How to Hide the Elephant– or the Donkey– in the Room: Practical Privacy Against Statistical Inference for Large Data

Salman Salamatian*, Amy Zhang†, Flavio du Pin Calmon‡, Sandilya Bhamidipati†,
Nadia Fawaz†, Branislav Kveton†, Pedro Oliveira†, Nina Taft†
* EPFL, Switzerland, † Technicolor, Palo Alto, CA, ‡ MIT, Cambridge, MA

***Abstract*—We propose a practical methodology to protect a user's private data, when he wishes to publicly release data that is correlated with his private data, in the hope of getting some utility. Our approach relies on a general statistical inference framework that captures the privacy threat under inference attacks, given utility constraints. Under this framework, data is distorted before it is released, according to a privacy-preserving probabilistic mapping. This mapping is obtained by solving a convex optimization problem, which minimizes information leakage under a distortion constraint. We address a practical challenge encountered when applying this theoretical framework to real world data: the optimization may become untractable and face scalability issues when data assumes values in large size alphabets, or is high dimensional. Our work makes two major contributions. We first reduce the optimization size by introducing a quantization step, and show how to generate privacy mappings under quantization. Second, we evaluate our method on a dataset showing correlations between political views and TV viewing habits, and demonstrate that good privacy properties can be achieved with limited distortion so as not to undermine the original purpose of the publicly released data, e.g. recommendations.**

I. Introduction

One of the central problems of managing privacy in the Internet lies in the simultaneous management of both public and private data. In order to receive services such as recommendations, users often willing release *some* data about themselves, such as their movie watching history. However users also have other data they consider private, such as income level, political affiliation, or medical conditions. In this work, we focus on a method in which a user can release her public data, but is able to prevent against inference attacks that may learn her private data from the public information. Our solution consists of a privacy mapping, which informs a user how to distort her public data, before releasing it, such that no inference attacks can successfully learn her private data. At the same time, the distortion should be bounded so that the original service (e.g., recommendations) can continue to be useful.

We adopt the general privacy framework in [1] that considers the privacy threat incurred by a user when a passive adversary attempts to infer the user's private information from the user's public (released) data. [1] argues that the privacy loss can be measured in terms of mutual information, which leads to an optimization formulation similar to rate-distortion theory. This formulation, albeit general and theoretically sound, faces a key challenge of scalability when applied to real world datasets. The scalability issue occurs when the size of the underlying alphabet of the user data is very large.

In this paper, we focus on this challenge. Our first contribution is to propose a quantization approach to limit the dimensionality of the problem, and to characterize the additional distortion introduced by quantization. We quantize the original data by clustering it and then distort the data in the space defined by the clusters. The privacy mapping is computed using a convex solver that minimizes privacy leakage subject to a distortion constraint. Our scheme is computationally efficient - we reduce the number of optimized variables from being quadratic in the size of the underlying feature alphabet to being quadratic in the number of clusters, and thus make the optimization independent of the number of observable data samples. For some real world datasets, such as those common in movie and TV recommender systems, this can lead to orders of magnitude reduction in dimensionality. This quantization step provides a fundamental extension to the original method in [1].

Our second contribution is a preliminary characterization of the impact and performance of our method, on a new dataset. In [1] the authors only proposed and reasoned about their framework but did not evaluate it. To the best of our knowledge, our paper is the first evaluation of this method. We designed and ran a survey to collect data that contains users TV show ratings and their political affiliation. In this case study, we consider TV show opinions as data to be released and a user's political affiliation to be kept private. The framework in [1] allows for different kinds of data distortions such as removing an element of the user's public data (called *erasure-distortions*), altering the contents of some elements in a public profile (called *exchange-distortions*), or other forms of distortion.

Our evaluations demonstrate multiple things. First, we quantify the threat in our dataset and show that an adversary can infer political affiliation with roughly 71% accuracy. Second, we illustrate the effects of quantization and distortion on privacy and show that we can steadily reduce the threat with increasing distortion. We show that we can achieve perfect privacy with a reasonable amount of additional distortion. Third, we show that using limited distortion, our method can render an example Democrat/Republican classifier no better than an uninformed random guess. Finally, we conduct a preliminary experiment to examine the impact of our distortion on matrix factorization, commonly used in recommender systems, and show that additional errors in recommendation are not significant. For more details, the reader is referred to our research report [2].

II. Problem Statement

Threat Model: We consider the setting described in [1], where a user has two types of data: some data that he would like to remain private, e.g. his income level, his political views, and some data that he is willing to release publicly and from which he will derive some utility, e.g. the release of his media preferences to a service provider would allow the user to receive content recommendations. We denote by $A \in \mathcal{A}$ the vector of personal attributes that the user wants to keep private, and by $B \in \mathcal{B}$ the vector of data he is willing to make public, where $\mathcal{A}$ and $\mathcal{B}$ are the sets from which A and B can assume values.

We assume that the user private attributes A are linked to his data B by the joint probability distribution $p_{A,B}$. Thus, an adversary who would observe B could infer some information about A. To reduce this inference threat, instead of releasing B, the user releases a *distorted version* of B, denoted $\hat{B} \in \hat{\mathcal{B}}$, generated according to a conditional probabilistic mapping $p_{\hat{B}|B}$, called the *privacy-preserving mapping*. Note that the set $\hat{\mathcal{B}}$ may differ from $\mathcal{B}$. The privacy mapping $p_{\hat{B}|B}$ should be designed in a way that renders any statistical inference of A based on the observation of $\hat{B}$ harder, yet, preserves some utility to the released data $\hat{B}$, by limiting the distortion caused by the mapping. This can be modeled by a constraint $\Delta \geq 0$ on the average distortion $E_{B,\hat{B}}[d(B,\hat{B})] \leq \Delta$, for some distortion metric $d : \mathcal{B} \times \hat{\mathcal{B}} \to \mathbb{R}^+$. It should be noted that any distortion metric can be used, such as the Hamming (resp. l_2) distance if B and $\hat{B}$ are binary (resp. real) vectors, or even more complex metrics modeling the variation in utility, e.g. recommendation quality, that a user would derive from the release of $\hat{B}$ instead of B.

We assume the following standard statistical inference threat model [1]: we model the average statistical inference gain ΔC of the adversary after he observes $\hat{B}$, i.e. how much an adversary can reduce

his expected loss for a given loss function on S, after observing $\hat{B}$. . This statistical gain represents the gain in terms of inference of the private attribute A. The goal of the privacy mapping is to minimize this gain. Note that this general framework does not assume a particular inference algorithm. Moreover, using the log-loss, it can be shown [1] that $\Delta C = I(A;\hat{B})$ (for a justification of the relevance and generality of the log-loss c.f. [1, Section IV.A]). Hence, the privacy leakage is captured by the mutual information between the private attributes A and the publicly released data $\hat{B}$. In the case of perfect privacy ($I(A;\hat{B}) = 0$), the privacy mapping $p_{\hat{B}|B}$ renders the released data $\hat{B}$ statistically independent from the private data A. It should be mentioned that, although we model the privacy threat using the average cost gain in this paper, [1] also proposed a worst-case model where the privacy threat is measured in terms of the most informative output, i.e. the output that gives the largest gain in cost. Note that in the case of perfect privacy $\Delta C = 0$, the average and the worst-case threat model are equivalent. Thus conclusions drawn on distortion to achieve perfect privacy under the average threat model also hold for the worst-case model.

Privacy-Accuracy Framework: The mutual information $I(A;\hat{B})$ is a function of the joint distribution $p_{A,\hat{B}}$, which in turn depends on both the prior distribution $p_{A,B}$ and the privacy mapping $p_{\hat{B}|B}$, since $A \to B \to \hat{B}$ form a Markov chain. To stress these dependencies, we will denote

$$\Delta C = I(A;\hat{B}) = J(p_{A,B}, p_{\hat{B}|B}).$$

Similarly, the average distortion $E_{B,\hat{B}}[d(B,\hat{B})]$ is a function of the joint distribution $p_{B,\hat{B}}$, which in turn depends both on $p_{A,B}$, through the marginal p_B, and on $p_{\hat{B}|B}$. Consequently, given a prior $p_{A,B}$ linking the private attributes A and the data B, the privacy mapping $p_{\hat{B}|B}$ minimizing the privacy leakage subject to a distortion constraint is obtained as the solution to the optimization problem

$$\underset{p_{\hat{B}|B}}{\text{minimize}} \;\; J(p_{A,B}, p_{\hat{B}|B}) \qquad \text{s.t.} \;\; \mathbb{E}_{p_{B,\hat{B}}}\Big[d(B,\hat{B})\Big] \le \Delta \qquad (1)$$
$$p_{\hat{B}|B} \in \text{Simplex},$$

where Simplex denotes the probability simplex ($\sum_x p(x) = 1, \;\; p(x) \ge 0 \;\forall x$). It was shown in [1] that this problem is convex. Note that it is similar to a modified rate distortion problem.

Practical Challenge in the Presence of Large Data: When applying the aforementioned privacy-accuracy framework to large data, we encounter a practical challenge of scalability. Designing the privacy mapping requires characterizing the value of $p_{\hat{B}|B}(\hat{b}|b)$ for all possible pairs $(b,\hat{b}) \in \mathcal{B} \times \hat{\mathcal{B}}$, i.e. solving the convex optimization problem over $|\mathcal{B}||\hat{\mathcal{B}}|$ variables. When $\hat{\mathcal{B}} = \mathcal{B}$, and the size of the alphabet $|\mathcal{B}|$ is large, solving the optimization over $|\mathcal{B}|^2$ variables may become intractable. In Section III, we propose a method based on quantization to reduce the number of optimization variables. We show that the reduction in complexity does not affect the privacy levels that can be achieved, but comes at the expense of a limited amount of additional distortion, that we characterize.

III. Privacy for Large-Scale Data

In real-world datasets, the alphabet $\mathcal{B}$ is often large. In particular, the number of symbols in $\mathcal{B}$ may be $\theta(n)$, linear in the number of samples n in the dataset. Suppose that $\hat{\mathcal{B}} = \mathcal{B}$. Then the number of optimization variables $p_{\hat{B}|B}(\hat{b} \mid b)$ in problem ((1)) is $\theta(n^2)$. Note that the distortion constraint is linear in $p_{\hat{B}|B}(\hat{b} \mid b)$ but the objective function is neither linear nor quadratic, so the optimization problem (1) cannot be solved using fast linear or quadratic programming solvers. In general, the problem is hard to solve when the size of alphabet $\mathcal{B}$ exceeds a few hundreds symbols. To address this issue, we show how to solve our problem approximately by optimizing fewer variables. Our solution comprises three steps. First, a quantization step maps the symbols in alphabet $\mathcal{B}$ to $|\mathcal{C}|$ representative examples in a smaller alphabet $\mathcal{C}$. Second, we learn a privacy preserving mapping $q_{\hat{C}|C}$ on the new alphabet, where $\hat{\mathcal{C}} = \mathcal{C}$. Third, the symbols in $\mathcal{B}$ are mapped to the representative examples $\hat{\mathcal{C}}$ based on the learned mapping $q_{\hat{C}|C}$. Our approach is summarized in Alg. 1.

Algorithm 1 Quantized privacy preserving mapping.

Input: prior $p_{A,B}$

$q_{A,C}(a,c) \leftarrow \sum_{b\sim c} p_{A,B}(a,b) \quad \forall (a,c) \in (\mathcal{A},\mathcal{C})$

Solve the convex optimization problem:

$$\underset{p_{\hat{C}|C}}{\text{minimize}} \;\; J(q_{A,C}, p_{\hat{C}|C}) \qquad \text{s.t.} \;\; \mathbb{E}_{p_{C,\hat{C}}}\Big[d(C,\hat{C})\Big] \le \Delta \qquad (2)$$
$$p_{\hat{C}|C} \in \text{Simplex},$$

$p_{\hat{C}|B}(\hat{c} \mid b) \leftarrow q_{\hat{C}|C}(\hat{c} \mid \psi(b)) \quad \forall (b,\hat{c}) \in (\mathcal{B},\hat{\mathcal{C}})$

Output: mapping $p_{\hat{C}|B}$

Our solution has several notable properties. To begin with, the privacy-preserving mapping $q_{\hat{C}|C}$ is obtained for the reduced alphabet $\mathcal{C}$. Thus, we need to solve the convex optimization (1) for only $|\mathcal{C}||\hat{\mathcal{C}}|$ variables. In practice, $|\mathcal{C}| \ll |\mathcal{B}|$ and this results in major computational savings. Second, quantization and privacy-preserving optimization are done separately. Therefore, any quantization method can be easily combined with our approach. In particular, we can minimize the quantization error in the quantization step, and then our privacy mechanism guarantees the optimal mapping in terms of additional distortion. Finally, quantization obviously yields a suboptimal privacy-accuracy tradeoff, since the quantization step is an additional source of distortion. However, in Theorem 1, we quantify how quantization affects the privacy-accuracy trade-off, and show that the levels of privacy that can be achieved are not affected, but come at the expense of a bounded amount of distortion.

We now analyze Alg. 1. Problem (2) solves a variant of problem (1), where alphabets $\mathcal{B}$ and $\hat{\mathcal{B}}$ are substituted for alphabets $\mathcal{C}$ and $\hat{\mathcal{C}}$, and the joint probability distribution over A and C is defined as

$$q_{A,C}(a,c) = \textstyle\sum_{b\sim c} p_{A,B}(a,b), \qquad (3)$$

where $b \sim c$ means that the symbol b is in the cluster represented by center c. The above equation aggregates the probability mass of all symbols in the cluster in its center. The symbols in $\mathcal{B}$ are mapped to $\hat{\mathcal{C}}$ according to

$$p_{\hat{C}|B}(\hat{c} \mid b) = q_{\hat{C}|C}(\hat{c} \mid \psi(b)), \qquad (4)$$

where $\psi : B \to C$ is a function that maps a symbol in $\mathcal{B}$ to a cluster center in $\mathcal{C}$. Note that the probability distributions that are associated with optimization (2) are marked by q. Now we state our main claim.

Theorem 1. *Let $q_{\hat{C}|C}$ be a solution to problem (2) and $p_{\hat{C}|B}$ be the corresponding mapping from $\mathcal{B}$ (Equation 4). Moreover, let $\mathcal{C}$ be an alphabet such that $\max_{b\in\mathcal{B}} \min_{c\in\mathcal{C}} d(b,c) \le r$. Then the privacy leakage $J(p_{A,B}, p_{\hat{C}|B})$ of the mapping $p_{\hat{C}|B}$ is equal to the value of the objective function of (2):*

$$J(p_{A,B}, p_{\hat{C}|B}) = J(q_{A,C}, q_{\hat{C}|C}),$$

and its total distortion rate is no more than r larger than the target Δ:

$$\mathbb{E}_{p_{B,\hat{C}}}\Big[d(B,\hat{C})\Big] \le \Delta + r.$$

Theorem 1 (proof in [2]) states that the information leakage of the mapping $p_{\hat{C}|B}$ is the same as that of the optimized mapping $q_{\hat{C}|C}$. So we optimize the quantity of interest $J(p_{A,B}, p_{\hat{C}|B})$ in a time which is independent of the size of the input alphabet $\mathcal{B}$. The distortion rate

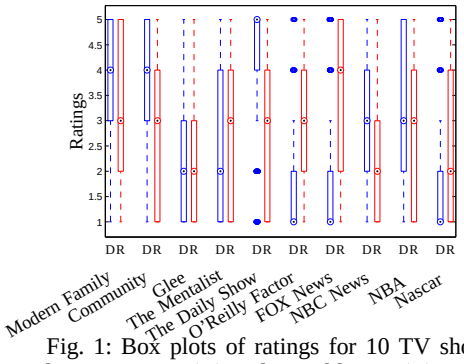

Fig. 1: Box plots of ratings for 10 TV shows by Democrats (D) and Republicans (R)

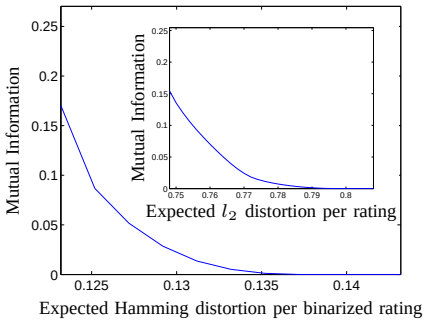

Fig. 2: Privacy-accuracy tradeoff after quantization on binarized ratings, and full ratings (inset)

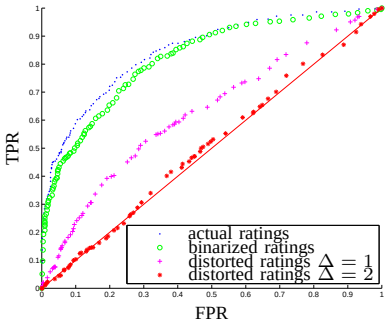

Fig. 3: ROC curve of a logistic regression of political views based on TV show ratings

increases due to quantization, linearly with the maximum distance r between any example b and its closest representative example $\psi(b)$.

The maximum distance r can be minimized by existing clustering methods, such as online k-center clustering [3]. It can be shown by a simple ball backing argument that this method clusters data such that $r \leq 8/|\mathcal{C}|^{-\frac{1}{d}}$, where d is the intrinsic dimension of data.

IV. Dataset

To evaluate our framework, we collected a dataset *Politics-and-TV*, on political views and TV preferences. The collection of such data was motivated by large scale surveys [4], [5], which illustrated that correlations exist between TV show ratings and political views.

Data Collection: We designed a survey that users take voluntarily. Users were first asked to provide demographic information (gender, age group, state they live in) as well as their political views (Democrat, Republican). Then users were asked to complete a sequence of 6 panels, each panel presenting the user with 6-8 TV shows of a certain genre, namely Sitcoms, Reality Shows, TV series, Talk Shows, News, and Sports, for a total of 50 TV shows. Users were asked to rate only those TV shows that they watched on a scale from 1 to 5— the usual star rating system. After providing their ratings, users were shown, for each genre, how their ratings compared with the average ratings given by Democrats and Republicans. In our privacy policy, users were informed that private information that can be used to identify an individual, e.g. cookies, IP addresses, was not stored.

We ran our survey in two phases. In phase 1 (October 2012), we ran it on Mechanical Turk requesting only US-based workers. In total, we obtained 854 surveys, with 518 Democrats and 336 Republicans. In phase 2 (November 2012), we launched our survey on the public web at www.PoliticsandMedia.org. We drove traffic to the survey website by running advertising campaigns on MyLikes.com and Google AdWords, shortly before the U.S. 2012 presidential election. From this, we obtained another 364 completed surveys, with 226 Democrats and 138 republicans.

Dataset: The dataset contains entries for 1218 users, broken into 744 Democrats, and 474 Republicans. For each user, the dataset entry is a vector $[\text{age}, \text{gender}, \text{state}, \text{politics}, r_1, \ldots..r_{50}]$ where $r_i \in \{0, 1, \ldots, 5\}$ is the user's star rating for show i if the user rated the show, and 0 otherwise. We consider two versions of the rating vector: the 5-star rating vector $R \in \{0, 1, \ldots, 5\}^{50}$, and the binarized rating vector $B \in \{0, 1\}^{50}$. The binarized rating b_i of show i is obtained by setting $b_i = 1$ if the original rating $r_i >= 4$ clearly indicating that the user likes the show, and $b_i = 0$ otherwise.

TABLE I: RMSEs of $|r - \hat{r}|$ and $|r - \hat{\hat{r}}|$

Set	1	2	3	4	5
RMSE1	1.2506	1.1820	1.2461	1.2155	1.2101
RMSE2	1.6972	1.6763	1.6215	1.7248	1.8036

V. Results

Consider the setting where a user wishes to release his TV show ratings $R \in \{0, 1, \ldots, 5\}^{50}$ (or $B \in \{0, 1\}^{50}$), in the hope of getting good recommendations, but is concerned about them leaking information about his political affiliation $A \in \{\text{Democrat}, \text{Republican}\}$. Note that although we focus on the case where the private data is a single variable representing political affiliation, the privacy-accuracy framework [1] can handle protecting a set of private variables, e.g. we could protect any subset of a user's three attributes [age,gender, politics]. The rating vector R (reps. B) lives in a large alphabet of size 6^{50} (resp. 2^{50}) over 6^{100} variables would be untractable, and justifies resorting to quantization. Note that the number of samples is small relative to the alphabet size, and the prior $p_{A,R}$ estimated from the dataset may be mismatched, issue addressed in [2].

Privacy threat: The threat comes from the underlying existence of TV shows that are highly correlated with political affiliation, e.g. *The Daily Show* is predominantly liked by Democrats, while *Fox News* is preferred by Republicans. Fig. 1 shows boxplots of ratings for 12 shows—two shows from each genre in the dataset— by Democrats and Republicans. Those shows for which there is little overlap in the opinions of Republicans and Democrats clearly demonstrate high correlation between political affiliation and opinion of those shows. Such shows have high discriminative power that inference algorithms can exploit. There exists a broad variety of shows in terms of their discriminative power - some are very much so, while others exhibit low correlation. Users who rate highly shows such as *The O'Reilly Factor*, or *The Daily Show*, may be facing a stronger threat than those who only watch and rate shows with little discriminating power. Broadly speaking, across our 50 shows, we found that roughly one third of them have strong correlation with political affiliation.

In order to understand the threat inherent in this dataset, we quantify the potential privacy leakage using mutual information $I(A; R)$. To provide an illustrative example, we thus consider a reduced set of our data for which we can compute the mutual information. We consider the top 5 most seen TV shows, and use the binarized version of the rating vector with ratings in $\{0, 1\}^5$. For this case, we observe that the mutual information between the observed features and the political orientation is already at 0.191 bits. An adversary, with this information on hand (the 5 tuple of binarized ratings), could use a maximum a posteriori (MAP) detector and guess the political affiliation of somebody with an accuracy of 71%. Hence, the privacy threat is real. Note that because mutual information is a non-decreasing function, as we add additional shows, the threat either stays the same or increases.

Privacy-accuracy trade-off: We now apply our quantization approach and investigate its impact on the privacy-distortion tradeoff. We first consider the full 50 show dataset with the binarized version of the ratings. Alg. 1 first quantizes the data using a K-means clustering algorithm with a Hamming distance metric and K=25 clusters; then

we apply the convex optimization on the quantized points. The resulting trade-off curve is depicted in Fig. 2. The curve shows that the quantization step alone introduces an average Hamming distortion of about 12% (leftmost point on x-axis) per rating, or 6.1 over all 50 shows, and results in a mutual information of 0.189 on the representative points (cluster centers). As this is still high, we are motivated to apply further distortion. Fig. 2 shows that using the optimal privacy preserving scheme resulting from convex optimization, we can steadily decrease the privacy threat with increasing distortion. Not only is our privacy-distortion curve properly behaved, but small increases in Hamming distance bring the privacy leakage down quickly. Moreover, we can achieve perfect privacy ($I = 0$) at the cost of an additional 3% in average Hamming distortion (beyond the clustering distortion). Perfect privacy is achieved at an overall Hamming Distortion of less than 7 out of 50; put alternatively, perfect privacy is obtainable if on average we change just less than 15% of a user's rating data before it is released. In Fig. **??**, we consider the same tradeoff with the actual 5-star ratings, and k-means clustering with L2 distance. The mutual information after quantization is 0.182. Perfect privacy is achieved with a total distortion of 0.8 per rating—0.75 of which are due to quantization. The actual ratings require slightly higher distortion to reach perfect privacy than with binarized ratings, since their range is $[0, 5]$ instead of $\{0, 1\}$.

Inference defeat: The previous plots show the reduction in privacy leakage achieved by our distortion. Another key performance metric is the reduction in accuracy of a Democrat/Republican classifier when distorted user ratings are used instead of actual ones. We consider the example of a logistic regression classifier to infer political affiliation ([6] used logistic regression to infer gender from movie ratings). We used 10-fold cross validation on our full dataset, considered both cases of actual and binarized ratings, and a distortion that achieves perfect privacy $I = 0$, at which any inference algorithm cannot perform better than an uninformed random guess. In Fig. 3 we plot the false positive rate (number Democrats falsely classified as Republicans), against the true positive rate (number of Republicans correctly classified). With a distortion bound $\Delta = 1$, we can significantly reduce the classifier's performance but not reach perfect privacy; however with $\Delta = 2$ the classifier is reduced to an uninformed random classifier. This demonstrates that our approach can successfully render inference attempts unsuccessful. Finally, note that logistic regression performs almost equally well with binarized and actual ratings, which means merely perturbing existing ratings is not enough, thus we must add or delete ratings to protect privacy.

Recommendation quality: As a final performance metric, we consider the impact of our distortion on recommendations that would be produced by a recommender system based on matrix factorization. In Table I, RMSE1 captures the root mean squared error in predicted ratings (compared to the true ratings) using unperturbed data $\hat{r}$, while RMSE2 captures the errors when ratings are predicted using the distorted data $\hat{\hat{r}}$ produced by our algorithm. The results were produced using 5-fold cross validation and randomly removing 10% of the ratings in each test set. We can see that any additional errors in TV recommendations, due to distorting ratings, is small. This preliminary result on the impact on a recommendation system is encouraging, yet requires further extensive testing.

VI. Related Work

The prevalent notion of privacy is differential privacy [7], [8]: a query over a database is differentially private if small variations in the entries of the database do not significantly change the output distribution of the query. Differential privacy does not take into account the distribution of the database entries, which makes the formulation mathematically tractable and simplifies its implementation, and it is robust against arbitrary side information from the attacker. However, differential privacy does not quantify the amount of information that is leaked by the system. Furthermore, for certain input distributions on the database entries, an adversary might able to infer with arbitrarily high precision the input database from a differentially private query [1]. More general and flexible frameworks similar to differential privacy exist such as the Pufferfish framework [9], which does not take into account distortion of the data, and requires knowing the adversary's belief of the input distribution.

Another existing trend in privacy research applies information-theoretic tools to quantify and design privacy-preserving mechanisms [1], [10]–[14]. Information theory provides a natural framework to measure the amount of private information that an adversary can learn by observing a user's public data. This was first noted by Reed [10]. One line of work, adopted in [11], [12], provides asymptotic and fundamental limits on rate-distortion-equivocation regions as the number of data samples grows arbitrarily. Non-asymptotic approaches to information-theoretic privacy were discussed, for example, in [1], [13], [14]. More recently, [1] introduced a general framework for privacy against statistical inference that takes into account distortion constraints for the user's public data. Information-theoretic approaches have also been used to quantify the information flow in security systems (e.g. [15] and references).

VII. Conclusion

In this paper we address a key scalability challenge that arises when applying an information theoretic framework to a privacy problem in which a user wants to release her public data while simultaneously protecting her private data from inference threats. The optimization formulation cannot scale when the underlying alphabet of the user's public data is large. Our main contribution is to propose a method, based upon quantization, that drastically reduces the dimensionality of the optimization. Rendering the optimization computationally efficient is a key step towards making our framework practical. We evaluated our approach on a novel dataset and demonstrated that in the use case of a recommendation system, a high level of privacy can be achieved without a significant impact on the recommendations.

References

[1] F. Calmon and N. Fawaz, "Privacy against statistical inference," in *Allerton Conf. on Communication, Control, Computing, Allerton*, 2012.

[2] S. Salamatian, A. Zhang, F. d. Calmon, S. Bhamidipati, N. Fawaz, B. Kveton, P. Oliveira, and N. Taft, "Managing your Private and Public Data: Bringing down Inference Attacks against your Privacy," *ArXiv e-prints*, 2013. [Online]. Available: http://arxiv.org/

[3] M. Charikar, C. Chekuri, T. Feder, and R. Motwani, "Incremental clustering and dynamic information retrieval," in *ACM STOC*, 1997.

[4] "What Your Favorite TV Shows And Networks Say About Your Politics," http://www.buzzfeed.com/rubycramer/what-your-favorite-tv-shows-say-about-your-politic, 2012.

[5] "Simmons Consumer Segmentations: PublicPersonas," http://www.experian.com/simmons-research/simmons-consumer-research.html, 2012.

[6] U. Weinsberg and S. Bhagat and S. Ioannidis and N. Taft, " BlurMe: Inferring and Obfuscating User Gender Based on Ratings," in *ACM Conference on Recommender Systems (RecSys)*, September 2012.

[7] C. Dwork, F. Mcsherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *TCC*, 2006.

[8] C. Dwork, "Differential privacy," in *Automata, Languages and Programming*. Springer, 2006, vol. 4052, pp. 1–12.

[9] D. Kifer and A. Machanavajjhala, "A rigorous and customizable framework for privacy," in *Proc. of ACM PODS*, 2012.

[10] I. S. Reed, "Information Theory and Privacy in Data Banks," in *Proc. of ACM AFIPS*, 1973.

[11] H. Yamamoto, "A source coding problem for sources with additional outputs to keep secret from the receiver of wiretappers," *IEEE Trans. Inf. Theory*, vol. 29, no. 6, 1983.

[12] L. Sankar, S. R. Rajagopalan, and H. V. Poor, "Utility-privacy tradeoff in databases: An information-theoretic approach," *IEEE Trans. Inf. Forensics Security*, 2013.

[13] D. Rebollo-Monedero, J. Forné, and J. Domingo-Ferrer, "From t-closeness-like privacy to postrandomization via information theory," *IEEE Trans. on Knowledge and Data Engineering*, 2010.

[14] A. Evfimievski, J. Gehrke, and R. Srikant, "Limiting privacy breaches in privacy preserving data mining," in *Proc. of ACM PODS*, 2003.

[15] S. Hamadou, V. Sassone, and C. Palamidessi, "Reconciling belief and vulnerability in information flow," in *IEEE Symposium on Security and Privacy*, 2010.