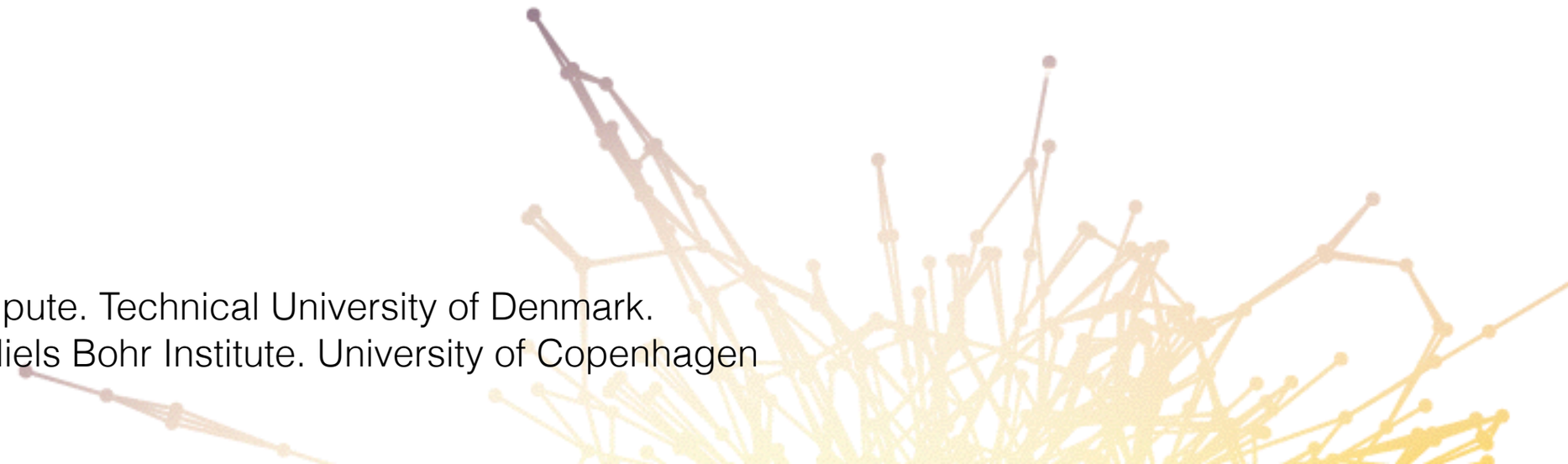
Complex Networks

connections, measurements, and social systems

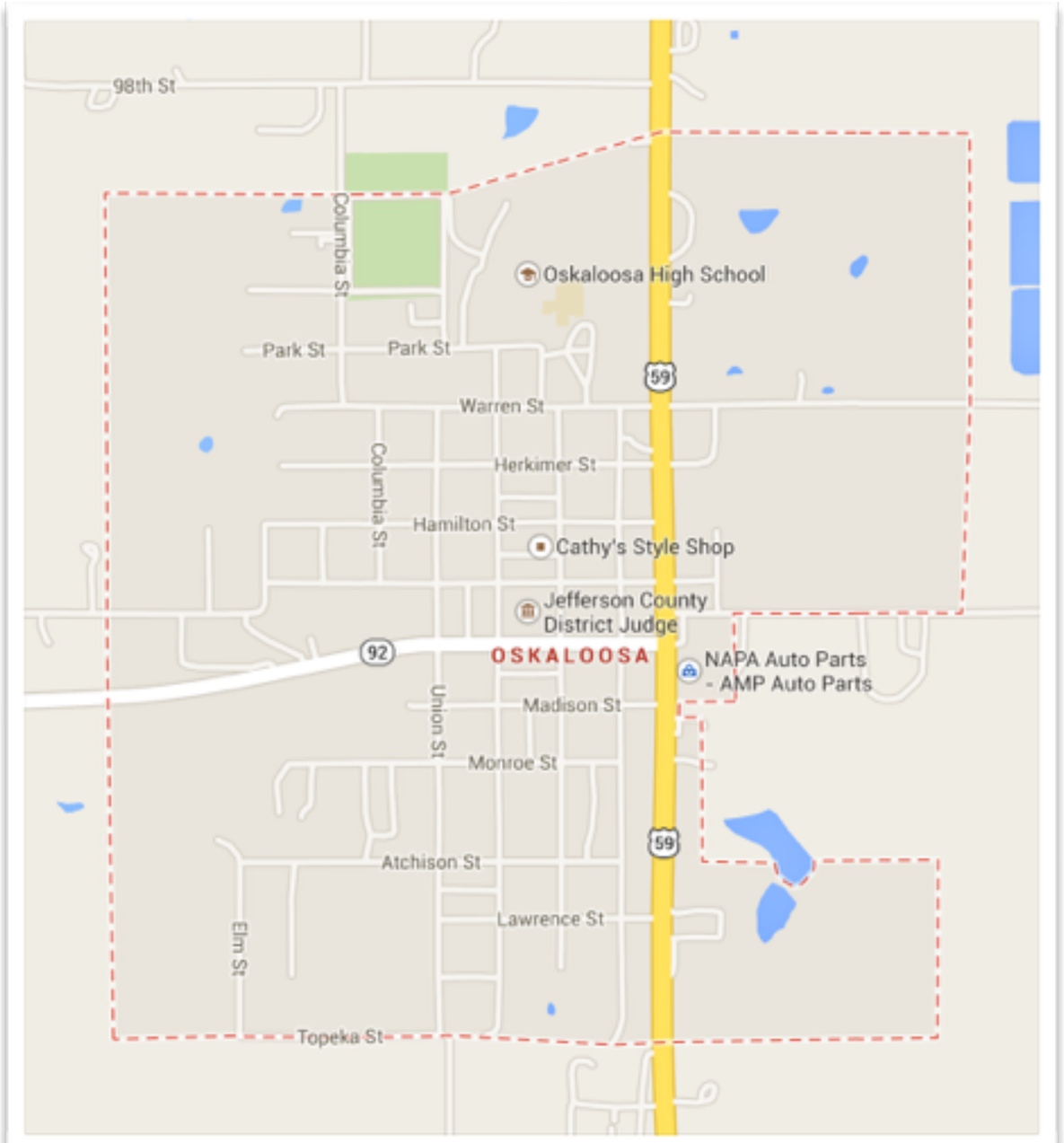
Sune Lehmann
- Associate Professor, DTU Compute. Technical University of Denmark.
- Adjunct Associate Professor, Niels Bohr Institute. University of Copenhagen
- @suneman




Pizza Pizza Pizza

I am slightly obsessed with data quality

Roger G. Barker



Monitored “Midwest” for 25 years, using grad students, research assistants, wives - equipped with pen and paper.

Deb Roy

Human Speechome Project

PHOTO: (C) PHILIP DECAMP AND DEB ROY 2012 FROM THE HUMAN FACE OF BIG DATA

Human Speechome Project

- 11 fisheye lens cameras + motion sensors.
- 14 omnidirectional microphones
- 1000m wires connect recorders to servers in basement
- Record from 8am -10pm every day for 3 years

PHOTO: (C) PHILIP DECAMP AND DEB ROY 2012 FROM THE HUMAN FACE OF BIG DATA

Human Speechome Project

- 11 fisheye lens cameras + motion sensors.
- 14 omnidirectional microphones
- 1000m wires connect recorders to servers in basement
- Record from 8am -10pm every day for 3 years

Human Speechome in numbers

- 90,000 hours of video recorded
- 140,000 hours of audio recordings
- Approx 200GB of data collected every day
- 150 TB of raw data collected over course of project
- 70% of infants waking hours captured
- 10 to 12m words spoken
- 4m words so far transcribed

number of bits per user

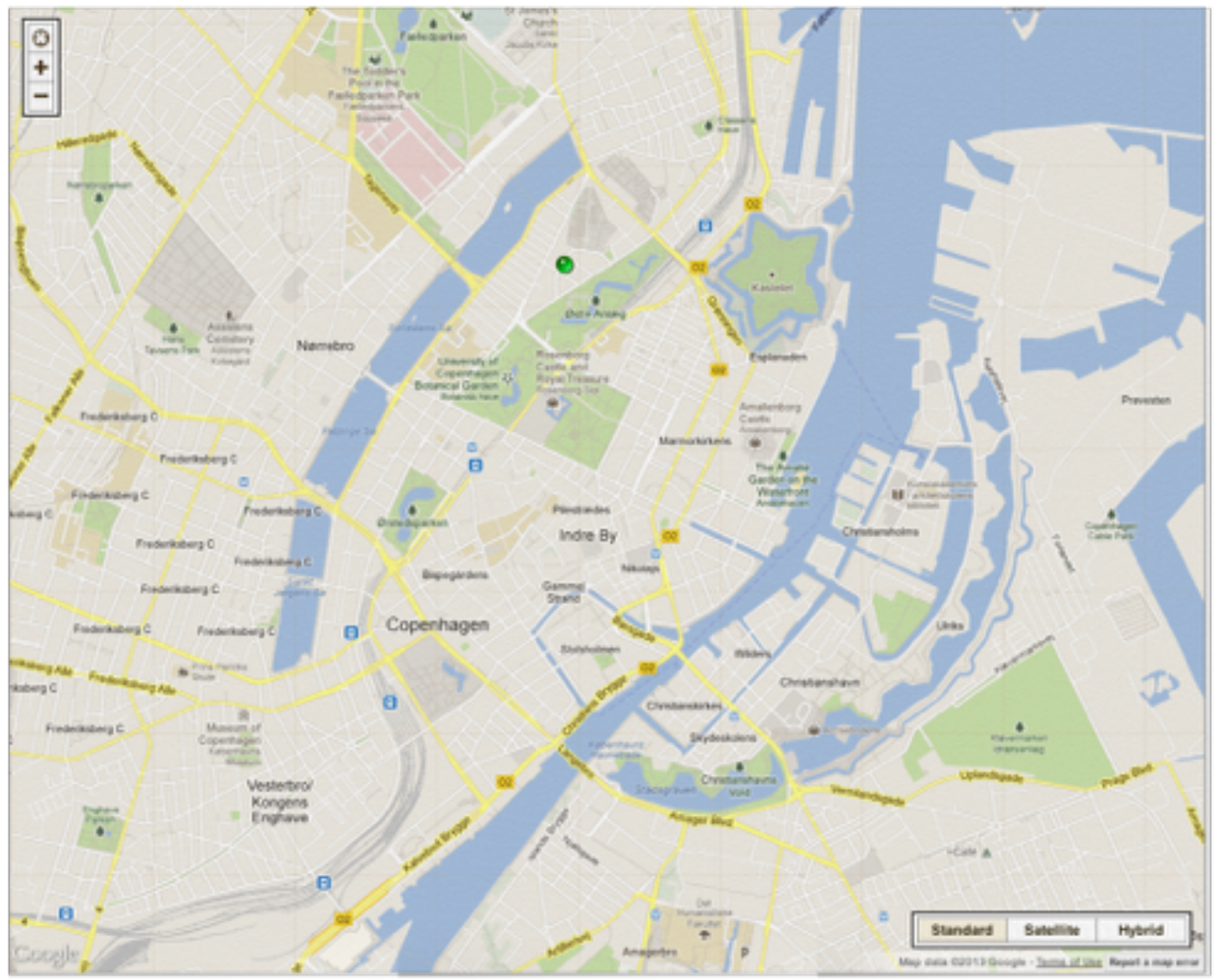
number of users

number of bits per user

number of users

Linked in


4:20
Google
Maps
Play Store
YouTube
Google


Nørrebro
Indre By
Copenhagen
Vesterbro/
Kongens
Enghave
Standard
Satellite
Hybrid

- proximity via Bluetooth
- telecommunication
- social networks
- geolocation

- demographics & personality
- live questionnaires

Credit goes to these guys (and others)

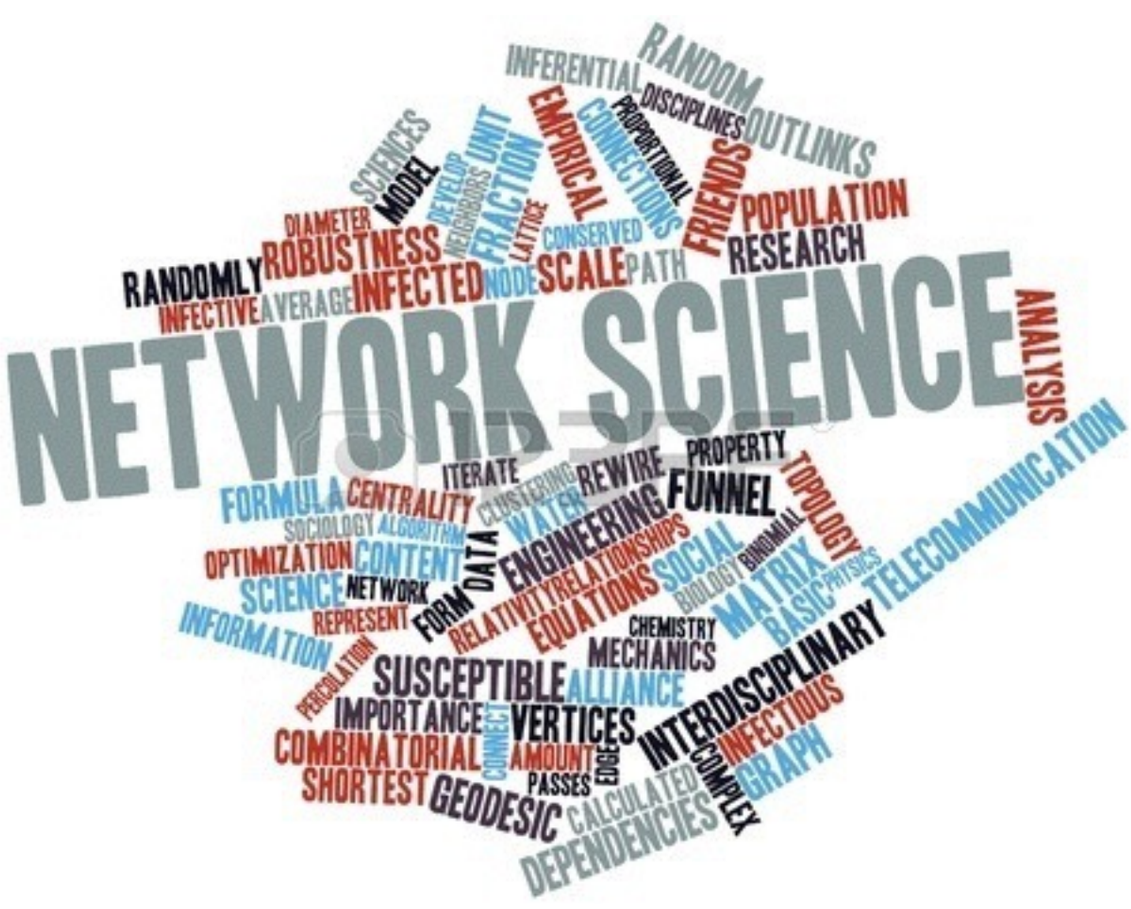
We started the project in January 2012,
so what do we have to show for it?

- Ran a 200 person beta-test in September 2012.
- Rolled out the full 1000 person experiment in September 2013.

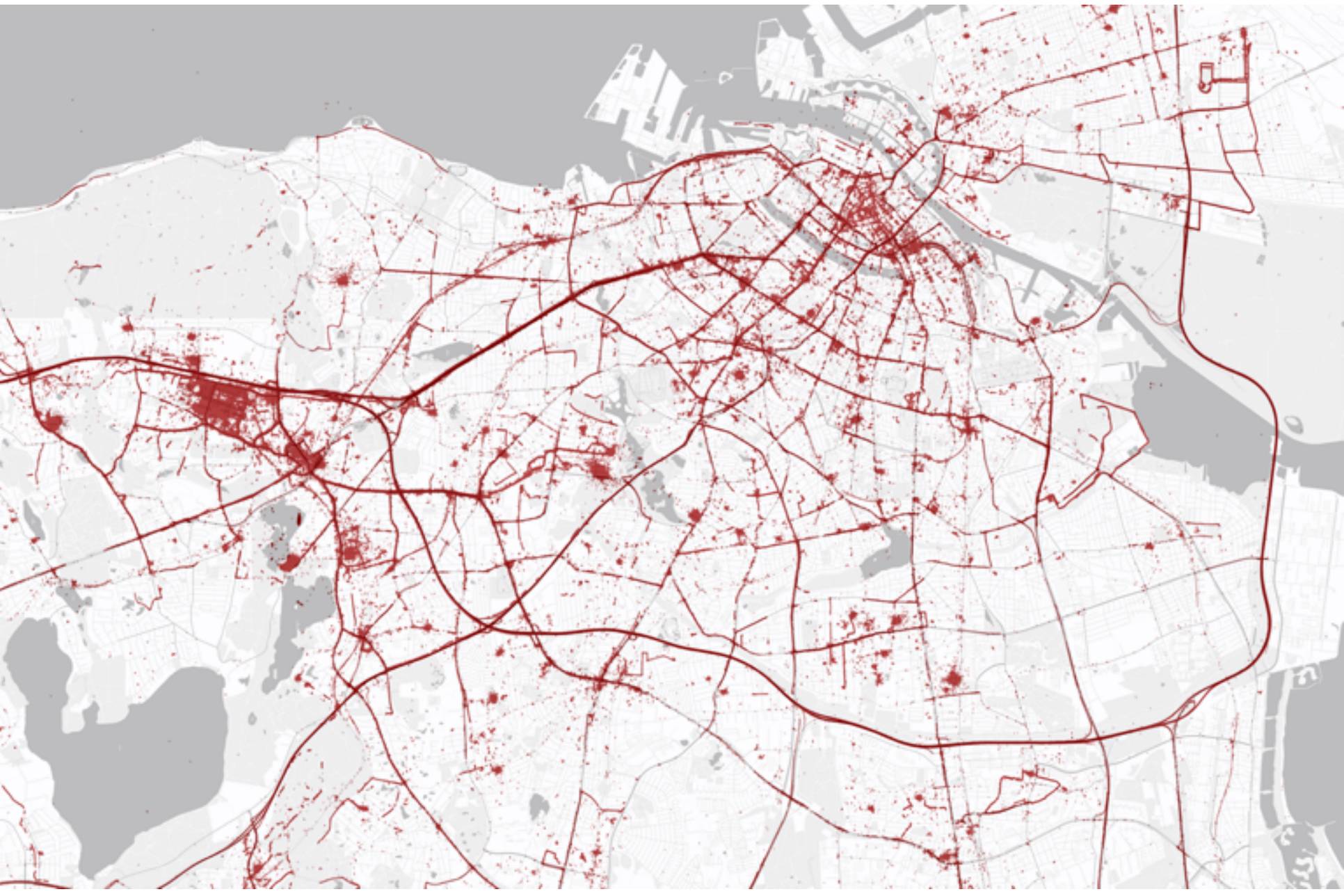
- Lots of initial work on technology: phone software, backend, visualizations
- Scientific focus so far has been
  - understanding probes,
  - privacy.
- Now we're slowly starting to do "real science".

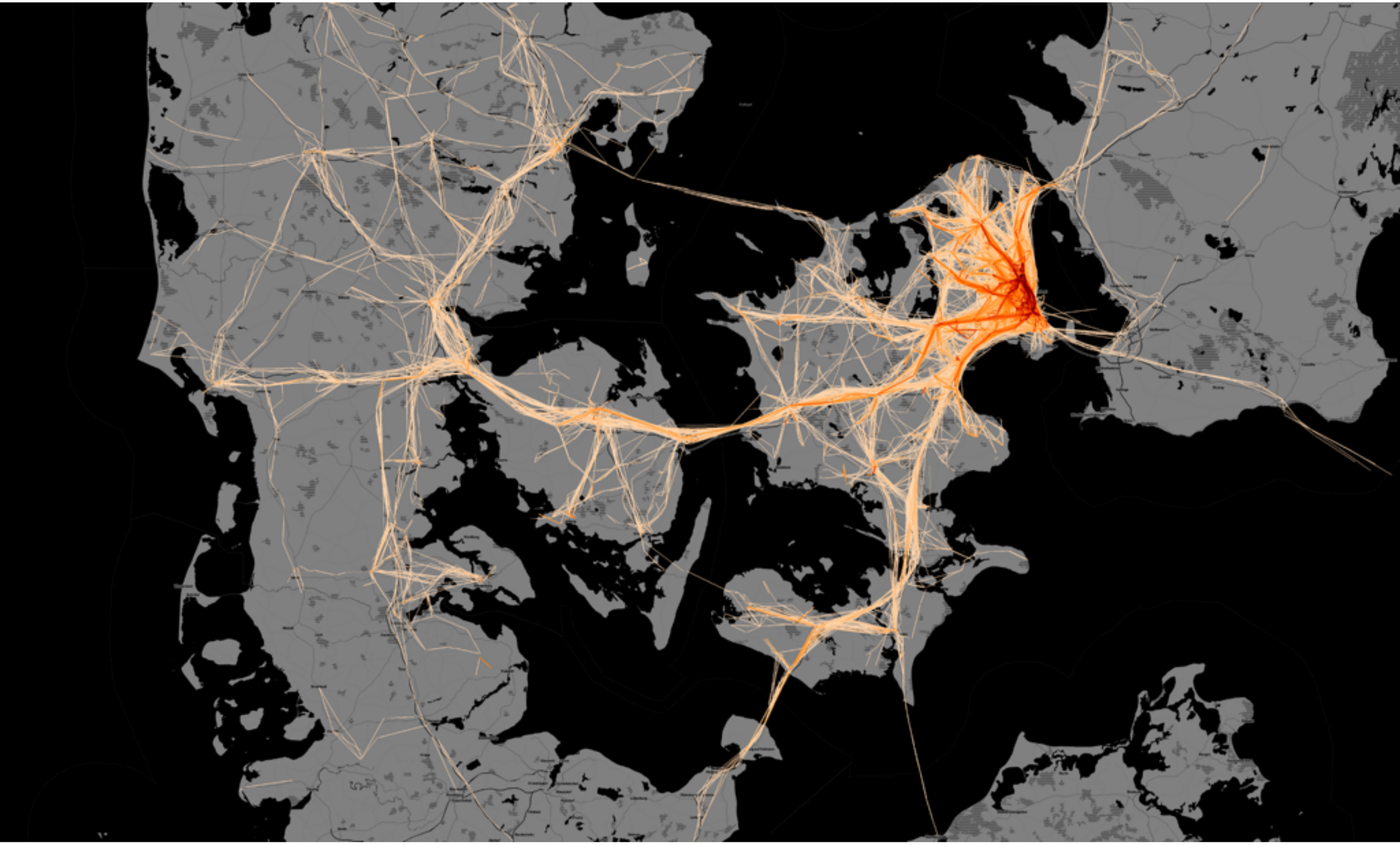
interests

- mathematical models for human interactions
- sampling/relationship between interaction channels
- spreading processes. information/epidemics
- privacy
- the technical component



the data is flowing






05:00
Lyngby Parkkirkegård
Lyngby Stadion

OPEN ACCESS Freely available online

PLOS | ONE

Measuring Large-Scale Social Networks with High Resolution

Arkadiusz Stopczynski[1]*, Vedran Sekara[1], Piotr Sapiezynski[1], Andrea Cuttone[1], Mette My Madsen[3], Jakob Eg Larsen[1], Sune Lehmann[1,2]

1 DTU Compute, Technical University of Denmark, Kgs. Lyngby, Denmark, **2** The Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark, **3** Department of Anthropology, University of Copenhagen, Copenhagen, Denmark

Abstract

This paper describes the deployment of a large-scale study designed to measure human interactions across a variety of communication channels, with high temporal resolution and spanning multiple years—the Copenhagen Networks Study. Specifically, we collect data on face-to-face interactions, telecommunication, social networks, location, and background information (personality, demographics, health, politics) for a densely connected population of 1 000 individuals, using state-of-the-art smartphones as social sensors. Here we provide an overview of the related work and describe the motivation and research agenda driving the study. Additionally, the paper details the data-types measured, and the technical infrastructure in terms of both backend and phone software, as well as an outline of the deployment procedures. We document the participant privacy procedures and their underlying principles. The paper is concluded with early results from data analysis, illustrating the importance of multi-channel high-resolution approach to data collection.

Citation: Stopczynski A, Sekara V, Sapiezynski P, Cuttone A, Madsen MM, et al. (2014) Measuring Large-Scale Social Networks with High Resolution. PLoS ONE 9(4): e95978. doi:10.1371/journal.pone.0095978

Editor: Yamir Moreno, University of Zaragoza, Spain

Received February 15, 2014; **Accepted** April 2, 2014; **Published** April 25, 2014

Funding: The SensibleDTU project was made possible by a Young Investigator Grant from the Villum Foundation (High Resolution Networks, awarded to SL). 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. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: arks@dtu.dk

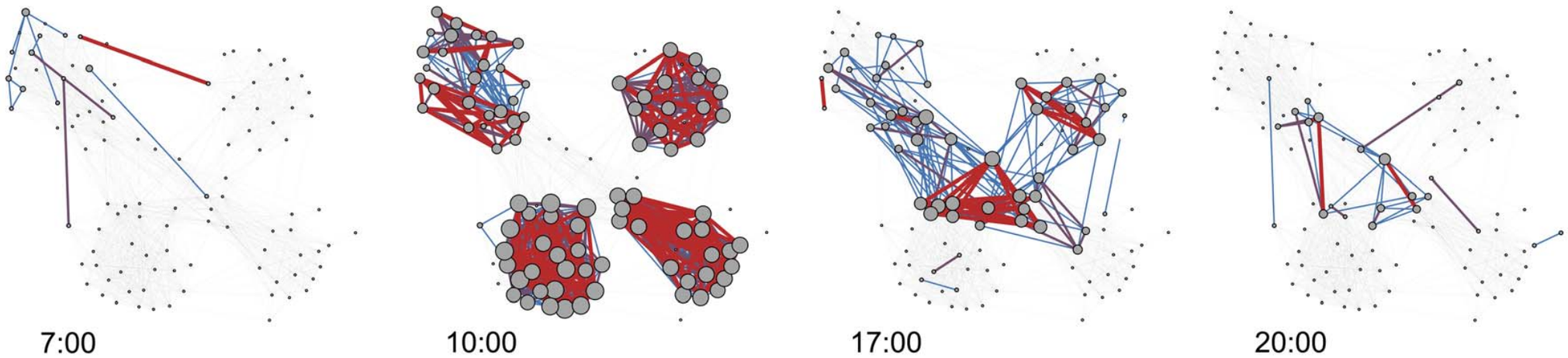


Figure 1. Dynamics of face-to-face interactions in the 2012 deployment. The participants meet in the morning, attend classes within four different study lines, and interact across majors in the evening. Edges are colored according to the frequency of observation, ranging from low (blue) to high (red). With 24 possible observations per hour, the color thresholds are respectively: blue ($0 <$ observations ≤ 6), purple ($6 <$ observations ≤ 12), and red (< 12 observations). Node size is linearly scaled according to degree.

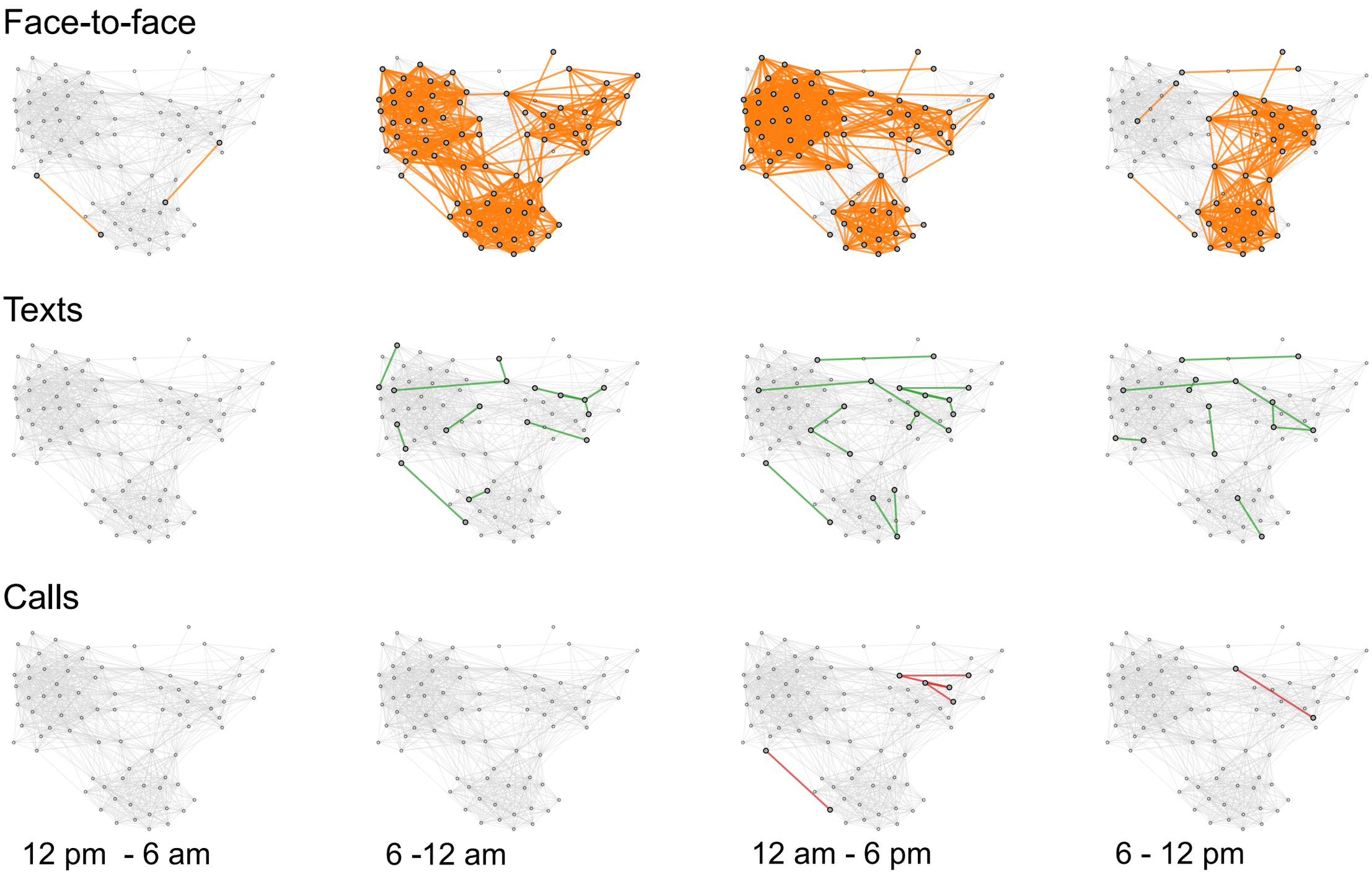


Figure 11. Daily activations in three networks. One day (Friday) in a network showing how different views are produced by observing different channels.

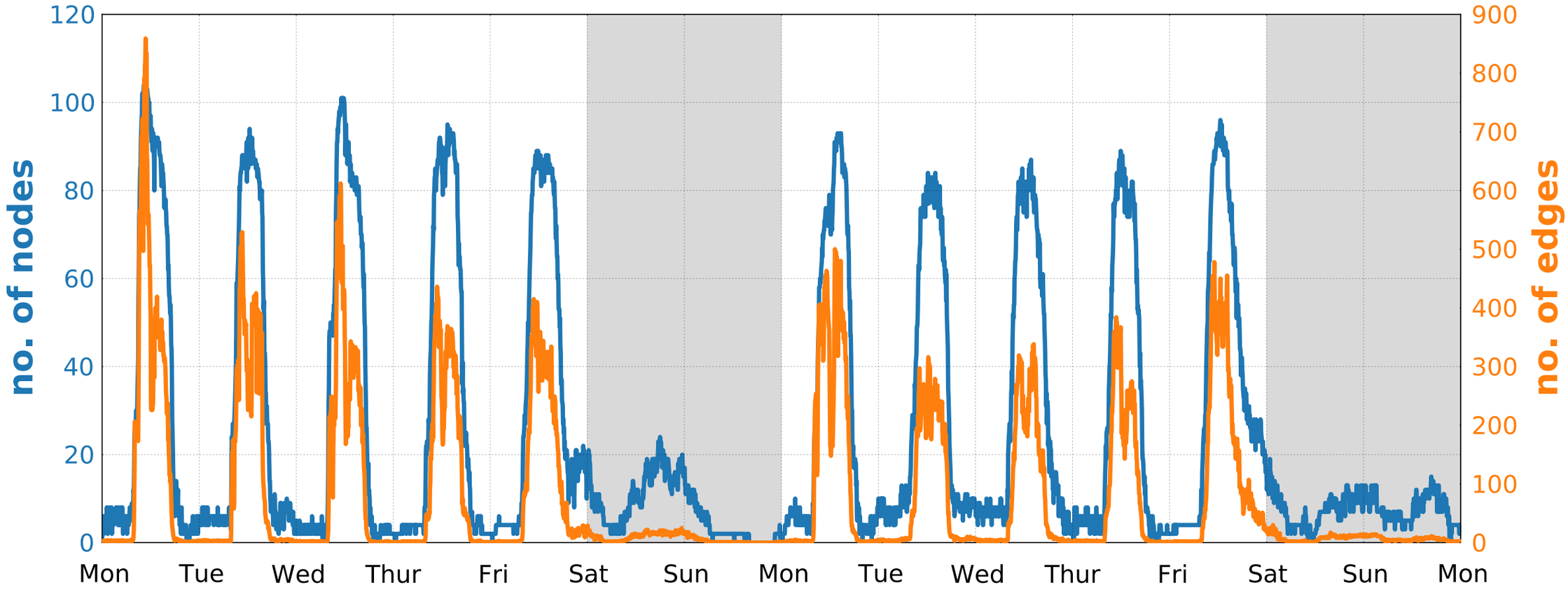


Figure 5. Weekly temporal dynamics of interactions. Face-to-face interaction patterns of participants in 5-minute time-bins over two weeks. Only active participants are included, i.e. users that have either observed another person or themselves been observed in a given time-bin. On average we observed 29 edges and 12 nodes in 5-minute time-bins and registered 10 634 unique links between participants.

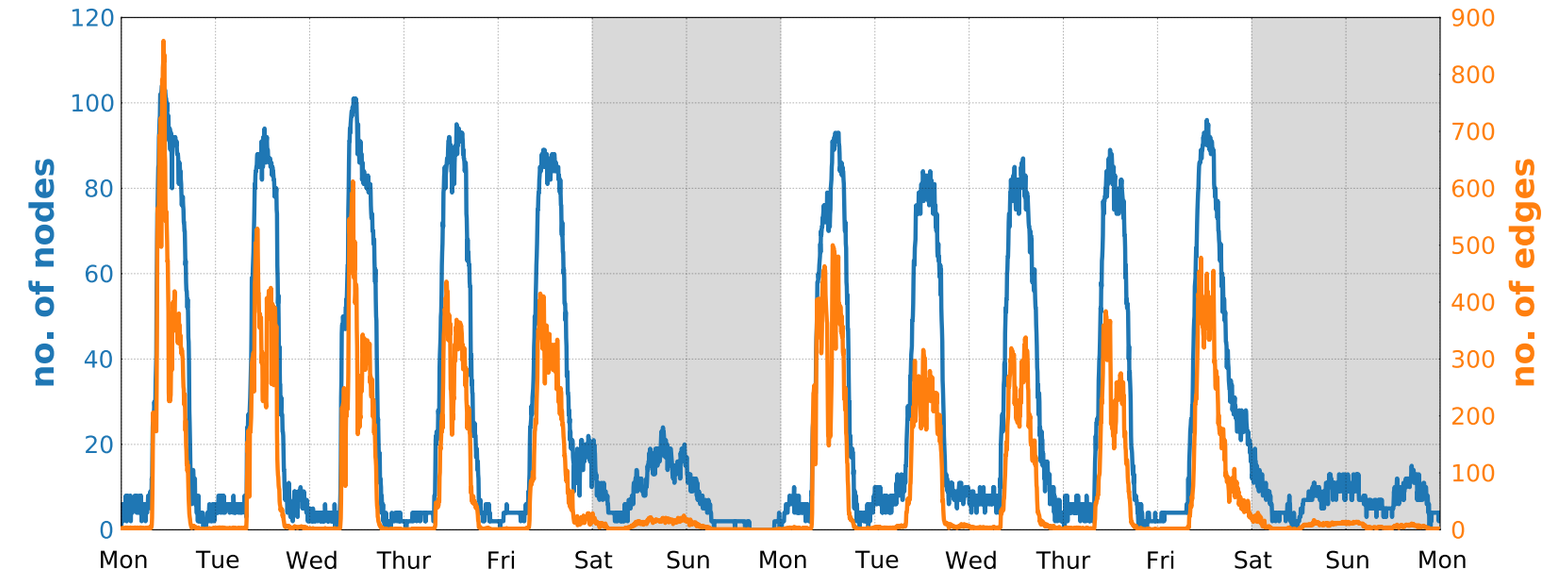


Figure 5. Weekly temporal dynamics of interactions. Face-to-face interaction patterns of participants in 5-minute time-bins over two weeks. Only active participants are included, i.e. users that have either observed another person or themselves been observed in a given time-bin. On average we observed 29 edges and 12 nodes in 5-minute time-bins and registered 10 634 unique links between participants.

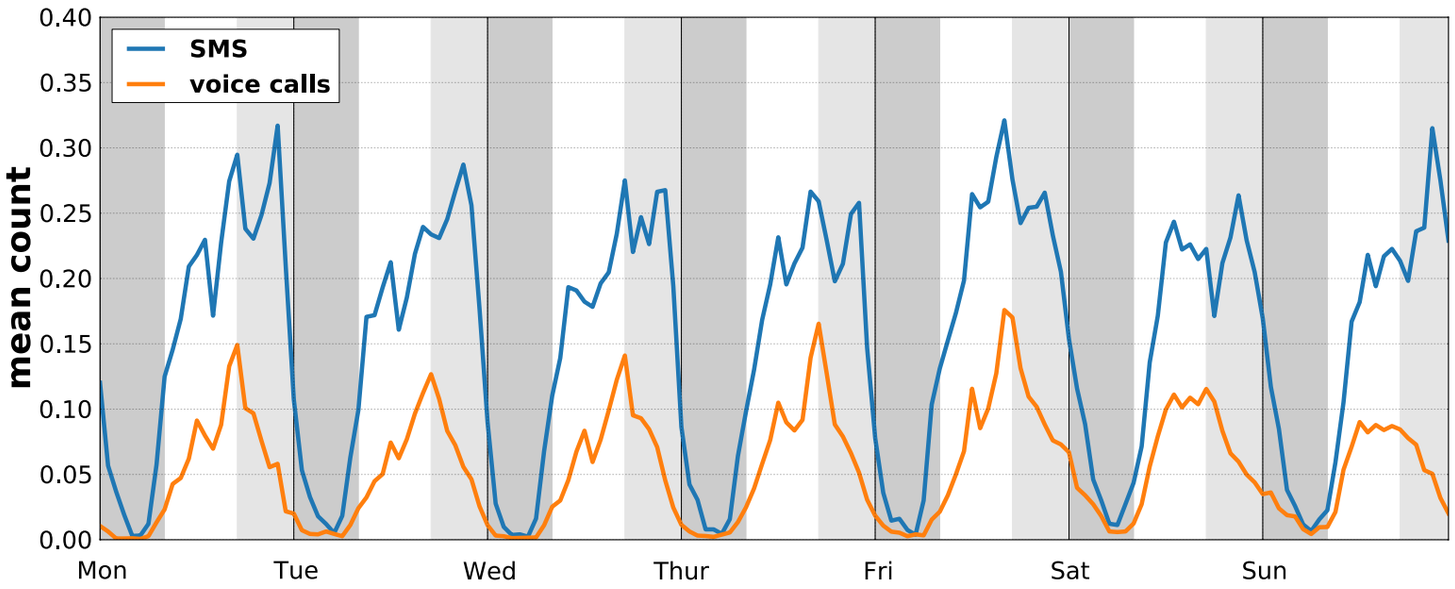


Figure 10. Weekly temporal dynamics of interactions. All calls and SMS', both incoming and outgoing calculated over the entire dataset and averaged per user and per week, showing mean number of interactions users had in given weekly bin. Light gray denotes 5pm, the end of lectures at the university, dark gray covers night between 12am and 8am. SMS is used more for communication outside regular business hours.

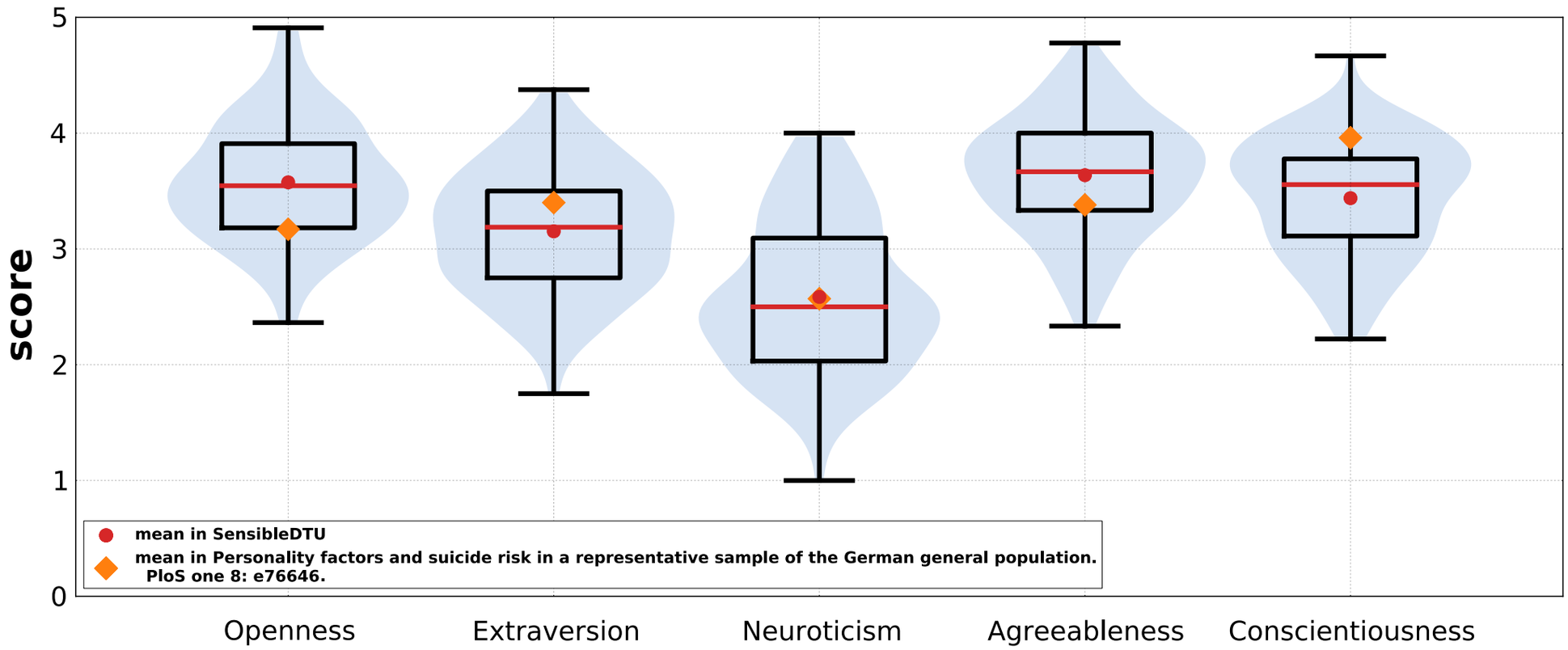


Figure 13. Personality traits Violin plot of personality traits. Summary statistics are: **openness** $\mu_O = 3.58$, $\sigma_O = 0.52$; **extraversion** $\mu_E = 3.15$, $\sigma_E = 0.53$; **neuroticism** $\mu_N = 2.59$ $\sigma_N = 0.65$; **agreeablenes** $\mu_A = 3.64$ $\sigma_A = 0.51$; **conscientiousness** $\mu_C = 3.44$ $\sigma_C = 0.51$. Mean values from our deployment (red circles) compared with mean values reported in [169] (orange diamonds).

technology & privacy


Harry Caul is
an invader of privacy.
The best in the business.
He can record
any conversation
between two people
anywhere.
So far,
three people are dead
because of him.
The Directors Company presents
GENE HACKMAN
in
"THE CONVERSATION"
Co-starring JOHN CAZALE · ALLEN GARFIELD · CINDY WILLIAMS · FREDERIC FORREST
Music scored by DAVID SHIRE
Co-producer FRED ROOS
Written, Produced and Directed by FRANCIS FORD COPPOLA
PG
PARENTAL GUIDANCE SUGGESTED
Color by TECHNICOLOR® A Paramount Pictures Release
74/142
"THE CONVERSATION"

Privacy in Sensor-Driven Human Data Collection:

A Guide for Practitioners

Working Paper

Arkadiusz Stopczynski[1,2,*], Riccardo Pietri[1], Alex 'Sandy' Pentland[2], David Lazer[3], Sune Lehmann[1,4]

1 Technical University of Denmark
2 MIT Media Lab
3 Northeastern University
4 The Niels Bohr Institute

arks@dtu.dk, riccardo.pietri@gmail.com, sandy@media.mit.edu, d.lazer@neu.edu, sljo@dtu.dk

Contents

what is privacy?

permission from appropriate legal entities

data security (attacks, physical, expiration)

informed consent (control & understanding)

right to choose (noise, anonymization, ownership)

sharing (friends & researchers)

what is privacy?

permission from appropriate legal entities

DATATILSYNET

Til forsiden | In English | Læs højt | aA | Ordbog | Udskriv | Sitemap | Kontakt

Hele websitet SØG

Nyheder Afgørelser Fortegnelsen Lovgivning Publikationer Internationalt Om Datatilsynet

Om Datatilsynet

Datatilsynet er den centrale uafhængige myndighed, der fører tilsyn med, at reglerne i persondataloven overholdes. Datatilsynet består af et råd - Datarådet - og et sekretariat.

Datatilsynet bl.a. rådgiver og vejleder, behandler klager og gennemfører inspektioner hos myndigheder og virksomheder.

Læs mere om Datatilsynet

Hvad Datatilsynet ikke kan

Persondatapolitik

Datatilsynets sagsbehandlingstider

Seneste afgørelser

17/01-2013
Transmission af gudstjenester

13/11-2012
Svar på ansøgning om tilladelse til "Dyrenes Vagtcentrals optagelse af opkald"

11/10-2012
Forslag til direktiv om beskyttelse af personoplysninger i forbindelse med straffesager

15/11-2012
Afgørelse i klagesag om tv-overvågning i en boligforening

19/10-2012
Kommunes offentliggørelse af byggesagsarkiv på internettet

30/08-2012
Kommunes ønske om brug af Skype

Læs flere afgørelser

Aktuelt

8/02-2013
Tak for besøget!

5/02-2013
Sikkerhedsbrist i Region Syddanmark

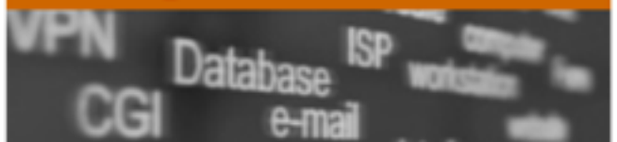
21/12-2012
Ændring af persondataloven fra 1. januar 2013

22/11-2012
Datatilsynet stiller afklarende spørgsmål til Radikale Venstre

Borger >>

Erhverv >>

Offentlig >>



Blanketter >>

what is privacy?

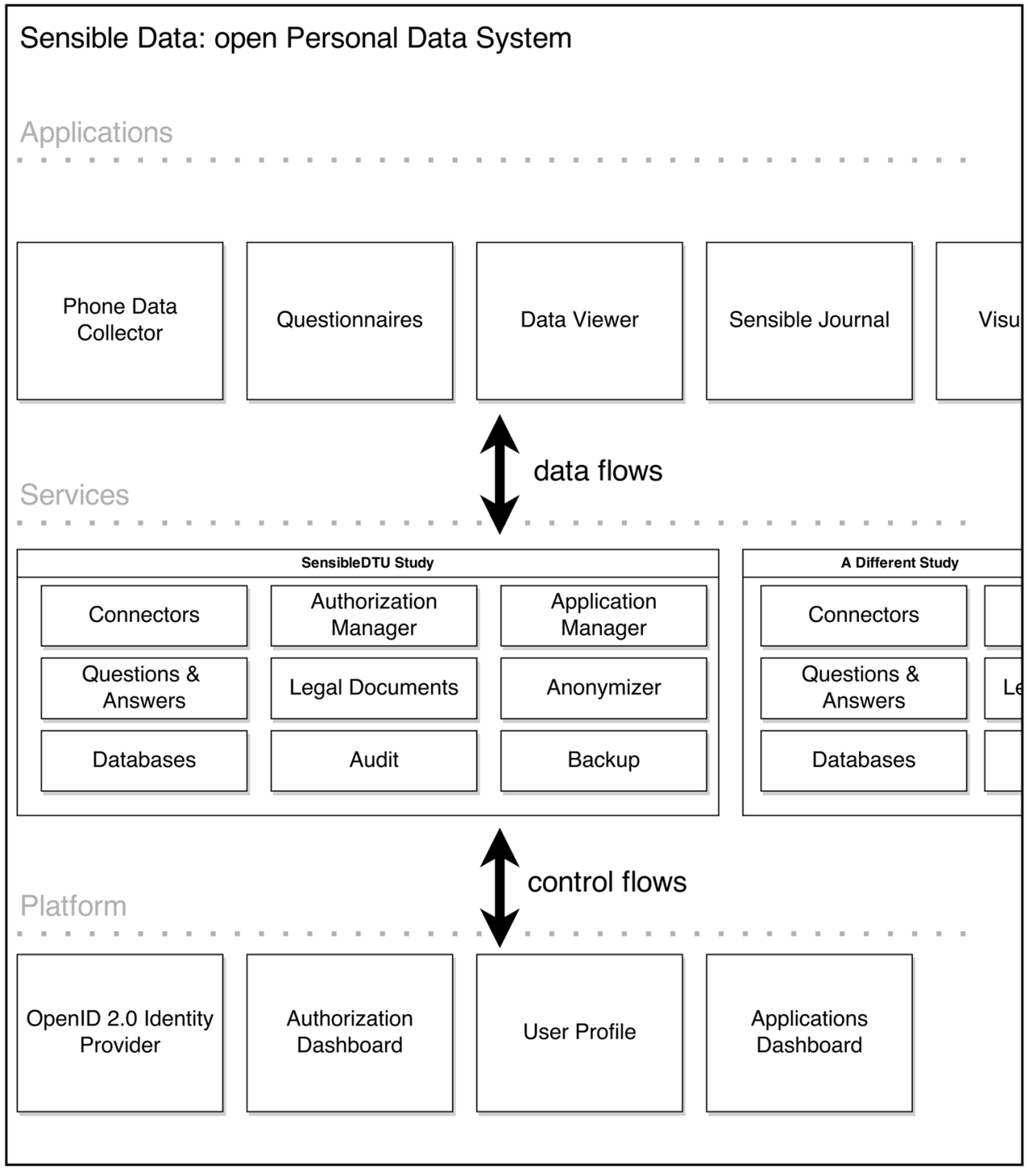
permission from appropriate legal entities

data security (attacks, physical, expiration)

what is privacy?

permission from appropriate legal entities

data security (attacks, physical, expiration)

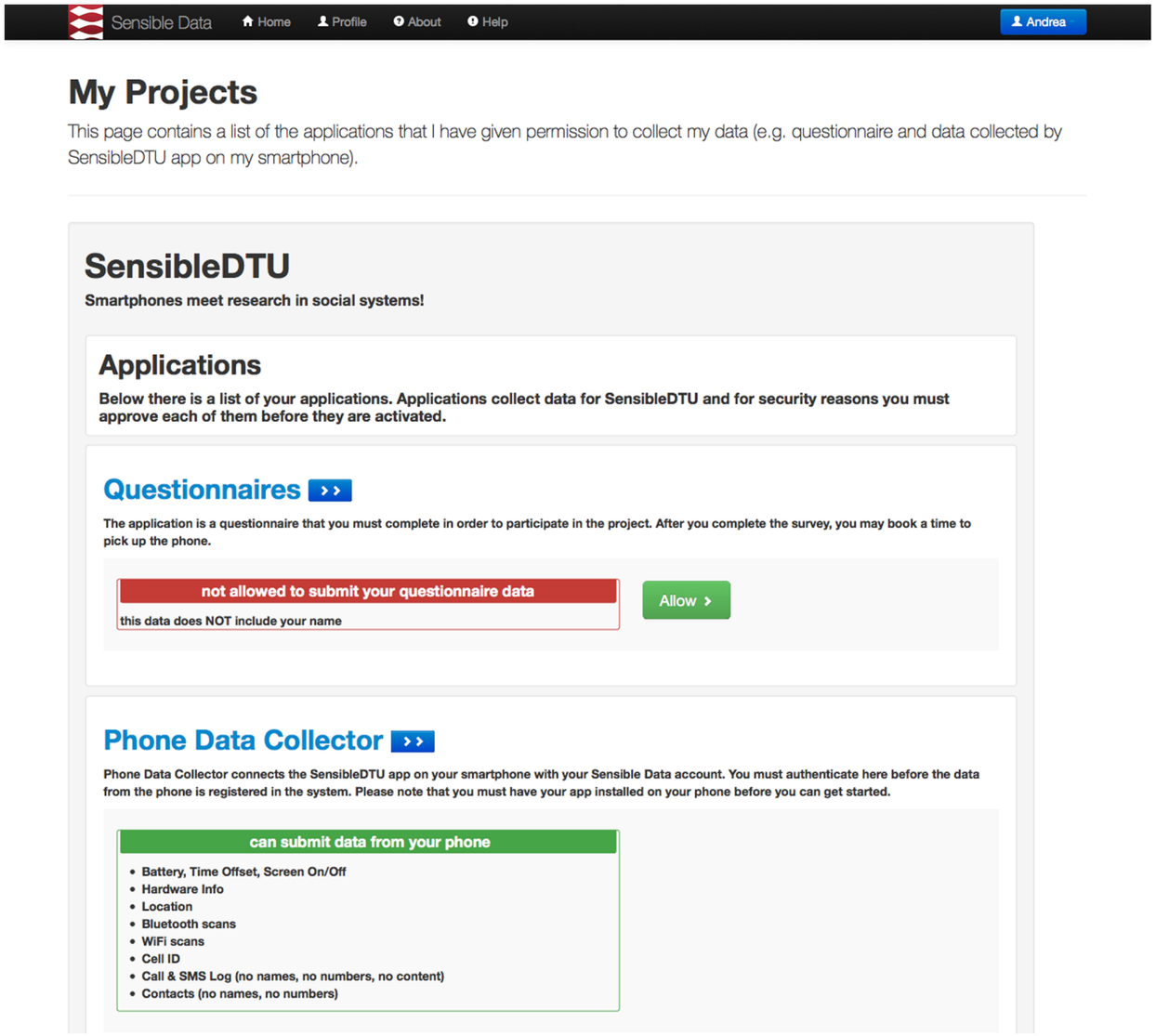


what is privacy?

permission from appropriate legal entities

data security (attacks, physical, expiration)

informed consent (**control** & understanding)



wha

i

Sensible Data | Home | Profile | About | Help | Andrea

My Projects

This page contains a list of the applications that I have given permission to collect my data (e.g. questionnaire and data collected by SensibleDTU app on my smartphone).

SensibleDTU

Smartphones meet research in social systems!

Applications

Below there is a list of your applications. Applications collect data for SensibleDTU and for security reasons you must approve each of them before they are activated.

Questionnaires ››

The application is a questionnaire that you must complete in order to participate in the project. After you complete the survey, you may book a time to pick up the phone.

not allowed to submit your questionnaire data

this data does NOT include your name

Allow ›

Phone Data Collector ››

Phone Data Collector connects the SensibleDTU app on your smartphone with your Sensible Data account. You must authenticate here before the data from the phone is registered in the system. Please note that you must have your app installed on your phone before you can get started.

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)**

what is privacy?

permission from appropriate legal entities

data security (attacks, physical, expiration)

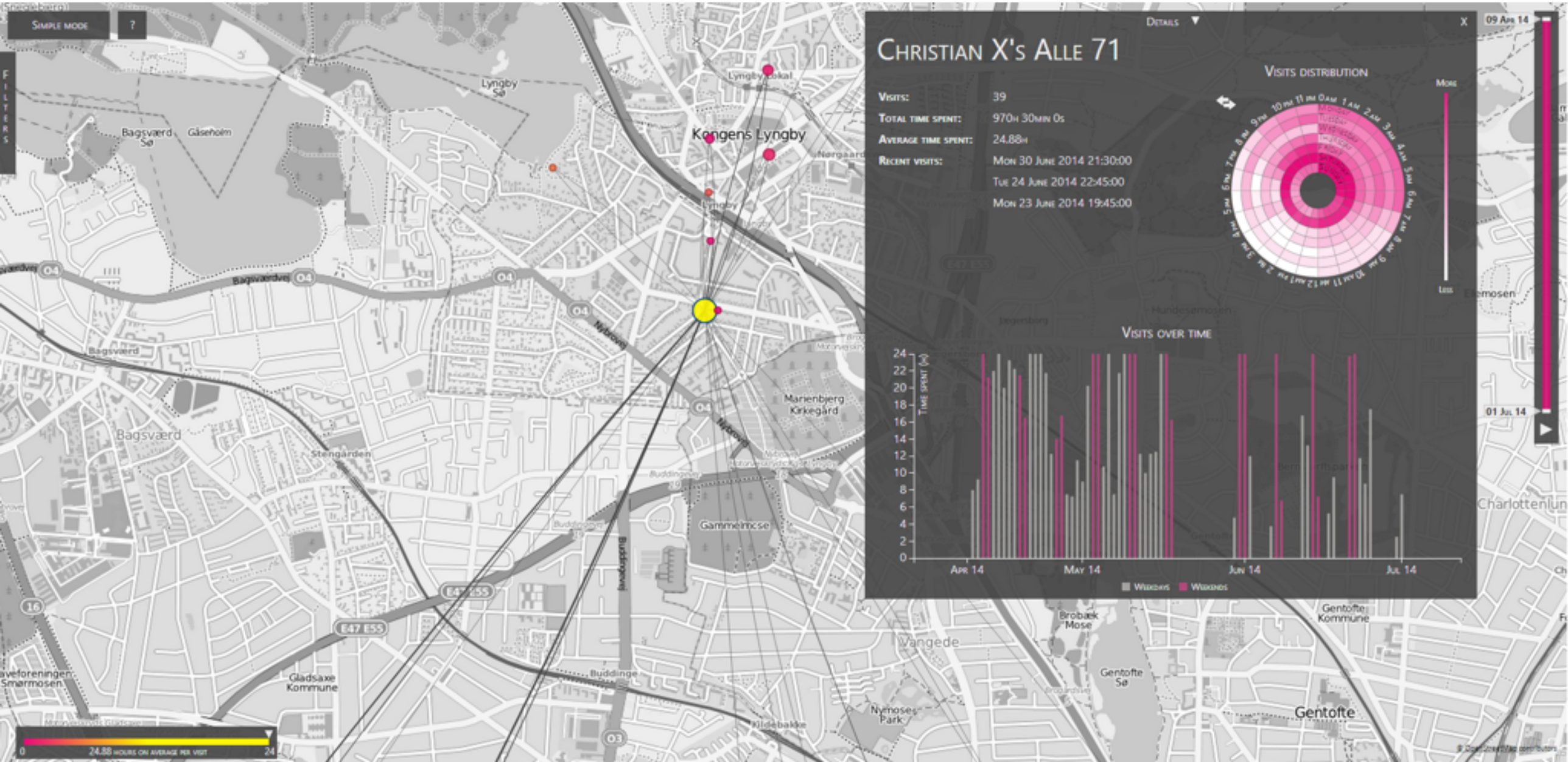
informed consent (control & **understanding**)

what is privacy?

permission from appropriate legal entities

data security (attacks, physical, expiration)

informed consent (control & **understanding**)



what is privacy?

permission from appropriate legal entities

data security (attacks, physical, expiration)

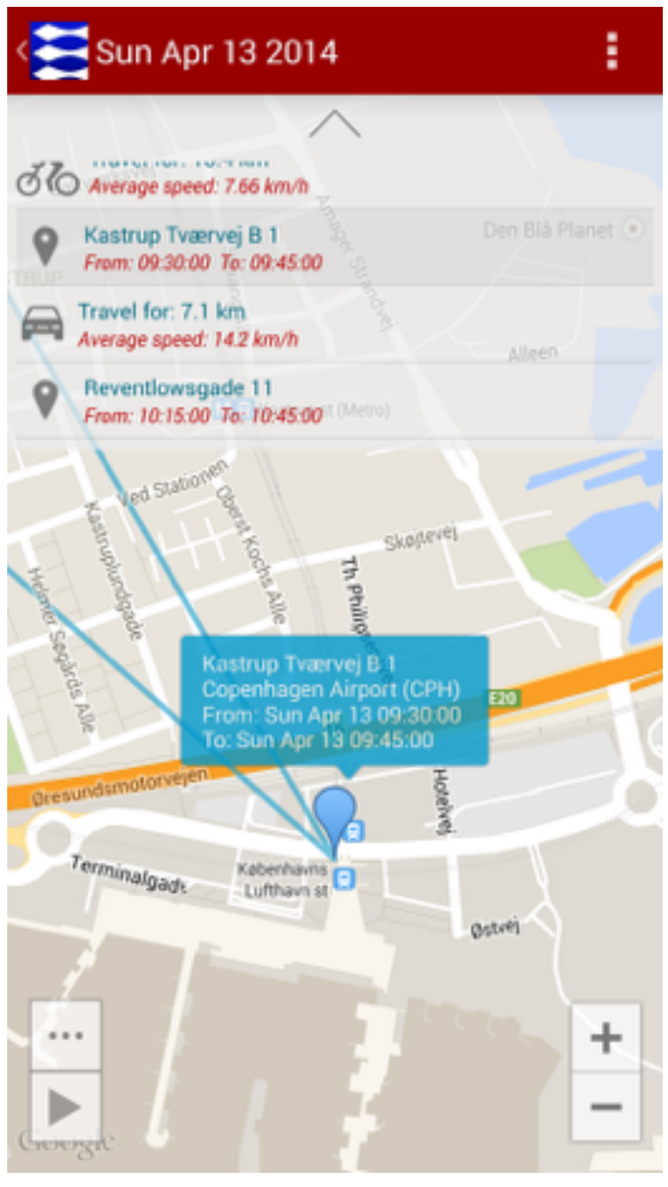
informed consent (control & **understanding**)

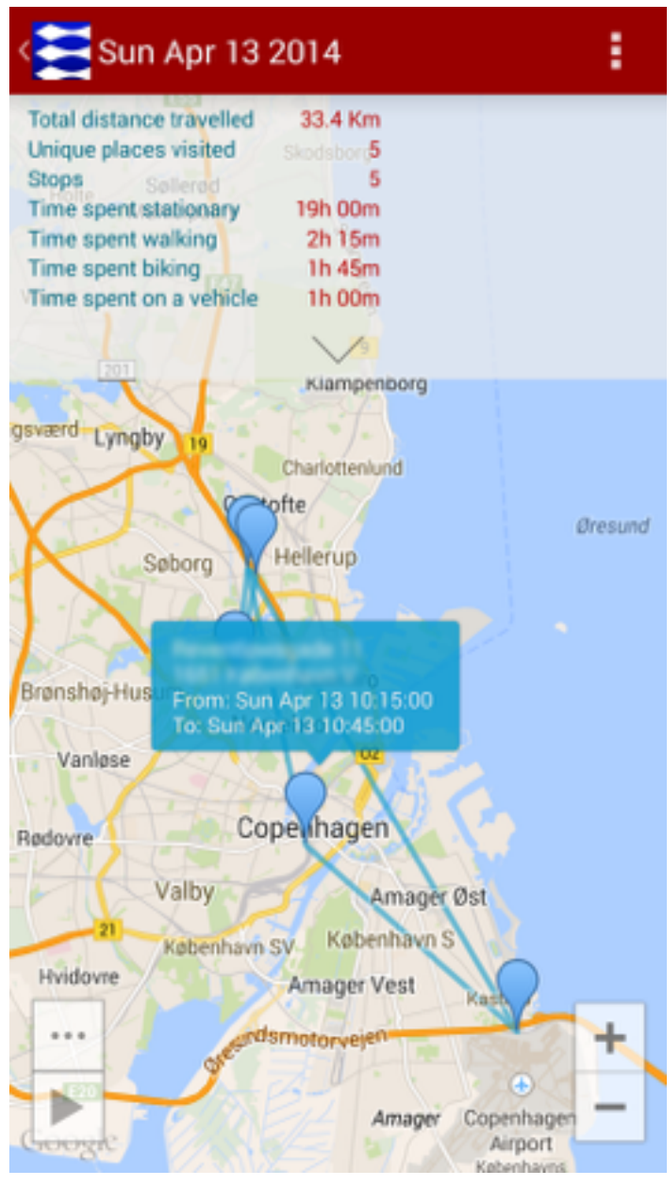
what is privacy?

permission from appropriate legal entities

data security (attacks, physical, expiration)

informed consent (control & **understanding**)





what is privacy?

permission from appropriate legal entities

data security (attacks, physical, expiration)

informed consent (control & understanding)

right to choose (noise, **anonymization**, ownership)

what is privacy?

permission from appropriate legal entities

data security (attacks, physical, expiration)

informed consent (control & understanding)

right to choose (noise, anonymization, ownership)

sharing (friends & researchers)

?

understanding the raw data

OPEN ACCESS Freely available online

PLOS ONE

The Strength of Friendship Ties in Proximity Sensor Data

Vedran Sekara[1]*, **Sune Lehmann**[1,2]

1 Cognitive Systems, Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kgs. Lyngby, Denmark, **2** Niels Bohr Institute, University of Copenhagen, Østerbro, Denmark

Abstract

Understanding how people interact and socialize is important in many contexts from disease control to urban planning. Datasets that capture this specific aspect of human life have increased in size and availability over the last few years. We have yet to understand, however, to what extent such electronic datasets may serve as a valid proxy for real life social interactions. For an observational dataset, gathered using mobile phones, we analyze the problem of identifying transient and non-important links, as well as how to highlight important social interactions. Applying the Bluetooth signal strength parameter to distinguish between observations, we demonstrate that weak links, compared to strong links, have a lower probability of being observed at later times, while such links—on average—also have lower link-weights and probability of sharing an online friendship. Further, the role of link-strength is investigated in relation to social network properties.

Citation: Sekara V, Lehmann S (2014) The Strength of Friendship Ties in Proximity Sensor Data. PLoS ONE 9(7): e100915. doi:10.1371/journal.pone.0100915

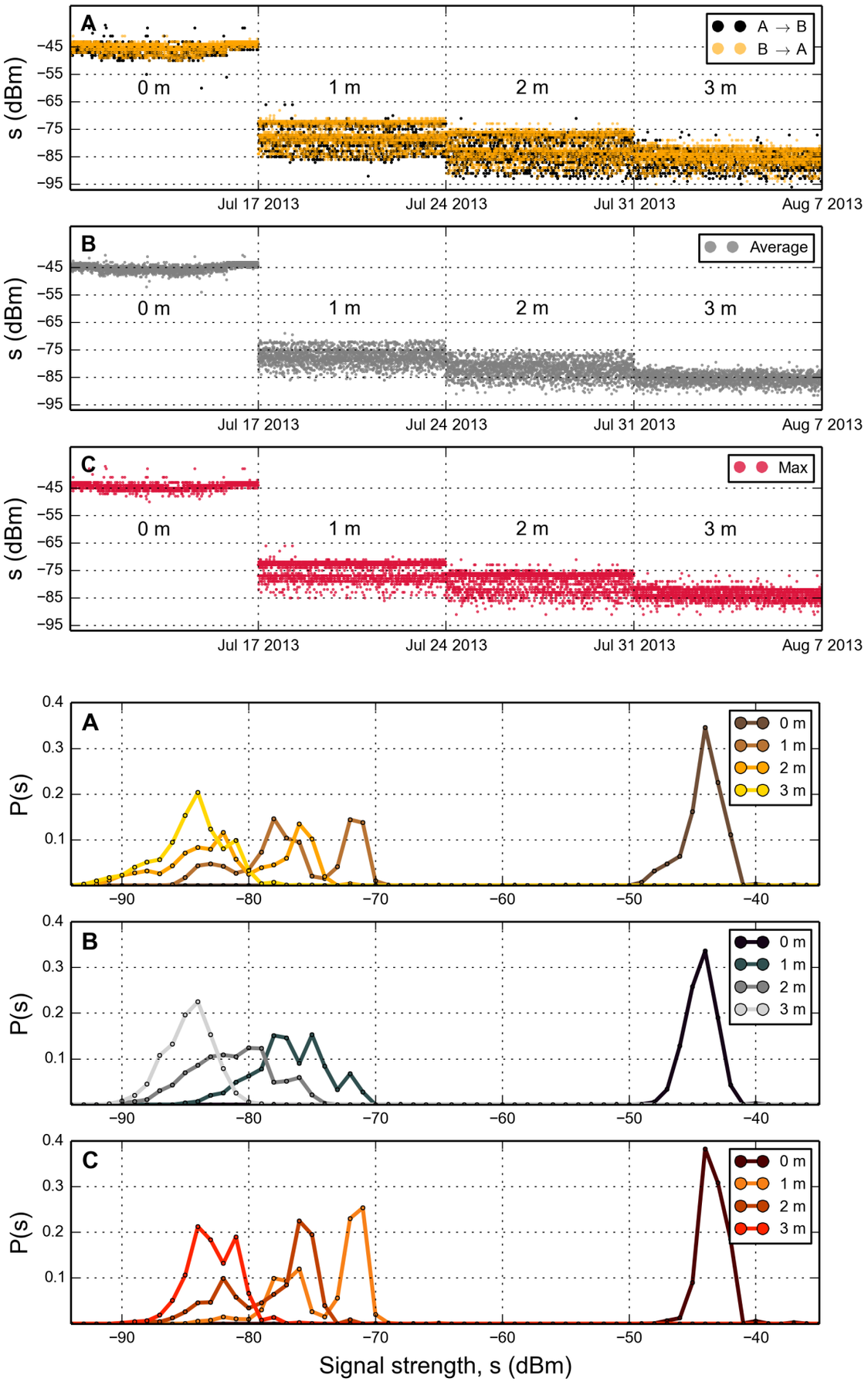
Editor: Christopher M. Danforth, University of Vermont, United States of America

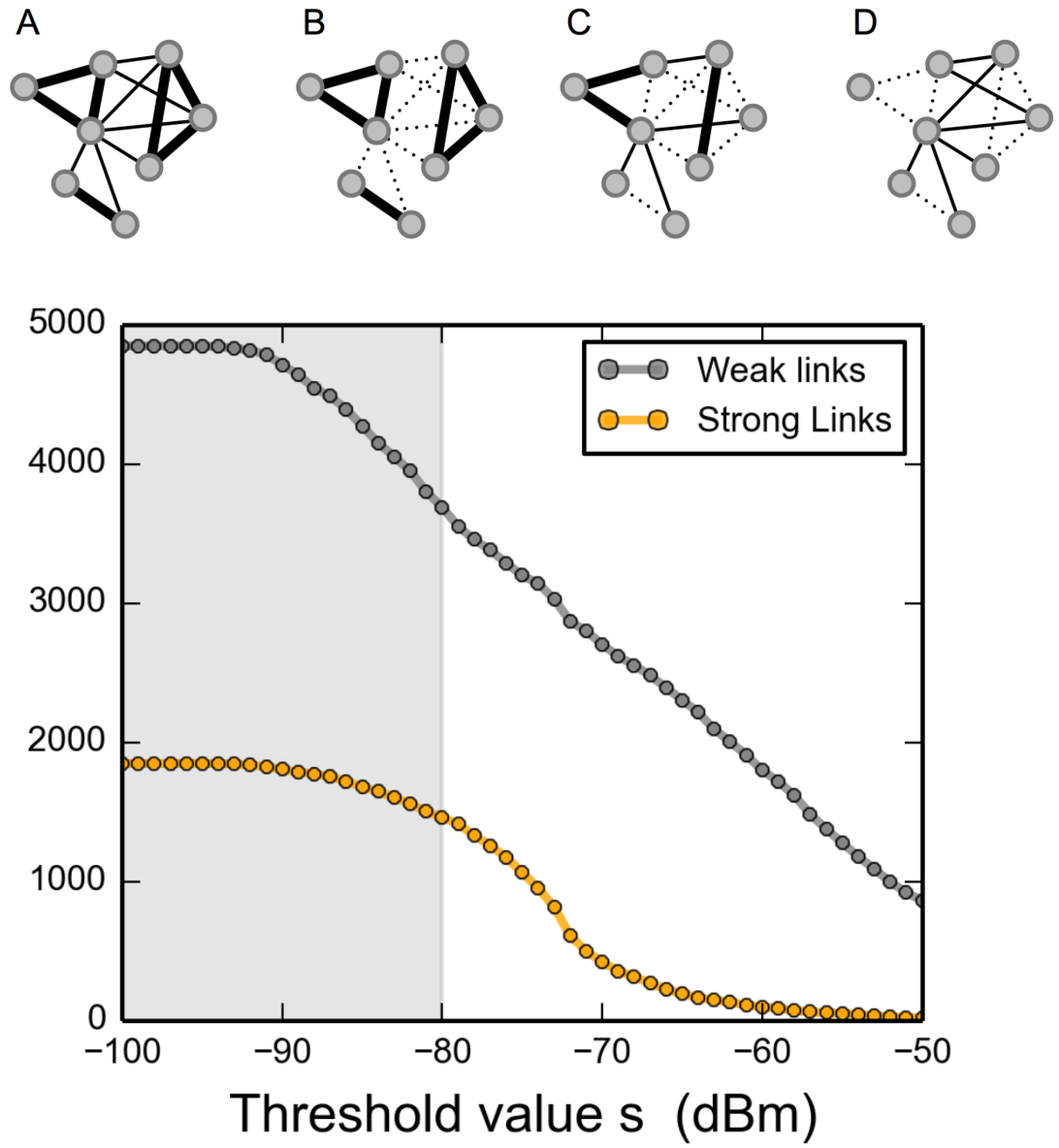
Received January 17, 2014; **Accepted** May 31, 2014; **Published** July 7, 2014

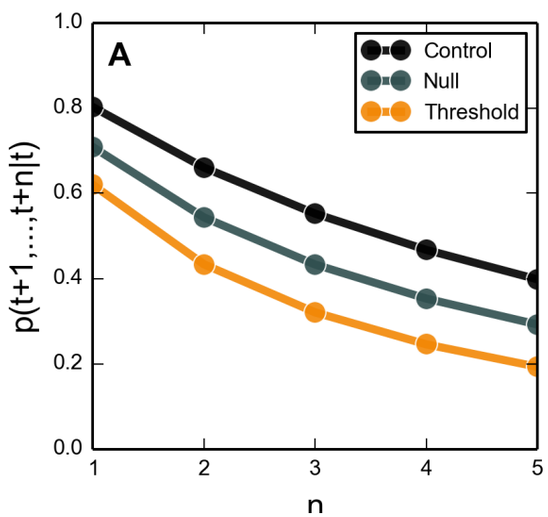
Funding: The authors were funded by a Young Investigator Grant from the Villum Fondation (High Resolution Networks, awarded to SL). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

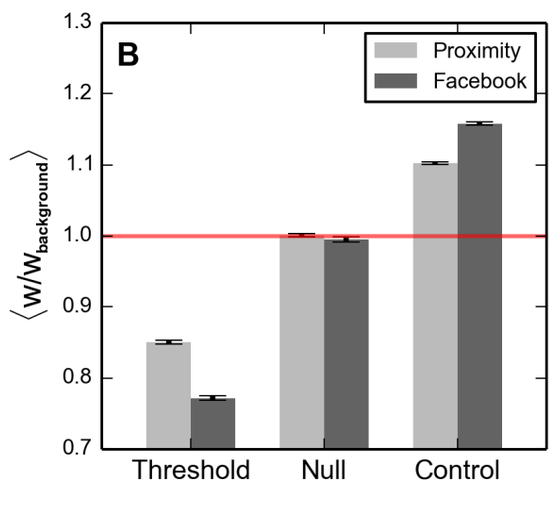
Competing Interests: The authors confirm that one of the authors, Sune Lehmann, is a PLOS ONE Editorial Board member. This does not alter the authors' adherence to PLOS ONE Editorial policies and criteria.

* Email: vese@dtu.dk


A
A → B
B → A
0 m
1 m
2 m
3 m
s (dBm)
−45
−55
−65
−75
−85
−95
Jul 17 2013
Jul 24 2013
Jul 31 2013
Aug 7 2013
B
Average
C
Max
P(s)
0.4
0.3
0.2
0.1
−90
−80
−70
−60
−50
−40
Signal strength, s (dBm)


A
B
C
D
Weak links
Strong Links
Count
5000
4000
3000
2000
1000
0
−100
−90
−80
−70
−60
−50
Threshold value s (dBm)


A
Control
Null
Threshold
p(t+1,...,t+n|t)
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
n


B
Proximity
Facebook
⟨W/W background⟩
1.3
1.2
1.1
1.0
0.9
0.8
0.7
Threshold
Null
Control

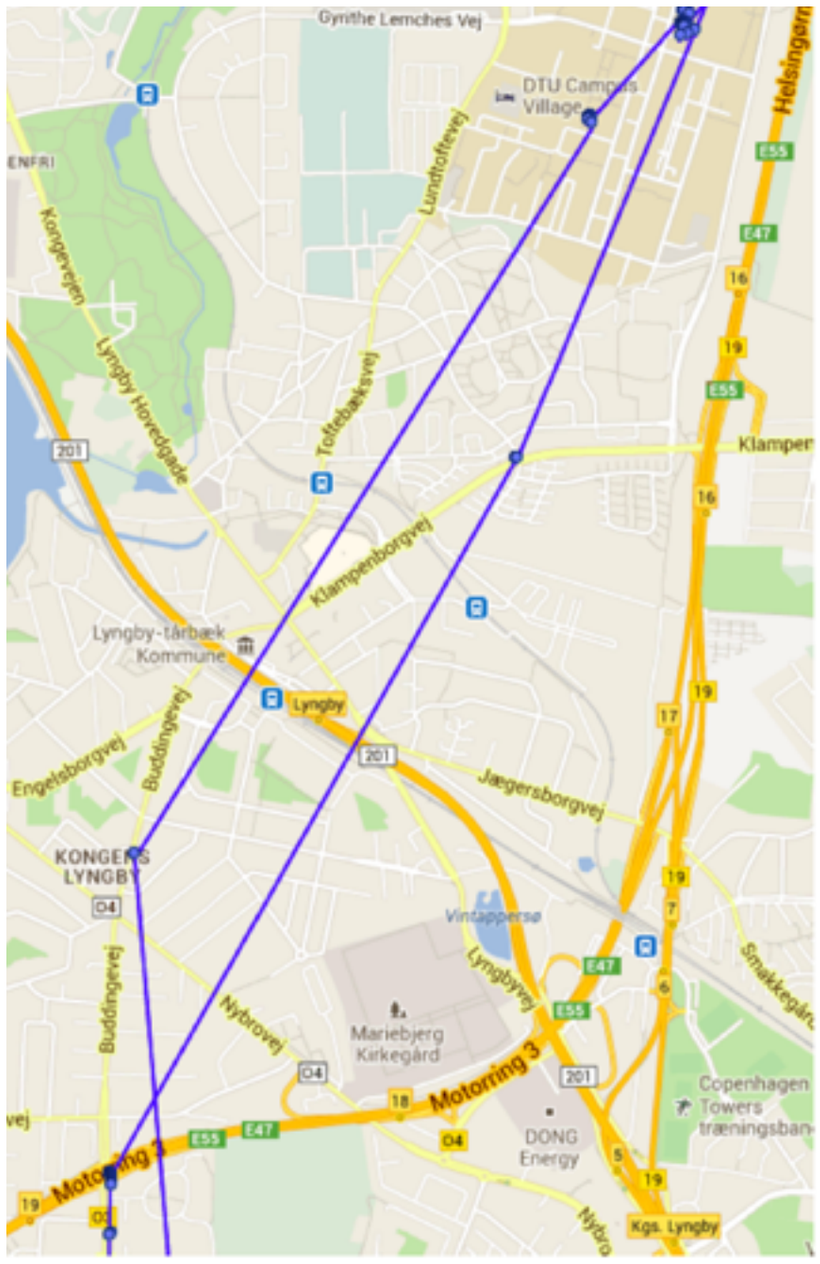
WiFi-Based Sensing of Social Systems Mobility

Piotr Sapiezynski[1,*], Arkadiusz Stopczynski[1,2], Radu Gatej[1], & Sune Lehmann[1,3]

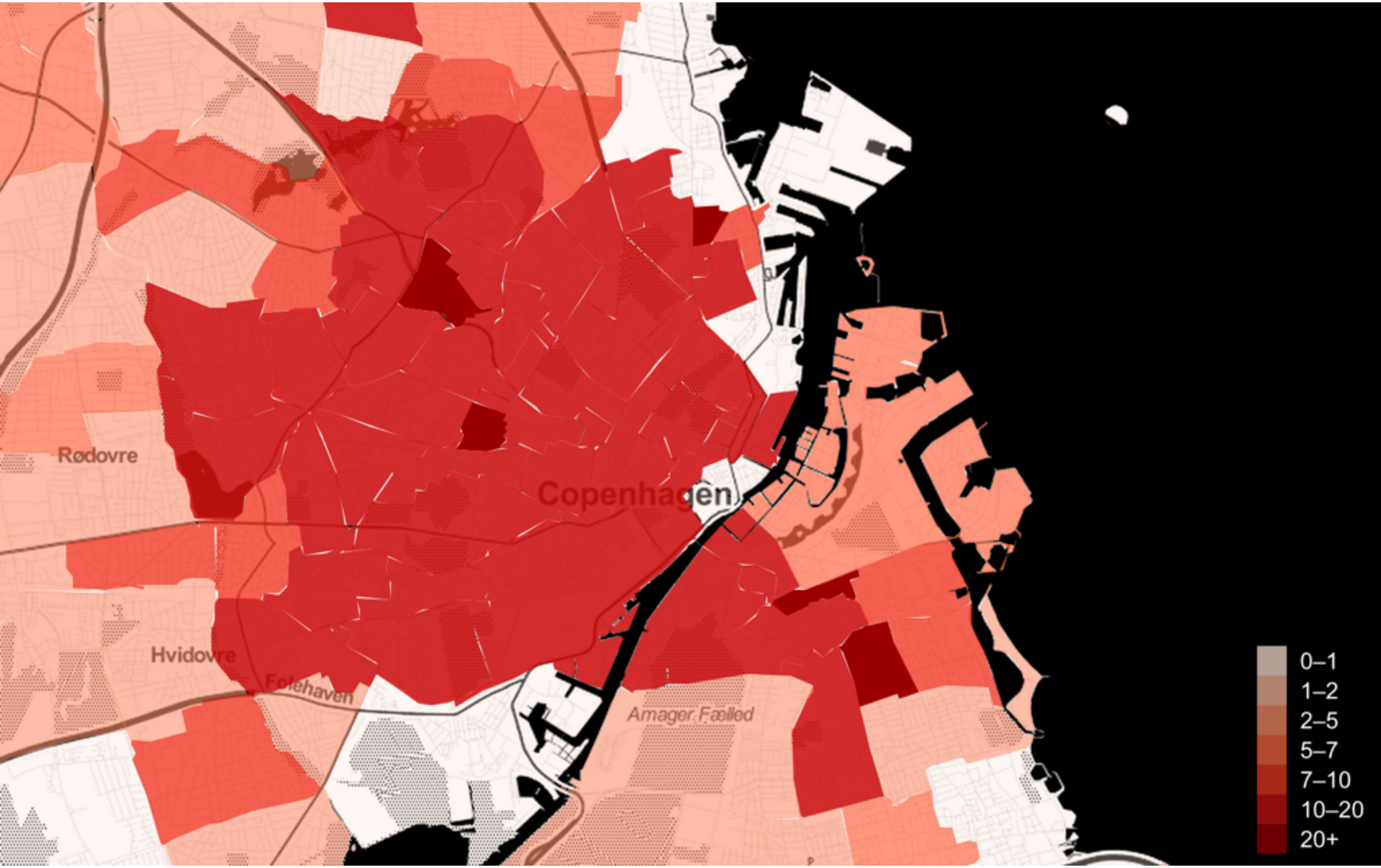
[1] *Department of Computer Science and Applied Mathematics, Technical University of Denmark*

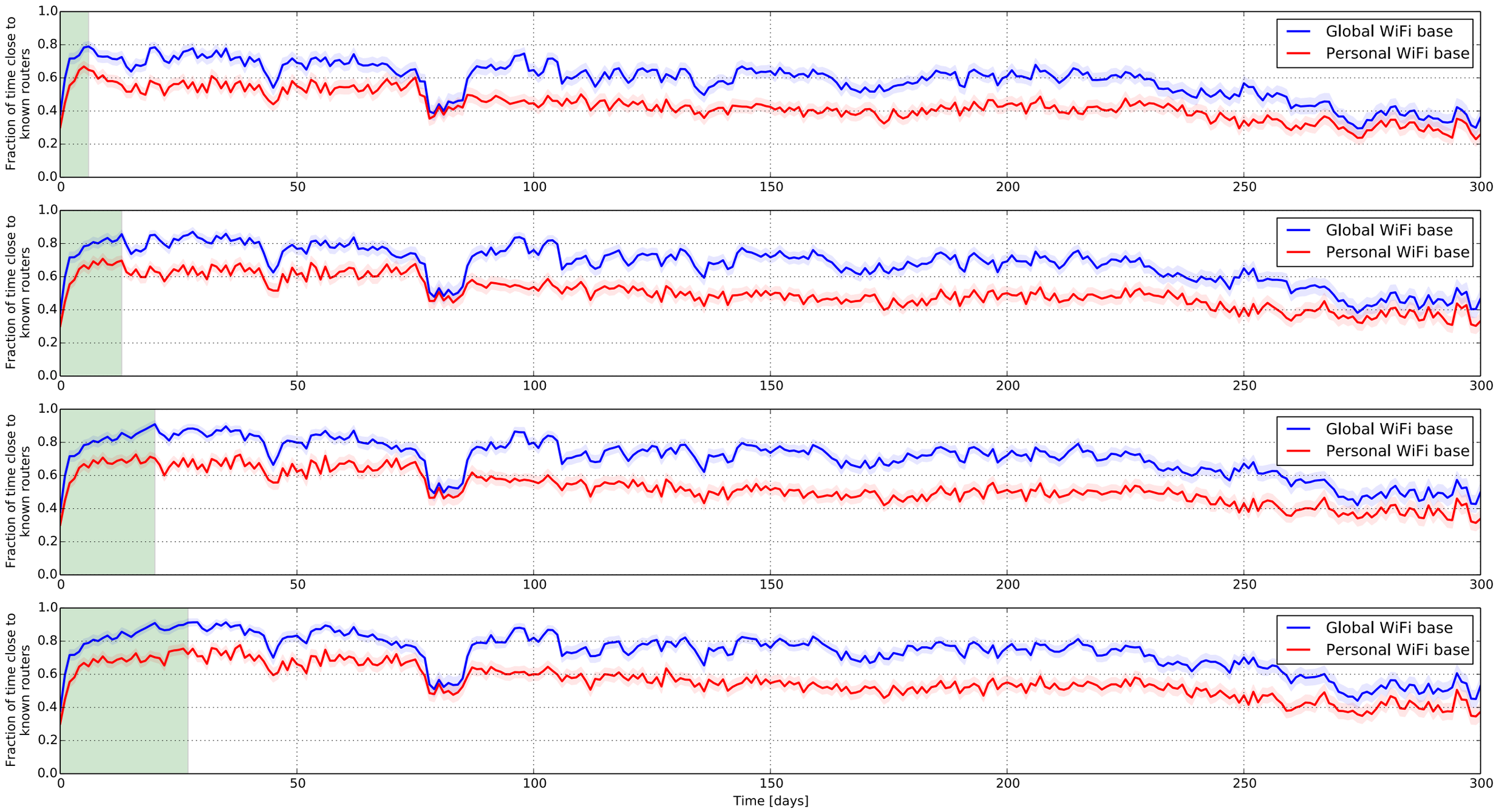
[2] *Media Lab, Massachusetts Institute of Technology*

[3] *Niels Bohr Institute, University of Copenhagen*



(a) (b)


Rødovre
Copenhagen
Hvidovre
Folehaven
Amager Fælled
0–1
1–2
2–5
5–7
7–10
10–20
20+


Fraction of time close to known routers
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
250
300
Global WiFi base
Personal WiFi base
Time [days]

questions and answers for **big data**

Inferring Human Mobility from Sparse Low Accuracy Mobile Sensing Data

Andrea Cuttone
DTU Compute
Dept. of Applied Mathematics and Computer Science
Richard Petersens Plads, Bldg. 324
2800 Kgs Lyngby, Denmark
ancu@dtu.dk

Sune Lehmann
DTU Compute
Dept. of Applied Mathematics and Computer Science
Richard Petersens Plads, Bldg. 324
2800 Kgs Lyngby, Denmark
sljo@dtu.dk

Jakob Eg Larsen
DTU Compute
Dept. of Applied Mathematics and Computer Science
Richard Petersens Plads, Bldg. 324
2800 Kgs Lyngby, Denmark
jaeg@dtu.dk

Abstract
Understanding both collective and personal human mobility is a central topic in Computational Social Science. Smartphone sensing data is emerging as a promising source for studying human mobility. However, most literature focuses on high-precision GPS positioning and high-frequency sampling, which is not always feasible in a longitudinal study or for everyday applications because location sensing has a high battery cost. In this paper we study the feasibility of inferring human mobility from sparse, low accuracy mobile sensing data. We validate our results using participants' location diaries, and analyze the inferred geographical networks, the time spent at different places, and the number of unique places over time. Our results suggest that low resolution data allows accurate inference of human mobility patterns.

ACM Classification Keywords
H.2.8 [Database Management]: Database Applications – data mining

Author Keywords
Spatial data mining; mobile sensing; place discovery; location sensing; mobility;

“real science”

SCIENTIFIC REPORTS

OPEN

The Strength of the Strongest Ties in Collaborative Problem Solving

SUBJECT AREAS:
APPLIED PHYSICS
COMPUTATIONAL SCIENCE
APPLIED MATHEMATICS

Received
28 March 2014

Accepted
14 May 2014

Published
20 June 2014

Correspondence and requests for materials should be addressed to

Yves-Alexandre de Montjoye[1], Arkadiusz Stopczynski[1,2], Erez Shmueli[1], Alex Pentland[1] & Sune Lehmann[2,3]

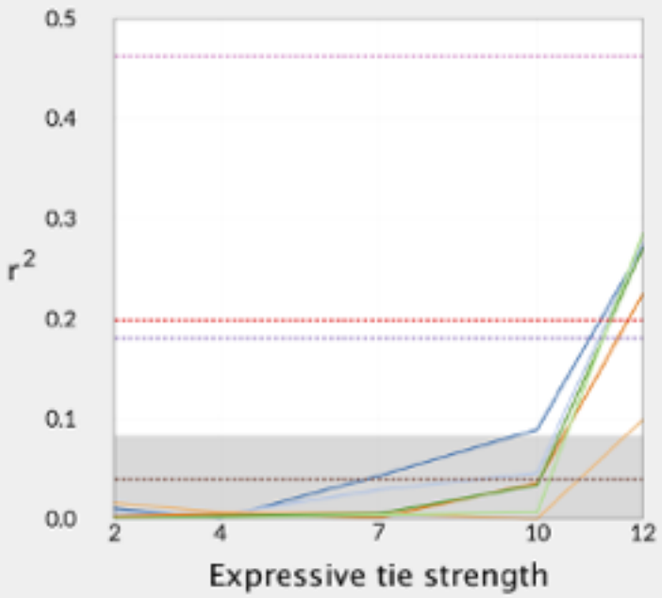
[1]Media Lab, Massachusetts Institute of Technology, [2]Department of Applied Mathematics and Computer Science, Technical University of Denmark, [3]Niels Bohr Institute, University of Copenhagen.

Complex problem solving in science, engineering, and business has become a highly collaborative endeavor. Teams of scientists or engineers collaborate on projects using their social networks to gather new ideas and feedback. Here we bridge the literature on team performance and information networks by studying teams' problem solving abilities as a function of both their within-team networks and their members' extended networks. We show that, while an assigned team's performance is strongly correlated with its networks of expressive and instrumental ties, only the strongest ties in both networks have an effect on performance. Both networks of strong ties explain more of the variance than other factors, such as measured or self-evaluated technical competencies, or the personalities of the team members. In fact, the inclusion of the network of strong ties renders these factors non-significant in the statistical analysis. Our results have consequences for the organization of teams of scientists, engineers, and other knowledge workers tackling today's most complex problems.

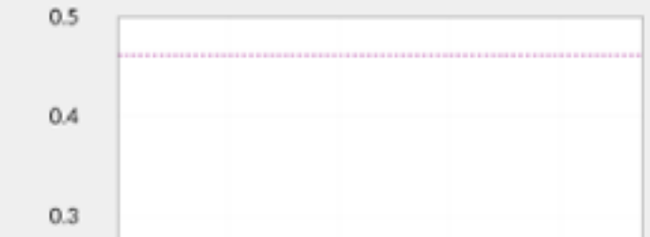
Why networking doesn't work?

Yves-Alexandre de Montjoye, Arkadiusz Stopczynski, Erez Shmueli, Alex 'Sandy" Pentland, and Sune Lehmann

"**Networking**," creating social ties in the workplace, generally increases **performance**. It is common career advice for executives, entrepreneurs, and academics. More companies are investing in common spaces and team building events to support networking.

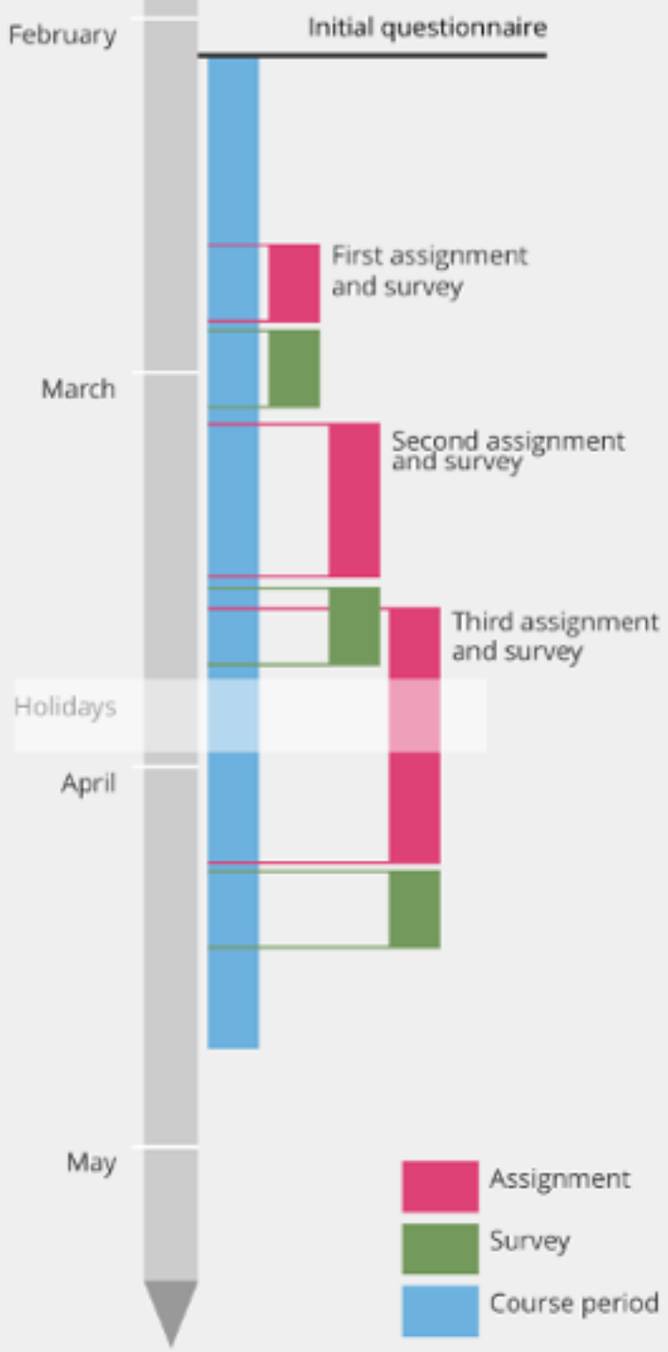


But networking doesn't seem to improve **performance in competitive environments**. Our experiment at a large western university showed that when faced with a complex problem, the **weak ties of team members did not have any effect** on their performance. Frequency of interactions with the people we'd consider to be acquaintances, and even with those we'd would consider friends, did not help your performance.



Network Measures
- Mean/Max team degree
- Mean/Max team in-degree

A Priori Competencies
- Self-evaluated technical competencies
- Measured technical competencies



Methodology of the experiment

We measured expressive ties, personality, and technical competencies through a questionnaire at the beginning of the experiment. We assigned participants to teams based on their expressive ties. Participants were assigned to three different teams over the course of the experiment. We used the logs of the university wifi system to infer the time participants were spending together. The effect is strongly non-linear: only the strongest ties in both expressive and instrumental networks have an effect on performance. Networks of both instrumental and expressive strong ties explain more of the variance than any other considered factors, such as measured or self-evaluated technical competencies, or the personalities of the team members; and the inclusion of the network of strong ties renders these factors non-significant in the statistical analysis.

The Strength of the Strongest Ties in Collaborative Problem Solving
de Montjoye, Y. A., Stopczynski, A., Shmueli, E., Pentland, A.S., Lehmann, S. (2014). The Strength of the Strongest Ties in Collaborative Problem Solving. Nature S.Rep., 4.

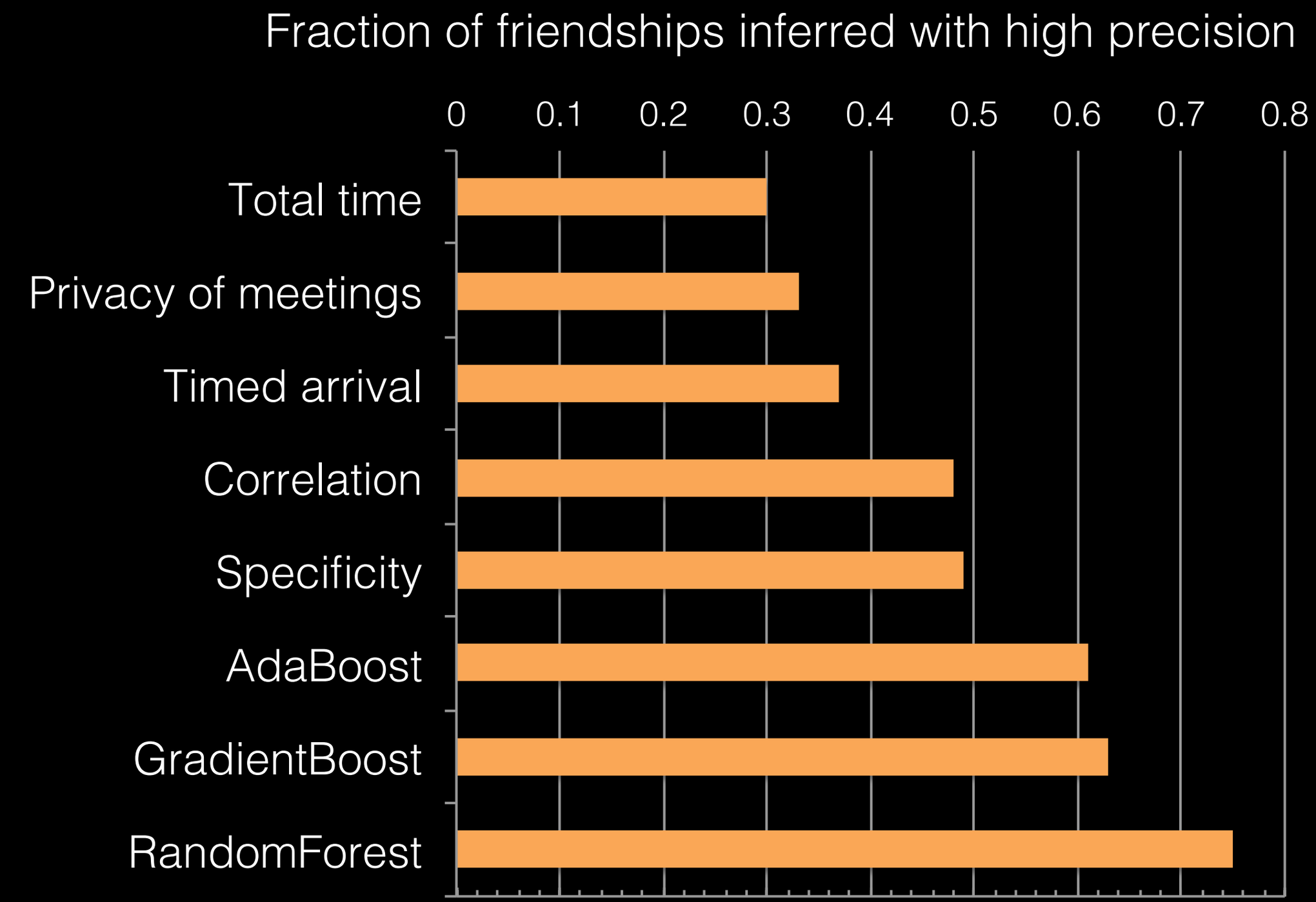
Contact: yva@mit.edu

MIT MEDIA LAB

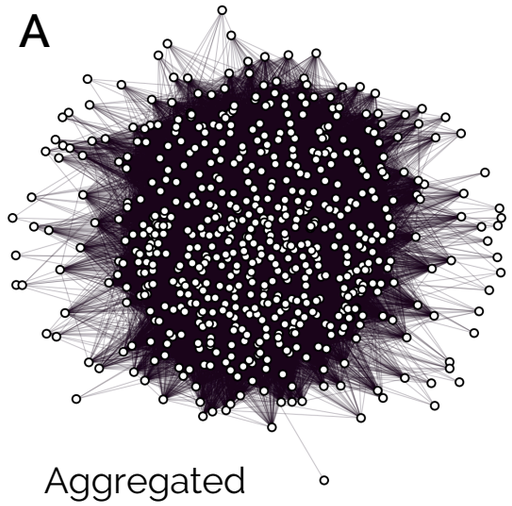
DTU Technical University of Denmark

work in progress

Inferring friendships from WiFi data

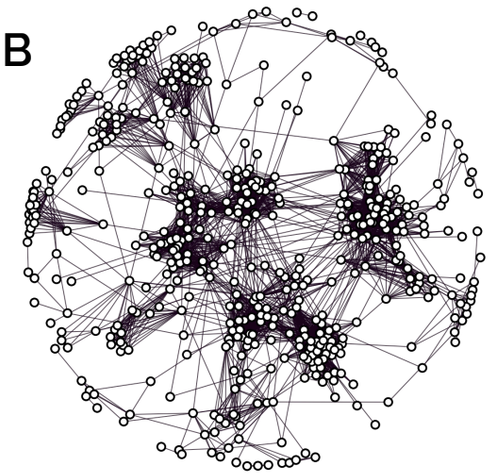


A



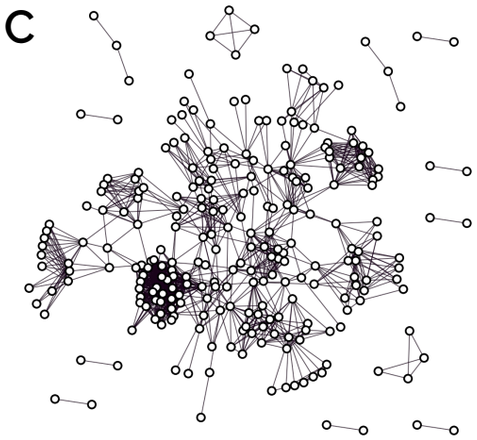
Aggregated

B



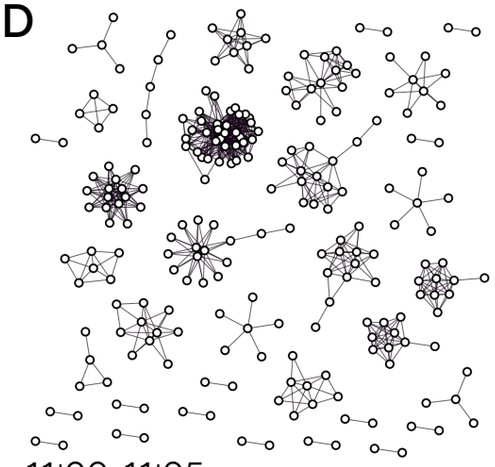
Thresholded

C



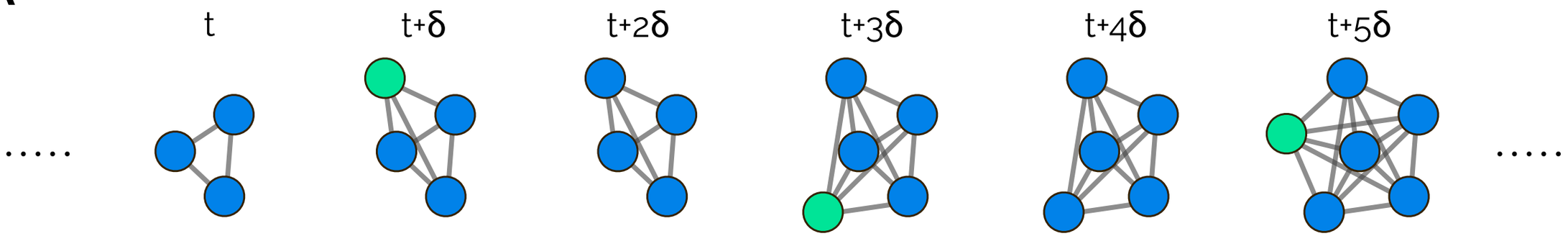
11:00-12:00

D

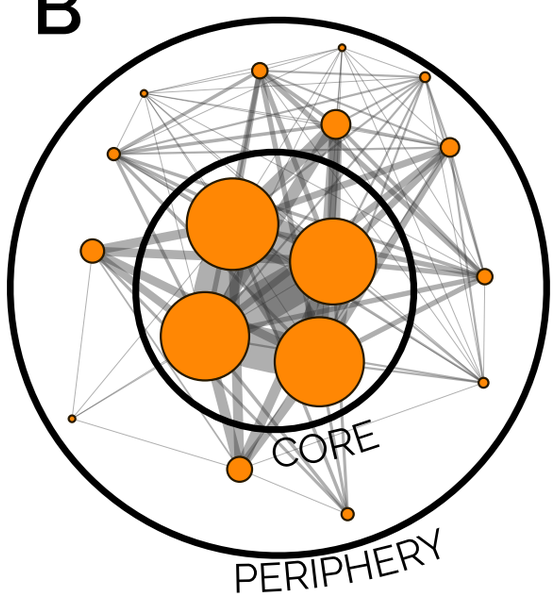


11:00-11:05

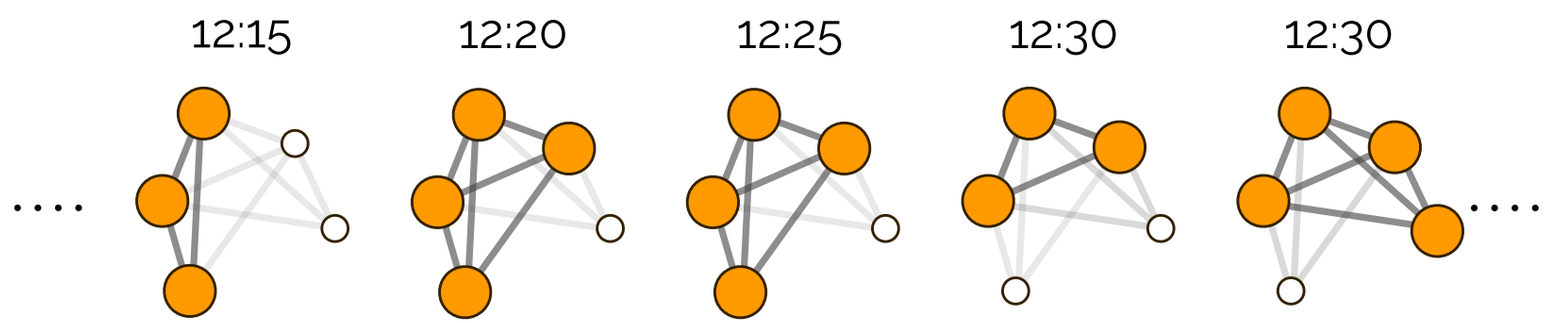
A



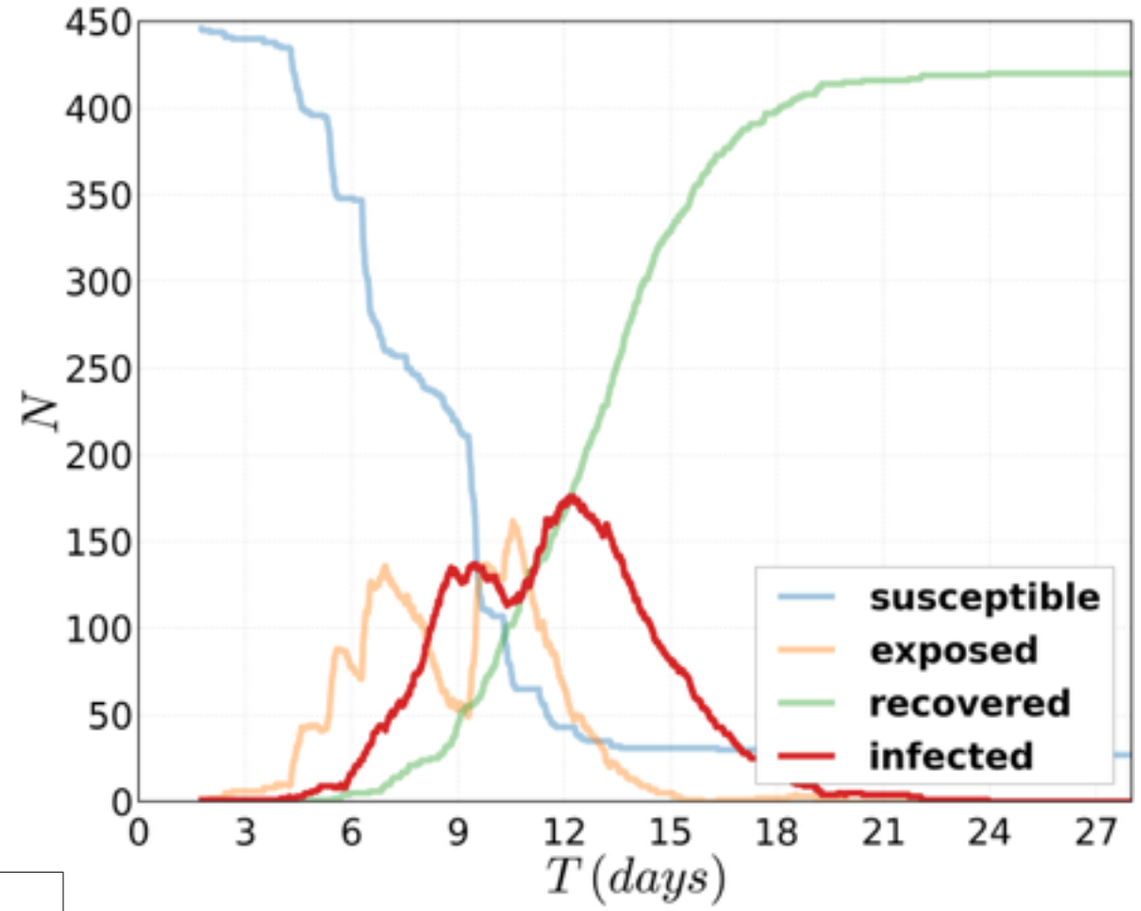
B

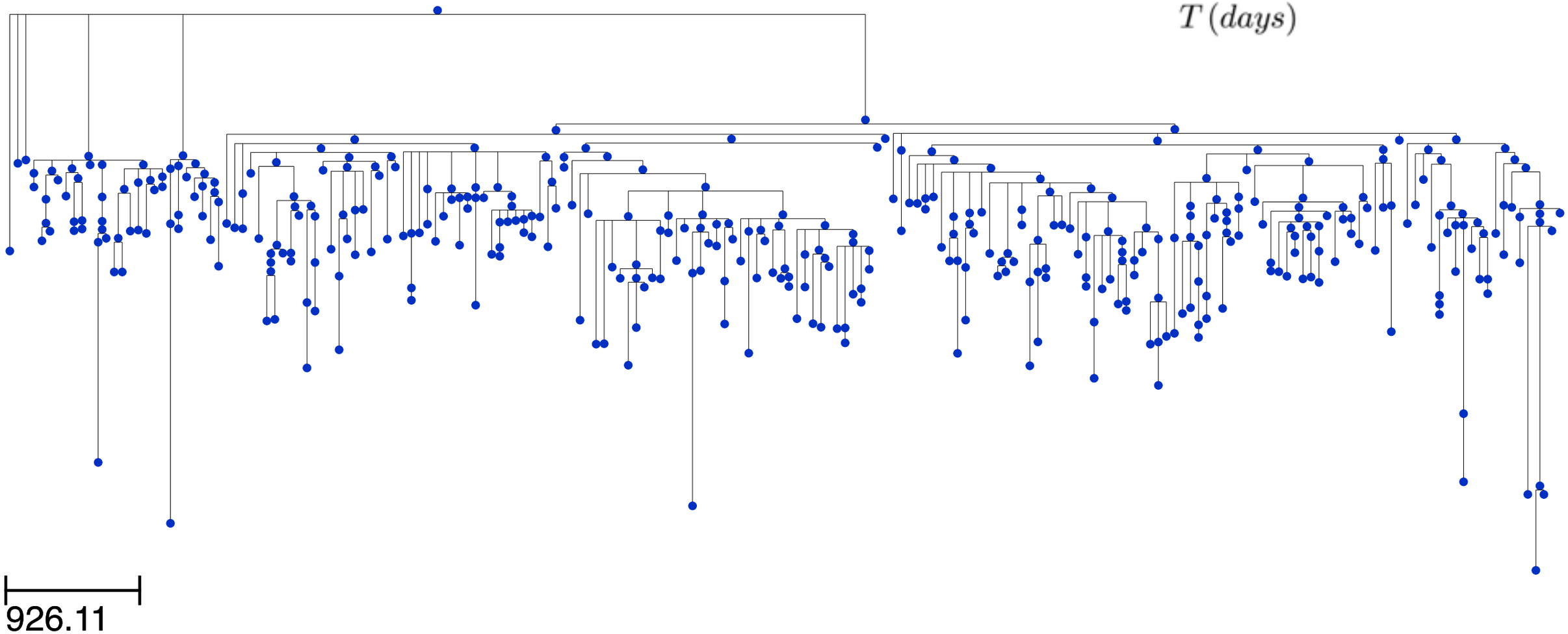


C



epidemics


450
400
350
300
250
200
150
100
50
0
N
0
3
6
9
12
15
18
21
24
27
T (days)
susceptible
exposed
recovered
infected


926.11