frontiers | Frontiers in Big Data

TYPE Mini Review
PUBLISHED 12 October 2023
DOI 10.3389/fdata.2023.1249997

OPEN ACCESS

EDITED BY
Yassine Himeur,
University of Dubai, United Arab Emirates

REVIEWED BY
Chaochao Chen,
Zhejiang University, China

*CORRESPONDENCE
Peter Müllner
✉ pmuellner@know-center.at;
✉ pmuellner@student.tugraz.at
Elisabeth Lex
✉ elisabeth.lex@tugraz.at
Dominik Kowald
✉ dkowald@know-center.at

RECEIVED 29 June 2023
ACCEPTED 25 September 2023
PUBLISHED 12 October 2023

CITATION
Müllner P, Lex E, Schedl M and Kowald D (2023) Differential privacy in collaborative filtering recommender systems: a review. *Front. Big Data* 6:1249997. doi: 10.3389/fdata.2023.1249997

Differential privacy in collaborative filtering recommender systems: a review

Peter Müllner[1,2]*, Elisabeth Lex[2]*, Markus Schedl[3,4] and Dominik Kowald[1,2]*

[1]Know-Center Gmbh, Graz, Austria, [2]Institute of Interactive Systems and Data Science, Graz University of Technology, Graz, Austria, [3]Institute of Computational Perception, Johannes Kepler University Linz, Linz, Austria, [4]Linz Institute of Technology, Linz, Austria

State-of-the-art recommender systems produce high-quality recommendations to support users in finding relevant content. However, through the utilization of users' data for generating recommendations, recommender systems threaten users' privacy. To alleviate this threat, often, differential privacy is used to protect users' data via adding random noise. This, however, leads to a substantial drop in recommendation quality. Therefore, several approaches aim to improve this trade-off between accuracy and user privacy. In this work, we first overview threats to user privacy in recommender systems, followed by a brief introduction to the differential privacy framework that can protect users' privacy. Subsequently, we review recommendation approaches that apply differential privacy, and we highlight research that improves the trade-off between recommendation quality and user privacy. Finally, we discuss open issues, e.g., considering the relation between privacy and fairness, and the users' different needs for privacy. With this review, we hope to provide other researchers an overview of the ways in which differential privacy has been applied to state-of-the-art collaborative filtering recommender systems.

KEYWORDS

differential privacy, collaborative filtering, recommender systems, accuracy-privacy trade-off, review

1. Introduction

Several previous research works have revealed multiple privacy threats for users in recommender systems. For example, the disclosure of users' private data to untrusted third parties (Calandrino et al., 2011), or the inference of users' sensitive attributes, such as gender or age (Zhang et al., 2023). Similarly, also the users themselves care more about their privacy in recommender systems (Herbert et al., 2021). For these reasons, privacy-enhancing techniques have been applied, most prominently *differential privacy (DP)* (Dwork, 2008). DP injects random noise into the recommender system and formally guarantees a certain degree of privacy. However, through this random noise, the quality of the recommendations suffers (Berkovsky et al., 2012). Many works aim to address this trade-off between recommendation quality and user privacy via carefully applying DP in specific ways. Friedman et al. (2016) show that in case of matrix factorization, DP can be applied to three different parts of the recommender system: (i) to the input of the recommender system, (ii) within the training process of the model, and (iii) to the model after training. However, a concise overview of works with respect to these three categories does not exist yet.

Therefore, in the paper at hand, we address this gap and identify 26 papers from relevant venues that deal with DP in collaborative filtering recommender systems. We briefly review these 26 papers and make two key observations about the state-of-the-art. Firstly, the vast majority of works use datasets from the same non-sensitive domain, i.e., movies. Secondly, research on applying DP after model training is scarce. Finally, we discuss our findings and present two open questions that may be relevant for future research: *How does applying DP impact fairness?* and *How to quantify the user's perceived privacy?*

Our work is structured as follows: In Section 2, we present threats to the privacy of users in recommender systems and additionally, introduce the DP framework. In Section 3, we precisely outline our methodology for obtaining the set of 26 relevant papers. In Section 4, we review these papers and group them into three groups according to the way in which they apply DP. In Section 5, we discuss our findings and propose open issues that we identified.

2. Background

In recent years, users of recommender systems have shown increasing concerns with respect to keeping their data private (Herbert et al., 2021). In fact, several research works (Bilge et al., 2013; Jeckmans et al., 2013; Friedman et al., 2015; Beigi and Liu, 2020; Majeed and Lee, 2020; Himeur et al., 2022) have revealed multiple privacy threats, for example, the inadvertent disclosure of users' interaction data, or the inference of users' sensitive attributes (e.g., gender, age).

Typically, a recommender system utilizes historic interaction data to generate recommendations. Ramakrishnan et al. (2001) show that in k nearest neighbors recommender systems, the recommendations could disclose the interaction data of the neighbors, i.e., users, whose interaction data is utilized to generate the recommendations. Similarly, Calandrino et al. (2011) inject fake users to make the recommendations more likely to disclose the neighbors' interaction data, and also, they can infer users' interaction data based on the public outputs of a recommender system, e.g., public interaction data or public product reviews. Furthermore, Hashemi et al. (2022) and Xin et al. (2023) aim to learn user behavior via observing many recommendations and, in this way, can disclose parts of a user's interaction data. Weinsberg et al. (2012) show that an adversary could infer sensitive attributes, in this case, gender, based on a user's interaction data. Their attack relies on a classifier that leverages a small set of training examples to learn the correlation between a user's preferences and gender. Likewise, Ganhör et al. (2022) show that recommender systems based on autoencoder architectures are vulnerable to infer the user's gender from the latent user representation. The authors also propose an adversarial training regime to mitigate this problem. Similarly, also Zhang et al. (2023) infer the age and gender of users in a federated learning recommender system. In summary, many of a user's sensitive attributes can be inferred via thoroughly analyzing the user's digital footprint (e.g., the behavior in a recommender system or social media platform) (Kosinski et al., 2013).

Overall, the utilization of users' interaction data for generating recommendations poses a privacy risk for users. Therefore, privacy-enhancing techniques, such as homomorphic encryption (Gentry, 2009), federated learning (McMahan et al., 2017), or most prominently, *differential privacy (DP)* (Dwork, 2008) have been applied to protect users' privacy. Specifically, DP is applied via injecting noise into the recommender system. This ensures that the recommender system uses noisy data instead of the real data. For example, an additive mechanism samples random noise from the Laplace or Gaussian distribution and adds it to the users' rating data (Dwork and Roth, 2014). Alternatively, the randomized responses mechanism flips a fair coin, which decides whether to use the real data or random data, and this way, ensures DP (Warner, 1965; Dwork and Roth, 2014). Overall, the degree of noise that is used is defined by the parameter ϵ, i.e., the privacy budget. Intuitively, the smaller the ϵ-value is, the better the privacy, but the stronger the expected accuracy drop. Therefore, choosing ϵ is non-trivial and depends on the specific use case (Dwork, 2008).

3. Review methodology

To conduct our review, we chose relevant conferences in the field, i.e., ACM SIGIR, TheWebConf, ACM KDD, IJCAI, ACM CIKM, and ACM RecSys and journals, i.e., TOIS, TIST, UMUAI, and TKDE. Adopting a keyword-based search, we identified relevant publications in the proceedings via querying the full-texts for "differential privacy" and "recommender system", "recommend", "recommendation", or "recommender". We manually checked the resulting papers for their relevance and retrieved 16 publications. In addition, we conducted a literature search on Google Scholar using the same keywords and procedure, which resulted in 10 publications. Overall, we considered 26 publications in the paper at hand.

4. Recommender systems with differential privacy

According to Friedman et al. (2016), DP can be applied via (i) adding noise to the input of a collaborative filtering-based recommender system, e.g., the user data or other user representations, (ii) adding noise to the training process of the model, i.e., the model updates, or (iii) adding noise to the model after training, i.e., to the resulting latent factors. In Table 1, we group the selected publications into these three categories.

4.1. Differential privacy applied to the user representation

In collaborative filtering recommender systems, the input to the system is typically given by interaction or rating data. However, more complex user representations exist, e.g., neural-based user embeddings.

Chen et al. (2020) protect POI (point of interest) interaction data of users, e.g., a user visited a restaurant, with DP. Specifically, they use this data to privately calculate POI features, i.e., the

TABLE 1 Overview of the reviewed 26 publications.

References	Domain(s)	DP applied to		
		User represent.	Model updates	After training
Long et al. (2023)	Location	•		
Müllner et al. (2023)	Movies, Music, Books, Social	•		
Neera et al. (2023)	Movies, Jokes, Dating	•		
Wang et al. (2023)	Movies, Music		•	
Chai et al. (2022)	Movies, Location	•		
Chen et al. (2022)	Movies, Music, Books	•		
Jiang et al. (2022)	Movies, Music, Location, Groceries		•	
Liu et al. (2022)	Social		•	
Ning et al. (2022)	Movies		•	
Ran et al. (2022)	Movies, Music			•
Ren et al. (2022)	Social	•		
Wu et al. (2022)	Advertisement	•		
Li et al. (2021)	Movies, Dating		•	
Minto et al. (2021)	Movies		•	
Zhang et al. (2021)	Movies	•		•
Chen et al. (2020)	Location	•		
Gao et al. (2020)	Movies, Smartphone	•		
Ma et al. (2019)	Health		•	
Meng et al. (2018)	Social		•	
Shin et al. (2018)	Movies, Dating		•	
Liu et al. (2017)	Movies	•		
Yang et al. (2017)	Movies	•		
Li et al. (2016)	Movies	•		
Hua et al. (2015)	Movies		•	•
Zhu et al. (2013)	Movies	•		
Zhao et al. (2011)	Movies	•		

We mark whether DP is applied to the user representation, to the model updates, or after training. Domain(s) refers to the domain(s) in which the recommendations are evaluated. We sort the publications with respect to recency.

number of visitors per restaurant, which are subsequently used for generating recommendations instead of the DP-protected interaction data. This way, they can increase recommendation accuracy. Similarly, Long et al. (2023) use DP to recommend POIs, but in a decentralized fashion. A central server collects public data to train a recommendation model and to privately identify groups of similar users. DP is used for privately calculating user-user similarities. Then, users locally use information from similar users, which leads to a better trade-off between recommendation quality and privacy than comparable approaches.

Liu et al. (2017) add noise to users' rating data and to the user-user covariance matrix to ensure DP of a KNN-based recommender system. They show that this leads to better privacy than in case only the covariance matrix is protected via DP. Besides revealing users' rating data, an attacker could also aim to infer sensitive attributes (e.g., gender) of the users. Therefore, Chai et al. (2022) propose an obfuscation model to protect gender information. After applying this obfuscation model, users protect their data via DP and send it to a central server. Yang et al. (2017) use the Johnson-Lindenstrauss transform (Blocki et al., 2012), i.e., they ensure DP via multiplying the original interaction matrix with a random matrix. Using this protected matrix, their approach guarantees differential privacy and also can even generate more accurate recommendations than a non-private approach. Neera et al. (2023) underline that adding Laplacian noise can lead to "unrealistic" rating values, i.e., outside the rating range, and through this, recommendation accuracy can drop severely. Therefore, they bound the noisy ratings to a "realistic" value range without harming DP. Plus, they use a Gaussian mixture model to estimate and then remove noise in the recommendation process to keep recommendation accuracy.

Cross-domain recommendation models can increase recommendation accuracy in the target domain by exploiting data from multiple source domains. To protect user privacy when data from the source domain is made available to the target domain,

Chen et al. (2022) use the Johnson-Lindenstrauss transform. Due to the high sparsity of the rating matrix, they employ a variant that performs better when applied to sparse matrices (Ailon and Chazelle, 2009). Ren et al. (2022) utilize data from different social network platforms to generate recommendations and apply DP to the user attributes and the connections in the social network graphs. Plus, they apply a variant of DP to protect textual data (Fernandes et al., 2019). Moreover, to increase the click-through rate for recommended advertisements, Wu et al. (2022) leverage user interaction data from multiple platforms. First, user embeddings are generated per platform and then protected with DP. Second, the recommender system collects and aggregates a user's DP-protected embeddings across platforms and then applies DP again to the aggregated user embedding. According to the authors, applying DP after aggregation allows for smaller noise levels when applying DP to the per-platform user embeddings, which results in higher accuracy. Typically, many users use a variety of different online platforms. Therefore, Li et al. (2016) leverage these multiple data sources per user to increase recommendation accuracy. Specifically, they combine DP-protected item-item similarities from dataset *B* as auxiliary data that helps to generate more accurate recommendations for users in dataset *A* (cf. Zhao et al., 2011).

Gao et al. (2020) compute item-item similarities by using DP-protected user interaction data. With these item-item similarities, users can locally generate recommendations on their own devices, therefore not harming their privacy. The item-based KNN recommender system proposed by Zhu et al. (2013) utilizes DP in two ways: First, they randomly rearrange the most similar neighbors to foster privacy. Second, they measure how the item-item similarity changes if a specific user interaction was not present, and with this, they add the necessary level of noise to the users' interactions. This way, recommendation accuracy can be better preserved than with approaches that apply the same level of noise to all user interactions. For user-based KNN, Müllner et al. (2023) identify neighbors that can be reused for many recommendations. This way, only a small set of users are used as neighbors for many recommendations and need to be protected with DP. Many users, however, are only rarely utilized as neighbors and therefore do not need to be protected with DP. Overall, this yields more accurate recommendations than in case DP needs to be applied to all users.

4.2. Differential privacy applied to the model updates

Some recommender systems do not process user data and create user representations on a central server, instead, they compute the model updates, i.e., gradients, locally on their users' device. Then, the recommender system collects these gradients to adapt its recommendation model. To prohibit the leakage of user data through these gradients (Bagdasaryan et al., 2020), DP can be applied.

For example, Hua et al. (2015) add noise to the gradients of the recommendation model to ensure DP. However, due to the sparsity of the gradients, the application of DP can be ineffective and information about what items have been rated by the user can be disclosed. To address this problem, Shin et al. (2018) use DP to mask whether a user appears in the dataset. Also, they formally show that the noise added to the gradients hinders a fast convergence of the recommendation model, and in this way, increases the training time. Therefore, they introduce a stabilization factor to enable better training of the recommendation model. Wang et al. (2023) propose a recommender system that uses a special DP-mechanism (Zhao et al., 2020) to simultaneously protect the rating values and the set of items that is rated by a user. The DP-protected item-vectors are then send to a central server, which performs dimensionality reduction to reduce the accuracy drop (cf. Shin et al., 2018). In Minto et al. (2021), users receive a global model from a central server and, then, compute their respective updates locally. These updates are protected via DP, before being sent back to the server. Plus, the number of updates per user are restricted to further improve privacy. Moreover, the authors highlight that high-dimensional gradients can negatively impact the recommendation quality, as they are especially prone to higher sparsity (cf. Hua et al., 2015; Shin et al., 2018). When DP is applied, the gradients become denser since noise is added to the entire gradient, including the zero-entries. This, in practice, leads to additional communication overhead, since all non-zero-entries need to be transmitted (Ning et al., 2022). Therefore, Ning et al. only add noise to the non-zero gradients. This way, the communication overhead is reduced; however, DP cannot be guaranteed anymore.

Jiang et al. (2022) reduce the accuracy drop via an adaptive DP mechanism that depends on the number of training steps. Intuitively, after many training steps, the model fine-tunes its predictions and the gradients need to be measured more accurately than during the beginning of the model training. Thus, they add more noise in the beginning and less noise in the end of the training process. This yields more accurate recommendations than a static DP mechanism that always adds the same level of noise. Li et al. (2021) also use noisy model updates to ensure DP. They observe that noise can lead to large values for the user embeddings, which increases the sensitivity and therefore also the level of noise that is required to ensure DP. To foster recommendation quality, they map the user embeddings to a certain range, which bounds the sensitivity and requires less noise. Liu et al. (2022) leverage user interactions and social connections to generate recommendations via a federated graph neural network. To ensure DP, they add noise to the gradients that are sent to a central server. However, gradients with different magnitudes have different sensitivities (cf. Li et al., 2021), and thus, need a different level of noise to ensure DP. Therefore, they fit the noise level to the gradient magnitudes to satisfy DP, but also, to preserve recommendation accuracy.

Ma et al. (2019) employ federated tensor factorization in the health domain. A global model is distributed to hospitals, which locally update the model based on their data. To protect privacy, a variant of DP is applied to the model updates, which are subsequently sent to the global server to adapt the global model. Meng et al. (2018) randomly divide users' ratings into non-sensitive and sensitive ratings. For sensitive ratings, they apply more noise than for non-sensitive ratings. With this, their approach can preserve higher recommendation accuracy than in case the same noise level is used for sensitive and non-sensitive data.

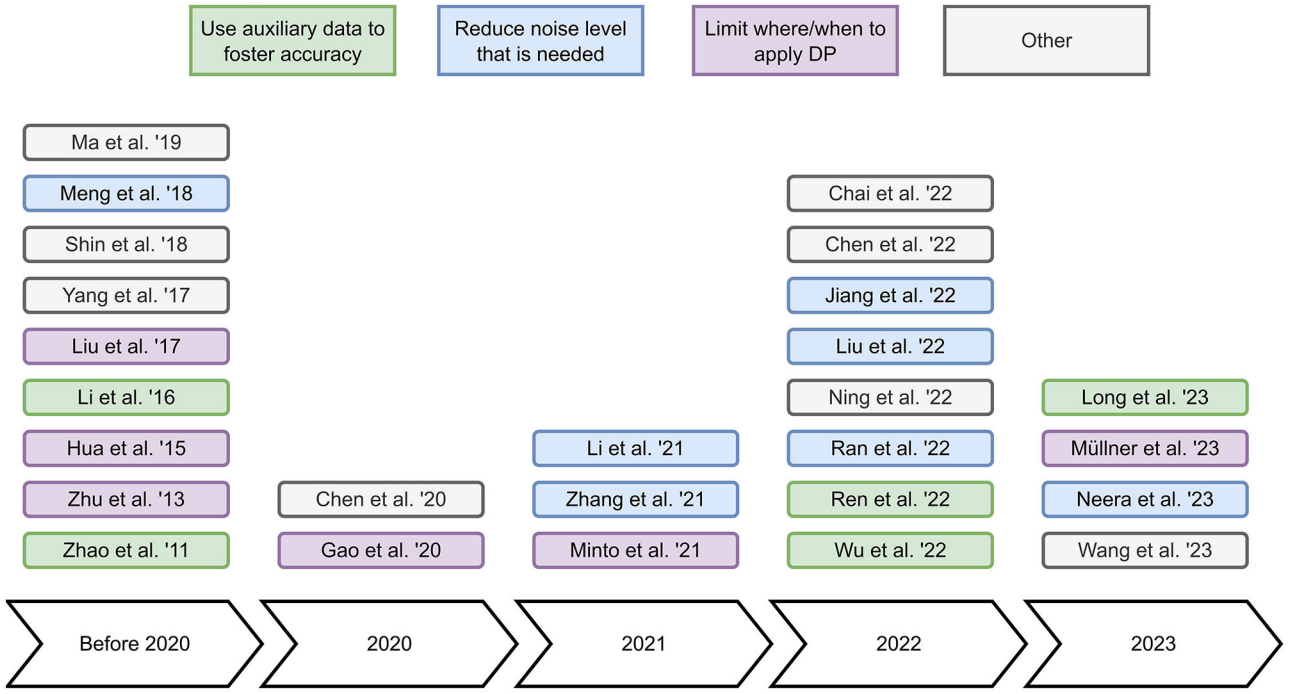

FIGURE 1
Overview of the main concepts of the reviewed papers. *Use auxiliary data to foster accuracy* refers to the incorporation of data from other domains, datasets or users to increase recommendation accuracy. *Reduce noise level that is needed* refers to designing recommender systems that require a minimal amount of noise to ensure DP. *Limit where/when to apply DP* refers to carefully minimizing the application of DP. *Other* refers to approaches that do not fit into the previous categories.

4.3. Differential privacy applied after training

Only few works apply DP to the recommendation model after training. In case of a matrix factorization approach, noise can be added to the learned user- and item-vectors to ensure DP. Our selected publications (see Section 3) do not include any works that apply DP exclusively to the model after training. Nevertheless, we describe works that apply DP to the user representation or the model updates, but also after training.

For example, Hua et al. (2015) consider a matrix factorization model, where the model sends item-vectors back to the users and this way, users' data can get leaked. To prohibit this, Hua et al. perturb the model's objective function after training via adding noise to the latent item-vectors. Similarly, Ran et al. (2022) also use DP to prohibit data leakage through the item-vectors that are sent to the users. Specifically, a trusted recommender system generates a matrix factorization model. Instead of publishing the item-vectors of this model, they learn new item-vectors on the DP-protected user-vectors. Through this, they can minimize the noise that is introduced and thus, can improve recommendation accuracy over comparable approaches. Zhang et al. (2021) apply DP to the user representation and also, to the model after training. Specifically, they use a polynomial approximation of the model's loss function to efficiently compute the sensitivity of the dataset and, accordingly, adapt the level of noise that is added to the loss function.

5. Summary and open questions

In this review, we investigate research works that apply DP to collaborative filtering recommender systems. We identify 26 relevant works and categorize these based on how they apply DP, i.e., to the user representation, to the model updates, or to the model after training (see Table 1). In addition, we briefly summarize these relevant works to obtain a broad overview of the state-of-the-art. Furthermore, we identify the main concepts of the relevant works in Figure 1 to help readers to understand in which diverse ways the reviewed papers apply DP to improve the accuracy-privacy trade-off. Our main findings from reviewing the discussed literature are two-fold: (i) The majority of works use datasets from the same non-sensitive domain, i.e., movies, and (ii) applying DP to the model after training seems to be an understudied topic.

Many research works use datasets from the movie domain, which, in general, does not include sensitive data. For research on DP in collaborative filtering recommender systems, however, datasets from sensitive domains may be better suited to resemble real-world privacy threats well. For example, datasets from the health, finance, or job domain. Moreover, the majority of research focuses on either applying DP to the user representation or to the model updates. Research on applying DP to the model after training is scarce, and therefore, this opens up the possibility of future work to fill this gap.

Our review of relevant work allows to grasp the state-of-the-art and to identify the following open research questions:

Q1: How does applying DP impact fairness? Dwork et al. (2012) and Zemel et al. (2013) suggest that in theory, privacy can lead to fairness and fairness can lead to privacy. The reason is that for both, a user's data shall be hidden, either to ensure privacy or to prohibit discrimination based on this data. However, in practice, correlations in private data can still lead to unfairness (Ekstrand et al., 2018; Agarwal, 2020). Only recently, Yang et al. (2023) and Sun et al. (2023) investigate the connection between privacy and fairness in recommender systems. For example, Sun et al. (2023) use DP-protected information to re-rank the items in the recommendation list and in this way, increase a more fair exposure of items. Nonetheless, the impact of DP on fairness remains an understudied topic.

Q2: How to quantify the user's perceived privacy? Users perceive privacy differently, e.g., some users tolerate disclosing their gender, while others refuse to do this (Joshaghani et al., 2018). This perceived privacy depends on many factors, e.g., context or situational factors (Knijnenburg and Kobsa, 2013; Mehdy et al., 2021). However, measuring users' perceived privacy is hard and is usually done via questionnaires (Knijnenburg and Kobsa, 2013). This is in stark contrast to how privacy is measured in the DP framework, i.e., via quantifying to what extent the data impacts the output of the recommender system. Therefore, developing methods to better quantify users' privacy is an important future research avenue.

Author contributions

PM: literature analysis, conceptualization, and writing. MS: conceptualization and writing. EL and DK: conceptualization, writing, and supervision. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the DDAI COMET Module within the COMET-Competence Centers for Excellent Technologies Programme, funded by the Austrian Federal Ministry for Transport, Innovation and Technology (BMVIT), the Austrian Federal Ministry for Digital and Economic Affairs (BMDW), the Austrian Research Promotion Agency (FFG), the province of Styria (SFG), and partners from industry and academia. The COMET Programme is managed by FFG. In addition, the work received funding from the TU Graz Open Access Publishing Fund and from the Austrian Science Fund (FWF): DFH-23 and P33526.

Conflict of interest

PM was employed by Know-Center Gmbh.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Agarwal, S. (2020). *Trade-offs between fairness, interpretability, and privacy in machine learning* (Master's thesis). University of Waterloo, Waterloo, ON, Canada.

Ailon, N., and Chazelle, B. (2009). The fast Johnson–Lindenstrauss transform and approximate nearest neighbors. *SIAM J. Comput.* 39, 302–322. doi: 10.1137/060673096

Bagdasaryan, E., Veit, A., Hua, Y., Estrin, D., and Shmatikov, V. (2020). "How to backdoor federated learning," in *International Conference on Artificial Intelligence and Statistics* (Palermo: PMLR), 2938–2948.

Beigi, G., and Liu, H. (2020). A survey on privacy in social media: identification, mitigation, and applications. *ACM Trans. Data Sci.* 1, 1–38. doi: 10.1145/3343038

Berkovsky, S., Kuflik, T., and Ricci, F. (2012). The impact of data obfuscation on the accuracy of collaborative filtering. *Expert Syst. Appl.* 39, 5033–5042. doi: 10.1016/j.eswa.2011.11.037

Bilge, A., Kaleli, C., Yakut, I., Gunes, I., and Polat, H. (2013). A survey of privacy-preserving collaborative filtering schemes. *Int. J. Softw. Eng. Knowledge Eng.* 23, 1085–1108. doi: 10.1142/S0218194013500320

Blocki, J., Blum, A., Datta, A., and Sheffet, O. (2012). "The Johnson-Lindenstrauss transform itself preserves differential privacy," in *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science* (New Brunswick, NJ: IEEE), 410–419.

Calandrino, J. A., Kilzer, A., Narayanan, A., Felten, E. W., and Shmatikov, V. (2011). ""You might also like:" privacy risks of collaborative filtering," in *Proc. of S&P'11* (Oakland, CA: IEEE), 231–246.

Chai, D., Wang, L., Chen, K., and Yang, Q. (2022). Efficient federated matrix factorization against inference attacks. *ACM Trans. Intell. Syst. Technol.* 13, 1–20. doi: 10.1145/3501812

Chen, C., Wu, H., Su, J., Lyu, L., Zheng, X., and Wang, L. (2022). "Differential private knowledge transfer for privacy-preserving cross-domain recommendation," in *Proc. of ACM WWW'22* (Lyon).

Chen, C., Zhou, J., Wu, B., Fang, W., Wang, L., Qi, Y., et al. (2020). Practical privacy preserving POI recommendation. *ACM Trans. Intell. Syst. Technol.* 11, 1455–1465. doi: 10.1145/3394138

Dwork, C. (2008). "Differential privacy: a survey of results," in *International Conference on Theory and Applications of Models of Computation* (Berlin: Springer), 1–19.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., and Zemel, R. (2012). "Fairness through awareness," in *Proc. of ITCS'12* (Cambridge, MA), 214–226.

Dwork, C., and Roth, A. (2014). The algorithmic foundations of differential privacy. *Found. Trends Theoret. Comput. Sci.* 9, 211–407. doi: 10.1561/0400000042

Ekstrand, M. D., Joshaghani, R., and Mehrpouyan, H. (2018). "Privacy for all: Ensuring fair and equitable privacy protections," in *Proc. of FAccT'18* (New York, NY: PMLR), 35–47.

Fernandes, N., Dras, M., and McIver, A. (2019). "Generalised differential privacy for text document processing," in *Principles of Security and Trust: 8th International Conference, POST 2019* (Prague: Springer International Publishing), 123–148.

Friedman, A., Berkovsky, S., and Kaafar, M. A. (2016). A differential privacy framework for matrix factorization recommender systems. *User Model. User Adapt. Interact.* 26, 425–458. doi: 10.1007/s11257-016-9177-7

Friedman, A., Knijnenburg, B. P., Vanhecke, K., Martens, L., and Berkovsky, S. (2015). "Privacy aspects of recommender systems," in *Recommender Systems Handbook*, eds F. Ricci, L. Rokach, and B. Shapira (Boston, MA: Springer), 649–688.

Ganhör, C., Penz, D., Rekabsaz, N., Lesota, O., and Schedl, M. (2022). "Unlearning protected user attributes in recommendations with adversarial training," in *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, eds E. Amigó, P. Castells, J. Gonzalo, B. Carterette, J. S. Culpepper, and G. Kazai (Madrid: ACM), 2142–2147.

Gao, C., Huang, C., Lin, D., Jin, D., and Li, Y. (2020). "DPLCF: differentially private local collaborative filtering," in *Proc. of SIGIR'20* (Xi'an), 961–970.

Gentry, C. (2009). *A Fully Homomorphic Encryption Scheme (Dissertation)*. Stanford University.

Hashemi, H., Xiong, W., Ke, L., Maeng, K., Annavaram, M., Suh, G. E., et al. (2022). Data leakage via access patterns of sparse features in deep learning-based recommendation systems. *arXiv preprint arXiv:2212.06264*.

Herbert, C., Marschin, V., Erb, B., Meißner, D., Aufheimer, M., and Bösch, C. (2021). Are you willing to self-disclose for science? Effects of privacy awareness and trust in privacy on self-disclosure of personal and health data in online scientific studies–an experimental study. *Front. Big Data* 4, 763196. doi: 10.3389/fdata.2021.763196

Himeur, Y., Sohail, S. S., Bensaali, F., Amira, A., and Alazab, M. (2022). Latest trends of security and privacy in recommender systems: a comprehensive review and future perspectives. *Comput. Sec.* 118, 102746. doi: 10.1016/j.cose.2022.102746

Hua, J., Xia, C., and Zhong, S. (2015). "Differentially private matrix factorization," in *International Joint Conference on Artificial Intelligence* (Buenos Aires).

Jeckmans, A. J., Beye, M., Erkin, Z., Hartel, P., Lagendijk, R. L., and Tang, Q. (2013). "Privacy in recommender systems," in *Social Media Retrieval*, eds N. Ramzan, R. van Zwol, J.-S. Lee, K. Clüver, and X.-S. Hua (London: Springer), 263–281.

Jiang, X., Liu, B., Qin, J., Zhang, Y., and Qian, J. (2022). "FedNCF: federated neural collaborative filtering for privacy-preserving recommender system," in *2022 International Joint Conference on Neural Networks (IJCNN)* (Padua: IEEE), 1–8.

Joshaghani, R., Ekstrand, M. D., Knijnenburg, B., and Mehrpouyan, H. (2018). "Do different groups have comparable privacy tradeoffs?" in *Workshop on Moving Beyond a "One-Size Fits All" Approach: Exploring Individual Differences in Privacy, in Conjunction with the ACM CHI Conference on Human Factors in Computing Systems, CHI 2018* (Montreal, QC).

Knijnenburg, B. P., and Kobsa, A. (2013). Making decisions about privacy: information disclosure in context-aware recommender systems. *ACM Trans. Interact. Intell. Syst.* 3, 1–23. doi: 10.1145/2499670

Kosinski, M., Stillwell, D., and Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proc. Nat. Acad. Sci. U.S.A.* 110, 5802–5805. doi: 10.1073/pnas.1218772110

Li, J., Yang, J.-J., Zhao, Y., Liu, B., Zhou, M., Bi, J., and Wang, Q. (2016). Enforcing differential privacy for shared collaborative filtering. *IEEE Access* 5, 35–49. doi: 10.1109/ACCESS.2016.2600258

Li, Z., Ding, B., Zhang, C., Li, N., and Zhou, J. (2021). Federated matrix factorization with privacy guarantee. *Proc. VLDB Endowment* 15, 900–913. doi: 10.14778/3503585.3503598

Liu, X., Liu, A., Zhang, X., Li, Z., Liu, G., Zhao, L., and Zhou, X. (2017). "When differential privacy meets randomized perturbation: a hybrid approach for privacy-preserving recommender system," in *International Conference on Database Systems for Advanced Applications* (Suzhou: Springer), 576–591.

Liu, Z., Yang, L., Fan, Z., Peng, H., and Yu, P. S. (2022). Federated social recommendation with graph neural network. *ACM Trans. Intell. Syst. Technol.* 13, 1–24. doi: 10.1145/3501815

Long, J., Chen, T., Nguyen, Q. V. H., and Yin, H. (2023). Decentralized collaborative learning framework for next POI recommendation. *ACM Trans. Inf. Syst.* 41, 1–25. doi: 10.1145/3555374

Ma, J., Zhang, Q., Lou, J., Ho, J. C., Xiong, L., and Jiang, X. (2019). "Privacy-preserving tensor factorization for collaborative health data analysis," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM '19* (New York, NY: Association for Computing Machinery), 1291–1300.

Majeed, A., and Lee, S. (2020). Anonymization techniques for privacy preserving data publishing: a comprehensive survey. *IEEE Access* 9, 8512–8545. doi: 10.1109/ACCESS.2020.3045700

McMahan, B., Moore, E., Ramage, D., Hampson, S., and Arcas, B. A. (2017). "Communication-efficient learning of deep networks from decentralized data," in *Artificial Intelligence and Statistics* (Fort Lauderdale, FL: PMLR), 1273–1282.

Mehdy, A. N., Ekstrand, M. D., Knijnenburg, B. P., and Mehrpouyan, H. (2021). "Privacy as a planned behavior: effects of situational factors on privacy perceptions and plans," in *Proc. of UMAP'21* (Utrecht).

Meng, X., Wang, S., Shu, K., Li, J., Chen, B., Liu, H., et al. (2018). "Personalized privacy-preserving social recommendation," in *Proceedings of the AAAI Conference on Artificial Intelligence* (New Orleans, LA).

Minto, L., Haller, M., Livshits, B., and Haddadi, H. (2021). "Stronger privacy for federated collaborative filtering with implicit feedback," in *Proceedings of the 15th ACM Conference on Recommender Systems* (Amsterdam), 342–350.

Müllner, P., Lex, E., Schedl, M., and Kowald, D. (2023). ReuseKNN: neighborhood reuse for differentially-private KNN-based recommendations. *ACM Trans. Intell. Syst. Technol.* 14, 1–29. doi: 10.1145/3608481

Neera, J., Chen, X., Aslam, N., Wang, K., and Shu, Z. (2023). Private and utility enhanced recommendations with local differential privacy and Gaussian mixture model. *IEEE Trans. Knowledge Data Eng.* 35, 4151–4163. doi: 10.1109/TKDE.2021.3126577

Ning, L., Chien, S., Song, S., Chen, M., Xue, Y., and Berlowitz, D. (2022). "EANA: reducing privacy risk on large-scale recommendation models," in *Proceedings of the 16th ACM Conference on Recommender Systems* (Seattle, WA), 399–407.

Ramakrishnan, N., Keller, B. J., Mirza, B. J., Grama, A. Y., and Karypis, G. (2001). When being weak is brave: privacy in recommender systems. *IEEE Internet Comput.* 5, 54–62. doi: 10.1109/4236.968832

Ran, X., Wang, Y., Zhang, L. Y., and Ma, J. (2022). A differentially private matrix factorization based on vector perturbation for recommender system. *Neurocomputing* 483, 32–41. doi: 10.1016/j.neucom.2022.01.079

Ren, J., Jiang, L., Peng, H., Lyu, L., Liu, Z., Chen, C., et al. (2022). "Cross-network social user embedding with hybrid differential privacy guarantees," in *Proceedings of the 31st ACM International Conference on Information &; Knowledge Management, CIKM '22* (New York, NY: Association for Computing Machinery), 1685–1695.

Shin, H., Kim, S., Shin, J., and Xiao, X. (2018). Privacy enhanced matrix factorization for recommendation with local differential privacy. *IEEE Trans. Knowledge Data Eng.* 30, 1770–1782. doi: 10.1109/TKDE.2018.2805356

Sun, J. A., Pentyala, S., Cock, M. D., and Farnadi, G. (2023). "Privacy-preserving fair item ranking," in *Advances in Information Retrieval*, eds J. Kamps, L. Goeuriot, F. Crestani, M. Maistro, H. Joho, B. Davis, C. Gurrin, U. Kruschwitz, and A. Caputo (Cham: Springer Nature), 188–203.

Wang, Y., Gao, M., Ran, X., Ma, J., and Zhang, L. Y. (2023). An improved matrix factorization with local differential privacy based on piecewise mechanism for recommendation systems. *Expert Syst. Appl.* 216, 119457. doi: 10.1016/j.eswa.2022.119457

Warner, S. L. (1965). Randomized response: a survey technique for eliminating evasive answer bias. *J. Am. Stat. Assoc.* 60, 63–69.

Weinsberg, U., Bhagat, S., Ioannidis, S., and Taft, N. (2012). "Blurme: inferring and obfuscating user gender based on ratings," in *Proceedings of the Sixth ACM Conference on Recommender Systems* (Dublin), 195–202.

Wu, C., Wu, F., Lyu, L., Huang, Y., and Xie, X. (2022). FedCTR: Federated native ad CTR prediction with cross-platform user behavior data. *ACM TIST* 13, 1–19. doi: 10.1145/3506715

Xin, X., Yang, J., Wang, H., Ma, J., Ren, P., Luo, H., et al. (2023). On the user behavior leakage from recommender system exposure. *ACM Trans. Inform. Syst.* 41, 1–25. doi: 10.1145/3568954

Yang, M., Zhu, T., Ma, L., Xiang, Y., and Zhou, W. (2017). "Privacy preserving collaborative filtering via the Johnson-Lindenstrauss transform," in *2017 IEEE Trustcom/BigDataSE/ICESS* (Sydney, NSW: IEEE), 417–424.

Yang, Z., Ge, Y., Su, C., Wang, D., Zhao, X., and Ying, Y. (2023). "Fairness-aware differentially private collaborative filtering," in *Companion Proceedings of the ACM Web Conference 2023, WWW '23 Companion* (Austin, TX: Association for Computing Machinery), 927–931.

Zemel, R., Wu, Y., Swersky, K., Pitassi, T., and Dwork, C. (2013). "Learning fair representations," in *Proc. of ICML'13* (Atlanta, GA: PMLR), 325–333.

Zhang, S., Yin, H., Chen, T., Huang, Z., Cui, L., and Zhang, X. (2021). "Graph embedding for recommendation against attribute inference attacks," in *Proceedings of the Web Conference 2021, WWW '21* (New York, NY: Association for Computing Machinery), 3002–3014.

Zhang, S., Yuan, W., and Yin, H. (2023). Comprehensive privacy analysis on federated recommender system against attribute inference attacks. *IEEE Trans. Knowledge Data Eng.* 1-13. doi: 10.1109/TKDE.2023.3295601

Zhao, Y., Feng, X., Li, J., and Liu, B. (2011). "Shared collaborative filtering," in *Proceedings of the Fifth ACM Conference on Recommender Systems* (Chicago, IL), 29–36.

Zhao, Y., Zhao, J., Yang, M., Wang, T., Wang, N., Lyu, L., et al. (2020). Local differential privacy-based federated learning for internet of things. *IEEE Internet Things J.* 8, 8836–8853. doi: 10.1109/JIOT.2020.3037194

Zhu, T., Li, G., Ren, Y., Zhou, W., and Xiong, P. (2013). "Differential privacy for neighborhood-based collaborative filtering," in *Proc. of IEEE/ACM ASONAM'13* (Niagara Falls, ON), 752–759.