

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 Gibbs¹, Mirco Musolesi², James Cheshire¹, Rosalind M. Eggo³

¹*UCL Department of Geography*

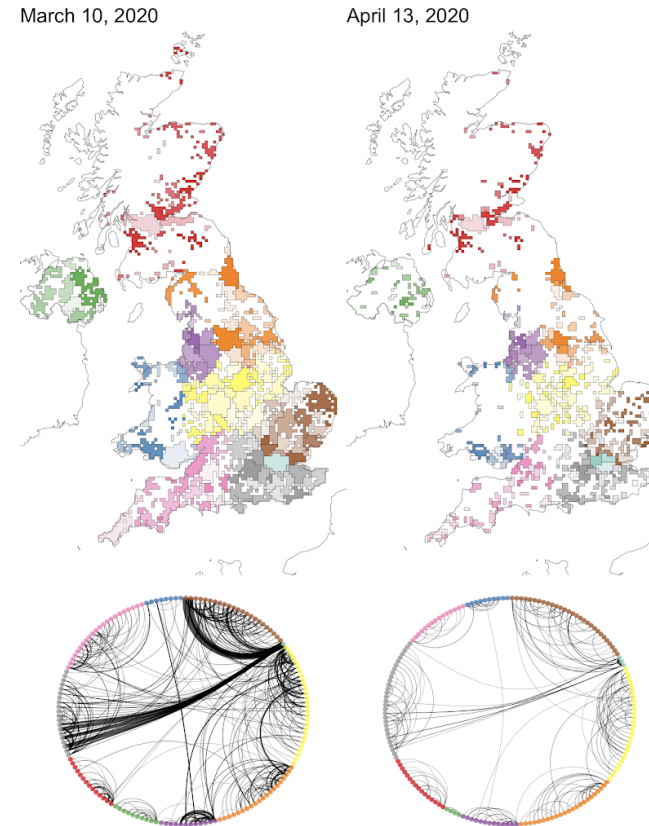
²*UCL Department of Computer Science*

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 official statistics
 - Much more...

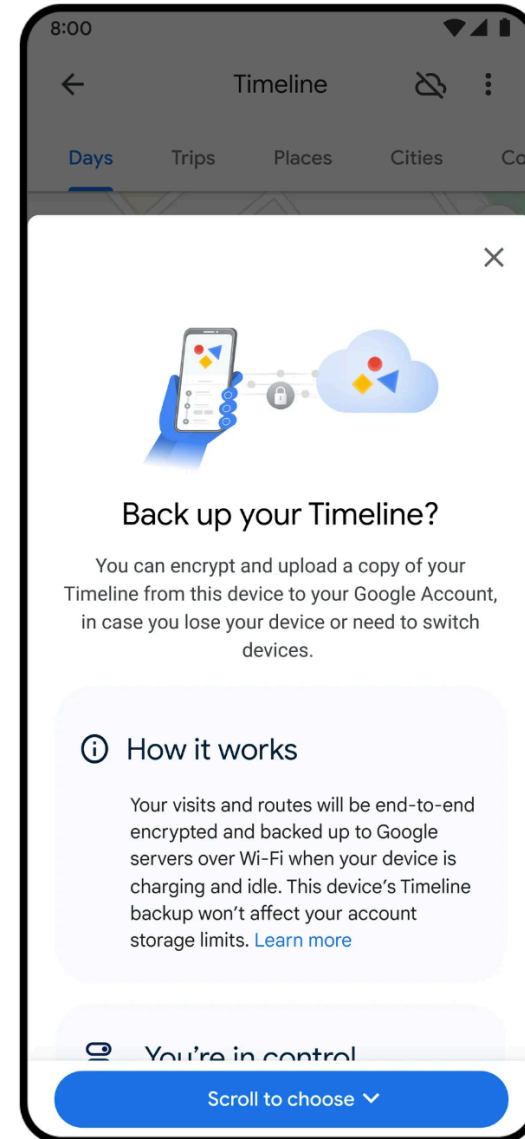


Effect of lockdown on mobility in the UK.

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

MOTHERBOARD
TECH BY VICE

Data Broker Is Selling Location Data of People Who Visit Abortion Clinics

NEWS & COMMENTARY

Catholic Group Buying Data to Out Gay Priests is Tip of Location-Tracking Iceberg

Data from hookup and dating apps is just one corner of a multibillion-dollar ecosystem of private information bought and sold without our permission.

Opinion | **THE PRIVACY PROJECT**

Twelve Million Phones, One Dataset, Zero Privacy

MOTHERBOARD
TECH BY VICE

How the U.S. Military Buys Location Data from Ordinary Apps

TECHNOLOGY

The hidden trackers in your phone, explained

How covert code enables your phone's apps to spy on you.

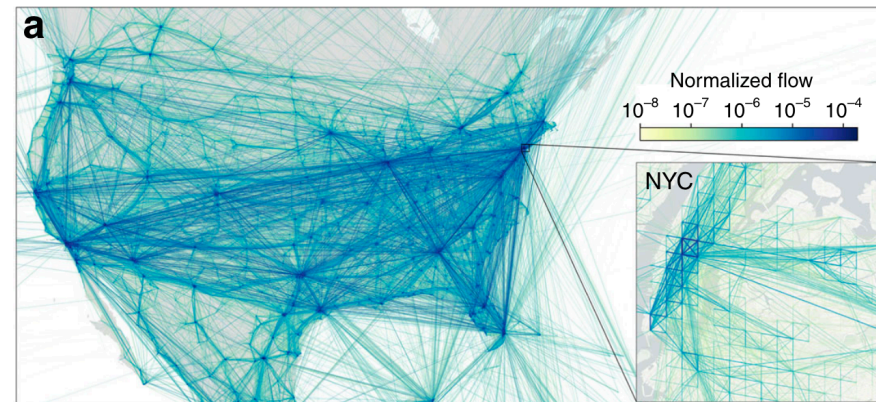
Your Apps Know Where You Were Last Night, and They're Not Keeping It Secret

Dozens of companies use smartphone locations to help advertisers and even hedge funds. They say it's anonymous, but the data shows how personal it is.

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*).

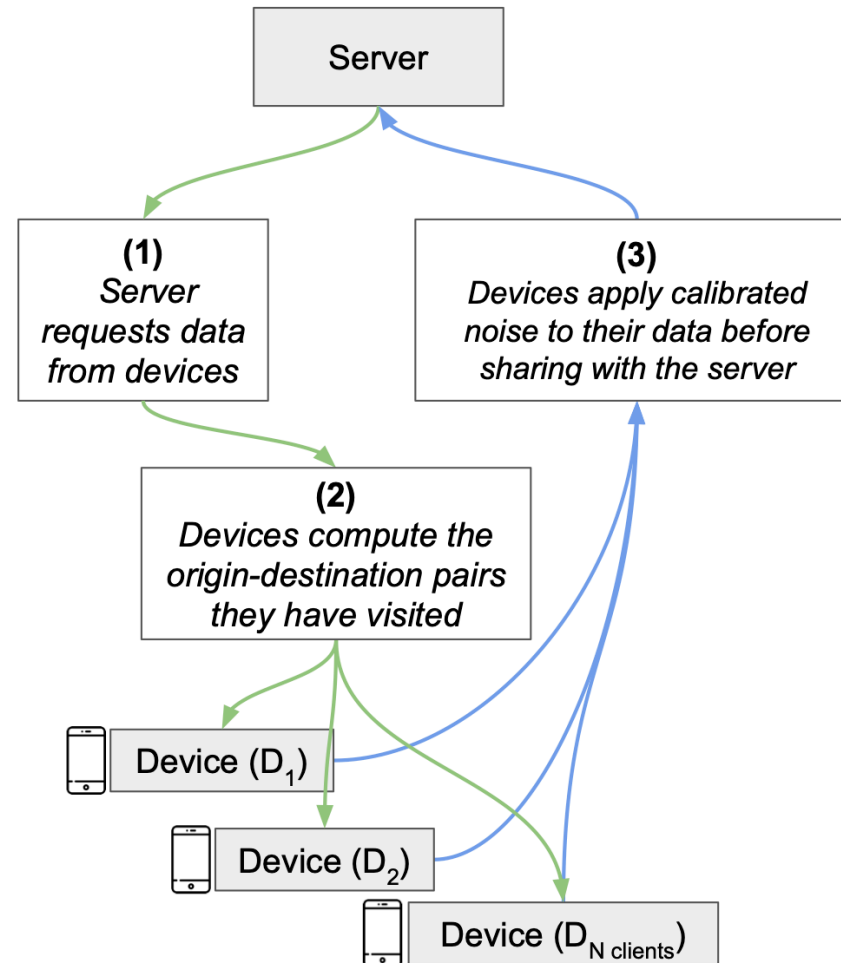


OD network with differential privacy.

From: [Bassolas et. al. \(2019\)](#).

Decentralized privacy

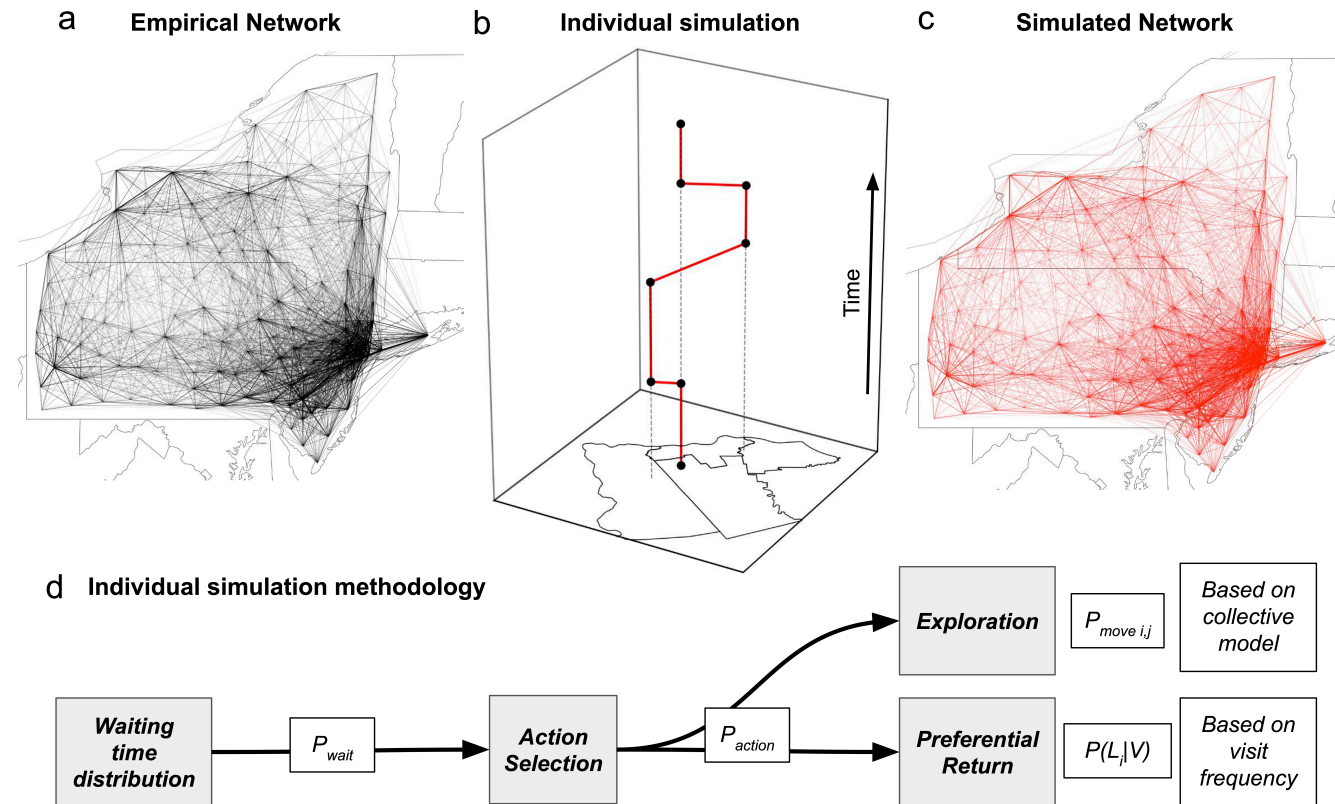
- Current privacy models require centralized collection of location data.
- **Alternative:** Federation with Local Differential Privacy (LDP).
- **Key question:** Does the noise required by LDP introduce too much error?



Methods

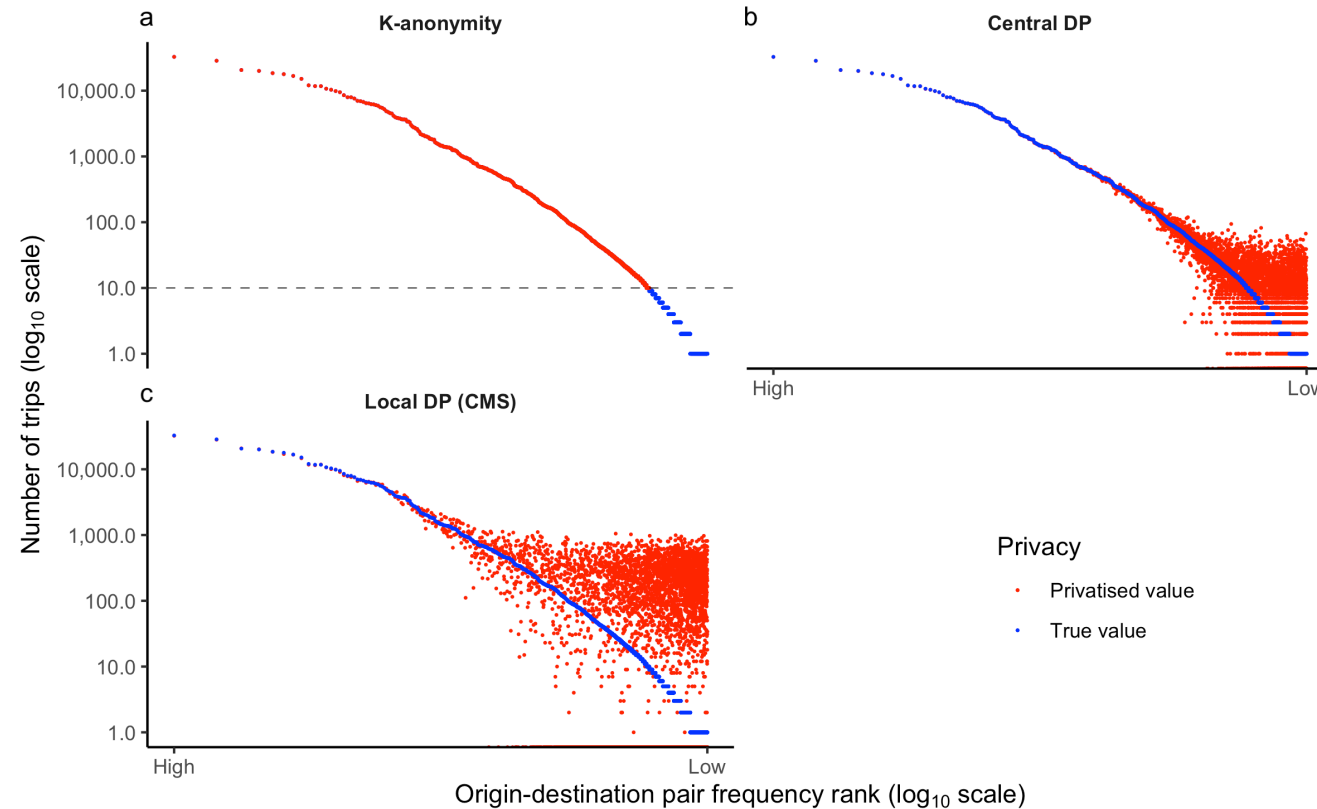
- Simulate a decentralized location dataset
- Apply privacy with three different models
 - *k-anonymity, Central DP, LDP.*
- Quantify impact on data accuracy of:
 - *Privacy model*
 - *Privacy model parameters*
 - *Units of spatial / temporal aggregation*

Methods



- Simulated individual mobility reproduces collective dynamics from empirical data.

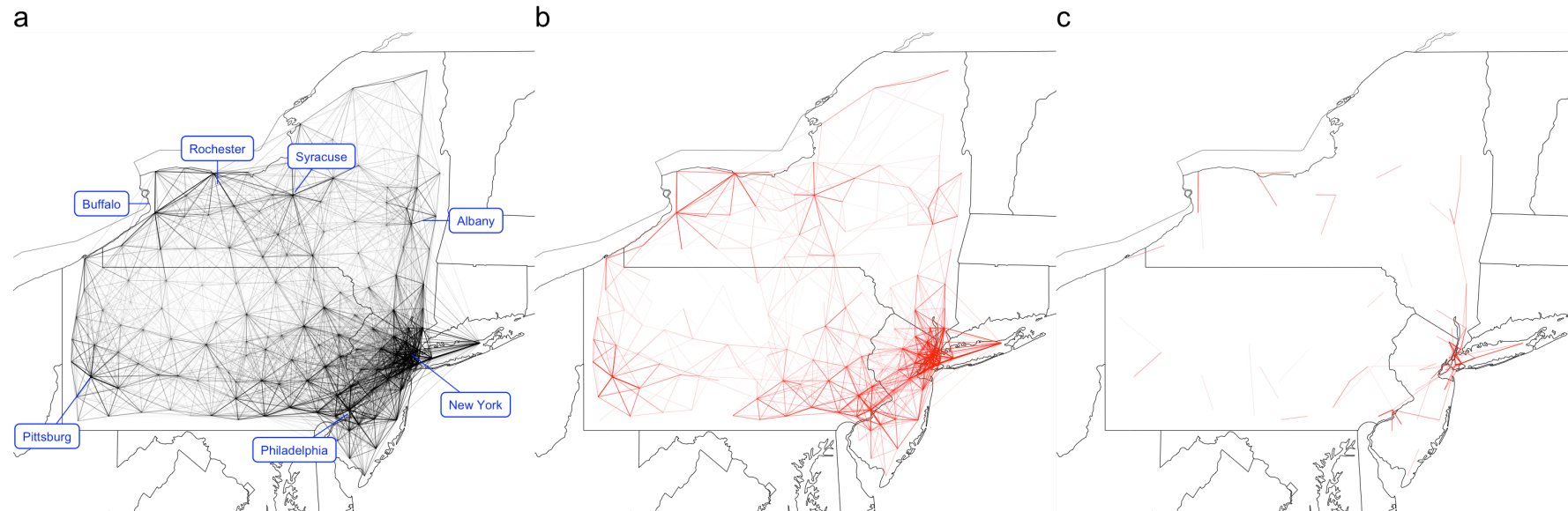
Results



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

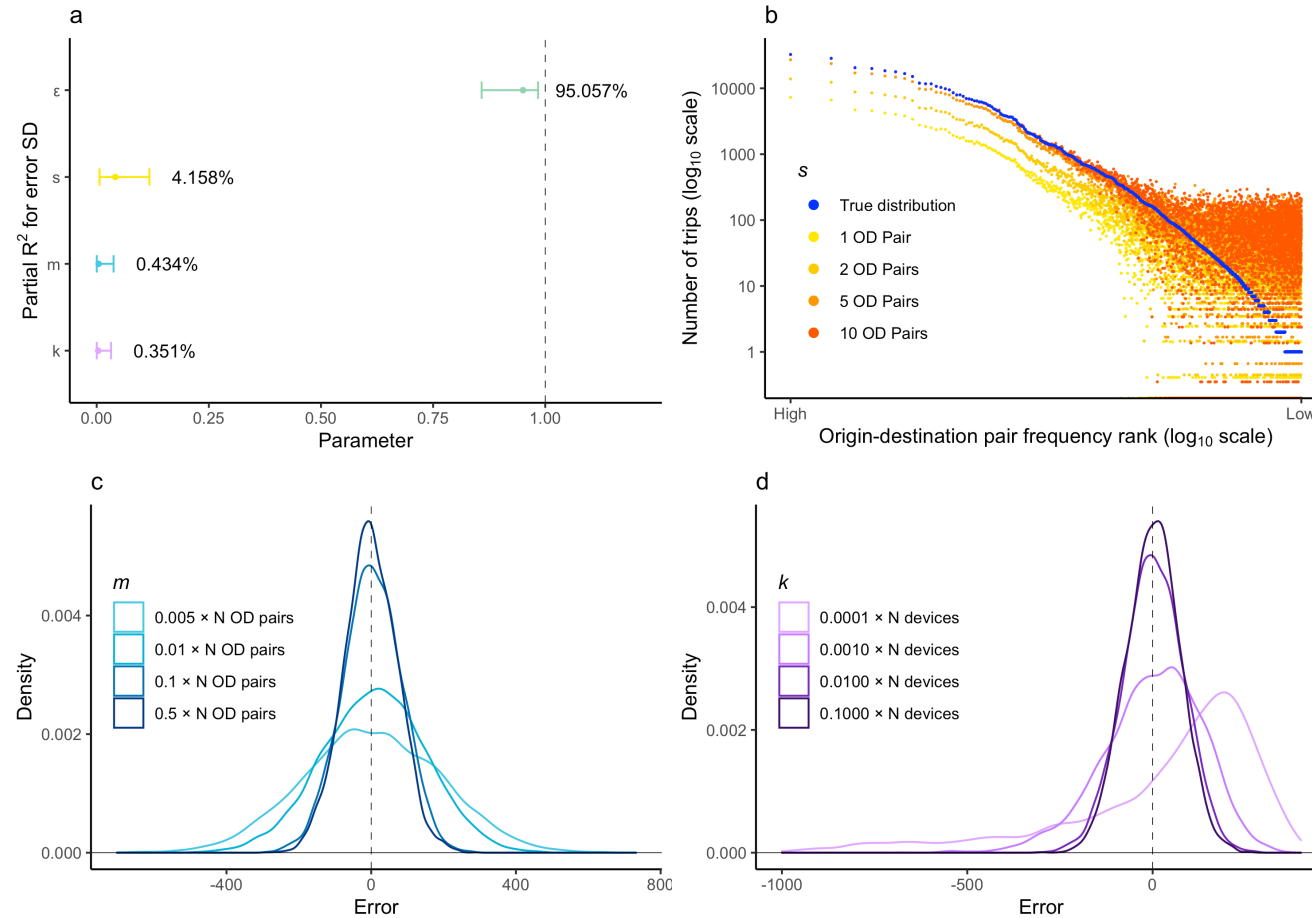
Privacy parameters: a) $k=10$, b) $\epsilon=1$, $s=10$, c) $m=2340$, $k=205$, $\epsilon=5$, $s=10$.

Results



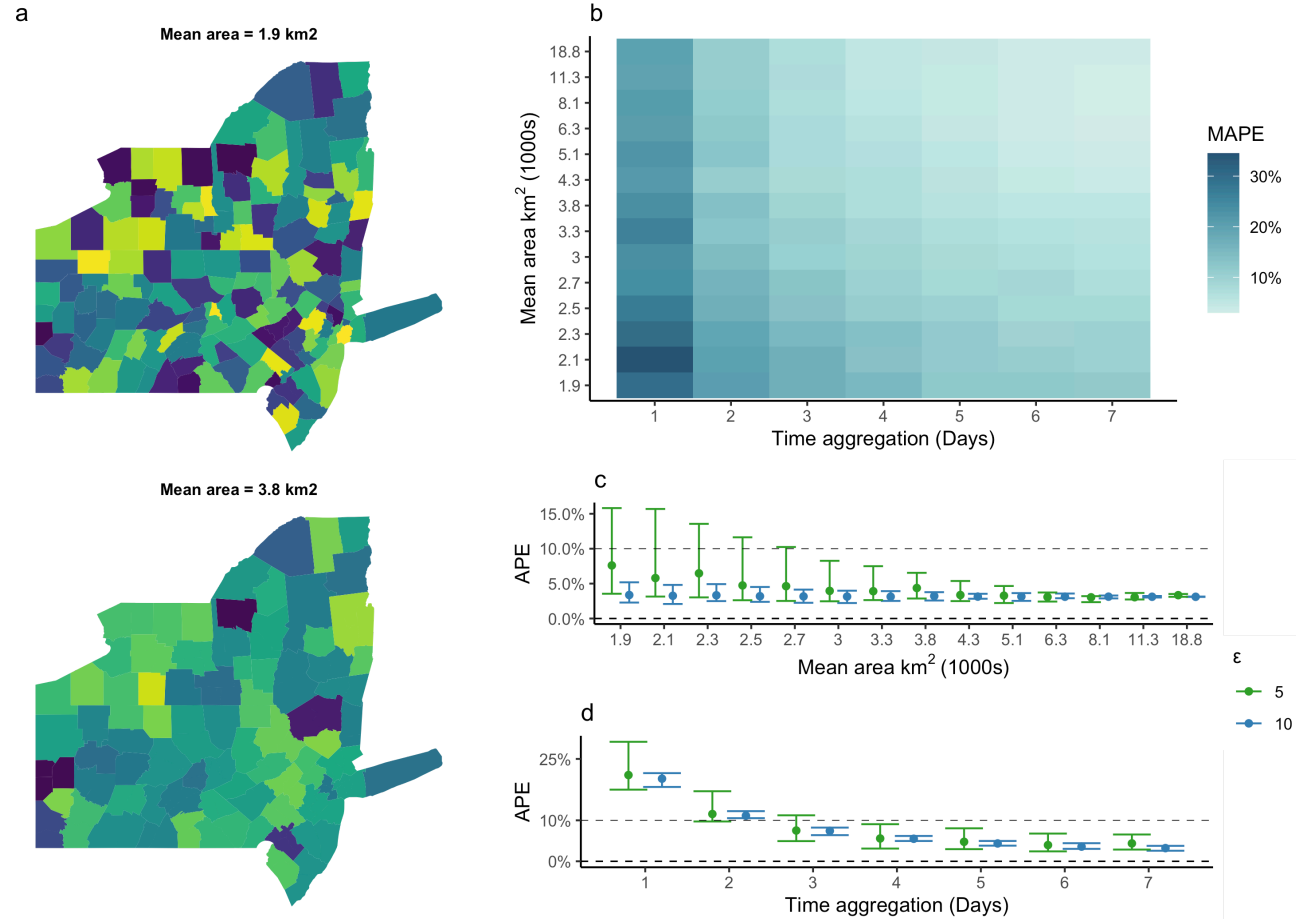
- Most connections have error $>10\%$ in an LDP network. a) Original data, b) Central DP network, c) LDP network.
- But, there are many “levers” to improve data accuracy.

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?