TOPICDP - ENSURING DIFFERENTIAL PRIVACY FOR TOPIC MINING

BY

JAYASHREE SHARMA

Submitted in partial fulfillment of the
requirements for the degree of
Master of Science in Computer Science
in the Graduate College of the
Illinois Institute of Technology

Approved _____
Adviser

Chicago, Illinois
December 2021

ACKNOWLEDGEMENT

I would like to acknowledge and give my warmest thanks to my supervisor Prof. Yuan Hong, who made this work possible. His guidance and advice carried me through all the stages of the research project. I would like to specially mention and thank for the support from Han Wang, who helped me understand the nuances of differential privacy during the course of my thesis. I would also like to thank my committee members for your brilliant comments and suggestions, thanks to you.

I would also like to give special thanks to my husband Amit Sharma, daughter Gargi and my family as a whole for their continuous support and patience when undertaking my research and writing my project. Your support has always motivated and inspired me to work hard and move forward in my accomplishments.Finally, I would like to thank God, for showing path through all the challenges.

AUTHORSHIP STATEMENT

I, Jayashree Sharma, attest that the work presented in this thesis is substantially my own.

In accordance with the disciplinary norm of Computer Science(See IIT Faculty Handbook Appendix S), the following collaborations occurred in the thesis:

Prof. Yuan Hong devised the project and the main conceptual ideas as is the norm of a thesis advisor. He laid the foundation plan for the design of the privacy mechanism, theoretical analyses and the experimental evaluations.

Han Wang of Illinois Institute of Technology, Chicago, collaborated on the design of experiments and analysis of experiment results, theoretical analyses as well as the paper writing. Han Wang and I finished and submitted a co-first author paper based on the contents of this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Topic mining enables applications to recognize patterns and draw insights from text data, which can be used for applications such as sentiment analysis, building of recommender systems and classifiers. The text data can be a set of documents or emails or product feedback and reviews. Each document is analysed using probabilistic models and statistical analysis to discover patterns that reflects underlying topics.

TopicDP is a differentially private topic mining technique, which injects well-calibrated Gaussian noise into the matrix output of the topic mining model generated from LDA algorithm. This method ensures differential privacy and good utility of the topic mining model. We derive smooth sensitivity for the Gaussian mechanism via sensitivity sampling, which resses the major challenges of high sensitivity in case of topic mining for differential privacy. Furthermore, we theoretically prove the differential privacy guarantee and utility error bounds of TopicDP. Finally, we conduct extensive experiments on two real-word text datasets (Enron email and Amazon Product Reviews), and the experimental results demonstrate that TopicDP can generate better privacy preserving performance for topic mining as compared against other state-of-the-art differential privacy mechanisms.

CHAPTER 1

INTRODUCTION

Billions of documents are generated everyday via personal computers, email servers, IoT devices, cloud, among others. These documents include considerable amounts of information, and they are frequently collected and shared for text analysis in various applications. One important type of applications is to extract topics from those documents, which can facilitate sentiment analysis [1], opinion summarization [2], recommender systems [3], and anomalous texts detection [4]. Topic mining statistically analyzes a corpus of documents to identify the discussion topics in them. The text data in each document is analyzed using probabilistic models and statistical analysis to discover patterns for the underlying topics. Thus, topic mining is widely used in the above real-world applications.

However, when topics of those documents are extracted and shared to untrusted third parties for further analysis, it raises severe privacy risks since the untrusted recipients may re-identify the owners of those documents from dataset with a diverse set of possible background knowledge related to the user and his/her documents (e.g., some keywords in the documents, and linguistic patterns). Thus, privacy preserving solutions for topic mining should be explored.

A simple privacy-enhancing technique is to replace the real user IDs of these documents with pseudonyms. This has been proven to be vulnerable to re-identification attacks [5, 6] (e.g., the AOL data leak incident [7]). As a rigorous privacy model against arbitrary background knowledge known to the adversaries, *differential privacy* (DP) has been extensively studied to address the privacy concerns in the text data [7–9]. However, such techniques only considers the privacy of term frequencies or related quantities in the documents. In practice, topic mining with differential privacy for documents should be a more complicated function rather than

calculating the frequency of terms. Consider the context of email mining, where a corpus of emails exchanged among a set of users can be analysed to identify the topics being discussed or assess the sentiment among the users. In such scenario, the privacy of an individual with high volume of email exchange can be at risk. The identity of a user can be exposed by a user with partial knowledge of the context in the email exchange. In this work, we use differential privacy principles to safeguard the privacy of individuals participating in topic mining models. We attempt to eliminate risk of exposing the identities of the individuals due to direct correlation between the topic model and identity of users, who have contributed to the text corpus used for analysis.

In the current work, I propose a differentially private topic mining technique (namely TopicDP) that protects the privacy of individuals involved in the documents used for any topic mining model (model-agnostic). It ensures indistinguishable analysis result, derived from the input data do not expose the identity of any individual whose documents are used to generate a topic model. Specifically, I attempt to inject well-calibrated Gaussian noise into the result of topic mining (usually as a matrix with probability entries for different keywords in different topics), which would work regardless of the topic mining models. Thus, the untrusted recipient cannot distinguish whether any user is included in the documents or not.

Topic mining generates a matrix output in which each row represents a topic and each entry in the row is a keyword and its probability of occurrence in the topic. Therefore, different from generic differential privacy mechanisms on generic queries [10–12] , TopicDP addresses three major challenges:

- Topic mining generates a matrix output rather than an aggregated value, e.g., count, max, average, and sum.
- A single user's documents may include a unique word, which is not found in

other users' documents, this puts the user at high risk of re-identification.

- A single user may include a large number of documents, removal of such user data from the text corpus would change the topic model to a great extent, resulting in high sensitivity of the topic model.

Given these challenges, I define a novel privacy notion for protecting the individuals in the documents for topic mining: "$(\epsilon, \delta, \gamma)$-random differential privacy", which is a relaxed notion extended from ϵ-differential privacy. First, δ (close to 0, e.g., 0.0001) is used to ensure that the probability of generating any unique word in the output matrix is bounded by δ. Second, γ (close to 0, e.g., 0.02) is used to smoothen the sensitivity such that at least (1-γ) portion of users can be protected with ϵ-differential privacy, which ensures indistinguishability (bounded by e^{ϵ}) for the topic mining results regardless of the presence or absence of each user's all the documents. Moreover, we design algorithms for ensuring $(\epsilon, \delta, \gamma)$-random differential privacy.

1.1 Contributions Following are the major contributions of this work:

- The first model-agnostic differentially private technique TopicDP, to protect the privacy of users in the documents used for topic mining. The well-calibrated noise is generated for the matrix output of topic mining, and thus works for any topic mining algorithm (model-agnostic). It can also be readily extended for other machine learning models which generate matrix outputs.

- A novel differential privacy mechanism that smoothens the sensitivity of topic mining with sensitivity sampler. This new relaxed privacy notion could significantly improve the utility of topic mining while achieving ϵ-differential privacy with very high probability.

- Formal proof of the privacy and utility error bound.

- Extensive experiments to validate the performance of TopicDP on two real text datasets.

CHAPTER 2
BACKGROUND

2.1 Topic Mining

Topic mining discovers the patterns of words to represent the topics, each of which is defined as a set of words that frequently occur together in the text corpus. For instance, while mining topics from a bunch of emails, the email bodies are analyzed to identify the topics to discover patterns of word usage. There are many existing algorithms [13] used for topic mining, e.g., Latent Dirichlet Allocation (LDA) [14], Latent Semantic Analysis (LSA) [15], Probabilistic Latent Semantic Analysis (PLSA) [16], and Correlated Topic Model (CTM) [17]. Our TopicDP algorithm can guarantee differential privacy for any of the topic mining models.

Some generative probabilistic models (e.g., LDA) define the topic as a group of words that have a high likelihood of co-occurrence in the document corpus. These words can be noun, verb, adjective and adverb. Therefore, topic mining generates a set of keywords and their corresponding probabilities in the topic. Thus, the output of topic mining can be denoted as a matrix of keywords and their probabilities. Each row in the matrix represents a topic discussed in the text documents, and the entries in the row refer to the probabilities of different keywords in the given topic (the sum of the entries in each row is 1).

Formally, we denote the output of topic mining as a matrix W (rows: top m topics, columns: n words) with $m \times n$ probabilities $W \in [0,1]^{m \times n}$. Since different topics may involve different sets of keywords, n words are the union of the keywords in all m top topics, and the probability in the matrix would be 0 if any topic does not include such word.

2.2 Topic Mining Algorithms

2.2.1 LDA Algorithm. [14] LDA is a generative probabilistic model for collections of discrete data such as text corpora. The goal is to find short descriptions of the members of a collection that enable efficient processing of large collections while preserving the essential statistical relationships that are useful for basic tasks such as classification, novelty detection and so on. In the current experiment we use LDA algorithm to identify the topics from a collection of emails and a corpus of product reviews. The short descriptions are bigrams, a pair of words occurring together in various emails. The long descriptions are the topics of discussion in the emails.

A Topic is defined to be a distribution over a fixed vocabulary. The vocabulary is defined from the keywords in the data set.The statistical model reflects the intuition that documents exhibit multiple topics. Each document in the data set exhibits the topic in different proportion. We describe that the goal of LDA topic modeling is to automatically discover the topics from the collection of documents.The documents are used to infer the hidden topic structure, which comprises of the topics, per-document topic distributions and per document, per word topic assignments.The utility of the topic model stems from the property that the discovered topic structure resembles the thematic structure of the collection

We can describe LDA more formally with the following notation. The topic are β_k, where each β_k is a distribution over the vocabulary. The topic proportions for the d_{th} document are θ_d, where $\theta_{d,k}$ is the topic proportion for topic k in document d. The topic assignments for the d^{th} document are z_d, where $z_{d,n}$ is the topic assignment for the n^{th} word in the document d. Finally the observed words for document d are w_d, where $w_{d,n}$ is the n^{th} word in document d, which is an element from the fixed vocabulary.

Given these notations, the LDA algorithm corresponds to the joint distribution of the words as follows:

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D}) = \prod_{i=1}^{k} p(\beta_i) \prod_{d=1}^{D} p(\theta_d)(\prod_{n=1}^{N} p(z_{d,n}/\theta_d)p(w_{d,n}/\beta_{1:K}, z_{d,n}))$$

The given distribution specifies various dependencies such as the topic assignment $z_{d,n}$ which in turn depends on the per document topic proportions θ_d, the observed word $w_{d,n}$ which in turn depends on the topic assignment $z_{d,n}$ and all of the topics.

To compute the conditional distribution of the topic structure given the observed documents:

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D}) = \frac{p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D})}{p(w_{1:D})}$$

The numerator is the joint distribution of all random variables. The denominator is probability of seeing the observed corpus under any topic model. In theory, it is the total joint distribution over every possible instantiation of the hidden topic structure.

2.3 Differential Privacy

To protect the matrix outputs of topic mining, we first consider two input document datasets D and D' that differ in any user as two neighboring datasets. It is worth noting that any user may have multiple documents (e.g., all his/her emails) in the dataset. The DP model in case of topic mining would be interpreted as: adding or removing any user's all the documents should not cause significant changes to the output of the topic mining. Thus, the privacy risks resulted from each user's documents can be bounded, even if the adversary possesses arbitrary background

knowledge on all the users. Therefore, the DP model can be defined as below:

Definition 2.3.1 (ϵ-Differential Privacy)**.** *[13] A randomization algorithm $\mathcal{A}$ satisfies ϵ-differential privacy if for any two input datasets D and D' that differ in any user u (including at least one document), and for any output $S \in range(\mathcal{A})$, we have*

$$e^{-\epsilon} \leq \frac{Pr[\mathcal{A}(D) \in S]}{Pr[\mathcal{A}(D') \in S]} \leq e^{\epsilon}.$$

2.3.1 Differential Privacy Mechanisms. Differential privacy is based on strong mathematical foundations and hence has been widely used to achieve privacy guarantees in various applications. Privacy mechanisms such as Laplace mechanism, Gaussian mechanism and Exponential mechanism, derive independent and identically distributed(i.i.d) noise based on the values of sensitivity of the data set. The i.i.d noise may not be efficient in scenarios where it cannot make use of the underlying structure of the domain to preserve privacy.

Laplace, Gaussian mechanisms form the foundation for a whole spectrum of derived mechanisms based on concepts such as smooth sensitivity, elastic sensitivity, lossy compression, spatial decomposition and so on. Sensitivity based mechanisms use smooth sensitivity and elastic sensitivity to avoid scenarios of worst-case sensitivity to devise differential privacy mechanisms. These mechanisms are required to be tuned for each domain of application.

Laplace Mechanism adds noise drawn from the Laplace distribution scaled to the L1-sensitivity of the query function. It was initially designed for scalar-valued query functions, but can be extended to a matrix-valued query function by adding i.i.d. Laplace noise to each element of the matrix. It provides strong ϵ-differential privacy guarantee.

Gaussian Mechanism uses i.i.d. additive noise drawn from the Gaussian distribution scaled to the L2-sensitivity. It guarantees (ϵ, δ)-differential privacy. Like the Laplace mechanism, it also does not automatically consider the structure of the matrices.

Exponential Mechanism uses noise introduced via the sampling process. It draws its query answers from a custom distribution designed to preserve ϵ-differential privacy. To provide reasonable utility, the distribution is chosen based on the quality function, which indicates the utility score of each possible sample. Due to its generality, it has been utilized for many types of query functions, including the matrix-valued query functions. We experimentally compare our approach to the Exponential mechanism, and show that, with slightly weaker privacy guarantee, our method can yield significant utility improvement.

2.4 Sensitivity

Global Sensitivity refers to the maximum difference of the output of a query function that can result from a single change in the input data set.

$$GS_F(f) = \max_{neighbors(x,x')} \left|\left|f(x) - f(y)\right|\right|_1$$

Local Sensitivity measures variability in neighbourhood of specific data set. While global sensitivity is more generic to a wide variety of data sets, local sensitivity is specific for certain data set. It allows us to place finite bounds on the sensitivity of some functions where it is difficult to set boundaries on global sensitivity.

$$LS_F(f) = \max_{y:d(x,y)=1} \left|\left|f(x) - f(y)\right|\right|_1$$

In case of global sensitivity, noise magnitude depends on the values GSF and privacy parameter ϵ.

Smooth Sensitivity adds instance-specific noise with smaller magnitude than the worst-case noise as in case of global sensitivity. We define a class of smooth upper bounds *Sf* and *LSf*, such that adding noise proportional to *Sf* preserves utility and privacy of the result.

2.5 Preserving Privacy in Matrix-valued Queries

Matrix-valued queries refer to the two-dimensional results from a data set. One property that distinguishes the matrix-valued query functions from the scalar-valued query functions is the relationship and interconnection among the elements of the matrix. Laplace and Gaussian mechanisms can be extended to a matrix-valued query function by adding i.i.d. noise to each element of the matrix, this method is often sub-optimal as it forfeits an opportunity to exploit the structural characteristics typically associated with matrix analysis.

Multivariate Gaussian Mechanism [18], adds Gaussian noise scaled to the L2-sensitivity of the matrix-valued query function, while ensuring differential privacy.

Definition 2.5.1 (Multi-Variate Gaussian(MVG) Mechanism)**.** *[18] Given a matrix-valued query function $f(X) \in R_{mn}$, and a matrix-valued random variable, $Z \approx MVG_{m,n}(0, \Sigma, \Psi)$, the MVG mechanism is defined as,*

$$MVG(f(x)) = f(x) + Z$$

where Σ is row-wise co-variance matrix and Ψ is column co-variance matrix.

The choice of the row-covariance matrix(Σ) and column-variance matrix(Ψ) has to satisfy certain constraints to ensure differential privacy.

CHAPTER 3

RELATED WORK

Privacy preserving text analysis has been extensively studied. For instance, [19] proposes a privacy-preserving classification technique for personal text messages based on the secure multiparty computation, which encompasses both private feature extraction from texts, and private classification with logistic regression and tree ensembles. It proves that when using the secure text classification method, the application cannot learn anything about the texts, and the author of the text cannot learn anything about the text classification model either. [20] proposes a privacy-preserving keyword search scheme for searching over encrypted data. To avoids the high computational cost of asymmetric encryption, this scheme employs symmetric encryption and Bloom filter.

Several other works focus on the text analysis with differential privacy. Specifically, [9] proposes an automated text anonymization approach that produces synthetic term frequency vectors for the input documents with differential privacy. [21] presents a formal approach to preserve privacy in text perturbation using the notion of d_χ-privacy which is also extended from differential privacy. It considers the input distance between any two inputs of the domain to achieve indistinguishability. Some other works address privacy concerns in the Latent Dirichlet Allocation (LDA) training process [14,22,23]. For instance, [14] mainly proposes a HDP-LDA algorithm to protect the entire training process on centralized datasets. However, in their privacy model, the neighboring datasets only differ in one record and the HDP-LDA algorithm is based on the collapsed Gibbs sampling which adds noise to the word count statistics in the each iteration of the sampling process. [7, 8, 24] release search query logs with differential privacy while ensuring differential privacy for the query keywords and URLs. However, none of the above DP models can be applied for

TopicDP since they cannot inject noise to matrix outputs.

Finally, besides the data analysis performed on the textual data, differential privacy has been applied to many different applications, such as video queries [25,26], trajectory data publishing [27,28], and machine learning [29,30].

CHAPTER 4

TOPIC MINING WITH DIFFERENTIAL PRIVACY

4.1 Threat Model

We adopt the standard threat model setting, where a trusted data owner collects a large number of users' documents and perturbs the topic mining algorithm with DP guarantee. Thus, other untrusted data recipients (adversaries) would request the topic mining analysis on the documents, but cannot infer if any user's document (e.g., any user's email) is included in the topic mining even if the adversaries possess arbitrary background knowledge (e.g., knowing the contents of all the users' emails). We assume that all the parties are honest-but-curious to follow the procedures without maliciously manipulating the data.

4.2 Defining a Differential Privacy Mechanism

Before designing a DP mechanism for topic mining, we need to explore its sensitivity. We start from the global sensitivity which refers to the maximum output difference of the function applied to the neighboring datasets. For example, the L_2-norm sensitivity of the query function f for Gaussian mechanism is:

Definition 4.2.1 (Global Sensitivity)**.** *Given a function f, its L_2-sensitivity is defined as:*

$$\Delta_{gs}(f) = \max_{D,D'} \left|\left|f(D) - f(D')\right|\right|_2$$

As mentioned earlier, we have to address three challenges while designing TopicDP. First, topic mining will be considered as a complex function that generates a matrix output with probability entries. Thus, we first define the sensitivity for such function that returns matrix entries:

Definition 4.2.2 (Sensitivity for Matrix-Output Function)**.** *Given a matrix-output*

function $f(D) \in \mathbb{R}^{m \times n}$*, define the* L_2*-sensitivity as,*

$$\Delta(f) = \max_{D,D'} \left|\left| f(D) - f(D') \right|\right|_F$$

where $||\cdot||_F$ *is the Frobenius norm.*

Second, every user may include some unique words. Given any output P that includes any unique keyword from any user, the probabilities of applying topic mining to D and D' to generate such output P cannot be bounded with $\frac{Pr[\mathcal{A}(D)=S]}{Pr[\mathcal{A}(D')=S]} \leq e^{\epsilon}$ and $\frac{Pr[\mathcal{A}(D')=S]}{Pr[\mathcal{A}(D)=S]} \leq e^{\epsilon}$ since one of $Pr[\mathcal{A}(D) = S]$ and $Pr[\mathcal{A}(D') = S$ is equal to 0. Thus, the probabilities of such extreme cases should be bounded by a small number δ to ensure (ϵ, δ)-differential privacy.

Definition 4.2.3 ((ϵ, δ)-Differential Privacy)**.** *A randomization algorithm* $\mathcal{A}$ *satisfies* (ϵ, δ)*-DP if for any two input datasets* D *and* D' *that differ in any user* u *(including at least one document), and for any output set* $S \subseteq range(\mathcal{A})$*, we have* $Pr[\mathcal{A}(D) \in S] \leq e^{\epsilon} Pr[\mathcal{A}(D') \in S] + \delta$*, and vice versa.*

Third, the global sensitivity of topic mining might be very high since each user may include a large number of documents. Thus, we relax the protection to $(\epsilon, \delta, \gamma)$-random differential privacy(RDP) where the confidence parameter of satisfying (ϵ, δ)-DP is denoted as $\gamma \in [0, 1)$. Then, $1 - \gamma$ portion of the dataset satisfies ϵ-DP while the probability of generating any unique keyword in the output topics is bounded by δ.

Definition 4.2.4 ($(\epsilon, \delta, \gamma)$-Random Differential Privacy)**.** *A randomized mechanism* $\mathcal{A}$*:* $D^N \rightarrow \mathbb{R}$ *responding with values in arbitrary response set* $\mathbb{R}$ *preserves* $(\epsilon, \delta, \gamma)$*-RDP, at privacy level* $\epsilon > 0$*,* $\delta \in [0, 1)$*, and confidence parameter* $\gamma \in [0, 1)$*, if* $Pr[\forall S \subset \mathbb{R}, Pr(\mathcal{A}(D) \in S) \leq e^{\epsilon} \cdot Pr(\mathcal{A}(D') \in S) + \delta] \geq Pr(|f(D) - f(D')| \leq \Delta) \geq$

$1-\gamma$, *with the inner probabilities over the mechanism's randomization, and the outer probability over neighboring datasets* D, D'.

Intuitively, given the sensitivity $\Delta > 0$, when neighboring datasets D and D' satisfy $|f(D) - f(D')| \leq \Delta$, the randomization mechanism $\mathcal{A}(D, \epsilon, \delta, \gamma)$ ensures (ϵ, δ)-DP. Then, we have the probability of holding ϵ-DP is at least $(1-\delta)(1-\gamma)$ since the maximum leakage occurs if two leakages are disjoint. Thus, TopicDP satisfies such DP notion with minor relaxations since very small δ and γ make $(1-\delta)(1-\gamma)$ very close to 1.

4.3 The TopicDP Framework

Since TopicDP requires (ϵ, δ)-DP, we adapt the Gaussian mechanism in TopicDP which is widely used to ensure (ϵ, δ)-DP by injecting a Gaussian noise $\mathcal{N}(0, \sigma^2)$ to the query (or analysis function) where $\sigma^2 = 2\Delta^2 \log(1.25/\delta)/\epsilon^2$ and Δ refers to the sensitivity of the query/function. The noise generated by Gaussian mechanism will be calibrated with the sampled sensitivity.

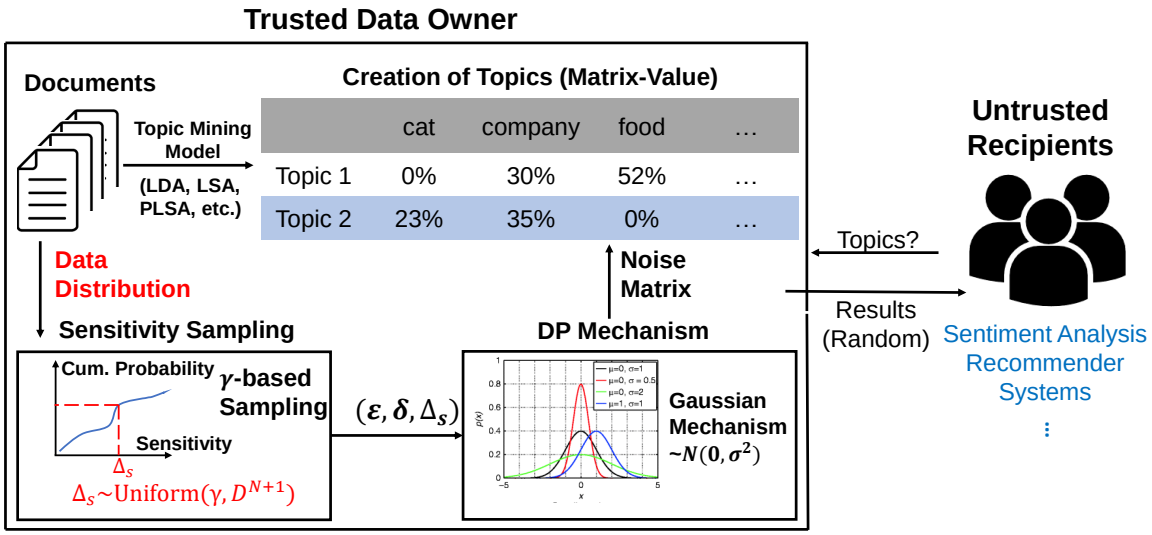


Figure 4.1. Overview of the TopicDP framework

Figure 4.1 illustrates the framework of TopicDP. We carefully inject Gaussian noise into the topic matrix (*output perturbation*) with the sampled sensitivity.

Thus, DP can be ensured for any topic mining algorithm (*model-agnostic*). TopicDP includes the following major steps.

Step 1 : The trusted data owner collects all the documents D from N users, and specifies the privacy parameters (ϵ, δ) and the confidence parameter γ.

Step 2 : The untrusted recipient requests to identify topics from all the documents D.

Step 3 : The trusted data owner applies any topic mining algorithm (e.g., LDA) to extract the topic matrix W with probability entries for a set of keywords.

Step 4 : The trusted data owner uniformly samples the smooth sensitivity Δ (denoted as Δ_s) of dataset D with parameter γ, and sequentially injects the noise with privacy parameters ϵ, δ and sampled Δ_s from Gaussian mechanism to the topic matrix W.

Step 5 : The noised topic matrix W' (normalized) is returned to the untrusted recipient for further analyses.

Note that the sensitivity sampling in Step 4 is extended from the Pain-Free algorithm [31]. We will discuss the details for smoothing the sensitivity using the distribution in the dataset D as follows.

4.4 Sensitivity Derivation

The sensitivity sampler in Pain-Free algorithm obviates the challenge of unbounded sensitivity in DP. With it, we can approximate the global sensitivity with

Input : Database size N, topic mining function f, the distribution P, the confidence parameter γ
Output: The sampling sensitivity Δ_s

Compute the sample size $h = \left\lceil \frac{\log(1/\rho)}{2(\gamma-\rho)^2} \right\rceil$ where $\rho = exp(W_{-1}(-\frac{\gamma}{2\sqrt{e}}) + 0.5)$
Compute the order statistic index $k = h(1 - \gamma + \rho + \sqrt{\log(1/\rho)/(2h)})$
$P \leftarrow Uniform()$
for $i = 1$ to h **do**
    Sample $D_{1,..,N-1} \sim P^N$, $D_N \sim P^N$ and $D_{N+1} \sim P^N$
    $D \leftarrow D_{1,..,N-1} \cup D_n$
    $D' \leftarrow D_{1,..,N-1} \cup D_{N+1}$
    $\Delta_i = \left|\left| f(D) - f(D') \right|\right|_F$
end for
Sort $\Delta_1, ..., \Delta_h$ with the ascending order
Return $\Delta_s = \Delta_k$

Algorithm 1: Sensitivity Sampling

very high probability, assuming only oracle access to the target query function f evaluations.

Algorithm 1 presents the detail of sampling sensitivity of topic mining. With the given confidence parameter γ, we compute the value of sampling size h and order index k, which can guarantee $(\epsilon, \delta, \gamma)$-RDP. These two parameters are involved in the the Lambert-W function [32]. The distribution P is chosen to match the desired distribution of dataset D. There are a number of natural choices for the dataset distribution P. We use the uniform distribution for our documents since the Pain-Free algorithm and its privacy guarantee are derived by assuming a uniform distribution defined over the domain D. However, the accuracy should be significantly boosted if we can approximate the distribution of dataset with some background knowledge (e.g., users' linguistic patterns).

With the distribution P, in each iteration, we independently sample the $N + 1$ records from the domain to construct the database D and D' which differ in one user. Then, the sensitivity can be computed for these two neighboring datasets. After h iterations, there are h sensitivities and sort them in an ascending order. The

final smooth sensitivity should be the kth sensitivity.

4.5 Algorithm

After computing the smooth sensitivity [33], the DP algorithm can be designed. Algorithm 2 presents the details for topic mining (matrix outputs) with $(\epsilon, \delta, \gamma)$-RDP. Since there are mn entries in the probability matrix, the privacy budget ϵ is equally allocated to all the entries to generate an $m \times n$ noise matrix. We then add the noise matrix to the topic mining output (the probability matrix) and release the noisy result to untrusted recipients. It is worth noting that the estimation of sampled sensitivity can always be the same for a specific domain. Thus, such sensitivity sampling could be performed entirely in an offline stage and executed once.

Input : Database D, the query size $m \times n$, the sampled sensitivity Δ_s, the privacy parameter ϵ and *gamma*
Output: The noisy matrix result W'
1: Apply the LDA algorithm to extract the m topics and keywords probabilities W of dataset D
2: Allocate the privacy budget to each matrix entry: $\epsilon' = \frac{\epsilon}{m \times n}$
3: **for** each element e_i in the matrix W **do**
4: $\quad \tilde{e}_i \leftarrow e_i + \mathcal{N}(0, \sigma^2)$ and $\sigma^2 = 2\Delta^2 \log(1.25/\delta)/(\epsilon')^2$
5: $\quad$ Update the e_i with $\tilde{e}_i$
6: **end for**
7: $W' \leftarrow W$
8: **Return** the noisy matrix W'

Algorithm 2: Generating Noise with Gaussian Mechanism

4.6 Privacy and Utility Analysis

We first analyze the privacy bound of TopicDP.

Theorem 4.6.1. *: Consider any non-private function f: $D^N \rightarrow \mathcal{B}$, any sensitivity-induced (ϵ, γ)-differentially private mechanism mapping $\mathcal{B}$ to (randomised) responses in $\mathbb{R}$, any database D of N records, privacy parameters $\epsilon > 0, \delta \in [0, 1], \gamma \in (0, 1)$, and sampling parameters size $h \in \mathbb{N}$, order statistic index $h \geq k \in \mathbb{N}$, approximation*

confidence $0 < \rho < min\{\gamma, 1/2\}$*, distribution* P *on* D*. If*

$$h \geq \left\lceil \frac{\log(1/\rho)}{2(\gamma - \rho)^2} \right\rceil$$

$$k \geq h\big(1 - \gamma + \rho + \sqrt{\log(1/\rho)/(2h)}\big)$$

then Algorithm 1 running with D, Δ_s, f, h, k, P*, preserves* $(\epsilon, \delta, \gamma)$*-random differential privacy.*

Theorem 4.6.1 proves that the probability of RDP of $\mathcal{A}_{\Delta_s}$ when run on fixed Δ_s is at least $Pr(G < \Delta_s)$ where G is the sampled sensitivity group $\Delta_1, ..., \Delta_h$ drawn from the Algorithm 1. Then, by the Dvoretzky-Kiefer-Wolfowitz inequality, we can prove the probability of RDP of $\mathcal{A}_{\Delta_s}$ is at least $1 - \gamma$. Next, we will prove that Algorithm 2 satisfies $(\epsilon, \delta, \gamma)$-RDP with Theorem 4.6.1.

Theorem 4.6.2. *The noisy matrix output from Algorithm 2 satisfies* $(\epsilon, \delta, \gamma)$*-random differential privacy.*

Proof. It is straightforward to prove that Algorithm 2 satisfies $(\epsilon, \delta, \gamma)$-RDP. Based on Theorem 4.6.1, each entry in the matrix W satisfies the $(\frac{\epsilon}{mn}, \delta, \gamma)$-RDP since the privacy budget is equally allocated to mn entries. With the sequential composition of DP [34], adding noise to mn entries of W ensures $(\epsilon, \delta, \gamma)$-RDP. □

Then, we analyze the utility error bound of TopicDP.

Theorem 4.6.3. *The expectation of the amplitude of noise in TopicDP is* $\frac{2\sigma}{\sqrt{2\pi}}$ *where* $\sigma = \sqrt{2\Delta_s^2 \log(1.25/\delta)/(\epsilon)^2}$.

Proof. See Appendix A.1.2. □

In the current work, we inject the same well-calibrated Gaussian noise to all the matrix entries to protect the worst case that the involved user records of each topic in the matrix W are all correlated. Indeed, in practice, some real-world topics might be disjoint (e.g., from completely different users), then the matrix entries may follow parallel composition along with sequential composition. Then, higher privacy budgets can be allocated to certain entries to further improve the utility.

CHAPTER 5
EXPERIMENTS OF VARIOUS DATASETS

The performance of TopicDP on a topic model generated by LDA algorithm is evaluated on two datasets - Enron email dataset [35] and Amazon product review dataset [36].

5.1 Experimental Setup

TopicDP is evaluated on two different datasets:

1. **Enron Email Dataset** [35] was collected by the CALO Project. It contains data from about 158 users, mostly senior management of Enron, organized into folders and contains thousands of mails exchanged among the employees. The dataset has been processed to create a Kaggle dataset, comprising of a .csv file. This .csv file has to be further processed to suit the application. The dataset has been pre-processed to remove the email headers and duplicate emails. Figure 2 shows the three attributes of the dataset ("body", "to", "from"). The attribute "to" is used to identify the email recipients (users). All the emails associated to each "to" email address will be considered as a specific user's emails. The topic mining algorithm (e.g., LDA) is applied to the "body" field of the dataset.

   We require the "to" mail address, "from" mail address and email "body". We treat each distinct email address of "from" attribute as a user and all the emails from this user as the email records. LDA algorithm creates the topic model using the "body" field of the dataset. The main goal of adding noise to the output of the topic model and remove any direct correlation between the topics in the topic model and 'from' field in the data set.

2. **Amazon Product Review Dataset** [36] is a collection of product reviews created by users on the product pages. It includes reviews (ratings, text,

body	to	from
Here is our forecast	tim.belden@enron.com	randall.gay@enron.com
Traveling to have a business meeting takes the...	john.lavorato@enron.com	phillip.allen@enron.con
test successful. way to go!!!	leah.arsdall@enron.com	tim.belden@enron.com
Randy, Can you send me a schedule of the salary..	randall.gay@enron.com	phillip.allen@enron.con
Congratulations!! Your guys played very well....	greg.piper@enron.com	randall.gay@enron.com

username	reviews
llyyue	I thought it would be as big as small paper bu...
Charmi	This kindle is light and easy to use especiall...
johnnyjojojo	Didnt know how much i'd use a kindle so went f...
Kdperry	I am 100 happy with my purchase. I caught it o...
Johnnyblack	Solid entry level Kindle. Great for kids. Gift...

Figure 5.1. The Attributes of Datasets

helpfulness votes), product metadata (descriptions, category information, price, brand, and image features), and links (also viewed/also bought graphs). We pre-processed the dataset to retain two attributes of this dataset. Figure 2 shows some retained attributes: user name and text reviews.

Each user has at least one review for a specific product. The topic mining algorithm (e.g., LDA) is applied to the "reviews" of the dataset.

The goal of the privacy model is to safeguard the privacy of the field 'reviews.username'. By adding noise to the topic model, we remove direct correlation between the topics and the username to safeguard the privacy of the user.

The performance of the noise generation model proposed is evaluated against MVG mechanism. The noise generated by PG mechanism and the MVG mechanism are compared on three parameters - L1 distance, Kendall's tau distance and Root Mean Square Error(RMSE). These parameters reflect the utility of the topic model after addition of noise.

Table 5.1. Characteristics of Experimental Datasets

Dataset	User #	Docs #	Avg Word #	Avg Docs #/User
Enron	158	24,151	154	152
Amazon	3,815	50,000	31	13

Table 5.1 presents the characteristics of the datasets. To evaluate the utility

of TopicDP, we perform three groups of experiments. First, we adopt *the L_1-distance* and *Root-Mean-Square Error (RMSE)* metric to quantify the noise. Specifically, we compare the original output probability matrix and noisy output probability matrix with these two metrics. Second, we use the *Kendall's Tau* distance to evaluate the misalignment between the keyword ranking before and after adding noise. Finally, we visualize the keyword distributions in some example topics before and after noise.

Moreover, there are some parameters in TopicDP: privacy parameters ϵ and δ, confidence γ, and output matrix size $m \times n$ (top m topics and n keywords in each topic). Thus, we study the effectiveness of TopicDP by varying ϵ, γ, and m, n by setting δ as a very small probability upper bound 0.0001 and $m = n$ since one of the benchmarks Multivariate Gaussian (MVG) mechanism only works for square matrix outputs.

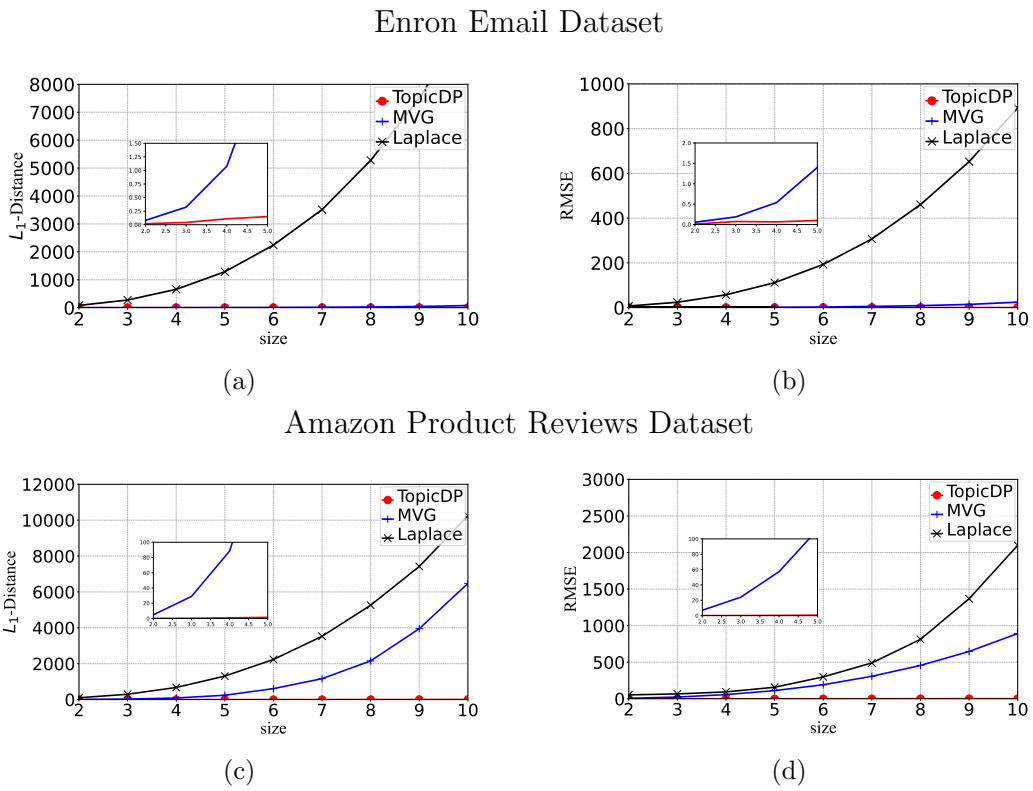


Figure 5.2. L_1-Distance and RMSE vs Matrix Size with fixed $\epsilon = 3$ on the Enron Dataset (a, b) and Amazon Dataset (c, d).

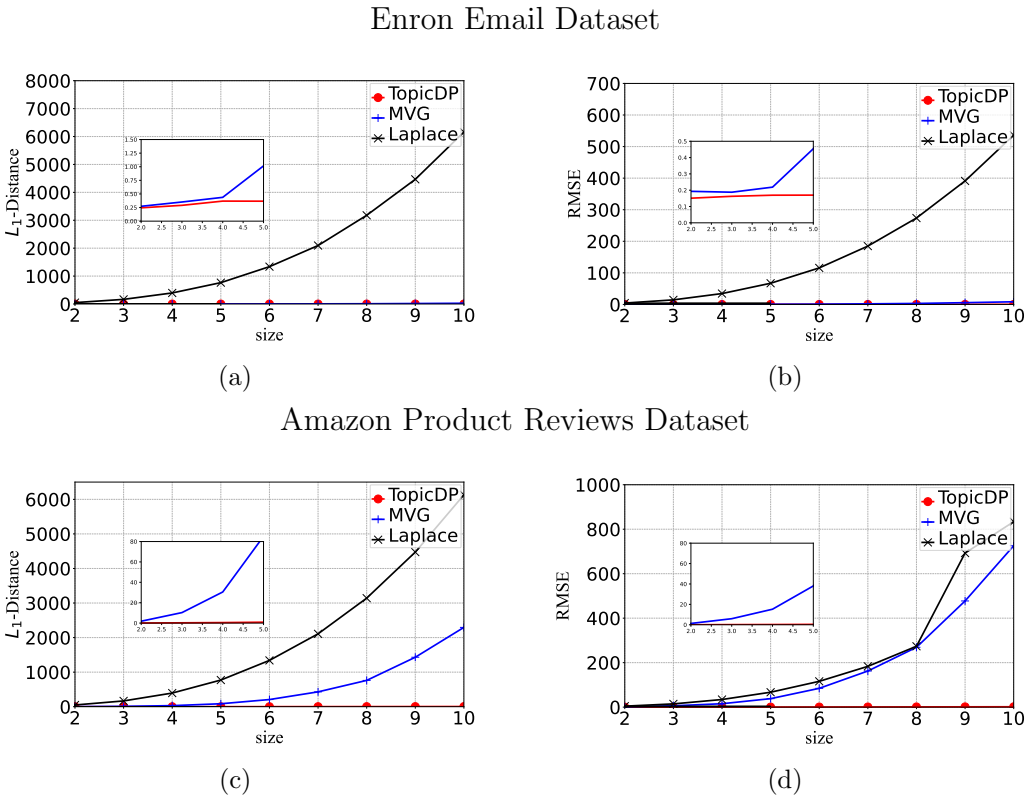


Figure 5.3. L_1-Distance and RMSE vs Matrix Size with fixed $\epsilon = 5$ on the Enron Dataset (a, b) and Amazon Dataset (c, d).

5.2 Evaluating TopicDP

In this section, we examine the utility of TopicDP. We compare the TopicDP with the well-known Laplace mechanism (adapted for matrix outputs) and MVG mechanism which also aims to protect privacy for matrix outputs.[1] Specifically, in the Laplace mechanism, we use the L_1 sensitivity and use an approximate global sensitivity $2m$ based on the extreme case that the topics and keywords are totally different (the probability difference of each topic should be 2). The setting of the MVG mechanism is the same as TopicDP in which the neighboring datasets differ by a single

[1]LDA Algorithm uses Laplace mechanism for the LDA algorithm, which cannot be model-agnostic for multiple topic mining algorithms (due to input/query perturbation). Thus, we adapt it (Laplace mechanism) to output perturbation for fair comparisons with TopicDP and the MVG mechanism (also model-agnostic).

user and the sensitivity may be unbounded. However, in the MVG mechanism, there is a threshold for the unbounded sensitivity. For a fair comparison, we also use the smooth sensitivity in the MVG mechanism.

First, Figure 5.6 demonstrates the L_1-distance and RMSE by varying the privacy bound ϵ from 1 to 10 with a step of 0.5 while fixing the confidence $\gamma = 0.1$ and output matrix size as 10×10. Figure 5.6(a) and 5.6(b) show the L_1-distance and RMSE results on the Enron Email Dataset, respectively. Similarly, Figure 5.6(c) and 5.6(d) demonstrate the results on the Amazon Product Review Dataset. As ϵ increases, the L_1-distance and RMSE decrease (noise gets smaller for all mechanisms while increasing ϵ). Second, the noise generated by Laplace mechanism (output perturbation) is far larger than other two mechanisms due to the very large global sensitivity. Third, in Figure 5.6(a), given a small ϵ (strong privacy), the L_1-distance between the actual output and the noisy output is greater than 2500 in MVG mechanism but less than 6 in our TopicDP. For large ϵ (weak privacy, e.g., $\epsilon = 10$), the L_1-distance for the MVG mechanism is still much higher than TopicDP. Similarly, we can also observe such trend from Figure 5.6(b), 5.6(c) and 5.6(d).

Second, Figure 5.7 shows the L_1-distance and RMSE by varying the confidence parameter γ from 0.02 to 0.22 with a step of 0.05 and fixing $\epsilon = 3$ and output matrix size 10×10. Figure 5.7(a), 5.7(b), 5.7(c), and 5.7(d) show the results for L_1-distance and RMSE, respectively. Since the global sensitivity is not related to γ, the sensitivity and the utility of Laplace mechanism should not be changed. We can observe that, as the γ increases, the L_1-distance and RMSE of other two mechanisms generate smaller noise. The main reason is that smaller γ gives stronger privacy by ensuring ϵ-differential privacy for a higher percent of records (thus the sampled sensitivity should be larger). Second, although the MVG mechanism decreases drastically as γ increases, the lowest result of MVG is still higher than the result of TopicDP

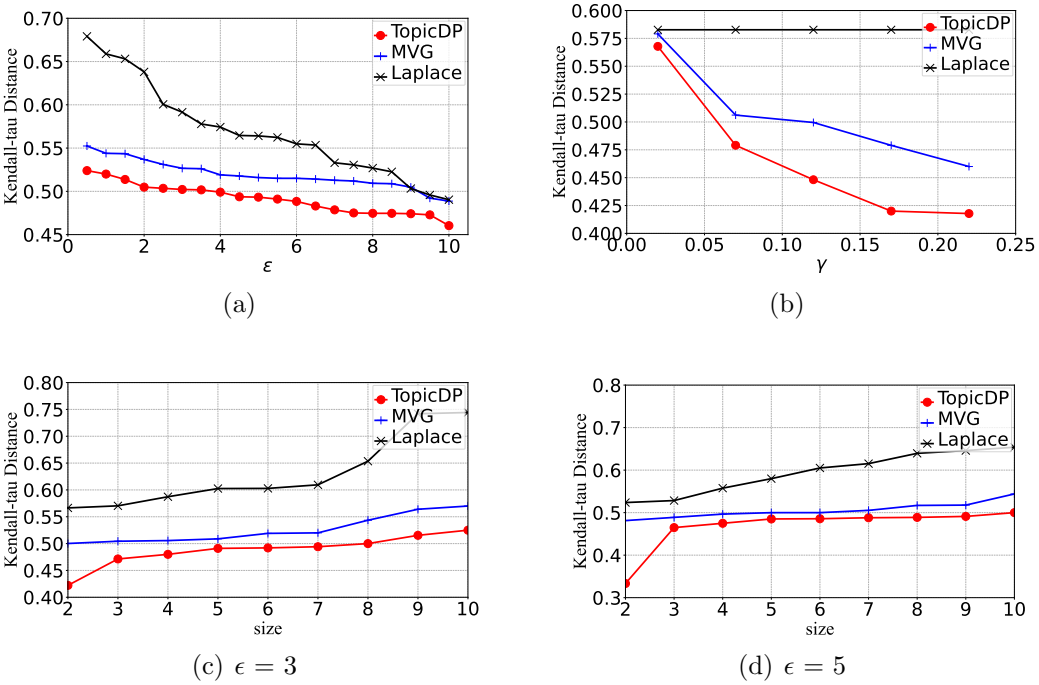
(e.g., the range of L_1 distance for TopicDP is from 0.1389 to 8.4426 whereas the range for MVG is from 4.8034 to 6475.0496, as shown in Figure 5.7(a)).

Third, Figure 5.2 shows the L_1-distance and RMSE by varying the matrix size $m \times n$. We fix $\epsilon = 3$ and $\gamma = 0.1$. Both m and n vary from 2 to 10 with a step of 1. For both datasets, as m, n get larger, the L_1-distance exponentially increases for both Laplace and MVG, whereas TopicDP increases much slower. The increasing trends on growing $m \times n$ is consistent with the the sequential composition for adding noise to the matrix entries. Larger m and n mean less privacy budget allocated for each entry and the utility should be worse. Clearly, the Laplace mechanism generates a much larger noise than other two mechanisms.

Thus, we will compare the utility of MVG and TopicDP. In Figure 5.2(a), the range of L_1-distance for TopicDP is from 0.0179 to 0.8954, whereas the range of L_1-distance for MVG is from 0.0794 to 77.1487. In Figure 5.2(b), the range of RMSE for TopicDP is from 0.0122 to 0.2838, whereas it is from 0.0562 to 24.3965 for MVG. When the matrix size is small (e.g., 2×2 and 3×3), the noise generated by these two mechanisms can be similar in terms of the L_1-distance and RMSE metric. However, when the matrix output size is large (e.g., greater than 5), TopicDP significantly outperforms MVG. We can draw same conclusions from Figure 5.2(b), 5.2(c) and 5.2(d) as well.

Furthermore, in Figure 5.3, we make the privacy bound $\epsilon = 5$, and compare the results with Figure 5.2. Then, all metric values get smaller, and this is consistent to the trends in Figure 5.6. Thus, we can conclude that TopicDP greatly outperforms MVG and Laplace mechanisms.

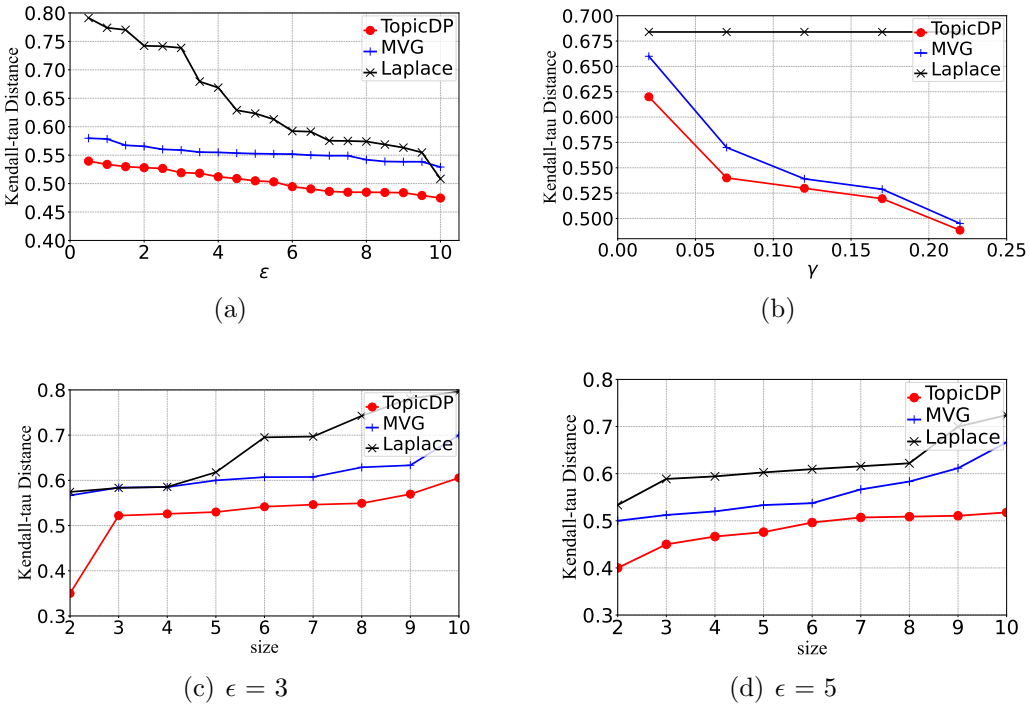
In the second group of experiments, we consider the utility by the rank of keywords in each topic before and after adding noise. To evaluate this, we use the



(a) (b)

(c) $\epsilon = 3$ (d) $\epsilon = 5$

Figure 5.4. Kendall's Tau Distance on the Enron Dataset.



(a) (b)

(c) $\epsilon = 3$ (d) $\epsilon = 5$

Figure 5.5. Kendall's Tau Distance on the Amazon Dataset.

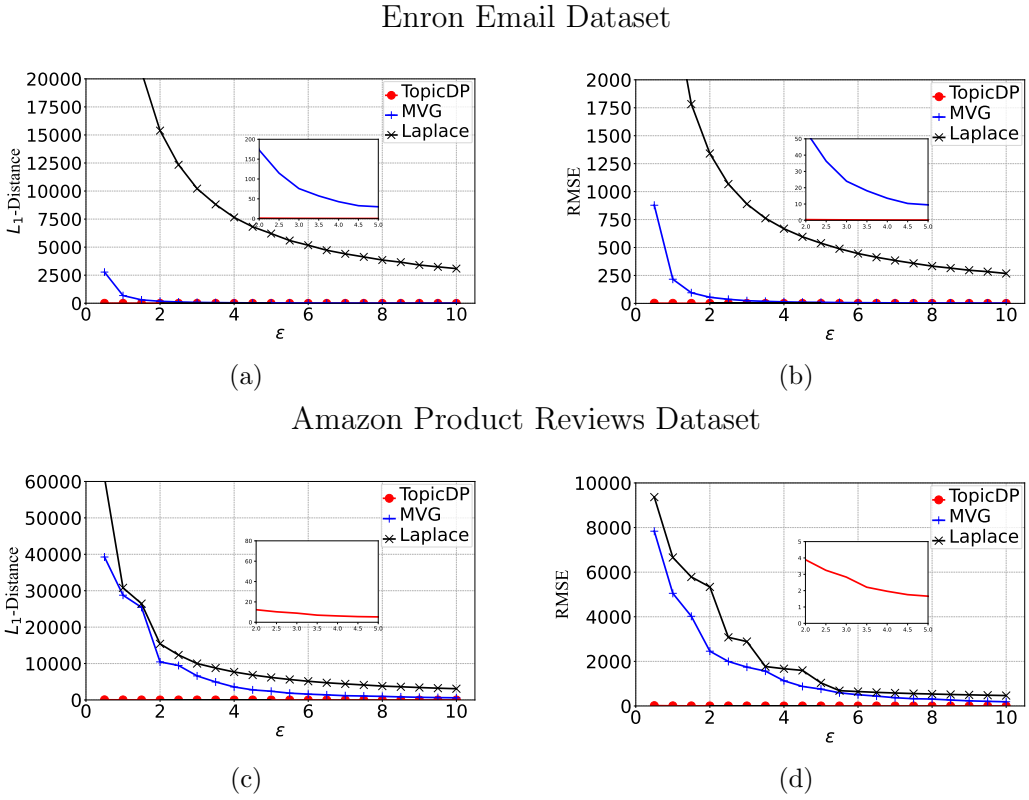


Figure 5.6. L_1-Distance and RMSE vs Privacy Bound ϵ on the Enron Dataset (a, b) and Amazon Dataset (c, d).

Kendall's Tau distance to measure the misalignment between every two sets of ranked keywords. The larger distance means higher dissimilarity. In this group, we also test how these three parameters affect the utility.

Figure 5.4 shows the Kendall's Tau distance on the Enron Email dataset. With a larger ϵ, the Kendall's Tau distance is slightly smaller (see Figure 5.4(a)). This is consistent what we observe in the results of L_1-distance and RMSE. We also observe that TopicDP outperforms MVG and Laplace mechnaisms as well. Furthermore, we can see the result gets smaller as γ goes larger in Figure 5.4(b). Finally, in Figure 5.4(c) and 5.4(d), the distance increases as the output matrix size gets larger, but TopicDP always has the smallest distance for the best utility. In Figure 5.5, four sub-figures illustrate the similar observations and trends on the Amazon review dataset.

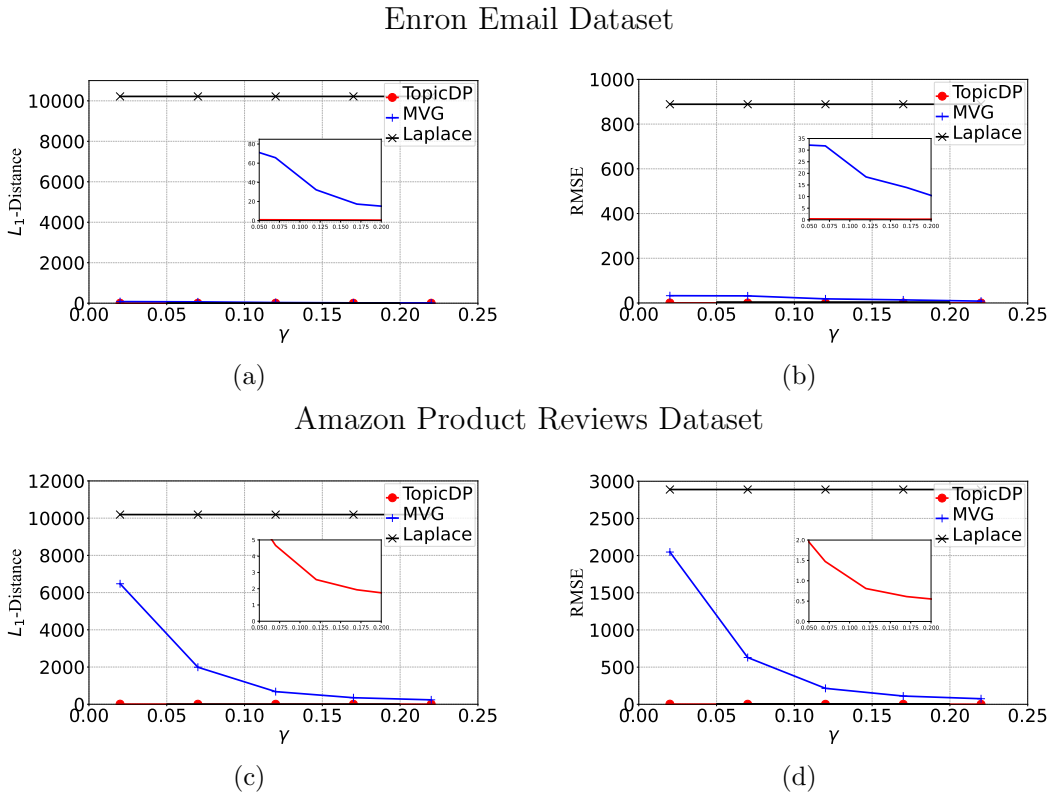
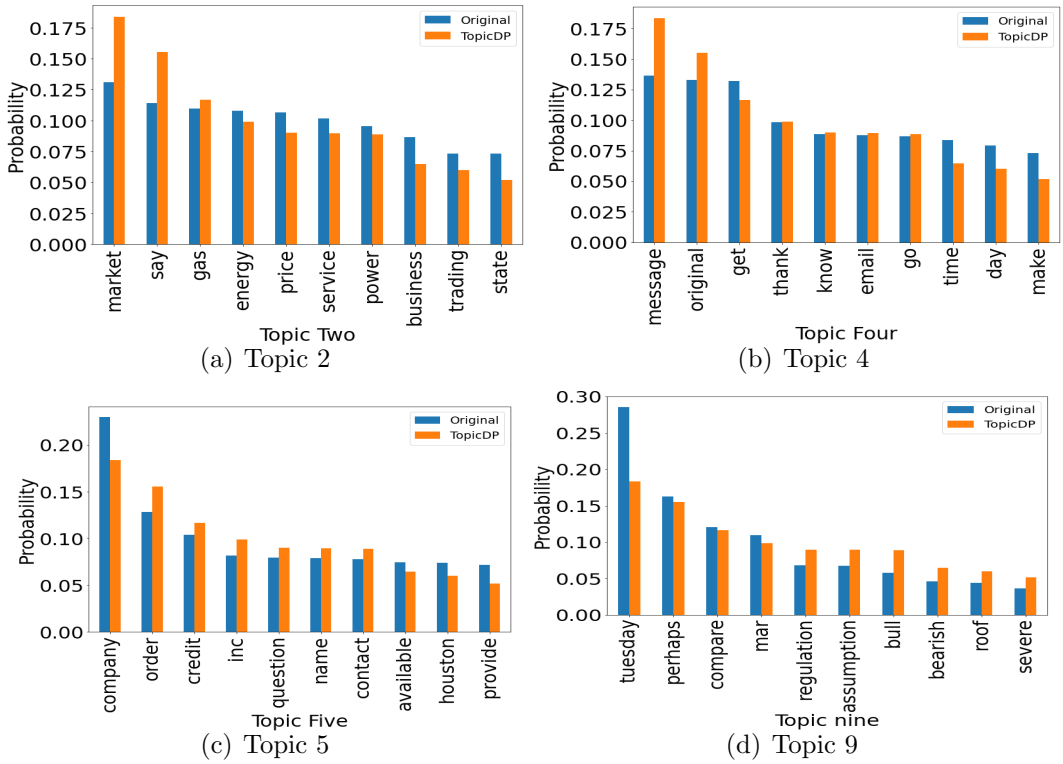


Figure 5.7. L_1-Distance and RMSE vs γ on the Enron Email Dataset (a, b) and Amazon Product Review Dataset (c, d).

In the third group of experiments, we visualize specific topics extracted from two datasets and show the keyword distributions of specific topics before and after adding noise. In Figure 5.8, we randomly pick four topics $2, 4, 5$ and 9 from the Enron dataset, and show their keyword distributions. Also, in Figure 5.9, we randomly pick four topics $2, 4, 6$ and 10 from the Amazon dataset, and show their keyword distributions. From these figures, we can see that the keyword distributions after injecting the noise using TopicDP are still close to the original distributions. This proves the practicality of the TopicDP. On the contrary, the set of keywords and the corresponding probabilities have been significantly obfuscated while injecting noise with the Laplace and MVG.

In summary, such experimental results validate that TopicDP retains good utility for topic mining (much better than the two benchmarks) with rigorous privacy

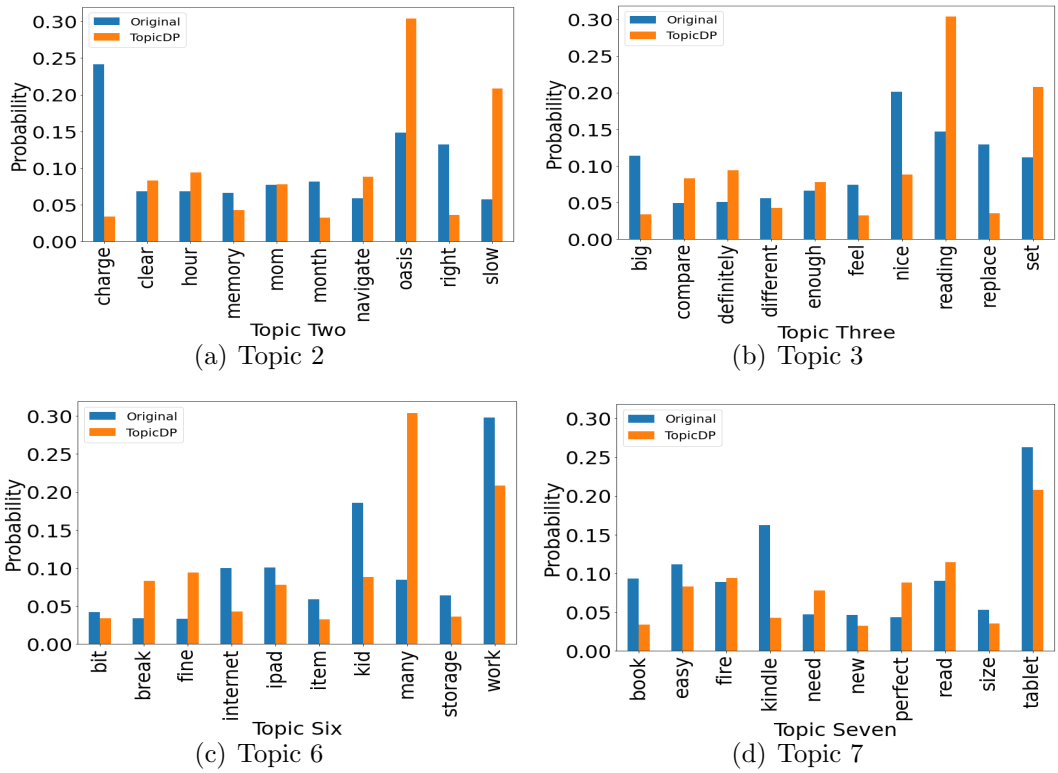


(a) Topic 2 (b) Topic 4

(c) Topic 5 (d) Topic 9

Figure 5.8. Keyword Distribution of Four Randomly Selected Topics in Enron Dataset

guarantees.



(a) Topic 2 (b) Topic 3

(c) Topic 6 (d) Topic 7

Figure 5.9. Keyword Distribution of Four Randomly Selected Topics in Amazon Dataset

CHAPTER 6

CONCLUSION AND FUTURE WORK

There is a high risk on re-identifying individuals from the topic mining on documents with certain background knowledge. To our best knowledge, we propose the first differentially private topic mining technique that injects well-calibrated Gaussian noise (with smooth sensitivity) to the output of any topic mining algorithm. It can ensure high utility and guarantee differential privacy for at least $1 - \gamma$ portion of records (γ is close to 0). The noisy result can be privately shared to any untrusted recipient for any further analysis.

In the future, first, we will empirically study TopicDP with multiple topic mining models (besides the LDA) to validate the model-agnostic property of the proposed DP mechanism. Second, we will experimentally validate the utility of the output matrices (generated by TopicDP) with more downstream analyses, such as sentimental analysis and recommender systems. Such experiments are expected to further confirm the high utility of TopicDP in a wide variety of applications. Third, we will investigate whether the recent reconstruction attack [37] to compromise privacy in the domain of natural language processing can be applied to the topic mining models, and we will evaluate how differential privacy can defend against such attack. Finally, we also plan to investigate the potential adversarial attacks on the topic mining models [38, 39] and propose the defense methods against them.

APPENDIX A

THEOREMS FOR PRIVACY GUARANTEE

A.1 Proofs

A.1.1 Proof of Theorem 4.6.1.

Proof. Assuming that the sampled sensitivities $\Delta_1, \cdots, \Delta_h$ are sorted as $\Delta_1 \leq \cdots, \leq \Delta_h$, given any $\rho' \in (0, 1)$ satisfying the following:

$$1 - \gamma + \rho + \rho' \leq 1 \Leftrightarrow \rho' \leq \gamma - \rho$$

Then, the random sensitivity $\Delta_s = \Delta_k$, where $h(1 - \gamma + \rho + \rho')$, is the smallest $\Delta \geq 0$ such that $\Phi_h(\Delta) \geq 1 - \gamma + \rho + \rho'$. Thus, we have:

$$\Phi_h(\Delta_s) = \frac{1}{h} \sum_{i=1}^{h} \mathbb{1}(\Delta_i \leq \Delta_s) \geq 1 - \gamma + \rho + \rho'$$

Define the events as

$$A_{\Delta_s} = \{\forall S \subset \mathbb{R}, Pr(\mathcal{A}_{\Delta_s}(D) \in S) \leq$$
$$e^{\epsilon} \cdot Pr(\mathcal{A}_{\Delta_s}(D') \in S) + \delta\}$$
$$B_{\rho'} = \left\{ \sup_{\Delta_s} (\Phi_h(\Delta_s) - \Phi(\Delta_s)) \leq \rho' \right\}$$

where $\Phi(\Delta_s)$ is the unknown CDF and $\Phi_h(\Delta_s)$ is the corresponding random empirical CDF. The former is the event that DP holds for a specific DB pair, when the mechanism is executed with (possibly random) sensitivity parameter Δ_s; the latter records the empirical CDF uniformly one-sided approximating the CDF to level ρ'. Moreover, per the definition of differential privacy, we have

$$\forall \Delta_s > 0, \quad Pr_{D,D' \sim P^{N+1}}(A_{\Delta_s}) \geq \Phi(\Delta_s).$$

The random D, D' on the left-hand side induce the distribution on Δ_s on the right-hand side under which $\Phi(\Delta_s) = Pr(\Delta \leq \Delta_s)$. The probability on the left-hand side is the level of random differential privacy of $\mathcal{A}_{\Delta_s}$ while running on the fixed Δ_s. By the Dvoretzky-Kiefer-Wolfowitz inequality [40], we have: for all $\rho' \geq \sqrt{(\log 2)/(2h)}$,

$$Pr_{\Delta_1,\cdots,\Delta_h}(B_{\rho'}) \geq 1 - e^{-2h\rho'^2}$$

Thus, we have

$$\begin{aligned}
&Pr_{D,D',\Delta_1,\cdots,\Delta_h}(A_{\Delta_s}) \\
=&\mathbb{E}(\mathbb{1}[A_{\Delta_s}]|B_{\rho'})Pr(B_{\rho'}) + \mathbb{E}(\mathbb{1}[A_{\Delta_s}]|\bar{B}_{\rho'})Pr(\bar{B}_{\rho'}) \\
\geq&\mathbb{E}[\Phi_h(\Delta_s)|B_{\rho'}]Pr(B_{\rho'}) \\
\geq&\mathbb{E}[\Phi_h(\Delta_s) - \rho'|B_{\rho'}](1 - e^{-2h\rho'^2}) \\
\geq&(1 - \gamma + \rho + \rho' - \rho')(1 - e^{-2h\rho'^2}) \\
\geq&(1 - \gamma + \rho)(1 - \rho) \\
\geq&1 - \gamma + \rho - \rho \\
=&1 - \gamma
\end{aligned}$$

The last inequality subjects to $\rho < \gamma$; the penultimate inequality subjects to the setting

$$\rho' \geq \sqrt{(\log \frac{1}{\rho})/(2h)}$$

and then the DKW condition, $\rho' \geq \sqrt{(\log 2)/(2h)}$, is satisfied given $\rho \leq 1/2$. □

A.1.2 Proof of Theorem 4.6.3.

Proof. In this paper, we use the Gaussian mechanism with sampled sensitivity. The Gaussian distribution has a probability density function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

where $\sigma = \sqrt{2\Delta_s^2 \log(1.25/\delta)/(\epsilon)^2}$. Then, the expectation of the amplitude of noise by Gaussian distribution is

$$\begin{aligned}
\mathbb{E}(V) &= \int_{-\infty}^{\infty} |x| f(x) dx \\
&= \int_{0}^{\infty} 2x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} dx \\
&= \frac{-2\sigma}{\sqrt{2\pi}} \left(\int_{0}^{\infty} de^{-\frac{x^2}{2\sigma^2}} \right) \\
&= 0 - \frac{-2\sigma}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} |_0 \\
&= \frac{2\sigma}{\sqrt{2\pi}}
\end{aligned}$$

We observe that the noise is related to the sampled sensitivity that is dependent on the confidence γ. □

BIBLIOGRAPHY

[1] B. Liu, "Sentiment analysis and opinion mining," *Synthesis lectures on human language technologies*, vol. 5, no. 1, pp. 1–167, 2012.

[2] V. B. Raut and D. Londhe, "Opinion mining and summarization of hotel reviews," in *Computational Intelligence and Communication Networks*, 2014, pp. 556–559.

[3] C. Wang, M. Zhang, W. Ma, Y. Liu, and S. Ma, "Modeling item-specific temporal dynamics of repeat consumption for recommender systems," *The World Wide Web Conference*, pp. 1977–1987, 2019.

[4] A. Manolache, F. Brad, and E. Burceanu, "DATE: detecting anomalies in text via self-supervision of transformers," *CoRR*, vol. abs/2104.05591, 2021.

[5] A. Narayanan and V. Shmatikov, "Robust de-anonymization of large sparse datasets," in *IEEE Symposium on Security and Privacy*, 2008, pp. 111–125.

[6] M. Jawurek, M. Johns, and K. Rieck, "Smart metering de-pseudonymization," in *Proceedings of the annual computer security applications conference*, 2011, pp. 227–236.

[7] A. Korolova, K. Kenthapadi, N. Mishra, and A. Ntoulas, "Releasing search queries and clicks privately," in *Proceedings of the World Wide Web*, 2009, pp. 171–180.

[8] Y. Hong, J. Vaidya, H. Lu, P. Karras, and S. Goel, "Collaborative search log sanitization: Toward differential privacy and boosted utility," *IEEE Trans. Dependable Secur. Comput.*, vol. 12, no. 5, pp. 504–518, 2015.

[9] B. Weggenmann and F. Kerschbaum, "Syntf: Synthetic and differentially private term frequency vectors for privacy-preserving text mining," in *Research & Development in Information Retrieval*, 2018, pp. 305–314.

[10] S. Ghane, L. Kulik, and K. Ramamohanarao, "Publishing spatial histograms under differential privacy," in *Proceedings of the Scientific and Statistical Database Management*, 2018, pp. 1–12.

[11] F. D. McSherry, "Privacy integrated queries: an extensible platform for privacy-preserving data analysis," in *SIGMOD*, 2009, pp. 19–30.

[12] N. Johnson, J. P. Near, and D. Song, "Towards practical differential privacy for sql queries," *Proceedings of the VLDB Endowment*, vol. 11, no. 5, pp. 526–539, 2018.

[13] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," in *Foundations and Trends in Theoretical Computer Science*, 2014, p. 9(3–4):211–407.

[14] F. Zhao, X. Ren, S. Yang, Q. Han, P. Zhao, and X. Yang, "Latent dirichlet allocation model training with differential privacy," *TIFS*, vol. 16, pp. 1290–1305, 2020.

[15] S. T. Dumais, "Latent semantic analysis," *Annual review of information science and technology*, vol. 38, no. 1, pp. 188–230, 2004.

[16] T. Hofmann, "Probabilistic latent semantic analysis," *arXiv:1301.6705*, 2013.

[17] D. M. Blei and J. D. Lafferty, "A correlated topic model of science," *The annals of applied statistics*, vol. 1, pp. 17–35, 2007.

[18] T. Chanyaswad, A. Dytso, H. V. Poor, and P. Mittal, "Mvg mechanism: Differential privacy under matrix-valued query," in *Proceedings of the Computer and Communications Security*, 2018, pp. 230–246.

[19] M. De Cock, A. C. Nascimento, D. Reich, R. Dowsley, and A. Todoki, "Privacy-preserving classification of personal text messages with secure multi-party computation," *Advances in Neural Information Processing Systems 32*, p. 3752, 2019.

[20] I. Elhenawy, S. H. Mahmoud, A. Moustafa *et al.*, "A lightweight privacy preserving keyword search over encrypted data in cloud computing," *Journal of Cybersecurity and Information Management*, vol. 3, no. 2, pp. 29–9, 2021.

[21] O. Feyisetan, B. Balle, T. Drake, and T. Diethe, "Privacy-and utility-preserving textual analysis via calibrated multivariate perturbations," in *Proceedings of the Web Search and Data Mining*, 2020, pp. 178–186.

[22] T. Zhu, G. Li, W. Zhou, P. Xiong, and C. Yuan, "Privacy-preserving topic model for tagging recommender systems," *Knowledge and information systems*, vol. 46, no. 1, pp. 33–58, 2016.

[23] C. Decarolis, M. Ram, S. Esmaeili, Y.-X. Wang, and F. Huang, "An end-to-end differentially private latent dirichlet allocation using a spectral algorithm," in *International Conference on Machine Learning*, 2020, pp. 2421–2431.

[24] Y. Hong, J. Vaidya, H. Lu, and M. Wu, "Differentially private search log sanitization with optimal output utility," in *15th International Conference on Extending Database Technology*, 2012, pp. 50–61.

[25] H. Wang, S. Xie, and Y. Hong, "Videodp: A flexible platform for video analytics with differential privacy," *Proc. Priv. Enhancing Technol.*, vol. 2020, no. 4, pp. 277–296, 2020.

[26] H. Wang, Y. Hong, Y. Kong, and J. Vaidya, "Publishing video data with indistinguishable objects," in *Proceedings of the 23rd International Conference on Extending Database Technology*, 2020, pp. 323–334.

[27] B. Liu, S. Xie, H. Wang, Y. Hong, X. Ban, and M. Mohammady, "Vtdp: Privately sanitizing fine-grained vehicle trajectory data with boosted utility," *IEEE Transactions on Dependable and Secure Computing*, pp. 1–1, 2019.

[28] L. Ou, Z. Qin, S. Liao, Y. Hong, and X. Jia, "Releasing correlated trajectories: Towards high utility and optimal differential privacy," *IEEE Trans. Dependable Secur. Comput.*, vol. 17, no. 5, pp. 1109–1123, 2020.

[29] J. Vaidya, B. Shafiq, A. Basu, and Y. Hong, "Differentially private naive bayes classification," in *2013 IEEE/WIC/ACM International Conferences on Web Intelligence*, 2013, pp. 571–576.

[30] M. Mohammady, S. Xie, Y. Hong, M. Zhang, L. Wang, M. Pourzandi, and M. Debbabi, "R2DP: A universal and automated approach to optimizing the randomization mechanisms of differential privacy for utility metrics with no known optimal distributions," in *CCS*, 2020, pp. 677–696.

[31] B. I. P. Rubinstein and F. Alda, "Pain-free random differential privacy with sensitivity sampling," in *IMCL*, 2017, p. 2950–2959.

[32] S. R. Valluri, D. J. Jeffrey, and R. M. Corless, "Some applications of the lambert w function to physics," *Canadian Journal of Physics*, vol. 78, no. 9, pp. 823–831, 2000.

[33] K. Nissim, S. Raskhodnikova, and A. D. Smith, "Smooth sensitivity and sampling in private data analysis," in *STOC*, 2007.

[34] P. Kairouz, S. Oh, and P. Viswanath, "The composition theorem for differential privacy," in *International conference on machine learning*, 2015, pp. 1376–1385.

[35] B. Klimt and Y. Yang, "The enron corpus: A new dataset for email classification research," in *Machine Learning: ECML*, 2004, vol. 3201, pp. 217–226. [Online]. Available: http://www.springerlink.com/content/q8g7blqvqyxrpvap/

[36] R. He and J. McAuley, "Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering," in *proceedings of the World Wide Web*, 2016, pp. 507–517.

[37] S. Xie and Y. Hong, "Reconstruction attack on instance encoding for language understanding," in *The 2021 Conference on Empirical Methods in Natural Language Processing*, 2021.

[38] S. Xie, H. Wang, Y. Kong, and Y. Hong, "Universal 3-dimensional perturbations for blackbox attacks on video recognition systems," in *IEEE Symposium on Security and Privacy*, 2022.

[39] J. Sun, B. Liu, and Y. Hong, "Logbug: Generating adversarial system logs in real time," in *The 29th ACM International Conference on Information and Knowledge Management*. ACM, 2020, pp. 2229–2232.

[40] P. Massart, "The tight constant in the dvoretzky-kiefer-wolfowitz inequality," *The annals of Probability*, pp. 1269–1283, 1990.