Privacy for co-trajectories

Joel Dahne

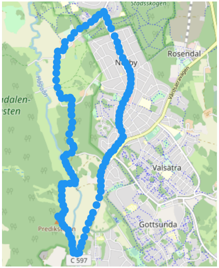
Uppsala university

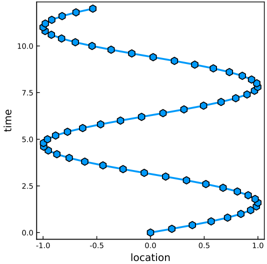
23 April, 2019

Trajectories

Definition (Trajectory)

A trajectory, r, is given by a set of datapoints where each datapoint, d, is given by a location and a time, $d = (x, t)$. We denote the set of all trajectories by $\mathcal{T}$.

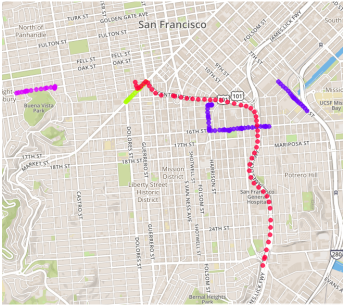


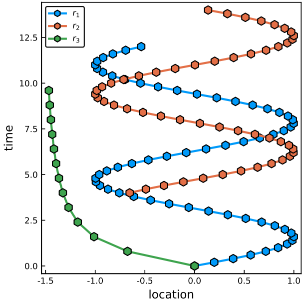


Co-Trajectories

Definition (Co-Trajectory)

A co-trajectory, R, is a collection of trajectories, $R = \{r_i\}_{i=1}^{N}$. The set of all co-trajectories is denoted by $\mathcal{R}$.





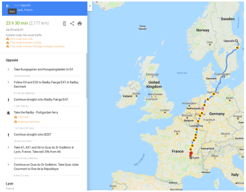
Decision problems

City planning

Selling ads

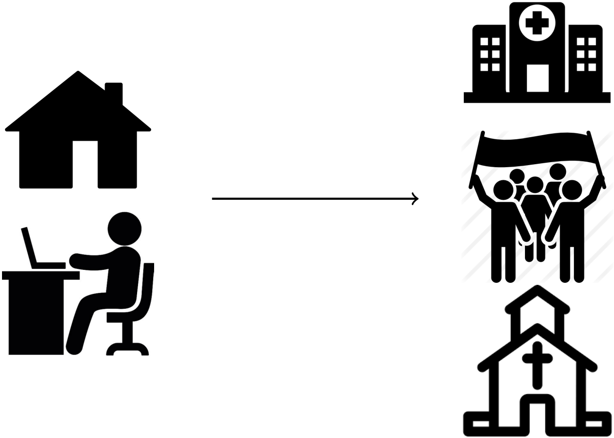
Google Ads

Driving directions

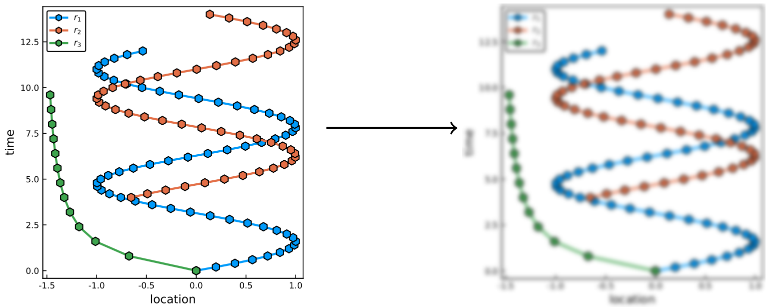


Trading oil

Re-Identification



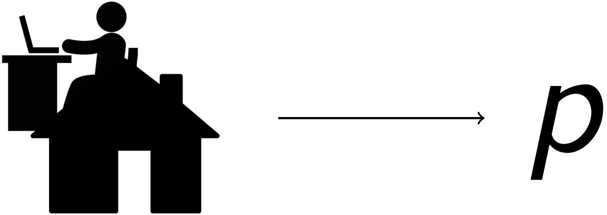
Improving privacy



Measuring privacy

Definition (Trajectory predicate)

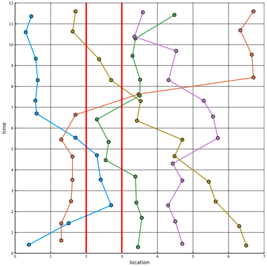
A predicate on trajectories is a function $p : \mathcal{T} \to \{\text{true}, \text{false}\}$. It returns true for trajectories satisfying the predicate and false for those that don't.



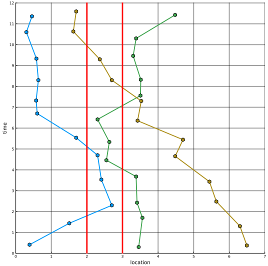
Measuring privacy

Re-identification with predicates

For a co-trajectory R and a predicate p let $p(R)$ be the set of trajectories in R satisfying p.



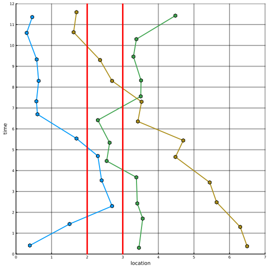
$$+ \quad p \quad \rightarrow$$



Measuring privacy

Definition (k-anonymity)

The k-anonymity for a co-trajectory R with respect to a predicate p is given by the number of trajectories satisfying the predicate, $|p(R)|$.
For a family of predicates, $\{p_i\}_{i \in I}$, it is given by $\min_{i \in I} |p_i(R)|$.



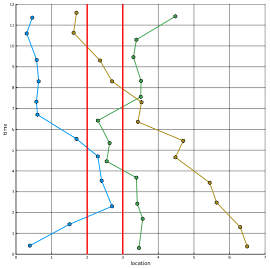
$$\longrightarrow \quad k = 3$$

Measuring privacy

Diversity

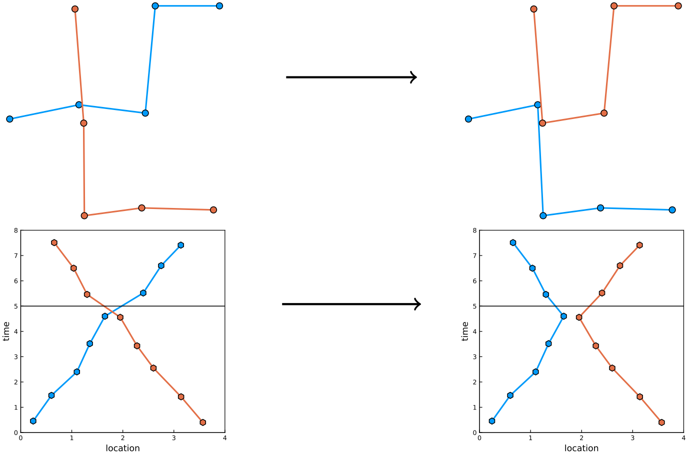
The diversity of a co-trajectory R with respect to a predicate p is given by the diversity of $p(R)$.

Problem: we have no clear way to measure diversity



$$\longrightarrow \quad \text{Diversity} = ?$$

Swap trajectories that meet



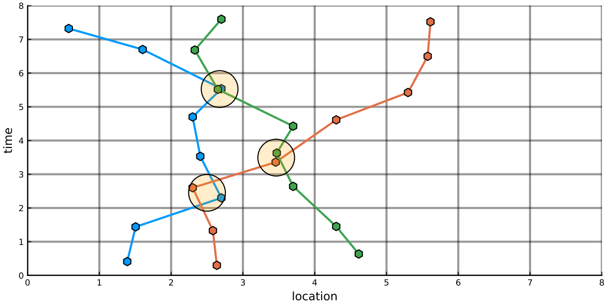
[Salas, Megías, Torra]

SwapMob

$$d = (x, t) \in \mathbb{L} \times \mathbb{R}_+$$

Definition (Similarity for datapoints)

Given an arbitrary partitioning of $\mathbb{L}$ and a partitioning of $\mathbb{R}_+$ into uniform intervals we get a partitioning of $\mathbb{L} \times \mathbb{R}_+$. Two datapoints, d_1, d_2, are considered similar, $d_1 \approx d_2$, if they lie in the same partition. Gives an equivalence relation.



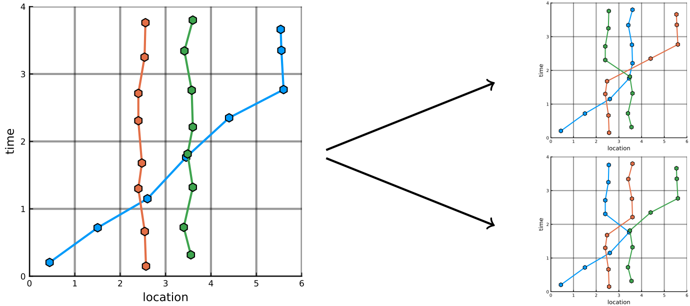
SwapMob

Proposition (1)

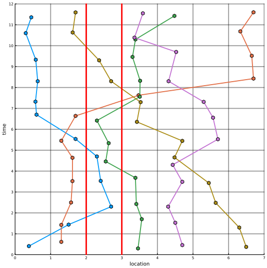
Ordering of swaps at different times does not affect the outcome.

Proposition (2)

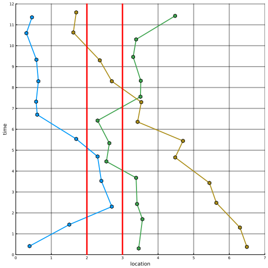
Ordering of swaps at the same time can affect the outcome.

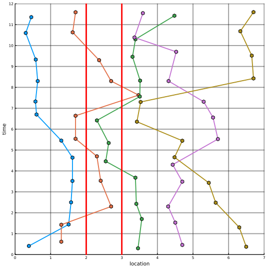


Analysing SwapMob

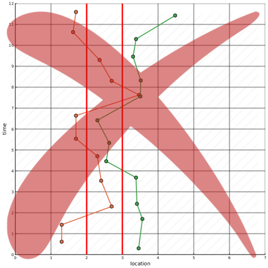


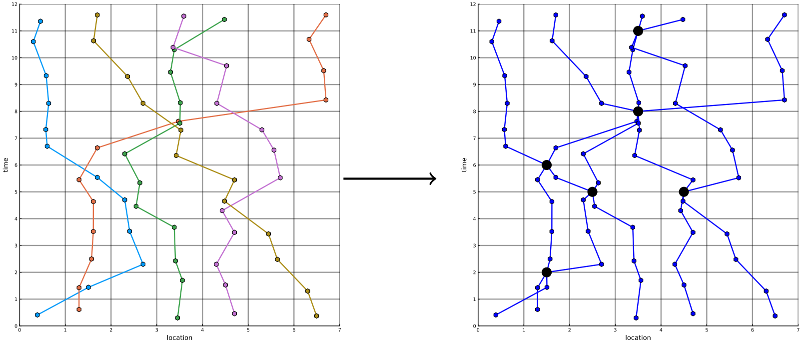
$$+ \quad p \quad \rightarrow$$





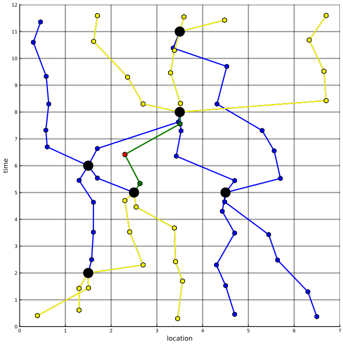
$$+ \quad p \quad \rightarrow$$

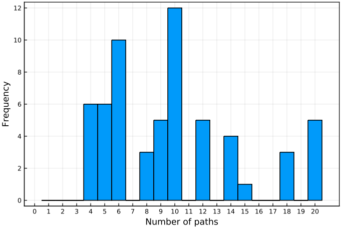




time
location
0
1
2
3
4
5
6
7
8
9
10
11
12

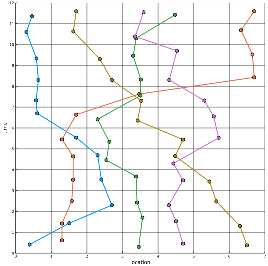
$$\text{Family: } \{p_d\}_{\{d \in r : r \in R\}}$$

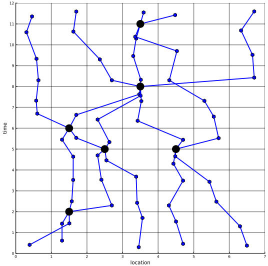




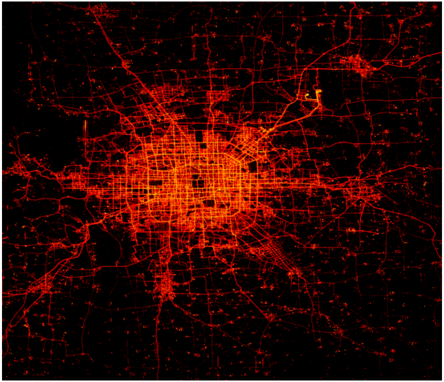
Family: $\{p_i\}_{1 \leq i \leq 5}$ start and end

$$|p_1(P)| = 2,\ |p_2(P)| = 3,\ |p_3(P)| = 2,\ |p_4(P)| = 2,\ |p_5(P)| = 1.$$





T-drive dataset



[Yuan, Zheng, Zhang, Xie, Xie, Sun, Huang]

10,357 Taxis in Beijing

February 2 to February 8, 2008

15 million datapoints

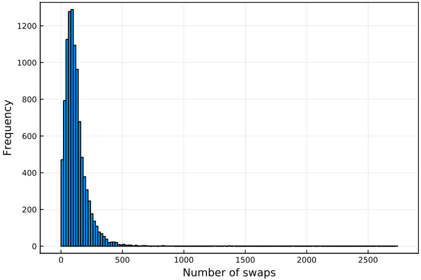
Average sampling 180 seconds

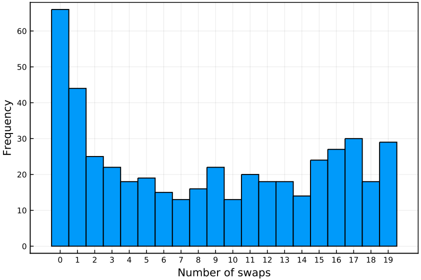
Average distance 620 meters

Number of swaps

Total: 540,812

Average: 119

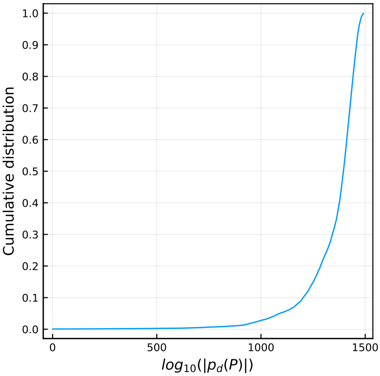




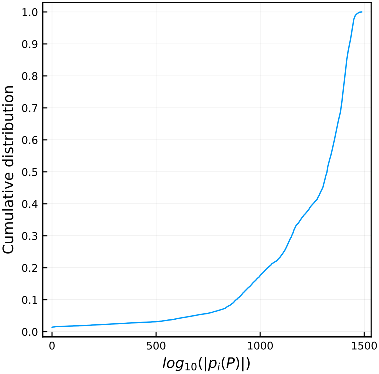
Number of paths in the graph: $6.7 \cdot 10^{1491}$

T-drive predicates

Family: $\{p_d\}_{\{d \in r : r \in R\}}$

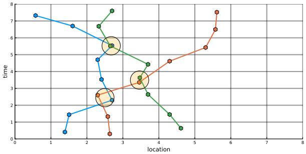


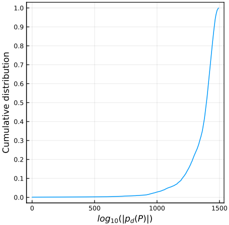
Family: $\{p_i\}_{i \in R}$




Gerland
Stade
de Gerland
Parc de
Gerland
A 7
le Rhône
Avenue Tony Garnier

Tack!


time
location


Cumulative distribution
$log_{10}(|p_d(P)|)$