PrivEval: a tool for interactive evaluation of privacy metrics in synthetic data generation

Frederik Marinus Trudslev
Aalborg University
Aalborg, Denmark
fmtr@cs.aau.dk

Matteo Lissandrini
University of Verona
Verona, Italy
matteo.lissandrini@univr.it

Juan Manuel Rodriguez
Aalborg University
Aalborg, Denmark
jmro@cs.aau.dk

Martin Bøgsted
Aalborg University and Aalborg University Hospital
Aalborg, Denmark
martin.boegsted@rn.dk

Daniele Dell'Aglio
Aalborg University
Aalborg, Denmark
dade@cs.aau.dk

ABSTRACT

Synthetic data generation (SDG) is the process of generating a new synthetic dataset based on the statistical properties of a confidential existing dataset. Differential privacy is the property of a SDG mechanism that establishes how protected individuals whose sensitive data is part of the confidential dataset are, when sharing such data. To ensure a SDG is differentially private, noise is injected into the statistics learned from the dataset. Depending on the amount of noise injected, we witness a trade-off between privacy and utility. Privacy is then measured via a set of privacy metrics that usually establish a lower bound on a few aspects of the privacy-utility trade-off. Therefore, it is not possible to assess privacy based only on one metric. To close this gap, we demonstrate PrivEval, a tool to assist users in evaluating the privacy properties of a synthetic dataset. PrivEval implements several privacy metrics and validates them on both a single user and the overall dataset. Besides, PrivEval checks assumptions behind each metric. Hence, PrivEval is a first step to bridge the gap between privacy experts and the general public to make privacy estimation more transparent.

PVLDB Reference Format:
Frederik Marinus Trudslev, Matteo Lissandrini, Juan Manuel Rodriguez, Martin Bøgsted, and Daniele Dell'Aglio. PrivEval: a tool for interactive evaluation of privacy metrics in synthetic data generation. PVLDB, 18(12): 5271 - 5274, 2025.
doi:10.14778/3750601.3750649

PVLDB Artifact Availability:
The source code is available at https://github.com/hereditary-eu/PrivEval.

1 INTRODUCTION

Differential Privacy (DP) [3] has emerged as a gold standard for protecting the privacy of individuals that contribute their data to research and data collections. Among the various applications of DP, privacy-preserving (PP) Synthetic Data Generation (SDG) has recently gained attention as a way to share data without revealing users' sensitive information [13, 15, 20, 21, 24, 26]. A PP-SDG mechanism generates synthetic datasets from confidential ones based on dataset statistics and noise perturbations. Hence, the process offers privacy guarantees allowing for data sharing and analysis without privacy concerns. This, for example, is crucial in the medical field, where secondary usage of data for research is essential, but has to adhere to strict legal and ethical frameworks given the sensitive nature of the data.

Despite the availability of PP-SGD methods, it is still an open problem to assess the privacy risks that one may incur in, should they release a synthetic dataset generated with this process. At its core, DP has a tuning parameter ε that controls the privacy-utility tradeoff: the lower the ε value, the more noise is injected lowering the utility of the dataset but increasing its privacy guarantees. Still, ε is a number that can hardly be translated to a practical notion of privacy risk that one may relate to, especially to non-experts in statistics or computer science. Moreover, the same value of ε may lead to different privacy and utility values depending on the dataset, the underlying PP-SDG algorithm, and the downstream tasks, making it difficult to properly tune this parameter [8, 11].

Previous systems have demonstrated limited privacy risk assessment [6, 7, 12]. They focus on utility and efficiency, rely on user-defined privacy parameterisation, or focus on PP-SDG evaluation without assessing the actual privacy risk. These gaps underline the necessity for novel tools to study privacy risk in PP-SDG.

To understand the impact of ε in a given PP-SDG instance, researchers and practitioners proposed various privacy metrics to quantify the empirical privacy risk associated with the generated dataset [2, 5, 18, 22, 23]. However, these privacy metrics' nature is probabilistic as they estimate the risk of how much a specific kind of attack can learn about an individual. Moreover, each metric takes specific assumptions on the confidential dataset structure and the information available to the adversary. Therefore, it is not possible to use a single privacy metric as a definition of the overall privacy risk associated with sharing a given synthetic dataset. Still, evaluating the privacy guarantees of a given dataset requires in-depth knowledge of the metrics and their assumptions. This constitutes a non-trivial issue for non-experts, such as medical doctors or lawyers, who also need to assess the privacy properties of a synthetic dataset.

Proceedings of the VLDB Endowment, Vol. 18, No. 12 ISSN 2150-8097.
doi:10.14778/3750601.3750649

To assist with the assessment of the privacy guarantees of a given dataset, we introduce PrivEval, a novel tool for the semi-automatic assessment of the privacy guarantees of a given synthetic dataset. Using PrivEval, the user is able to explore the result of the application of different metrics and their applicability to the given use-case. Further, it is possible to simulate a given individual and also evaluate how their specific confidential data is being taken into account by the metrics. During the demonstration of PrivEval, users compare a confidential dataset in which their data is present to the synthetic dataset generated based on it. For each privacy metric the tool will present both their applicability to the data, based on automatic checks for the set of assumptions required by the metric, and then the implications of the results proposed by each metric. Therefore, PrivEval is a first step in making privacy metrics more accessible to the general public.

2 PRIVACY-PRESERVING SYNTHETIC DATA GENERATION

PrivEval is based on two core concepts: differential privacy and privacy-preserving synthetic data generation .

Let $A=\{a_1, a_2, \ldots, a_m\}$ be a set of attributes in a dataset, and let $\mathcal{D}$ be the set of all datasets. Differential privacy is a property that ensures a randomised algorithm (*mechanism*) applied to two similar datasets in $\mathcal{D}$ will produce outcomes that are also similar.

Definition 2.1 (ε-Differential Privacy ((ε, δ)-DP) [3]). A randomised mechanism $\mathcal{M}$ is (ε, δ)-DP if for all $S \subseteq range(\mathcal{M})$, and any pair of datasets $D, D' \in \mathcal{D}$ that differ in one record, it holds that $\Pr(\mathcal{M}(D) \in S) \leq e^{\varepsilon} \Pr(\mathcal{M}(D') \in S) + \delta$. Further, when $\delta = 0$, $\mathcal{M}$ is ε-DP.

The parameter ε is called *privacy budget* and controls the trade-off between privacy and utility, while δ is the probability that DP does not hold. From definition 2.1, it follows that when ε decreases, the distributions should get closer, i.e. the privacy is stronger. The parameter δ should be a small number to reduce privacy risk.

In practice, DP mechanisms inject noise to obfuscate the impact that a record can have on the result. For example, the Laplace and Gaussian mechanisms are DP mechanisms where noise is sampled from Laplace and Gaussian distributions, respectively [3]. The former is ε-DP, while the latter is (ε, δ)-DP.

Hence, a SDG mechanism is a randomised algorithm that transforms an input dataset $Y \in \mathcal{D}$ (the confidential dataset) into an output dataset $Z \in \mathcal{D}$ (the synthetic dataset). A PP-SDG mechanism is a mechanism that leverages differential privacy.

Definition 2.2 (Privacy-preserving synthetic data generation). A PP-SDG mechanism is an (ε, δ)-DP mechanism $\mathcal{M} : \mathcal{D} \times \Theta \rightarrow \mathcal{D}$, where Θ is the parameter space of the mechanism and $\mathcal{M}$ maps a dataset Y with attributes A to a corresponding synthetic version Z.

There are many PP-SDG mechanisms [10] that can be roughly categorised as methods based on probabilistic graphical models and methods based on neural networks. In this article, we use a representative of the first category, PrivBayes [26], but our system can be easily adapted to include any other method. PrivBayes [26] is an ε-DP PP-SDG mechanism that synthesises data by a three-step process: (1) it samples a Bayesian Network $\mathcal{B}$ from a probability distribution using the Exponential mechanism [3] on a scoring of possible attribute-parent pairs maximising the mutual information between attribute-parent pairs in a privacy-preserving approach, (2) it poses Laplacian noise on the conditional distributions of the sampled network, and (3) it generates synthetic records by sampling from the network. Given the stochastic nature of the process and given a synthetic dataset, the question is how much a given generated dataset is effectively preserving the privacy of the individuals whose data was present in the confidential dataset.

3 PRIVACY METRICS

PrivEval allows analysing the result of 17 privacy metrics [19], so to help users assess whether they are in a position to share such a dataset. While the ε parameter controls the trade-off between privacy and utility, it is not directly evaluating the privacy guarantees of the generated dataset in a way that is easily interpretable and relatable. Researchers and practitioners have therefore introduced utility and privacy metrics. Utility metrics measure to what extent a synthetic dataset can replace the confidential dataset in analytics and learning tasks. Typical utility metrics measure the similarity between the datasets' distributions (e.g. KL-divergence) or performance in downstream tasks (e.g. AUC-ROC in predictions) [10].

Conversely, privacy metrics quantify the privacy risk associated with a synthetic dataset, i.e. how likely confidential information present in the confidential dataset can be leaked when the synthetic dataset is shared. Privacy metrics often consider an opaque-box model [18], as they assume that an adversary has access to the synthetic data but not to the PP-SDG mechanism that generated it. The adversary may also have background (or prior) knowledge on part of the confidential dataset, modelling the fact that an adversary may know part of the confidential data. For example, in a medical scenario, an attacker may aim to discover the medical condition of their neighbour, knowing their age, gender, and height.

Definition 3.1 (Privacy Metric). A privacy metric p quantifies the risk of privacy associated with a synthetic dataset, such that $p : \mathcal{D} \times \mathcal{D} \times \mathcal{D} \rightarrow [0, 1]$, where p maps a confidential dataset, a synthetic dataset, and the auxiliary information to a score in the range $[0, 1]$, where 0 indicates complete privacy and 1 indicates no privacy.

Intuitively, privacy metrics aim to quantify how much information an adversary can gain if they know part of the confidential dataset and they get access to the synthetic dataset. Technically, privacy metrics vary based on the privacy attack they target, i.e. the confidential information that the adversary wants to infer, the adversary's background knowledge and the approach to quantify the metric. Privacy metrics are usually categorised into reconstruction attacks, tracing attacks, and re-identification attacks [4, 16].

In a *reconstruction attack* the goal is to reconstruct Y from the information obtained from Z, optionally knowing part of Y (for example, part of the records of a group of individuals). Examples of privacy metrics addressing the reconstruction risk are Attribute Inference Risk [2] and ZeroCAP [14].

In a *tracing attack* (also known as *Membership Inference Attack*), the goal is to determine whether an individual y^* is part of the confidential dataset Y that led to the synthetic dataset Z., i.e. whether $y^* \in Y$. The usual assumption on background knowledge is that the attacker knows the record y^*. Privacy metric examples used in

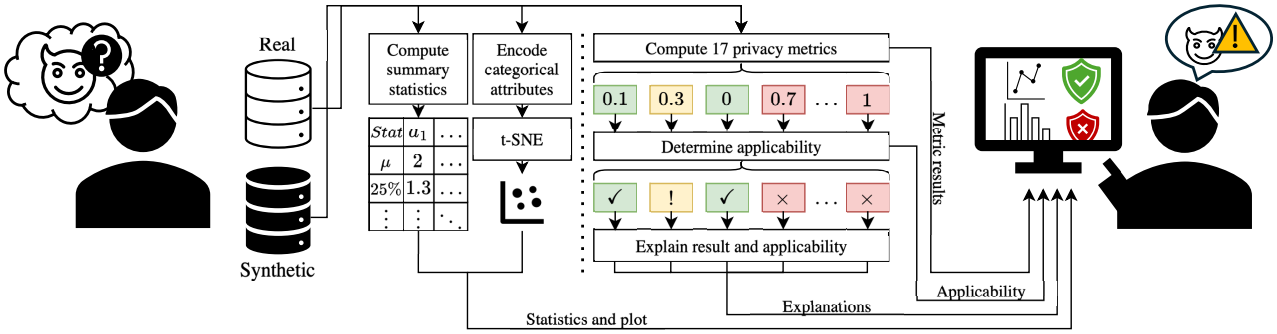

Figure 1: General structure of PrivEval.

this case are Hidden Rate [5] and Nearest Neighbour Adversarial Accuracy [9].

In a *re-identification attack*, the adversary aims to discover which records of the confidential dataset led to the generation of a synthetic record. The usual background knowledge consists in part of the confidential dataset. Privacy metrics associated with this attack include Authenticity [1] and Distance to Closest Record [5].[1]

4 PRIVEVAL

PrivEval is an interactive tool to support data curators in evaluating and exploring privacy metrics. PrivEval is agnostic to the SDG mechanism, as it receives as input a confidential dataset and a synthetic datasets. Currently, PrivEval supports datasets in the CSV format, as nowadays there are many tools that allow exporting data in this format. PrivEval has three main steps, shown in Figure 1.

Generate dataset statistics. As the first step, PrivEval computes statistics and visualisations of the input datasets, to support the user in understanding the dataset content. In the case of categorical attributes, PrivEval computes the distinct attribute values and the mode. For numerical attributes, PrivEval computes the mean, standard deviation, minimum and maximum values, and the quartile values. It also visualises the records of the two datasets in a plot by applying t-distributed stochastic neighbor embedding (t-SNE).

Estimate privacy of the synthetic dataset. Next, PrivEval computes the privacy metrics, which the user can inspect to assess the overall synthetic dataset privacy. Since privacy metrics also assume information about the background knowledge of the adversary, the user must provide the attribute name that is deemed sensitive. Further, PrivEval introduces the notion of *shareability*, i.e. an assessment based on the outcome of a privacy metric on the possibility to share a synthetic dataset. We assign to shareability three values: (1) shareable, (2) conflicts, and (3) unshareable. Conflict is introduced when the result of the privacy metric value is considered inconclusive, requiring the data curator to further investigate the privacy risk associated with specific records. For each metric, we defined rules to determine which shareability value should be assigned.

To support the user in navigating the privacy metrics, PrivEval also provides explanations about them and their implications. PrivEval provides an intuitive explanation of how the metrics estimate privacy, including illustrations of such computation using examples from the dataset, along with technical details for expert users.

Determine metric applicability. PrivEval determines which privacy metrics are applicable by using the statistics of the input datasets and information about the metric computation. Applicability indicates whether a metric estimates the risk as intended, and whether the metric estimation process is distorted by the characteristics of the datasets. Examples of applicability criteria are to check whether a metric is compatible with continuous attributes, or whether a metric can handle a high number of attributes.

Furthermore, PrivEval checks whether there are records of the confidential dataset that can be easily singled out using the synthetic dataset. To determine this, PrivEval finds the three nearest neighbors in the synthetic dataset for each record in the confidential dataset, then it compares the distances of these neighbors and identifies those neighbors that can be deemed problematic as they may reveal the existence in the confidential dataset of a given record.

These methods for determining how the assumptions of a metric influence the applicability, allow for categorisation of the applicability into three categories: *No assumption is compromised*, *Potential insufficient measurement of risk* and *Unreliable privacy estimation*. When investigating a single metric, the user can also gather insight into why the metric falls into the specific category as well as how to proceed in order to improve the privacy estimation.

5 DEMONSTRATION

To demonstrate PrivEval, we created a web application where the attendee pretends to provide data for a study. At the same time, one of the authors acts as the data curator, and during the demo, they will guide the attendees through two questions: (i) what are the privacy risks for the attendee, if they join the study? And (ii) how confident should the data curator be from a privacy perspective, if the synthetic data is released? PrivEval helps novice users to get an overall understanding of these topics, while experts will inspect and compare different privacy metrics and their applicability.

Setup details. We implemented the demo with the Python framework Streamlit. In order to keep the demonstration scenario completely safe, we created a confidential dataset with nine attributes and 1500 records. It contains some common user data including *First* and *LastName*, *Nationality*, and *Height*, as well as some more "sensitive" data, like *Favorite Icecream*, *Times Been To Italy*,

[1]The full list of privacy metrics and a description of these are present in the repository and in the technical report associated to this article [19].

Like Liquorice, *First Time In London*, and *Steps Per Day*. For this dataset, we use PrivBayes as the PP-SDG mechanism to generate multiple synthetic datasets with ε values from 0.02 to 5.

Overall, PrivEval can manage any pair of tabular confidential and synthetic datasets, and we showcase that using synthetic data generated by the SDG TabSyn [25] and the PP-SDG PrivBayes with different values of ε on the publicly available Shoppers dataset [17].

Demonstration scenario. At the beginning of the demonstration, we ask the user some (harmless) information and assign them a record in the confidential dataset that closely matches their answer. From that point, this data will be considered the confidential information of the attendee that should be protected.

The next step is the application of the PP-SDG mechanism. PrivEval analyses the corresponding synthetic dataset (see Section 4) and shows the corresponding summary statistics. PrivEval allows comparing the statistics of the synthetic datasets generated with different values of ε.

In the third and central step, PrivEval allows one to investigate the result of different privacy metrics from two different perspectives. Firstly, it allows us to estimate the risks for the attendee. In practice, different metric computations offer insights on how easy it is for an attacker to establish the attendee's contribution and the value of sensitive attributes. For example, if a user profile corresponds to an outlier, the metrics for the given user will suggest that it is easy to establish the membership in the confidential dataset, even though for all other users this risk does not materialise. Secondly, PrivEval allows one to establish the risks for the data curator, that is, how the metric estimates the overall privacy for all records in the dataset.

To help in this task, for each metric, PrivEval shows a description of the metric, the values for the attendee's record and for the whole dataset, along with the shareability assessments. The description changes based on the privacy metric value, the given user record, and the characteristics of the dataset.

5.1 Advantages

To our knowledge, PrivEval is the first tool to illustrate the impact of differential privacy in PP-SDG and assess the effectiveness of various privacy metrics in this context. By computing 17 privacy metrics at both the dataset and record levels, it provides insights into the shareability of synthetic datasets. PrivEval also offers insights into the underlying assumptions behind the metrics and informs about the dataset's correspondence with these. In the future, we aim to develop PrivEval into a fully automated reporting tool for data curators to support open science, tailoring it for specific users, e.g. medical doctors or law scholars.

ACKNOWLEDGMENTS

This work is partially supported by the HEREDITARY Project, as a part of the European Union's Horizon Europe research and innovation programme under Grant Agreement No GA 101137074.

REFERENCES

[1] Ahmed M. Alaa, Boris van Breugel, Evgeny Saveliev, and Mihaela van der Schaar. 2021. How Faithful is your Synthetic Data? Sample-level Metrics for Evaluating and Auditing Generative Models. *CoRR* abs/2102.08921 (2021). arXiv:2102.08921

[2] Edward Choi, Siddharth Biswal, Bradley Malin, Jon Duke, Walter F. Stewart, and Jimeng Sun. 2018. Generating Multi-label Discrete Patient Records using Generative Adversarial Networks. arXiv:1703.06490 [cs.LG]

[3] Cynthia Dwork and Aaron Roth. 2014. *The Algorithmic Foundations of Differential Privacy*. Vol. 9.

[4] Cynthia Dwork, Adam Smith, Thomas Steinke, and Jonathan Ullman. 2017. Exposed! A Survey of Attacks on Private Data. *Annual Review of Statistics and Its Application* (2017).

[5] Morgan Guillaudeux, Olivia Rousseau, Julien Petot, Zineb Bennis, Charles-Axel Dein, Thomas Goronflot, Nicolas Vince, Sophie Limou, Matilde Karakachoff, Matthieu Wargny, and Pierre-Antoine Gourraud. 2023. Patient-centric synthetic data generation, no reason to risk re-identification in biomedical data analysis. *npj Digital Medicine* 6, 1 (10 Mar 2023), 37.

[6] Michael Hay, Ashwin Machanavajjhala, Gerome Miklau, Yan Chen, Dan Zhang, and George Bissias. 2016. Exploring Privacy-Accuracy Tradeoffs using DPComp. In *Proceedings of the 2016 International Conference on Management of Data*. 2101–2104.

[7] Zhiqi Huang, Ryan McKenna, George Bissias, Gerome Miklau, Michael Hay, and Ashwin Machanavajjhala. 2019. PSynDB: accurate and accessible private data generation. *Proc. VLDB Endow.* 12, 12 (Aug. 2019), 1918–1921.

[8] James Jordon, Lukasz Szpruch, Florimond Houssiau, Mirko Bottarelli, Giovanni Cherubin, Carsten Maple, Samuel N. Cohen, and Adrian Weller. 2022. Synthetic Data – what, why and how? arXiv:2205.03257 [cs.LG]

[9] Anton Danholt Lautrup, Tobias Hyrup, Arthur Zimek, and Peter Schneider-Kamp. 2024. SynthEval: A Framework for Detailed Utility and Privacy Evaluation of Tabular Synthetic Data. arXiv:2404.15821 [cs.LG]

[10] Anton Danholt Lautrup, Tobias Hyrup, Arthur Zimek, and Peter Schneider-Kamp. 2025. Systematic Review of Generative Modelling Tools and Utility Metrics for Fully Synthetic Tabular Data. *ACM Comput. Surv.* 57, 4 (2025), 90:1–90:38.

[11] Jaewoo Lee and Chris Clifton. 2011. How Much Is Enough? Choosing ϵ for Differential Privacy. In *Information Security*. Berlin, Heidelberg, 325–340.

[12] Haoran Li, Li Xiong, Lifan Zhang, and Xiaoqian Jiang. 2014. DPSynthesizer: differentially private data synthesizer for privacy preserving data sharing. *Proc. VLDB Endow.* 7, 13 (2014), 1677–1680.

[13] Noseong Park, Mahmoud Mohammadi, Kshitij Gorde, Sushil Jajodia, Hongkyu Park, and Youngmin Kim. 2018. Data synthesis based on generative adversarial networks. *Proceedings of the VLDB Endowment* 11, 10 (2018), 1071–1083.

[14] Neha Patki, Roy Wedge, and Kalyan Veeramachaneni. 2016. The Synthetic Data Vault. In *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. 399–410.

[15] Bjarne Pfitzner and Bert Arnrich. 2022. DPD-fVAE: Synthetic Data Generation Using Federated Variational Autoencoders With Differentially-Private Decoder. arXiv:2211.11591 [cs.LG]

[16] Maria Rigaki and Sebastian Garcia. 2024. A Survey of Privacy Attacks in Machine Learning. *Comput. Surveys* 56, 4 (April 2024), 1–34. arXiv:2007.07646 [cs].

[17] C. Sakar and Yomi Kastro. 2018. Online Shoppers Purchasing Intention Dataset. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5F88Q.

[18] Theresa Stadler, Bristena Oprisanu, and Carmela Troncoso. 2022. Synthetic Data – Anonymisation Groundhog Day. In *31st USENIX Security Symposium*. 1451–1468.

[19] Frederik Marinus Trudslev, Matteo Lissandrini, Juan Manuel Rodriguez, Martin Bøgsted, and Daniele Dell'Aglio. 2025. A Review of Privacy Metrics for Privacy-Preserving Synthetic Data Generation. arXiv:2507.11324 [cs.CR] https://arxiv.org/abs/2507.11324

[20] Benjamin Weggenmann, Valentin Rublack, Michael Andrejczuk, Justus Mattern, and Florian Kerschbaum. 2022. DP-VAE: Human-Readable Text Anonymization for Online Reviews with Differentially Private Variational Autoencoders. In *Proceedings of the ACM Web Conference 2022*. 721–731.

[21] Liyang Xie, Kaixiang Lin, Shu Wang, Fei Wang, and Jiayu Zhou. 2018. Differentially Private Generative Adversarial Network. arXiv:1802.06739 [cs.LG]

[22] Andrew Yale, Saloni Dash, Ritik Dutta, Isabelle Guyon, Adrien Pavao, and Kristin P. Bennett. 2020. Generation and evaluation of privacy preserving synthetic health data. *Neurocomputing* 416 (2020), 244–255.

[23] Jinsung Yoon, Lydia N. Drumright, and Mihaela van der Schaar. 2020. Anonymization Through Data Synthesis Using Generative Adversarial Networks (ADSGAN). *IEEE journal of biomedical and health informatics* 24, 8 (2020), 2378–2388.

[24] Jinsung Yoon, James Jordon, and Mihaela van der Schaar. 2019. PATE-GAN: Generating Synthetic Data with Differential Privacy Guarantees. In *International Conference on Learning Representations*.

[25] Hengrui Zhang, Jiani Zhang, Balasubramaniam Srinivasan, Zhengyuan Shen, Xiao Qin, Christos Faloutsos, Huzefa Rangwala, and George Karypis. 2024. Mixed-Type Tabular Data Synthesis with Score-based Diffusion in Latent Space. In *The twelfth International Conference on Learning Representations*.

[26] Jun Zhang, Graham Cormode, Cecilia M. Procopiuc, Divesh Srivastava, and Xiaokui Xiao. 2017. PrivBayes: Private Data Release via Bayesian Networks. *ACM Trans. Database Syst.* 42, 4 (2017).