SpatialEpi Workshop

30th ACM SIGSPATIAL, November 1 - 4, 2022, Seattle, Washington, USA

Using mobile network data to color epidemic risk maps

Elisa Cabana
elisa.cabana@imdea.org
IMDEA Networks Institute
Madrid, Spain

Andra Lutu
andra.lutu@telefonica.com
Telefonica Research
Barcelona, Spain

Enrique Frias-Martinez
enrique.frias@ucjc.edu
Universidad Camilo Jose Cela
Madrid, Spain

Nikolaos Laoutaris
nikolaos.laoutaris@imdea.org
IMDEA Networks Institute
Madrid, Spain

Acknowledgments: REACT-CONTACT-CM-23479 project (RGM and ERDF)

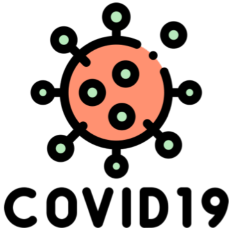
Motivation

Population structure
+
Patterns of interaction
+
Mobility

→ COVID-19

→ Spatio-temporal information from Mobile Network Data

→ Tools & Solutions for:
-Individuals
-First responders
-Decision makers

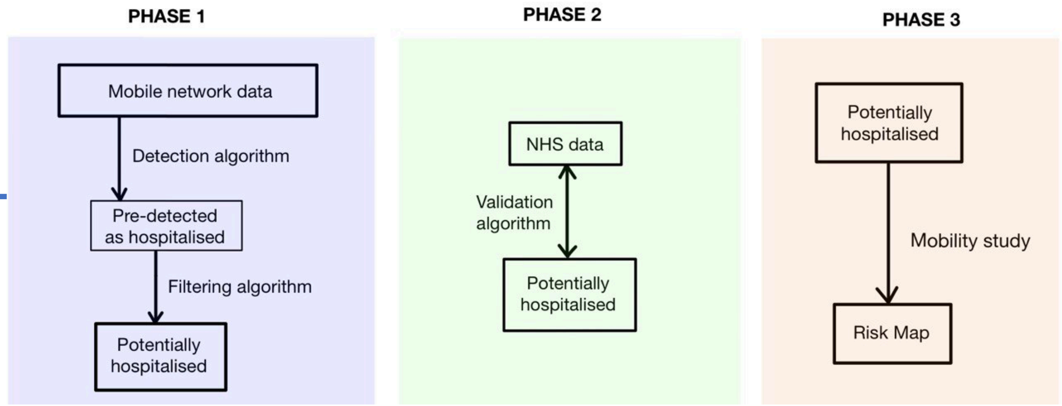


Data

D_S **Mobile network (signaling) data:** additional data generated in communication networks to support the successful transfer of data

- GDPR-compliant and anonymized.
- More than 2 million users.
- March and April 2020.
- Region of Greater London, United Kingdom.
- Spatio-temporal information:
  - h_i: home locations.
  - l_i: top location at night (00:00h – 08:00h).
  - t_i: time spent in that location.

Methodology



Objective: detect potential COVID-19 hospitalizations by looking for individuals whose phone started appearing during the night in the same area or very close to a hospital that received COVID-19 patients.

Filters:

- **Home filter:** Individuals living near the Hospital locations.
- **Work filter:** Individuals working at the Hospital (e.g., nurses, doctors, security guards, etc.)

PHASE 1

Input: D_S, φ, r, η
foreach $s_i \in S$ **do**
  if $h_i^{(\varphi)} \notin L^{(\varphi,r)}$ & $l_i^{(\varphi)} \in L^{(\varphi,r)}$, *for* η *nights s.t.* $\eta_0 + \eta = \eta_{last}$ **then**
    $s_i \in \Phi$
  else
    $s_i \notin \Phi$
Output: Φ

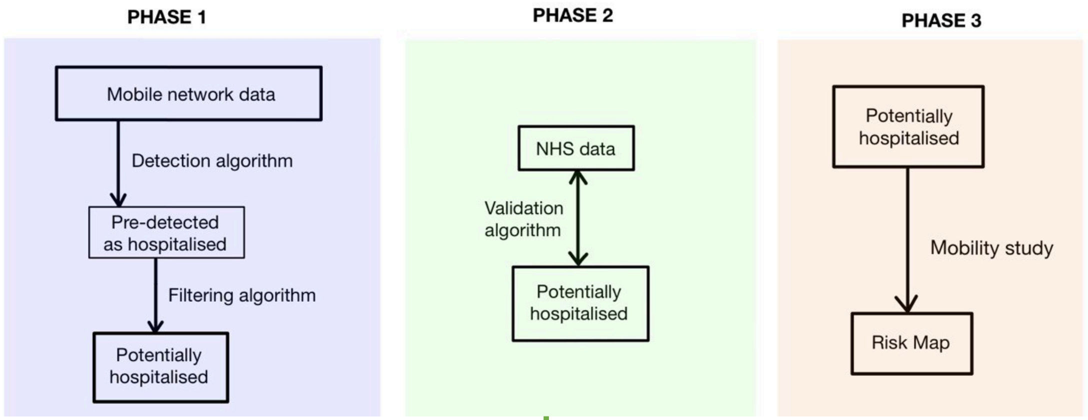
Algorithm 1: Detection and Filtering algorithm.

Parameters

Concept	Notation	Values	Meaning
Location granularity	φ	P T	Postcode Cell-tower
Cell-tower ratio	r	0.5 0.75 1.0	500 m 750 m 1000 m
Time granularity	η	≥ 1	# of consecutive nights

Result: a set of “potentially hospitalized” individuals.

Methodology



Parameters

Concept	Notation	Values	Meaning
Location granularity	φ	P T	Postcode Cell-tower
Cell-tower ratio	r	0.5 0.75 1.0	500 m 750 m 1000 m
Time granularity	η	≥ 1	# of consecutive nights

PHASE 2

Validation results:

$$\varphi = T$$

$$r = 0.75$$

$$\eta = 4$$

Validation and fine-tuning: find the parameter configuration that consistently matches the validation data (from NHS) across different settings, while maintaining a high value of correlation.

Risk Maps

Objective: creating risk maps based on the mobility of the final set of individuals detected as hospitalized, during the period of 2 weeks prior to their first day of hospitalization.

PHASE 3

Input: Φ, λ, D_λ
foreach $t = 1, \ldots, T$ **do**
  foreach $k = 1, \ldots, A$ **do**
    foreach $i = 1, \ldots, |\Phi|$ **do**
      if *s_i is in area l_k at time t* **then**
        $a_i^{(l_k,t)} = 1$
      else
        $a_i^{(l_k,t)} = 0$
    $p^{(l_k,t)} = \sum_{s_i \in \Phi} a_i^{(l_k,t)}$
  Risk Map (λ, t)
Output: Risk Map Movie (λ)

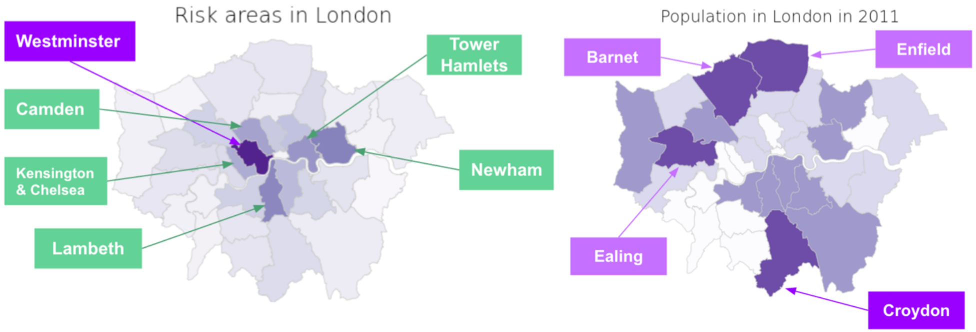
Algorithm 2: Algorithm for obtaining the risk maps.

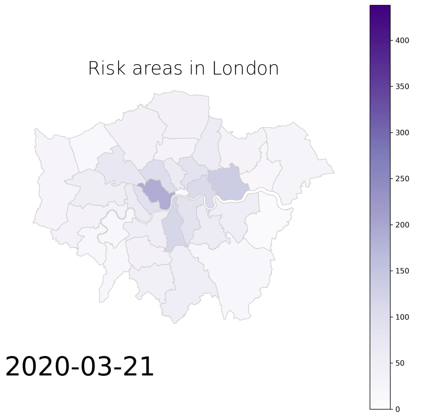
The parameters involved in Algorithm 2 are the following:

- T: the total number of days.
- $t = 1, \ldots, T$: the daily time steps.
- λ: the geography to plot the risk map.
- D_λ: the polygon geometries depending on the selected λ.
- A: the total number of areas in the map.
- $l_k \in D_\lambda$ $(k = 1, \ldots, A)$: the location areas in the map.
- $a_i^{(l_k,t)}$, for $i = 1, \ldots, |\Phi|$: counts the number of users (detected as hospitalized) located in area l_k at time t.
- $p^{(l_k,t)}$: the measure of risk of area l_k at time t.

A risk map provides a visual representation of risks that vary through time and space at different granularities

Results

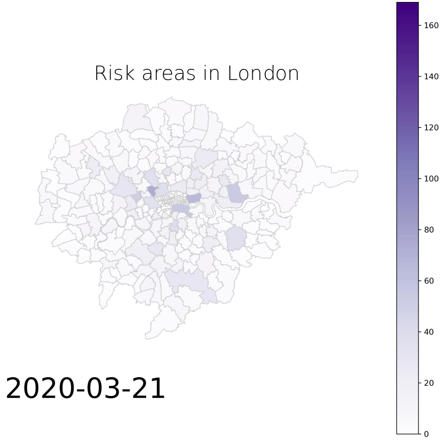




1. The areas of highest risk are not necessarily the most densely populated ones.
2. Time-related variability within the group of individual risk maps.
3. Hospitalized individuals tended to have a higher average mobility than non-hospitalized ones.

The proposed method yields a 98.6% agreement with released public records of patients admitted to hospitals in London within the same time frame.

The multidimensional characteristic of the risk of an area is better reflected when taking spatio-temporal information of high granularity



Limitations and Advantages

Limitations:

1. Severe cases vs asymptomatic cases.
2. Sensitive data and privacy concerns.

Advantages:

1. **User involvement** is not required to collect the data.
2. No need of healthcare providers' **interventions or digital certificates**.
3. **Base-station antenna is less privacy-invasive** than GPS coordinates or near-field contacts of a few meters (Contact Tracing).
4. Generic and configurable approach that can be **adapted** to other areas, countries or diseases.
5. A **visual representation of risks** is useful for: individuals, first responders, decision makers, evaluating the stress on the healthcare system, forecasting future risks, etc.
6. Telecommunication companies could implement **fully automated services** to estimate the dangerous areas and send an **SMS to warn individuals** when they enter a high-risk area as a reminder to take extra protective measures.

Conclusions

- New approach to **detect potential COVID-19 hospitalizations** and epidemic **risk maps**, based on mobile network data containing detailed spatio-temporal information about millions of cellphones at various scales.
- **Three phases**: (i) detection and filtering, (ii) validation and fine-tuning, and (iii) mobility patterns for obtaining risk maps.
- **High agreement with public records.**
- **The risk maps provide insightful information about the spatio-temporal variability of the risk of an area.**

Extension of the work

1. Data from other areas:
   - GDPR-compliant and anonymized GPS data.
   - Several millions of users.
   - March and April 2020.
   - New York, United States of America.
   - Very granular spatio-temporal information:
     - l_i^u: location pings for user u at time i.
2. Explore additional filters, parameters and scenarios:
   - Improve the home / work filter using also data from the post-period, for false positives reduction.
   - Fine-tune the incubation period parameter for detecting home confinements.
3. Once we have the estimated hospitalizations: use time series or ML/AI models to forecast the future hospitalizations.
4. Same for forecasting risks for each area.
5. Develop an automated contagion risk alert system.

Thank you!

Watch a ***risk map movie*** obtained through the time-lapse set of daily risk maps:
https://networks.imdea.org/mobile-network-data-an-efficient-method-for- assessing-the-spread-of-epidemics

elisa.cabana@imdea.org

@elisacabana

@elisacabana