The Human Mobility Data Landscape

Call Detail Records (CDRs)

NBC/Instagram

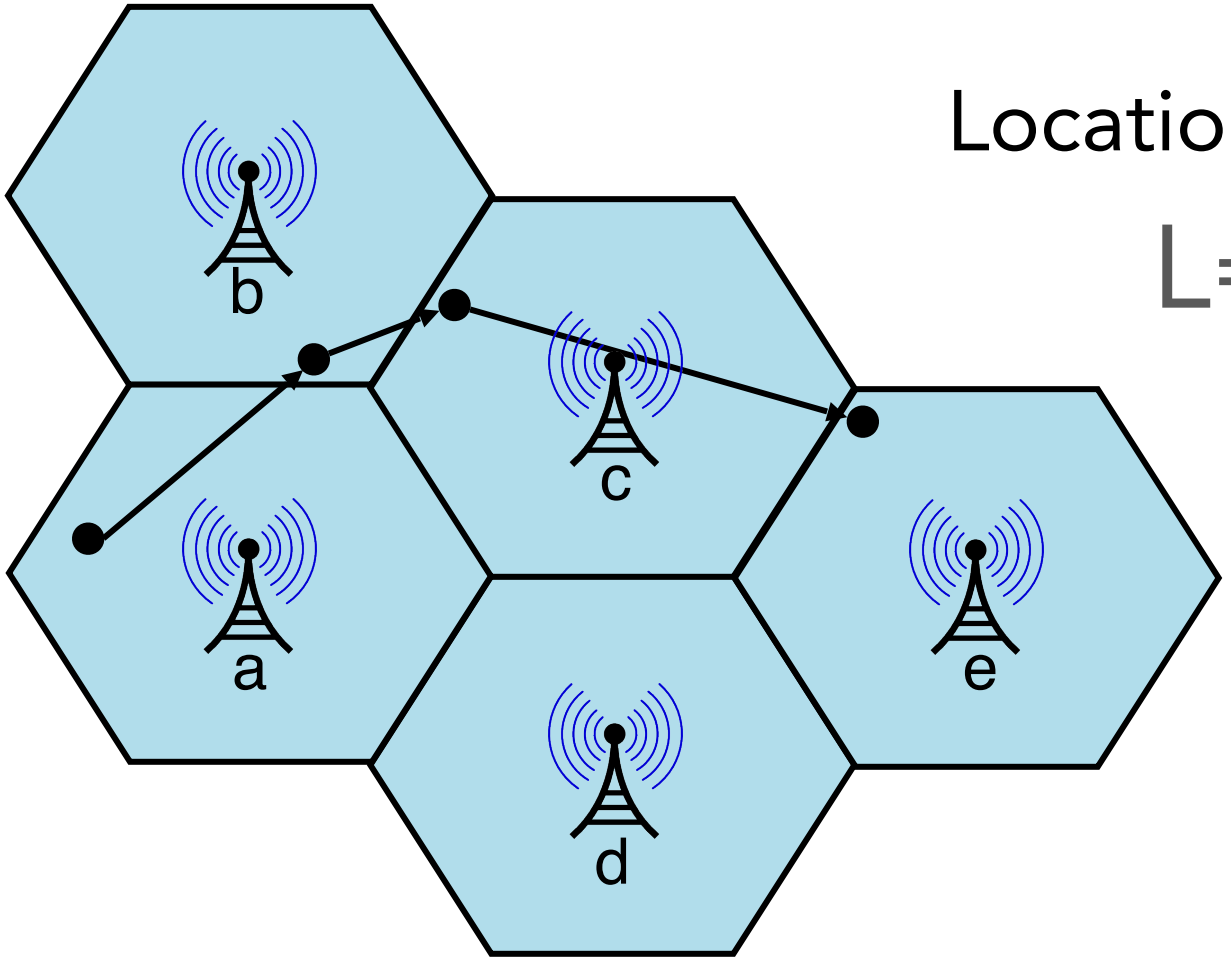
Call Detail Records (CDRs)

- Produced by a phone carrier to store **details of calls** passing through a device

- They contain various attributes of the call:

timestamp	source	destination	tower	duration
10/04/2008 10:12:00	A8563XY	C1123YT	36256	10:32
10/04/2008 10:12:32	B6766CS	H2223IU	43860	03:07
...	...	...	...	...

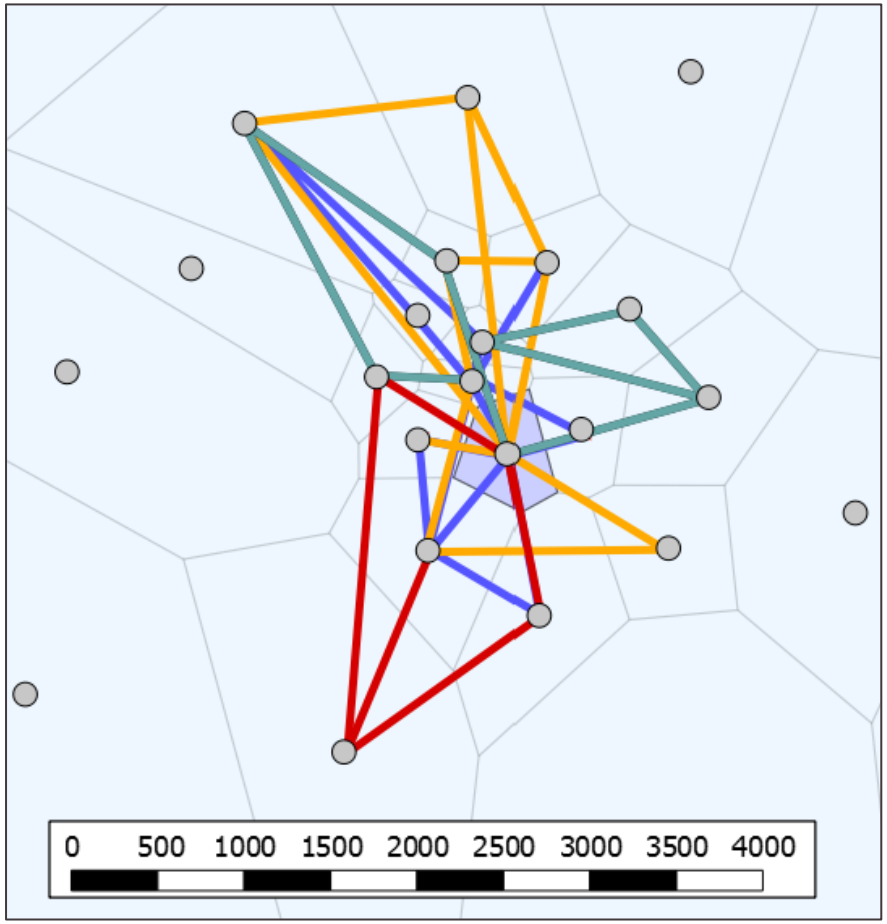
Call Detail Records (CDRs)



Location history

L=abce

Inferring mobility trajectories from CDRs



- **Call**
  Location/tower
- **Trajectory**
  Sequence of locations
- **Home**
  Most nighttime calls

Pappalardo et al., An analytical framework to nowcast well-being with mobile phone data, JDSA, 2016.

PROs and CONs of CDRs

PROs

- mobile phones are **ubiquitous**
- **large** (huge) sample **size**
- **rich** and **multidimensional**
  (social, mobile, time, demographics)

CONs

- **not publicly available**
- position is **partially detected**
  (only when calls are made)
- position is known at **tower level** only
  (resolution is variable from m to km)
- ping-pong effect can create **noise**

Preprocessing on CDR data

- Non-relevant locations are discarded (e.g., if $n/N < 0.005$)
- Users calling from just one locations are discarded (they do not move!)
- Users with too low or high calling activity are discarded (e.g. by imposing min and max average calls per day)

Available data extracted from CDRs

- A multi-source dataset of urban life in the city of Milan and the Province of Trentino

Official Repository

Kaggle Repository

- D4D Challenge - available upon request

References for CDRs

- Understanding individual human mobility patterns *(*Gonzalez et al., Nature, 2008).

- Human mobility prediction based on individual and collective geographical preferences.
(Calabrese et al., IEEE ITSC, 2010).

- Friendship and mobility: user movement in location-based social networks.
(Cho et al., ACM SIGKDD, 2011).

- A survey of results on mobile phone datasets analysis (Blondel et al., EPJ DS, 2015).

Location Based Social Networks (LBSN)

Social Network + Location Data

Social Networks

A social network is a social structure made up of individuals connected by one or more specific types of interdependency.

Location Data

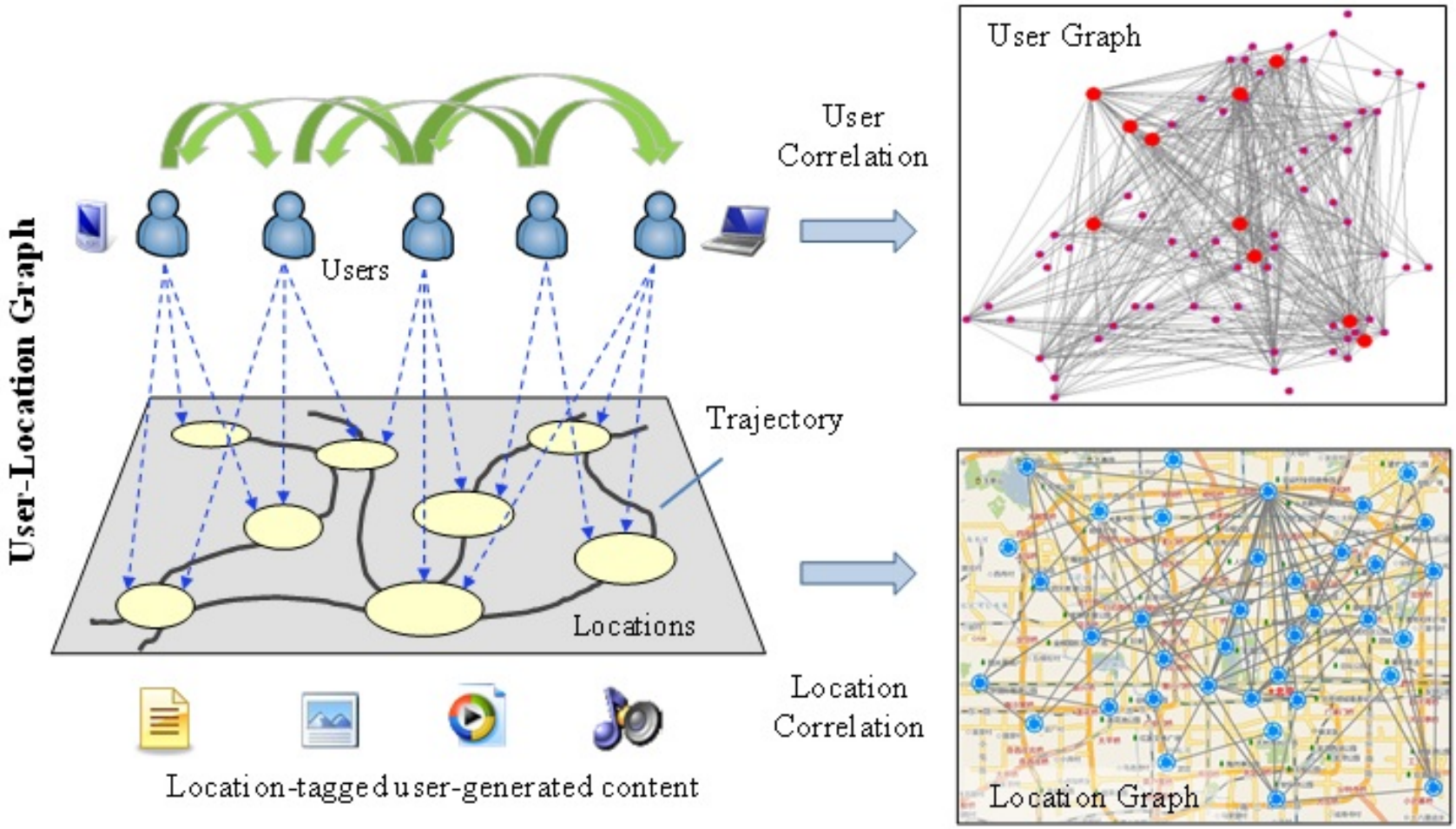
A location can be represented in absolute (latitude-longitude coordinates), relative (100 meters north of the Space Needle), and symbolic (home, office, or shopping mall) form.

Location-Based Social Network (LBSN)

A new social structure made up of **individuals** connected by the interdependency derived from their **locations** in the physical world as well as their location-tagged media content.

Location-Based Social Networks: Users
(Y. Zheng, Computing with Spatial Trajectories. Springer.)

Location-Based Social Network (LBSN)



PROs and CONs of LBSNs

PROs

- often **publicly available**
- **objective** location and **semantic** information (e.g., restaurant, mall, etc.)

CONs

- movement is **partially detected** (only when checkins are made)
- data **sparsity** (more sparse than CDRs)
- position is known at **location level** only
- **self-selection bias**

References for LBSNs

- Location-Based Social Networks: Users (Y. Zheng, Computing with Spatial Trajectories. Springer.)
- Mobility and geo-social networks (Spinsanti et al., Mobility Data, Cambridge press, 2013).
- A tale of many cities: Universal Patterns in Human Urban Mobility (Noulas et al., PLoS One, 2012).
- Understanding Human Mobility from Twitter (R. Jurdak, PLoS One, 2015).

Public datasets

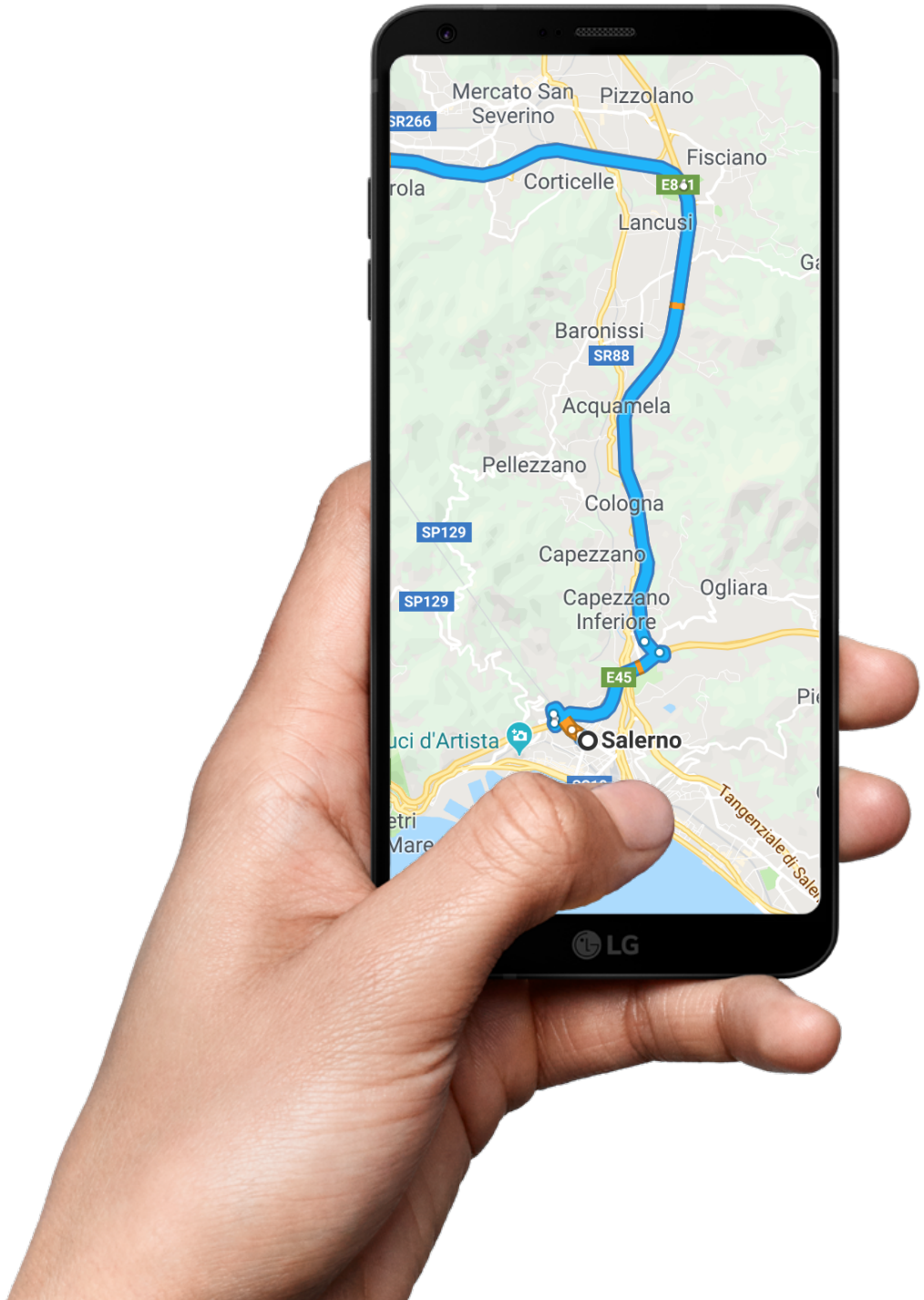
- YFCC100M: geotagged photos from Flickr

- Brightkite: checkins in different countries

Public datasets

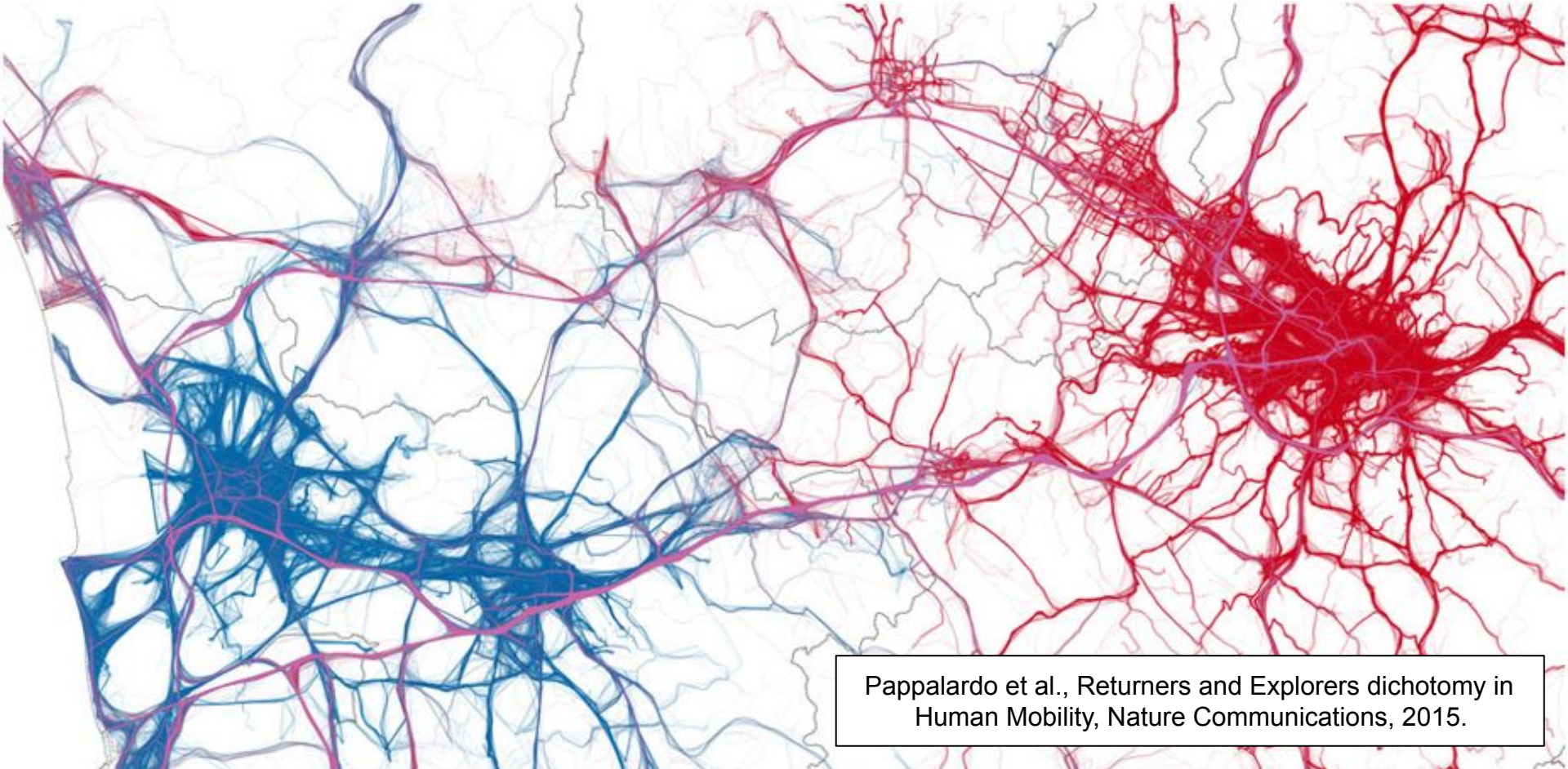
- Foursquare: checkins in different countries

GPS data


Mercato San Severino
Pizzolano
SR266
Fisciano
Corticelle
E841
Lancusi
Baronissi
SR88
Acquamela
Pellezzano
Cologna
SP129
Capezzano
SP129
Capezzano Inferiore
Ogliara
E45
Salerno
Tangenziale di Salerno
LG

GPS data

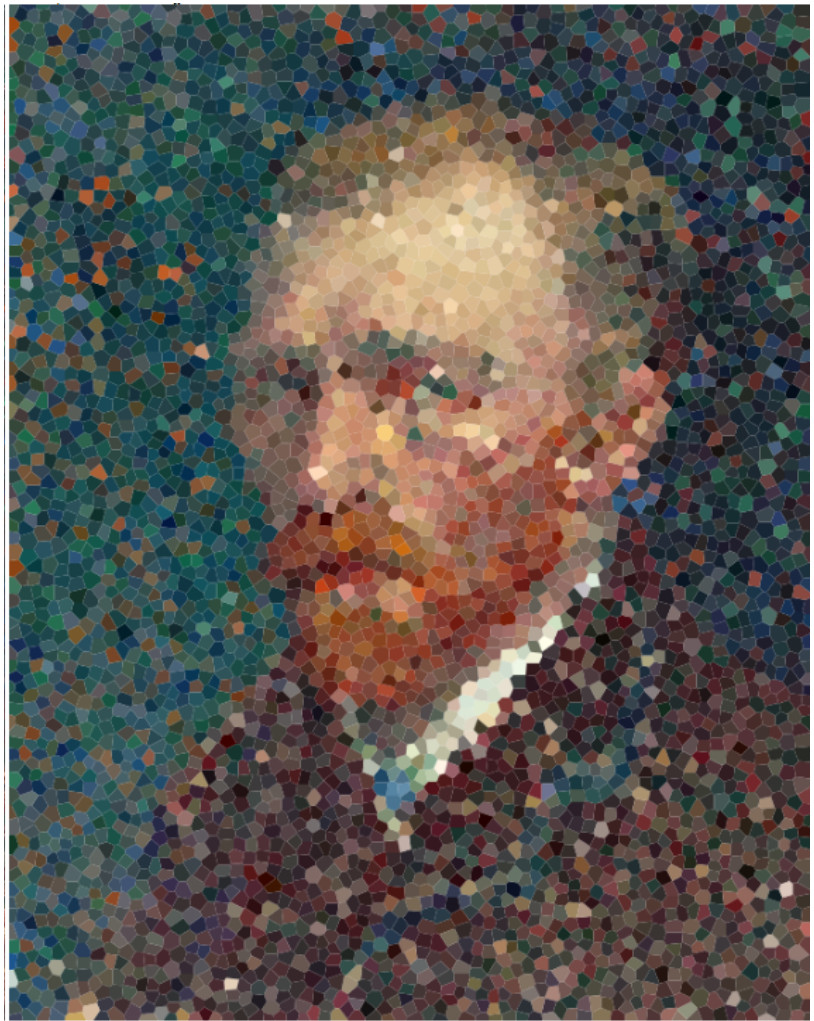
Produced by **GPS devices** embedded in *smartphones*, *private vehicles*, *boats*, etc.



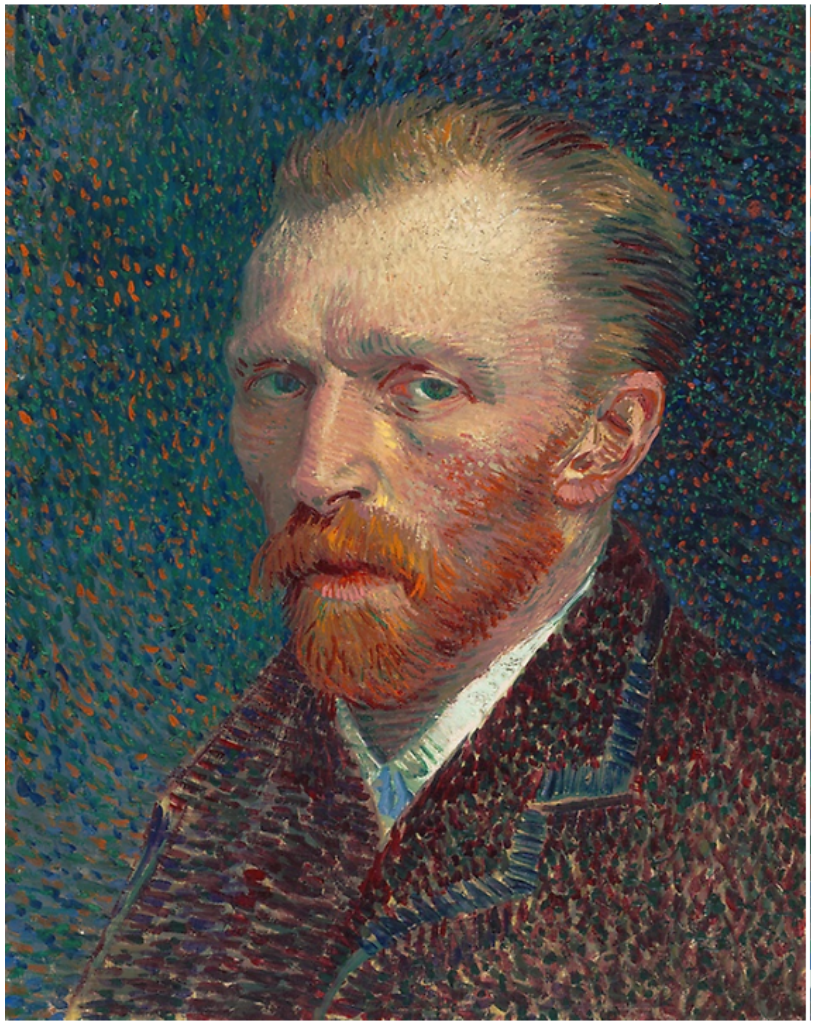
Pappalardo et al., Returners and Explorers dichotomy in Human Mobility, Nature Communications, 2015.

GPS data vs CDRs

CDRs



GPS



PROs and CONs of GPS data

PROs

- trajectory at **high** spatial and temporal **resolution**
- track the **full trajectory** (e.g., all cars movements)

CONs

- **no objective** definition of location (preprocessing is needed)
- **errors** when signal is noisy

Preprocessing on GPS data

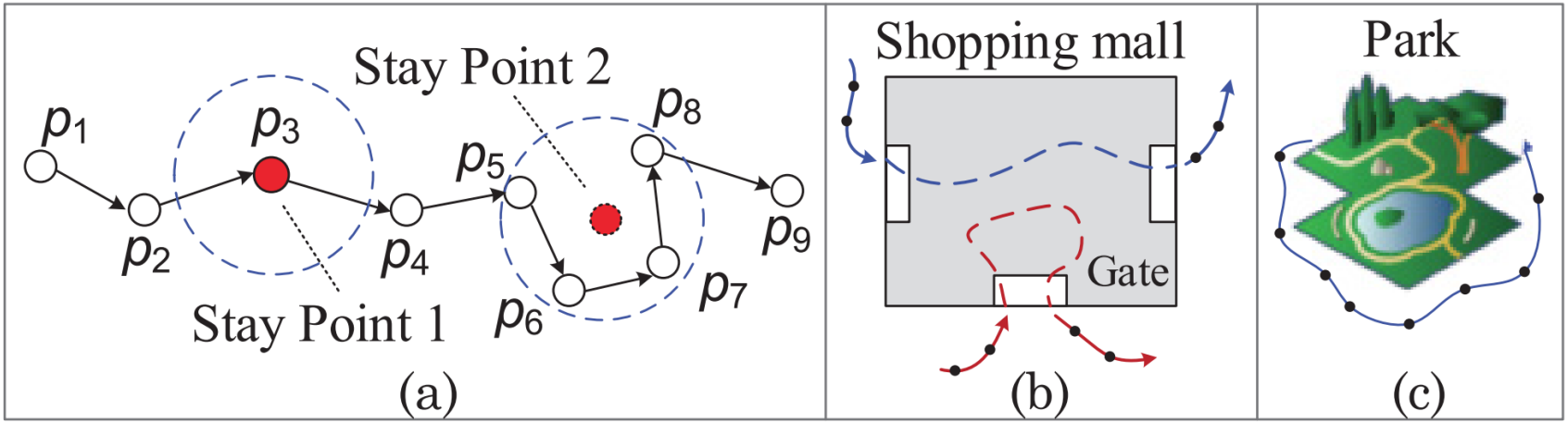
Some common steps to apply on GPS data:

- **location detection**
  group points into one meaningful location

- **trajectory segmentation**
  split a trajectory in sub-trajectories

Location detection in GPS data

Locations (or stay points) denote places where people have stayed for a while



Zheng, Trajectory Data Mining: an overview, ACM TIST, 2015.

$$P = p_1 \rightarrow p_2 \rightarrow \cdots \rightarrow p_n \implies$$

$$S = s_1 \rightarrow s_2 \rightarrow \cdots \rightarrow s_n$$

Stay Point detection algorithm

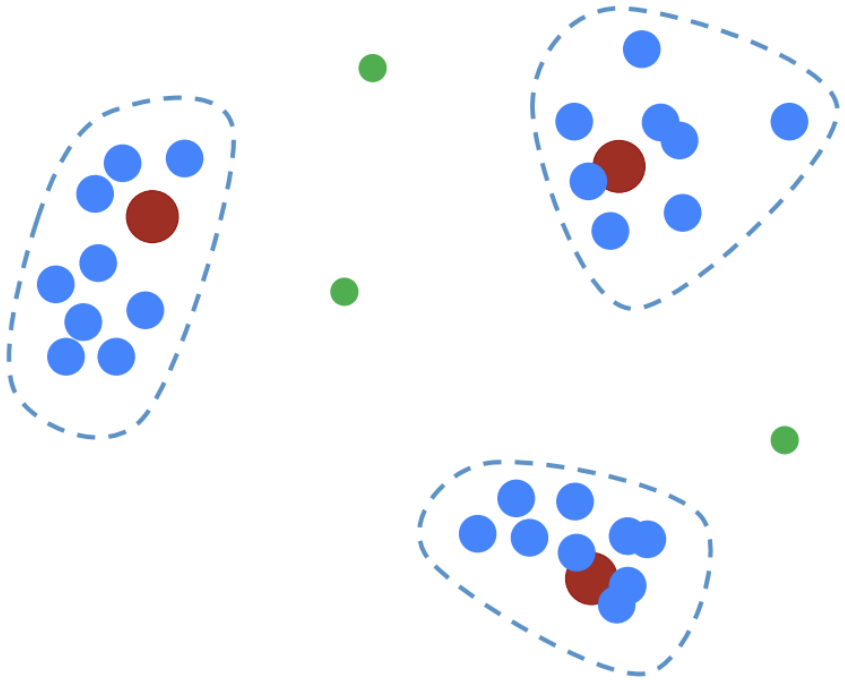
The points are scanned in temporal order, then if the following condition holds:

$$d(p_{i+1}, p_i) < \delta \wedge t(p_{i+1}, p_i) < \Delta$$

the points belong to the same stay point

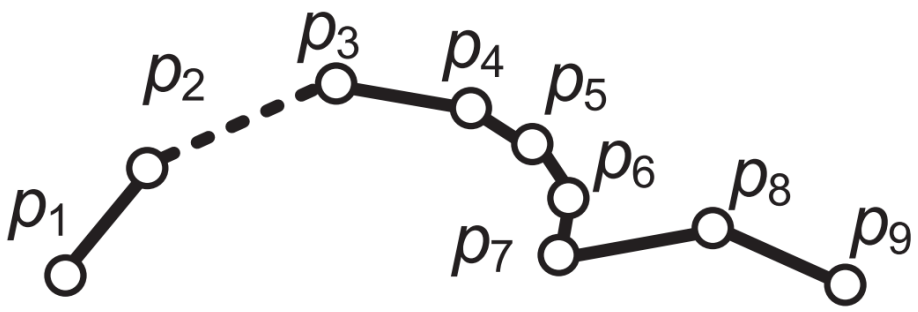
Semantic Location detection

Density clustering algorithm, like DBSCAN or OPTICS (time is not considered)

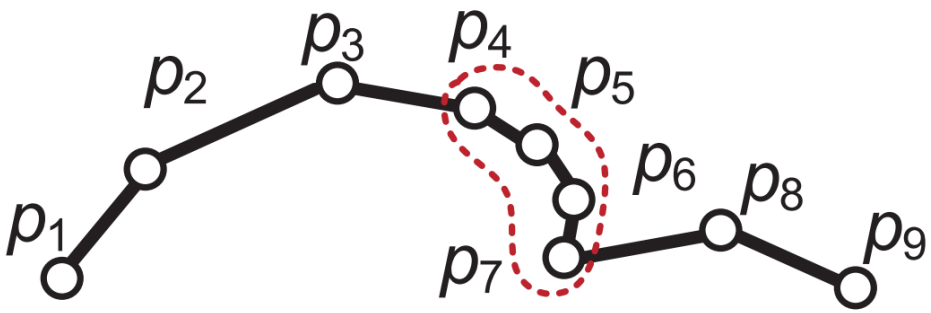


Trajectory segmentation

a trajectory is split into two or more sub-trajectories, with several techniques:



time-interval



stay-point

Zheng, Trajectory Data Mining: an overview, ACM TIST, 2015.

Publicly available GPSs datasets

- Geolife GPS trajectory dataset

Official Repository

- Taxi trajectory dataset
  (warning: not individual)

References for GPS data

- Understanding the patterns of car travel (Pappalardo et al., EPJ ST, 2013).
- Learning travel recommendations from user-generated GPS traces (Zheng et al., ACM TIST, 2011).
- Trajectory Data Mining: an Overview (Zheng, ACM TIST, 2015.)