

# Private Data Release Via Learning Thresholds

Speaker: Moritz Hardt (IBM Almaden)

Joint work with

Guy Rothblum (MSR SVC)

Rocco Servedio (Columbia)

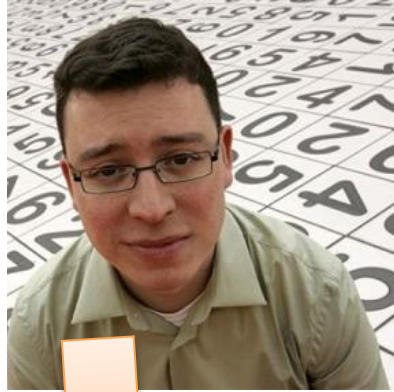
Advertisers



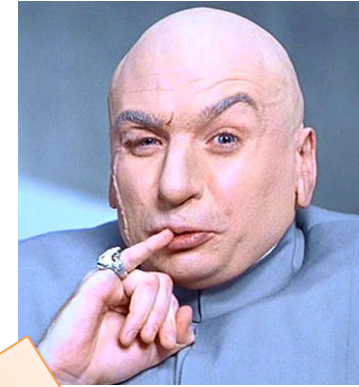
Banks, insurance companies



Scientists, data miners



Miscellaneous, possibly evil people



Sensitive Data

Country	Male	%	Female	% Female
United States	18,071,700	7,249,020	10,822,680	59.9
United Kingdom	6,825,600	2,436,400	4,389,200	64.3
Canada	6,850,860	2,224,840	4,626,020	67.5
Australia	1,920,300	684,940	1,235,360	64.3
Turkey	1,633,760	491,240	1,142,520	69.9
Sweden	960,620	313,620	647,000	67.3
Norway	822,560	287,860	534,700	65.0
France	771,580	233,960	537,620	69.6
South Africa	603,960	222,560	381,360	63.1
Colombia	488,520	133,100	355,420	72.7
Mexico	435,780	123,240	312,540	71.7
Egypt	398,200	147,080	248,120	62.3
Germany	366,000	122,020	243,980	66.6
India	333,560	113,000	220,560	66.1
Singapore	286,360	81,860	204,500	71.4
New Zealand	250,760	71,600	179,160	71.4
Spain	223,380	87,220	136,160	60.9
Israel	209,620	61,140	148,480	70.8
UAE	196,880	75,040	121,840	61.9
Malaysia	169,680	47,680	122,000	71.9
Lebanon	169,300	58,600	110,700	65.3
Ireland	167,660	45,760	121,900	72.6
Italy	161,600	47,620	114,180	70.5
Switzerland	144,400	45,740	98,660	68.3
Netherlands	139,020	41,080	97,940	70.4
China	132,460	37,560	94,900	71.6
Saudi Arabia	120,520	47,840	72,680	60.3
Pakistan	117,960	44,880	73,080	61.9
Japan	108,700	36,520	72,180	66.5
Korea	55,060	19,280	35,780	65.0
Dominican Republic	32,100	9,600	22,500	70.1
Totals	42,966,760	15,113,000	27,853,760	64.8

e.g., census data, medical records, educational data, financial data, social network profiles, search logs, commuting data, web traffic, tracking information etc.

Goal: Protect privacy of individuals



while allowing useful analyses

# Differential Privacy

[Dwork-McSherry-Nissim-Smith-06]

- **Rigorous privacy guarantee** with important properties (e.g., handles adversaries with auxiliary information, composition, robustness)
  - hundreds of papers in theory, databases, statistics, and systems
- **Intuition:** Probability of any (undesired) event can increase only slightly when an individual enters data set.

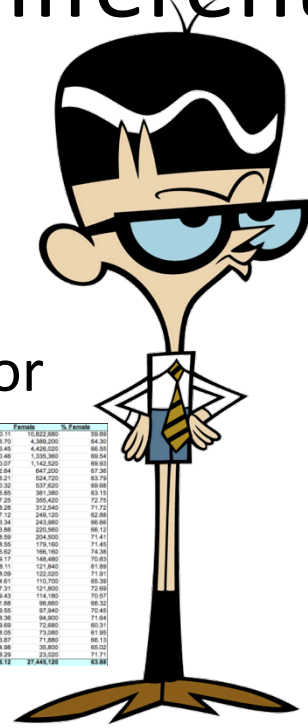
# Differentially private data

Intended answer  
 $f(D) := \mathbb{E}_{u \in D} f(u)$

Curator

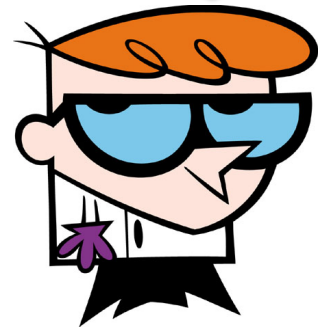
Country	Members	Men	% Male	Female	% Female
United States	18,211,700	7,280,200	40.16	10,931,500	59.84
United Kingdom	6,628,600	2,436,400	36.76	4,192,200	63.24
Canada	6,600,800	2,214,800	33.48	4,386,000	66.52
Russia	1,520,300	584,900	38.46	935,400	61.54
Turkey	1,513,700	691,200	45.66	822,500	54.34
Sweden	960,800	313,200	32.60	647,600	67.40
Norway	622,000	227,800	36.62	394,200	63.38
France	771,580	233,980	30.32	537,600	69.68
South Africa	603,900	222,800	36.89	381,100	63.11
Colombia	488,520	133,100	27.25	355,420	72.75
Mexico	430,700	123,200	28.60	307,500	71.40
Spain	346,200	141,200	40.78	205,000	59.22
Uganda	346,000	122,000	35.26	224,000	64.74
Germany	313,500	110,000	35.09	203,500	64.91
Singapore	286,300	81,800	28.57	204,500	71.43
New Zealand	200,700	71,000	35.38	129,700	64.62
Spain	223,300	87,200	39.05	136,100	60.95
Israel	208,600	81,100	38.92	127,500	61.08
USA	199,800	70,200	35.18	129,600	64.82
Malaysia	189,000	67,800	35.87	121,200	64.13
Libanon	189,300	68,800	36.35	120,500	63.65
India	187,600	67,800	36.14	119,800	63.86
Italy	181,800	67,600	37.24	114,200	62.76
Netherlands	158,400	57,400	36.24	101,000	63.76
Belgium	118,200	41,200	34.86	77,000	65.14
China	130,400	37,800	29.00	92,600	71.00
South Korea	130,000	47,800	36.77	82,200	63.23
Pakistan	117,800	44,800	38.05	73,000	61.95
Japan	108,700	38,800	35.69	69,900	64.31
Korea	85,000	32,800	38.59	52,200	61.41
Democratic Republic	82,100	30,800	37.51	51,300	62.49
Totals	42,868,700	16,521,800	38.57	26,346,900	61.43

$D \subseteq U$



Statistical query  
 $f: U \rightarrow \{0, 1\}$

Randomized answer A



Analyst

**Privacy:** Curator satisfies differential privacy

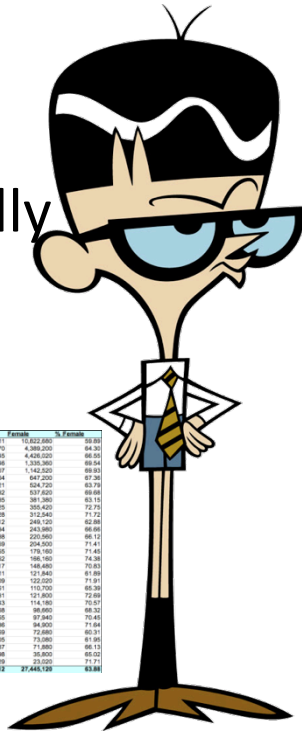
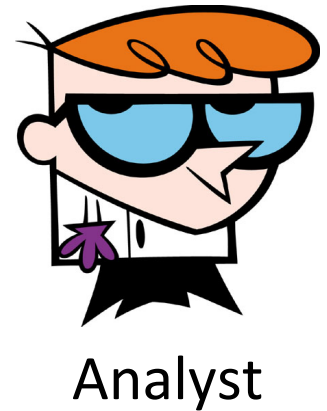
**Accuracy:**  $|A - f(D)| < \alpha$

# Multiple queries

Queries drawn from some distribution  $G$  over class  $Q$



$$S: Q \rightarrow [0, 1]$$



Differentially private  
Curator

Country	Members	Male	% Male	Female	% Female
United States	18,311,700	7,280,200	40.1%	11,031,500	59.9%
United Kingdom	6,625,600	2,436,400	36.7%	4,189,200	63.3%
Canada	6,600,800	2,214,800	33.6%	4,386,000	66.4%
Australia	1,920,300	584,900	30.4%	1,335,400	69.6%
Turkey	1,631,700	691,200	42.3%	940,500	57.7%
Russia	960,800	313,200	32.6%	647,600	67.4%
Germany	822,000	287,800	35.1%	534,200	65.0%
France	711,500	233,900	32.9%	477,600	67.1%
South Africa	600,900	202,900	33.8%	398,000	66.2%
Colombia	488,500	133,100	27.2%	355,400	72.8%
Mexico	436,700	123,200	28.2%	313,500	71.8%
Spain	366,200	141,200	38.6%	225,000	61.4%
Japan	346,000	122,000	35.3%	224,000	64.7%
Italy	291,700	113,000	38.8%	178,700	61.2%
Singapore	266,300	81,900	30.8%	184,400	69.2%
New Zealand	200,700	71,000	35.4%	129,700	64.6%
Spain	223,300	87,200	39.0%	136,100	61.0%
Israel	208,600	81,100	39.0%	127,500	61.0%
USA	198,800	70,200	35.3%	128,600	64.7%
Malaysia	189,000	47,900	25.3%	141,100	74.7%
Libanon	189,300	68,800	36.4%	120,500	63.6%
India	187,600	49,700	26.5%	137,900	73.5%
Italy	181,800	47,600	26.2%	134,200	73.8%
South Korea	154,000	47,100	30.6%	106,900	69.4%
Belgium	136,200	41,200	30.3%	95,000	69.7%
China	132,400	37,900	28.6%	94,500	71.4%
South Korea	131,000	47,900	36.6%	83,100	63.4%
Pakistan	117,900	44,800	38.0%	73,100	61.9%
Japan	108,700	38,800	35.7%	69,900	64.3%
Poland	92,000	34,200	37.2%	57,800	62.8%
Democratic Republic of Congo	82,100	30,900	37.6%	51,200	62.4%
Totals	42,868,700	15,521,800	36.2%	27,346,900	63.8%

$D \subseteq U$

Accurate over distribution:

$$\Pr_{f \sim G} \{ |S(f) - f(D)| \leq \alpha \} \geq 1 - \beta$$

# (PAC-)Learning

Unknown concept

Class of examples  $Q$   
 labeled examples drawn from some distribution  $G$

Country	Members	Men	% Men	Women	% Women
United States	18,271,702	7,246,323	39.71	11,025,380	60.29
United Kingdom	6,826,626	2,436,420	35.70	4,390,206	64.30
Canada	6,690,860	2,224,843	33.40	4,466,017	66.60
Australia	1,900,300	864,343	45.48	1,035,957	54.52
Turkey	6,820,790	491,243	7.20	6,329,547	92.80
Sweden	950,820	313,820	33.00	637,000	67.00
Norway	822,080	297,880	36.23	524,200	63.77
France	771,080	233,980	30.34	537,100	69.66
South Africa	460,080	222,080	48.27	238,000	51.73
Colombia	446,020	131,100	29.39	314,920	70.61
Malawi	435,780	123,240	28.28	312,540	71.72
Spain	366,260	147,960	40.39	218,300	59.61
Germany	366,000	122,020	33.34	243,980	66.66
Italy	333,960	113,020	33.84	220,940	66.16
Singapore	286,360	81,860	28.58	204,500	71.42
New Zealand	290,760	71,620	24.63	219,140	75.37
Spain	223,380	57,220	25.62	166,160	74.38
Israel	290,620	81,140	27.92	209,480	72.08
UAE	196,880	70,840	35.99	126,040	64.01
Malaysia	188,680	47,680	25.27	141,000	74.73
Latvia	180,300	66,600	37.22	113,700	62.78
Hungary	107,060	40,760	37.98	66,300	62.02
Bel	101,800	47,020	46.24	54,780	53.76
Switzerland	144,400	45,740	31.68	98,660	68.32
Netherlands	130,020	41,060	31.58	88,960	68.42
China	132,480	37,880	28.59	94,600	71.41
Israel/Anolis	120,020	47,040	39.19	72,980	60.81
Pakistan	117,960	44,860	38.03	73,100	61.97
Japan	106,700	36,820	34.52	69,880	65.48
Korea	85,080	19,280	22.66	65,800	77.34
Democratic Republic	30,100	6,080	20.19	24,020	79.81
Totals	42,964,780	15,921,680	36.12	27,043,100	63.88

Learner



Hypothesis



$$h: Q \rightarrow \{0, 1\}$$

Accurate over distribution:

$$\Pr_{(x,l) \sim G} \{h(x) = l\} \geq 1 - \beta$$

## Privacy

- Sensitive **database**
  - privacy-preserving access
- **Queries** labeled by answer on DB
- **Synopsis**
  - approximates DB on distribution over queries

## Learning

- Unknown **concept**
  - limited access via examples
- **Examples** labeled by concept
- **Hypothesis**
  - approximates target concept on distribution over examples

**Is this a coincidence?**

# In this talk

- Turn intuitive similarity between **private data release and learning theory** into formal connection
- **Efficient reduction** from (any) private data release problem to related learning problem
- Instantiate with existing learning algorithms to obtain first *subexponential* private data release algorithms in several settings

# Main Result

**Informal Theorem:** Distribution-free learning algorithm for thresholds of sums of predicates in  $Q$  implies differentially private release mechanism for  $Q$

- efficient reduction
- Extends to the distribution-specific setting
- Interfaces nicely with existing learning algorithms:
  - Learning based on polynomial threshold functions [Klivans-Servedio]
  - Harmonic Sieve [Jackson] and extension [Jackson, Klivans, Servedio]

# Applications

# State of the Art in DP

Laplace Mechanism [DMNS06]:  
Add Laplacian noise to  
each answer

Good:  
Highly efficient

Bad:  
Need  
 **$\#queries \ll |D|$**   
for good accuracy

Advanced mechanisms:  
[HR10,RR10,DRV10,BLR08]

Good:  
Can have  
 **$\#queries \gg |D|$**   
with good accuracy!

Bad:  
Running time  
***exponential*** in data  
dimensionality!

Sometimes necessary [DNRRV09]

Question: Can we have best of both worlds?

# Example: Boolean Conjunctions

Universe  $U = \{0,1\}^d$

Salary > \$50k	Syphilis	Height > 6'1	Weight < 180	Male
True	False	True	False	True
True	True	True	True	True
False	False	False	True	False
True	False	False	True	True
False	False	False	False	False

Example Conjunction: “(Salary > \$50k) AND (Male)”

Evaluates to 3 on this database

Allows us to learn **covariances** in a data set

Important class of queries in differential privacy

[BCDKMT07,KRSU10,GHRU11,HMT11,...]

**Informal Corollary** (Subexponential algorithm for conjunctions).

There is a differentially private release algorithm with running time  $\text{poly}(|D|)$  such that for any distribution over Boolean conjunctions the algorithm is w.h.p.  $\alpha$ -accurate provided that:

$$|D| \geq 2^{\tilde{O}(d^{1/3} \log(1/\alpha))}$$

Previous:  
 $2^{O(d)}$

**Informal Corollary** (Small width).

There is a differentially private release algorithm with running time  $\text{poly}(|D|)$  such that for any distribution over width- $k$  Boolean conjunctions the algorithm is w.h.p.  $\alpha$ -accurate provided that:

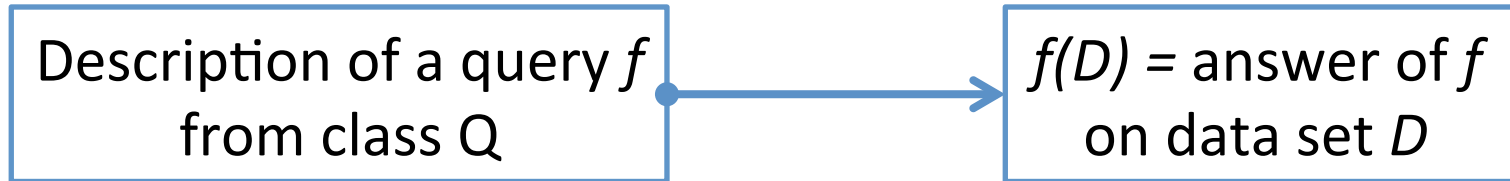
$$|D| \geq 2^{O\left(\sqrt{k \log(k \log d/\alpha)}\right)}$$

Previous:  
 $d^{O(k)}$

Main ideas

# Database as a function

Consider the mapping  $F: Q \rightarrow [0, 1]$

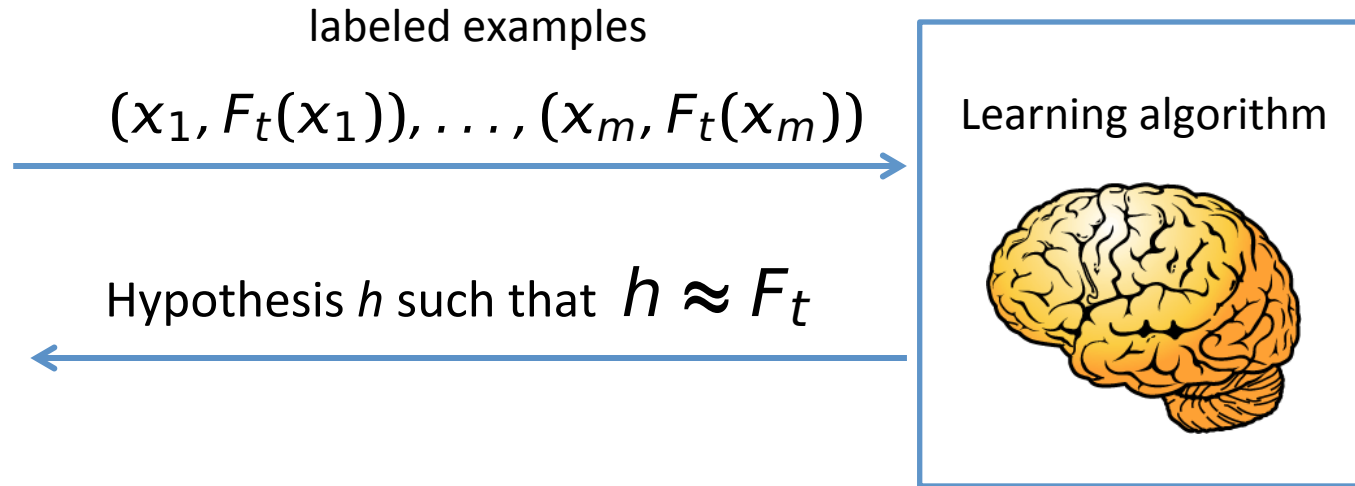


$$F_t: Q \rightarrow \{0, 1\}, \quad F_t = \mathbf{1}\{F \geq t\}$$

## Observation:

Enough to learn  $F_t$  for  $t = \alpha, 2\alpha, \dots, (1-\alpha)$  in order to approximate  $F$

# High-level idea



**Observation:** If all labels are privacy-preserving, then so will be hypothesis  $h$

# Main hurdles

- Privacy requires **noise**, noise might defeat learning algorithm
  - avoid noisy labels
- Can only generate  $|D|$  examples efficiently before running out of privacy
  - avoid **dependence on  $|D|$**  in learning algorithm!

## Threshold Oracle

Compute  $a = F(x) + N$

If  $|a - t|$  tiny:

output "fail"

Else if  $a > t$ :

output 1

Else if  $a < t$ :

output 0

Ensures:

1. Privacy
2. "Removes" noise
3. Complexity independent of  $|D|$

## Generate samples:

1. Pick  $x_1, x_2, \dots, x_m$
2. Receive  $b_1, b_2, \dots, b_m$  from TO
3. Remove all "failed" examples
4. Pass on remaining labeled examples to learner

" $F(x) > t$ "?

$b$  in  $\{0, 1, \text{fail}\}$

$(y_1, l_1), \dots, (y_r, l_r)$

Learning algorithm



$h \approx F_t$

# Conclusion

- Learning-theoretic approach to **breaking curse of dimensionality in differential privacy**
- Reduction from private data release to learning thresholds
  - no need to solve noisy/agnostic learning problem
- Helps to explain why learning techniques are so useful in differential privacy

# Awesome follow-up work

- Ullman-Vadhan (forthcoming): Can remove distributional relaxation and get same results for *all* Boolean conjunctions
  - e.g.  $\exp(O(d^{1/3}))$  running time, data complexity for all conjunctions, constant accuracy

# Open problems

- **Extend connection to learning theory**
  - inverse connection?
- **Harness implicit assumptions in privacy**
- No barriers for answering queries on graphs (e.g., **graph cuts**). Better upper bounds?
  - See discussion in Gupta-Roth-Ullman'11
- Efficient query release for **2-way, 3-way, 4-way** conjunctions?

Thank you

