

Article

A Privacy Measurement Framework for Multiple Online Social Networks against Social Identity Linkage

Xuefeng Li ^{1,2,*} , Yixian Yang ^{1,2}, Yuling Chen ² and Xinxin Niu ^{1,2}

¹ National Engineering Laboratory for Disaster Backup and Recovery, Information Security Center, School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing 100876, China; yxyang@bupt.edu.cn (Y.Y.); xxniu@bupt.edu.cn (X.N.)

² Guizhou Provincial Key Laboratory of Public Big Data, Guizhou University, Guizhou 550025, China; ylchen3@gzu.edu.cn

* Correspondence: lixuefeng3710@gmail.com; Tel.: +86-188-1152-6536

Received: 22 August 2018; Accepted: 17 September 2018; Published: 1 October 2018



Abstract: Recently, the number of people who are members of multiple online social networks simultaneously has increased. However, if these people share everything with others, they risk their privacy. Users may be unaware of the privacy risks involved with sharing their sensitive information on a network. Currently, there are many research efforts focused on social identity linkage (SIL) on multiple online social networks for commercial services, which exacerbates privacy issues. Many existing studies consider methods of encrypting or deleting sensitive information without considering if this is unreasonable for social networks. Meanwhile, these studies ignore privacy awareness, which is rudimentary and critical. To enhance privacy awareness, we discuss a user privacy exposure measure for users who are members of multiple online social networks. With this measure, users can be aware of the state of their privacy and their position on a privacy measurement scale. Additionally, we propose a straightforward method through our framework to reduce information loss and foster user privacy awareness by using spurious content for required fields.

Keywords: multiple social networks; privacy security; measurement; profile; social identity linkage; attribute content

1. Introduction

With the progress of society, the rapid development of the Internet, and increasingly onerous work, increasing numbers of people prefer to communicate on the Internet because it is efficient and inexpensive. As a result, in recent years, Online social networks (OSNs) have expanded tremendously and emerged as an indispensable part of human life. People can chat and share messages, news, pictures, videos, and other resources via OSNs. Moreover, various types of social networks are currently used, each with its own unique features [1,2]. People make use of OSNs to various degrees according to their needs. Inevitably, people who use these sites create an online role. Additionally, for various reasons including peer pressure, conformity, lack of privacy awareness, and blind trust in social networking sites and other users, users are encouraged to disclose personally identifiable information (PII). Furthermore, social networking sites encourage users to disclose personal information so that other users can find them more easily, which promotes user stickiness [3]. It seems that users blindly trust OSN service providers to handle user data in a fair and conscientious way and to continue to do so in the future. However, the reality is that Uber acknowledged in November 2017 that for more than a year it covered up a hacking attack that stole personal information about more than 57 million customers and drivers; US officials say they'll examine claims that a data analysis firm mishandled

Facebook users' information, in order to support Donald Trump's election campaign. Facebook knew of this in the past two years, but no measures were taken; in 2017, Twitter publicly announced that it is abandoning the DNT(Do not track) privacy protection standard; according to a recent study, more than 6.05 billion pieces of personal information have been disclosed in China [4]. Thus, privacy protection ultimately depends on the users.

In the field of computer security, the basic principle of protecting privacy is preventing information from escaping its intended boundaries. However, privacy on OSNs is contrary to the goal of people using them. The only way to mitigate this paradox is to find a reasonable boundary between protecting privacy and disclosing PII. However, although people find it inherently easy to understand physical concepts, they have difficulty with virtual concepts, and privacy is a virtual concept. Initially, most OSNs offer their users privacy control, which is simple to use but limited; for example, a privacy control may enable users to set their entire profile as public, visible to friends only, or private (visible only to the user). With growing demand from users and increasing attention to privacy in the media, many OSNs (e.g., Facebook) have started offering their users more control, such as the ability to set the visibility of individual items. However, if interfaces become overly complicated, then users will not understand the settings or find them too cumbersome, and thus, they might set them in an unreasonable manner or ignore them. In a case study, Gross and Acquisti [5] show that most users do not change the default privacy settings provided by the OSN when sharing a large amount of information on their profile. In another case study, Tufecki [6] concludes that privacy-aware users are more reluctant to join social networks, but once they join, they still disclose a vast amount of information. In other words, an overwhelming majority of people have considerable difficulty understanding privacy settings [7], especially now that most users are using multiple social networks [8–11]. Moreover, many researchers focus on social identity linkage (SIL) across multiple OSNs, which is an effort to identify users from multiple heterogeneous OSNs and integrate the various networks. The compelling nature of the field has motivated many studies [12–20]; however, few people pay attention to the security issues this type of research introduces. Given this trend, malicious attackers can integrate a complete online role via profiles across multiple OSNs, and serious harm can be caused to the real individual in various ways [21]. Given this background, privacy protection on a single platform is far more than enough to manage [8–11,22,23].

At present, there is no perfect solution to this problem because users have varied requirements for privacy protection that depend on the context. Therefore, the best solution is to provide a method to quantify the privacy of individuals, transform the virtual concept of privacy into a visible physical space, help users accurately recognize the state of their privacy and help users improve their privacy. We are deeply aware that only by letting users understand their privacy leakage can we better protect user privacy. In this paper, we propose a measurement framework based on multiple OSNs to ensure that users can understand the state of their privacy and rationally adjust their privacy settings to improve it.

2. Related Work

The role of OSNs represents social relationships that exist in real life, which is called the real-life social network (RSN). The PII stored in OSNs can be modelled as an online social graph, and there is a one-to-one mapping from the RSN to the online social graph model. If multiple OSNs are used, this mapping can be found especially quickly, accurately, and inexpensively due to SIL research. Therefore, the disclosure of PII may lead to malicious attacks from the cyberspace and real world [24,25]; examples of these attacks include, but are not limited to, tracking, defamation, spamming, phishing, identity theft, profile cloning, Sybil attack, etc.

In recent years, many studies have been performed on privacy preservation via data mining and publishing; additionally, some privacy protection methods have been proposed for specific scenarios, such as RFID(Radio Frequency Identification) and smart grids [26,27], but not much has been explored related to user privacy awareness, which can be defined as the individual's awareness of the actions

and behaviours required to protect their personal information [28,29]. In contrast, SIL is developing rapidly [12–20].

In the existing research on privacy metrics, some researchers measure a single aspect. For example, Dey et al. [30] studied the amount of work that has been done regarding the harm to users caused by the disclosure of age information. Liang et al. [31] conducted an in-depth study on the privacy disclosure of a social network from the perspective of image deletion delay. Srivastava et al. [32] discussed the issue of privacy disclosure from the dissemination of character information.

Meanwhile, others have tried to solve the problem based on the overall consideration of the user. Maximilien et al. [33] discussed the concepts of attribute sensitivity and profile visibility, then evaluated these two values using a Bayesian method to evaluate privacy. Liu et al. [34] broadened their study in a different way: Using item response theory (IRT) in combination with sensitivity and visibility to provide an intuitive and mathematical approach for calculating privacy metrics of OSNs. Fang et al. [7] devised a template for privacy wizards to help users complete profile settings; however, they did not explain why. In the opinion of Zeng et al. [35], personal privacy disclosure levels are based on the protection of information that public friends disclose; they proposed a framework to assess the privacy disclosure in a community. Similarly, Li Minghui et al. [36] believed that attackers could use the background knowledge of public neighbours to obtain the privacy of victims; they used K-anonymity and L-diversity to approach this challenge, but these two approaches do not completely solve the issue. Ruggero G. Pensa et al. [23] used a circle-based definition of privacy score to measure privacy leakage and applied a learning approach to help users change privacy settings.

Currently, users are not confined to using only one social network. Thus, the above studies, which are based on a single platform, cannot accurately quantify privacy leakage. However, studies based on multiple OSNs are extremely rare and immature. Irani et al. [37] revealed that attackers can aggregate a user's PII on multiple OSNs for identity and password recovery attacks. However, they did not propose an effective method for solving this problem, only suggesting that users should disclose their PII as little as possible. Patsakis et al. [25] proposed a framework based on scenarios with multiple social networks that can achieve the goal of protecting user privacy; however, it is impossible for each social network to interact with each other in practice. Erfan et al. [38] combined attribute sensitivity and visibility, and used statistical fuzzy systems to solve the problem of privacy metrics on multiple OSNs. Nonetheless, statistical fuzzy systems, which include fewer quantitative components and more qualitative components, are difficult to use to convince people.

Over the course of our research, we noticed that many researchers are working on SIL. The results of SIL can benefit many applications, such as building interest models, providing a better view of expertise, improving social recommendations, and improving the ability to search for people across websites. However, these studies ignored various privacy and security threats, such as identity thefts and profile cloning, which can lead to compromised accounts, directed spam, phishing, online profiling by advertisers and attackers, and online stalking. Although the security issue is serious, an inevitable situation must be faced, namely, the vast number of users on social networks. Given an SIL problem instance of two social network platforms with N_1 and N_2 users, the number of all possible pairs of users to examine is given by [13]:

$$\sum_{n=1}^{\min(N_1, N_2)} \frac{N_1! N_2!}{n! (N_1 - n)! (N_2 - n)!} \quad (1)$$

where $N! = \prod_{k=1}^N k$.

However, because N represents billions of users, (1) is impossible to calculate. Existing studies have applied heuristic knowledge regarding overlapping PII, such as username, avatar, or email address, to reduce this scope. Therefore, a reasonable PII setting can be effectively prevented by SIL. Additionally, the privacy protection method we proposed is inspired by this.

As outlined above, we consider and encourage that users deliberately fill in spurious PII, especially for items that are sensitive and mandatory. The work in this paper is inspired by the privacy score defined by Liu and Terzi [34], for which we have greatly improved the measurement method in [34] to adapt it to scenarios with multiple social networks to reflect how OSNs are currently used by most users. Additionally, the discussion and usage of attribute content and privacy awareness are heuristically added to better combine the behaviour and psychology of users in social networks and more accurately measure privacy leakage. Based on these improvements, we proposed a new approach to prevent SIL. Our main contributions are as follows:

- (1) We consider the use of spurious contents to protect privacy. Therefore, we propose the quantification of an attribute's content.
- (2) We improve the method of visibility quantification. Meanwhile, we heuristically propose a quantitative method to measure a user's privacy awareness.
- (3) We use and simplify the half-suppressed fuzzy C-means clustering algorithm [39] to quantify visibility, which can still obtain an excellent result.
- (4) We found that user behaviour and consciousness are out of sync; thus, we use questionnaires to measure attribute sensitivity and real OSNs settings to calculate visibility.
- (5) We experimented with the data in a previous study [38], and original data obtained from real OSN users for comparison with the existing study.

3. Problem Descriptions and Notation

In actual scientific research, there is not an effective way to warn users how much their privacy will be exposed when PII is divulged or certain changes are made to their privacy settings. Without a practical and effective approach to quantify, measure, and evaluate privacy, it is hard for users to determine how much information they are willing to share and understand the risk involved. It is impossible for OSN service providers to make appropriate policies to protect user privacy. Meanwhile, privacy measurement, as a virtual concept, is a challenging issue because the definition of privacy is subjective. Users have different opinions and expectations about privacy. Social networking sites protect privacy by profile setting, which includes attributes consisting of structured data; privacy can be identified by this structured information, such as name, hometown, birthdate, etc. Therefore, using the attributes of a profile is a good way to measure the privacy of individuals.

However, the usage of attributes also introduces problems. Platforms vary regarding the attributes they require users to provide [1,2,40], and each attribute has a different impact on individual privacy; for example, birthdate, phone number, and address cause unequal levels of privacy leakage. Therefore, we first must determine a method of quantifying the degree of leakage for attributes.

Here, we introduce the mathematical notation we will use in the rest of our paper. We use a set of n users $\mu = \{u_1, u_2, \dots, u_n\}$ corresponding to the individuals participating in OSNs. Each user has a set of m attributes or profile items $\alpha_n = \{a_1, a_2, \dots, a_m\}$; for instance, the PII includes items such as username, gender, birthdate, and hometown. In addition, for convenience, we consider that $\beta = \{b_1, b_2, \dots, b_s\}$ corresponds to the same attribute on s OSNs. More specifically, $\beta_m^s = \{b_1, b_2, \dots, b_s\}$ corresponds to the extraction difficulty of attribute m on s OSNs.

Other general notations that are used in our framework are presented in Table 1.

Table 1. Notation.

Notation	Description
p	privacy score
μ	user of OSNs
s	number of OSNs
n	number of users
m	number of attributes
ϵ	extraction difficulty of attribute
δ	accessibility to a certain attribute
ω	individual privacy awareness
γ	reliability of attribute
v	visibility of attribute
θ	Sensitivity of attribute
λ	attribute content

4. The Measurement Method

To calculate a user’s privacy score, we need to select the attributes that affect a user’s privacy on a social network. These attributes (such as birthday, email address, and address) can be obtained from the user’s profile, messages, pictures, and status updates posted by users. It should be noted that our research does not include methods for obtaining the content of these attributes from the above sources, but rather ensures the information actually exists and can be obtained in practice.

Inspired by the privacy score defined by Liu and Terzi [34], we add a fine-grained approach for a more accurate result. We calculate ϵ , δ , ω , and γ for each attribute on each platform and then calculate visibility. Finally, we can calculate privacy score p with sensitivity θ .

4.1. Extraction Difficulty

The difficulty of extraction ϵ represents the degree of difficulty associated with obtaining the attribute contents. It is relatively easy to obtain attributes from a profile that a user provided, but it is possible that some attributes are not provided or are not required. Therefore, attributes must be obtained or inferred from messages, images, videos, and other media, which is relatively complicated. For example, it is difficult to recognize a user’s hometown in a picture or a video. To calculate the difficulty of the data extraction, we define the following formula:

$$\epsilon_m = \frac{\sum_{i=1}^s (\epsilon_m^i)}{s} \tag{2}$$

where ϵ_m^i is the value that expresses the extraction difficulty of attribute m on the *i*th platform. We sum all the values of ϵ_m^i and average it to obtain the ϵ_m that indicates the total extraction difficulty of attribute m on s social networks.

To measure ϵ_m^i , we defined three levels: 1 is difficult, 2 is relatively easy, and 3 is easy. The specific definitions are as follows: 1 represents content obtained from pictures or other approaches, 2 represents content obtained from character messages, and 3 represents content directly acquired from the profile. This approach is taken because not every platform provides all attributes in the profile. In our experimental data, we document in detail the extraction difficulty of each attribute on each OSN.

4.2. Accessibility

In general, OSN operators provide a way to protect privacy by allowing users to set accessibility (i.e., visibility to specific users) for each attribute in their profile, which includes making the attribute visible to only the user or to everyone. Accessibility signifies how many people can see the attribute content. According to popular OSN settings, we define four different levels of accessibility: 1 represents content access by only the owner of the information; 2 represents content access by friends; 3 represents

content access by a specific group of people, such as colleagues and schoolmates; and 4 represents publicly available content.

Because our research is based on multiple platforms, we calculate each attribute separately for each platform. Moreover, we consider that not all users fill in the same content for one attribute on each platform; therefore, we developed statistics of user profiles, as shown in Table 2, where 1–4 indicate the accessibility of the attribute, and A–Z represent the content of the attribute; we use letters to represent attribute content to protect user privacy. In addition, the same letter in the same attribute of different users does not represent the same content, and 0 means that we cannot access the attribute content, which means a user does not disclose any information for this attribute.

Table 2. Attribute content and accessibility.

Attribute	Platform 1	Platform 2	Platform 3	Platform s
Attribute 1	1A	3A	4B	1C
Attribute 2	2A	1A	2B	3E
Attribute 3	1A	4B	3B	2A
Attribute 4	1A	3B	0	2C
...
Attribute m				

Through Table 2, we can use Algorithm 1 to calculate the accessibility for all attributes. $\beta^{\delta,\lambda} = (1A, 3B, 1A, 0, 4B)$, for instance, is the value of an attribute, and the final result is $(1 + 3 + 4)/3 \approx 2.67$. Unlike other researchers, we consider the situation in which users fill in different content for the same attribute, which means they fill in spurious content to protect privacy. The existing algorithm can be used to achieve character consistency, while content extracted from images and long text information is manually performed; in the near future, we expect to use deep learning to remove the need for this manual step.

Algorithm 1. Calculation of accessibility

Input: $\beta_m^{\delta,\lambda}, s$
Output: δ_m
1 for i in $\beta_m^{\delta,\lambda} \in u_n \in \mu$ do
2 if i is 0 or repetitive in $\beta_m^{\delta,\lambda}$ do
3 delete i
4 $s = s - 1$
5 $\delta_m = \text{sum}(\beta_m^{\delta})/s$
6 end

Notably, different content may have the same accessibility, which causes privacy leakage. Repeated items are removed because they do not provide additional privacy losses, i.e., they are redundant. Values of 0 are excluded to address cases like $(4A, 0, 0, 0, 0)$, where if we do not rule out 0, the result is 2, which is unreasonable because this attribute has been fully disclosed. The accessibility should not be lower than $(3A, 3B, 3C, 0, 0)$. The average is used to compare the differences between $(4A, 4B, 4C, 4D, 4E)$ and $(4A, 4A, 4A, 4A, 4A)$; if this approach is not used, the former result is 16 and the latter is 4, although the accessibility should be exactly the same.

4.3. Reliability

Attribute reliability quantization, one of the most important aspects, determines whether privacy metrics are accurate because spurious content does not have the same impact as real content.

Previous research showed that people are willing to share real information with others on social platforms. Therefore, [28] adopts the following strategy: As the number of sources of disclosure

increases, reliability increases. Because we considered attribute content and users are more likely to fill in real PII, for our improvement, we use the maximum number of the repeated content to measure reliability.

While collecting and processing our experimental data, we analysed the data to comprehend user behaviour. We found that content used to fill an attribute for various platforms is very diverse in terms of reliability. In our 279 samples, the content reliability of several platforms is shown in Table 3, from which we can see that the reliability on a single platform is much less than 50%, but has a linear growth rate when the same content is provided on more platforms. Then, as more platforms show the same content, growth will slow accordingly so that the global trend forms an S-curve.

Table 3. Attribute reliability.

Platforms	Attribute Reliability
1	0.32
2	0.67
3	0.91
4	0.98
5	0.99

It can be seen from the curve in Figure 1 that the function used by Erfan [38], (3), does not fit the actual situation well. Meanwhile, Liu and Terzi [34] use the item response theory (IRT) theory model to calculate the overall privacy score without considering reliability.

$$\gamma = \frac{2}{1+e^{-q}} - 1 \tag{3}$$

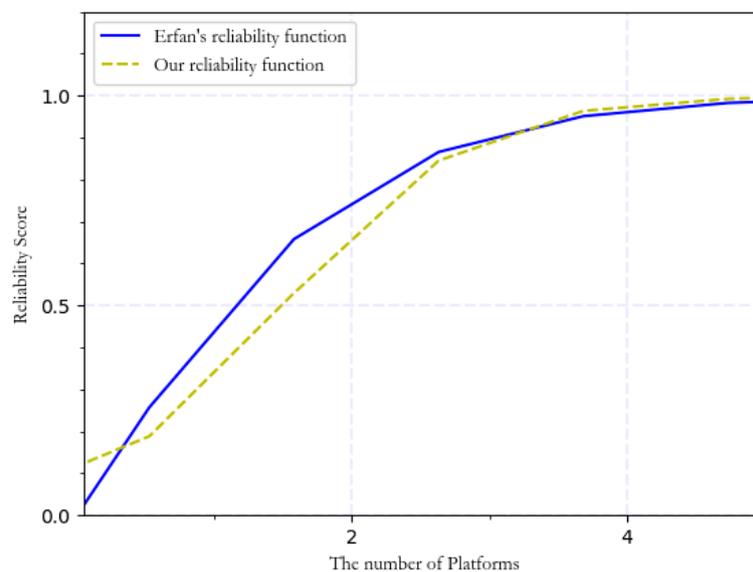


Figure 1. Data reliability.

Because IRT (also known as latent trait theory, strong true score theory, or modern mental test theory) is a paradigm for the design, analysis, and scoring of tests, questionnaires, and similar instruments that measure abilities, attitudes, or other variables [41], we use IRT as our theoretical basis to design the function; the original IRT function is shown in (4). First, we set the value of b to $3/2$ based on Table 3. Because the difference in reliability is largest between 1 and 2, the inflection point is between 1 and 2. In addition, q is the number of platforms that provide the same attribute content the maximum number of times; then, we use the least square method and the data in Table 3

to fit parameter a , and finally construct the function, as shown in (5). The function curve is shown in Figure 1.

$$\gamma = \frac{1}{1+e^{-a(s-b)}} \tag{4}$$

$$\gamma = \frac{1}{1+e^{-1.5(s-1.5)}} \tag{5}$$

The range of the function is (0, 1). The larger the value of s , the higher the reliability.

4.4. Privacy Awareness

In the measurement of privacy leakage, privacy awareness refers to the users' knowledge and understanding of the privacy options available to them on the social networking site. Users with higher privacy protection awareness usually hide sensitive information or fill in spurious attribute content to confuse attackers and increase the difficulty of malicious behaviour. Therefore, it is necessary to measure privacy awareness (although it was not considered in previous studies).

Our approach for measuring privacy awareness is counting the number of different attributes that users choose to fill with the same content on multiple OSNs. For example, user λ filled in (A, B, A, C, B) for an attribute, in which at least two of the values are spurious. Moreover, privacy leakage is minimal in the case (0, 0, 0, 0, 0), which indicates that the user has not disclosed any information about this attribute in the profile or other media. The higher the user's privacy awareness, the lower the possibility that their privacy is exposed. To measure privacy awareness, we used the following function:

$$\omega = 2 - \frac{2}{1 + e^{-0.5q}} \tag{6}$$

We design the function in this manner because it is a 3-parameter logistic model (3PL), which is a variant of IRF (item response function). IRF provides the probability that a person with a given ability level will answer correctly. People with lower ability have less chance of answering correctly, while persons with high ability are very likely to answer correctly [42]; thus, this method can be an excellent expression of the human sense of privacy.

As seen from the function curve in Figure 2, the greater the number of instances of spurious content, the greater the difficulty for a malicious attacker to attack and the lower the user privacy leakage, which expresses stronger user privacy awareness.

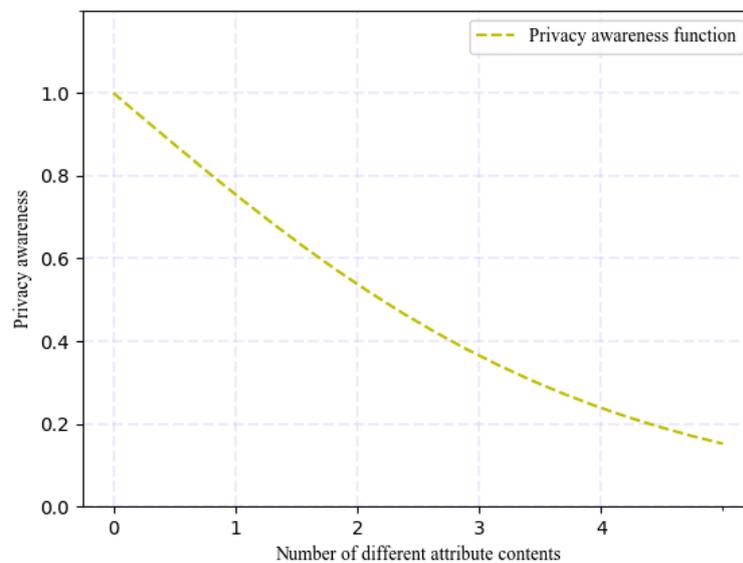


Figure 2. Privacy awareness.

4.5. Visibility

Through the abovementioned work, we can quantify extraction difficulty, accessibility, reliability, and privacy awareness. To calculate visibility, we carefully choose to use the half-suppressed fuzzy C-means clustering algorithm [39], which is a clustering algorithm based on FCM (Fuzzy C-means Algorithm).

However, the original algorithm has great time complexity and space complexity, so we simplified it to improve these problems. First, we remove the step for training an SVM (Support Vector Machine) because our input is only four dimensions and removing the SVM can enormously increase efficiency; meanwhile, the clustering result is sufficient. Second, we change the iteration of the algorithm: If the last step determines that the iteration does not end, we restore the new cluster centre to our pre-set cluster centre.

We chose the half-suppressed fuzzy C-means clustering method because it can achieve sufficient results using very few training samples, especially when considering the great need for an open large-scale sample set and data with unique needs such as ours. In addition, this method divides the sample space into several categories according to the value of the sample cluster membership to convert the scoring problem into a classification problem that uses an A–F rating instead of the hundred-mark system, as in educational systems, which can yield better generalization.

Moreover, the algorithm can manually specify the clustering centre, which is extremely easy to perform in our research. According to common sense, we can formulate a perfect clustering centre that is similar to the difference between (1A, 1B, 1C, 1D, 1E) and (4A, 4A, 4A, 4A, 4A). When we specify the cluster centre, we can determine the visibility of a sample using the cluster centre to which the sample is aggregated.

In a sample space of algorithm, $X = \{x_j | j = 1, 2, \dots, n; x_j \in R^p\}$ is a vector of n dimensions. $v_i (i = 1, 2, \dots, c)$ represents the center of each cluster. u_{ij} is the membership of sample x_j belonging to the class which satisfies:

$$\begin{aligned} \sum_{i=1}^c u_{ij} &= 1; j = 1, 2, \dots, n, \\ u_{ij} &\geq 0; i = 1, 2, \dots, c; j = 1, 2, \dots, n, \\ 0 < \sum_{j=1}^n u_{ij} &< n; i = 1, 2, \dots, c. \end{aligned}$$

The specific process of simplified algorithm is as follows:

- (1) Initialize cluster centers $v_i^{(0)}$, the inhibiting factor is α , inhibition threshold is β , prime index factor is m , error threshold is $\varepsilon > 0$ and the maximum number of iterations is K , set the number of iterations $k = 0$.
- (2) Use $u_{ij} = \frac{(\frac{1}{\|x_j - v_i\|^2})^{\frac{1}{m-1}}}{\sum_{k=1}^n (\frac{1}{\|x_j - v_k\|^2})^{\frac{1}{m-1}}}$ to calculate $U^{(k)} = [u_{ij}^{(k)}]$.
- (3) According to the above correction equation to get $U^{(k)'}$ by fuzzy classification matrix u .
- (4) Calculate new cluster center from $v_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m}$ and $U^{(k)'}$.
- (5) If $\left(\sum_{i=1}^n \|v_i^{(k+1)}\|^2\right)^{\frac{1}{2}} < \varepsilon$ or $k < K$, the iteration is over; Otherwise $k = k + 1$, $v_i = v_i^{(0)}$, return to the step 2.

In this paper, we use accessibility, reliability, extraction difficulty, and privacy awareness as the four-dimensional sample input; then, we can obtain the visibility using the class into which the data are clustered. After we conduct a number of experiments, the better model parameters obtained are as follows: The inhibiting factor was 0.5, the inhibiting threshold was 0.5, the index prime factor

$m = 2$, the error threshold $D = 0.01$, and the maximum number of iterations $K = 30$. We used the above parameters in the subsequent comparative analysis and set the outcome to six categories. The initial clustering centres have preset values.

4.6. Sensitivity

On social networking sites, there are far more attributes that affect user privacy than those we surveyed, and choosing reasonable attributes is crucial to the results. When calculating the final privacy score, the influence of various attributes on privacy leakage must be considered. This influence is called the sensitivity of the attribute.

In previous studies, Erfan et al. [38] used 11 attribute sensitivities calculated by Srivastava et al. [32], who used the naïve approach proposed by Liu and Terzi [34]. However, we do not adopt this approach. They used the profile data that a user actually fills in to calculate the sensitivity, which is unreasonable; in real life, due to the complexity of settings, the actual settings are not consistent with a user’s expectation, which means that using the profile setting cannot reflect the actual sensitivity of attributes.

In our study, we launched an online questionnaire about privacy sensitivity (<https://www.wjx.cn/report/1647730.aspx>). We divided attribute sensitivity into five levels: L1, not worried at all; L2, not worried; L3, no idea; L4, worried; L5, extremely worried. L1–L5 are given as the percent of people who select that level. Excluding the large number of duplicate questionnaires, we obtained 364 valid questionnaires. Then, we calculated sensitivity using (7); the results are shown in Table 4. We set L4 as the benchmark for sensitivity and use a weight coefficient adjustment for L3 and L5.

$$\theta = \frac{0.5 * L3 + L4 + 1.5 * L5}{1.5} \tag{7}$$

The other difference between our approach and that of Srivastava et al. [32] is that we add the username and avatar to guard against SIL, which is not considered in the previous work. Meanwhile, because our research background is in China, we abandon the attribute of religious views and political views.

Table 4. Attribute sensitivity.

Attribute	Sensitivity
Username	0.2381
Avatar	0.3553
Phone number	0.5669
E-Mail	0.3260
Address	0.4212
Birthdate	0.2748
Hometown	0.2253
Job Details	0.2024
Relationship Status	0.1731
Interests	0.1255
Education	0.1575

4.7. Privacy Score

Formula (8) can be used to calculate a user’s privacy score through visibility and sensitivity.

$$p = \frac{\sum_{i=1}^m v_i * \theta_i}{m} \tag{8}$$

The higher privacy score p , the more severe privacy leakage.

5. Experimental Evaluation

To fully prove the advantages of our method, our experiment is divided into three parts. The data in [38] are used in the first part. The second part uses the data obtained through our research. Then, we use these data to compare the performance of the other two existing methods. Finally, we illustrate methods of improving user privacy through our method and prove that our method is effective at preventing SIL.

5.1. Experiment 1

In [38], the authors collected the data of 15 users (represented as user a to user o in the original paper) who were involved in four different online social networks (i.e., Facebook, ResearchGate, LinkedIn, and Google+); 11 attributes for each user were used to measure the information disclosure and privacy risk of those users. They reported that the chosen number of users covers a diverse range of values from user profiles, which is needed to show the effectiveness of the proposed privacy scoring method. To guarantee the fairness of the comparison, we use the data of two of the users (represented as user b and user o in the original paper), which were fully published in their paper, and the sensitivity they reported. The data of the remaining 13 users is not public. Study [38] compared the privacy disclosure score of all users with that provided by the privacy scoring model by Liu & Terzi [34] to evaluate the superiority of their proposed model; we also include the model of Liu & Terzi [34] in our experiment.

Figure 3 shows the normalized privacy score calculated using the three methods. Because user1 has a higher accessibility for most attributes and a lower extraction difficulty, the privacy score of this user is higher than that of the others, which means privacy leakage is severe. The result of Liu et al. [34] is lower because extraction difficulty and reliability have not been considered. Moreover, only binary values are used to represent accessibility, where 1 means accessible and 0 means inaccessible; therefore, the final value is low. When our method and Erfan's method [38] are compared, the final value in our method is relatively low, as Erfan's method [38] does not consider attribute content. Users may use spurious attribute content on a highly accessible platform and fill in real information on a more confidential platform. Our method outputs a higher score for user2 because user2 fills in the same content for most attributes (see Table 5), which suggests that the privacy awareness of user2 is very poor. Although extraction difficulty and accessibility are high, peep screen, Trojans, and phishing attacks cannot be prevented; thus, accessibility is not as effective as imagined. Therefore, we believe that the privacy leakage of user2 is not as optimistic as shown by other methods.

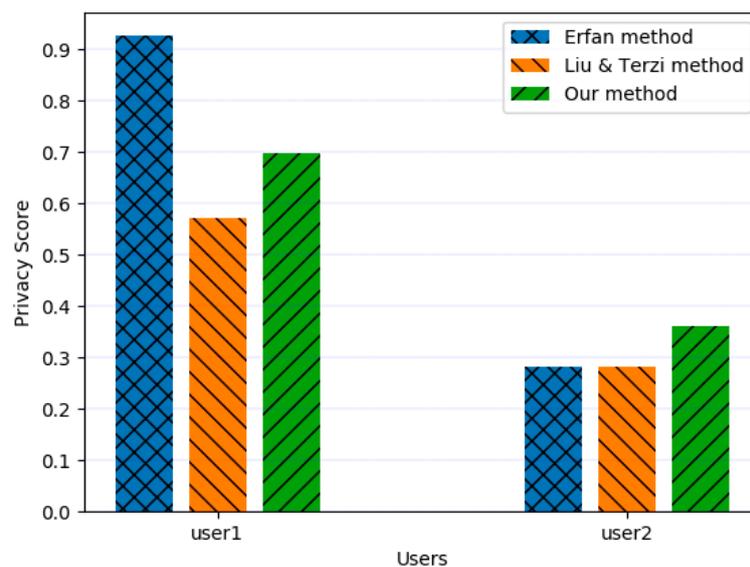


Figure 3. Experiment 1 privacy score.

Table 5. Attribute accessibility and content.

Attribute	User1 Accessibility	A	B	User2 Accessibility	A	B
Contact number	2A,2B,4A,3A	2.75	2.75	1A,1A,1A,2B	2.75	1.5
E-Mail	2A,2A,3B,4B	2.75	3	1A,1A,1A,2B	1.25	1.5
Address	2A,2A,2A,3B	2.5	2.5	2A,1B,2A,1B	1.5	1.5
Birthdate	3A,2A,3A,4A	3	3	1A,1A,1A,1A	1	1
Hometown	3A,2B,3A,4C	3	3	2B,1A,1A,1A	1.25	1.5
Current town	3A,3A,2A,4A	3	3	3A,1B,2A,1B	1.75	2
Job Details	2A,4A,4A,4A	3	3	2A,1A,4B,1A	3	2.33
Relationship Status	3A,2A,2A,4B	2.75	3	2A,1A,1A,1A	1.25	1.5
Interests	3A,3A,2A,3B	2.75	2.67	2A,1A,3B,1A	1.75	2
Religious Views	3A,2A,2A,4A	2.75	3	1A,1A,1A,1A	1	1
Political	2A,2A,1A,1A	1.5	1.5	1A,1A,1A,1A	1	1

It is worth noting that we considered the effect of the attribute content, which is not considered in other research; thus, we include attribute content to support our method (A is the method in [38] and B is our method). The final result is shown in Table 6.

Table 6. Visibility and privacy score.

Attribute	User1 Visibility (A)	User1 Visibility (B)	User2 Visibility (A)	User2 Visibility (B)	User1 Score	User2 Score
Contact number	7.32	4	1.5	3		
E-Mail	7.32	4	1.52	3		
Address	7.3	4	4.11	2		
Birthdate	7.32	6	0	2		
Hometown	7.32	4	1.5	3	(A)	(A)
Current town	7.32	6	4.11	3	2.582	0.790
Job Details	7.32	6	7.32	3	(B)	(B)
Relationship Status	7.32	5	1.5	3	1.592	0.825
Interests	7.32	5	4.11	3		
Religious Views	7.32	5	0	1		
Political	4.11	2	0	1		

5.2. Data Collection

In our second experiment, we used the dataset we collected. We choose QQ, Sina Microblog, Tencent Microblog, and Kaixin.com as our experimental platforms. These four platforms are the most widely used OSNs in China. To record the user’s multiplatform data and ensure data authenticity, as shown in Figure 4, we conducted a random survey of 279 people in our school, including undergraduates, graduate students, doctoral students, teachers, and staff, and performed statistical analysis on the profiles of the frequently used social networking sites. Subsequently, with permission and by guaranteeing not to divulge personal information, we asked those surveyed to log into their social network account, open the profile page, and truthfully record their personal information. The specific data format is similar to the data in Table 2. However, it is worth noting that letters A–Z represent different content in different user’s profiles, even if they are the content of the same attribute.

As expected, not every user used all four social networks. Only 28 of 279 people used the four social platforms. We performed a more detailed statistical analysis on these 28 users, including profile content, messages, pictures, and videos. The data of the other 251 users are used as samples to train the clustering algorithm.

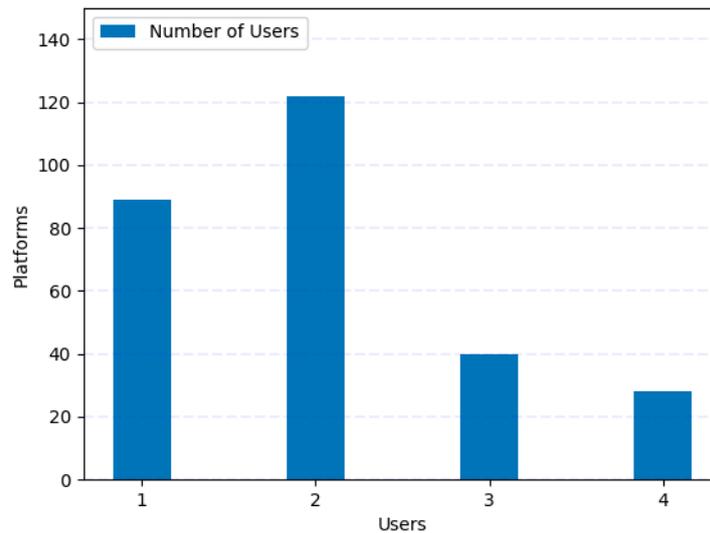


Figure 4. User data statistics.

5.3. Experiment 2

Due to space limitations in this paper, we selected seven students with great diversity from the 28 users, as shown in Figure 5, to show the detail in Table 7. In this experiment, we use the sensitivity in Table 4.

Table 7. Attribute content and accessibility.

	User1	User2	User3	User4	User5	User6	User7
Username	4A,4A,4A,4B	4A,4B,4A,4A	4A,4A,4A,4A	4A,4A,4B,4C	4A,4B,4B,4B	4A,4A,4A,4A	4A,4A,4A,4A
Avatar	4A,4A,4A,4B	4A,4B,4C,4A	4A,4A,4B,4B	4A,4A,4B,4B	4A,4A,4B,4A	4A,4B,4B,4B	4A,4A,4A,4A
Phone number	2A,0, 4B,1C	2A,0,4B,1A	0,0,0,0	2A,0,4B,2A	0,0,0,1A	4A,0,4A,2A	4A,4A,4A,4A
E-Mail	4A,2B, 1B,1B	0,1A, 1A,2B	4A,2B,0,1A	4A,1A,1A,2A	4A,4A,4B,2B	4A,2A,4A,4A	4A,4A,4A,4B
Address	2A,2A,1A,1A	2A,2A,0,1B	1A,1A,1B,0	2A,2A, 4B,1B	2A,2B, 0,2A	2A,4A, 4B,2B	2A,2A,4A,4B
Birthdate	4A,4A,4B,4A	4A,4B,4C,2D	2A,2A, 4B,2B	4A,1B, 4A,2B	4A,4B,0,2C	4A,4A,4A,4A	4A,4A,4A,4A
Hometown	4A,4B,4B,2B	4A,2B,4A,2B	4A,2B,0,2A	4A,0,0,2A	4B,2A,4A,2A	4A,2A,4B,4A	4A,4A,4A,2A
Job Details	4A,2B,4A,2B	2A,4A,2B,2A	1A,1A,0,2B	2A,2A,1A,2B	4A,2A,4A,2B	4A,4A,4B,4A	4A,4A,4A,4A
Relationship Status	0,4A,0,2A	2A,4A,0,2A	0,1A, 0,1A	0,2A, 0,2A	0,2A,2A,2A	0,0,0,2A	0,2A,0,4A
Interests	4A,4B,4A,2B	4A,4B,4C,2B	0,0,0,2A	0,4A,4B,2C	0,4A,2B,4C	0,4A,4A,2B	4A,4A,4A,4A
Education	0,4A,1A,2A	2A,0,1A,2B	2A,0, 1A,0	4A,4A, 1A,2A	4A,0,1A,2A	4A,4A,4A,4A	4A,4A,4A,4A

As shown in Figure 5, we counted the privacy scores of the 28 users using the four social networks mentioned in the previous section; the blue “●” indicates the seven users selected in our experiment, and the yellow “×” indicates the remaining 21 users. Moreover, the privacy score has been normalized so that we can see that the majority are higher than 0.5. Although it is inevitable that partial privacy will be compromised, people’s privacy is very serious and more extensive attention should be paid to it. Notably, these results are based on members of a university where user privacy awareness is generally high; thus, it can be expected that the privacy status in other places is more worrisome.

Figure 6 shows the score of seven people under the three methods. It can be seen that the score calculated by our method is lower in most cases because we consider cases of users providing false content, in particular user2 and user5, who prefer to provide different content on different platforms. In addition, since user7 provides the same content on most platforms, we believe that user7’s privacy leakage is more serious than that calculated by the other two methods.

To reflect the advantages of our approach on multiple social networks, we calculated privacy scores on both two social platforms and three social platforms in our samples for comparison.

As shown in Figure 7, scores increase with the number of platforms, which means that the more platforms the users use, the more privacy exposure occurs. Liu et al. [34] do not consider the multiplatform scenario, and so the score in their method does not change much. Meanwhile,

because the reliability and privacy awareness of our method will increase as the number of platforms increases, users may obtain a lower score when they use more platforms, which means that malicious attackers will encounter more difficulty and higher costs in the implementation of violations, thus making the privacy status of users more optimistic.

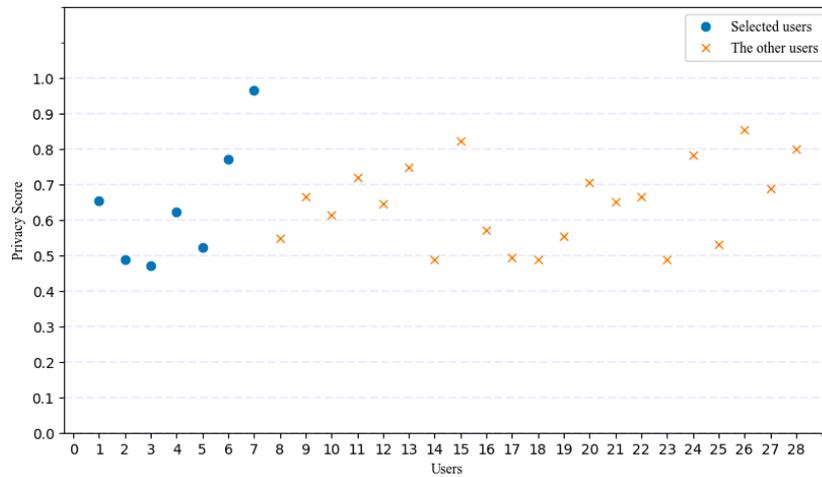


Figure 5. User privacy score statistics.

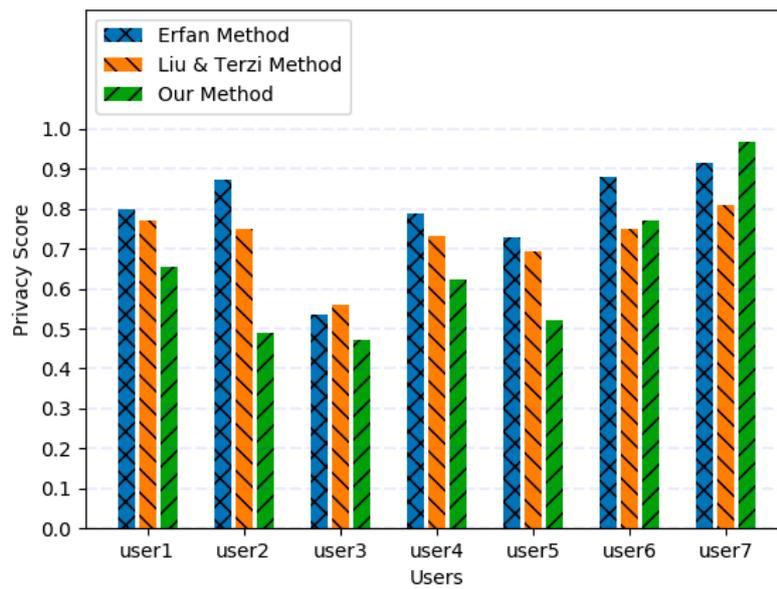


Figure 6. Experiment 2 privacy score.

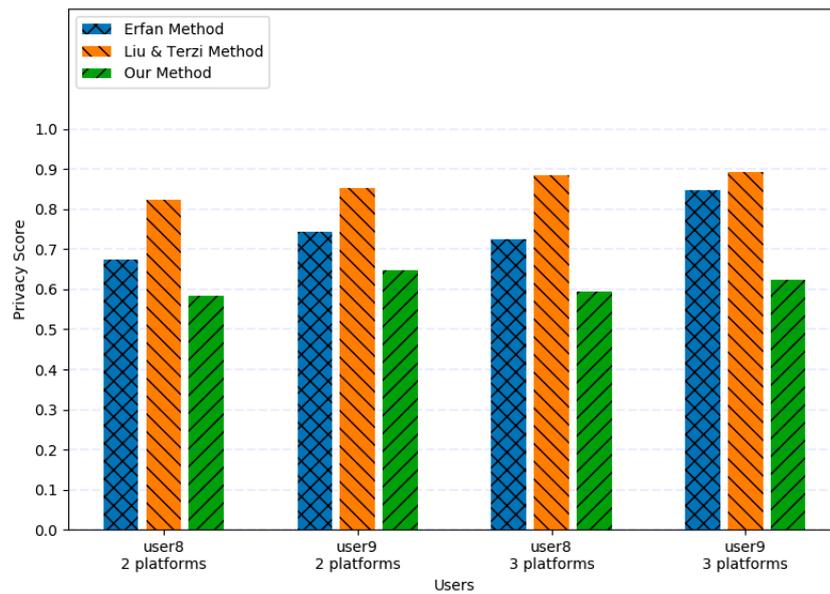


Figure 7. Privacy score on fewer online social networks (OSNs).

5.4. Experiment 3

In this section, we will use User7 as an example to show how users reasonably change profile settings to reduce privacy leakage and guard against SIL using our method.

In Figure 8, we can see that the privacy score of each attribute significantly decreases after the settings change based on Table 8. Here, we provide an example of how to reduce privacy leakage using our method. Because people have varying requirements for privacy protection in the real world, there is no universal protection method. The fundamental purpose of the privacy score in this paper is to stimulate and cultivate the privacy awareness of users to ensure that the privacy of users is not violated from the beginning. Users can intuitively understand their privacy according to the privacy score.

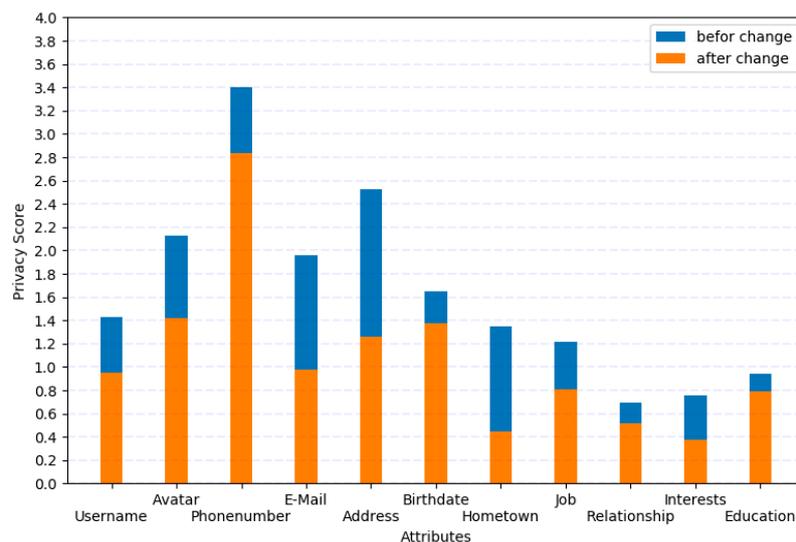


Figure 8. Privacy score after profile changes.

Table 8. Profile change.

	User7 (Before Change)	User7 (After Change)
Username	4A,4A,4A,4A	4A,4B,4C,4A
Avatar	4A,4A,4A,4A	4A,4B,4A,4C
Phone number	4A,4A,4A,4A	2A,2A,4B,2A
E-Mail	4A,4A,4A,4B	4A,2B,2C,4B
Address	2A,2A,4A,4B	2A,2A,2B,4C
Birthdate	4A,4A,4A,4A	2A,4A,4B,4A
Hometown	4A,4A,4A,2A	2A,1B,0,2A
Job Details	4A,4A,4A,4A	2A,2A,2A,4B
Relationship Status	0,2A,0,4A	0,2A,0,4A
Interests	4A,4A,4A,4A	2A,2A,4B,2C
Education	4A,4A,4A,4A	4A,2A,2A,2A

To demonstrate the effectiveness of our approach to prevent SIL, we chose three methods of SIL that use profile matching [16,17,20] (represented by method 1, method 2, and method 3). First, we formed a dataset by randomly crawling profiles on Sina Microblog; this dataset includes 2000 profiles. Then, we put the profiles of the 28 people in our study into the Sina Microblog and Kaixin.com dataset. Finally, we use their Tencent Microblog and QQ profiles as testing data to match the 2056 profiles in the dataset.

In this experiment, we ran each method 1000 times and recorded the precision of successful matches. Moreover, the changes of the profiles in our testing data settings did not deliberately differentiate the testing data from the dataset. Therefore, some profiles can still be matched after changing the settings. In Figure 9, because most people do not deliberately set different usernames, avatars, and other highly recognizable attributes on different OSNs, the possibility of a user being successfully matched is very high. After changing the settings, this probability is greatly reduced.

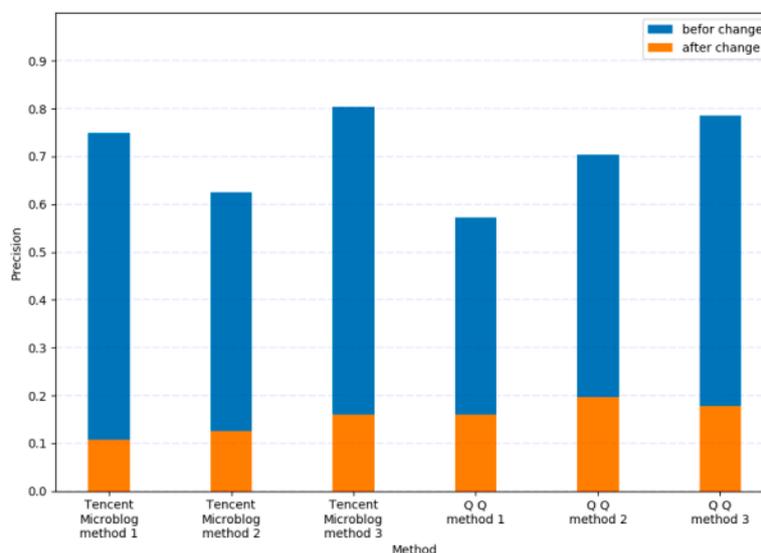


Figure 9. Social identity linkage (SIL) precision comparison.

6. Conclusions and Future Work

With the growing number of users and the influence of social networks, the protection of privacy is an urgent problem that needs to be solved. People are no longer satisfied with a single social network, but different social networks expose different aspects of PII according to their purposes. These PII can be used to integrate a user’s real identity, which can lead to serious harm. Methods of avoiding such harm are challenging to develop. In this paper, we consider accessibility, extraction difficulty,

reliability, and privacy awareness. Then, we use the simplified half-suppressed fuzzy C-means clustering algorithm to calculate visibility; using sensitivity we can calculate a final privacy score for users in a multiplatform scenario. Through the privacy score, a user's personal information disclosure status, including the degree of leakage of various attributes, can be seen visually. Finally, users can choose which attributes to hide or disclose according to their expectations and overall privacy status.

Nevertheless, in our experimental data, few people simultaneously use four social networks; thus, we take advantage of the data of users that use fewer social platforms to train the clustering algorithm. Although the result is still a considerable achievement, if we use more data from multiple social networks, the result can be further improved. In future research, we will collect more comprehensive data to optimize our results. Meanwhile, because a social network is a large, converged community, we will consider the privacy score of friends in a community in our future work.

Author Contributions: X.L. put forward privacy measurement framework and the innovation method, Y.C. implemented the graph with Python, X.L. wrote the original manuscript and X.N. and Y.Y. improved the writing.

Funding: This work was supported in part by the National Key R&D Program of China under Grant 2017YFB0802300, Foundation of Guizhou Provincial Key Laboratory of Public Big Data under Grant 2017BDKFJJ015 and in part by Major Scientific and Technological Special Project of Guizhou Province under Grant 20183001.

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

References

1. He, Z.; Cai, Z.; Yu, J. Latent-data privacy preserving with customized data utility for social network data. *IEEE Trans. Veh. Technol.* **2018**, *67*, 665–673. [CrossRef]
2. He, Z.; Cai, Z.; Wang, X. Modeling propagation dynamics and developing optimized countermeasures for rumor spreading in online social networks. In Proceedings of the 35th IEEE International Conference on Distributed Computing Systems, (ICDCS), Columbus, OH, USA, 29 June–2 July 2015; pp. 205–214.
3. Ge, J.; Peng, J.; Chen, Z. Your privacy information are leaking when you surfing on the social networks: A survey of the degree of online self-disclosure (DOSD). In Proceedings of the 2014 IEEE 13th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC), London, UK, 18–20 August 2014; pp. 329–336.
4. Jiang, J. Personal Information Leakage on The Rise in China: Report. 2017. Available online: <http://en.people.cn/n3/2017/0331/c90000-9197748.html> (accessed on 31 March 2017).
5. Gross, R.; Acquisti, A. Information revelation and privacy in online social networks. In Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society, WPES'05, Alexandria, VA, USA, 7 November 2005; ACM: New York, NY, USA, 2005; pp. 71–80.
6. Tufekci, Z. Can you see me now? Audience and disclosure regulation in online social networksites. *Bull. Sci. Technol. Soc.* **2008**, *28*, 20–36. [CrossRef]
7. Fang, L.; LeFevre, K. Privacy wizards for social networking sites. In Proceedings of the 19th international Conference on World Wide Web, Raleigh, NC, USA, 26–30 April 2010; pp. 351–360.
8. Cai, Z.; He, Z.; Guan, X.; Li, Y. Collective data-sanitization for preventing sensitive information inference attacks in social networks. *IEEE Trans. Depend. Sec. Comput.* **2018**, *15*, 577–590. [CrossRef]
9. Han, M.; Li, J.; Cai, Z.; Han, Q. Privacy reserved influence maximization in GPS-enabled cyber-physical and online social networks. In Proceedings of the 9th IEEE International Conference on Social Computer Networking, Sustainable Computing and Communications, Atlanta, GA, USA, 8–10 October 2016; pp. 284–292.
10. He, Z.; Cai, Z.; Yu, J.; Wang, X.; Sun, Y.; Li, Y. Cost-efficient strategies for restraining rumor spreading in mobile social networks. *IEEE Trans. Veh. Technol.* **2017**, *66*, 2789–2800. [CrossRef]
11. Zheng, X.; Cai, Z.; Yu, J.; Wang, C.; Li, Y. Follow but no track: Privacy preserved profile publishing in cyber-physical social systems. *IEEE Internet Things J.* **2017**, *4*, 1868–1878. [CrossRef]

12. 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 August 2015; pp. 1485–1494.
13. Liu, S.; Wang, S.; Zhu, F.; Zhang, J.; Krishnan, R. Hydra: Large-scale social identity linkage via heterogeneous behavior modeling. In Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, Snowbird, UT, USA, 22–27 June 2014; pp. 51–62.
14. Wang, Y.; Feng, C.; Chen, L.; Yin, H.; Guo, C.; Chu, Y. User identity linkage across social networks via linked heterogeneous network embedding. In *World Wide Web*; Springer: Berlin, Germany, 2018; pp. 1–22.
15. Mu, X.; Zhu, F.; Lim, E.P.; Xiao, J.; Wang, J.; Zhou, Z.H. User identity linkage by latent user space modelling. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13 August 2016; pp. 1775–1784.
16. Iofciu, T.; Fankhauser, P.; Abel, F.; Bischoff, K. Identifying Users across Social Tagging Systems. In Proceedings of the 5th International AAAI Conference on Weblogs and Social Media, ICWSM, Barcelona, 17–21 July 2011.
17. Vosecky, J.; Hong, D.; Shen, V.Y. User identification across multiple social networks. In Proceedings of the First International Conference on Networked Digital Technologies, Ostrava, Czech Republic, 28–31 July 2009; pp. 360–365.
18. Nie, Y.; Jia, Y.; Li, S.; Zhu, X.; Li, A.; Zhou, B. Identifying users across social networks based on dynamic core interests. *Neurocomputing* **2016**, *210*, 107–115. [[CrossRef](#)]
19. Shu, K.; Wang, S.; Tang, J.; Zafarani, R.; Liu, H. User identity linkage across online social networks: A review. *ACM SIGKDD Explor. Newslett.* **2017**, *18*, 5–17. [[CrossRef](#)]
20. Zafarani, R.; Liu, H. Connecting corresponding identities across communities. In Proceedings of the ICWSM 2009, San Jose, CA, USA, 17–20 May 2009; pp. 354–357.
21. Krishnamurthy, B.; Wills, C. Privacy diffusion on the web: A longitudinal perspective. In Proceedings of the 18th international conference on World Wide Web, Geneva, Switzerland, 1 April 2009; pp. 541–550.
22. He, Z.; Cai, Z.; Han, Q.; Tong, W.; Sun, L.; Li, Y. An energy efficient privacy-preserving content sharing scheme in mobile social networks. *Pers. Ubiquitous Comput.* **2016**, *20*, 833–846. [[CrossRef](#)]
23. Pensa, R.G.; Di Blasi, G. A privacy self-assessment framework for online social networks. *Expert Syst. Appl.* **2017**, *86*, 18–31. [[CrossRef](#)]
24. Lam, I.F.; Chen, K.T.; Chen, L.J. Involuntary information leakage in social network services. In *International Workshop on Security*; Springer: Berlin, Germany, 2008; pp. 167–183.
25. Patsakis, C.; Zigomitos, A.; Papageorgiou, A.; Galván-López, E. Distributing privacy policies over multimedia content across multiple online social networks. *Comput. Netw.* **2014**, *75*, 531–543. [[CrossRef](#)]
26. Gope, P.; Lee, J.; Quek, T.Q. Lightweight and Practical Anonymous Authentication Protocol for RFID Systems Using Physically Unclonable Functions. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 2831–2843. [[CrossRef](#)]
27. Gope, P.; Sikdar, B. An Efficient Data Aggregation Scheme for Privacy-Friendly Dynamic Pricing-based Billing and Demand-Response Management in Smart Grids. *IEEE Internet Things J.* **2018**, *5*, 3126–3135. [[CrossRef](#)]
28. Wang, K.; Chen, R.; Fung, B.; Yu, P. Privacy-preserving data publishing: A survey on recent developments. *ACM Comput. Surv.* **2010**, *42*. [[CrossRef](#)]
29. Yin, D.; Shen, Y.; Liu, C. Attribute Couplet Attacks and Privacy Preservation in Social Networks. *IEEE Access* **2017**, *5*, 25295–25305. [[CrossRef](#)]
30. Dey, R.; Tang, C.; Ross, K.; Saxena, N. Estimating age privacy leakage in online social networks. In Proceedings of the INFOCOM, Orlando, FL, USA, 25–30 March 2012; pp. 2836–2840.
31. Liang, K.; Liu, J.K.; Lu, R.; Wong, D.S. Privacy concerns for photo sharing in online social networks. *IEEE Internet Comput.* **2015**, *19*, 58–63. [[CrossRef](#)]
32. Srivastava, A.; Geethakumari, G. Measuring privacy leaks in online social networks. In Proceedings of the 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Mysore, India, 22–25 August 2013; pp. 2095–2100.
33. Maximilien, E.M.; Grandison, T.; Sun, T.; Richardson, D.; Guo, S.; Liu, K. Privacy-as-a-service: Models, algorithms, and results on the Facebook platform. In Proceedings of the Web 2.0 Security and Privacy Workshop, Oakland, CA, USA, 21 May 2009; Volume 2.
34. Liu, K.; Terzi, E. A framework for computing the privacy scores of users in online social networks. *ACM Trans. Knowl. Discov. Data* **2010**, *5*, 6. [[CrossRef](#)]

35. Zeng, Y.; Sun, Y.; Xing, L.; Vokkarane, V. Trust-aware privacy evaluation in online social networks. In Proceedings of the 2014 IEEE International Conference on Communications (ICC), Sydney, Australia, 10–14 June 2014; pp. 932–938.
36. Li, M.; Liu, Z.; Dong, K. Privacy Preservation in Social Network against Public Neighborhood Attacks. In Proceedings of the Trustcom/BigDataSE/I SPA, Tianjin, China, 23–26 August 2016; pp. 1575–1580.
37. Irani, D.; Webb, S.; Li, K.; Pu, C. Modeling unintended personal-information leakage from multiple online social networks. *IEEE Internet Comput.* **2011**, *2011*, 13–19. [[CrossRef](#)]
38. Erfan, A.; Garg, S.; Gao, L.; Yu, S.; Montgomery, J. Scoring Users' Privacy Disclosure Across Multiple Online Social Networks. *IEEE Access* **2017**, *5*, 13118–13130.
39. Zhao, Q.H.; Ha, M.H.; Peng, G.B.; Zhang, X.K. Support vector machine based on half-suppressed fuzzy c-means clustering. In Proceedings of the 2009 International Conference on Machine Learning and Cybernetics, Baoding, China, 12–15 July 2009; Volume 2, pp. 1236–1240.
40. Clearinghouse, Privacy Rights. Social Networking Privacy: How to Be Safe, Secure and Social. Available online: <https://www.privacyrights.org/consumer-guides/social-networking-privacy-how-be-safe-secure-and-social> (accessed on 1 June 2018).
41. Zwick, R.; Thayer, D.T.; Wingersky, M. Effect of Rasch Calibration on Ability and DIF Estimation in Computer-Adaptive Tests. *J. Educ. Meas.* **1995**, *32*, 341–363. [[CrossRef](#)]
42. Bock, R.D.; Aitkin, M. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika* **1981**, *46*, 443–459. [[CrossRef](#)]



© 2018 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 (<http://creativecommons.org/licenses/by/4.0/>).