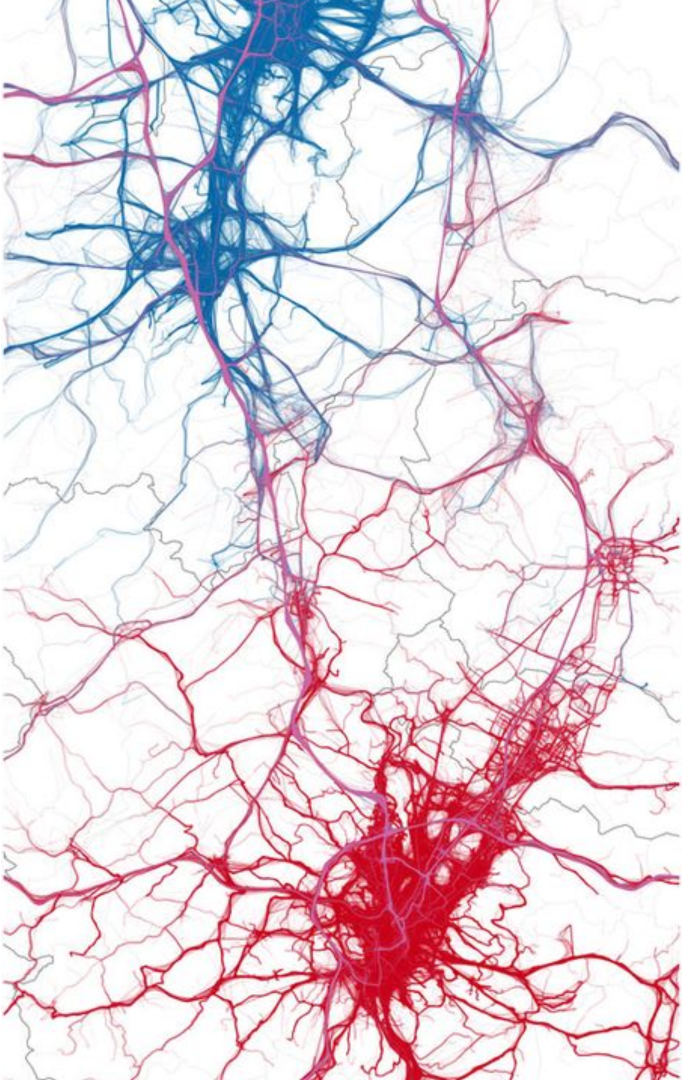


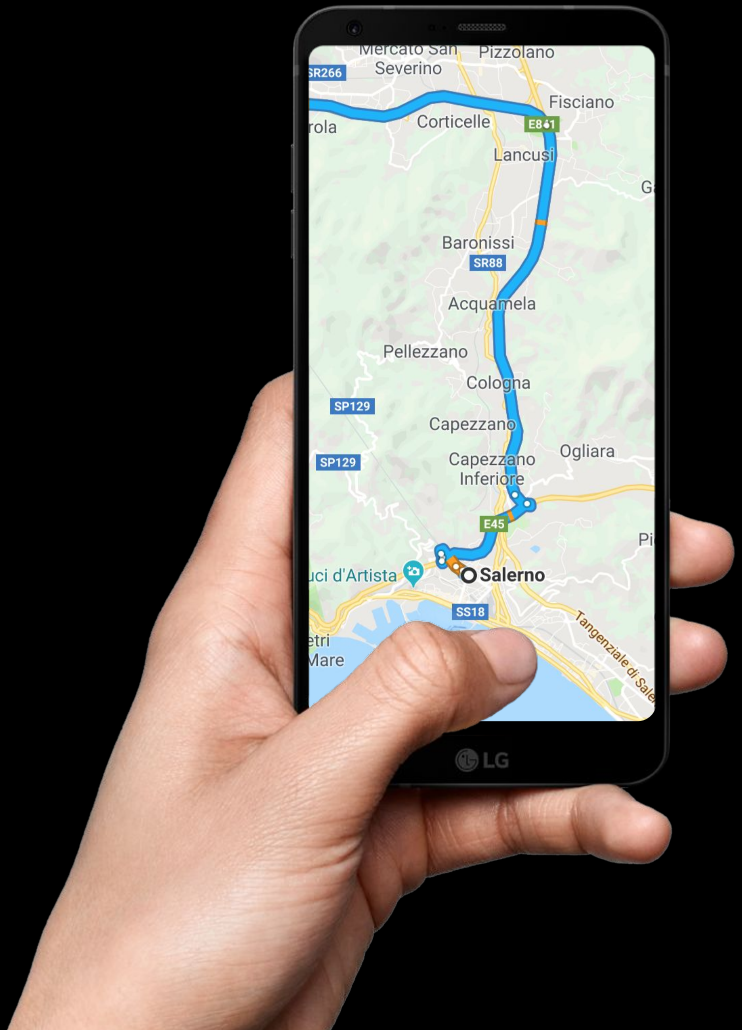
IN SUPREMÆ DIGNITATIS
1343

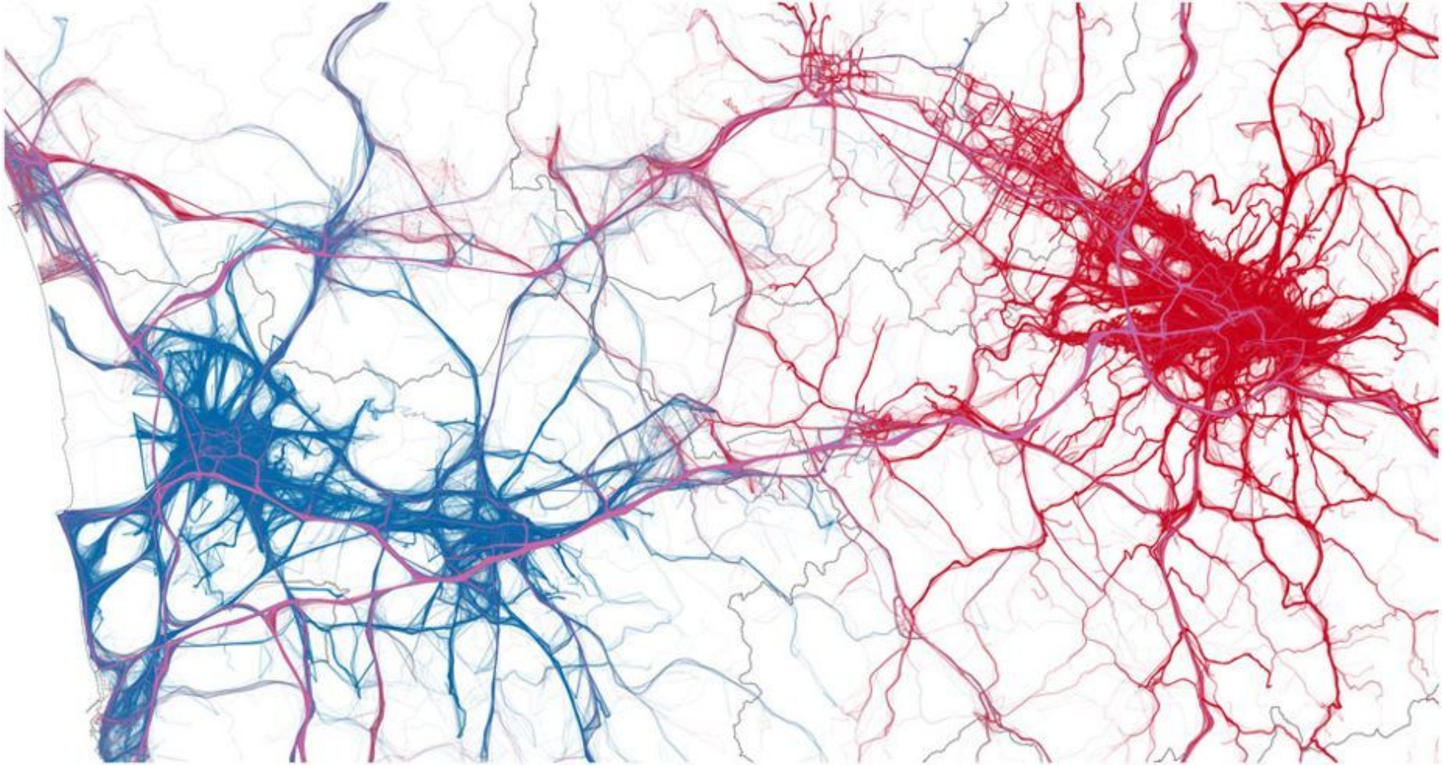
Consiglio Nazionale
delle Ricerche

SCUOLA
NORMALE
SUPERIORE

Spatial and Mobility Data

GPS data


Mercato San Severino
Pizzolano
SR266
Fisciano
Corticelle
E841
Lancusi
Baronissi
SR88
Acquamela
Pellezzano
Cologna
SP129
Capezzano
Ogliara
SP129
Capezzano Inferiore
E45
Salerno
SS18
Mare
Tangenziale di Sale
LG



Global Positioning System (GPS)

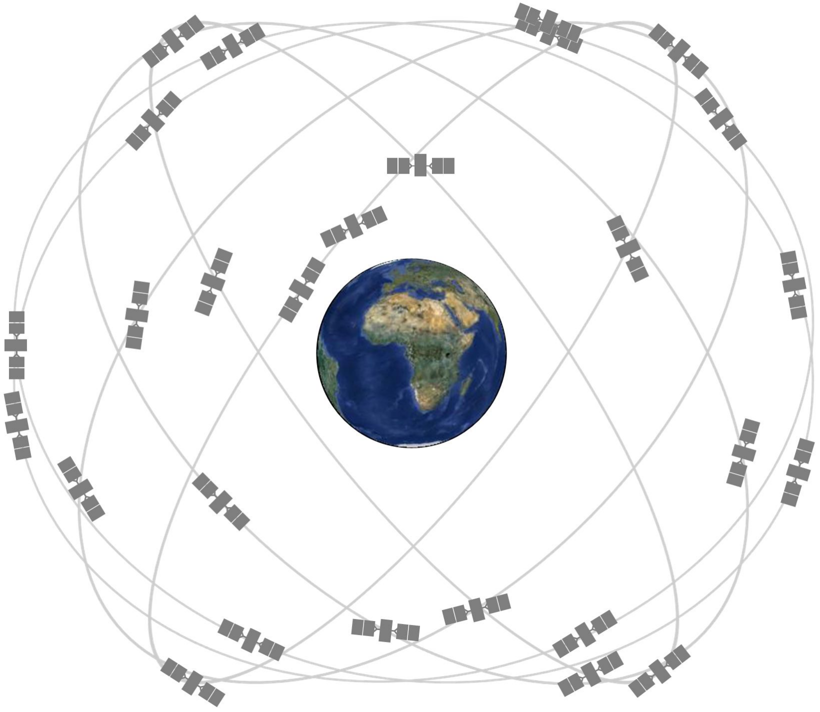
- Most famous GNSS (global navigation satellite system)
- Complete name: NAVTEC GPS

- U.S.-owned utility providing users with positioning, navigation, and timing (PNT) services

- A system made of three main components:
  a. Space segment
  b. Control segment
  c. User segment

From latin *satellus* (pl. *satellites*)

1) Space segment



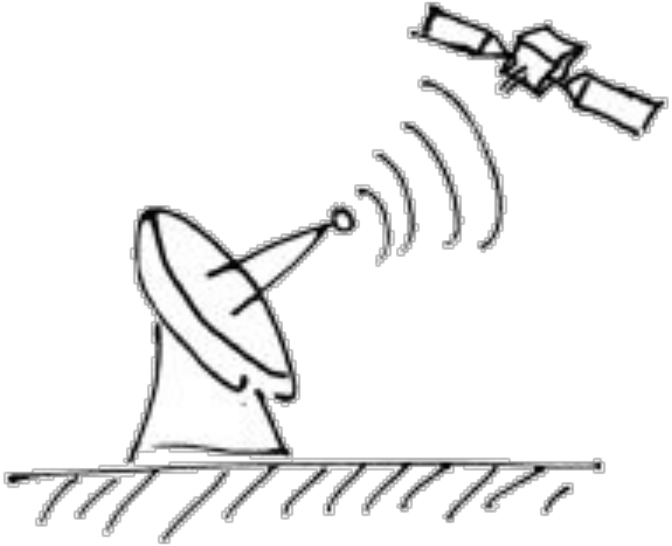
https://www.gps.gov/systems/gps/space/

24+ satellites, each

- orbiting at 20k km
- circling Earth twice a day in six equally-spaced orbits
- broadcasting a GPS radio signal

- each orbit contains four satellites
- users view at least four satellites from any point

2) Control segment



Ground facilities that:

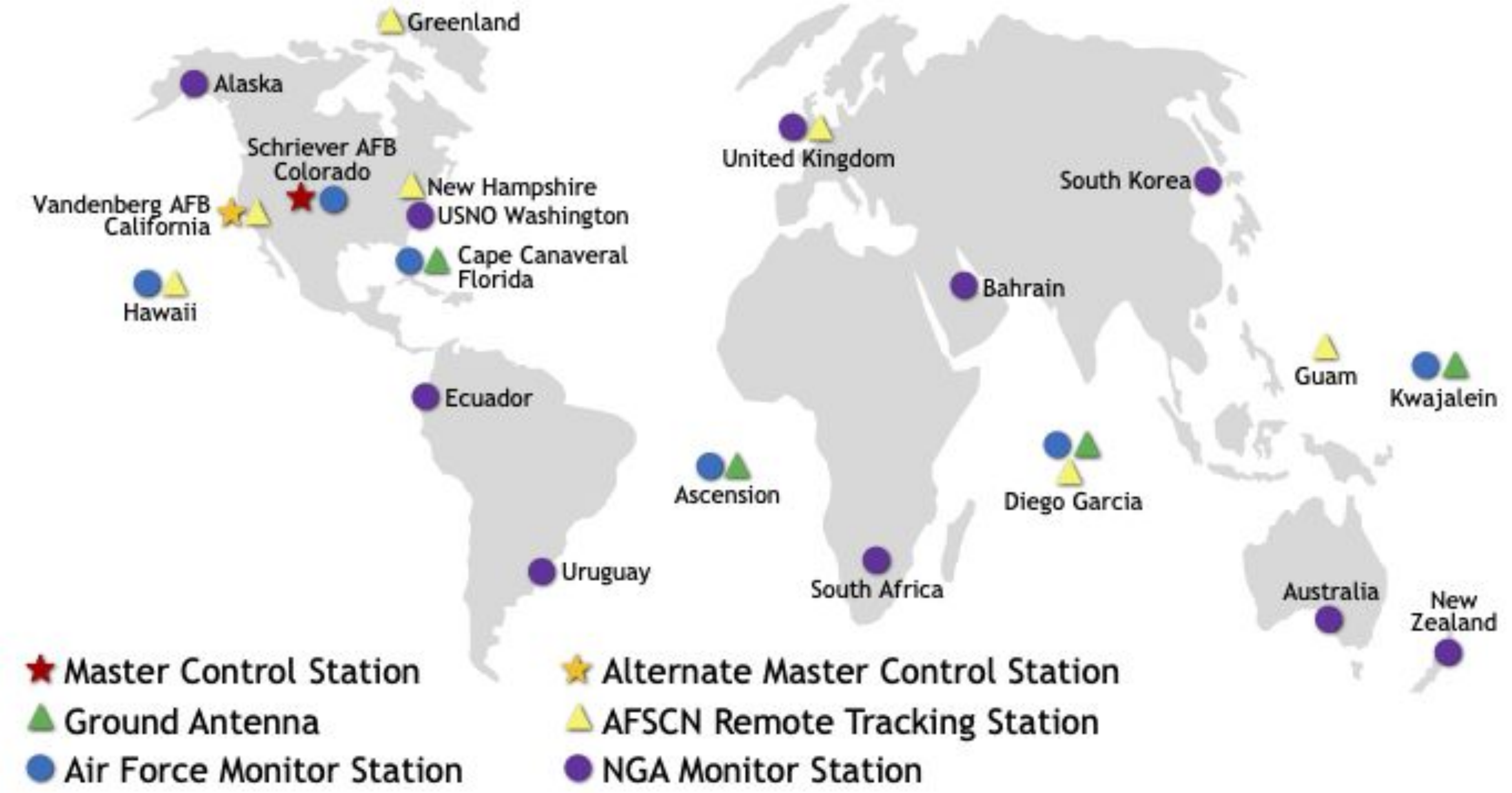
1. track the GPS satellites
2. monitor their transmissions
3. perform analyses
4. send commands and data to the constellation

2) Control segment

The current Operational Control Segment (OCS) includes:

1. a master control station
2. an alternate master control station
3. 11 command and control antennas
4. 16 monitoring sites

The GPS constellation delivers high performance thanks to the dedicated efforts of its operators (nicknamed Team Blackjack)


Greenland
Alaska
Schriever AFB
Colorado
New Hampshire
Vandenberg AFB
California
USNO Washington
Cape Canaveral
Florida
Hawaii
United Kingdom
South Korea
Bahrain
Guam
Kwajalein
Ecuador
Ascension
Diego Garcia
Uruguay
South Africa
Australia
New
Zealand
Master Control Station
Alternate Master Control Station
Ground Antenna
AFSCN Remote Tracking Station
Air Force Monitor Station
NGA Monitor Station

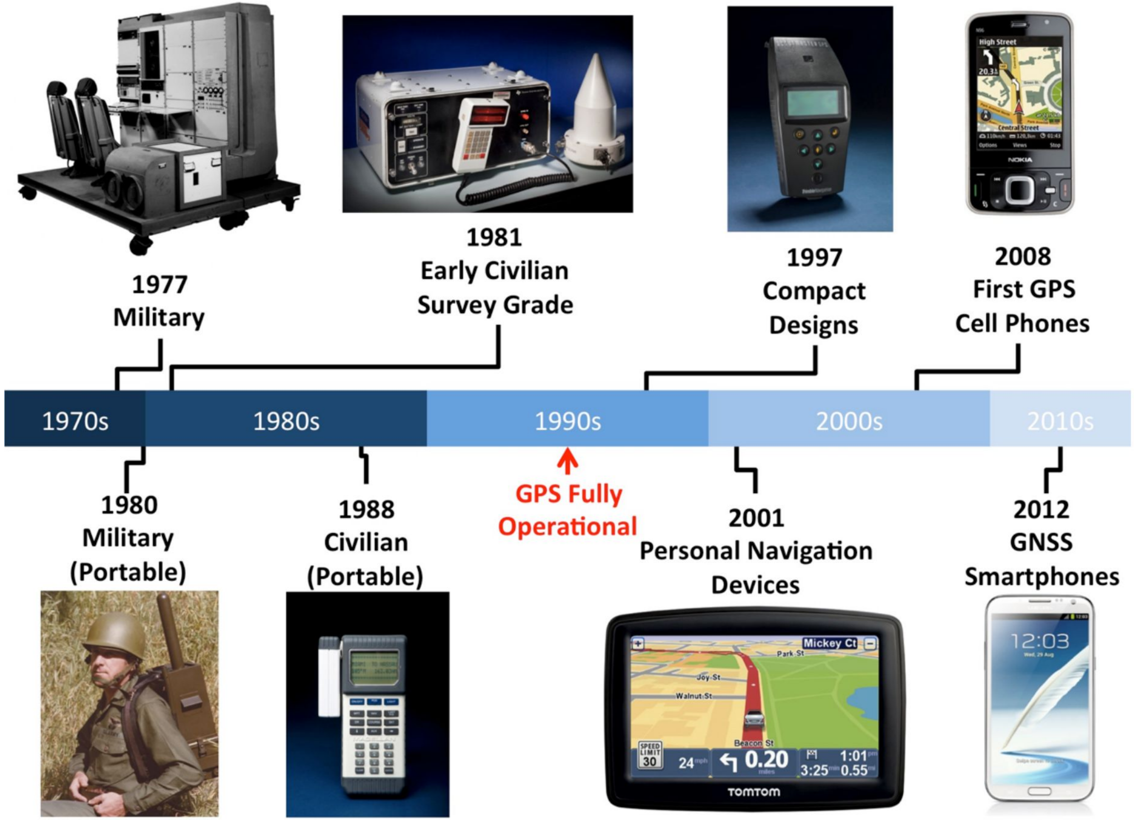
3) User segment

GPS is everywhere:

- cell phones
- wristwatches
- bulldozers
- shipping containers
- ATM's
- sports bras
- ...

3) User segment

GPS has not always been that small…



How does GPS work?

1. GPS satellites broadcast radio signals at the speed of light `{c}` providing their locations, status, and precise time `{t1}` from on-board atomic clocks

2. A GPS device receives the signals, noting exact time of arrival `{t2}` to calculate its distance from the satellite

3. Once the device knows its distance from at least four satellites, it can determine its location on Earth in three dimensions

Trilateration

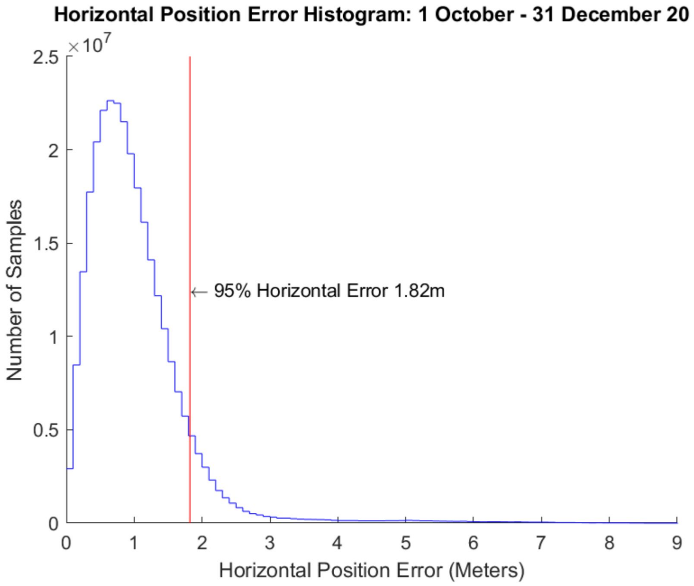
- Each satellite broadcasts a signal that eventually hits the GPS receiver
- The distance between the device and the satellite forms a sphere
- Where all spheres intersect determines the device's position
- At least three satellites needed to accurately position a device

GPS traces format

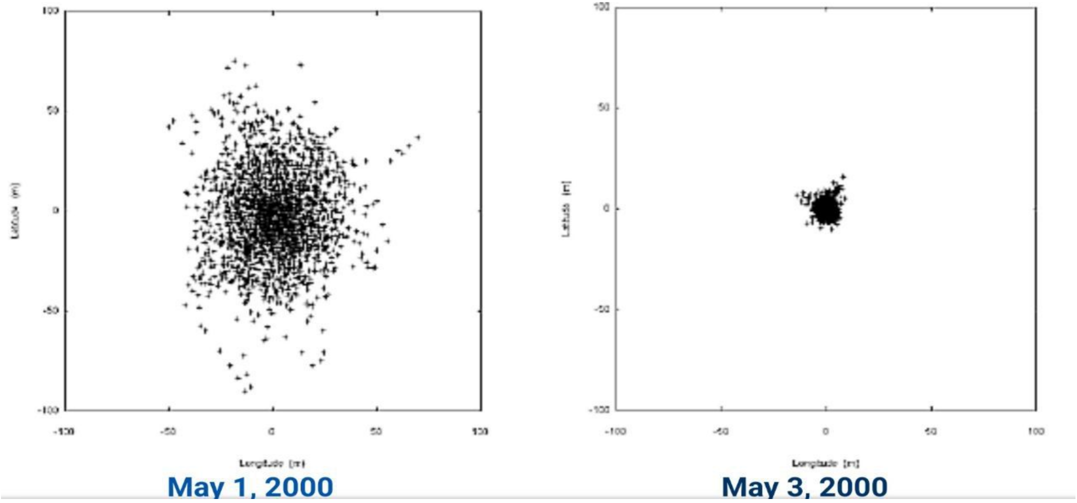
$$(\underset{\text{device}}{u}, \underbrace{lat, lng, alt}_{\text{position}}, \underset{\text{time}}{t})$$

- Sampling rate from few secs to 1-2 minutes
- Spatial precision 10 m

GPS precision



https://www.nstb.tc.faa.gov/reports/2020_Q4_SPS_PAN_v2.0.pdf#page=38



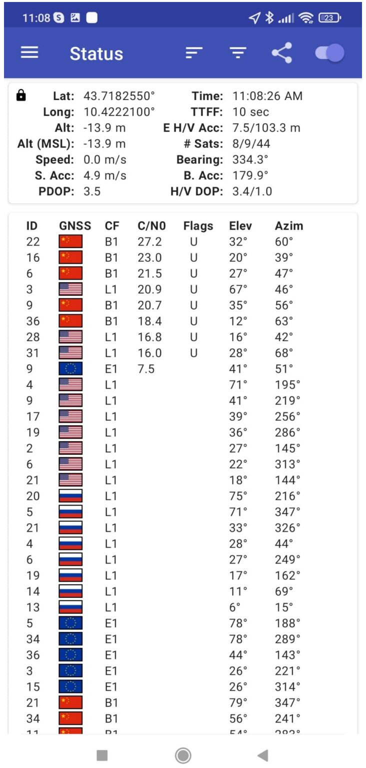
GPS errors before and after "Selective Availability" was abolished by President Clinton, in 2000.

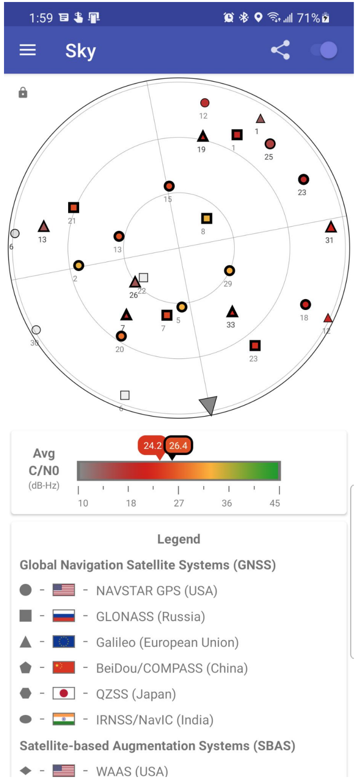
https://www.gps.gov/systems/gps/modernization/sa/data/

GPS is not the only GPS

		Present (2017)			Future		
Core Const.	Owner	Total Number of Operational Satellites	Orbit Type	Configuration	Total Number of Operational Satellites	Orbit Type	Configuration
GPS	USA	31	MEO	6 Planes	24	MEO	6 Planes
GLONASS	Russia	24	MEO	3 Planes	24	MEO	3 Planes
Galileo	Europe	15	MEO	3 Planes	24	MEO	3 Planes
BeiDou	China	19	MEO	6 Satellites in 2 Planes	35	MEO	27 Satellites in 3 Planes
			IGSO	8 Satellites		IGSO	3 Satellites
			GEO	5 Satellites		GEO	5 Satellites

Source: https://searchworks.stanford.edu/view/12064597





GPSTest Android app

https://gnssplanning.com/#/settings

PROs and CONs

PROs

- ubiquitous
- high resolution
- dense traces

CONs

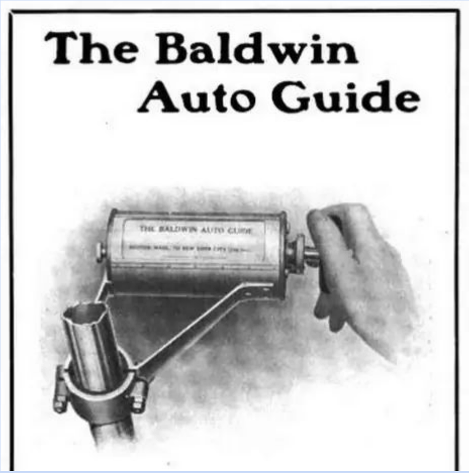
- rarely public
- no semantic location (preprocessing is needed)
- errors when signal is noisy
- small samples

References

- [webpage] GPS Overview, gps.gov, 2021
- [webpage] Trilateration, GIS Geography, 2022
- [paper] A survey of deep learning for human mobility, Pappalardo et al., ACM Computing Surveys, 2021
  - Appendix C.2
- [Book] **Pinpoint**, Greg Milner, W. W. Norton & Company; 1st edition, 2017
- [Episode] Watch history, Netflix, season 2 - GPS

INTERVALLO

Who needs GPS? Early automobile navigation systems



+70 years before commercials uses of GPS for navigation, inventors developed automobile navigation devices:

- 1909: The Baldwin Auto Guide describes a scrolled map that attached to the steering wheel

INTERVALLO

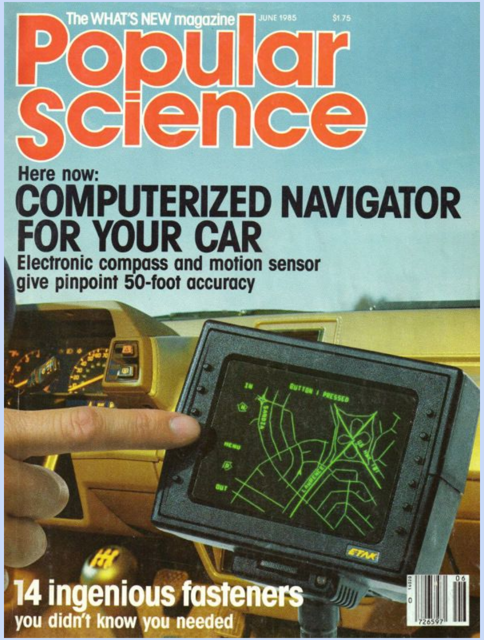
Who needs GPS? Early automobile navigation systems

+70 years before commercials uses of GPS for navigation, inventors developed automobile navigation devices:

- 1930: Iter Avto (Touring Club Italiano) added a level of automation by tethering the map scroll to the car’s speedometer.

INTERVALLO

Who needs GPS? Early automobile navigation systems



+70 years before commercials uses of GPS for navigation, inventors developed automobile navigation devices:

- 1985: Story of Etak's Amazing 1985 Car Navigation System

References

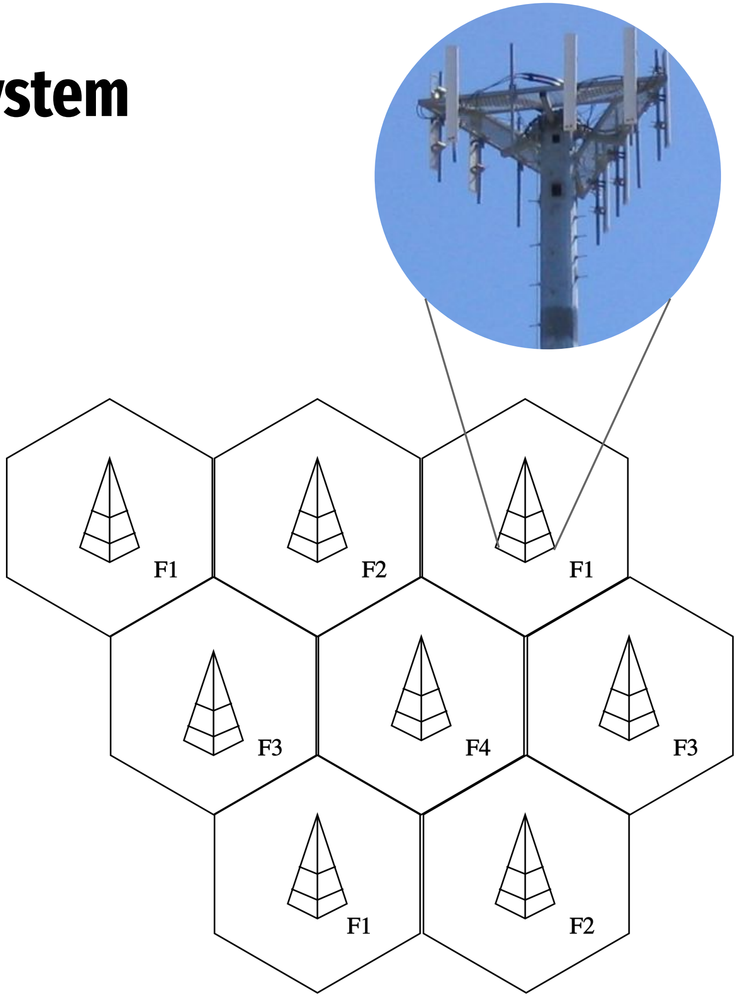
- [Video] How did Polynesian wayfinders navigate the Pacific Ocean? YouTube, TED-Ed channel
- [Article] Before there was the GPS, GIS Lounge, 2013
- [Article] Who Needs GPS? The Forgotten Story of Etak's Amazing 1985 Car Navigation System, FastCompany, 2015

Mobile Phone Records

NBC/Instagram

Mobile Network System

The land area is divided into **cells**:

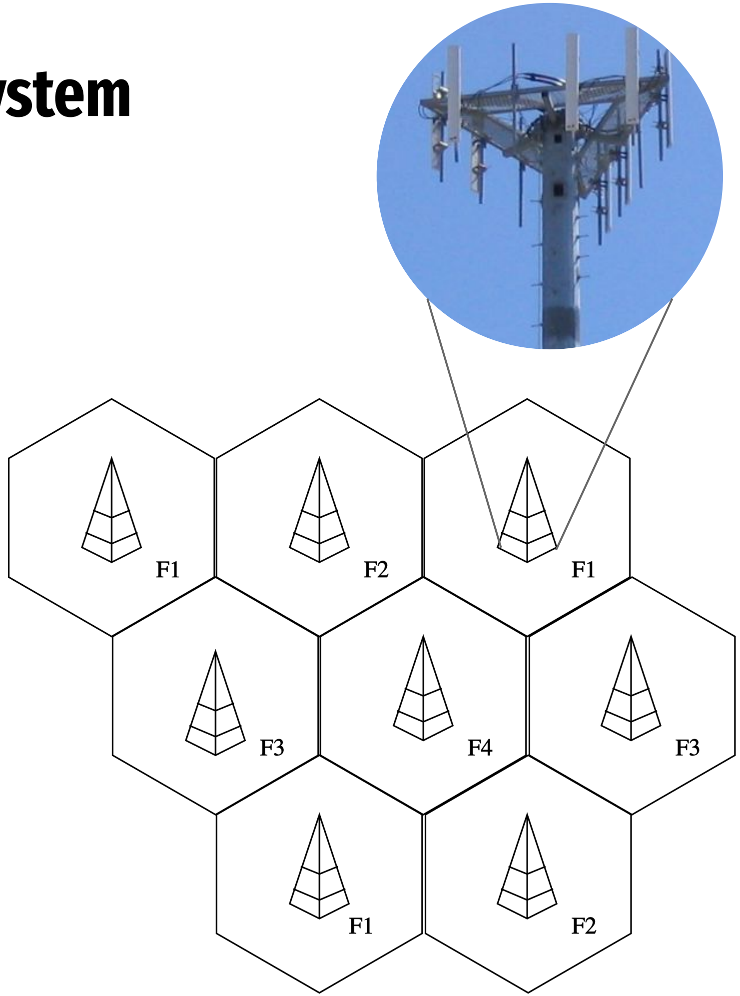


Mobile Network System

The land area is divided into **cells**:

- served by one base transceiver

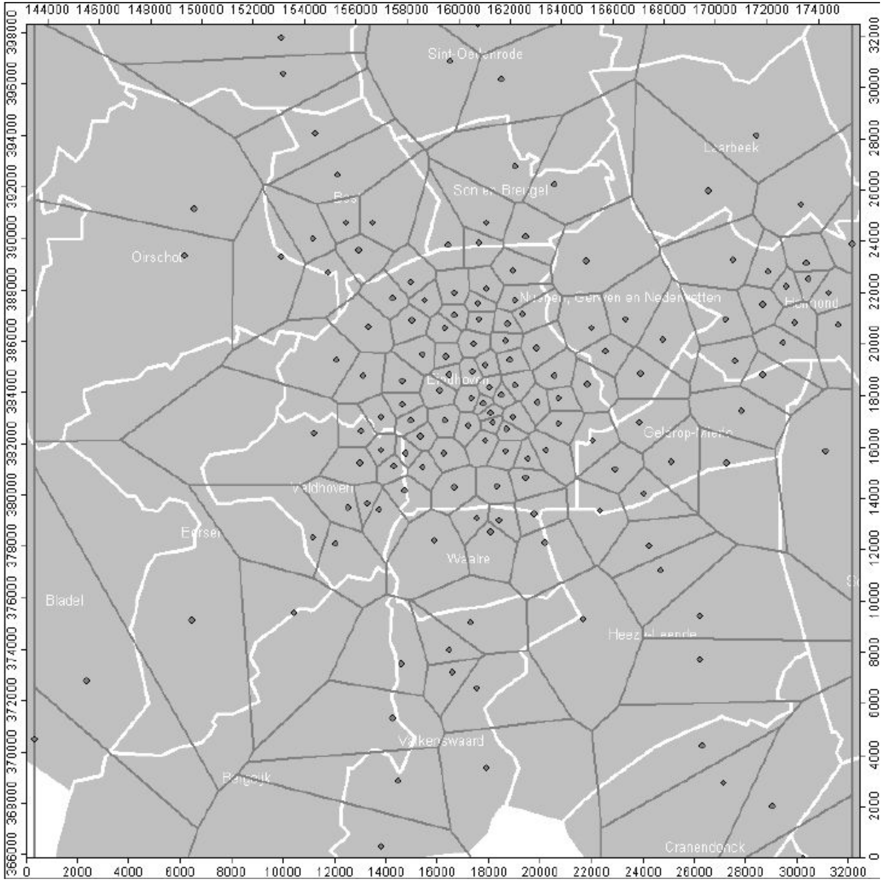
- using a different set of frequencies from neighboring cells
  (to avoid interference)



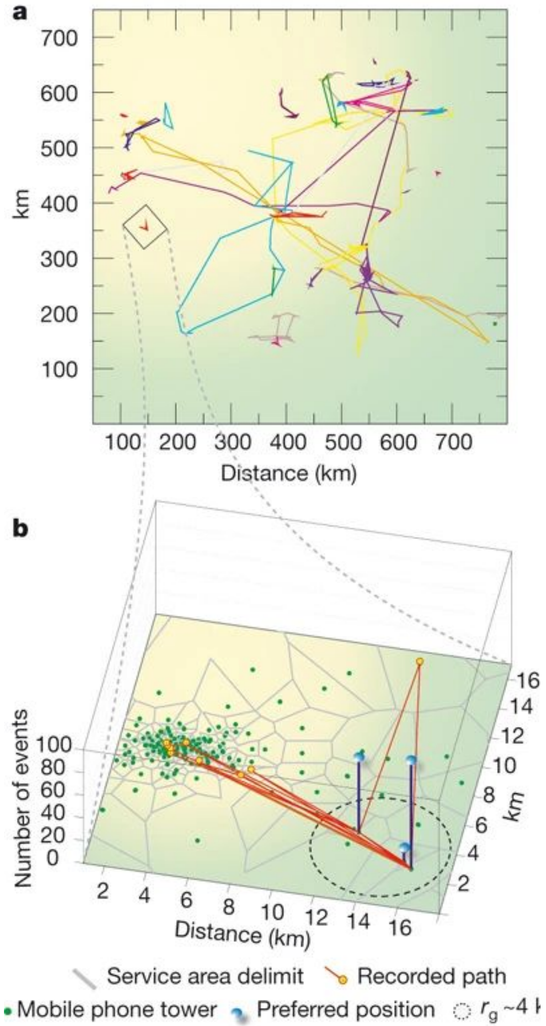
Mobile Phone Network

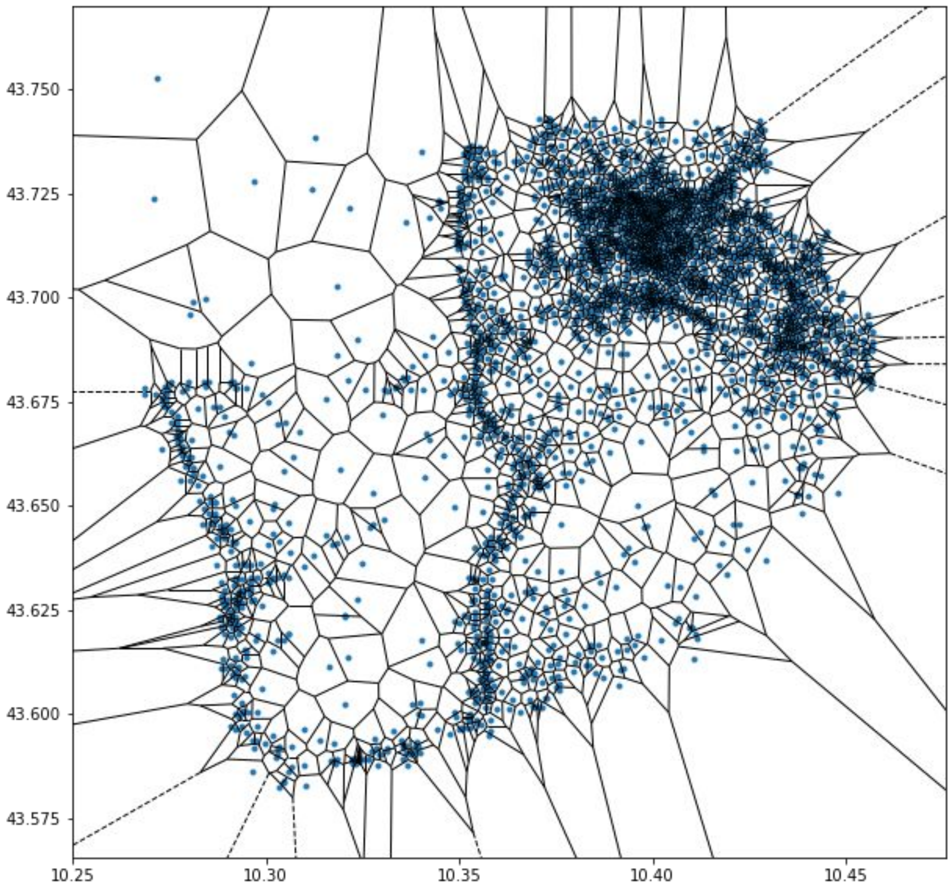
- Used by mobile phone operators for coverage and capacity
- Large areas split into smaller cells to to support more users

- Range of cells:
  - cities: up to 1 km
  - rural areas: up to 8 km
  - unpopulated areas: up to 40 km

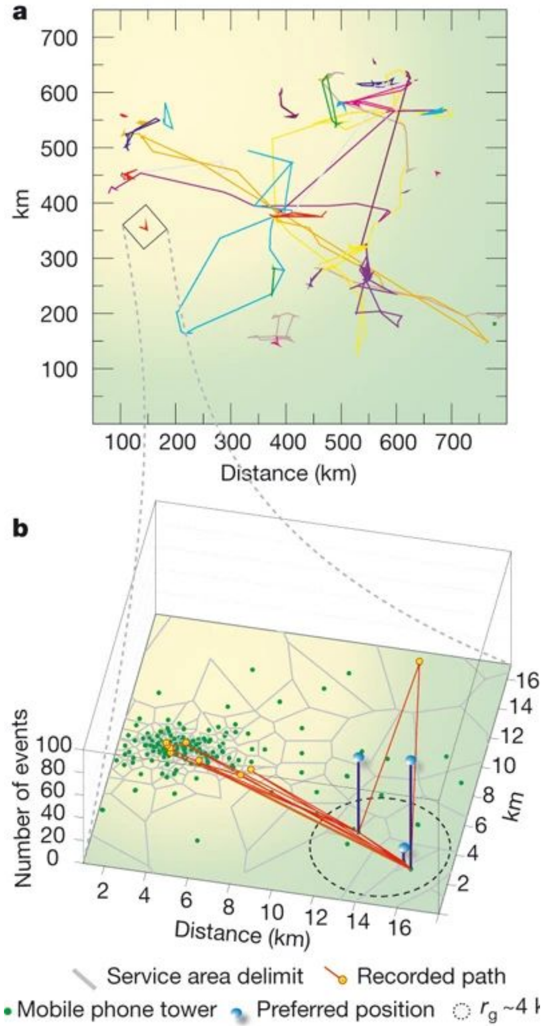


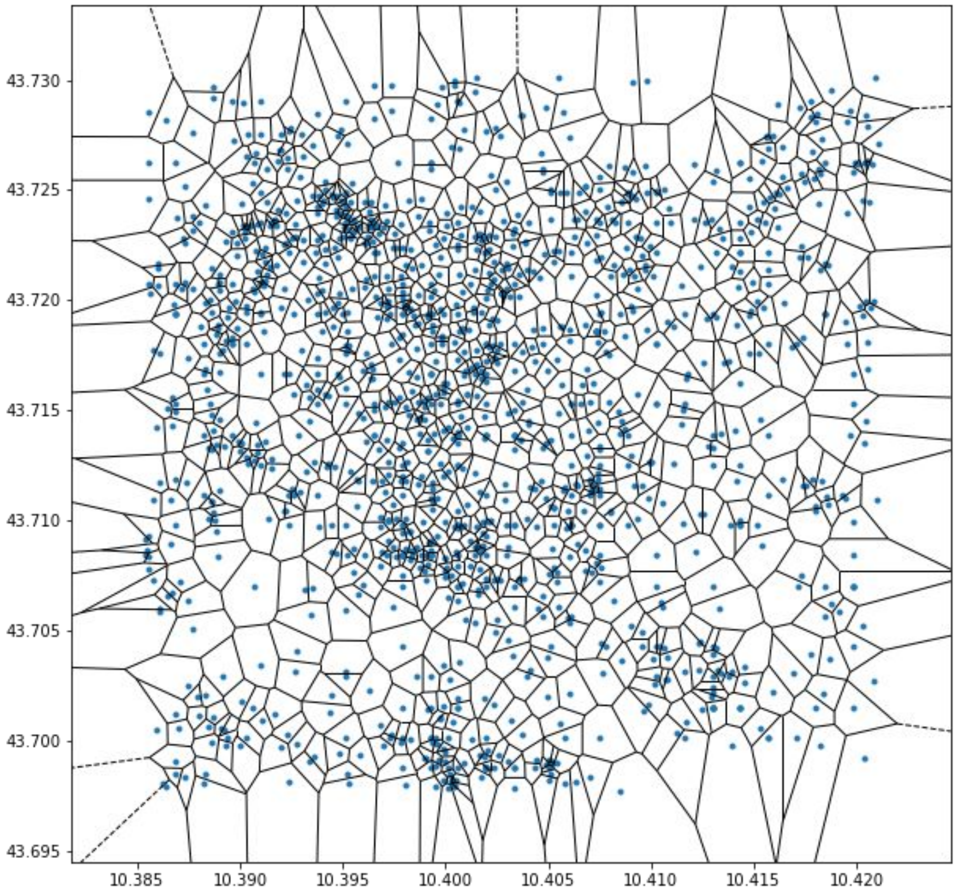
Voronoi tessellation of GSM cells in Eindhoven



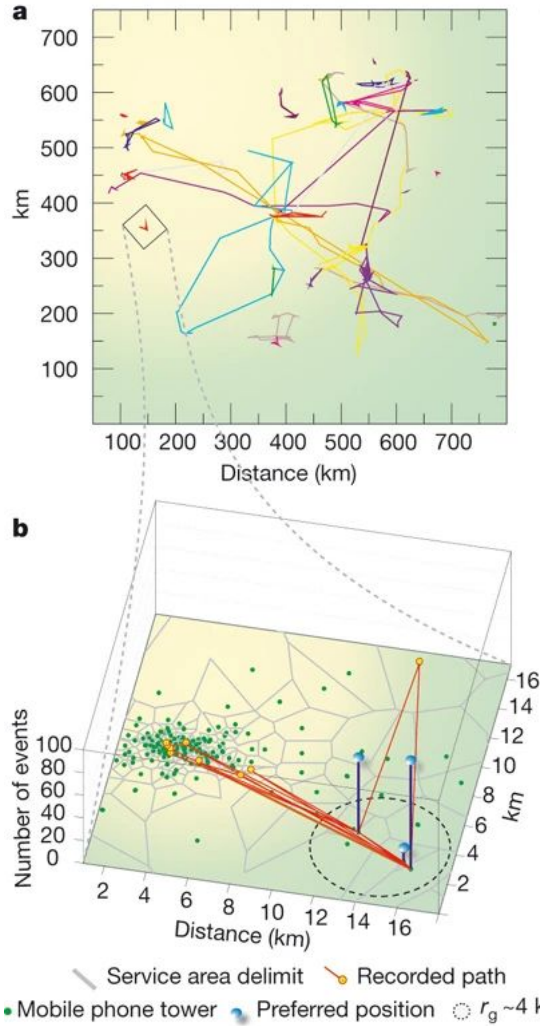


Voronoi tessellation of GSM cells in Pisa



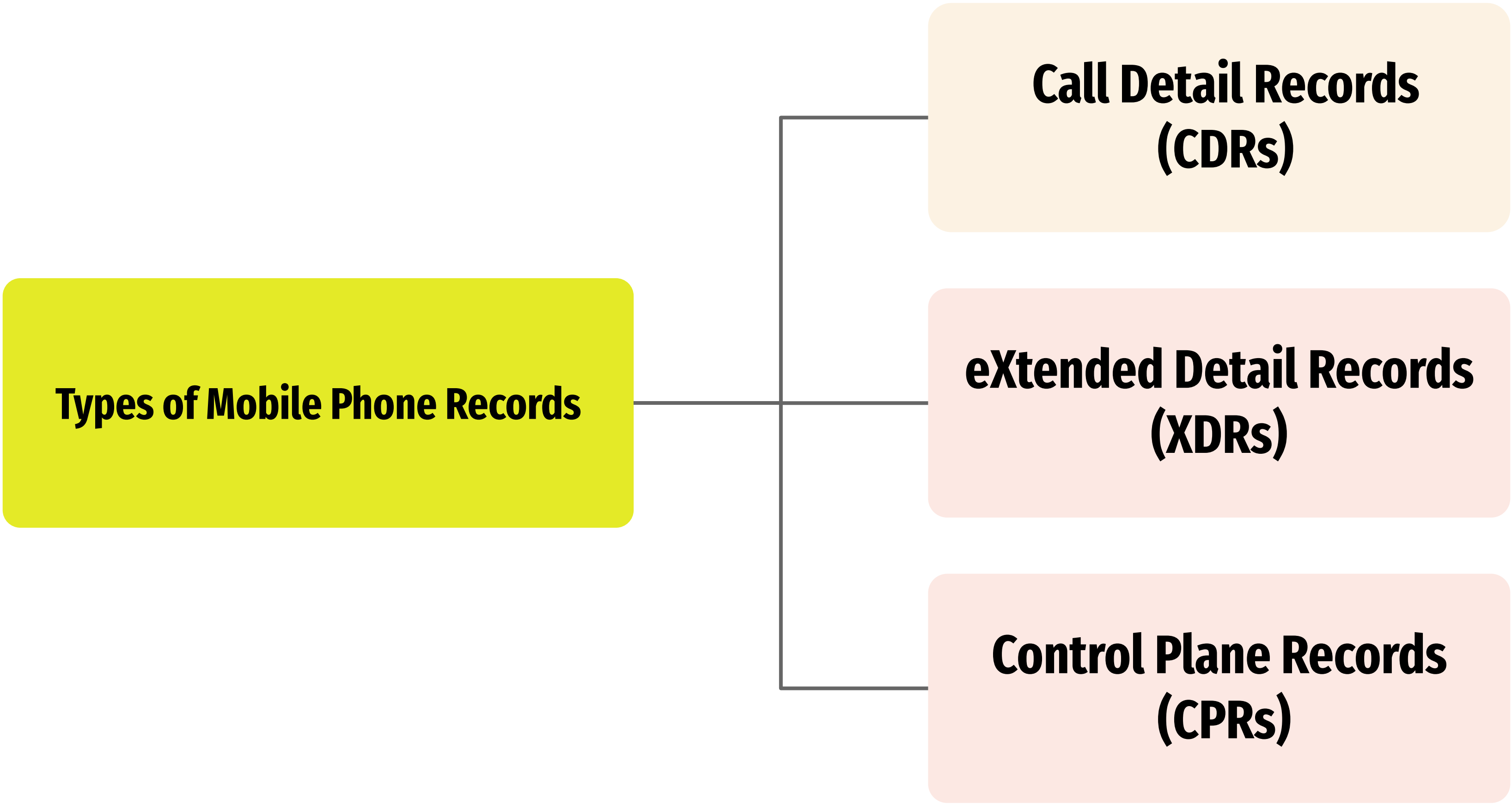


Voronoi tessellation of GSM cells in Pisa



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)


Types of Mobile Phone Records
Call Detail Records (CDRs)
eXtended Detail Records (XDRs)
Control Plane Records (CPRs)

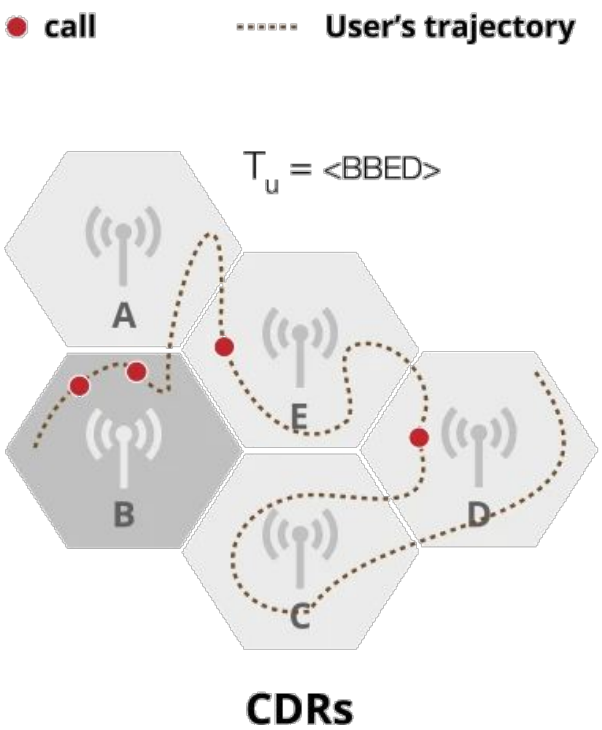
Call Detail Records (CDRs)

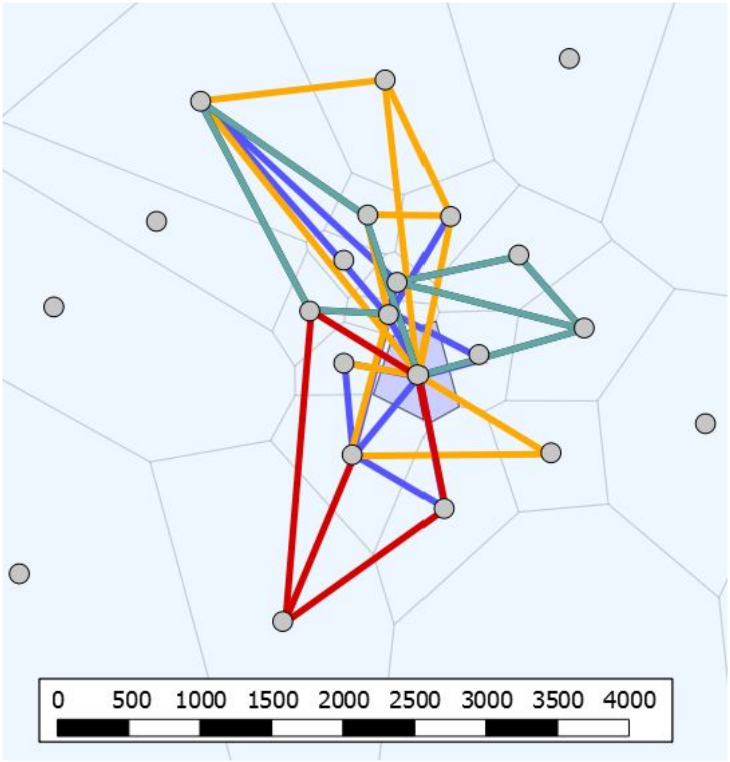
- collected for billing purposes
- purely **human-triggered**, thus sparse
- record created at every call/SMS

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

caller callee time caller antenna callee antenna duration

Call Detail Records (CDRs)





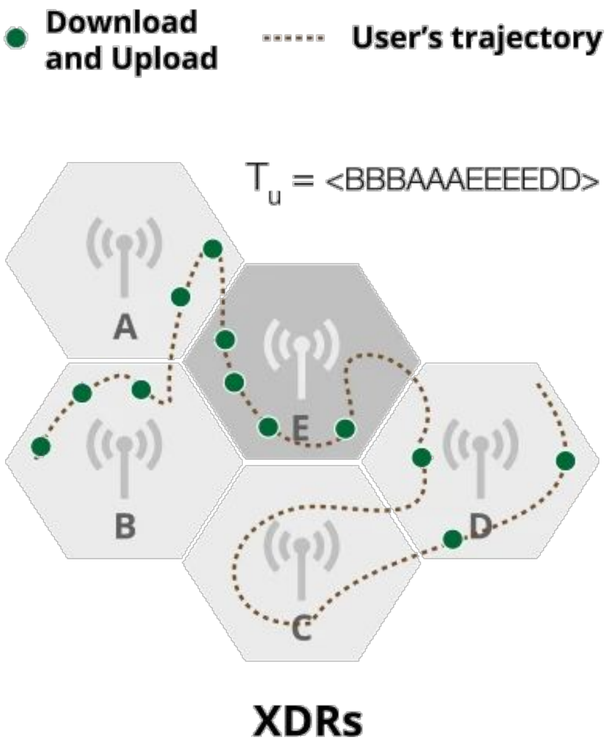
eXtended Detail Records (XDRs)

- 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

eXtended Detail Records (XDRs)



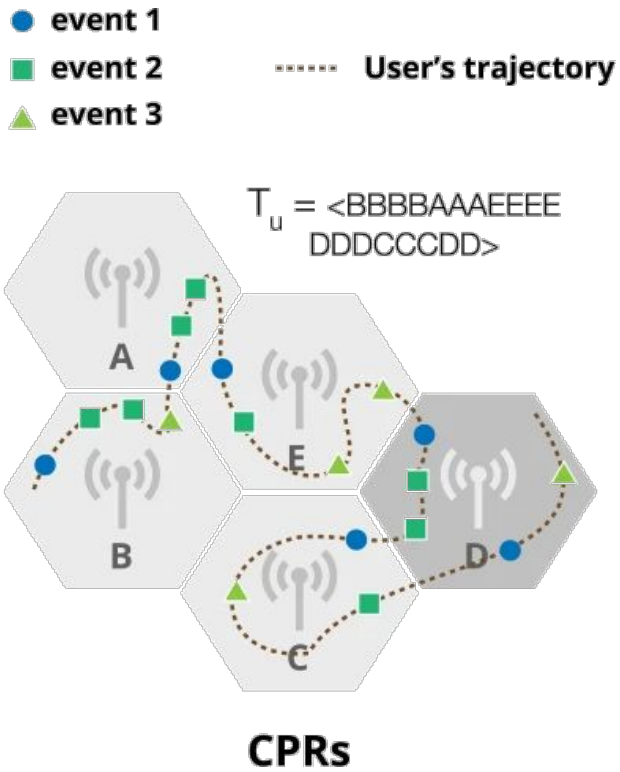
- better temporal resolution than CDRs
- more suitable than CDRs nowadays

Control Plane Records (CPRs)

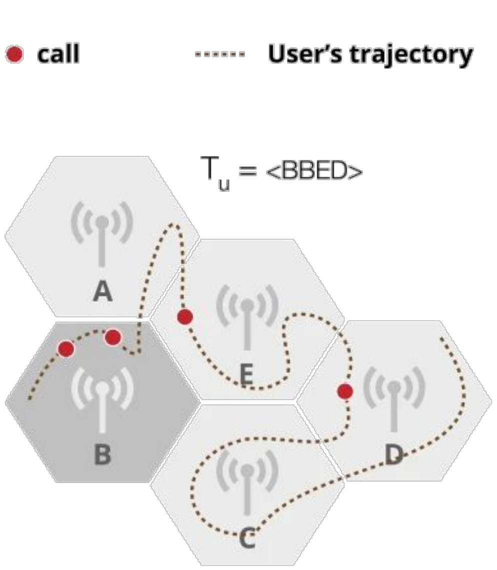
- used to check network “health”
- network-triggered
- record is created, e.g., when a device is attached to a new antenna (“handover”)

$$(\underset{\text{user}}{n}, \underset{\text{time}}{t}, \underset{\text{antenna}}{A}, \underbrace{e_1, e_2, \ldots, e_n}_{\text{events}})$$

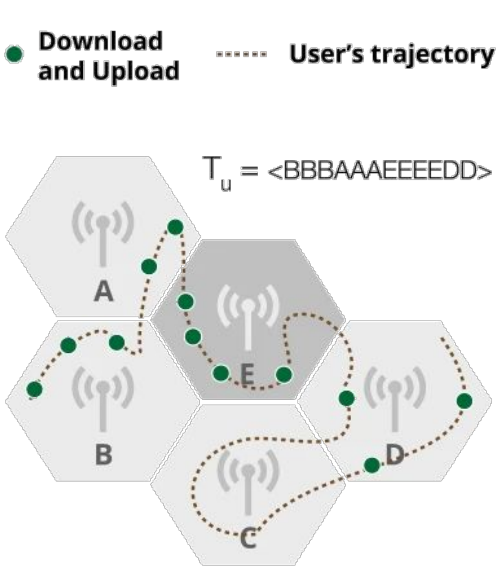
Control Plane Records (CPRs)



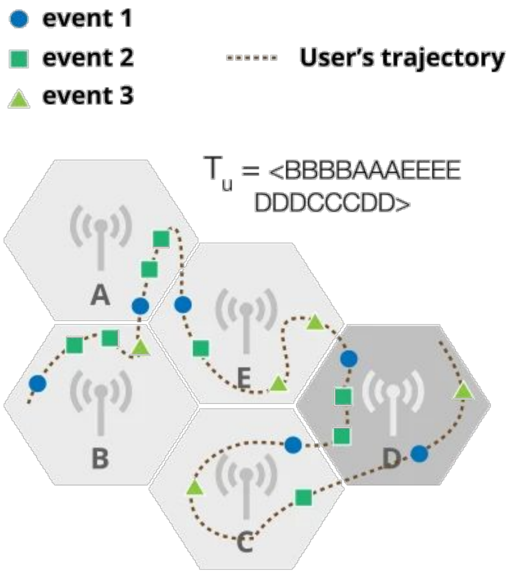
- better temporal resolution than CDRs
- rarely available
- may suffer from the ping-pong effect



CDRs



XDRs

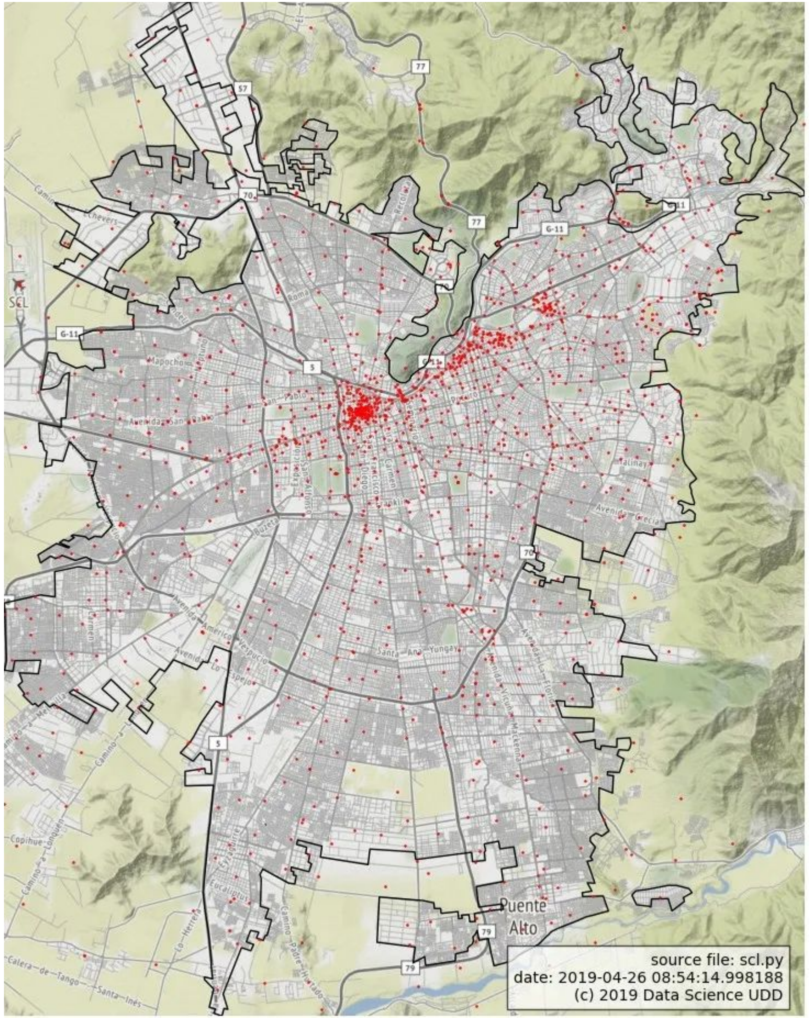


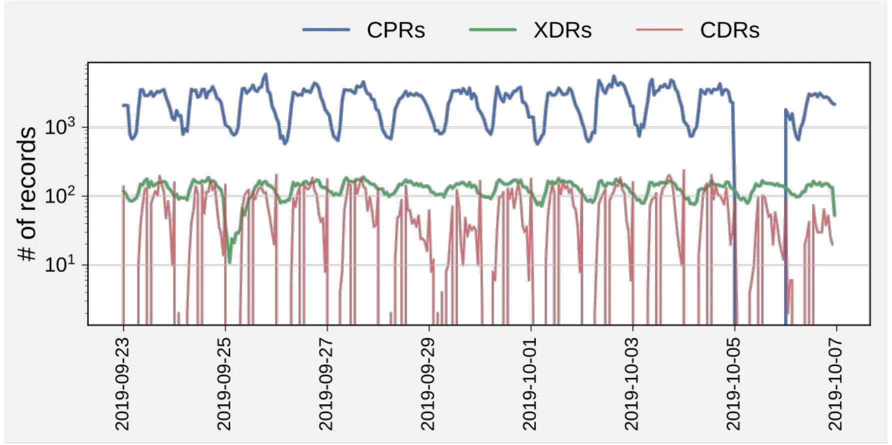
CPRs

4 records

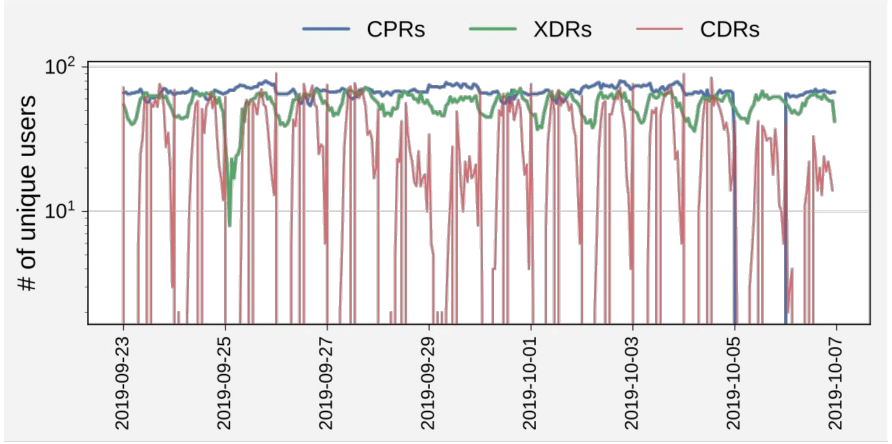
12 records

19 records

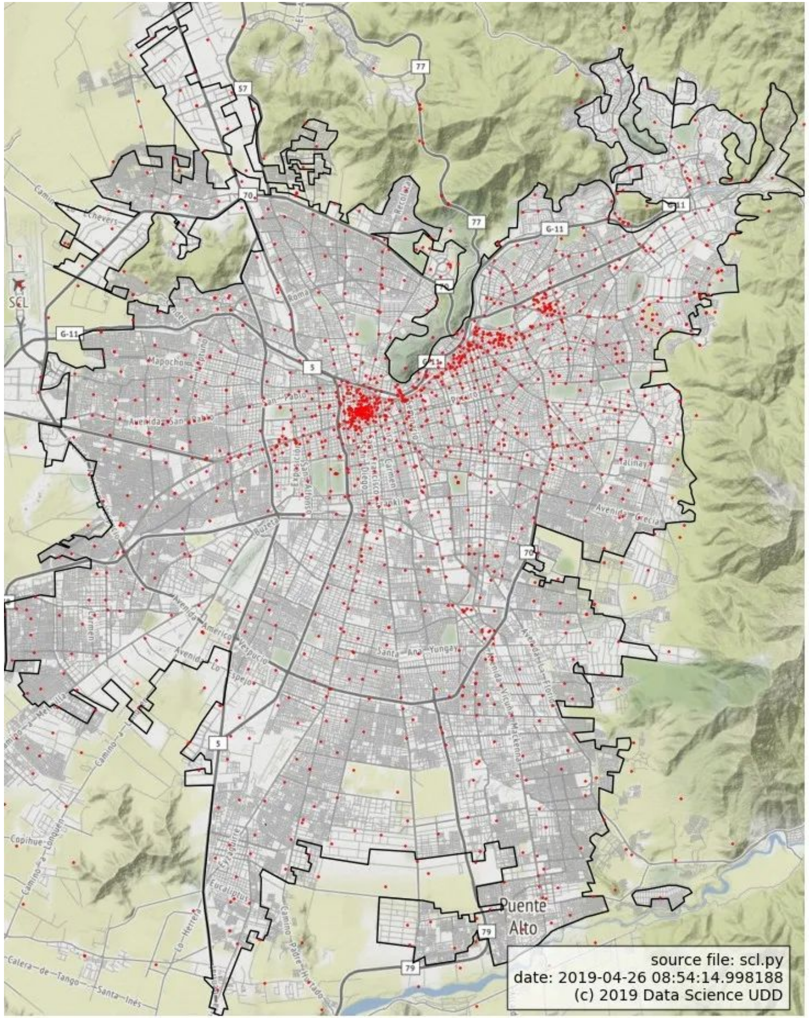


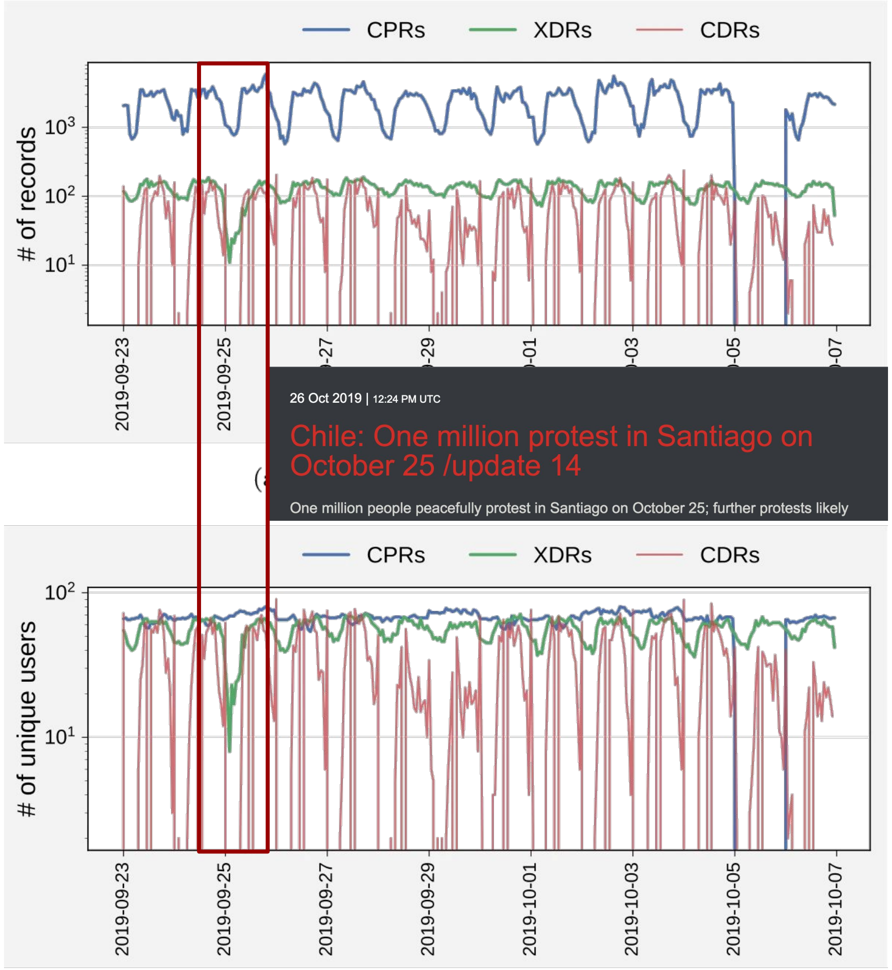


(a) log number of events per hour



(b) log number of unique users per hour





(b) log number of unique users per hour

PROs vs CONs

PROs

- ubiquity
- huge sample size
- rich and multidimensional
    - space, time, social

CONs

- not publicly available
- sparsity
- low spatial precision
- ping-pong effect

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

Location-based social networks

Social Network + Location Data

Social Networks

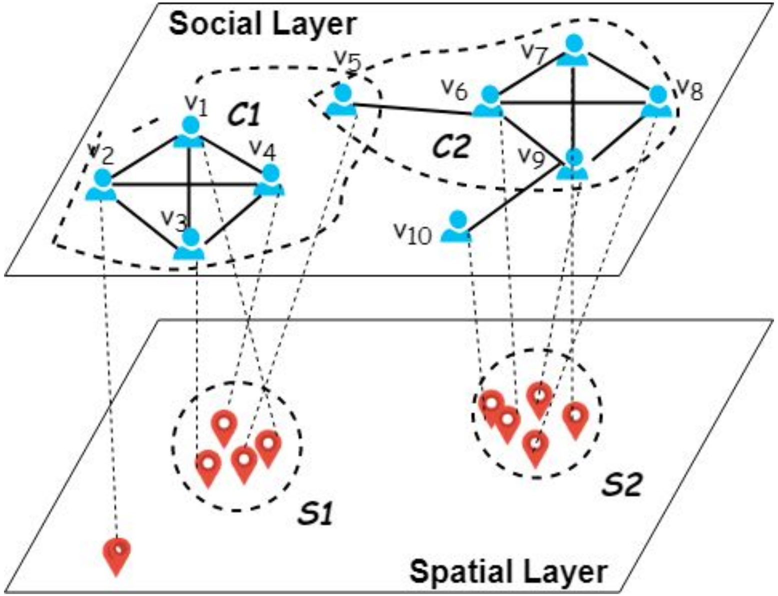
Social structures made up of individuals connected

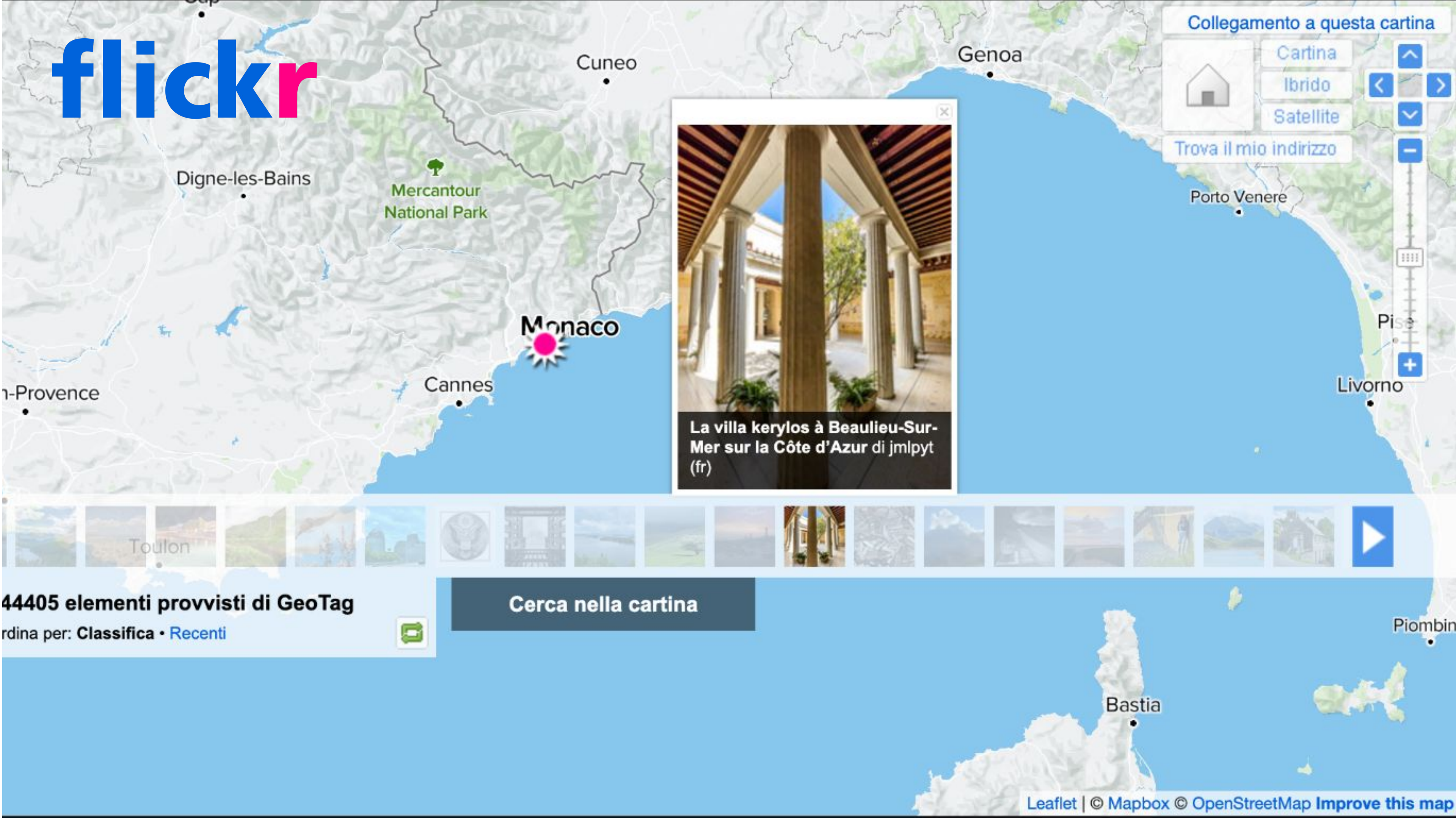
Location Data

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

Location-based social networks (LBSN)

New social structure made up of individuals connected derived from their locations in the physical world as well as their location-tagged media content.




flickr
Cuneo
Genoa
Collegamento a questa cartina
Cartina
Ibrido
Satellite
Trova il mio indirizzo
Digne-les-Bains
Mercantour
National Park
Porto Venere
Monaco
Cannes
-Provence
Livorno
La villa kerylos à Beaulieu-Sur-Mer sur la Côte d’Azur di jmlpyt (fr)
Toulon
44405 elementi provvisti di GeoTag
rdina per: Classifica • Recenti
Cerca nella cartina
Piombin
Bastia
Leaflet | © Mapbox © OpenStreetMap Improve this map


Patrick Torsney
@ilegal
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 1
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

PROs vs CONs

PROs

- ubiquity
- precise location (error <10m)
- semantic information
- social dimension

CONs

- sparsity (more than MPRs)
- location level only
- self-selection bias (age)

INTERVALLO

Null Island: the busiest place that doesn’t exists

station 13010

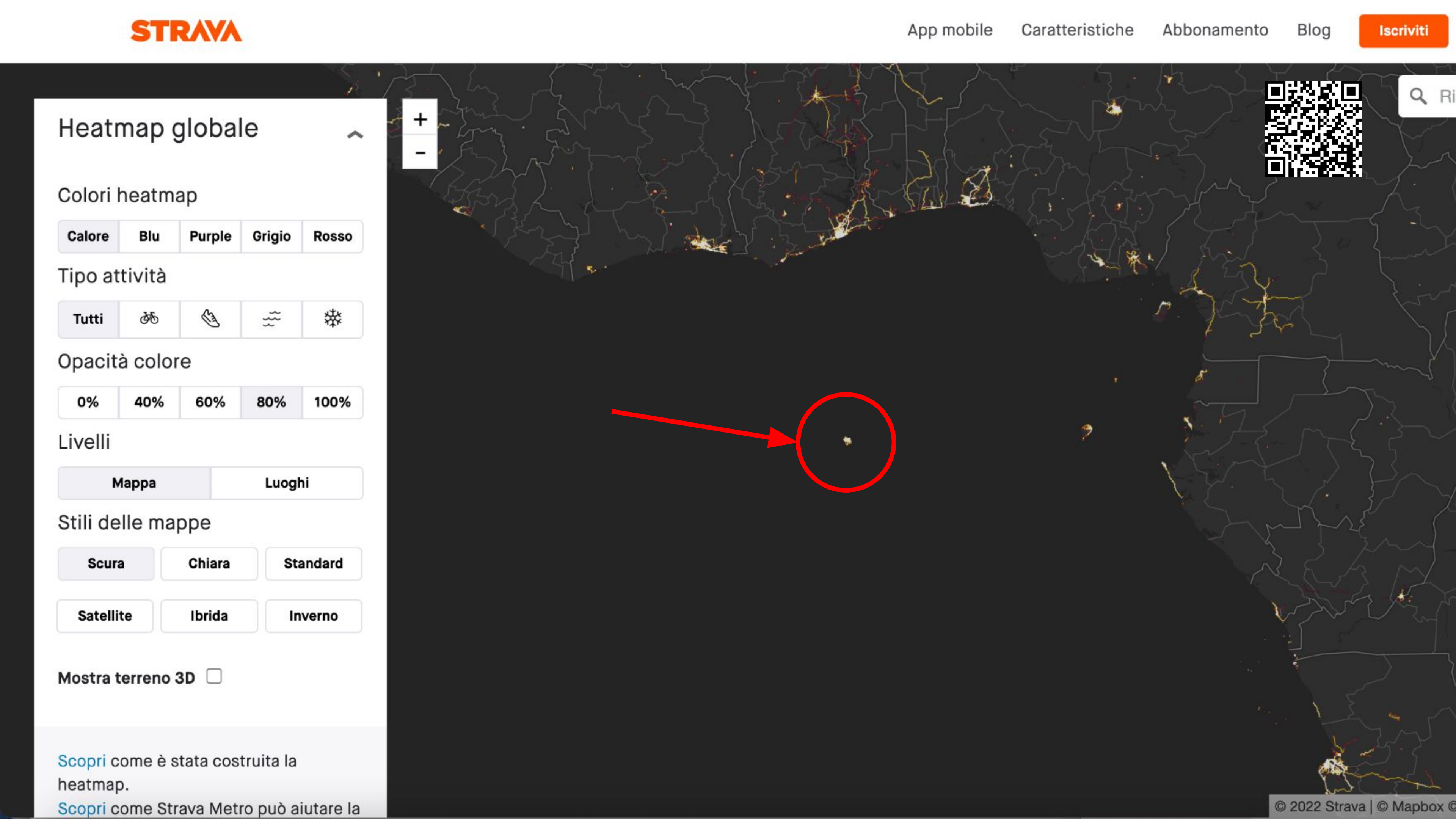
- it’s at coordinates (0,0) in WGS84
- nothing there…but a buoy
- developer Steve Pellegrin first coined the term in 2008

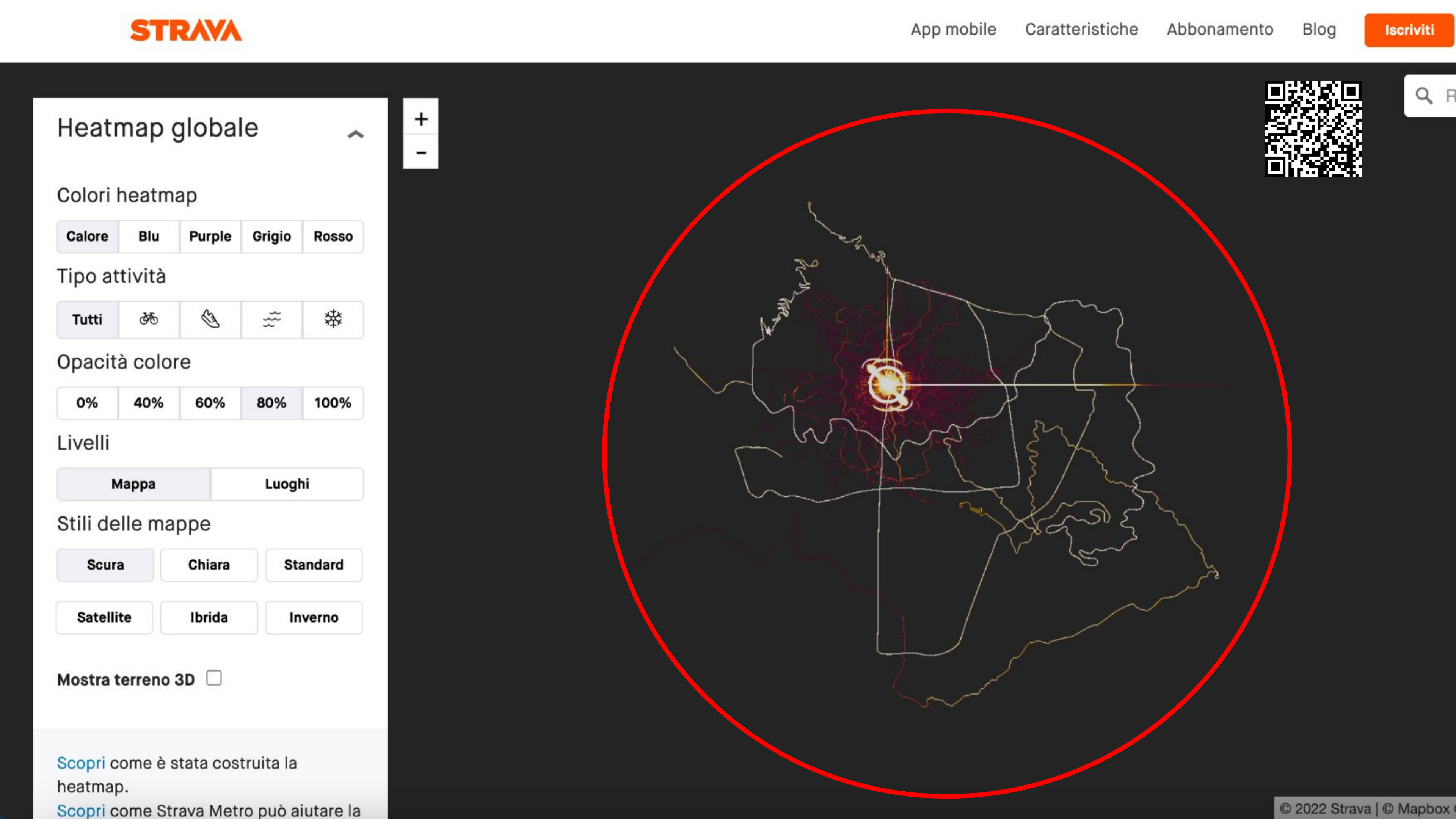
INTERVALLO

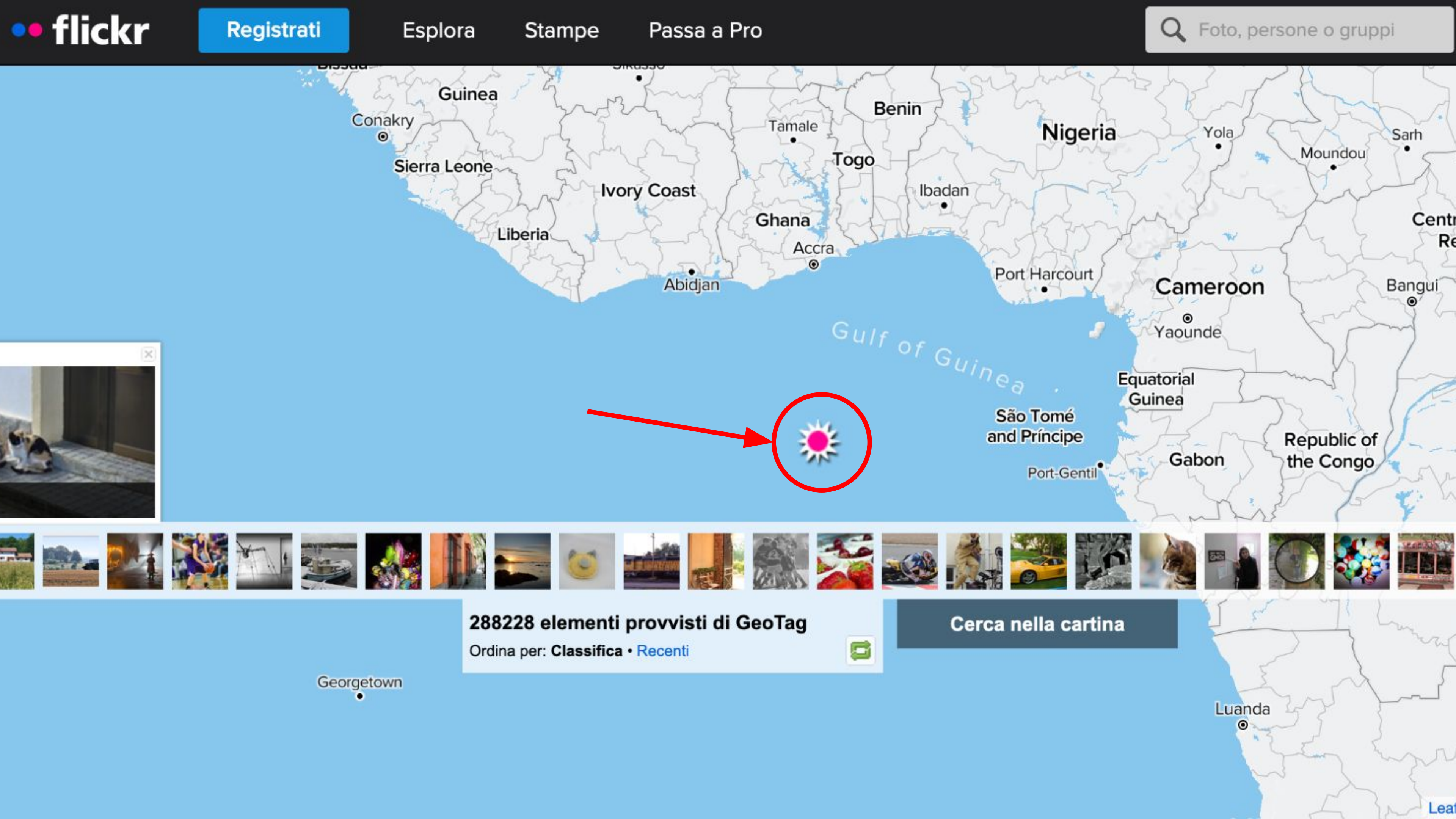
Null Island: the busiest place that doesn’t exists

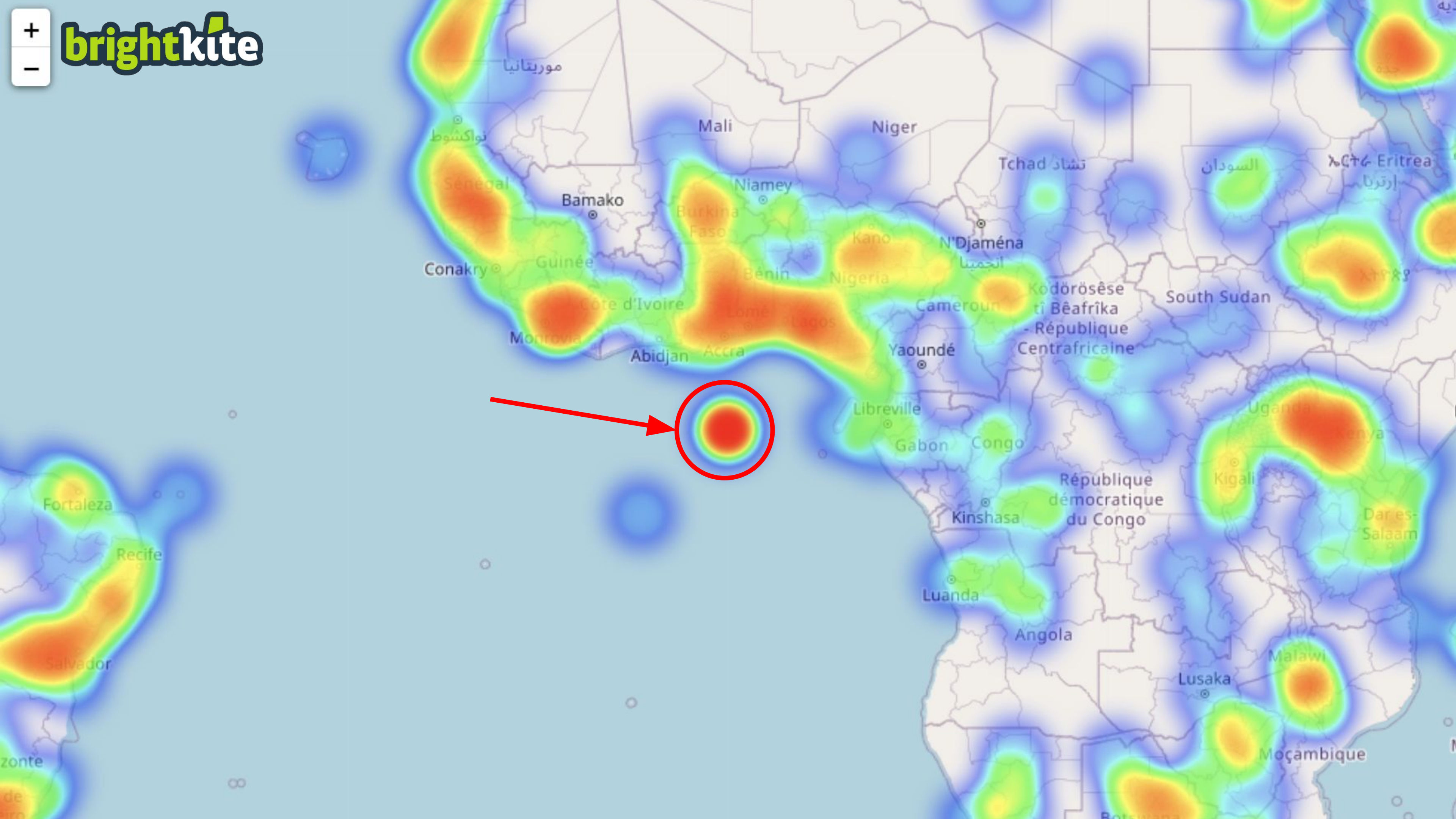
station 13010

- 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”


STRAVA
App mobile
Caratteristiche
Abbonamento
Blog
Iscriviti
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
© 2022 Strava | © Mapbox


STRAVA
App mobile
Caratteristiche
Abbonamento
Blog
Iscriviti
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
© 2022 Strava | © Mapbox


flickr
Registrati
Esplora
Stampe
Passa a Pro
Foto, persone o gruppi
Guinea
Conakry
Sierra Leone
Liberia
Ivory Coast
Abidjan
Ghana
Accra
Tamale
Togo
Benin
Nigeria
Ibadan
Port Harcourt
Yola
Moundou
Sarh
Cameroon
Yaounde
Bangui
Gulf of Guinea
Equatorial
Guinea
São Tomé
and Príncipe
Port-Gentil
Gabon
Republic of
the Congo
Georgetown
Luanda
288228 elementi provvisti di GeoTag
Ordina per: Classifica • Recenti
Cerca nella cartina



brightkite
موريتانيا
نواكشوط
Mali
Niger
Tchad تشاد
السودان
Eritrea
Sénégal
Bamako
Niamey
Burkina Faso
N'Djaména
Kano
Conakry
Guinée
Bénin
Nigeria
Kodörösêse tî Bêafrîka - République Centrafricaine
South Sudan
Côte d'Ivoire
Lomé
Lagos
Cameroun
Monrovia
Abidjan
Accra
Yaoundé
Libreville
Gabon
Congo
Uganda
Kenya
Kigali
Fortaleza
Recife
République démocratique du Congo
Kinshasa
Dar es Salaam
Luanda
Angola
Salvador
Malawi
Lusaka
Moçambique

REFERENCES

@nullisland

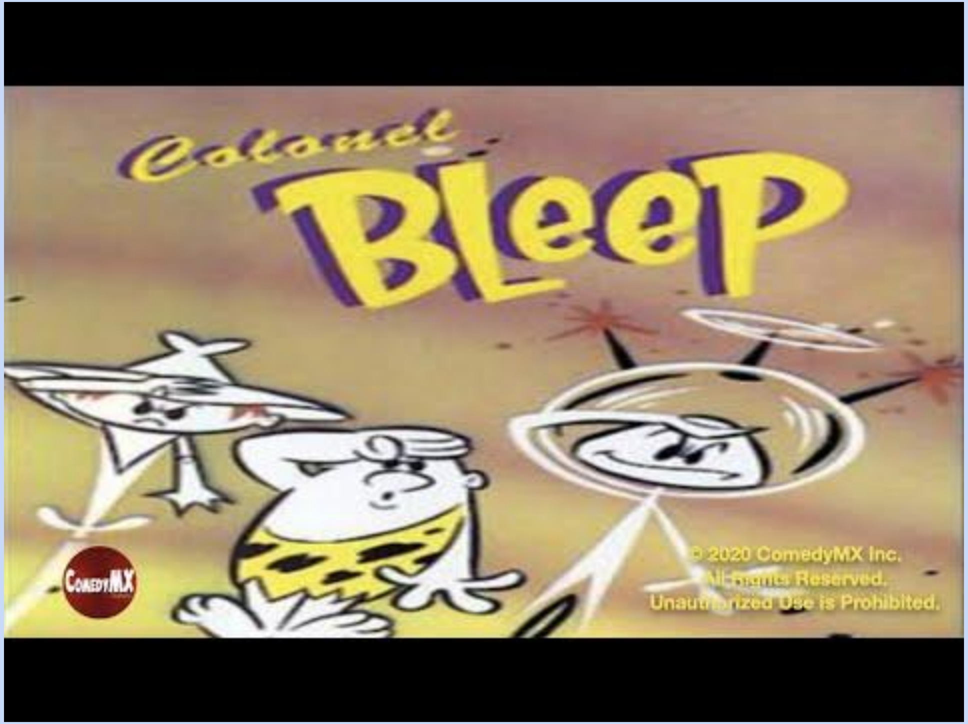
- [article] MapLab: Inside Null Island, 'the Most Real of Fictional Places', Bloomberg, CityLab, 2022
- [article] L'isola dell'Atlantico creata dai dati sbagliati, Il Post, 2022
- [social media post] I think I discovered a secret Chinese military base in the middle of the ocean, Reddit post, 2021
- [blog 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

@nullisland

- [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] Nill points: map of all null islands, lakes and holes, Kenneth Field

REFERENCES



Lesson 03 - 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

Mobility Data: a comparison

GPS traces

Pros

- high resolution
- dense traces
- small samples

Cons

- rarely public
- no semantics
- localization errors

Mobile Phone Records

Pros

- multidimensional
- large samples

Cons

- non-public
- sparse traces
- low resolution
- localization errors

Checkins

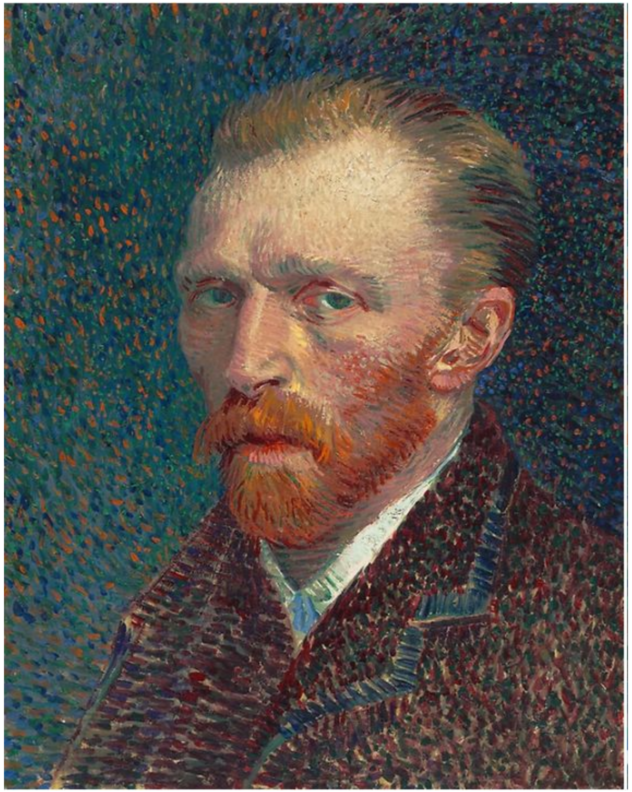
Pros

- precise localization
- semantic info
- multidimensional
- available (e.g., APIs)

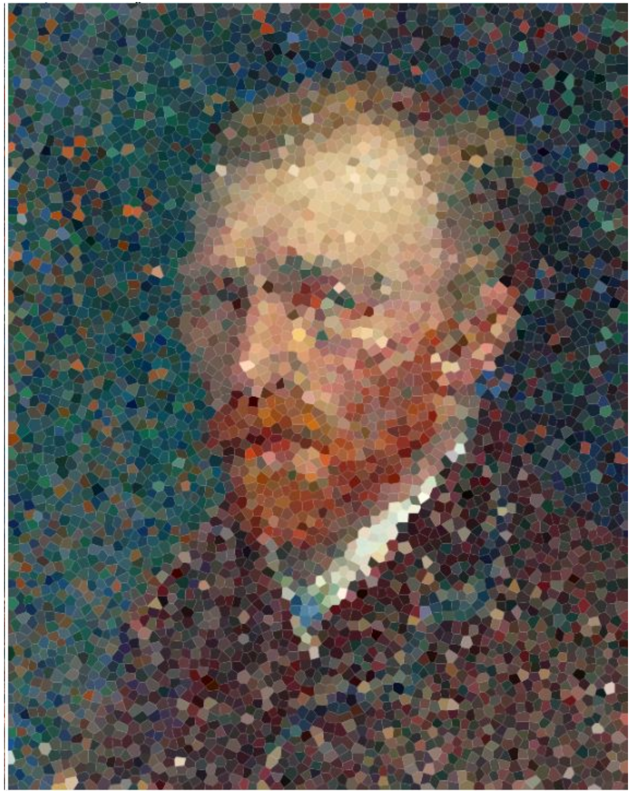
Cons

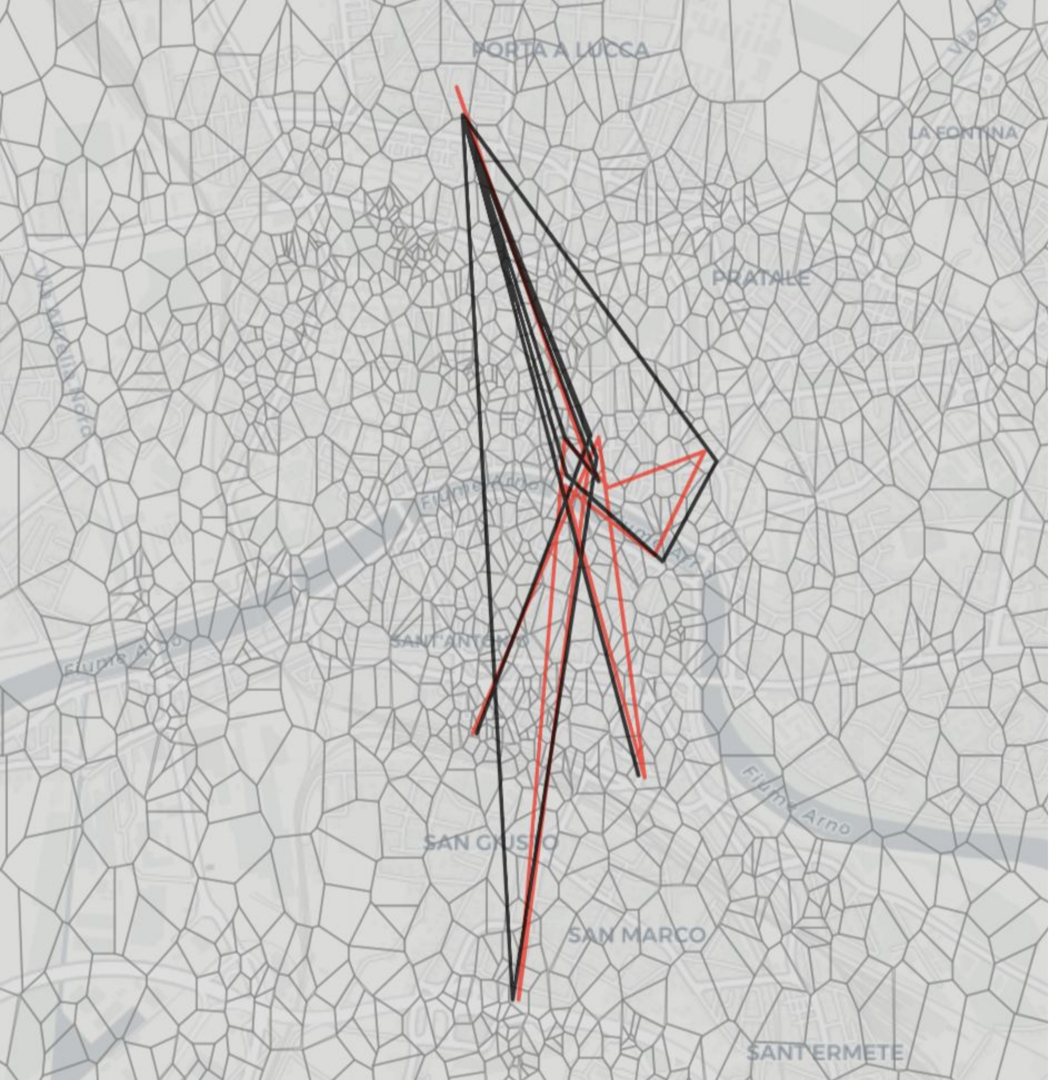
- sparse traces
- self-selection bias

GPS traces

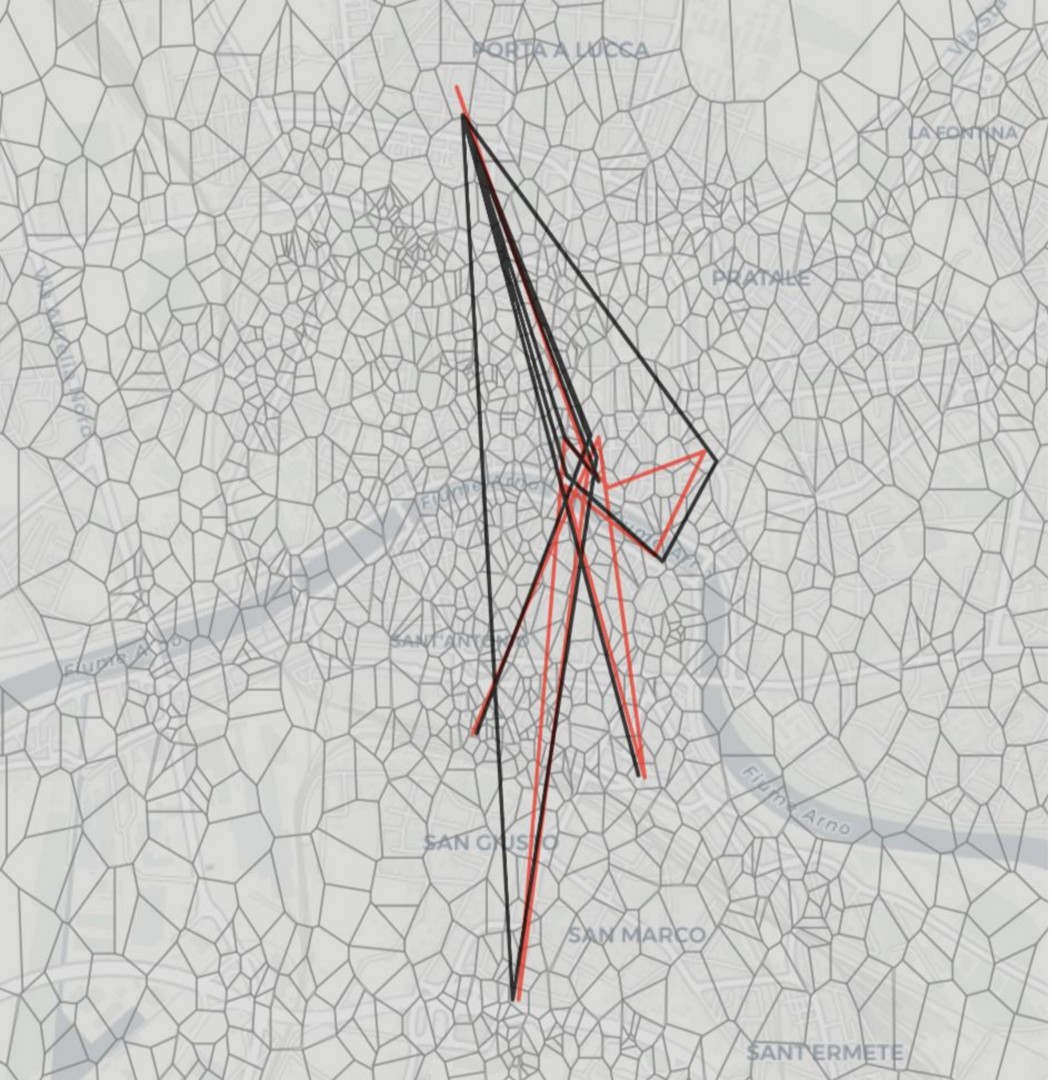


Phone & LBS data





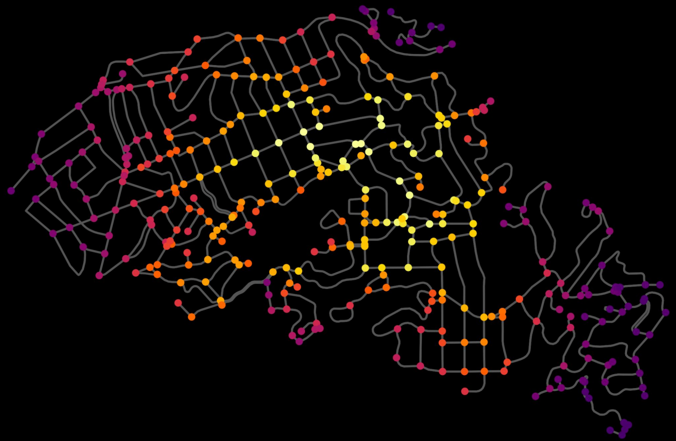
Black: GPS trajectory
Red: CDR-like trajectory



Black: GPS trajectory
Red: CDR-like trajectory

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

Road Networks and POIs



Road Networks

- Describe streets and intersections in cities
- Typically represented as a multigraph, where nodes are intersections and edges are roads
- Nodes’ and edges’ shapes are geometries (Points, Lines, or Polygons)

Data format: intersections (nodes)

$$(id, lat, lng, streets, geometry)$$

- **id**: intersection identifier
- **lat, lng**: geographic coordinates
- **streets**: number of crossing streets at that intersection
- **geometry**: geometric shape (typically a Point)

Data format: roads (edges)

$$(id, u, v, \underbrace{f_1, \ldots, f_k}_{\text{characteristics}}, geometry)$$

- **id**: road identifier
- **u**: origin node
- **v**: destination nodes
- **f_1, ..., f_k**: road characteristics, such as if the road is oneway and reversed, its length, number of lanes, name, and max speed
- **geometry**: geometric shapes (typically a Line)

Road Networks

They allow to:

- associate GPS traces or locations to the street they belong to
- find routes (e.g., shortest path) between two places
- simulate traffic in cities (e.g., with SUMO)

OSMnx allows downloading and analyzing them easily.

Points of Interest (POIs)

- Describe geospatial entities in the city
  - E.g.: grocery stores, transit stops, restaurants, building footprints

- Typically represented as Points or Polygons

- Used to describe activity in cities and infer the associated human behaviour

Data format

$$(id, f_1, \ldots, f_n, geometry)$$

- **id**: POI identifier
- **f_1, …, f_k**: POI characteristics, such as the road nodes involved, the type, the type of building, name.
- **geometry**: geometric shapes (typically a Point or Polygon)

Quiz

How many satellites are needed at minimum to determine a device’s position in 3D?

a) 2

b) 3

c) 4

d) 6

Quiz

What is the main purpose of Call Detail Records (CDRs)?

a) To measure signal strength

b) For billing purposes ✓

c) For geolocation tracking

d) To monitor network health

Quiz

Which record type is network-triggered rather than user-triggered?

a) CDR

b) XDR

c) CPR

d) SMS

Quiz

What is the “ping-pong effect”?

a) The fluctuation of the GPS signal

b) Frequent switching between antennas

c) Data duplication between users

d) The latency of data transmission

Quiz

Which of the following is not typically a data attribute in LBSNs?

a) User identifier

b) Location coordinates

c) Antenna ID

d) Post text

Quiz

What is “Null Island”?

a) An island used as a reference point for GPS calibration

b) A fictional location at (0°, 0°) often caused by geocoding errors

c) The world’s first geolocated social network dataset

d) A satellite base for global positioning

Homeworks

Homework 3.2

Download Cell Towers data about France from OpenCellID. *Then, create and visualize (in folium) a Voronoi tessellation of GSM towers in Lyon, France. Repeat for another French city.*

- Use the geovoronoi package to create the Voronoi tessellation
- Plot the tessellation with folium
- (Use just a random sample if the towers if they are too many to be visualized and/or to compute the tessellation)

- Submit a (well commented) python notebook

Homework 3.3

What would the Italian regions look like if the territorial division were done using a Voronoi tessellation based on the position of the regional capitals? What about provinces?

- Download the positions of Italian cities (comuni)
- Use the geovoronoi package to compute Voronoi shapes
- Plot the tessellation with folium

- Submit a (well commented) python notebook

Homework 3.4

Download your positions from Google Maps *for a short period and one city. Plot the corresponding GPS trajectory. How would the trajectory be if the resolution would be imposed by mobile phone towers? Compute an error metric between your GPS and CDRs-like trajectories*

- Use the geovoronoi package to compute Voronoi shapes
- Tip: map the trajectory to the Voronoi tessellation, and take centroids as visited points
- Plot the tessellation, the original trajectory and the transformed trajectory with folium
- (As error metric, you may use, e.g., the RMSE)
- Submit a (well commented) python notebook

Homework 3.5

Can you find any null island, lake, or black hole

in your country of origin?

- Use the nill points map to find them (if any)
- What are the coordinate systems associated with them, and by which countries are they managed?

- Write a blog post (max 2 pages) about it!

Homework 3.6

Find other digital data sources that are not in scikit-mobility and create a `json` file to add them using the `data` module (doc)

- Make a notebook where you explore the dataset using scikit-mobility functionalities
    - example of processing script

Homework 3.7

Define a trajectory in a dead reckoning way, i.e., as a sequence of visited Points of Interest (so without coordinates) and times to reach a point from the previous one (so, not absolute datatimes) and then compare it with a GPS real trajectory (with coordinates and datetimes).

- What are the pros and cons of using POIs versus GPS traces?

Homework 3.8

Download from figshare this flows dataset, create a tessellation and a `FlowDataFrame`; plot them together using skmob.

- This article describes the dataset.
- Create a json file using the data module (doc) to upload the dataset on skmob

Homework 3.9

Create your own road network: a grid-like city with 10*10 intersections

- Use OSMnx for that, i.e., create appropriate lines and points to represent edges and nodes
- Save it in a proper format

Homework 3.10

Use OSMnx to download the shapes of all cinemas in Rome

- Plot in folium the shape of the municipality of Rome
- Plot the smallest polygon that contains all cinemas in Rome
- Plot the cinemas in folium with a pop-up showing their name
- Compute the cinemas with the lowest and highest average distance to the others, and draw a circle (of radius 100m)

Homework 3.11

Use OSMnx to download the shapes of all pharmacies in Pisa

- Plot in folium the shape of the municipality of Pisa and the pharmacies with a pop-up showing their name
- Compute the distance of each of these pharmacies to the Department of Computer Science (DPC) and to the CNR.
- Compute and visualize the pedestrian shortest path from DPC and its closest pharmacy. Do the same for CNR.

Homework 3.12

Select a data paper from the Scientific Data journal that regards geospatial or human mobility data (tessellations, trajectories, flows, or other useful info) and:

- Make a notebook to explore the dataset, i.e., load it in the proper skmob or geopandas data structure and visualize the data. A possible list of such datasets is here.
- Create a json file using the data module (doc) to upload the dataset on skmob

Homework 3.13

Take all football clubs in London that play in Premier League or the Football League Championship and visualize the position of their official headquarters (as a Marker) and the polygon representing their stadiums.

- Each marker's tooltip should visualize the club's name
- Visualize the shortest path between each club's headquarter and it stadium
- For each club, what's the total distance the club should travel (from its headquarter) to reach the stadiums of all the opponents in the same league?

Homework 3.14

Take the position of your home in Pisa (point A) and the position of the Department of Computer Science (point B). Compute the shortest path p(A, B) between A and B. Now select the roads e_1, e_2 that are the closest to the points P_1 and P_2, which are at a (earth) distance *d* on a straight line *R* that intersects the segment *S* that connects A and B at its middle and that is perpendicular to S.

- Visualize the three paths
- Compute the difference between the two new paths and path p(A, B), in terms of geographic distance and number of roads