

# An Analysis of the Impact of Differentially Private Data on Recommendation

Ferrara Antonio<sup>1,\*</sup>, Alberto Carlo Maria Mancino<sup>1,\*</sup>, Angela Di Fazio<sup>1</sup>, Vito Walter Anelli<sup>1</sup>, Tommaso Di Noia<sup>1</sup> and Eugenio Di Sciascio<sup>1</sup>

<sup>1</sup>Polytechnic University of Bari, via Orabona 4, Bari, 70126, Italy

## Abstract

The success of recommender systems heavily relies on having access to public datasets. However, there is growing concern about users' privacy, which makes the publication of such datasets a challenging task. One potential solution to this issue is to use differential privacy (DP), a well-established framework for maintaining users' privacy in machine learning. Nevertheless, applying DP to release recommendation datasets may negatively impact the performance of the recommender systems, given the statistical properties of these datasets. To further explore this issue, we aim to investigate the impact of DP on recommendation performance, considering the dataset characteristics. We draw inspiration from previous studies that highlight the relationship between data characteristics and recommendation performance. In our research, we propose using randomized response as a straightforward mechanism for releasing implicit recommendation datasets privately. We generate over 1800 sub-datasets and build an explanatory framework<sup>1</sup> that estimates the performance degradation due to privatization while considering the dataset characteristics and the privacy budget. Our study provides researchers with statistically validated and reproducible results and contributes to a deeper understanding of the interplay between data characteristics and the impact of data privatization on recommender system performance.

## 1. Introduction


Over the past few years, the issue of privacy has emerged as a major concern in the context of big data applications. The general public has become increasingly aware of the significance of this issue, particularly in the wake of major data breaches like the one that occurred in 2018 involving Cambridge Analytica. This incident involved the unauthorized sharing and harvesting of data from numerous users for political campaigning purposes, without their consent, which has served as a catalyst for heightened public scrutiny of data privacy practices.[1]. However, machine learning and data mining algorithms heavily rely on data pertaining to identity, biometrics, health, facial recognition, smartphones, transportation and vehicles, and video surveillance, as it serves as their essential fuel. This includes recommender systems, which are trained to suggest unexplored items that users are likely to prefer. When experiencing a recommender system, users want to receive accurate predictions, while still being concerned about sharing personal information, raising the well-known *personalization-versus-privacy paradox*. Numerous studies in the field use robust assurances of differential privacy [2] to create confidential mod-

els [3, 4, 5, 6]. This involves the service provider adding random noise to the model to safeguard the users' privacy. In cases where the service provider is not deemed trustworthy, certain studies suggest utilizing local differential privacy in distributed or federated learning settings [7, 8, 9, 10]. Despite their flexibility, differential privacy techniques are usually applied at the parameter or gradient-level, while rarely applied directly in the data collection or release phase. These aspects are highly relevant since, even though the data collector applies anonymization techniques before sharing data with third parties, privacy violations may still occur [11, 12]. In the recommender systems domain, the randomized response can be conceived as a simple mechanism to privately release binary datasets. This privacy-preserving mechanism ensures that users' true responses to sensitive questions remain confidential while providing plausible deniability[13] through a perturbed response returned with a certain probability. One such approach is RAPPOR, which uses randomized response to collect binary data while guaranteeing differential privacy. However, while public datasets are crucial for recommendation research, techniques like RAPPOR are rarely used to release such data. Since recommendation datasets are characterized by peculiar characteristics such as high levels of sparsity and skewness, we investigate whether and how mechanisms based on the randomized response distort the dataset, thus differently impacting the recommendation performance based on the data characteristics. Adomavicius and Zhang [14] demonstrated that a notable relationship between recommender systems performance and dataset characteristics exists. Based on these findings, we deem that, when evaluating the effect of randomized re-

*Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29–31, 2023, Pisa, Italy*

\*Corresponding author.

✉ antonio.ferrara@poliba.it (F. Antonio);  
alberto.mancino@poliba.it (A. C. M. Mancino);  
angela.difazio@poliba.it (A. D. Fazio); vitowalter.aneli@poliba.it  
(V. W. Anelli); tommaso.dinoia@poliba.it (T. D. Noia);  
eugenio.disciascio@poliba.it (E. D. Sciascio)

 © 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

sponse on the recommendation utility, both the amount of perturbation and the dataset characteristics should be taken into account. This paper makes a step in this direction, with a two-fold contribution: (1) we propose a formalization of a simple **randomized-response-based mechanism** for privatizing user data and releasing differentially private implicit recommendation datasets; (2) we perform an extensive **explanatory study** to systematically analyze which dataset characteristics are more prone to degrade the performance of different recommendation models when randomized response is applied. In particular, we assess the impact of the original dataset characteristics and the chosen privacy strength on the accuracy and popularity bias of four recommendation models, ranging from unpersonalized to neighborhood-, autoencoder-, and graph-based models.

## 2. Background

**Differential Privacy.** Differential privacy (DP) [2] represents a formal mathematical definition for quantifying and limiting information disclosure about individuals.

**Definition 1** ( $\epsilon$ -Differential Privacy). *A randomized mechanism  $\mathcal{M} : \mathbb{N}^{|\mathcal{X}|} \rightarrow \mathcal{R}$  preserves  $\epsilon$ -differential privacy if given any two adjacent datasets  $\mathcal{K}_1$  and  $\mathcal{K}_2$  (i.e., they differ by only one record), for all  $\mathcal{S} \subseteq \mathcal{R}$ :*

$$\Pr[\mathcal{M}(\mathcal{K}_1) \in \mathcal{S}] \leq e^\epsilon \Pr[\mathcal{M}(\mathcal{K}_2) \in \mathcal{S}]. \quad (1)$$

Differential privacy is noteworthy for its ability to ensure that the results of a function remain consistent even when a record is removed or altered. This similarity is dependent on the level of the privacy budget, represented by the value  $\epsilon$ , where lower values correspond to greater privacy. Typically, a mechanism called randomized mechanism  $\mathcal{M}$  is used, which involves adding noise to the function's output in proportion to its sensitivity (i.e., the largest impact a single record has on the output) or based on an exponential distribution of a set of distinct values. As the value of  $\epsilon$  decreases, privacy increases, but it becomes more challenging to maintain accuracy with such a mechanism, as smaller values of  $\epsilon$  lead to decreased accuracy.

**Randomized Response.** Randomized response [15, 16] is a mechanism that respondents to a survey can use to protect their privacy when asked about a sensitive attribute, e.g., «Did you visit Venice?».

**Definition 2** (Randomized Response). *Let  $x \in \{x_1, \dots, x_r\}$  be the variable containing the answer to a sensitive question. The randomized response privatizes the true answer reporting the value of a variable  $\tilde{x}$  instead of  $x$ , based on the following perturbation matrix  $\mathbf{P}$ :*

$$\mathbf{P} = \begin{pmatrix} p_{x_1 x_1} & \cdots & p_{x_1 x_r} \\ \vdots & \ddots & \vdots \\ p_{x_r x_1} & \cdots & p_{x_r x_r} \end{pmatrix}, \quad (2)$$

where  $p_{uv} = \Pr[\tilde{x} = v \mid x = u]$ , for  $u, v \in \{x_1, \dots, x_r\}$ .

The intuition behind randomized response is that it provides *plausible deniability*. For instance, a response «Yes» may have been provided because of the true value or because of the perturbation. In general, randomized response allows a user to deny an original value  $u$  when providing a perturbed value  $v$ . Indeed, denoting  $\hat{p}_{uv} = \Pr[x = u \mid \tilde{x} = v]$ , by the Bayes' formula we have:

$$\hat{p}_{uv} = \frac{\Pr[\tilde{x} = v \mid x = u] \Pr[x = u]}{\sum_{u' \in \{x_1, \dots, x_r\}} \Pr[\tilde{x} = v \mid x = u'] \Pr[x = u']}. \quad (3)$$

Therefore, as long as  $\hat{p}_{uv} < 1$ , a user has a chance to deny that the true value is  $x = u$  given the released value  $\tilde{x} = v$ . Given a reported value, the more similar the probabilities, the higher the deniability. Specifically, Domingo-Ferrer and Soria-Comas [13] define the deniability in terms of Shannon entropy as:

$$H(x \mid \tilde{x} = v) = - \sum_{u \in \{x_1, \dots, x_r\}} \hat{p}_{uv} \log_2 \hat{p}_{uv} \quad (4)$$

whose maximum value is 1, corresponding to the case  $\hat{p}_{uv} = 1/r$  for any  $u \in \{x_1, \dots, x_r\}$ . This value intuitively measures how much an attacker is confused about the true response  $x$  of a user, when provided with the perturbed value  $\tilde{x} = v$ .

**Connection with Differential Privacy.** Randomized response can be analyzed under the lens of differential privacy. According to Eq. (1), the ratio of the probabilities to have the same output given different inputs should be upper-bounded by  $e^\epsilon$ . In randomized response, this corresponds to having the maximum ratio between the probabilities in the same column of  $\mathbf{P}$  always bounded by  $e^\epsilon$ :

$$\max_{v \in \{x_1, \dots, x_r\}} \frac{\max_{u \in \{x_1, \dots, x_r\}} p_{uv}}{\min_{u \in \{x_1, \dots, x_r\}} p_{uv}} \leq e^\epsilon. \quad (5)$$

## 3. Privatizing User Data

In the work at hand, we explore and analyze a simple privatization method for recommendation datasets built on the pillars of randomized response and differential privacy. For the sake of simplicity and without loss of generality, an implicit feedback scenario is considered. However, randomized response can be extended to any categorical variable, including explicit ratings, dealing with them as integer-valued categories. Notably, we properly apply randomized response to the original user rating matrix to perturb the private user-item interactions. This approach guarantees users' privacy thanks to the connection with differential privacy.

### Randomized Perturbation of User-Item Feedback.

Let  $\mathbf{X} \in \{0, 1\}^{|\mathcal{U}| \times |\mathcal{I}|}$  be a matrix containing the negative/positive feedback of each user in  $\mathcal{U}$  to each item in  $\mathcal{I}$ . Each element of the matrix  $\mathbf{X}$  has to be independently perturbed, as it inherently represents a private answer to a different sensitive question, e.g., «Do you like this restaurant?», and is not affected by the answer to other questions. Therefore, in the following, we focus on the perturbation of one feedback of a single user, whose true value can be either 0 or 1, meaning the user responded «No» or «Yes» to the sensitive question. Randomized response can offer users plausible deniability about their answers. To this aim, we build a  $2 \times 2$  perturbation matrix  $\mathbf{P}$ , representing the transition probability of their feedback from 0 to 1, and vice versa:

$$\mathbf{P} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}. \quad (6)$$

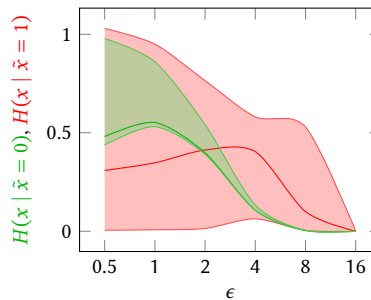
Assuming that the we build  $\mathbf{P}$  favoring the true values, i.e.,  $p_{00}, p_{11} > 0.5$ , we can write the condition in Eq. (5) as:

$$\max \left\{ \frac{p_{00}}{p_{10}}, \frac{p_{11}}{p_{01}} \right\} \leq e^\epsilon \quad (7)$$

If this inequation holds, the reported value can be released with limited disclosure of the real value and guaranteeing  $\epsilon$ -differential privacy. In particular, Wang et al. [17] prove that, given a fixed of  $\epsilon$ , we can satisfy at the same time  $\epsilon$ -differential privacy and maximize  $p_{00} + p_{11}$ , thus avoiding useless perturbations. To get this result, the perturbation matrix should have the following pattern:

$$\mathbf{P} = \begin{pmatrix} \frac{e^\epsilon}{1+e^\epsilon} & \frac{1}{1+e^\epsilon} \\ \frac{1}{1+e^\epsilon} & \frac{e^\epsilon}{1+e^\epsilon} \end{pmatrix}. \quad (8)$$

**Plausible Deniability of Feedback.** Even though the value of  $\epsilon$  guarantees differential privacy to a certain extent, its implications in Eq. (4) should be carefully considered. By doing so, when releasing randomized data, we can get an idea of the plausible deniability we guarantee to our users with respect to their individual interactions. Indeed, this value strictly depends on the density of the recommendation dataset, given that it follows the proportion of the true yes/no responses. As an example, in Figure 1, we show the values of plausible deniability for the answer 0 (green plot) and for the answer 1 (red plot) when the recommendation dataset MovieLens 1M is perturbed with different  $\epsilon$  values. Along with the average deniabilities, the plots show both the minimum and the maximum value over all the sensitive questions. The sparse nature of MovieLens 1M – and, in general, of all the recommendation datasets – implies that for a user-item interaction  $x$ , it holds  $Pr[x = 0] \gg Pr[x = 1]$ . Then, returning 1 makes the attacker very little confused about



**Figure 1:** Minimum, maximum, and average plausible deniability for  $\tilde{x} = 0$  (green plot) and  $\tilde{x} = 1$  (red plot) at values of  $\epsilon$ .

the original value, especially if it is returned for questions where this value is usually unexpected (i.e., long-tail items). However, returning 1 for popular items makes the attacker highly confused about the true feedback of the user. This contrast clearly explains the wide width of the red plot. Returning 0 provides instead higher deniability when choosing a strong privacy budget (e.g.,  $\epsilon = 0.5$ ), given the higher frequency of true 0 values. On the contrary, when  $\epsilon$  increases and the privacy constraints are relaxed, we observe that the plausible deniability falls both for the responses 0 and 1, but more slowly when the user responds 1. This is still due to the sparsity of the dataset: even increasing  $\epsilon$ , i.e., decreasing the transition probability, the large number of 0s still often mutate to 1, thus making the attacker confused about the origin of the observed 1. Instead, the small number of 1s implies little transitions to 0 when increasing  $\epsilon$ , thus making the attacker quite sure that an observed 0 comes from a true 0.

## 4. Explanatory Analysis

A higher privacy level obtained with noise injection, e.g., with the randomized response, necessarily causes inherent performance degradation of the recommendation algorithm using the noisy data. Inspired by Adomavicius and Zhang [14], we realize a regression model to analyze which dataset characteristics are more prone to influence the outcome of different recommendation models when the randomized response is applied to the original dataset, as proposed in Section 3. Notably, if a recommendation model performs  $\mu$  on a dataset  $\mathbf{X}$  for a specific metric, we argue that a perturbed dataset  $\tilde{\mathbf{X}}$  with  $\epsilon$ -differential privacy causes a degradation of the performance  $\Delta\mu$ , which we estimate as:

$$\Delta\mu(\mathbf{X}, \epsilon) = \theta_0 + \sum_{i \in \mathcal{C}} \theta_i g_i(\mathbf{X}) + \theta_\epsilon \epsilon, \quad (9)$$

where  $\mathcal{C}$  is the set of the selected statistical characteristics, and the  $g_i(\mathbf{X})$ 's represent the same characteristics

**Table 1**

Results of the explanatory model on three datasets: Movielens 1M (ML1M), Amazon Digital Music (ADM), and Library Thing (LT). The cells report the weights,  $p$ -values and the  $R^2$  of the model.

		$\Delta$ Precision@10			$\Delta$ ARP@10		
		ML1M	ADM	LT	ML1M	ADM	LT
<b>MostPop</b>	$R^2$	0.82979	0.56968	0.54467	0.88818	0.91515	0.90205
	$\theta_{\text{SpaceSize}}$	-0.64251***	-0.39873***	-0.6265***	0.22146***	-0.82756***	-0.57115***
	$\theta_{\text{Shape}}$	-0.19213***	0.73963***	0.15363**	-0.08774***	0.18594***	0.20732***
	$\theta_{\text{UserRatings}}$	0.69568***	0.28625***	0.18444***	0.06276**	0.22991***	0.22402***
	$\theta_{\text{ItemRatings}}$	0.31384***	-0.14931	-0.11729	-0.87571***	-0.17379***	-0.54339***
	$\theta_{\text{ItemGini}}$	-0.02678	0.47820**	0.47271***	-0.00986	0.00370	0.09390***
	$\theta_{\epsilon}$	-0.78355***	-0.07379***	-0.59126***	0.49497***	0.40423***	0.46219***
<b>ItemKNN</b>	$R^2$	0.94533	0.46682	0.55700	0.93794	0.91163	0.91563
	$\theta_{\text{SpaceSize}}$	0.25873***	-0.16068**	0.51065***	-0.2053***	-0.79519***	-0.48621***
	$\theta_{\text{Shape}}$	-0.00830	-1.31884***	0.07239	-0.03561**	0.23152***	0.2711***
	$\theta_{\text{UserRatings}}$	0.64500***	-0.74571***	0.25576***	0.13604***	0.26218***	0.20821***
	$\theta_{\text{ItemRatings}}$	-0.12319***	0.50660***	0.50000	-0.77877***	-0.21905***	-0.64315***
	$\theta_{\text{ItemGini}}$	0.11027***	0.63739***	-0.05162	0.05531***	-0.00743	0.06457***
	$\theta_{\epsilon}$	-0.45170***	-0.05388**	-0.35984***	0.37732***	0.39550***	0.42191***
<b>EASER</b>	$R^2$	0.86210	0.52585	0.67834	0.88996	0.92051	0.90981
	$\theta_{\text{SpaceSize}}$	-0.53847***	-0.14922***	-0.31296***	0.21301***	-0.82758***	-0.57363***
	$\theta_{\text{Shape}}$	-0.20757***	-1.42374***	-0.47790***	-0.08559***	0.18872***	0.20678***
	$\theta_{\text{UserRatings}}$	0.76275***	-0.79616***	0.36638***	0.06738**	0.23271***	0.22056***
	$\theta_{\text{ItemRatings}}$	0.30958***	0.81163***	0.47182***	-0.88313***	-0.18788***	-0.54813***
	$\theta_{\text{ItemGini}}$	-0.04132**	0.39370***	0.36040***	-0.00571	0.00419	0.08408***
	$\theta_{\epsilon}$	-0.73270***	-0.08127***	-0.46949***	0.48379***	0.39152***	0.44191***
<b>RP3<math>\beta</math></b>	$R^2$	0.93757	0.54796	0.74282	0.91886	0.90823	0.91086
	$\theta_{\text{SpaceSize}}$	0.23621***	-0.11209**	-0.04328	-0.12050**	-0.81584***	-0.52753***
	$\theta_{\text{Shape}}$	-0.02467*	-1.40264***	-0.37873***	-0.0646***	0.21221***	0.24107***
	$\theta_{\text{UserRatings}}$	0.65072***	-0.75766***	0.35893***	0.12073**	0.25509***	0.21959***
	$\theta_{\text{ItemRatings}}$	-0.05555*	0.72418***	0.47977***	-0.78237***	-0.23406***	-0.59419***
	$\theta_{\text{ItemGini}}$	0.00136	0.43725***	0.19951***	0.03231**	0.02788	0.06346***
	$\theta_{\epsilon}$	-0.49532***	-0.04955**	-0.37098**	0.40914***	0.38137***	0.42439***

\*\*\* $p \leq .001$ , \*\* $p \leq 0.01$ , \* $p \leq 0.05$

measured on the original dataset  $\mathbf{X}$ . The regression model has been computed with the ordinary least squares method based on  $m$  sub-datasets  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m$ , which have been randomly subsampled from a bigger dataset following the sampling procedure proposed in [14]. Each sub-dataset is used for (i) training a recommendation model, then (ii) perturbed and used to (iii) re-train the recommendation model and (iv) estimate the performance degradation.

**Dataset Characteristics.** To build the regression model, we choose a subset of the recommendation dataset characteristics adopted in [18], namely  $\mathcal{C} = \{\text{SpaceSize}, \text{Shape}, \text{UserRatings}, \text{ItemRatings}, \text{ItemGini}\}$ . We denote with  $\mathcal{U}$  the set of users in  $\mathbf{X}$ , with  $\mathcal{I}$  the set of items, and with  $\mathcal{R}$  the set of positive ratings.

**Structure of the user rating matrix.** We select four properties related to the structure of the rating matrix  $\mathbf{X}$ . The first one indicates the maximum number of preferences that can be collected in  $\mathbf{X}$ :

$$g_{\text{SpaceSize}}(\mathbf{X}) = |\mathcal{U}| \cdot |\mathcal{I}|, \quad (10)$$

The second characteristic denotes if there are more candidate neighbor users than candidate neighbor items ( $g_{\text{Shape}} \gg 0.001$ ) or the opposite scenario ( $g_{\text{Shape}} \ll 0.001$ ):

$$g_{\text{Shape}}(\mathbf{X}) = \frac{|\mathcal{U}|}{|\mathcal{I}| * 1000}, \quad (11)$$

Finally, the last two characteristics are measured as follows:

$$g_{\text{UserRatings}}(\mathbf{X}) = \frac{|\mathcal{R}|}{|\mathcal{U}|} \quad \text{and} \quad g_{\text{ItemRatings}}(\mathbf{X}) = \frac{|\mathcal{R}|}{|\mathcal{I}|}. \quad (12)$$

Intuitively, UserRatings measures how active the users are in the system while ItemRatings relates to how long-tail items compare to popular ones.

**Rating and feedback frequency.** Ratings and feedback are usually distributed over catalog items following the well-known long-tail distribution, with a small number of items being highly popular. Examining this characteristic helps in understanding how biased the algorithms

could be toward popular items:

$$g_{\text{ItemGini}}(\mathbf{X}) = 1 - 2 \sum_{i=1}^{|\mathcal{I}|} \frac{|\mathcal{I}| + 1 - i}{|\mathcal{I}| + 1} \times \frac{|\mathcal{R}_i|}{|\mathcal{R}|} \quad (13)$$

where  $|\mathcal{R}_i|$  is the number of ratings in  $\mathbf{X}$  associated with item  $i$ . ItemGini coefficient measures the rating frequency distribution for items. A value  $g_{\text{ItemGini}} = 0$  represents the situation where all the items got the same number of ratings while  $\text{ItemGini} = 1$  indicates that only one item received all the ratings.

**Experimental Protocol.** In order to cover a wide range of values for each characteristic, we sampled 600 random sub-datasets<sup>1</sup> from three well-known datasets: MovieLens 1M [19], Amazon Digital Music, and LibraryThing<sup>2</sup>. Each of the 1800 generated sub-datasets was privatized using the randomized response applied with three different  $\epsilon$  values (3, 2, and 0.5). The resulting 7200 datasets have been used to train four recommendation models, under different approaches: popularity-based (*MostPop*), distance-based (*ItemKNN* [20]), autoencoder (*EASER* [21]), and graph (*RP3 $\beta$*  [22]). The explanatory model (see Eq. (9)) is trained on 28,800 experiments to analyze the variation of accuracy and popularity bias with respect to the characteristics of the non-privatized datasets. The popular Precision and Average Recommendation Popularity (ARP) [23] metrics have been calculated on the top-10 recommendation lists. The regression variables have been normalized.

**Discussion.** Table 1 reports the learned coefficients of the regression model. Each coefficient represents the relationship between the data characteristic and the variation of accuracy and popularity bias in the recommendation. Higher values of the regressor  $R^2$  indicate that the set of variables can explain the variation of the target variable. Conversely, for each model-variable-characteristic combination, the model weight expresses the importance and direction of the relationship, i.e., direct or inverse. Higher values entail higher variations in the target metric, while the higher the delta, the greater the performance degradation. For completeness, a degradation of precision implies a lower accuracy, and a degradation of ARP stands for a more downward popularity bias. Finally, the  $p$ -values reported with the stars notation indicate when the result is statistically significant. *When considering the variation of precision, the regression model can explain the 89% of the variation for MovieLens 1M, the 53% for Amazon Digital Music, and 63% for LibraryThing, along with the four models. Even higher when considering ARP, with 91% for MovieLens 1M, Amazon Digital Music, and*

*LibraryThing on average. This demonstrates that the selected six characteristics effectively explain the variation of performance.*

It is worth noticing that the explanatory capabilities seem to depend more on the dataset choice than on the recommender model. Indeed, the lowest values of  $R^2$  are found for the Amazon Digital Music dataset. Conversely, MovieLens 1M characteristics are very informative. This could relate to the substantial difference in data sparsity: a higher relative number of interactions in MovieLens 1M could help the regressor unveil hidden relationships. The *privacy budget*  $\epsilon$  has a key role in the explanatory model: its decrease is always related to a worsening of accuracy and an increase in popularity bias. Indeed, small  $\epsilon$  values, i.e., strong anonymization of the dataset, tend to preserve popular items. The finding is confirmed by the statistical hypothesis tests. Furthermore, the dataset SpaceSize shows an inverse dependency on the degradation of accuracy. Differently from  $\epsilon$ , it shows an inverse relationship with ARP. The larger the SpaceSize, the lower the probability of returning a popular item in the randomized dataset. For UserRatings and ItemRatings, they respectively have a negative and a positive impact on the increase of the popularity bias. A higher ItemRatings makes the dataset more affected by popular items and the randomized response reinforces this behavior. On the contrary, a high UserRatings drives the dataset to be less affected by cold users, whose recommendations are strongly influenced by bias. Finally, the results for Shape and ItemGini are largely statistically significant and show the ability to explain both the accuracy and the popularity bias variations.

## 5. Conclusion and Future Work

This study has extensively analyzed which dataset characteristics are more prone to influence the accuracy and popularity bias of different recommendation models when randomized response is applied to the original dataset. Under the lens of plausible deniability, privacy budget, and explanatory modeling, this investigation has unveiled several insights and provided interesting suggestions to the researcher interested in protecting users' privacy. However, the relationship between dataset characteristics across different datasets deserves further attention. Future studies could unveil these aspects and extend our analysis considering the effects of applying other differential privacy techniques.

## References

- [1] C. Cadwalladr, E. Graham-Harrison, Revealed: 50 million facebook profiles harvested for cambridge

<sup>1</sup>Their densities were forced within [0.0007, 0.04] to ensure realistic values.

<sup>2</sup><https://cseweb.ucsd.edu/~jmcauley/datasets.html>

- analytica in major data breach, *The guardian* 17 (2018) 22.
- [2] C. Dwork, A. Roth, The algorithmic foundations of differential privacy, *Foundations and Trends in Theoretical Computer Science* 9 (2014) 211–407. URL: <https://doi.org/10.1561/0400000042>. doi:10.1561/0400000042.
- [3] A. Friedman, S. Berkovsky, M. A. Kâafar, A differential privacy framework for matrix factorization recommender systems, *User Model. User Adapt. Interact.* 26 (2016) 425–458.
- [4] A. Machanavajjhala, A. Korolova, A. D. Sarma, Personalized social recommendations - accurate or private?, *Proc. VLDB Endow.* 4 (2011) 440–450.
- [5] T. Guo, J. Luo, K. Dong, M. Yang, Differentially private graph-link analysis based social recommendation, *Inf. Sci.* 463-464 (2018) 214–226.
- [6] J. Zhang, C. Chow, Enabling probabilistic differential privacy protection for location recommendations, *IEEE Trans. Serv. Comput.* 14 (2021) 426–440.
- [7] J. Hua, C. Xia, S. Zhong, Differentially private matrix factorization, in: *IJCAI*, AAAI Press, 2015, pp. 1763–1770.
- [8] F. Zhang, V. E. Lee, K. R. Choo, Jo-dpmf: Differentially private matrix factorization learning through joint optimization, *Inf. Sci.* 467 (2018) 271–281.
- [9] J. S. Kim, J. W. Kim, Y. D. Chung, Successive point-of-interest recommendation with local differential privacy, *IEEE Access* 9 (2021) 66371–66386.
- [10] T. Qi, F. Wu, C. Wu, Y. Huang, X. Xie, Privacy-preserving news recommendation model learning, in: *EMNLP (Findings)*, volume *EMNLP 2020 of Findings of ACL*, Association for Computational Linguistics, 2020, pp. 1423–1432.
- [11] A. Narayanan, V. Shmatikov, How to break anonymity of the netflix prize dataset, *CoRR abs/0610105* (2006). URL: <http://arxiv.org/abs/cs/0610105>. arXiv:cs/0610105.
- [12] P. Samarati, L. Sweeney, Generalizing data to provide anonymity when disclosing information (abstract), in: A. O. Mendelzon, J. Paredaens (Eds.), *Proceedings of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, June 1-3, 1998, Seattle, Washington, USA, ACM Press, 1998, p. 188. URL: <https://doi.org/10.1145/275487.275508>. doi:10.1145/275487.275508.
- [13] J. Domingo-Ferrer, J. Soria-Comas, Connecting randomized response, post-randomization, differential privacy and t-closeness via deniability and permutation, *CoRR abs/1803.02139* (2018).
- [14] G. Adomavicius, J. Zhang, Impact of data characteristics on recommender systems performance, *ACM Trans. Management Inf. Syst.* 3 (2012) 3:1–3:17. URL: <https://doi.org/10.1145/2151163.2151166>. doi:10.1145/2151163.2151166.
- [15] S. L. Warner, Randomized response: A survey technique for eliminating evasive answer bias, *Journal of the American Statistical Association* 60 (1965) 63–69. URL: <https://www.tandfonline.com/doi/abs/10.1080/01621459.1965.10480775>. doi:10.1080/01621459.1965.10480775. arXiv:<https://www.tandfonline.com/doi/pdf/10.1080/01621459.1965.10480775>. PMID: 12261830.
- [16] B. G. Greenberg, A.-L. A. Abul-Ela, W. R. Simmons, D. G. Horvitz, The unrelated question randomized response model: Theoretical framework, *Journal of the American Statistical Association* 64 (1969) 520–539. URL: <https://www.tandfonline.com/doi/abs/10.1080/01621459.1969.10500991>. doi:10.1080/01621459.1969.10500991. arXiv:<https://www.tandfonline.com/doi/pdf/10.1080/01621459.1969.10500991>.
- [17] Y. Wang, X. Wu, D. Hu, Using randomized response for differential privacy preserving data collection, in: *EDBT/ICDT Workshops*, volume 1558 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2016.
- [18] Y. Deldjoo, A. Bellogín, T. D. Noia, Explaining recommender systems fairness and accuracy through the lens of data characteristics, *Inf. Process. Manag.* 58 (2021) 102662. URL: <https://doi.org/10.1016/j.ipm.2021.102662>. doi:10.1016/j.ipm.2021.102662.
- [19] F. M. Harper, J. A. Konstan, The movielens datasets: History and context, *TiiS* 5 (2016) 19:1–19:19.
- [20] Y. Koren, Factor in the neighbors: Scalable and accurate collaborative filtering, *TKDD* 4 (2010) 1:1–1:24.
- [21] H. Steck, Embarrassingly shallow autoencoders for sparse data, in: *WWW*, ACM, 2019, pp. 3251–3257.
- [22] B. Paudel, F. Christoffel, C. Newell, A. Bernstein, Updatable, accurate, diverse, and scalable recommendations for interactive applications, *ACM Trans. Interact. Intell. Syst.* 7 (2017) 1:1–1:34.
- [23] H. Abdollahpouri, R. Burke, B. Mobasher, Managing popularity bias in recommender systems with personalized re-ranking, in: R. Barták, K. W. Brawner (Eds.), *Proceedings of the Thirty-Second International Florida Artificial Intelligence Research Society Conference*, Sarasota, Florida, USA, May 19-22 2019, AAAI Press, 2019, pp. 413–418. URL: <https://aaai.org/ocs/index.php/FLAIRS/FLAIRS19/paper/view/18199>.