computational social media

lecture 5: moving

daniel gatica-perez

idiap RESEARCH INSTITUTE

EPFL

announcements

reading #6 will be discussed today:

T. Gebru, J. Krause, Y. Wang, D. Chen, J. Deng, E. Lieberman Aiden, and L. Fei-Fei, Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the US, PNAS, 2017

reminder: assignment #3 is due next Monday (16.05, 7pm)

reminder: project schedule

1. **team building: DONE**
   email the list of your team members on **Week 2: Fri 04.03.2022**
   each team will have a designated project mentor

2. **project pitch: DONE**
   5-minute presentation of your project on **Week 5: Fri 25.03.2022**
   structure: title, problem, goals, approach, evaluation

3. **project progress presentation: : DONE**
   5-minute presentation per team on **Week 10: Fri 29.04.2022**

4. **final project presentation** on **Fri 10.06.2022**
   talk: 25-minute presentation + 20-minute questions
   schedule: 09:00-15:30

5. **final project report** by **Fri 17.06.2022**
   ACM conference paper format (6 pages + references + appendix)

final project presentation day (friday 10.06.2022)

09:00-09:45	group 1
09:45-10:30	group 2
10:30-10:45	break
10:45-11:30	group 3
11:30-12:15	group 4
12:15-13:00	lunch break
13:00-13:45	group 5
13:45-14:30	group 6
14:30-14:45	break
14:45-15:30	group 7

+ everybody is invited to attend the full day
+ please reserve the slot for your team
+ room to be confirmed (most likely ELD020)

this lecture

1. introduction

why location matters

2. motivations

why people use geolocalized social media

3. large-scale human mobility

understanding collective mobility from geo-localized media

4. human geography

discovering patterns related to land use

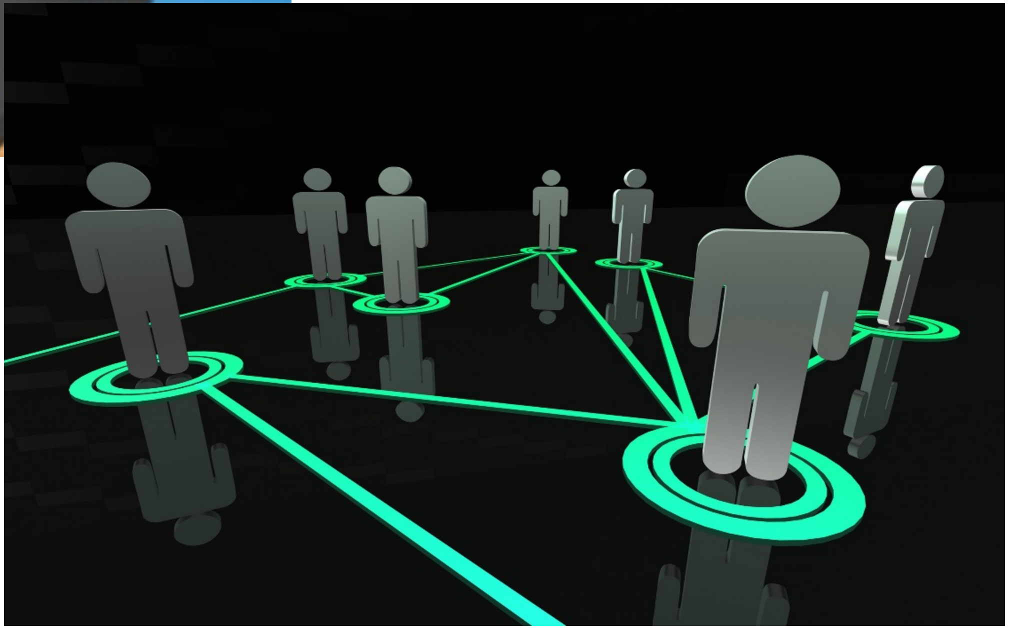
5. place perception

perceiving environments from social media places

6. biases in mobility data

urban and rural differences

location & social media: physical & online



foursquare (May 2016)

launched in 2009

predecessor: dodgeball (sold to google in 2005)

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

	2015	2016
monthly users	55M	50M
check-ins	7B	8B
places	65M	65M
employees	170	180

https://foursquare.com/about

FOURSQUARE

Products Resources Partners Company Log In Contact Us

If it tells you where, it's probably built on Foursquare

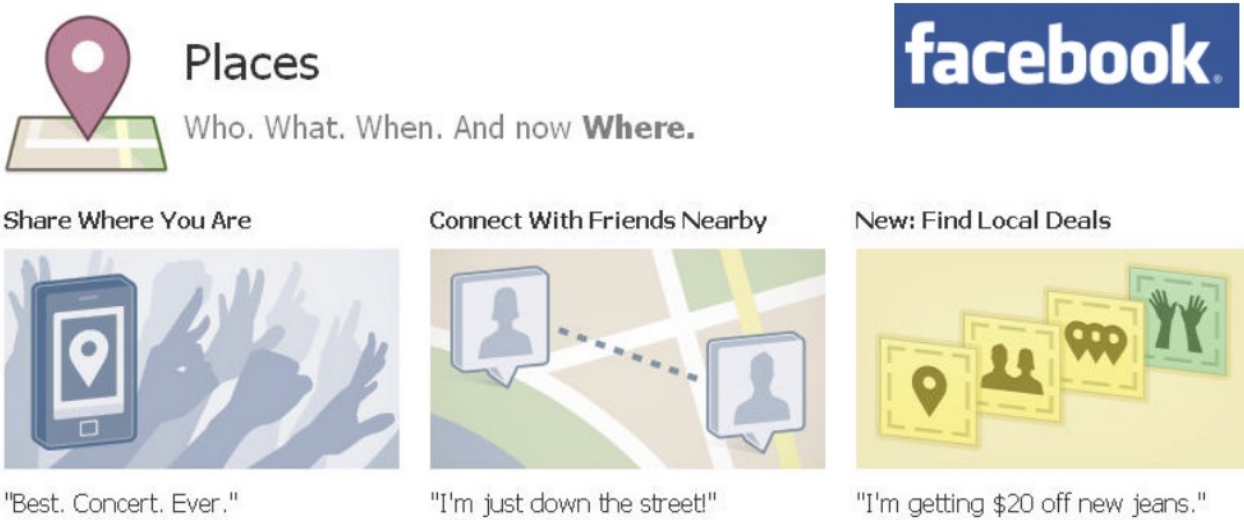
We believe in the power of location. What people experience in the real world and the places they go are powerful reflections of who they are and what they care about. We help leading global companies tap into this intelligence to create better customer experiences and smarter business outcomes, all based on the world's leading platform for understanding people, places, and the interactions between them.

https://foursquare.com/about/

Google latitude

2009-2013

facebook



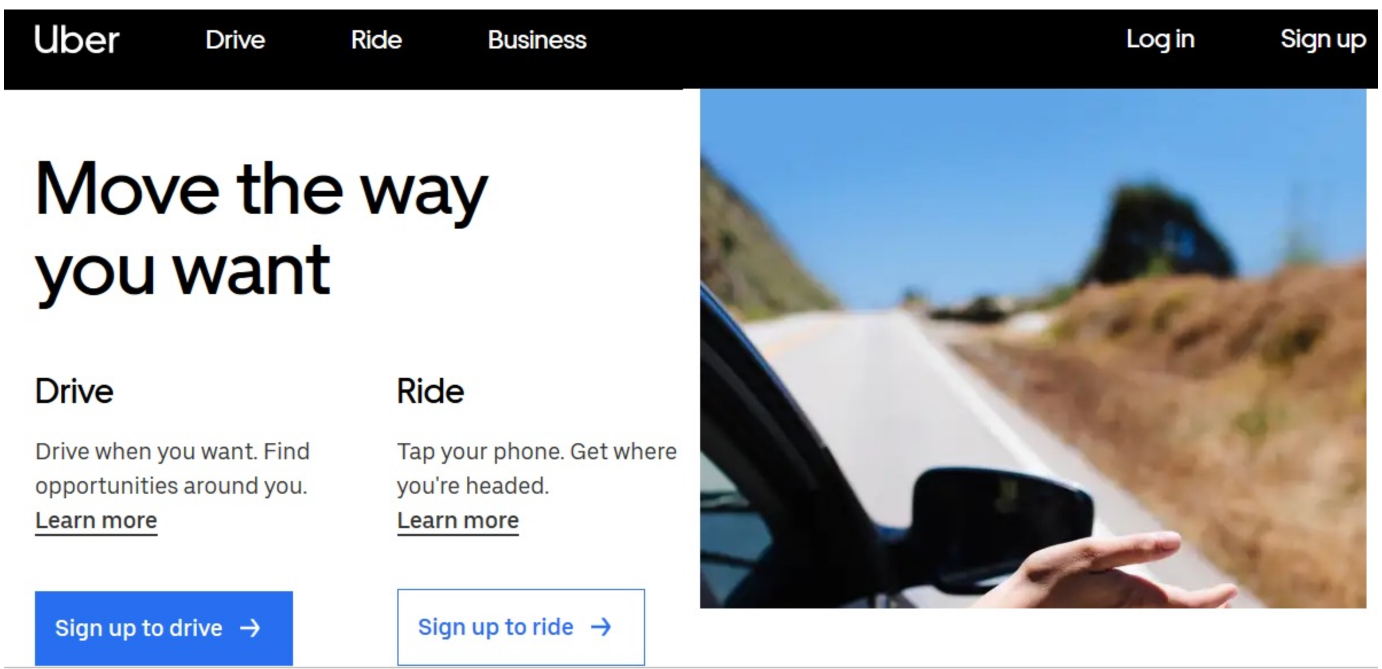
2010-2011,
revived 2014

yelp

Real People. Real Reviews.®

2005-present
check-ins: 2010

founded 2008



founded 2009

this lecture

1. introduction

why location matters

2. motivations

why people use geolocalized social media

3. large-scale human mobility

understanding collective mobility from geo-localized media

4. human geography

discovering patterns related to land use

5. place perception

perceiving environments from social media places

6. biases in mobility data

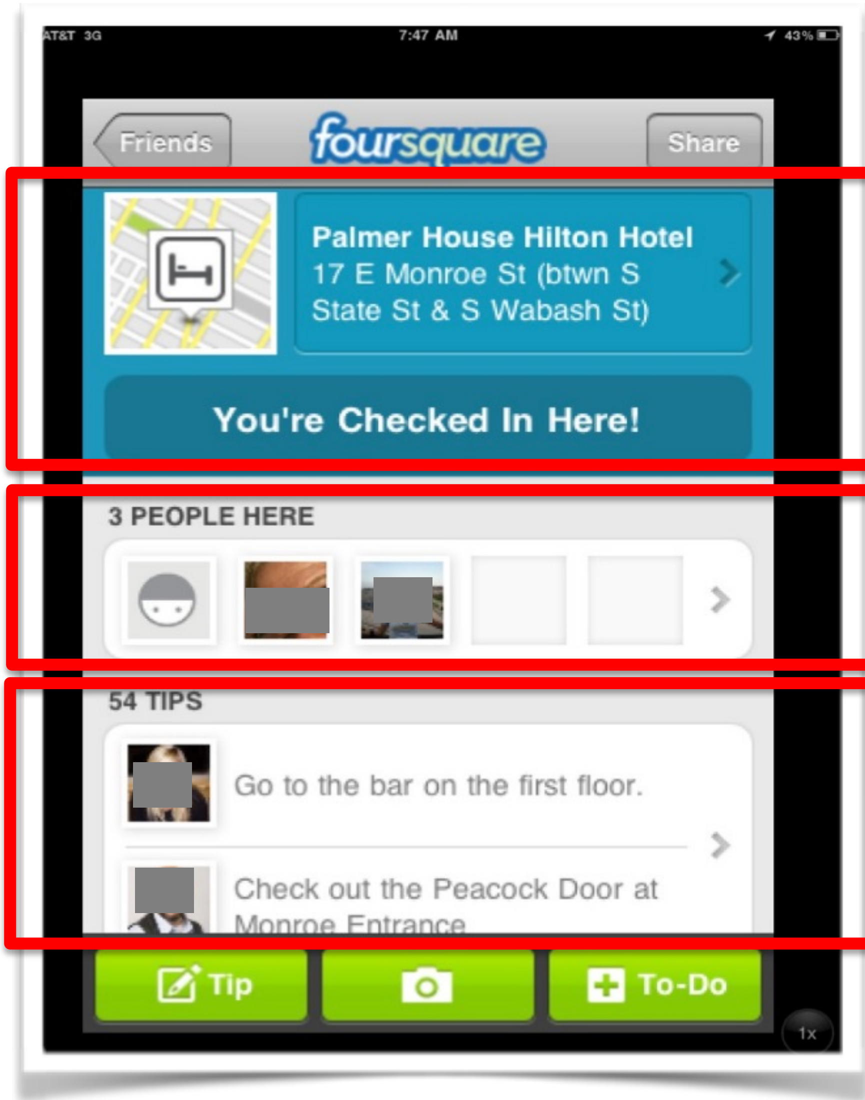
urban and rural differences

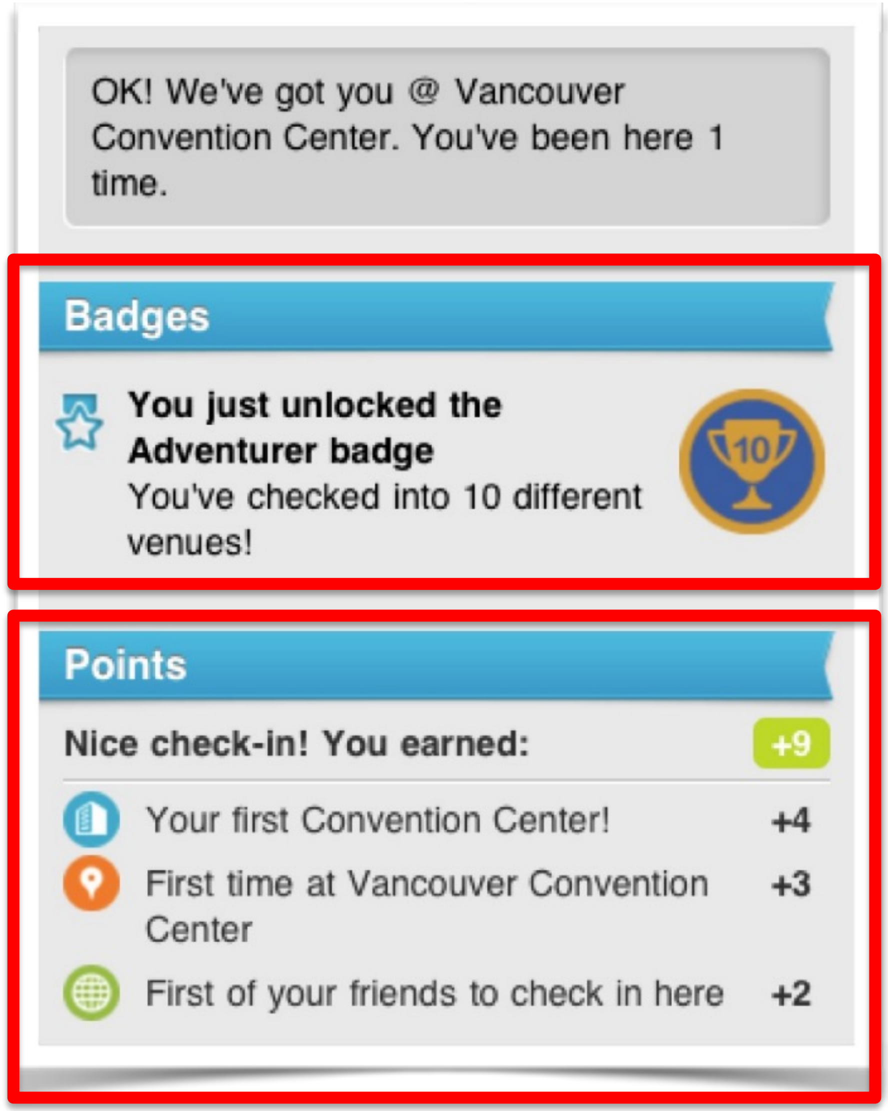
foursquare by foursquare

"mobile application that makes **cities easier to use** and more interesting to **explore**. It is a **friend-finder**, a **social city guide** and a **game** that challenges users to **experience new things**, and **rewards** them for doing so. Foursquare lets users **'check in'** to a place when they're there, **tell friends where they are** and **track the history** of where they've been and who they've been there with."

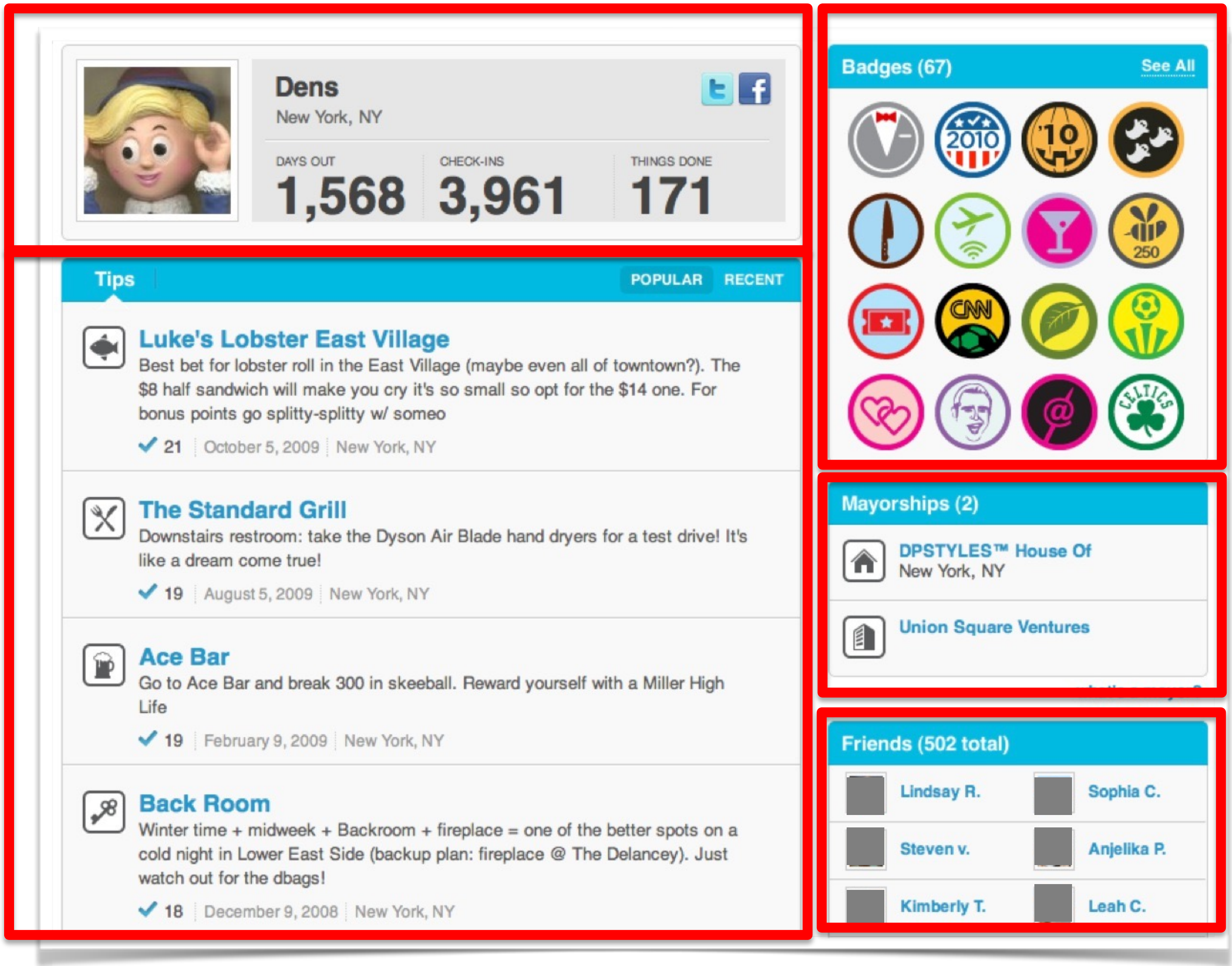
Source: foursquare & Janne Lindqvist http://www.winlab.rutgers.edu/iab/2011-02/documents/winlab-iab-nov-2011.pdf
Additional slides: Alexandra Olteanu

functionalities





4sq profile page



characterizing motivations to use 4sq

5 user studies (3 surveys + 2 interviews)

- **I1 (N=6)** interviews with early adopters
- **I2 (N=20)** interviews with typical foursquare users

- **S1 (N=18)** survey to qualitatively examine usage patterns
- **S2 (N=219)** survey to quantitatively probe questions about usage
- **S3 (N=47)** survey to qualitatively examine motivations for check-in

Janne Lindqvist, Justin Cranshaw, Jason Wiese, Jason Hong, John Zimmerman, I'm the mayor of my house: examining why people use foursquare - a social driven location sharing application, in Proc. CHI 2011
Henriette Cramer, Mattias Rost, Lars Erik Holmquist, Performing a Check-in: Emerging Practices, Norms and ‘Conflicts’ in Location-Sharing Using Foursquare, in Proc. Mobile HCI 2011

I1

I2

why do people use 4sq?

Personal tracking

Intimate sharing at a distance

Discovery of new people

Running into friends

Gaming aspect

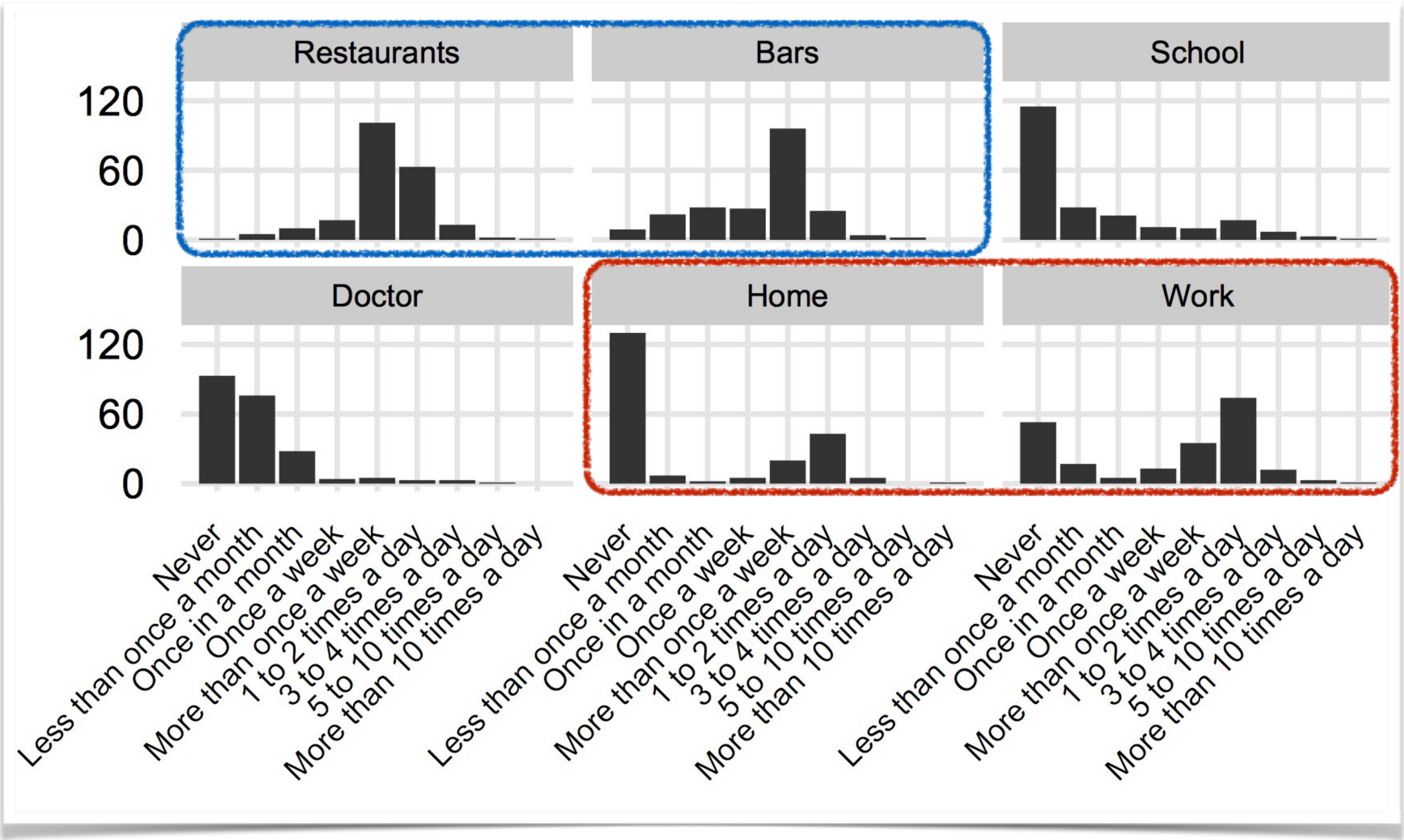
Seeing where friends have been

Routine vs. non-routine places

At large events

S2

where do people check-in?



S1 privacy

no privacy concerns (50%)

- good mental model of how 4sq worked

privacy concerns (50%)

- misalignment in how people understood 4sq
- concerns about stalkers and strangers

S1

S2

S3

managing privacy

74% used recognizable profile photos

58% friended people they never met

32% used 4sq to verify friends reached destination safely

Fun factor seemed more important than privacy

S1

I2

being mindful about 4sq

Self-representation issues

(don’t check @ fast food, doctors, banks, boring places)

Spam & interruptions to others

(avoid sending too many notification to friends)

Safety reasons

(indicate safe arrival after leaving a place)

To signal availability

(when alone at home)

this lecture

1. introduction

why location matters

2. motivations

why people use geolocalized social media

3. large-scale human mobility

understanding collective mobility from geo-localized media

4. human geography

discovering patterns related to land use

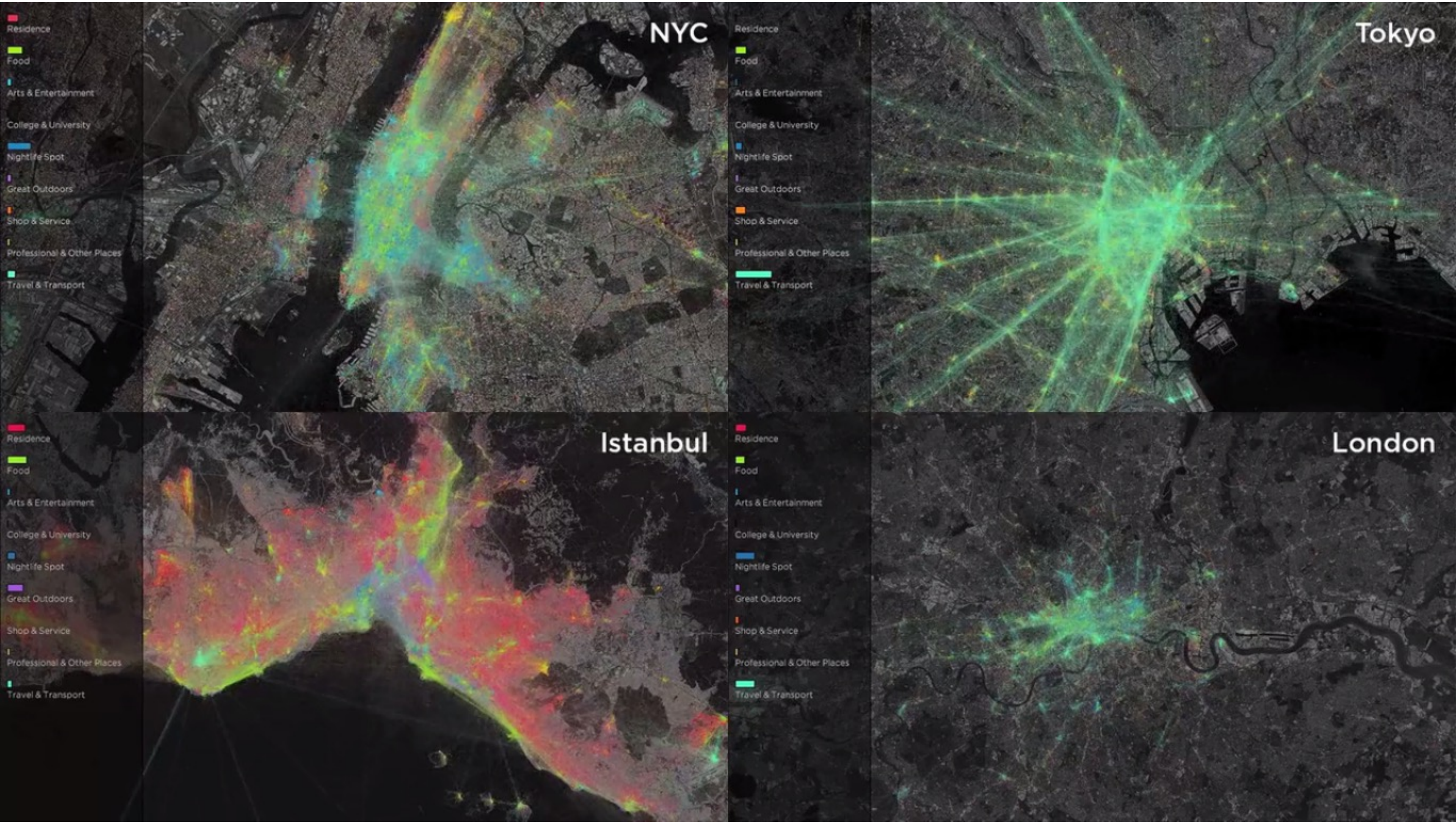
5. place perception

perceiving environments from social media places

6. biases in mobility data

urban and rural differences

mapping mobility in cities



credit: foursquare
https://vimeo.com/144409527

models for human mobility from check-ins

Gowalla: 6.4M check-ins, 196k users (02.2009-10.2010)
Brightkite: 4.5M check-ins, 58k users (04.2008-10.2010)

Foursquare: 35.2M check-ins, 925k users (05.2010-11.2010)

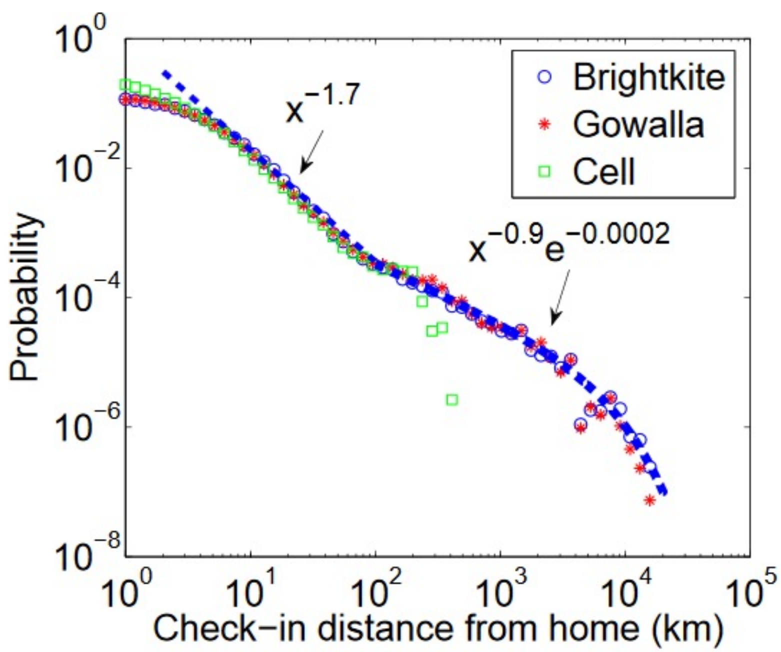


Figure 1: Fraction of check-ins as a function of distance traveled from home. Note the change in slope at around 100km.

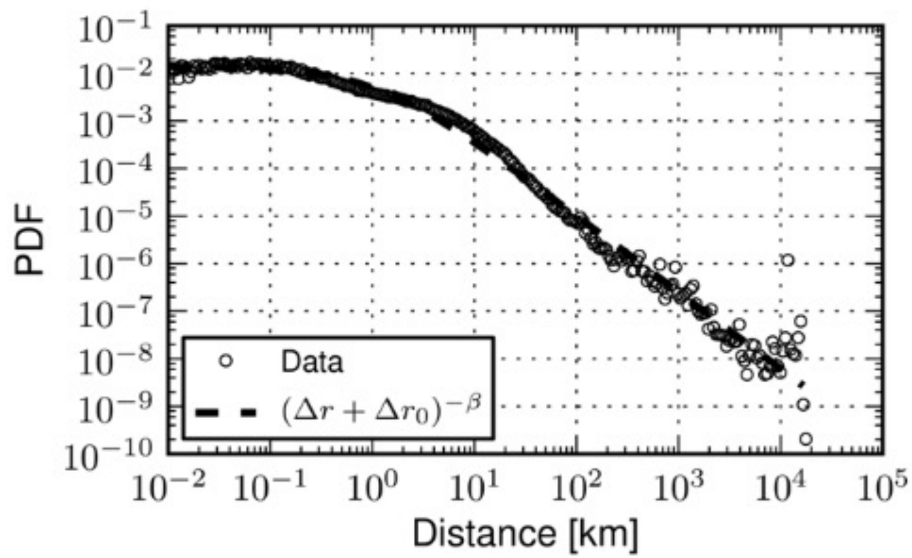
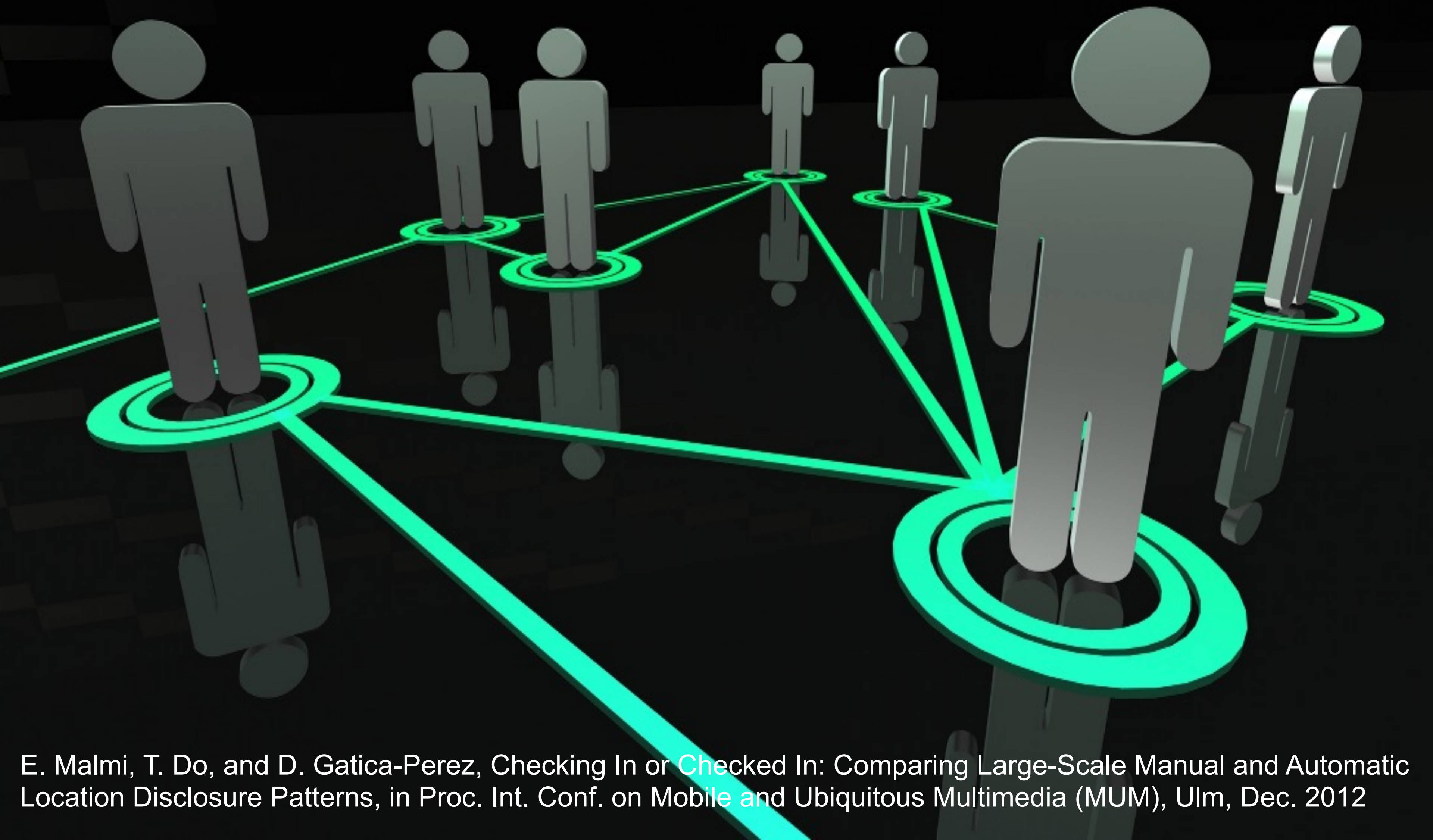


Figure 1. Global movements. The probability density function (PDF) of human displacements as seen through 35 million location broadcasts (check-ins) across the planet. The power-law fit features an exponent $\beta = 1.50$ and a threshold $\Delta r_0 = 2.87$ confirming previous works on human mobility data.

E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: user movement in location-based social networks. In Proc. ACM KDD 2011.

A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, C. Mascolo A Tale of Many Cities: Universal Patterns in Human Urban Mobility. PLoS ONE 7(5), 2012

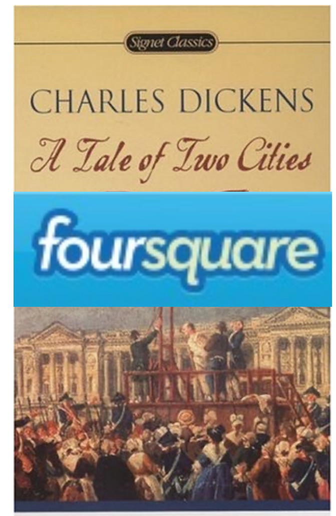
are check-ins a good proxy to understand large-scale mobility?



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

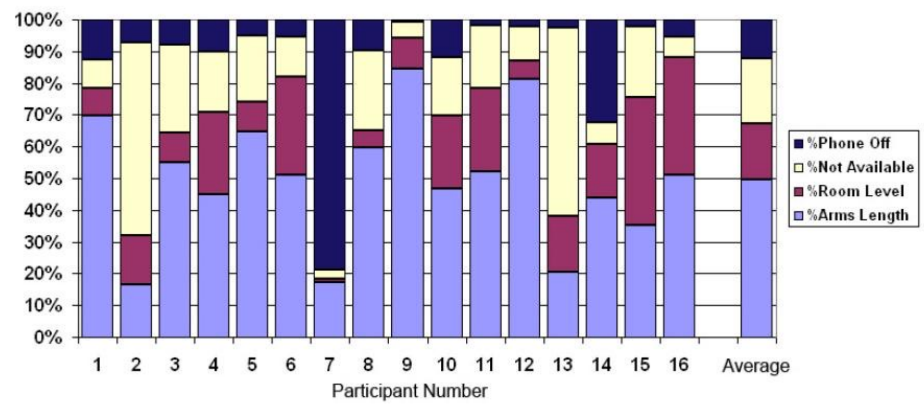
what about the assumptions?

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

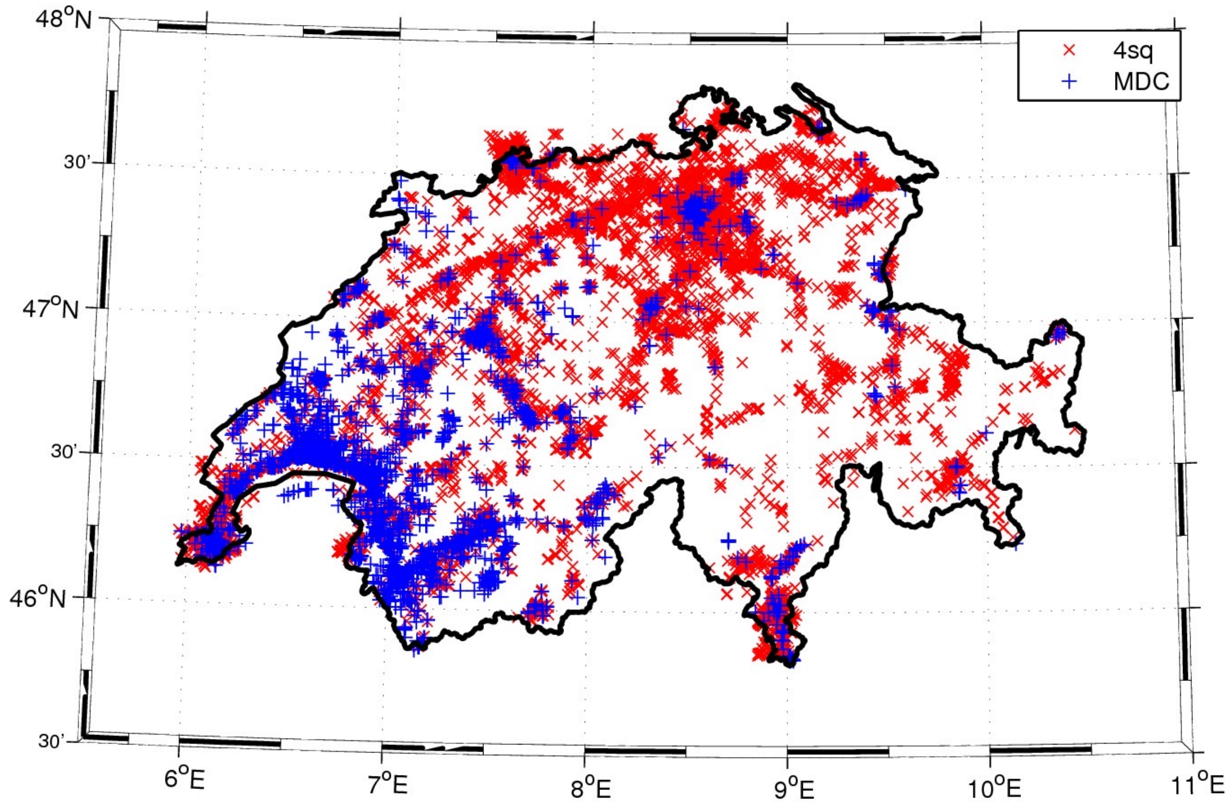


(Gonzalez, Nature 2008)
cell phone records (CDRs)
100 000 users, 6 months
0.91 call/sms per day

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



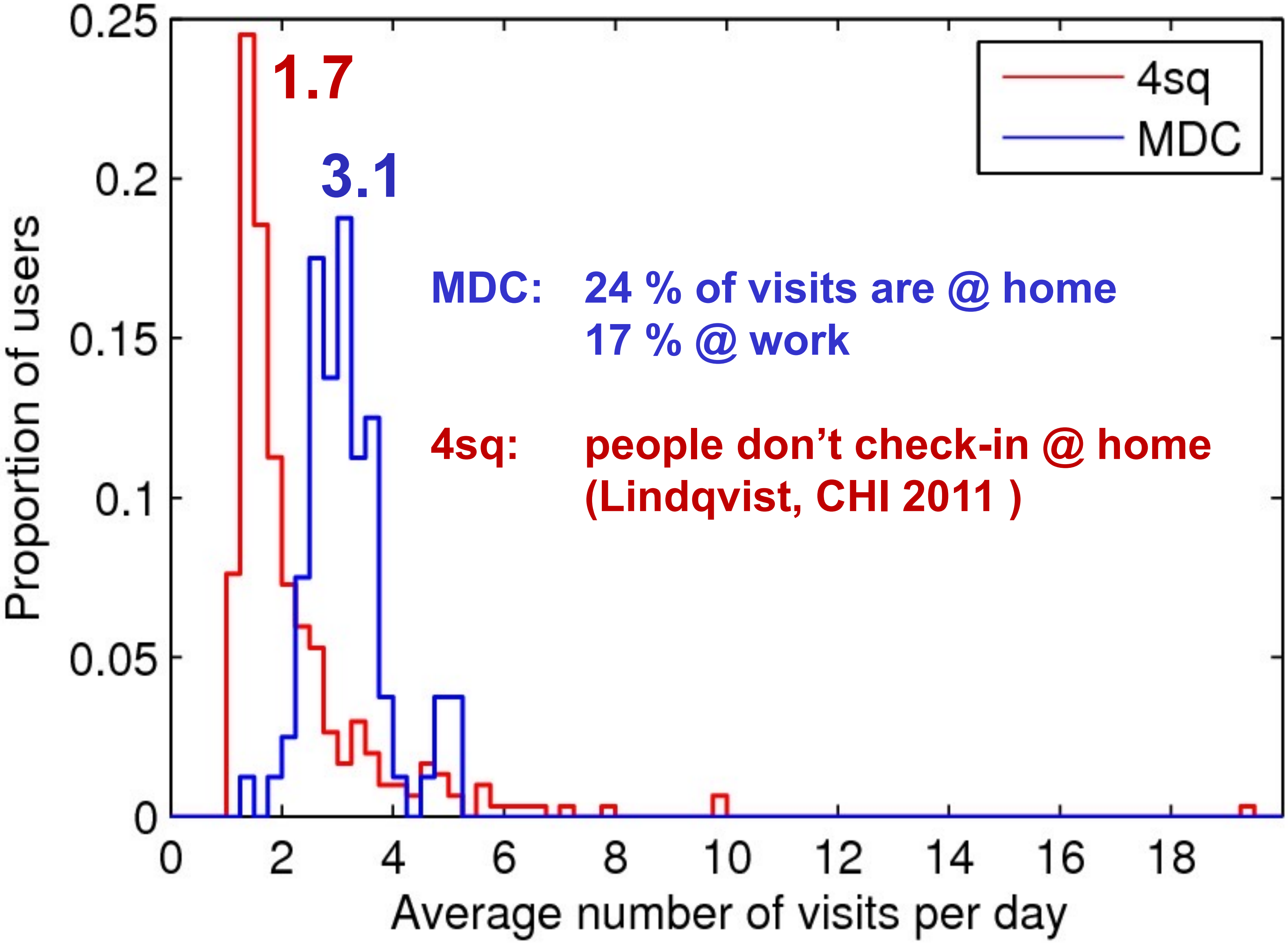
data: inferred check-ins vs. actual check-ins



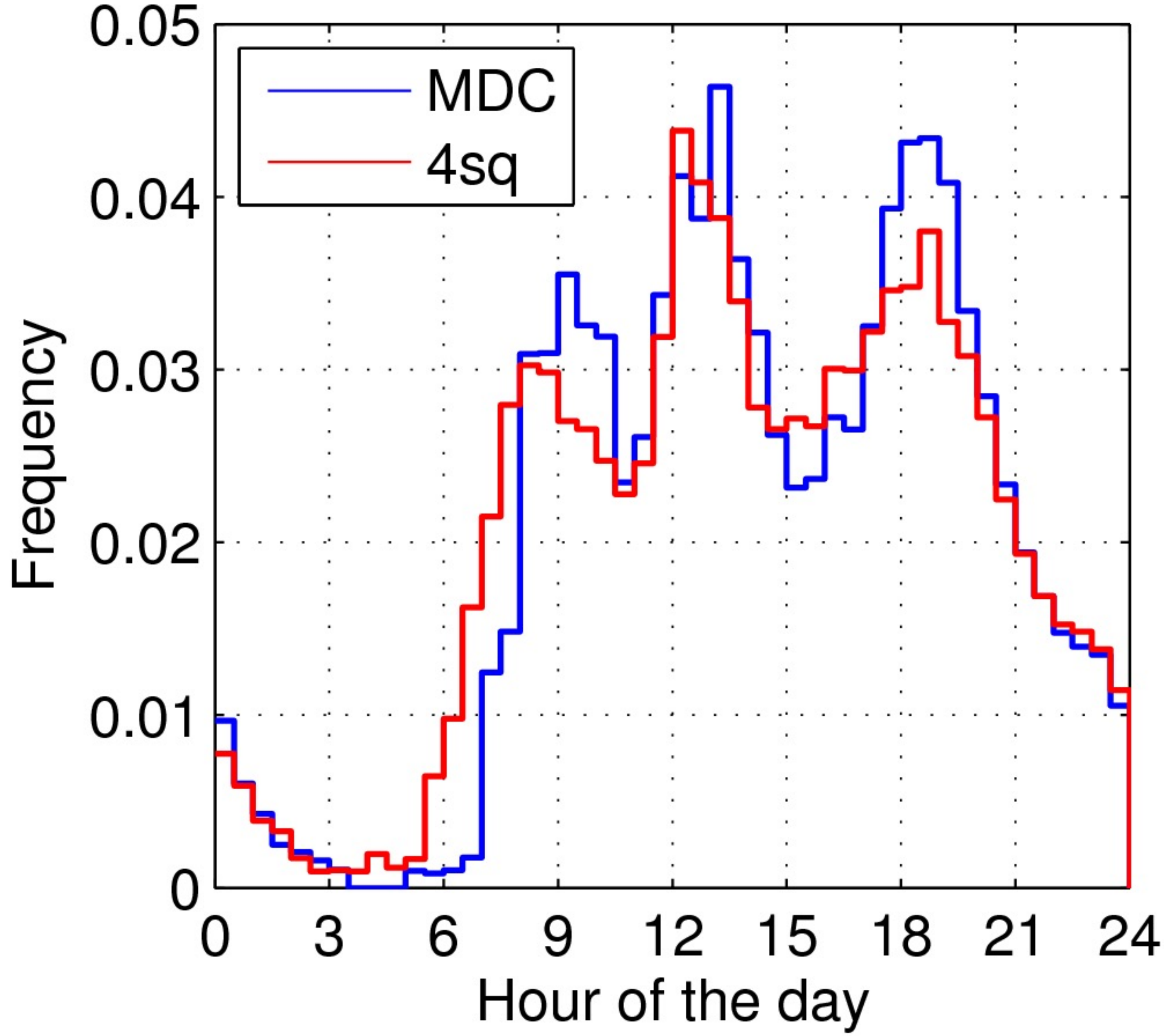
MDC: Mobile Data Challenge
(inferred 'check-ins')
80 active users
51,600 'check-ins'

4sq
(check-ins linked to tweets)
300 active users
40,600 check-ins

results (1)
daily check-in distributions



results (2)
the rhythm of daily activity



this lecture

1. introduction

why location matters

2. motivations

why people use geolocalized social media

3. large-scale human mobility

understanding collective mobility from geo-localized media

4. human geography

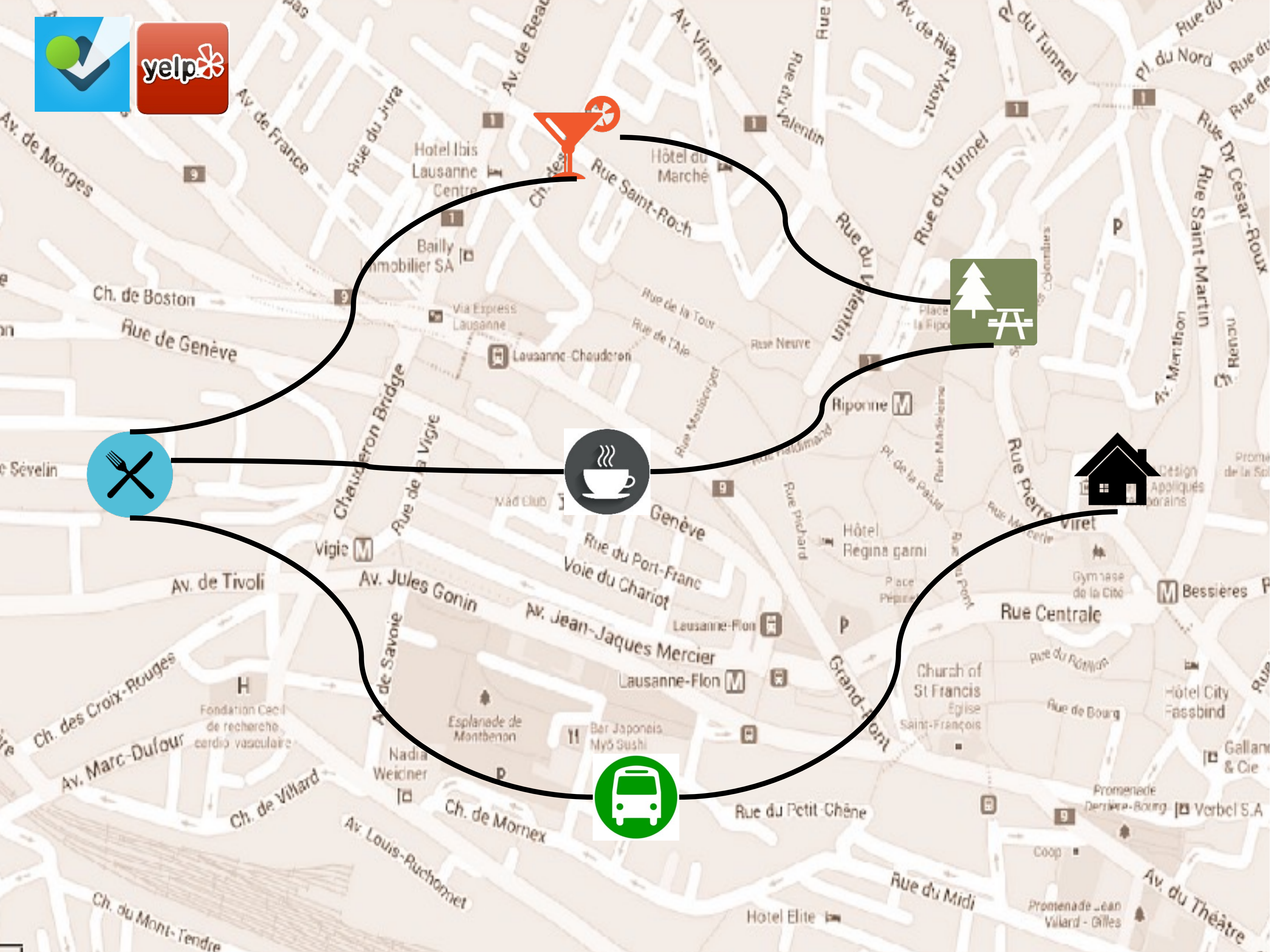
discovering patterns related to land use

5. place perception

perceiving environments from social media places

6. biases in mobility data

urban and rural differences



yelp
Av. de Morges
Av. de France
Rue du Jura
Av. de Beau
Hotel Ibis
Lausanne
Centre
Av. Vinet
Rue du Valentin
Pl. du Tunnel
Pl. du Nord
Rue Dr César-Roux
Rue Saint-Martin
Hôtel du
Marché
Rue Saint-Roch
Rue du Tunnel
Rue du Valentin
Bailly
Immobilier SA
Ch. de Boston
Rue de Genève
Via Express
Lausanne
Lausanne-Chauderon
Rue de la Tour
Rue de l'Ale
Rue Neuve
Riponne
Av. Marthon
Chauderon Bridge
Rue de la Vigie
Sévelin
Vigie
Genève
Rue du Port-Franc
Voie du Chariot
Rue Pierre Viret
Hôtel
Regina garni
Av. de Tivoli
Av. Jules Gonin
Av. Jean-Jaques Mercier
Lausanne-Flon
Bessières
Rue Centrale
Gymnase
de la Cité
Church of
St Francis
Église
Saint-François
Rue de Bourg
Hôtel City
Fassbind
Ch. des Croix-Rouges
Fondation Cecil
de recherche
cardio vasculaire
Av. Marc-Dufour
Esplanade de
Montbenon
Bar Japonais
Myô Sushi
Nadia
Weidner
Ch. de Villard
Ch. de Mornex
Rue du Petit-Chêne
Grand-Pont
Promenade
Derrière-Bourg
Verbel S.A
Av. Louis-Ruchonnet
Rue du Midi
Coop
Hotel Elite
Promenade Jean
Villard - Gilles
Av. du Théâtre
Ch. du Mont-Tendre

urban patterns of land use from 4sq data

Livehoods

Home Maps About Research Press Contact

livehoods.org

Re-Imagining the City in the Age of Social Media

Livehoods offer a new way to conceptualize the dynamics, structure, and character of a city by analyzing the social media its residents generate. By looking at people's checkin patterns at places across the city, we create a mapping of the different dynamic areas that comprise it. Each Livehood tells a different story of the people and places that shape it.

> MORE

Using Machine-Learning to Study Cities

Our research hypothesis is that the character of an urban area is defined not just by the the types of places found there, but also by the people that make it part of their daily life. To explore this idea, we use data from approximately 18 million check-ins collected from the location-based social network foursquare, and apply clustering algorithms to discover the different areas of the city.

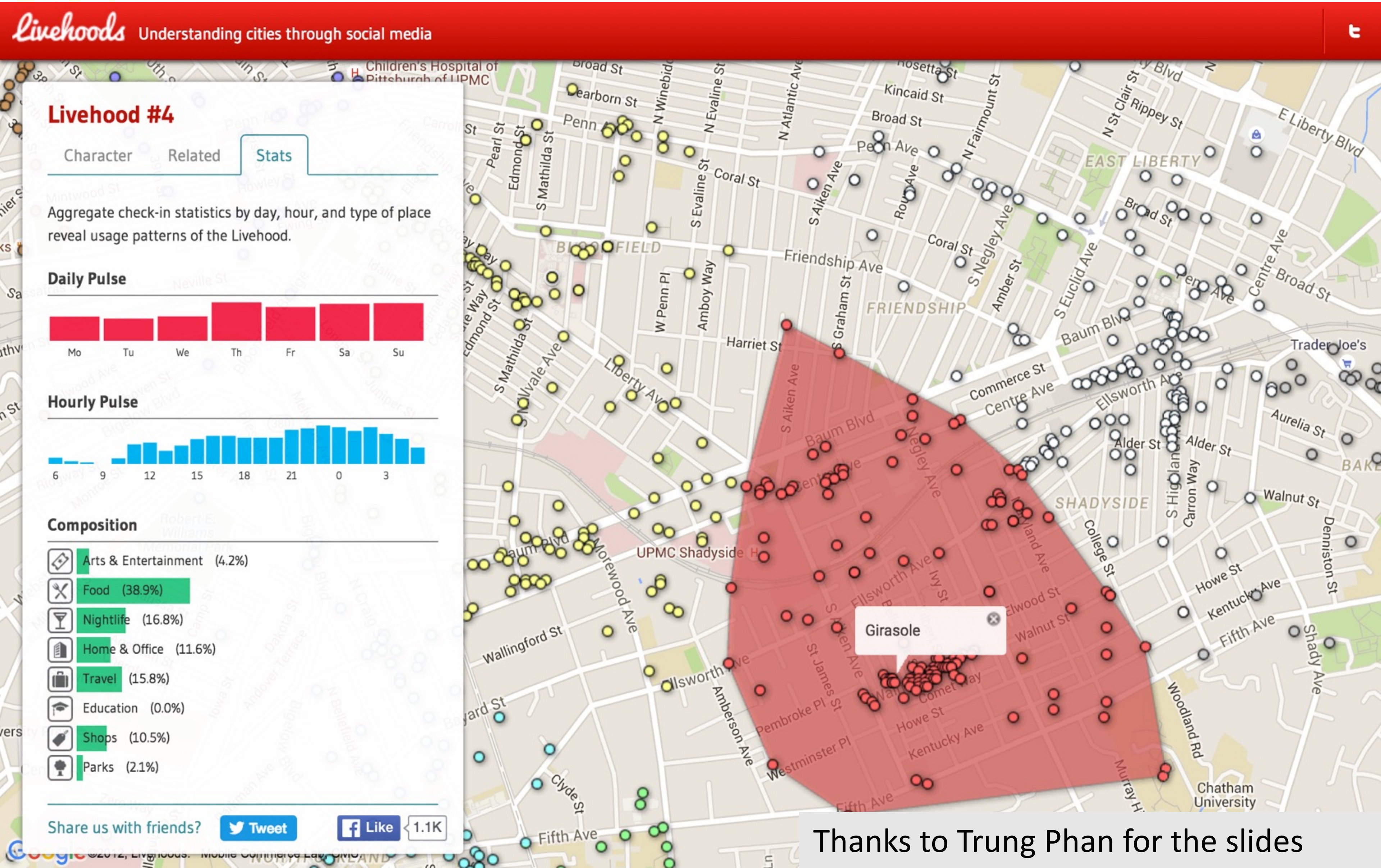
> MORE

Current Maps

NYC > New York City   SFO > San Francisco   PIT > Pittsburgh   > More Maps

J. Cranshaw, R. Schwartz, J. I. Hong, N. Sadeh, The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City, in Proc. AAAI Int. Conf. on Weblogs and Social Media (ICWSM), 2012.

"livehoods": urban regions with similar activities & users



Thanks to Trung Phan for the slides

clustering 4sq venues:
pairwise distance between venues

- **geographic distance**

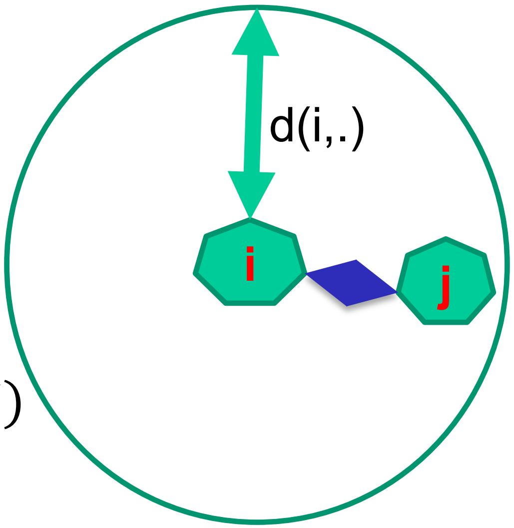
  d(i,j): physical distance between venues i &j using latitude & longitude

- **social distance**

  set of venues V, $n_V = 5349$
  set of users U, $n_U = 3840$
  set of check-in vectors per venue C:
  at each venue *v*, build vector c_v having n_U dimensions, each dimension is # of check-ins of user u^{th} at venue *v*

$s(i,j) = \frac{c_i . c_j}{\|c_i\| . \|c_j\|}$, c_i ,c_j are vectors for each venue (cosine distance)

clustering 4sq venues (2): build a venue graph



- **venue graph**

  1. at venue i, choose $\boldsymbol{N_m(i)}$: ***m closest venues*** using geographic distance d(i,j):

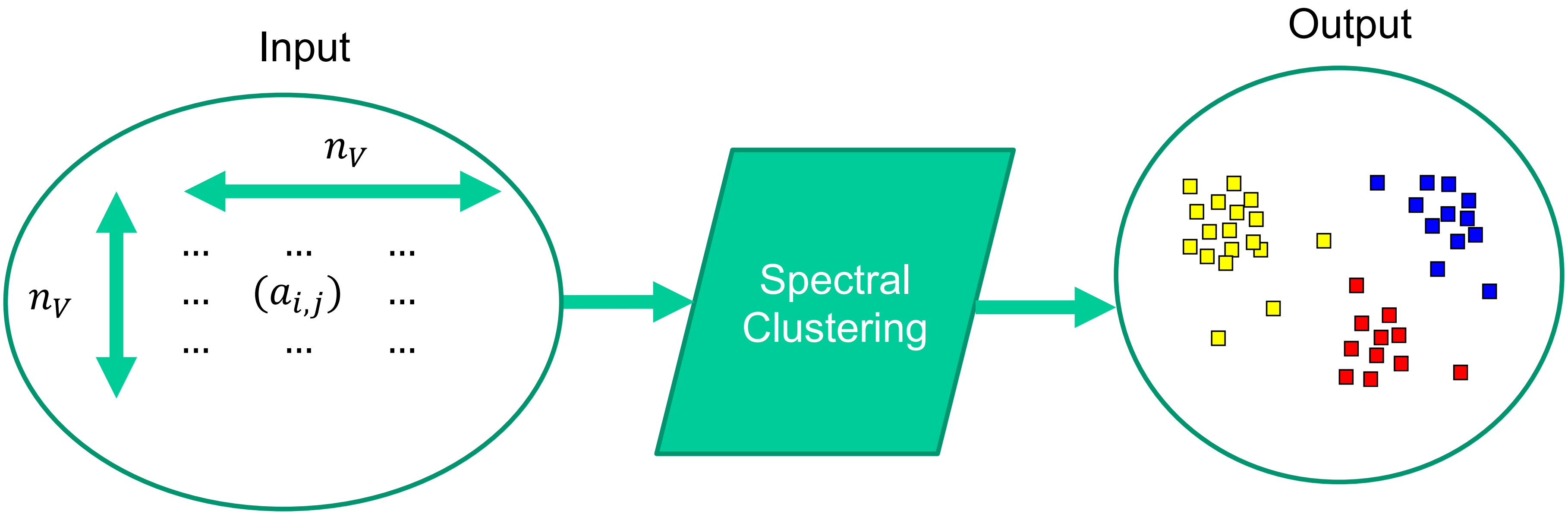
$$a_{i,j} = \begin{cases} s(i,j) + \alpha & if\ j \in N_m(i) or\ i \in N_m(j) \\ 0 & otherwise \end{cases}$$

  where α is a constant and $s(i,j)$ is social distance

  2. build graph G(A) with matrix size n_V x n_V

$$\begin{array}{cc} & n_V \\ n_V & \begin{array}{ccc} \dots & \dots & \dots \\ \dots & (a_{i,j}) & \dots \\ \dots & \dots & \dots \end{array} \end{array}$$

spectral clustering of venues



A. Y. Ng, M. I. Jordan,; and Y. Weiss, On spectral clustering: Analysis and an algorithm. In Proc. NIPS 2001.

discovered livehoods in NYC

https://hcii.cmu.edu/news/2012/cmu-researchers-use-foursquare-check-data-create-dynamic-view-cities

this lecture

1. introduction

why location matters

2. motivations

why people use geolocalized social media

3. large-scale human mobility

understanding collective mobility from geo-localized media

4. human geography

discovering patterns related to land use

5. place perception

perceiving environments from social media places

6. biases in mobility data

urban and rural differences

which place feels louder?


image: shbs@Pixabay


image: PublicDomainPictures@Pixabay

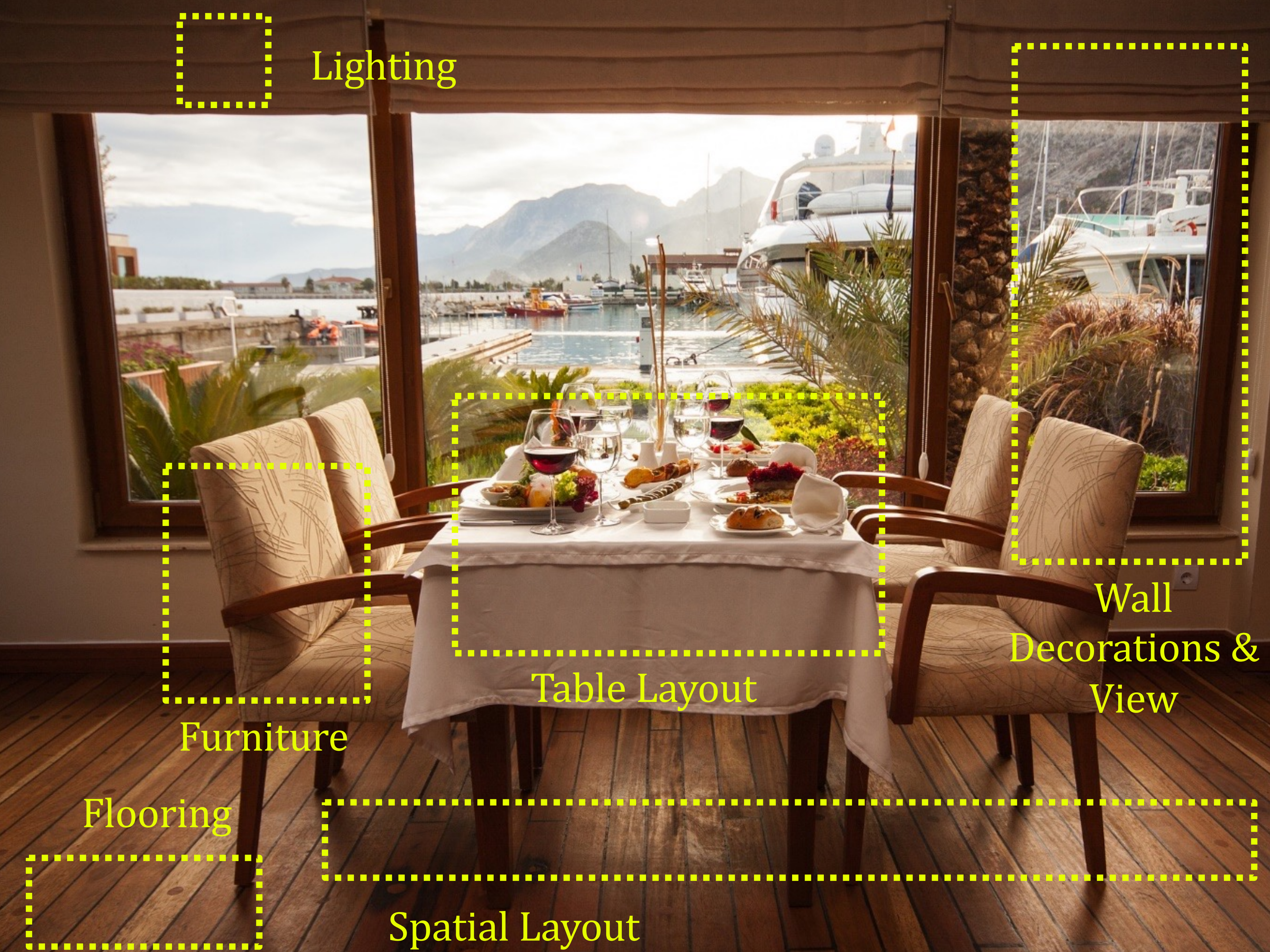
ambiance: “the mood or feeling associated with a particular place” or “the character and atmosphere of a place”

Journal of Personality and Social Psychology
1980, Vol. 38, No. 2, 311-322

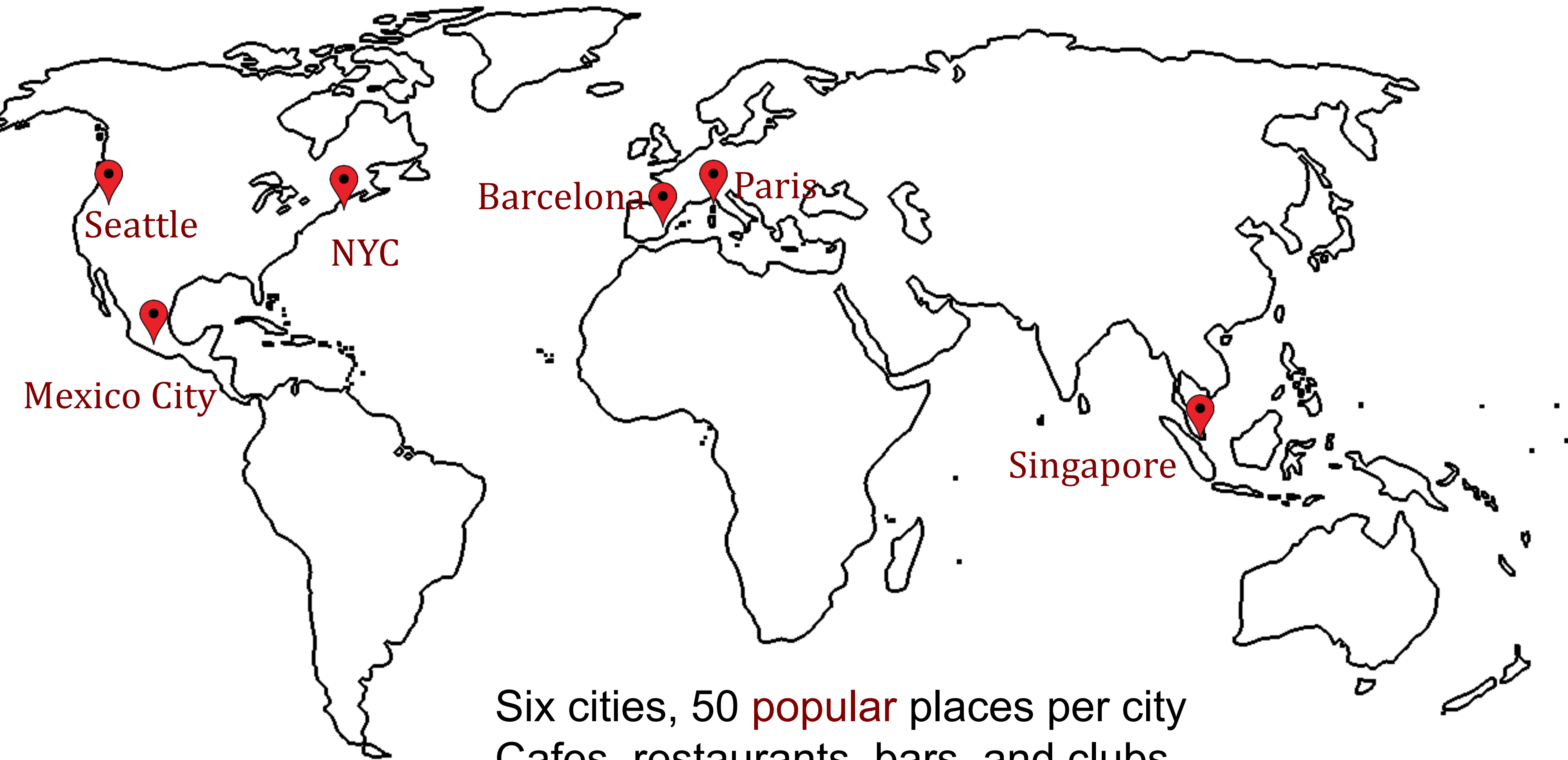
A Description of the Affective Quality Attributed to Environments

James A. Russell and Geraldine Pratt
University of British Columbia, Vancouver, Canada

The meaning that persons attribute to environments is divided into perceptual-cognitive meaning and affective meaning. Affective meaning is then conceptualized as a two-dimensional bipolar space that can be defined by eight variables falling in the following circular order around the perimeter: pleasant (arbitrarily set at 0°), exciting (45°), arousing (90°), distressing (135°), unpleasant (180°), gloomy (225°), sleepy (270°), and relaxing (315°, which is thus 45° from pleasant). Alternatively, the same space can be defined by two orthogonal bipolar dimensions of pleasant–unpleasant and arousing–sleepy—or equally well by exciting–gloomy and distressing–relaxing. Reliable verbal scales for these eight variables are developed and shown to approximate the proposed theoretical structure.


Lighting
Wall
Decorations &
View
Table Layout
Furniture
Flooring
Spatial Layout

ambiance dataset: popular places in 4sq



Six cities, 50 popular places per city
Cafes, restaurants, bars, and clubs
50,000 images via 4sq API

D. Santani and D. Gatica-Perez, Loud and Trendy: Crowdsourcing Impressions of Social Ambiance in Popular Indoor Urban Places, in Proc. ACM Int. Conf. on Multimedia, 2015

how do people perceive ambiance?

N=300 places, 10 MTurk raters per place, 5-point Likert scale

Label	Combined ICC
Artsy	0.76
Bohemian	0.62
Conservative	0.76
Creepy	0.59
Dingy	0.74
Formal	0.91
Sophisticated	0.86
Loud	0.80
Old-fashioned	0.72
Off the beaten path	0.58
Romantic	0.82
Trendy	0.69
Up-scale	0.86

[0.8, 1.0)

[0.6, 0.8)

[0.5, 0.6)

Ambiance types

1. Feature extraction using CNN pre-trained on ImageNet

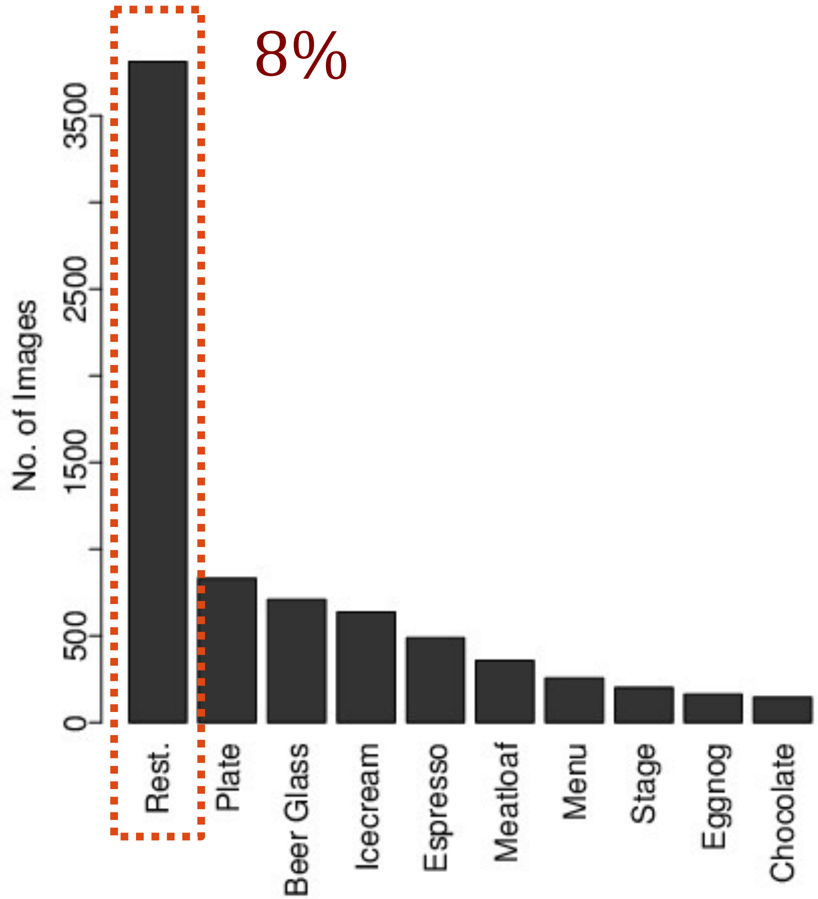
2. Regression with Random Forest

- Artsy
- Bohemian
- Conservative
- Creepy
- Dingy
- Formal
- Sophisticated
- Loud
- Old Fashioned
- Off the beaten path
- Romantic
- Trendy
- Upscale

D. Santani, R. Hu, D. Gatica-Perez, InnerView: Learning Place Ambiance from Social Media Images, in Proc. ACM Int. Conf. on Multimedia, 2016

results

CNN-extracted features & regression



50K Corpus

Visual categories recognized by CNN

Variables	R^2
Artsy	0.22
Bohemian	0.24
Conservative	0.30
Creepy	0.14
Dingy	0.17
Formal	0.37
Loud	0.52
Off-the-beaten-path	0.17
Old fashioned	0.22
Romantic	0.39
Sophisticated	0.38
Trendy	0.32
Upscale	0.40

this lecture

1. introduction

why location matters

2. motivations

why people use geolocalized social media

3. large-scale human mobility

understanding collective mobility from geo-localized media

4. human geography

discovering patterns related to land use

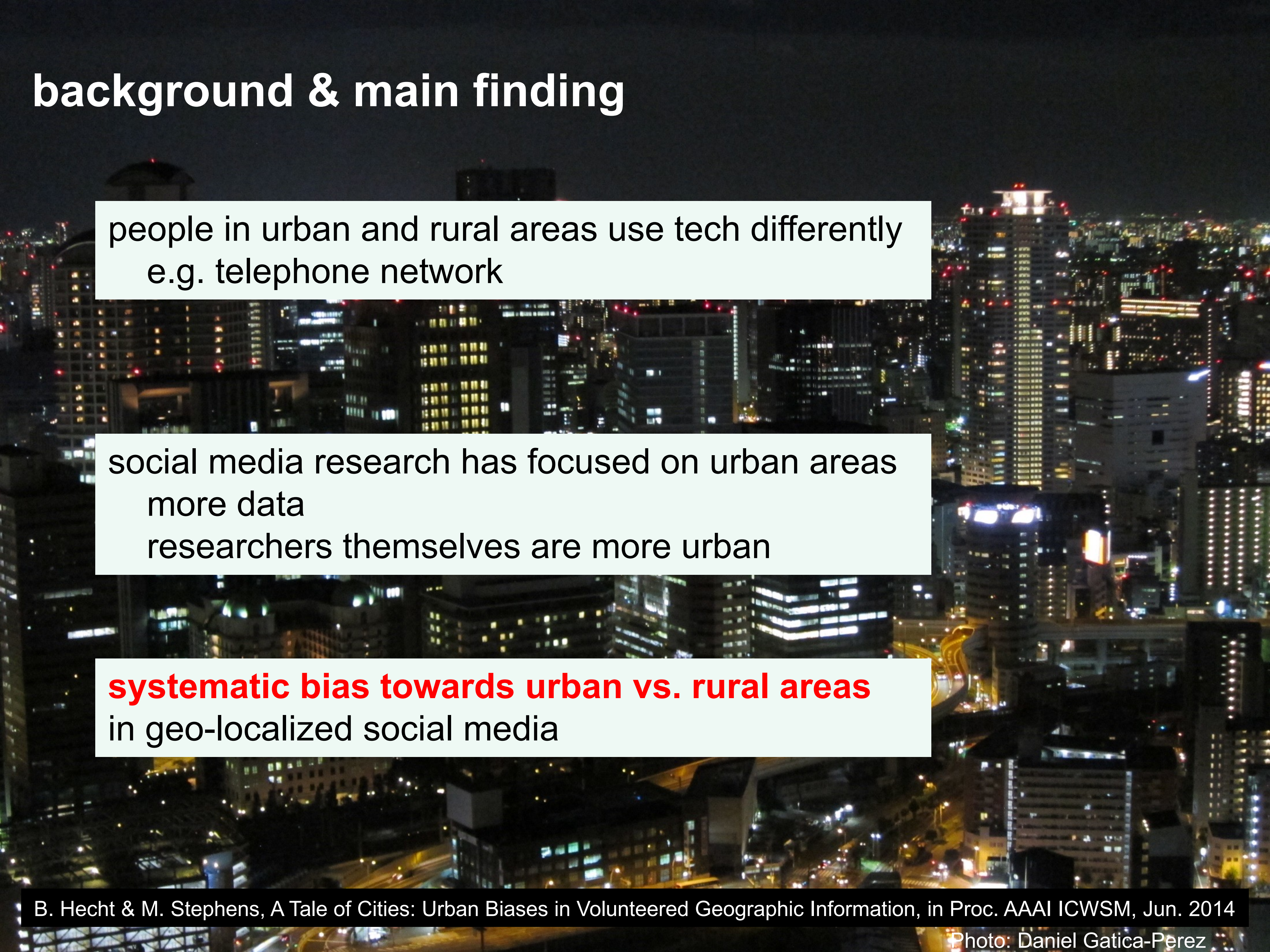
5. place perception

perceiving environments from social media places

6. biases in mobility data

urban and rural differences

background & main finding



people in urban and rural areas use tech differently
- e.g. telephone network

social media research has focused on urban areas
- more data
- researchers themselves are more urban

systematic bias towards urban vs. rural areas
in geo-localized social media

B. Hecht & M. Stephens, A Tale of Cities: Urban Biases in Volunteered Geographic Information, in Proc. AAAI ICWSM, Jun. 2014

Photo: Daniel Gatica-Perez

method

Geolocalized social media

Twitter:	56.7M tweets; 1.6M users
Flickr:	52M photos; 522K users
4sq:	11.1M checkins; 122K users

US Census Bureau (2010, county-level)

Urban area: Big cities and towns of population 2,500 or more
https://www.census.gov/geo/reference/ua/uafaq.html

"Urbanness": percentage of population in a county who lives in an urban area

Aggregrating social media at county level & comparing with US Census

- + Identify "local" users
- + Account for spatial autocorrelation (adjacent measures are correlated)
  - + Clifford's correction (effective sample size)
- + Correlation analysis
  - + Spearman rank correlation
  - + Significance test with Bonferroni correction

correlation results (social media vs. “urbanness”)

NOTE: N is not reported,
US census 2013: 3,143 counties

2 ways to infer
who are local users

Property	n-days	plurality
Users per Capita	0.46***	0.54***
Number of Total Tweets per Capita	n/a	0.53***
Sample Period Tweets per Capita	0.49***	0.50***
Median Total Tweets	n/a	0.28***
@ Mentions per Tweet	0.19***	0.21***

*Table 1: Attributes of Twitter VGI and their correlation with the percent of a population that lives in urban area. Significance is calculated using the Clifford et al. “effective sample size” method that controls for spatial autocorrelation in spatial datasets. † (marginally) significant at p < .10; * significant at p < .05; ** significant at p < .01 *** significant at p < .001 (with Bonferroni correction)*

Property	n-days	cluster
Median Number of Photos Per User	0.41***	0.38***
Tags per Photo	0.11***	0.26***
Photos per Capita	0.20***	0.26***
Users per Capita	-0.05 (n.s.)	0.10 (n.s.)

Table 2: Spearman’s correlations between the percent urban population in a county and properties of our Flickr data assigned to that county.

Attribute	n-days	cluster
Check-Ins Per Capita	0.61***	0.63***
Foursquare Users per Capita	0.51***	0.61***
Median Number of Check-Ins Per User	0.51***	0.43***

Table 3: Spearman’s correlations coefficients between the percent urban population in a county and properties of the Foursquare data assigned to each county.

implications

Big bias towards urban areas

Results per capita
+ 24 times more 4sq users
+ 3 times more Twitter users
+ 5 times more tweets

Studies using geo-localized social media **"are less studies of human behavior than studies of urban human behavior"**

Do not call results "universal laws" or "general mobility laws"

Photo: Daniel Gatica-Perez

what to remember

motivations for use of geo-localized social media

- many positive ones (fun, social, safety, local search)
- but also privacy implications

human mobility

- informative data source, but not holy grail
- limitations w.r.t. temporal resolution

human geography

- potential to inform specific urban aspects
- bias towards cities, rural areas not well represented
- bias towards economically developed areas

place perception

- environmental psychology research at scale
- deep learning as a tool to support visual analysis

questions?

daniel.gatica-perez@epfl.ch