high similarity. Then, the dynamic community clustering was performed by resetting the different community clustering numbers. Finally, the iteration is repeated until no new matching user pairs are generated, or the set number of iterations is reached. Experiments conducted on real-world social networks Twitter-Foursquare datasets demonstrate that compared with the global user matching method and hidden label node method, the average accuracy of the proposed UIUTDC algorithm is improved by 33% and 26.8%, respectively. In the case of only user topology information, the proposed UIUTDC algorithm effectively improves the accuracy of identity recognition in practical applications.

## 1. Introduction

With the rapid development and application of artificial intelligence technology, the application scope of artificial intelligence technology is also expanding. Artificial intelligence technology represented by deep learning and breadth learning is becoming more mature than before. Among them, data integration is the premise and foundation of breadth learning. In order to achieve multisource data integration, user identification across social network has become a very valuable research hotspot.

Social networks connect users on the Internet, allowing users to communicate and interact, forming a virtual social behavior similar to reality. According to statistics, 42% of users have more than one social network account, and 93% of Instagram users use Facebook at the same time [1].

Different social network platforms have different functions, and these platforms are independent of each other. User information is scattered in different networks, and the same real user information cannot be shared between different networks. Each network forms an “island,” which makes it impossible to integrate data between networks. In order to break the phenomenon of information “islands” and achieve multisource data integration, cross-social network identification is a necessary premise and basis. Cross-social network user identification has strong research value and practical application significance in many fields such as user portraits, commercial advertising, friend recommendation, and maintenance of online public opinion security.

At present, cross-social network user identification methods mainly include methods based on user attribute information, user behavior information, and user topology

information, and the integration of three different characteristic information methods. In social networks, user topology, namely, friend relationships, is authentic and difficult to forge [2]. Therefore, this article decides to use user topology information for user identification.

## 2. Related Work

Among the methods based on user attributes, Zafarani et al. [3] first proposed the method of user name mapping to identify users. Peritio et al. [4] proposed a method to calculate the similarity between user names based on the uniqueness of user names. Liu Dong et al. [5] extracted the hidden features of usernames from multiple angles and integrated the statistical results of the probability distribution of various features to infer the identity of the corresponding usernames. Vosecky et al. [6] first proposed using vectors to represent user profile information and then calculated the similarity between the vectors. Although the accuracy of the user-based attribute method is very high, the attribute information belongs to the user's privacy and is so difficult to obtain. In addition, due to the user's high awareness of network security, the user may provide wrong content when filling in the attribute information. Therefore, the universality of user-based attribute methods is not high [7].

Among the methods based on user behavior information, Kong and Zhang et al. [8] match users by calculating the similarity of users in different networks with respect to time, space, and text information. Liu et al. [9] proposed a method to identify users by integrating information such as the content of the user's published content, writing habits, and behavior trajectory. Roedler et al. [10] established the user's own unique social behavior pattern by using the time information carried by the social network and the geographical information recorded by the device, which was used as an identification mark. The method based on user behavior information faces the problem that the user's geographic and spatial information has sparse characteristics in social networks, and it is difficult to apply to large-scale social networks.

Among the methods based on user topology information, Narayanan et al. [11] proved for the first time that only user relationships are used, relying on a small number of initial matching seed nodes and iteratively updating to continuously find new nodes, but the recognition accuracy of this method is not high. Nitish et al. [12] proposed an identification algorithm for multiple social networks based on the node degree and the number of common neighbors. Zhou et al. [13] took the number of seed nodes shared by the nodes to be matched as the cross-network similarity of the nodes and selected those with greater similarity for matching. But only simply using user topology information, when there are many nodes, the efficiency and the accuracy are not high.

Among the methods based on multidimensional information fusion, Peled et al. [14] extracted two aspects of user topology and user attributes, established a 27-dimensional feature vector, and finally judged whether the user identity

matches through the similarity of the feature vector. Liu et al. [9] used three-dimensional information to train the model in a semisupervised learning manner to complete the matching. Zhang et al. [15] also used the information of all the above three dimensions. First, the network structure information was used as users to be matched to select the set of potential matching nodes, and then the user name and spatiotemporal trajectory were used to train the classifier, which is a kind of unsupervised learning algorithm. Xing et al. [16] firstly used entropy to assign weights to user name features, then analyzed user interests, combined with the user name and user published content to identify users across social networks. These algorithms fully consider multidimensional information, so the overall performance of the algorithms is better. Although the methods based on multidimensional information fusion are effective, it is difficult to obtain comprehensive data in specific social networks. Moreover, the multidimensional information model is complex, difficult to model, inefficient, and prone to "overfitting" when the amount of data is not enough.

Although the method based on user topology information is not efficient and accurate, the user topology information is authentic and difficult to forge. Wu Zheng et al. [17] used potential relationship information to improve the recognition of nodes to be matched by clustering social networks. However, the clustering of communities in this method is fixed, resulting in the loss of information of nodes outside the community, and the efficiency and accuracy are not high.

Based on the above literature analysis, this paper proposes a dynamic community clustering two-stage user identification algorithm based on user topological relationships to solve the problem of node information loss during user matching between fixed communities. Firstly, perform community clustering on different social networks, calculate the similarity between different network communities, and filter larger similar community pairs. Secondly, match users between larger similar community pairs on different networks. Finally, add the matched node pair to the seed node user pair. Reset the number of different community clusters (decrease by a certain level), redo dynamic community clustering, and then match users in larger similar communities. Repeat iterations until no new matching user pairs are generated or reach the set number of iterations.

## 3. UIUTDC Algorithm

*3.1. Related definitions.* Identity recognition based on user relationships uses user topological structure relationships to identify the accounts of the same natural person on different community platforms. A formal description of this problem is as follows:

**Definition 1** (cross-social network user matching). There are two different social network platforms,  $G_A$  and  $G_B$ , where  $G_A = (U_A, E_A)$ ,  $G_B = (U_B, E_B)$ .  $U_A$  and  $U_B$  represent the set of all users in social networks A and B, respectively, and  $E_A$  and  $E_B$  represent the set of user topological relationships in social networks A and B,

respectively. The cross-social network user matching relationship is  $M$ , where  $M = \{(u, v) \mid u \in U\_A, v \in U\_B\}$ .  $M$  is a pair of users belonging to the same natural person in the A and B networks.

**Definition 2** (known user matching node pair). The known user matching node refers to the network user matching node that is found in advance through specific methods such as URL address information. This article uses Seed\_User to indicate known user matching node pairs.

**3.2. UIUTDC Algorithm Principle.** Since there are many users in social networks and the relationship of friends is relatively complicated, if the similarity calculation is performed on the user nodes of the network one by one, the cost of similarity calculation is very high. Except for a small number of friends (compared to the number of users in the entire social network) of a user in a social network, most other users rarely contact this user. According to the principle of clustering of things and groups of people, a user and his friends are likely to be in a cluster (community) in a social network, while in reality, a user belongs to a cluster in a different social network (community), these clusters have a large degree of similarity. Taking into account the problem of fixed community clustering and node information loss, first, the UIUTDC algorithm uses multiple rounds of dynamic community clustering method to cluster different networks. Set a different number of clustered communities in each round (decreasing according to a certain level), cluster from different angles, cover the entire network, and more fully match users through multiple iterations. After each round of community clustering, calculate the similarity of the communities in different networks, as shown in Figure 1(a), and filter the larger similar community pairs, then calculate the similarity of the user node in the community pair with higher similarity, and the node pair with higher matching similarity is the matching user pair, as shown in Figure 1(b).

**3.3. TSUIBUTDC Algorithm Framework.** The specific framework of the UIUTDC algorithm is shown in Figure 2.

Firstly, we initialize the number of community clusters in social networks A and B, perform community clustering on social networks A and B, respectively, and calculate and filter out larger similar community pairs in social networks A and B. Secondly, we select any community of A network in the larger similar community pair, make each user match with any user of B network in the larger similar community pair, and calculate and screen out the user pairs with a high similarity between A and B network communities. The user pairs with large similarities are bidirectionally matched, and the matched nodes are added to the seed node pair set. Loop iteratively until all communities in the community pair with greater similarity in A and B networks are matched. Judging whether the iteration is over, that is, whether it has reached the maximum number of iterations or whether it converges (The maximum number of iterations of the judgment condition here is obtained through experiments, and whether it converges

is when there is no new seed node generated). If it reaches the maximum number of iterations or has converged, the newly generated seed node pairs are output, and the program ends. Otherwise, we reset the number of clustering communities of A and B social networks (decrease according to a certain level), repeat the above process of community clustering, screening large similar community pairs, and matching user nodes until the maximum number of iterations is reached or no new seed node pairs are generated.

**3.3.1. Calculate and Filter Out the Community Pairs with Greater Similarity between A and B Networks.** The calculation of community similarity is based on the common a priori seed node relations in the community. The calculation is shown in formula (1), where  $c_A^i$  and  $c_B^j$  represent the  $i$ -th community in A social network and the  $j$ -th community in B social network, respectively.  $c_A^i \text{SU}$  represents the a priori seed node of the  $i$ -th community in A social network,  $c_B^j \text{SU}$  is the prior seed node of the  $j$ -th community in B social network.

$$\text{Csim}(c_A^i, c_B^j) = 2 \times \frac{|c_A^i \text{SU} \cap c_B^j \text{SU}|}{|\text{total nodes of } c_A^i| + |\text{total nodes of } c_B^j|}. \quad (1)$$

In order to store the community pairs with greater similarity, we design the Com\_pair set. Its element data structure includes the community pair sequence number attribute Com and the community pair similarity attribute Sim, where the Com attribute contains Com\_A and Com\_B. Com\_A stores the community number of the A network and Com\_B stores the community ordinal number of the B network. The Sim attribute stores the similarity between the A network community and the B network community. The structure is shown in Figure 3.

Com\_pair [m]. [(Com\_A, Com\_B)] represents the  $m$ th similar community pair in Com\_pair and Com\_pair [m]. Sim represents the similarity between the  $m$ th similar community pair.

The community pairs with greater similarity between A and B networks are calculated as follows:

**3.3.2. Calculate and Filter out User Pairs with Greater Similarity between Social Networks A and B.** After obtaining the community pairs with high similarity, the user similarity between communities with high similarity in different networks is calculated; that is, the ratio of the number of the same prior seed nodes in the neighbor nodes of two users to the total number of the neighbor nodes of two users is calculated. The specific calculation is shown in formula (2), where  $u_A^i N$  represents the set of neighbor nodes of the  $i$ th node of the community in the A network, and  $u_B^j N$  represents the set of neighbor nodes of the  $j$ th node of the community in the B network. NCSU represents the number of common seed node pairs in neighbor nodes.

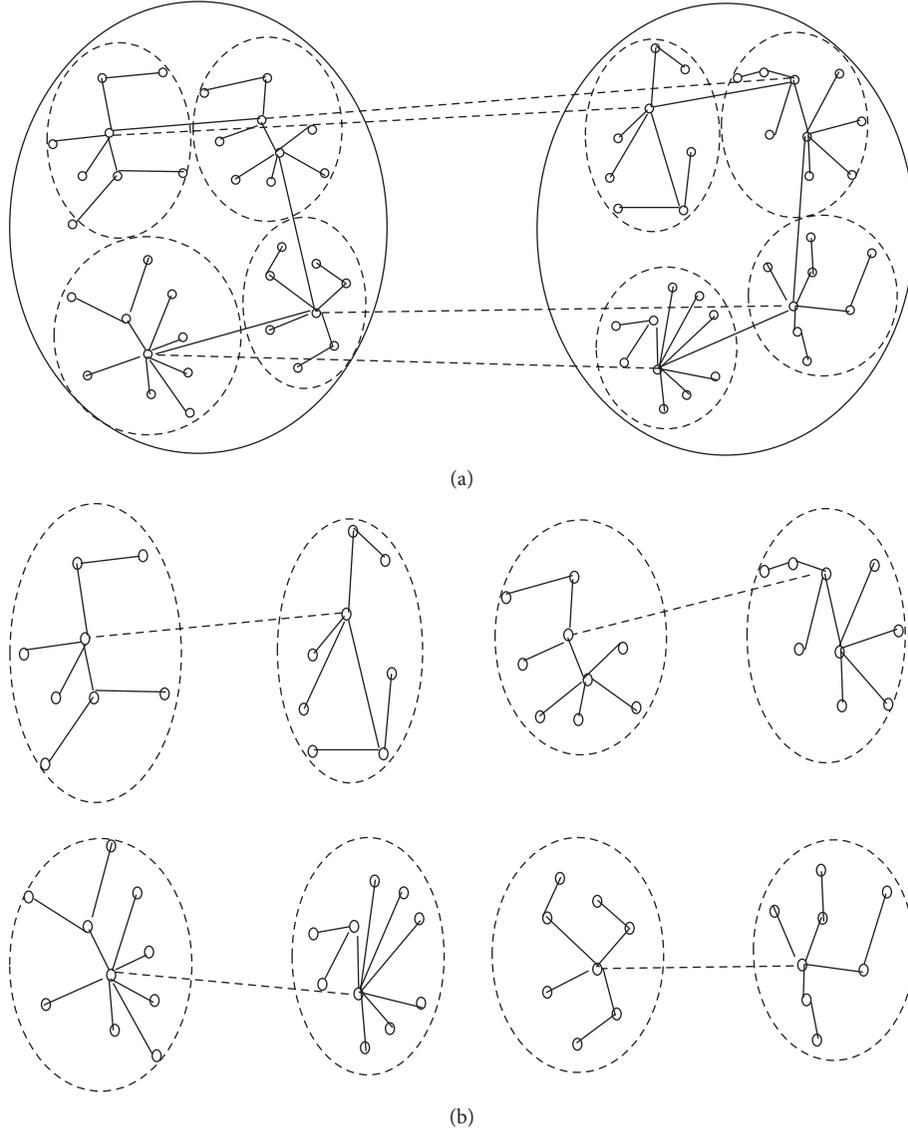


FIGURE 1: (a): Cluster the network and identify similar communities. (b): User identification in similar communities.

$$U\text{sim}(u_A^i, u_B^j) = 2 \times \frac{\text{NCSU}}{|u_A^i N| + |u_B^j N|}. \quad (2)$$

In order to store large similarity user pairs and their similarity, the *User\_sim* set is designed. Its element data structure includes *User* attribute and *Sim* attribute, among which *User* attribute contains user [0] and user [1], user [0] stores A network user node, user [1] stores B network user node. *Sim* stores the similarity between A network user node and B network user node. The structure is shown in Figure 4:

*User\_sim*[k]. user [0] represents the A network user in the *k*th user pair of *User\_sim* set, *User\_sim*[k]. user [1] represents the B network user in the *k*th user pair of *User\_sim* set, *User\_sim*[k]. sim represents the similarity of the *k*th user pair in the *User\_sim* set.

Calculate and filter out user pairs with greater similarity between social networks A and B:

**3.3.3. User Pair Two-Way Matching.** Considering the accuracy of user-pair matching, this paper uses the user-pair screening mechanism (user two-way matching). The user in the A network and the user in the B network are selected with the greatest similarity, and the user similarity in the B network and the A network is also the largest. The user pair with the highest similarity is selected in both directions as a result, and the rest wait for matching.

As shown in Figure 5, it is the two-way matching process. After two-way matching, two user matching pairs are generated, and the remaining two users wait for the next match. User matching is mainly judged by the similarity of the user pairs. We sort the similar user pairs *User\_sim* obtained in Section 3.3.2 according to the similarity *sim* from large to small and match the sorted similar users.

The two-way matching process is as follows:

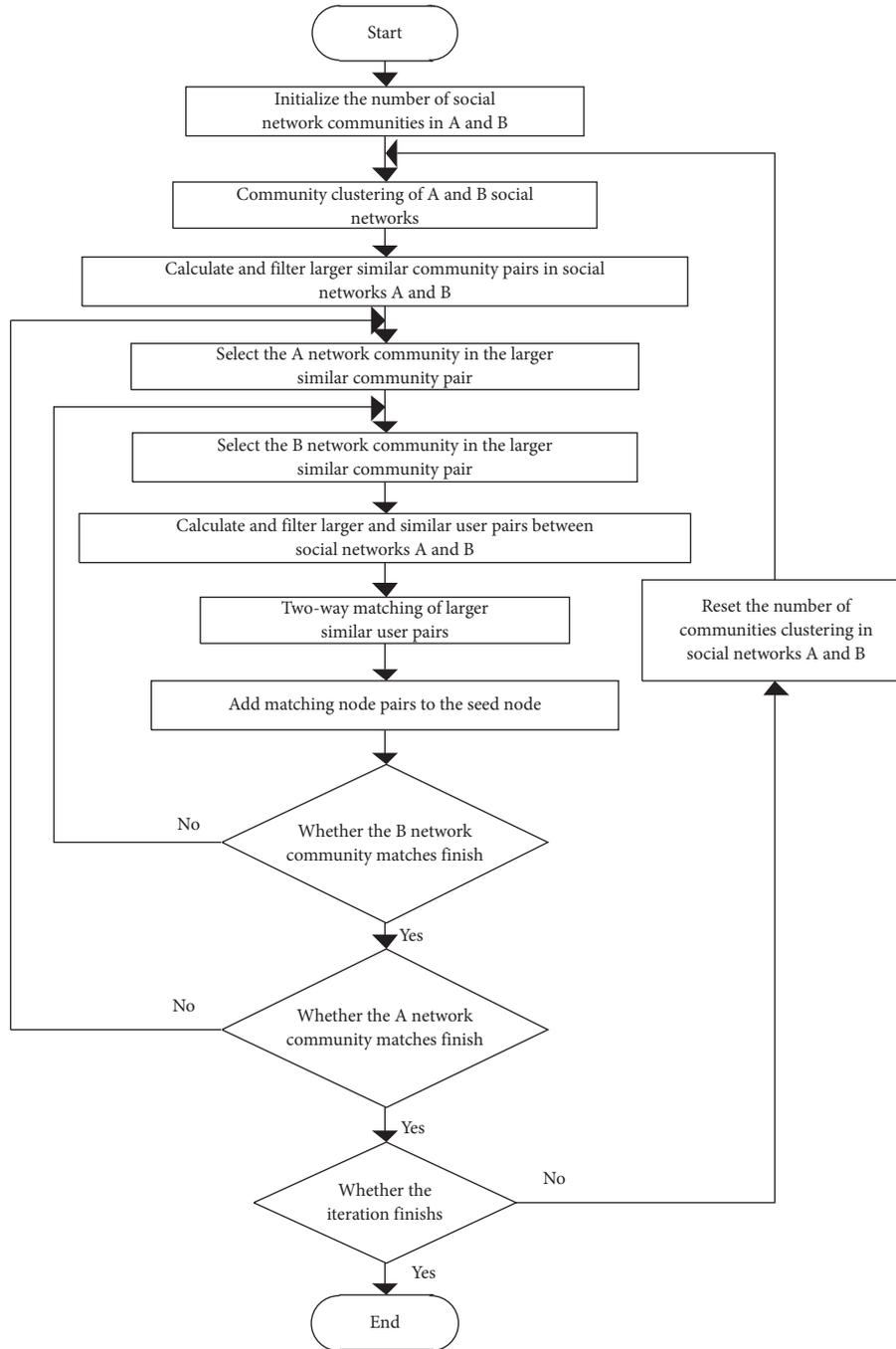


FIGURE 2: Flow chart of UIUTDC algorithm.

## 4. Experimental Results and Analysis

**4.1. Experimental Dataset.** In this paper, the Twitter-Foursquare [18] dataset is selected for the experiment. First, the user's homepage in Twitter is found according to the URL link in the user's homepage in Foursquare to determine the seed nodes. Then, the two social networks are processed, respectively, according to the user's node degree, and the nodes whose user node degree was less than 1 were deleted. The dataset is shown in Table 1, which shows the relevant information of two real-world

social network datasets, among which the number of anchor links between the two social networks is 1862. Here, the nodes connected by anchor links in the two networks are regarded as seed nodes. The percentage of seed nodes in Twitter and Foursquare is 69.6% and 61.7%, respectively.

**4.2. Evaluation Criterion.** Since in the experimental data, 1862 node pairs are known to be matched, it is uncertain whether there are other matching nodes except for these

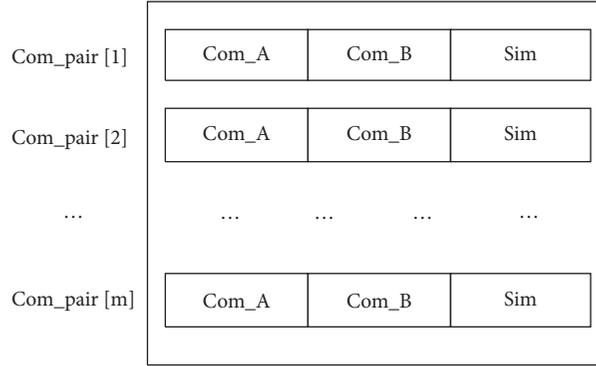


FIGURE 3: Com\_pair data structure.

---

Pseudocode for calculating community pairs with greater similarity between A and B networks

---

Input: Initialize the set of similar communities Com\_pair, the a and B social networks of the divided communities, the community similarity threshold  $\epsilon$

Output: Similar community set Com\_pair

- 1: For each ap in A network//Ap is a community in a network
  - 2: For each bq in B network//Bq is a community in B network
  - 3: Calculate the similarity csim between ap and bq according to formula (1)
  - 4: If csim >  $\epsilon$
  - 5: Add the ap and bq communities and their corresponding user node sets to Com\_pair
  - 6: End for 7: End for 8: Return Com\_pair
- 

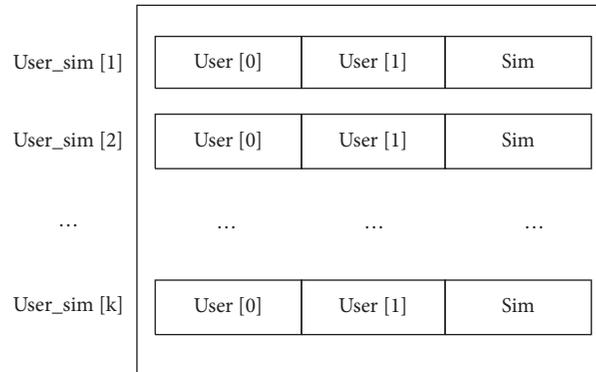


FIGURE 4: User\_sim data structure.

---

Pseudocode for calculating and filtering user pairs with greater similarity between social networks a and B:

---

Input: Large similarity community pair set Com\_pair, initialize large similarity user pair set User\_sim, user pair similarity threshold  $\theta$

Output: User\_sim with greater similarity

- 1: For  $k=0$  to length (Com\_pair)-1//k is the number of Com\_pair community pairs
  - 2: For each ACu in Com\_pair (k). com\_A//Acu is all users of the a network community in Com\_pair (k)
  - 3: Get the neighbor node set of the ACu user node--acu\_neighbor
  - 4: For each BCu in Com\_pair (k). com\_B//BCu is all users of the B network community in Com\_pair (k)
  - 5: Get the neighbor node set of the BCu user node--BCu\_neighbor
  - 6: Calculate the similarity usim between user ACu and BCu according to formula (2)
  - 7: If usim >  $\theta$
  - 8: Add [(ACu, BCu), Usim] to User\_sim
  - 9: End if
  - 10: End for 11: End for 12: End for 13: Return User\_sim
-

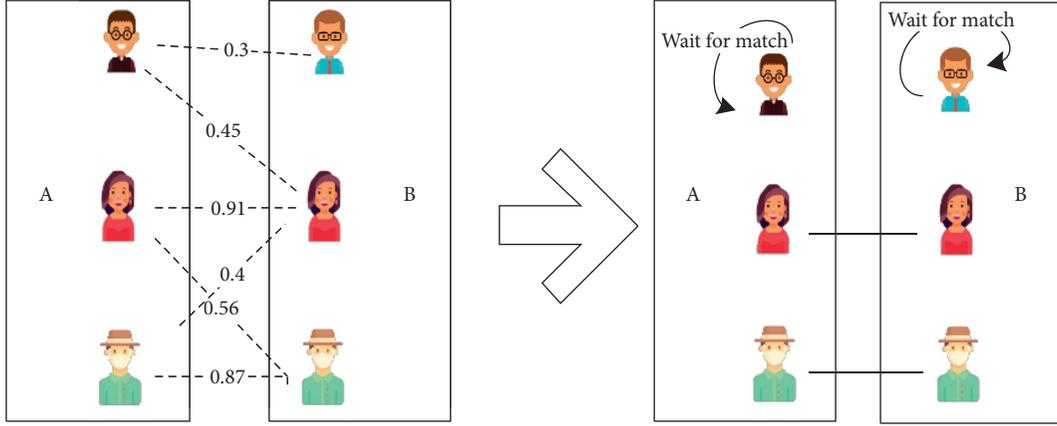


FIGURE 5: The process of user pair two-way matching.

## Two-way matching process pseudocode

Input: User\_sim with greater similarity

Output: User\_sim after two-way matching

1: Sort the User\_sim collection according to sim from largest to smallest

2: For  $i=0$  to length (User\_sim)-23: If User\_sim [i]. sim  $\neq 0$ 4: For  $j=i+1$  to length (User\_sim)-1

5: If User\_sim [i]. user [0] = User\_sim [j]. user [0] or User\_sim [i]. user [1] = User\_sim [j]. user [1]//The same user has multiple matches with another network user with high similarity

6: User\_sim [j]. sim = 0//Since the match is sorted by sim from largest to smallest, the match appears after User\_sim [i], and the user-pair similarity is marked as 0, which is the deletion mark

7: End if

8: End for 9: End if

10: End for 11: Return User\_sim

TABLE 1: Dataset description.

Property	Foursquare	Twitter
Nodes	3018	2672
Link	34920	32787
Anchors	1862	1862

nodes. Therefore, it can only be judged how many matching node pairs are found out of the 1862 node pairs except the prior seed nodes. It is impossible to determine whether the node pair that is not judged as a matching node pair is correct, so this article only uses the accuracy rate (that is, how many seed node pairs are correctly found out of 1862 seed node pairs) as the evaluation criterion. The specific calculation is shown in formula (3). Where Acc represents the accuracy rate,  $F\_seed$  represents the number of matching node pairs found in the final iteration, and  $SU$  represents the known user matching node pairs, that is, the number of 1862 anchor link (matching) node pairs in this experiment,  $R\_seed$  is the number of prior seed node pairs randomly selected from 1862 matching node pairs.

$$Acc = \frac{F\_seed - R\_seed}{SU - R\_seed}. \quad (3)$$

This paper randomly selects 100 and 200 prior seed node pairs and uses the global user matching method (GUMM), the hidden label node method (HLNM), and the UIUTDC algorithm to Conduct Experiments. The comparative analysis of the results is as follows.

4.2.1. Comparison and Analysis of 100 Prior Seed Node Pairs Iterative Matching Node Pairs. As shown in Figure 6, when 100 a priori seed nodes are randomly selected, the experimental results can be seen that the result of the UIUTDC

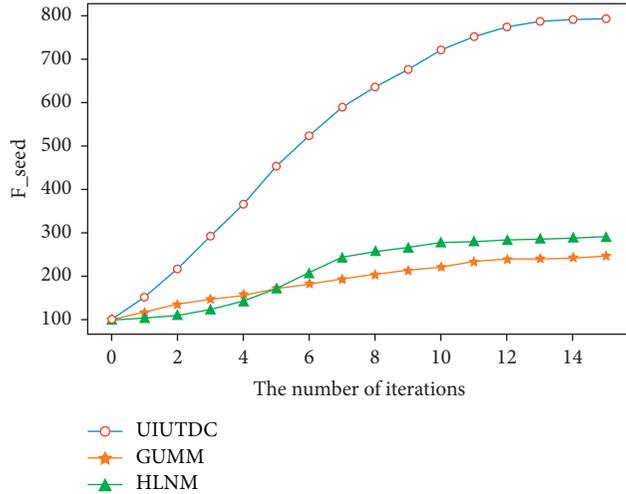


FIGURE 6: Comparison diagram of 100 prior seed nodes to experiment.

method proposed in this paper from the beginning of the iteration to the end of the iteration is far greater than the hidden label node method and the global node matching method. The reason is that the UIUTDC method uses dynamic community division, matching in two stages to make the matching process more comprehensive and the matching result better. The number of pairs of nodes matched by the hidden label node method in the early stage is less than that of the global matching method. The reason is that the method of hiding the label node is based on the degree of the node (that is, the number of friends the user has in the social network) from large to small. In the early stage, nodes with a larger degree are selected to participate in the matching, and there are fewer nodes with a larger degree in the network, resulting in fewer nodes participating in the matching in the early stage, and the result is lower than the result of the global matching method. It is not until the fifth iteration that more nodes from the hidden label node method participate in the matching process that the number of nodes matched is more than that of the global matching method.

**4.2.2. Comparison and Analysis of 100 Prior Seed Node Pairs Iterative Matching Node Pairs.** As shown in Figure 7, when 200 seed nodes are randomly selected, the experimental result graph shows that as the number of iterations increases, the number of node pairs generated by the UIUTDC algorithm is always greater than the number of node pairs generated by the hidden label node method and the global matching method. The hidden label node method generates fewer seed nodes before the 7th iteration than the global matching method. The reason is that the method of hiding the label nodes selects the nodes participating in the matching according to the degree of the nodes from large to small, resulting in fewer nodes participating in the matching in the early stage, and the result is lower than the result of the global matching method. Until the seventh iteration, all the nodes in the hidden-tag node method participate, and the result is better than that of the global matching method.

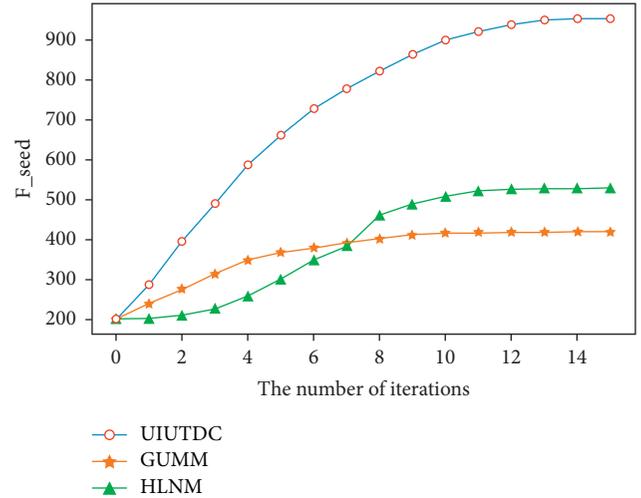


FIGURE 7: Comparison diagram of 200 prior seed nodes to experiment.

**4.2.3. Comparison and Analysis of Accuracy of Different Methods.** It can be seen from the accuracy graphs of different seed nodes in Figure 8, on the one hand, the accuracy of the UIUTDC method is much higher than the accuracy of global user matching and the accuracy of hidden label nodes. On the other hand, it can be seen that when there are few seed nodes, the accuracy of the global matching method and the hidden label node method is not very high. However, with the increase of seed nodes, the accuracy of the UIUTDC method, global user matching method, and hidden label node method are all improving, which shows that prior seed nodes have a certain influence on the experimental results. The more points, the higher the accuracy rate, and the most obvious are hidden label nodes. The experimental results show that the average accuracy of the UIUTDC method is 42.33% higher than the average accuracy of global user matching by 33% and 26.8% higher than the average accuracy of the hidden label node method.

**4.2.4. Comparison and Analysis of Time Consumption of Different Methods.** This paper uses the same computer to verify the UIUTDC method, the global user matching method, and the hidden label node method on the real network data set and obtains the running time comparison chart of different seed nodes, as shown in Figure 9. It can be seen from the time comparison graph that the overall time consumption of the UIUTDC method is much lower than that of the global user matching algorithm and the hidden label node method. That is, its time complexity is better than the global matching method and the hidden label node method. The global user matching method requires users in two social networks to perform a one-to-one matching calculation (assuming that both network user nodes are  $n$ ), which is extremely complicated. The time complexity is  $O(n^2)$ . In the UIUTDC method, the number of nodes in the cluster is much less than that of the entire network (assuming the number of clusters is  $isk$ , the average node in each cluster is  $l, l = n/k$ ), which greatly reduces the overall cost of similarity calculation for all nodes in

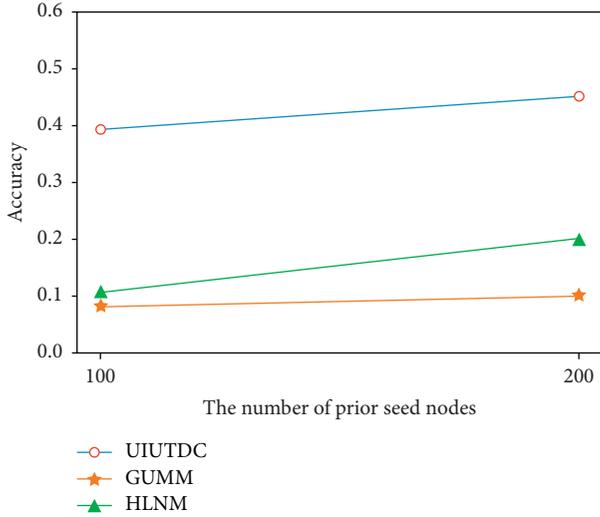


FIGURE 8: The accuracy of different seed nodes.

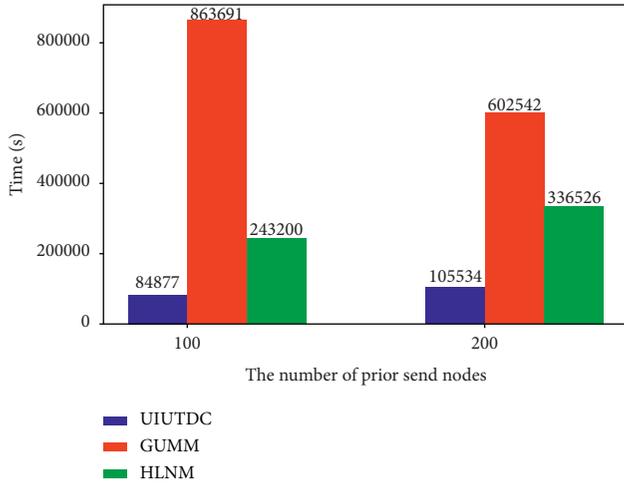


FIGURE 9: Comparison of running time of different seed nodes.

the network. The cost includes two parts: one is the cluster similarity calculation,  $K$  clusters match each other, the calculation cost is  $O(k^2)$ , the other is the user node matching within the cluster, the calculation cost is  $O(l^2)$ . The total cost is  $O(n) + O(k^2) + O(kl^2)$ , which is  $O(k^2 + kl^2 + n)$ . Since  $l = n/k$ , the total cost is  $O(k^2 + (n^2/k) + n)$ ,  $1 < k < n$ . It can be calculated that when  $n$  is very large, the appropriate value of  $k$  can make  $O(k^2 + (n^2/k) + n) < O(n^2)$ , that is, the computational complexity of the UIUTDC method is less than the global user matching. The hidden label node method has fewer participating nodes in the early stage, so the early time consumption is less, but as the number of nodes increases, the number of calculations increases, resulting in longer time consumption in the later stage. The UIUTDC method is based on the division of communities and the selection of communities with high similarity, which greatly reduces the calculation of user matching to a certain extent. The UIUTDC method is superior to the global user matching method and hidden label node method in terms of time.

## 5. Conclusions

This paper proposes a dynamic community clustering two-stage user identification algorithm based on user topological relationships. The algorithm uses the user topological structure information in the social network to match by dividing the social network into communities and selecting communities with greater similarity in different networks to match the nodes in the community. Based on this method, we can reduce the time complexity of the matching algorithm while improving the accuracy of node matching. In order to prevent the loss of node information when the community is divided according to the fixed number of communities for node matching, dynamic community division is adopted. The number of community divisions in each iteration is different, and the nodes in the network community are fully matched from different angles, which can improve the accuracy of node matching. Applying the algorithm in this paper to a real social network data set, the results show that the effect of the algorithm is 33% and 26.8% higher than the global user matching algorithm and the hidden label node algorithm, respectively. In terms of time, the algorithm in this paper reduces on average 637911 seconds and 1,94657 seconds than the global user matching algorithm and the hidden label node algorithm.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the Key Research Projects of Humanities and Social Sciences in Colleges and Universities of Anhui Province under Grant number SK2019A0664, Provincial Natural Science Research Projects of Universities in Anhui Province-General Projects under Grant numbers KJ2019JD24 and KJ2019JD17, Provincial Natural Science Research Projects of Universities in Anhui Province-Key Projects under Grant number KJ2019A0783, and 2019 Provincial Quality Engineering Project of Anhui Province under Grant number 2019jyxm1146.

## References

- [1] Y. Yang, H. Yu, Y. Huang, and Z. Liu, "Cross-social network user identity matching based on fusion representation learning," *Computer Engineering*, vol. 44, no. 9, pp. 45–51, 2018.
- [2] J. B. Liu, S. Wang, C. Wang, and S. Hayat, "Further results on computation of topological indices of certain networks," *IET Control Theory & Applications*, vol. 11, no. 13, pp. 2065–2071, 2017.
- [3] R. Zafarani and H. Liu, "Connecting corresponding identities across communities," in *Proceedings of the Third International*

- Conference on Weblogs and Social Media, ICWSM*, pp. 354–357, San Jose, CA, USA, May 2009.
- [4] D. Perito, C. Castelluccia, M.A. Kaafar, and P. Manils, “How unique and traceable are usernames,” in *Proceedings of the International Symposium on Privacy Enhancing Technologies Symposium*, pp. 1–17, Waterloo, CA, USA, July 2011.
  - [5] D. Liu, Q. Y. Wu, W. H. Han, and B. Zhou, “A user identity identification method based on user name feature,” *Chinese Journal of Computers*, vol. 038, no. 10, pp. 2028–2040, 2015.
  - [6] J. Vosecky, D. Hong, and V. YS. Shen, “User identification across multiple social networks,” in *Proceedings of the International Conference on Networked Digital Technologies*, pp. 360–365, Dubai, UAE, April 2009.
  - [7] M. M. Luan, T. Zhao, X. H. Yang, X. Y. Li, and J. Zhang, “An overview of cross-social network user identity recognition,” *Journal of Qilu University of Technology*, vol. 4, pp. 55–60, 2020.
  - [8] X. Kong, J. Zhang, and P. S. Yu, “Inferring anchor links across multiple heterogeneous social networks,” in *Proceedings of the Acm International Conference on Information & Knowledge Management*, pp. 179–188, San Francisco, CA, USA, October 2013.
  - [9] S. Liu, S. Wang, F. Zhu et al., “Hydra: Large-scale social identity linkage via heterogeneous behavior modeling,” in *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, pp. 51–62, Snowbird, UT, USA, June 2014.
  - [10] R. Roedler, D. Kergl, and G. D. Rodosek, “Profile matching across online social networks based on geo-tags,” in *Advances in Nature and Biologically Inspired Computing*, pp. 417–428, Springer, Berlin, Germany, 2016.
  - [11] A. Narayanan and V. Shmatikov, “De-anonymizing social networks,” in *Proceedings of the 2009 30th IEEE Symposium on Security and Privacy*, pp. 173–187, Berkeley, CA, USA, May 2009.
  - [12] N. Korula and S. Lattanzi, “An efficient reconciliation algorithm for social networks,” pp. 1307–1690, 2013, <https://arxiv.org/abs/1307.1690>.
  - [13] X. Zhou, X. Liang, H. Zhang, and Y. Ma, “Cross-platform identification of anonymous identical users in multiple social media networks,” *Institute of Electrical and Electronics Engineers Transactions on Knowledge and Data Engineering*, vol. 28, no. 2, pp. 411–424, 2016.
  - [14] O. Peled, M. Fire, L. Rokach, and Y. Elovici, “Entity matching in online social networks,” in *Proceedings of the International Conference on Social Computing*, pp. 339–344, Washington, DC, USA, September 2013.
  - [15] J. Liu, F. Zhang, X. Y. Song, Y. I. Song, C. Y. Lin, and H. W. Hon, “What’s in a name? an unsupervised approach to link users across communities,” in *Proceedings of the Sixth Acm International Conference on Web Search & Data Mini*, pp. 495–504, Rome, Italy, February 2013.
  - [16] L. Xing, K. Deng, H. Wu, P. Xie, and J. P. Gao, “Behavioral habits-based user identification across social networks,” *Symmetry*, vol. 11, no. 9, pp. 1134–1152, 2019.
  - [17] Z. Wu, H. T. Yu, R. Y. Huang, and S. X. Liu, “Cross-network user identity recognition based on hiddentag node mining,” *Application Research of Computers*, vol. 35, no. 04, pp. 1191–1196, 2018.
  - [18] J. W. Zhang and S. Yu. Philip, “Integrated anchor and social link predictions across social networks,” in *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, pp. 2125–2132, Buenos Aires, Argentina, July 2015.