Privacy Implications of Measuring Large-Scale Social Networks With High Resolution

Arkadiusz Stopczynski, Technical University of Denmark & MIT Media Lab
arks@dtu.dk, @h0pbeat, https://sensible.dtu.dk

Measuring Large-Scale Social Networks With High Resolution

Copenhagen Networks Study

x 1 000

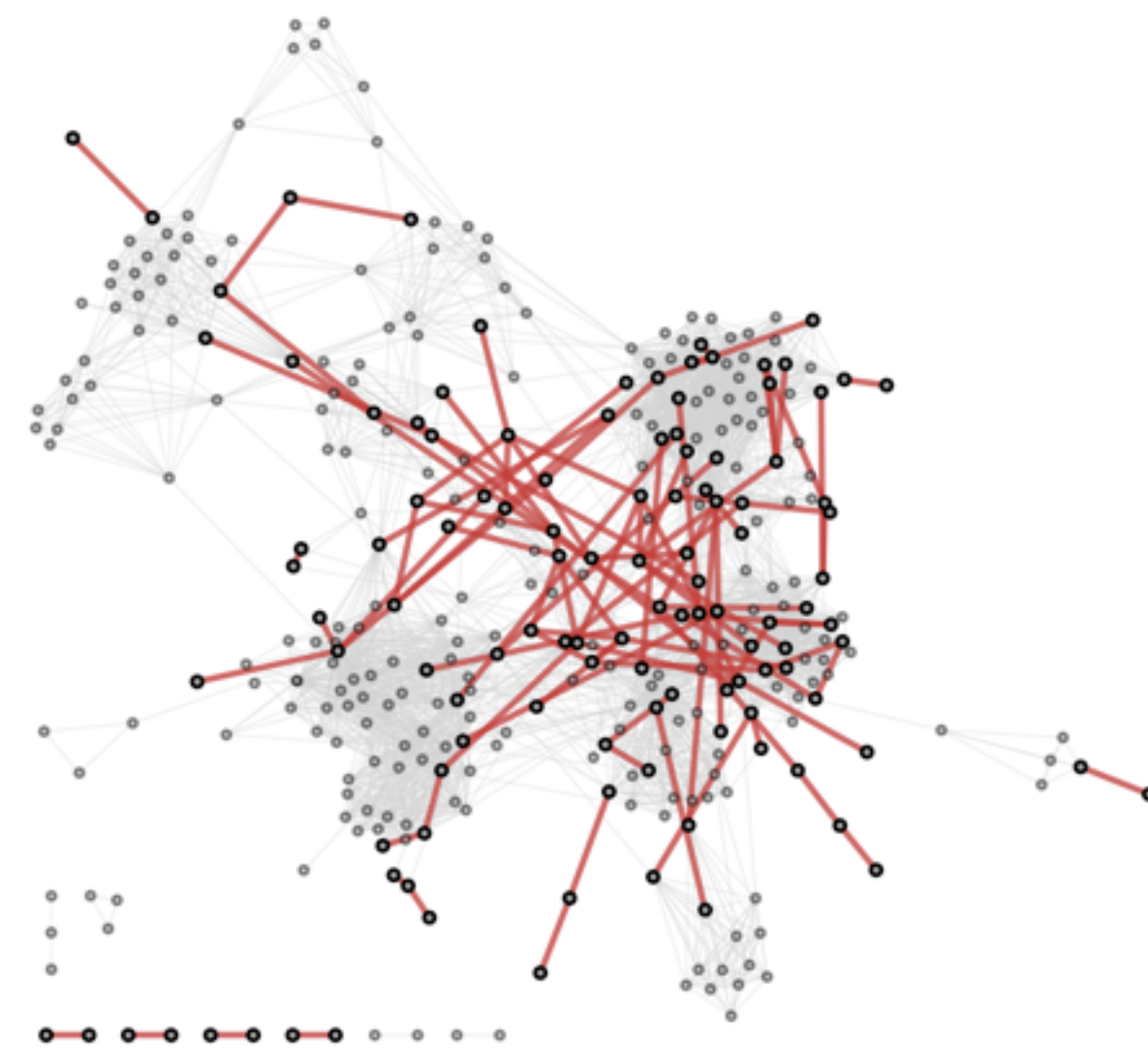|		
face to face	health	friends
calls	psychology	interactions
messages	sociology	activity
mobility	demographics	demographics

fünf
Open Sensing Framework
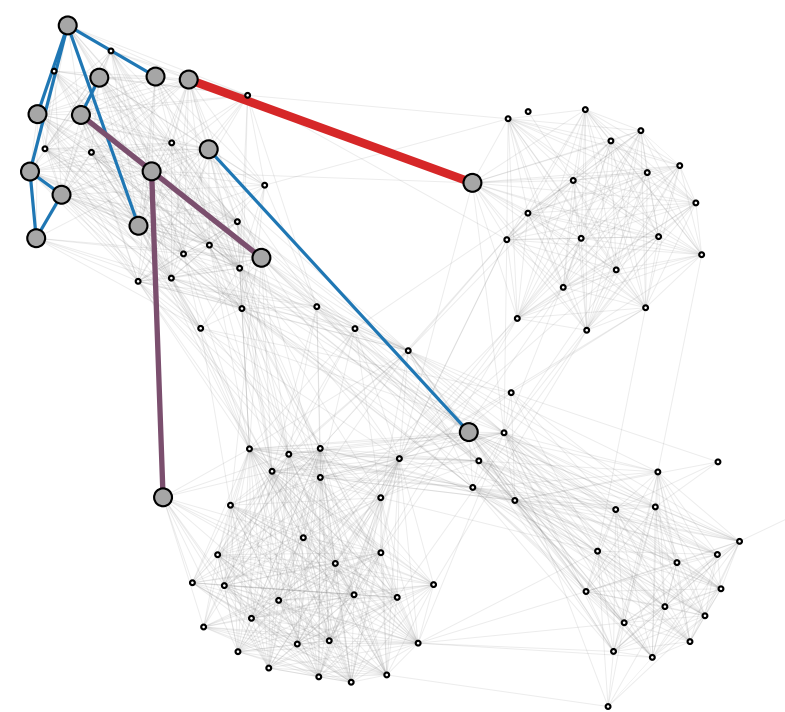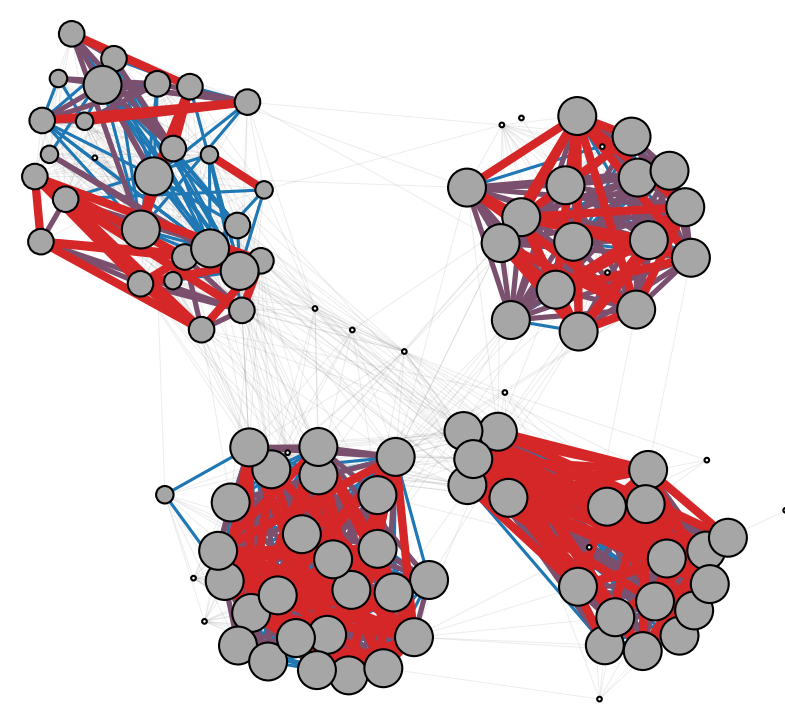
Copenhagen Networks Study
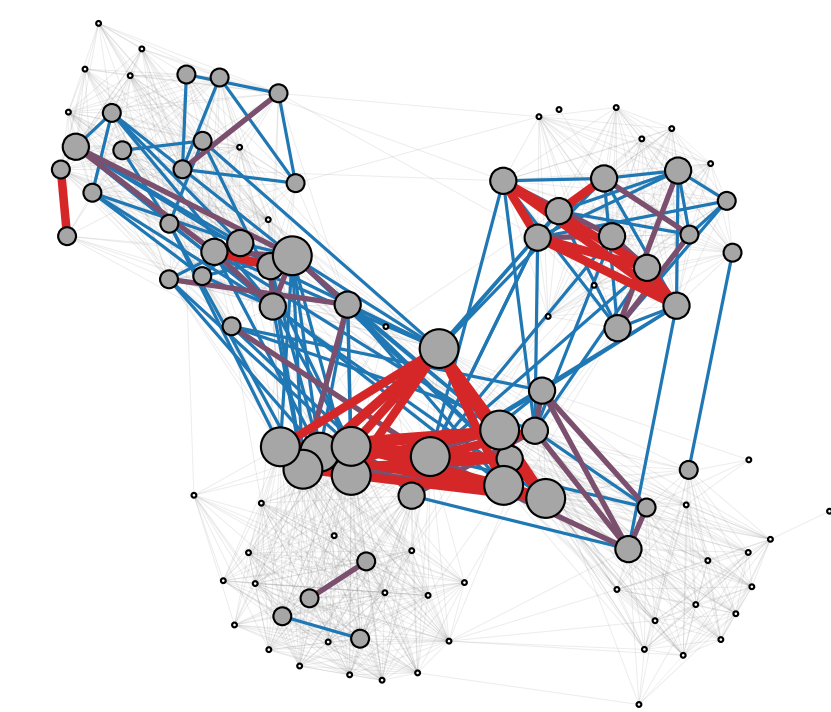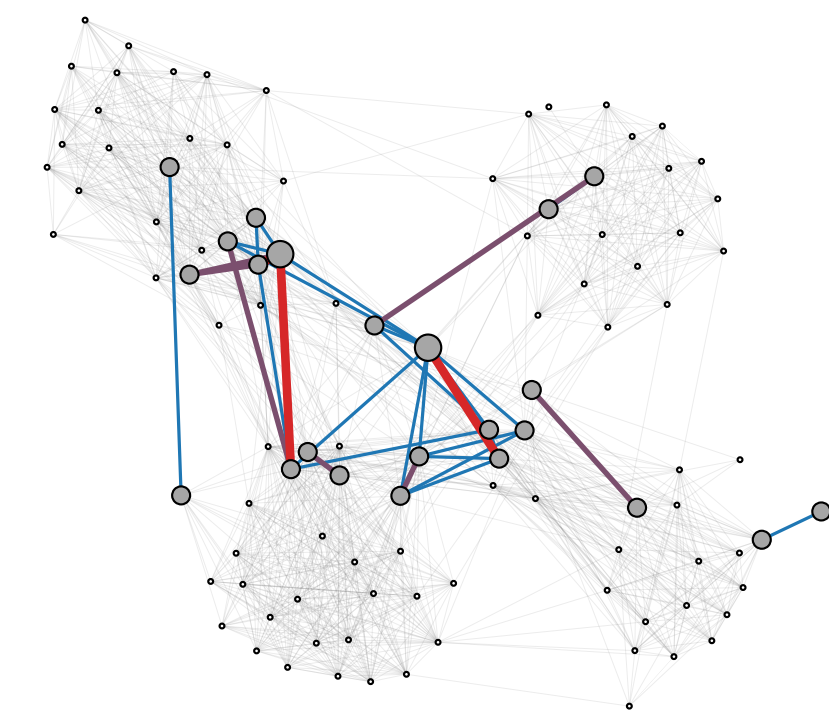

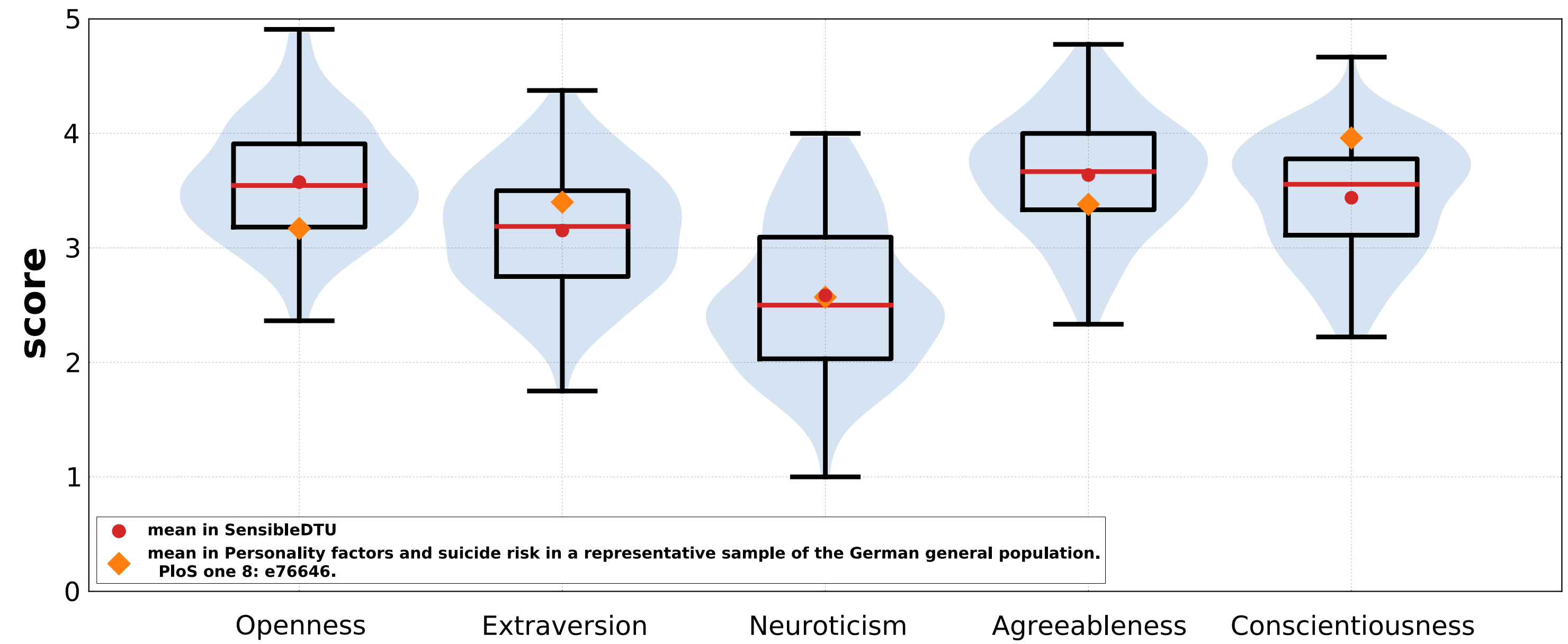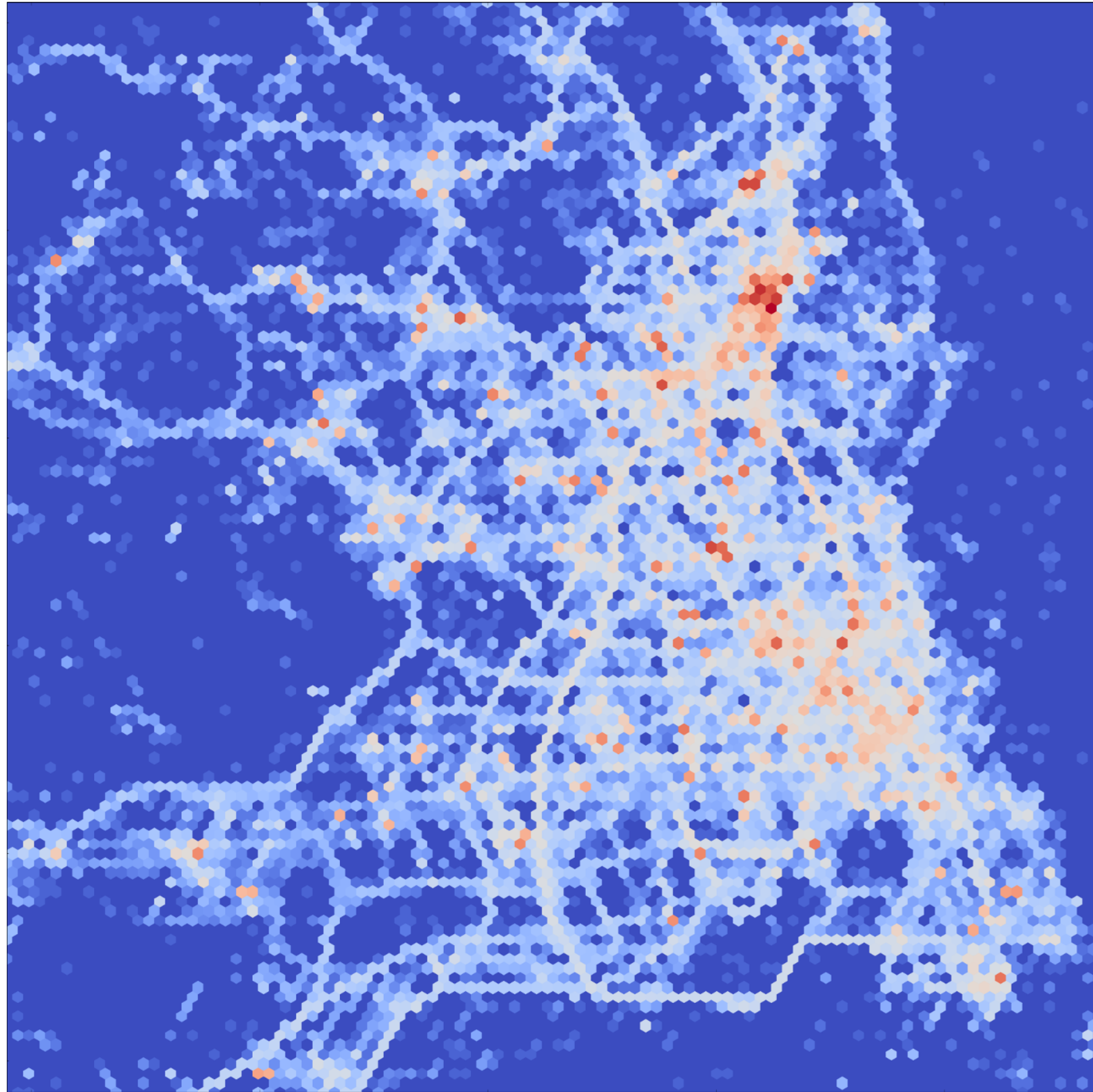Online   Face-to-face   Extra Info. Online   Extra Info. F2F

7:00   10:00   17:00   20:00

Copenhagen Networks Study

Copenhagen Networks Study

Privacy: Users' Perspective

- OpenID 2.0 Identity Provider
- pseudonyms, optionally PPIDs
- verified identity attributes via linked account

Privacy: Users' Perspective

- simplified language
- accepted online
- creating OAuth2 token between study and user service
- text stored, revisioned, and pulled from git
- acceptance stored including git commit

Privacy: Users' Perspective

- ◉ Living Informed Consent
- ◉ OAuth2 authorizations dashboard
- ◉ interactive contract
- ◉ data flow dashboard in development

Privacy: Users' Perspective

- applications using one IdP (but don't have to)
- OAuth2 authorizations for in and out
- manually registered in the system by admin

Privacy: Users' Perspective

- data viewer
- access to all collected data
- same API as used by applications & researchers
- building URIs
- simple visualizations
- but of course this is not enough…
- users' perspective privacy research testbed

Privacy: Security

DTU

Amazon AWS
(Ireland)

OAuth2 bearer tokens

API + Applications

encrypted 'obvious PIIs'

DB
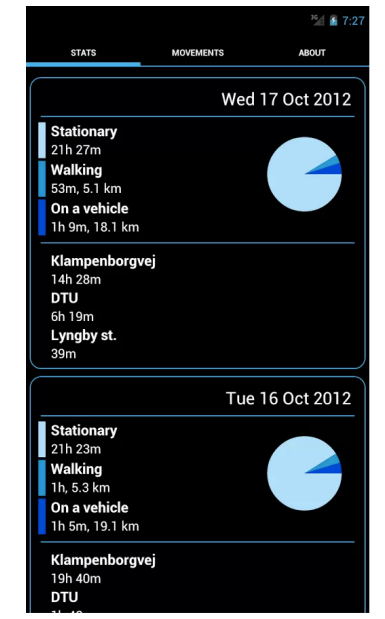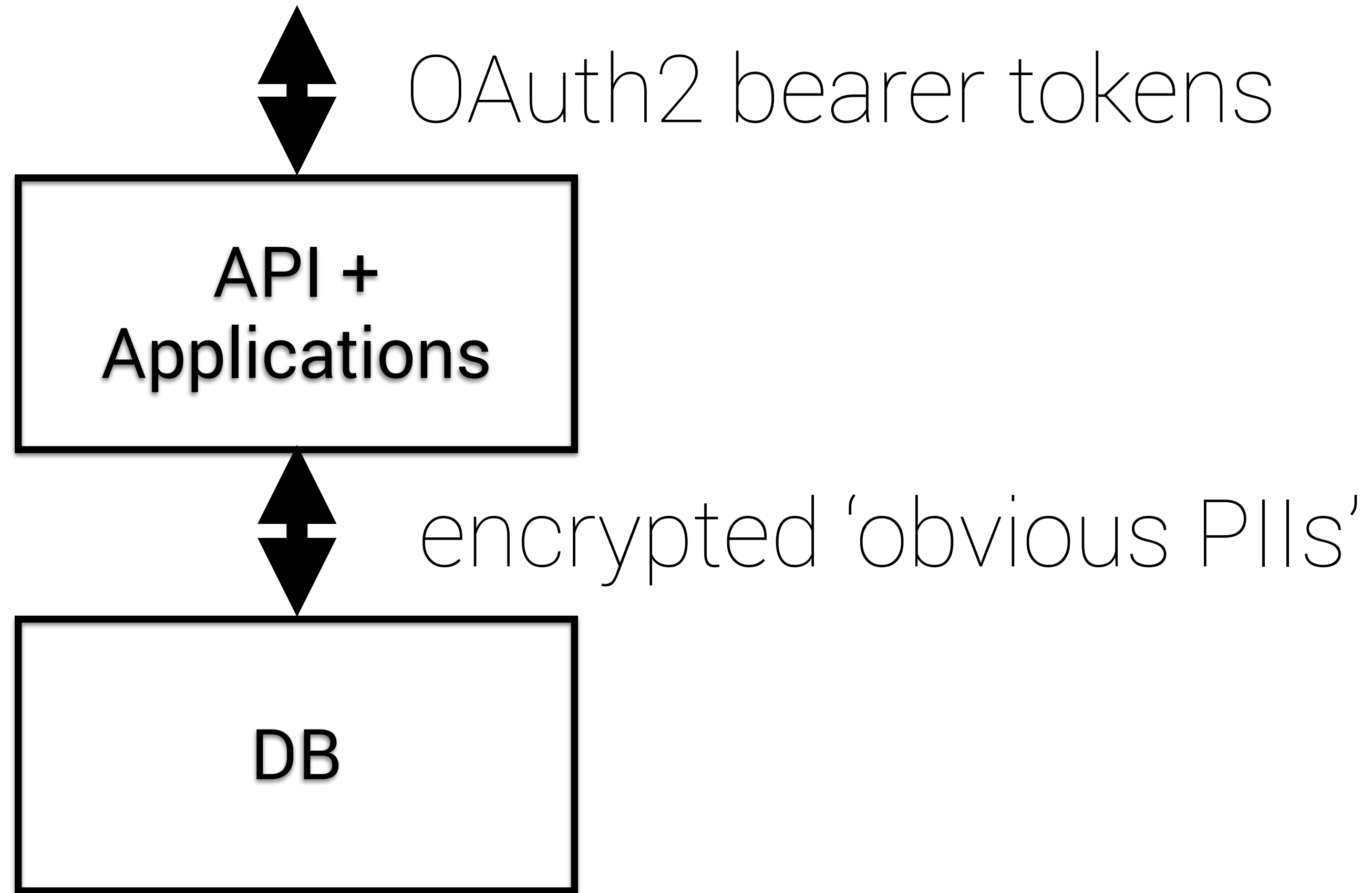
Privacy: Security
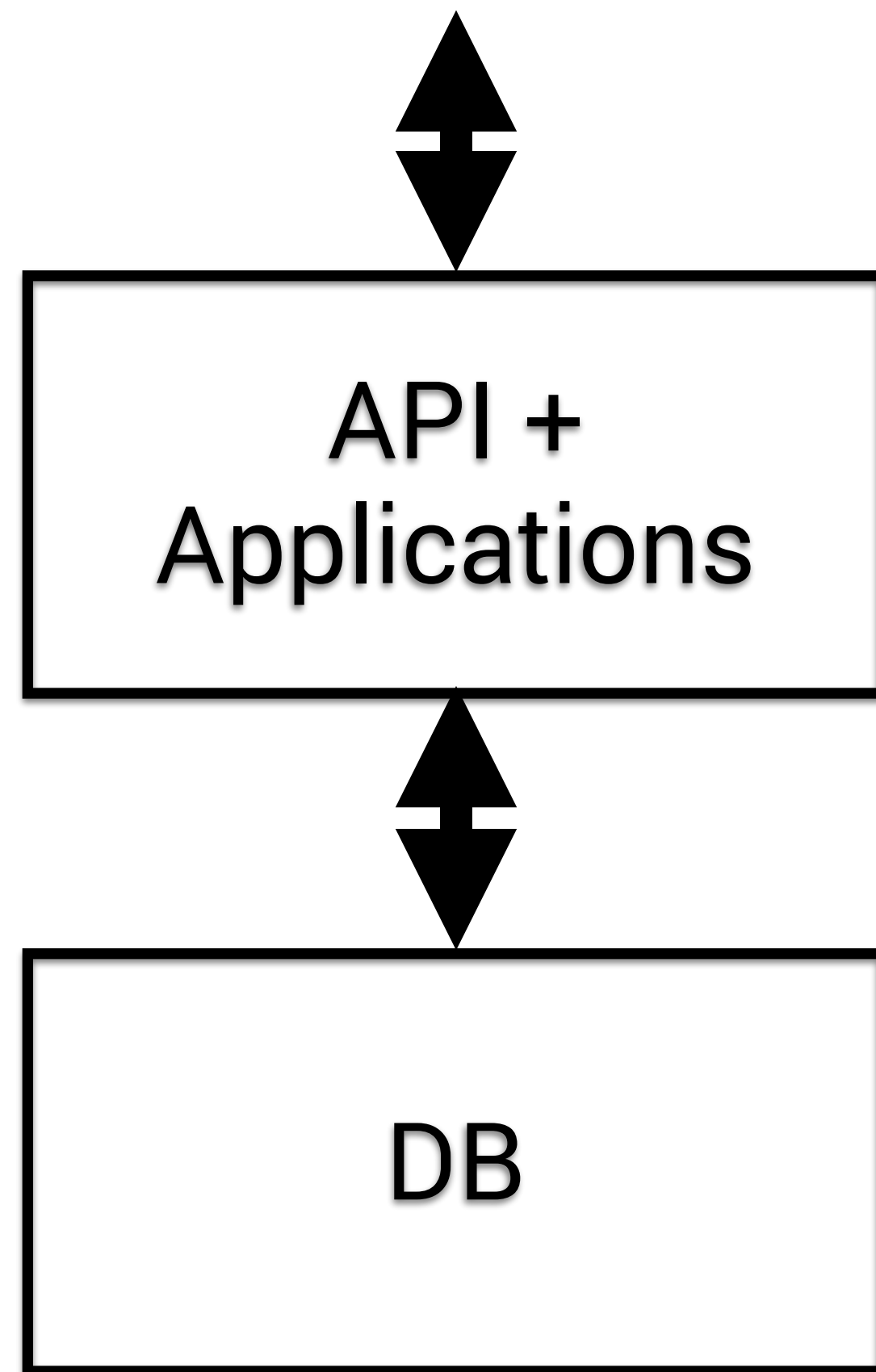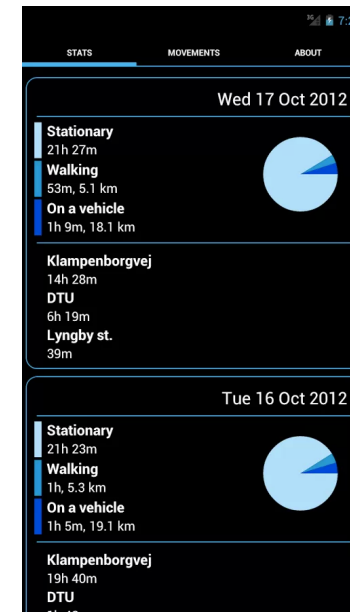
- the only large study in DK approved by Danish Data Protection Agency to use cloud storage
- accounts, keys stay at uni server
- TLS everywhere
- disk encryption on AWS
- short-lived OAuth2 tokens (app & researcher)
- encrypted database files on the phone, encrypted archive files for backup (Amazon Glacier)

API + Applications

↕

DB

Privacy: Security

Sensible Data: open Personal Data System

Applications

Phone Data Collector	Questionnaires	Data Viewer	Sensible Journal	Visu

↕ **data flows**

Services

SensibleDTU Study

Connectors	Authorization Manager	Application Manager
Questions & Answers	Legal Documents	Anonymizer
Databases	Audit	Backup

A Different Study

Connectors	
Questions & Answers	Le
Databases	

↕ **control flows**

Platform

OpenID 2.0 Identity Provider	Authorization Dashboard	User Profile	Applications Dashboard

- 3-layer architecture
- multiple studies with common dashboard
- extensible code
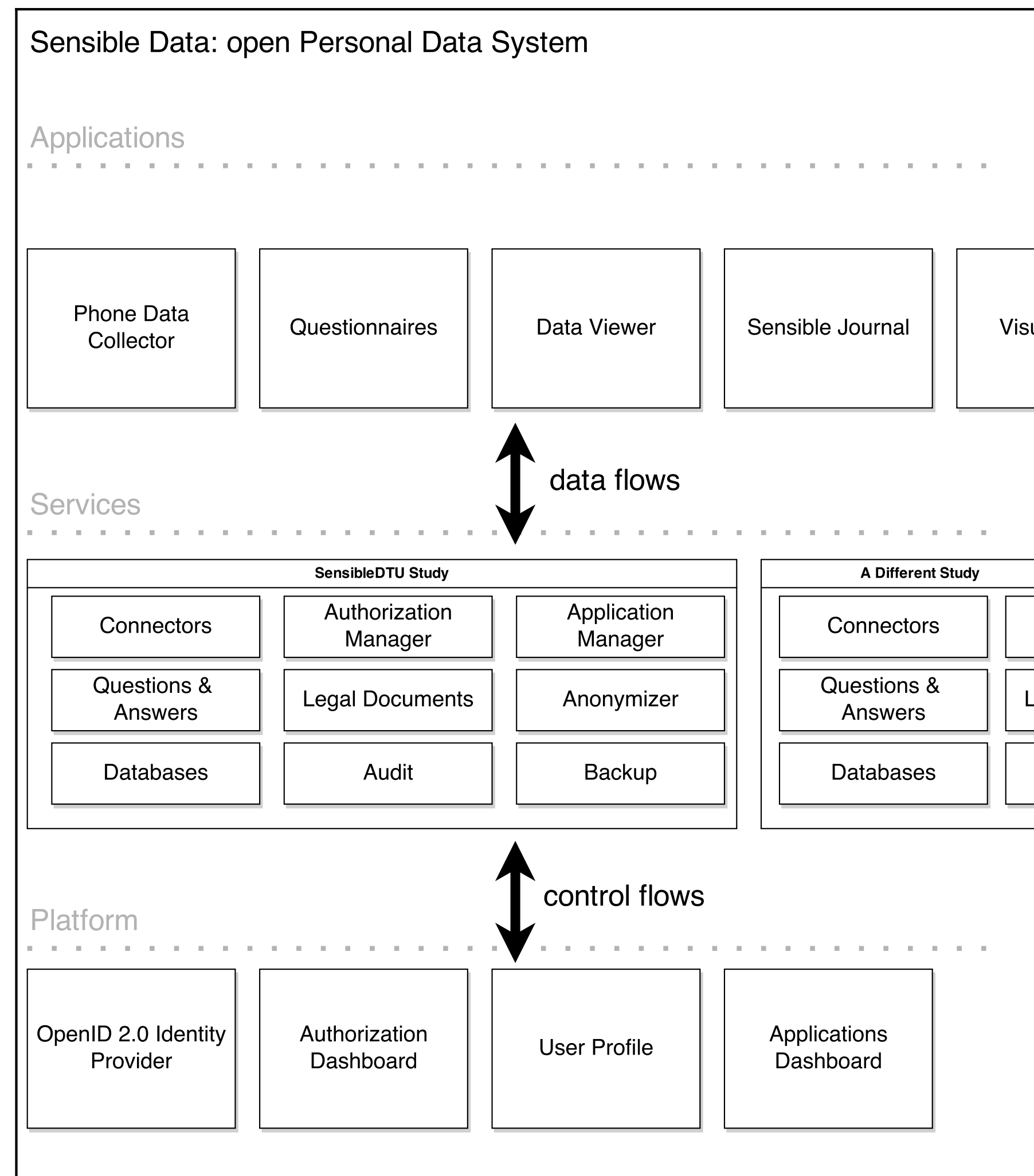
Privacy: High-Resolution Data

- Bluetooth (5 minutes)
- location (10 minutes)
- Wifi (20 seconds)
- cell ID
- screen state
- call/SMS log
- contacts

- Big Five Inventory
- Rosenberg Self-Esteem Scale
- Narcissism NAR-Q
- Satisfaction With Life Scale
- Rotters Locus of Control Scale
- ...

- friends
- interactions
- checkins
- posts
- work
- eduction
- demographics

Privacy: High-Resolution Data
Location
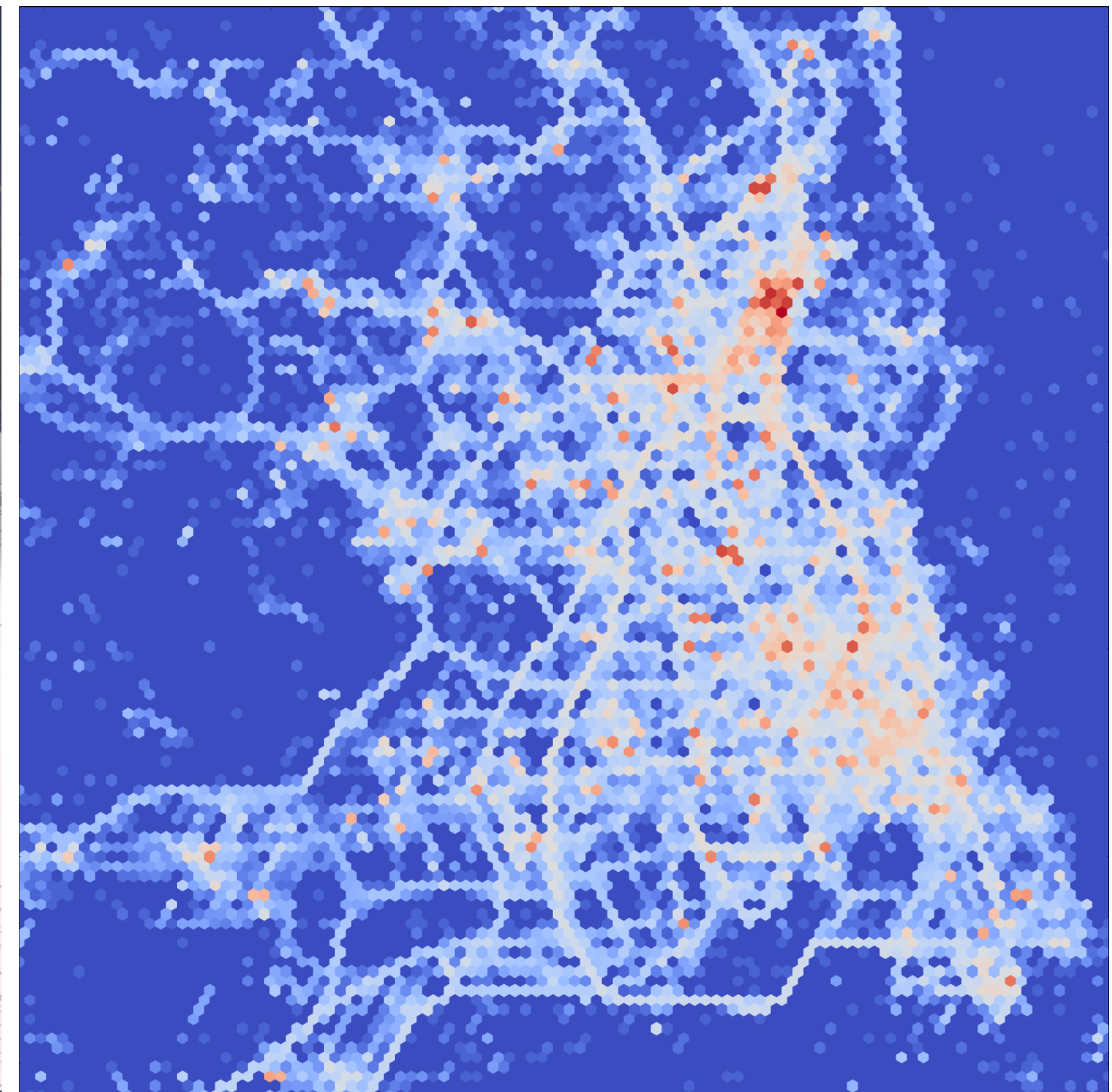
Privacy: High-Resolution Data
Location
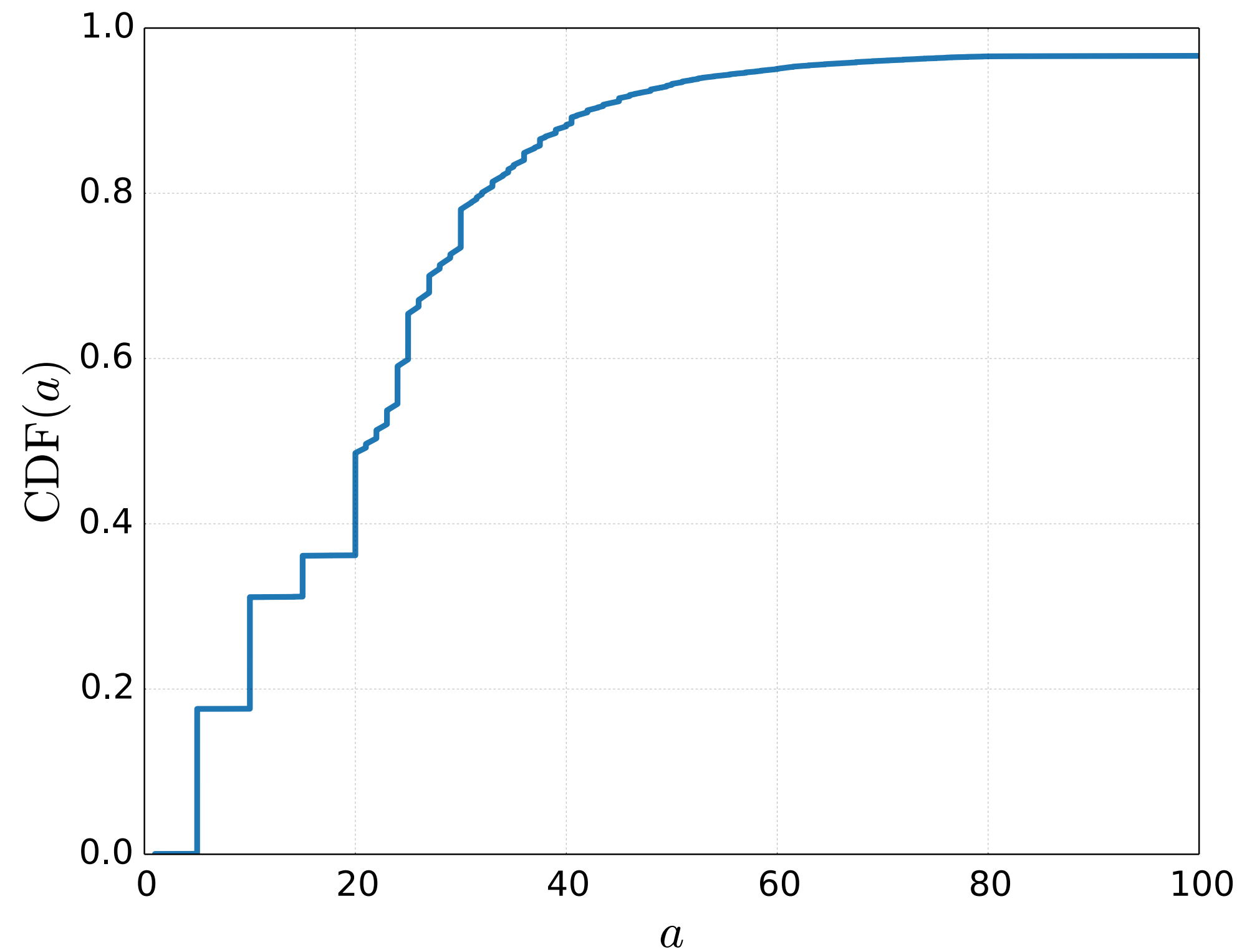
- high accuracy
- storing raw & real-time data for applications
- dense social system

Privacy: High-Resolution Data
Social

- ⊙ de-identification
- ⊙ deletion
- ⊙ feedback
  impact

Face-to-face

Texts

Calls

12 pm - 6 am   6 -12 am   12 am - 6 pm   6 - 12 pm

Privacy: High-Resolution Data
Social

- de-identification
- deletion
- feedback
  impact

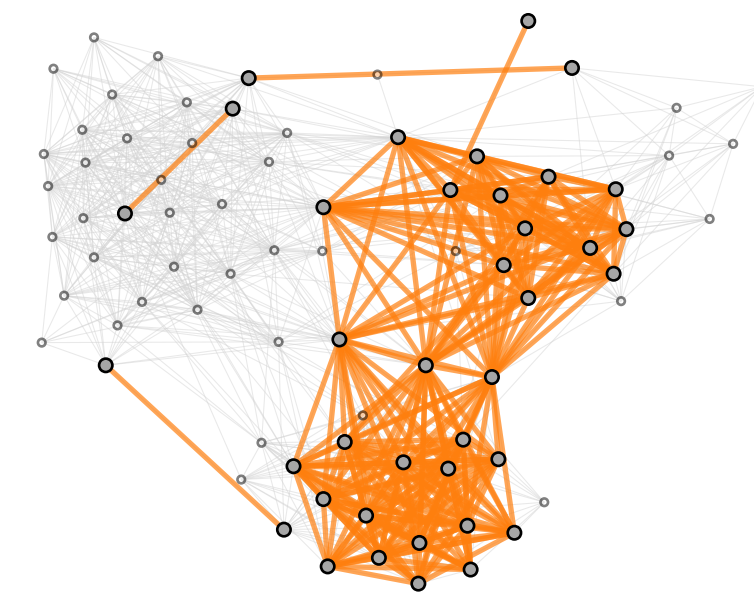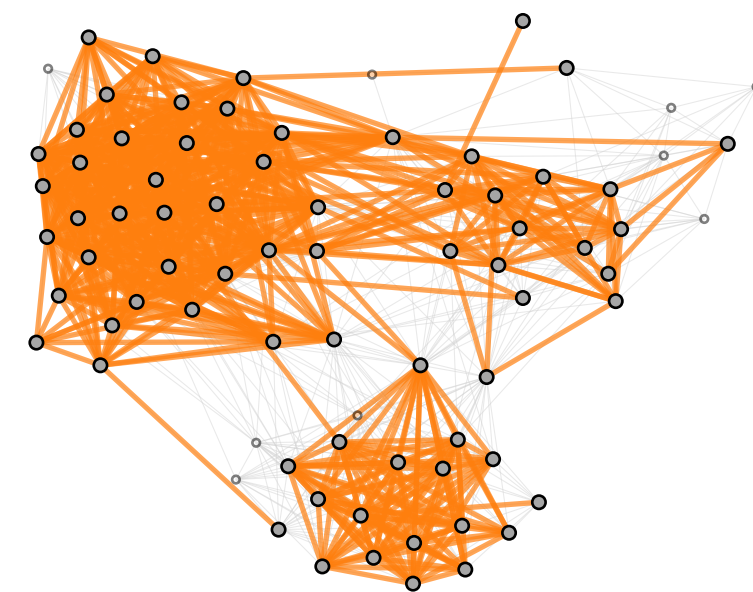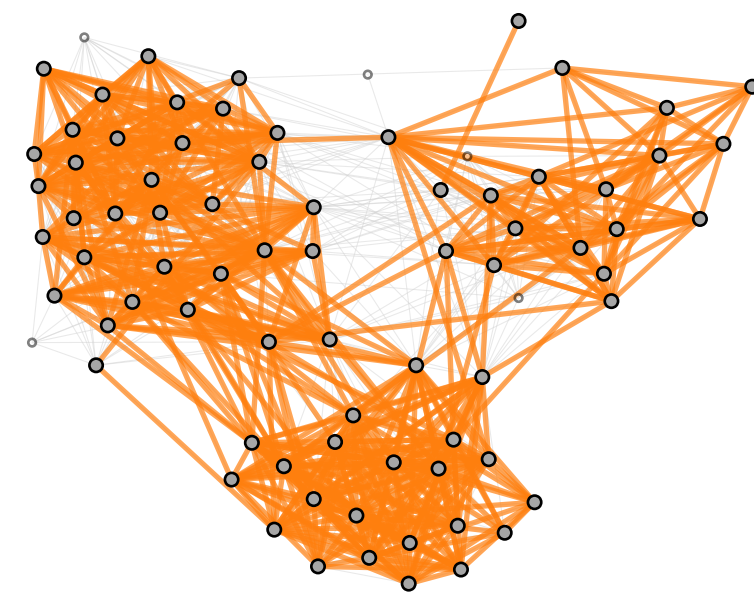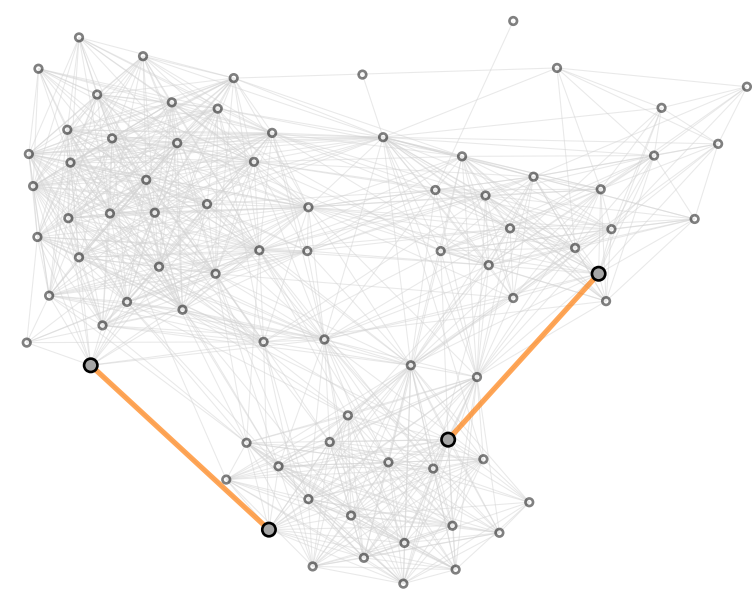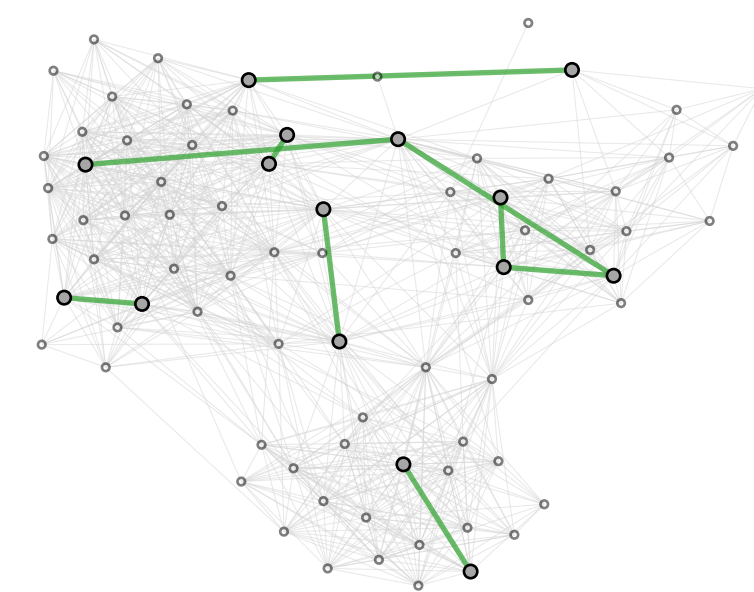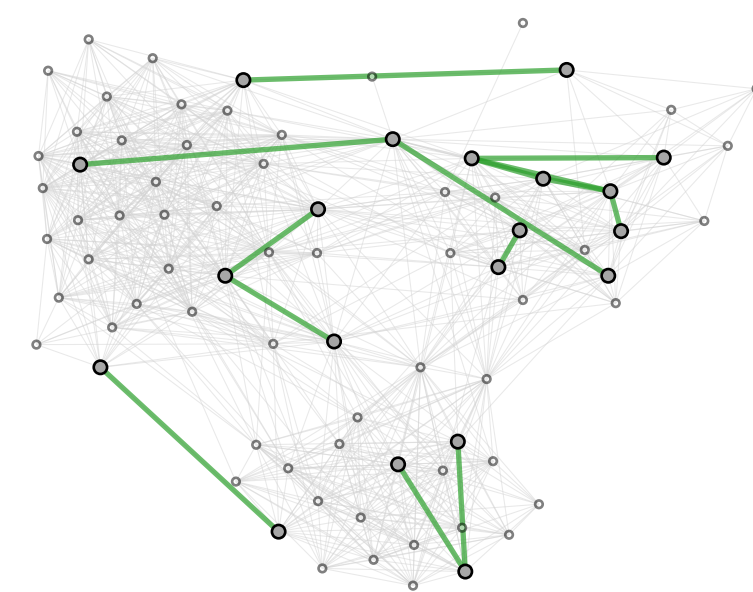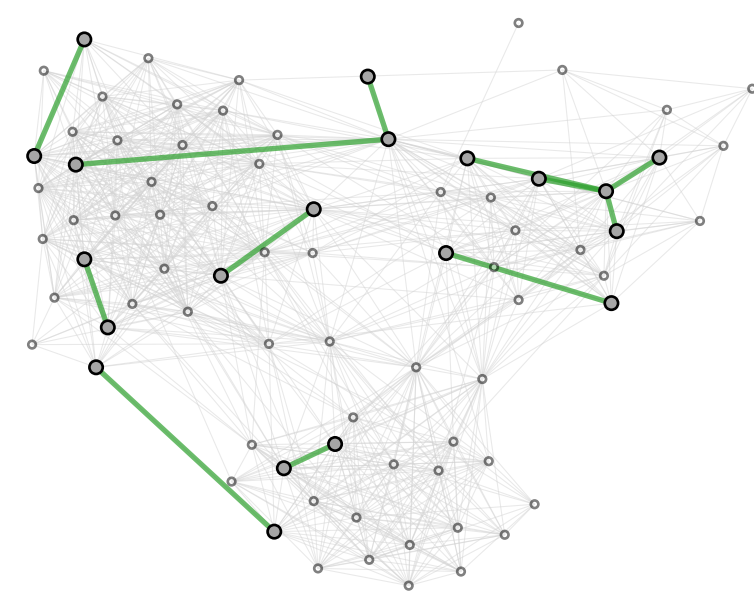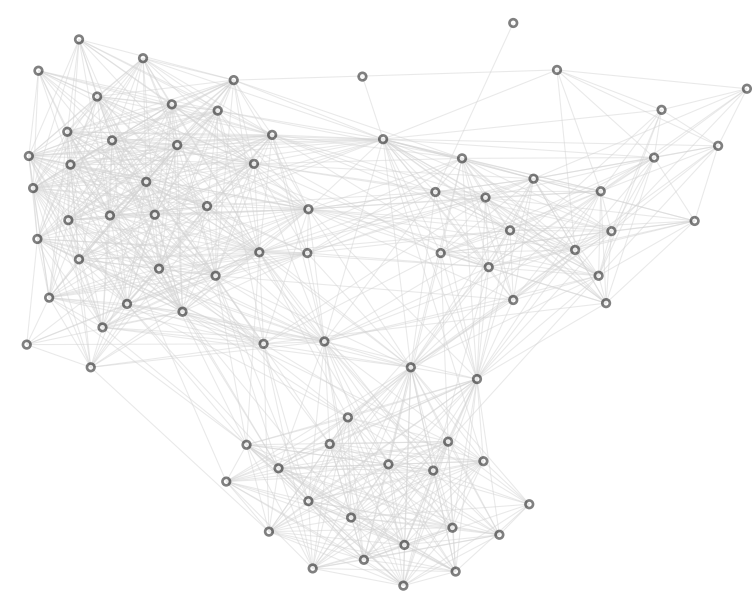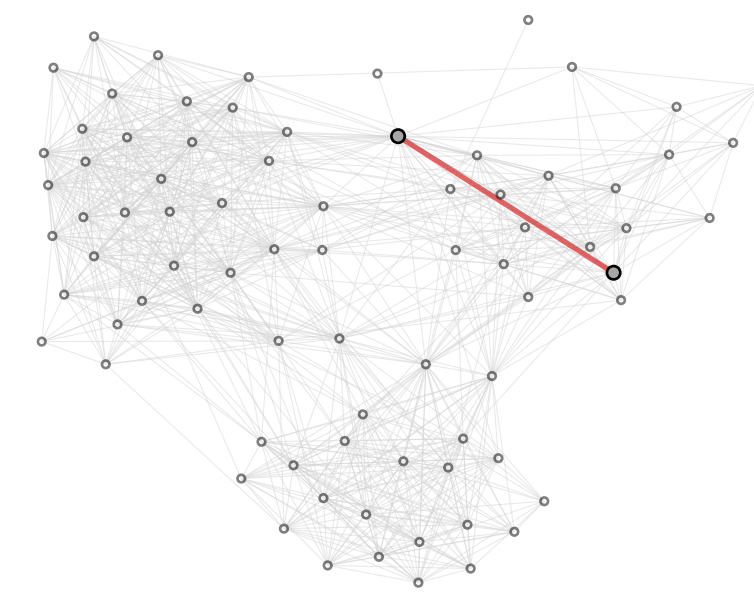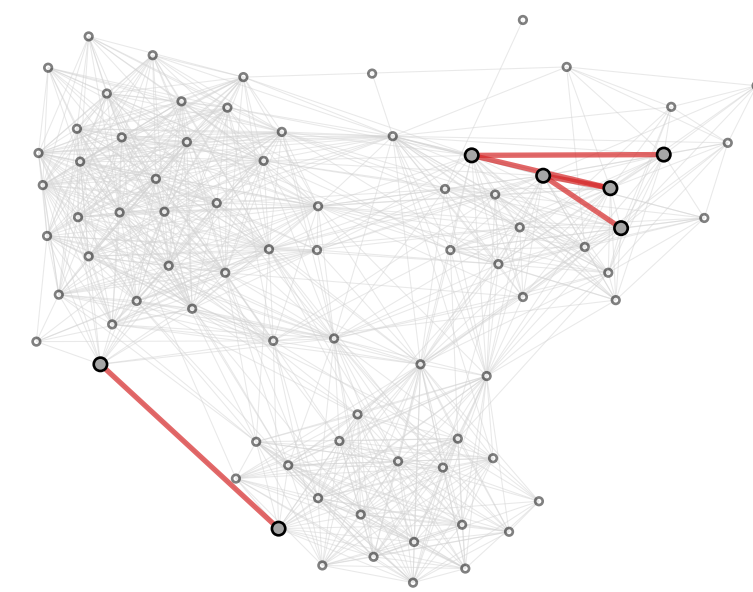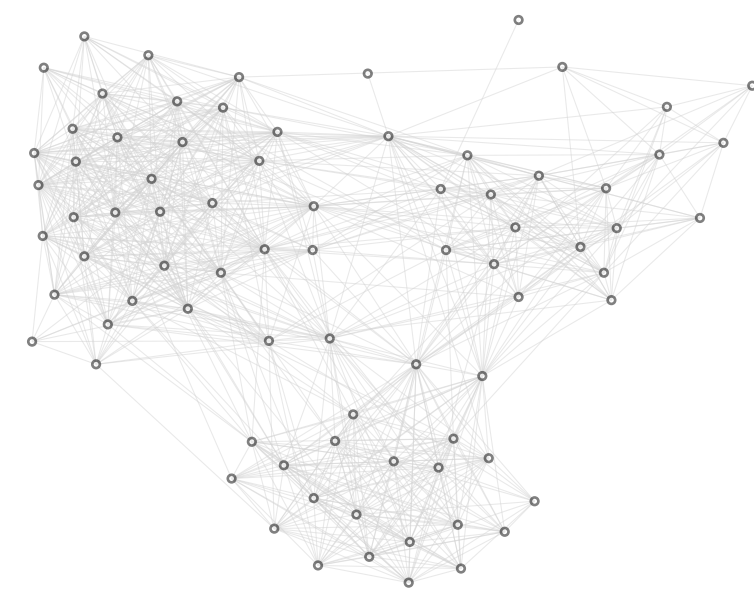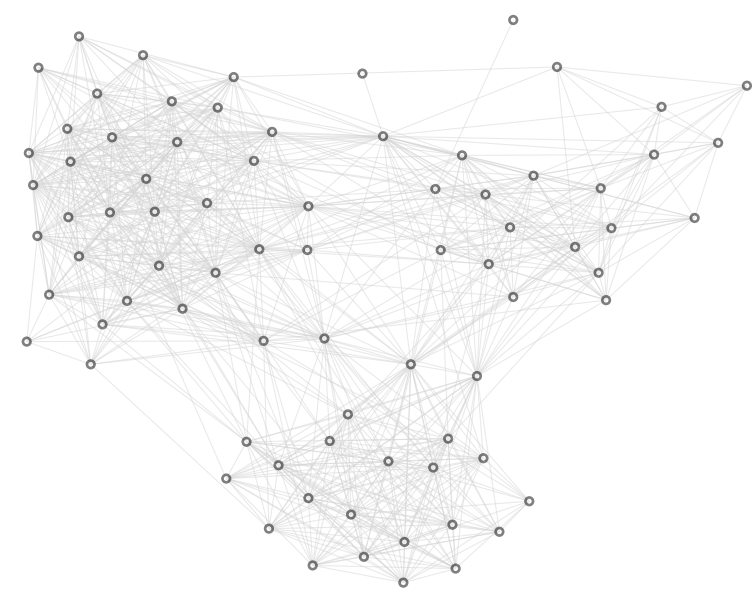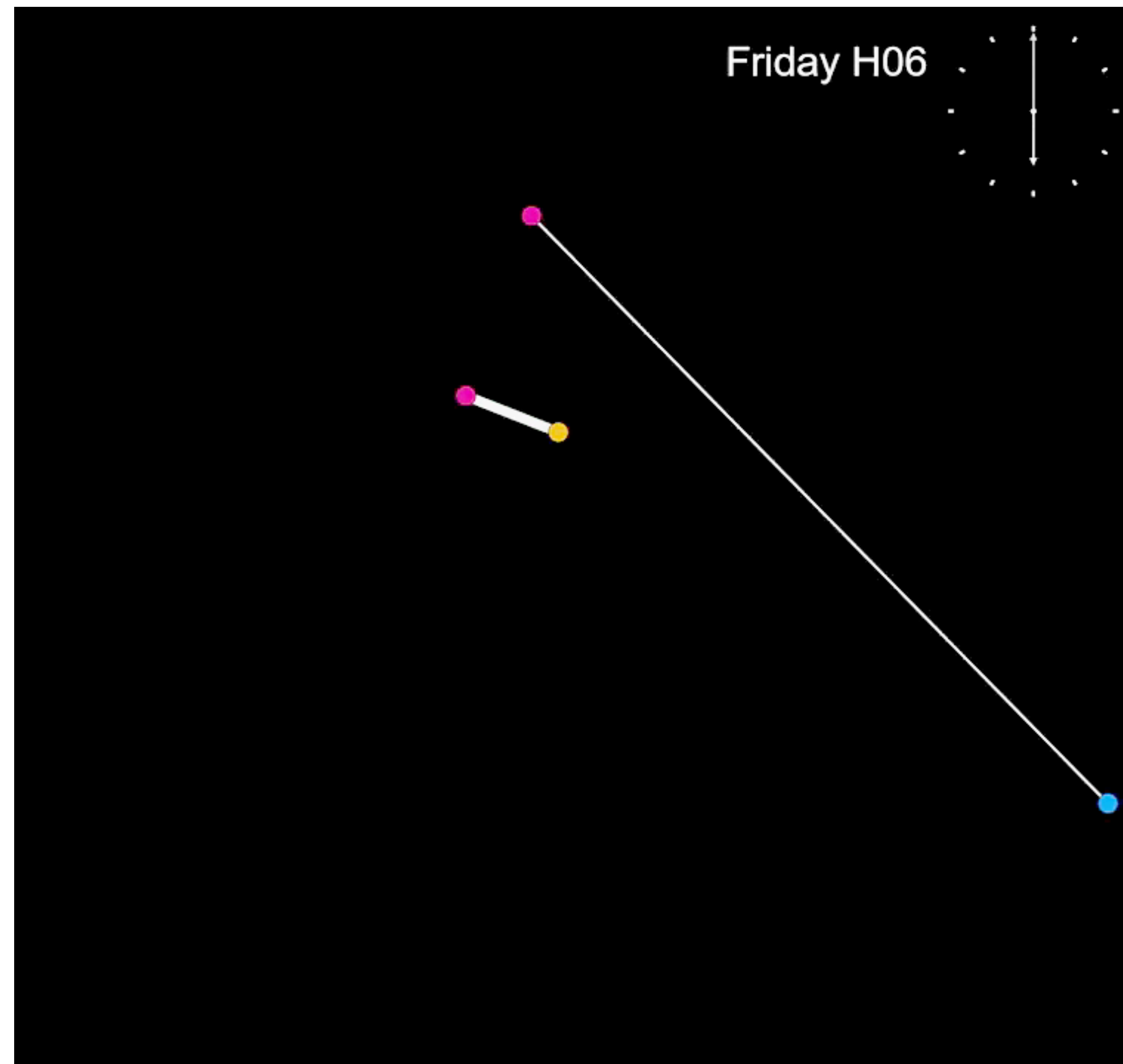# Privacy: High-Resolution Data
Network Privacy

- ⊙ neighborhood attacks
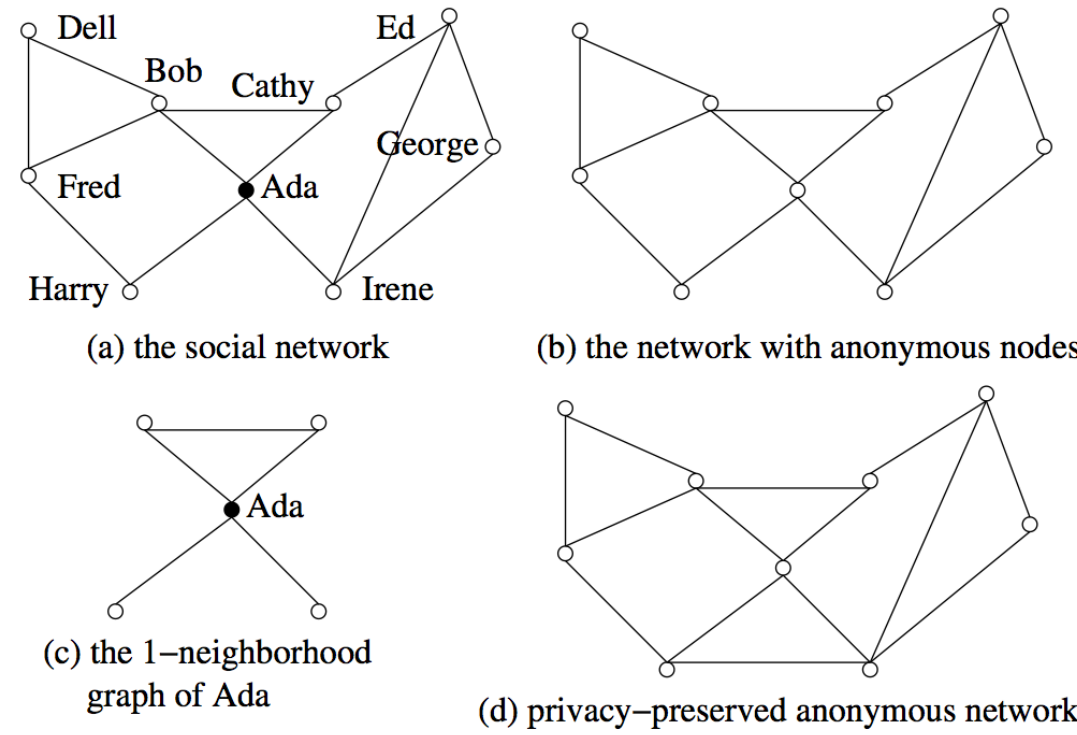- ⊙ NodeInfo, LinkInfo
- ⊙ α-proximity

Dell · Bob · Cathy · Ed · George · Fred · Ada · Harry · Irene

(a) the social network

(b) the network with anonymous nodes

Ada

(c) the 1−neighborhood graph of Ada

(d) privacy−preserved anonymous network

Fig. 1. Neighborhood attacks in a social network.

from 1.

(a) (x) (e) (a) (e')

(b) (f) (c') (r)

(c) (y) (r) (b) (f)

(d) (z) (s) (d) (s')

G   Gk
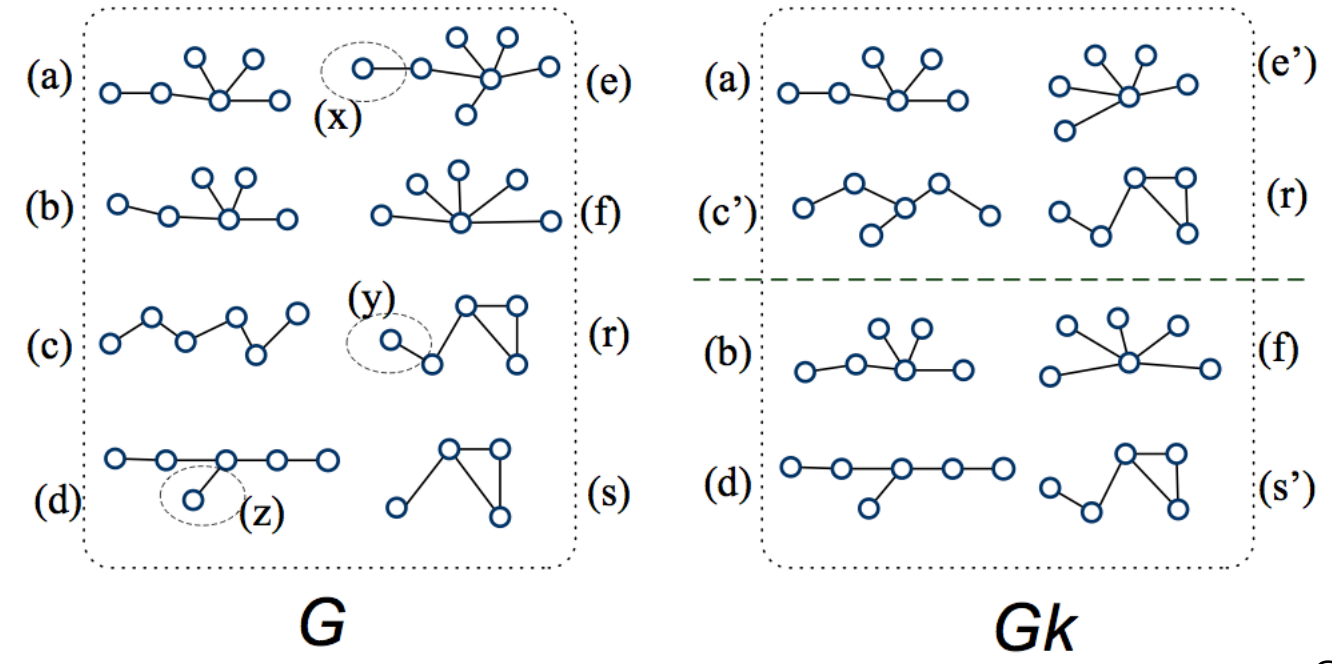
Figure 7: Graph G and Anonymized Graph G_k

from 2.

(a) A small social network

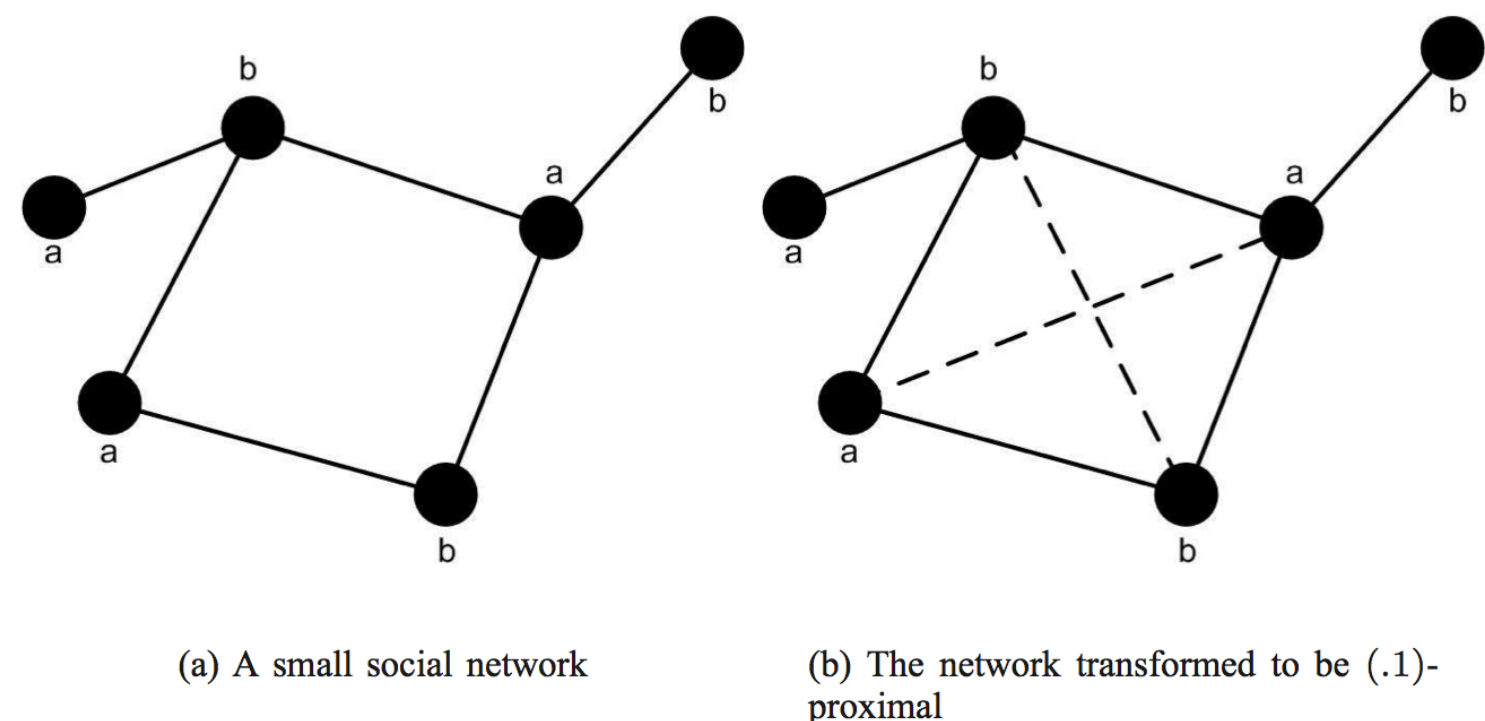(b) The network transformed to be (.1)-proximal

Fig. 1. Example of α-proximity

from 3.

1. Zhou, Bin, and Jian Pei. "Preserving privacy in social networks against neighborhood attacks." Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on. IEEE, 2008.
2. Cheng, James, Ada Wai-chee Fu, and Jia Liu. "K-isomorphism: privacy preserving network publication against structural attacks." Proceedings of the 2010 ACM SIGMOD International Conference on Management of data. ACM, 2010.
3. Chester, Sean, and Gautam Srivastava. "Social network privacy for attribute disclosure attacks." Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on. IEEE, 2011.

Privacy: High-Resolution Data
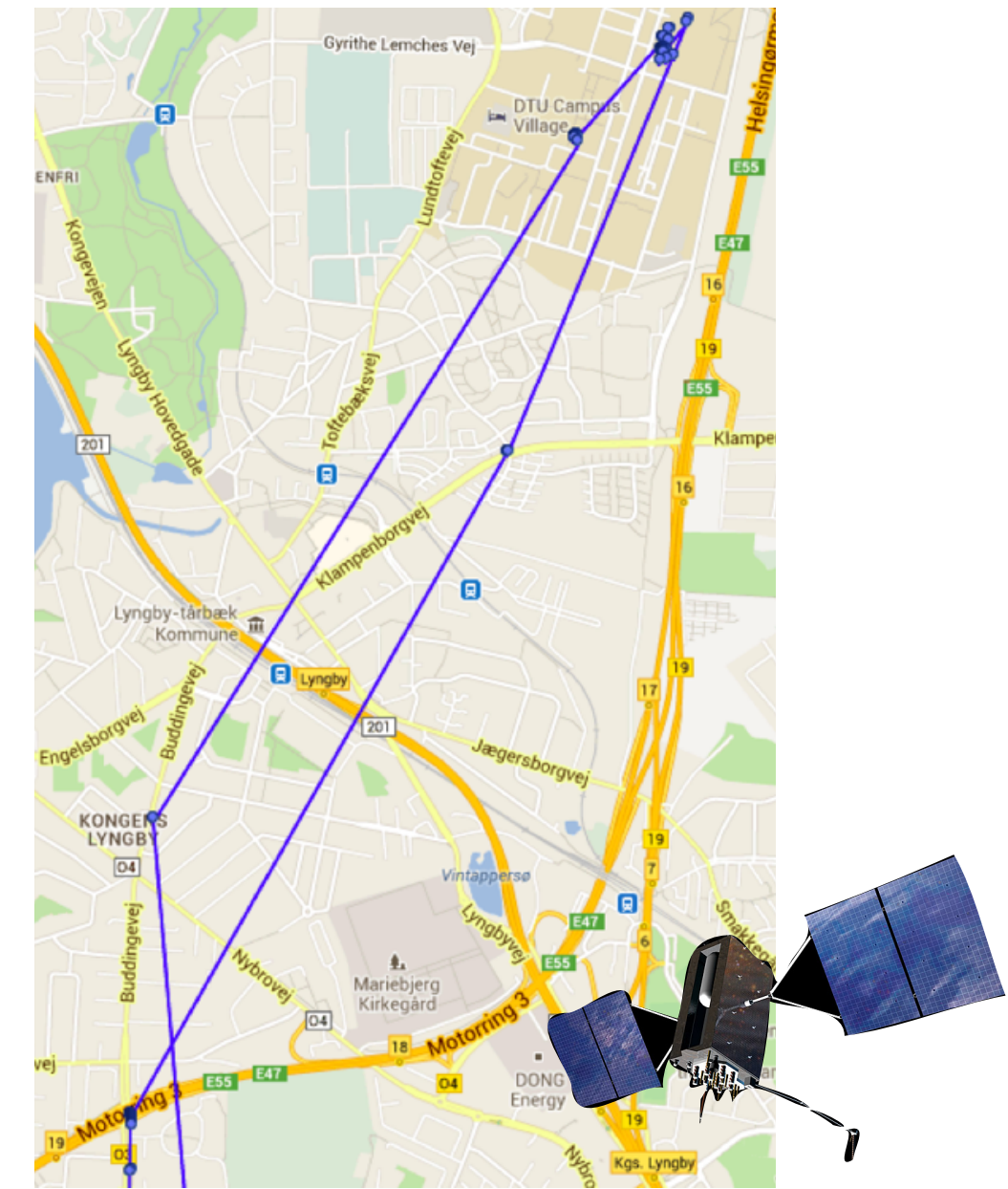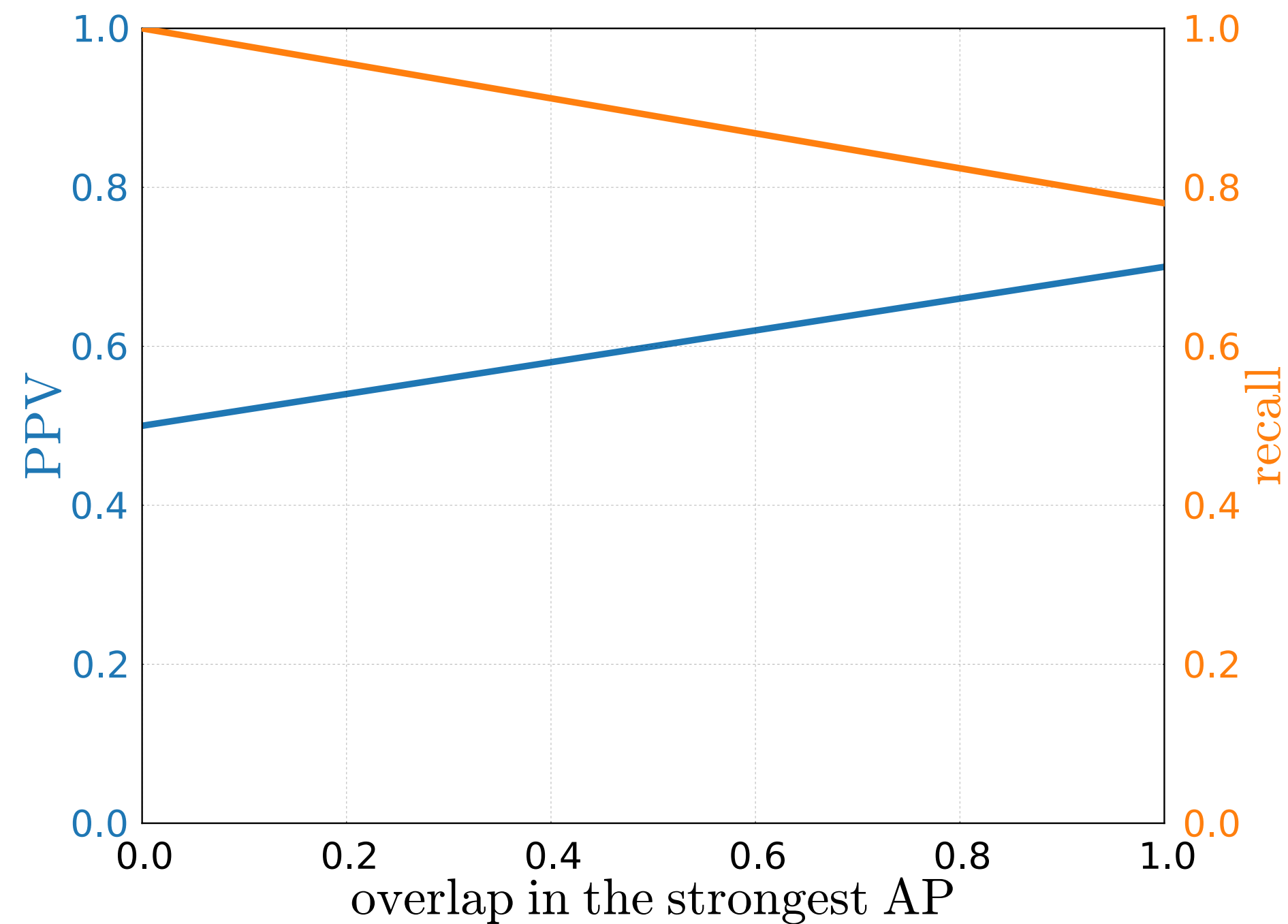## **Social Mobility**

- highly-linked data
- reconstructing individual traces
- distributed sensing
- quantification?
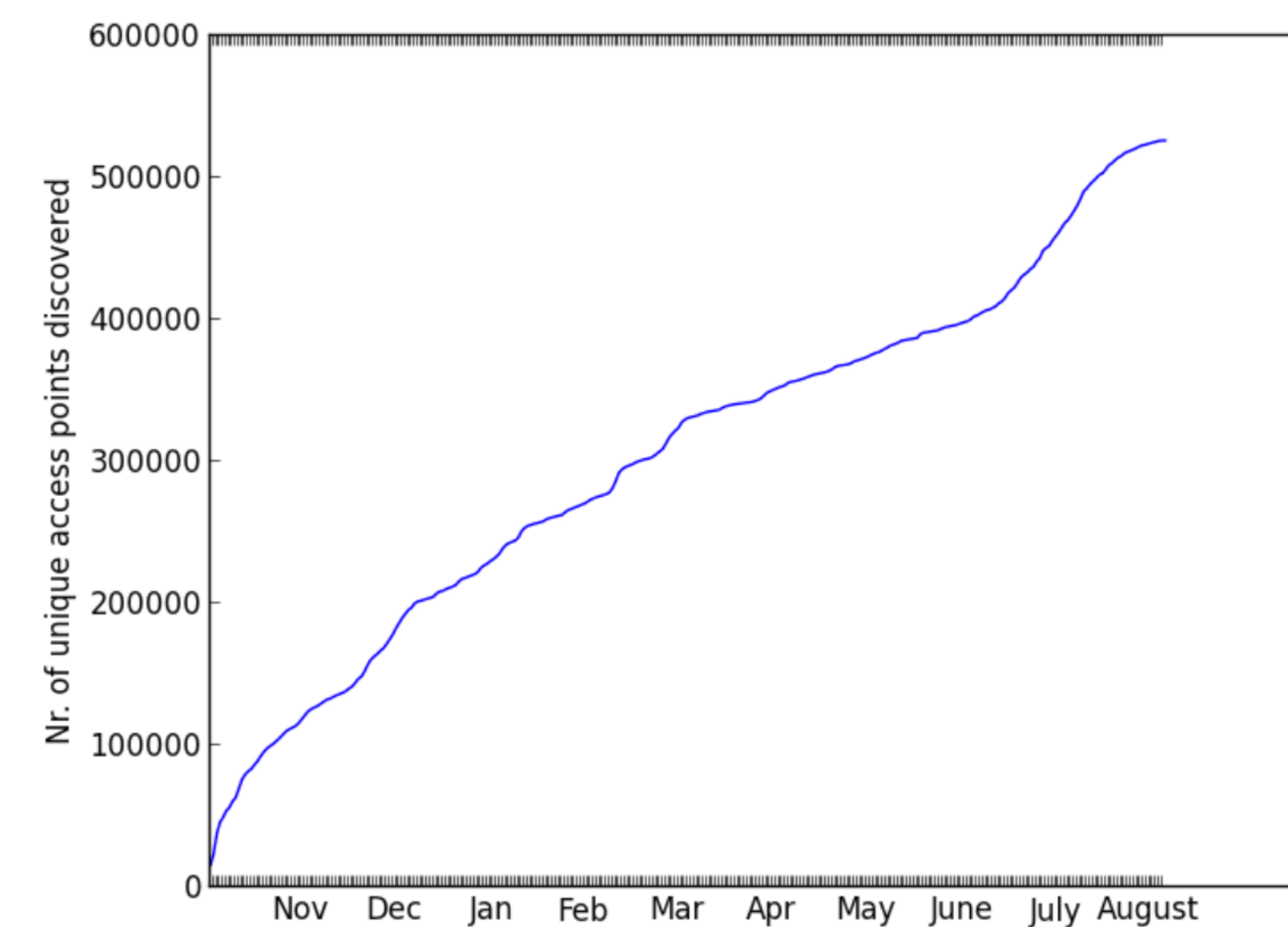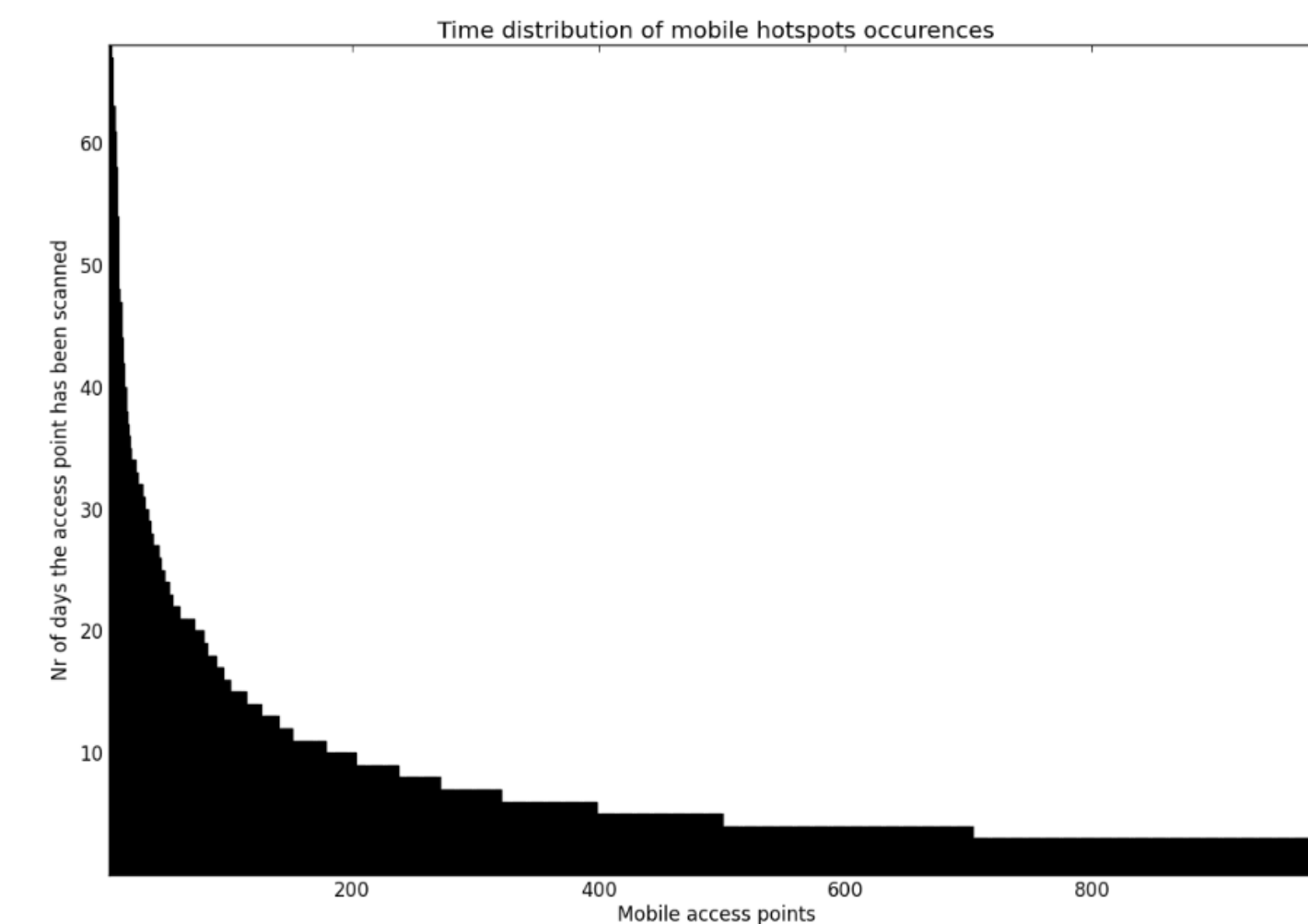
Privacy: High-Resolution Data
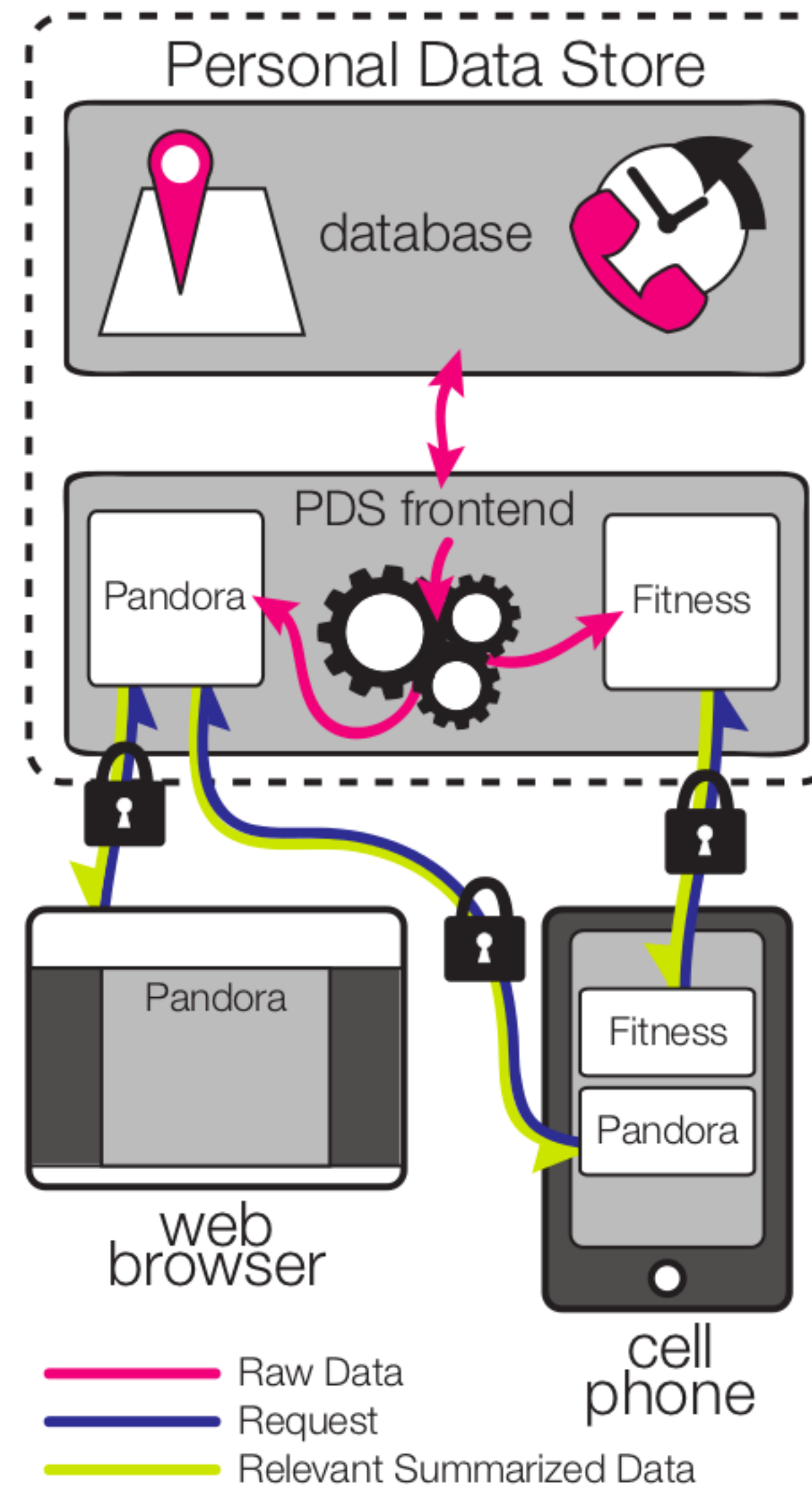# **Channel Redundancy**

- face-to-face from Wifi...

- ...or Facebook photos...

- ...or Facebook checkins

- location from Facebook...

- ...or Wifi

Privacy: High-Resolution Data
PDS

- ◉ Big Data is terrible to handle
- ◉ questions & answers
- ◉ researchers interested in aggregates
- ◉ providing value for researchers
- ◉ contracts

Personal Data Store

database

PDS frontend

Pandora

Fitness

Pandora

web browser

Fitness

Pandora

cell phone

Raw Data
Request
Relevant Summarized Data

from http://openpds.media.mit.edu/

DTU

Acknowledgments

Privacy Implications of Measuring Large-Scale Social Networks With High Resolution

Let's discuss...

Arkadiusz Stopczynski, Vedran Sekara, Piotr Sapiezynski, Andrea Cuttone, Mette My Madsen, Jakob Eg Larsen, and Sune Lehmann. *Measuring large-scale social networks with high resolution*. arXiv preprint arXiv:1401.7233 (2014)

Arkadiusz Stopczynski, Technical University of Denmark & MIT Media Lab
arks@dtu.dk, @h0pbeat, https://sensible.dtu.dk