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 calls messages mobility
health psychology sociology demographics
friends interactions activity 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
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

encrypted ‘obvious PIIs’
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

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

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

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

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 inventions 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 1000 individuals in 2013 was made possible by an 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, Technical University of Denmark & MIT Media Lab, arks@dtu.dk, @h0pbeat, https://sensible.dtu.dk