# Identification of Anonymous User in Social Media Network through Profile and Friend Relationship

**K.Shanthi[1] Mr.G.Rajarajacholan[2]**
[1]M.Phil Research Scholar [2]Assistant Professor
[1,2]Department of Computer Science & Engineering
[1]Prist University, Thanjavur [2]Centre for Knowledge Management, Prist University, Thanjavur

*Abstract*— In social media networks (SMN), profile details of one user can be used by others to create account with original user identity or the original user may have multiple accounts in multiple social media sites. Discovery of multiple accounts that belong to the same person is an interesting and challenging work in social media analysis. Contents and network structures can be used for user identification in social media sites. The main idea of this paper is to identify alias and identical accounts by merging multiple SMN in order to get complete information about a particular user. This paper develops a methodology Profile and Friend Relationship-Based Anonymous User Identification (PFRAUI) algorithm for mapping individuals on cross application SMN's. The friend cycle of every individual differs therefore, accuracy of this result will be maintained if use friend list as a key component to analyze cross application social media networks. It also combines profile attributes to match the users. The combination of profile and friends relationship based method efficiently identifies the anonymous users. Results of extensive experiments demonstrate that PFRAUI performs much better than current network structure-based algorithms.

***Key words:*** Social Media Network, Anonymous Identical Users, Friend Relationship, User Profile, User Identification

## I. INTRODUCTION

Today, most of the people use social media sites. It is obvious that people tend to use different social media application for different purpose. Facebook, is a profit corporation and most popular social media application in the world, has more than 1.7 billion users. Twitter is an online social networking service that allows users to send and read small 140-character messages called "tweets". At the second half of 2016 the number of registered users was more than 313 million users. Only registered users can read and post tweets, otherwise unregistered users can only read them.

Due to this diversity of online social media networks (SMNs), people tend to use different SMNs for different purposes. For instance, face book is used to connect with people all over the world and exchange their thoughts through messaging. Twitter provides micro-blog service where people tweet or share their opinion. Every existent SMN satisfies some user needs. In terms of SMN management, matching anonymous users across different SMN platforms can provide integrated details on each user and inform corresponding regulations, such as targeting services provisions.

With the growth of SMN platforms on the Internet, the cross-platform approach has merged various SMN platforms to create richer raw data and more complete SMNs for social computing tasks. SMN users form the natural bridges for these SMN platforms. The primary topic for cross-platform SMN research is user identification for different SMNs. User identification is also called user recognition, user identity resolution, user matching, and anchor linking. Although no solution can identify all identical anonymous SMN users, some SMN elements may be used to identify a portion of users across multiple SMNs.

A Novel Technique for investigating privacy and confidentiality in social networks and implement a novel re-identification algorithm targeting anonymized social-network graph was obtainable in [1] for data protection and confidentiality. A novel technique for Conditional Random Fields based user profile matching proposed by Bartunov et al [2] that broadly merges usage of outline attributes and social linkage.

SMN connections fall into two categories: single-following connections and mutual-following connections. Single-following connections are also called following relationships or following links. If user A follows user B, then user A and user B have a following relationship (single-way fans in which one knows the other, but not vice versa). Following relationships are common in micro-blogging SMNs, such as Twitter and Sina Micro-blog. Likewise, mutual following connections are called friend relationships. In micro-blogging SMNs, a friend relationship refers to the mutual following relationships between two users. In most other SMNs, such as Face book, RenRen and Wechat, a friend relationship forms only if a friend request is sent by one user and confirmed by the other user. Friend relation-ships are difficult to fake by malicious users, and therefore reflect real-world relationships much better. Due to their reliability and consistency, friend relationships are more robust in user identification tasks. Moreover, since unified friend relationships are formed, our algorithm can also be applied to SMNs with a heterogeneous network structure, such as Twitter and Face book.

This paper focused on profile and friend relationships in SMNs and developed a new algorithm based on network structures. This algorithm can only identify a portion of the identical users in SMNs. However, it can be applied jointly with other feature-based user identification algorithms for more accurate identification results.

This paper is organized as follows: The section 2 discusses related work of user identification. The Section 3 describes the proposed work of profile and friend relationship-based anonymous user identification. Section 4 presents the implementation details of proposed work and section 5 provides the conclusion of this work

## II. RELATED WORK

This section discussed some related work of user identification in social network.

 

Perito et al [3] investigate the feasibility of using usernames to trace or link multiple profiles across services that belong to the same individual. The perception is that the possibility that two usernames refer to the similar physical person powerfully depends on the "entropy" of the username string itself. It uses merely username to discover profile.

Liu et al. [4] connect user's crossways many online communities. They spotlight on the alias-disambiguation to distinguish users with similar usernames. They proposed using the n-gram probabilities of usernames to estimate the rareness or commonness of username.

Zafarani et al. [5] provide the evidence on the existence of a mapping among identities across multiple communities, providing a method for connecting the community websites. They used username to discovery corresponding identities across communities.

Zafarani et al. [6] further develop behavioral-modeling approach for effective user identification using machine leaning technique. It identifies users' unique behavioral patterns that lead to information redundancies across sites.

Acquisti et al. [7] addressed the user identification task using profile photos. They conducted the experiment on Facebook using face recognition algorithm. Facebook profile photos are visible to all by default. Most of the members use photos of them as primary profile image and use real first and last names on their profiles. Face recognition of everyone or everywhere or all the time is not yet feasible. Both screen name and profile image can identify users but they cannot be applied to large social media networks. This is because some users may have the same screen name and profile images. Iofciu et al [8] identifies users based on their user ids and social tags. Motoyama et al [9] gathered attributes as sets of words and matched users by calculating the similarity of users.

Zheng et al [10] developed a structure for authorship detection of online messages to tackle the identity-tracing trouble. Four kinds of writing-style features (syntactic, lexical, content-specific features and structural) are mined and inductive learning algorithms are used to construct feature-based classification replicas to recognize authorship of online messages.

Almishari et al [11] anticipated a technique to connect community based on user reviews. Kong et al [12] proposed anchor link prediction across multiple heterogeneous social networks, i.e., discovering the correspondence among different accounts of the same us. Goga et al [13] proposed a technique for correlating user accounts across sites, based on otherwise innocuous information like location and timing patterns.

Geo-location appears to have forceful features for user recognition. However this information is often sparse in SMNs, since only a small portion of users is willing to post their locations. Although writing style solutions perform well in scenarios involving long content, these techniques are not applicable to SMS.

There are so many network structure based user identification methods are used to recognize identical users. Korula et al [14] proposed social networks reconciliation. It use many-to-many mapping algorithm based on the degree of unmapped users and the number of common neighbors.

Zhou et al. [15, 16] analyzed the neighborhood attacks of de-anonymization and proposed privacy preservation approaches using k-anonymity and l-diversity. Hay et al [17] propose three models of external information used by an adversary to attack naively-anonymized networks.

Jain et al [18] proposed Finding Nemo which uses all the three dimensions (profile, content and network) of an identity to search for a user on multiple social networks. The system exploits a known identity on one social network to search for her identities on other social networks.

JLA [2] attempts to match unmapped nodes from different graphs by comparing the mapped neighbors of each node. It calculates a network distance between any two unmapped nodes in two undirected networks. Similarity, studies show that certain profiles can be matched based solely on the network structure using JLA. This paper proposes an innovative approach to address the challenges faced by previous studies. This new approach focuses on the profile and friendship structure, and develops the Profile and Friend relationship based used identification (PFRUI) algorithm.

## III. ANONYMOUS USER IDENTIFICATION IN SMN

### *A. Problem Statement*

The users of internet are increasing and a large number of them are an active member of a social network. People rely on different social media networks for news, information and opinion of other people about different subjects. People have multiple accounts in multiple social media site. Discovering the same user accounts that belong to the same user is becoming a growing interest among researchers. Though it is more challenging, it is useful in developing many applications. Identifying the same user accounts among multiple sites is also useful in the application automatic contacts' merging that happens almost in most of the mobile phones. So identifying the same user accounts among multiple online social media sites is a challenging research area.

The SMN proves to be the best platform for information retrieval. However, identifying unknown and identical users on multiple social media application is still an unsolved problem. People use different social media for different purpose; the idea of integrating multiple social media application can take the research a step forward. To identify identical accounts by merging multiple SMN in order to get complete information about a particular user, need a novel algorithm for mapping individuals on cross application SMN's.

For various reasons, some people hold many accounts in the same SMN, yet often assume that these multiple accounts are independent and belong to different individuals. In other words, only identify one of these accounts.

To solve the user identification problem the profile and friend relationship based user identification methods was proposed. It identifies anonymous users in multiple SMN.

 

### *B. Proposed Methodology*

In this section, the proposed profile and friend relationship-based anonymous user identification is described. Figure 1 shows the flow of proposed system model.

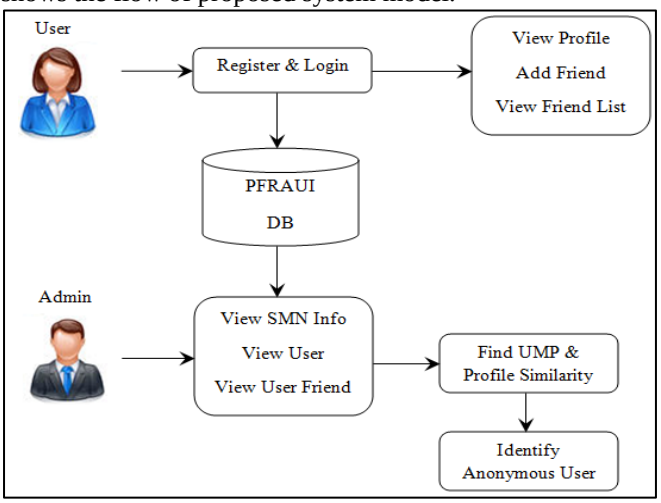


Fig. 1: Proposed System Model

This paper proposes Profile and Friend Relationship-Based Anonymous User Identification (PFRAUI) to match a degree of all candidates User Matched Pairs (UMP), and only UMP's with top ranks are considered as identical users. It also scrutinizes the identical profiles and find out the common attributes to improve the accuracy of our algorithm. The network structure based user identification first obtains a Prior UMP'S through pre-processor, and then identifies more UMPs through the Identifier in an iteration process.

An SMN is defined as SMN= {U, P, C, I}, where U, P, C and I denote the users, profile connections and interactions among users, respectively.

A User Entity (UE) is a user in combination with his or her profile, connections and interaction content. An SMN is a set of UEs which has the same number as the accounts in the SMN. Similarly, $UE_A$ is used to indicate the UE list of $SMN_A$, and $UE_{Ai}$ is taken as the token of the i-th element in $UE_A$.

Given that $SMN_A$ and $SMN_B$, if $UE_{Ai}$ and $UE_{Bj}$ belong to the same individual in real-life, which is denoted as Ψ, then we hold that $UE_{Ai}$ and $UE_{Bj}$ match on Ψ, and they compose a User Matched Pair $UMP_\Psi$. $UMP_\Psi$ can also be expressed as $UMP_{A\sim B}(i, j)$ or $UMP(UE_{Ai}, UE_{Bj})$,equivalently.

The user identification problem of two $SMN_A$ and $SMN_B$ can be defined as

$$F(U_{Ai}, U_{Bj}) = g(M_{ij}) \;\&\; ProSim(U_{Ai}, U_{Bj})$$

$$g(M_{ij}) = \begin{cases} 1, & if\ U_{Ai}\ and\ U_{Bj}\ belong\ to\ the\ same\ individual \\ 0, & Otherwise \end{cases}$$

Where Mij denotes the match degree of $U_{Ai}$ and $U_{Bj}$'s known (identified) friends.

$$Pr\,oSim(U_{Ai}, U_{Bj}) = \begin{cases} 1, & if\ U_{Ai}\ and\ U_{Bj}\ profile\ matched \\ 0, & Otherwise \end{cases}$$

Where ProSim find the similarity score of $U_{Ai}$ and $U_{Bj}$ . if the similarity score is > threshold value (like 0.6 or 0.7) then ProSim value is 1 otherwise 0.

The proposed PFRUI algorithm is summarized in Figure 2

**Algorithm: PFRUI**

Input: $SMN_A$, $SMA_B$

Output: Identified Users

1. Extract User Info in both $SMN_A$ and $SMA_B$
2. For each User $U_A$ in $SMN_A$
3. For each User $U_B$ in $SMN_B$
4. Extract friend list of $U_A$ and $U_B$
5. Match $U_A$ and $U_B$
6. $g(M) = \begin{cases} 1, & if\ U_A\ and\ U_B\ belong\ to\ the\ same\ individua \\ 0, & Otherwise \end{cases}$
7. For each User $U_A$ in $SMN_A$
8. For each User $U_B$ in $SMN_B$
9. Extract Profile Info of $U_A$ and $U_B$
10. Compute Similarity based on LCS (Longest common substring)
11. if (similarity is > threshold)
12. Profile Matched.
13. $ProSim(U_A, U_B) = \begin{cases} 1, & if\ U_A\ and\ U_B\ profile matched \\ 0, & Otherwise \end{cases}$
14. Display Matched User Information

Fig. 2: PFRUI Algorithm

## IV. IMPLEMENTATION

This section explains the performance evaluation of proposed approach. The Profile and Friend Relationship based Anonymous User Identification was implemented using Java (version 1.8), and the experiments are performed on a Intel(R) Pentium machine with a speed 2.13 GHz and 2.0 GB RAM using Windows 7 32-bit Operating System.

In this paper, an application was developed through which can perform user identification. To register this application, user selects type of social network (Facebook, Twitter), gives their user name, password and basic information of user (i.e, mobile no, email, gender and location). After the login process, the user can view their profile, add friends and view their friend list. The administrator has right to identifies the anonymous user.

Table 1 and figure 3 shows the comparison of FRUI and PFRUI Precisions values.

| Percentage of Prior UMP | Precision | |
|---|---|---|
| | FRUI | PFRUI |
| 0.1 | 0.623 | 0.783 |
| 0.2 | 0.659 | 0.797 |
| 0.3 | 0.668 | 0.856 |
| 0.4 | 0.697 | 0.874 |
| 0.5 | 0.717 | 0.888 |
| 0.6 | 0.732 | 0.891 |
| 0.7 | 0.759 | 0.912 |

Table 1: FRUI and PFRUI Precisions

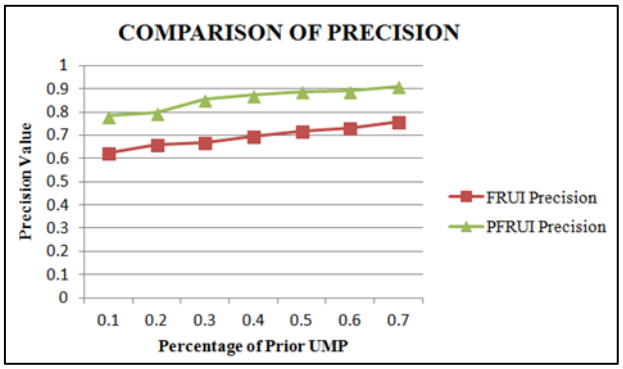


Fig. 3: Comparison of FRUI and PFRUI Precisions

Table 2 and figure 4 shows the comparison of FRUI and PFRUI recall rates

| Percentage of Prior UMP | Recall Rate | |
|---|---|---|
| | FRUI | PFRUI |
| 0.1 | 0.39 | 0.52 |
| 0.2 | 0.418 | 0.542 |
| 0.3 | 0.431 | 0.561 |
| 0.4 | 0.452 | 0.62 |
| 0.5 | 0.487 | 0.658 |
| 0.6 | 0.493 | 0.679 |
| 0.7 | 0.523 | 0.69 |

Table 2: FRUI and PFRUI Recall Rates

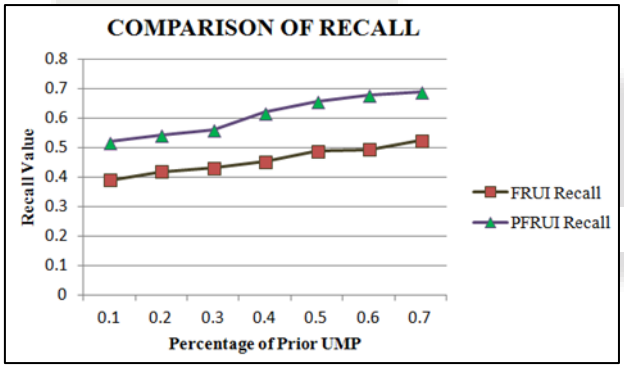


Fig. 4: Comparison of FRUI and PFRUI Recall Rates

Table 3 and figure 5 shows the running time comparison of FRUI and PFRUI

| Percentage of Prior UMP | Running Time in Seconds | |
|---|---|---|
| | FRUI | PFRUI |
| 0.1 | 32 | 24 |
| 0.2 | 40 | 31 |
| 0.3 | 38 | 26 |
| 0.4 | 37 | 28 |
| 0.5 | 39 | 25 |
| 0.6 | 38 | 27 |
| 0.7 | 36 | 29 |

Table 3: FRUI and PFRUI Running Times in seconds

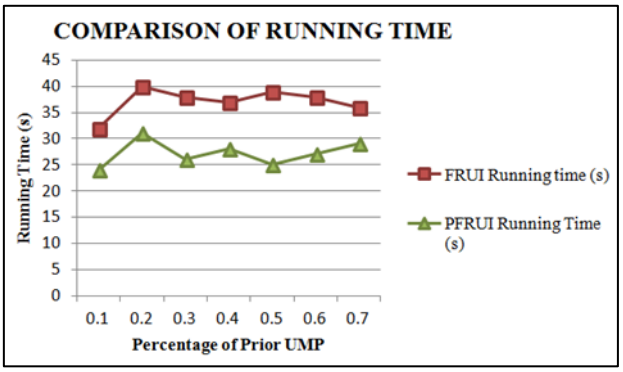


Fig. 5: Comparison of FRUI and PFRUI Running Times in seconds

Results show that PFRUI is more efficient in practice, which is consistent with the theoretical analysis.

## V. Conclusion

Identifying same user accounts across multiple social media sites is interesting and also it is very challenging. This study addressed the problem of user identification across multiple social media network. This paper developed a novel profile and friend relationship based algorithm called PFRUI. This algorithm is simple, yet efficient and performed much better than existing state of art network structure-based user identification solution. Identifying anonymous users across multiple social media network is challenging work. Therefore, only a portion of identical users with different nicknames can recognized with this method.

## References

[1] A. Narayanan and V. Shmatikov,(2009) "De-anonymizing social net-works," Proc. Of the 30th IEEE Symposium on Security and Privacy (SSP'09), pp. 173-187, 2009.

[2] S. Bartunov, A. Korshunov, S. Park, W. Ryu, and H. Lee,(2012) "Joint link-attribute user identity resolution in online social net-works," The 6th SNA-KDD Workshop '12, 2012.

[3] D. Perito, C. Castelluccia, M. A. Kaafar, and P. Manils,(2011) "How unique and traceable are usernames?" in Proc. 11th Int. Conf. Privacy Enhancing Technol., 2011, pp. 1–17

[4] J. Liu, F. Zhang, X. Song, Y. I. Song, C. Y. Lin, and H. W. Hon, (2013) "What's in a name?: An unsupervised approach to link users across communities," in Proc. 6th ACM Int. Conf. Web Search Data Mining, 2013, pp. 495–504.

[5] R. Zafarani and H. Liu,(2009) "Connecting corresponding identities across communities," in Proc. 3rd Int. ICWSM Conf., 2009, pp. 354–357.

[6] R. Zafarani and H. Liu,(2013) "Connecting users across social media sites: a behavioral-modeling approach," in Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2013, pp. 41–49

[7] A. Acquisti, R. Gross, and F. Stutzman,(2011), "Privacy in the age of augmented reality," in Proc. Nat. Acad. Sci.,2011,pp.3653

[8] T. Iofciu, P. Fankhauser, F. Abel, and K. Bischoff,(2011), "Identifying users across social

 

tagging systems," in Proc. 5th Int. AAAI Conf. Weblogs Social Media, 2011, pp. 522–525

[9] M. Motoyama and G. Varghese,(2009) "I seek you: searching and matching individuals in social networks," in Proc. 11th Int. Workshop Web Inf. Data Manage., 2009, pp. 67–75.

[10] R. Zheng, J. Li, H. Chen, and Z. Huang,(2006) "A framework for au-thorship identification of online messages: writing style features and classification techniques," J. of the American Society for Information Science and Technology, vol. 57, no. 3, pp. 378-393, 2006.

[11] M. Almishari and G. Tsudik, (2012) "Exploring linkability of user re-views," Computer Security–ESORICS 2012 (ESORICS'12), pp. 307-324, 2012.

[12] X. Kong, J. Zhang, and P.S. Yu,(2013) "inferring anchor links across multiple heterogeneous social networks," Proc. of the 22nd ACM International Conf. on Information and Knowledge Management (CIKM'13), pp. 179-188, 2013

[13] O.Goga, H.Lei, S.H.K. Parthasarathi, G. Friedland, R.Sommer, and R.Teixeira, (2013), "Exploiting innocuous activity for correlating users across sites", Proc. 22nd international conference on world wide web (www'13), pp. 447-548, 2013

[14] N. Korula and S. Lattanzi,(2013) "An efficient reconciliation algorithm for social networks," arXiv preprint arXiv:1307.1690, 2013.

[15] B. Zhou and J. Pei,(2008) "Preserving privacy in social networks against neighborhood attacks," Proc. Of the 24th IEEE Interna-tional Conference on Data Engineering (ICDE'08), pp. 506–515, 2008.

[16] B. Zhou and J. Pei,(2011) "The k-anonymity and l-diversity ap-proaches for privacy preservation in social networks against neighborhood attacks," Knowl. Inf. Syst, vol. 28, no. 1,pp. 47-77, 2011.

[17] M. Hay, G. Miklau, D. Jensen, and D. Towsley,(2008) "Resisting struc-tural identification in anonymized social networks," Proc. of the 34th International Conference on Very Large Databases (VLDB'08), pp. 102-114, 2008.

[18] P. Jain and P. Kumaraguru,(2012) "Finding Nemo: searching and re-solving identities of users across online social networks," arXiv preprint arXiv:1212.6147, 2012.

 