18734: Foundations of Privacy

Differentially Private Recommendation Systems

Giulia Fanti

Slides by Anupam Datta

Fall 2019

Administrative

- Proposals graded on Gradescope
  - Please look at comments
  - Even on questions where you got full points

- Piazza
  - StackOverflow guidelines
  - NEW: 5 students with the most answers endorsed by an instructor get a bonus point on final grade

- Recitation this week (Sruti): Friday @ 12.30 ET/9.30 am PT
  - Differential privacy practice

- James Office Hours
  - Wed (today) at 4pm ET/1 pm ET

- Giulia Office Hours
  - Friday at 5.30 pm ET/2.30 pm PT

Canvas Quiz

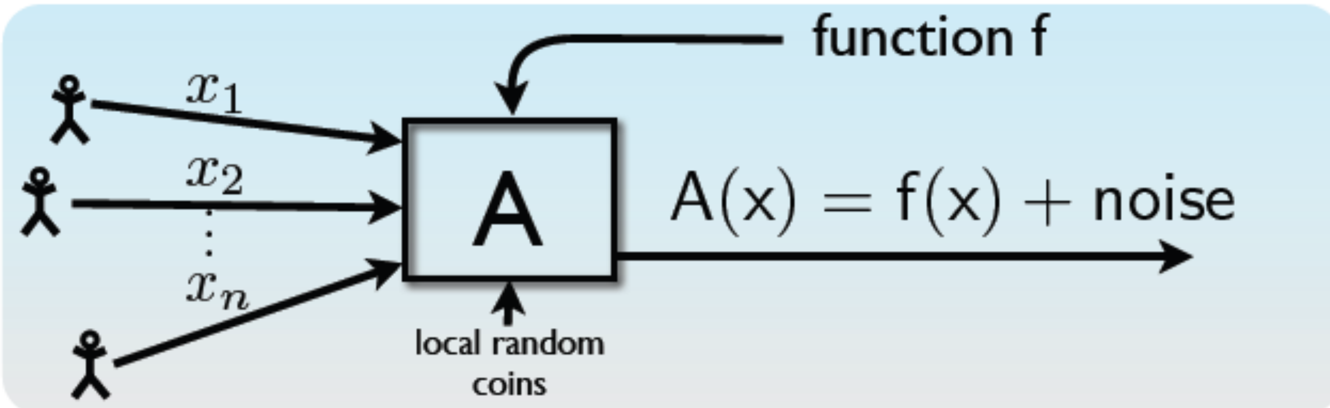
- 10 minutes

Last time: Differential Privacy

Randomized sanitization function κ has ε-differential privacy if for all data sets D_1 and D_2 differing by at most one element and all subsets S of the range of κ,

$$\Pr[\kappa(D_1) \in S] \leq e^{\varepsilon} \Pr[\kappa(D_2) \in S]$$

Laplace Mechanism



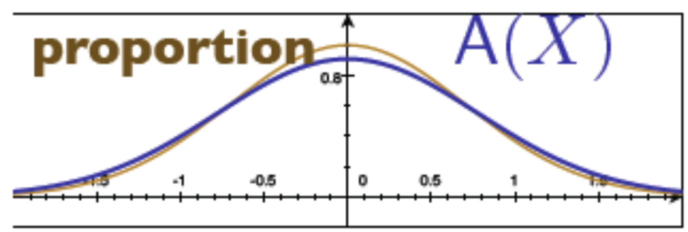
- **Global Sensitivity:** $GS_f = \max_{\text{neighbors } x,x'} \|f(x) - f(x')\|_1$
  - Example: $GS_{\text{proportion}} = \frac{1}{n}$

Theorem: If $A(x) = f(x) + \text{Lap}\left(\frac{GS_f}{\epsilon}\right)$, then A is ϵ-differentially private.

Slide: Adam Smith

Example: Noise Addition

- Example: proportion of diabetics
  - $GS_{\text{proportion}} = \frac{1}{n}$
  - Release $A(x) = \text{proportion} \pm \dfrac{1}{\epsilon n}$
- Is this **a lot**?
  - If x is a random sample from a large underlying population, then **sampling noise** $\approx \dfrac{1}{\sqrt{n}}$
  - A(x) "as good as" real proportion



Using Global Sensitivity

- Many natural functions have low global sensitivity
  - Histogram, covariance matrix, Lipschitz optimization problems

- Different mechanisms have different privacy-utility tradeoffs
  - Laplace noise can add more noise than necessary

Composition Theorem

Repeated querying degrades privacy; degradation is quantifiable

- **Theorem.** If A_1 is ε_1-differentially private and A_2 is ε_2-differentially private and they use independent random coins then the composition of A_1 and A_2 is $(\varepsilon_1+\varepsilon_2)$-differentially private
- Work with your neighbor to prove this.

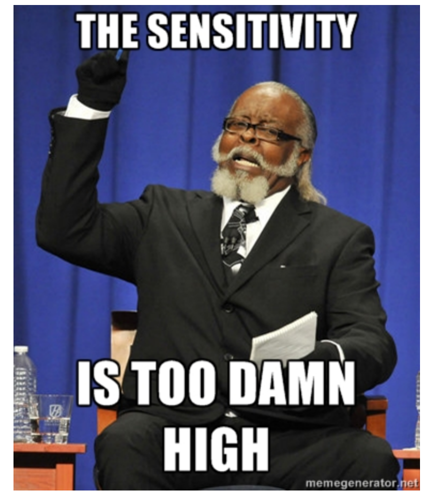
Applications

- Netflix data set [McSherry, Mironov 2009; MSR]
  - Accuracy of differentially private recommendations (wrt one movie rating) comparable to baseline set by Netflix
- Network trace data sets [McSherry, Mahajan 2010; MSR]

		High accuracy
Packet-level analyses		
Packet size and port dist.	(§5.1.1)	strong privacy
Worm fingerprinting [27]	(§5.1.2)	weak privacy
Flow-level analyses		
Common flow properties [30]	(§5.2.1)	strong privacy
Stepping stone detection [33]	(§5.2.2)	medium privacy
Graph-level analyses		
Anomaly detection [13]	(§5.3.1)	strong privacy
Passive topology mapping [9]	(§5.3.2)	weak privacy

Challenge: High Sensitivity

- Approach: Add noise proportional to sensitivity to preserve ε-differential privacy



- Improvements:
  - Smooth sensitivity [Nissim, Raskhodnikova, Smith 2007; BGU-PSU]
  - Restricted sensitivity [Blocki, Blum, Datta, Sheffet 2013; CMU]

Differential Privacy: Summary

- An approach to releasing privacy-preserving statistics
- A rigorous privacy guarantee
  - Significant activity in theoretical CS community
- Several applications to real data sets
  - Recommendation systems, network trace data,..
- Some challenges
  - High sensitivity -> high noise
  - Repeated querying

So far, you have seen:

- Definition of differential privacy
- How to make a scalar query differentially private
  - Laplacian noise

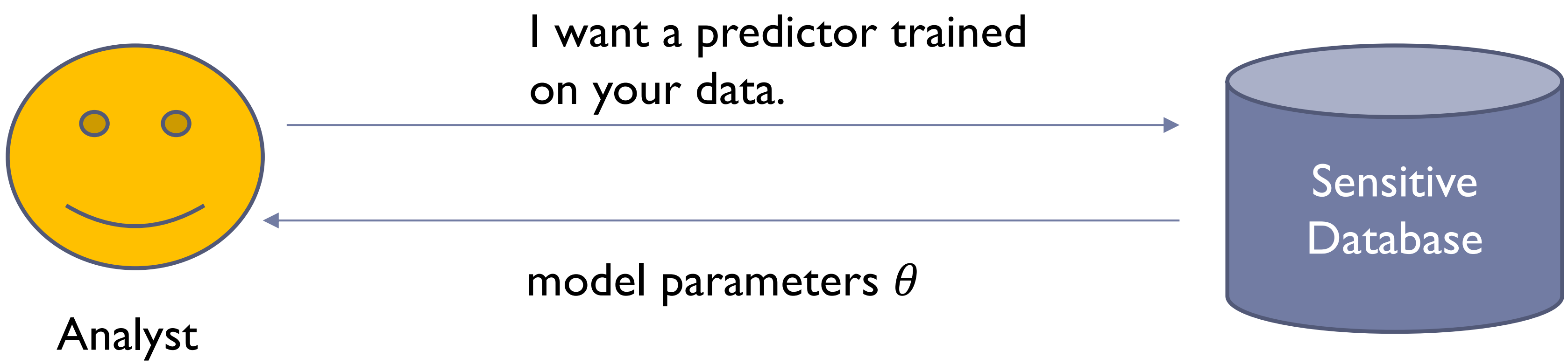
- What about training machine learning models on sensitive data?
  - How do we use differential privacy?

Differentially Private Recommender Systems: Building Privacy into the Netflix Prize Contenders

Frank McSherry and Ilya Mironov

KDD 2009

Machine learning model



How can we make this differentially private?

Netflix $1,000,000 Prize Competition

User/Movie	….	300	The Notebook	….
…	…	…	…	…
John		4	Unrated	
Mary		Unrated	Unrated	
Sue		2	5	
Joe		5	1	
…	…	…	…	…

Queries: On a scale of 1 to 5 how would John rate "The Notebook" if he watched it?

Netflix Prize Competition

User/Movie	….	13,537	13,538	….
…	…	…	…	…
258,964		(4, 10/11/2005)	Unrated	
258,965		Unrated	Unrated	
258,966		(2, 6/16/2005)	(5, 6/18/2005)	
258,967		(5, 9/15/2005)	(1,4/28/2005)	
…	…	…	…	…

Note: N x M table is very sparse (M = 17,770 movies, N = 500,000 users)

To Protect Privacy:

- Each user was randomly assigned to a globally unique ID
- Only 1/10 of the ratings were published
- The ratings that were published were perturbed a little bit

Root Mean Square Error

$$RMSE(P) = \sqrt{\frac{\sum_{i=1}^{k} (p_i - a_i)^2}{k}}$$

$p_i \in [1,5]$ - predicted ratings

$a_i \in [1,5]$ - actual ratings

Netflix Prize Competition

Goal: Make accurate predictions as measured by Root Mean Squared Error (RMSE)

$$RMSE(\vec{P}) = \sqrt{\frac{\sum_{i=1}^{k}(p_i - a_i)^2}{k}}$$

$p_i \in [1,5]$ - predicted ratings

$a_i \in [1,5]$ - actual ratings

Algorithm	RMSE
BellKor's Pragmatic Chaos	0.8567 < 0.8572
Challenge: 10% Improvement	0.8572
Netflix's Cinematch (Baseline)	0.9525

Leaderboard

Showing Test Score. Click here to show quiz score

Display top 20 leaders.

Rank	Team Name	Best Test Score	% Improvement	Best Submit Time
Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos				
1	BellKor's Pragmatic Chaos	0.8567	10.06	2009-07-26 18:18:28
2	The Ensemble	0.8567	10.06	2009-07-26 18:38:22
3	Grand Prize Team	0.8582	9.90	2009-07-10 21:24:40
4	Opera Solutions and Vandelay United	0.8588	9.84	2009-07-10 01:12:31
5	Vandelay Industries !	0.8591	9.81	2009-07-10 00:32:20
6	PragmaticTheory	0.8594	9.77	2009-06-24 12:06:56
7	BellKor in BigChaos	0.8601	9.70	2009-05-13 08:14:09
8	Dace_	0.8612	9.59	2009-07-24 17:18:43
9	Feeds2	0.8622	9.48	2009-07-12 13:11:51
10	BigChaos	0.8623	9.47	2009-04-07 12:33:59
11	Opera Solutions	0.8623	9.47	2009-07-24 00:34:07
12	BellKor	0.8624	9.46	2009-07-26 17:19:11
Progress Prize 2008 - RMSE = 0.8627 - Winning Team: BellKor in BigChaos				
13	xiangliang	0.8642	9.27	2009-07-15 14:53:22
14	Gravity	0.8643	9.26	2009-04-22 18:31:32
15	Ces	0.8651	9.18	2009-06-21 19:24:53
16	Invisible Ideas	0.8653	9.15	2009-07-15 15:53:04
17	Just a guy in a garage	0.8662	9.06	2009-05-24 10:02:54
18	J Dennis Su	0.8666	9.02	2009-03-07 17:16:17
19	Craig Carmichael	0.8666	9.02	2009-07-25 16:00:54
20	acmehill	0.8668	9.00	2009-03-21 16:20:50
Progress Prize 2007 - RMSE = 0.8723 - Winning Team: KorBell				
Cinematch score - RMSE = 0.9525				

Netflix Privacy Woes

3/12/2010 @ 12:35PM | 2,590 views

Netflix Settles Privacy Lawsuit, Cancels Prize Sequel

Taylor Buley , Contributor

\+ Comment Now   + Follow Comments

On Friday, Netflix announced on its corporate blog that it has settled a lawsuit related to its Netflix Prize, a $1 million contest that challenged machine learning experts to use Netflix's data to produce better recommendations than the movie giant could serve up themselves.

The lawsuit called attention to academic research that suggests that Netflix indirectly exposed the movie preferences of its users by publishing anonymized customer data. In the suit, plaintiff Paul Navarro and others sought an injunction preventing Netflix from going through the so-called "Netflix Prize II," a follow-up challenge that Netflix promised would offer up even more personal data such as genders and zipcodes.

:)

-data,
es are
edge.
mous
. We
easily

Outline

- *Approximate Differential Privacy*
- Prediction Algorithms
- Privacy Preserving Prediction Algorithms
- Remaining Issues

Privacy in Recommender Systems

- Netflix might base its recommendation to me on both:
  - My own rating history
  - The rating history of other users

- Goal: not leak other users' ratings to me

- Basic recommendation systems leak other users' information
  - Calandrino, et al. Don't review that book: Privacy risks of collaborative filtering, 2009.

Recall Differential Privacy [Dwork et al 2006]

Dual-sided restatement: for all data sets D_1 and D_2 differing by at most one element and all outputs s in the range of κ,

$$e^{-\epsilon} \leq \frac{\Pr[\kappa(D_1) = s]}{\Pr[\kappa(D_2) = s]} \leq e^{\epsilon}$$

and more generally, for all subsets S of the range of κ

$$e^{-\epsilon} \leq \frac{\Pr[\kappa(D_1) \in S]}{\Pr[\kappa(D_2) \in S]} \leq e^{\epsilon}$$

Review: Laplacian Mechanism

$\mathrm{K}(D) =$ (GS_f)

Thm: K

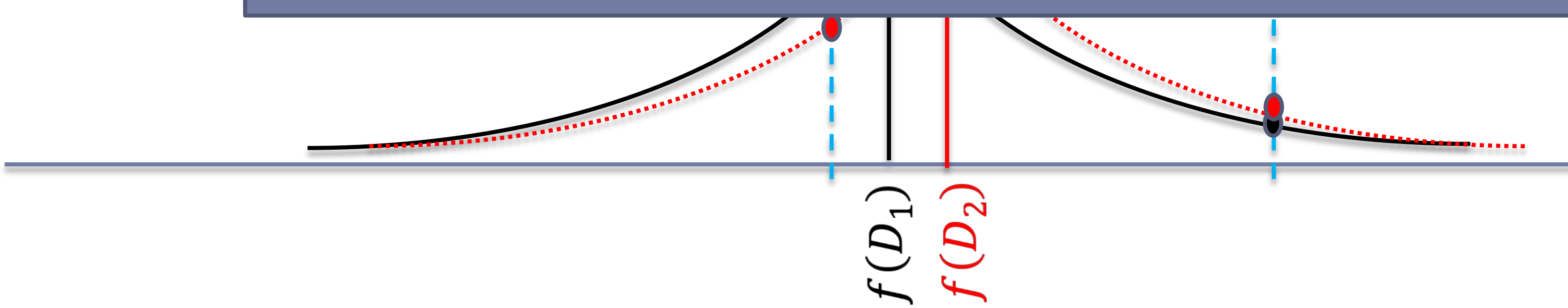
Probability Density Function

$\frac{|x|}{\sigma}\Big)$

Picture P

$e^{-\varepsilon} \leq$

Question: The Gaussian (Normal) distribution is nicer because it is more tightly concentrated around its mean. Can we use that distribution instead?



Gaussian Mechanism

$$\kappa(D) = f(D) + \mathrm{N}\left(\frac{GS_f}{\varepsilon}\right)$$

Thm? K is ε differentially private?

Probability Density Function

$$\mathrm{N}(x, 0, \sigma) \propto \frac{1}{2\sqrt{2\ldots}} \exp\left(\frac{-x^2}{2\sigma^2}\right)$$

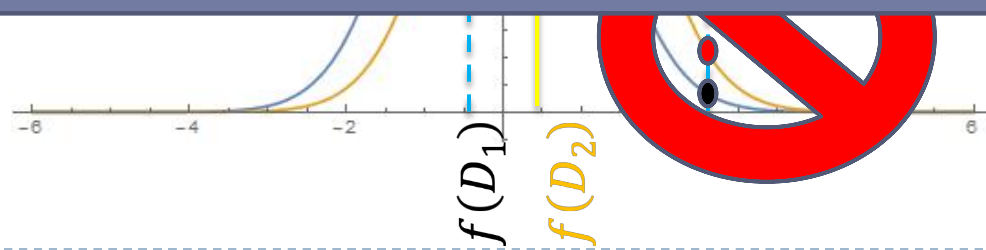
Pic

Problem: The ratio can be huge at the tails!

$e^{-\varepsilon} \le Ratio = \frac{\ldots}{Red} \le e^{\varepsilon}$

Red

But these events are very unlikely…



Approximate Differential Privacy

Randomized sanitization function κ has *(ε, δ)-differential privacy* if for all data sets D_1 and D_2 differing by at most one element and all subsets S of the range of κ,

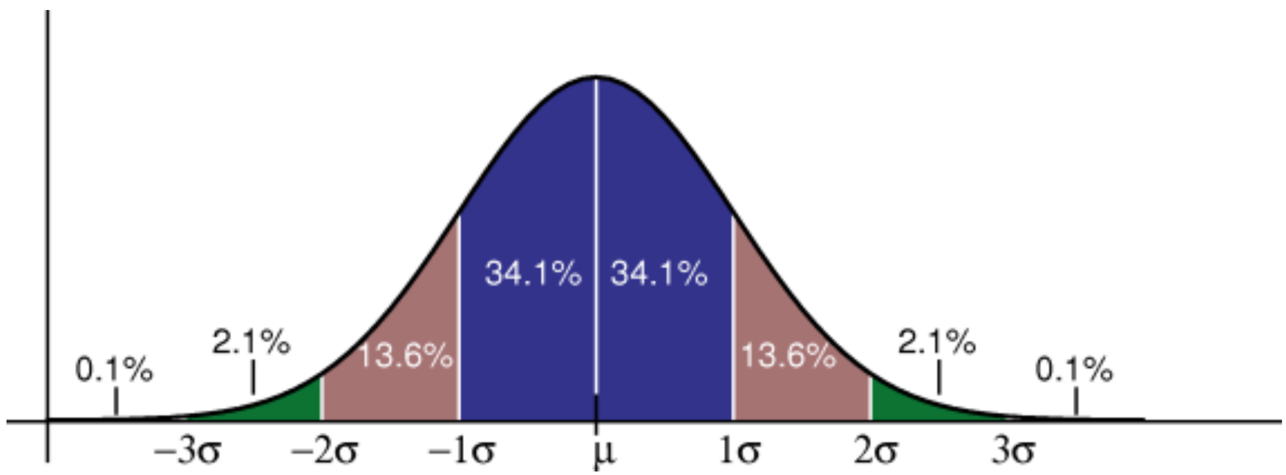
$$\Pr[\kappa(D_1) \in S] \leq e^{\epsilon} \Pr[\kappa(D_2) \in S] + \delta$$

Gaussian Mechanism

$K(D) = f(D) + \mathrm{N}(\sigma^2)$

Thm K is ($\boldsymbol{\varepsilon}$, δ)-differentially private as long as $\sigma \geq \frac{\sqrt{2\ln(2/\delta)}}{\varepsilon} \times GS_f$

Idea Use δ to exclude the tails of the gaussian distribution



Multivariate Gaussian Mechanism

Suppose that f outputs a length d vector instead of a number

$\mathrm{K}(D) = f(D) + \mathrm{N}(\sigma^2)^d$

Thm K is ($\boldsymbol{\varepsilon}$, δ)-differentially private as long as

$$\sigma \geq \frac{\sqrt{2\ln(2/\delta)}}{\varepsilon} \times \max_{D1 \approx D2} \|f(D_1) - f(D_2)\|_2$$

Remark: Similar results would hold with the Laplacian Mechanism, but we would need to add noise proportional to the larger L_1 norm

Approximate Differential Privacy

- Key Difference
  - Approximate Differential Privacy does NOT require that:

$$\text{Support}(\kappa(D_1)) = \text{Support}(\kappa(D_2))$$

- The privacy guarantees made by *(ε,δ)-differential privacy* are not as strong as ε-differential privacy, but less noise is required to achieve *(ε,δ)-differential privacy*.

Approximate Differential Privacy

- Key similarity
  - Composition still holds!

  - If M_1 and M_2 satisfy (ϵ_1, δ_1) and (ϵ_2, δ_2) differential privacy, respectively, then their linear composition satisfies $(\epsilon_1 + \epsilon_2, \delta_1 + \delta_2)$ differential privacy.

Differential Privacy for Netflix Queries

- What level of granularity to consider? What does it mean for databases D_1 and D_2 to differ on at most one element?
  - One user (column) is present in D_1 but not in D_2
  - One rating (cell) is present in D_1 but not in D_2

- Issue 1: Given a query "how would user i rate movie j?" Consider: $\kappa(D - u[i])$ - how can it possibly be accurate?

- Issue 2: If the definition of differing in at most one element is taken over cells, then what privacy guarantees are made for a user with many data points?

Netflix Predictions – High Level

- Q(i,j) – "How would user i rate movie j?"

- Predicted rating may typically depend on
  - Global average rating over all movies and all users
  - Average movie rating of user i
  - Average rating of movie j
  - Ratings user i gave to *similar* movies
  - Ratings *similar users* gave to movie j

- Sensitivity may be small for many of these queries

Personal Rating Scale

- For Alice a rating of 3 might mean the movie was really terrible.
- For Bob the same rating might mean that the movie was excellent.
- How do we tell the difference?

Check the following:

$$r_{im} - \bar{r}_i > 0?$$

How do we tell if two users are similar?

Pearson's Correlation is one metric for similarity of users i and j

- Consider all movies rated by both users
- Negative value whenever i likes a movie that j dislikes
- Positive value whenever i and j agree

$$S(i,j) = \sum_{m \in L_i \cap L_j} (r_{im} - \bar{r}_i)(r_{jm} - \bar{r}_j)$$

We can use similar metrics to measure the similarity between two movies.

Netflix Predictions Example

- Collaborative Filtering
  - Find the k-nearest neighbors of user i who have rated movie j by Pearson's Correlation:

$S(i, \ell)$ similarity of users i and ℓ

$N_i(k, j) = \{u_1, u_2, \dots, u_k\}$ k most similar users who rated movie j

  - Predicted Rating

$$p_{ij} = \bar{r}_i + \frac{1}{k} \sum_{u \in N_i(k,j)} (r_{uj} - \bar{r}_u)$$

Netflix Prediction Sensitivity Example

$$p_{ij} = \bar{r}_i + \frac{1}{k} \sum_{u \in N_i(k,j)} (r_{uj} - \bar{r}_u)$$

- Pretend the query Q(i,j) included user i's rating history
- At most one of the neighbors ratings changes, and the range of ratings is 4 (since ratings are between 1 & 5). The L_1 sensitivity of the prediction is:

$$\Delta p = \frac{4}{k}$$

Similarity of Two Movies

- Let U be the set of all users who have rated both movies i and j then

$$S(i,j) = \sum_{u \in U} (r_{uj} - \bar{r}_u)(r_{ui} - \bar{r}_u)$$

K-Nearest Users or K-Nearest Movies?

Find k most similar *users* to i that have also rated movie j?

Find k most similar *movies* to j that user i has rated?

Either way, after some pre-computation, we need to be able to find the k-nearest users/movies quickly!

Covariance Matrix

Movie-Movie Covariance Matrix

- (MxM) matrix
- Cov[i][j] measures similarity between movies i and j
- M ≈ 17,000
- More accurate

User-User Covariance Matrix?

- (NxN) Matrix to measure similarity between users
- N ≈ 500,000
- Less accurate

What do we need to make predictions?

For a large class of prediction algorithms it suffices to have:

- Gavg – average rating for all movies by all users
- Mavg – average rating for each movie by all users
- Average Movie Rating for each user
- Movie-Movie Covariance Matrix (COV)

Differentially Private Recommender Systems (High Level)

To respect approximate differential privacy publish

- Gavg + NOISE
- Mavg + NOISE
- COV + NOISE

- GS(Gavg), GS(Mavg) are very small so they can be published with little noise (e.g., Laplacian)
- GS(COV) requires more care (our focus)

- Don’t publish average ratings for users (used in per-user prediction phase using k-NN or other algorithms)

Source: *Differentially Private Recommender Systems(McSherry* and Mironov)