Introduction to Differential Privacy

Jeremiah Blocki
CS-555
11/22/2016

Credit: Some slides are from Adam Smith

Google Scholar

differential privacy

About 3,000,000 results (**0.06** sec)

Articles
Case law
My library

Any time
Since 2016
Since 2015
Since 2012
Custom range...

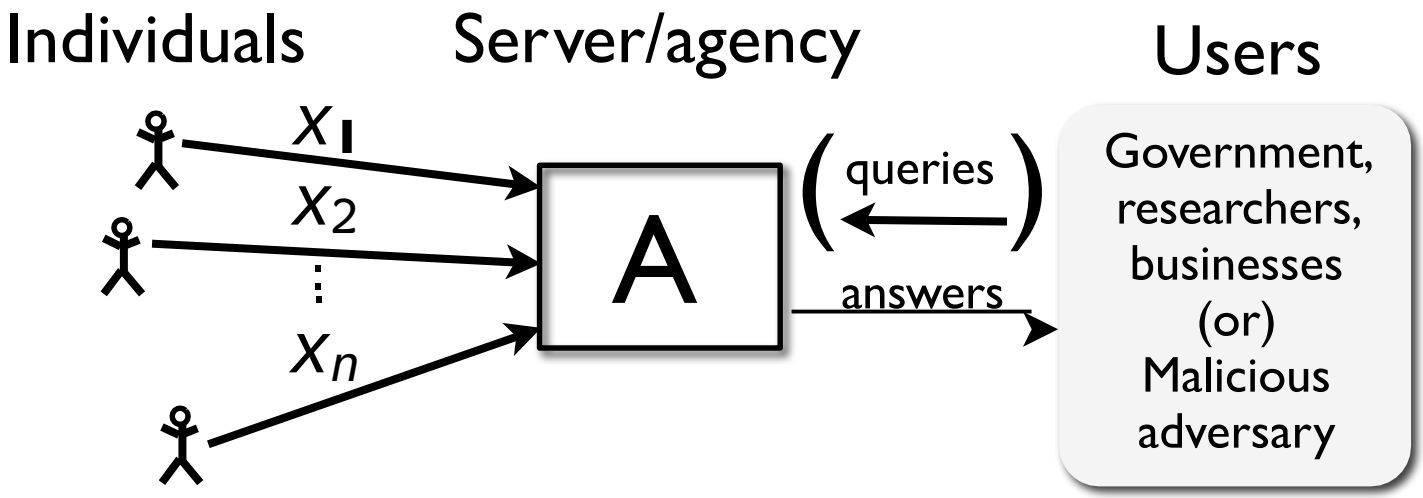
Differential privacy: A survey of results
C Dwork - International Conference on Theory and Applications of …, 2008 - Springer
Abstract Over the past five years a new approach to **privacy**-preserving data analysis has born fruit [13, 18, 7, 19, 5, 37, 35, 8, 32]. This approach differs from much (but not all!) of the related literature in the statistics, databases, theory, and cryptography communities, in that **...**
Cited by 2557 Related articles All 32 versions Web of Science: 365 Cite Save More

Mechanism design via **differential privacy**
F McSherry, K Talwar - … of Computer Science, 2007. FOCS'07. …, 2007 - ieeexplore.ieee.org
Abstract We study the role that **privacy**-preserving algorithms, which prevent the leakage of specific information about participants, can play in the design of mechanisms for strategic agents, which must encourage players to honestly report information. Specifically, we **...**
Cited by 708 Related articles All 25 versions Cite Save

Microsoft®
Research

Differential Privacy

Privacy in Statistical Databases



- What information can be released?
- Two conflicting goals
  - Utility: Users can extract "global" statistics
  - Privacy: Individual information stays hidden
- How can these be made **precise**?
  - (How context-dependent **must** they be?)

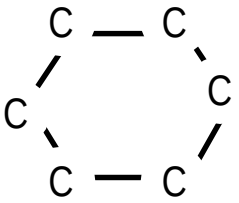
Secure Function Evaluation

a.k.a. "multi-party computation"

- Several parties, each with input x_i, want to compute a function $f(x_1, x_2, ..., x_n)$
- **Ideal world**: all parties hand their inputs to a trusted party who computes $f(x_1, ..., x_n)$ and releases the result
- There exist secure protocols for this task
  - Idea: a simulator can geneerate a dummy transcript given only the value of f
- Privacy: use SFE protocols to jointly data mine
  - Horizontal vs vertical
  - Lots of papers

Why not use crypto definitions?

- **Attempt #1**:
  - **Def'n**: For every entry i, no information about x_i is leaked (as if encrypted)
  - **Problem**: no information at all is revealed!
  - Tradeoff privacy vs utility
- **Attempt #2**:
  - Agree on summary statistics f(DB) that are safe
  - **Def'n**: No information except f(DB)
  - **Problem**: why is f(DB) safe to release?
  - Tautology trap
  - (Also: how do you figure out what **f** is?)



A Problem Case

Question 1: How many people in this room have cancer?

Question 2: How many students in this room have cancer?

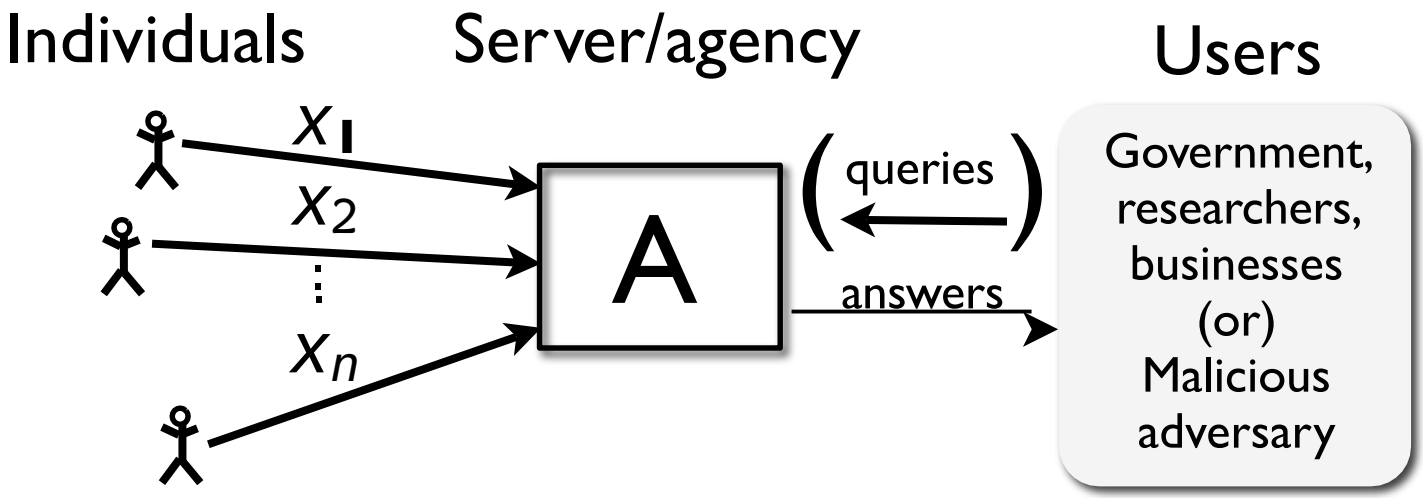
The difference (A1-A2) exposes my answer!

Why not use crypto definitions?

- **Problem**: Crypto makes sense in settings where the line between "inside" and "outside" is well-defined
  - E.g. psychologist:
    - "inside" = psychologist and patient
    - "outside" = everyone else

- Statistical databases: fuzzy line between inside and outside

Privacy in Statistical Databases



- What information can be released?
- Two conflicting goals
  - Utility: Users can extract "global" statistics
  - Privacy: Individual information stays hidden
- How can these be made **precise**?
  - (How context-dependent **must** they be?)

Straw Man #0

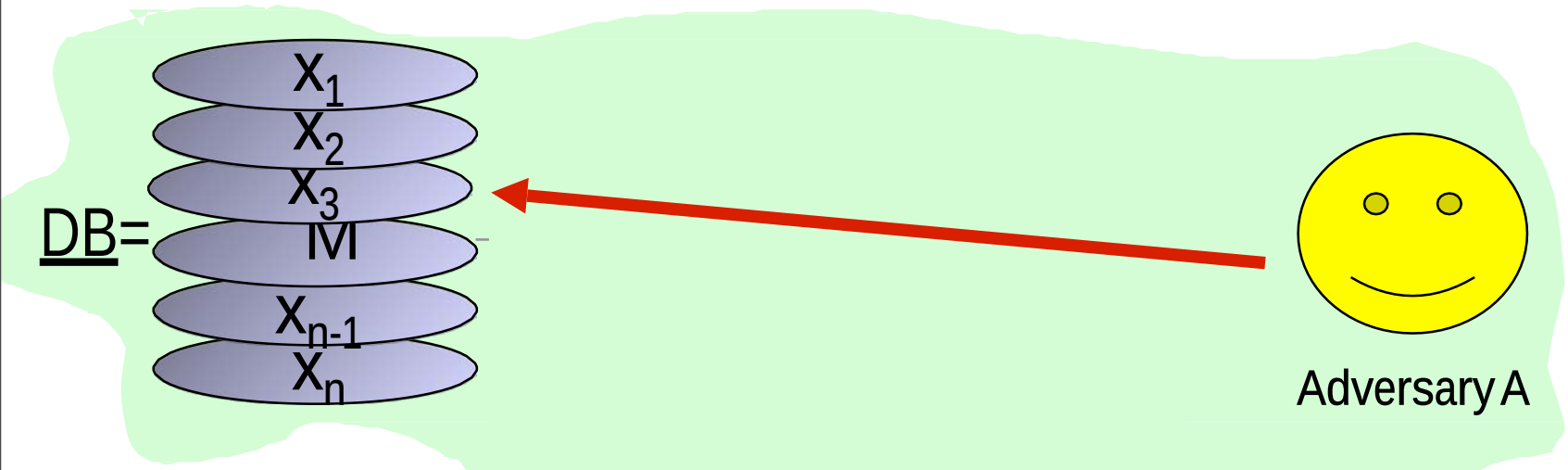
L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

Simple Demographics Often Identify People Uniquely

1. Abstract

In this document, I report on experiments I conducted using 1990 U.S. Census summary data to determine how many individuals within geographically situated populations had combinations of demographic values that occurred infrequently. It was found that combinations of few characteristics often combine in populations to uniquely or nearly uniquely identify some individuals. Clearly, data released containing such information about these individuals should not be considered anonymous. Yet, health and other person-specific data are publicly available in this form. Here are some surprising results using only three fields of information, even though typical data releases contain many more fields. It was found that 87% (216 million of 248 million) of the population in the United States had reported characteristics that likely made them unique based only on {5-digit ZIP, gender, date of birth}. About half of the U.S. population (132 million of 248 million or 53%) are likely to be uniquely identified by only {place, gender, date of birth}, where place is basically the city, town, or municipality in which the person resides. And even at the county level, {county, gender, date of birth} are likely to uniquely identify 18% of the U.S. population. In general, few characteristics are needed to uniquely identify a person.

Straw man #1: Exact Disclosure



- **Def'n:** safe if adversary cannot learn any entry **exactly**
  - leads to nice (but hard) combinatorial problems
  - Does not preclude learning value with 99% certainty or narrowing down to a small interval
- Historically:
  - Focus: auditing interactive queries
  - Difficulty: understanding relationships between queries
  - E.g. two queries with small difference

Two Intuitions for Data Privacy

- "If the release of statistics *S* makes it possible to determine the value [of private information] more accurately than is possible without access to *S*, a disclosure has taken place." [Dalenius]
  - Learning more about me should be hard

- Privacy is "protection from being brought to the attention of others." [Gavison]
  - Safety is blending into a crowd

A Problem Example?

Suppose adversary knows that I smoke.

Question 0: How many patients smoke?

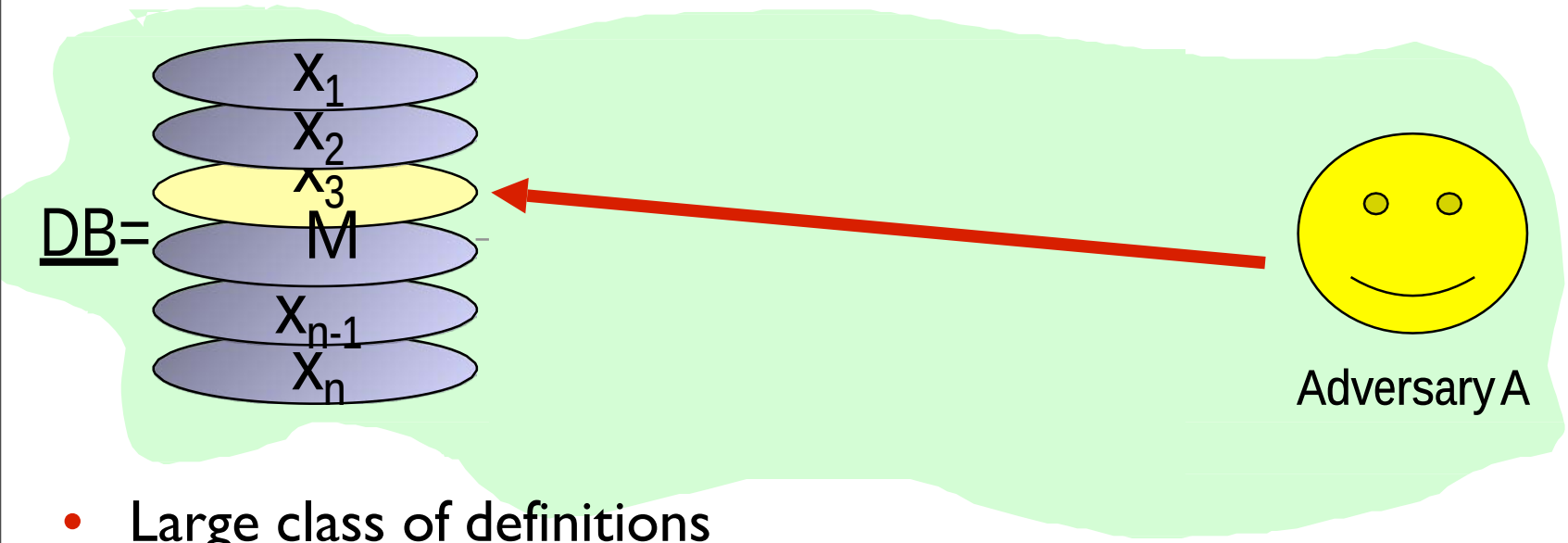
Question1: How many smokers have cancer?

Question 2: How many patients have cancer?

If adversary learns that smoking → cancer then he learns my health status.

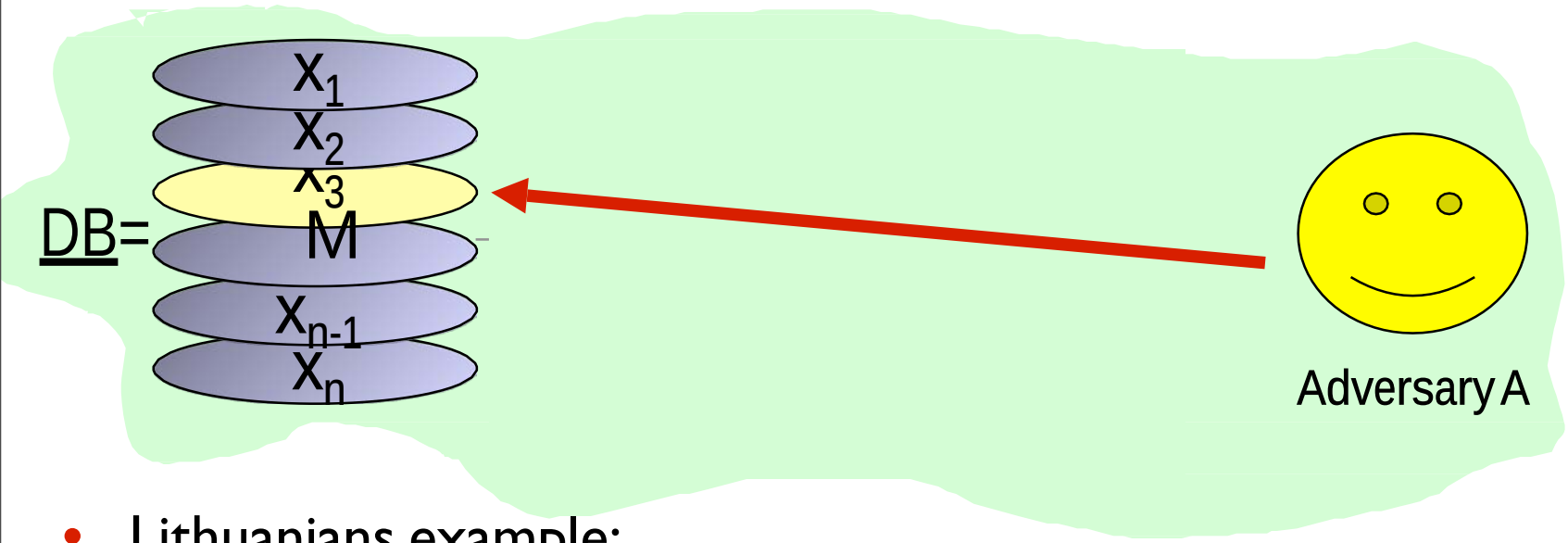
Privacy Violation?

Preventing Attribute Disclosure



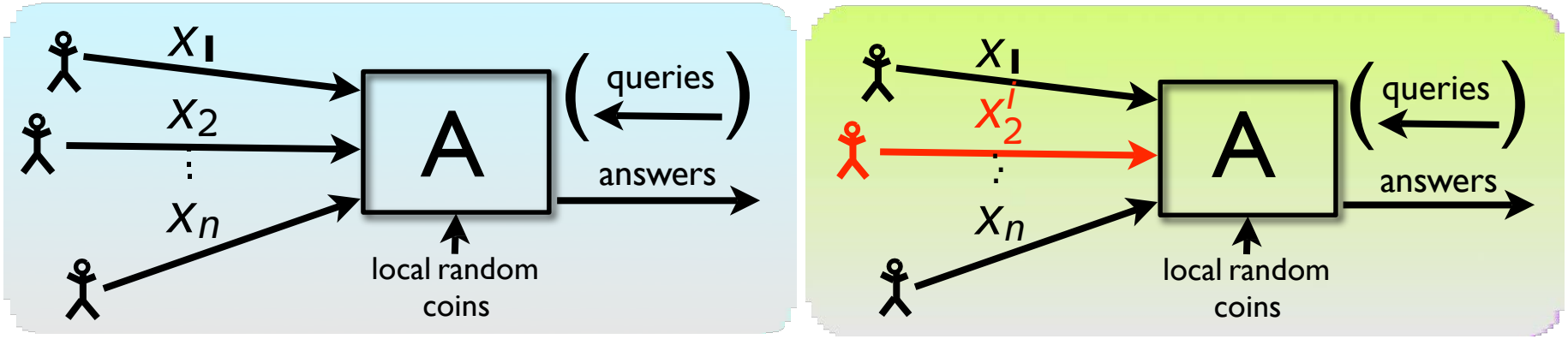
- Large class of definitions
  - safe if adversary can't learn "too much" about any entry
  - E.g.:
    - Cannot narrow $\mathbf{X_i}$ down to small interval
    - For uniform $\mathbf{X_i}$, mutual information $I(X_i; San(DB)) \cdot \varepsilon$
- How can we decide among these definitions?

Differential Privacy



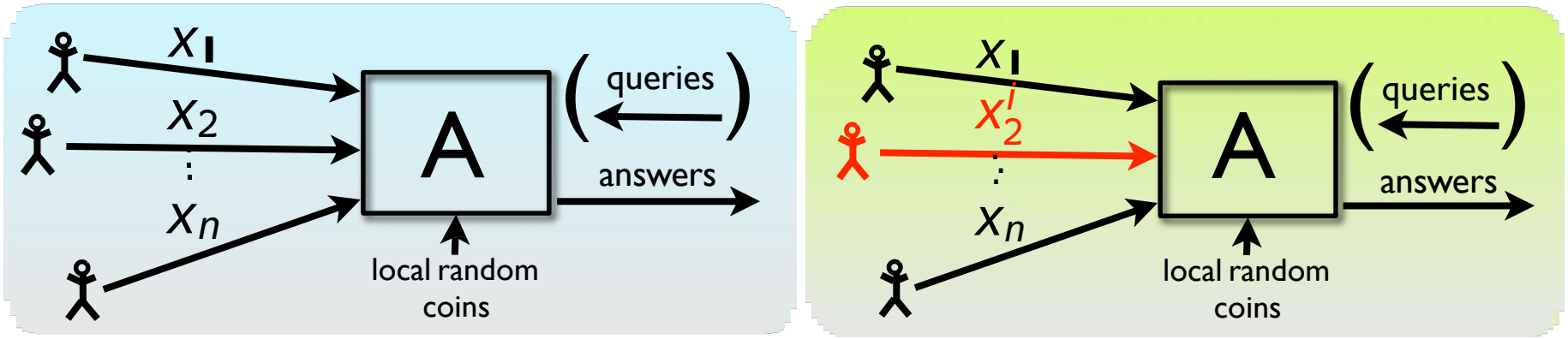
- Lithuanians example:
  - Adv. learns height even ifAlice not in DB
- Intuition [DM]:
  - "Whatever is learned would be learned regardless of whether or not Alice participates"
  - Dual:Whatever is already known, situation won't get worse

Approach: Indistinguishability



x' is a neighbor of x
if they differ in one row

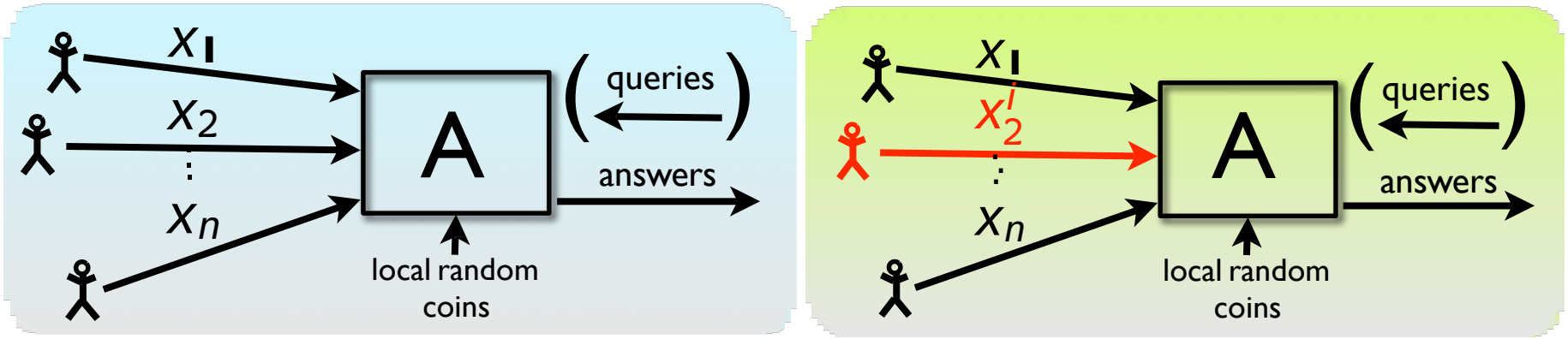
Approach: Indistinguishability



x' is a neighbor of x
if they differ in one row

Neighboring databases induce **close** distributions on transcripts

Approach: Indistinguishability



x' is a neighbor of x
if they differ in one row

Definition: A is ε –differentially private
if, for all neighbors x,x',
for all subsets S of transcripts

$$\Pr[A(x) \in S] \leq e^{\varepsilon} \Pr[A(x') \in S]$$

Neighboring databases induce **close** distributions on transcripts

Approach: Indistinguishability

- Note that ε has to be non-negligible here
  - Triangle inequality: **any** pair of databases at distance $< \varepsilon n$
  - If $\varepsilon < 1/n$ then users get no info!
- Why this measure?
  - Statistical difference doesn't make sense with $\varepsilon > 1/n$
  - E.g. choose random i and release i, x_i
  - This compromises someone's privacy w.p. 1

Definition: A is ε –differentially private
if, for all neighbors x,x',
for all subsets S of transcripts

$$\Pr[A(x) \in S] \leq e^{\varepsilon} \Pr[A(x') \in S]$$

Neighboring databases induce **close** distributions on transcripts

Differential Privacy

- Another interpretation [DM]:

  You learn the same things about me
  regardless of whether I am in the database

- Suppose you know I am the height of median Canadian
  - You could learn my height from database!
    But it didn't matter whether or not my data was part of it.
  - Has my privacy been compromised? No!

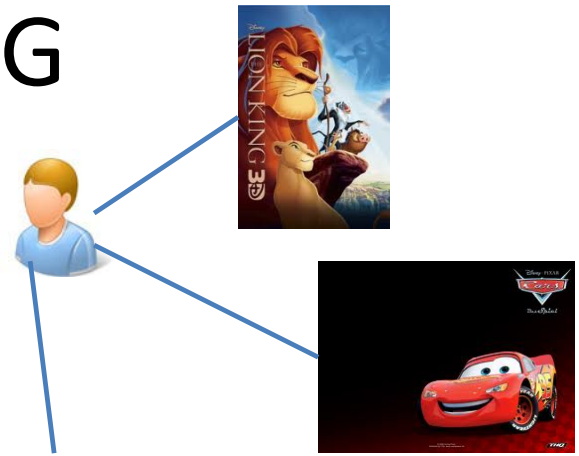
Definition: A is ε –differentially private
if, for all neighbors x,x',
for all subsets S of transcripts

$$\Pr[A(x) \in S] \leq e^{\varepsilon} \Pr[A(x') \in S]$$

Neighboring databases induce **close** distributions on transcripts

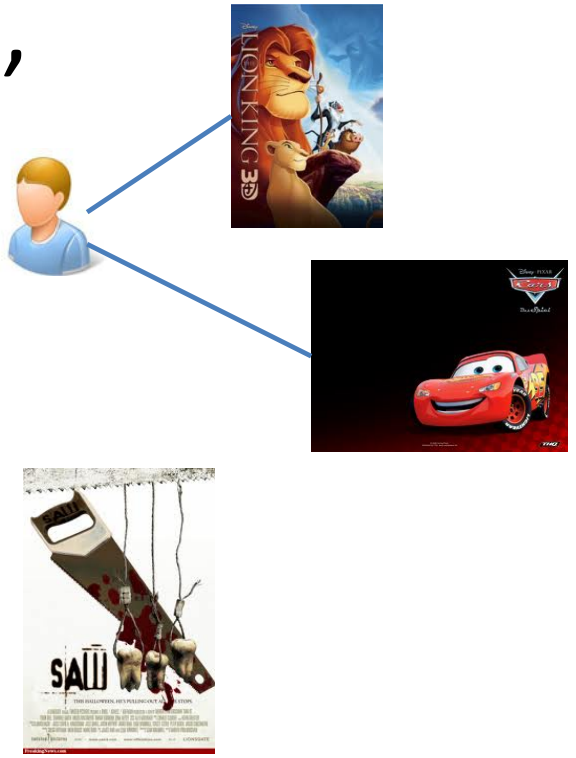
Graphs: Edge Adjacency

G



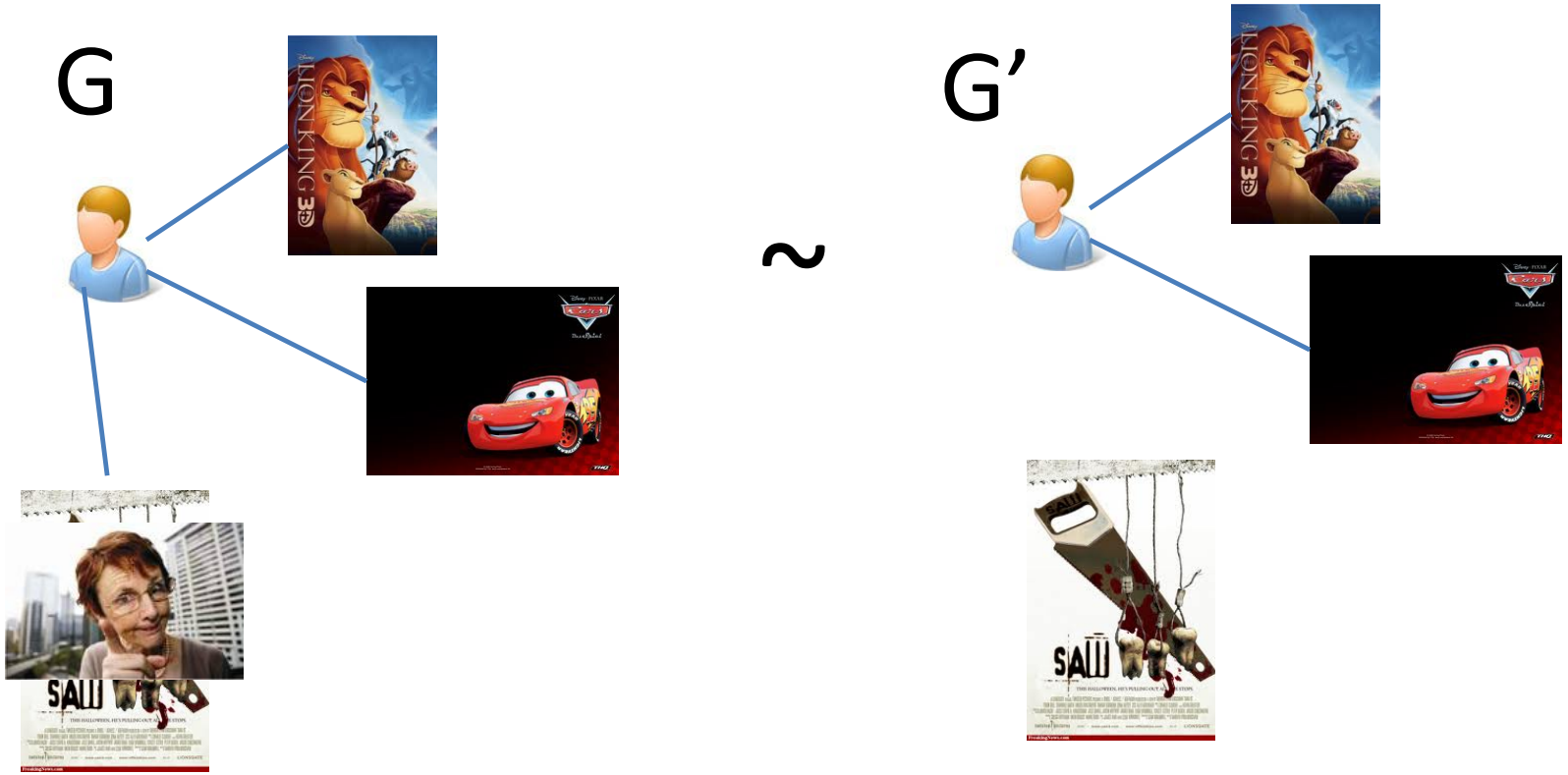
~

G’



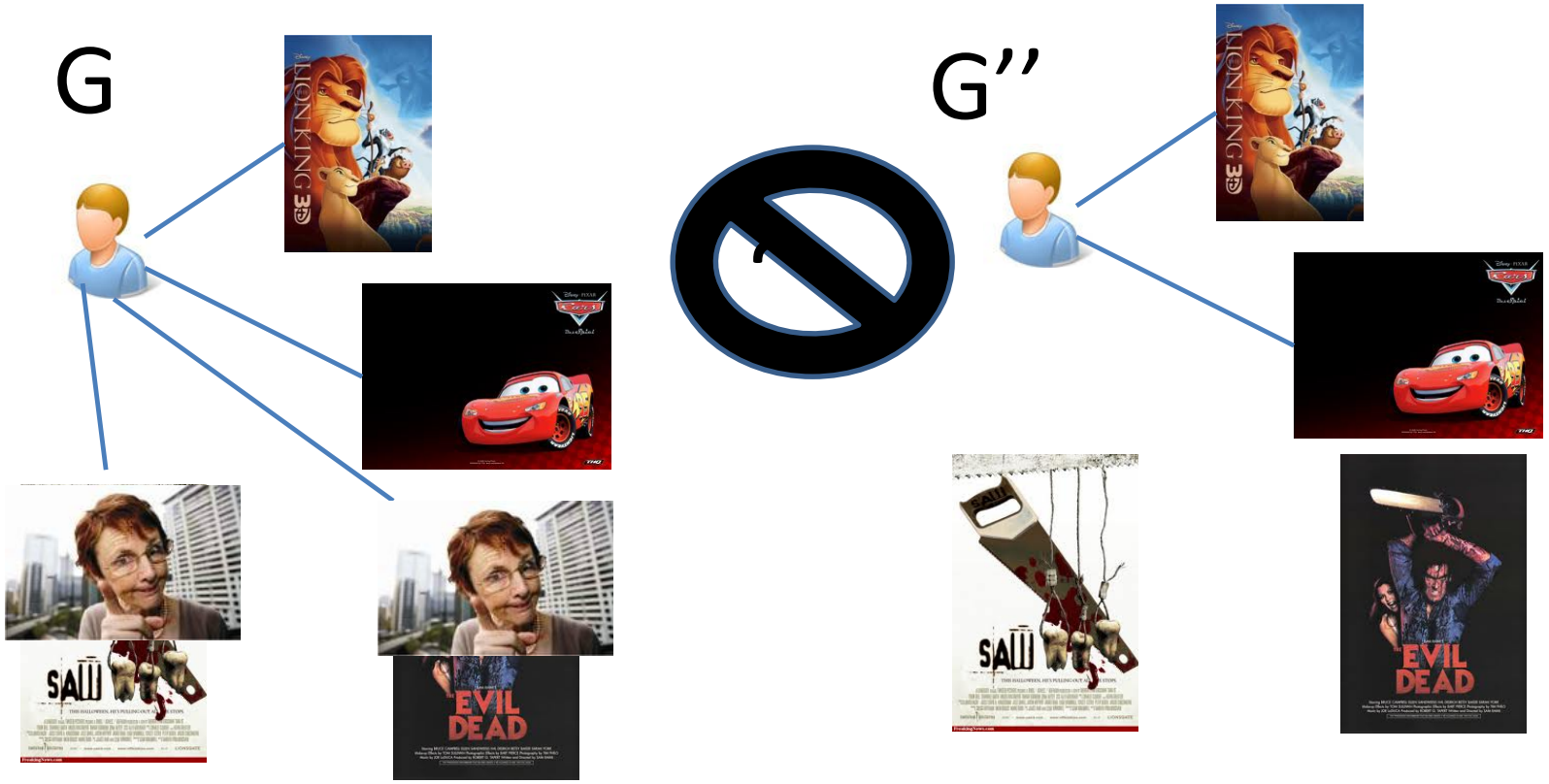
$$\Pr[A(G) \in \text{  }] \leq e^{\varepsilon} \Pr[A(G') \in \text{  }] + \delta$$

Graphs: Edge Adjacency



Johnny’s mom does not learn if he watched Saw from the output A(G).

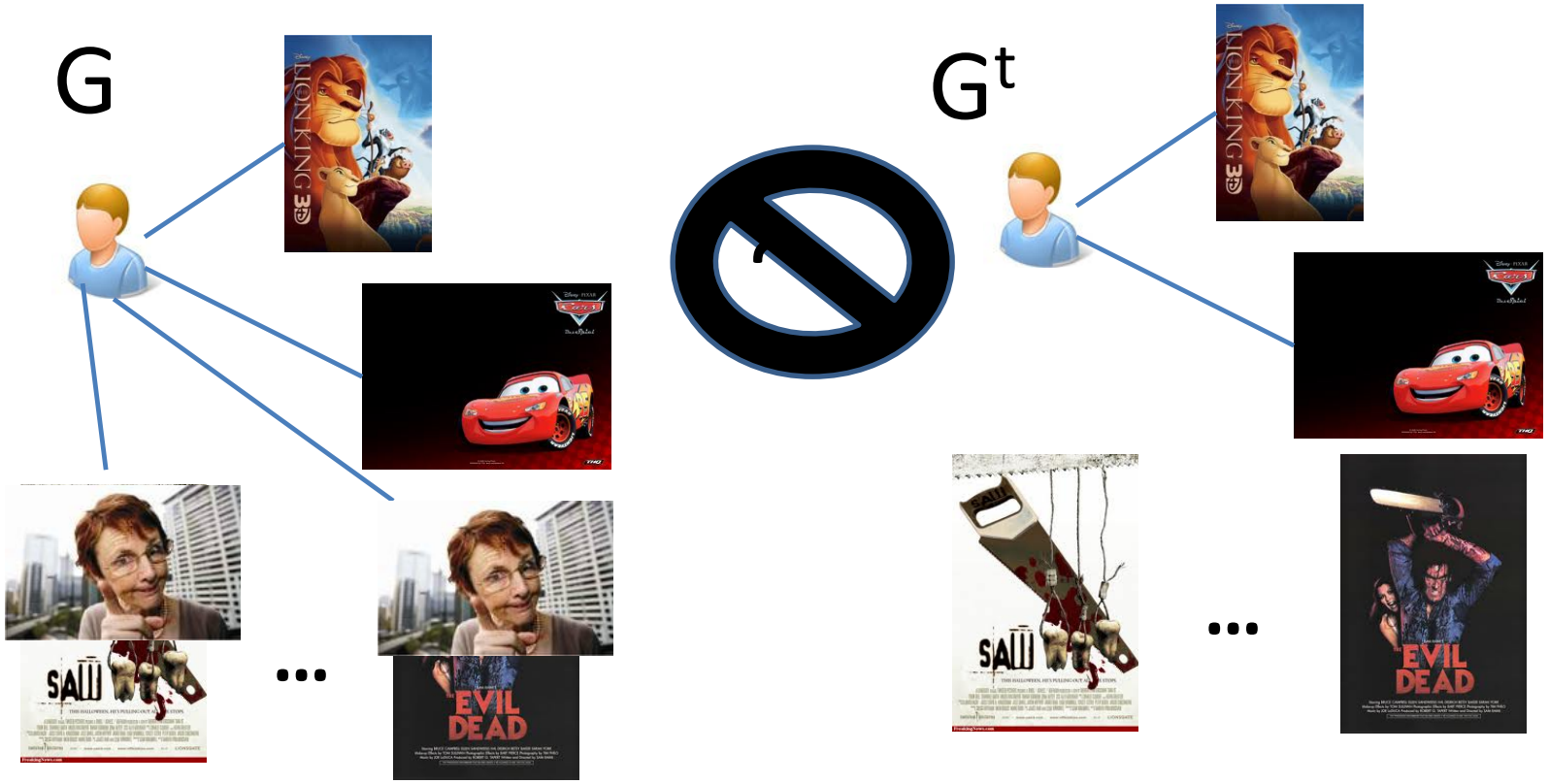
Privacy for Two Edges?



$$\Pr[A(G) \in \text{[image]}] \leq e^{2\varepsilon} \Pr[A(G'') \in \text{[image]}] + 2\delta$$

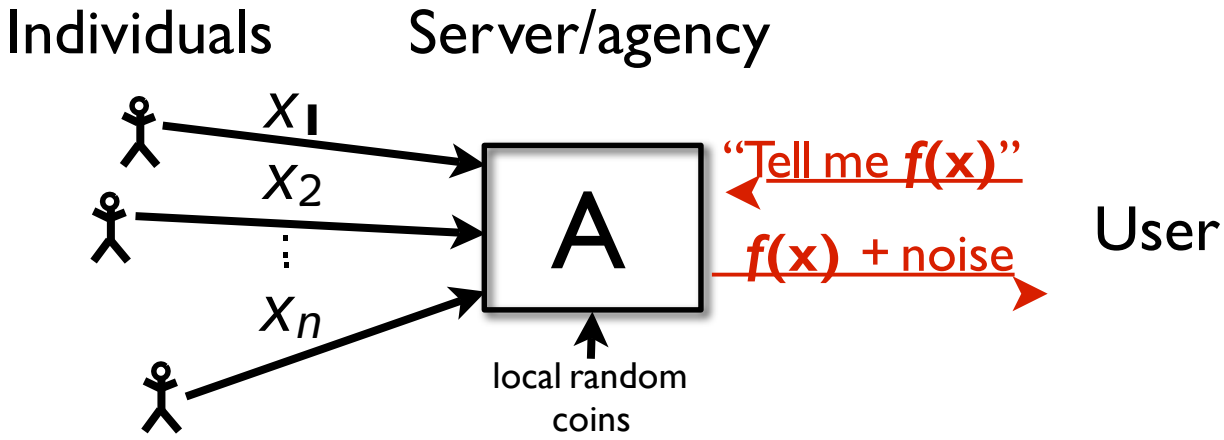
23

Limitations



Johnny's mom may now be able tell if he watches R-rated movies from A(G).

Output Perturbation



- Intuition: $f(x)$ can be released accurately when f is insensitive to individual entries $x_1, x_2, \ldots, x_n$

Global Sensitivity

$$\Delta Q \coloneqq \max_{G \sim G'} |Q(G) - Q(G')|$$

Global Sensitivity

$$\Delta Q \coloneqq \max_{G \sim G'} |Q(G) - Q(G')|$$

- What does G~G’ mean?
- Example: Change one attribute

- Q_1(G) = #users who watched Lion King

- ΔQ_1 = ?

Global Sensitivity

$$\Delta Q \coloneqq \max_{G \sim G'} |Q(G) - Q(G')|$$

- What does G~G’ mean?
- Example: Change one attribute

- Q_2(G) = #users who watched Toy Story

- $\Delta Q_2 = 1$

Global Sensitivity

$$\Delta Q := \max_{G \sim G'} |Q(G) - Q(G')|$$

- What does G~G’ mean?
- Example: Change one attribute

- $Q(G) = Q_1(G) + Q_2(G)$

- ΔQ_2 = ?

Global Sensitivity

$$\Delta Q \coloneqq \max_{G \sim G'} |Q(G) - Q(G')|$$

- What does G~G’ mean?
- Example: Change one attribute

- Q_1(G) = #users who watched Lion King

- ΔQ_1 = ?

Global Sensitivity

$$\Delta Q := \max_{G \sim G'} |Q(G) - Q(G')|$$

- What does G~G’ mean?
- Example: Add/delete one row?

Global Sensitivity

$$\Delta Q := \max_{G \sim G'} |Q(G) - Q(G')|$$

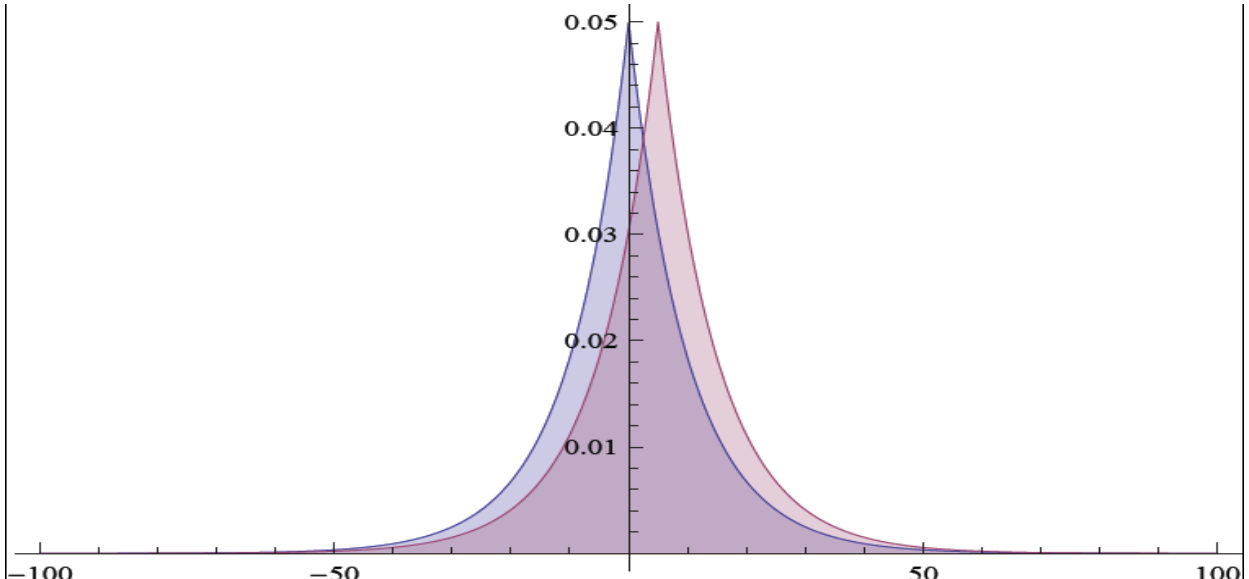
- Example: Add/delete one row?
- $Q(G) = Q_1(G) + Q_2(G)$
- $\Delta Q = ?$

Traditional Differential Privacy Mechanism

Fact: The Laplacian Mechanism:

$$\mathrm{A}(G) = Q(G) + \mathrm{Lap}\left(\frac{\Delta Q}{\varepsilon}\right),$$

satisfies $(\varepsilon, 0)$-differential privacy.

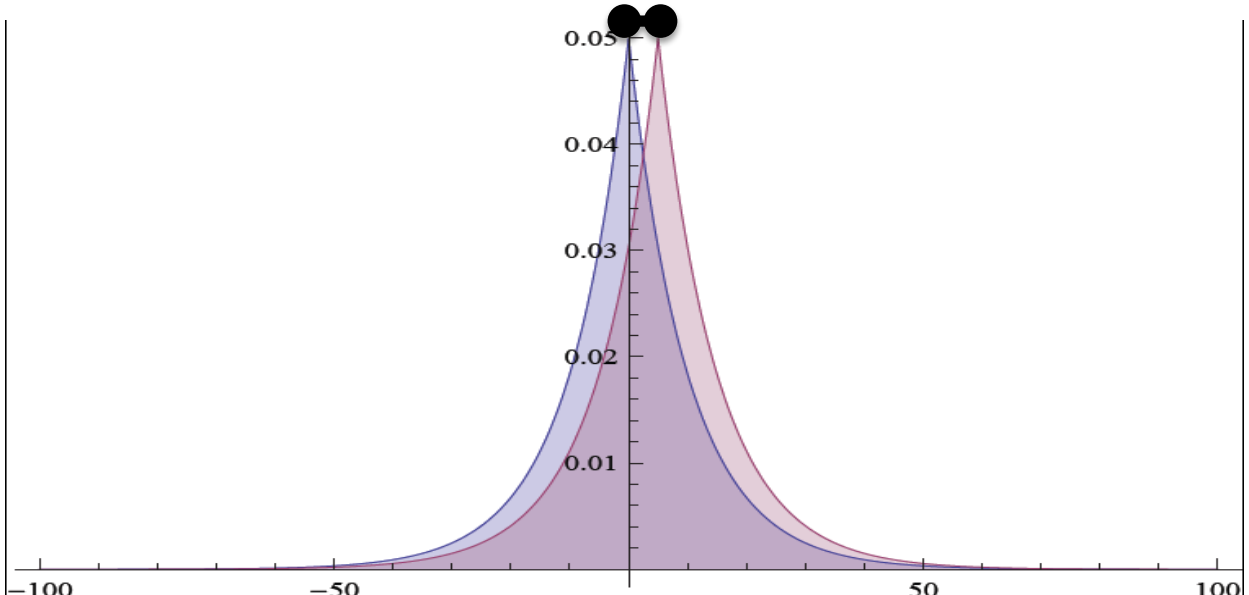


Traditional Differential Privacy Mechanism

$$PDF_G(x) \propto e^{-|x\varepsilon|}$$

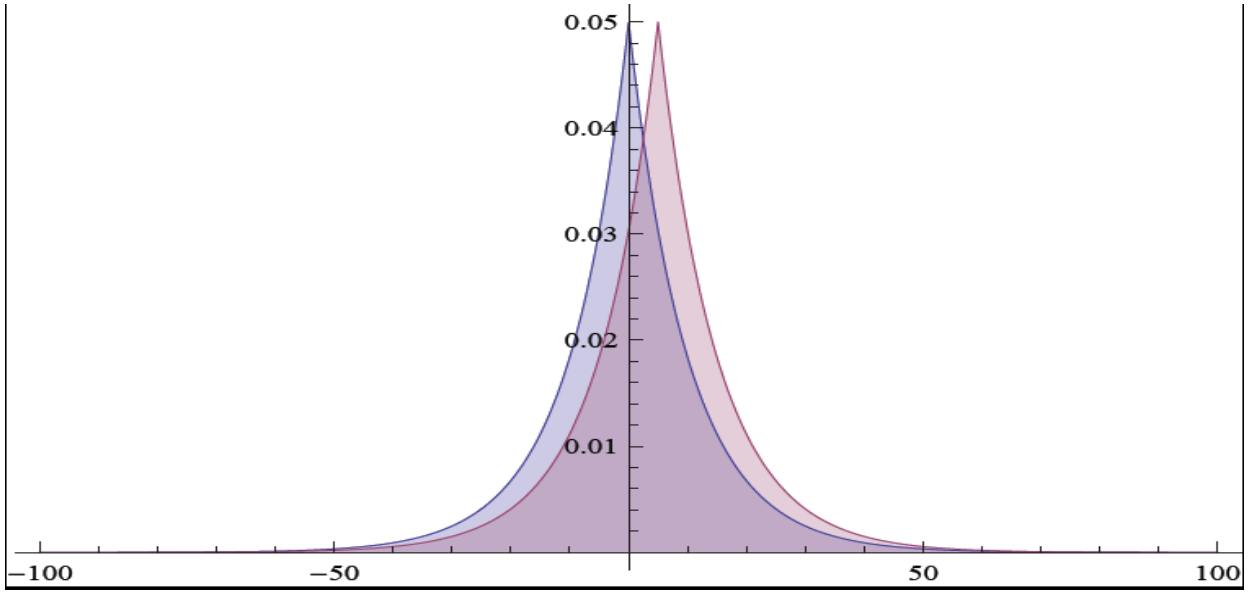
$$PDF_{G'}(x) \propto e^{-|(x-1)\varepsilon|}$$

ΔQ=1

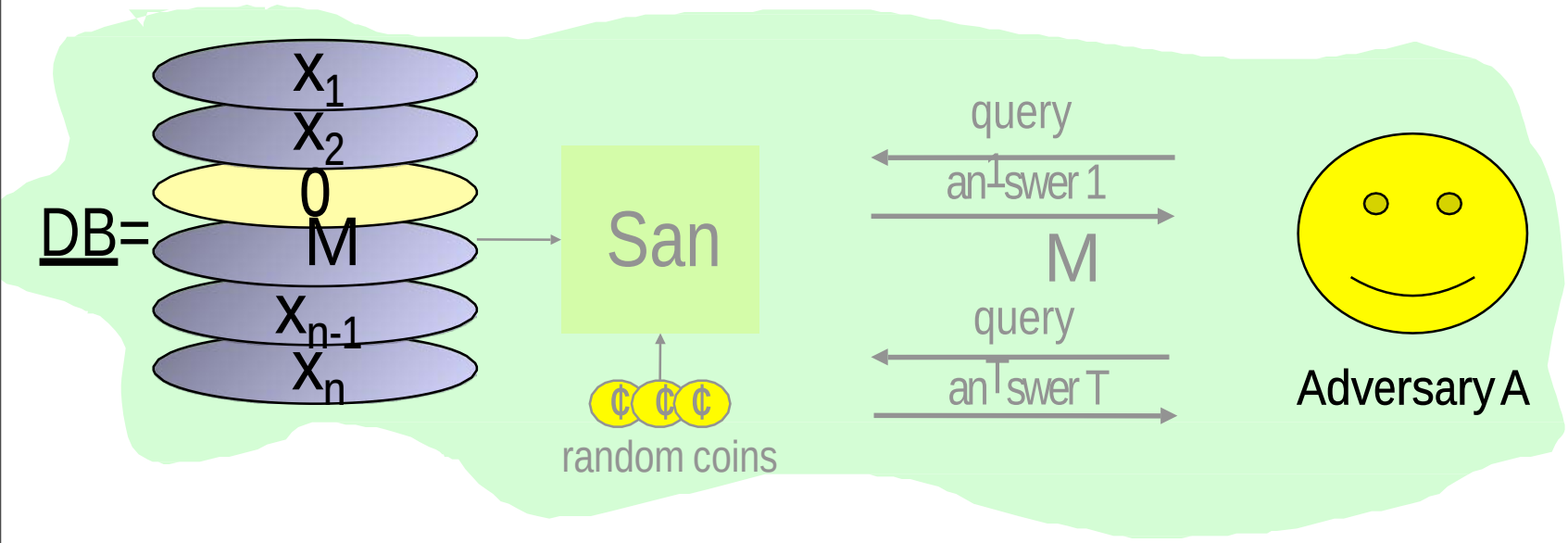


Traditional Differential Privacy Mechanism

$$\forall x, \frac{PDF_G(x)}{PDF_{G'}(x)} = \frac{e^{-|x\varepsilon|}}{e^{-|(x-1)\varepsilon|}} \leq e^{-\varepsilon}$$



Differential Privacy



Examples of low global sensitivity

- Example: $GS_{average} = \frac{1}{n}$ if $x \in [0,1]^n$
  - Add noise Lap($\frac{1}{n}$)
  - Comparison: to estimate a frequency (e.g. proportion of diabetics) in underlying population, get sampling noise $\frac{1}{\sqrt{n}}$
- Many natural functions have low GS, e.g.:
  - Histograms and contingency tables
  - Covariance matrix
  - Distance to a property
  - Functions that can be approximated from a random sample
- [BDMN] Many data-mining and learning algorithms access the data via a sequence of low-sensitivity questions
  - e.g. perceptron, some "EM" algorithms, SQ learning algorithms

Why does this help?

With relatively little noise:

- Averages
- Contingency tables
- Matrix decompositions
- Certain types of clustering
- …

Differential Privacy

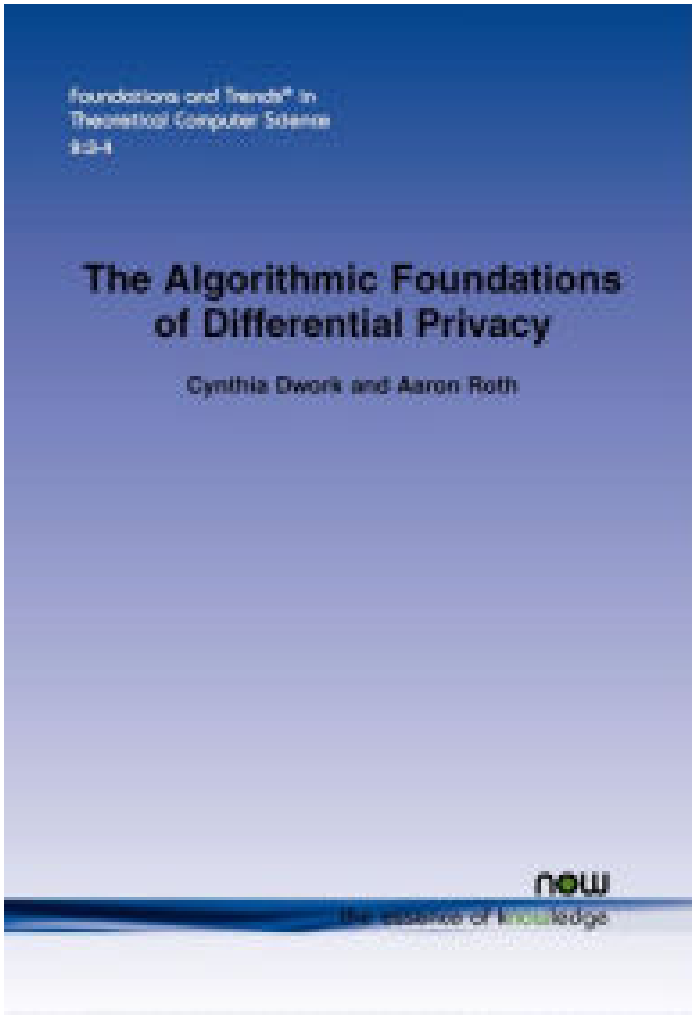
Protocols

- Output perturbation
  (Release f(x) + noise)
  - Sum queries
    - [DiN'03,DwN'04,BDMN'05]
  - "Sensitivity" frameworks
    - [DMNS'06, NRS'07]
- Input perturbation
  ("randomized response")
  - Frequent item sets [EGS'03]
  - (Various learning results)

Lower bounds

- Limits on communication models
  - Noninteractive [DMNS'06]
  - "Local" [NSW'07]
- Limits on accuracy
  - "Many" good answers allow reconstructing database
    - [DiNi'03,DMT'07]
- Necessity of "differential" guarantees [DN]

Resources



BARNES & NOBLE

$99

Free PDF:
https://www.cis.upenn.edu/~aaroth/Papers/privacybook.pdf