Mobile Health

Location Tracking and Public Health

Cecilia Mascolo

UNIVERSITY OF CAMBRIDGE

Radios

- Wearable/mobile device sensors have multiple radios which are used to detect presence in a location as well as contact.
    - Global Positioning System
    - Bluetooth
    - WiFi
    - Cellular
    - RFID
    - …
- Technologies differ broadly in range, architecture and energy efficiency.

Global Positioning System

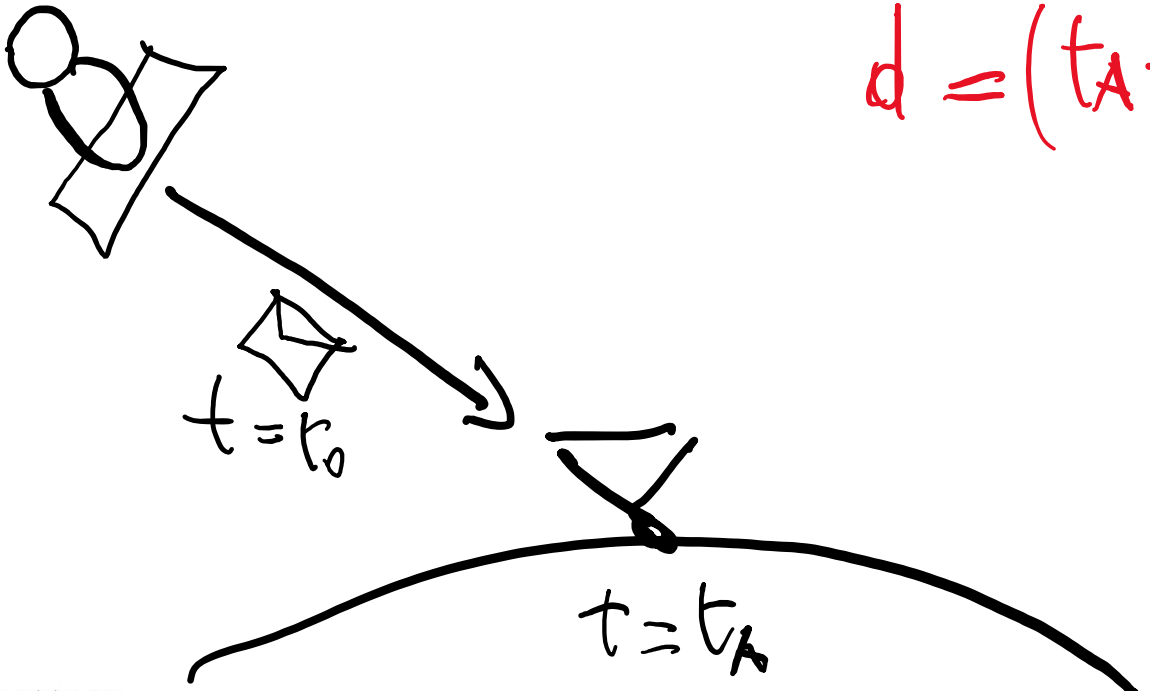
- Global Positioning System (GPS) is US governed.
- Other Global Navigation Satellite Systems (GNSS) exist, e.g., GLONASS (Russian Federation), Beidou (China) and Galileo (European Union).

Satellites

- There are 30+ GPS Satellites.
- Satellites have atomic clock (receivers do not: but can deduce time precisely from the various satellites).
- Their position is known very precisely.
- The time delay between when the satellite transmits and the receiver receives is proportional to the distance between the satellite and the receiver.
- A satellite broadcasts a message (also containing its transmit time ToT).
- Receiver can calculate the time delay of the signal (ie T).

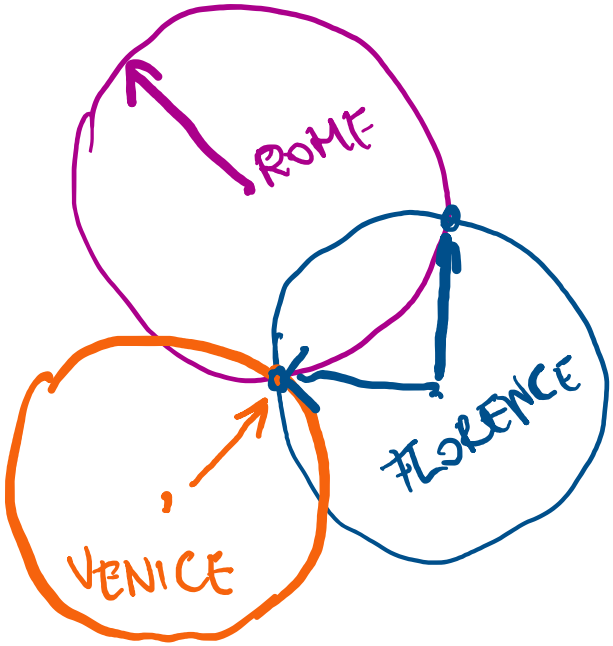
Time of travel x speed of light = distance D

Simplification of Distance Calculation



$$d = (t_A - t_{t_0}) \cdot c$$

2-D and 3-D Triangulation



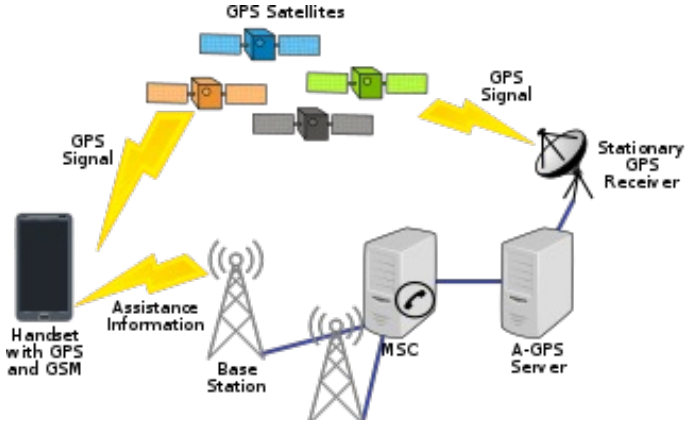
- 3-D triangulation is similar but with spheres.
- 3 Satellites intersect in exactly 2 points.
- The Earth is the 4th sphere and reduces this to exactly 1 point.

How does a receiver know the position of satellites?

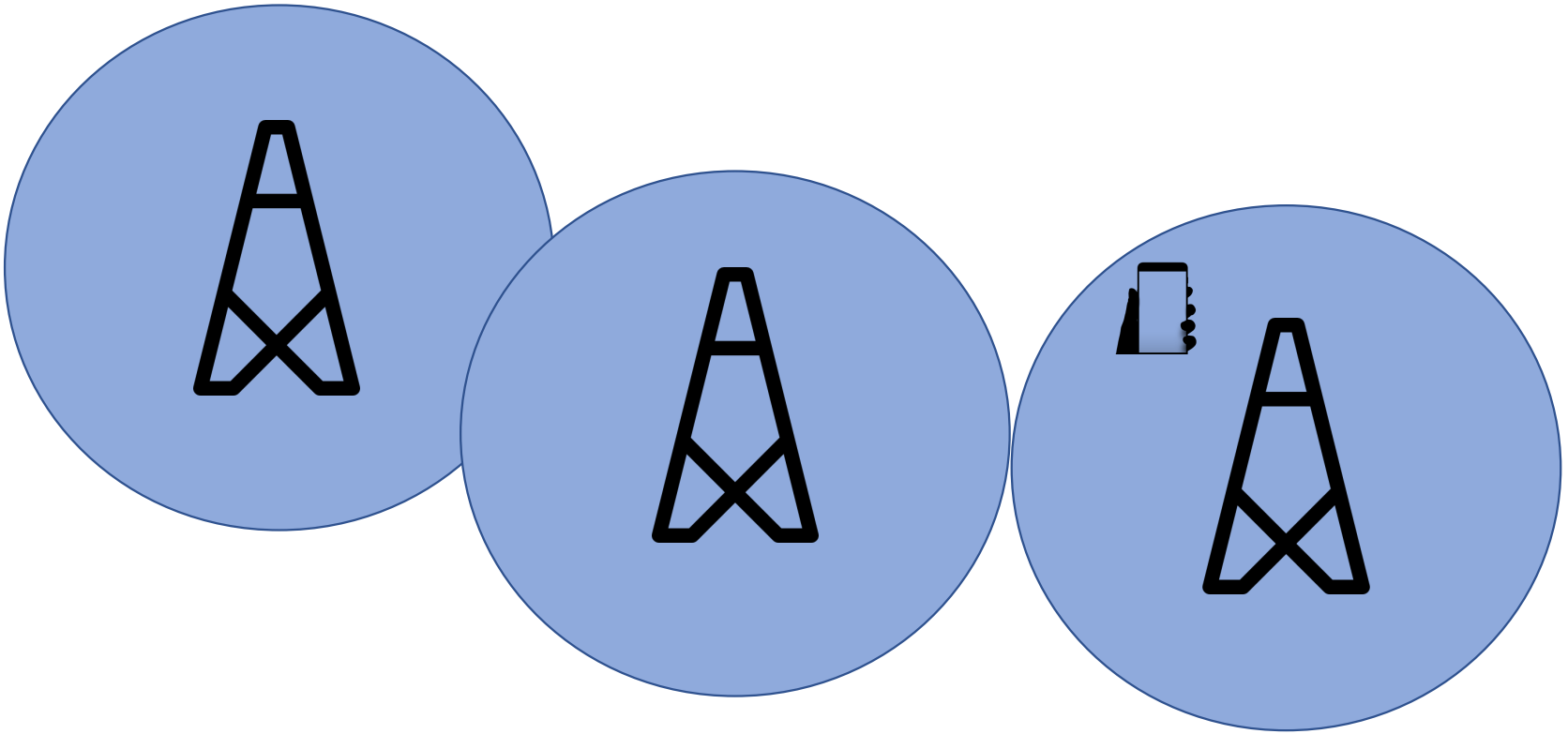
- Satellites are on fixed trajectories.
- An almanac of satellites and approximate orbital information are downloaded by the receivers.
- Ground Control station computes corrections to satellite trajectories (these are also downloaded by receivers regularly).

Assisted GPS

- GPS receivers take time to start as they have to download the almanac, or download trajectories (if almanac has already been downloaded).
- Trajectory corrections updated every 4 hours.
- **Assisted GPS** simply sends this additional information through a data link (cellular or WiFi) to the receiver making it faster and making the GPS always ready.



Cellular Networks



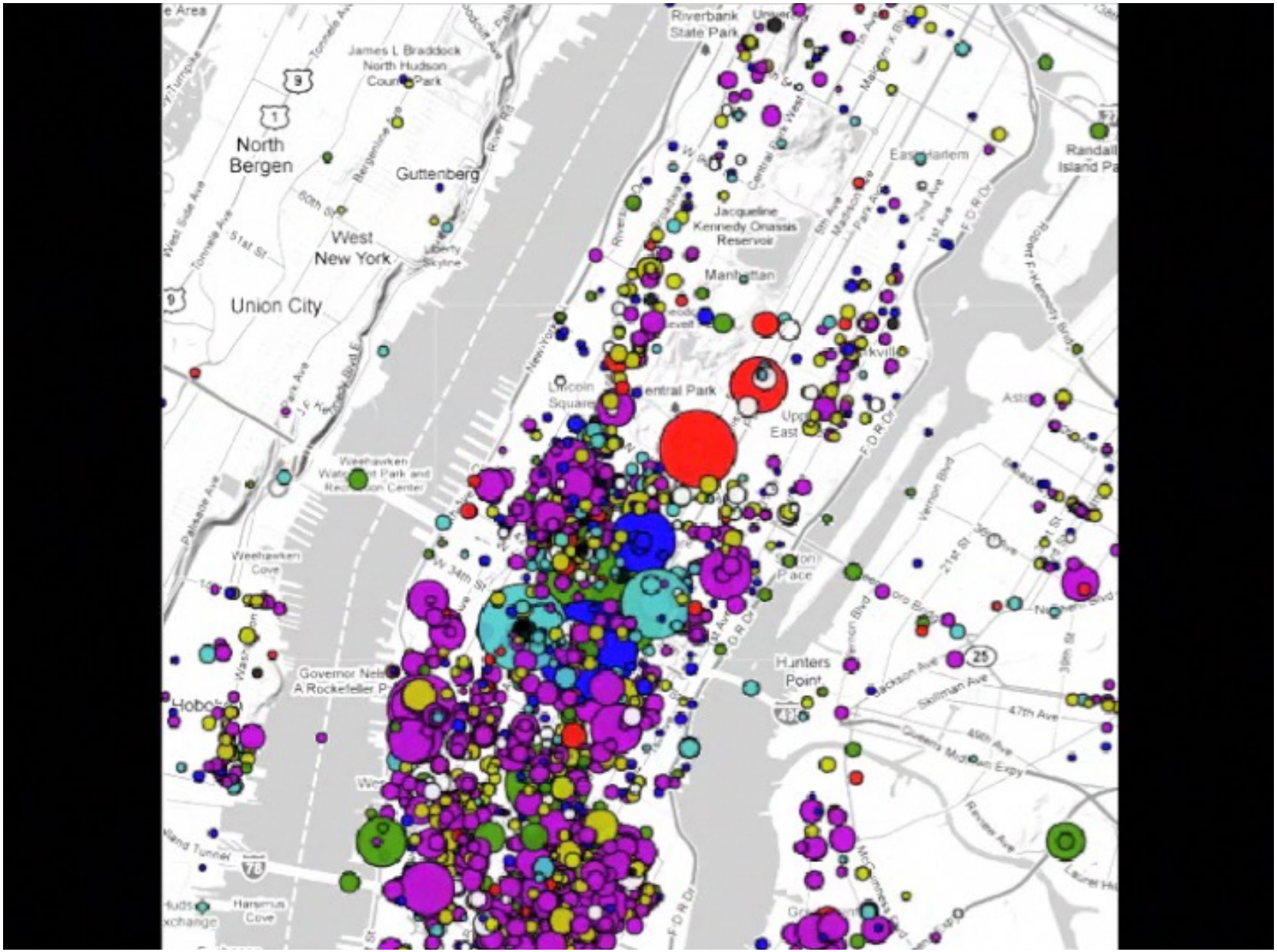
Cellular Networks and Location

- Triangulation can be done if you know where the basestations are, to infer phone locations

- Basestation density changes depending on areas (eg more dense in cities)

FOURSQUARE CHECK-INS SHOW THE PULSE OF **TOKYO**

UNIVERSITY OF CAMBRIDGE



UNIVERSITY OF CAMBRIDGE

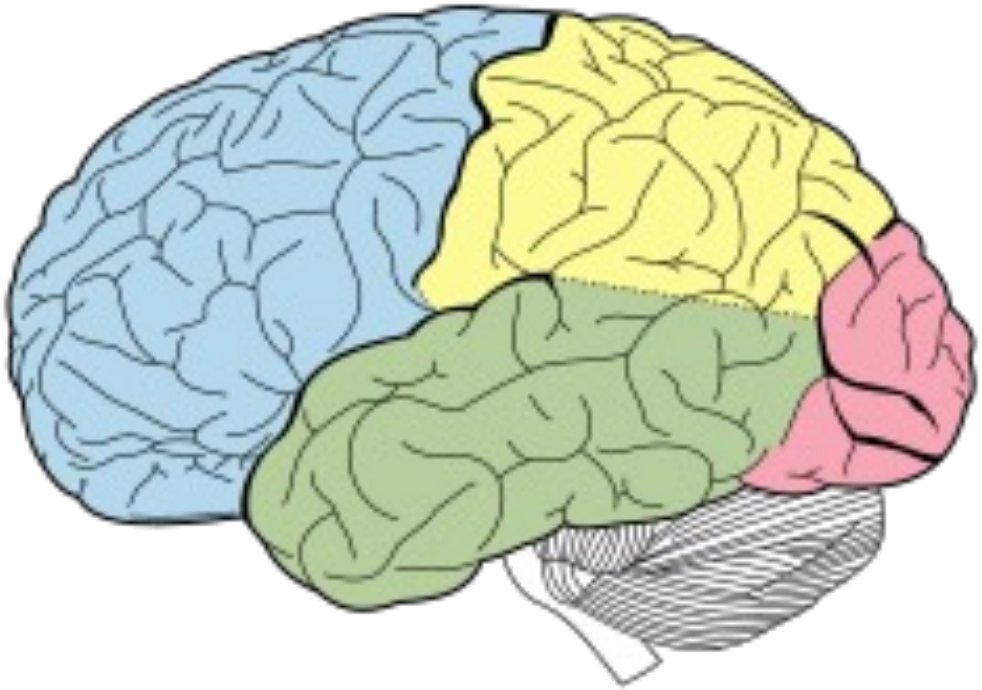
Applications

- Navigations and dementia
- COVID-19
- Malaria tracking

Detecting Alzheimer's Disease (AD) Early from Outdoor Mobility Traces

UK has 680k AD patients costing £23B to the economy

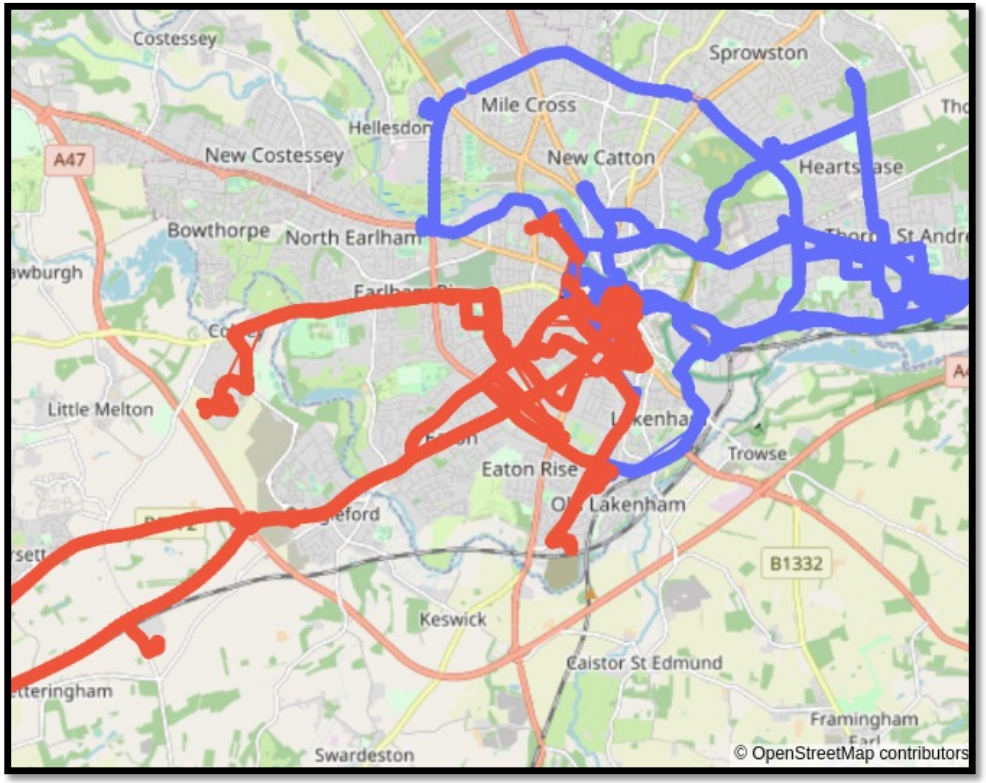
Navigation and AD



UNIVERSITY OF CAMBRIDGE

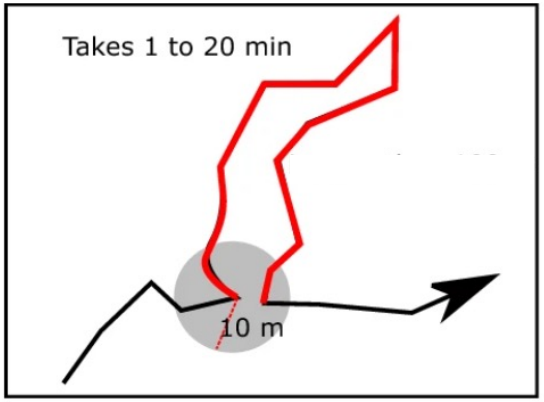
Outdoor Navigation

- 33 participants at MCI stage
- Mini-ACE test results: 15 AD patients and 18 controls
- Used GPS tracker devices
- Tracked for 15 days



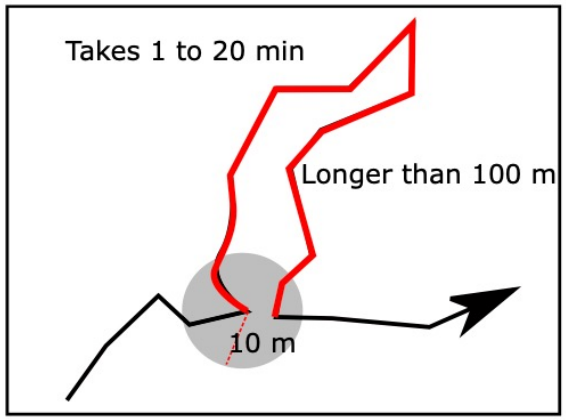
Preprocessing: Extract returning segments

- Returning to a place is driven by Entorhinal cortex
- This is affected early by AD



Trajectories

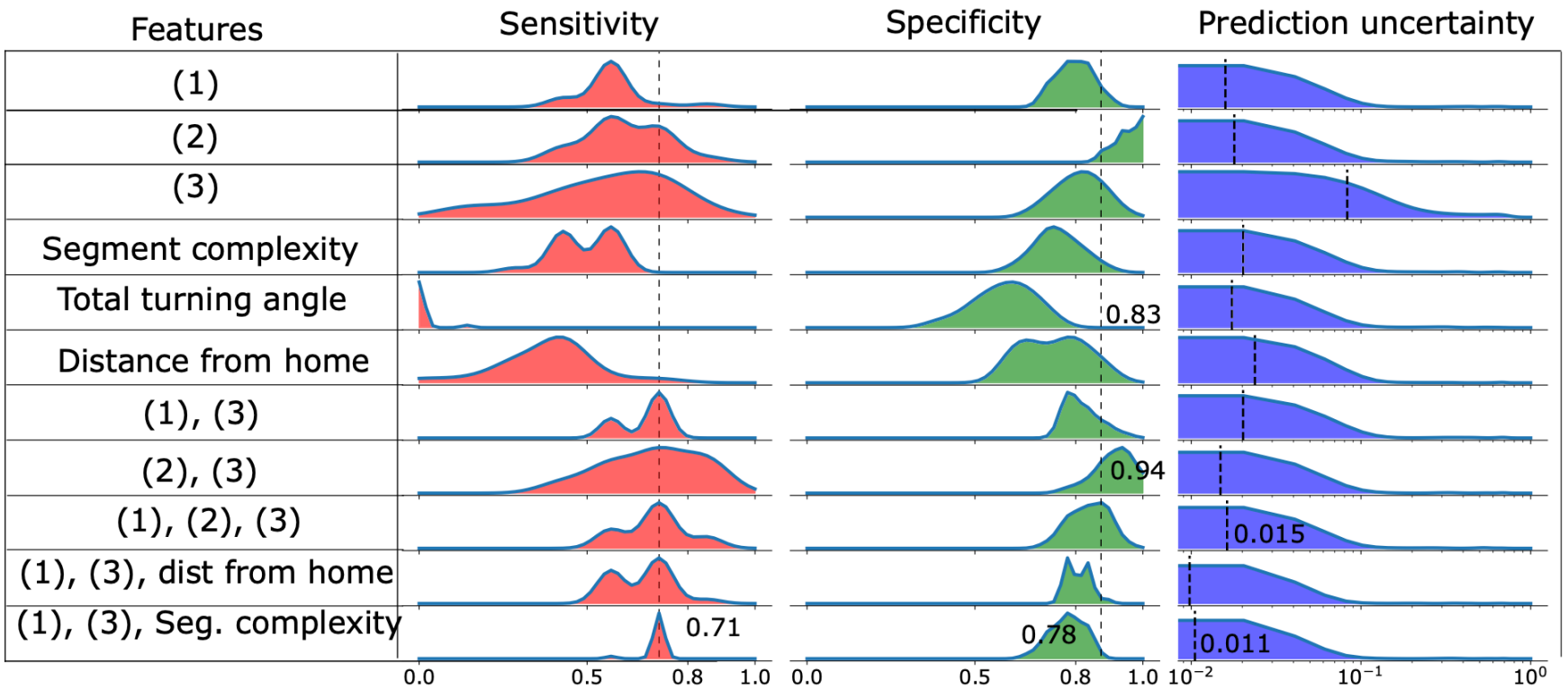
- Segments



- Features

Spatial Feature	Description	Intuitive interpretation
Mobility domain characterization		
Segment similarity	# of segments that crosses more than half the same cells as this segment	Weather the same route is used in navigation
Entropy	Shannon entropy of the distribution of the cells visited by a person	Coverage of the movement environment
Distance from home	Euclidean distance from segment centroid to home location	Affinity towards home
Temporal characterization		
Duration of stops	Duration of stops (more than a min) in a segment	
Shape characterization		
Total turning angle	Sum of the sin of the turning angles	Measures the complexity of the shape of a segment
Segment complexity	# of times a person turned more than 120^oin a segment	Another measure of complexity

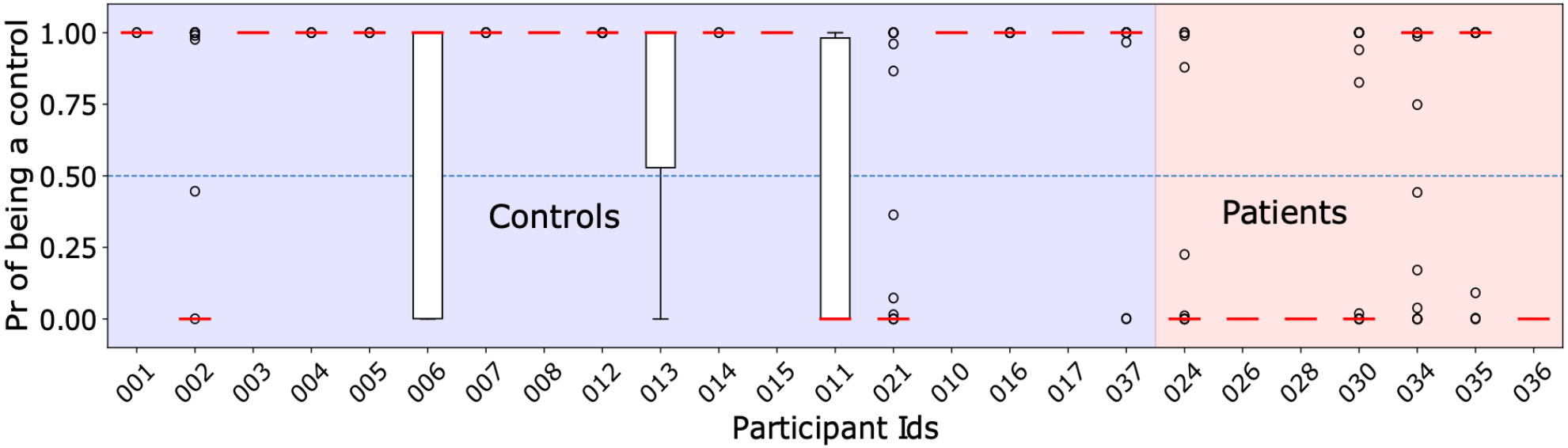
Classification of Alzheimer's Patients



(a)

(1) - Segment similarity, (2) - Duration of stops, (3) - entropy

Classification of Alzheimer's Patients



UNIVERSITY OF CAMBRIDGE

Interesting questions

- How many samples we need?
- What are the discriminating features?
- At what stage of AD can we discriminate?
- Do we need to send the data elsewhere for analysis? Or just features?
- Can the model run on device?

COVID-19 Mobility, Inequality and Reopening

- Anonymized location data from mobile apps
  - Census Block Groups (CBGs) contain from 300 to 6000 people
  - Points of interests (POIs): restaurants, gyms,…
- March 2020 to May 2020

Metro area	CBGs	POIs	Hourly edges	Total modeled pop	Total visits
Atlanta	3,130	39,411	540,166,727	7,455,619	27,669,692
Chicago	6,812	62,420	540,112,026	10,169,539	33,785,702
Dallas	4,877	52,999	752,998,455	9,353,561	37,298,053
Houston	3,345	49,622	609,766,288	7,621,541	32,943,613
Los Angeles	8,904	83,954	643,758,979	16,101,274	38,101,674
Miami	3,555	40,964	487,544,190	6,833,129	26,347,947
New York City	14,763	122,428	1,057,789,207	20,729,481	66,581,080
Philadelphia	4,565	37,951	304,697,220	6,759,058	19,551,138
San Francisco	2,943	28,713	161,575,167	5,137,800	10,728,090
Washington DC	4,051	34,296	312,620,619	7,740,276	17,898,324
All metro areas combined	**56,945**	**552,758**	**5,411,028,878**	**97,901,278**	**310,905,313**

Article

Mobility network models of COVID-19 explain inequities and inform reopening

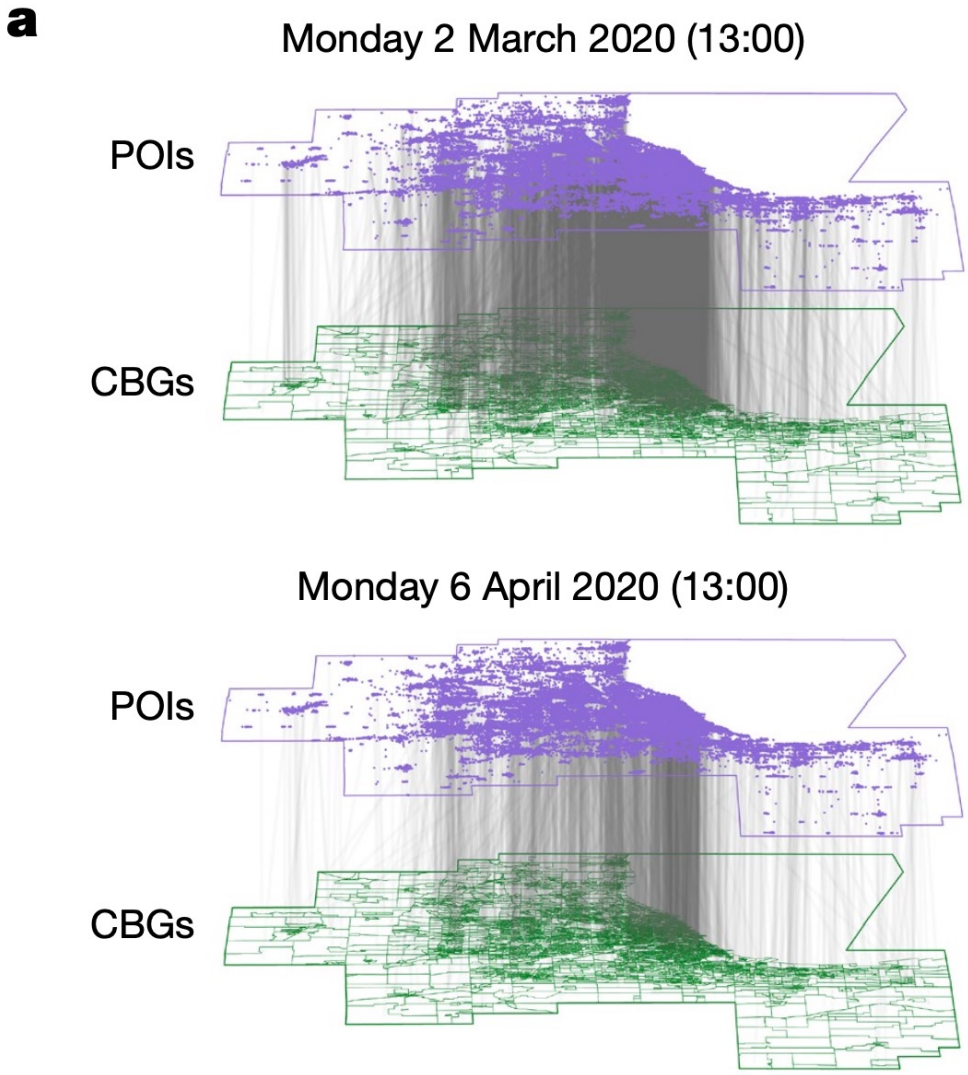
https://doi.org/10.1038/s41586-020-2923-3
Received: 15 June 2020
Accepted: 21 October 2020
Published online: 10 November 2020
Check for updates

Serina Chang[1,9], Emma Pierson[1,2,9], Pang Wei Koh[1,9], Jaline Gerardin[3], Beth Redbird[4,5], David Grusky[6,7] & Jure Leskovec[1,8]✉

The coronavirus disease 2019 (COVID-19) pandemic markedly changed human mobility patterns, necessitating epidemiological models that can capture the effects of these changes in mobility on the spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)[1]. Here we introduce a metapopulation susceptible–exposed–infectious–removed (SEIR) model that integrates fine-grained, dynamic

Mobile Network

- CBGs (info on population, race and age demographics, income…)
- POI (label, eg restaurant, size and dwell time averages)
- Bipartite graph: G (V,E)
  - V= CBGs U POIs
  - E= weights (number of people moving from CBG to POI per time interval)



UNIVERSITY OF CAMBRIDGE

Epidemics Model

- SEIR Model
  - Susceptible
  - Exposed: infected but not infectious
  - Infectious
  - Recovered

- The model has only three free parameters: (1) transmission rates at POIs, (2) transmission rates at CBGs and (3) the initial proportion of exposed individuals.

The SEIR Model applied…

$$N^{(t)}_{S_{c_i}\to E_{c_i}} \sim \text{Pois}\left(\frac{S^{(t)}_{c_i}}{N_{c_i}} \sum_{j=1}^{n} \lambda^{(t)}_{p_j} w^{(t)}_{ij}\right) + \text{Binom}\left(S^{(t)}_{c_i}, \lambda^{(t)}_{c_i}\right)$$

$$\lambda^{(t)}_{p_j} = \beta^{(t)}_{p_j}(I^{(t)}_{p_j}/V^{(t)}_{p_j}),$$

$$N^{(t)}_{E_{c_i}\to I_{c_i}} \sim \text{Binom}\left(E^{(t)}_{c_i}, 1/\delta_E\right)$$

$$N^{(t)}_{I_{c_i}\to R_{c_i}} \sim \text{Binom}\left(I^{(t)}_{c_i}, 1/\delta_I\right),$$

$$\beta^{(t)}_{p_j} = \psi d^2_{p_j} \frac{V^{(t)}_{p_j}}{a_{p_j}},$$

The SEIR Model applied…

$$N^{(t)}_{S_{c_i}\to E_{c_i}} \sim \text{Pois}\left(\frac{S^{(t)}_{c_i}}{N_{c_i}} \sum_{j=1}^{n} \lambda^{(t)}_{p_j} w^{(t)}_{ij}\right) + \text{Binom}\left(S^{(t)}_{c_i}, \lambda^{(t)}_{c_i}\right)$$

$$N^{(t)}_{E_{c_i}\to I_{c_i}} \sim \text{Binom}\left(E^{(t)}_{c_i}, 1/\delta_E\right)$$

$$N^{(t)}_{I_{c_i}\to R_{c_i}} \sim \text{Binom}\left(I^{(t)}_{c_i}, 1/\delta_I\right),$$

$$\lambda^{(t)}_{p_j} = \beta^{(t)}_{p_j}(I^{(t)}_{p_j}/V^{(t)}_{p_j})$$

$$\beta^{(t)}_{p_j} = \psi d^2_{p_j} \frac{V^{(t)}_{p_j}}{a_{p_j}},$$

The SEIR Model applied…

$$N^{(t)}_{S_{c_i}\to E_{c_i}} \sim \text{Pois}\left(\frac{S^{(t)}_{c_i}}{N_{c_i}} \sum_{j=1}^{n} \lambda^{(t)}_{p_j} w^{(t)}_{ij}\right) + \text{Binom}\left(S^{(t)}_{c_i}, \lambda^{(t)}_{c_i}\right)$$

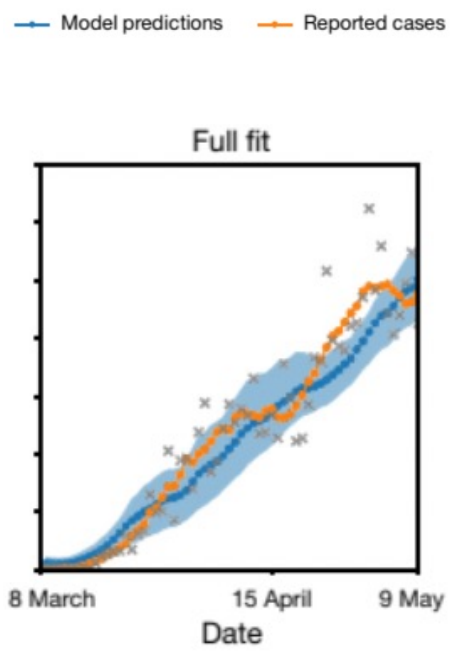
$$N^{(t)}_{E_{c_i}\to I_{c_i}} \sim \text{Binom}\left(E^{(t)}_{c_i}, 1/\delta_E\right)$$

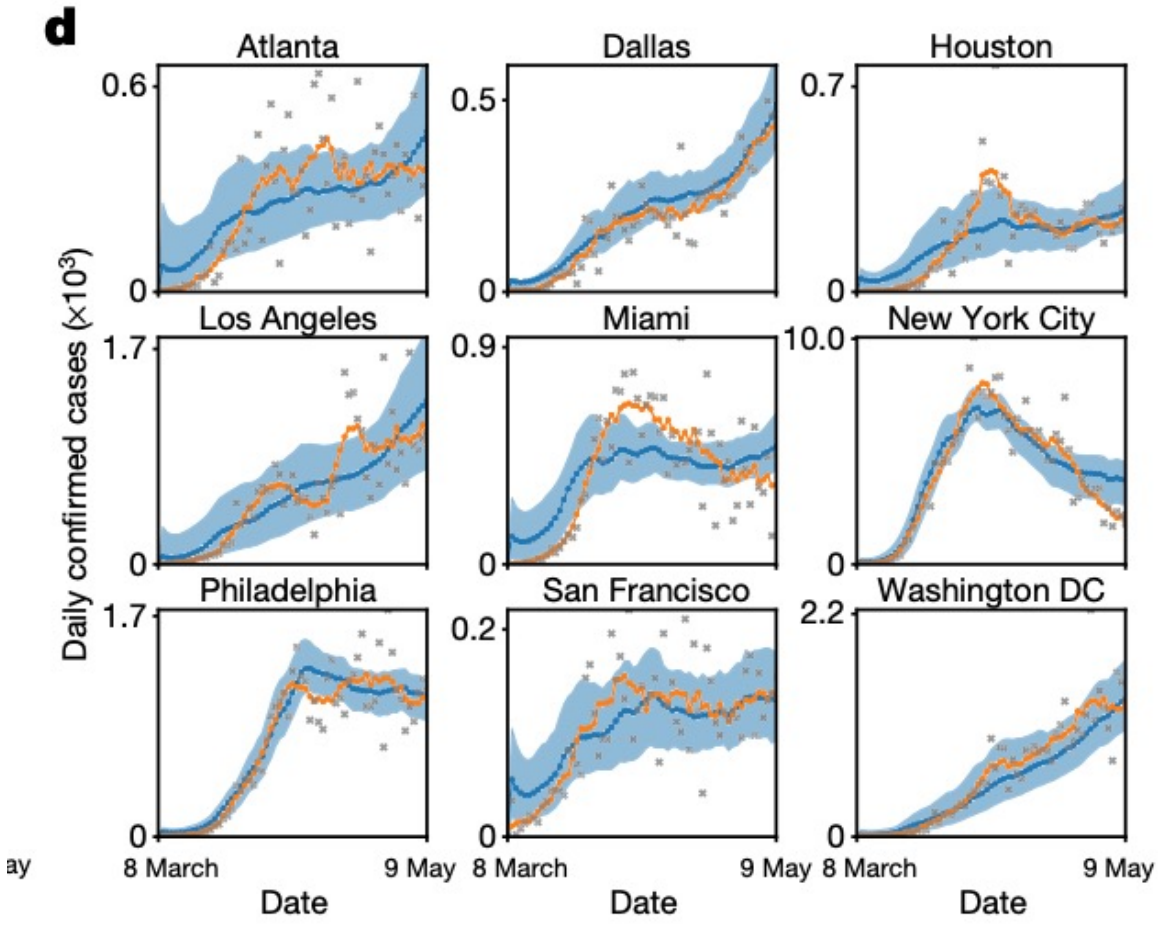
$$N^{(t)}_{I_{c_i}\to R_{c_i}} \sim \text{Binom}\left(I^{(t)}_{c_i}, 1/\delta_I\right),$$

$$\lambda^{(t)}_{p_j} = \beta^{(t)}_{p_j}\left(I^{(t)}_{p_j}/V^{(t)}_{p_j}\right),$$

$$\beta^{(t)}_{p_j} = \psi d^2_{p_j} \frac{V^{(t)}_{p_j}}{a_{p_j}},$$

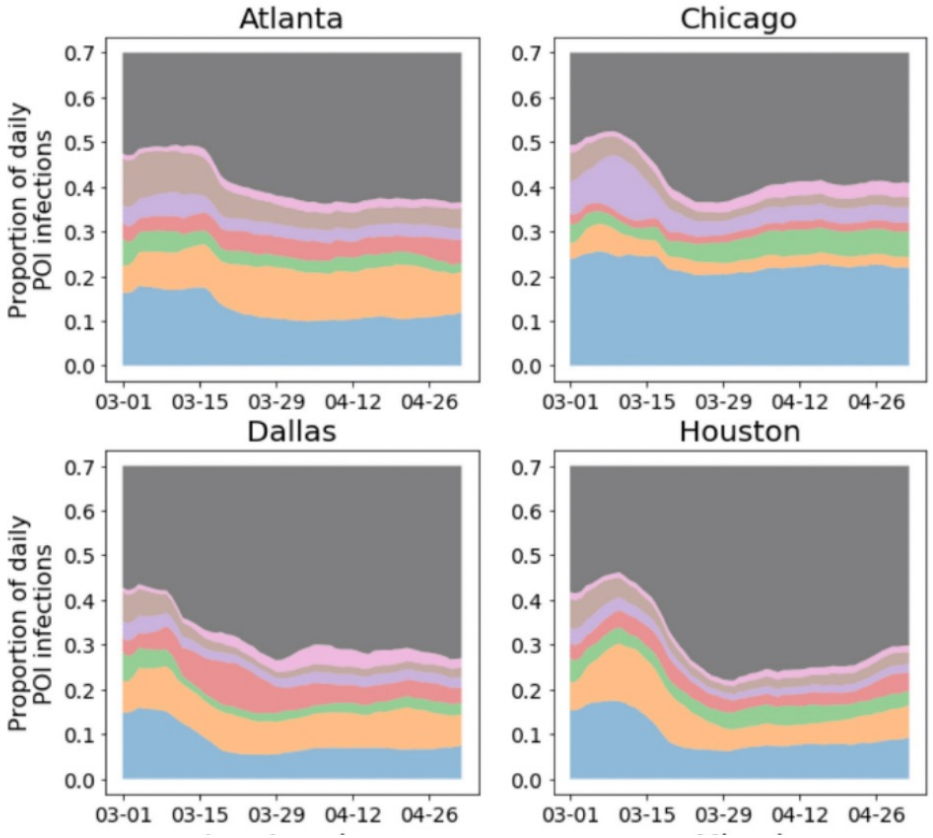
Model Fitting

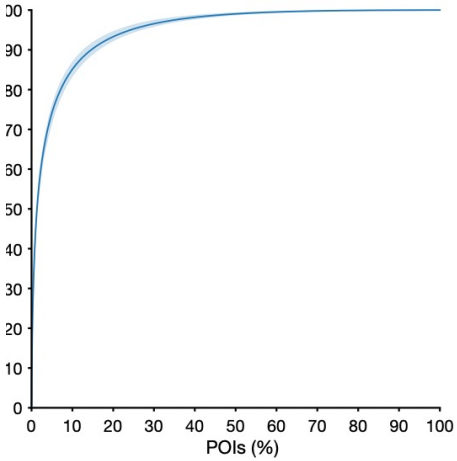




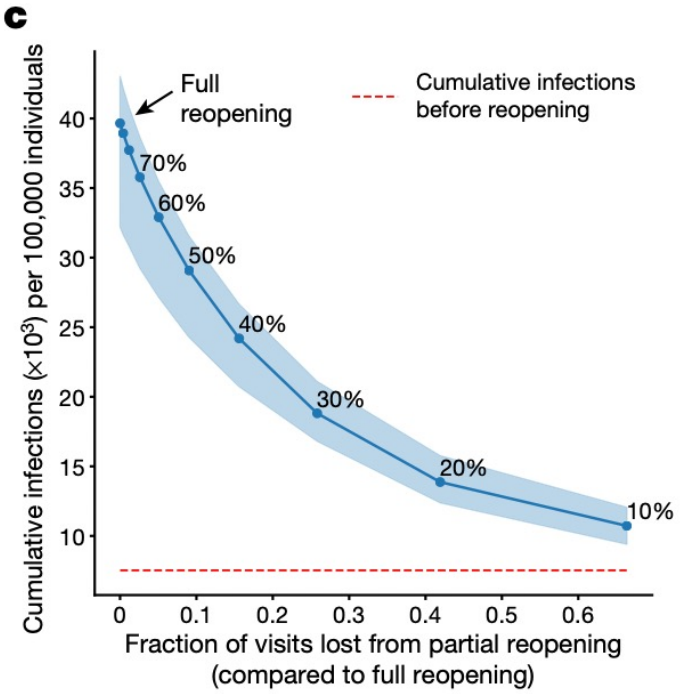
Infections per POIs

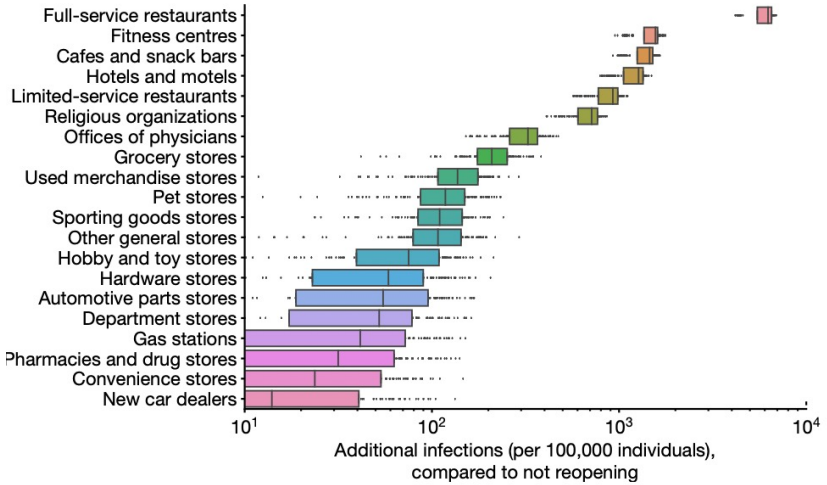
- in the Chicago metro area, 10% of POIs accounted for 85% of infections.





Reopening…

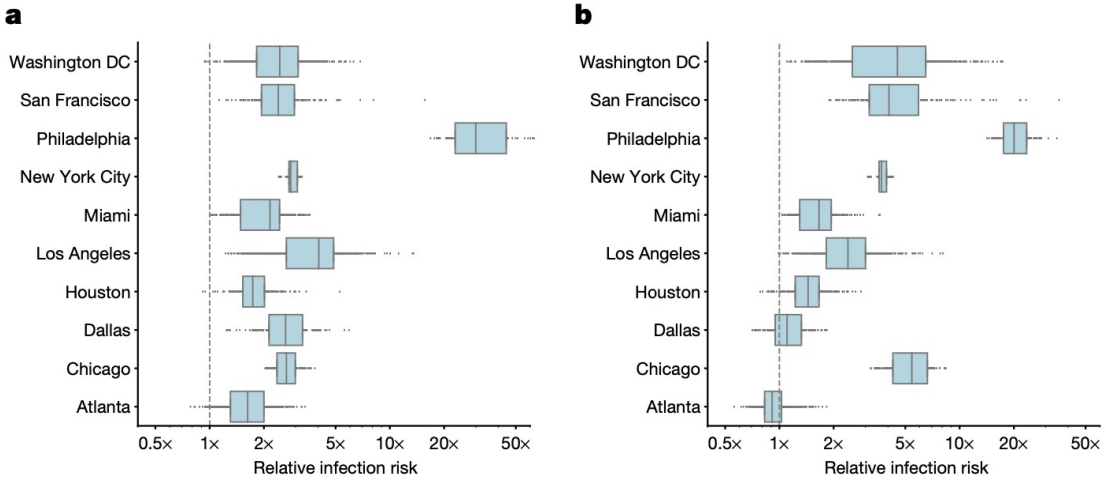


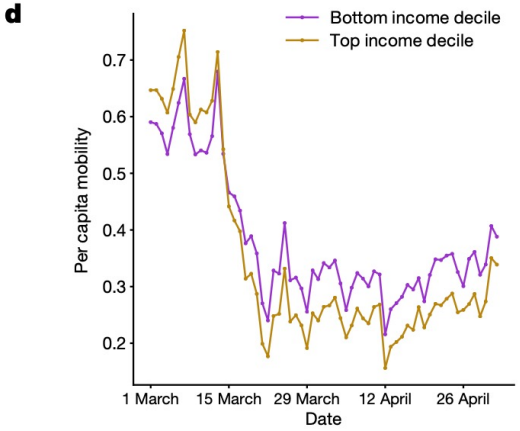


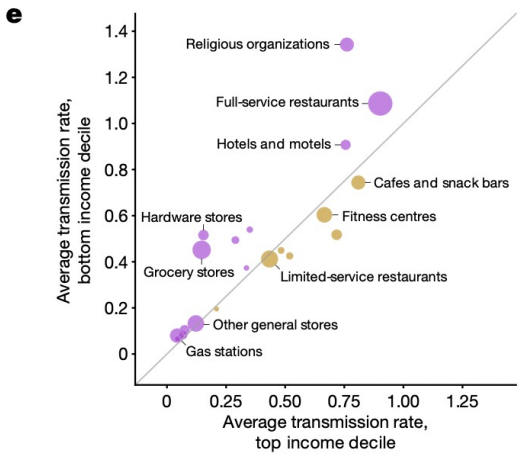
Infections predicted after 1 month of reopening over fractions of visits lots due to partial reopening

UNIVERSITY OF CAMBRIDGE

Race, Income effects







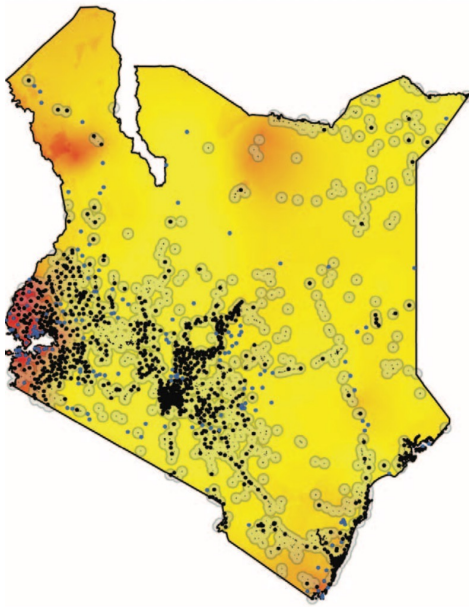
Malaria

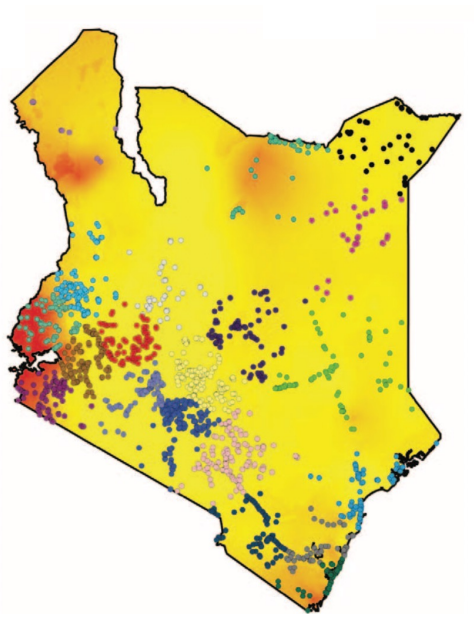
- The malaria parasite is carried by an infected person and spreads through mosquito bites.
- Human movements increases malaria spreading

Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, Snow RW, Buckee CO. Quantifying the impact of human mobility on malaria. Science. 2012 Oct 12;338(6104):267-70.

Cellular Data vs Malaria Areas

- Cellular base station location
- Cluster of similar malaria risks and geography



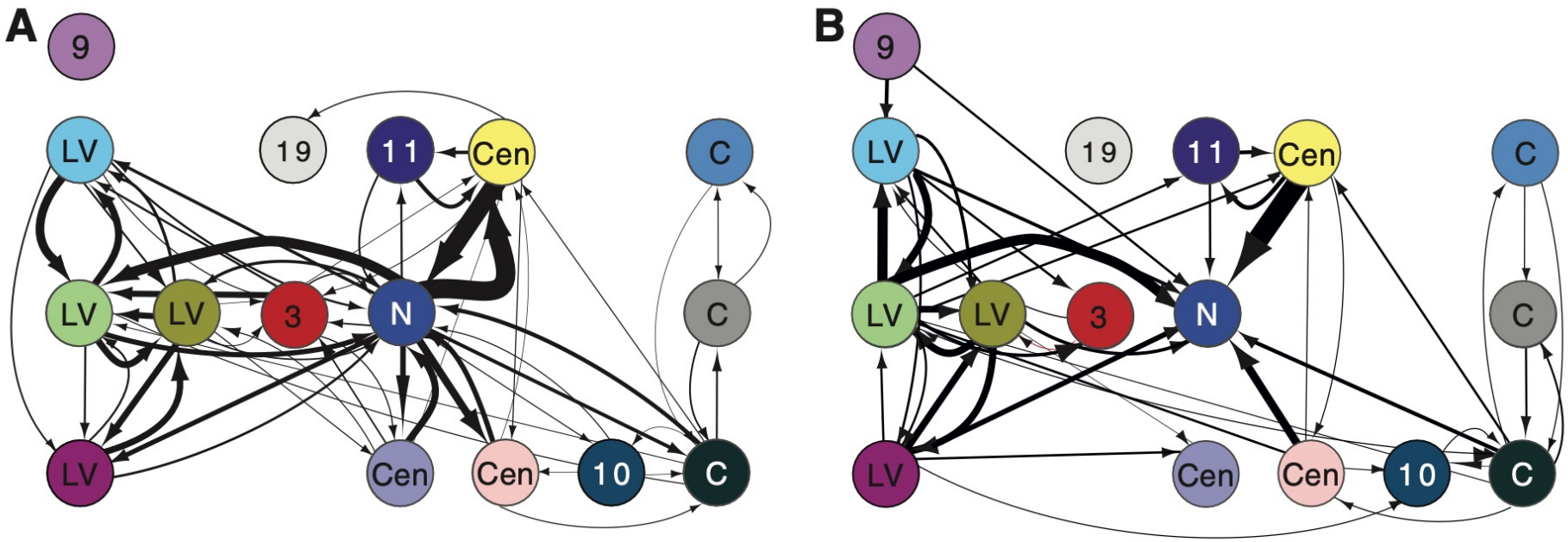


Black dots: base stations
Gray dots: settlements
Red: high malaria prevalence

UNIVERSITY OF CAMBRIDGE

Quantifying Human Movement between Risk Regions

- A: movement
- B: returning residents (importing parasites)



UNIVERSITY OF CAMBRIDGE

Ethics and Privacy!

- Location data is sensitive
- Anonymized aggregations offer some protection
- One paper says “exclude census block groups if there are only less than five devices recorded from that CBG to a POI”

Location Privacy Breaches

- If the approximate locations of an anonymous user's home and workplace can be deduced from a location trace, then the median size of the individual's anonymity set in the U.S. working population is 1, 21 and 34,980, for locations known at the granularity of a census block, census tract and county respectively.

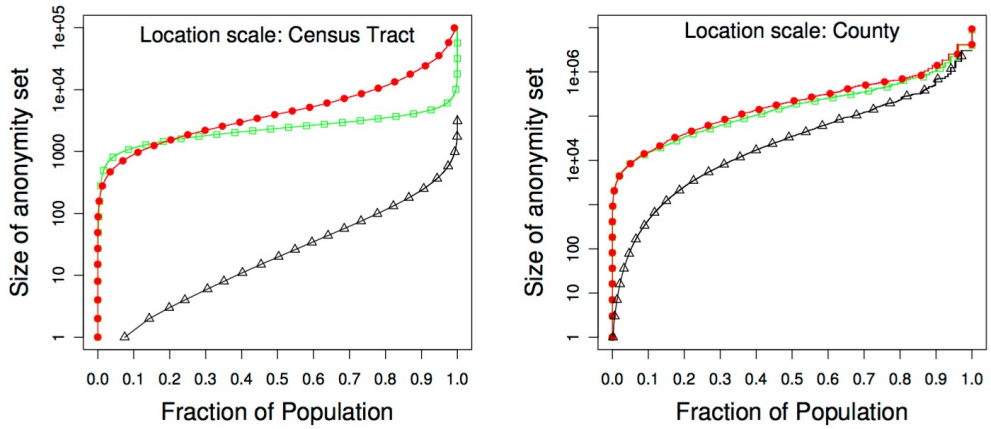


Fig. 1. Size of anonymity set under disclosure of work location (red circles), home location (green squares) or both (black triangles). Location granularity is either census tract (left graph) or county (right graph). Note the different scales on the Y-axes.

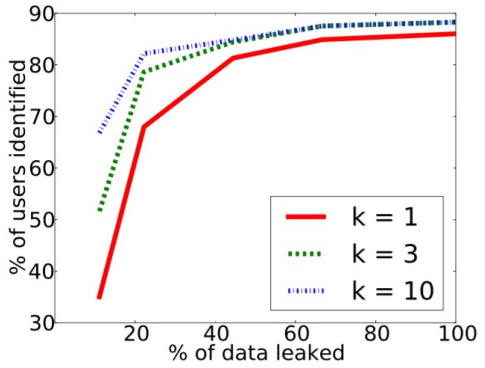
P. Golle, K. Partridge. 2009. On the Anonymity of Home/Work Location Pairs. In Procs of the 7th International Conference on Pervasive Computing (Pervasive '09). Springer-Verlag, 390–397.

De-anonimyzation

- Netflix type datasets contains narrow range of behaviour (preferences of movies)
- Sensing data contain a wider range of behaviours
- Correlation of activities can be a “key”
  - Eg you go to the gym at a certain time after a train ride
  - This increases data sparsity
  - Broad range of auxiliary information which can be used by an adversary

Example

- Auxiliary information of the adversary is a collection of activities by a user.
- Aim: Identify which of the anonymized activity streams belongs to the target user.
- Auxiliary information may be collected by observing the user or from available public sources.



N. Lane, J. Xie, T. Moscibroda, and F. Zhao, “On the feasibility of user de-anonymization from shared mobile sensor data,” in Proceedings of the Third International Workshop on Sensing Applications on Mobile Phones. ACM, 2012.

Other Location Data Public Health Applications

- Gordon-Larsen, P., Nelson, M.C., Page, P. and Popkin, B.M., 2006. Inequality in the built environment underlies key health disparities in physical activity and obesity. *Pediatrics*, *117*(2), pp.417-424.
- Zenk, S.N., Schulz, A.J., Matthews, S.A., Odoms-Young, A., Wilbur, J., Wegrzyn, L., Gibbs, K., Braunschweig, C. and Stokes, C., 2011. Activity space environment and dietary and physical activity behaviors: a pilot study. *Health & place*, *17*(5), pp.1150-1161.
- Wesolowski, A., O'Meara, W.P., Tatem, A.J., Ndege, S., Eagle, N. and Buckee, C.O., 2015. Quantifying the impact of accessibility on preventive healthcare in sub-Saharan Africa using mobile phone data. *Epidemiology (Cambridge, Mass.)*, *26*(2), p.223.
- Peak, C.M., Wesolowski, A., zu Erbach-Schoenberg, E., Tatem, A.J., Wetter, E., Lu, X., Power, D., Weidman-Grunewald, E., Ramos, S., Moritz, S. and Buckee, C.O., 2018. Population mobility reductions associated with travel restrictions during the Ebola epidemic in Sierra Leone: use of mobile phone data. *International journal of epidemiology*, *47*(5), pp.1562-1570.
- Bajardi, P., Poletto, C., Ramasco, J.J., Tizzoni, M., Colizza, V. and Vespignani, A., 2011. Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic. *PloS one*, *6*(1), p.e16591.
- Gao, S., Rao, J., Kang, Y., Liang, Y., Kruse, J., Dopfer, D., Sethi, A.K., Reyes, J.F.M., Yandell, B.S. and Patz, J.A., 2020. Association of mobile phone location data indications of travel and stay-at-home mandates with COVID-19 infection rates in the US. *JAMA network open*, *3*(9), pp.e2020485-e2020485.
- Kraemer MUG, Hill V, Ruis C, Dellicour S, Bajaj S, McCrone JT, Baele G, Parag KV, Battle AL, Gutierrez B, Jackson B, Colquhoun R, O'Toole Á, Klein B, Vespignani A; COVID-19 Genomics UK (COG-UK) Consortium, Volz E, Faria NR, Aanensen DM, Loman NJ, du Plessis L, Cauchemez S, Rambaut A, Scarpino SV, Pybus OG. Spatiotemporal invasion dynamics of SARS-CoV-2 lineage B.1.1.7 emergence. Science. 2021 Aug 20;373(6557):889-895. doi: 10.1126/science.abj0113. Epub 2021 Jul 22. PMID: 34301854.

Questions