Decentralized privacy for modeling human mobility

Hamish Gibbs

May 31, 2024

Overview

Impact of federated data with local differential privacy for human mobility modeling

Hamish Gibbs[1], Mirco Musolesi [2], James Cheshire [1], Rosalind M. Eggo [3]

[1]*UCL Department of Geography*

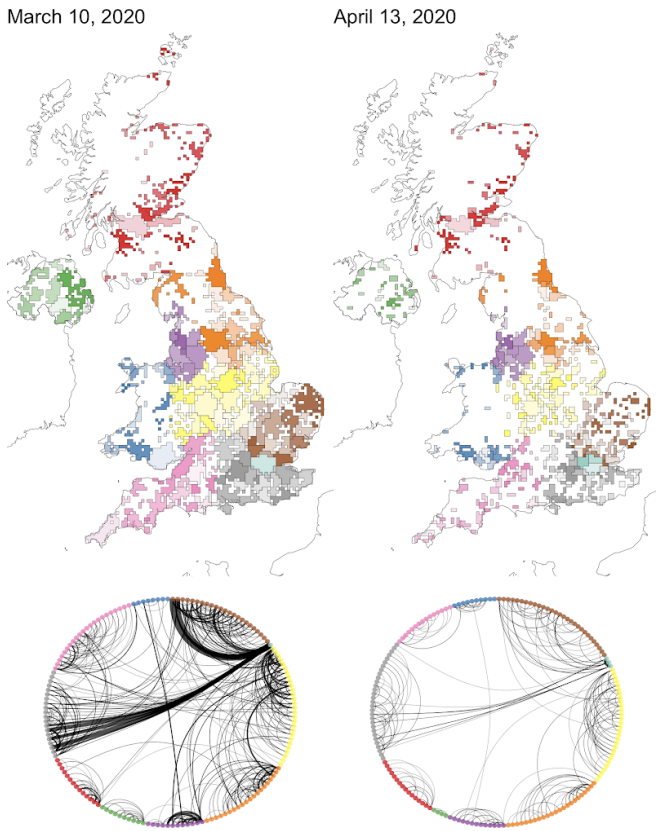
[2]*UCL Department of Computer Science*

[3]*LSHTM Department of Infectious Disease Epidemiology*

Topics: **Human Mobility**, **Data Privacy**, **Decentralized Data**

Mobility data

- Location data from mobile phones is used for:
  - Epidemic modelling
  - Urban planning
  - Natural Disaster response
  - Augmenting offical statistics
  - Much more…

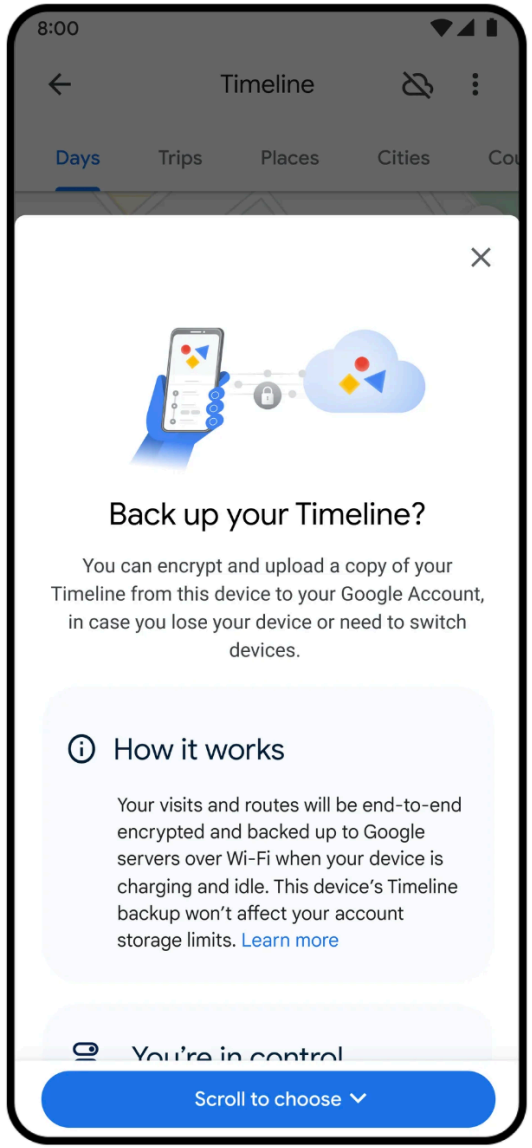


Effect of lockdown on mobility in the UK.

From: Gibbs et. al. (2021).

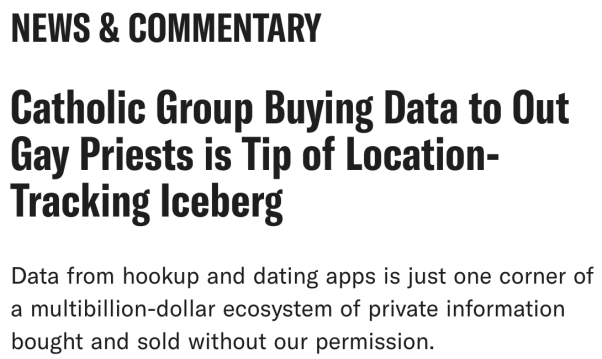
Decentralized mobility data

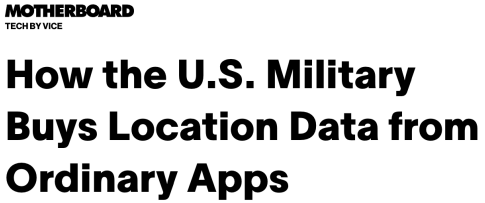
- Major changes are coming to systems for generating mobility data.
  - **Previously:** Individual-level mobility data was stored in a single database.
  - **Increasingly:** Mobility data are stored and processed on the device that collected them.



Privacy risks

- Unique mobility patterns, “linking” with spatial context

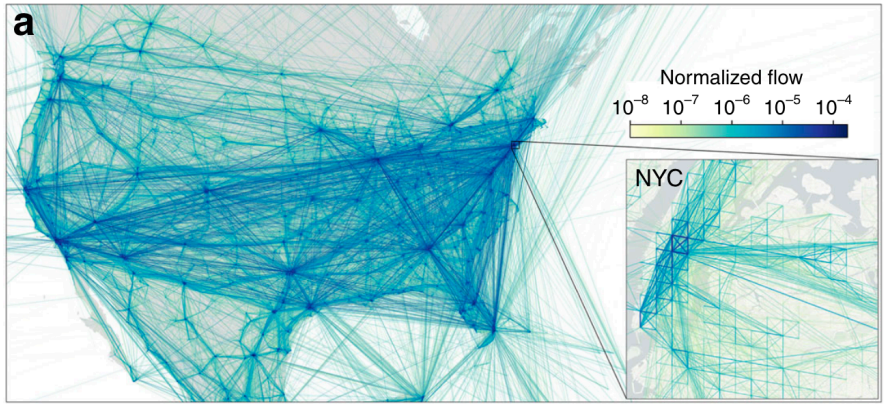




Sources (clockwise from top-left): Vice, New York Times, Vox, ACLU, Vice, New York Times.

Current privacy models

- We focus on: *origin-destination (OD) networks.*
- Two common approaches to privacy in OD networks:
  - **K-anonymity** *(low count suppression).*
  - **Differential privacy (DP)** *(calibrated noise defined by a privacy budget ε).*

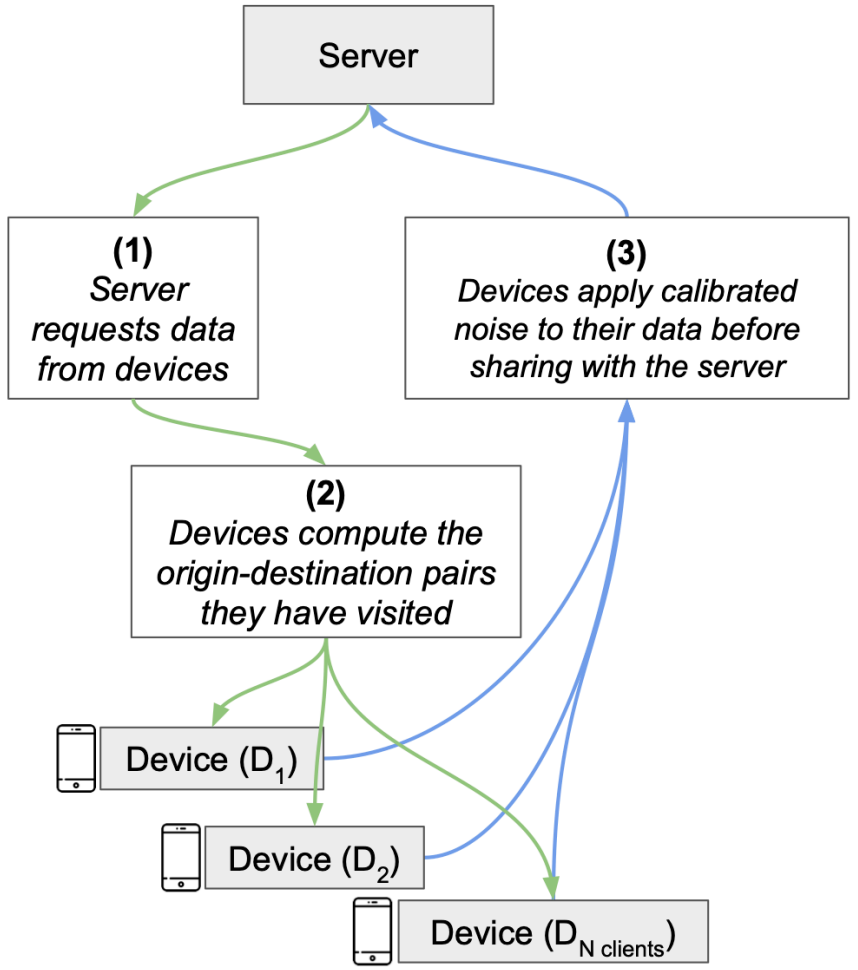


OD network with differential privacy.

From: Bassolas et. al. (2019).

Decentralized privacy

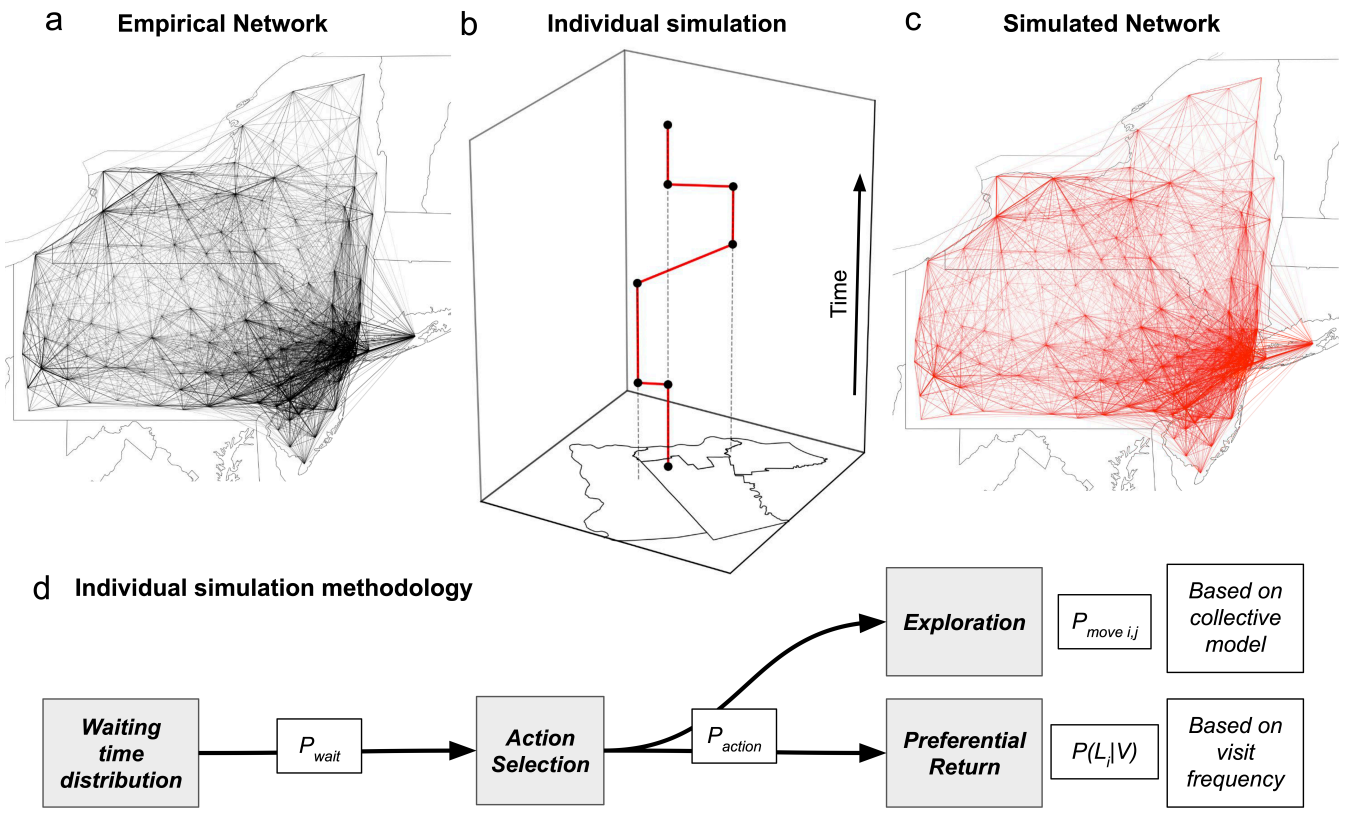
- Current privacy models require centralized collection of location data.
- **Alternative:** Federation with Local Differential Privacy (LDP).
- **Key question:** Does the noise required by LDP introduce too much error?



Methods

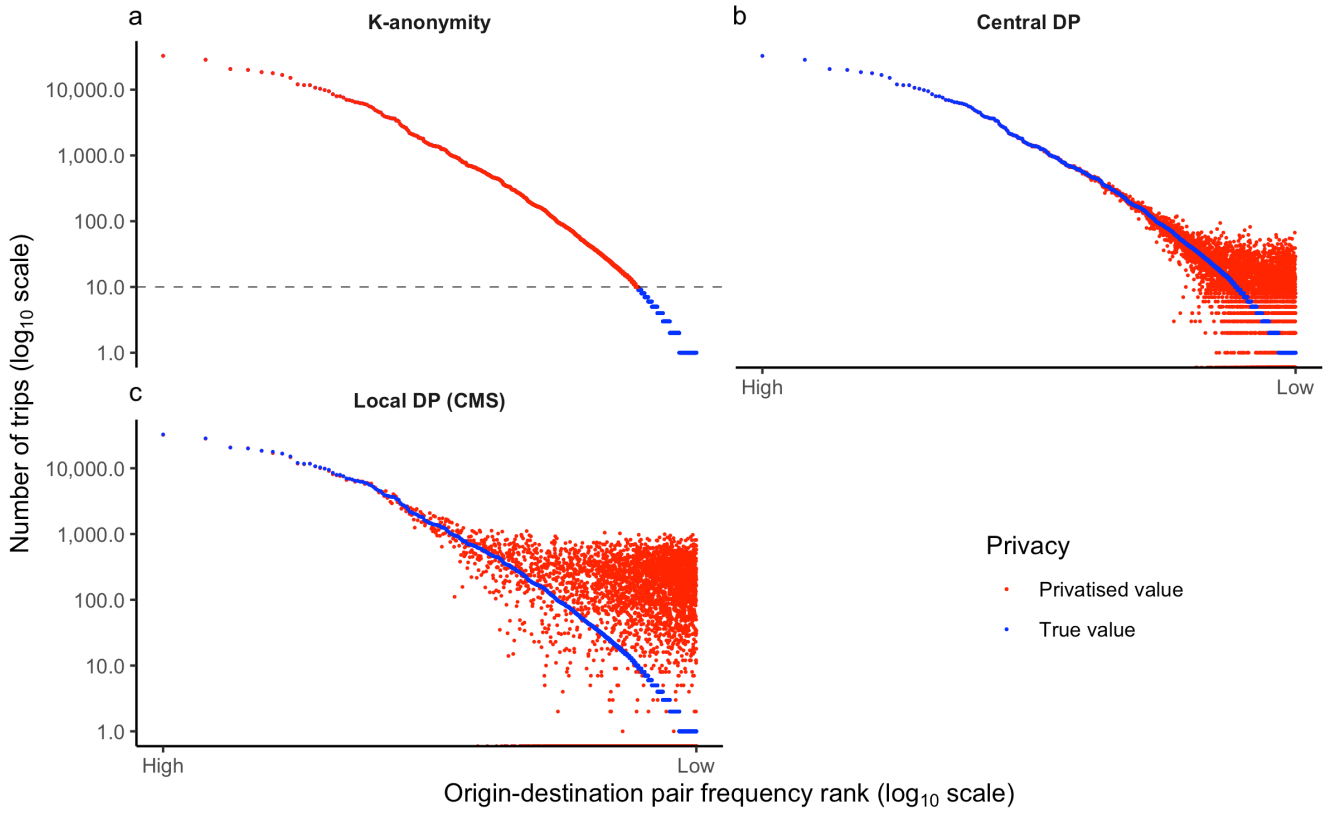
- Simulate a decentralized location dataset
- Apply privacy with three different models
  - *k-anonymity, Central DP, LDP.*
- Quantify impact on data accuracy of:
  - *Privacy model*
  - *Privacy model parameters*
  - *Units of spatial / temporal aggregation*

Methods



- Simulated individual mobility reproduces collective dynamics from empirical data.

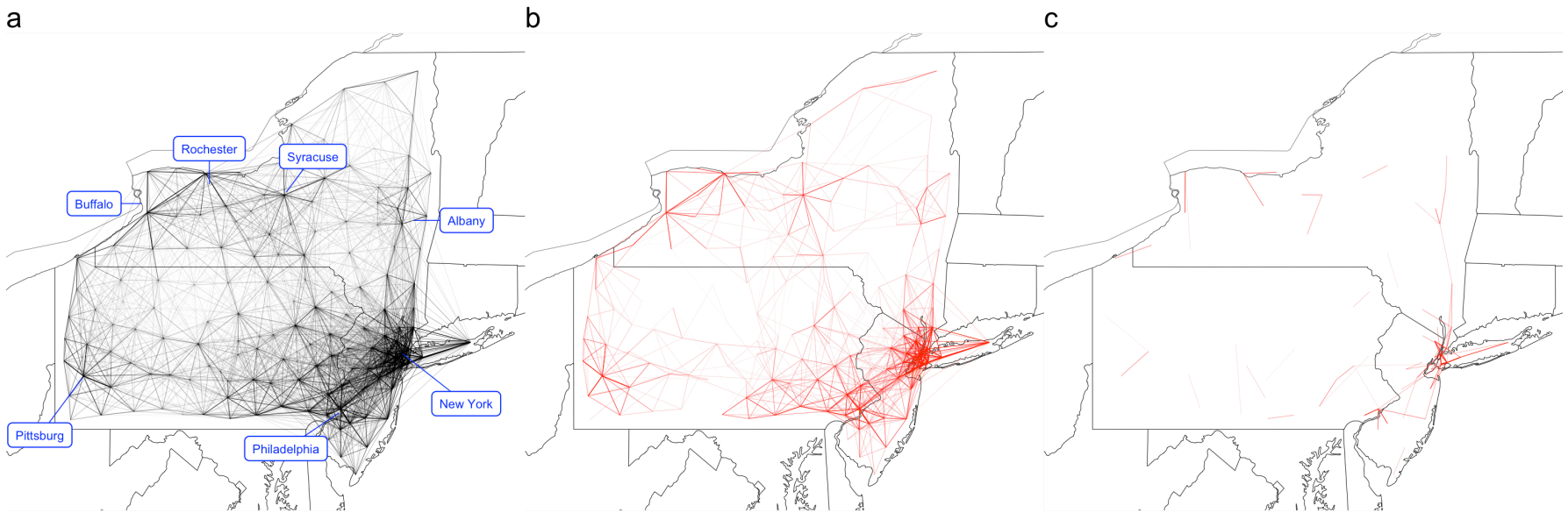
Results



- “Compounding” noise required for LDP introduces high error for low frequency edges.

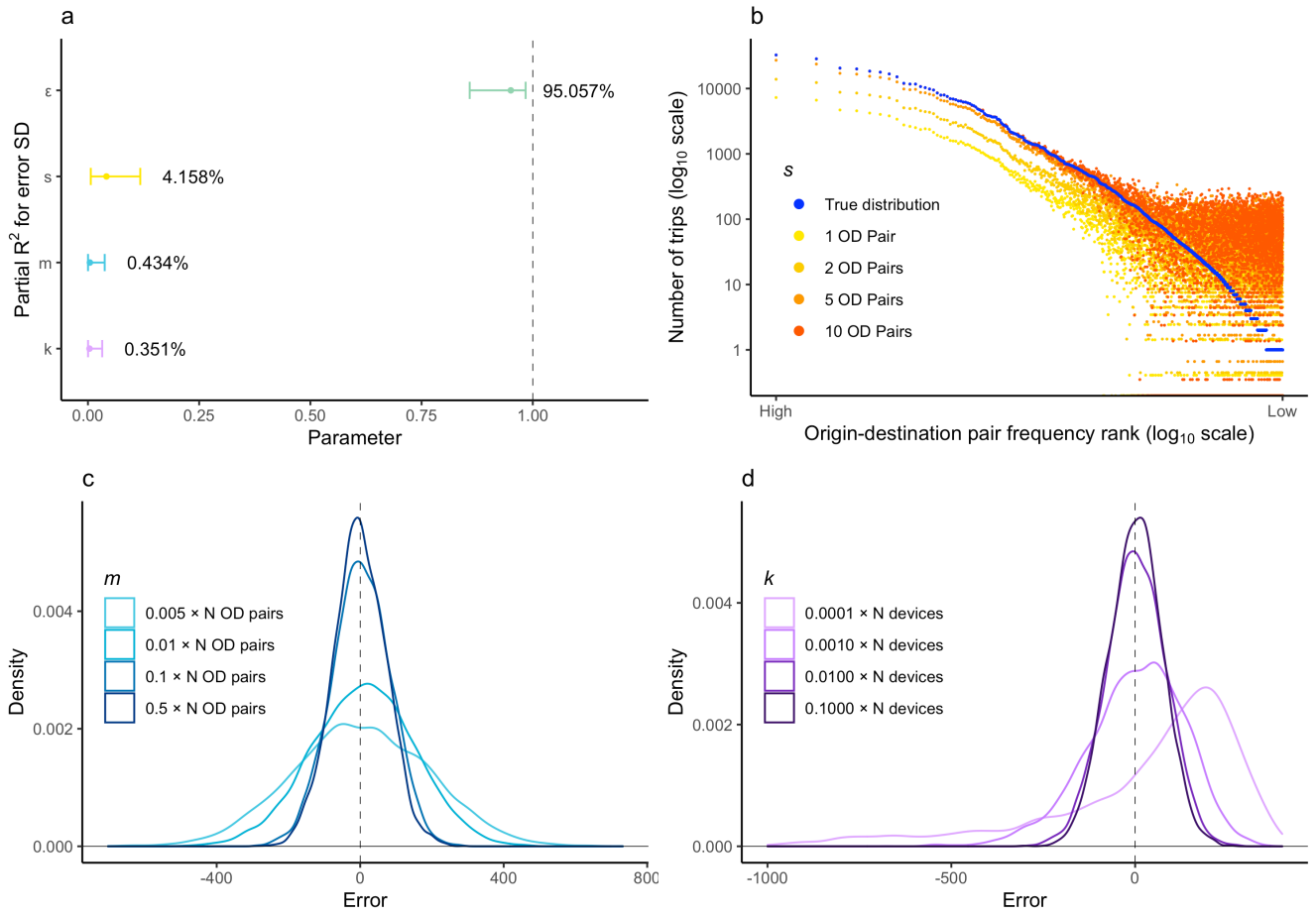
  Privacy parameters: a) k=10, b) ε=1 , s=10, c) m=2340, k=205, ε=5, s=10.

Results



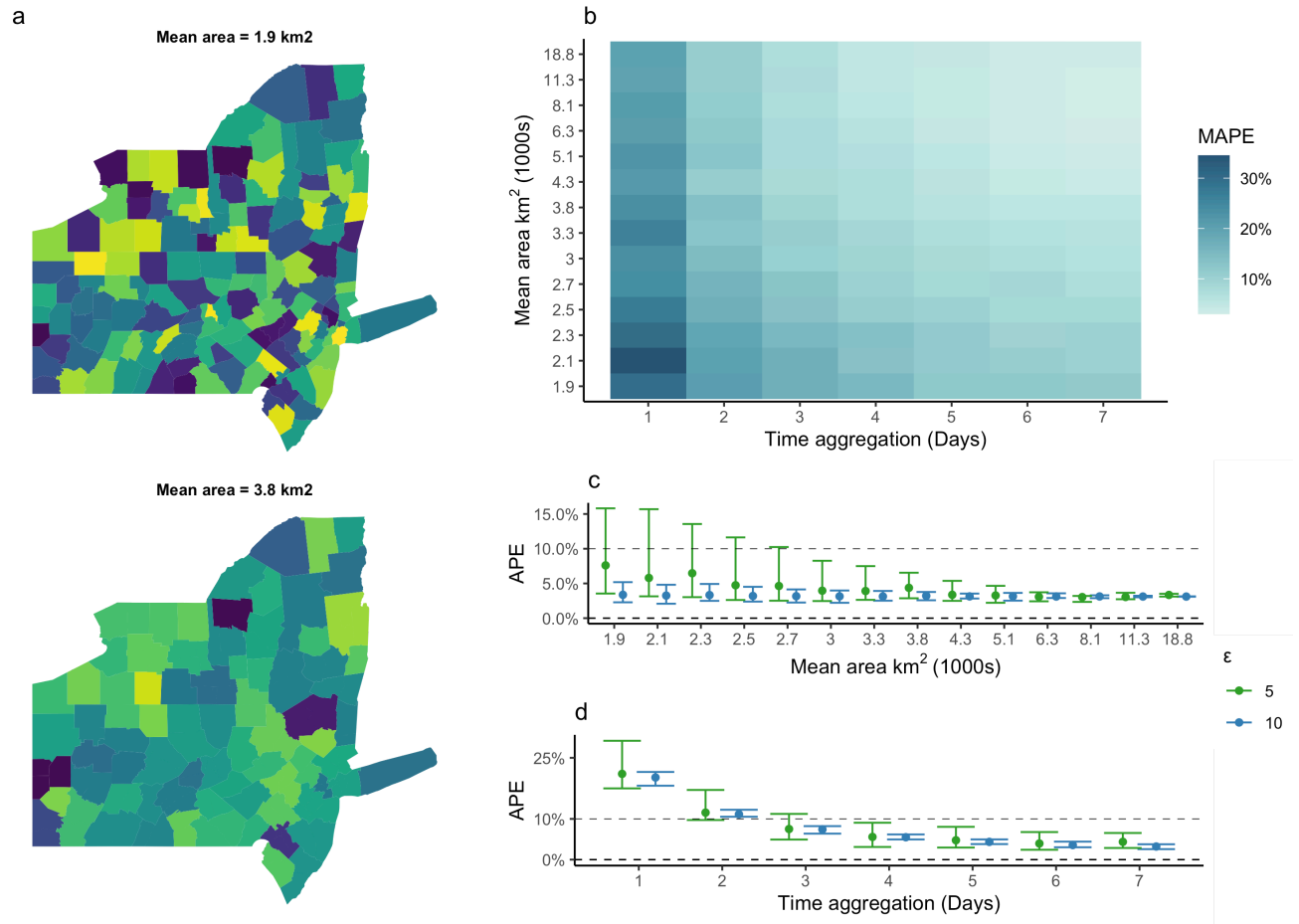
- Most connections have error >10% in an LDP network. a) Original data, b) Central DP network, c) LDP network.
- But, there are many “levers” to improve data accuracy.

Results



- **One ‘lever’:** changing algorithm-specific privacy parameters.

Results



- **Another 'lever':** choosing units of spatial/temporal aggregation.

Conclusions

- Simulating individual-level mobility data allows full transparency into effect of privacy choices.
- There are many opportunities to improve data accuracy.
- Decentralized data with LDP could allow continued use of mobility data.
  - Also: new opportunities for understanding human behavior *(on-device data linkage, complex analytics).*

Questions?