

social media, crowdsourcing, and human behavior

lecture 2: mobile social data

daniel gatica-perez

01.09.2015

A photograph of two young women with long, dark hair standing outdoors. The woman on the left is wearing a black cable-knit sweater and a blue and white plaid scarf, and is holding a smartphone with both hands, looking at the screen with a slight smile. The woman on the right is wearing a light green top and a white scarf with a floral pattern, and is also holding a smartphone, looking down at it. The background is a blurred outdoor setting with warm, golden light. In the bottom left corner, there is a yellow text credit.

**credit: garryknight
@flickr (cc)**

phones: the ultimate sensor

- GPS
- Wifi
- Bluetooth
- Accelerometer
- Phone call logs
- SMS logs
- Web
- Audio
- Photos
- Video
- More and more apps...

individual cues

- + location
- + motion

physical daily life

relational cues

- + proximity
- + communication

social daily life

online cues

- + internet
- + apps

online daily life

one world: physical and online



foursquare 1

12:07

PEOPLE TALK ABOUT "PORK DUMPLINGS" IN CHINATOWN

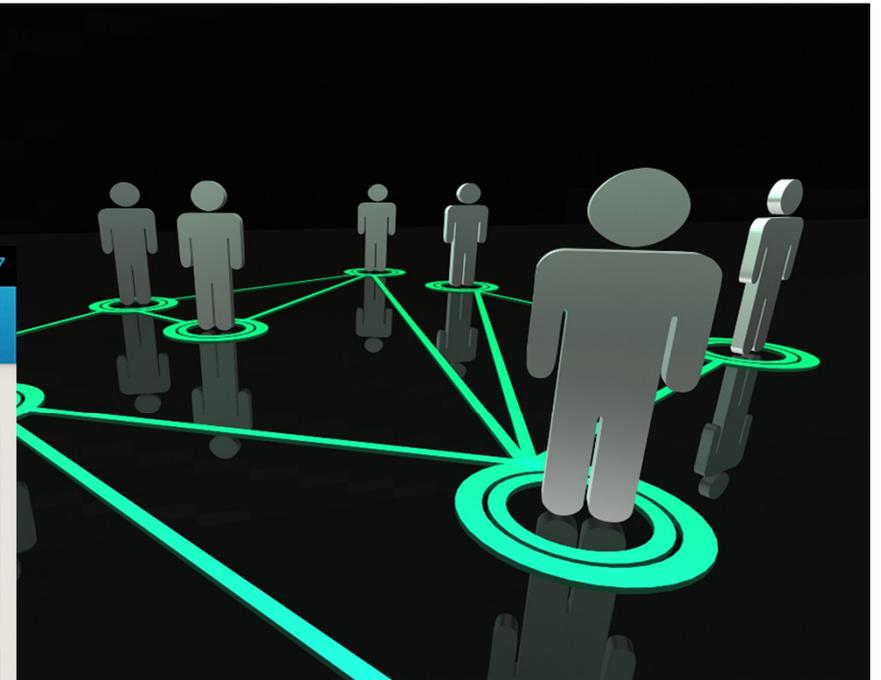
Vanessa's Dumpling House
DUMPLINGS
0.9mi | \$\$\$\$ | 2

This tiny little venue has tiny little prices for their delightful dumplings. The sesame pancakes are also great, and perfect to grab to-go on your way to picnic in the park, just a block west.

epi Epicurious

8 friends have
+3

VIEW MAP



foursquare (May 2015)



launched in 2009

predecessor: dodgeball (sold to google in 2005)

2014: split into two apps: foursquare (local search)
& swarm (location and social network)

55 million people

70 million tips

7 billion check-ins

2 million businesses

65 million places

170 employees

<https://foursquare.com/about>



Google
latitude

2009-2013



Places

Who. What. When. And now **Where.**

facebook

Share Where You Are



"Best. Concert. Ever."

Connect With Friends Nearby



"I'm just down the street!"

New: Find Local Deals



"I'm getting \$20 off new jeans."

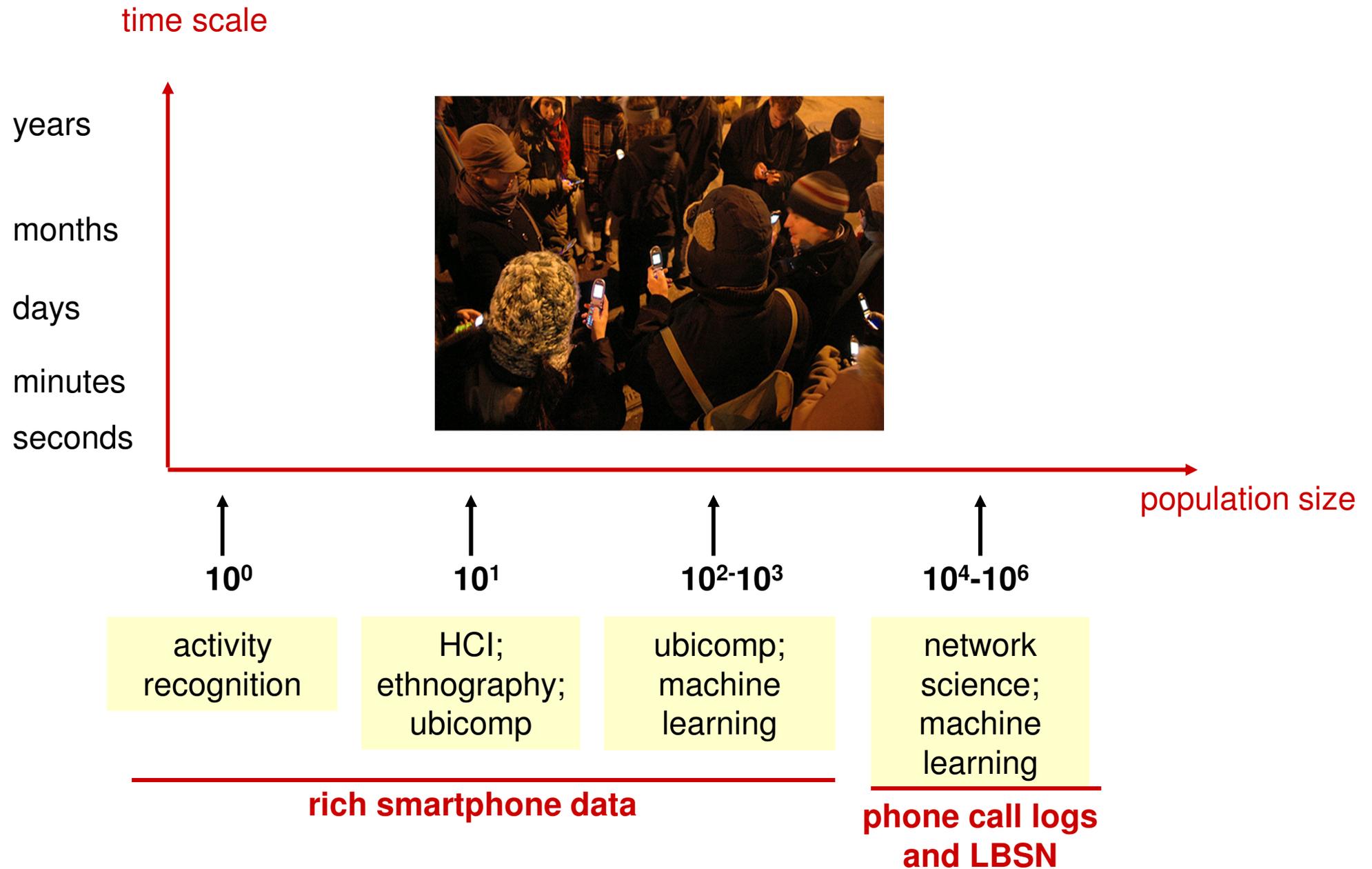
2010-2011,
revived 2014



Real People. Real Reviews.

2005-
check-ins: 2010

who is researching this?





1. human mobility

manual and inferred
check-ins to understand
collective mobility

2. human geography

language use in LBSNs

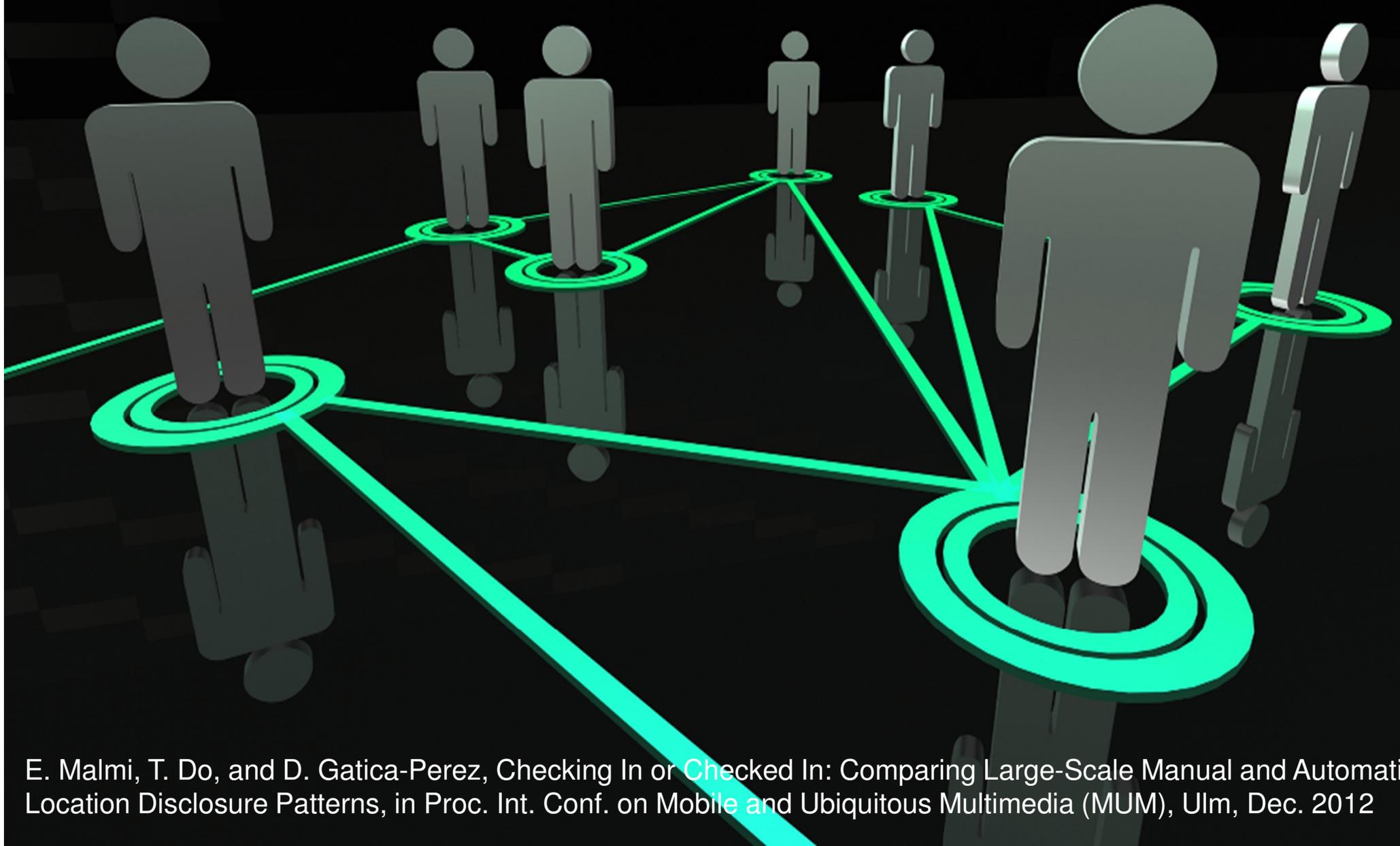
3. personality

big-five and phone data

1.

mobility analysis

are check-ins a good proxy to understand large-scale mobility?



E. Malmi, T. Do, and D. Gatica-Perez, Checking In or Checked In: Comparing Large-Scale Manual and Automatic Location Disclosure Patterns, in Proc. Int. Conf. on Mobile and Ubiquitous Multimedia (MUM), Ulm, Dec. 2012

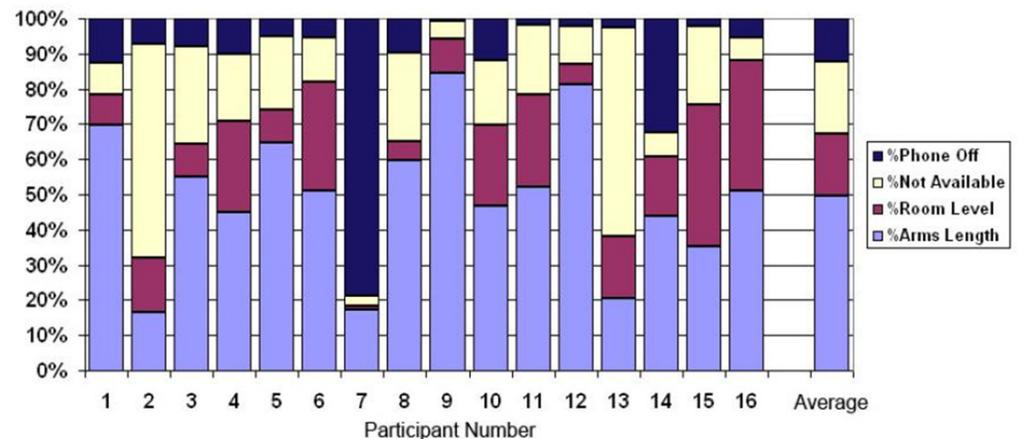
previous work



(Gonzalez, Nature 2008)
cell phone records (CDRs)
100 000 users, 6 months
0.91 call/sms per day

(Patel, Ubicomp 2006)
Bluetooth connectivity
only 70% of time user &
phone are in same room

(Noulas, PLoS ONE 2012)
foursquare data
925 000 users, 6 months
0.21 check-ins per day



A photograph of a crowd of people at night, many holding up their smartphones. The scene is dimly lit, with the primary light source being the screens of the phones. The people are dressed in winter clothing, including jackets and hats. The overall atmosphere is one of a public gathering or event.

inferred check-ins from smartphone sensors: Lausanne Data Collection Campaign (LDCC)

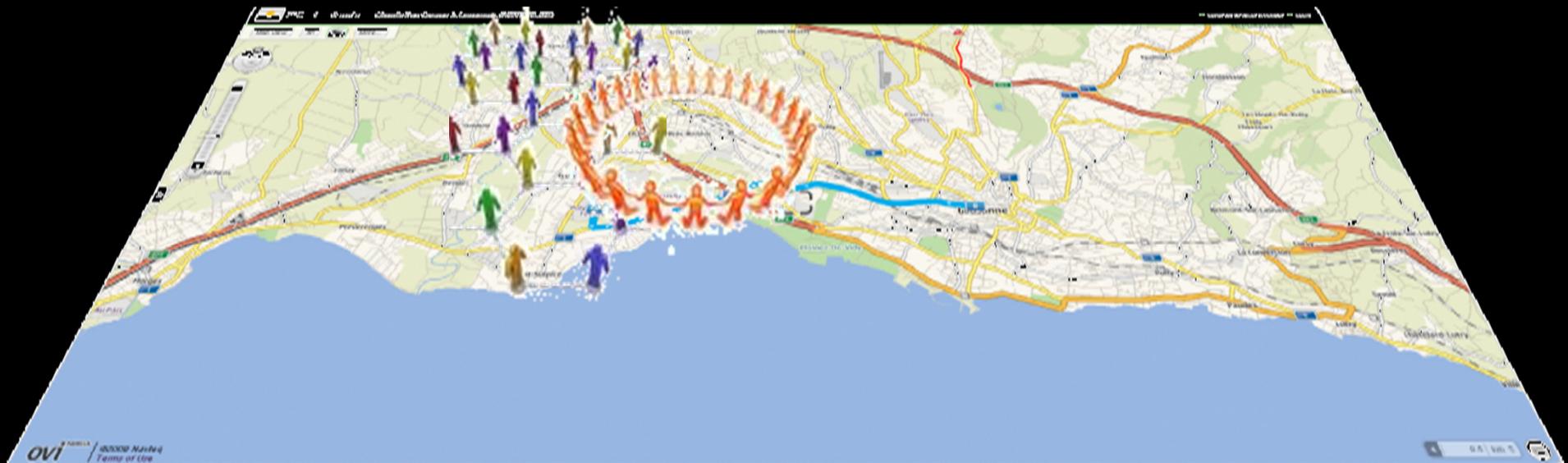
180 volunteers, 18 months, French-speaking Switzerland

GPS, bluetooth, accelerometer, media, apps

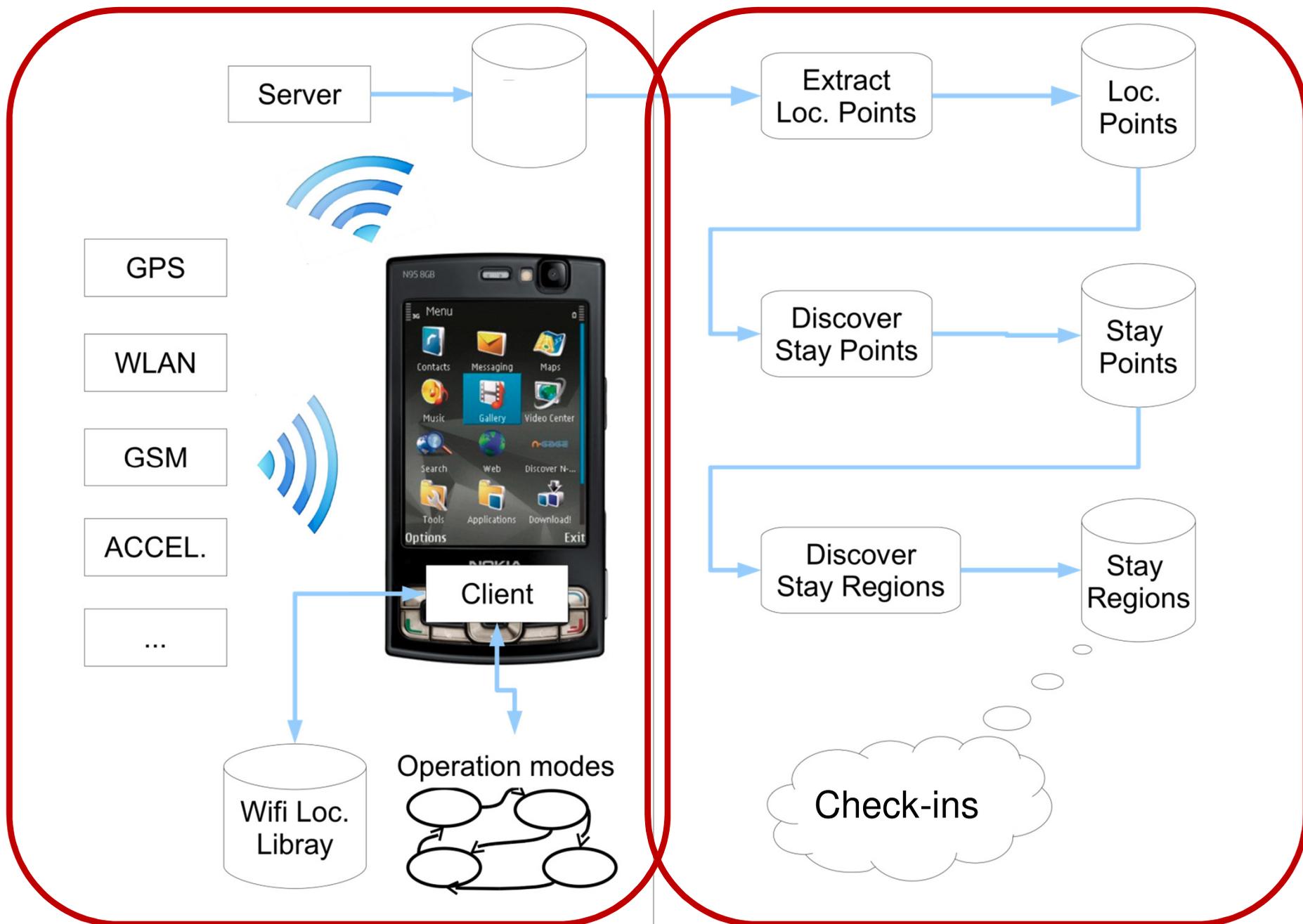
credit: [icopythat@flickr](#)

a few numbers

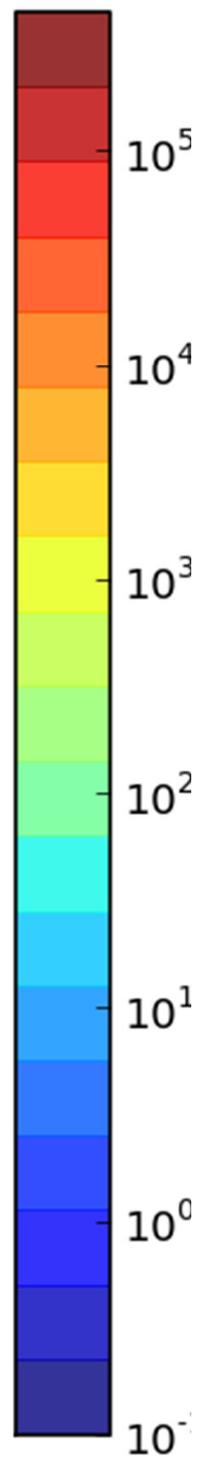
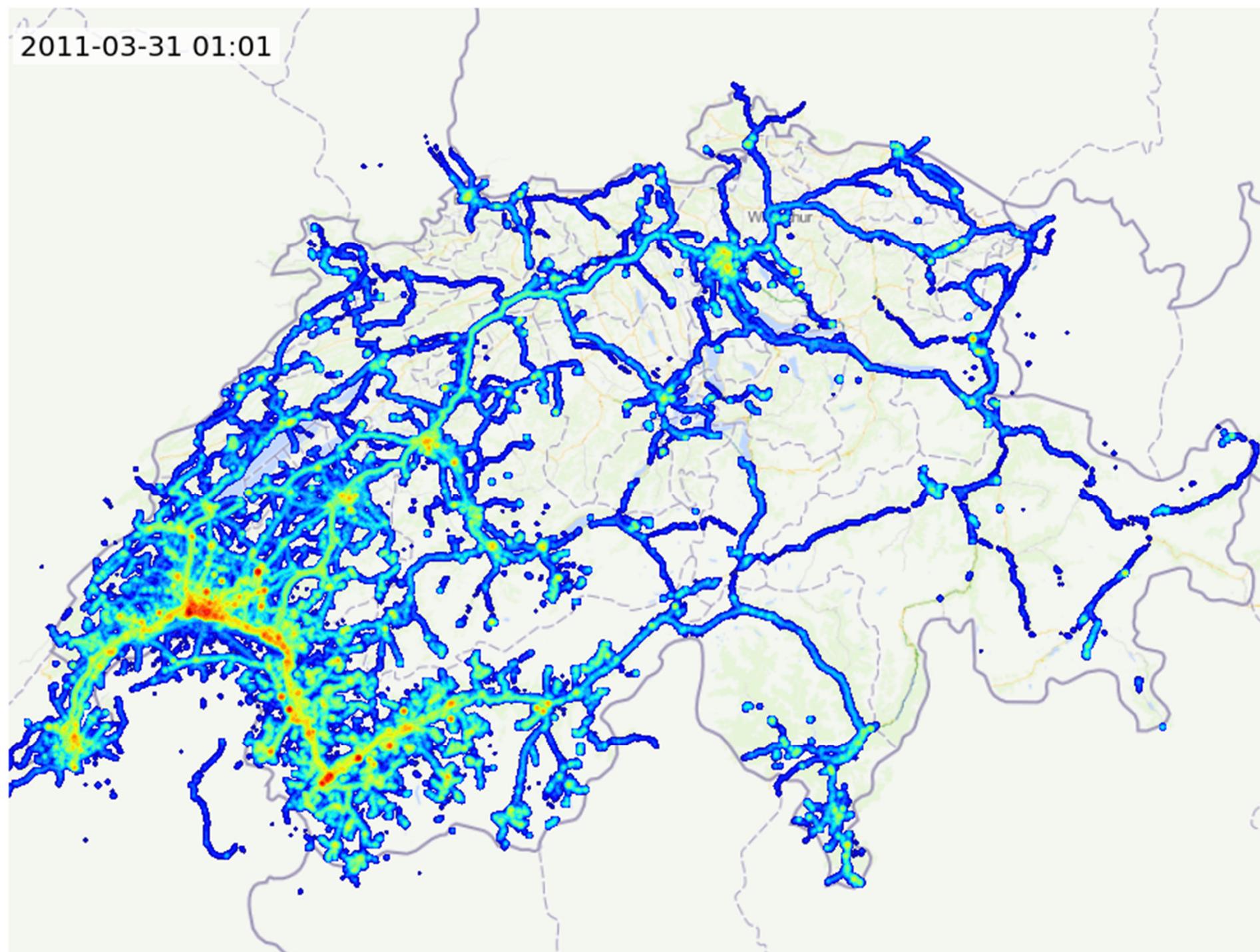
- 26 M location points
- 38 M bluetooth readings
- 31 M Wifi observations
- 1 M accelerometer samples
- 240 K phone call events
- 175 K sms events
- 40 K photos/videos



inferring check-ins from smartphone sensors



2011-03-31 01:01



manual check-ins from Foursquare

foursquare I'm looking for... Current Map View LOG IN SIGN UP

Suggestions for **Top Picks** TOP PICKS

Show me places ...

- I haven't been to
- My friends have been to
- I have been to before
- With Foursquare specials

Place Saint-François
1 **8.5** Place Saint-François
Plaza / Square

During weekend there are some fruit markets - Gabriel G.

This spot is popular on Foursquare Save Like

Visual Players
2 **6.9** 49 avenue du Grey
Design

best design and quality ! - Aegerter G.

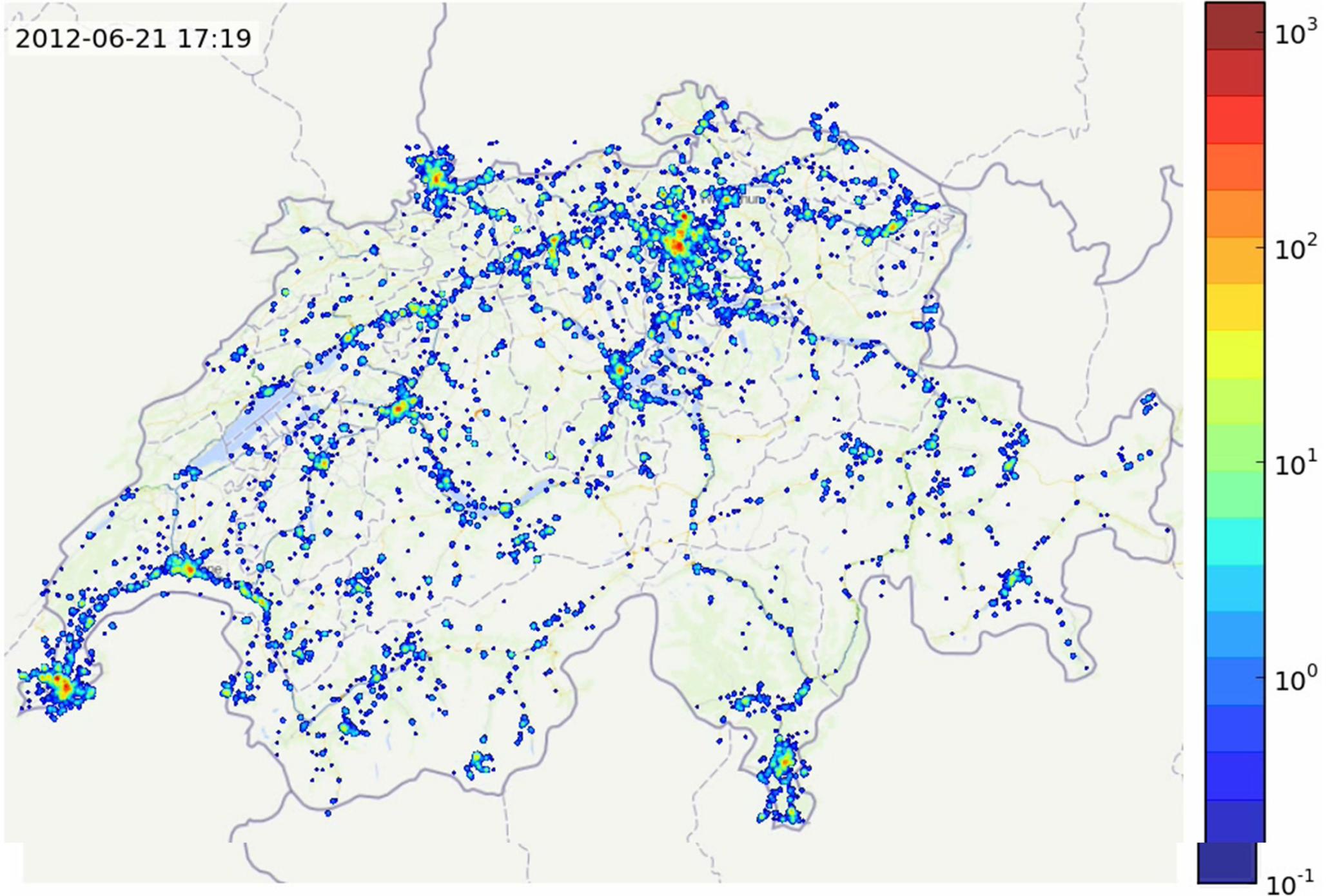
There's a special here! Save Like

Holy Cow!
3 **8.6** Rue des Terreaux 10
Burgers

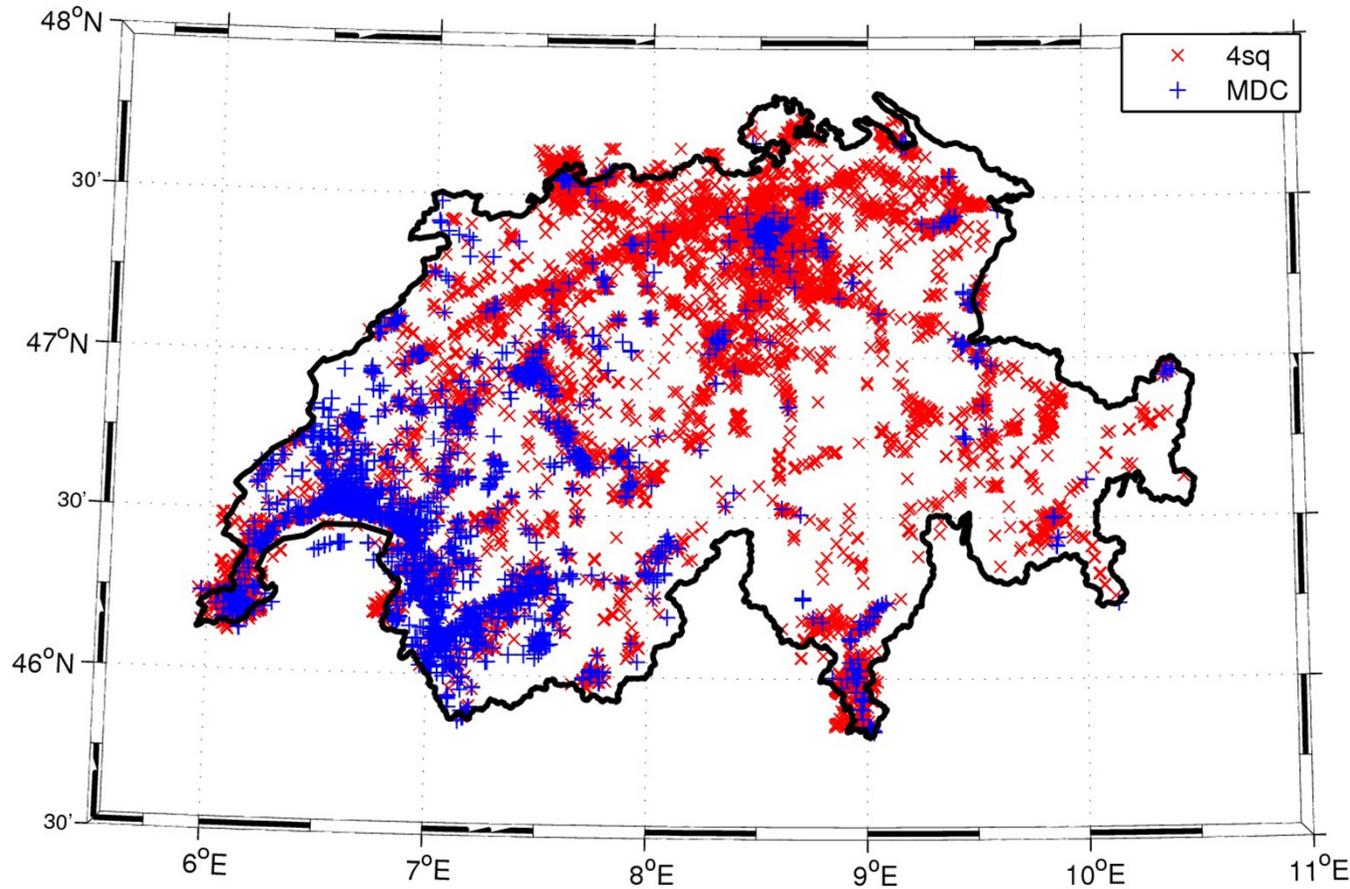
Holy Cow!, à Lausanne, c'est déjà une institution. Le Temple du Bon Burger. Ce resto est le second d'une

credit: foursquare.com

4sq in Switzerland (6 months)



data sets summary

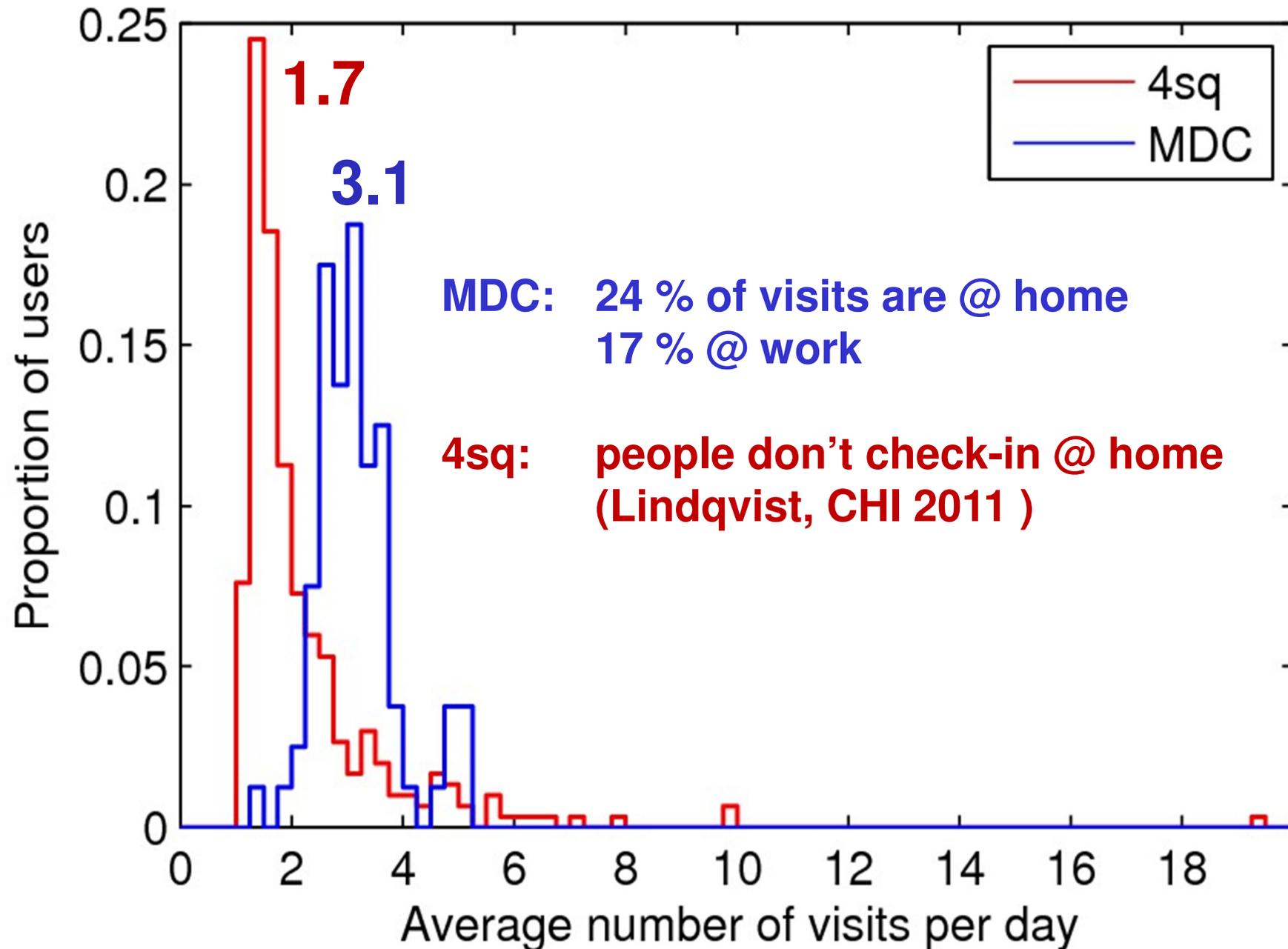


80 active users
51,600 check-ins
7,200 places

300 active users
40,600 check-ins
17,400 places

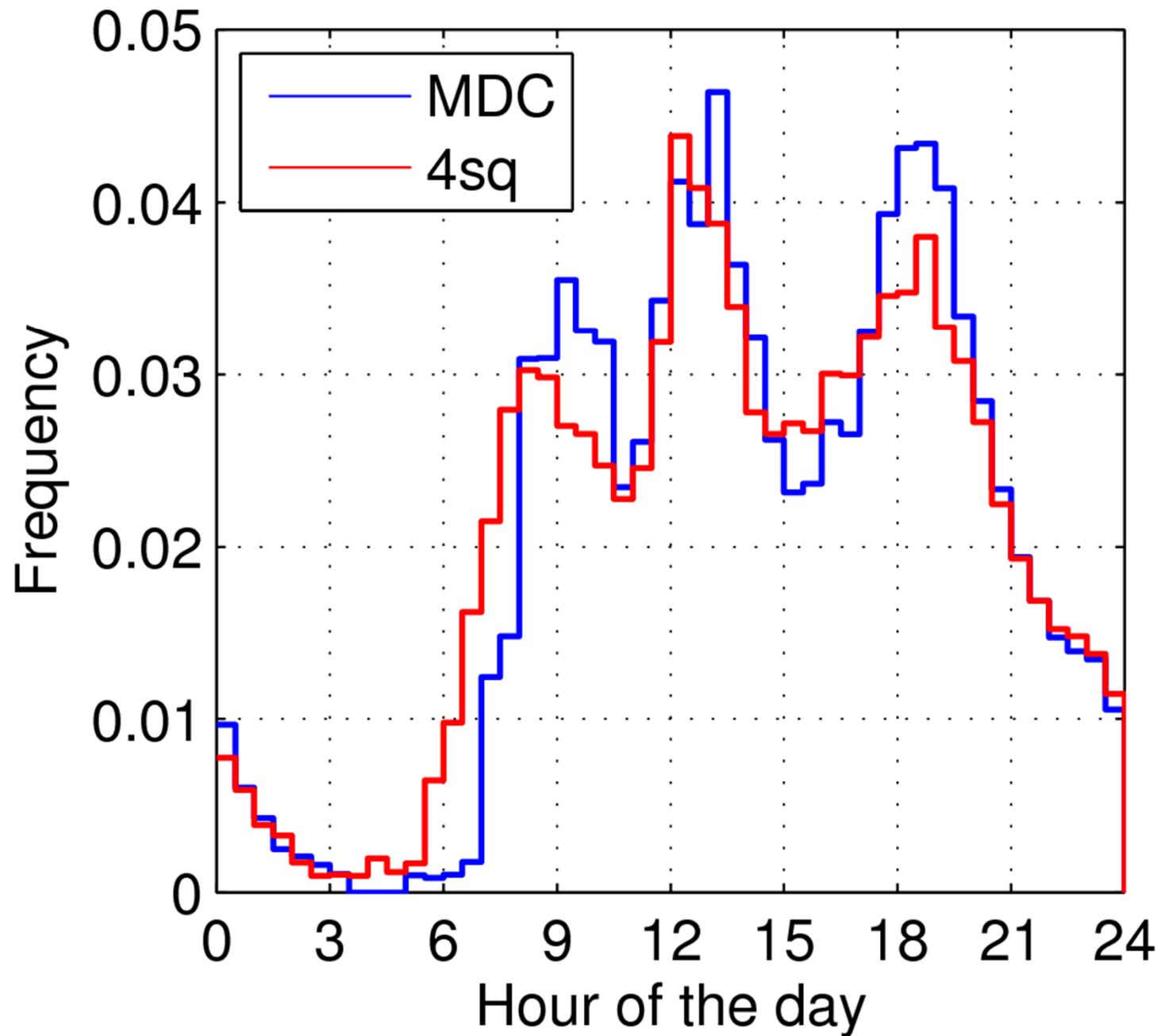
results (1)

daily check-in distributions



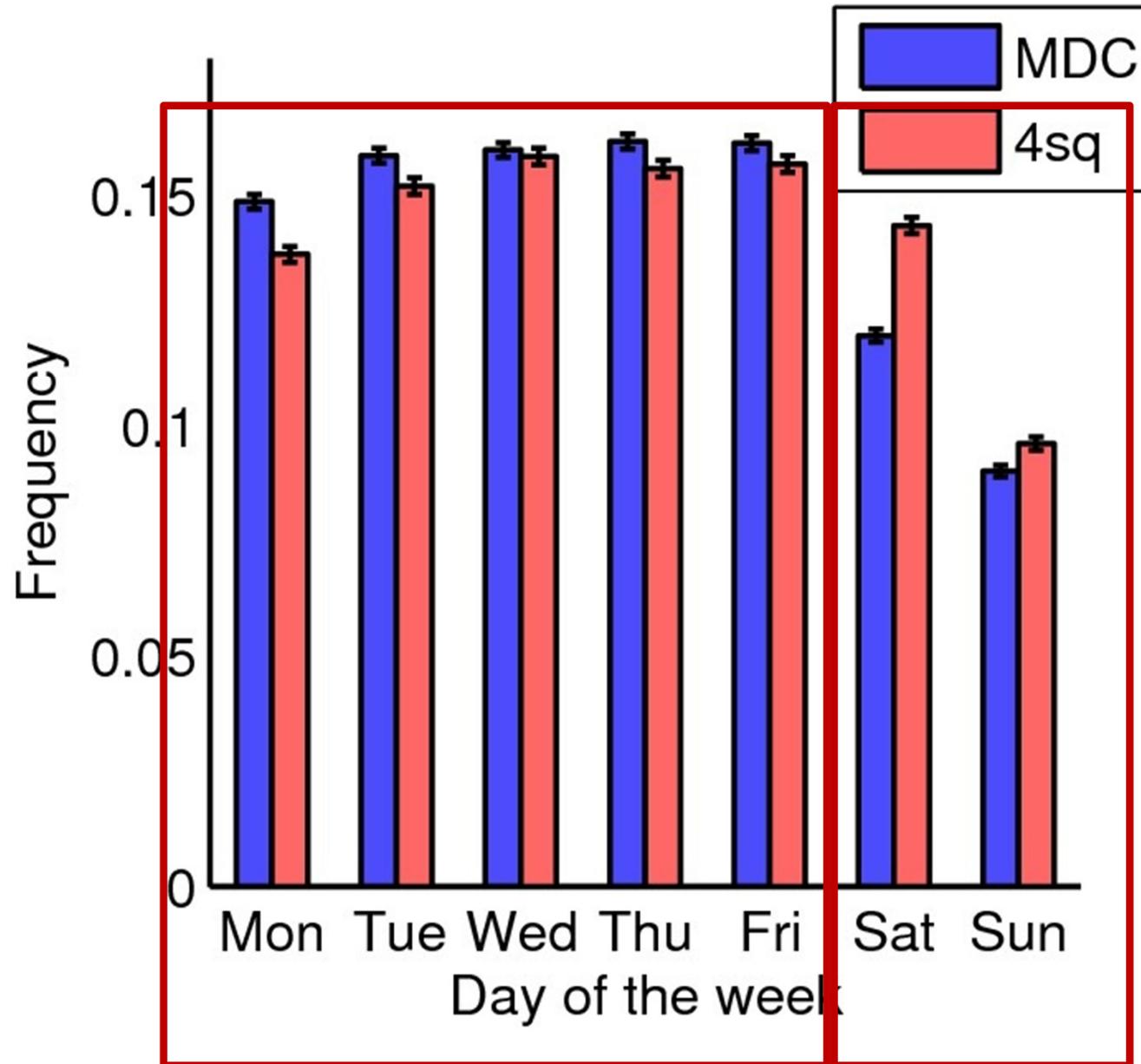
results (2)

the rhythm of daily activity

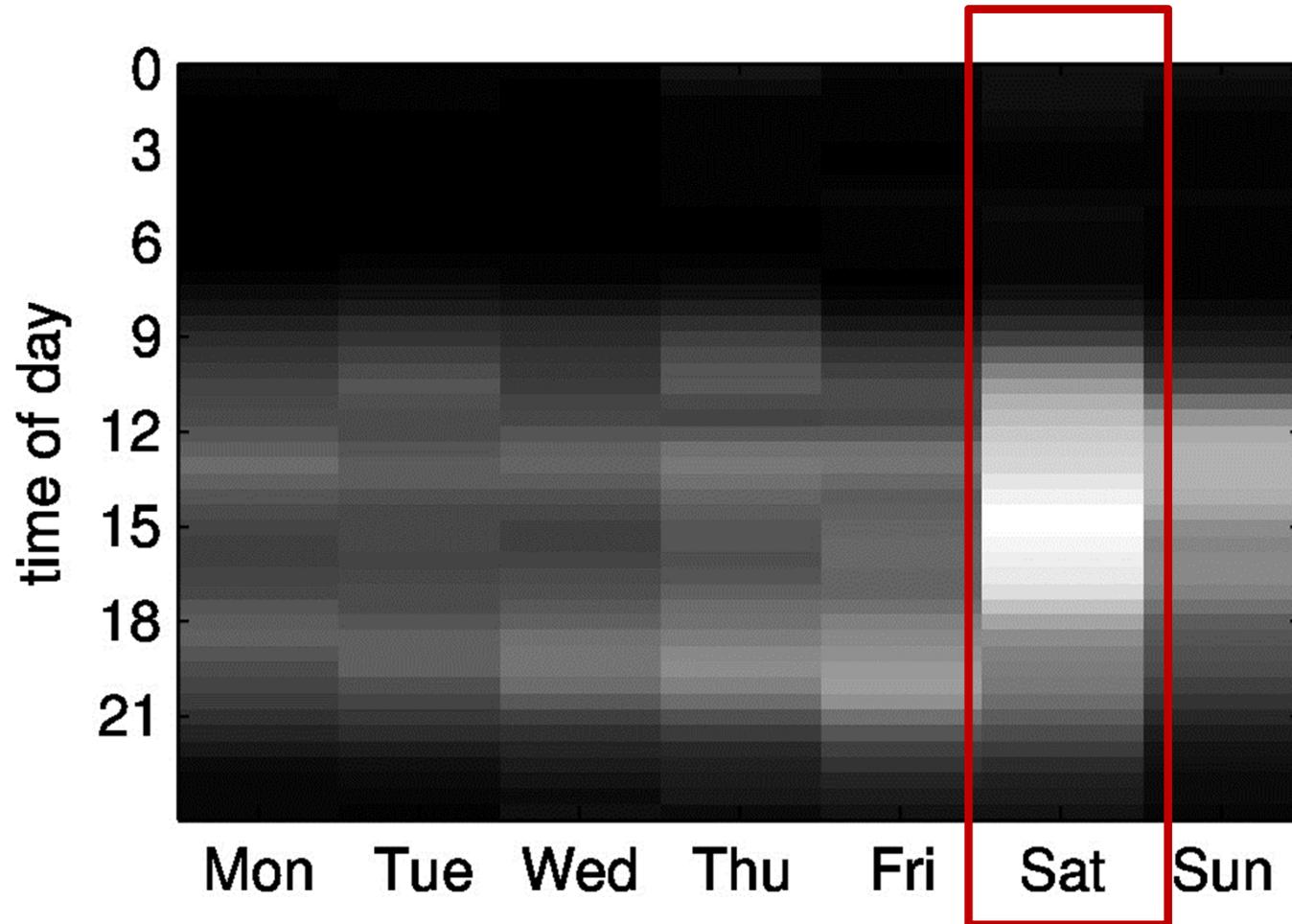


results (3)

weekly check-in time distributions



the pursuit of new places



the biases of mobile social data

socio-economic bias

not everybody is represented

activity bias

not everything is represented

spatial bias

not every place is represented

credit: Devin Smith @ flickr (cc)

2. LBSNs and human geography

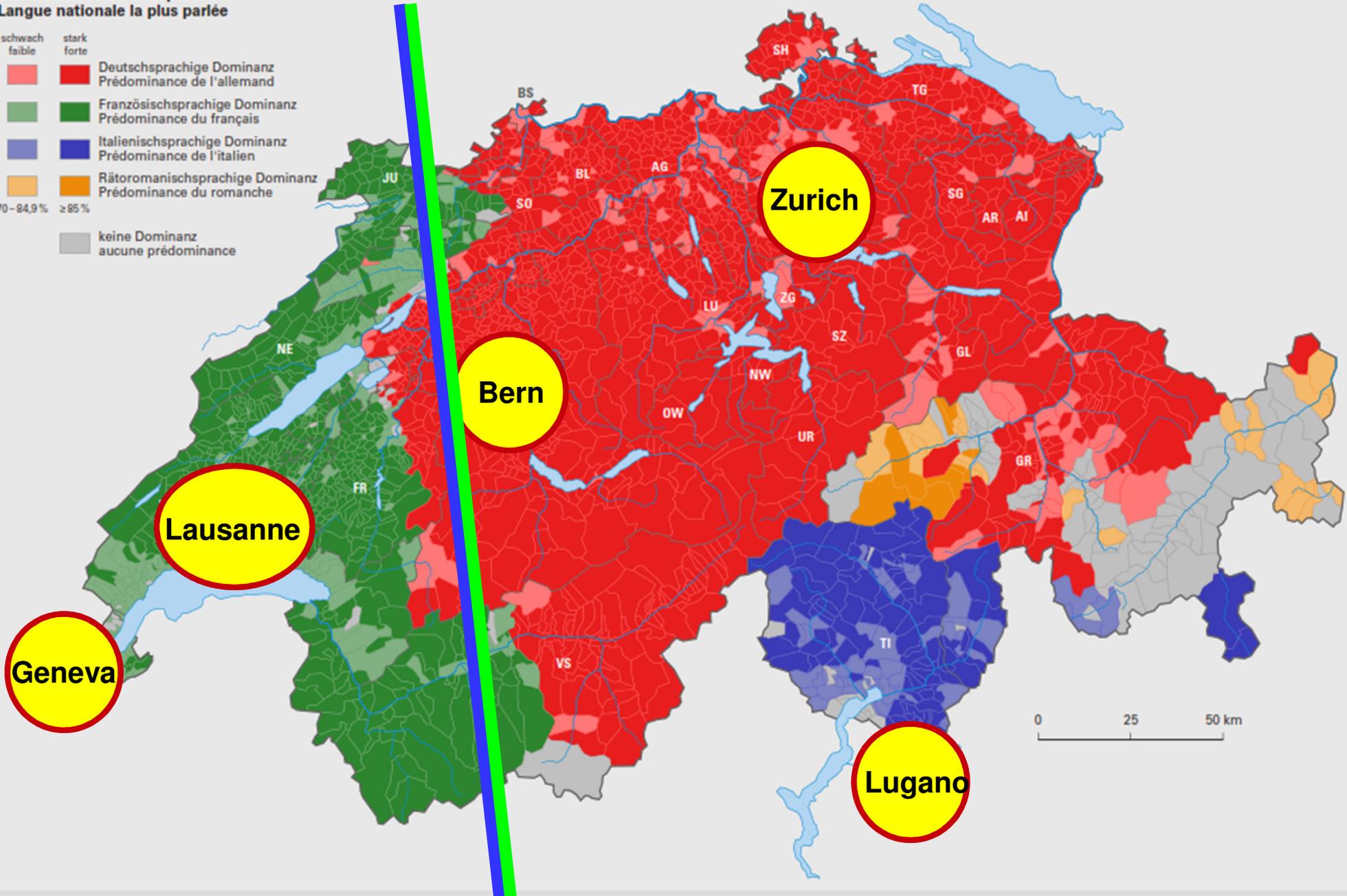
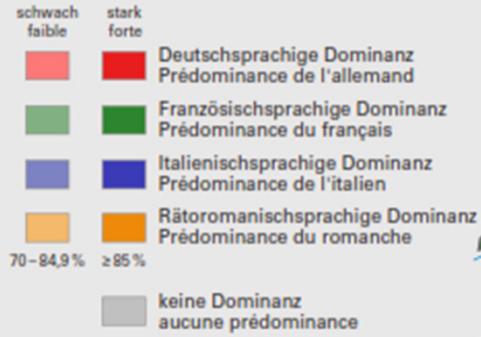
language use in foursquare



D. Santani and D. Gatica-Perez, Speaking Swiss: Languages and Venues in Foursquare, in Proc. ACM Int. Conf. on Multimedia, Barcelona, Oct. 2013

credit: gato-gato-gato
@flickr (cc)

Dominierende Landessprache
Langue nationale la plus parlée





COMMON LANGUAGE

English challenges multilingual Switzerland

Culture

Reuse article

↑

LINGUA FRANCA
Should Switzerland
preserve its
multilingualism?

- ! Language figures
- ! Language Rich Europe
- ! Swiss initiative

63 comments

Readers Recommend

by Sonia Fenazzi, swissinfo.ch

June 1, 2012 - 11:00

The language debate is reaching new heights in Switzerland with French and Italian speakers uneasy about the progressive abandonment of their languages by the German-speaking community in favour of English.

German-speaking cantonal governments have been responding to what they say is pressure from the globalised economy by gradually giving English priority over the national languages at school; English - not French - is taught as a second language in 14 of the 17 cantons.



Switzerland possesses an enviable range of language resources (Keystone)

source:

http://www.swissinfo.ch/eng/culture/English_challenges_multilingual_Switzerland_.html?cid=32806320

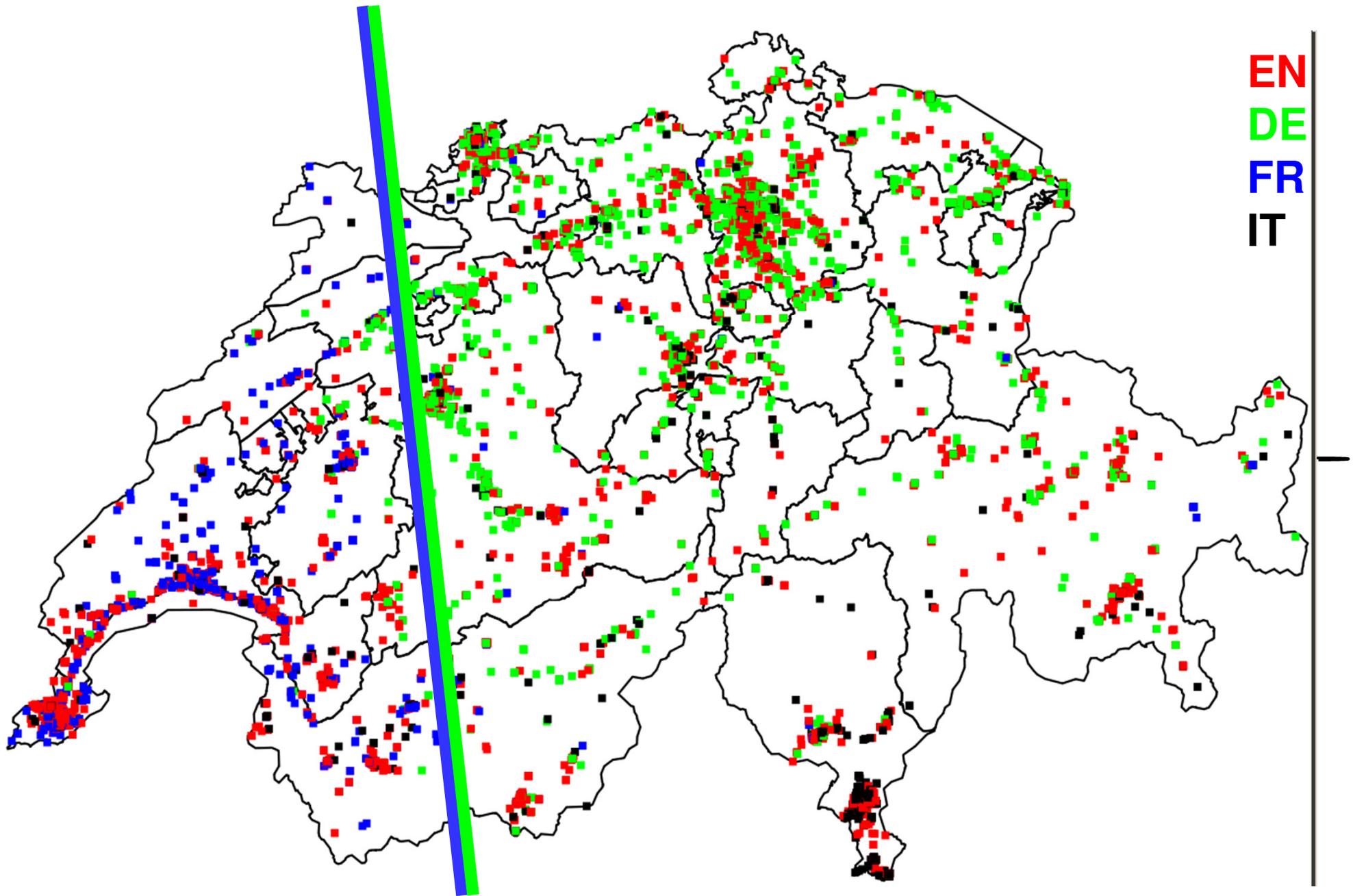
swiss 4sq comments dataset

21,780 comments – 8,748 users – 53 languages

Language Used	% of Total Comments	% of Users
English (EN)	58.22	61.28
German (DE)	21.91	20.52
French (FR)	10.68	9.28
Italian (IT)	3.41	3.64

English is dominant

*national languages
follow rankings*



spatial distribution of languages used for 4sq comments

multilingualism through 4sq language use

Languages Used	1	2	3	4
% of Users	81.0%	14.77%	3.05%	0.83%
% of Places	72.0%	21.71%	5.00%	0.98%
Language Pairs	EN-DE	EN-FR	EN-IT	DE-FR
% of Bilingual Users	42.57%	18.89%	8.36%	1.32%

comparing language proportions in specific cantons

official language →

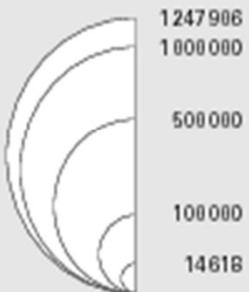
Canton	EN	DE	FR	IT
Zurich (DE)	4215 (63.20%)	1912 (28.67%)	84 (1.26%)	114 (1.71%)
Geneva (FR)	2429 (71.74%)	37 (1.09%)	654 (19.31%)	61 (1.80%)
Bern (DE)	1091 (53.74%)	735 (36.21%)	58 (2.86%)	38 (1.87%)
Vaud (FR)	1117 (49.06%)	27 (1.19%)	947 (41.59%)	44 (1.93%)
Ticino (IT)	451 (50.79%)	54 (6.08%)	18 (2.02%)	319 (35.92%)

*rankings of languages in each
canton match census*

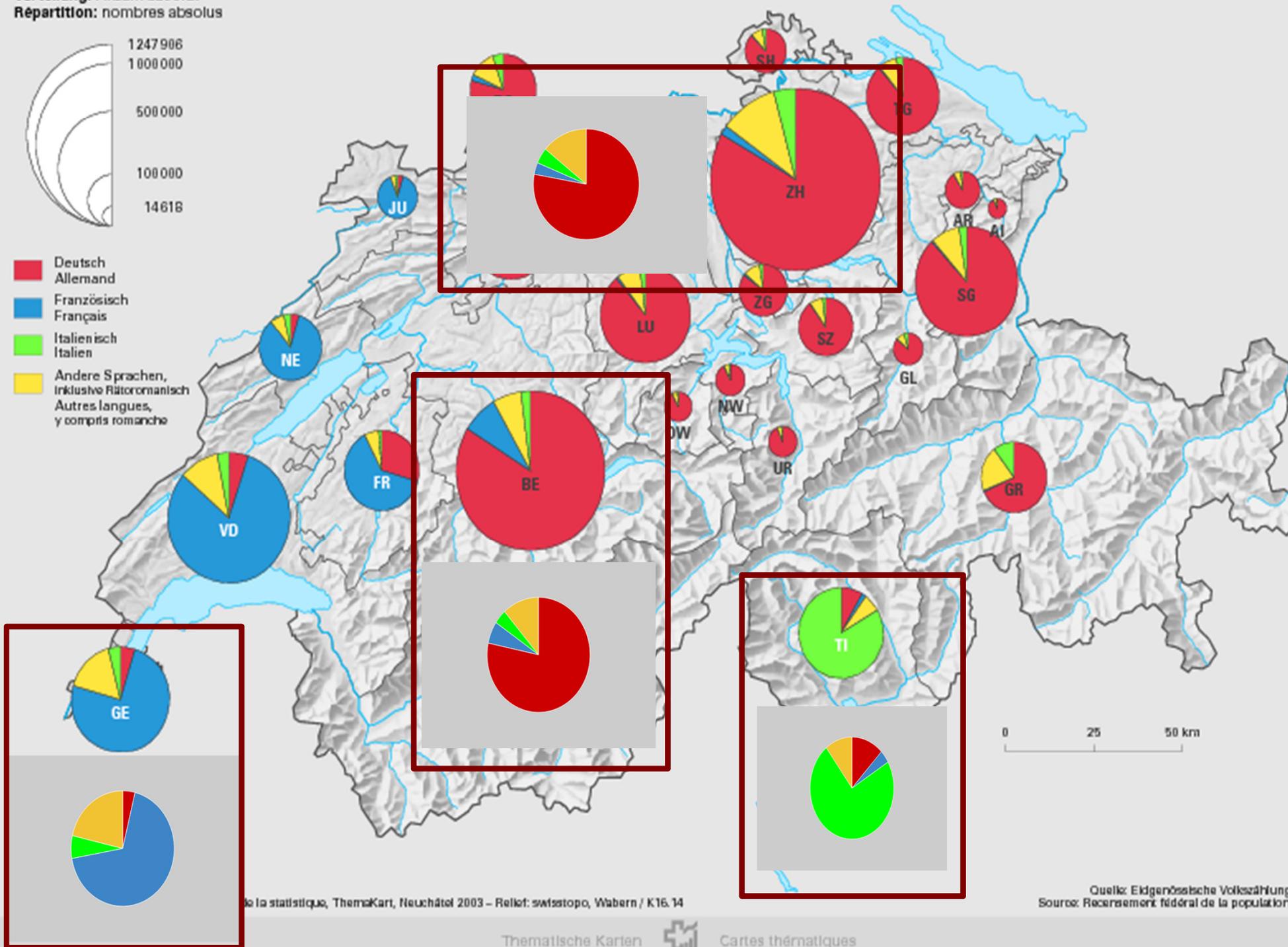
Landessprachen in den Kantonen, 2000
Langues nationales parlées dans les cantons, en 2000

credit: Swiss Federal Statistical Office

Verteilung: Anzahl absolut
 Répartition: nombres absolus



- Deutsch
Allemand
- Französisch
Français
- Italienisch
Italien
- Andere Sprachen,
inklusive Rätoromanisch
Autres langues,
y compris romanche



© 2003, swissstat, Neuchâtel 2003 – Relief: swisstopo, Wabern / K16.14

Quelle: Eidgenössische Volkszählung 2000, BFS
 Source: Recensement fédéral de la population 2000, OFS

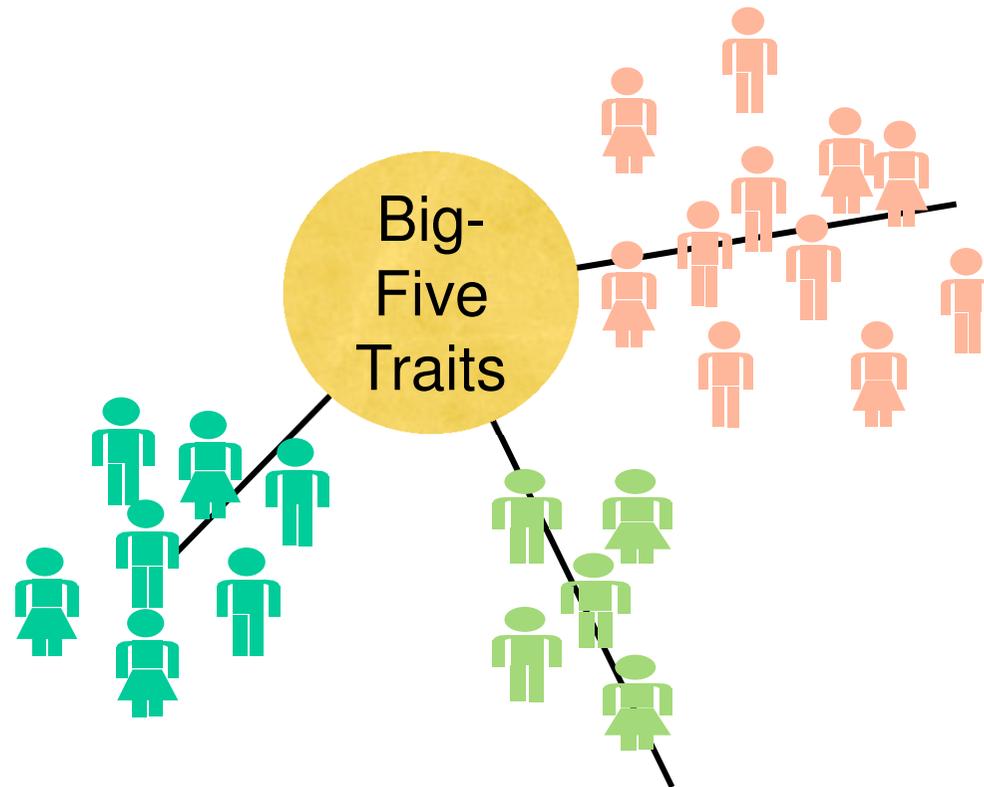
3. personality and smartphone data

G. Chittaranjan, J. Blom and D. Gatica-Perez, "Mining Large-Scale Smartphone Data for Personality Studies," *Personal and Ubiquitous Computing*, published online Dec 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



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

S. Butt and J.G. Phillips, 2008

the study



117 subjects (73 male)
age range: 19 - 63 years ($\mu = 30.2$)
17 months of phone data
all were previous mobile phone users
most did not 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.

tasks:

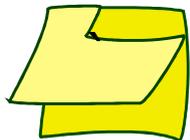
Correlation analysis and classification

features extracted and aggregated for user-months from anonymous logs

Apps

Counts of

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



Bluetooth

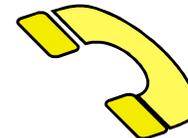
Unique ID counts
Duration



Calls

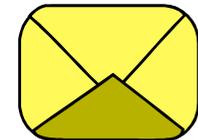
Call duration I/O
Call counts I/O
Missed call
counts
Known contacts
I/O

I : Incoming
O : Outgoing



SMS

Message count
(inbox, sent)
Word length



correlation analysis for extraversion ($p < 0.01$)

	Total duration of Incoming Calls	0.21
	Avg. duration of Incoming Calls	0.20
	Number of Incoming Calls	0.14



more extraverted are slightly more likely...
to **speak longer** in incoming calls...
to **receive calls**

	Use of Games	-0.43
	Use of Internet	-0.13



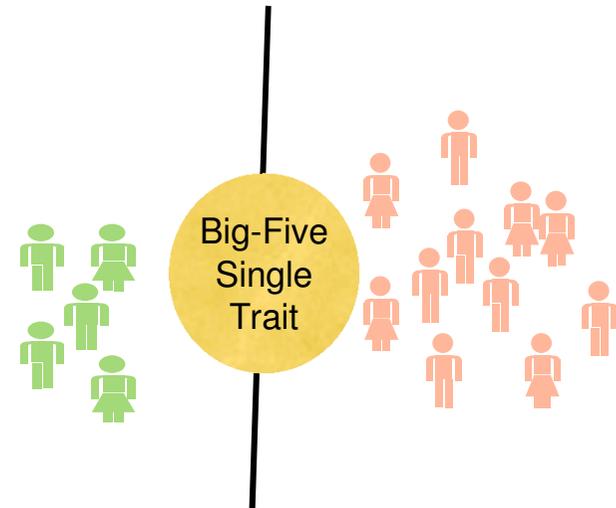
less extraverted are slightly more likely...
to **use the web**
to **play games**

big-five user-month classification experiments

2-class problem for each trait
high vs. low trait levels
(population split across median)

classifier: SVM, RBF kernel
leave 1-user out cross-validation

performance: F-measure



+ blue: better than majority classifier

Female Only

Male Only

Trait	Performance	Trait	Performance
Extraversion	0.80	Extraversion	0.72
Agreeableness	0.81	Agreeableness	0.83
Conscientiousness	0.77	Conscientiousness	0.75
Emotional Stability	0.78	Emotional Stability	0.73
Openness	0.72	Openness	0.76

conclusion



smartphones & mobile social data

- allow large-scale studies
- real-life data: from mobility to personality

open problems

- research questions are asked a-posteriori
- datasets have biases
- experiment design: custom-made data?
- larger scales and geographic diversity

mobile data challenge

- some data is available
- <http://www.idiap.ch/project/mdc/>
- email: mobiledata.challenge@idiap.ch

questions?

gatica@idiap.ch, daniel.gatica-perez@epfl.ch