# Local Differential Privacy: Tools, Challenges, and Opportunities

Qingqing Ye[1,2][0000−0003−1547−2847] and Haibo Hu[2,3][0000−0002−9008−2112]

[1] School of Information, Renmin University of China, China
[2] Department of Electronic and Information Engineering,
Hong Kong Polytechnic University, Hong Kong, China
[3] Shenzhen Research Institute, Hong Kong Polytechnic University, Hong Kong, China
yeqq@ruc.edu.cn, haibo.hu@polyu.edu.hk

**Abstract.** Local Differential Privacy (LDP), where each user perturbs her data locally before sending to an untrusted party, is a new and promising privacy-preserving model. Endorsed by both academia and industry, LDP provides strong and rigorous privacy guarantee for data collection and analysis. As such, it has been recently deployed in many real products by several major software and Internet companies, including Google, Apple and Microsoft in their mainstream products such as Chrome, iOS, and Windows 10. Besides industry, it has also attracted a lot of research attention from academia. This tutorial first introduces the rationale of LDP model behind these deployed systems to collect and analyze usage data privately, then surveys the current research landscape in LDP, and finally identifies several open problems and research directions in this community.

**Keywords:** Data collection · data analysis · local differential privacy.

## 1 Motivation

With the prevalence of world wide web and big data, manufacturers and service providers at various layers are increasingly enthusiastic in collecting and analyzing user data to improve their services, lower the operation costs, and provide better personalization. The following are several examples.

1. Social networks and media such as Facebook and Instagram keep track of users' time and frequency spending on various pages in order to profile their favorites and interests. Such profile can be used for better personalized recommendation and advertisement.
2. Insurance company collects policy holders' usage data (for example, health and medical data for medical insurance, driving data for vehicle insurance) to improve actuarial model and personalized pricing.
3. Mobile phone manufacturers collect user' phone usage statistics such as app screen-on time and click frequency in order to design better battery optimization policy.

4. Wearable and IoT devices such as smart watches and smart bulbs collect usage statistics to provide personalized services to users.

However, such data collection comes at the price of privacy risks, not only to users but also to service providers who are vulnerable to internal and external data breaches. With General Data Protection Regulation (GDPR) enforced in EU from May 2018 and California Consumer Privacy Act (CCPA) will be in effect on January 1, 2020, there is a compelling need to provide strong privacy guarantees to users when collecting and analyzing their usage data.

As an answer to privacy-preserving data collection and analysis, Local Differential Privacy (LDP)[18] [11] [22] is a privacy model where each user perturbs her data locally before sending them to an untrusted party. In that sense, no one else can get access to the original data except the data owners themselves. So far, LDP has been deployed in many real products by several major software and Internet companies, including Google's RAPPOR system [14], Apple's iOS and macOS [33], and Microsoft Windows 10 [9].

The aim of this tutorial is to familiarize the audience with LDP as a popular privacy-preserving technique for data collection and analysis. This talk first introduces the rationale of LDP model and the basic analytical queries built on it, based on which the key technical underpinnings of the above deployed systems are then presented. As for academic research, this tutorial not only surveys the current research landscape in LDP, but also identifies open problem and research directions in this community.

## 2 Outline of the Tutorial

This tutorial is intended to introduce to WISE participants the new and yet rapidly developing topic of Local Differential Privacy (LDP). Inspired by the real LDP deployments, we structure the core of this tutorial by first describing the rationale of LDP and then introducing the key technical underpinnings of the deployed systems. Finally, we give an overview of the current research landscape and point out some research directions. The detailed outline of this tutorial is given below.

### 2.1 Preliminaries

Centralized differential privacy [13] assumes a trusted party that does not steal or leak data owner's private information. However, this does not hold in many real-world applications, particularly after the Facebook privacy scandal [31]. To this end, local differential privacy [18] [11] is proposed for the setting where each data owner locally perturbs her data using a randomized algorithm, and then sends the perturbed version to a third party. As such, a trusted party is no longer needed for data collection and analysis.

In the LDP setting, the whole database $D$ consists of a set of users, each possessing a private value $v$ in some domain. These users interact with an untrusted aggregator who is allowed to learn some statistics (e.g., mean, frequency, and even the distribution) of the private value in the whole population. LDP ensures the information leakage for each individual is bounded. Specifically, each user perturbs her private value $v$ by a randomized algorithm which takes $v$ as the input and outputs $v^*$. Upon receiving perturbed values $v^*$ from all users, the aggregator restores the statistics of $v$ in the whole population. $\epsilon$-local differential privacy (or $\epsilon$-LDP) is defined on $\mathcal{A}$ and a privacy budget $\epsilon > 0$ as follows.

*Definition 1 ($\epsilon$-local differential privacy).* A randomized algorithm $\mathcal{A}$ satisfies $\epsilon$-local differential privacy, if and only if for any two inputs $v, v' \in D$ and for any output $v^*$, the following inequation always holds.

$$\Pr[\mathcal{A}(v) = v^*] \leq e^{\epsilon} \times \Pr[\mathcal{A}(v') = v^*]$$

Intuitively, $\epsilon$-LDP means by observing the output $v^*$, the aggregator cannot infer whether the input is $v$ or $v'$ with high confidence, which is different from the centralized differential privacy defined on two neighboring databases that only differ in one record.

### 2.2 Randomized Response and Basic Queries

**Randomized Response** Randomized Response (RR) [41] is a technique developed for the interviewees in a survey to give random answer to a sensitive boolean question so that they can achieve plausible deniability. Specifically, each interviewee gives the genuine answer with probability $p$ and gives the opposite answer with probability $1 - p$. To adopt RR to satisfy $\epsilon$-LDP, we set $p$ as $\frac{e^{\epsilon}}{1+e^{\epsilon}}$, so that $\frac{p}{1-p} = e^{\epsilon}$. Note that the percentage of "true" answers (denoted by $f$) directly obtained from all perturbed answers is biased and thus needs to be calibrated to $\frac{p-1+f}{2p-1}$. Currently, RR and its variants have become the predominant perturbation mechanism for LDP to answer two basic queries — frequency estimation over categorical data and mean estimation over numerical data. We briefly describe the corresponding perturbation mechanism $\mathcal{A}$ for these two queries to achieve $\epsilon$-LDP.

**Frequency estimation over categorical data** This is a core problem in LDP, on which many research problems are investigated, including heavy hitter identification [35], frequent itemset mining [38], marginal release [8], spatiotemporal data aggregation [7] and range query [20].

As RR only targets at boolean variables, Generalized Randomized Response (GRR) with a domain size of $k$ is introduced in [17] as follows:

$$\Pr[\mathcal{A}(v) = v^*] = \begin{cases} \frac{e^{\epsilon}}{k-1+e^{\epsilon}}, & \text{if } v = v^* \\ \frac{1}{k-1+e^{\epsilon}}, & \text{if } v \neq v^* \end{cases}$$

GRR is a generalized form of RR, and when $k = 2$ it degenerates to RR. The estimation accuracy of GRR degrades as the domain size $k$ increases, because the probability that a value is correctly reported is approximately inversely proportional to $k$. Several other perturbation mechanisms have been proposed, including RAPPOR [14], Random Matrix Projection [4] and a mechanism based on the

Count sketch with Hadamard transform [3]. Wang *et al.* present a framework to compare different perturbation mechanisms in terms of estimation variance [40]. In the conclusion, they provide guidelines about which mechanism to choose based on the privacy budget the domain size.

**Mean estimation over numerical data** There are two state-of-the-art numerical value perturbation mechanisms. The first one is to add Laplace noise [13] to the value and provides the same plausible deniability as with the centralized differential privacy. However, different from the centralized setting where the Laplace noise is added to the query result on a set of users, in the local setting this noise is directly added to the value of each user. This mechanism has been adopted in works of [36] [26]. The second one takes RR as the building block, and decompose the mechanism into three major steps — discretization, perturbation and calibration. Specifically, a numerical value is first discretized to a binary one and is then perturbed by RR to satisfy $\epsilon$-LDP. As the perturbation causes the mean estimation to be biased, a calibration of the perturbed value is also needed. This mechanism has been adopted in works of [12] [9] [44] [10].

### 2.3 State-of-the-art Deployment

To illustrate the practical use of LDP, we describe three cases of deployment in major software and Internet companies, and introduce the key ideas behind them.

**RAPPOR from Google** The first large scale deployment of LDP in industry is RAPPOR. It is deployed in Google Chrome to enable a privacy-preserving collection of statistics in Chrome usage (e.g., about how a malware is hijacking users' settings in Chrome). Such statistics are crucial to improve browser security and user experience. The key idea is to transform a sensitive string into a Bloom Filter and then apply RR method to perturb it. A follow-up work [15] from the same team extends RAPPOR to more complex statistics collection without explicit dictionary knowledge.

**Apple's LDP deployment** This deployment was announced in 2016, and documented in a patent application [33] and a subsequent white paper [1]. The technique exploits several techniques — a Fourier transformation to spread out signal information, and a sketching algorithm to reduce the dimensionality of the domain. The deployed system now runs on hundreds of millions of iOS and macOS devices, performing a variety of data collection tasks, such as identifying popular emojis, popular health data types, and media playback preferences in Safari.

**Telemetry collection from Microsoft** The deployment of LDP in Microsoft collects telemetry data for mean and histogram estimation over time. A rounding technique has been applied to address the problem that privacy guarantee degrades rapidly when telemetry is collected regularly. This deployment has been rolled out since Windows 10 Fall Creator Update in 2017 and is now running on millions of devices to collect application usage statistics in a privacy-preserving manner [9].

### 2.4 Current Research Landscape

We will briefly describe the current research landscape in LDP for privacy-preserving data collection and analysis.

**Heavy hitter identification** The goal is to identify the values that are frequent globally. When the size of the domain is small, this problem can be directly solved by frequency estimation. That is, one simply queries the frequency of every value in the range, and then identifies those ones with the highest frequency counts. However, if the domain is very large (e.g., 128 bits or larger), finding the most frequent values in this way is computationally infeasible. The method proposed by Thakurta *et al.* [33] identifies the frequent byte at each location, and uses semantic analysis to filter out meaningless combinations. There are some other works in the pure LDP setting. Hsu *et al.* [16] and Mishra *et al.* [23] propose efficient protocols for heavy hitters, but the error bound is higher than that of the method proposed by Bassily and Smith [4]. A follow-up work by Bassily *et al.* [3] proposes *TreeHish*, which is shown more efficient and accurate than that of the Bassily and Smith method [4]. Bun *et al.* [5] proposes *PrivateExpanderSketch* with state-of-the-art theoretical performance. A concurrent work implements the first real protocol *PEM* [37] .

**Itemset mining** This problem considers the setting where each user has a set of items from a domain. For example, when Apple wants to estimate the frequencies of the emojis typed everyday by the users, each user has a set of emojis they have typed [34]. The problem is challenging because the simple method of encoding each itemset as a single value in the domain (power set of the original domain) and applying frequency estimation does not work. A frequency estimation method can only identify items that are very frequent in the population, but it is possible that all items in an infrequent itemset are very frequent. If no itemset in the power set domain is frequent enough, a direct encoding only sends noise to an aggregator. To solve this problem, the *LDPMiner* protocol [25] uses a technique called "padding and sampling". That is, each user first pads her itemset with dummy items to a fixed size $l$, and then randomly samples one item from the padded set, and finally uses a frequency estimation method to report the item. When estimating the frequency of an item, one multiples the estimated frequency by $l$. A very recent work [38] further improves the accuracy of *LDPMiner* within the same privacy constraints. The advantage comes from several key observations including privacy amplification under sampling, which is known to hold in the centralized setting [21].

**Marginal release** Marginal statistics, which captures the correlations among a set of attributes, are the building block of many data analysis tasks. A major challenge of marginal release with LDP is that heavy perturbation needs to be introduced in high dimension. Intuitively, marginal statistics can be derived from a space which consists of a noisy frequency of each value in the domain of all attributes. However, this space grows exponentially with the number of attributes, which leads to overwhelming noise in the computed marginals. To address this problem, Cormode *et al.* [8] apply Fourier Transformation to calculate $k$-way marginals. It only requires a few coefficients in the Fourier domain. Therefore, each user only needs to submit noisy Fourier coefficients to aggregator to compute the desired $k$-way marginals, instead of all values in those marginals. Another work [46] proposes to choose sets of attributes to reconstruct all $k$-way marginals in a local setting, which is inspired by *PriView* [24], a work for marginal release under a centralized setting.

**Graph data analysis** Sensitive information among users are embedded in a social graph, e.g., the intimate relationship of two users. As graph data mining has become an important means of knowledge discovery, safeguarding graph data of individuals becomes imperative. Qin *et al.* proposes *LDPGen* [26] to build synthetic social networks by a graph generation model named *BTER* [29]. For each user, a Laplace noise is added to the group-based node degree to satisfy $\epsilon$-LDP. Various mining tasks can then be carried out on the generated synthetic graph. Though this seems to be a general way for graph data analysis, it suffers from low data utility. The most recent work [32] proposes decentralized differential privacy (DDP) on local graph structures to ensure each individual protects not only her own privacy, but also the privacy of her neighbors. Based on DDP, a recursive framework with multi-phase is developed for subgraph counting, including triangles, three-hop paths and $k$-cliques. Though it is a dedicated solution for a specific graph analysis task, it can achieve better data utility than graph generation method.

**Key-value data collection** Key-value pair is a popular data model and pervasive in big data analytics. Intuitively, we could apply existing LDP methods for categorical data to perturb keys, and methods for numerical data to perturb values. However, this solution is not able to retain the key-value correlation that inherently exists between keys and values. A recent work *PrivKV* [44] proposes an efficient local perturbation protocol and an iterative model to collect key-value data for frequency and mean estimation. The main idea of *PrivKV* is to perturb the key first, and then apply perturbation on values based on the perturbed results of that key. Therefore, the correlation can be retained. In order to improve the estimation accuracy and reduce the network latency, an optimization strategy called “virtual iteration” is further proposed.

**Spatiotemporal data aggregation** As location-based service becomes a necessary part in our daily lives, collection spatial data with rigorous privacy guarantee is an urgent need. This problem has been extensively studied in a centralized DP setting. In an LDP setting, Chen *et al.* [7] are the first to propose a personalized LDP, i.e., each user may individually set her privacy requirements, to learn the user distribution over a spatial domain. It is yet an open problem to evaluate this method on more sophisticated user movement models. With the rapid development of indoor positioning technologies, privacy preserving of users' indoor location information has received increasing attention. To this end, Kim *et al.* [19] propose to apply LDP to the domain of indoor positioning systems to estimate the population density of the specified indoor area. A recent work [6] indicates that existing location privacy-preserving mechanisms may not protect the sensitive information about users' spatiotemporal activities, when a user's locations are released continuously. To address this, a framework called *PriSTE* is proposed to transform an existing location privacy-preserving mechanism into one protecting spatiotemporal privacy.

### 2.5 Open Problems and New Directions

Finally, based on emerging trends in the literature, we will point to some directions for future work.

**Iterative interactions** Many data analysis tasks need to access the original data multiple times to improve the accuracy of the results. In a local setting, this requires users to perturb their data and then send to the aggregator with multiple rounds of interactions. To exploit such interactions, in each round the aggregator poses new queries in the light of previous responses. This approach has been adopted in several works, including machine learning model [36], heavy hitters estimation [25], synthetic graph modeling [26] and key-value data collection [44]. It is yet an open problem to have a deep understanding of the effectiveness and consequences of multiple rounds of interactions [30].

**High-dimensional data analysis** A large amount of potential knowledge and patterns can be extracted through high-dimensional data analysis. To enable privacy-preserving high-dimensional data analysis, some existing works propose to enable LDP for joint distribution estimation [28] or to handle some queries with different types of predicates and aggregation functions (e.g., SUM, AVG and STDEV) [39], with a focus on improving data utility. Besides the effect on data utility, achieving LDP on high-dimensional data analysis raises great challenges in terms of privacy guarantee, computation efficiency and communication overhead. For high-dimensional data, correlations among different dimensions should be taken into account as it may lead to privacy vulnerability [45] [43], thus increasing the success ratio of many reference attacks. In addition, computation efficiency and communication overhead are also concerns, especially for applications that can only afford lightweight operations, such as IoT scenarios and real-time applications.

**Privacy-preserving data mining and machine learning** Privacy-preserving data mining tasks and machine learning models in a centralized DP setting have been widely investigated, e.g., frequent subgraph mining [42], random forest [27] and deep learning [2]. However, there are very few works in a local setting. To the best of our knowledge, only some of the simple machine learning models [36] [47] (e.g., linear regression, logistic regression and support vector machine) and data mining tasks (e.g., itemset mining on set-valued data [38], and community detection on graph data [26]) have been studied in the context of LDP. The challenges come from two aspects. First, data mining tasks and machine learning models are often based on a global view of the data, and thus need several interactions between users and aggregator, which deviates from the local setting of LDP. Second, comparing with the centralized DP, LDP has stronger privacy model and yet incurs heavier noise, which makes it less practical for complex data mining tasks and machine learning models.

**Theoretical underpinnings** Several works on LDP have addressed questions about the theoretical power and limitation of LDP. For example, what are the lower bounds on the accuracy guarantees (as a function of privacy parameter and population size) [4]? Is there any benefit from adding an additive "relaxation" $\delta$ to the privacy definition [5]? And how to minimize the amount of data collected from each user to a single bit [11]?

## 3 Biography

**Qingqing Ye** is a PhD candidate in School of Information, Renmin University of China and a research assistant in the Hong Kong Polytechnic University. Her research interests include data privacy and security, with a focus on local differential privacy.

**Haibo Hu** is an associate professor in the Department of Electronic and Information Engineering, Hong Kong Polytechnic University. His research interests include cybersecurity, data privacy, internet of things, and machine learning. He has published over 70 research papers in refereed journals, international conferences, and book chapters. As principal investigator, he has received over 10 million HK dollars of external research grants from Hong Kong and mainland China. He is the recipient of a number of titles and awards, including IEEE MDM 2019 Best Paper Award, WAIM Distinguished Young Lecturer, VLDB Distinguished Reviewer, ACM-HK Best PhD Paper, Microsoft Imagine Cup, and GS1 Internet of Things Award.

## Acknowledgment

This work was supported by National Natural Science Foundation of China (Grant No: 61572413, U1636205, 91646203, 61532010, 91846204 and 61532016), the Research Grants Council, Hong Kong SAR, China (Grant No: 15238116, 15222118 and C1008-16G).

## References

1. Differential Privacy Team, Apple. Learning with privacy at scale. 2017
2. Abadi, M., Chu, A., Goodfellow, I., McMahan, H.B., Mironov, I., Talwar, K., Zhang, L.: Deep learning with differential privacy. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. pp. 308–318. ACM (2016). https://doi.org/10.1145/2976749.2978318
3. Bassily, R., Nissim, K., Stemmer, U., Thakurta, A.G.: Practical locally private heavy hitters. In: Advances in neural information processing systems (NIPS). pp. 2288–2296 (2017)
4. Bassily, R., Smith, A.: Local, private, efficient protocols for succinct histograms. In: Proceedings of the 47th Annual ACM on Symposium on Theory of Computing (STOC). pp. 127–135. ACM (2015). https://doi.org/10.1145/2746539.2746632
5. Bun, M., Nelson, J., Stemmer, U.: Heavy hitters and the structure of local privacy. In: Proceedings of the 35th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems (PODS). pp. 435–447. ACM (2018). https://doi.org/10.1145/3196959.3196981
6. Cao, Y., Xiao, Y., Xiong, L., Bai, L.: Priste: from location privacy to spatiotemporal event privacy. In: IEEE 35th International Conference on Data Engineering (ICDE). pp. 1606–1609. IEEE (2019). https://doi.org/10.1109/icde.2019.00153
7. Chen, R., Li, H., Qin, A., Kasiviswanathan, S.P., Jin, H.: Private spatial data aggregation in the local setting. In: IEEE 32nd International Conference on Data Engineering(ICDE). pp. 289–300. IEEE (2016). https://doi.org/10.1109/icde.2016.7498248
8. Cormode, G., Kulkarni, T., Srivastava, D.: Marginal release under local differential privacy. In: Proceedings of the 2018 International Conference on Management of Data (SIGMOD). pp. 131–146. ACM (2018). https://doi.org/10.1145/3183713.3196906
9. Ding, B., Kulkarni, J., Yekhanin, S.: Collecting telemetry data privately. In: Advances in neural information processing systems (NIPS). pp. 3574–3583 (2017)
10. Ding, B., Nori, H., Li, P., Allen, J.: Comparing population means under local differential privacy: with significance and power. In: 32nd AAAI Conference on Artificial Intelligence (2018)
11. Duchi, J.C., Jordan, M.I., Wainwright, M.J.: Local privacy and statistical minimax rates. In: IEEE 54th Annual Symposium on Foundations of Computer Science (FOCS). pp. 429–438. IEEE (2013). https://doi.org/10.1109/focs.2013.53
12. Duchi, J.C., Jordan, M.I., Wainwright, M.J.: Minimax optimal procedures for locally private estimation. Journal of the American Statistical Association **113**(521), 182–201 (2018). https://doi.org/10.1080/01621459.2017.1389735
13. Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: Theory of cryptography conference. pp. 265–284. Springer (2006). https://doi.org/10.29012/jpc.v7i3.405
14. Erlingsson, Ú., Pihur, V., Korolova, A.: Rappor: Randomized aggregatable privacy-preserving ordinal response. In: Proceedings of the 2014 ACM SIGSAC conference on computer and communications security (CCS). pp. 1054–1067. ACM (2014). https://doi.org/10.1145/2660267.2660348

15. Fanti, G., Pihur, V., Erlingsson, Ú.: Building a rappor with the unknown: Privacy-preserving learning of associations and data dictionaries. Proceedings on Privacy Enhancing Technologies **2016**(3), 41–61 (2016). https://doi.org/10.1515/popets-2016-0015
16. Hsu, J., Khanna, S., Roth, A.: Distributed private heavy hitters. Automata, Languages, and Programming pp. 461–472 (2012). https://doi.org/10.1007/978-3-642-31594-7_39
17. Kairouz, P., Oh, S., Viswanath, P.: Extremal mechanisms for local differential privacy. In: Advances in neural information processing systems (NIPS). pp. 2879–2887 (2014)
18. Kasiviswanathan, S.P., Lee, H.K., Nissim, K., Raskhodnikova, S., Smith, A.: What can we learn privately? SIAM Journal on Computing **40**(3), 793–826 (2011). https://doi.org/10.1137/090756090
19. Kim, J.W., Kim, D.H., Jang, B.: Application of local differential privacy to collection of indoor positioning data. IEEE Access **6**, 4276–4286 (2018). https://doi.org/10.1109/access.2018.2791588
20. Kulkarni, T.: Answering range queries under local differential privacy. In: Proceedings of the 2019 International Conference on Management of Data. pp. 1832–1834. ACM (2019). https://doi.org/10.1145/3299869.3300102
21. Li, N., Qardaji, W., Su, D.: On sampling, anonymization, and differential privacy or, k-anonymization meets differential privacy. In: Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security (ASIACCS). pp. 32–33. ACM (2012). https://doi.org/10.1145/2414456.2414474
22. Li, N., Ye, Q.: Mobile data collection and analysis with local differential privacy. In: IEEE International Conference on Mobile Data Management (MDM). https://doi.org/10.1109/access.2018.2791588
23. Mishra, N., Sandler, M.: Privacy via pseudorandom sketches. In: Proceedings of the twenty-fifth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems (PODS). pp. 143–152. ACM (2006). https://doi.org/10.1145/1142351.1142373
24. Qardaji, W., Yang, W., Li, N.: Priview: practical differentially private release of marginal contingency tables. In: Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data (SIGMOD). pp. 1435–1446. ACM (2014). https://doi.org/10.1145/2588555.2588575
25. Qin, Z., Yang, Y., Yu, T., Khalil, I., Xiao, X., Ren, K.: Heavy hitter estimation over set-valued data with local differential privacy. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS). pp. 192–203. ACM (2016). https://doi.org/10.1145/2976749.2978409
26. Qin, Z., Yu, T., Yang, Y., Khalil, I., Xiao, X., Ren, K.: Generating synthetic decentralized social graphs with local differential privacy. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS). pp. 425–438 (2017). https://doi.org/10.1145/3133956.3134086
27. Rana, S., Gupta, S.K., Venkatesh, S.: Differentially private random forest with high utility. In: 2015 IEEE International Conference on Data Mining (ICDM). pp. 955–960. IEEE (2015). https://doi.org/10.1109/icdm.2015.76
28. Ren, X., Yu, C.M., Yu, W., Yang, S., Yang, X., McCann, J.A., Philip, S.Y.: LoPub: High-dimensional crowdsourced data publication with local differential privacy. IEEE Transactions on Information Forensics and Security **13**(9), 2151–2166 (2018). https://doi.org/10.1109/tifs.2018.2812146

29. Seshadhri, C., Kolda, T.G., Pinar, A.: Community structure and scale-free collections of erdős-rényi graphs. Physical Review E **85**(5), 056109 (2012). https://doi.org/10.1103/physreve.85.056109
30. Smith, A., Thakurta, A., Upadhyay, J.: Is interaction necessary for distributed private learning? In: 2017 IEEE Symposium on Security and Privacy (SP). pp. 58–77. IEEE (2017). https://doi.org/10.1109/sp.2017.35
31. Stephanie, B.: (Facebook Scandal a 'Game Changer' in Data Privacy Regulation), *Bloomberg*, Apr 8, 2018
32. Sun, H., Xiao, X., Khalil, I., Yang, Y., Qin, Z., Wang, H.W., Yu, T.: Analyzing subgraph statistics from extended local views with decentralized differential privacy. In: Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security (CCS). pp. 703–717. ACM (2019). https://doi.org/10.1145/3319535.3354253
33. Thakurta, A.G., Vyrros, A.H., Vaishampayan, U.S., Kapoor, G., Freudiger, J., Sridhar, V.R., Davidson, D.: Learning new words (Mar 14 2017), uS Patent 9,594,741
34. Thakurta, A.G., Vyrros, A.H., Vaishampayan, U.S., Kapoor, G., Freudinger, J., Prakash, V.V., Legendre, A., Duplinsky, S.: Emoji frequency detection and deep link frequency (Jul 11 2017), uS Patent 9,705,908
35. Wang, N., Xiao, X., Yang, Y., Hoang, T.D., Shin, H., Shin, J., Yu, G.: Privtrie: Effective frequent term discovery under local differential privacy. In: IEEE 34th International Conference on Data Engineering (ICDE). pp. 821–832. IEEE (2018). https://doi.org/10.1109/icde.2018.00079
36. Wang, N., Xiao, X., Yang, Y., Zhao, J., Hui, S.C., Shin, H., Shin, J., Yu, G.: Collecting and analyzing multidimensional data with local differential privacy. In: IEEE 35th International Conference on Data Engineering (ICDE) (2019). https://doi.org/10.1109/icde.2019.00063
37. Wang, T., Blocki, J., Li, N., Jha, S.: Locally differentially private protocols for frequency estimation. In: USENIX Security Symposium. pp. 729–745 (2017)
38. Wang, T., Li, N., Jha, S.: Locally differentially private frequent itemset mining. In: IEEE Symposium on Security and Privacy (SP). pp. 127–143. IEEE (2018). https://doi.org/10.1109/sp.2018.00035
39. Wang, T., Ding, B., Zhou, J., Hong, C., Huang, Z., Li, N., Jha, S.: Answering multi-dimensional analytical queries under local differential privacy. In: Proceedings of the 2019 International Conference on Management of Data (SIGMOD). pp. 159–176. ACM (2019). https://doi.org/10.1145/3299869.3319891
40. Wang, T., Li, N., Jha, S.: Locally differentially private heavy hitter identification. IEEE Transactions on Dependable and Secure Computing (TDSC) (2019). https://doi.org/10.1109/tdsc.2019.2927695
41. Warner, S.L.: Randomized response: A survey technique for eliminating evasive answer bias. Journal of the American Statistical Association **60**(309), 63–69 (1965). https://doi.org/10.2307/2283137
42. Xu, S., Su, S., Xiong, L., Cheng, X., Xiao, K.: Differentially private frequent subgraph mining. In: IEEE 32nd International Conference on Data Engineering (ICDE). pp. 229–240. IEEE (2016). https://doi.org/10.1109/icde.2016.7498243
43. Yang, B., Sato, I., Nakagawa, H.: Bayesian differential privacy on correlated data. In: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data (SIGMOD). pp. 747–762. ACM (2015). https://doi.org/10.1145/2723372.2747643

44. Ye, Q., Hu, H., Meng, X., Zheng, H.: PrivKV: Key-value data collection with local differential privacy. In: IEEE Symposium on Security and Privacy (SP). pp. 317–331. IEEE (2019). https://doi.org/10.1109/sp.2019.00018
45. Zhang, J., Cormode, G., Procopiuc, C.M., Srivastava, D., Xiao, X.: Privbayes: Private data release via bayesian networks. In: Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data (SIGMOD). pp. 1423–1434. ACM (2014). https://doi.org/10.1145/2588555.2588573
46. Zhang, Z., Wang, T., Li, N., He, S., Chen, J.: CALM: Consistent adaptive local marginal for marginal release under local differential privacy. In: Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security (CCS). pp. 212–229. ACM (2018). https://doi.org/10.1145/3243734.3243742
47. Zheng, H., Ye, Q., Hu, H., Fang, C., Shi, J.: BDPL: A boundary differentially private layer against machine learning model extraction attacks. In: European Symposium on Research in Computer Security. pp. 66–83. Springer (2019)