Karlsruhe Institute of Technology

P.E.Ts: Trajectory privacy

Patricia Guerra-Balboa

July 11, 2022

Mechanism Achieving Differential Privacy

Mechanism Achieving Differential Privacy

Figure 8: Static context

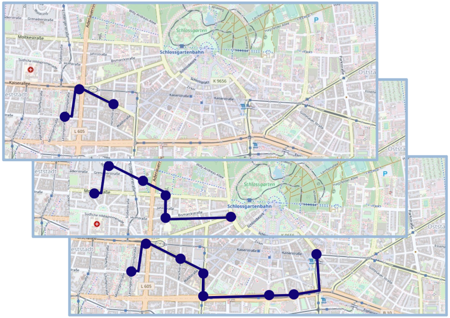


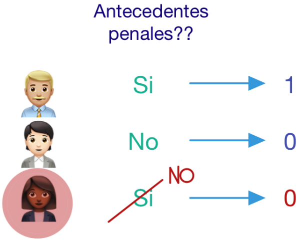
Figure 9: Dynamic or streaming

Algorithms Achieving Differential Privacy

ℓ_1-sensitivity

The ℓ_1-sensitivity of a function $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^n$ is:

$$\Delta(f) := \max_{\|D, D'\|_1 = 1} \|f(D) - f(D')\|_1$$



$\Delta f = 1$

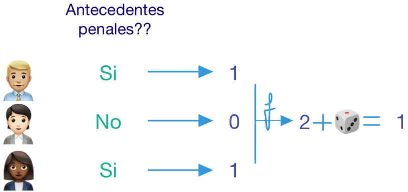
Algorithms Achieving Differential Privacy
Laplace Mechanism

Laplace Mechanism

Given any function $f : \mathbb{N}^{|\mathcal{X}|} \rightarrow \mathbb{R}^n$ the Laplace mechanism is defined as:

$$ML(D, f(\cdot), \epsilon) = f(D) + (Y_1, \ldots, Y_n)$$

where Y_i are i.i.d. random variables drawn from $Lap(\frac{\Delta f}{\epsilon})$.



Algorithms Achieving Differential Privacy
Exponential Mechanism

Abstract range

$$u: \mathbb{N}^{|\mathcal{X}|} \times \mathcal{R} \longrightarrow \mathbb{R}$$

Score function

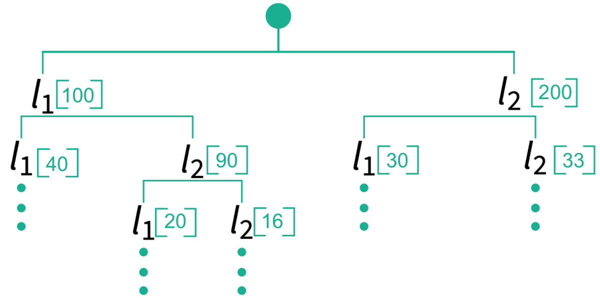
H	SMOKES	HAS_CANCER	DRINKS_SODA
	•	•	
		•	•
			•
			•

$$\Delta(u) := \max_{r \in \mathcal{R} \| D, D' \|_1 = 1} \| u(D, r) - u(D', r) \|_1$$

$M_E(D, u, \mathcal{R})$ selects and outputs an element $r \in \mathcal{R}$ with probability proportional to $exp(\frac{\epsilon u(D,r)}{2\Delta(u)})$. 2u

Mechanism Achieving Differential Privacy
Static Context

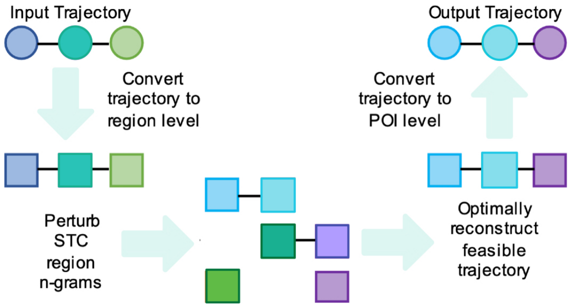
Noisy counts | Clustering | Perturbing semantic trajectories



Mechanism Achieving Differential Privacy

Static Context

Noisy counts | Clustering | **Perturbing semantic trajectories**

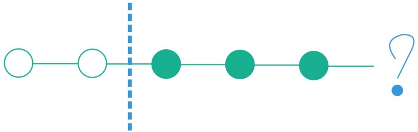


Mechanism Achieving Differential Privacy

Discrete-time Markov process of order l

Given a sequence of random variables $X_1, X_2, X_3, \ldots$. We say that they follow a Markov process of order l iff probability of moving to the next state depends only on the l previous states :

$$\Pr(X_{n+1} = x \mid X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \Pr(X_{n+1} = x \mid X_{n-l} = x_{n-l}, \ldots, X_n = x_n)$$



Mechanism Achieving Differential Privacy
Synthetic Data

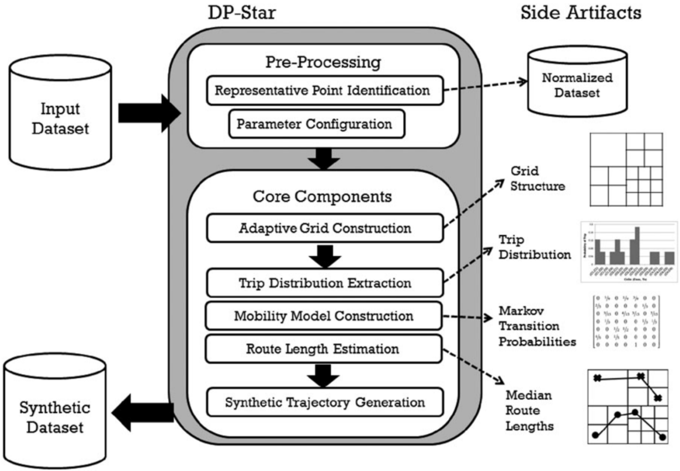
Mechanisms	n-grams	DPT	DP-STAR
Time variable	✗	✗	✗
Prefix tree	✓	✓	✗
Neighboring databases	$D_1 = D_2 \cup \{T\}$	$D_1 = D_2 \cup \{PT\}$	$D_1 = D_2 \cup \{T\}$
Main mechanism	Laplacian mechanism	Laplacian mechanism	Laplacian and exponential mechanism
Sensitivity bound	l_{max} truncation	Normalization by #transitions in PT	Normalization by $\|T\|$
Markov process	order $n-1$	order $l < k$	order 1

Mechanism Achieving Differential Privacy

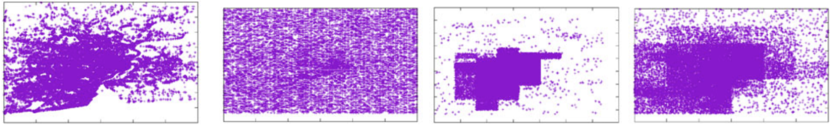
Synthetic Data

Mechanisms	n-grams	DPT	DP-STAR
Time variable	✗	✗	✗
Prefix tree	✓	✓	✗
Neighboring databases	$D_1 = D_2 \cup \{T\}$	$D_1 = D_2 \cup \{PT\}$	$D_1 = D_2 \cup \{T\}$
Main mechanism	Laplacian mechanism	Laplacian mechanism	Laplacian and exponential mechanism
Sensitivity bound	l_{max} truncation	Normalization by $\#$transitions in PT	Normalization by $\lvert T\rvert$
Markov process	order $n-1$	order $l < k$	order 1

Mechanism Achieving Differential Privacy
Synthetic Data



Mechanism Achieving Differential Privacy
Synthetic Data



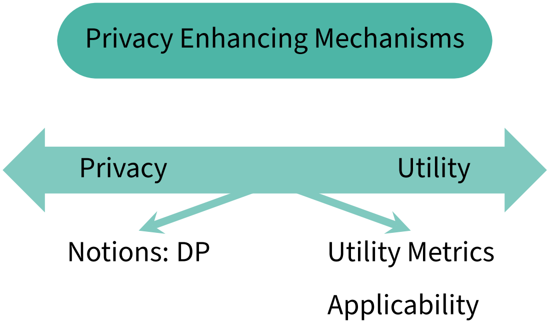
(a) Original distribution (b) DPT (c) ngram (d) DP-Star

$T = \langle l_1, \ldots, l_n \rangle$ where each l_i is a location

TIME?

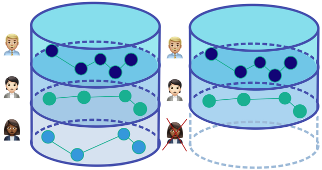
Mechanisms Analysis

Mechanisms Analysis



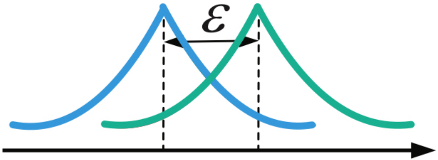
Mechanisms Analysis

Privacy Analysis



Privacy Loss (by observing r)

$$\mathcal{L}^{r}_{M(D)||M(D')} = ln\left(\frac{\mathbb{P}(M(D) = r)}{\mathbb{P}(M(D') = r)}\right) \leq \epsilon$$

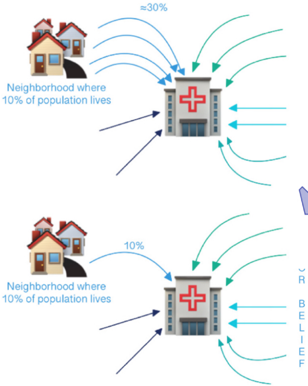


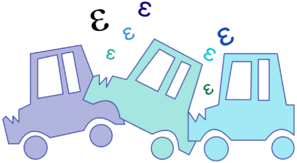
Mechanisms Analysis

Limitations on Differential Privacy

Correlation

Bayesian inference





Infinity streaming

Mechanisms Analysis

Limitations on Differential Privacy

- $$\mathbb{P}(r|_{X_1=x_1}) = \sum_{y\in\{0,1\}} \mathbb{P}(r|x_1,y)\cdot\mathbb{P}(y|x_1)$$

  X_1, X_2 independent $\Rightarrow \mathbb{P}(y/x_1) = \mathbb{P}(y)$

  $$\Rightarrow \mathbb{P}(r/_{X_1=x_1}) = \sum_{y\in\{0,1\}} \mathbb{P}(r|x_1,y)\cdot\mathbb{P}(y) \leq$$
  $$\leq \sum_{y\in\{0,1\}} e^{\varepsilon}\mathbb{P}(r|\tilde{x}_1,y)\cdot\mathbb{P}(y) \leq$$
  $$\leq e^{\varepsilon}\,\mathbb{P}(r|_{\tilde{x}_1}).$$

- $$\mathbb{P}(r|_{X_1=x_1}) = \sum_{y\in\{0,1\}} \mathbb{P}(r|x_1,y)\cdot\mathbb{P}(y|x_1)$$

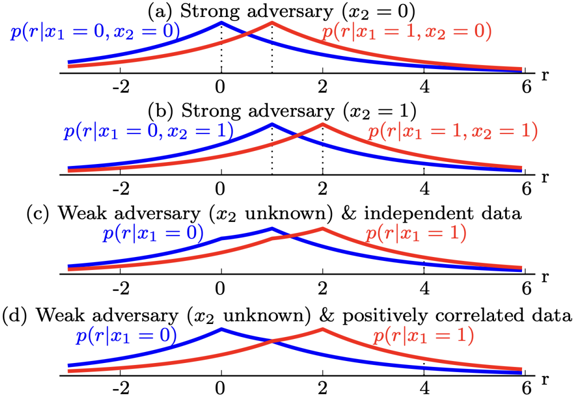
  $\left.\mathbb{P}(X_1=X_2)=0'9\right]$ correlation!!

  $$\Rightarrow \mathbb{P}(r/_{X_1=x_1}) = \sum_{y\in\{0,1\}} \mathbb{P}(r|x_1,y)\cdot\mathbb{P}(y|x_1) \approx$$
  $$\approx 0'8\cdot\mathbb{P}(r|_{x_1,x_1}) + \mu_{\approx 0}$$

  distance $=2$!! not neighbouring $\Downarrow$ $\cancel{\leq}\,\varepsilon$

  $$\mathbb{P}(r/_{X=\tilde{x}_1}) \approx 0'8\,\mathbb{P}(r|_{\tilde{x}_1,\tilde{x}_1}) + \tilde{\mu}_{\approx 0}$$

Mechanisms Analysis
Limitations on Differential Privacy



Mechanisms Analysis

Classification of Utility Metrics

Utility metrics

- **Total preservation of data**
  - *Location preservation, number of suppressed points, ...*
- **Close preservation of data**
  - *Use of similarity measures, preservation range query, discernability...*
- **Preservation of semantic information**
  - *Most visited places, frequent sequential patterns, trajectory length preservation...*
  - *Query error distortion function:*

$$\text{error}(q) = \frac{|q(D) - q(D')|}{\max\{q(D), b\}}.$$

- **Assurance of realism**
  - *Reachability, geo-spatial consistency...*

Mechanisms Analysis

Classification of Utility Metrics

Utility metrics

- **Total preservation of data**
  - *Location preservation, number of suppressed points, ...*
- **Close preservation of data**
  - *Use of similarity measures, preservation range query, discernability...*
- **Preservation of semantic information**
  - *Most visited places, frequent sequential patterns, trajectory length preservation...*
  - *Query error distortion function:*

$$\mathrm{error}(q) = \frac{|q(D) - q(D')|}{\max\{q(D), b\}}.$$

- **Assurance of realism**
  - *Reachability, geo-spatial consistency...*

Mechanisms Analysis
Similarity Measures

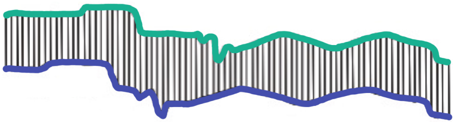
	Euclidean distance	Hausdorff & Fréchet	DTW	TWED	LCSS	EDR
Can compare different lengths	✗	✓	✓	✓	✓	✓
Considers time & allows for local time shifting	✗	✗	✓	✓	✓	✓
Is robust to noise	✗	✗	✗	✗	✓	✓
Is a metric	✓	✓	✗	✓	✗	✗

Table 3: DTW: Dynamic time warping, TWED: Time warping edit distance, LCSS: Longest common subsequence, EDR: Edit distance on real sequences

Mechanisms Analysis
Euclidean distance

	Euclidean distance
Can compare different lengths	✗
Considers time & allows for local time shifting	✗
Is robust to noise	✗
Is a metric	✓

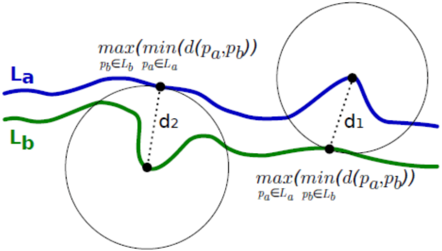
$$\mathrm{Eu}(T, T') = \sqrt{\sum_{i=1}^{n} d((x_i, x_i'), (y_i, y_i'))^2}$$



Mechanisms Analysis

Hausdorff & Fréchet distances

	Hausdorff & Fréchet
Can compare different lengths	✓
Considers time & allows for local time shifting	✗
Is robust to noise	✗
Is a metric	✓

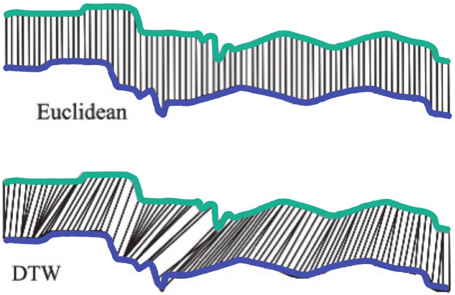


Mechanisms Analysis

Dynamic time warping (DTW) & variations

	DTW	TWED
Can compare different lengths	✓	✓
Considers time & allows for local time shifting	✓	✓
Is robust to noise	✗	✗
Is a metric	✗	✓

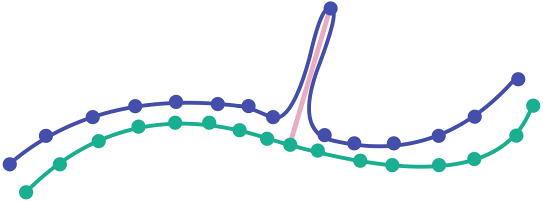
TWED: Time warping edit distance



Mechanisms Analysis

Longest common subsequence (LCSS)

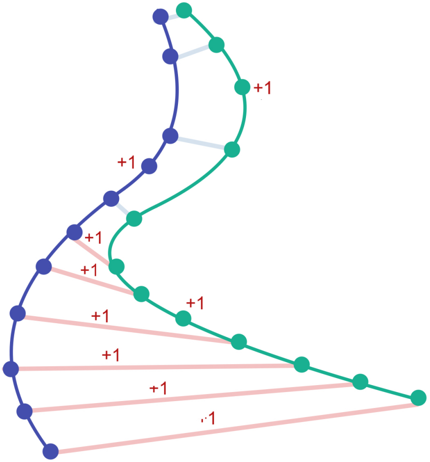
	LCSS
Can compare different lengths	✓
Considers time & allows for local time shifting	✓
Is robust to noise	✓
Is a metric	✗



Mechanisms Analysis

Edit distance on real sequences (EDR)

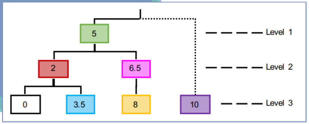
	EDR
Can compare different lengths	✓
Considers time & allows for local time shifting	✓
Is robust to noise	✓
Is a metric	✗



Mechanisms Analysis

Similarity Measures Divisions

		Dimension-wise		
		Spatial	Temporal	Categorical
	DTW, LCSS, EDR...	Euclidean, Hausdorff, Fréchet distances...	Global Time difference	Hamming distance-like function...
Point-wise	Distance functions	Euclidean distance, Haversine formula, road maps...	Time difference	Hierarchical tree...



Mechanisms Analysis

Limitations on Utility

Inherent properties of trajectory data

Sparseness

leads to
Unavoidable data lost
&
high sensitivities

High dimensionality

Problems of current proposals

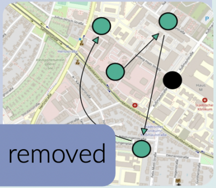
Impossible trajectories

Small universe of locations

Utility metrics are not representative

Ignoring the temporal dimension

Weird trajectory patterns



Stops are removed

Conclusions and Future Research

Conclusions and Future Research