

I'm Here, So What?

**The Value for Statistics of Swiss
Geo-Localized Social Media**

Daniel Gatica-Perez

Darshan Santani & Trung Phan Thanh

Mobile social media



credit: garryknight
@flickr (cc)

Linking the physical world and online activity





Google
latitude



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."



yelp
*Real People. Real Reviews.**

The talk

The background of the slide is a stylized illustration of a hand holding a smartphone. The phone's screen displays a map with various icons and lines. The hand is rendered in a light skin tone. The overall background is a solid blue color. Three semi-transparent purple text boxes are overlaid on the image, each containing a numbered point and a question.

1. Geo-localized social data

What is the data about?

2. Opportunities

How does social data add value to research?

3. Challenges

What are the inherent limitations?

1.
Geo-localized social data



Barack Obama ✓

@BarackObama

This account is run by Organizing for Action staff. Tweets from the President are signed -bo.

Joined March 2007



Barack Obama ✓

@BarackObama



Follow

If we don't #ActOnClimate now, the sea level could rise by more than six feet by the end of the century: ofa.bo/uOB



Antarctic loss could double expected sea level rise by 2100, scientists say

If carbon emissions continue unabated, expanding oceans and massive ice melt would threaten global coastal communities, according to new projections.

[washingtonpost.com](https://www.washingtonpost.com)

RETWEETS

1,417

LIKES

2,482



5:55 PM - 6 Apr 2016



Tweet Attributes

```
[u'contributors',  
  u'truncated',  
  u'text',  
  u'is_quote_status',  
  u'in_reply_to_status_id',  
  u'id',  
  u'favorite_count',  
  u'source',  
  u'retweeted',  
  u'coordinates',  
  u'timestamp_ms',  
  u'entities',  
  u'in_reply_to_screen_name',  
  u'id_str',  
  u'retweet_count',  
  u'in_reply_to_user_id',  
  u'favorited',  
  u'retweeted_status',  
  u'user',  
  u'geo',  
  u'in_reply_to_user_id_str',  
  u'possibly_sensitive',  
  u'lang',  
  u'created_at',  
  u'filter_level',  
  u'in_reply_to_status_id_str',  
  u'place',  
  u'extended_entities']
```

Entities

Media

Urls

User Mentions (@)

Hashtags (#)

Places are named locations with geo coordinates.

Tweets associated with places are not necessarily issued from that location

Sources:

1. <https://dev.twitter.com/overview/api/tweets>
2. <https://dev.twitter.com/overview/api/entities>
3. <https://dev.twitter.com/overview/api/places>

Getting Data from Twitter

REST API

Public, Non-persistent, imposes clear rate limits

Streaming API

Public, Persistent

Monitor & process in real-time

Up to 1% of all tweets produced on Twitter
(exact % unknown)

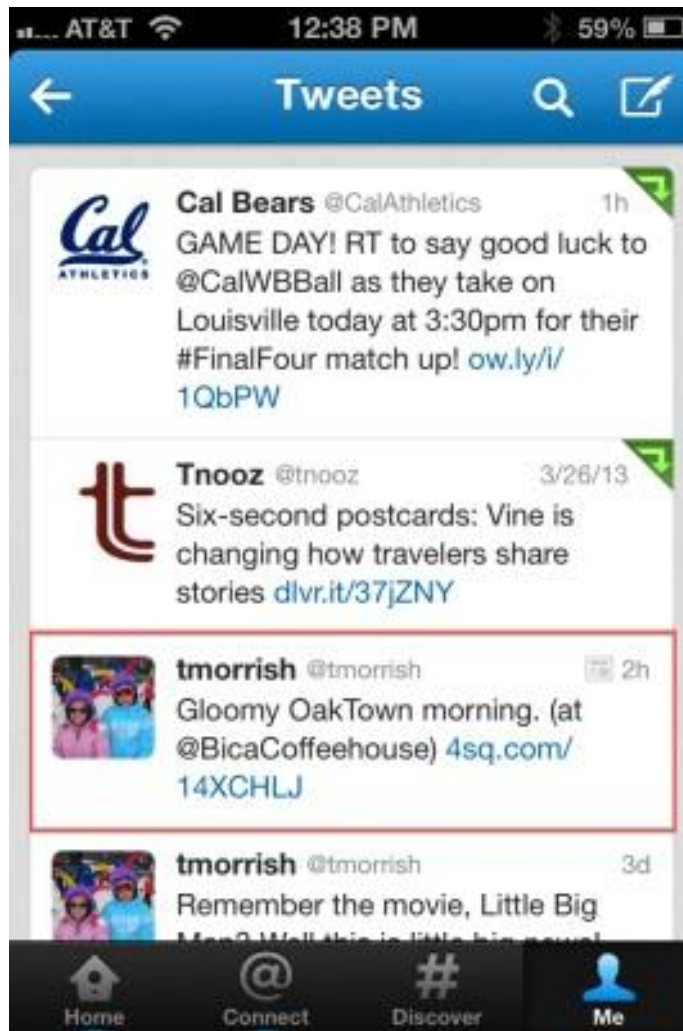
Representativeness – Sample bias

Firehose

100% of all public tweets

Expensive: financial and computing resources

Geo-localized tweets and Foursquare check-ins



credit: <https://www.tnooz.com/article/twitter-cards-travel-use-cases-you-need-to-know-about-and-how-to-do-it/>

Foursquare

The screenshot displays the Foursquare website interface. At the top, there is a blue navigation bar with the Foursquare logo, a search bar containing "I'm looking for...", a dropdown menu, a "Current Map View" button, a search icon, and "Log In" and "Sign Up" buttons. Below the navigation bar, there are filter tabs: "Filters:", "Specials", "Haven't Been", "Following", "Price", "Open Now", "Saved", and "Liked".

The main content area is split into two columns. The left column shows search results for hotels in Lausanne:

- 1 Beau-Rivage Palace**
Place du Port 17-19, Lausanne
Hotel
9.6
Very good hotel! Onyx bar, spa, view, service, food - this is the level! Don't hesitate to ask for everything traditional and try it! You'd love it!
- 2 Lausanne Palace Hotel & Spa**
Rue du Grand-Chêne 7, Lausanne
Hotel
9.4
THE 5 star hotel of Lausanne. Great bar and lobby area for drinks, fantastic club sandwiches, expensive but nice spa. Worth a visit.

The right column features a map of the Lausanne area with numerous blue location pins numbered 1 through 28. The map includes labels for "Bois Mermet", "Forêt de Sauvabelin", "CHAILLY", and "Pully". A search bar on the map says "Search this area".

At the bottom of the map, there is a footer with "Privacy Terms © Mapbox © OpenStreetMap Improve this map".

credit: foursquare.com

Foursquare venues

FOURSQUARE I'm looking for... Renens, CH

Beau-Rivage Palace
Hotel
Lausanne

Tips 26 | Photos 368

9.6 / 10 143 ratings

credit: foursquare.com

Example: geo-localized tweets in Switzerland

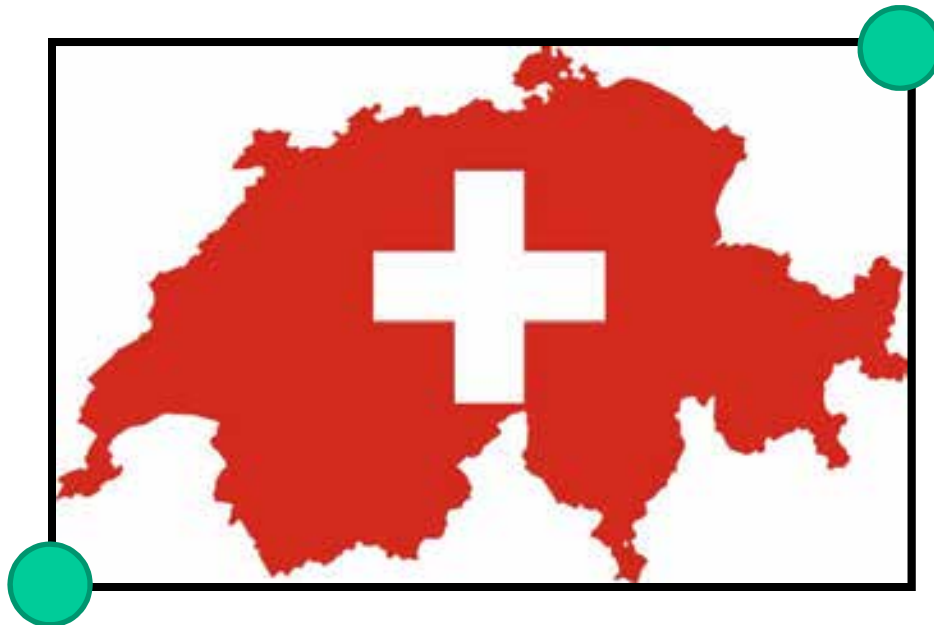
Data collection

Streaming API: bounding box parameters

22 million tweets (Dec 2011- Jan 2015)

614,000 distinct users

Geo-localized Tweets are a **fraction** of all tweets (1-2%)

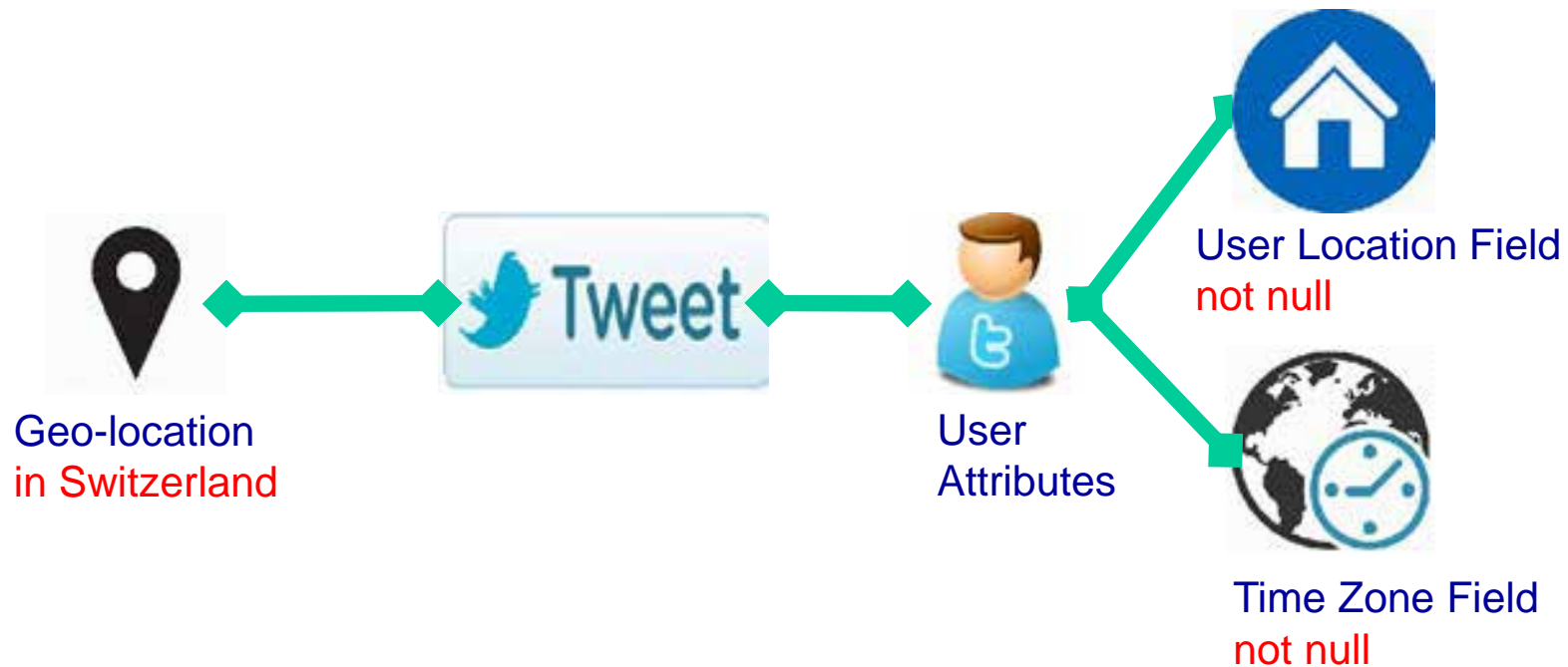


Data Filtering

Keep tweets having geo-location (latitude/longitude) in Switzerland polygon, non-null user location & time zone

4.6 million tweets

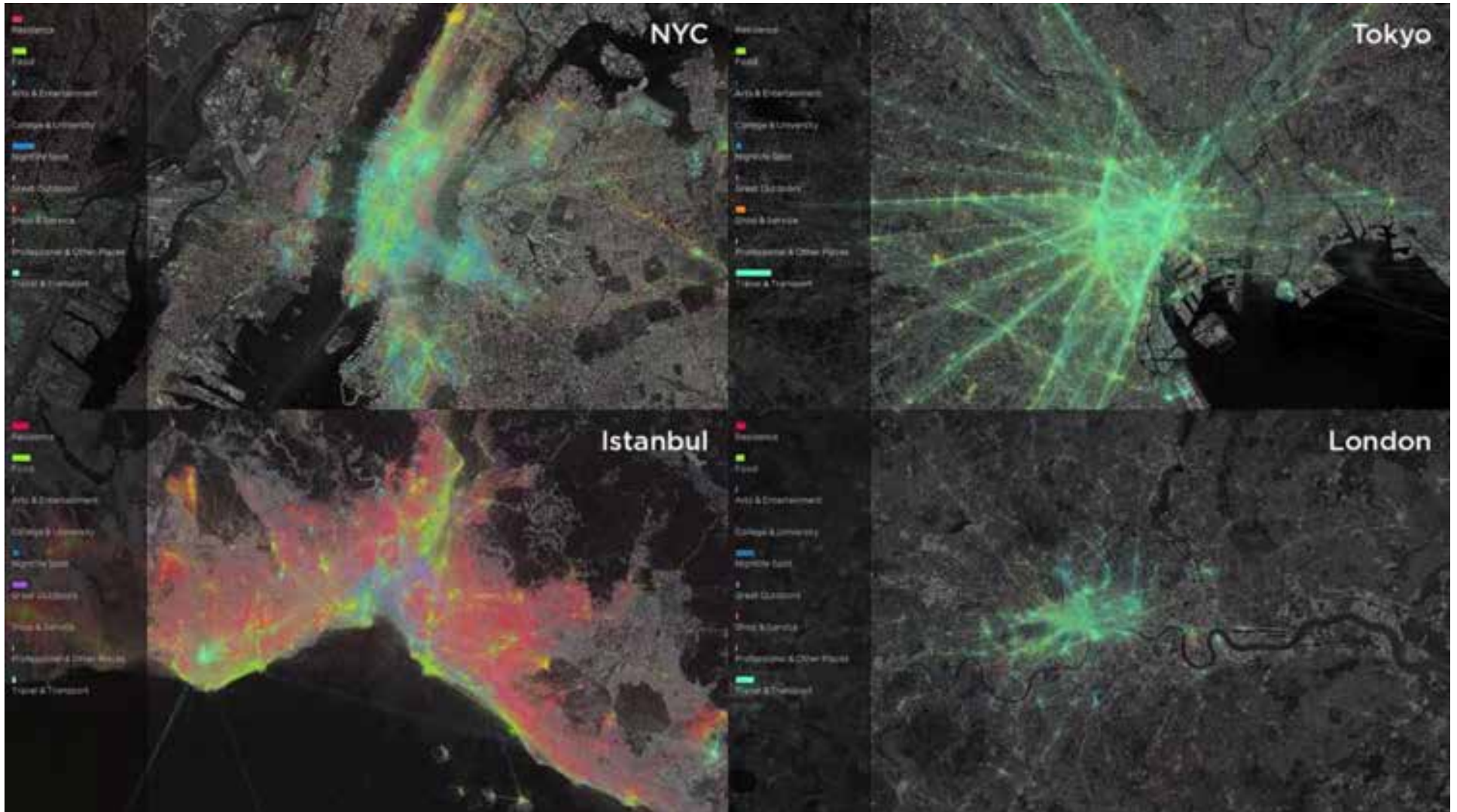
89,000 distinct users



2.

Opportunities & Challenges
Example: Mobility analysis

Mapping mobility in cities



credit: foursquare

<https://vimeo.com/144409527>

Human mobility models 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)

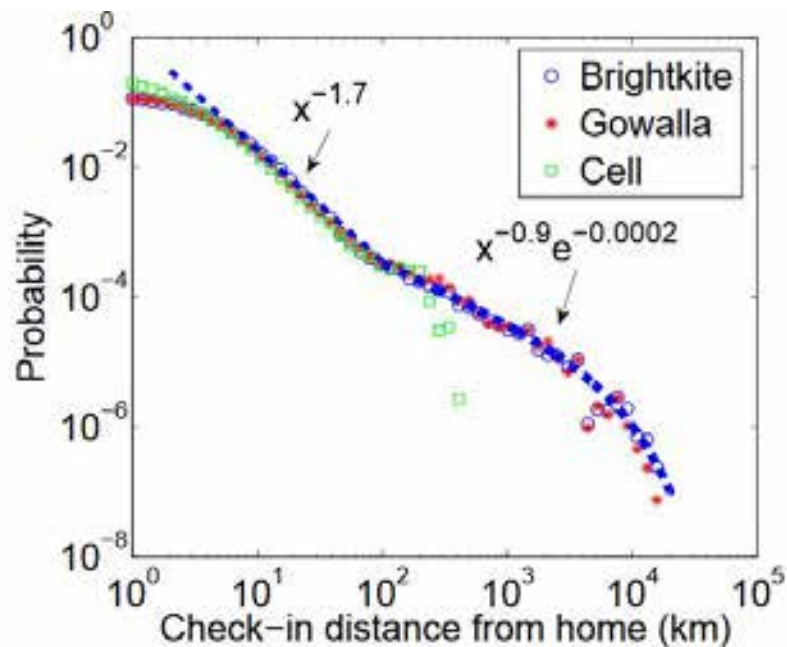


Figure 1: Fraction of check-ins as a function of distance traveled from home. Note the change in slope at around 100km.

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

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

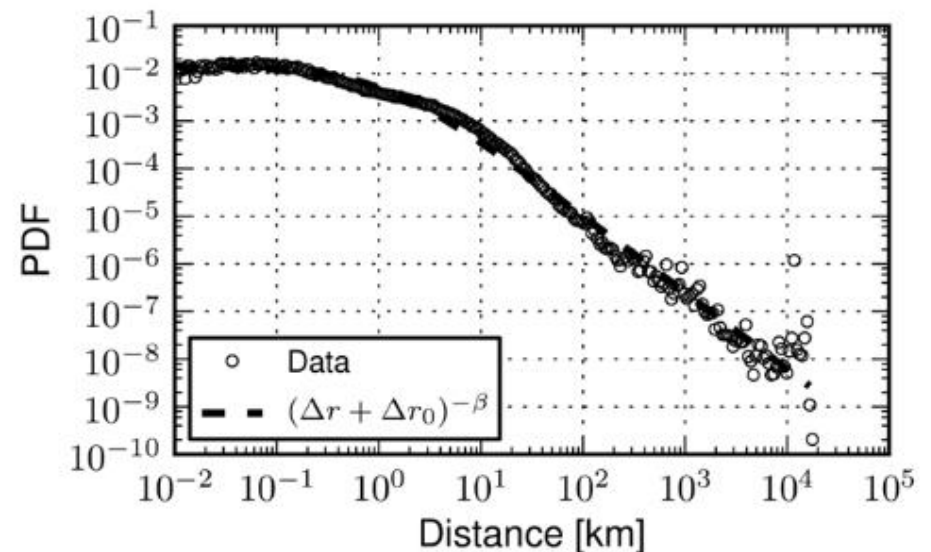
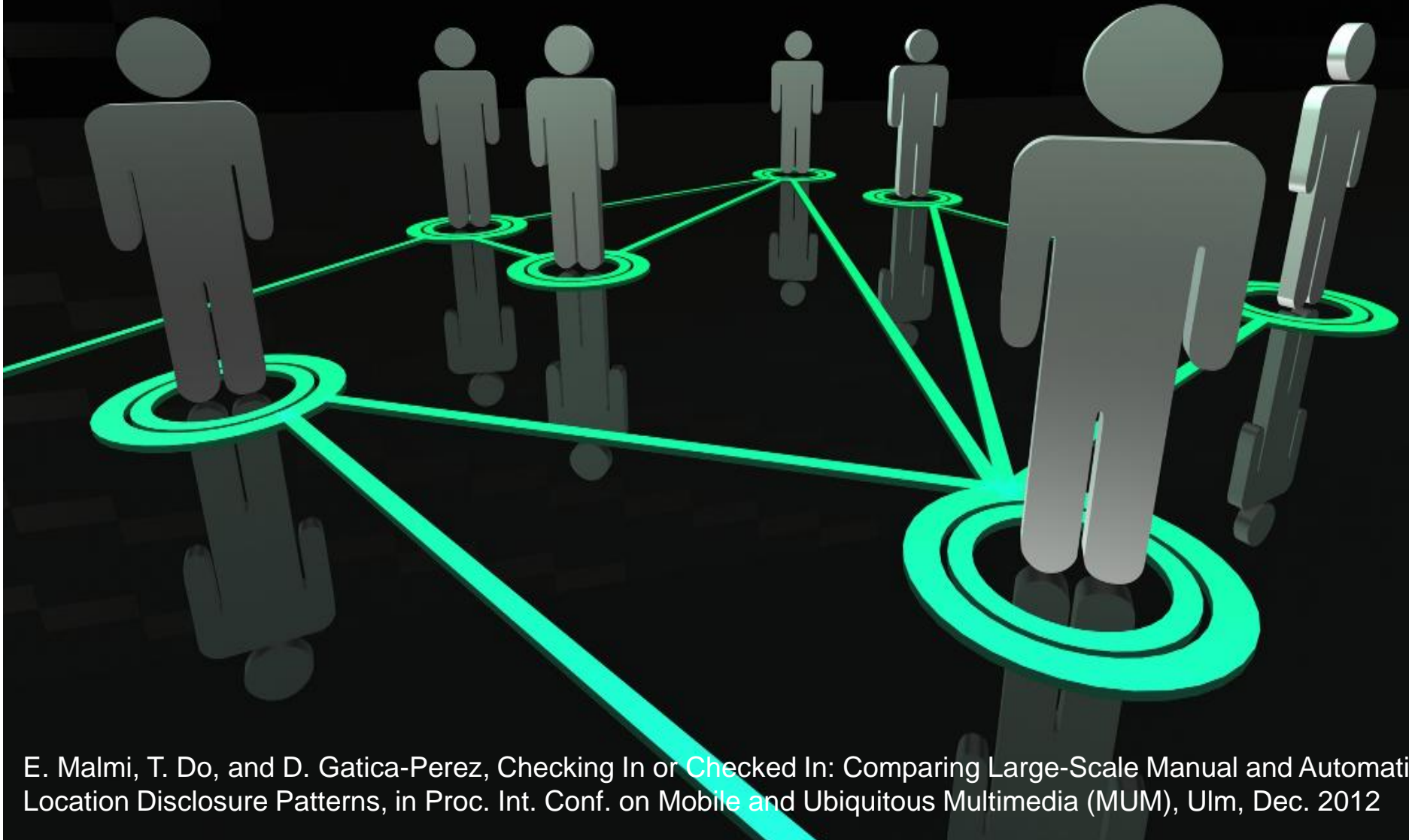


Figure 1. Global movements. The probability density function (PDF) of human displacements as seen through 35 million location broadcasts (check-ins) across the planet. The power-law fit features an exponent $\beta=1.50$ and a threshold $\Delta r_0=2.87$ confirming previous works on human mobility data.

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

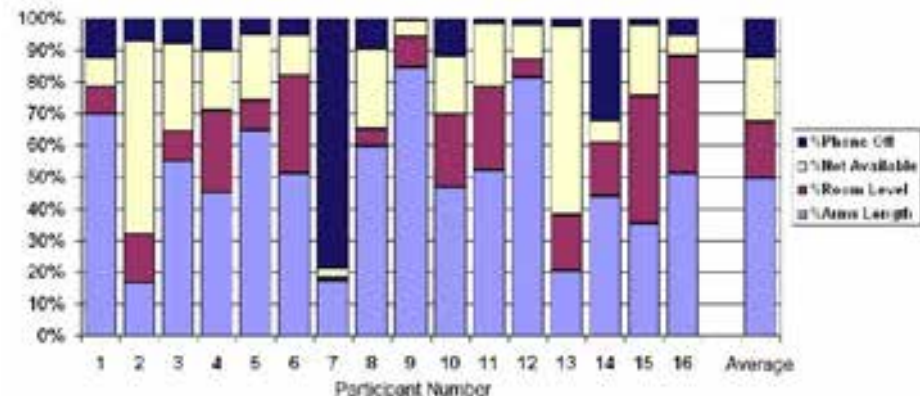
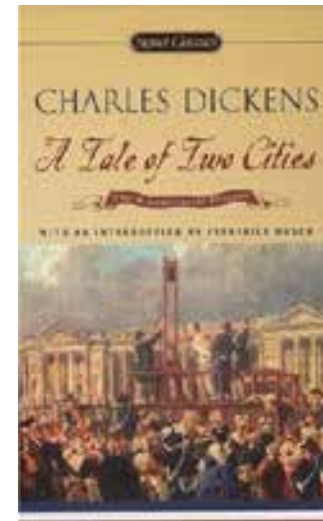
What about the assumptions?



(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 large crowd of people at night, many of whom are holding up their smartphones. The scene is illuminated by warm, yellowish light, possibly from streetlights or event lighting. The people are dressed in winter clothing, including hats and jackets. The overall atmosphere suggests a public event or a large-scale data collection activity.

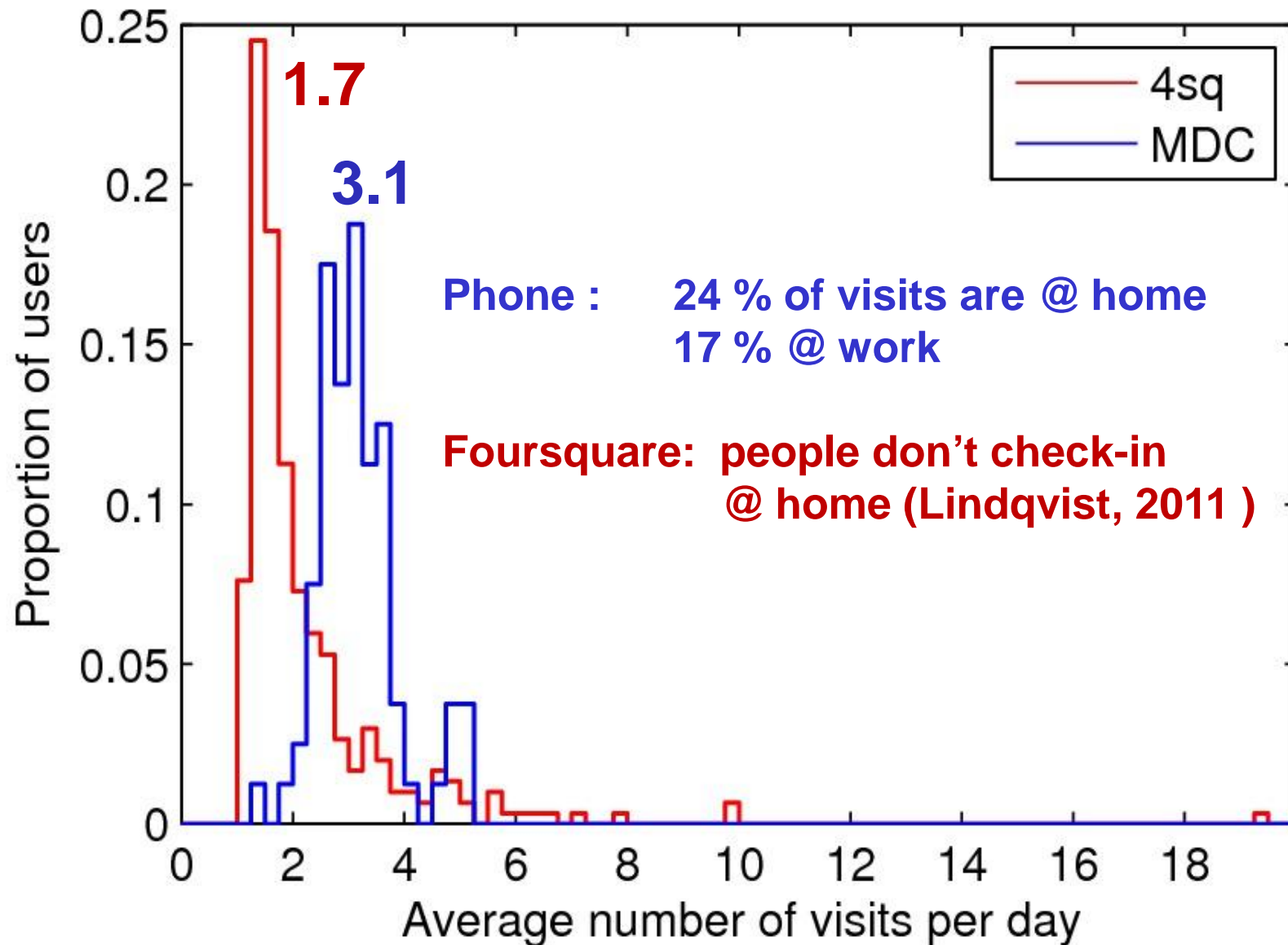
Inferred check-ins from smartphone sensors: Mobile Data Challenge (MDC)

180 volunteers, 12 months, Suisse Romande

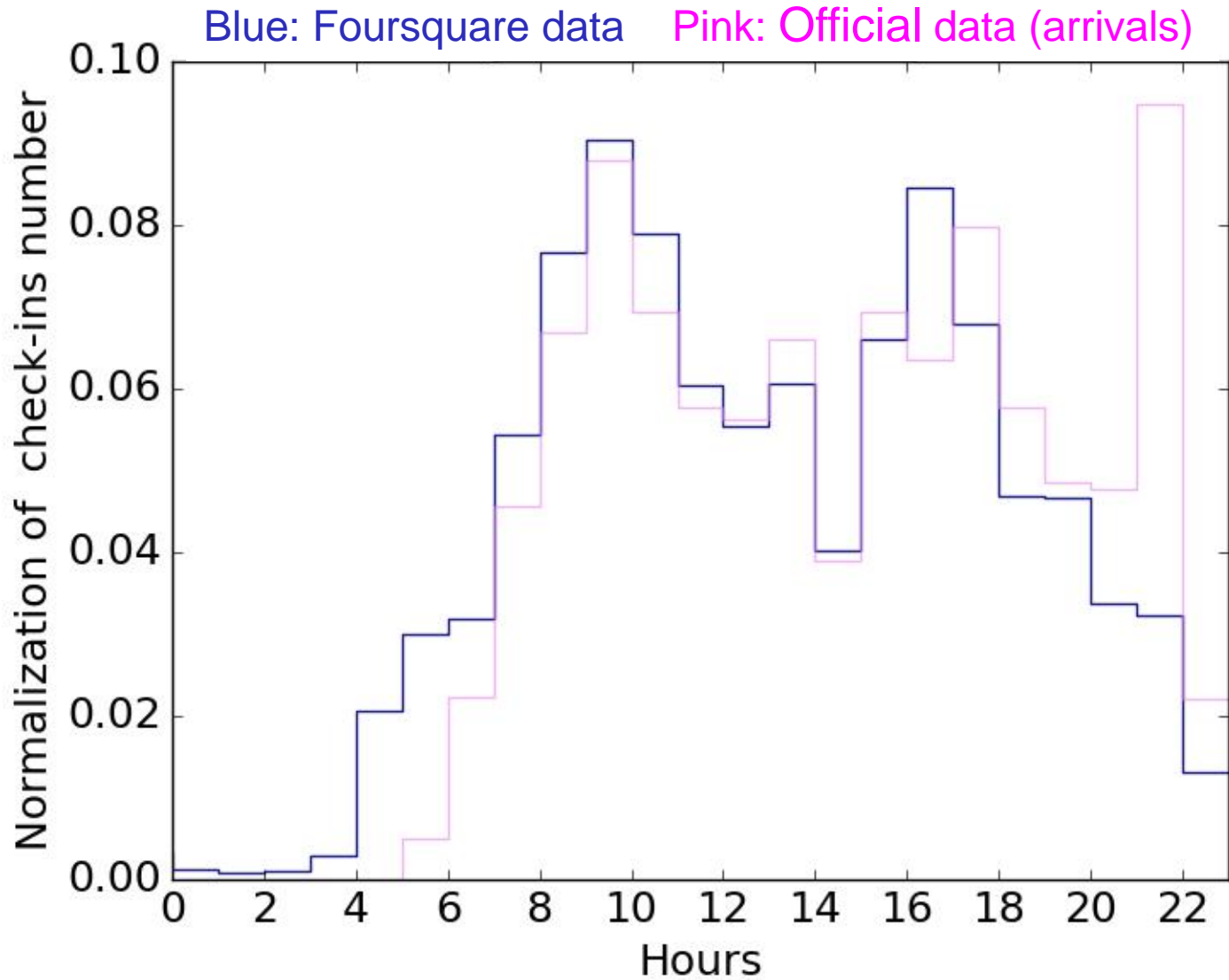
Location and other sensors

credit: [icopythat@flickr](#)

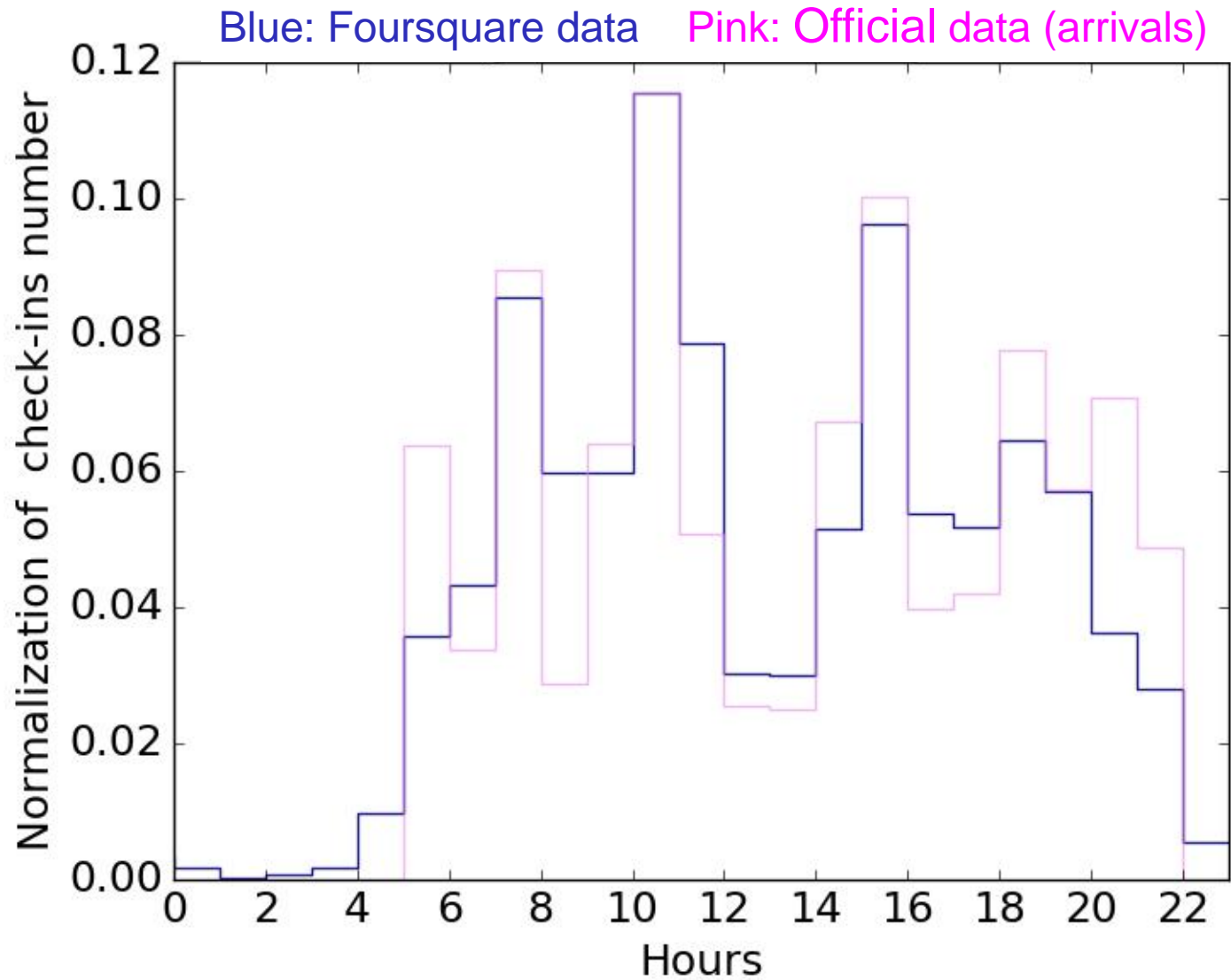
Daily “check-in” distributions: comparing phone sensors and Foursquare



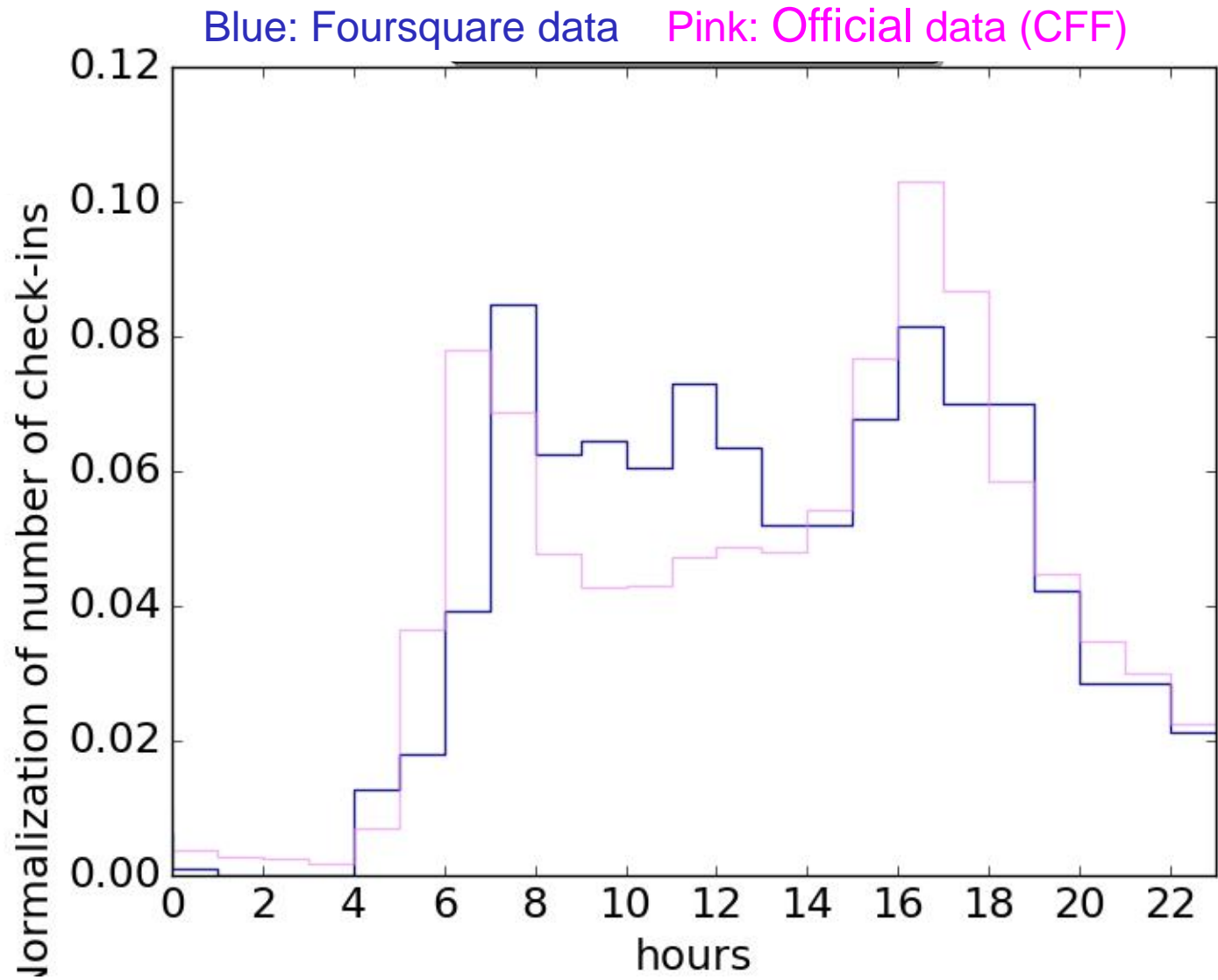
Check-in hourly activity, Geneva airport



Check-in hourly activity, Zurich airport



Check-in activity, Zurich train station



Youth@Night project



200 young volunteers (16-25 yo) documenting Swiss nightlife
Recruitment on the street was done using “social media maps”

D. Santani, J.-I. Biel, F. Labhart, J. Truong, S. Landolt, E. Kunstche, and D. Gatica-Perez
The Night is Young: Urban Crowdsourcing of Nightlife Patterns
in Proc. ACM Int. Joint Conf. on Pervasive and Ubiquitous Computing (UbiComp), Sep. 2016

Urban biases in geo-localized social media

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

Social media research has focused on urban areas
More data produced in cities

**Study: aggregate geo-localized social media at
county level & compare with US Census**

**Systematic bias towards urban vs. rural areas +
5 times more** Tweets per capita

Conclusion

Geo-localized social data

Informative source of certain phenomena

Increasingly used in social media research

Challenges: biases, privacy implications

Social media research needs official statistics

Understand quality of social data

Assess validity of findings

Integrate multiple sources

Could statistical offices benefit from social data?

Thank you

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

