Blind De-anonymization Attacks using Social Networks

Wei-Han Lee
Princeton University
weihanl@princeton.edu

Changchang Liu
Princeton University
cl12@princeton.edu

Shouling Ji
Zhejiang University, China and
Georgia Tech, USA
sji@gatech.edu

Prateek Mittal
Princeton University
pmittal@princeton.edu

Ruby B. Lee
Princeton University
rblee@princeton.edu

ABSTRACT

It is important to study the risks of publishing privacy-sensitive data. Even if sensitive identities (e.g., name, social security number) were removed and advanced data perturbation techniques were applied, several de-anonymization attacks have been proposed to re-identify individuals. However, existing attacks have some limitations: 1) they are limited in de-anonymization accuracy; 2) they require prior seed knowledge and suffer from the imprecision of such seed information.

We propose a novel structure-based de-anonymization attack, which does not require the attacker to have prior information (e.g., seeds). Our attack is based on two key insights: using multi-hop neighborhood information, and optimizing the process of de-anonymization by exploiting enhanced machine learning techniques. The experimental results demonstrate that our method is robust to data perturbations and significantly outperforms the state-of-the-art de-anonymization techniques by up to 10× improvement.

KEYWORDS

De-anonymization; Graph Anonymity; Machine Learning;

1 INTRODUCTION

Privacy-sensitive data (social relationships, mobility traces, medical records, etc.) are increasingly becoming public to facilitate data-mining researchers and applications. To protect users' privacy, data anonymization techniques have been the focus of extensive investigations [7, 15, 22].

Most privacy-sensitive data are closely related to individual behavior, and thus contain rich structural/graph-theoretic characteristics. For example, social networks can be modeled as graphs in a straightforward manner. Mobility traces can also be modeled as graph topologies [25]. However, even equipped with sophisticated anonymization techniques [4, 13, 14], the privacy of structural data still suffers from de-anonymization attacks assuming that the adversaries have access to rich auxiliary information (also called background information) from other channels [1, 2, 6, 8–12, 15, 17, 18, 20, 25].

Today, many individuals have accounts in various social networks such as Facebook, Twitter, Google+, Myspace and Flickr. Based on the inherent cross-site correlations, Narayanan et al. [18] effectively de-anonymized a Twitter dataset by utilizing a Flickr dataset as auxiliary information. Furthermore, Nilizadeh et al. [20] exploited the community structure of a graph to de-anonymize social networks. Other public datasets may also contain individual behavior information. For instance, Srivatsa et al. [25] proposed to de-anonymize a set of location traces based on a social network. They demonstrated that a contact graph identifying meetings between anonymized users in the location traces can be structurally correlated with the corresponding social network graph.

However, previous work on de-anonymization attacks have several limitations: 1) most previous works [17, 18, 20] rely on a seed-identification process. To obtain the useful seeds, they assume that the attacker possesses detailed information about a small number of members of the target network. They also assume that the attacker can determine if these members are also present in his auxiliary network (e.g., by matching user names and other contextual information). Furthermore, these methods may suffer from the imprecision of the adversary's background knowledge (misidentified seeds); 2) existing seed-free de-anonymization techniques [8, 21] have limited accuracy because they only utilize limited structural information of the data. In this paper, we aim to solve these problems by proposing a novel blind (i.e., seed-free) de-anonymization technique and exploring fine-grained structure information of graph topologies. Overall, we make the following contributions:

- We present a novel de-anonymization technique, which does not require adversaries to have any prior information (e.g., seeds). In our method, 1) we propose the **nK**-*series* to incorporate multi-hop neighbors' information in graph structures as novel features in our de-anonymization attack; 2) we jointly optimize the matching for users between the anonymized graph and the auxiliary graph by leveraging a machine learning technique: pseudo relevance feedback support vector machine (PRF-SVM).
- We show that our method is practical and effective: our attack is robust to data perturbations and has significant de-anonymization advantages over existing approaches with up to 10× improvement. Our method demonstrates that structural data can be effectively de-anonymized even without any seed information.

WPES'17, October 30, 2017, Dallas, TX, USA.

ACM ISBN 978-1-4503-5175-1/17/10…$15.00
https://doi.org/10.1145/3139550.3139562

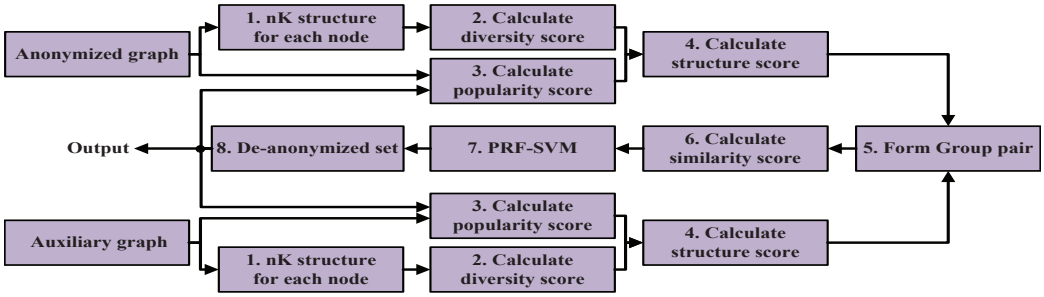


Figure 1: Mechanism for our blind de-anonymization attack.

2 BLIND DE-ANONYMIZATION ATTACKS

Previous works on structure-based de-anonymization do not fully utilize fine-grained graph-theoretic information. For instance, only one-hop neighbors have been utilized in [17], and very limited structural information has been leveraged in [8]. Also, most existing work rely on assumptions that the adversary has prior information or some ground truth (e.g., the seed information in [18]). However, seed-based de-anonymization attacks have some issues in practice: 1) the seed identification process usually requires heavy computational complexity [18]; 2) misidentified seeds may seriously decrease the de-anonymization capability.

We propose a general, blind (seed-free) de-anonymization attack. Figure 1 outlines our method, which consists of the following key steps on each of the anonymized graph, G_a, and the auxiliary graph, G_u. Our proposed **nK**-series aims to capture fine-grained structural information of each node, and the PRF-SVM aims to jointly de-anonymize the nodes. By exploring richer and finer-grained structural information of the graphs, our method can achieve better de-anonymization performance without requiring the adversaries to have any prior information (e.g., seeds).

Step 1: We first explore multi-hop neighbor information by proposing our new **nK**-*series* structural features for each node.

nK-*series:* Inspired by the idea of the **dK**-series [16] for characterizing structural statistics of a graph, we propose the **nK**-series to describe structural features of each node in a fine-grained manner, by incorporating the structural information of its multi-hop neighbors. **nK0** represents the degree of the node, i.e., the number of its neighbors. **nK1** captures the degree histogram of its neighbors and **nK2** captures the degree histogram of its 2-hop neighbors. Here, we focus our research on **nK0**, **nK1**, and **nK2** to construct the **nK** structural features of each node a as $\mathbf{v}(a) = [\mathbf{nK0}(a), \mathbf{nK1}(a), \mathbf{nK2}(a)]^T$.

Step 2: Based on the **nK** structural features, we calculate the *diversity score* for each node a, which measures the richness of the structural characteristics of this node and is defined as $DS(a) = \frac{\sum_i \tilde{v}_i(a) \log \tilde{v}_i(a)}{\log(\dim(\tilde{\mathbf{v}}(a)))}$. where $\tilde{\mathbf{v}}(a)$ denotes the normalized structural feature vector of a, i.e., $\tilde{\mathbf{v}}(a) = \frac{\mathbf{v(a)}}{\|\mathbf{v(a)}\|_2}$. $\dim(\tilde{\mathbf{v}}(a))$ denotes the dimension for $\tilde{\mathbf{v}}(a)$. Here, $\sum_i \tilde{v}_i(a) \log \tilde{v}_i(a)$ is actually similar to *entropy* in information theory [3], which evaluates the amount of information stored in $\tilde{v}_i(a)$, and $\log(\dim(\tilde{\mathbf{v}}(a)))$ is just for normalizing the diversity score so that $DS(a) \in [0, 1]$. A higher diversity score means that this node has more distinguishable structural characteristics.

Next, we start de-anonymizing the anonymized data in an iterative manner.

Step 3: For each round, we calculate the *popularity score* for each node, which evaluates its relationships with the set of de-anonymized nodes in the previous round (the set of de-anonymized nodes is empty in the initial round). We denote $\mathcal{N}^t$ as the set of nodes that have been de-anonymized after the t-th iteration, where $\mathcal{N}^t$ is an empty set for the first round. We define the popularity score of node a, $PS(a)$, as the Jaccard similarity [24] between the set of neighbors $N(a)$ for each node a and $\mathcal{N}^t$ as:

$$PS(a) = J(\mathcal{N}^t, N(a)) = |\mathcal{N}^t \cap N(a)| / |\mathcal{N}^t \cup N(a)| \tag{1}$$

where $J(A, B)$ is the Jaccard similarity, $J(A, B) = \frac{|A \cap B|}{|A \cup B|}$, and $PS(a) \in [0, 1]$. A higher popularity score represents a closer relationship between this anonymized node and those previously de-anonymized nodes. In addition to the diversity score, the popularity score can also be leveraged to evaluate the structural characteristics of each anonymized node.

Step 4: Subsequently, we compute the *structure score* (SS) for each node a as $SS(a) = DS(a) + c \cdot PS(a)$. where c is a pre-defined parameter to balance the diversity score and the popularity score.

Step 5: Next, we group the nodes in the anonymized graph and the auxiliary graph according to their structure scores. Our grouping process works as follows: for the t-th iteration, we select N_{group} nodes with higher SS from the anonymized graph and the auxiliary graph to form the group pair C_a^t (for the anonymized graph) and C_u^t (for the auxiliary graph). Note that for the first iteration, we select those nodes with higher DS (since $PS = 0$ for the initial round).

Step 6: For each group pair, we rank each potential pair of nodes according to the similarities between their **nK** structural features.

For each node a in C_a^t and node b in C_u^t, we evaluate the similarity between their **nK** structural features by computing the cosine similarity [5] between $\mathbf{v}(a)$ and $\mathbf{v}(b)$ as $\mathrm{Sim}(a, b) = \frac{\langle \mathbf{v}(a), \mathbf{v}(b) \rangle}{\|\mathbf{v}(a)\|_2 \|\mathbf{v}(b)\|_2}$. Larger cosine similarity score means two nodes are more similar. Furthermore, to emphasize the differences between node pairs and thus to improve the node matching performance, we can transform the above similarity linearly as

$$S(a, b) = \max_{d \in C_u^t} (\mathrm{Sim}(a, d)) - \frac{\max_{d \in C_u^t} (\mathrm{Sim}(a, d)) - \mathrm{Sim}(a, b)}{var(\mathrm{Sim}(a, :))} \tag{2}$$

where $\text{Sim}(a,:)$ is a vector consisting of $\text{Sim}(a,d)$ for $d \in C_u^t$ and $var(\cdot)$ is the variance of a vector.

Step 7: Next, we leverage machine learning techniques: pseudo relevance feedback with support vector machine (PRF-SVM), to re-rank these potential pairs of nodes.

Specifically, we view this node-matching process from a classification perspective, i.e., we aim to classify all the possible pairs of nodes as two categories: *matched* or *unmatched*. For each SVM iteration, we select the top N_{train} node pairs with the highest similarity scores and the bottom N_{train} node pairs with the lowest similarity scores as the training data, labeling them as *matched* and *unmatched*, respectively.

With these training node pairs, we apply SVM to classify the remaining node pairs. The SVM method would result in a classification hyperplane. Based on this hyperplane, each possible node pair would be given a value $dis(a,b)$ derived according to its distance from the hyperplane. We define a confidence score $\text{SVM}(a,b)$ for each potential node pair (a,b), which is linearly normalized as $\text{SVM}(a,b) = \frac{|dis(a,b)|-d_{min}}{d_{max}-d_{min}}$, where d_{max}, d_{min} represent the maximum and minimum distance from the hyperplane computed over all the remaining node pairs. The updated similarity score $\hat{S}(a,b)$ is obtained by integrating the original similarity score $S(a,b)$ in Eq. 2 with the confidence score $\text{SVM}(a,b)$ as

$$\hat{S}(a,b) = S(a,b) \cdot \text{SVM}(a,b)^{\alpha} \tag{3}$$

where α is a parameter that emphasizes the importance of the confidence score $\text{SVM}(a,b)$ in Eq. 3. A new ranking list is thus generated based on these updated similarity scores, $\hat{S}(a,b)$. This process of classification and re-ranking can be conducted iteratively until a stable classification result is obtained.

Step 8: Finally, we extract the matched pairs of nodes based on the classification result of PRF-SVM in *Step 7* and then update the set of de-anonymized nodes $\mathcal{N}^t$. We iteratively repeat *Step 3-Step 7* until we cannot de-anonymize any more nodes.

Note that although our method is seed-free, it can be directly generalized to incorporate seed knowledge if the adversary has such prior information. Given a set of known seeds, these seeds could be considered as the matched result in the first group of our algorithm, and the iteration for finding more matched nodes can be implemented consequently as shown in Figure 1.

3 EXPERIMENTAL ANALYSIS

In this section, we compare our attack with the state-of-the-art de-anonymization techniques [8, 20], to show the significant advantage of our approach (up to 10× improvement in de-anonymization accuracy). For fair comparison, we use the default parameters in the code these authors provided or the optimal parameters they utilized in their papers. We experiment on the collaboration dataset, the Twitter dataset and the Gowalla dataset (discussed below) for fair comparison with the method of Ji et al. and the method of Nilizadeh et al. since these are also the datasets they utilized [8, 20].

3.1 Datasets and General Settings

The Collaboration dataset [19] is a network of co-authorships between scientists who have posted preprints on the Condensed Matter E-Print Archive, which consists of *36,458* users and *171,735* edges. The Twitter dataset [20] captures the connections between users who mentioned each other at least once between March 24th, 2012 and April 25th, 2012, and contains two different graphs named Twitter (small) with *9,745* users and *50,164* edges, and Twitter (large) with *90,331* users and *358,422* edges.

The Gowalla dataset consists of a social graph and a mobility trace dataset [23]. The social graph contains *196,591* users with *950,327* edges. The mobility trace consists of *6.44*M checkins generated by these users. To better evaluate the performance of our method, we leverage the techniques in [23] to construct four graphs from the mobility trace dataset with different recalls and precisions, denoted by M1, M2, M3, and M4. All the four mobility trace graphs contain *196,591* users, and the corresponding number of edges are *659,186*, *829,375*, *919,671*, *1,070,790*, respectively.

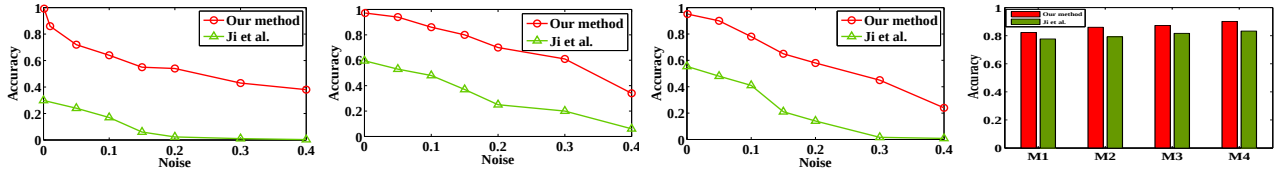
To evaluate the performance of our de-anonymization attack, we consider a popular perturbation method of Hay et al. [7], which applies a sequence of r random edge deletions followed by r random edge insertions (a similar perturbation process has been utilized for the de-anonymization attacks in [20]). Here, we define *noise* (perturbations) as the extent of edge modification, i.e., the ratio of altered edges r to the total number of edges M, i.e., $noise = \frac{r}{M}$. Note that we add the same amount of *noise* to the original graph of the Collaboration, Twitter datasets to obtain the anonymized graph and the auxiliary graph, respectively. For Gowalla mobility trace, we utilize its social network structure for de-anonymization attacks. Furthermore, we vary the system parameters of our method and set $c = 2, N_{group} = 1000, N_{train} = 1250, \alpha = 1$ for achieving the best performance in our experiments.

We utilize *Accuracy* to evaluate the de-anonymization performance. Accuracy is the ratio of the correctly de-anonymized nodes out of all the overlapped nodes between the anonymized graph and the auxiliary graph, i.e., $Accuracy = \frac{N_{cor}}{|V_a \cap V_u|}$, where N_{cor} is the number of correctly de-anonymized nodes and V_a, V_u represent the sets of nodes in the anonymized and auxiliary graph, respectively.

3.2 Comparison with Ji et al. [8]

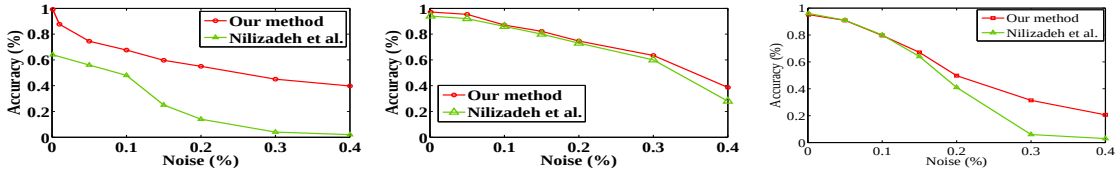
Ji et al. [8] proposed a cold-start optimization-based de-anonymization attack. Although they utilized four structural attributes for each node: degree, 1-hop neighborhood, top-K reference distance and sampling closeness centrality, these attributes only represent coarse-grained structure information of the graphs.

We compare our approach with the method of Ji et al. in the Collaboration dataset, the Twitter dataset and the Gowalla dataset in Figure 2. We can see that our approach has much higher accuracy than their method: we can achieve up to 10× improvement for collaboration dataset, and about 6× improvement for two Twitter graphs. Furthermore, we utilize the Gowalla social dataset to de-anonymize the Gowalla mobility trace dataset, in order to compare with the method of Ji et al. for fairness (they experimented on this data in [8]). In Figure 2(d), the de-anonymization results of our method for Gowalla mobility trace datasets (M1)(M2)(M3)(M4) are 81.3%, 84.8%, 85.3% and 89.1%, respectively. By utilizing finer-grained and richer structural information, our method also outperforms the method of Ji et al. for the Gowalla dataset.



(a) Collaboration (b) Twitter: small (c) Twitter: large (d) Gowalla

Figure 2: The comparison of our approach with the method of Ji et al. [8].



(a) Collaboration (b) Twitter: small (c) Twitter: large

Figure 3: The comparison of our approach with the method of Nilizadeh et al. [20].

3.3 Comparison with Nilizadeh et al. [20]

Nilizadeh et al. [20] leveraged community detection techniques to partition the networks into separate components. Then, they applied existing network alignment methods to the nodes inside the communities for more seed knowledge. However, their method has the following limitations: 1) it requires prior knowledge (seeds) to boot up their attack, which is a strong assumption and may suffer from misidentified seeds; 2) their performance may be influenced by the inconsistency problem of community detection methods [26]. We experiment on the collaboration dataset and the two Twitter graphs for fair comparison with the method of Nilizadeh et al. since these are also the data they used in [20].

Figure 3(a) compares our method with the approach of Nilizadeh et al. on the collaboration dataset. Our method can de-anonymize much more authors and is also more stable to data perturbations. For $noise = 0.4$, our method significantly outperforms the method of Nilizadeth et al. by more than 10× for de-anonymization accuracy.

Figure 3(b) and 3(c) compare our method with the method of Nilizadeh et al. on the Twitter datasets. Our method is more robust to noise, and has higher accuracy especially when the noise is high. For $noise = 0.4$, we have almost 1.25× improvement for Twitter (small) dataset, and 9× improvement for Twitter (large) dataset.

4 CONCLUSION

We presented a novel blind (seed-free) de-anonymization method by utilizing the **nK**-series that we define to capture fine-grained structure features, and proposing a new variant of the SVM machine learning technique called PRF-SVM to do concurrent matching of the nodes between the anonymized graph and the auxiliary graph. Experimental results demonstrate the significant advantages (up to 10× improvement in de-anonymization accuracy) of our method over the state-of-the-art de-anonymization attacks.

REFERENCES

[1] Lars Backstrom, Cynthia Dwork, and Jon Kleinberg. 2007. Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography. In *WWW*.

[2] Carla-Fabiana Chiasserini, Michele Garetto, and Emilio Leonardi. 2015. Impact of Clustering on the Performance of Network De-anonymization. In *COSN*.

[3] Thomas M Cover and Joy A Thomas. 2012. *Elements of information theory*. John Wiley & Sons.

[4] Cynthia Dwork. 2008. Differential privacy: A survey of results. In *International Conference on Theory and Applications of Models of Computation*. Springer, 1–19.

[5] E Garcia. 2006. Cosine similarity and term weight tutorial. *Information retrieval intelligence* (2006).

[6] Michael Hay, Gerome Miklau, David Jensen, Don Towsley, and Philipp Weis. 2008. Resisting structural re-identification in anonymized social networks. In *VLDB*.

[7] Michael Hay, Gerome Miklau, David Jensen, Philipp Weis, and Siddharth Srivastava. 2007. Anonymizing social networks. *Computer Science Department Faculty Publication Series* (2007).

[8] Shouling Ji, Weiqing Li, Mudhakar Srivatsa, and Raheem Beyah. 2014. Structural data de-anonymization: Quantification, practice, and implications. In *ACM CCS*.

[9] Ehsan Kazemi, S Hamed Hassani, and Matthias Grossglauser. 2015. Growing a graph matching from a handful of seeds. In *VLDB*.

[10] Wei-Han Lee, Changchang Liu, Shouling Ji, Prateek Mittal, and Ruby Lee. 2017. How to Quantify Graph De-anonymization Risks. In *Information Systems Security and Privacy*. Springer.

[11] Wei-Han Lee, Changchang Liu, Shouling Ji, Prateek Mittal, and Ruby B Lee. 2017. Quantification of De-anonymization Risks in Social Networks. In *Information Systems Security and Privacy*. IEEE.

[12] Hong Li, Cheng Zhang, Yunhua He, Xiuzhen Cheng, Yan Liu, and Limin Sun. 2016. An enhanced structure-based de-anonymization of online social networks. In *WASA*.

[13] Changchang Liu, Supriyo Chakraborty, and Prateek Mittal. 2016. Dependence Makes You Vulnberable: Differential Privacy Under Dependent Tuples.. In *NDSS*.

[14] Changchang Liu and Prateek Mittal. 2016. LinkMirage: Enabling Privacy-preserving Analytics on Social Relationships.. In *NDSS*.

[15] Kun Liu and Evimaria Terzi. 2008. Towards identity anonymization on graphs. In *SIGMOD*.

[16] Priya Mahadevan, Dmitri Krioukov, Kevin Fall, and Amin Vahdat. [n. d.]. Systematic topology analysis and generation using degree correlations. In *SIGCOMM*. ACM.

[17] Arvind Narayanan and Vitaly Shmatikov. 2008. Robust de-anonymization of large sparse datasets. In *IEEE S&P*.

[18] Arvind Narayanan and Vitaly Shmatikov. 2009. De-anonymizing social networks. In *IEEE S&P*.

[19] Mark EJ Newman. 2001. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences* (2001).

[20] Shirin Nilizadeh, Apu Kapadia, and Yong-Yeol Ahn. 2014. Community-enhanced de-anonymization of online social networks. In *ACM CCS*.

[21] Pedram Pedarsani, Daniel R Figueiredo, and Matthias Grossglauser. 2013. A bayesian method for matching two similar graphs without seeds. In *Allerton*

Conference on Communication, Control, and Computing.

[22] Pedram Pedarsani and Matthias Grossglauser. 2011. On the privacy of anonymized networks. In *SIGMOD*.

[23] Huy Pham, Cyrus Shahabi, and Yan Liu. 2013. EBM: an entropy-based model to infer social strength from spatiotemporal data. In *SIGMOD*.

[24] Raimundo Real and Juan M Vargas. 1996. The probabilistic basis of Jaccard's index of similarity. *Systematic biology* (1996).

[25] Mudhakar Srivatsa and Mike Hicks. 2012. Deanonymizing mobility traces: Using social network as a side-channel. In *ACM CCS*.

[26] Jierui Xie, Stephen Kelley, and Boleslaw K Szymanski. 2013. Overlapping community detection in networks: The state-of-the-art and comparative study. *Acm computing surveys (csur)* 45, 4 (2013), 43.