# POSTER: Evaluating Privacy Metrics for Graph Anonymization and De-anonymization

Yuchen Zhao
De Montfort University
Leicester, UK
yuchen.zhao@dmu.ac.uk

Isabel Wagner
De Montfort University
Leicester, UK
isabel.wagner@dmu.ac.uk

## ABSTRACT

Many modern communication systems generate graph data, for example social networks and email networks. Such graph data can be used for recommender systems and data mining. However, because graph data contains sensitive information about individuals, sharing or publishing graph data may pose privacy risks. To protect graph privacy, data anonymization has been proposed to prevent individual users in a graph from being identified by adversaries. The effectiveness of both anonymization and de-anonymization techniques is usually evaluated using the adversary's success rate. However, the success rate does not measure privacy for individual users in a graph because it is an aggregate per-graph metric. In addition, it is unclear whether the success rate is monotonic, i.e. whether it indicates higher privacy for weaker adversaries, and lower privacy for stronger adversaries. To address these gaps, we propose a methodology to systematically evaluate the monotonicity of graph privacy metrics, and present preliminary results for the monotonicity of 25 graph privacy metrics.

**ACM Reference Format:**
Yuchen Zhao and Isabel Wagner. 2018. POSTER: Evaluating Privacy Metrics for Graph Anonymization and De-anonymization. In *ASIA CCS '18: 2018 ACM Asia Conference on Computer and Communications Security, June 4–8, 2018, Incheon, Republic of Korea.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3196494.3201586

## 1 INTRODUCTION

Data that contains relationships between individuals is often represented in graphs where each node represents an individual and each edge represents a relationship between two individuals. Many graph data sets, for example from online social networks [4] and email networks [6], have already been published for scientific or commercial use. Graph data can help us understand social networks [3] and improve recommendations [5], but can also violate the privacy of individuals because relationships can reveal sensitive information.

Many anonymization algorithms have been proposed to protect privacy in published graph data. Advanced adversaries, however, may be able to re-identify nodes by analyzing the structure of the anonymized graph, and by drawing on knowledge about the original graph and the relationships between the anonymized and original graphs. Thus, anonymization algorithms and adversaries who use de-anonymization algorithms are two parts of a protect-attack relationship. To evaluate the effectiveness of both anonymization and de-anonymization, researchers usually use *privacy metrics*.

The most common metric used to measure graph privacy is the adversary's success rate [2], i.e. the percentage of nodes the adversary was able to re-identify correctly. However, the success rate does not reveal much detail about the privacy of individual nodes in the graph – they are either re-identified or not – which may not be fine-grained enough to inform the development of new privacy-enhancing technologies (PETs).

Many different privacy metrics have been proposed for other domains [10]. These metrics associate privacy with different quantities such as the adversary's uncertainty, success rate, or error rate. However, when evaluating a new PET, the choice of privacy metrics is often arbitrary because it is not clear how strong different privacy metrics are in specific domains such as graph privacy. Thus understanding the strengths of different metrics will help us choose suitable privacy metrics when evaluating new PETs.

In this poster, we argue that strong privacy metrics should be monotonic, that is, that they indicate higher privacy levels for weaker adversaries, and we evaluate the monotonicity of 25 metrics for graph anonymization and de-anonymization.

Our preliminary results suggest that, although some metrics are stronger than others, there is no single metric that is monotonic in all situations. The best metrics are per-graph metrics, such as the adversary's success rate, which indicates that further research is needed to develop strong per-node metrics for graph privacy.

## 2 PROPOSED METHODOLOGY

To evaluate the strength of privacy metrics for graph privacy, we follow a similar methodology as in [8]. As Figure 1 shows, we first define *scenarios* consisting of user data and adversary behavior. Using the results of the adversary's de-anonymization algorithm, we then compute the values of different *privacy metrics*. Finally, we analyze monotonicity as the *strength indicator* for each metric. We have implemented this methodology in Python using NumPy, SciPy, and scikit-learn, and use the anonymization and de-anonymization algorithms implemented in SecGraph [2].

### 2.1 User Data

User data consists of published, real-world graphs where nodes represent users and edges represent connection or interaction between users. We use four datasets for our experiments: a social network (Facebook) [4], online contacts (PGP) [1], email communication


*ASIA CCS '18, June 4–8, 2018, Incheon, Republic of Korea*

ACM ISBN 978-1-4503-5576-6/18/06.
https://doi.org/10.1145/3196494.3201586

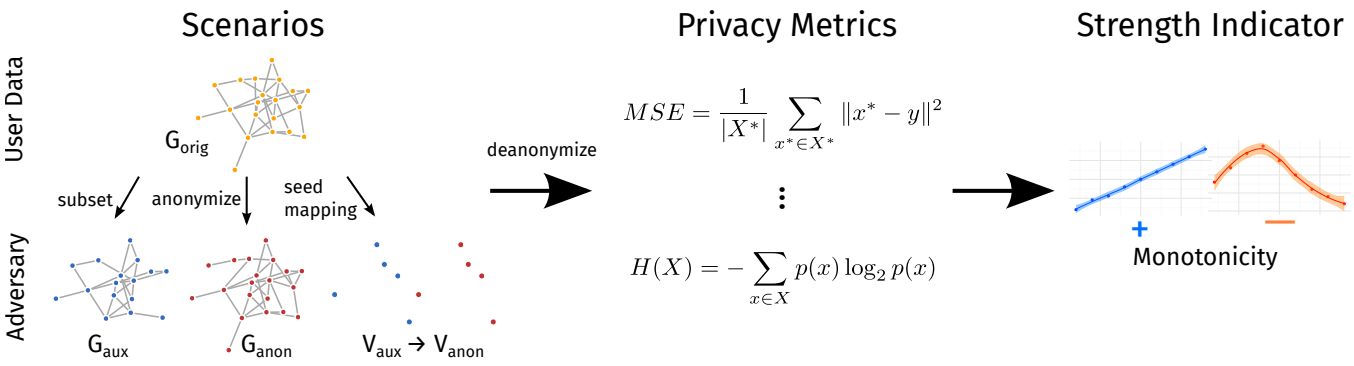



**Figure 1: The original graph $G_{orig}$ is processed by anonymization algorithms to produce the anonymized graph $G_{anon}$. The auxiliary graph $G_{aux}$ is a sub-graph of $G_{orig}$. The adversary relies on seed mappings between some nodes ($V_{aux}$ and $V_{anon}$) in $G_{aux}$ and $G_{anon}$ to bootstrap its mapping of the remaining nodes in $G_{aux}$ to the nodes in $G_{anon}$.**

(Manufacturing) [6] and message board communication (Irvine) [7]. We anonymize each dataset with five different anonymization algorithms that are implemented in SecGraph: k-degree anonymity (k-DA), differential privacy (DP), Switch, t-mean, and random walk (RW).

## 2.2 Adversary

The adversary aims to re-identify users in the anonymized graph. We use six de-anonymization algorithms from SecGraph: Adaptive De-Anonymization (ADA), Distance Vector (DV), Ji/Li/Srivatsa/Beyah (JLSB), Korula/Lattanzi (KL), Narayanan/Shmatikov (NS), and Yartseva/Grossglauser (YG). By default, SecGraph outputs the adversary's success rate for each algorithm. To be able to compute other privacy metrics, we have modified SecGraph to additionally output the adversary's estimated probability distribution.

To evaluate the monotonicity of privacy metrics, we need to calculate the metric values for a series of adversaries with ordered strength levels. We study two types of adversary strength levels: first, we vary the overlap of the auxiliary graph with the original graph between 60% and 95%, and second, we vary the number of seed mappings between 5 and 100.

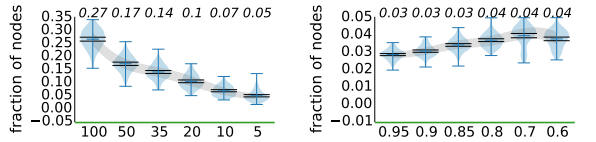
## 2.3 Metrics for Graph Privacy

Based on our survey of privacy metrics [10], we selected 25 privacy metrics to evaluate in our experiments. These metrics include metrics that have already been used in graph privacy and metrics from other domains. We refer to [10] for formal definitions.

## 2.4 Monotonicity

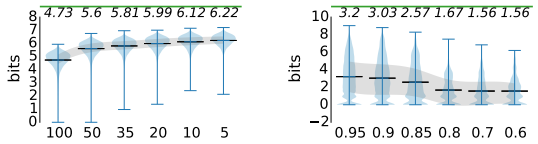
To evaluate the strength of privacy metrics, we require that they are monotonic, i.e. that they indicate higher privacy levels for weaker adversaries. For example, we expect that, with increasing levels of adversary strength, the values of the adversary's success rate (as a lower-better metric) increase, and that the values of entropy-based metrics (as higher-better metrics) decrease.

Our algorithm to score the monotonicity of metrics [8] is based on statistical tests for differences between the mean values of two samples. We apply these tests to the metric values for each pair of successive adversary strengths. If the difference between the means




(a) Adversary's success rate, Facebook, seed, KL, t-mean

(b) Adversary's success rate, Facebook, aux, JLSB, RW




(c) Entropy, Manufacturing, seed, YG, DP

(d) Entropy, PGP, aux, NS, k-DA

**Figure 2: Detailed results for success rate and entropy.**

is statistically significant and indicates a change in the expected direction, the algorithm increases the metrics' monotonicity score by 1. If the difference indicates a change in the wrong direction, the algorithm subtracts 1 from the monotonicity score. If the changes in metric values change direction, e.g. increasing for one pair and decreasing for the next, the algorithm subtracts 2 because such a peak may indicate the same privacy levels for both strong and weak adversaries and is thus undesirable. A metric's final monotonicity score is the average of the scores for two statistical tests (t-test and rank-sum statistic), normalized to $[0, 1]$.

# 3 PRELIMINARY FINDINGS

We have performed 100 replications for each combination of user data, anonymization algorithm, de-anonymization algorithm, and adversary strength, and applied the algorithm described in Section 2.4 to summarize our experimental results into monotonicity scores.

Figure 2 shows detailed results for two metrics, adversary's success rate and entropy, in two example scenarios each (our experiments generate results for 240 such plots per metric). Each violin shows the distribution of metric values for one adversary strength level, ordered from strongest (left) to weakest (right). In addition,

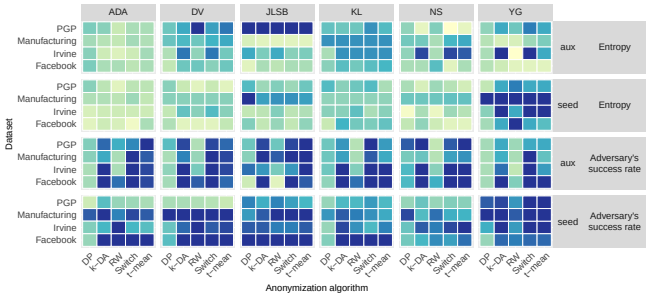



**Figure 3: Heat map visualizing monotonicity scores for two privacy metrics, grouped by dataset, anonymizer, de-anonymizer, and adversary strength type. Light yellow colors indicate low monotonicity (weak metric), and dark blue colors indicate high monotonicity (strong metric).**

black horizontal lines indicate confidence intervals for the mean, italic values on top of each violin indicate the mean value, and the green line at the top (resp. bottom) indicates whether higher (resp. lower) values indicate higher privacy. The metrics on the left (Figures 2a and 2c) behave as expected by indicating higher privacy for weaker adversaries, whereas the metrics on the right (Figures 2b and 2d) indicate higher privacy for stronger adversaries. This behavior is undesirable because it may lead to misjudging the strength of a new PET.

Figure 3 summarizes all results for these two metrics in a heat map. Each field in the heat map represents the monotonicity score of one metric in one scenario. For example, Figure 2a corresponds to the bottom-right field in the KL column. The heat map shows that on average, the adversary's success rate has much higher monotonicity than entropy. However, there are some combinations of dataset, anonymization, and de-anonymization algorithms, for which entropy is stronger than the adversary's success rate. The fact that monotonicity can vary depending on the dataset and algorithms indicates that there is no single metric that is always best to measure graph privacy.

To illustrate this point further, we aggregate all monotonicity scores for each metric into a box plot (Figure 4). The ranking clearly shows that the adversary's success rate and amount of leaked information are the strongest metrics. However, the figure also shows that the whiskers for almost all metrics extend both to the highest and to the lowest monotonicity. This indicates that even the strongest metrics are non-monotonic in some scenarios.

## 4 DISCUSSION AND OPEN ISSUES

Our evaluation of the strength of 25 graph privacy metrics shows that the strongest metrics are per-graph metrics, that is, metrics that indicate privacy for the entire graph, but not for individual nodes. Entropy-based metrics give detailed measurements of privacy for individual nodes, but although entropy is monotonic in other domains, e.g., vehicular communication [9], it is not monotonic in most graph privacy scenarios.

Our preliminary results thus point to important future research directions for graph privacy metrics: to study what conditions will result in monotonic privacy metrics, to develop strong metrics that

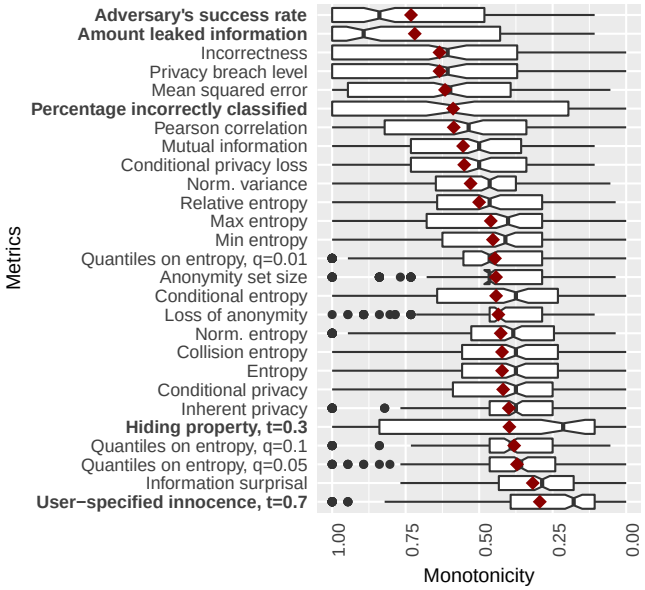



**Figure 4: Ranking of privacy metrics according to their monotonicity scores. Each box summarizes 240 data points (4 datasets × 2 types of adversary strength × 5 anonymizers × 6 de-anonymizers). Notches indicate 95% confidence intervals. Per-graph metrics are bold.**

measure the privacy of individual nodes, and to analyze whether monotonicity is the only requirement for strong graph privacy metrics.

## ACKNOWLEDGEMENTS

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grant EP/P006752/1 and used the ARCHER UK National Supercomputing Service.

## REFERENCES

[1] Marián Boguñá, Romualdo Pastor-Satorras, Albert Díaz-Guilera, and Alex Arenas. 2004. Models of Social Networks Based on Social Distance Attachment. *Physical Review E* 70, 5 (Nov. 2004), 056122.
[2] Shouling Ji, Weiqing Li, Prateek Mittal, Xin Hu, and Raheem Beyah. 2015. SecGraph: A Uniform and Open-source Evaluation System for Graph Data Anonymization and De-anonymization. In *Proc. 24th USENIX Security Symposium*. Washington, D.C., 303–318.
[3] Ravi Kumar, Jasmine Novak, and Andrew Tomkins. 2010. Structure and Evolution of Online Social Networks. In *Link Mining: Models, Algorithms, and Applications*. 337–357.
[4] Jure Leskovec and Julian J. Mcauley. 2012. Learning to Discover Social Circles in Ego Networks. In *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc., 539–547.
[5] Hao Ma, Dengyong Zhou, Chao Liu, Michael R. Lyu, and Irwin King. 2011. Recommender Systems with Social Regularization. In *Proc. 4th ACM Intl. Conference on Web Search and Data Mining (WSDM)*. Hong Kong, China, 287–296.
[6] Radosław Michalski, Sebastian Palus, and Przemysław Kazienko. 2011. Matching Organizational Structure and Social Network Extracted from Email Communication. In *Business Information Systems*. Springer, Berlin, Heidelberg, 197–206.
[7] Tore Opsahl and Pietro Panzarasa. 2009. Clustering in Weighted Networks. *Social Networks* 31, 2 (May 2009), 155–163.
[8] Isabel Wagner. 2017. Evaluating the Strength of Genomic Privacy Metrics. *ACM Transactions on Privacy and Security (TOPS)* 20, 1 (Jan. 2017), 2:1–2:34.
[9] Isabel Wagner. 2017. Measuring Privacy in Vehicular Networks. In *Proceedings of the 42nd IEEE Conference on Local Computer Networks (LCN)*. Singapore, 183–186.
[10] Isabel Wagner and David Eckhoff. 2018. Technical Privacy Metrics: A Systematic Survey. *ACM Computing Surveys (CSUR)* (2018). (to appear).