A Comment on "Cross-Platform Identification of Anonymous Identical Users in Multiple Social Media Networks"

Yongjun LI and Zhaoting Su

Abstract—The Friend Relationship-Based User Identification (FRUI) algorithm is considered to be the ideal method. However, if the seed User Matched Pairs (UMPs) were not suitable, FRUI would stop early due to Controversial UMPs. We highlight this gap and propose a minor change to make FRUI a more general identification algorithm.

Index Terms—Cross-platform, social media network, user identification, friend relationship

1 Introduction

A Friend Relationship-Based User Identification (FRUI) algorithm is presented in [1]. As we know that identical users [2] tending to set up partial similar friendships in different Social Media Networks (SMNs), FRUI calculates match degree of friendships for all candidate User Matched Pairs (UMPs), and selects UMP with top rank as identical user. In FRUI, the match degree is the key step, and it considers two cases. The first case considers the number of common identical users. When Controversial UMPs exist, the second case introduces the similarity of known identical users to match degree. FRUI is considered being ideal for user identification solely based on friend relationships. However, if the seed UMPs were not suitable, FRUI would stop early due to Controversial UMPs. Some works [3], [4] have also noticed this problem, and propose novel solutions. Different from the existing works, our object is to highlight this gap and propose a minor change to make FRUI a more general user identification algorithm.

2 Discussions of FRUI

Assume $\mathrm{SMN_A}$ and $\mathrm{SMN_B}$ are two SMNs. $\mathrm{UE}_{\mathrm{A}i}$ is the i^{th} user of $\mathrm{SMN_A}$. $\mathrm{UE}_{\mathrm{A}i}$ and $\mathrm{UE}_{\mathrm{B}j}$ compose a UMP and be denoted as $\mathrm{UMP}_{\mathrm{A}\sim\mathrm{B}}(i,j)$. For convenience, we denote $\mathrm{UMP}_{\mathrm{A}\sim\mathrm{B}}(i,j)$ as (i,j). The match degree of (i,j) is expressed as M_{ij}, and defined by Eq. (1).

$$M_{\mathrm{ij}} = \left|\mathrm{F}_{\mathrm{A}i} \cap \mathrm{F}_{\mathrm{B}j}\right|, \tag{1}$$

where F_u denotes the identified friends of user u. As mentioned in [1], $|\mathrm{F}_{\mathrm{A}i} \cap \mathrm{F}_{\mathrm{B}j}|$ and $|\mathrm{F}_{\mathrm{A}k} \cap \mathrm{F}_{\mathrm{B}j}|$ are equal in some cases, so (i,j) and (k,j) compose Controversial UMPs. To address this problem, FRUI introduces the similarity of known friends to M_{ij}. M_{ij} is further defined by Eq. (2). To describe convenience, we denote this new metric as M^2_{ij}.

$$M^2_{ij} = |F_{\mathrm{A}i} \cap F_{\mathrm{B}j}| + \frac{|F_{\mathrm{A}i} \cap F_{\mathrm{B}j}|}{\min\left(|F_{\mathrm{A}i}|, |F_{\mathrm{B}j}|\right)}, \tag{2}$$

M^2_{ij} still generates some Controversial UMPs with high probability when the seed UMPs are not suitable. This case is explained using an illustration. Let us consider the example shown in Fig. 1, which is illustrated in Fig. 3 of [1].

As illustrated in [1], if we choose (1,1) and (2,2) as seed UMPs, FRUI works very well. However, if (4,4) and (5,5) were selected as seed UMPs, FRUI would ultimately generate some Controversial UMPs as shown in Table 1. In this case, FRUI would stop immediately without identifying any other identical users. The reason for this situation is that every UMP has Controversial UMPs.

We conduct the further study on the friend relationships as shown in Fig. 1, and find that the adjacent relationships of every user are completely different. In other words, if we were to introduce suitable metric to supplement match degree M_{ij}, these above Controversial UMPs would not exist. To overcome the limitation of FRUI, we introduce the Closeness Centrality [5] of user to supplement M_{ij}. The new metric M^3_{ij} is expressed by Eq. (3).

$$M^3_{ij} = |F_{\mathrm{A}i} \cap F_{\mathrm{B}j}| + \frac{\min(CC(\mathrm{A}i), CC(\mathrm{B}j))}{\max(CC(\mathrm{A}i), CC(\mathrm{B}j))}, \tag{3}$$

where $CC(u)$ is closeness centrality of user u.

For Controversial UMPs that cannot be eliminated by FRUI, we calculate their M^3_{ij}, select UMPs with top n rank as identical users and new seed UMPs, continue the identification algorithm. Compared with original FRUI, our suggestion only makes a minor change to M_{ij}, and the rest of the algorithm remains the same. We call FRUI with the proposed suggestion as *p*-FRUI. We also take Fig. 1 as example to illustrate the effectiveness of suggested change. In Fig. 1, $CC(\mathrm{A2}) = 0.5454$, $CC(\mathrm{A6}) = 0.6$, $CC(\mathrm{B2}) = 0.5454$, $CC(\mathrm{B6}) = 0.6$. *p*-FRUI could accurately identify the identical users (2, 2), (6, 6) (7, 7), (1, 1), (3, 3) in turn.

Another example favoring our suggestion is shown in Fig. 2. We consider 12 users and their friendships obtained from Facebook and Twitter, respectively.

If we were to choose (29, 29) and (48, 48) as seed UMPs, FRUI would generates some Controversial UMPs and could not accurately identify any other identical users. However, *p*-FRUI can identify 10 pairs of identical users (81, 81) (71, 71), (164, 164), (52, 52), (39, 39), (49, 273), (8, 49), (31, 31), (273, 8), (378, 378). In the identification results, 7 of 10 pairs are accurate. We conduct the further study on the misidentified UMPs (8, 49), (49, 273) and (273, 8), and find none of three UMPs has Controversial UMPs. That is, the number of common identical users defined by Eq. (1) is not ideal metric for three UMPs.

If we were to select (8, 8), (164, 164), (378, 378) as seed UMPs, FRUI also could not work, while *p*-FRUI could identify 7 pairs of identical users (49, 49), (71, 71), (52, 52), (273, 273), (31, 31), (81, 81), (29, 29). All of identified users are accurate. In this case, two UMPs, (48, 48) and (39, 39), are not identified. We analyze the results and find four UMPs (48, 48), (48, 39), (39, 48) and (39, 39) are Controversial UMPs. This illustrates *p*-FRUI is still not perfect.

Similarly, If we were to select (8, 8), (31, 31), (48, 48) and (52, 52) as UMPs, FRUI could not work, while *p*-FRUI could identify 8 pairs of identical users (49, 49), (39, 39), (29, 29), (71, 71), (273, 273), (81, 81), (378, 378), (164, 164). All of identified users are accurate.

Two examples show that the proposed suggestion can improve the performance of FRUI, and make FRUI more general. Compared with FRUI, *p*-FRUI increases the computational complexity. However, we think it is worthwhile to make this minor change. In most of cases, FRUI works very well, and no additional computational complexity is introduced. Only when FRUI generates Controversial UMPs, we employ *p*-FRUI to identify user, so the added computational complexity is limited. *p*-FRUI can improve the identification performance with a little bit of computational complexity.

• *The authors are with the School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China. E-mail: lyj@nwpu.edu.cn, Sutingnwpu@163.com.*

Manuscript received 19 Dec. 2017; revised 16 Apr. 2018; accepted 17 Apr. 2018. Date of current version 1 June 2018.
(Corresponding author: Yongjun Li.)
Recommended for acceptance by Y. Zhang.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TKDE.2018.2828812

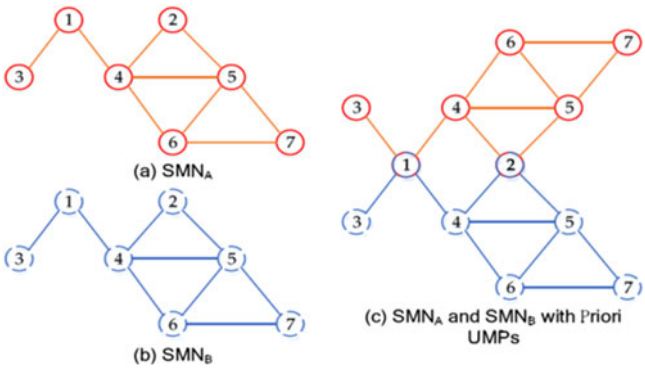
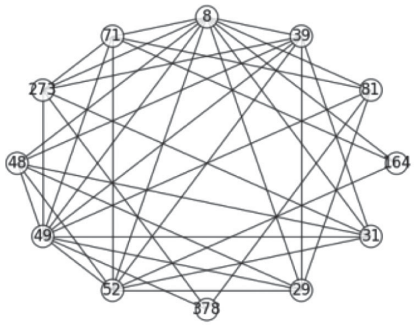


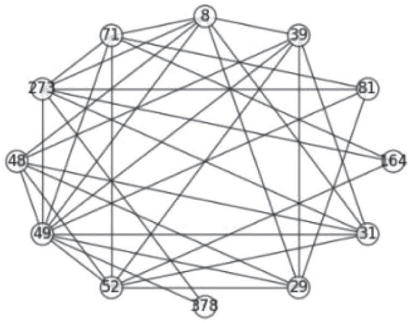
Fig. 1. Examples of two SMNs with FRUI.

TABLE 1
Controversial UMPs Generated by FRUI

M_{ij}	$M_{ij}{}^2$	Controversial UMPs ($UMP_{A\sim B}$(i,j))
2	3	(2,2), (2,6), (6,2), (6,6)
1	2	(1,1), (1,2), (1,6), (2,1), (2,7), (6,1), (6,7), (7,2), (7,6), (7,7)



(a) example of Facebook



(b) example of Twitter

Fig. 2. Examples of facebook and twitter.

3 Conclusion and Future Work

FRUI can not work very well, when the seed UMPs are not suitable. *p*-FRUI addresses this problem, and two examples illustrate the effectiveness of our suggestion. From the examples, we find that the suggested change can be a definite improvement over FRUI. In existing works, the community-enhanced [4] solution is a novel idea, and we will consider introducing this idea in our future work.

Acknowledgments

The work is partly supported by the Shaanxi Natural Science Foundation, China (2014JM2-6104).

References

[1] X Zhou, X Liang, H Zhang, et al., "Cross-platform identification of anonymous identical users in multiple social media networks," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 2, pp. 411–424, Feb. 2016.
[2] Y Li, Zh Zhang, Y Peng, et al., "Matching user accounts based on user generated content across social networks," *Future Generation Comput. Syst.*, vol. 83, pp. 104–115, 2018.
[3] X Zhou, X Liang, X Du, et al., "Structure based user identification across social networks", *IEEE Trans. Knowl. Data Eng.*, to be published, doi: 10.1109/TKDE.2017.2784430.
[4] S Nilizadeh, A. Kapadia, Y. Ahn, "Community-enhanced de-anonymization of online social networks", in *Proc. ACM SIGSAC Conf. Comput. Commun. Security*, pp. 537–548, 2014.
[5] L Lü, D Chen, X Ren, et al., "Vital nodes identification in complex networks," *Phys. Rep.*, vol. 650, pp. 1–63, 2016.