

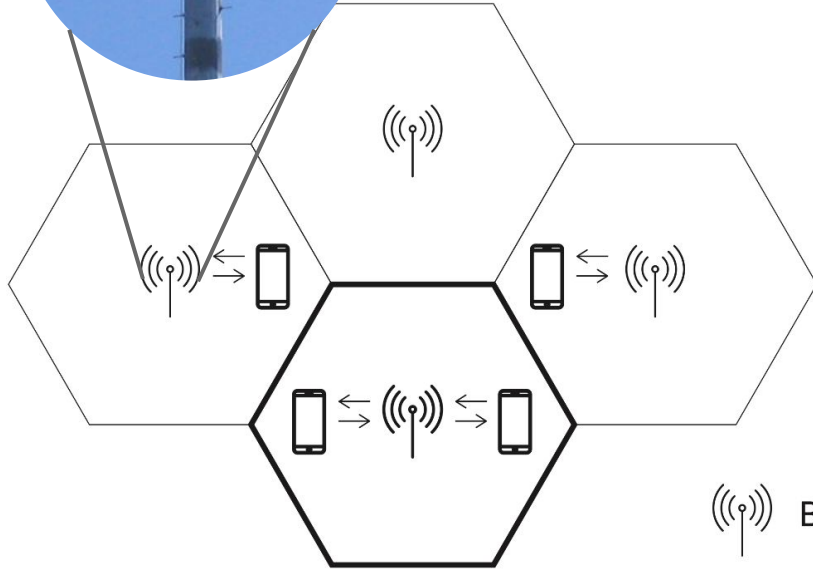
Mobile Phone Records



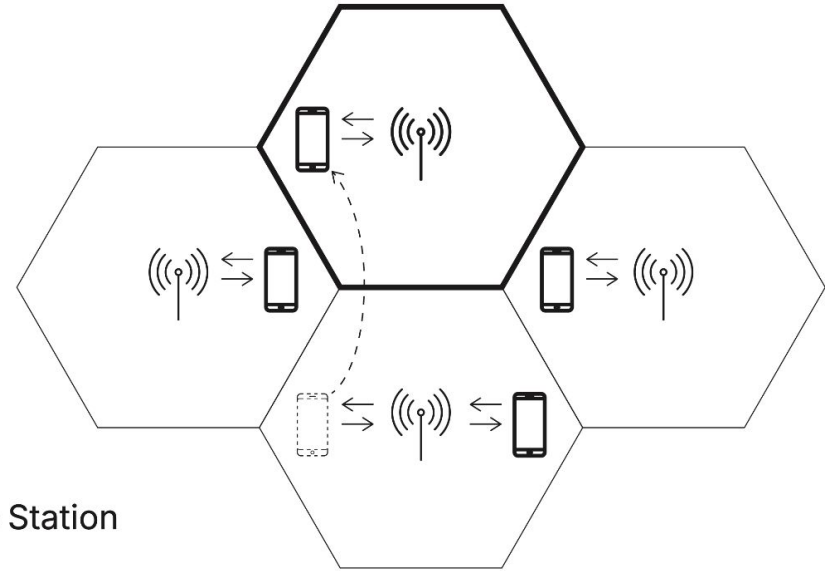




ular coverage



(b) handover event



Mobile Phone Network

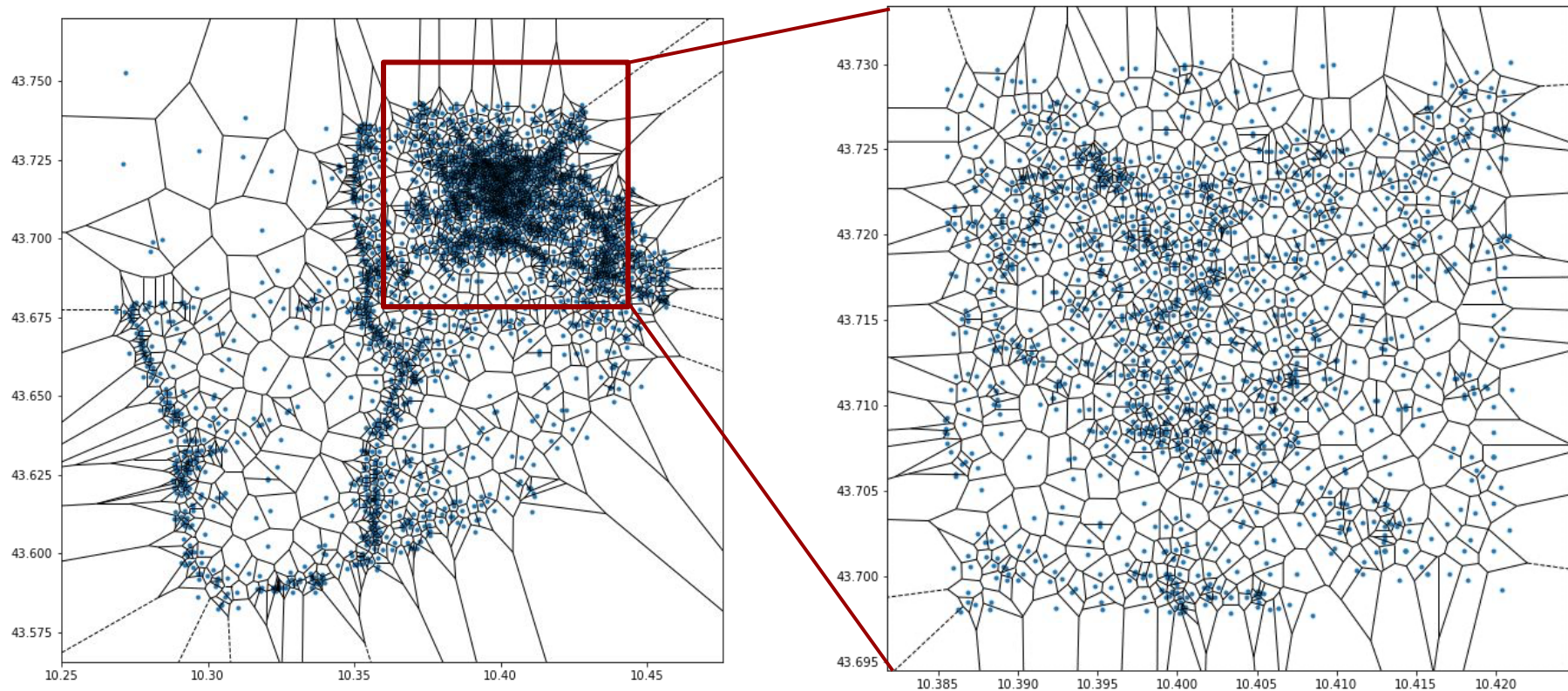
- Used by mobile phone operators for coverage and capacity
- Large areas split into smaller cells to support more users
- Range of cells:
 - cities: up to 1 km
 - rural areas: up to 8 km
 - unpopulated areas: up to 40 km



Voronoi tessellation generated from a subset of mobile phone antennas from the OpenCellID database, used as seed points.

Voronoi cells are smaller in densely covered urban areas, such as the city centre, and larger in less densely covered peripheral areas.

Voronoi tessellation of GSM cells in Pisa

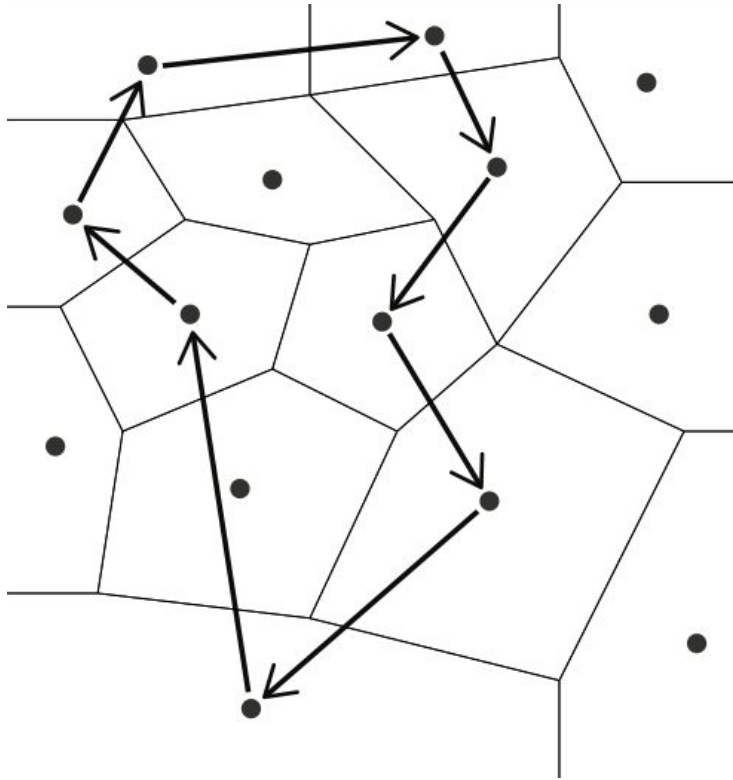


Exercise

How can we exploit the mobile phone network to describe the mobility of mobile phone users?

No Internet nor AI.

10 minutes

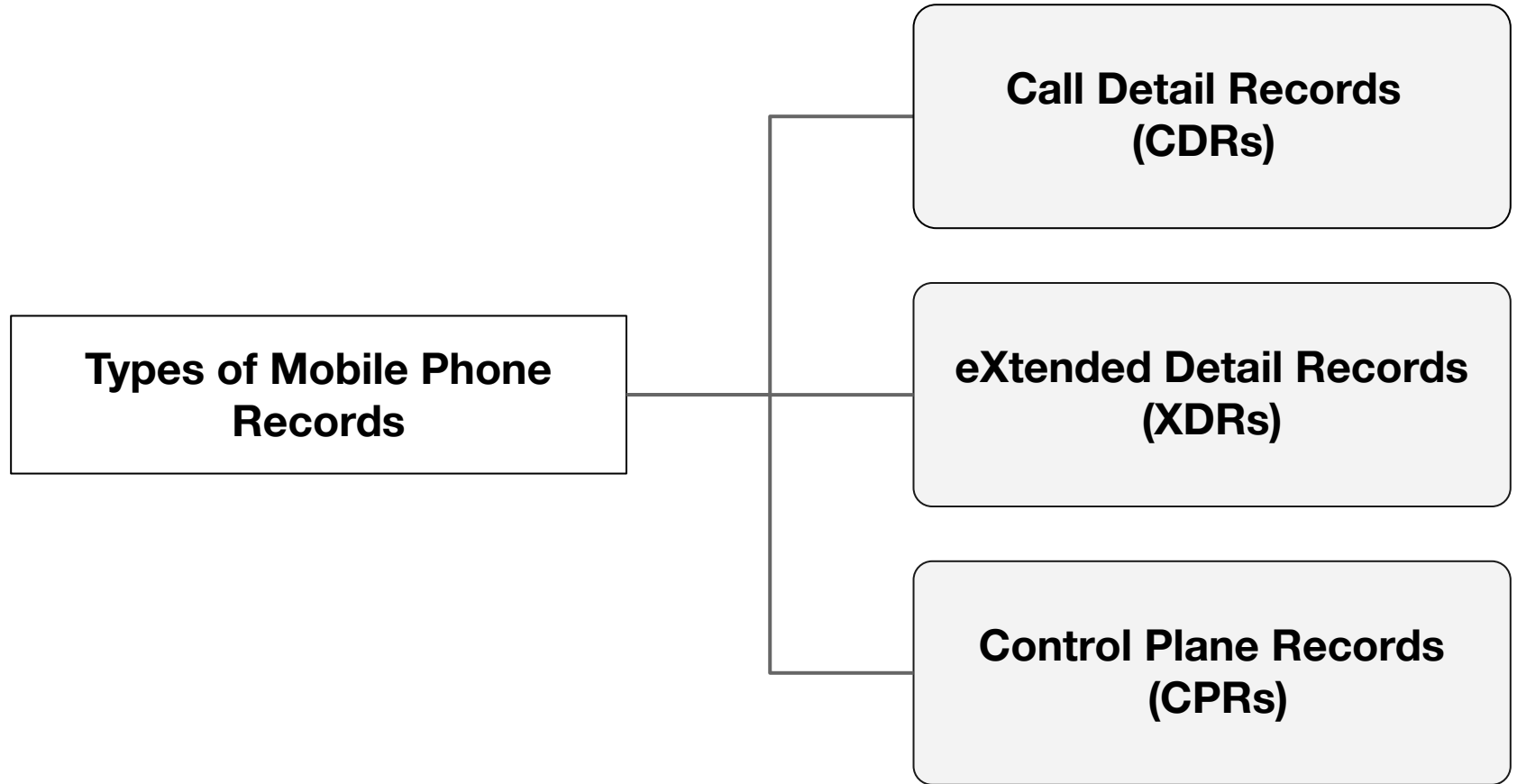


The mobility trajectory of a user extracted from mobile phone records.

The phone towers are shown as grey dots, and the Voronoi tessellation approximates each tower's reception area.

Mobile Phone Records

- Data records that documents details of:
 - calls, SMS, data exchange, or other events
- Attributes:
 - time, duration, completion status, source number, destination number, amount of data, etc.
- Automatically recorded by mobile phone providers (carriers)



Call Detail Records (CDRs)

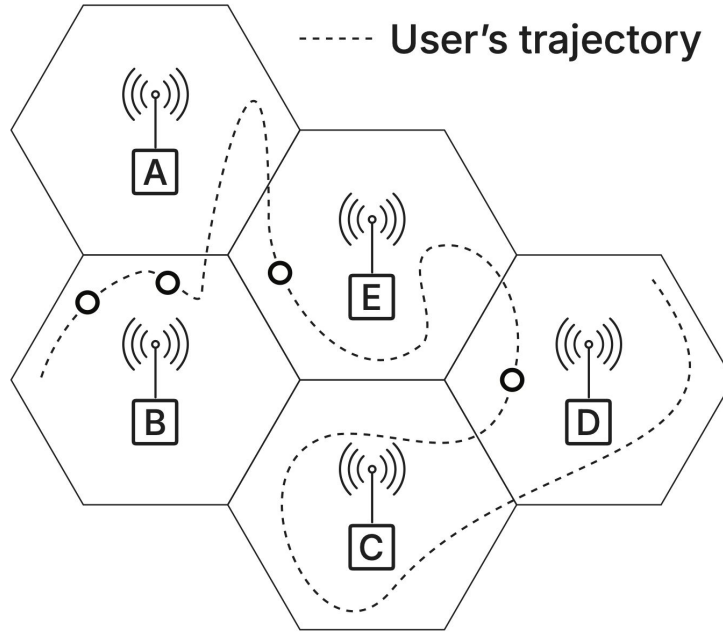
- collected for billing purposes
- purely **human-triggered**, thus sparse
- record created at every call/SMS

$$(n_o, n_i, t, A_o, A_i, d)$$

caller callee time caller antenna callee antenna duration

○ Call

----- User's trajectory



(BBED)

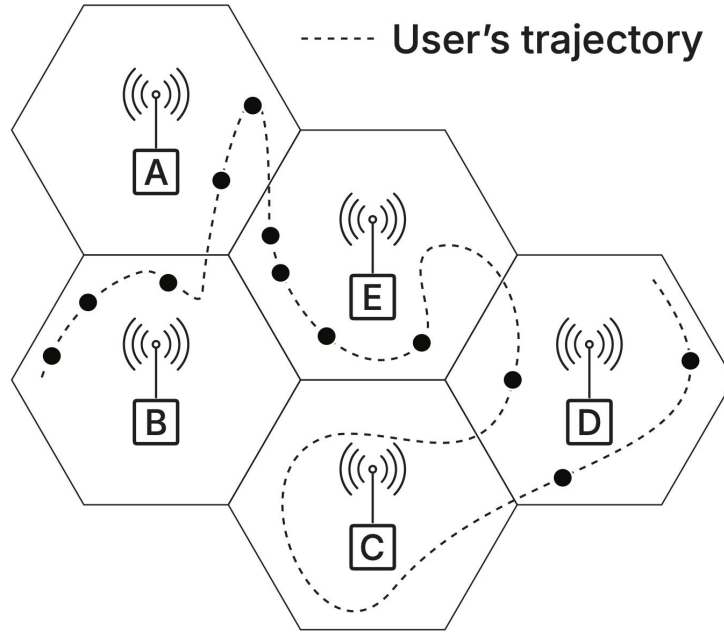
eXtended Detail Records (XDRs)

- collected for billing purposes
- mixture of **human- and device-triggered**
- record created when down- or up-loading internet content

$$(n, t, A, k)$$

user time antenna KB

- Download and Upload



(BBBAAEEEEEDDD)

Control Plane Records (CPRs)

- used to check network “health”
- **network-triggered**
- record is created, e.g., when a device is attached to a new antenna (“handover”)

$$(n, t, A, e_1, e_2, \dots, e_n)$$

user time antenna  events

o Event 1

□ Event 2

Δ Event 3

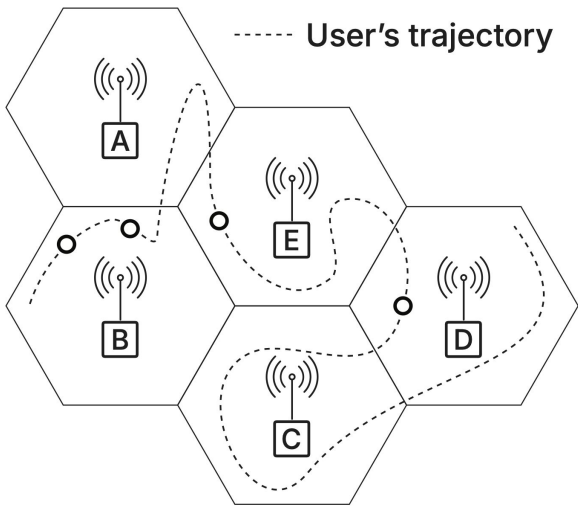


(BBBBAAEEEEDDCCDD)

CDRs

○ Call

----- User's trajectory

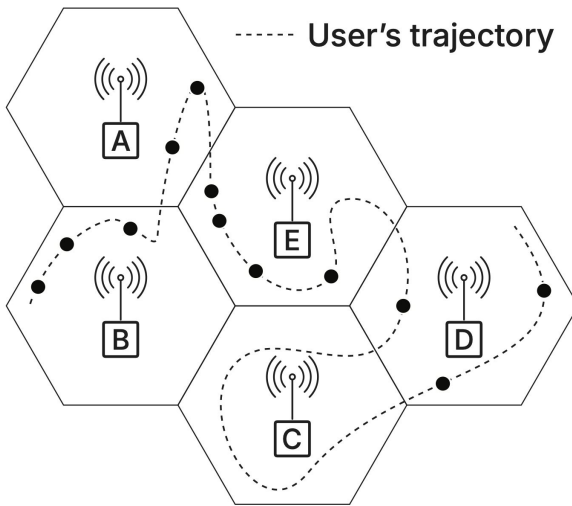


(BBED)

XDRs

● Download and Upload

----- User's trajectory



(BBBAAEEEEEDDD)

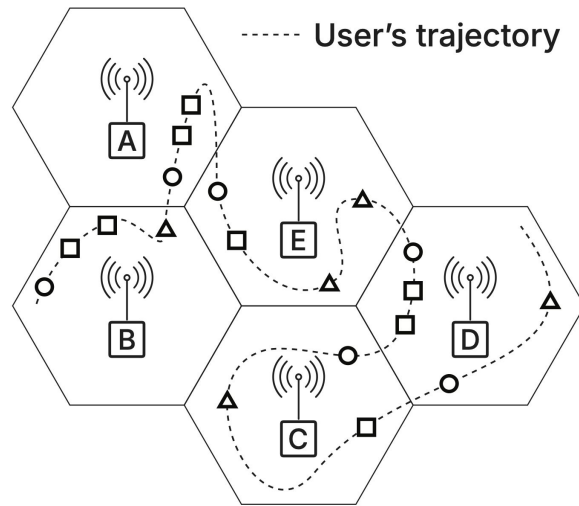
CPRs

○ Event 1

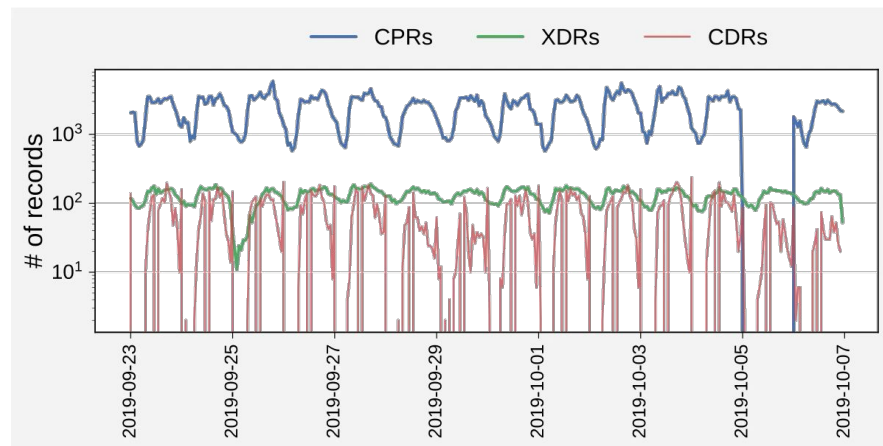
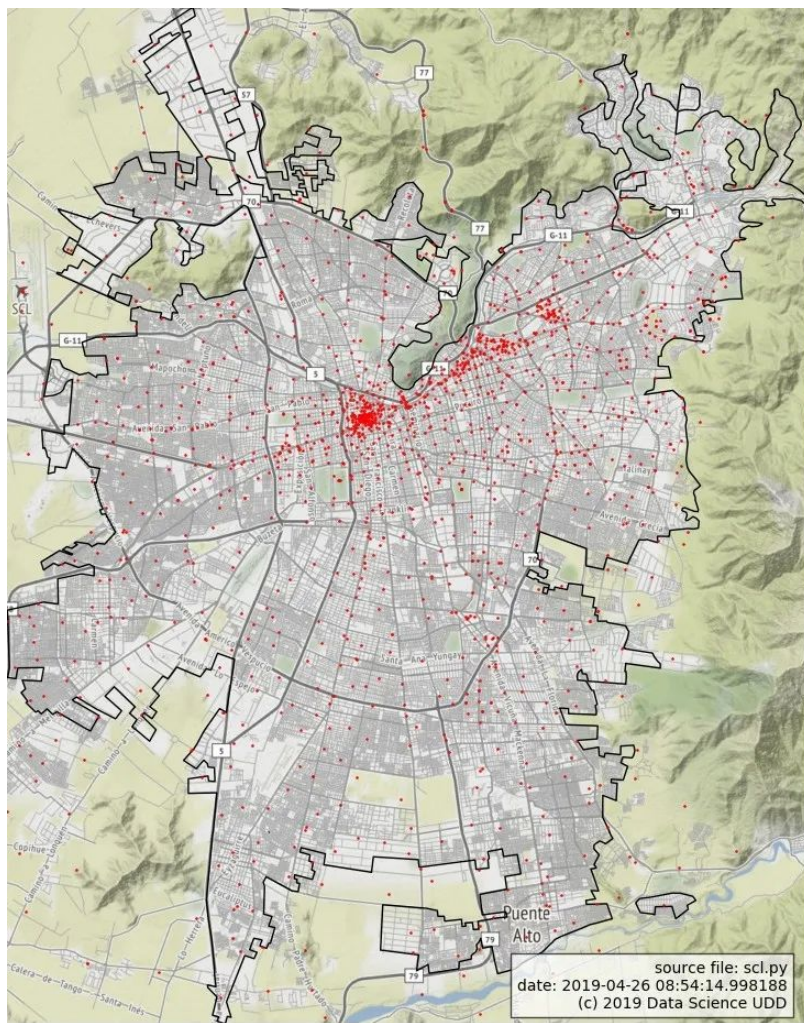
□ Event 2

△ Event 3

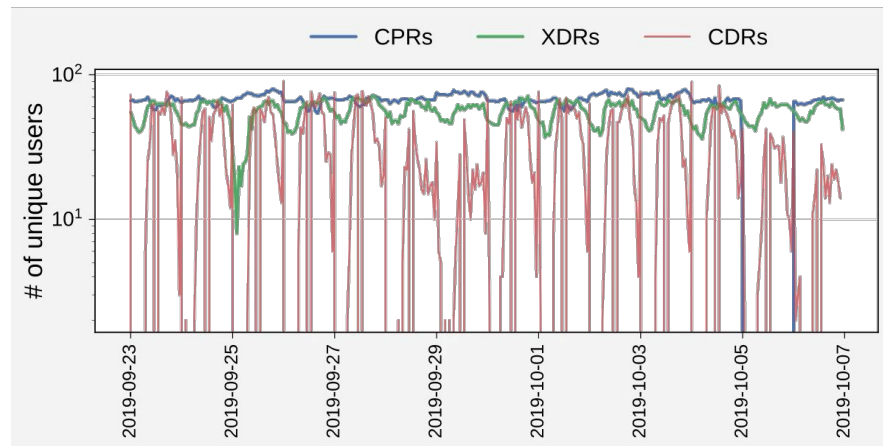
----- User's trajectory



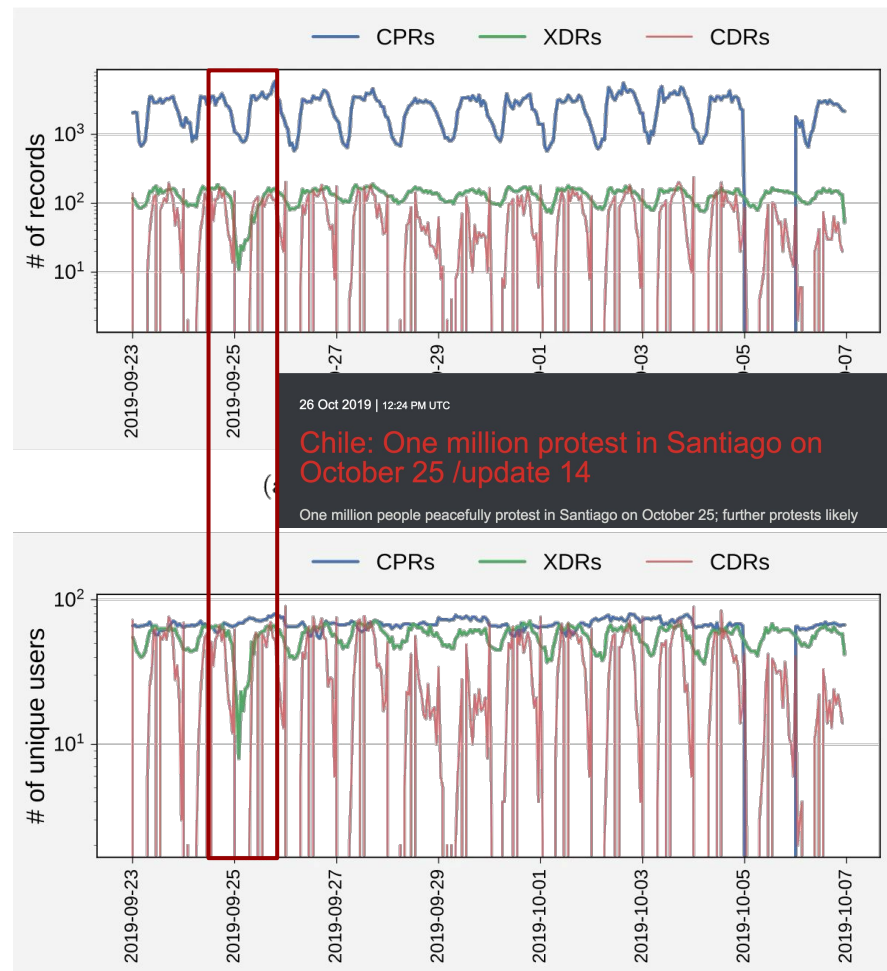
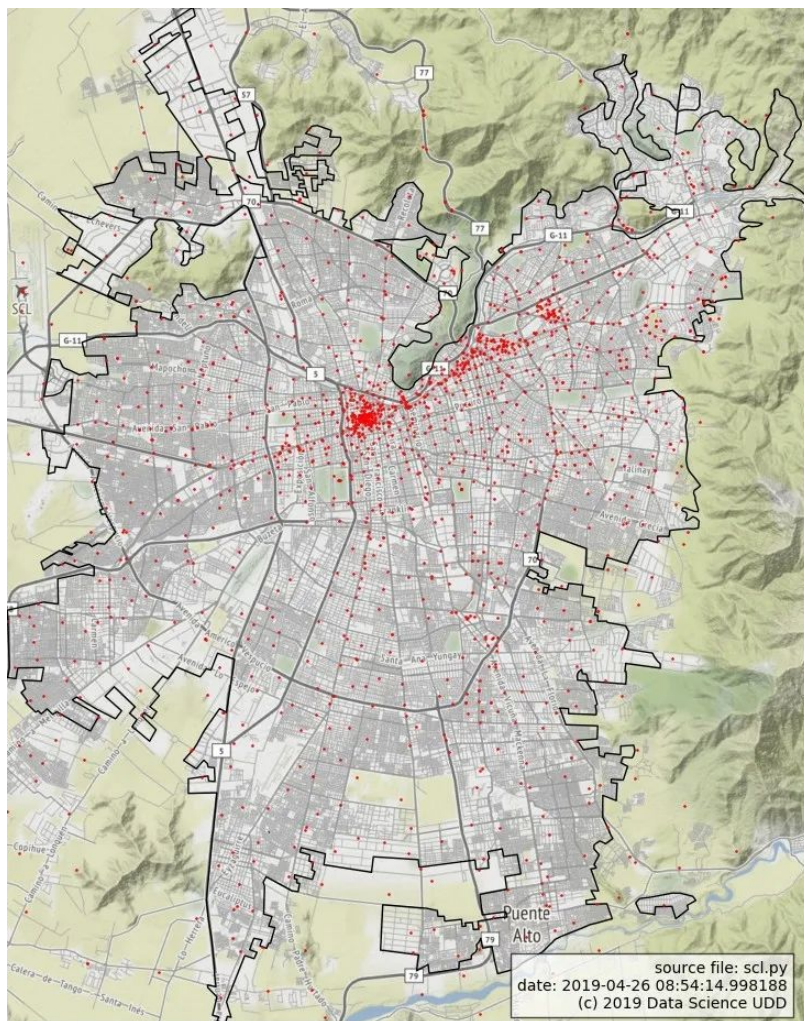
(BBBBAAAEEEEEDDDCCDD)



(a) log number of events per hour



(b) log number of unique users per hour

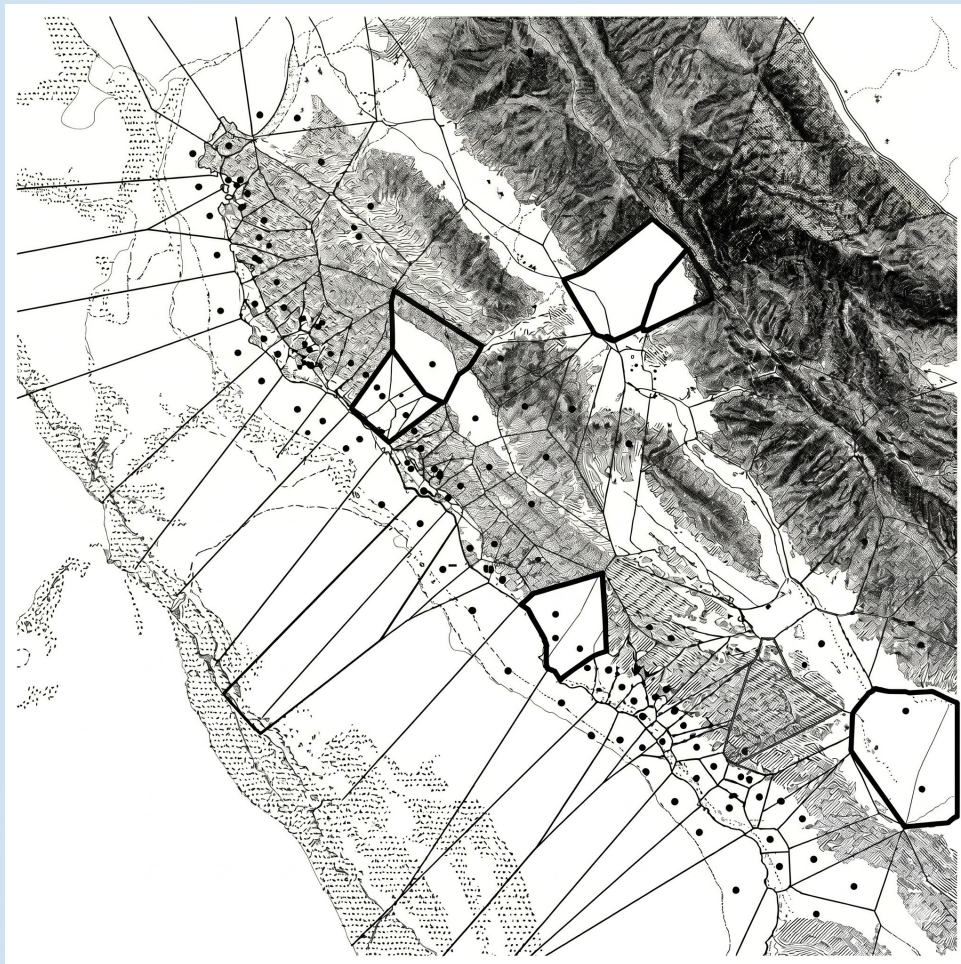


(b) log number of unique users per hour

	CDRs	XDRs	CPRs
Triggering mechanism	User - triggered (calls, SMS)	User - and device - triggered (data sessions)	Network - triggered (control and handover events)
Typical temporal density	Low (sparse events)	Medium to high (depends on data usage)	High (continuous tracking of cell transitions)
Spatial resolution	Antenna - level (coarse; cell - based)	Antenna - level (coarse; cell - based)	Antenna - level (coarse; cell - based)
Semantic content	Contains relational information (caller-receiver)	-	-
Main temporal bias	Large gaps when no calls/SMS occur	Activity - dependent sampling frequency	More complete coverage of movement
Main behavioral bias	Overrepresents frequent callers/SMS users	Overrepresents heavy data users	Less behaviour - dependent (system - driven)
Main spatial bias	Coarse localization; missing intermediate cells	Coarse localization; possible missing cells	Coarse localization; ping - pong effect possible

References

- [paper] [A survey of deep learning for human mobility](#), Pappalardo et al., ACM Computing Surveys, 2021
 - Appendix C.1
- [paper] [Understanding individual human mobility patterns](#)
Gonzalez et al., Nature, 2008
- [paper] [A survey of results on mobile phone datasets analysis](#)
Blondel et al., EPJ Data Science, 2015
- [paper] [Evaluation of home detection algorithms on mobile phone data using individual-level ground truth](#)
Pappalardo et al., EPJ Data Science, 2021



DID YOU KNOW THAT?

Just four mobile
phone records are
enough to identify you

1, 2, 3..4!

Knowing only **four** spatio-temporal points is enough to uniquely identify 95% of individuals in mobile phone data.

This surprising result emerged from the analysis of Call Detail Records of about 1.5 million individuals collected over 15 months.

Not surprising, after all

Suppose someone knows that a person was:

- near their workplace on Monday morning
- at a gym on Tuesday evening
- at a restaurant on Friday night
- at home on Sunday afternoon.

It is unlikely that another person was exactly at the same places at the same times.

Privacy vs information

There is a fundamental tension in mobility data analysis:

Although mobility datasets are often anonymised by removing names or phone numbers, mobility patterns themselves can act as powerful personal fingerprints

Yves-Alexandre de Montjoye, et al. 2013. Unique in the Crowd: The privacy bounds of human mobility. Scientific Reports 3, 1376.

Quiz

A phone automatically syncs email in the background. Which type of record is created?

A)
CDRs



B)
XDRs

C)
CPRs

Quiz

A subscriber makes a voice call, and the operator records the call time, duration, and associated cell information. What type of record is this?

A) 
CDRs

B)
XDRs

C)
CPRs

Quiz

A subscriber lives at the border of two cells. In what format the ping pong effect could be more evident?

A)
CDRs

B)
XDRs

 C)
CPRs

Location-based social networks



Social Network + Location Data

Social Networks

Social structures made up of individuals connected

Location Data

Represented in absolute (lat-long), relative (100 m north of the Space Needle), and symbolic (home, office, shopping mall) form

flickr



Collegamento a questa cartina



Cartina

Ibrido

Satellite

Trova il mio indirizzo



La villa kerylos à Beaulieu-Sur-Mer sur la Côte d'Azur di jmlpyt (fr)



44405 elementi provvisti di GeoTag

Ordina per: [Classifica](#) • [Recenti](#)



Cerca nella cartina





Patrick Torsney

@illegal

Right, doing a little experiment with geolocation. Blog to follow shortly. Here I am at Buckingham Palace..

[← Reply](#) [🗑 Delete](#) [★ Favorite](#) [⋮ More](#)

8:55 AM - 18 May 14 [📍 from Westminster, London](#)

Data Format

- records collected for each post
- location records at city or specific venue level

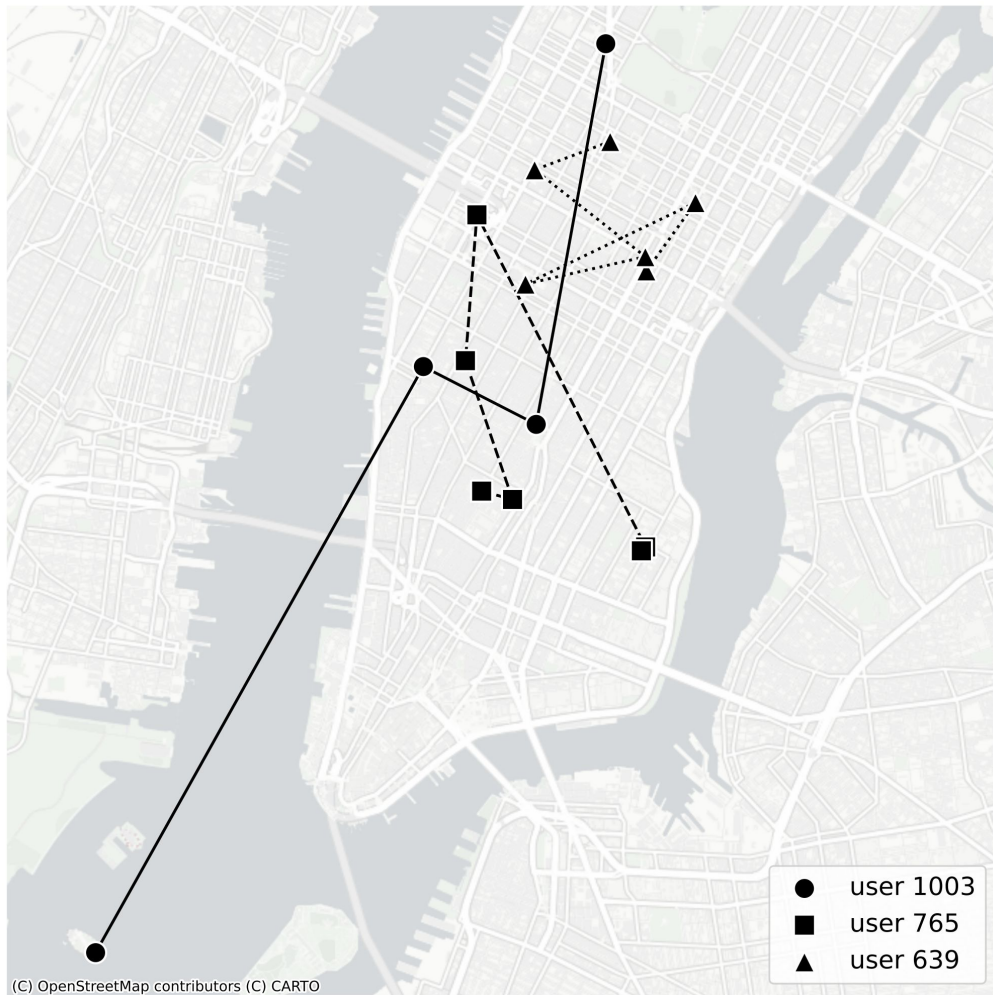
$$(u, t, lid, lpos, post)$$

user time

location
identifier

location
position

post
text



Three samples of trajectories from a dataset of checkins (denoted by markers).

Data available at
<https://snap.stanford.edu/data/loc-Brightkite.html>

Dimension

Main characteristics

Implications for analysis

Spatial resolution

Location precision varies from GPS coordinates to broad place labels or venue centroids. Locations may be user-selected, altered, or obscured.

Locations may be uncertain, and observations reflect where people post, not where they spend time overall.

Temporal resolution

Posts are sparse and irregular; posting frequency varies across users. Platform and API changes create gaps or discontinuities.

Complete routes cannot be reconstructed; inferred movements are approximate. Data need cleaning for anomalous rates and implausible jumps.

Semantic resolution

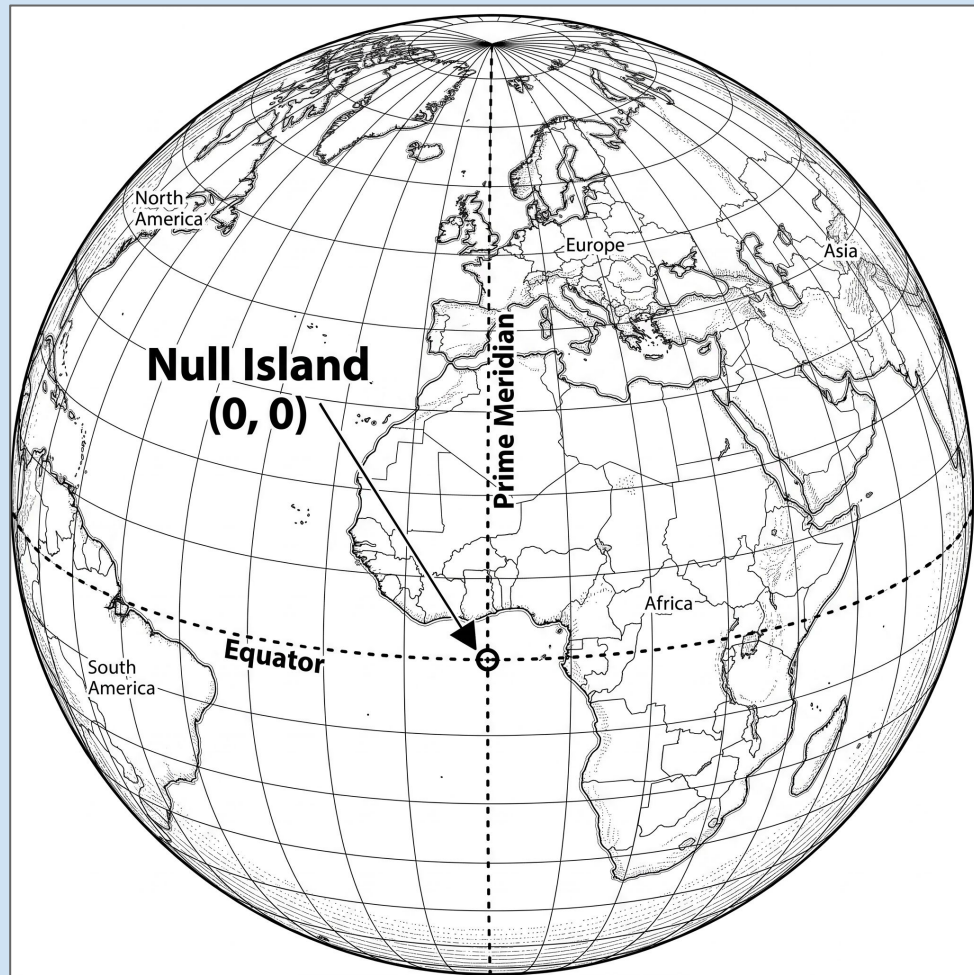
Venue labels, text, hashtags, images, and metadata can suggest activity type or purpose. These interpretations are indirect.

Rich contextual clues support activity inference, but sharing is selective: leisure and special events may appear more often than routine activities.

Representativeness

Geotagged users are a selective subset of the population. Younger, urban, and digitally engaged users are often overrepresented; many users do not share locations.

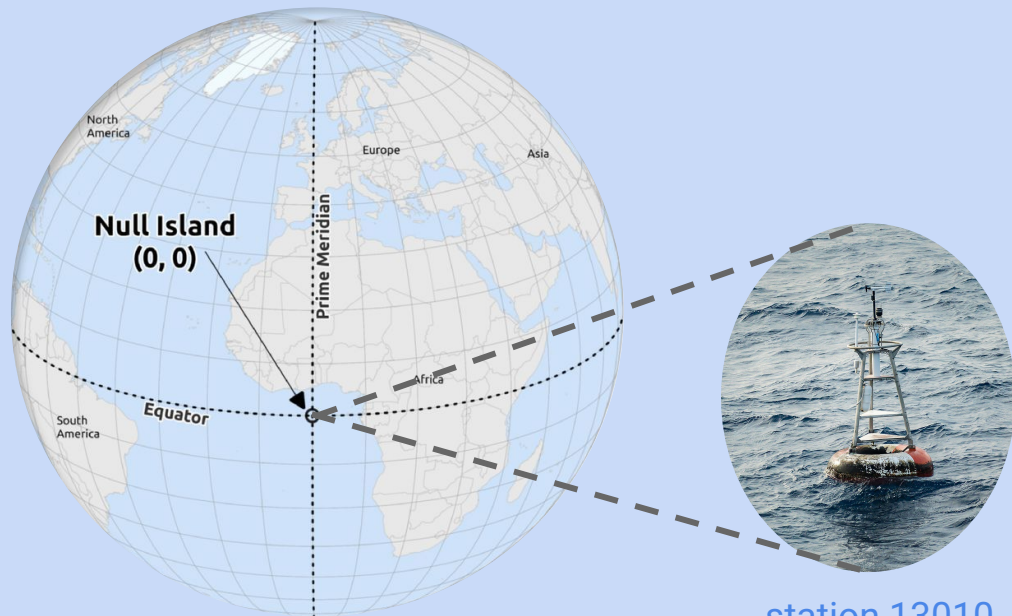
Treat the data as a selective, behaviorally rich sample—not a population-wide mobility record. Combine with other data sources where possible.



DID YOU KNOW THAT?

The busiest place in the world does not really exist

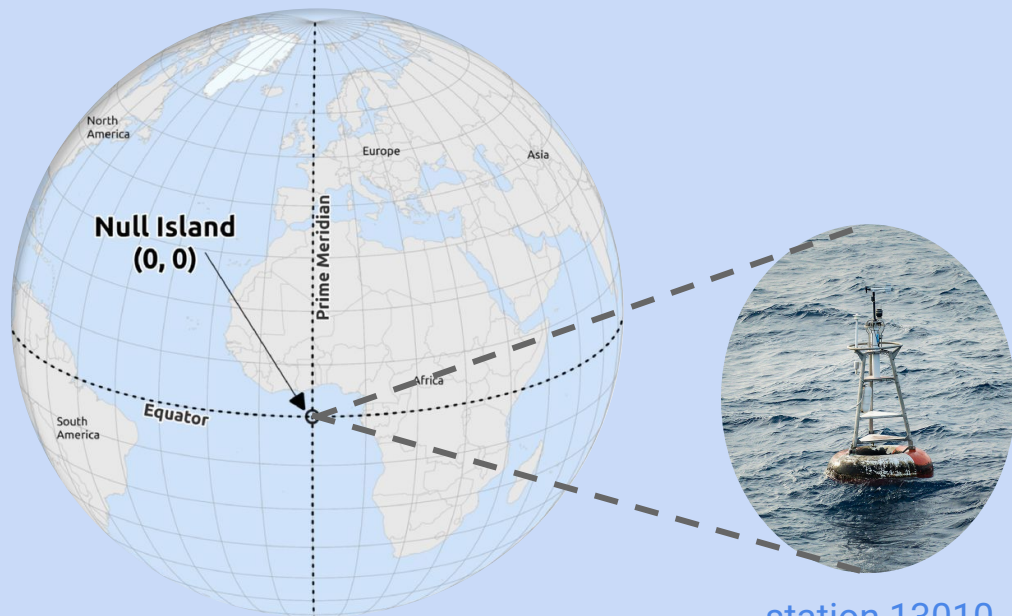
Null Island



station 13010

- it's at coordinates (0,0) in WGS84
- nothing there...but a buoy
- developer Steve Pellegrin first coined the term in 2008

Null Island



station 13010

- human typos & messy data (misspelled street names, non-existent building numbers), or glitches in the geocoder can create invalid addresses...
- ...so that the output is "0,0"

Heatmap globale

Colori heatmap

Calore Blu Purple Grigio Rosso

Tipo attività

Tutti    

Opacità colore

0% 40% 60% 80% 100%

Livelli

Mappa

Luoghi

Stili delle mappe

Scura

Chiara

Standard

Satellite

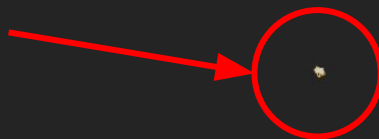
Ibrida

Inverno

Mostra terreno 3D ☐

[Scopri](#) come è stata costruita la heatmap.

[Scopri](#) come Strava Metro può aiutare la



Heatmap globale



Colori heatmap

Calore Blu Purple Grigio Rosso

Tipo attività

Tutti

Opacità colore

0% 40% 60% 80% 100%

Livelli

Mappa

Luoghi

Stili delle mappe

Scura

Chiara

Standard

Satellite

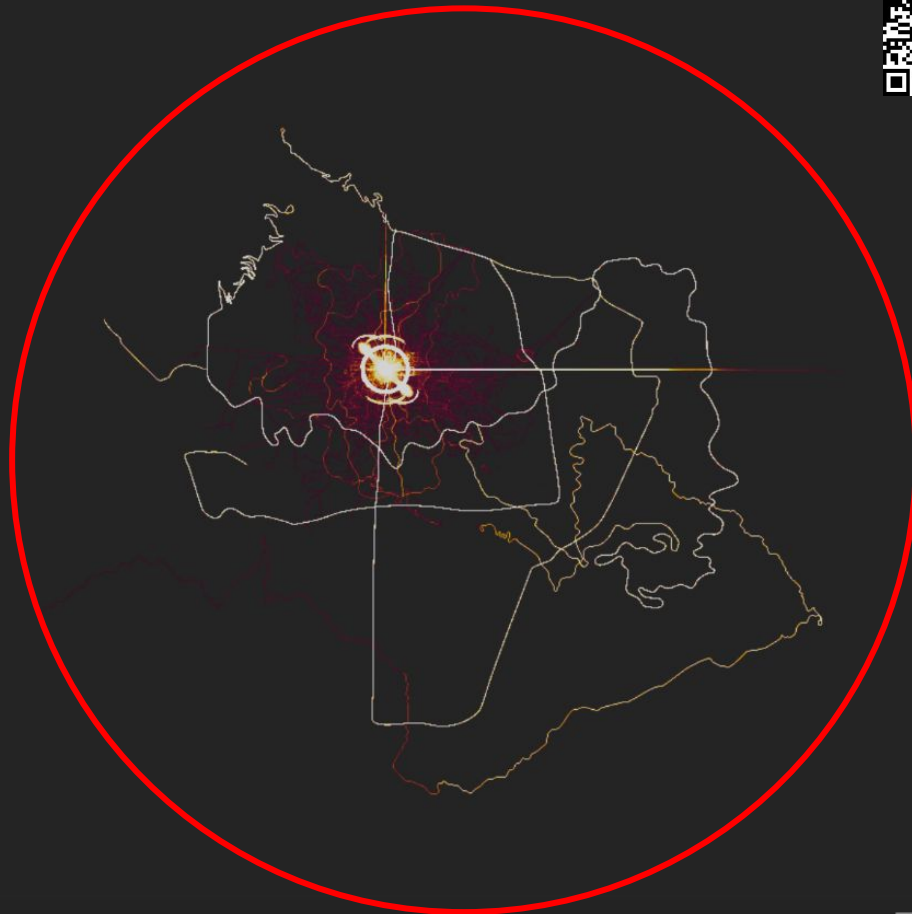
Ibrida

Inverno

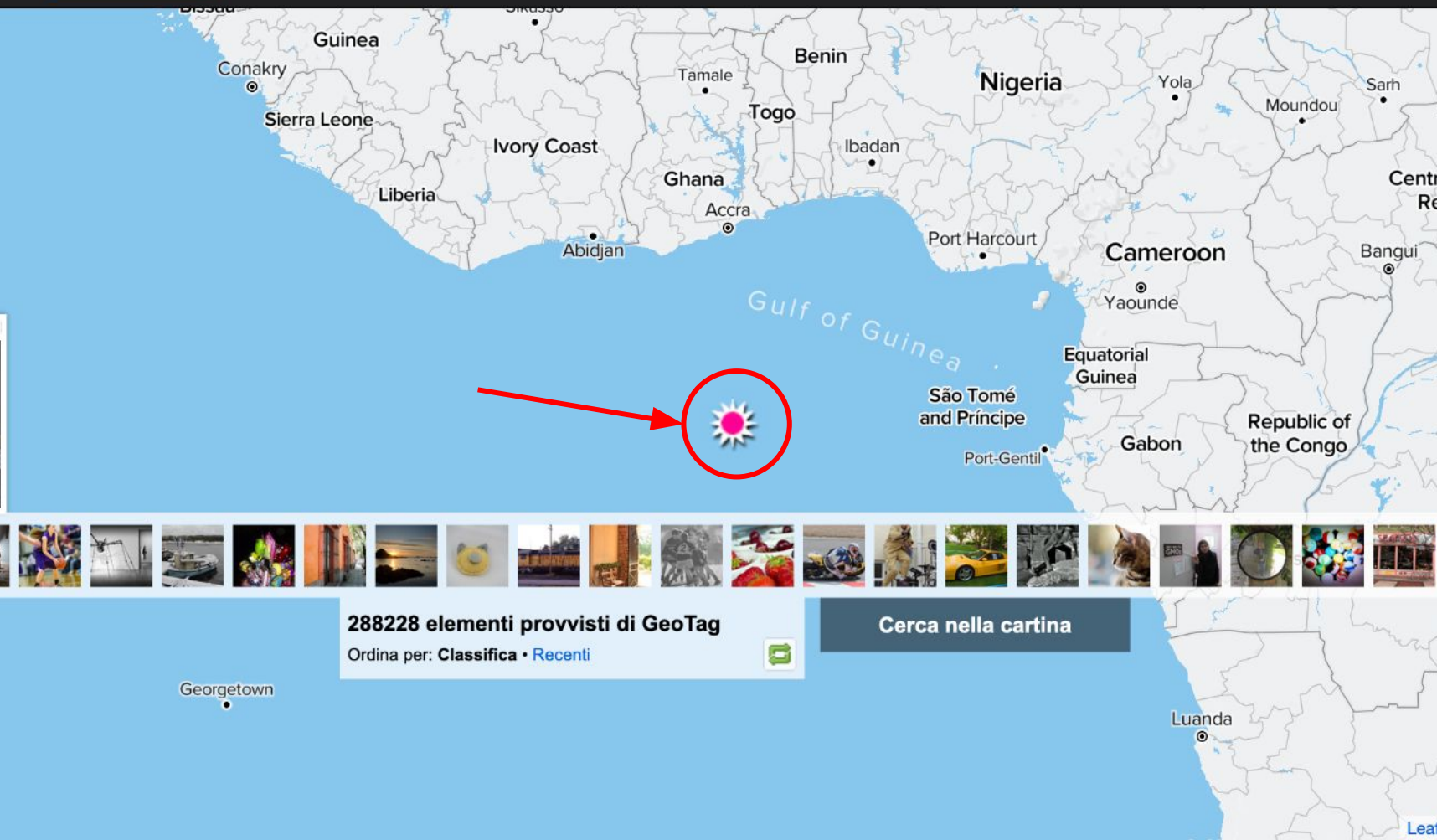
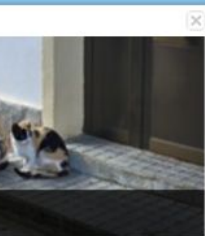
Mostra terreno 3D ☐

[Scopri](#) come è stata costruita la heatmap.

[Scopri](#) come Strava Metro può aiutare la



R

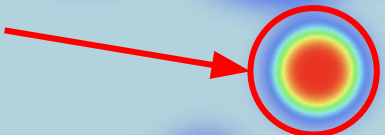


288228 elementi provvisti di GeoTag

Ordina per: [Classifica](#) • [Recenti](#)



Cerca nella cartina



REFERENCES



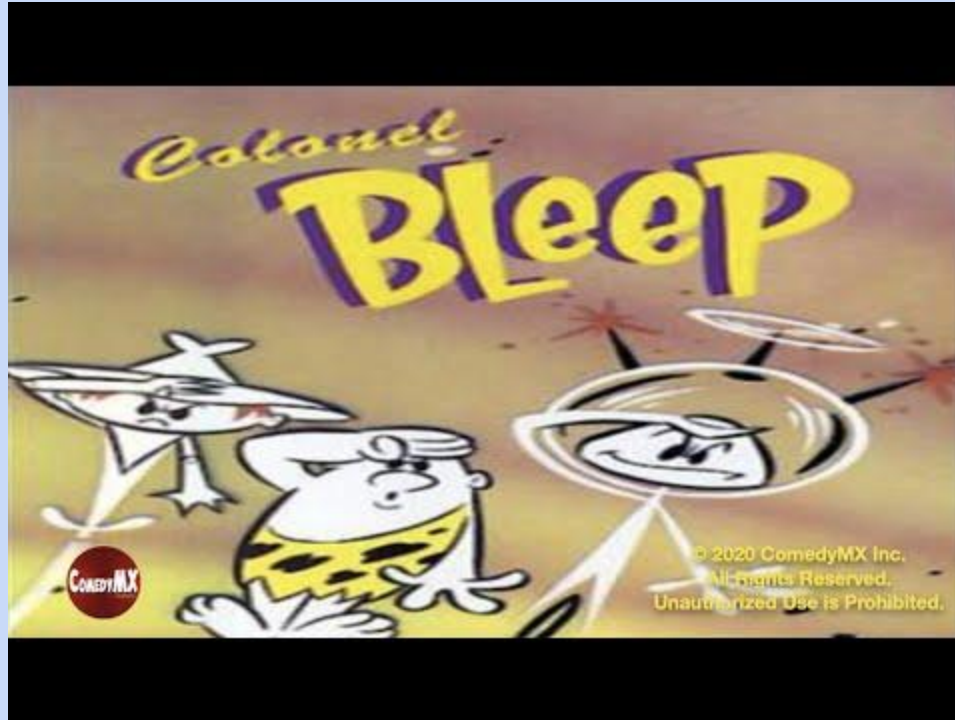
- [article] [MapLab: Inside Null Island, 'the Most Real of Fictional Places'](#), Bloomberg, CityLab, 2022
- [article] [L'isola dell'Atlantico creata dai dati sbagliati](#), Il Post, 2022
- [social media post] [I think I discovered a secret Chinese military base in the middle of the ocean](#), Reddit post, 2021
- [blog post] [The Geographical Oddity of Null Island](#), The Library of Congress, 2016
- [article] [If You Can't Follow Directions, You'll End Up on Null Island](#) The Wall Street Journal, 2016

REFERENCES



- [video] [Null Island: The Busiest Place That Doesn't Exist](#), YouTube video, 2016
- [blog post] [Null Island: Where Geocoding Errors Go to Live](#), GIS Lounge, 2016
- [paper] ["I think I Discovered a Military Base in the Middle of the Ocean" – Null Island, the Most Real of Fictional Places](#), ArXiv, 2022, <https://arxiv.org/pdf/2204.08383.pdf>
- [dataviz] [Null points: map of all null islands, lakes and holes](#), Kenneth Field

REFERENCES



References

- [paper] [A survey of deep learning for human mobility](#), Pappalardo et al., ACM Computing Surveys, 2021
 - Appendix C.3
- [paper] [Social media and mobility landscape: uncovering spatial patterns of urban human mobility with multi-source data](#), Cui et al., Frontiers of Environmental Science & Engineering, 2018
- [paper] [Urban Human Mobility: Data-Driven Modeling and Prediction](#), Wang et al., ACM SIGKDD Explorations Newsletter, 2019
 - Section 2.2

Quiz

After a major transport strike, a city wants to measure how aggregate mobility changes compared with previous weeks. Which data source is the best fit?

A) GPS

B) MPRs

C) Social Media

D) Census data



Quiz

A researcher wants to map the precise routes drivers take through a city and identify where they turn or change speed. Which data source is the best fit?

A) GPS



B) MPRs

C) Social Media

D) Census data

Quiz

A transport agency wants to estimate daily flows between districts and compare commuting patterns across a metropolitan area. Which data source is the best fit?

A) GPS

B) MPRs

C) Social Media

D) Census data



Quiz

A city planner wants to compare the age structure and household composition of residents across neighborhoods. Which data source is the best fit?

A) GPS

B) MPRs

C) Social Media

D) Census data



Quiz

A researcher wants to study which restaurants or parks people choose to visit and what those places mean to them. Which data source is the best fit?

A) GPS

B) MPRs

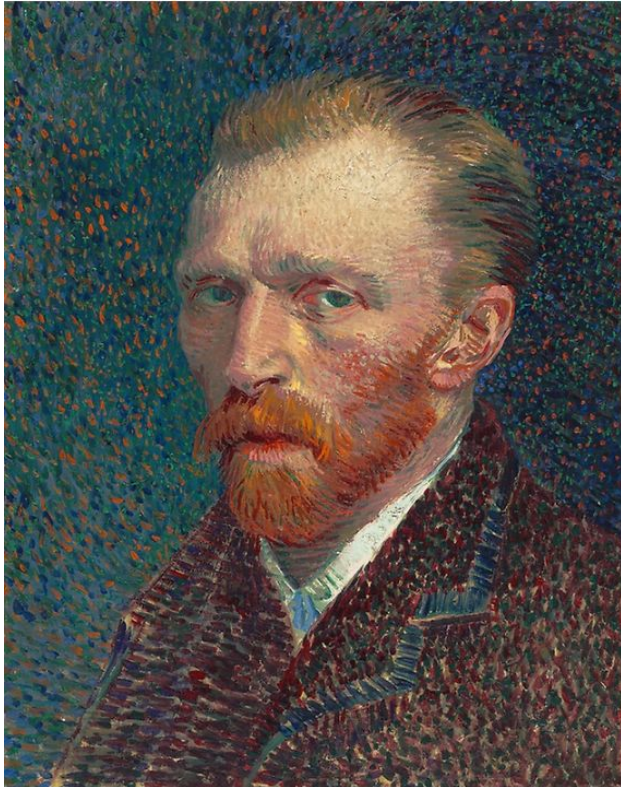
C) Social Media

D) Census data

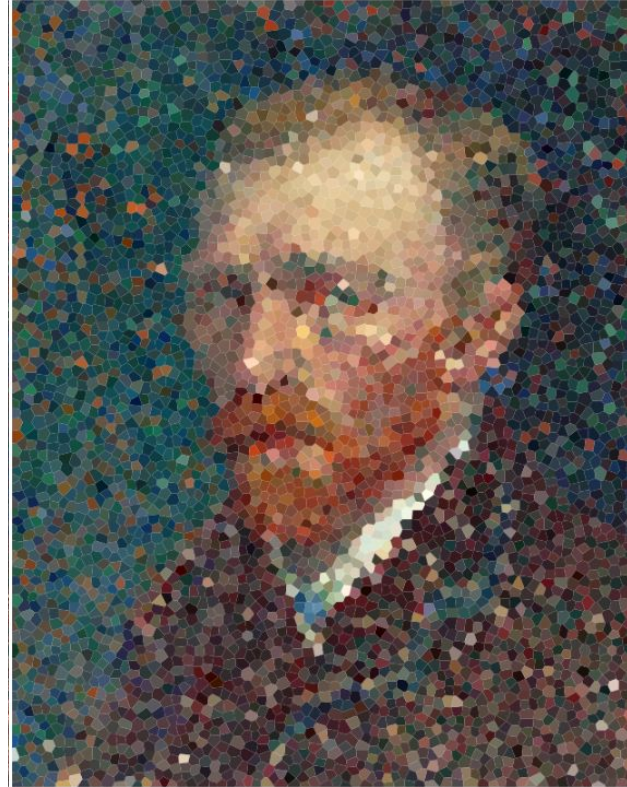


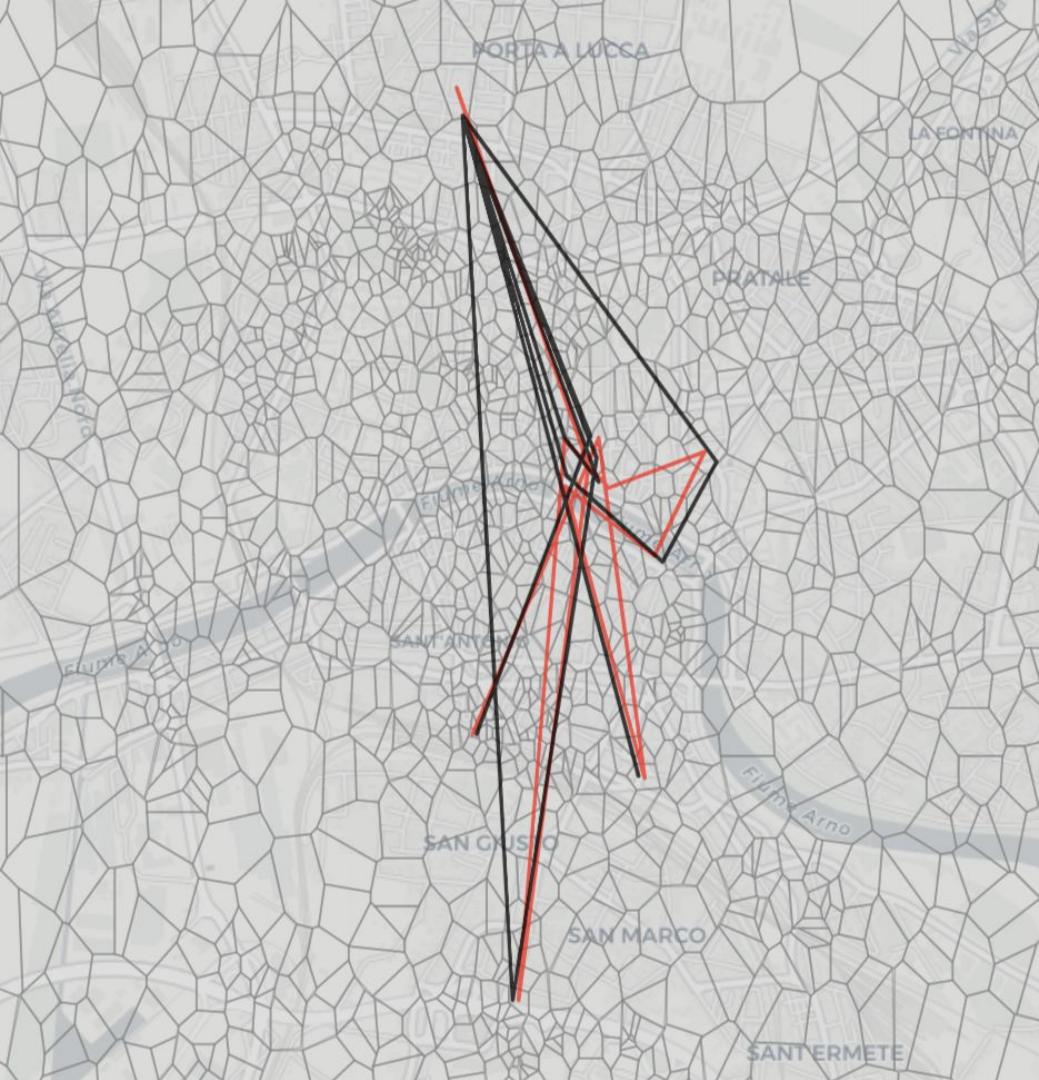
Analytical objective	Best-suited data source	Why it fits	Main limitation / useful complement
Reconstruct routes and movements	GPS	Frequent, precise coordinates support path reconstruction, map-matching, and analysis of speed and turns.	GPS samples may be socially or behaviorally selective.
Estimate mobility flows	MPRs	Broad coverage and repeated observations support OD flows and aggregate trends.	Cell-based locations are coarse; GPS can add route-level detail.
Study place preferences and socially meaningful visits	Social media	Posts, check-ins, text, and images reveal places people choose to share and their associated meanings.	Sharing is selective. GPS enriched with points of interest can help identify visits, but presence alone does not show preference.
Measure responses to short-term disruptions	MPRs (CPRs and XDRs)	Frequent observations across many users help track aggregate changes before, during, and after an event.	GPS can complement with route-level changes; social media can reveal reactions, but not reliably quantify population-wide mobility.

GPS traces

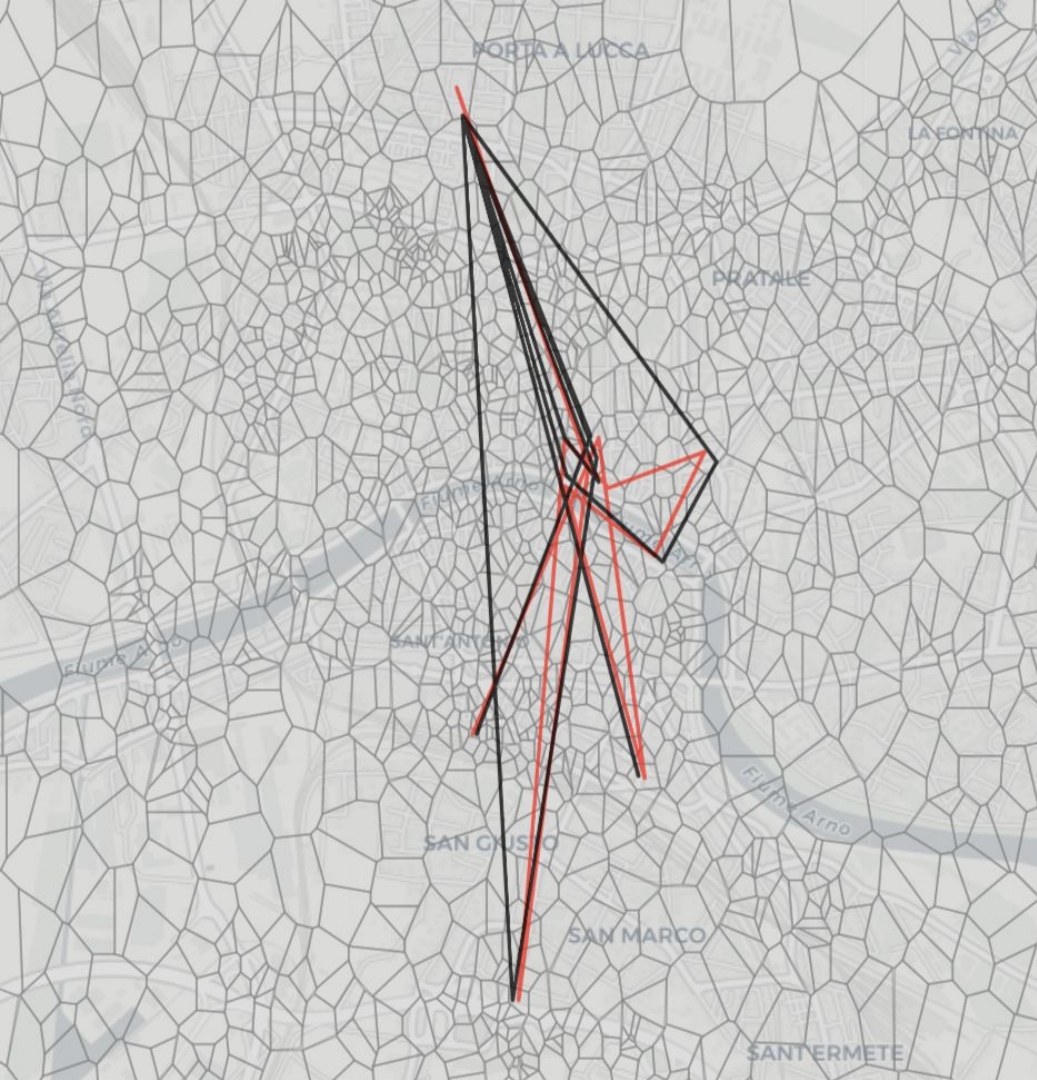


Phone & LBS data





Black: GPS trajectory
Red: CDR-like trajectory



Black: GPS trajectory
Red: CDR-like trajectory

- GPS trajectory from Google Maps
- Voronoi tessellation from OpenCellID
- GPS trajectory “transformed” into CDRs with skmob