Reality Mining for Real: Large-Scale Human Behavior and Smartphone Data

Daniel Gatica-Perez

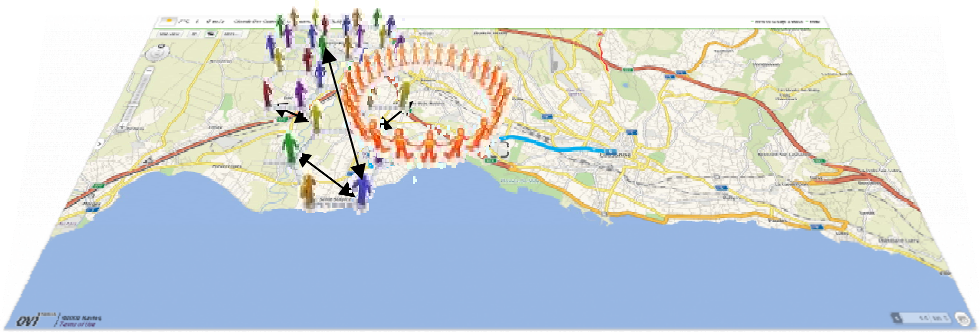
Idiap Research Institute

Switzerland

03.09.2011

photo: lusofox@flickr

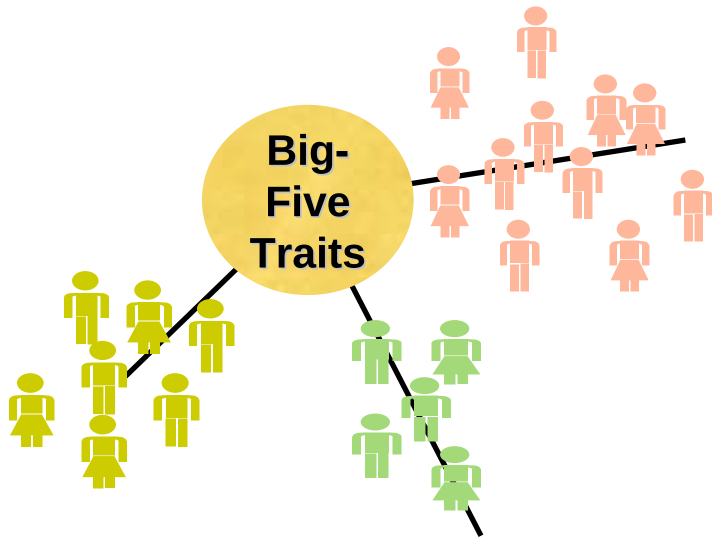
today's talk



large-scale smartphone data

discovering interaction types from phone data

smartphone usage in the wild



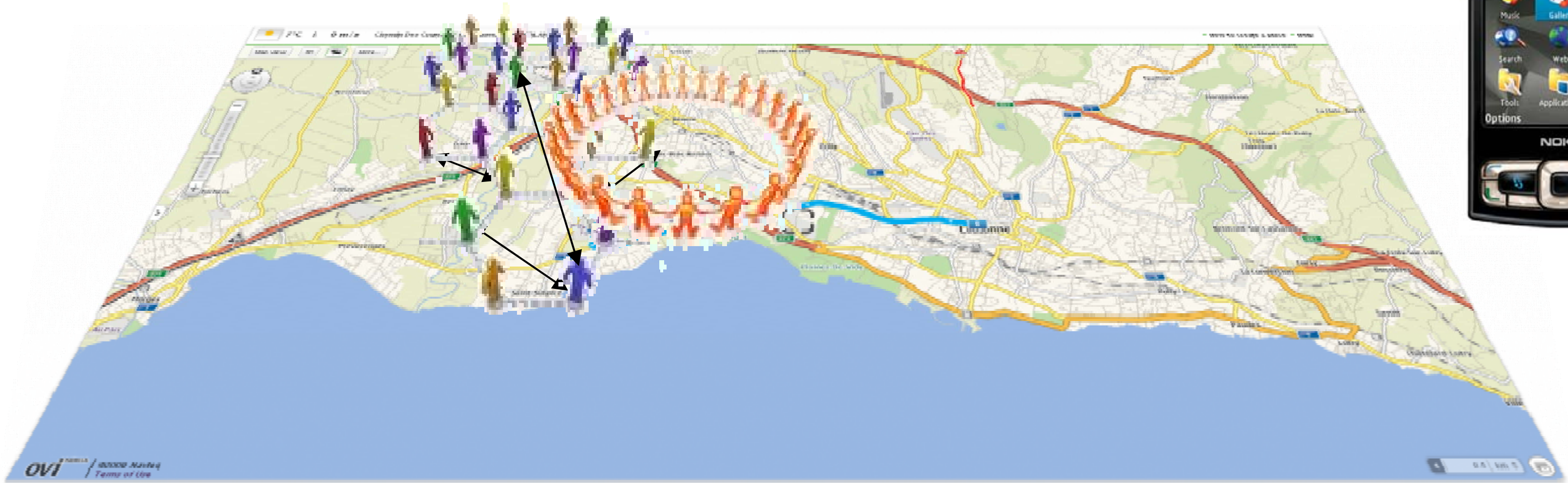
personality and phone data

part 1:
large-scale smartphone data

(joint work with Olivier Bornet, Niko Kiukkonen, Juha Laurila, Olivier Dousse, and Jan Blom)

N. Kiukkonen, J. Blom, O. Dousse, D. Gatica-Perez, and J. Laurila., "Towards Rich Mobile Phone Datasets: Lausanne Data Collection Campaign," ***in Proc. Int. Conf. on Pervasive Services (ICPS),*** **Berlin, Jul. 2010.**

large-scale data collection



+ 185 voluntary participants

+ strong social aspect

+ 18 months of duration

+ privacy preserving protocol

state-of-the-art sensing
+ N95 smartphones
+ 24/7 sensing

a European population
+ public transportation
+ several sub-communities

what does the phone sense?

- GPS
- Cell Tower ID + signal strength
- WLAN access points
- Bluetooth
- Accelerometer
- Phone call logs
- SMS logs
- Audio
- Photos
- Video
- Apps…

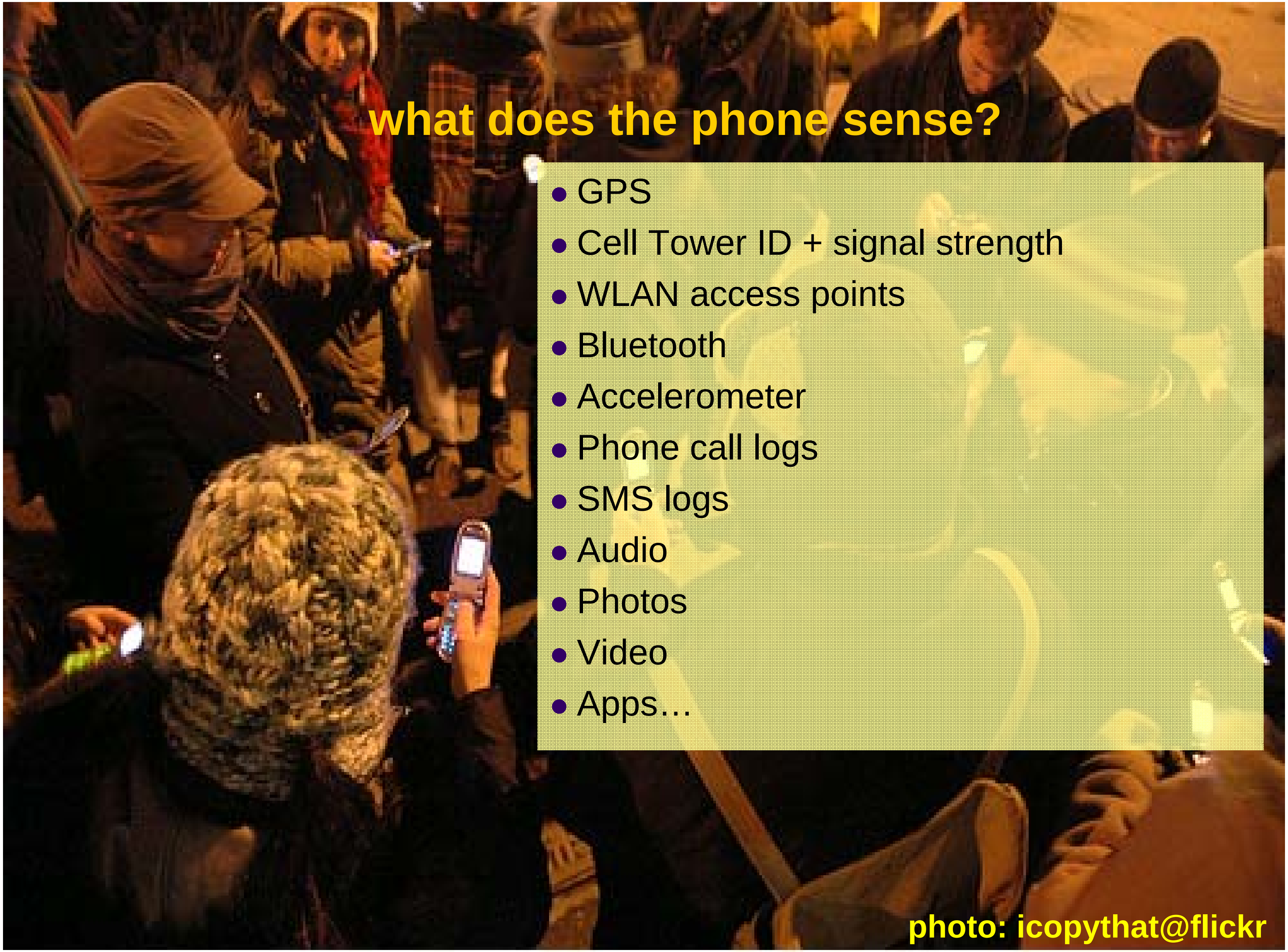
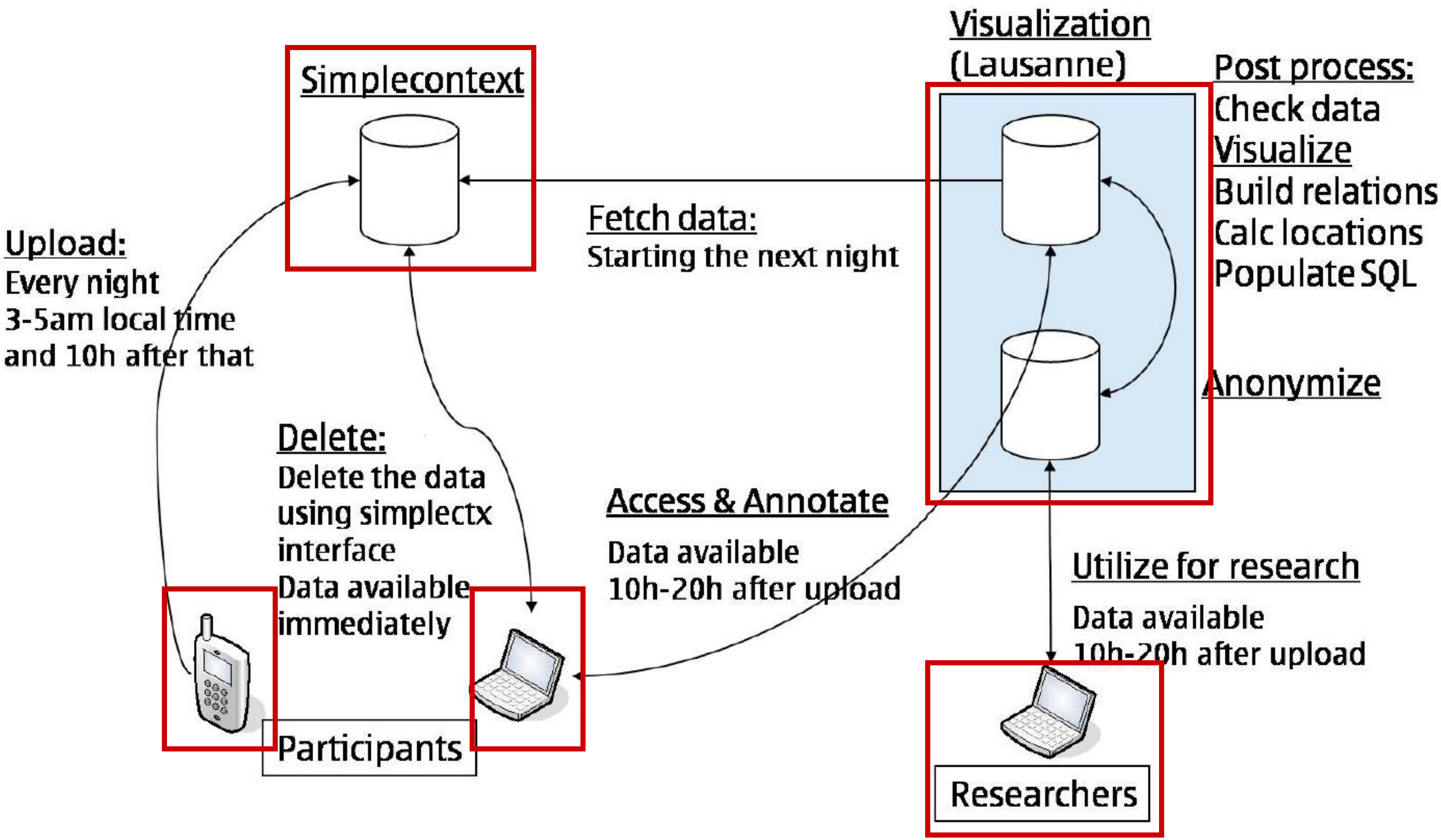


photo: icopythat@flickr

overall architecture



data rights

all users are volunteers

users **own** their data

users can **visualize, research, or share** their own data

anonymized data is used for research

technical challenges: battery life and sampling rate

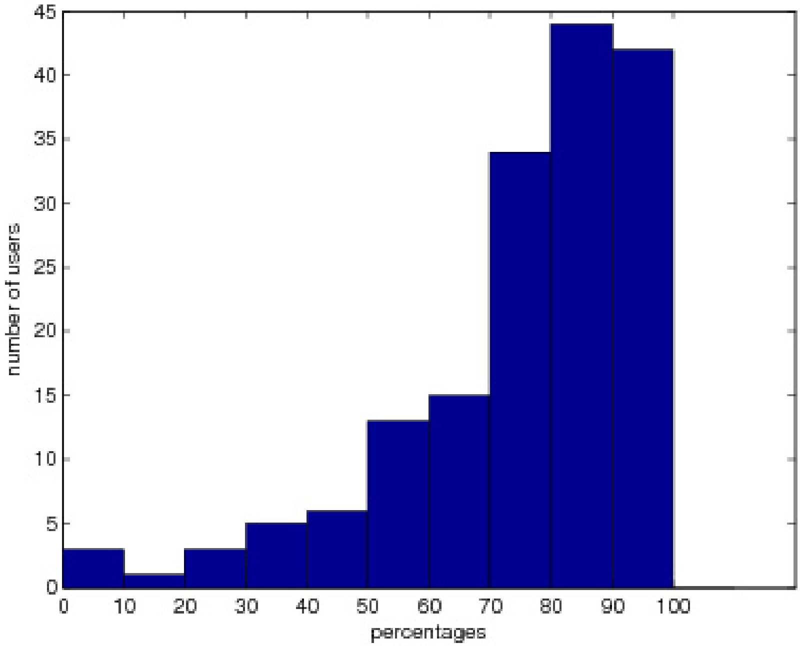
all sensors on: battery lasts 3 hours

design criteria

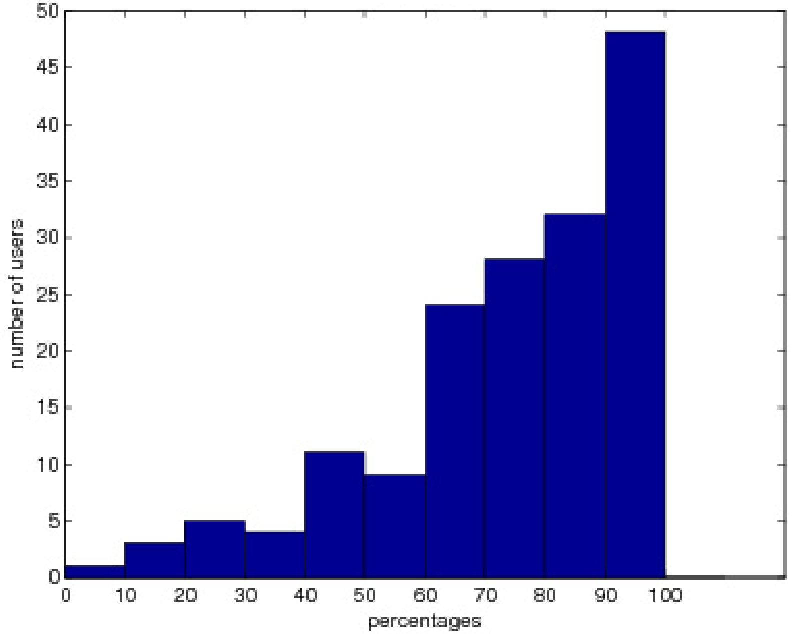
+ current sensor status
+ status of user activity
+ past observations
+ battery status

technical challenges: missing data

Location



Proximity



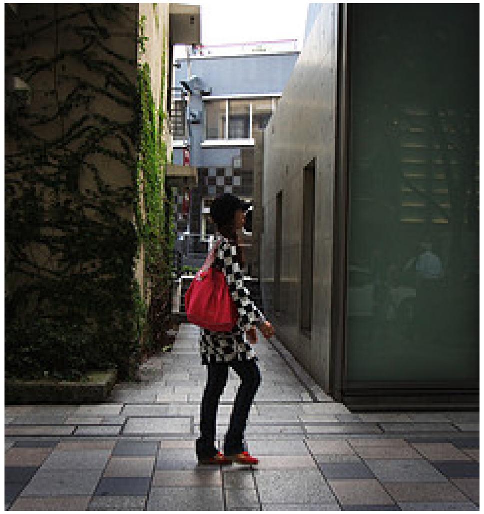
Percentage of days for which users have location or proximity data

part 2:
discovering interaction types with topic models

(joint work with Trinh-Minh-Tri Do)

T. Do and D. Gatica-Perez, "GroupUs: Smartphone Proximity Data Proximity Data and Human Interaction Type Mining," ***in Proc. IEEE Int. Symposium on Wearable Computers (ISWC)*****, San Francisco, Jun. 2011.**

human routines and interaction types



routine: temporal regularities related to location, proximity, and communication

interaction types:

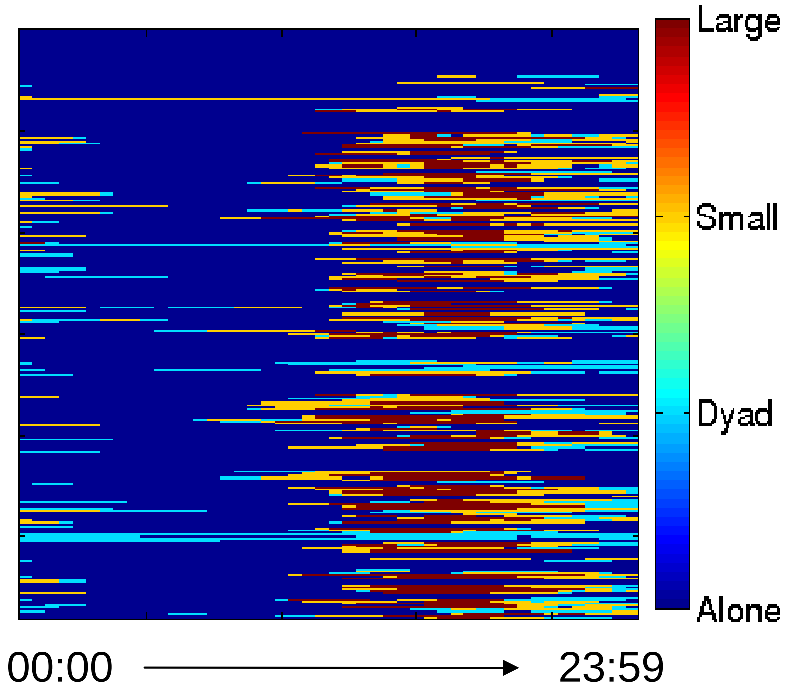
- what groups exist in an organization?
- when do people spend time together?

challenges: data is massive, noisy, and incomplete

annotation is sparse (when available)

bluetooth proximity data

Bluetooth devices detected within 10m radius



a relational model to discover interaction types

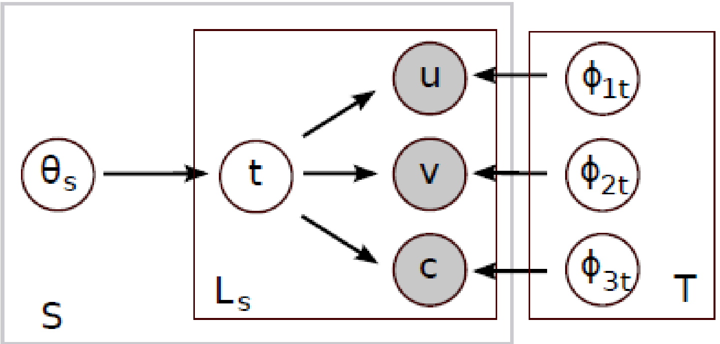


Figure 1. Graphical model.

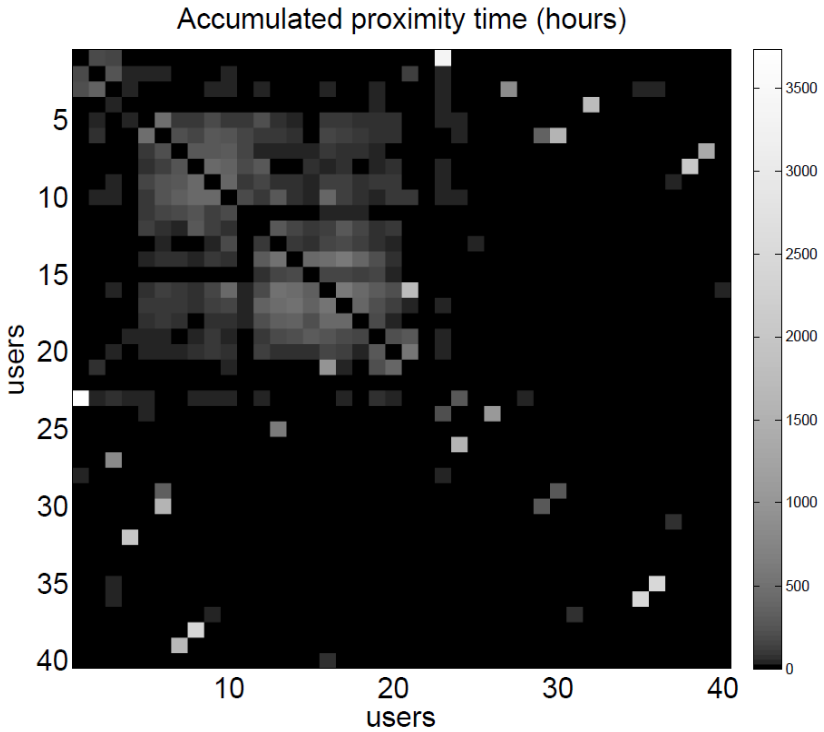
Initialization:
- Draw distribution $\theta_s \sim Dirichlet(\boldsymbol{\alpha})$ for each slice s.
- Draw distribution $\phi_t \sim Dirichlet(\boldsymbol{\beta})$ for each interaction type t.

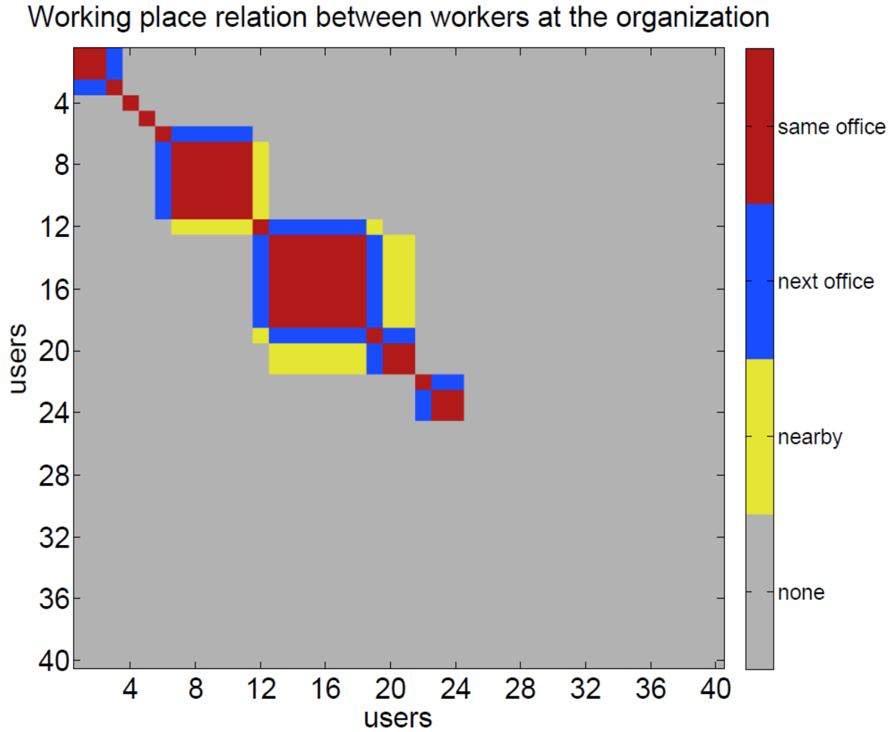
For each link of the slice s:
- Draw an interaction type $t|s \sim Multinomial(\theta_s)$.
- Draw a first person $u|t \sim Multinomial(\phi_{1t})$.
- Draw a second person $v|t \sim Multinomial(\phi_{2t})$.
- Draw a temporal context $c|t \sim Multinomial(\phi_{3t})$.

Table 1. Generative process.

- interaction types are characterized by
  - prominent people participating in the interaction (Φ1,Φ2)
  - temporal context (Φ3)
- learning with MCMC (Gibbs sampling)

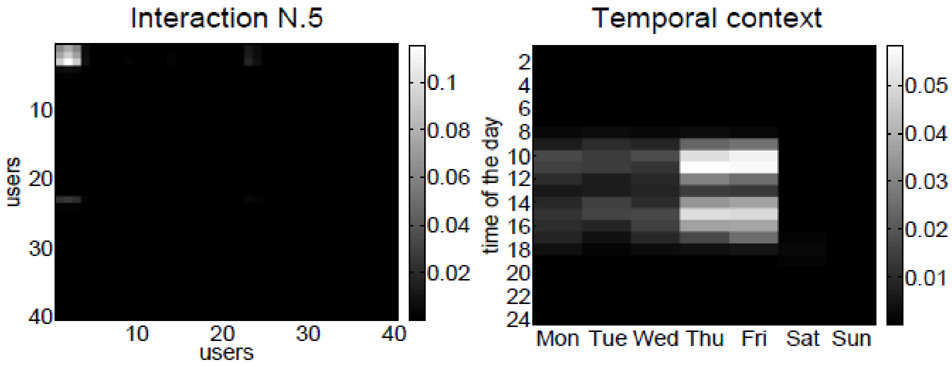
Bluetooth data in Lausanne data



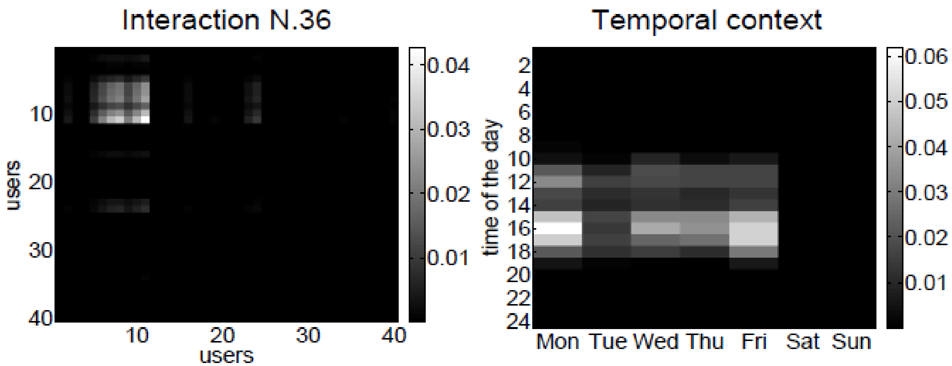


- 40 people with social connections (24 of them are co-workers)
- 1 year of real life
- BT scans every 1-3 minutes depending on the state of the client
- 2 million non-empty Bluetooth scans

discovered interaction types in Lausanne data (1)



(a) Small group



(b) Middle-size group

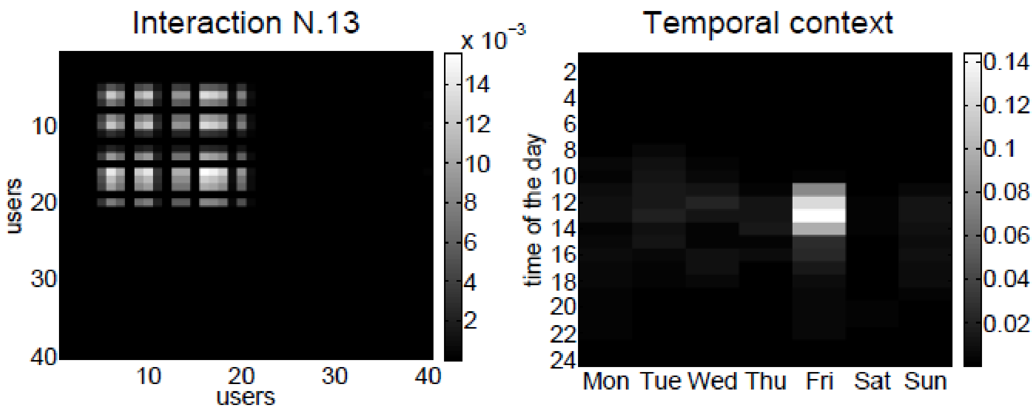
visualizing interaction types

- prominent participants
- temporal context in weekly calendar

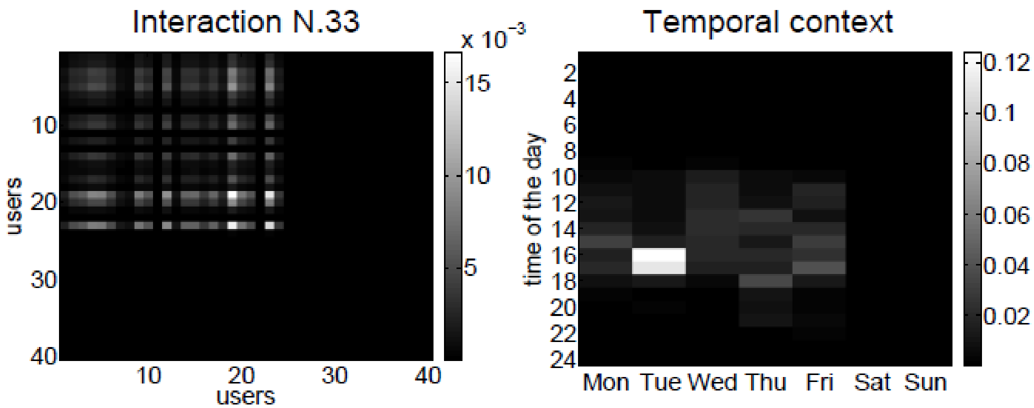
- interaction type 5
  - small group of 3 people who are together on Thu and Fri during office hours

- interaction type 36
  - co-workers in the same building region are in proximity most days

discovered interaction types (2)



(c) Weekly group meeting



(d) Big meeting of the organization

part 3:
smartphone usage in the wild

(joint work with Trinh-Minh-Tri Do and Jan Blom)

T. Do, J. Blom and D. Gatica-Perez, "Smartphone Usage in the Wild: A Large-Scale Study of Applications and Context," ***in Proc. Int. Conf. on Multimodal Interaction (ICMI)*****, Alicante, Nov. 2011.**

mining application logs given contextual cues

- **app logs**
  - all applications: system, pre-installed, user-downloaded apps
- **location**
  - automatic place discovery integrating multiple sensor data (GPS, GSM, WiFi, accelerometer)
- **bluetooth**
  - number of nearby BT devices used as a proxy for social context

how do people use their phones?

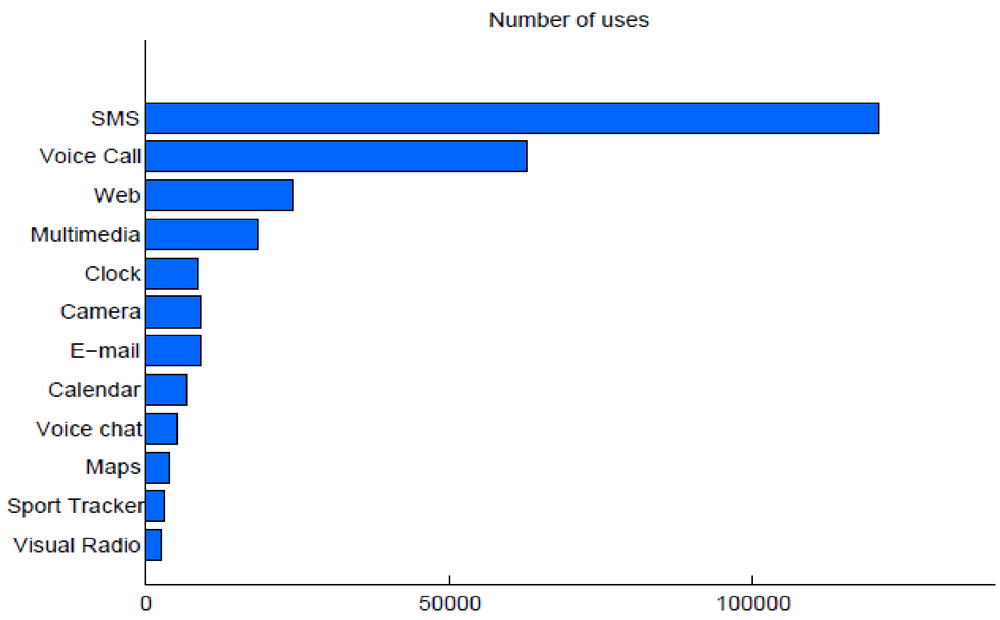


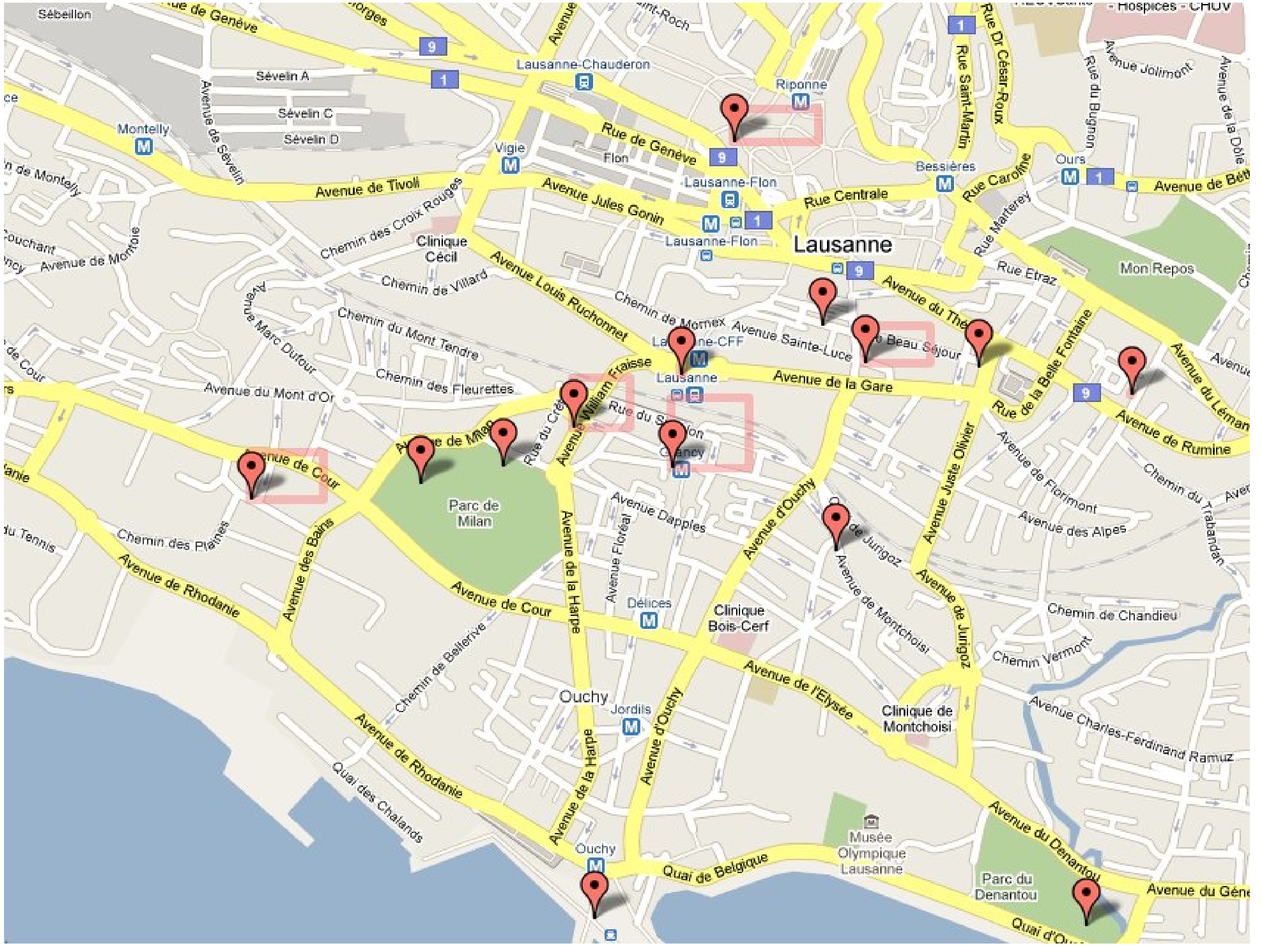
Figure 2. Number of events for each application.

77 users, 9 months

extracting places of interest from phone data

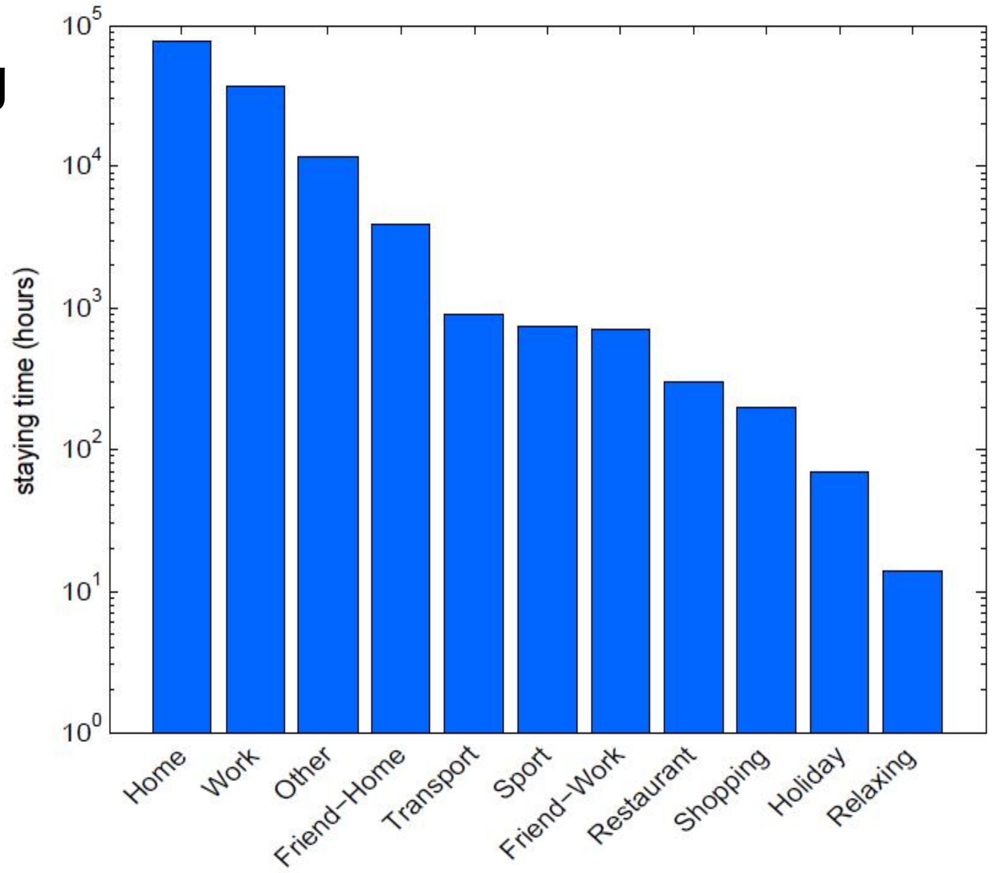
R. Montoliu and D. Gatica-Perez, "Discovering Human Places of Interest from Multimodal Mobile Phone Data,"*in Proc. ACM Int. Conf. on Mobile and Ubiquitous Multimedia (MUM)*, Limassol, Dec. 2010.

my own private Lausanne

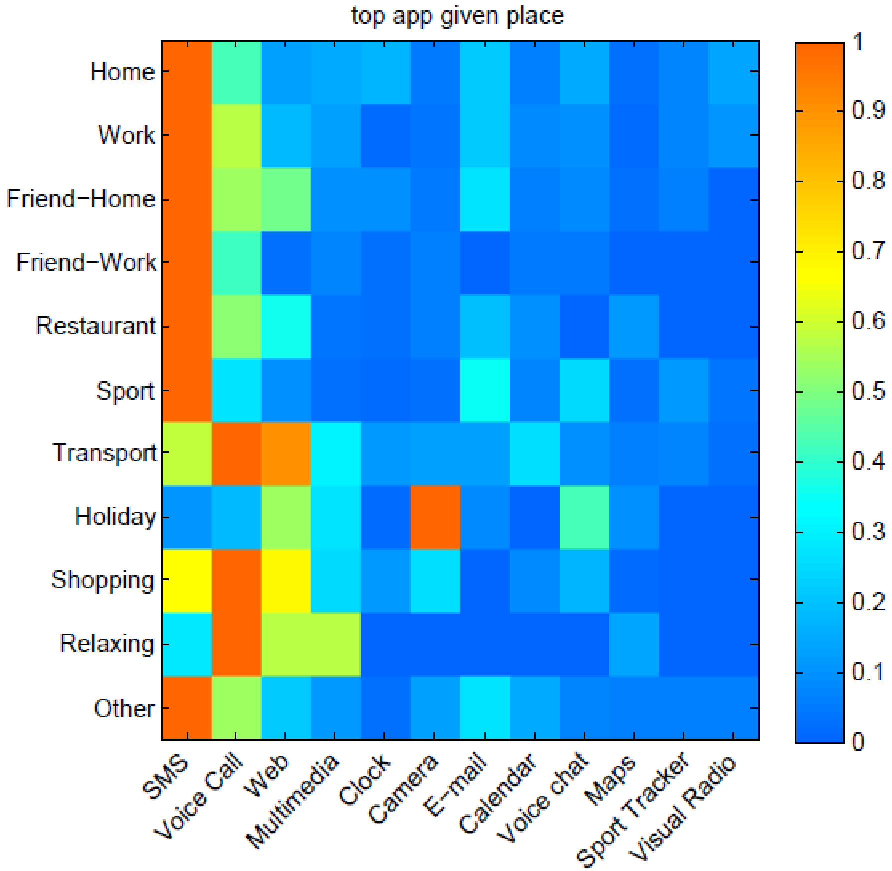


from geographic locations to semantic places

- personal place online tagging
  - users tag 5 most visited places and 3 randomly selected places
  - predefined 22-tag set
  - 616 annotated places



application usage conditioned on places



- SMS is highly used in many indoor locations

- Voice is highly used in moving contexts (e.g. waiting at the train station, shopping, relaxing outdoors)

- on holidays, people have a preference for using camera

part 4:
personality and smartphone data

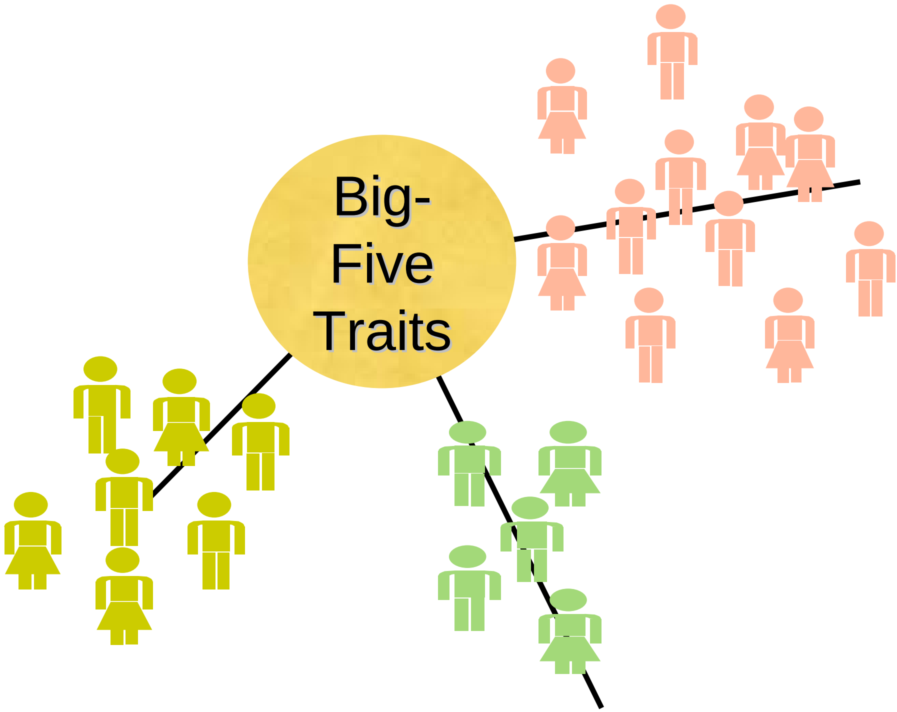
(joint work with Gokul Chittaranjan and Jan Blom)

G. Chittaranjan, J. Blom and D. Gatica-Perez, "Who's Who with Big-Five: Analyzing and Classifying Personality Traits with Smartphones," ***in Proc. IEEE Int. Symp. on Wearable Computers (ISWC)*****, San Francisco, Jun. 2011.**

personality traits

“the Big-Five traits have been broadly accepted as a way of presenting all the major traits of a person at the highest level of abstraction”

Gosling et al., 2003



(N)euroticism/ Emotional Stability
(E)xtraversion
(O)penness to Experience
(A)greeableness
(C)onscientiousness

McCrae and John, 1992

“...since **mobile phones** also **mediate social interactions**, phone usage could reflect an individual’s personality...”

S. Butt and J.G. Phillips, 2008

understanding the Big-Five

Extravert

Active
Assertive
Energetic
Enthusiastic
Outgoing
Talkative

Agreeable

Appreciative
Forgiving
Generous
Kind
Sympathetic

Conscientious

Efficient
Organized
Planful
Reliable
Responsible
Thorough

Neurotic

Anxious
Self-pitying
Tense
Touchy
Unstable
Worrying

Open

Artistic
Curious
Imaginative
Insightful
Original
Wide Interests

the study

83 subjects (53 Male)
8 months of everyday phone data
Age range 19 - 63 years (μ = 29.7, σ= 7.6)
All were previous mobile phone users
Most didn't own smartphones before

TIPI Questionnaire (Gosling et al., 2003)

1. _____ Extraverted, enthusiastic.
2. _____ Critical, quarrelsome.
3. _____ Dependable, self-disciplined.
4. _____ Anxious, easily upset.
5. _____ Open to new experiences, complex.
6. _____ Reserved, quiet.
7. _____ Sympathetic, warm.
8. _____ Disorganized, careless.
9. _____ Calm, emotionally stable.
10. _____ Conventional, uncreative.

Correlation analysis and classification

features extracted and aggregated for user-months from anonymous logs

Apps

Counts

Office
Internet
Video/Audio
Maps
Mail
Youtube
Calendar
Camera
Chat
SMS
Games

Bluetooth

Unique (IDs)
Duration (IDs)
Duration
(Most common ID)

Phone

Calls

duration I/O
counts I/O
known contacts I/O
missed counts

SMS

counts
sent
received

I : Incoming
O : Outgoing

SMS

Word length
Message count
(inbox, sent)

correlation analysis: extraversion

Extraversion	
Feature	Correlation
Uses of Internet	-0.26
Total duration of Incoming Calls	0.20
Avg. duration of Incoming Calls	0.18
Uses of Camera	-0.15
SMS Word Length (Sent)	-0.15
Calls Received	0.13
SMS Sent	-0.13

All correlations are significant (p<0.01)

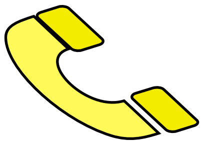
comparison of results to previous work
(Butt & Phillips, 2008, self-reported phone use)

similar results:

\+ **extraverts** are more likely…

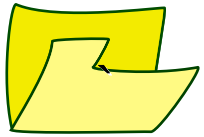
to receive calls

to speak longer in incoming calls



\+ **introverts** are more likely…

to use the web



\+ outgoing calls do not describe personality significantly

different result:

\+ **agreeable** people receive more incoming calls

\+ differences in self-perception of phone use?

user-month classification

Population split across median to obtain a 2-class problem.
SVMs with RBF and C4.5 decision trees used.
% Accuracy across all folds with Leave-one-out training reported.

Trait	Performance (%)		
	C4.5	SVM	Baseline*
Extraversion	61	75	59
Agreeableness	69	70	58
Conscientiousness	68	74	62
Emotional Stability	60	72	52
Openness to Experience	66	69	59

* Choosing majority class all the time.

concluding remarks

smartphones and human modeling

- phones allow large-scale studies
- rich, massive, real-life data
- our work: from places to personality

open problems

- sensing at larger scales
- enrich pattern description, from low-level features to social constructs

acknowledgments

@ Idiap

Trinh-Minh-Tri Do
Gokul Chittaranjan
Raul Montoliu
Olivier Bornet

@ Nokia Research

Niko Kiukkonen
Juha Laurila
Jan Blom
Olivier Dousse