Attribute Inference: Social Network De-anonymous

Hui Liu 517030910365

June 2020

1 Introduction

Online Social Networks(OSNs) like Facebook, Google+, Weibo are becoming increasingly important part in our life, which can provide us a platform for socialing and sharing. There are billions of users using a variety of OSNs, sharing information they want to public and interacting with friends. However, there are also information, we called private attributes here, users do not want to share at least with strangers, for example, schools, age, and even gender. Therefore, OSNs are platforms composed of user's private and public information.

With the popularity of deep learning, OSNs become platforms providing huge dataset, most of which removed the unique user id for privacy concerns. However, such anonymous does not work very well. Nexflix holds a competition in 2006 to improve the recommendation algorithm and public an anonymous dataset, which was de-anonymous after just a month. [10] estimate the likelihood of a specific person to be correctly specified, showing that 99.98% of Americans can be inference correctly in any dataset using only 15 attributes, which can be easily obtained through the OSN. We discuss the privacy risks of user's attribute, *attribute inference* attack in this paper.

In the attribute inference attack, attackers aims to infer the missing attributes users hidden for strangers or does not fill with knowing public data such as friend lists, public attributes and behaviors of users, which can be obtained by public dataset or spider through the network. Apart from privacy risks, there are also security-sensitive problems. For instance, cyber criminals may attack personal information based user authentication using the infer attributes [4]. Advertiser may provide the targeted advertisement to earn more money [2]. Moreover, attacker may link the online records with your offline identity, which results in security danger [11].

Existing attribute inference attacks are mostly based on the explicit data, which means they ignored the correlation between features. [12, 1, 6] leverage users' attribute through their behavior and public attribute while [8, 13, 9] using users' social connection with others. They just infer from the most obvious side. [3] and [5] can infer attributes combining both user behavior and social structures, but they also ignored the relationship between attributes.

We proposed a relevance based inference method to perform the inference attack. The intuition is that attributes may have some correlation. For instance, a male user is more likely to be higher than a female user, a user studied in Shanghai Jiao Tong University may have more probability to live in Shanghai. So we proposed the relevance based method on the basic of Social-behavior-attribute networks.

Section 2 states the traditional attribute inference method and introduce the state-of-the-art SBA method. Section 3 introduce the proposed model and the algorithm of random walk with restart while Section 4 shows the experiment. The last section concludes the paper and states some future work.

2 Related Work

[12] use Logistic Regression to inference the gender of users through the user scoring behavior. [1] leverage the users' like list of music information to infer attributes, showing that users like similar music they like. [6] infer user attribute using the Facebook dataset. They all infer the user attributes leveraging the user behavior.

Another mainstream method based on homophily, meaning that friends may sharing the similar attributes. [8] leverage the user' social links to infer the shared attributes by Naive Bayes. However, it cannot infer users who does not share any attributes at all. [13] improved the social links with the groups, which reduce the exceptional. What's more, [9] proposed the seed user method that could identify local community user attribute.

[3] proposed state-of-the-art method: social behavior attribute network(SBA). They construct a Graph $G = (V, E)$, where $V = V_{social} \cup V_{attribute} \cup V_{behavior}$, the union of social nodes, attribute nodes and behavior nodes. Each user is formulate as a social node while each attribute and behavior is formulated as a binary node respectively. If a user u has attribute a, there is an edge between nodes u, a called attribute links. Similarly, there is also behavior links between user node and behavior node. What's more, if user u_1, u_2 follows each other, then there is a social link between them, shown as Figure 1. SBA provides a "vote" scheme for infer attribute, which consist of *dividing*, *backtracking* and *aggregating* parts, shown as Figure 2. The dividing part is the main "vote" part: for each user node u, it will divide all scores it have to his social neighbors, those u follow and follow u, attribute neighbors, those who share the same attributes with u and behavior neighbors similar to attribute neighbor. Backtracking part is based on the intuition that nodes close to target nodes u_t need to have more capacity. So each node u divide part of its score to u_t. As for aggregating part, that is because nodes who have more neighbors should have more capacity, result in each nodes have to give another part of its votes to the direct and un-direct neighbors. After these three parts, each social node has some capacity so that we can divide all of the votes to the attributes. At the end, the nodes who earns the most votes should be our infered attribute.

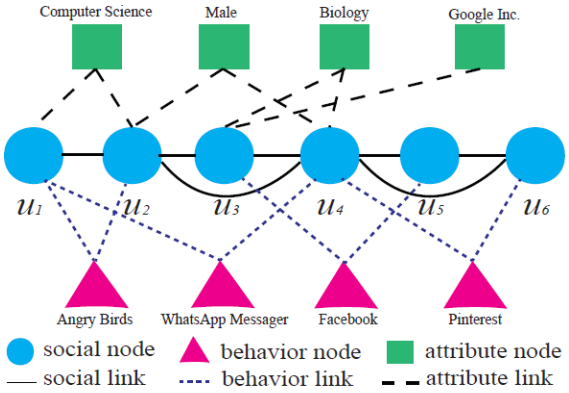

Figure 1: SBA

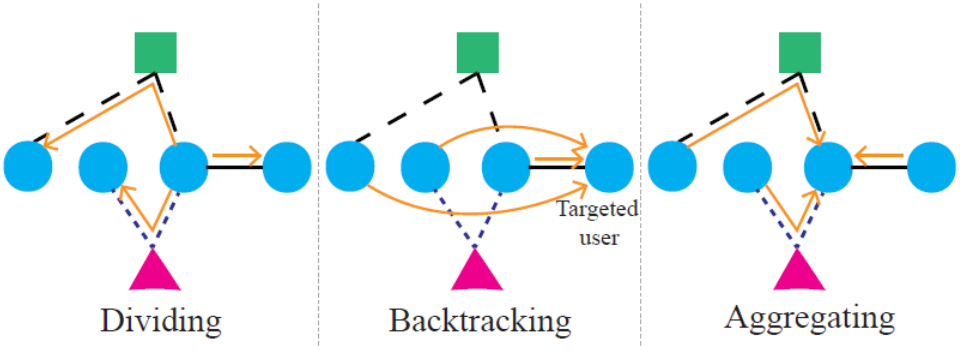

Figure 2: Dividing, backtracking and aggregating

3 Threat Model

Our proposed model is based on the SBA model. However, we found that the SBA model ignored the relevance between attributes and social nodes. That is, difference attributes have difference relevance and different social nodes also does. For example, gender effects height more than schools and close friend effects more than other speaking of acquaintance, so we proposed a relevance correlation based on Kulczynki measurement. For

example, a relevance between attribute a_i, a_j should be:

$$Relevance(a_i, a_j) = \frac{1}{2}(P(a_i|a_j) + P(a_j|a_i)) = \frac{1}{2}(\frac{U_{ij}}{U_i} + \frac{U_{ij}}{U_j}) \tag{1}$$

in which U_{ij} represent the number of users who have both attribute a_i, a_j while U_i represent the number of users who have attribute a_i. Equation (1) represent the probability of infer a_i while have a_j and infer a_j while have a_i.

Similar with relevance between attributes, there are also relevance between users:

$$Relevance(u_i, u_j) = \frac{1}{2}(R_{ij} + R'_{ij}) \tag{2}$$

represents the probability of the shared friends and the shared attribute.

$$\begin{aligned} R_{ij} &= \frac{1}{2}(\frac{N_{ij}}{N_i} + \frac{N_{ij}}{N_j}) \\ R'_{ij} &= \frac{1}{2}(\frac{A_{ij}}{A_i} + \frac{A_{ij}}{A_j}) \end{aligned} \tag{3}$$

Then we can construct a graph $G = (V, E, W)$ with the wights $w_{ij} = Relevance(i, j)$. That means there are weights on social links and also add attribute links with attribute relevance as the weight.

We also use "vote" method for attribute inference. However, we use a more mature way to represent the "vote" procedures, the random walk method. That means at first, we calculate the relevance as weights in the graph and then we only give some initial capacity to the target users. Then we repeatedly divide the votes to adjacent neighbors at the ratio of weights until convergence rather than just assign in average, which is random walk. Then we just divide the capacity of each user to their attributes. Finally, we run random walk again to assign votes of the attributes. The random walk can be easily formulate as:

$$R^{(t)} \Leftarrow W \times R^{(t-1)} \tag{4}$$

However, such random walk method can just result in an average image of the dataset. That means, we can just get the attribute which the occurs the most time in the dataset, due to the convergence proporty of the random walk. To highlight the given information of the target user, we reference the previous *backtracking* part and proposed the relevance random walk with restart, shown as Algorithm 1. The random walk with restart can be formulated by:

$$R^{(t)} \Leftarrow \alpha \times W \times R^{(t-1)} + (1 - \alpha) \times R^{(0)} \tag{5}$$

where α is the restart parameter. If α tends to 1, it becomes custom random walk and we get the portrait of the dataset while α tends to 0, we can just get the attribute occurs most times in the target users' neighbor. The value of α need to be calculate specifically and we find 0.85 reaches the best results through the experiment.

Algorithm 1 Relevance Random Walk with restart

$AA_{i,j} \Leftarrow relevance(a_i, a_j)$
$UU_{i,j} \Leftarrow relevance(u_i, u_j)$
Initialize $\vec{R}^{(0)}$ with target user index
repeat
  $\vec{R}^{(t)} \Leftarrow \alpha \cdot UU \cdot \vec{R}^{(t-1)} + (1 - \alpha)\vec{R}^{(0)}$
until Convergence
Assign $\vec{R}^{(t)}$ to attributes
Initialize $\vec{r^{(0)}}$ with target user's attribute
repeat
  $\vec{r}^{(t)} \Leftarrow \alpha \cdot AA \cdot \vec{r}^{(t-1)} + (1 - \alpha)\vec{r}^{(0)}$
until Convergence

4 Experiment

We use the dataset of Facebook[7] to show the results of proposed method. This dataset was composed of 22470 nodes with 171002 edges. All the attributes are binary labeled. Though their may be some multi-label attributes, we just divide them into binary attribute for easily calculating. Considering that the density of the represented matrix is 0.001, we use *sparse* package in spicy to accelerate the calculating.

The experiment results are shown as Figure 3. we calculate the accuracy of $top - k$ infered attributes and we can see the proposed method have higher accuracy than the state-of-the-art SBA model about 25% - 50%.

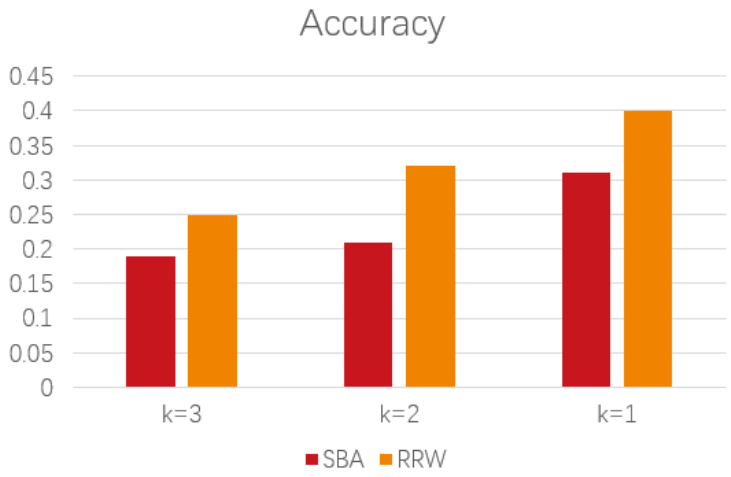

Figure 3: Experiment results

5 Conclusion

We proposed a relevance based random walk with restart method for attribute inference attack. The proposed method are based on state-of-the-art SBA model but adding relevance of attributes to complete the relationship and graph weights, gets some success. However, the dataset is too large in the real world so we need to perform some dimension reduction to reduce the calculation time. That could be the future work.

References

[1] Abdelberi Chaabane, Gergely Acs, Mohamed Ali Kaafar, et al. You are what you like! information leakage through users' interests. In *Proceedings of the 19th Annual Network & Distributed System Security Symposium (NDSS)*. Citeseer, 2012.

[2] Federal Commission. *Data brokers: A call for transparency and accountability*, pages 1–101. 01 2014.

[3] Neil Zhenqiang Gong and Bin Liu. You are who you know and how you behave: Attribute inference attacks via users' social friends and behaviors. In *25th {USENIX} Security Symposium ({USENIX} Security 16)*, pages 979–995, 2016.

[4] Payas Gupta, Swapna Gottipati, Jing Jiang, and Debin Gao. Your love is public now: Questioning the use of personal information in authentication. In *Proceedings of the 8th ACM SIGSAC Symposium on Information, Computer and Communications Security*, ASIA CCS '13, page 49–60, New York, NY, USA, 2013. Association for Computing Machinery.

[5] Jinyuan Jia, Binghui Wang, Le Zhang, and Neil Zhenqiang Gong. Attriinfer: Inferring user attributes in online social networks using markov random fields. In *Proceedings of the 26th International Conference on World Wide Web*, pages 1561–1569, 2017.

[6] Michal Kosinski, David Stillwell, and Thore Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the national academy of sciences*, 110(15):5802–5805, 2013.

[7] Jure Leskovec and Julian J Mcauley. Learning to discover social circles in ego networks. In *Advances in neural information processing systems*, pages 539–547, 2012.

[8] Jack Lindamood, Raymond Heatherly, Murat Kantarcioglu, and Bhavani Thuraisingham. Inferring private information using social network data. In *Proceedings of the 18th international conference on World wide web*, pages 1145–1146, 2009.

[9] Alan Mislove, Bimal Viswanath, Krishna P Gummadi, and Peter Druschel. You are who you know: inferring user profiles in online social networks. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 251–260, 2010.

[10] Luc Rocher, Julien M Hendrickx, and Yves-Alexandre De Montjoye. Estimating the success of re-identifications in incomplete datasets using generative models. *Nature communications*, 10(1):1–9, 2019.

[11] Latanya Sweeney. K-anonymity: A model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 10(5):557–570, October 2002.

[12] Udi Weinsberg, Smriti Bhagat, Stratis Ioannidis, and Nina Taft. Blurme: Inferring and obfuscating user gender based on ratings. In *Proceedings of the sixth ACM conference on Recommender systems*, pages 195–202, 2012.

[13] Elena Zheleva and Lise Getoor. To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In *Proceedings of the 18th international conference on World wide web*, pages 531–540, 2009.