BlurM(or)e: Revisiting Gender Obfuscation in the User-Item Matrix

Christopher Strucks
Radboud University
Netherlands
chr@strucks.org

Manel Slokom
TU Delft
Netherlands
m.slokom@tudelft.nl

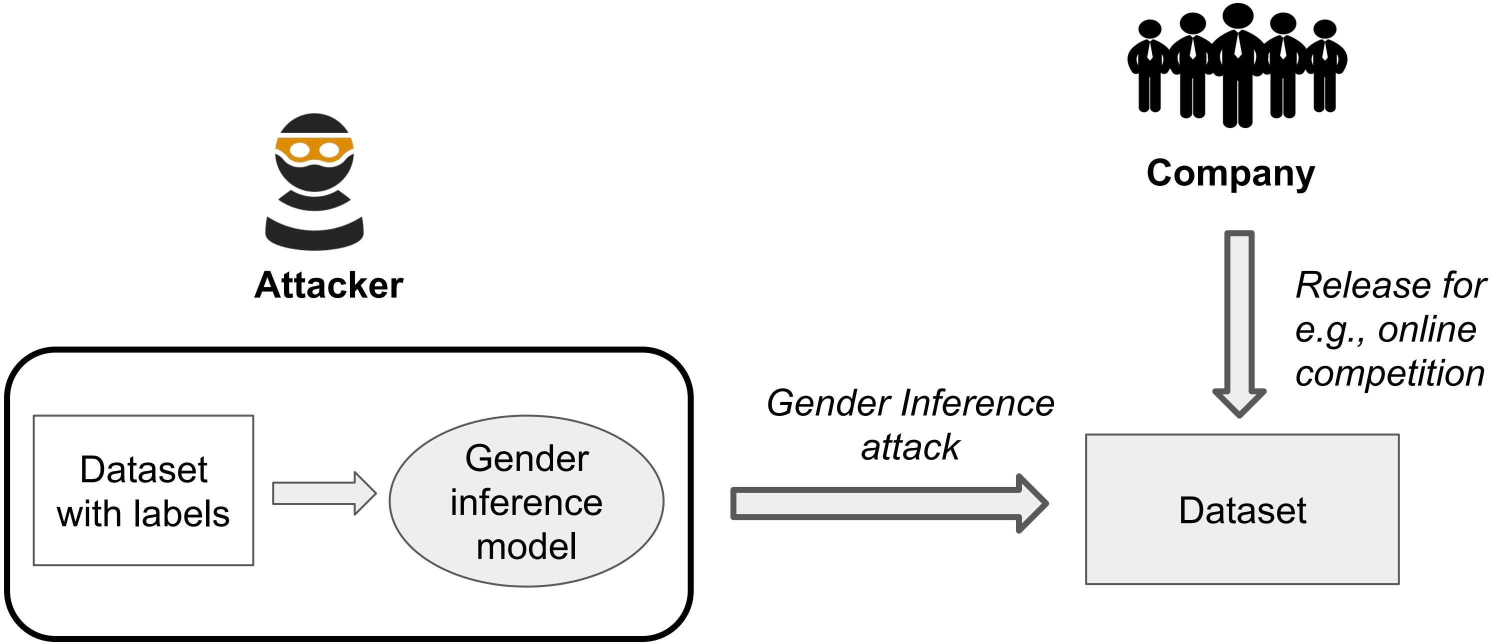
Martha Larson
Radboud University & TU Delft
Netherlands
m.larson@cs.ru.nl

User-Item Matrix: Privacy Issues

- Netflix Prize Challenge
  - User-Item matrix reveals identities (Narayanan and Shmatikov, 2008)

- Inference Attacks
  - Gender inference (**Weinsberg et al., 2012;** Liu, Qu, Chen and Mahmud, 2019)
  - Political orientation inference (Salamatian et al., 2013)

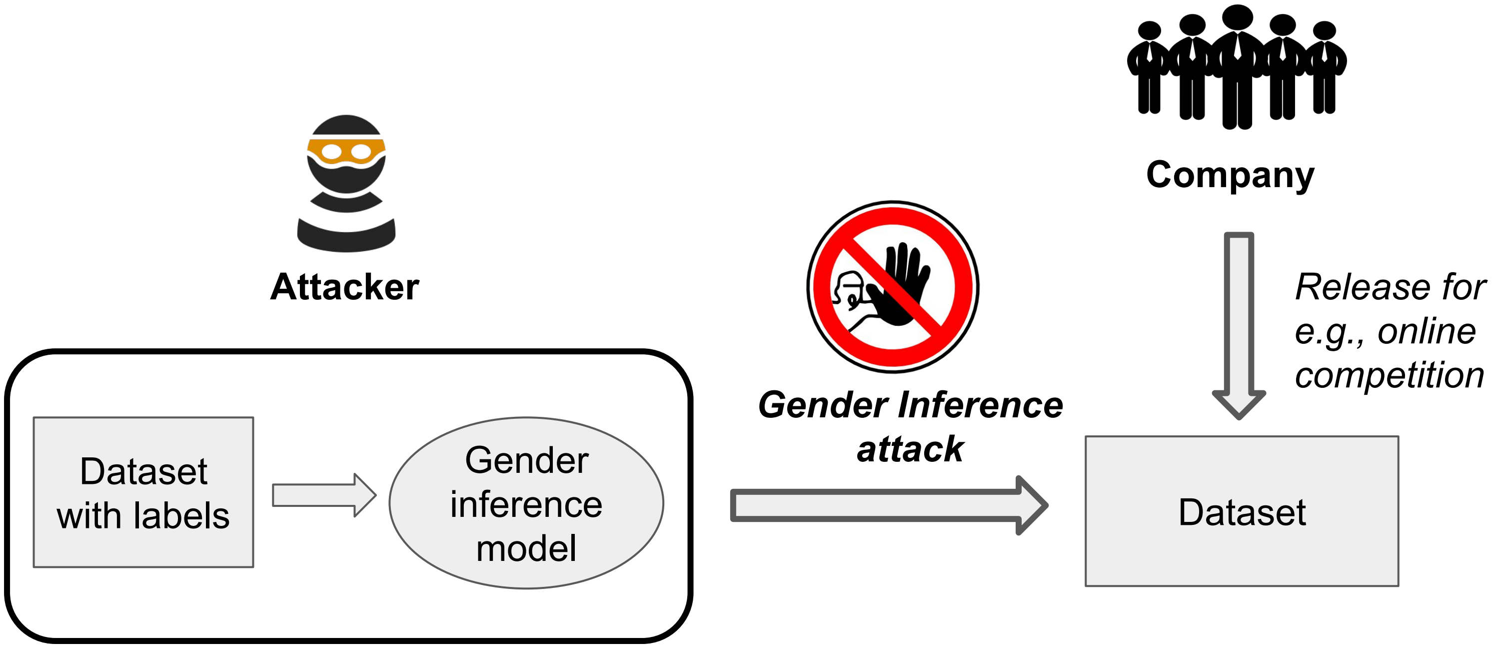
Gender Inference Attack



Research Goal

How can we protect users' demographic information from gender inference attacks while maintaining the accuracy of recommender systems?

Gender Inference Attack



Data Obfuscation & BlurMe

- Data Obfuscation
  - Hide implicit sensitive information by modifying the data

- BlurMe (Weinsberg et al., 2012):
  - add fake ratings from the *opposite* gender to hide gender information

BlurMe Data Obfuscation Algorithm

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1 (M)	5	0	5	0	3
User 2 (M)	4	0	3	0	5
User 3 (F)	2	5	0	4	1
User 4 (M)	5	0	4	0	5
User 5 (F)	0	5	1	5	3

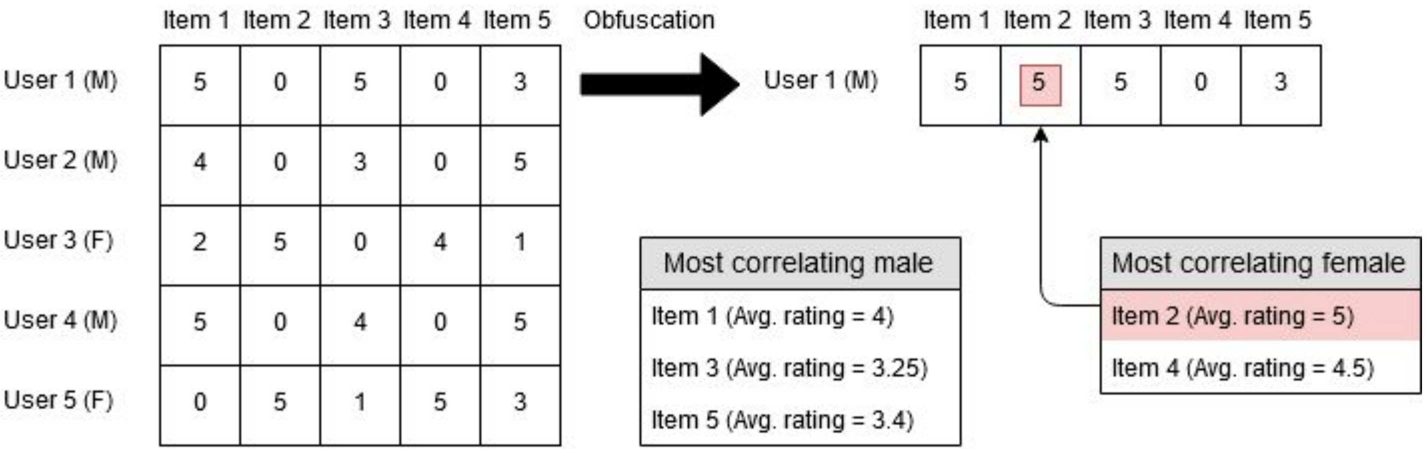
BlurMe Data Obfuscation Algorithm

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1 (M)	5	0	5	0	3
User 2 (M)	4	0	3	0	5
User 3 (F)	2	5	0	4	1
User 4 (M)	5	0	4	0	5
User 5 (F)	0	5	1	5	3

Most correlating male
Item 1 (Avg. rating = 4)
Item 3 (Avg. rating = 3.25)
Item 5 (Avg. rating = 3.4)

Most correlating female
Item 2 (Avg. rating = 5)
Item 4 (Avg. rating = 4.5)

BlurMe Data Obfuscation Algorithm



BlurMe Data Obfuscation Algorithm

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1 (M)	5	0	5	0	3
User 2 (M)	4	0	3	0	5
User 3 (F)	2	5	0	4	1
User 4 (M)	5	0	4	0	5
User 5 (F)	0	5	1	5	3

Obfuscation

	Item 1	Item 2	Item 3	Item 4	Item 5
	5	5	5	0	3
User 2 (M)	4	$\boxed{5}$	3	0	5

Most correlating male
Item 1 (Avg. rating = 4)
Item 3 (Avg. rating = 3.25)
Item 5 (Avg. rating = 3.4)

Most correlating female
Item 2 (Avg. rating = 5)
Item 4 (Avg. rating = 4.5)

BlurMe Data Obfuscation Algorithm

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1 (M)	5	0	5	0	3
User 2 (M)	4	0	3	0	5
User 3 (F)	2	5	0	4	1
User 4 (M)	5	0	4	0	5
User 5 (F)	0	5	1	5	3

Obfuscation

	Item 1	Item 2	Item 3	Item 4	Item 5
	5	5	5	0	3
	4	5	3	0	5
User 3 (F)	2	5	3	4	1

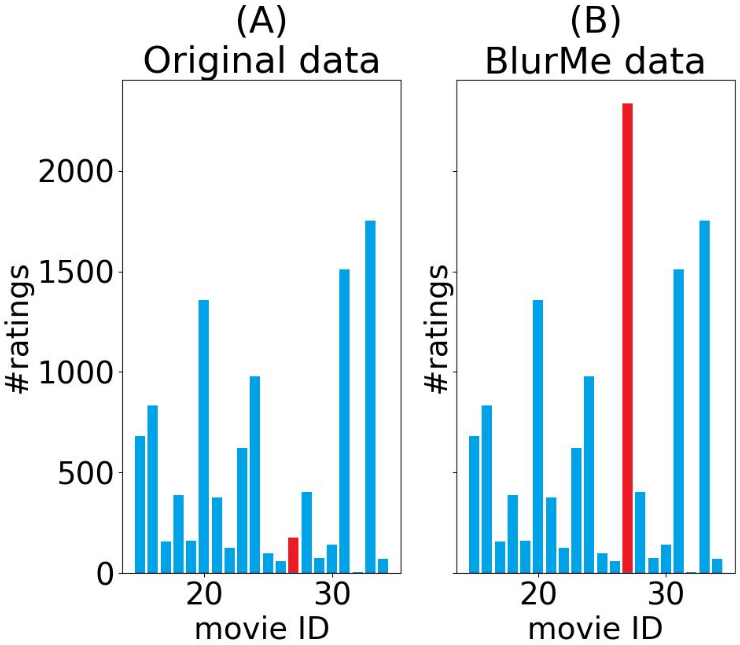
Most correlating male
Item 1 (Avg. rating = 4)
Item 3 (Avg. rating = 3.25)
Item 5 (Avg. rating = 3.4)

Most correlating female
Item 2 (Avg. rating = 5)
Item 4 (Avg. rating = 4.5)

Gender Obfuscation - BlurMe

	Classifier	Strategy	Accuracy with extra ratings			
			0%	1%	5%	10%
Flixster	Logistic Regression	Random	76.5	65.8	46.2	28.5
		Sampled	76.5	60.8	36.6	19.6
		Greedy	76.5	15	1.7	0.1
	Multinomial	Random	71.5	69.3	67	63.5
		Sampled	71.5	68.6	66	61.1
		Greedy	71.5	62	54.3	42.1
Movielens	Logistic Regression	Random	80.2	77.6	71.5	61.1
		Sampled	80.2	75.2	58.6	35.5
		Greedy	80.2	57.7	17.3	2.5
	Multinomial	Random	76.4	75.1	72.9	70.1
		Sampled	76.4	74.9	72.3	68.4
		Greedy	76.4	72.3	66.6	60.4

Table 4: Accuracy of gender inference for different strategies, when rating assignment is average movie rating



BlurM(or)e

Similar to BlurMe, but…

- ➢ Limit the number of extra ratings per movie
- ➢ Randomly remove ratings from users with 200 or more ratings

BlurM(or)e Algorithm

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1 (M)	5	0	5	0	3
User 2 (M)	4	0	3	0	5
User 3 (F)	2	5	0	4	1
User 4 (M)	5	0	4	0	5
User 5 (F)	0	0	1	4	3
Initial Rating Count:	4	1	4	2	5

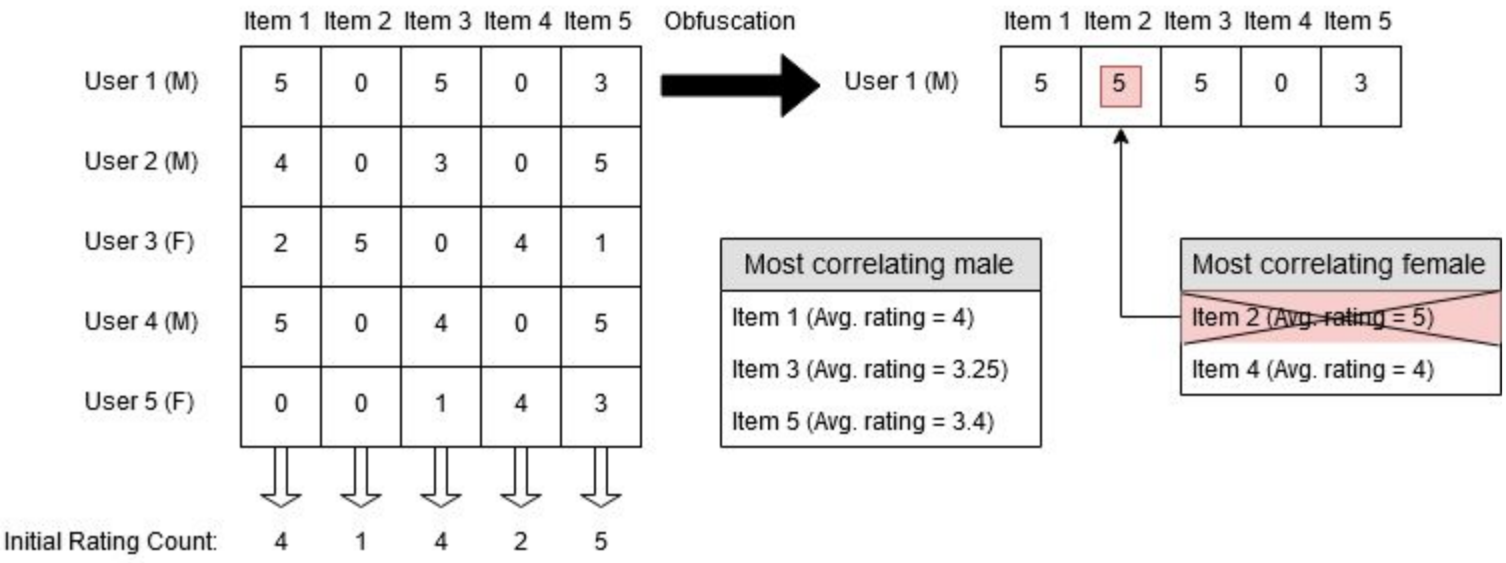
Obfuscation

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1 (M)	5	5	5	0	3

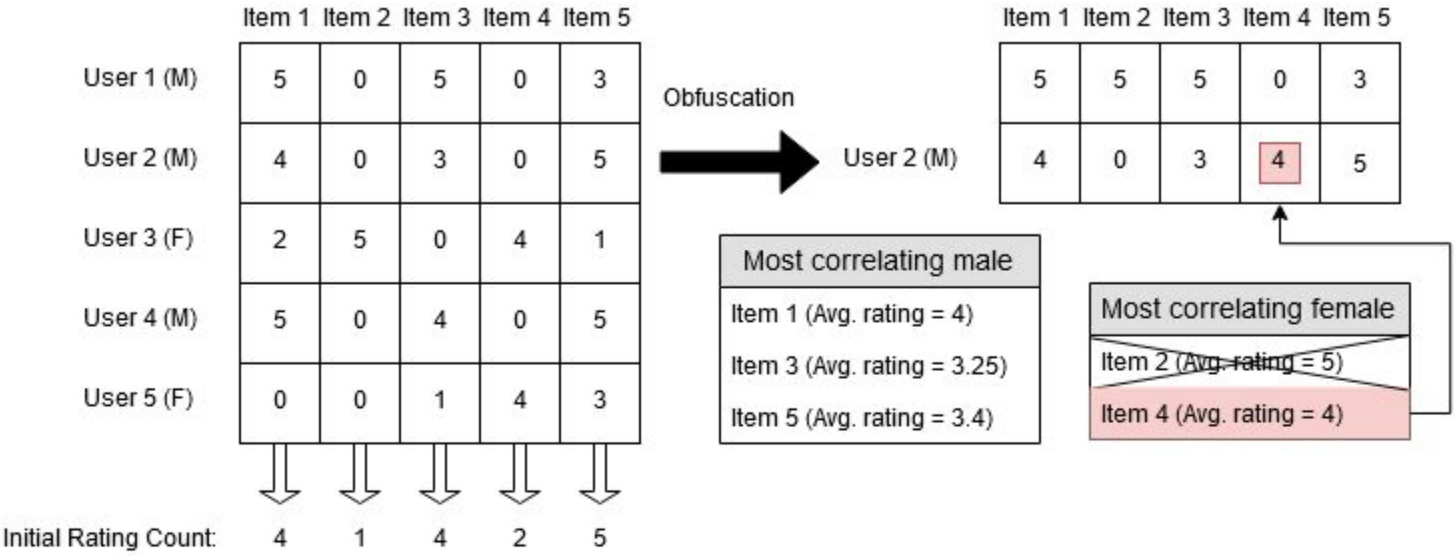
Most correlating male
Item 1 (Avg. rating = 4)
Item 3 (Avg. rating = 3.25)
Item 5 (Avg. rating = 3.4)

Most correlating female
Item 2 (Avg. rating = 5)
Item 4 (Avg. rating = 4)

BlurM(or)e Algorithm



BlurM(or)e Algorithm



BlurM(or)e Algorithm

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1 (M)	5	0	5	0	3
User 2 (M)	4	0	3	0	5
User 3 (F)	2	5	0	4	1
User 4 (M)	5	0	4	0	5
User 5 (F)	0	0	1	4	3
Initial Rating Count:	4	1	4	2	5

Obfuscation

Item 1	Item 2	Item 3	Item 4	Item 5
5	5	5	0	3
4	0	3	4	5
2	5	3	4	1
2	0	4	4	5
4	0	1	4	3

Most correlating male

- Item 1 (Avg. rating = 4)
- Item 3 (Avg. rating = 3.25)
- Item 5 (Avg. rating = 3.4)

Most correlating female

- ~~Item 2 (Avg. rating = 5)~~
- ~~Item 4 (Avg. rating = 4)~~

BlurM(or)e Algorithm

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1 (M)	5	0	5	0	3
User 2 (M)	4	0	3	0	5
User 3 (F)	2	5	0	4	1
User 4 (M)	5	0	4	0	5
User 5 (F)	0	0	1	4	3
Initial Rating Count:	4	1	4	2	5

Obfuscation →

	Item 1	Item 2	Item 3	Item 4	Item 5
	5	5	$\boxed{0}$	0	3
	$\boxed{0}$	0	3	4	5
	2	5	3	$\boxed{0}$	1
	2	0	4	4	$\boxed{0}$
	4	0	$\boxed{0}$	4	3

Most correlating male
Item 1 (Avg. rating = 4)
Item 3 (Avg. rating = 3.25)
Item 5 (Avg. rating = 3.4)

Most correlating female
~~Item 2 (Avg. rating = 5)~~
~~Item 4 (Avg. rating = 4)~~

BlurMe vs. BlurM(or)e

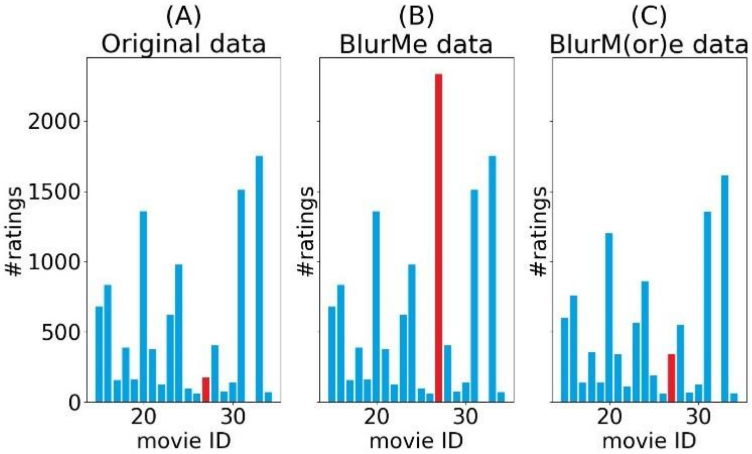


Figure 6: #ratings per movie for the movies 15 to 35. The red bar indicates an example of obvious data obfuscation after BlurMe is applied. The BlurMe data was created with the greedy strategy and with 10% extra ratings. The BlurM(or)e data contains also 10% extra ratings.

		Extra ratings			
Dataset	Classifier	0%	1%	5%	10%
BlurMe	Logistic Regression	0.76±0.02	0.54±0.03	0.15±0.03	0.02±0.01
BlurM(or)e	Logistic Regression	0.76±0.02	0.64±0.03	0.36±0.07	0.19±0.07
Original	Random Classifier	0.50	0.50	0.50	0.50

Table 9: Gender inference results measured in accuracy on BlurMe (reproduction) and BlurM(or)e. The datasets BlurMe and BlurM(or)e are created using the greedy strategy and the MovieLens dataset.

	Extra ratings			
Obfuscation	0%	1%	5%	10%
No	0.8766	—	—	—
BlurMe	0.8766	0.8686	0.8553	0.8385
BlurM(or)e	0.8766	0.8711	0.8640	0.8468

Table 11: The RMSE performance with Matrix Factorization on the original MovieLens dataset, BlurMe data and on BlurM(or)e data.

Discussion & Future Work

- What did we do?
  - Novel approach for obfuscating gender in a user-item matrix,
  - Maintain data quality
  - Difficult to detect
  - Step toward data sharing without privacy concerns

- What can be done next?
  - Other adding or removal strategies
  - Obfuscate other attributes

References

[1] Arvind Narayanan and Vitaly Shmatikov. Robust de-anonymization of large sparse datasets. In 2008 ieee symposium on security and privacy, pages 111–125. IEEE, 2008

[2] Udi Weinsberg, Smriti Bhagat, Stratis Ioannidis, and Nina Taft. 2012. BlurMe: Inferring and Obfuscating User Gender Based on Ratings. In Proceedings of the 2012 ACM Conference on Recommender Systems (RecSys '12). ACM, 195–202

[3] Yongsheng Liu, Hong Qu, Wenyu Chen, and SM Hasan Mahmud. 2019. An Efficient Deep Learning Model to Infer User Demographic Information From Ratings. IEEE Access 7 (2019), 53125–53135

[4] Salman Salamatian, Amy Zhang, Flavio du Pin Calmon, Sandilya Bhamidipati, Nadia Fawaz, Branislav Kveton, Pedro Oliveira, and Nina Taft. How to hide the elephant-or the donkey-in the room: Practical privacy against statistical inference for large data. In 2013 IEEE Global Conference on Signal and Information Processing, pages 269–272. IEEE, 2013.

Thank You!