Differential Privacy and Applications

Stefanović Tamara   Ghilezan Silvia

University of Novi Sad, Serbia

LAP, September 2021

Outline

1. Differential Privacy
   - Motivation
   - Definition
2. Applications
   - Location Privacy
   - Graph Analysis

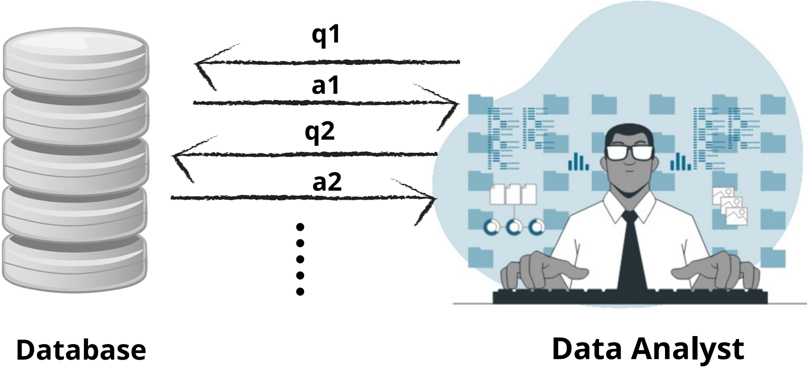

NSA
HACKERS
facebook
verizon
Google
twitter
iCloud
ADVERTISERS
MINE!

Data MINE...

Table of Contents

1. Differential Privacy
   - Motivation
   - Definition
2. Applications
   - Location Privacy
   - Graph Analysis

Statystical Analysis



How to give an appropriate answer to the data analyst while preserving privacy of individuals in the database?

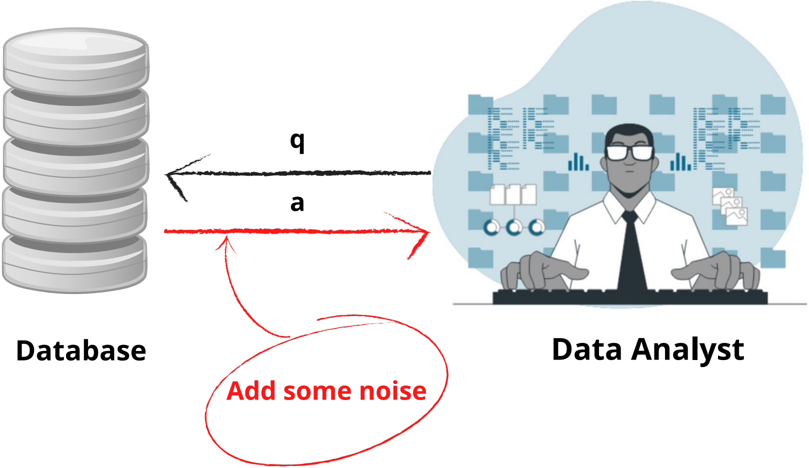


Figure: Adding Noise

- $D \in \mathcal{D}$ - a database/dataset;
- q - a query, function applied on a database;
- $\mathcal{M}$ - a mechanism which for every query q creates a new randomized query by adding noise $\mathcal{M}(D) = q(D) + \textit{noise}$.

Neighbouring Databases

$D \sim D'$ - adjacent/neighbouring datasets (differs in at most one entry)

Definition: ε-differential privacy

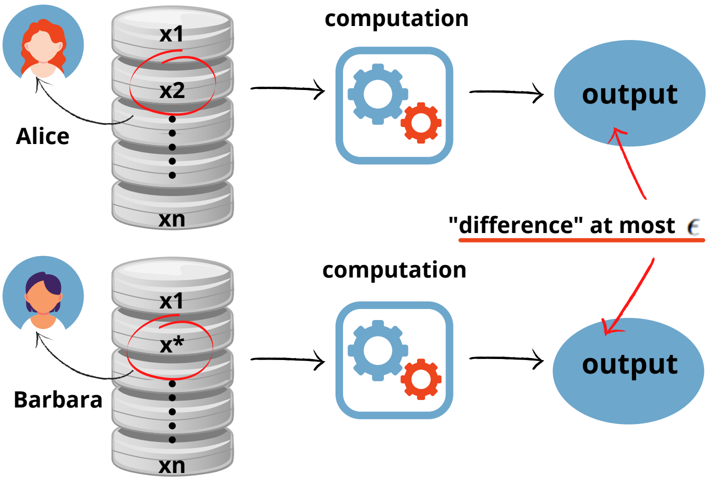
Let $\varepsilon > 0$. A mechanism $\mathcal{M}$ is ε-differentially private iff for every pair of adjacent databases D, D' and for every $S \subseteq range(\mathcal{M})$:

$$Pr[\mathcal{M}(D) \in S] \leq \exp(\varepsilon) Pr[\mathcal{M}(D') \in S],$$

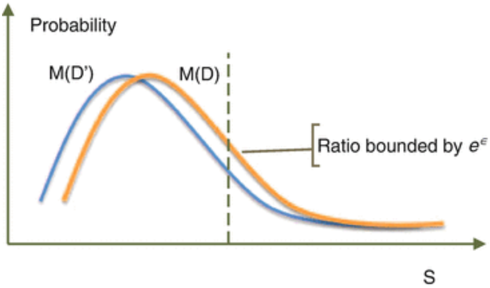
where the probability space is over the coin flips of the mechanism $\mathcal{M}$.

Dwork., C. Differential privacy: A survey of results. In Manindra Agrawal, Dingzhu Du, Zhenhua Duan, and Angsheng Li, editors, Theory and Applications of Models of Computation, pages 1–19, Springer, Berlin, Heidelberg, 2008.

What differential privacy promises?

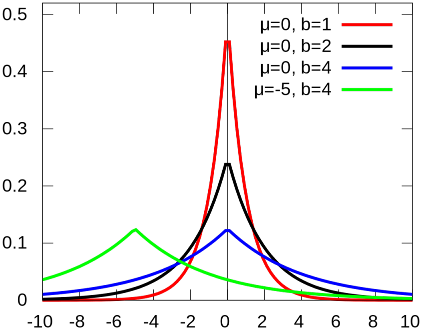


What kind of noise to use?



$$\frac{Pr[\mathcal{M}(D) \in S]}{Pr[\mathcal{M}(D') \in S]} \leq \exp(\varepsilon)$$

Laplace Distribution?



$$f(x;\mu,b) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right)$$

Laplace Mechanism

Definition: Global Sensitivity

The global sensitivity of a query $q : \mathcal{D} \to \mathbb{R}$ is

$$\Delta(q) = \max_{D,D'} ||q(D) - q(D')||_1$$

for all neighbouring D and D'.

Theorem: Laplace Mechanism

For a query q, a mechanism $\mathcal{M}(x) = q(x) + Y$ satisfies ε-differential privacy, where Y is a random variable with Laplace distribution with mean 0 and scales $\dfrac{\Delta(q)}{\varepsilon}$.

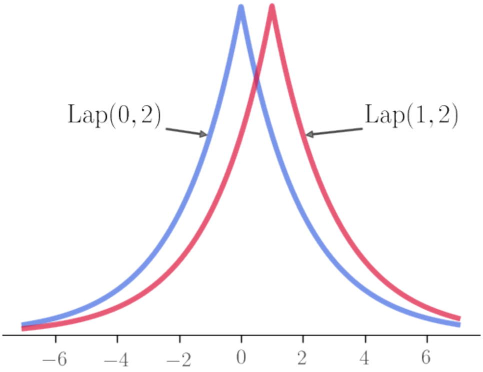


Figure: Laplace mechanism offering 0.5-differential privacy for a query with sensitivity 1

$$\frac{f(x;1,2)}{f(x;0,2)} = \frac{\frac{1}{4}\exp\left(-\frac{|x-1|}{2}\right)}{\frac{1}{4}\exp\left(-\frac{|x|}{2}\right)}$$

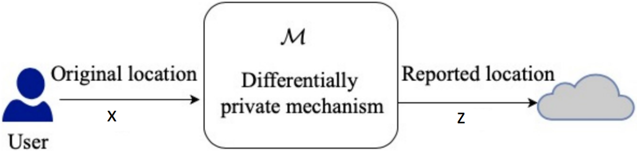
$$= \exp\left(\frac{|x| - |x-1|}{2}\right)$$

$$\leq \exp\left(\frac{1}{2}\right)$$

Table of Contents

1. Differential Privacy
   - Motivation
   - Definition
2. Applications
   - Location Privacy
   - Graph Analysis

Scenario: Raw Location Sharing

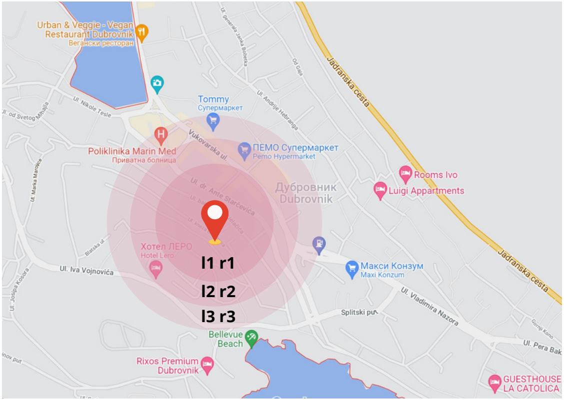


Methods:

- Distance-based Method;
- Obfuscation-based Method;
- Anonymity-based Method.

Distance-based Method

- $\mathcal{X}$ - a set of user's possible locations
- $\mathcal{Z}$ - a set of possible reported locations
- $d_{\mathcal{X}}$ - a distance metrics



Definition: ε-geo-indistinguishability

A mechanism $\mathcal{M}$ satisfies ε-geo-indistinguishability iff for every $r > 0$ and for every pair $x, x' \in \mathcal{X} : d_{\mathcal{X}}(x, x') < r$ and every $\mathcal{S} \subseteq \mathcal{Z}$:

$$Pr[\mathcal{M}(x) \in \mathcal{S}] \leq \exp(\varepsilon r) Pr[\mathcal{M}(x') \in \mathcal{S}]$$

Differential Privacy $\Longleftrightarrow$ Geo-indistinguishability ??

Geo-indistinguishability

$$Pr[\mathcal{M}(x) \in \mathcal{S}] \leq \exp\left(\varepsilon d_{\mathcal{X}}(x, x')\right) Pr[\mathcal{M}(x') \in \mathcal{S}]$$

Differential Privacy

$$Pr[\mathcal{M}(D) \in \mathcal{S}] \leq \exp\left(\varepsilon \cdot 1\right) Pr[\mathcal{M}(D') \in \mathcal{S}]$$

Differential Privacy $\Longleftrightarrow$ Geo-indistinguishability ??

Geo-indistinguishability

$$Pr[\mathcal{M}(x) \in \mathcal{S}] \leq \exp(\varepsilon d_{\mathcal{X}}(x, x'))Pr[\mathcal{M}(x') \in \mathcal{S}]$$

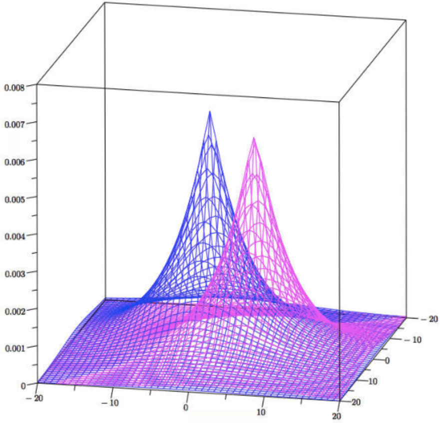
Differential Privacy

$$Pr[\mathcal{M}(D) \in \mathcal{S}] \leq \exp(\varepsilon \cdot 1)Pr[\mathcal{M}(D') \in \mathcal{S}]$$

$1 = d_h(D, D')$ for adjacent databases

d_h- the Hamming distance (number of records at which corresponding databases differ)

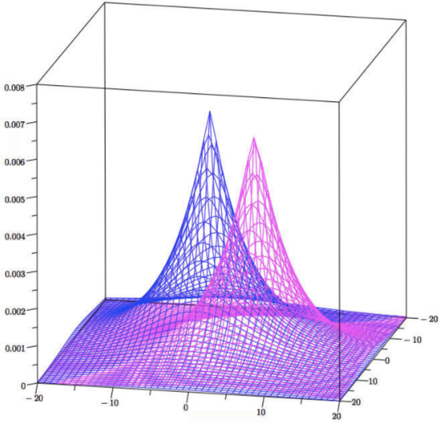
What kind of noise to use?



$$f(x;\mu,b) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right)$$

Laplace distribution

What kind of noise to use?



$$f(x;\mu,b) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right)$$

Laplace distribution

$$D_\varepsilon(\mu) = \frac{\varepsilon^2}{2\pi}\exp\left(-\varepsilon d(x,\mu)\right)$$

Planar Laplace Distribution

Datasets as graphs

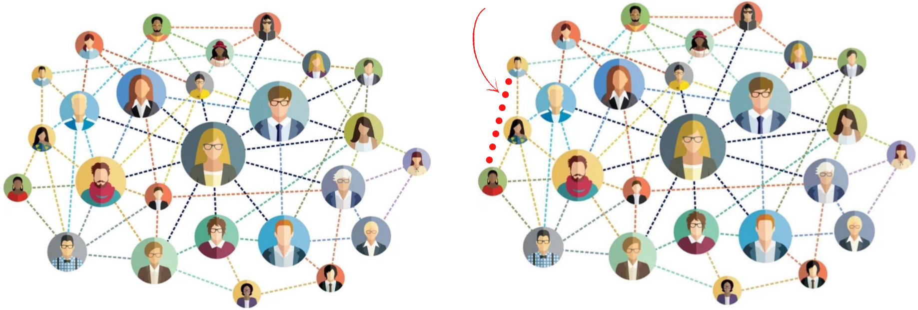
Many datasets can be represented as graphs:

- friendships in online social networks;
- financial transactions;
- e-mail communacations and so on.

Social Graph $G(V, E)$

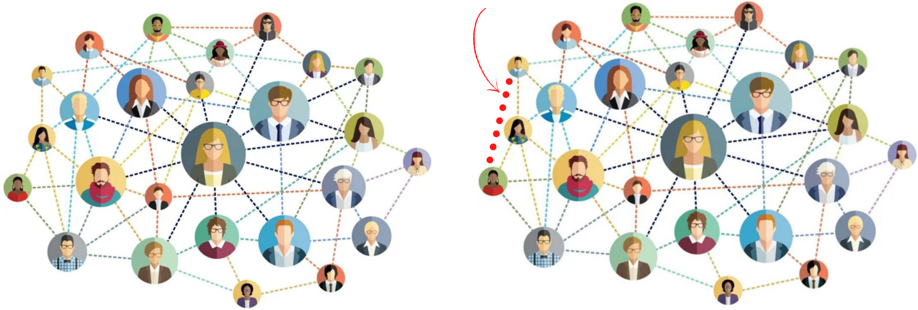
- V-set of vertices or nodes;
- E-set of edges.

Edge Differential Privacy



Two graphs are neighbors if they differ in one edge.

Edge Differential Privacy



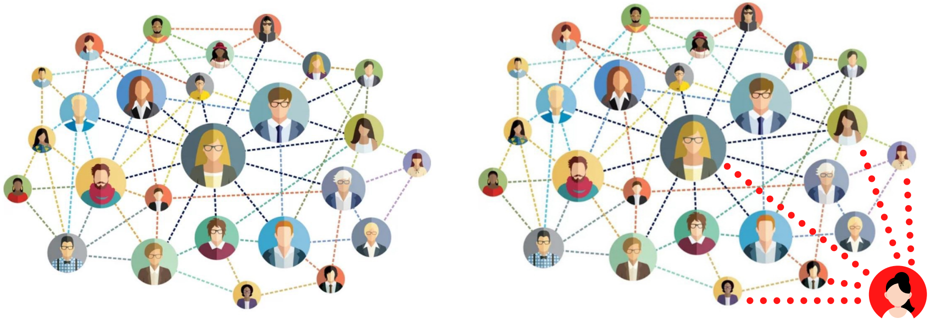
Two graphs are neighbors if they differ in one edge.

$$d_{edge}(G, G') = 1$$

Node Differential Privacy

Two graphs are neighbors if one can be obtained from the other by deleting a node and its adjacent edges (or by adding a node).

Node Differential Privacy



Two graphs are neighbors if one can be obtained from the other by deleting a node and its adjacent edges (or by adding a node).

$$d_{node}(G, G') = 1$$

Definition: Node (edge)-differential privacy

Let $\varepsilon > 0$. A mechanism $\mathcal{M}$ is ε-node (edge)-differentially private iff for every pair of neighbouring graphs G, G' and for every $S \subseteq range(\mathcal{M})$:

$$Pr[\mathcal{M}(G) \in S] \leq \exp(\varepsilon) Pr[\mathcal{M}(G') \in S].$$

Kasiviswanathan S.P., Nissim K., Raskhodnikova S., Smith A. : Analyzing Graphs with Node Differential Privacy. In: Sahai A. (eds) Theory of Cryptography. TCC 2013. Lecture Notes in Computer Science, vol 7785. Springer, Berlin, Heidelberg, 2013.

Edge differential privacy

$$Pr[\mathcal{M}(G) \in S] \leq \exp(\varepsilon \cdot 1) Pr[\mathcal{M}(G') \in S]$$

$$d_{edge}(G, G') = 1$$

Node differential privacy

$$Pr[\mathcal{M}(G) \in S] \leq \exp(\varepsilon \cdot 1) Pr[\mathcal{M}(G') \in S]$$

$$d_{node}(G, G') = 1$$

Different applications - different distance metrics!

Concluding Remarks

- Differential privacy - motivation and definition;
- Applications of differential privacy.

Ongoing Work:

- Impact of new metrics on differential privacy;
- Application of differential privacy in blockchain technology.

Publications

- Stefanović, T., Ghilezan, S. : An Overview of Mathematical Models for Data Privacy, LAP2020-8th Conference on Logic and Applications, September 21-25, 2020, Dubrovnik, Croatia
- Stefanović T., Ghilezan S. (2021) Preserving Privacy in Caller ID Applications. In: Friedewald M., Schiffner S., Krenn S. (eds) Privacy and Identity Management. Privacy and Identity 2020. IFIP Advances in Information and Communication Technology, vol 619. Springer, Cham.
- Ghilezan, S., Stefanović, T. : Privacy-preserving contact tracing, Mathematics for Human Flourishing in the Time of COVID-19 and Post COVID-19, Niš, October, 2020, Niš, Serbia
- Ghilezan, S., Kašterović, S., Liquori, L., Marinković, B., Ognjanović, Z., Stefanović, T. : Federating Digital Contact Tracing using Structured Overlay Networks. 2021. hal-03127890v3