The background features a complex network graph with nodes and edges. The graph is divided into three main color-coded clusters: a blue cluster on the left, a large orange cluster in the center, and a red cluster on the right. The nodes are small circles, and the edges are thin lines connecting them. The overall appearance is that of a large-scale social network visualization.

# **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

Stopczynski, A., Sekara, V., Sapiezynski, P., Cuttone, A., Madsen, M.M., Larsen, J.E., and Lehmann, S., arXiv preprint arXiv:1401.7233 (2014).

## Copenhagen Networks Study



**x 1 000**

face to face  
calls  
messages  
mobility

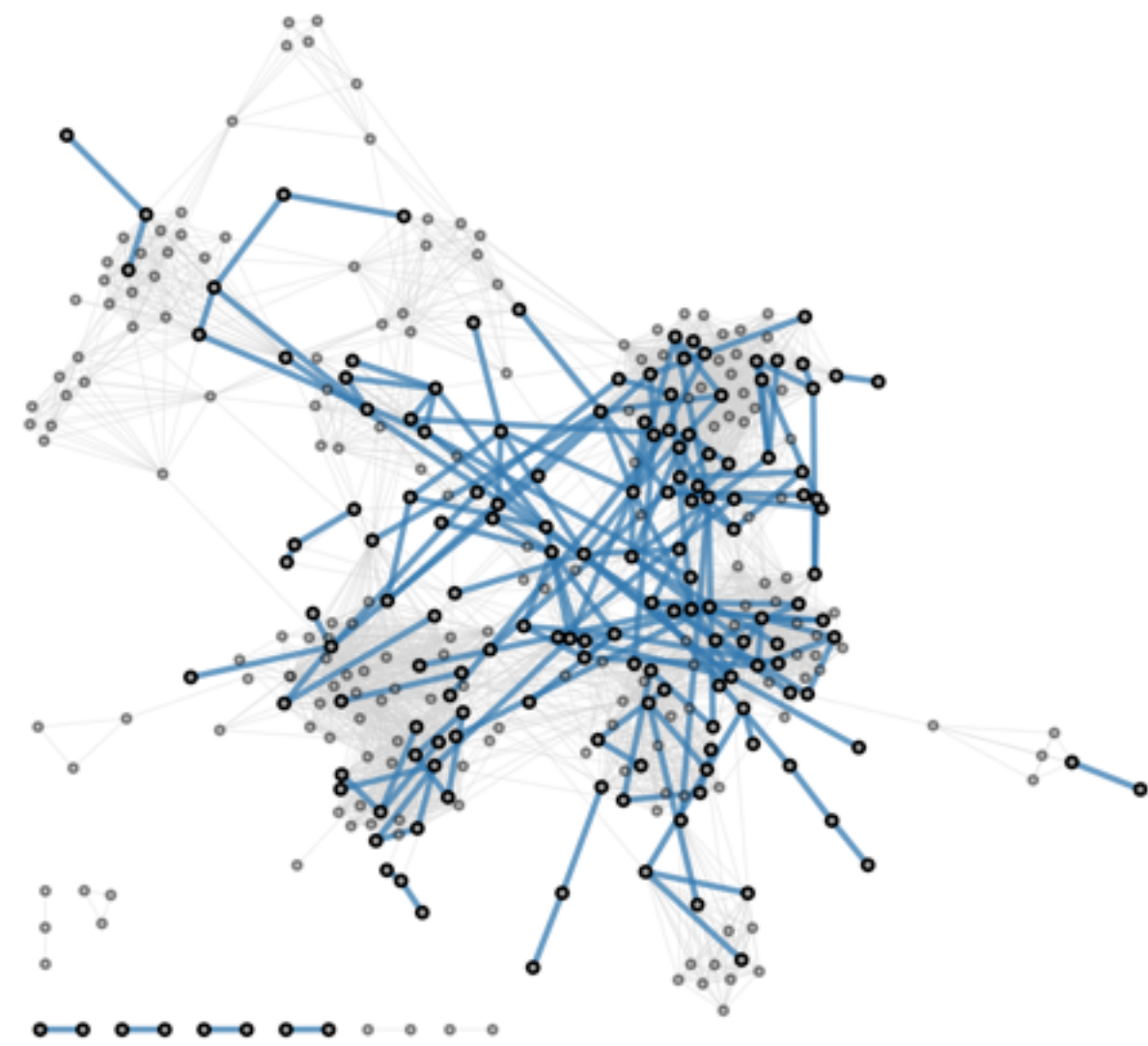


health  
psychology  
sociology  
demographics



friends  
interactions  
activity  
demographics

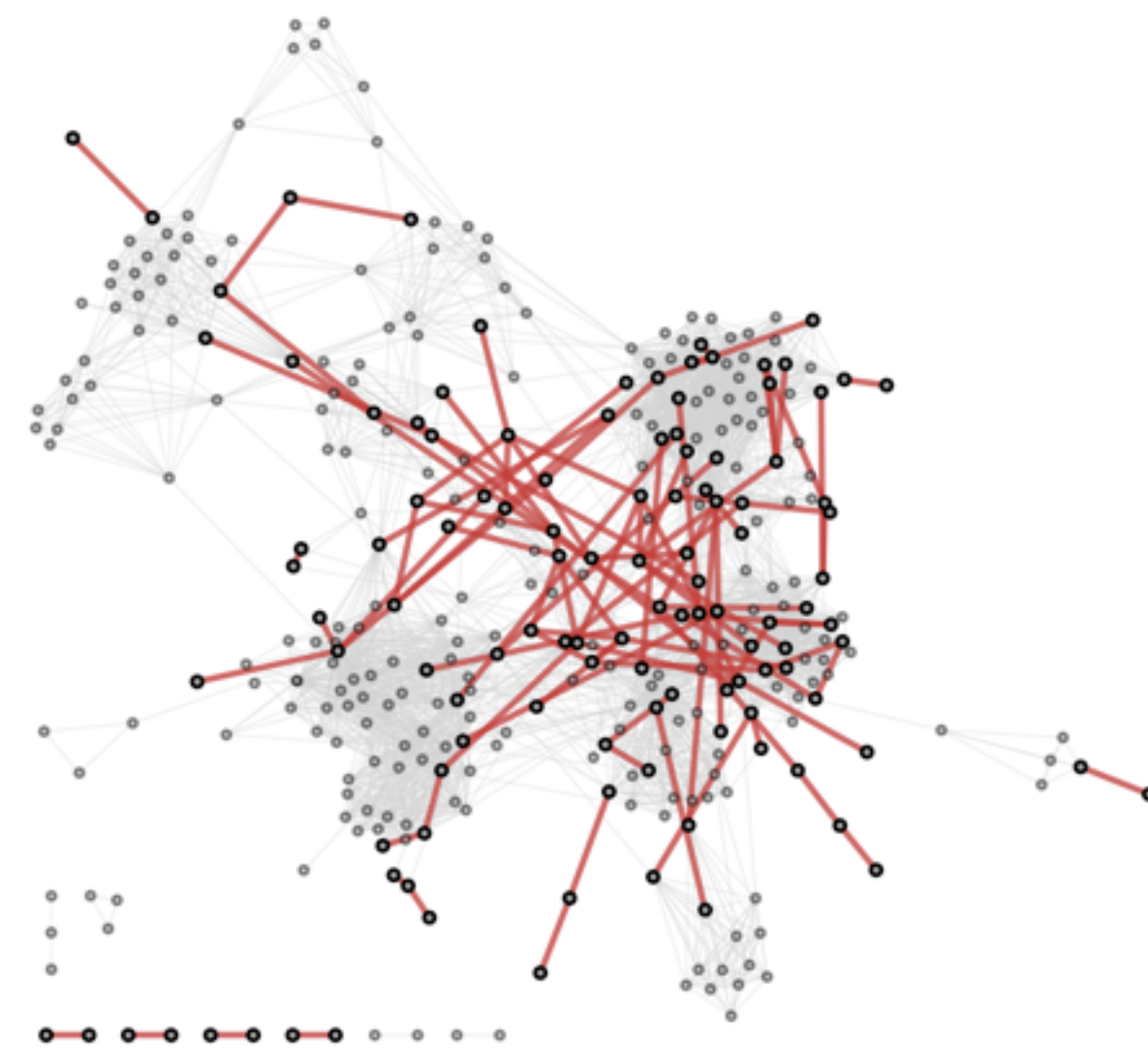
# Copenhagen Networks Study



Online



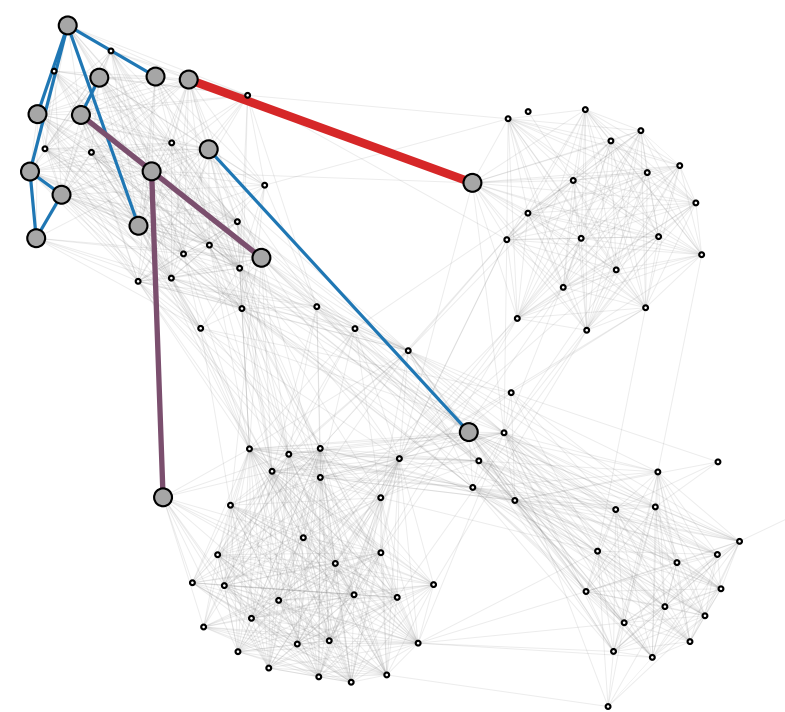
Face-to-face



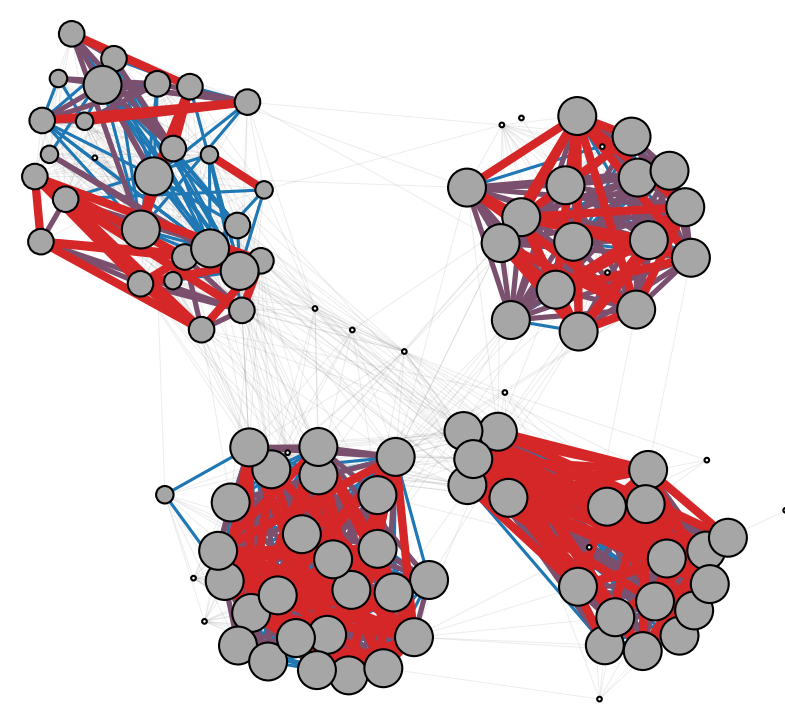
Extra Info. Online



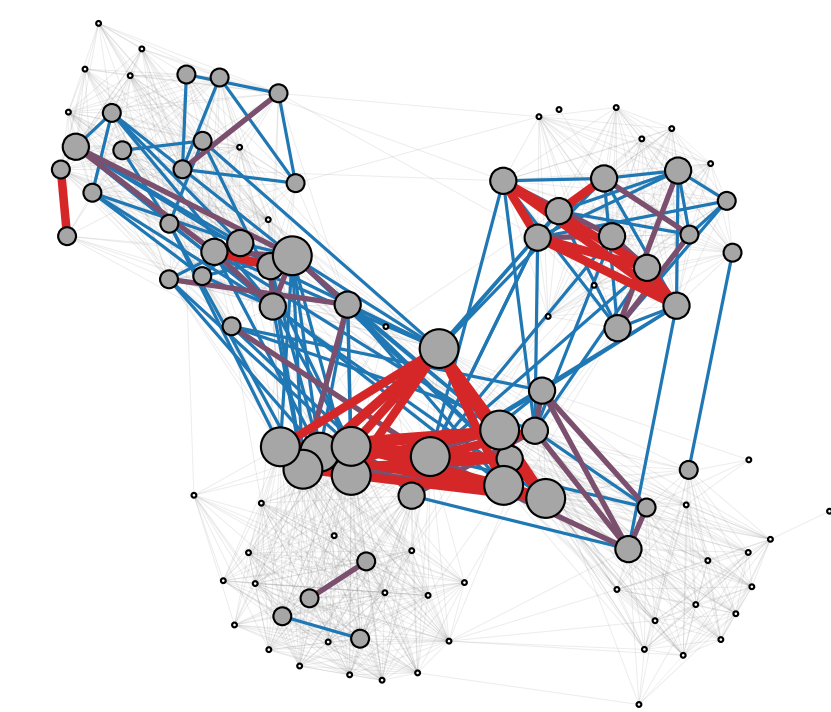
Extra Info. F2F



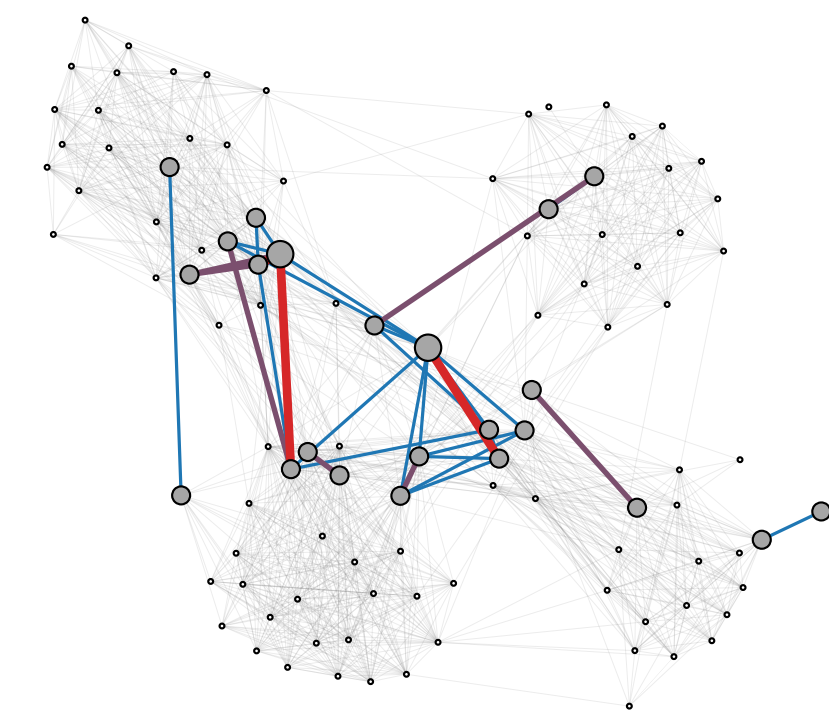
7:00



10:00

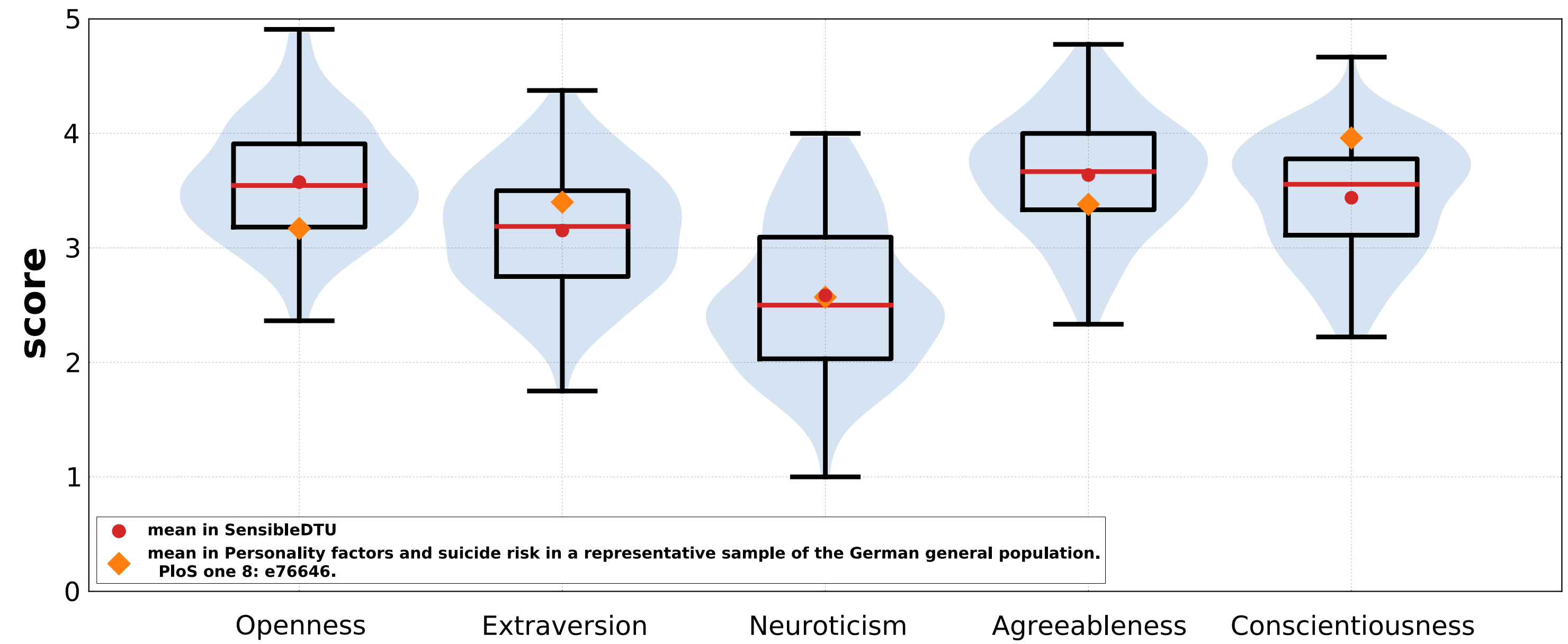
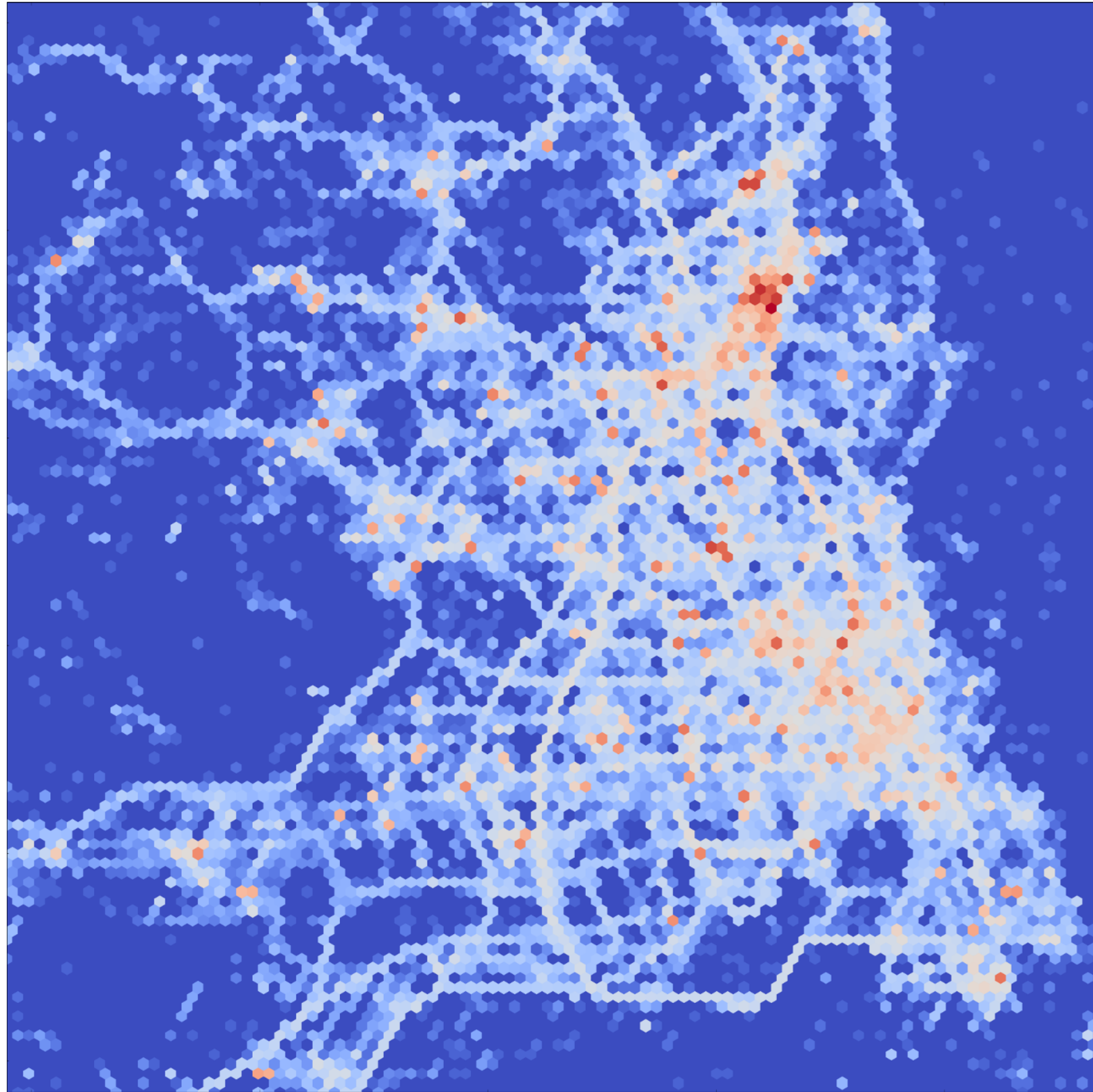


17:00



20:00

# Copenhagen Networks Study



# Copenhagen Networks Study



# Privacy: Users' Perspective



Sensible Data

About

Help

## Please sign in

Email

Password

Sign in

Register

Forgot your password?

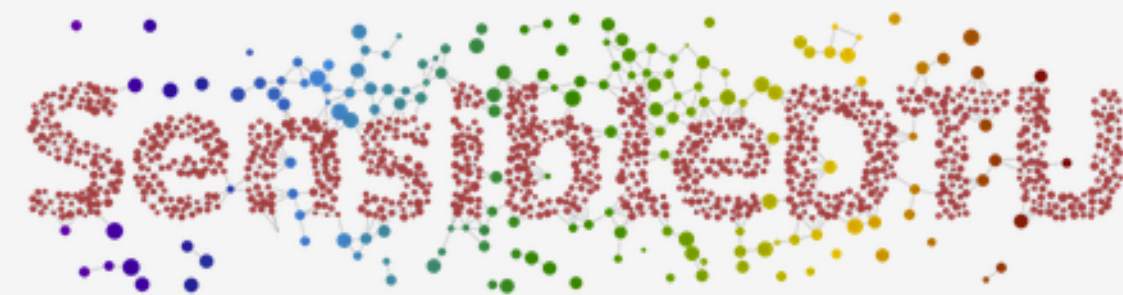
[Contact us](#) with your student number.

This is [Sensible Data](#), an [openPDS system 0.1.6](#)

licensed under a [Creative Commons Attribution 3.0 Unported License](#).

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

## Min profil



Username

First Name

Last Name

Phone Number

Save

DTU



Student ID

Email

Given Name

Family Name

# Privacy: Users' Perspective

The screenshot shows a mobile application interface for 'Sensible Data'. At the top, there is a navigation bar with a home icon, 'Home', 'Profile', 'About', and 'Help' links, and a user profile icon labeled 'Arkadiusz'. Below the navigation bar, the main content area contains several paragraphs of text in Danish, which are translations of the consent form terms. The text discusses the user's agreement to use a mobile phone for the project, the collection of data (Bluetooth, location, SMS, contacts, social media), and the user's right to withdraw consent at any time. At the bottom of the screen, there are two buttons: 'No' and 'Godkend' (Accept).

Sensible Data Home Profile About Help Arkadiusz

at jeg skal henvende mig til [support@sensible.dtu.dk](mailto:support@sensible.dtu.dk) hvis jeg har spørgsmål til Projektet eller ønsker at udtræde af Projektet.

at jeg til enhver tid kan udtræde af Projektet og levere mobiltelefonen tilbage.

Yderligere er undertegnede indforstået med, at det er en betingelse for min deltagelse i Projektet og for udlevering af telefon til brug under Projektet:

at jeg ved udlevering af telefon af mærket LG Nexus 4 til min brug under min deltagelse i Projektet, betaler et depositum på kr. 300,00. Dette depositum kan DTU bruge som sikkerhed for, at jeg afleverer telefonen tilbage i god stand ved min afslutning af deltagelse i Projektet eller når DTU beder mig om at tilbagelevere

at jeg får mit depositum retur når jeg returnerer telefonen og denne er i en almindelig god stand.

at jeg indsætter et SIM kort i telefonen og benytter den som min primære (mobil) telefon.

at jeg er indforstået med, at DTU til brug for Projektet, registrerer og indsamler data angående Bluetooth, mobil-master, WiFi netværk, opkaldte (men hashede/anonymiserede) telefonnumre og opkaldstid (intet indhold), afsendte SMS og tidspunkt (intet indhold; numre hashet som ovenfor), kontakter i telefonens telefonbog (ligeledes hashede informationer). Derudover indsamles oplysninger om forbindelser på online sociale networks (fx Facebook).

at jeg vil blive anmodet om at deltage i en række spørgeskemaundersøgelser. Hvis jeg vælger at deltage i disse, er jeg indforstået med, at resultaterne fra disse anvendes i anonymiseret form i forskningsprojektet på lige fod med den data indsamlet via min telefonbrug under Projektet.

at jeg er indforstået med, at den indhentede og registrerede data anvendes i det forskningsprojekt som Projektet er en del af med henblik på publicering af forskningsprojektets konklusioner.

at DTU ikke er ansvarlig for min brug af mobiltelefonen under Projektet. Det betyder, at hvis jeg eksempelvis via mobiltelefonen uploader, downloader eller videreformidler ophavsretligt beskyttet materiale på ulovlig vis, uploader, downloader eller videreformidler andet ulovlig materiale eller får eller spreder vira via mobiltelefonen, kan jeg ikke holde DTU ansvarlig for disse handlinger.

at mobiltelefonen jeg har fået udleveret til brug for Projektet er til låns og DTU vil kræve telefonen retur når forskningsprojektet afsluttes. DTU vil dog give mig et rimeligt varsel om tilbagelevering af telefonen på mindst 2 måneder, medmindre DTU's krav om tilbagelevering er begrundet i mit brug af eller behandling af telefonen.

at DTU kan rejse krav om erstatning overfor mig, hvis jeg med vilje eller ved grov uagtsomhed ødelægger eller beskadiger mobiltelefonen eller eksempelvis anvender mobiltelefonen til ulovlige aktiviteter eller andet, som påfører DTU et tab.

at DTU ikke kan holdes ansvarlig for eventuelle skattemæssige konsekvenser i forbindelse med mit lån og brug af mobiltelefonen under Forsøget.

Ved min godkende godkendelse tiltræder jeg hermed denne Tiltrædelseserklæring.

Arkadiusz Stopczynski. [Not you?](#)

No Godkend

- 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

## Mine Projekter

Denne side indeholder en oversigt over de projekter jeg deltagaer i. I øjeblikket er det kun SensibleDTU, men der kunne potentielt være andre.

### SensibleDTU

Stort projekt hvor smartphones møder forskning i sociale systemer!

#### Applikationer

Nedenfor ser du en oversigt over de applikationer du har godkendt til studiet. Indtil videre har vi kun spørgeskemaet klar, men det bliver fx her at du skal godkende de apps der kører på telefonen for at indsamle data, osv.

#### Questionnaire >>

Denne applikation er et spørgeskema du skal udfylde inden du kan få telefonen udleveret. Når du har udfyldt spørgeskemaet kan du booke en tid til at få telefonen udleveret.

is not allowed to submit your answers

Allow >

#### Phone Data Collector

##### can submit data from your phone

- Battery, Time Offset, Screen On/Off
- Hardware Info
- Location
- Bluetooth scans
- WiFi scans
- Cell ID
- Call & SMS Log (no names, no numbers, no content)
- Contacts (no names, no numbers)

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

### Allow Sensible Explorer to

access your location data

access your Bluetooth data

Bluetooth is used to see people you are meeting

access your wifi data

No

Yes



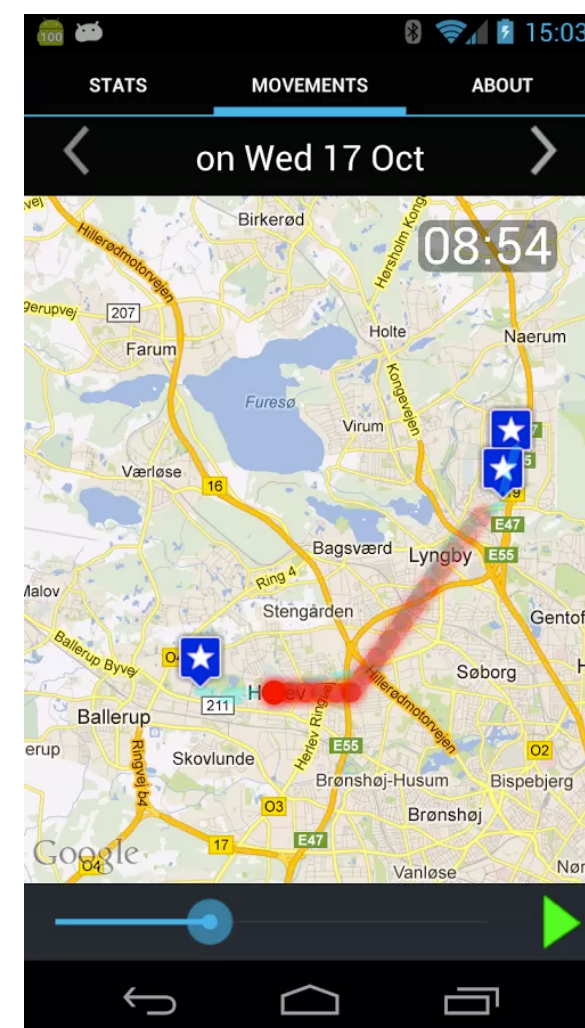
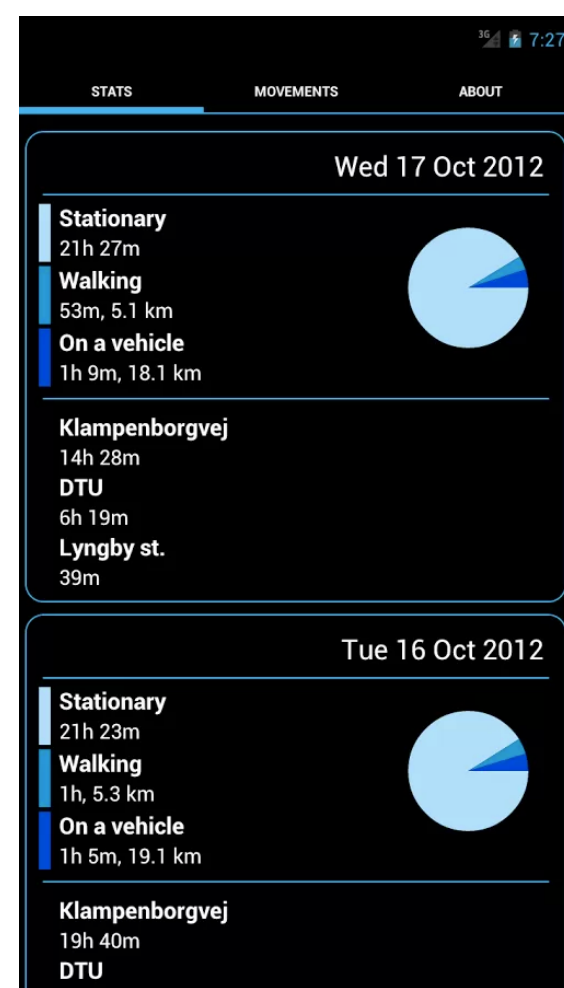
# Privacy: Users' Perspective

The screenshot shows the SensibleDTU: Questionnaires web application. At the top, there is a navigation bar with a logo, the text "SensibleDTU: Questionnaires", and links for "Home", "About", and "Quit". A user profile button labeled "Piotr" is on the right. Below the navigation bar, a progress bar indicates "ca. 64%". The main content area displays a survey question: "Har du inden for de sidste 14 dage taget piller eller medicin..?". Below the question is a table with seven columns representing different frequency options: "Ja, flere gange dagligt", "Ja, 1 gang dagligt", "Ja, 2-6 gange om ugen", "Ja, 1 gang om ugen", "Ja, sjældnere", and "Nej, slet ikke". The rows list various symptoms: "mod hovedpine", "mod smerter andre steder", "mod søvnproblemer", "mod nedtrykthed", "mod nervøsitet", "der fremmer din koncentrationsevne", and "der fremmer din fysiske præstation". Each cell in the table contains a radio button. At the bottom of the survey area, there are navigation buttons: "Næste ubesvarede >>", "< Forrige", and "Næste >".

	Ja, flere gange dagligt	Ja, 1 gang dagligt	Ja, 2-6 gange om ugen	Ja, 1 gang om ugen	Ja, sjældnere	Nej, slet ikke
mod hovedpine	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
mod smerter andre steder	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
mod søvnproblemer	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
mod nedtrykthed	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
mod nervøsitet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
der fremmer din koncentrationsevne	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
der fremmer din fysiske præstation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

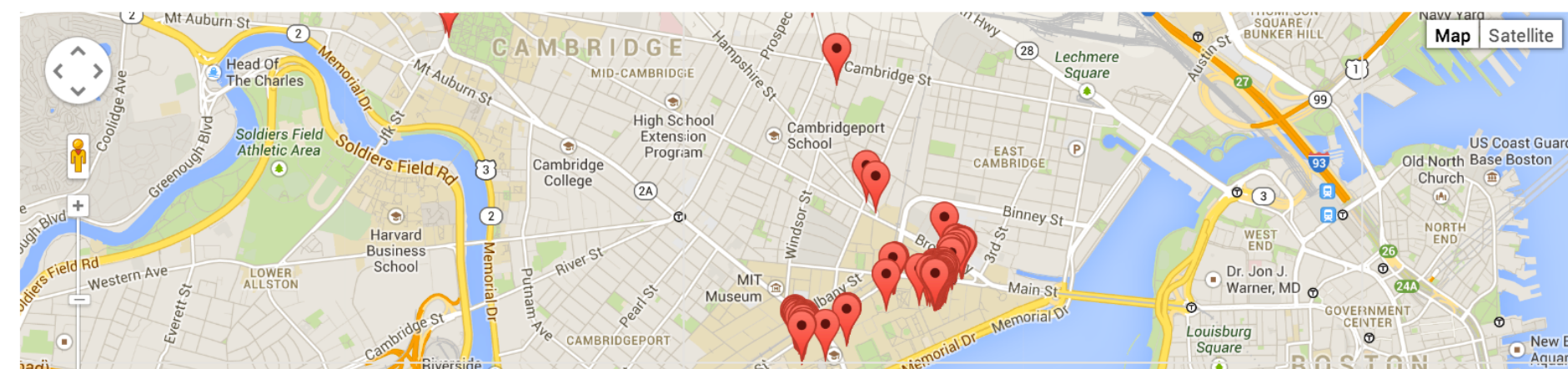
This is SensibleDTU: Questionnaires **0.1.8** © Technical University of Denmark 2013 | software license  
licensed under a [Creative Commons Attribution 3.0 Unported License](#).

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



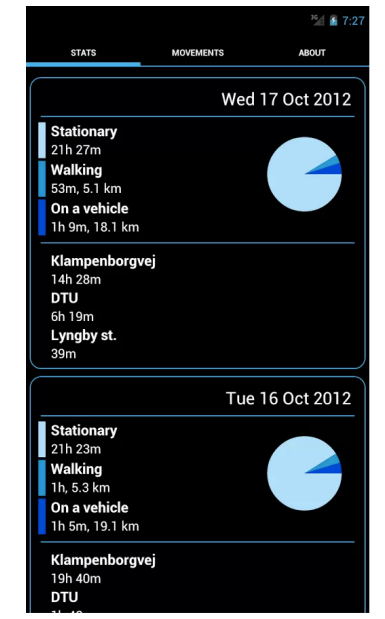
# Privacy: Users' Perspective

The screenshot shows the SensibleDTU: Data Viewer interface. At the top, there is a navigation bar with a home icon, 'About', 'Quit', a language dropdown set to 'English (en)', and a 'Set language' button. A user profile for 'Arkadiusz' is visible. Below the navigation bar, there are two tokens: 'Researcher token 79da9d9c251c6fac8bbf30630d4f79' and 'User token 4dc6e202586f22ec2970e339336b78'. The main content area is titled 'RAW DATA' and 'Location'. It includes a sidebar with categories: USER, QUESTIONNAIRES, PHONE DATA (with sub-items: Location, Bluetooth, Call Log, SMS, Wifi), FACEBOOK DATA (with sub-items: Birthday, Education, Friends, Friend Lists, Groups, Hometown, Interests, Likes, Location, Political, Religion), and ANSWERS. The 'Location' section has a 'Decrypted' checkbox checked, a 'Sort by' dropdown set to 'Descending', and a 'Fields to return' input field. There are also 'Start Date' and 'End Date' dropdowns set to 'epoch timestamp', and a 'No. of documents (max. 1000)' dropdown set to '1000'. The 'Visualization' section has radio buttons for 'Json Pretty', 'Json Format', 'CSV format', and 'Map'. A 'Submit' button is present, and below it, a message says 'took 3379 ms'. At the bottom, there is a 'Next' button and a 'Get Text' button. A progress bar shows '999 of 1,000 points selected.' Below the progress bar is a timeline visualization showing a period from 06 PM on Fri 15 to 06 AM on Sat 16.



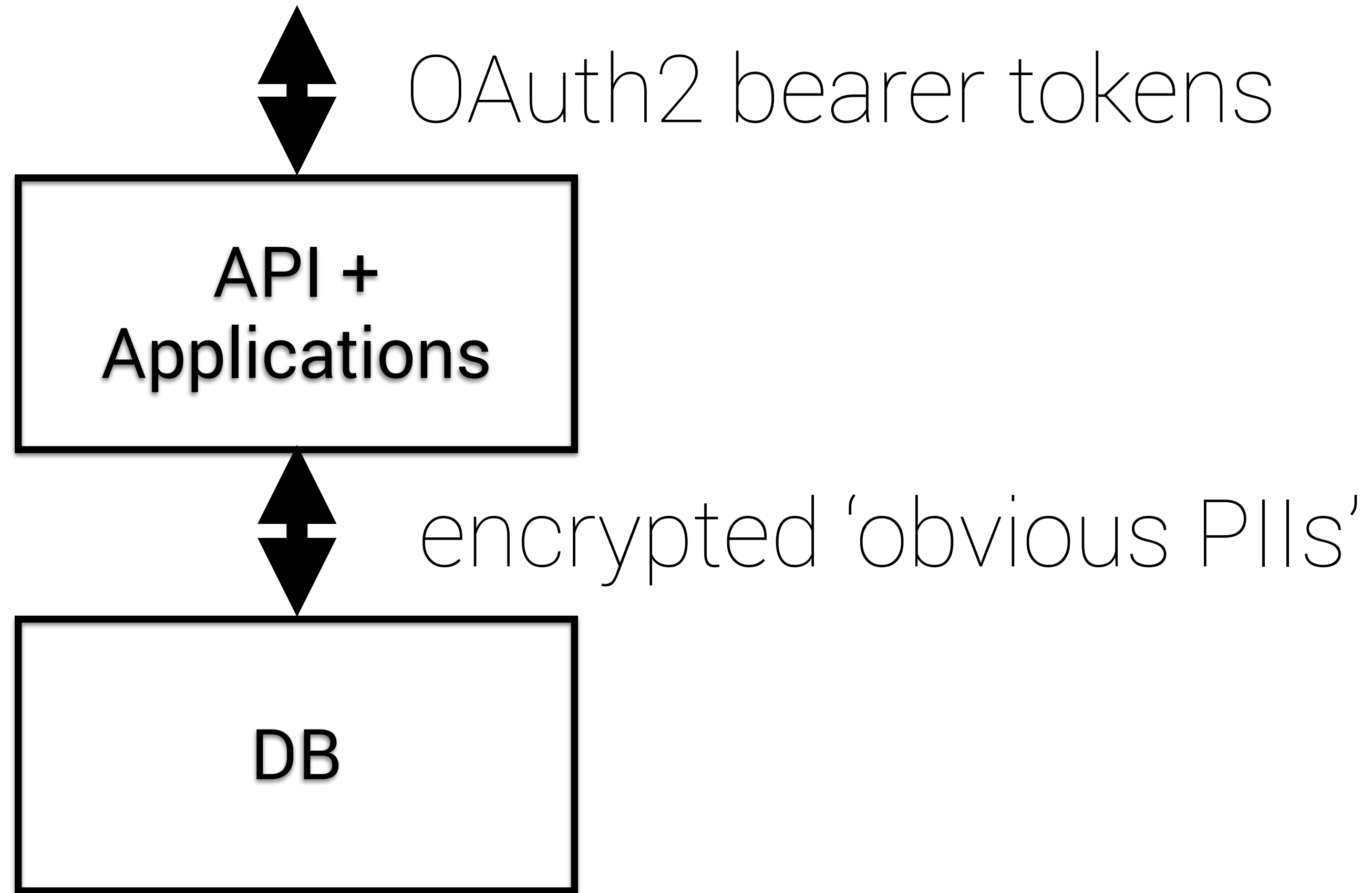
- data viewer
- access to all collected data
- same API as used by applications & researchers
- building URIs
- simple visualizations
- but of course this is not enough...
- users' perspective privacy research testbed

# Privacy: Security

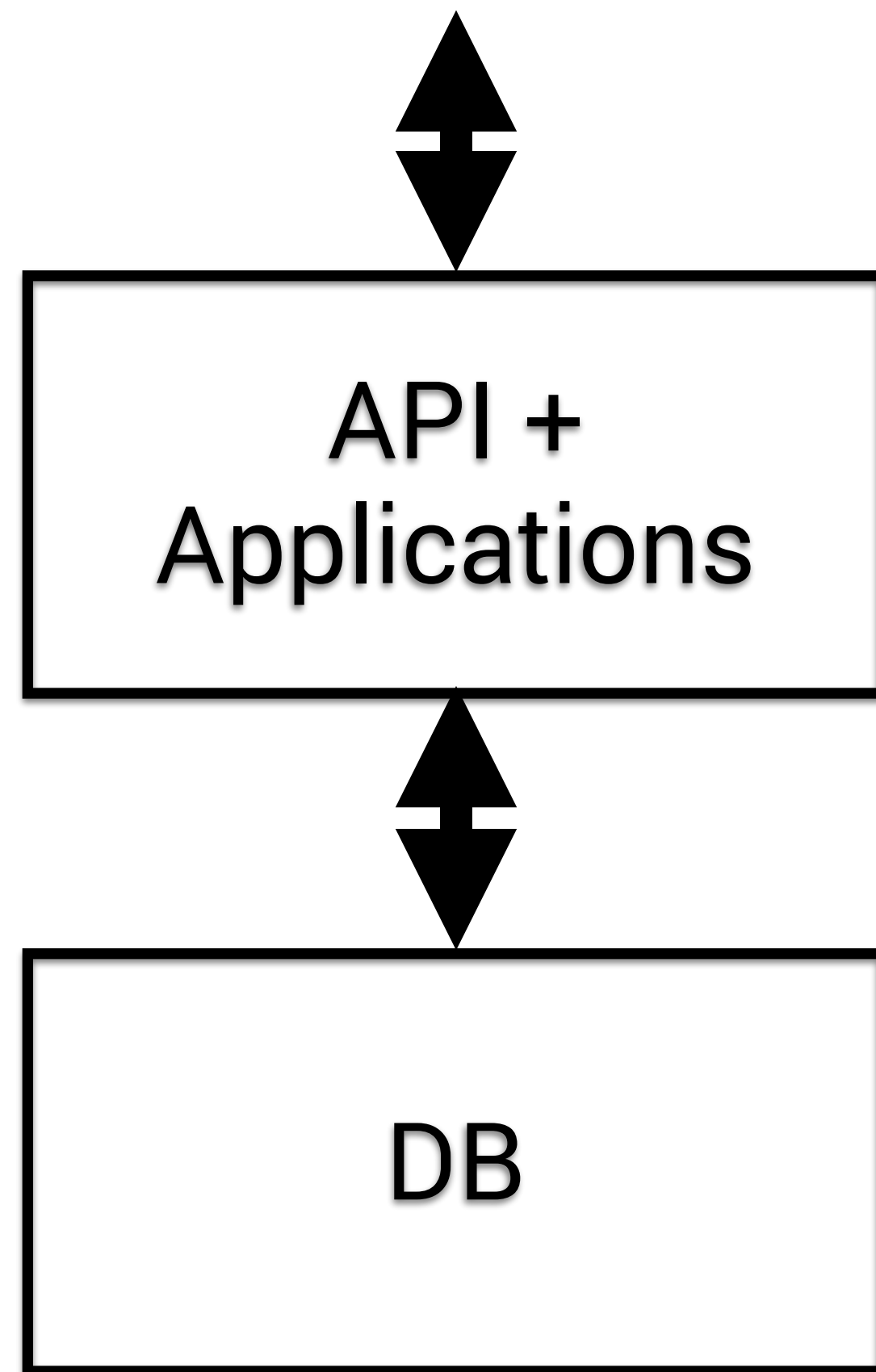
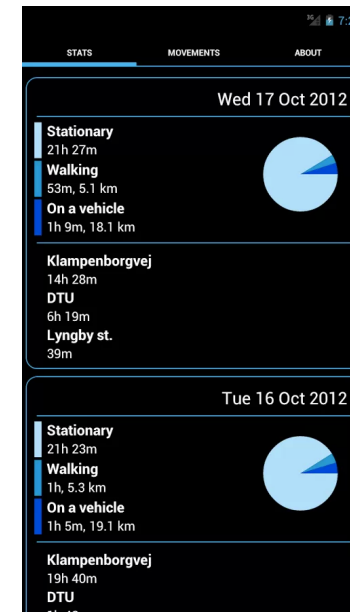


DTU

Amazon AWS  
(Ireland)



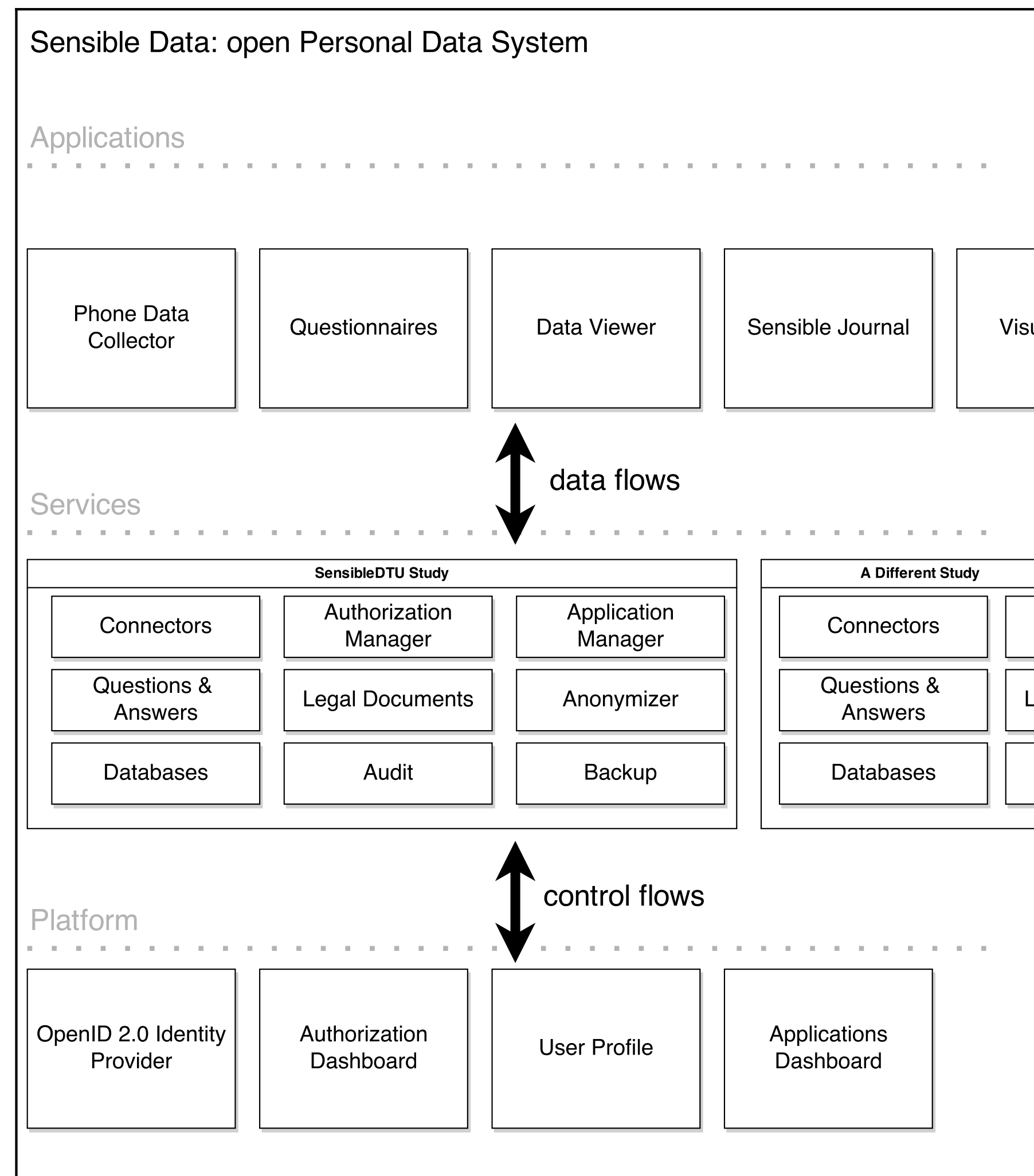
# Privacy: Security



- 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)

# Privacy: Security

- 3-layer architecture
- multiple studies with common dashboard
- extensible code



# Privacy: High-Resolution Data



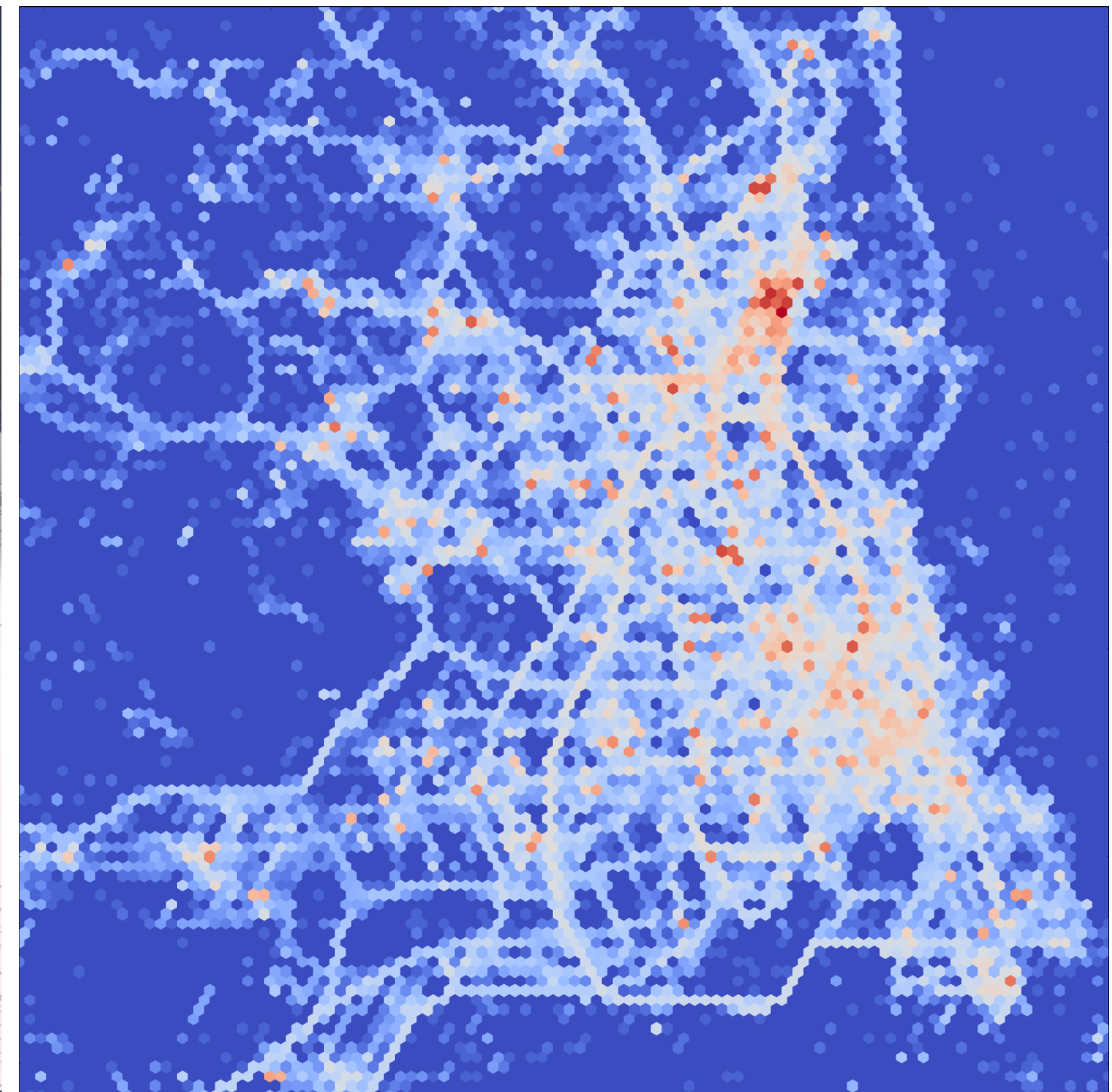
- Bluetooth (5 minutes)
- location (10 minutes)
- Wifi (20 seconds)
- cell ID
- screen state
- call/SMS log
- contacts

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

- friends
- interactions
- checkins
- posts
- work
- education
- demographics

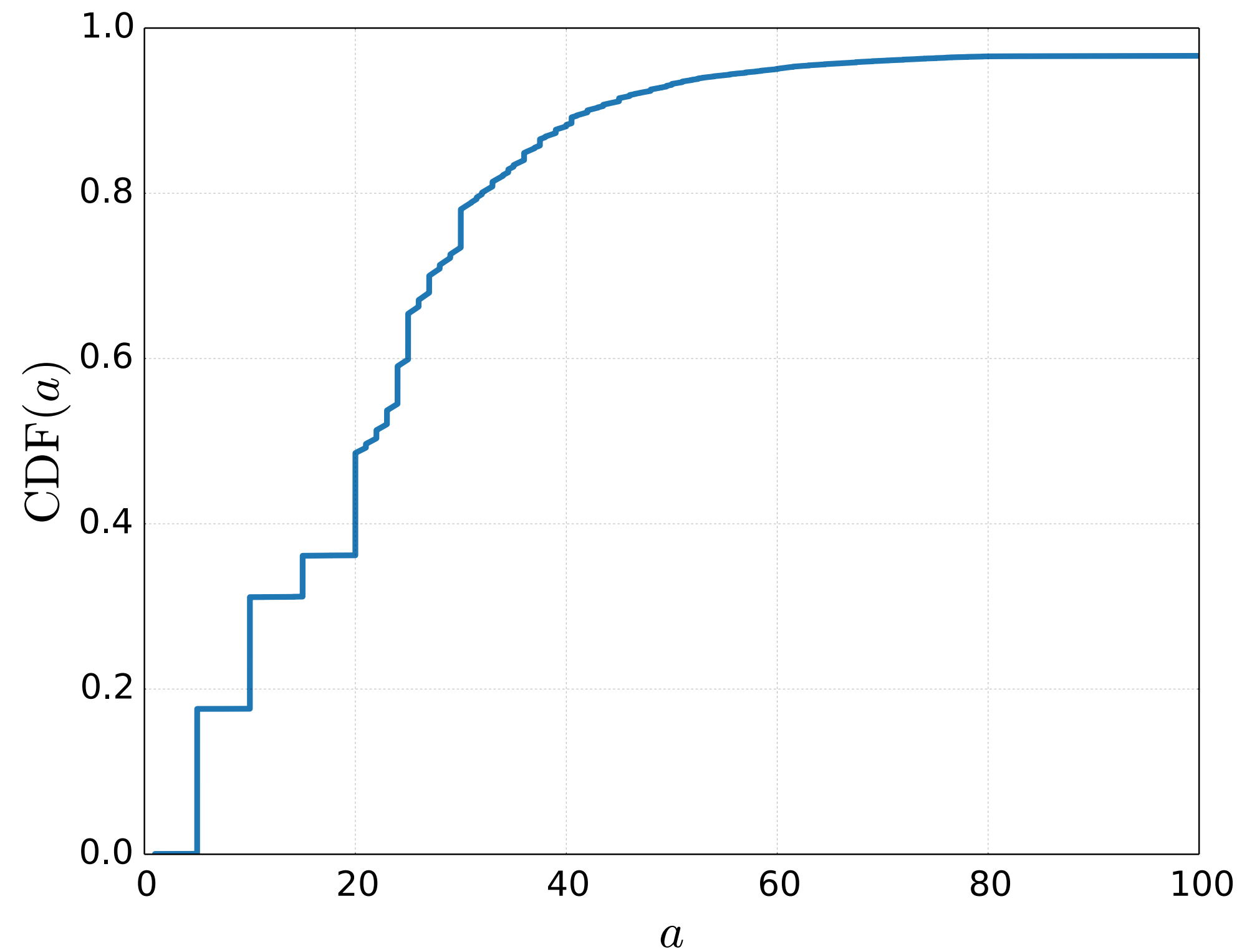
# Privacy: High-Resolution Data

## Location



# Privacy: High-Resolution Data

## Location



- high accuracy
- storing raw & real-time data for applications
- dense social system



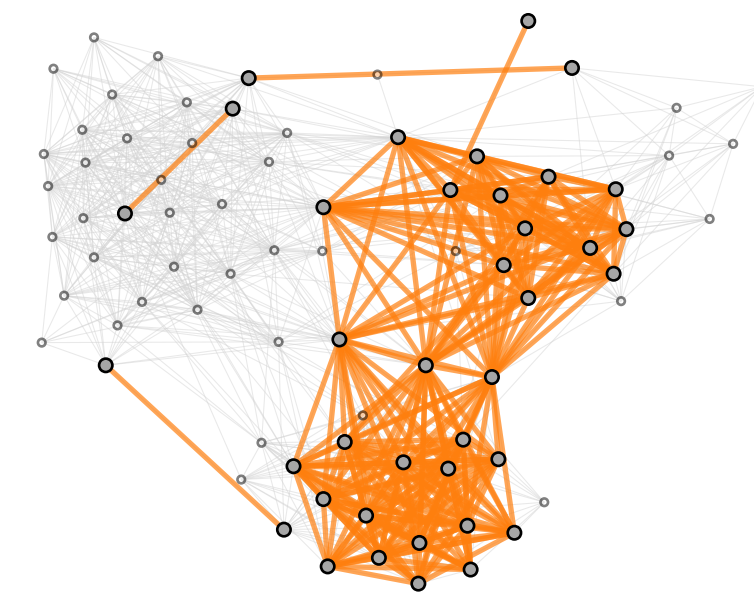
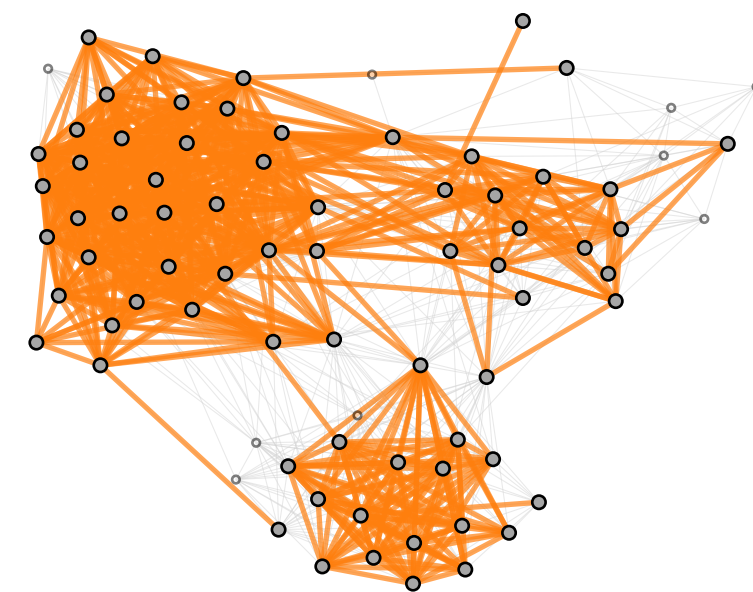
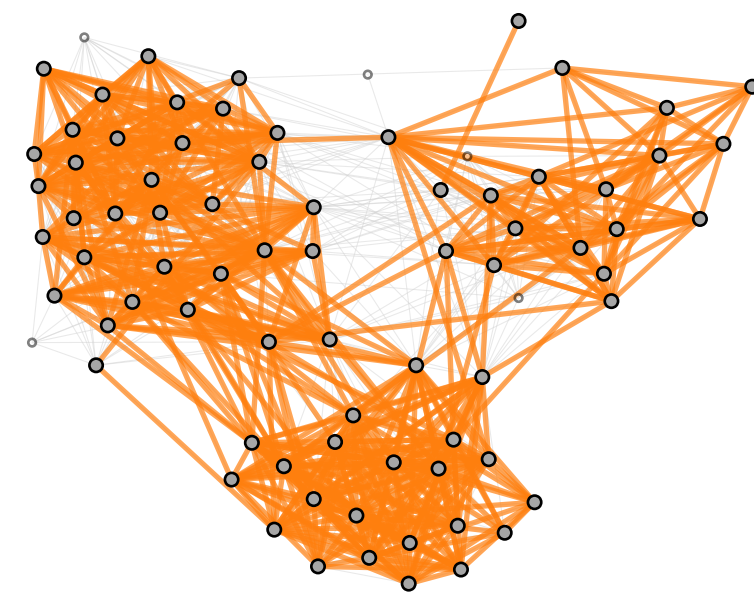
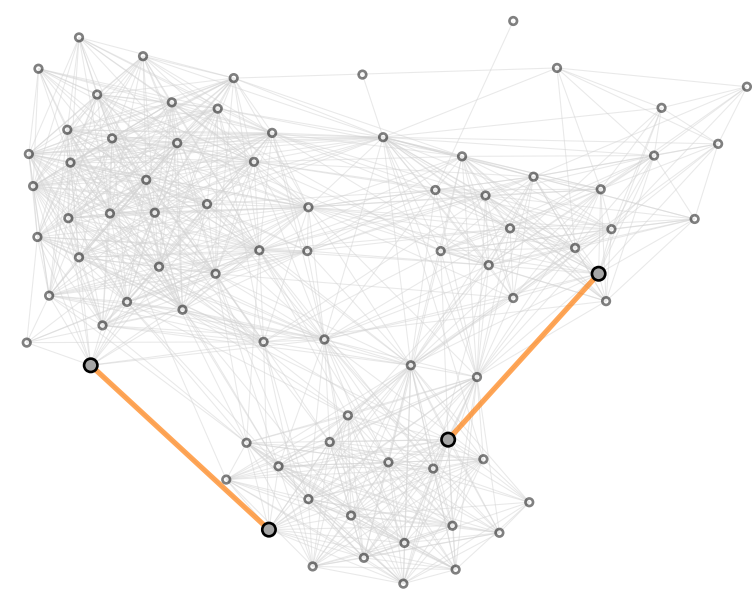


# Privacy: High-Resolution Data

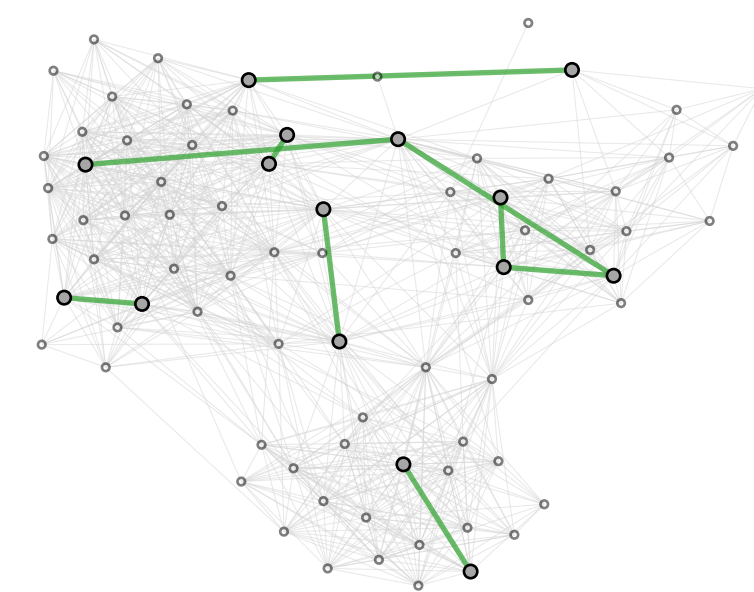
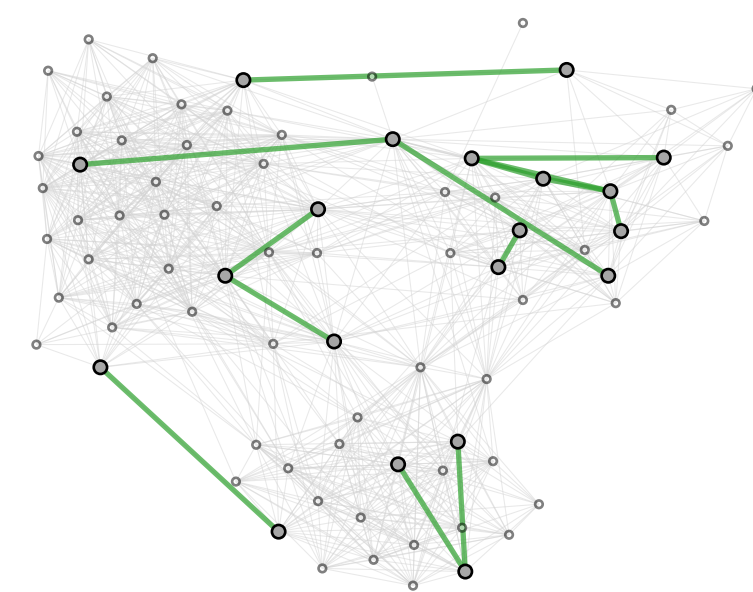
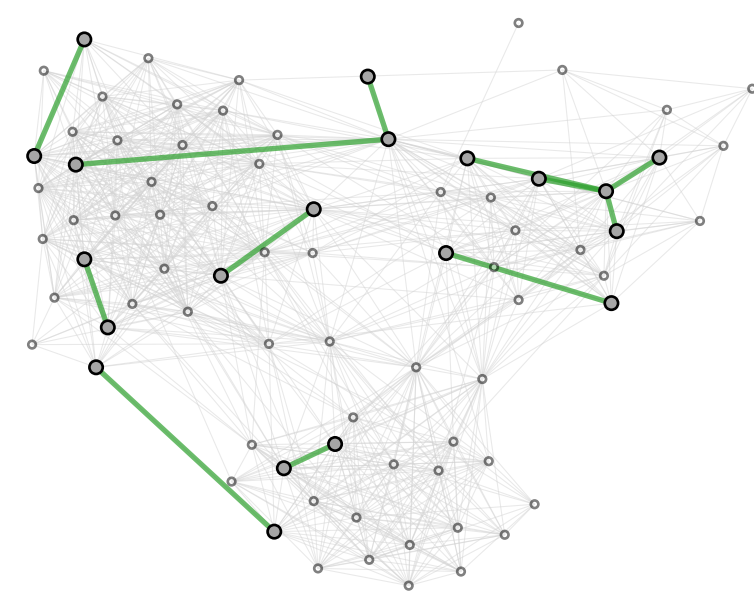
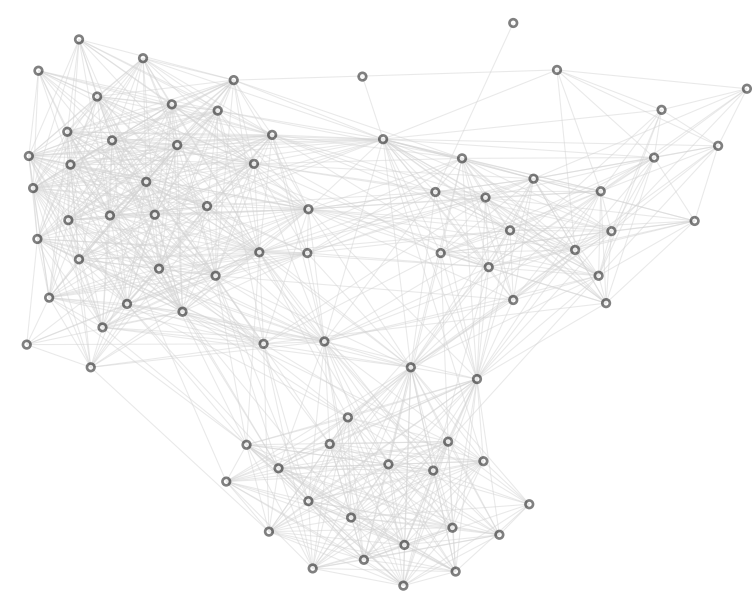
## Social

- de-identification
- deletion
- feedback impact

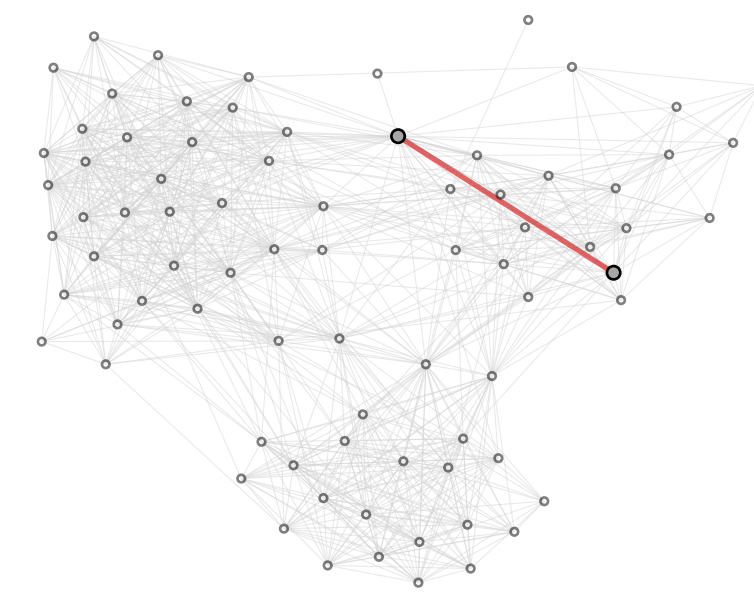
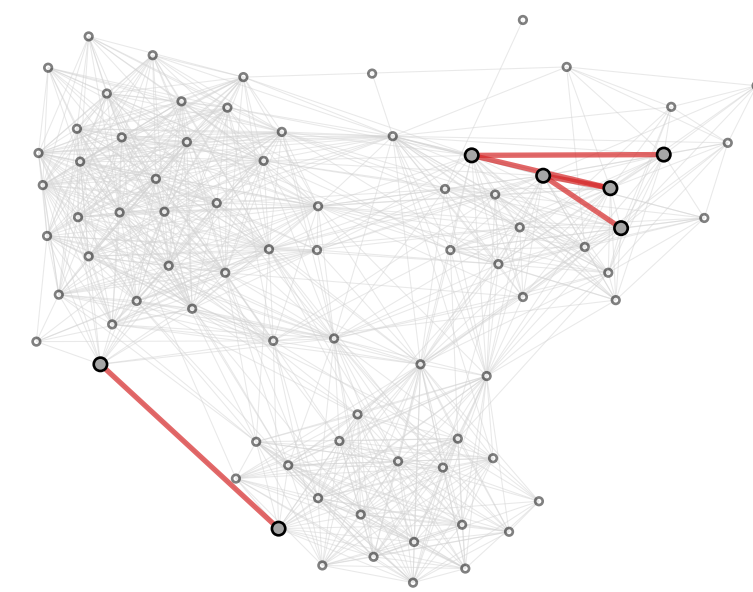
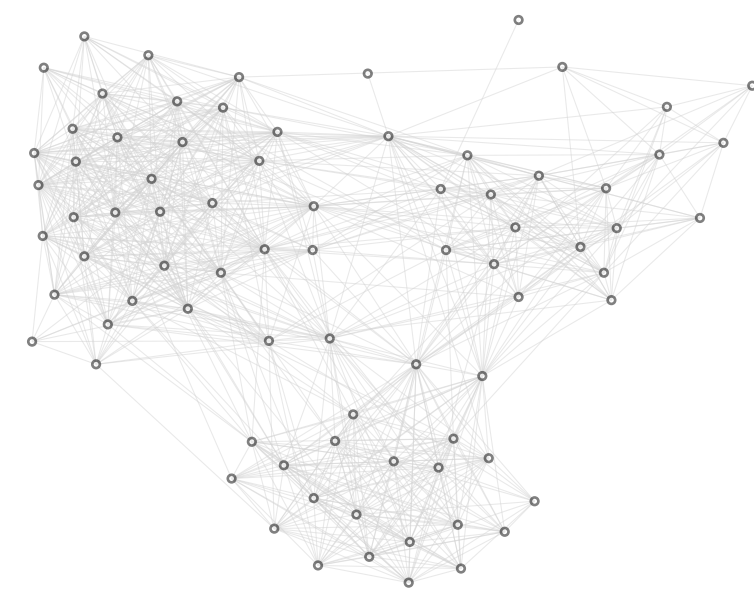
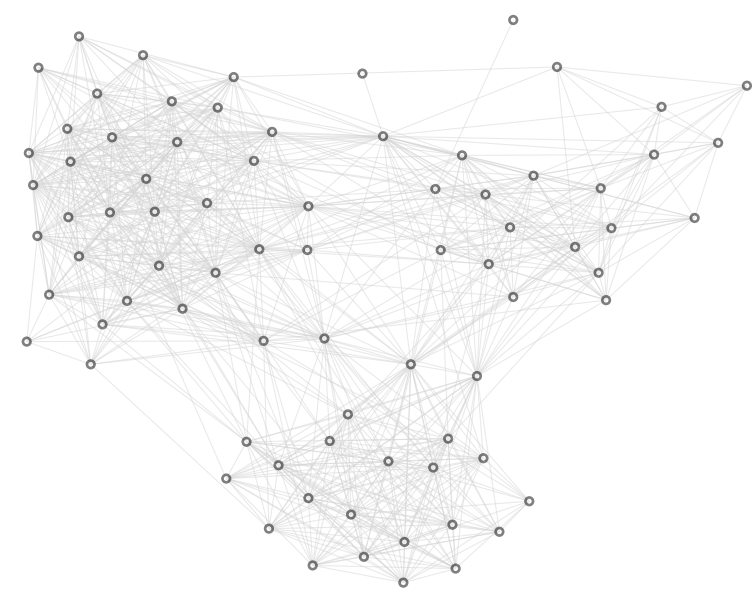
### Face-to-face



### Texts



### Calls



12 pm - 6 am

6 - 12 am

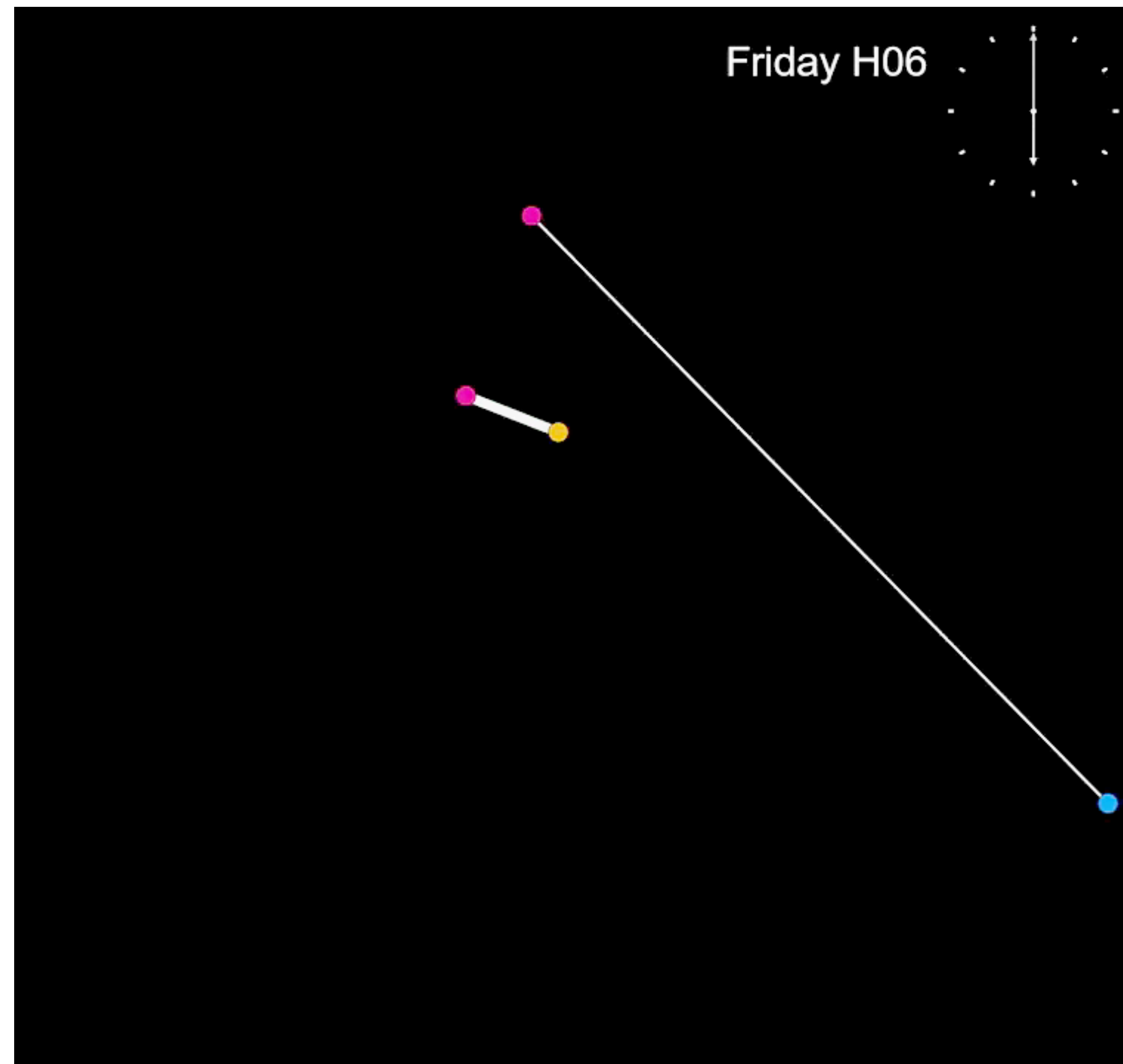
12 am - 6 pm

6 - 12 pm

# Privacy: High-Resolution Data

## **Social**

- de-identification
- deletion
- feedback impact



# Privacy: High-Resolution Data

## Network Privacy

- neighborhood attacks
- NodeInfo, LinkInfo
- $\alpha$ -proximity

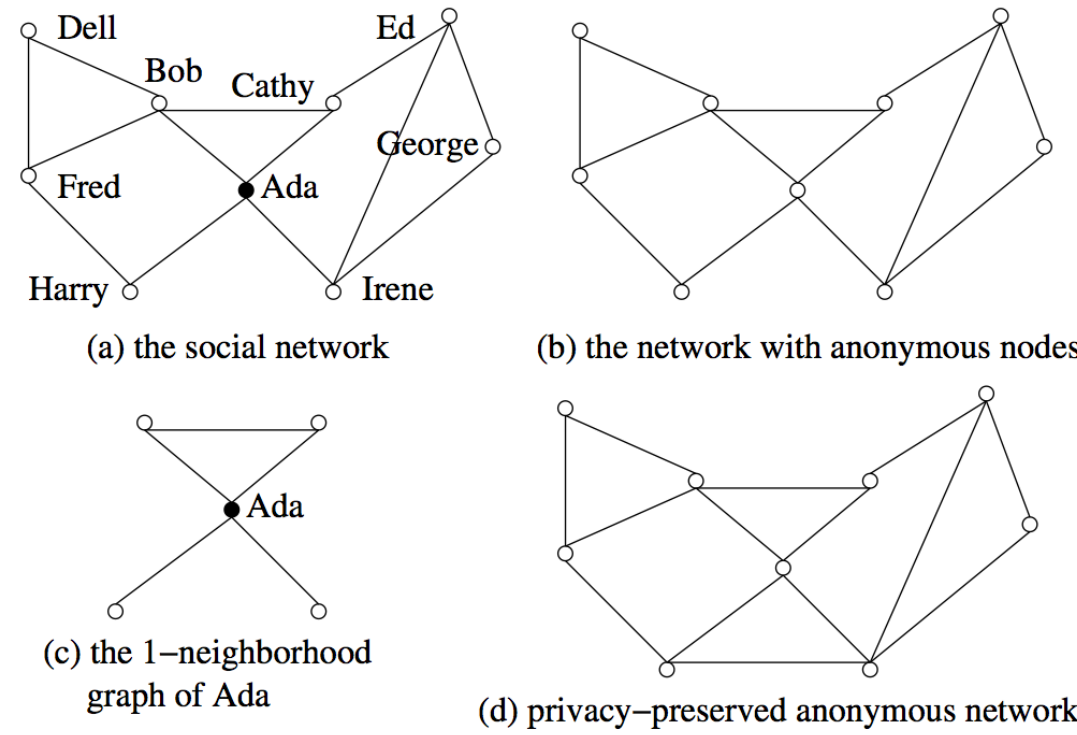


Fig. 1. Neighborhood attacks in a social network.

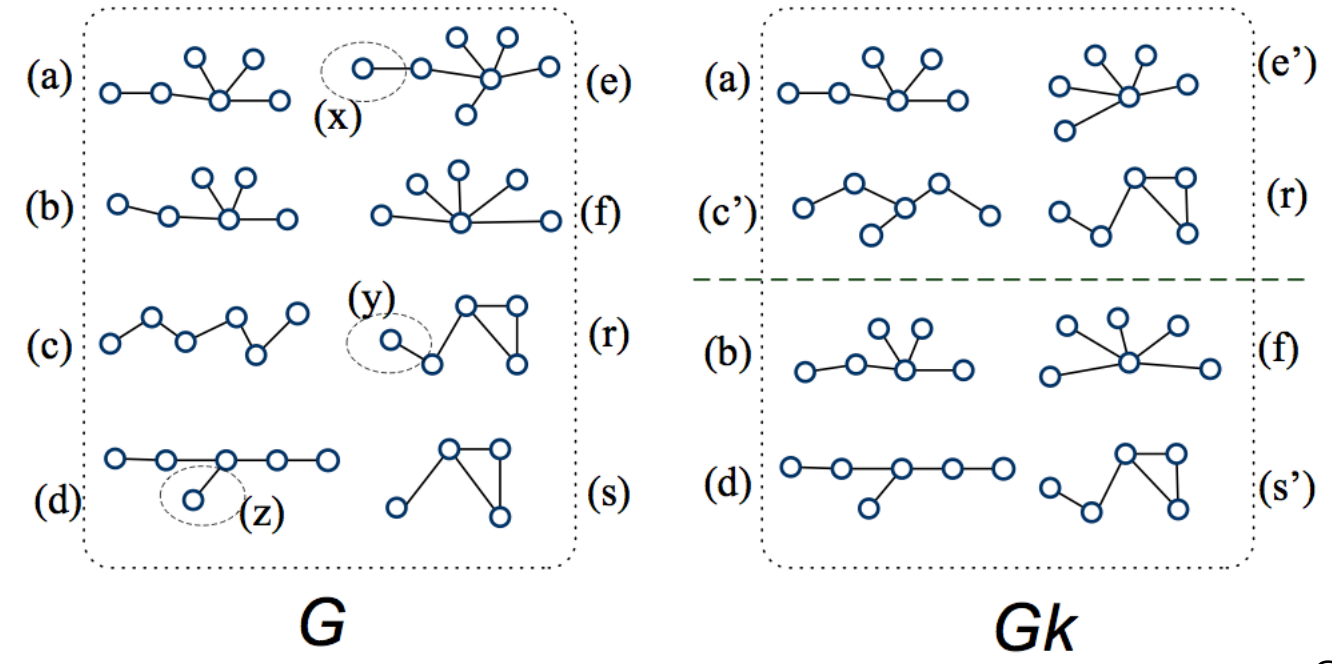


Figure 7: Graph  $G$  and Anonymized Graph  $G_k$

from 1.

from 2.

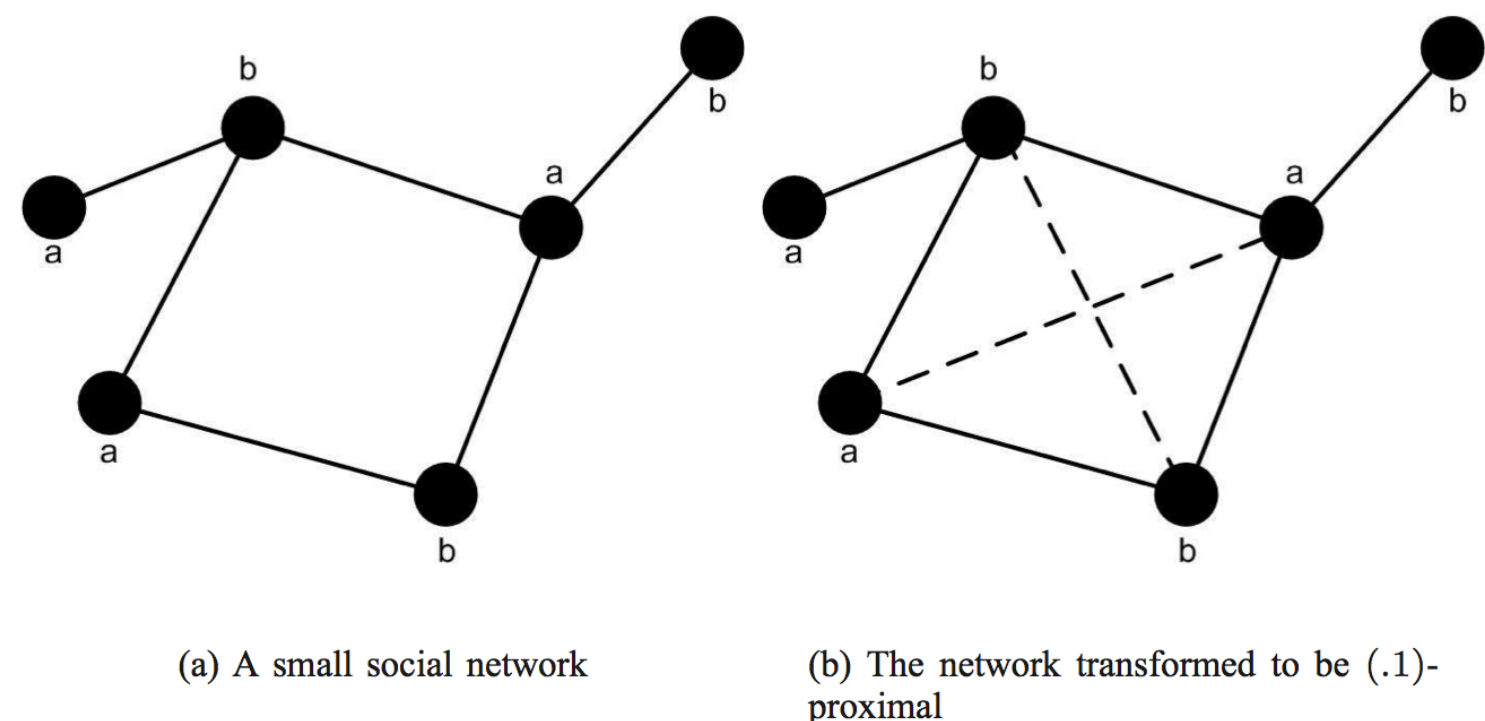


Fig. 1. Example of  $\alpha$ -proximity

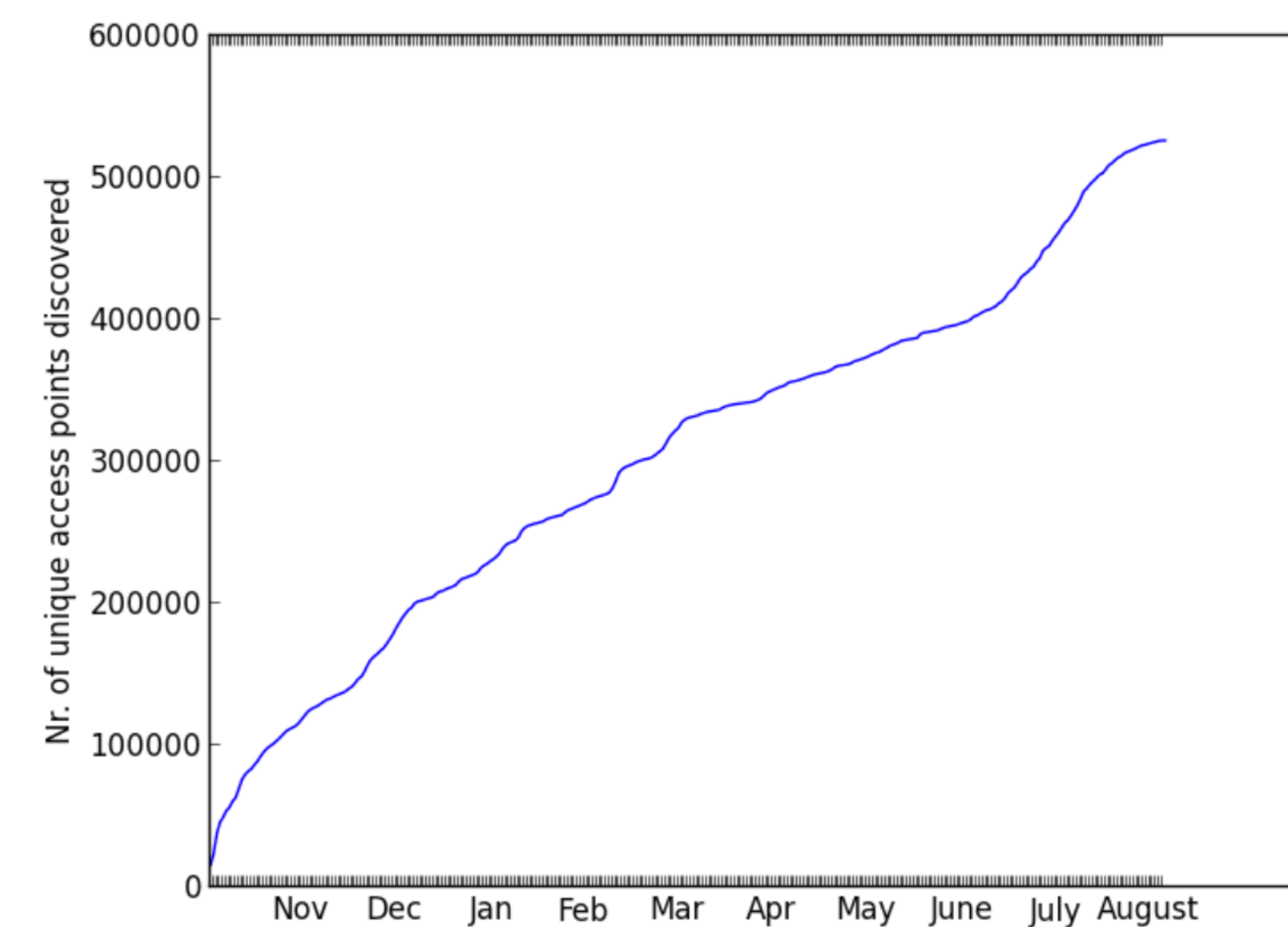
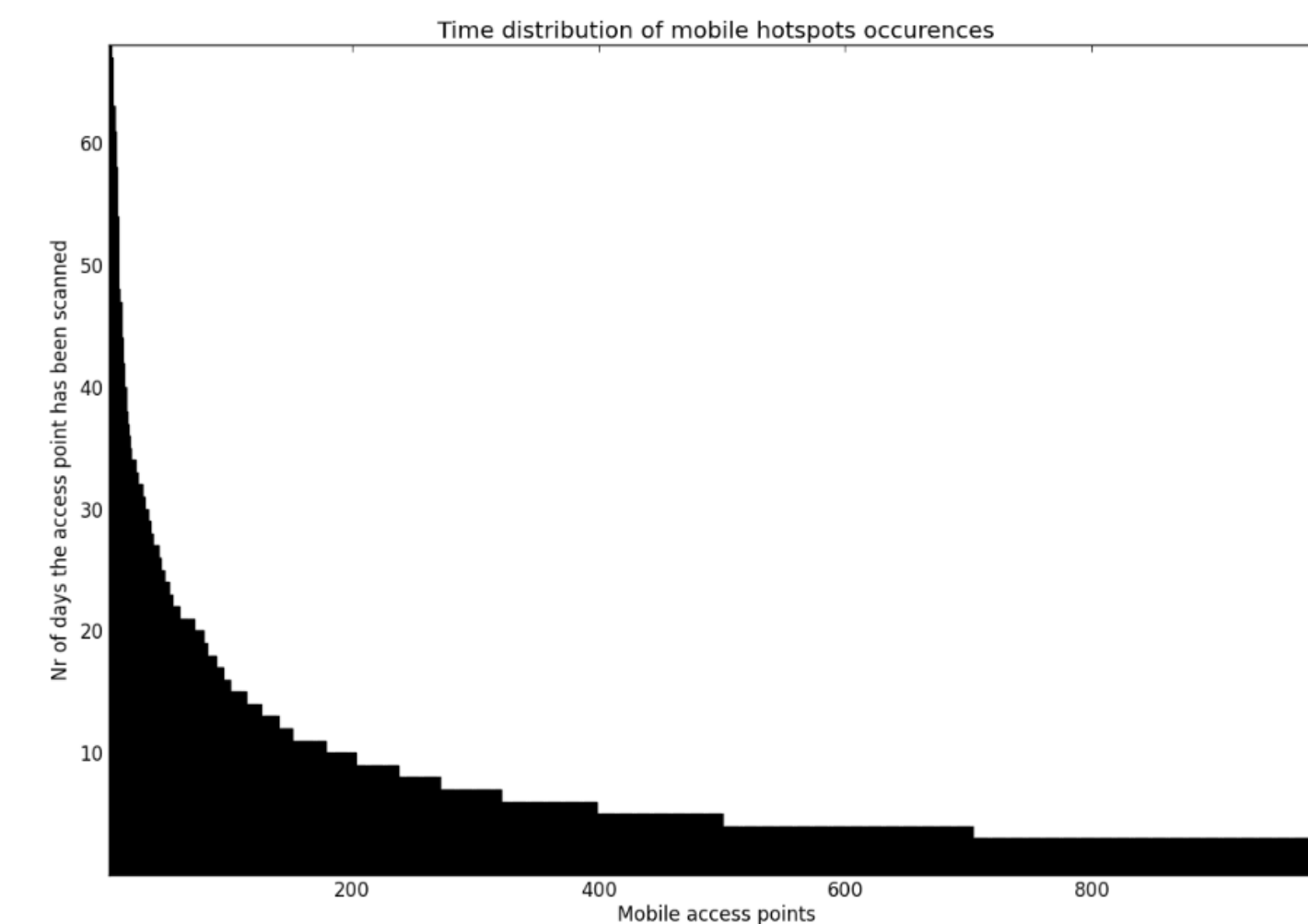
from 3.

- 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.
- 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.
- 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

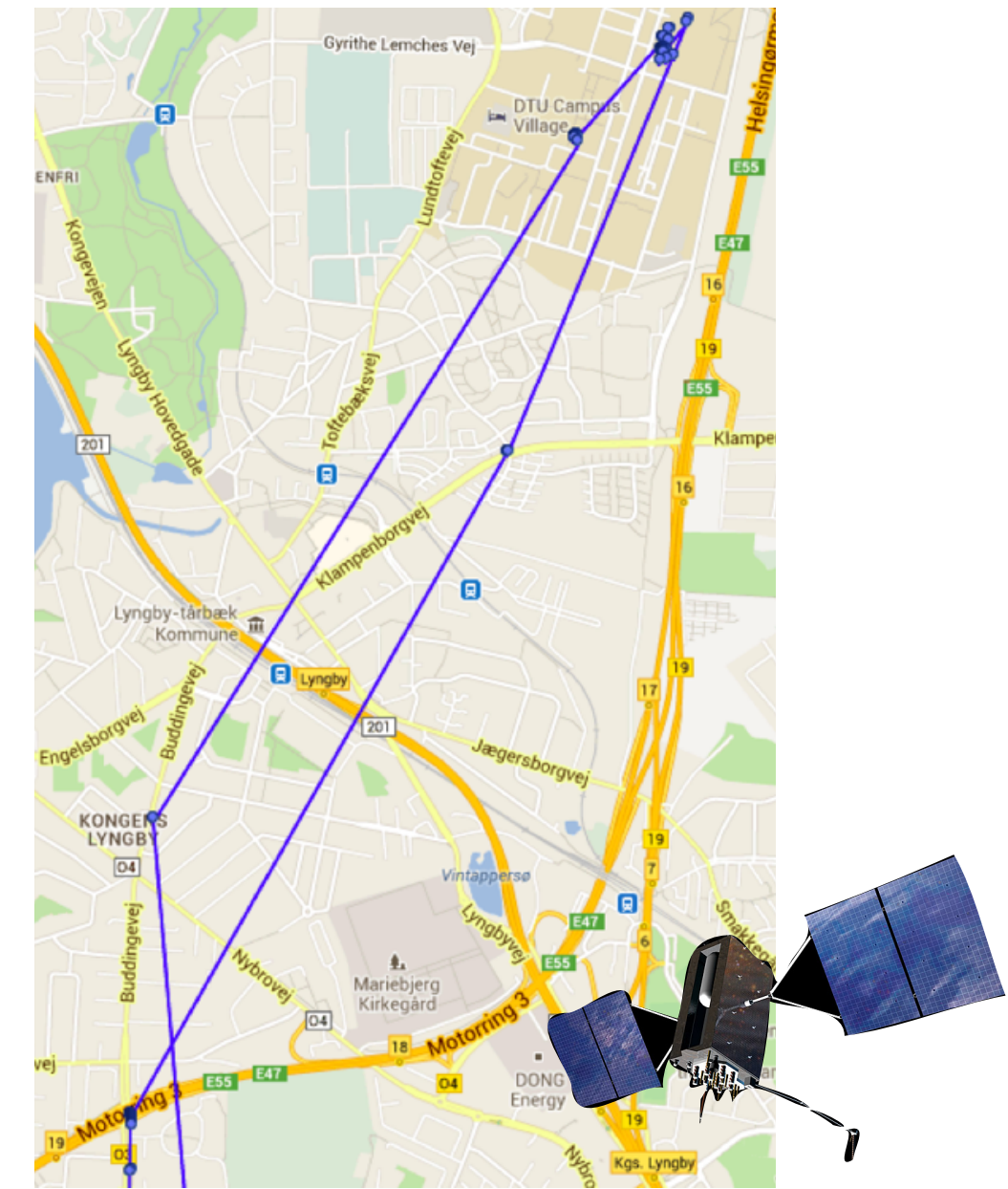
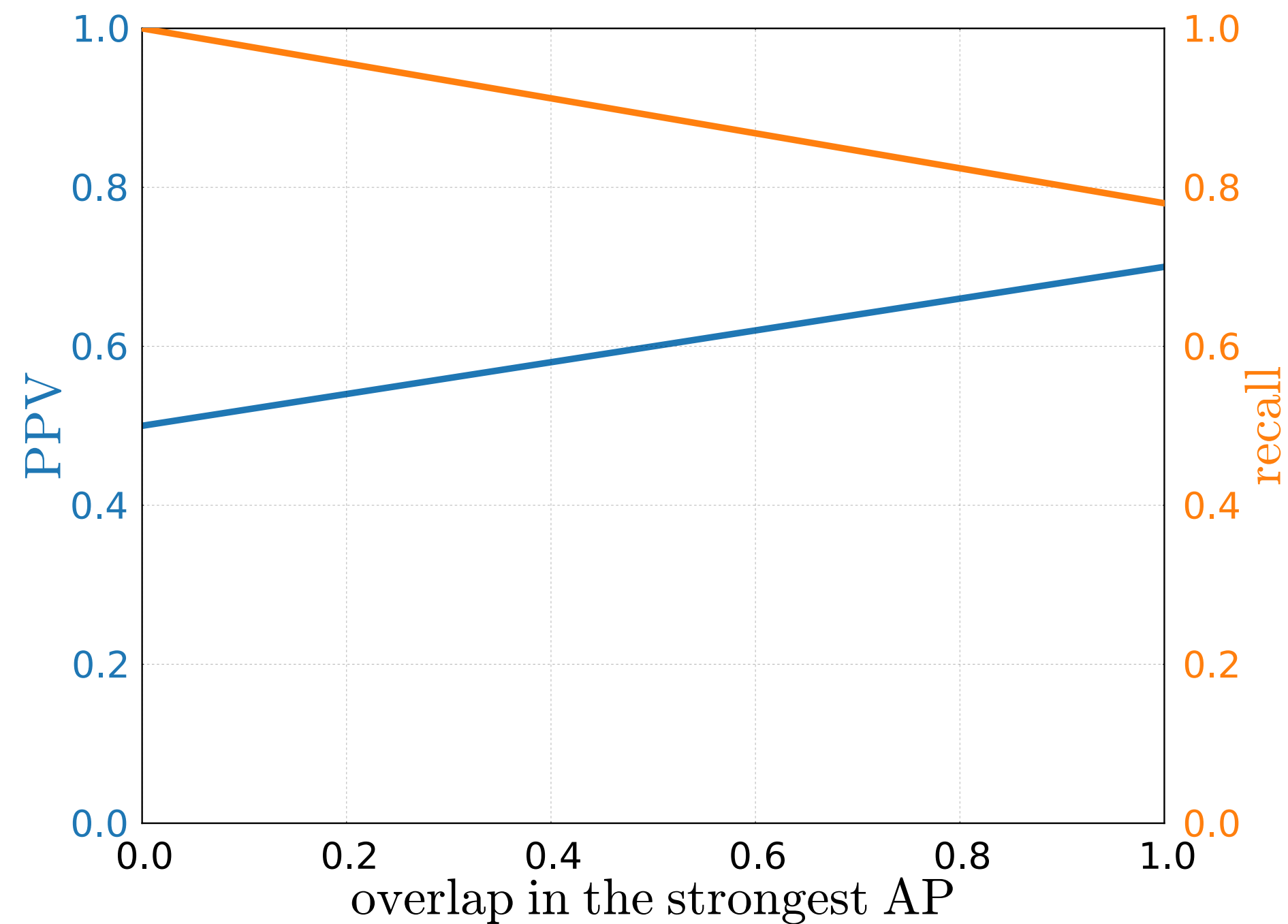
- highly-linked data
- reconstructing individual traces
- distributed sensing
- quantification?



# Privacy: High-Resolution Data

## Channel Redundancy

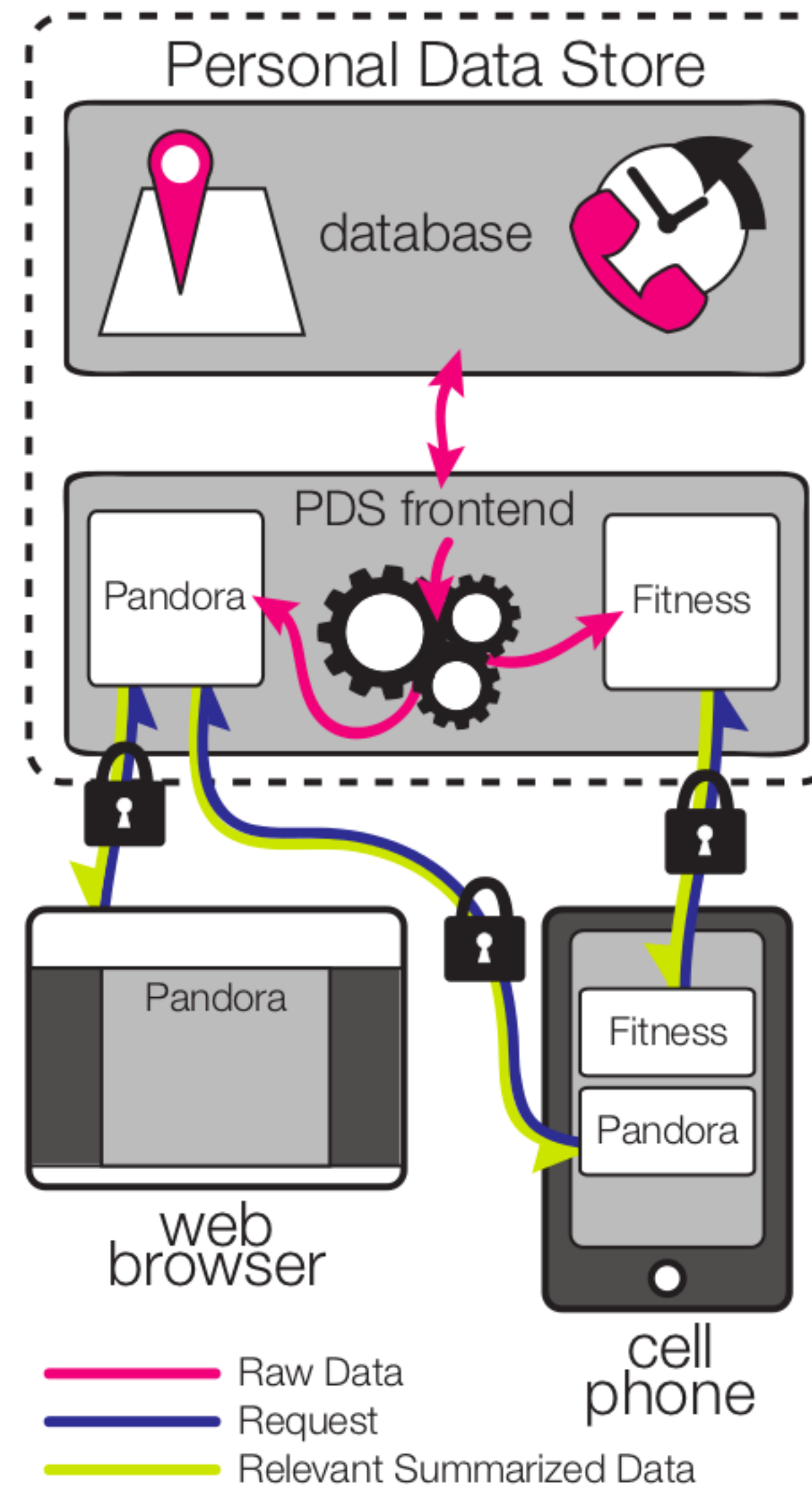
- face-to-face from Wifi...
- ...or Facebook photos...
- ...or Facebook checkins
- location from Facebook...
- ...or Wifi



# Privacy: High-Resolution Data

## PDS

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



### Collaboration with DTU

#### Confidentiality Agreement

For the purpose of participating in the Research Project: SensibleDTU/High Resolution Networks/Social Fabric, supervised by Associate Professor, Sune Lehmann, DTU Compute (hereinafter the "Project Supervisor"):

I, the undersigned,

- Title (mr/mrs/ms/dr/prof): \_\_\_\_\_
- Name: \_\_\_\_\_
- Student Number: \_\_\_\_\_
- Address: \_\_\_\_\_
- Phone: \_\_\_\_\_
- Personal email: \_\_\_\_\_

hereby declare my understanding and full acceptance of the conditions listed below subject to which I participate in the Research Project:

1. all information regarding the Department's know-how and research results, including but not limited to the Research Project and other research projects, written as oral, exchanged between any of the Research Project participants, **shall be treated as strictly confidential**
2. no information regarding the Research Project may be published or disclosed to any third party without the express written consent of the Project Supervisor, with the exception of information which:
  - a. was in the public domain or otherwise made available to the general public at the time of receipt;
  - b. was in my possession without any restrictions as to confidentiality before the Research Project was initiated;
  - c. was received from a third party with no relation to the Research Project.
3. publication of own Research Project results of any kind in dissertations articles or oral presentations or in any other form is subject to the written approval of the Project Supervisor. The Project Supervisor shall ensure that publication is not in any way detrimental to the possibilities of commercial exploitation of invention or software made in connection with the Research Project, as well as privacy of study participants. The Project Supervisor may as a condition for approval request that a dissertation which is to be used as an exam project be provided wholly or in part with a confidentiality clause in respect of parties other than the examiners.
4. no efforts will be made to determine the identity of individuals in a de-identified dataset.

# Acknowledgments

The SensibleDTU project was made possible by a Young Investigator Grant from the Villum Foundation.

Scaling the project up to 1 000 individuals in 2013 was made possible by a interdisciplinary UCPH 2016 grant, Social Fabric (PI David Dreyer Lassen, SL is co-PI) focusing mainly on the social and basic science elements of the project.

This grant has funded purchase of the smartphones, as well as technical personnel.

We are indebted to our University of Copenhagen partners on a number of levels:

All instrumentation on the 2013 questionnaires, as well as embedded anthropologist, are courtesy of the Social Fabric group, and most importantly the Social Fabric consortium has made valuable contributions through discussion and insight regarding nearly every aspect of the study.

For an overview of the Social Fabric project, see <http://socialfabric.ku.dk/>.

We thank group leaders in the Social Fabric Project: Professor David Dreyer Lassen, Professor Morten Axel Pedersen, Associate Professor Anders Blok, Assistant Professor Jesper Dammeyer, Associate Professor Joachim Mathiesen, Assistant Professor Julie Zahle, and Associate Professor Rikke Lund.

From Institute of Economics: David Dreyer Lassen, Andreas Bjerre Nielsen, Anne Folke Larsen, and Nikolaj Harmon.

From Institute of Psychology: Jesper Dammeyer, Lars Lundmann, Lasse Meinert Jensen, and Patrick Bender.

From Institute of Anthropology: Morten Axel Pedersen, My Madsen.

From Institute of Sociology: Anders Blok, Tobias Bornakke.

From Institute of Philosophy: Julie Zahle.

From Institute of Public Health: Rikke Lund, Ingelise Andersen, Naja Hulvej Rod, Ulla Christensen, and Agnete Skovlund Dissing.

From the Niels Bohr Institute (Physics): Joachim Mathiesen, and Mogens Høgh Jensen.



# 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>