Poster: Towards Plausible Graph Anonymization

Yang Zhang
CISPA Helmholtz
Center for Information Security

Mathias Humbert
Swiss Data Science Center
ETH Zurich and EPFL

Bartlomiej Surma
CISPA Helmholtz
Center for Information Security

Praveen Manoharan
CISPA Helmholtz
Center for Information Security

Jilles Vreeken
CISPA Helmholtz
Center for Information Security

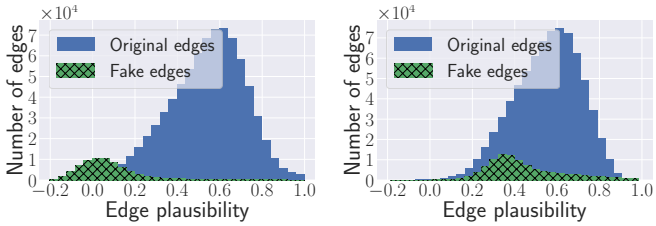
Michael Backes
CISPA Helmholtz
Center for Information Security

***Abstract*—Social graphs derived from online social interactions contain a wealth of information that is nowadays extensively used by both industry and academia. However, as social graphs contain sensitive information, they need to be properly anonymized before release. Most of the existing graph anonymization mechanisms rely on the perturbation of the original graph's edge set. In this paper, we identify a fundamental weakness of these mechanisms: They neglect the strong structural proximity between friends in social graphs, thus add implausible fake edges for anonymization. To exploit this weakness, we first propose a metric to quantify an edge's plausibility by relying on graph embedding. Extensive experiments on three real-life social network datasets demonstrate that our plausibility metric can very effectively differentiate fake edges from original edges with AUC values above 0.95 in most of the cases. We then rely on a Gaussian mixture model to automatically derive the threshold on the edge plausibility values to determine whether an edge is fake, which enables us to recover to a large extent the original graph from the anonymized graph. Then, we demonstrate that our graph recovery attack jeopardizes the privacy guarantees provided by the considered graph anonymization mechanisms. To mitigate this vulnerability, we propose a method to generate fake yet plausible edges given the graph structure and incorporate it into the existing anonymization mechanisms. Our evaluation demonstrates that the enhanced mechanisms decrease the chances of graph recovery and reduce the success of graph de-anonymization (up to 30%).**

I. Introduction

The rapid development of online social networks (OSNs) has resulted in an unprecedented scale of social graph data available. Access to such data is invaluable for both the industrial and academic domains. For instance, Amazon or Netflix have leveraged graph data to improve their recommendation services. Moreover, researchers have been using graph data to gain a deeper understanding of many fundamental societal questions, such as people's communication patterns [1], [2] and information propagation [3], [4]. These examples demonstrate that the sharing of large-scale graph data can bring significant benefits to the society.

On the downside, graph data also inherently contains very sensitive information about individuals [5], such as their social relations [6], and it can be used to infer private attributes [7]. In order to mitigate privacy risks, it is crucial to properly anonymize the graph data before releasing it to third parties. The naive approach of replacing real identifiers by random numbers has been proven ineffective by Backstrom et al. about a decade ago already [8]. From then on, the research community has been working on developing more robust graph anonymization mechanisms [10]–[12]. The majority of the proposed mechanisms focus on perturbing the original edge set of the graph (instead of perturbing the node set) by adding *fake edges* between users, such that the *perturbed graph* satisfies well-established privacy guarantees (such as k-anonymity [14] and differential privacy [15]).



(a) k-DA ($k = 100$) (b) Enhanced k-DA ($k = 100$)

Fig. 1: Plausibility distributions of fake and original edges in the NO dataset anonymized by (a) the original k-DA and (b) by our enhanced k-DA mechanisms.

Contributions. In this paper, we identify a fundamental weakness of the most prominent graph anonymization mechanisms: When creating fake edges, they do not take into account key characteristics of the underlying graph structure, such as the higher structural proximity between friends [16], which results in fake edges not being plausible enough compared to the original ones. To exploit this weakness, we first assess the plausibility of each edge by relying on graph embedding [17], [18]. We show that this approach can very effectively detect fake edges (see Figure 1a for an example), and thus can eventually help recover the original graph to a large extent. We then demonstrate that our graph recovery attack jeopardizes the anonymization mechanisms' privacy guarantees. Finally, we develop enhanced versions of the existing graph anonymization mechanisms that: (i) create plausible edges (Figure 1b), (ii) reduce the risk of graph recovery and graph de-anonymization and (iii) preserve the initial privacy criteria provided by the mechanisms. We concentrate on two of the best established graph anonymization mechanisms, which pro-

vide k-anonymity [9] and differential privacy [10] guarantees, respectively.

Edge Plausibility. We measure the plausibility of an edge as the structural proximity between the two users it connects. In the field of link prediction [16], structural proximity is normally measured by human-designed metrics, which only capture partial information of the proximity. Instead, we rely on graph embedding [17], [18] to map users in the anonymized graph into a continuous vector space, where each user's vector comprehensively reflects her structural properties in the graph. Then, we define each edge's plausibility as the similarity between the vectors of the two users this edge connects, and postulate that lower similarity implies lower edge plausibility.

Graph Recovery. We first show the effectiveness of our approach in differentiating fake edges from original ones without determining a priori a specific decision threshold on the plausibility metric. For this case, we adopt the AUC (area under the ROC curve) value as the evaluation metric. Extensive experiments performed on three real-life social network datasets show that our plausibility metric achieves excellent performance (corresponding to AUC values greater than 0.95) in most of the cases. Then, observing that the fake and real edges' empirical plausibility follow different Gaussian distributions, we rely on a Gaussian mixture model and maximum a posteriori probability estimate to automatically determine the threshold on the edge plausibility values to detect fake edges. Our experimental results show that this approach achieves strong performance with F1 scores above 0.8 in multiple cases. Deleting the fake edges let's us recover, to a large extent, the original graph from the anonymized one.

Privacy Damage. To precisely quantify the privacy impact of our graph recovery, we propose privacy loss measures tailored to each mechanism we target. As the first anonymization mechanism assumes the adversary uses the users' degrees to conduct her attack, we evaluate the corresponding privacy impact as the difference between users' degrees in the original, anonymized, and recovered graphs. For the differential privacy mechanism, we measure the magnitude and entropy of noise added to the statistical measurements of the graph. our experimental results show that the privacy provided by both mechanisms significantly decreases, which demonstrates the vulnerabilities of existing graph anonymization techniques.

Enhancing Graph Anonymization. In order to improve the privacy situation, we propose a method that generates plausible edges while preserving the original privacy guarantees of each mechanism. We rely on statistical sampling to select potential fake edges that follow a similar plausibility distribution as the edges in the original graph. Our experimental results show that our enhanced anonymization mechanisms are less prone to graph recovery (AUC dropping by up to 35%). More importantly, we show that our enhanced mechanisms reduce the state-of-the-art graph de-anonymization [19] attack's performance significantly.

In summary, we make the following contributions in this paper:

- We perform a graph recovery attack on anonymized social graphs based on graph embedding that captures the structural proximity between users and thus unveils fake edges (i.e., relations) between them.
- We show through extensive experimental evaluation on three different datasets that our graph recovery attack jeopardizes the privacy guarantees provided in two prominent graph anonymization mechanisms.
- We propose enhanced versions of these graph anonymization mechanisms that improve both their privacy and utility provisions.

REFERENCES

[1] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, "Structure and Tie Strengths in Mobile Communication Networks," *Proceedings of the National Academy of Sciences*, vol. 104.0, no. 18, pp. 7332–7336, 2007.

[2] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and Mobility: User Movement in Location-based Social Networks," in *Proc. KDD*, pp. 1082–1090, 2011.

[3] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the Spread of Influence through a Social Network," in *Proc. KDD*, pp. 137–146, 2003.

[4] D. M. Romero, B. Meeder, and J. Kleinberg, "Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter," in *Proc. WWW*, pp. 695–704, 2011.

[5] G. Beigi and H. Liu, "Privacy in Social Media: Identification, Mitigation and Applications," *CoRR abs/1808.02191*, 2018.

[6] M. Backes, M. Humbert, J. Pang, and Y. Zhang, "walk2friends: Inferring Social Links from Mobility Profiles," in *Proc. CCS*, pp. 1943–1957, 2017.

[7] J. Jia, B. Wang, L. Zhang, and N. Z. Gong, "AttriInfer: Inferring User Attributes in Online Social Networks Using Markov Random Fields," in *Proc. WWW*, pp. 1561–1569, 2017.

[8] L. Backstrom, C. Dwork, and J. Kleinberg, "Wherefore Art Thou R3579X? Anonymized Social Networks, Hidden Patterns, and Structural Steganography," in *Proc. WWW*, pp. 181–190, 2007.

[9] K. Liu and E. Terzi, "Towards Identity Anonymization on Graphs," in *Proc. SIGMOD*, pp. 93–106, 2008.

[10] A. Sala, X. Zhao, C. Wilson, H. Zheng, and B. Y. Zhao, "Sharing Graphs using Differentially Private Graph Models," in *Proc. IMC*, pp. 81–98, 2011.

[11] P. Mittal, C. Papamanthou, and D. Song, "Preserving Link Privacy in Social Network Based Systems," in *Proc. NDSS*, 2013.

[12] Q. Xiao, R. Chen, and K.-L. Tan, "Differentially Private Network Data Release via Structural Inference," in *Proc. KDD*, pp. 911–920, 2014.

[13] S. Ji, P. Mittal, and R. Beyah, "Graph Data Anonymization, De-Anonymization Attacks, and De-Anonymizability Quantification: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 19.0, no. 2, pp. 1305–1326, 2016.

[14] L. Sweeney, "k-Anonymity: A Model for Protecting Privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10.0, no. 5, pp. 557–570, 2002.

[15] C. Dwork and A. Roth, "The Algorithmic Foundations of Differential Privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9.0, no. 3-4, pp. 211–407, 2014.

[16] D. Liben-Nowell and J. Kleinberg, "The Link-prediction Problem for Social Networks," *Journal of the American Society for Information Science and Technology*, vol. 58.0, no. 7, pp. 1019–1031, 2007.

[17] B. Perozzi, R. Al-Rfou, and S. Skiena, "DeepWalk: Online Learning of Social Representations," in *Proc. KDD*, pp. 701–710, 2014.

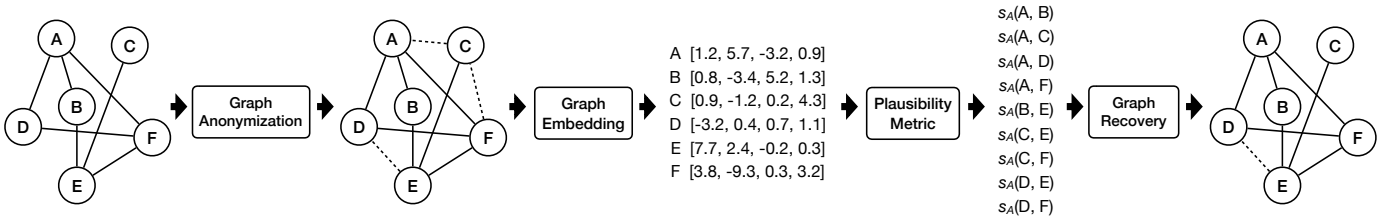
[18] A. Grover and J. Leskovec, "node2vec: Scalable Feature Learning for Networks," in *Proc. KDD*, pp. 855–864, 2016.

[19] A. Narayanan and V. Shmatikov, "De-anonymizing Social Networks," in *Proc. S&P*, pp. 173–187, 2009.

Towards Plausible Graph Anonyization

Yang Zhang, Mathias Humbert, **Bartlomiej Surma**, Praveen Manoharan, Jilles Vreeken, Michael Backes

Schematic view of the social network graph recovery attack



Graph Anonymization

Why

- OSN providers share their networks with advertisement companies or academia
- To protect user's privacy, graph needs to be anonymized
- Replacing user's names with random ids is not enough

How

- **k-anonymity** each user should share her node degree with k-1 other users, so add connections between users who's degree is to low
- **differential privacy** perturb users' degree distribution to achieve DP, add edges to the graph to get perturbed degree distribution

→ **To anonymize a graph (mostly) add fake edges (friendships, followings)**

Graph Embedding

Random Walk

- From one node jump to one of his neighbors randomly
- Save traces of such jumps
- Treat them as natural language sentences

Word2vec

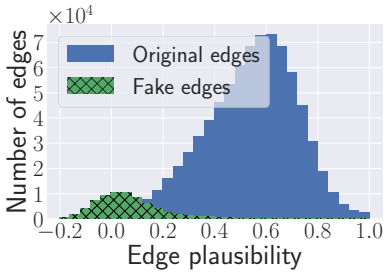
- Generated sentences feed to word2vec algorithm
- which embeds each word (node) into a vector based on:
  - their semantic similarity (neighborhood that given node is in)
  - gramatical role (structural role of a given node)

→ **Each node is embedded into a vector based on its proximity to other nodes and structural role**

Plausibility Metric

Plausible edges

- Calculate cosine similarities between node's vectors
- High cosine similarity means high edge plausibility
- This allows us to distinguish between original and fake edges



→ **Each edge we can transform into a number that tells us how likely it is to be one of original edges**

Graph recovery

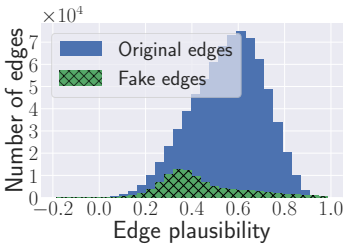
Removing fake edges

- Use Gaussian Mixture Model to find two bell curves (of fake and real edges) and their parameters
- Find a (plausibility) threshold, where the curves intersect each other. Different thresholds can be used for achieving higher true positives or true negatives ratio
- Remove all edges below given threshold

Plausible Graph Anonymization

Fixing graph anonymization

- Find target node degrees or a degree distribution to satisfy given anonymity definition (no change from original designs)
- Sample fake edges following real edges plausibility distribution
- Final anonymized graph is resilient to our graph recovery attack



This poster is based on the following publications:
1. Lars Backstrom and Cynthia Dwork and Jon Kleinberg
   „Wherefore Art Thou R3579X? Anonymized Social Networks, Hidden Patterns, and Structural Steganography" **(WWW 2007)**
2. Aditya Grover and Jure Leskovec
   "node2vec: Scalable Feature Learning for Networks" (**KDD 2016**)
3. Alessandra Sala, Xiaohan Zhao, Christo Wilson, Haitao Zheng, and Ben Y. Zhao
   "Sharing Graphs using Differentially Private Graph Models" (**IMC 2011**)
4. Kun Liu and Evimaria Terzi
   "Towards Identity Anonymization on Graphs" (**SIGMOD 2008**)

CISPA
HELMHOLTZ CENTER FOR INFORMATION SECURITY