

Differential Privacy and Fairness: Foundations and New Frontiers

Toniann Pitassi

Outline

1. Differential Privacy: The Basics
2. Differential Privacy in New Settings
 - Pan Privacy
 - Privacy in multi-party settings
 - Fairness

Outline

Differential Privacy: The Basics

Differential Privacy in New Settings

Pan Privacy

Privacy in multi-party settings

Fairness

Privacy in Statistical Data Analysis

Finding correlations

E.g. medical: genotype/phenotype correlations

Providing

Im

Publishing

Census data

Datamining

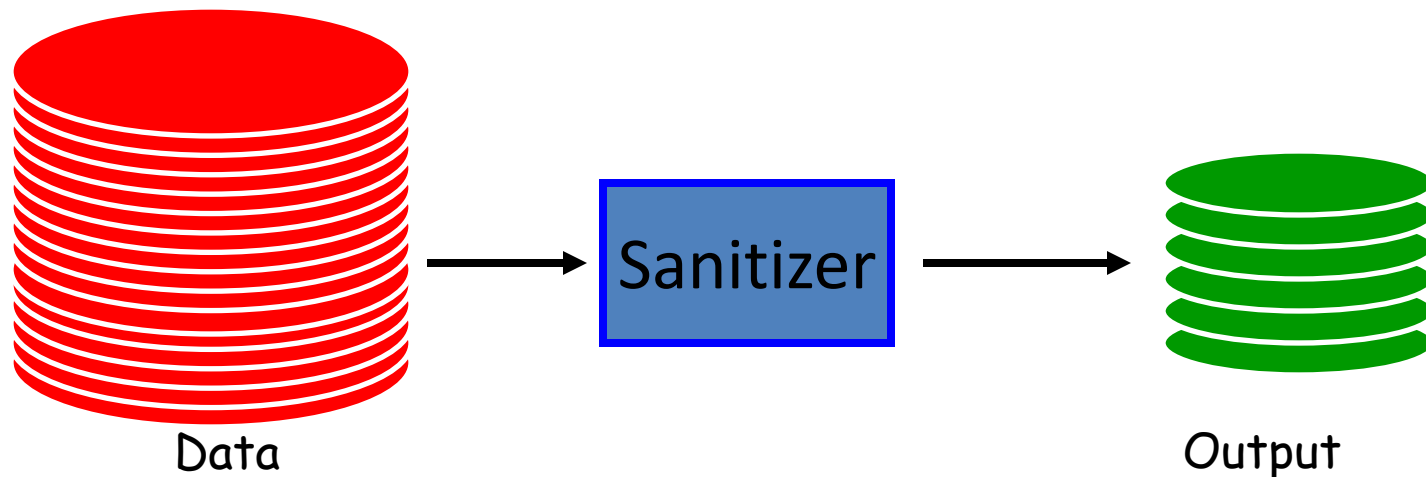
WHAT ABOUT PRIVACY?

However: data contains confidential information

The Basic Scenario

- Database with rows $x_1 \dots x_n$
- Each row corresponds to an individual in the database
- Columns correspond to fields, such as "name", "zip code"; some fields contain sensitive information.

Goal: Compute and release information about a sensitive database without revealing information about any individual



Typical Suggestions

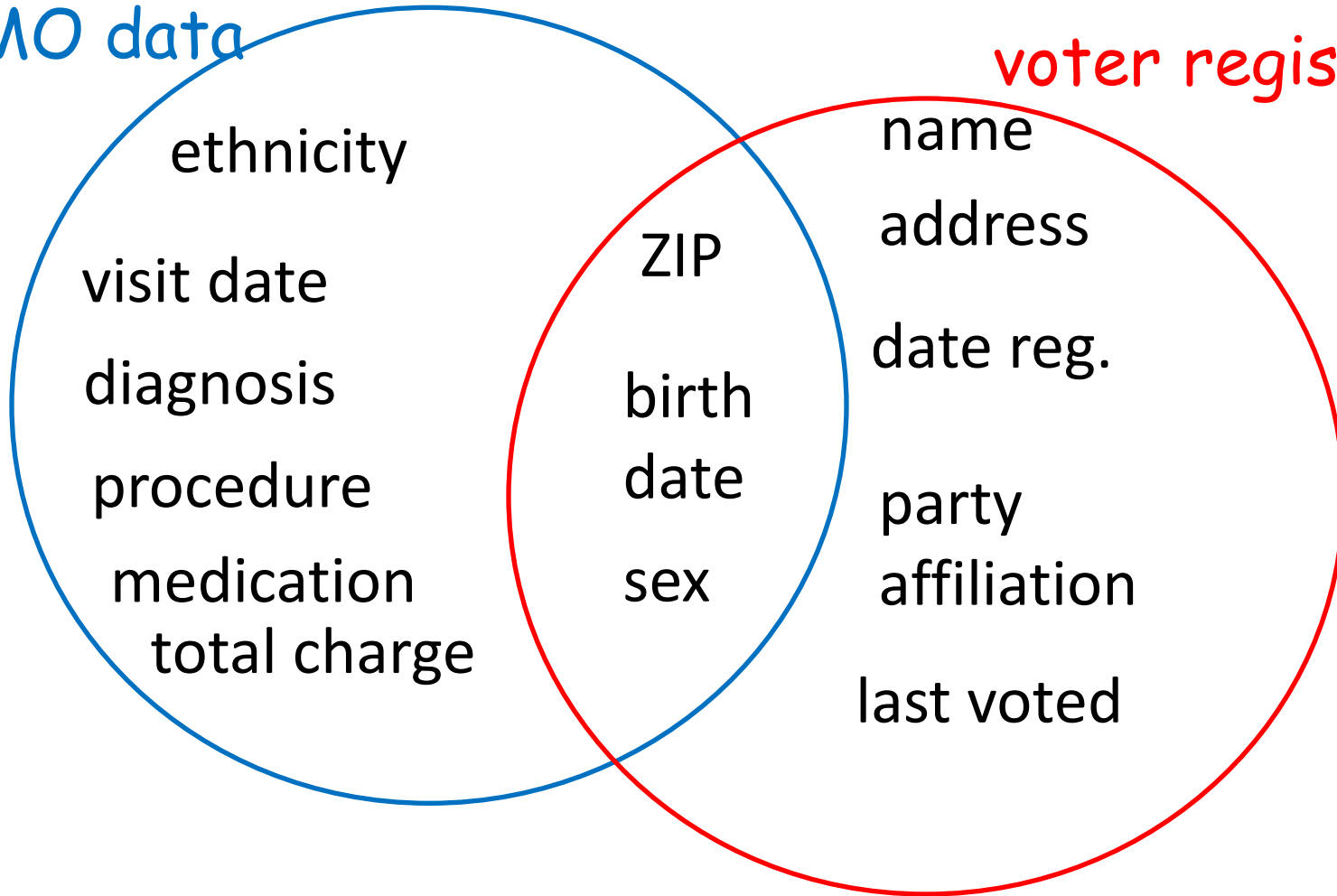
- Remove from the database any information which obviously identifies an individual.
 - i.e. remove "name" and "social security number"
 - ad hoc; propose-and-break cycle
- Only allow "large" set queries.
 - i.e. "How many females with initials TP are in theory?")
 - ad hoc; often not private
- Add random noise to true answer
 - if question is asked many times, privacy is lost
- Cryptography-inspired definition: Learn nothing about an individual that you didn't know otherwise
 - Limits utility

William Weld's Medical Record [S02]



HMO data

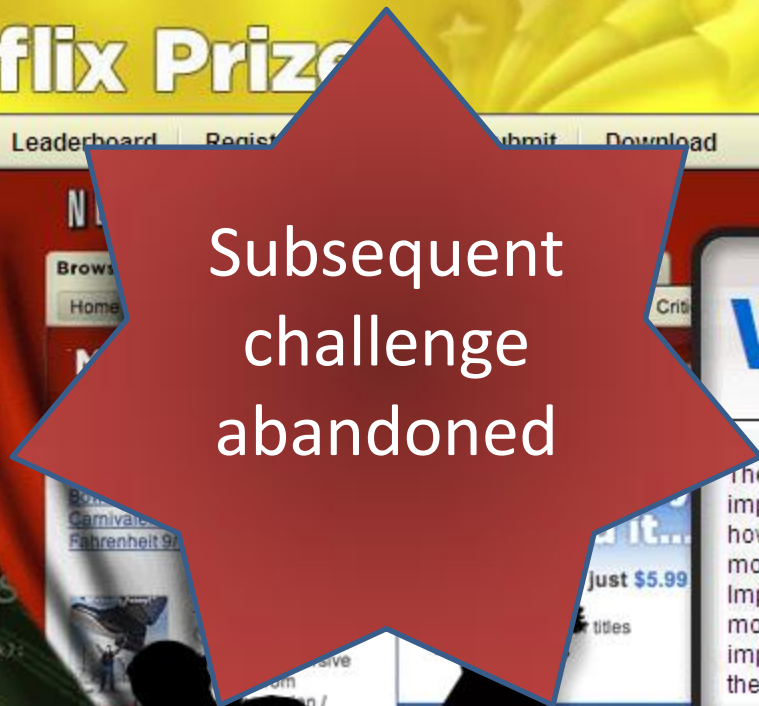
voter registration
data



NETFLIX

Netflix Prize

Home Rules Leaderboard Register Submit Download



Subsequent challenge abandoned

Welcome!

The Netflix Prize seeks to substantially improve the accuracy of predictions about how much someone is going to love a movie based on their movie preferences. Improve it enough and you win one (or more) Prizes. Winning the Netflix Prize improves our ability to connect people to the movies they love.

Read the [Rules](#) to see what is required to win the Prizes. If you are interested in joining the quest, you should [register a team](#).

You should also read the [frequently-asked questions](#) about the Prize. And check out how various teams are doing on the [Leaderboard](#).

Good luck and thanks for helping!

Code snippets on the left include: `startDataAS`, `istanceBD->`, `istance($sBan`, `startDataAS`, `BD->Recover`.

Movie recommendation: **Carnivale: Season 2** Disc Series. Daniel Kraus rivetingly cre series conti document t

Navigation: [Home](#), [Rules](#), [Leaderboard](#), [Register](#), [Submit](#), [Download](#)

AOL Search History Release (2006)

A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr. *The New York Times*

Published: August 9, 2006

Buried in a list of 20 million Web pages recently released on the Internet, AOL and assigned by the company to but it was not much of a shield.

Heads
Rolled



No. hundreds of search for a month period on topics ranging from “numb fingers” to “60 single men” to “dog that urinates on

Name: Thelma Arnold
Age: 62
Widow
Residence: Lilburn, GA

Differential Privacy

[Dwork, McSherry, Nissim, Smith 2006]

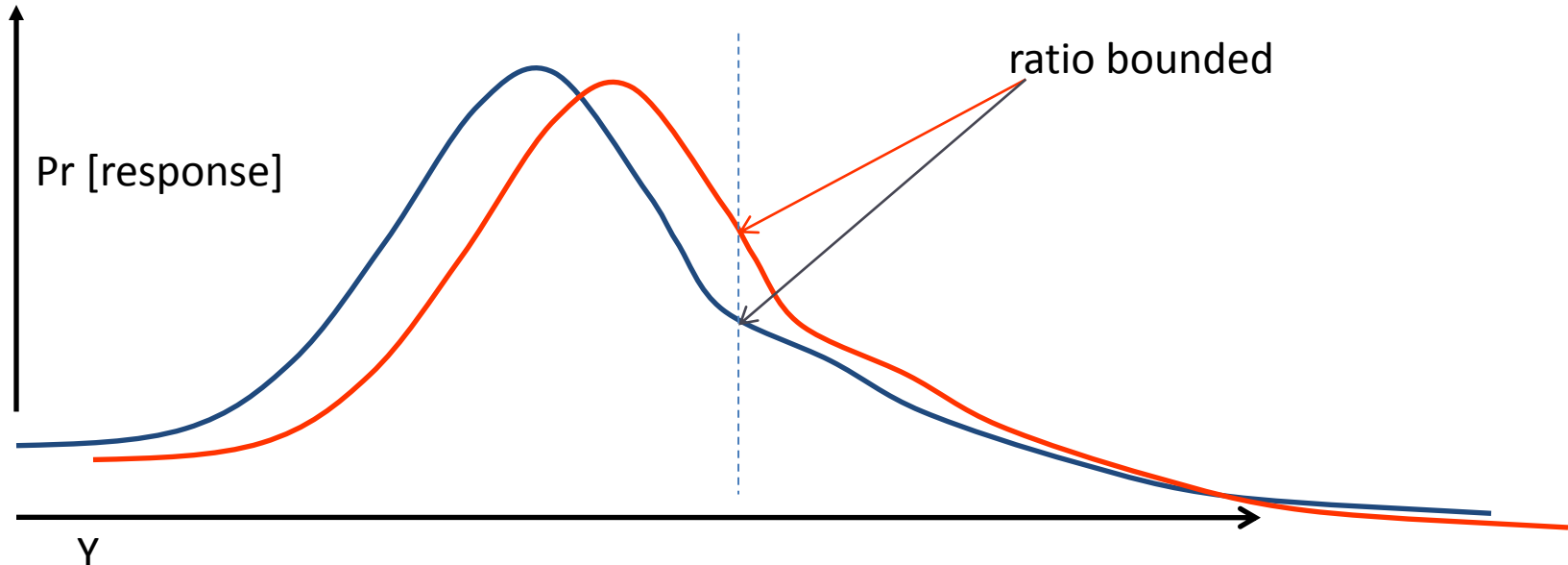
Q = space of queries; Y = output space; X = row space

Mechanism M : $X^n \times Q \rightarrow Y$ is ϵ -differentially private if:

for all q in Q , for all adjacent x, x' in X^n , the distributions $M(x, q)$, $M(x', q)$ are similar: $\forall y$ in Y, q in Q :

$$e^{-\epsilon} \leq \frac{\Pr[M(x, q) = y]}{\Pr[M(x', q) = y]} \leq e^{\epsilon}$$

Note: Randomness is crucial



Three Key Results

- Add Laplacian noise to answer
 - Works for numeric queries of low sensitivity
- Exponential mechanism
 - Extends Laplacian noise to work for non-numeric queries
- Handling many queries without compromising error too much

Achieving DP: Add Noise proportional to Sensitivity of the Query

$$\Delta q = \max_{\text{adj } x, x'} |q(x) - q(x')|$$

Sensitivity captures how much one person's data can affect the output

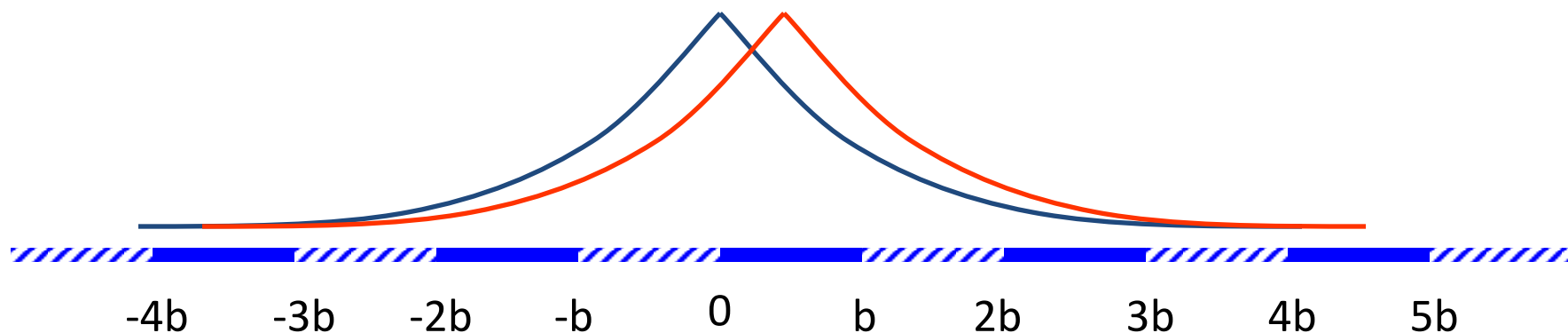
Counting queries have sensitivity 1.

Why Does it Work ?

$$\Delta q = \max_{D, D'} |q(x) - q(x')|$$

Theorem: To achieve ϵ -differential privacy, add scaled symmetric noise [Lap(b)] with $b = \Delta q / \epsilon$.

$$P(y) \sim \exp(-|y - q(x)|/b)$$



$$\frac{\Pr [M(x, q) = y]}{\Pr [(M(x', q) = y)]} = \frac{\exp(-|y - q(x)| \epsilon / \Delta q)}{\exp(-|y - q(x')| \epsilon / \Delta q)} \in [\exp(-\epsilon), \exp(\epsilon)]$$

Dealing with General Discrete-Valued Functions

- $f(x) \in S = \{y_1, y_2, \dots, y_k\}$
 - Strings, experts, small databases, ...
 - Each $y \in S$ has a utility for x , denoted $u(x, y)$
- Exponential Mechanism [McSherry-Talwar'07]

Output y with probability $\propto e^{u(x, y)\epsilon/\Delta u}$

$$\left[\frac{\exp(u(x, y))}{\exp(u(x', y))} \right]^{\epsilon/\Delta u} = \left[e^{u(x, y) - u(x', y)} \right]^{\epsilon/\Delta u} \leq e^\epsilon$$

Composition

- Simple k -fold composition of ϵ -differentially private mechanisms is $k\epsilon$ -differentially private
- Advanced: $\sqrt{k} \epsilon$, rather than $k\epsilon$
- This is tight if we want very small error
For counting queries, can't achieve $o(\sqrt{n})$ additive error with $O(n)$ queries.
- For larger error, much better results exist.

Hugely Many Queries

Blum, Ligett, Roth

- Proof of Concept: approach the problem within a learning framework.
- Handle exponentially many queries with low error, but infeasible
- Associate Q with a concept class C . For each x , output a probability distribution over synthetic databases.

Dwork, Rothblum, Vadhan

- Apply Boosting (continually re-weight the queries). Base learner using Laplacian mechanism.
- More efficient, better error.

Hardt-Rothblum

- Multiplicative Weight update method to handle the online setting.

Hugely Many Queries

	Counting Queries	Arbitrary Low-Sensitivity Queries
Offline	Error $n^{2/3}$ [Blum-Ligett-Roth'08] Runtime Exponential in $ U $ $(\epsilon, 0)$ -dp	Error \sqrt{n} [D.-Rothblum-Vadhan'10] Runtime Exp($ U $)
Online	Error \sqrt{n} [Hardt-Rothblum'10] Runtime Polynomial in $ U $	Error \sqrt{n} [Hardt-Rothblum] Runtime Exp($ U $)

Omitting **polylog**(various things, some of them big, like $|Q|$) terms

Differential Privacy: Summary

- **Resilience to All Auxiliary Information**
 - Past, present, future data sources and algorithms
- **Low-error high-privacy DP techniques exist for many problems**
 - datamining tasks (association rules, decision trees, clustering, ...), contingency tables, histograms, synthetic data sets for query logs, machine learning (boosting, statistical queries learning model, SVMs, logistic regression), various statistical estimators, network trace analysis, recommendation systems, ...
- **Programming Platforms**
 - <http://research.microsoft.com/en-us/projects/PINQ/>
 - http://userweb.cs.utexas.edu/~shmat/shmat_nsd10.pdf

Privacy in New Settings

- Pan Privacy
- Privacy in Multi-party settings
- Fairness

Privacy in New Settings

- Pan Privacy

[Dwork, Naor, Pitassi, Rothblum, Yekhanin]

- Privacy in Multi-party settings

- Fairness

How Can We Compute Without Storing Data?

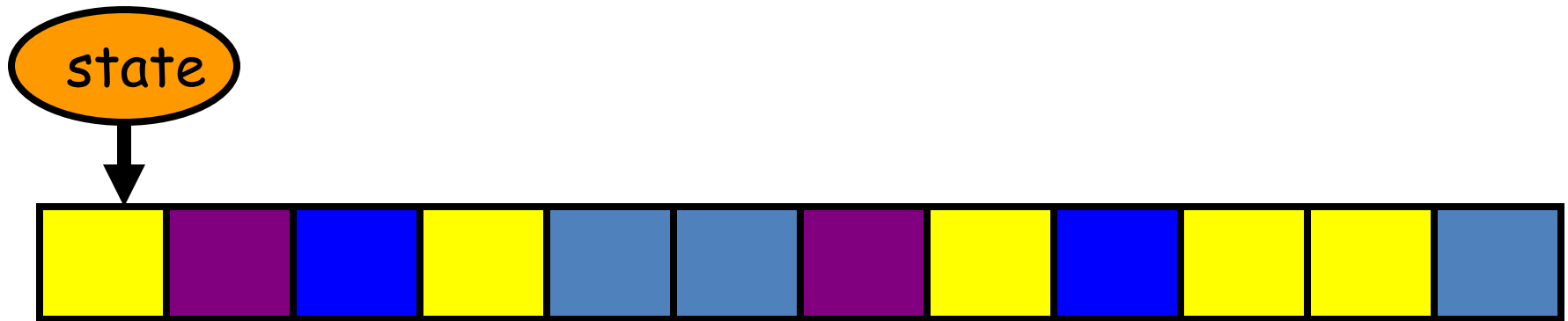
Pan Privacy:

- Input arrives continuously (a stream).
- A users data has many appearances, arbitrarily interleaved
- Queries need to be answered repeatedly
- Private "inside and out" :
 - query answers as well as the entire state of the computation should be differentially private!
- Protects against mission creep, subpoenas, intrusions

Pan-Private Streaming Model

[DNPRY]

- Data is a **stream** of items; each item belongs to a user. Sanitizer sees each item and updates internal state. Generates output at end of the stream (**single pass**).



Pan-Privacy: For every two **adjacent streams**, at any **single point in time**, the **internal state** (and final output) are differentially private.

What statistics have pan-private algorithms?

We give pan-private streaming algorithms for:

- Stream density / number of distinct elements
- t -cropped mean: mean, over users, of $\min(t, \#appearances)$
- Fraction of users appearing exactly k times
- Fraction of users appearing exactly 0 times modulo k
- Fraction of heavy-hitters, users appearing at least k times

What statistics do not have pan-private algorithms?

- How to prove negative results?
- By analogy to streaming, a nice approach uses communication complexity.
- This motivates the development of **differentially private communication complexity**:
 - Interesting in its own right.
 - Surprising connections to standard cc concepts
 - New lower bounds for pan-privacy

Privacy in New Settings

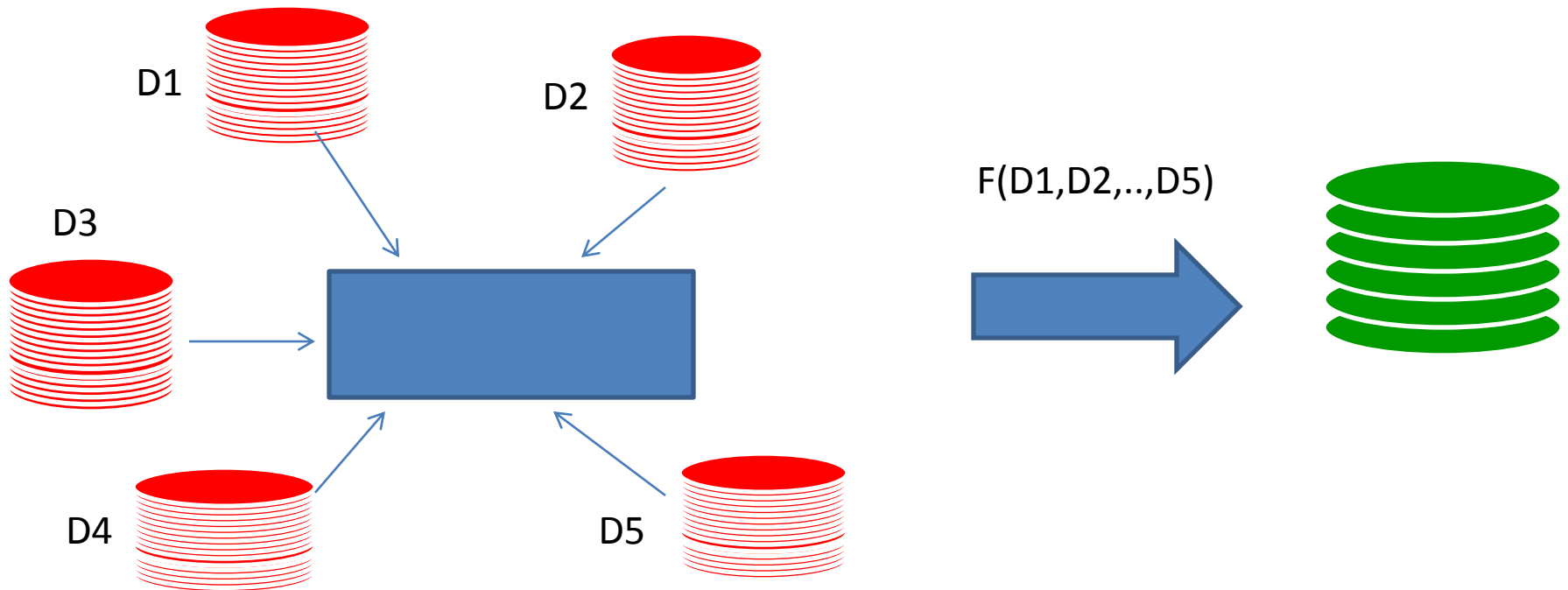
- Pan Privacy
- Privacy in Multi-party settings
- Fairness

Privacy in New Settings

- Pan Privacy
- **Privacy in Multiparty Settings**
[McGregor, Mironov, Pitassi, Reingold, Talwar, Vadhan]
- **Fairness**

Differentially Private Communication Complexity: A Distributed View

Multiple databases, each with private data.

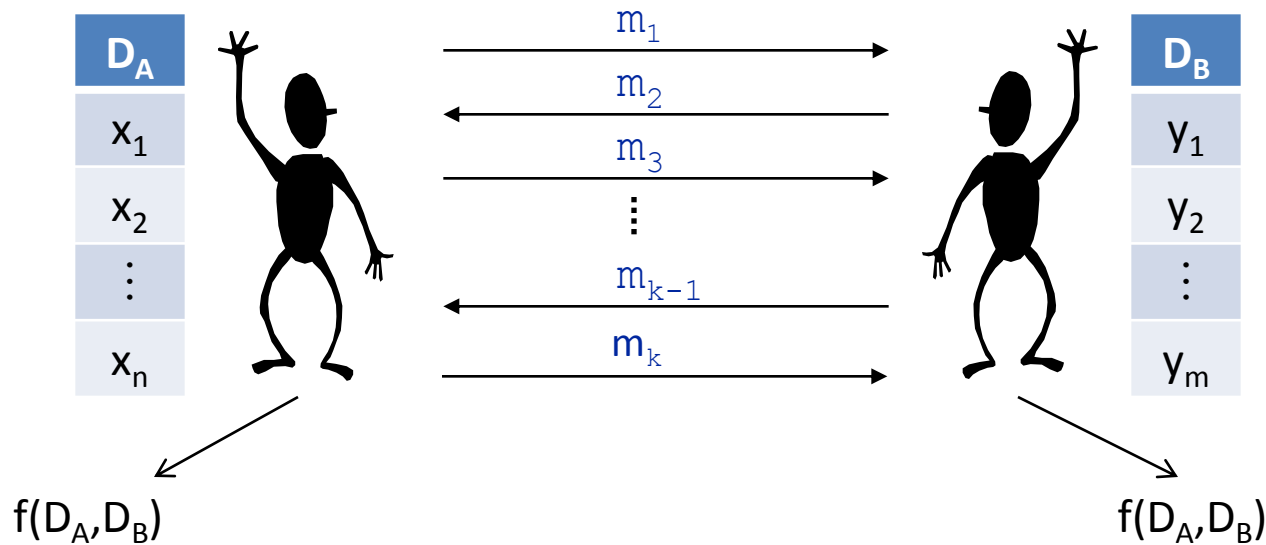


Goal: compute a joint function while maintaining privacy for any individual, with respect to both the outside world and the other database owners.

2-Party Communication Complexity

2-party communication: each party has a dataset.

Goal is to compute a function $f(D_A, D_B)$

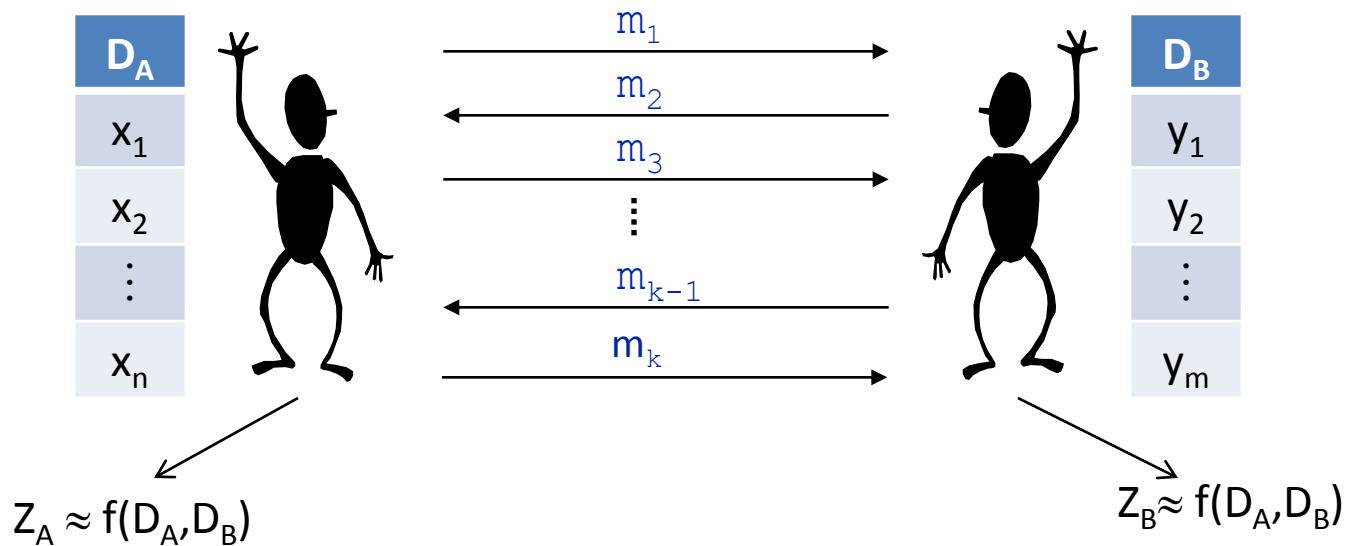


Communication complexity of a protocol for f is the number of bits exchanged between A and B .

In this talk, all protocols are assumed to be randomized.

2-Party Differentially Private CC

2-party (& multiparty) DP privacy: each party has a dataset; want to compute a joint function $f(D_A, D_B)$



A's view should be a **differentially private** function of D_B (even if A deviates from protocol), and vice-versa

Two-Party Differential Privacy

Let $P(x,y)$ be a 2-party protocol. P is ϵ -DP if:

(1) for all y , for every pair x, x' that are neighbors, and for every transcript π ,

$$\Pr[P(x,y) = \pi] \leq \exp(\epsilon) \Pr[P(x',y) = \pi]$$

(2) symmetrically, for all x , for every pair of neighbors y,y' and for every transcript π

$$\Pr[P(x,y)=\pi] \leq \exp(\epsilon) \Pr[P(x,y') = \pi]$$

Examples

1. **Ones(x,y)** = the number of ones in xy
 $\text{Ones}(00001111, 10101010) = 8.$

$$CC(\text{Ones}) = \log n.$$

There is a low error DP protocol.

2. **Hamming Distance HD(x,y)** = the number of positions i where $x_i \neq y_i$.
 $\text{HD}(00001111, 10101010) = 4$

$$CC(\text{HD}) = n.$$

No low error DP protocol

Is this a coincidence? Is there a connection between low cc and low-error DP protocols?

DP Protocols for Hamming Distance must have large error

Theorem. Let P be a 2-party ε -DP protocol, $\delta > 0$. Then with very high probability, P 's output differs from $IP(x,y)$ by at least $\Omega(\sqrt{n}/e^\varepsilon \log n)$

Notes:

- This lower bound is close to tight.
(There is an $O(\sqrt{n})$ error ε -dp protocol)
- Our result reveals strong connections between: DP protocols, low information cost protocols, and low complexity (short) protocols.

Implications of Lower bound for Hamming Distance

[MPRV] defined **computational ϵ -DP** protocols.

- Now the probability distribution over the transcripts for neighboring x, x' is e^ϵ - indistinguishable to a polytime algorithm.
- Via fully homomorphic encryption, any low sensitivity $f(x, y)$ has a $O(1)$ error computational ϵ -DP protocol, including Hamming distance.
- Thus our lower bound shows that in the context of distributed protocols, there can be a huge gain by relaxing DP to computational DP.

Applications to Pan Privacy

- Lower Bounds for ϵ -DP communication protocols imply pan privacy lower bounds for density estimation (via Hamming distance lower bound).
- Lower bounds also hold for multi-pass pan-private models

Privacy in New Settings

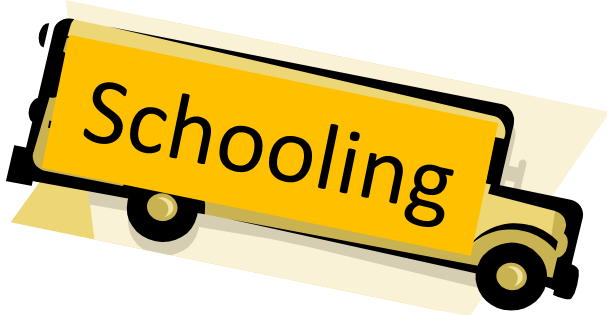
- Pan Privacy
- Privacy in Multi-party settings
- Fairness

Privacy in New Settings

- Pan Privacy
- Privacy in Multi-party settings
- **Fairness**

[Dwork, Hardt, Pitassi, Rothblum, Zemel]

Fairness in classification



Credit Application (WSJ 8/4/10)



More miles
and **no annual fee**

Earn trips faster with VentureOneSM

Get Started 

only at  **CARD LAB**

Capital One Card Lab
Platinum Prestige Credit Card

Capital One Card Lab
VentureOne Card

Savings Accounts
Earn With Great Rates

The advertisement features a yellow Capital One VentureOne Visa Signature credit card. The card displays the name 'VENTURE', the number '4000 1234 5678 9010', the expiration date '12/12 1/08', and the Visa Signature logo. The card is set against a background of a tropical island with palm trees and a blue sky.

User visits capitalone.com

Capital One uses tracking information provided by the tracking network [x+1] to personalize offers *

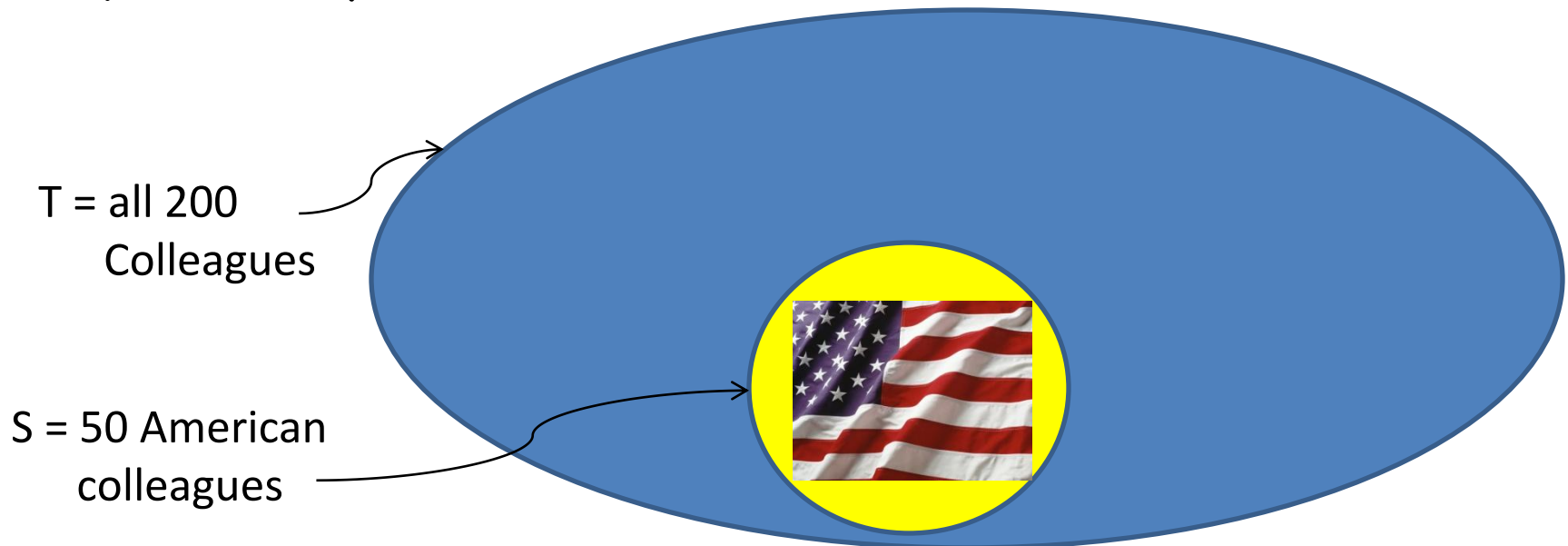
Concern: Steering minorities into higher rates (illegal)

Here: A CS Perspective


- Versatile *framework* for obtaining and understanding fairness
- An individual-based notion of **fairness-fairness through awareness**
- Lots of open problems/directions
 - Can Fairness Imply **Privacy** (beyond DB setting)?

First attempt: Group Fairness (Statistical Parity)

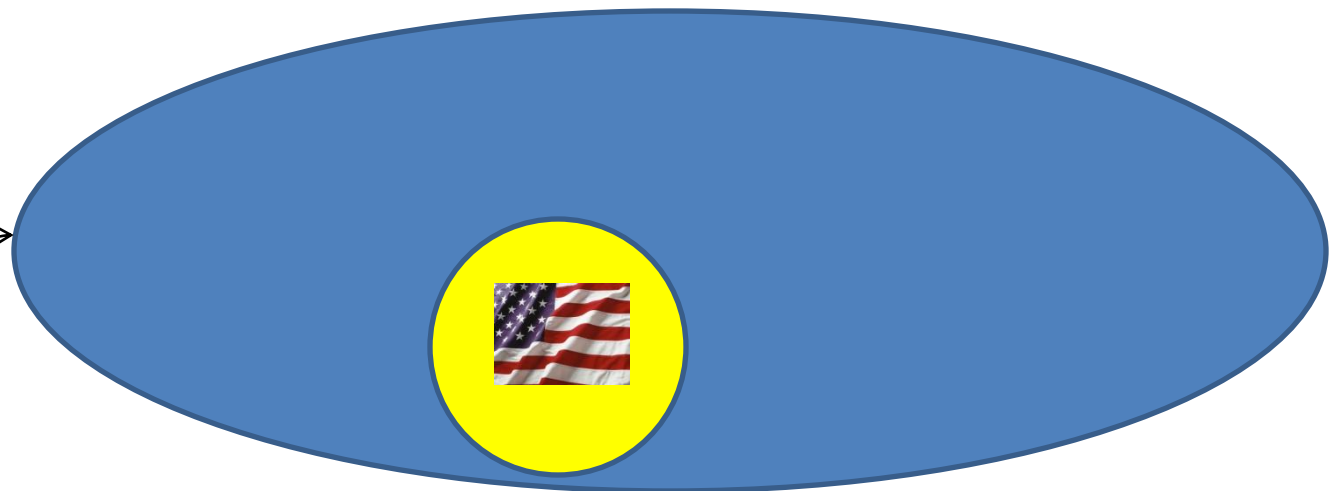
- Running Example: Pick DCS all-star departmental hockey team. (20 players out of 200), using machine learning
- Fairness: don't discriminate against your foreign American colleagues (50 people)
- Statistical Parity: $\Pr[\text{outcome} | S] = \Pr[\text{outcome} | T]$
equivalently: $\Pr[S | \text{outcome}] = \Pr[S]$



Statistical Parity may not be sufficient

- **Self-fulfilling prophecy:** Pick 5 of the worst American players. Then pick 15 best of the remaining.
- **Subset targeting:** Pick 5 from those who are  fans to satisfy the quota; Pick remaining 15 from rest.
- **Multiculturalism:** Best Americans are good at football; best non-Americans are good at soccer

200 Colleagues



Lesson: Fairness is Task Specific

- Fairness requires an understanding of the classification task
- In addition to statistical parity, we require that **similar** individuals are treated **similarly**

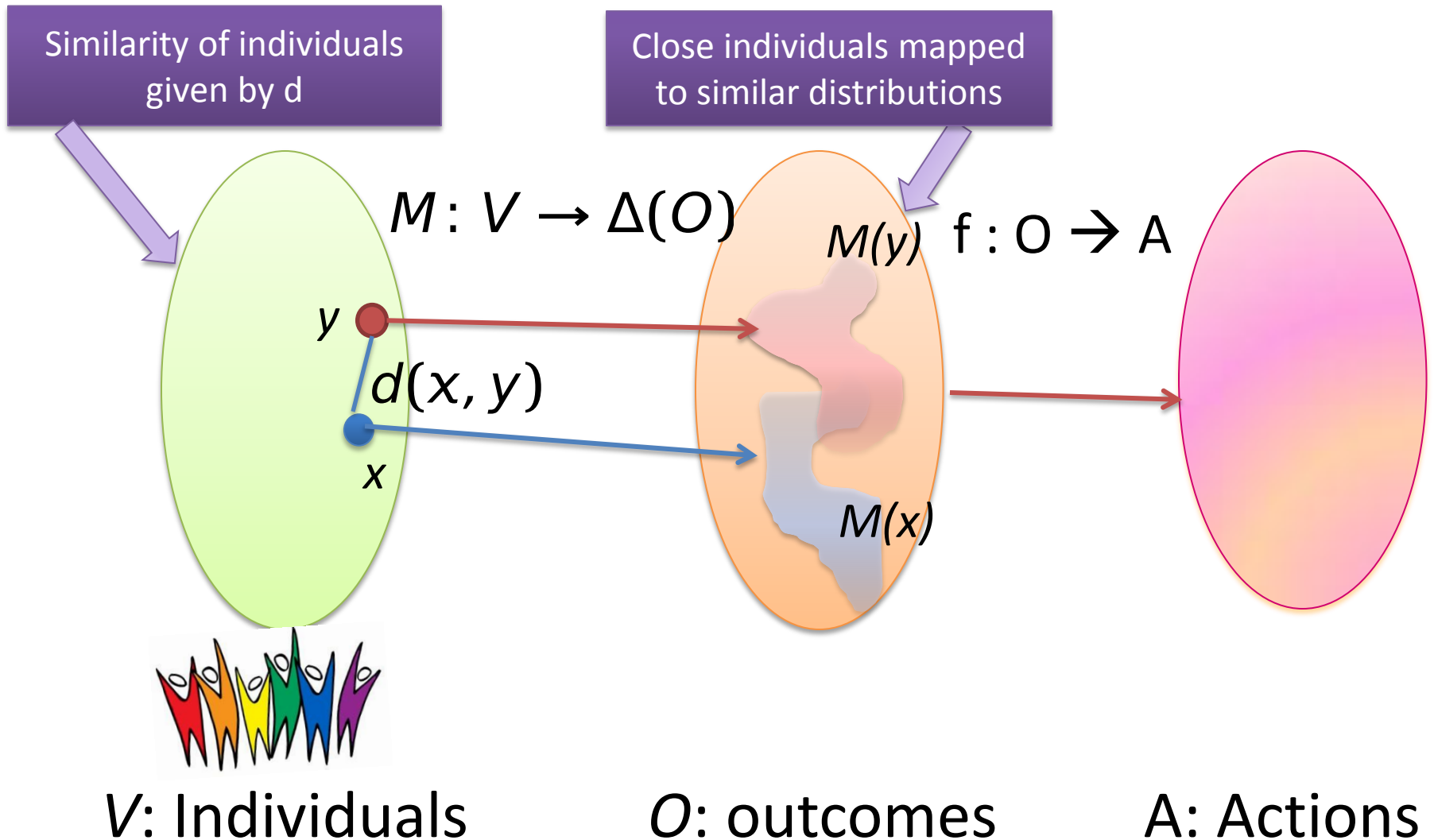


Similar for the
purpose of
classification task



Similar
distribution
over outcomes

Our Approach: Define a randomized mapping that “blends people with the crowd”



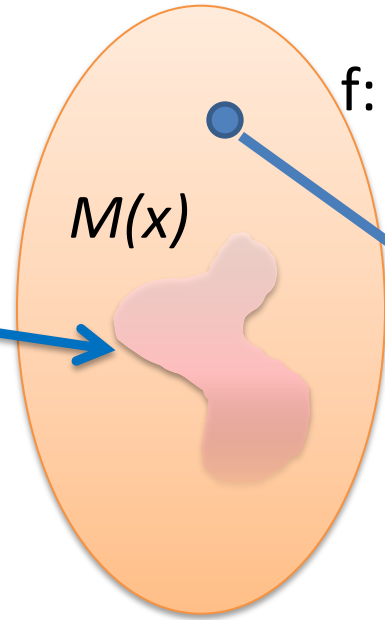
EXAMPLE: DCS All-Star Hockey Team

$$M: V \rightarrow \Delta(O)$$



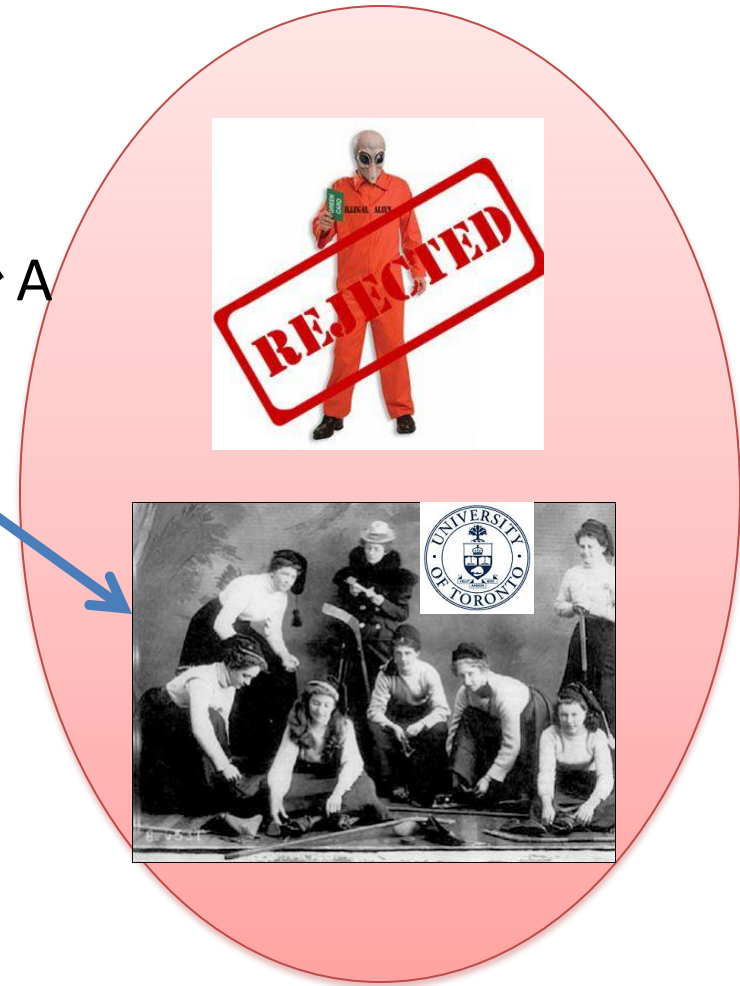
V: Individuals

$M(x)$



O: outcomes

$$f: O \rightarrow A$$



A: actions

Fairness versus Privacy

- Fairness is a measure of privacy: The mapping M is a differentially private mechanism (where databases are people).
- Privacy does not imply fairness.

An Algorithm for Fair Classification



utility
function
 $U: V \times O \rightarrow \mathcal{R}$

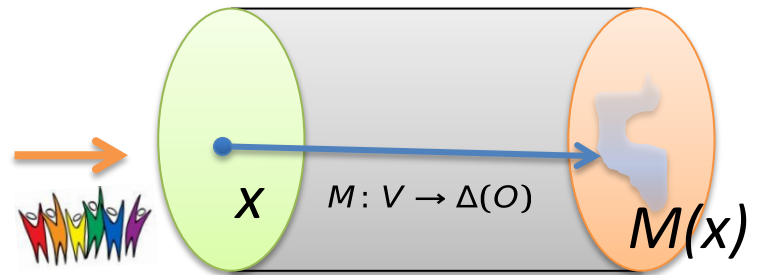


Metric

$d: V \times V \rightarrow \mathcal{R}$



d -fair mapping M



V : Individuals

O : outcomes

LP maximizing vendor's expected utility
subject to fairness condition



Analysis: Is the distance metric compatible with statistical parity?

Suppose we enforce individual fairness w.r.t. similarity metric d .

Question: Which pairs of groups of individuals receive (approximately) equal outcomes?

Theorem: Answer is given by the **Earthmover distance** (w.r.t. d) between the two groups.



Open Problems

- Is differential privacy the right definition?
Not many competing definitions at present (PAR)
- Axiomatic basis for differential privacy?
- Develop a large-scale application
- Privacy for other types of data
handwritten notes, images, etc.
- Fairness
Just the beginning...
What can be done without a metric?
Case study (health care?)

Thanks!