The Impact of Differential Privacy on Recommendation Accuracy and Popularity Bias

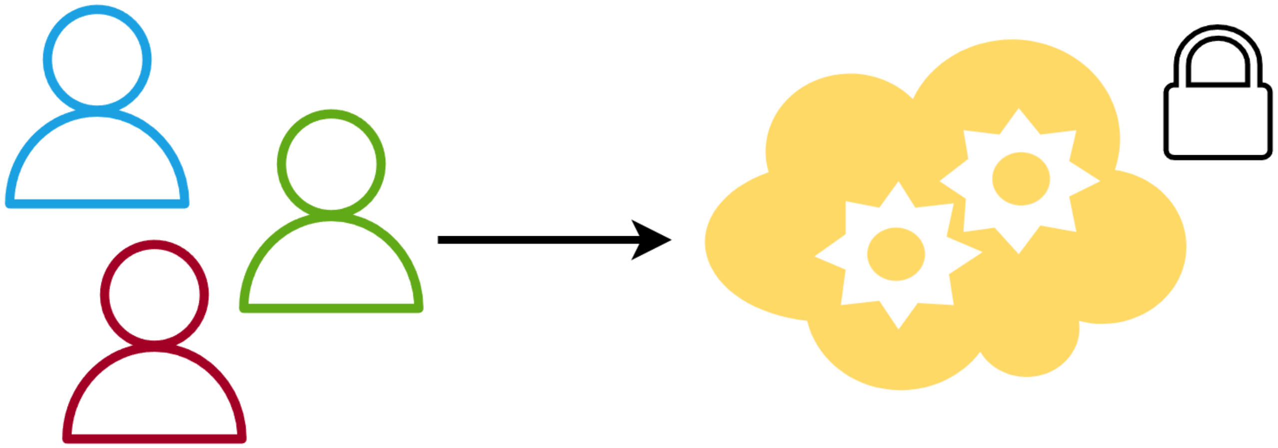
KNOW-CENTER GMBH

RESEARCH CENTER FOR TRUSTWORTHY AI AND DATA

Peter Müllner, Elisabeth Lex, Markus Schedl, Dominik Kowald
(pmuellner@know-center.at)

Motivation

- Recommender sytems utilize user data to generate recommendaions
- Recommendations can **leak user data**! [1, 4, 5]
- **Differential Privacy** to prohibit this leakage



The Problem with DP

- For DP, random noise is injected into the training process
- Level of noise/privacy regulated by privacy budget ϵ
- **How does this random noise impact recommendations?**
  - How many users are impacted?
  - How strong are they impacted?
  - How does privacy budget ϵ influence the accuracy?
  - How does DP impact popularity bias?

Our Approach

- DP mechanism[1] for binary implicit feedback data [2, 3]
- Train model *A* on data without DP
- Train model *B* on data with DP
- For each user, compare both models' recommendations

[1] Randomly substitutes positive feedback data with negative or missing feedback data with probability $1 / (e^{\epsilon} + 1)$

How many users are impacted?

- For each user, we compare recommendations with and without DP
- If at least one recommended item different → "impacted"
- **Nearly all users are impacted!**

ϵ	Model	*MovieLens 1M*	*LastFM User Groups*	*Amazon Grocery & Gourmet*
2	*ENMF*	99.85%	99.64%	**100.00%**
	LightGCN	99.86%	99.92%	99.99%
	MultVAE	99.93%	**100.00%**	**100.00%**
1	*ENMF*	99.99%	99.95%	**100.00%**
	LightGCN	**99.99%**	99.99%	**100.00%**
	MultVAE	**100.00%**	**100.00%**	**100.00%**
0.1	*ENMF*	**100.00%**	**100.00%**	**100.00%**
	LightGCN	**99.99%**	**100.00%**	**100.00%**
	MultVAE	**100.00%**	**100.00%**	**100.00%**

Table: No. of impacted users

How strong are those users impacted?

- Jaccard distance between recommendations with and without DP
- **Recommendations change a lot!**
- Worse for small ϵ values

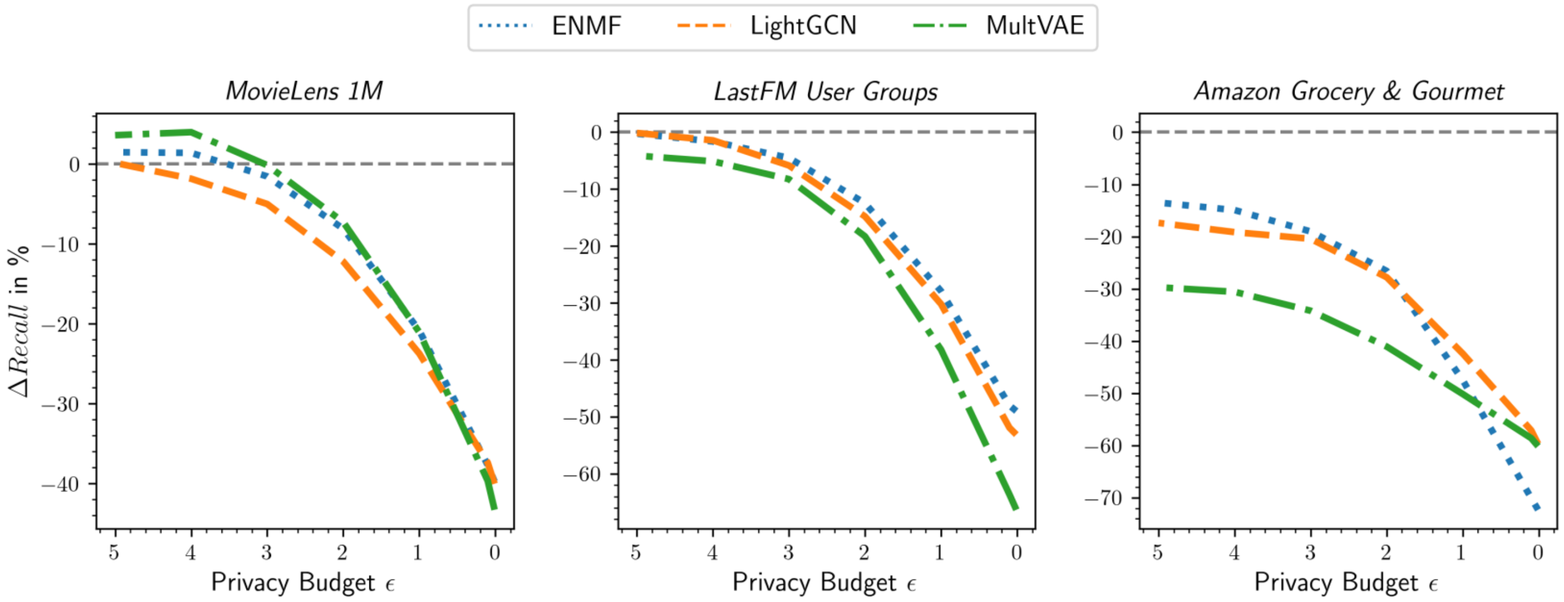
ϵ	Model	*MovieLens 1M*	*LastFM User Groups*	*Amazon Grocery & Gourmet*
2	*ENMF*	0.5974	0.5757	0.8620
	LightGCN	0.6252	0.6518	0.8132
	MultVAE	0.6828	0.7950	0.9447
1	*ENMF*	0.7006	0.6858	0.9253
	LightGCN	0.7352	0.7464	0.8775
	MultVAE	0.7592	0.8408	0.9567
0.1	*ENMF*	**0.8183**	**0.8058**	**0.9743**
	LightGCN	**0.8300**	**0.8490**	**0.9360**
	MultVAE	**0.8447**	**0.9250**	**0.9635**

Table: Average Jaccard distance

How does this affect accuracy and popularity bias?

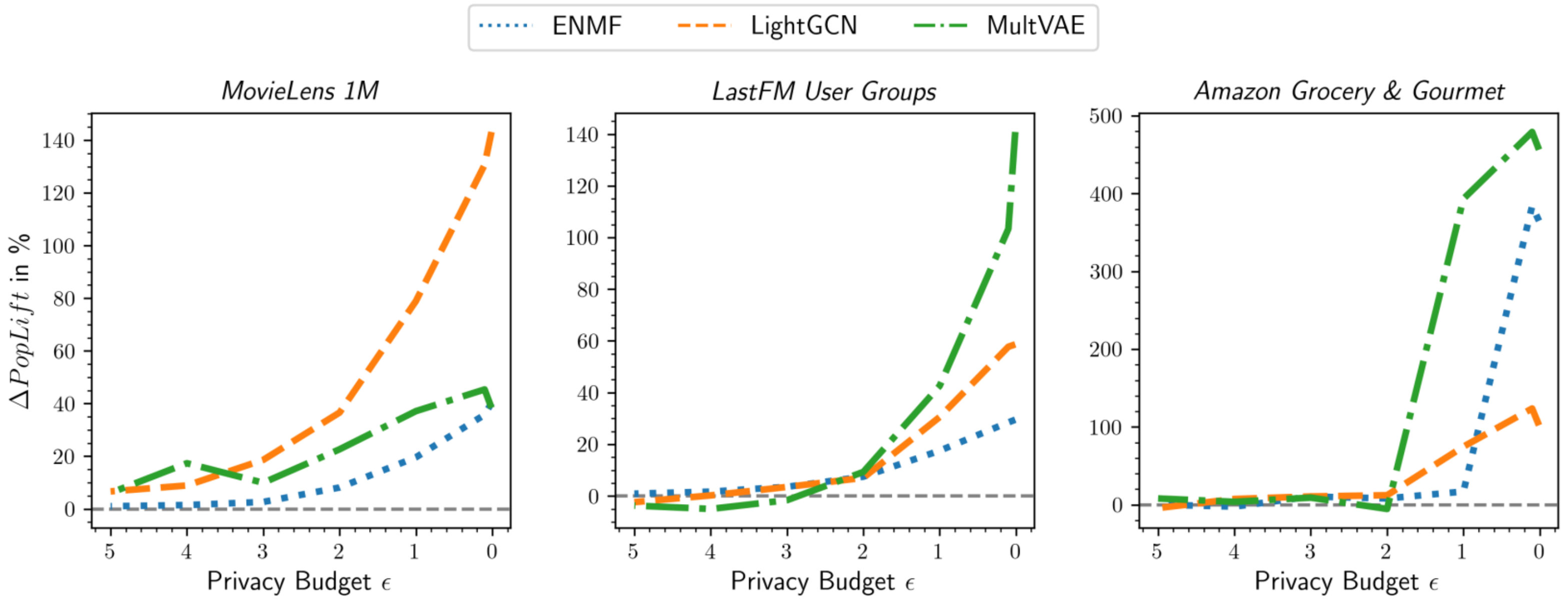
How does ϵ influence accuracy?

- Measure $\Delta Recall = (Recall_{DP} - Recall)/Recall$
- **Recall drops, especially for small ϵ values!**



How does DP impact popularity bias?

- Measure $\Delta PopLift = (PopLift_{DP} - PopLift)/PopLift$
- **PopLift increases, especially for small ϵ values!**



How does DP impact popularity bias?

- Users that prefer popular (U_{high}) and unpopular items (U_{high})
- Disparate Impact = $PopLift(U_{low}) - PopLift(U_{high})$
- **Popularity bias for U_{low} tends to be stronger than for U_{high}!**

ϵ	Method	*MovieLens 1M*	*LastFM User Groups*	*Amazon Grocery & Gourmet*
2	*ENMF*	+0.6941	+3.3158	+1.3597
	LightGCN	+0.1241	+1.8623	+1.3215
	MultVAE	+0.2595	-0.8629	+1.2161
1	*ENMF*	+0.8046	+3.9001	+1.4206
	LightGCN	+0.4368	+2.2851	+1.8303
	MultVAE	+0.3811	**-0.9744**	+4.0117
0.1	*ENMF*	**+0.8831**	**+4.2769**	**+3.7277**
	LightGCN	**+0.9352**	**+4.0773**	**+2.9563**
	MultVAE	**+0.5225**	-0.7113	**+5.2092**

Table: Disparate Impact

Conclusions and Future Work

- DP has substantial impact on entire user base
  - Different recommendations than without DP
  - Severe accuracy drop
  - Sharp increase of popularity bias
- Users are impacted differently w.r.t. popularity bias
- **Lower ϵ → more privacy, but stronger impact**
- Increase ϵ to decrease impact? → No privacy!

How to balance recommendation accuracy, popularity bias, and privacy?

(e.g., by applying popularity bias mitigation strategies)

Thank you!

Source code: github.com/pmuellner/ImpactOfDP/

Contact: pmuellner@know-center.at or pmuellner.github.io

KNOW-CENTER GMBH
Research Center for Data-Driven Business and Big Data Analytics

Sandgasse 36/4, 8010 Graz, Austria
Firmenbuchgericht Graz
FN 199 685 f
UID: ATU 50367703
www.know-center.at

References

1. J. A. Calandrino, A. Kilzer, A. Narayanan, E. W. Felten, and V. Shmatikov, *"you might also like:" privacy risks of collaborative filtering*, in 2011 IEEE Symposium on Security and Privacy (S&P), 2011, pp. 231–246.
2. C. Chen, J. Zhou, B. Wu, W. Fang, L. Wang, Y. Qi, and X. Zheng, *Practical privacy preserving poi recommendation*, ACM Transactions on Intelligent Systems and Technology (TIST), 11 (2020), pp. 1–20.
3. B. Ding, J. Kulkarni, and S. Yekhanin, *Collecting telemetry data privately*, in Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS), 2017, pp. 3574–3583.
4. H. Hashemi, W. Xiong, L. Ke, K. Maeng, M. Annavaram, G. E. Suh, and H.-H. S. Lee, *Data leakage via access patterns of sparse features in deep learning-based recommendation systems*, Workshop on Trustworthy and Socially Responsible Machine Learning (TSRML), in conjunction with the 36th Conference on Neural Information Processing Systems (NeurIPS), (2022).
5. X. Xin, J. Yang, H. Wang, J. Ma, P. Ren, H. Luo, X. Shi, Z. Chen, and Z. Ren, *On the user behavior leakage from recommender system exposure*, ACM Transactions on Information Systems (TOIS), 41 (2023), pp. 1–25.

Binary DP mechanism by Ding et al. [3]

- **Randomize positive feedback data used for model training!**
- For a user feedback $f_{u,i}$ for user u, item i, and training data D^+
- Now, randomly substitute positive feedback D^+ with negative/missing feedback D^-

$$Pr[f_{u,i} \in \mathcal{D}^+_{DP}] = \begin{cases} \frac{e^\epsilon}{e^\epsilon+1} & \text{if } f_{u,i} \in \mathcal{D}^+ \\ 1 - \frac{e^\epsilon}{e^\epsilon+1} & \text{if } f_{u,i} \in \mathcal{D}^- \end{cases}$$

- Privacy-parameter ϵ regulates how much leakage can be tolerated