Smart Mobility Services
Navigating from Tourist Flow Prediction to Song Propagation

Andrés Muñoz / Fernando Terroso-Sáenz

June 29, 2024

UCA Universidad de Cádiz

SummerSoc Service Oriented Computing

Agenda

1. Introduction.
2. Where’s my data?
3. Human mobility as a timeseries service
4. Human mobility as a graph-based service
5. Real-time? Movement Data Analysis
6. Real-time, again? Traffic alerts
7. Mobility in the digital world.
8. Conclusion

Agenda

1. ***Introduction.***
2. Where's my data?
3. Human mobility as a timeseries service
4. Human mobility as a graph-based service
5. Real-time? Movement Data Analysis
6. Real-time, again? Traffic alerts
7. Mobility in the digital world.
8. Conclusion

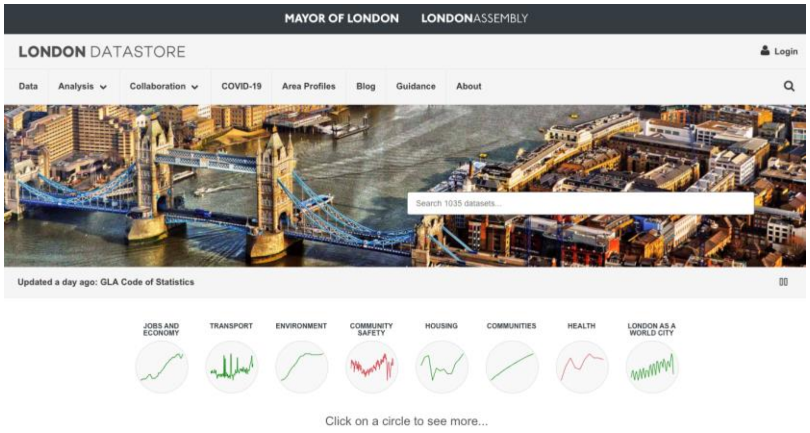
1. Introduction

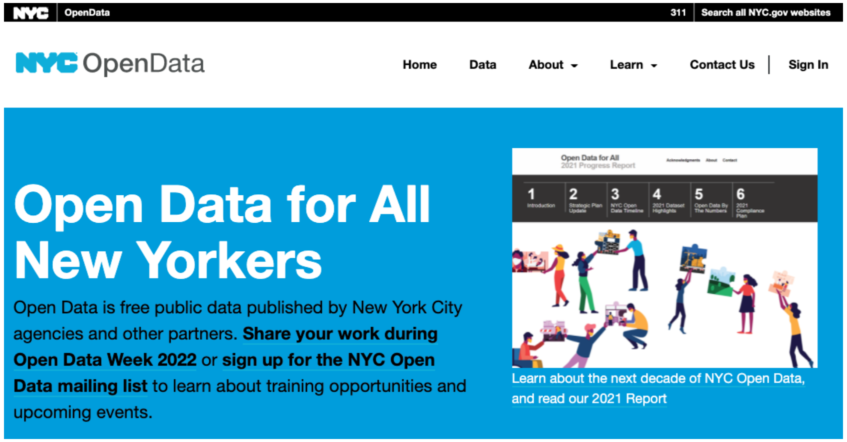
- In the last decade, a variety of location-powered technologies has emerged.
  - Bluetooth, RFDIs, WiFi, GPS.
- They allow to capture human movement at different scales and accuracies.
- **BUT** (always a “but…”) they access is rather limited.
  - A few GPS datasets available for the research community.

Microsoft Geolife GPS dataset

1. Introduction

- At the same time, there is an increasing open-data movement.
  - EU commission principle: *“As open as possible, as closed as necessary”*
- Many municipalities already have open data sites.
- **BUT** (again!) open location data comes at a cost.





1. Introduction

- **Here's an IDEA!**

  To implement Smart Mobility Services based on open data and ML/DL models

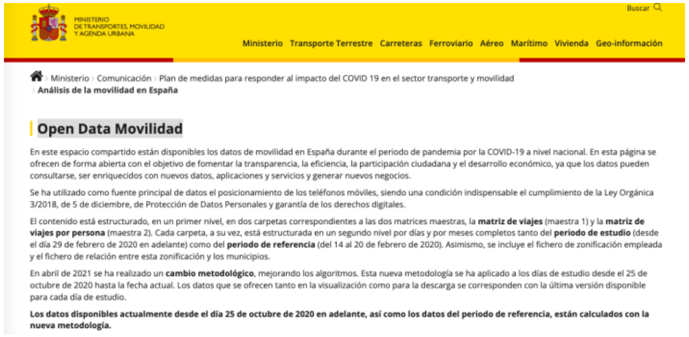
  - Predict use of the land
  - Forecast water consumption
  - Predict tourism flows
  - Monitor traffic incidents
  - Anticipate the next hit in Spotify even before that the record company???

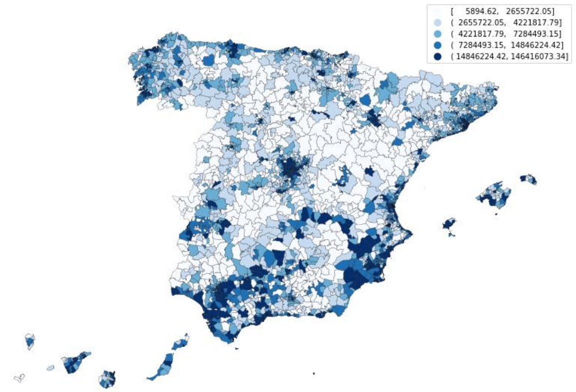
Agenda

1. Introduction.
2. ***Where's my (open) data?***
3. Human mobility as a timeseries service
4. Human mobility as a graph-based service
5. Real-time? Movement Data Analysis
6. Real-time, again? Traffic alerts
7. Mobility in the digital world.
8. Conclusion

2. Open data feeds for human mobility

- [Open Data Movilidad](https://www.mitma.es/ministerio/covid-19/evolucion-movilidad-big-data/opendata-movilidad) (https://www.mitma.es/ministerio/covid-19/evolucion-movilidad-big-data/opendata-movilidad)
  - Dataset released by Spanish Ministry of Transportation.
  - Movement of people across (2000+) districts every hour from February'20 to December'21.(local-based flows)
  - Split based on home district, age, gender, etc.
  - Flows extracted from different mobile phone carriers.

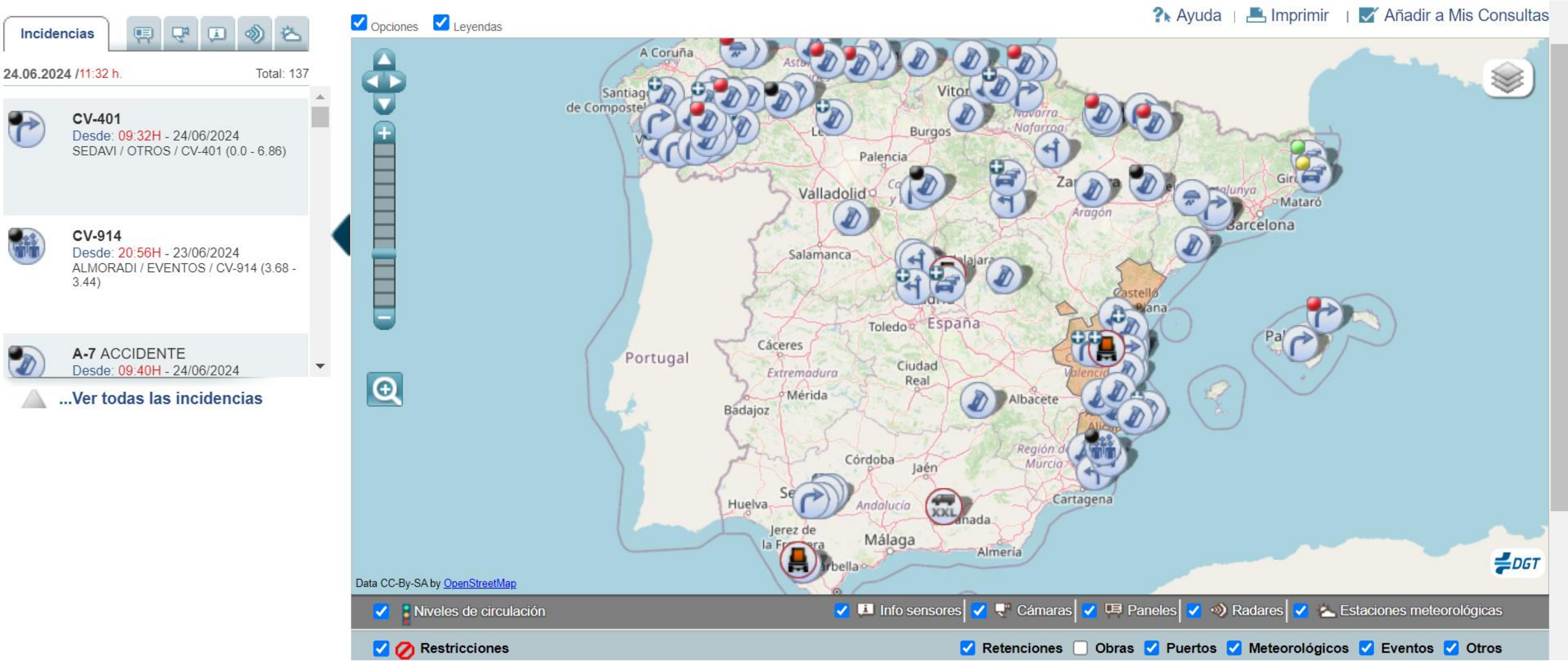




- [Spanish pilot study of nation-wide mobility](https://observatoriotransporte.mitma.gob.es/estudio-experimental) (https://observatoriotransporte.mitma.gob.es/estudio-experimental)
  - Similar to the previous one but defined at a higher administrative level (region-based flows).
  - Narrow period of time: June, July, October 2017.
  - It distinguishes among means of transport.
  - Flows extracted from Orange provider.

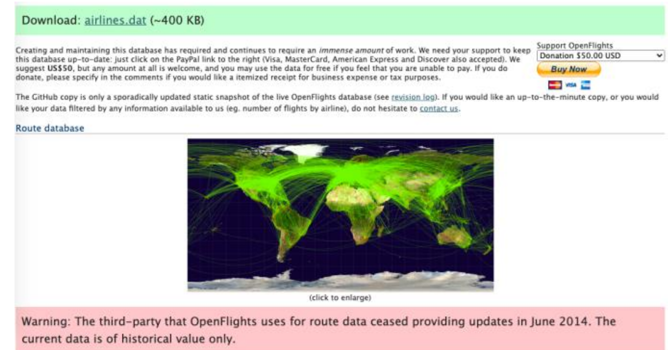
2. Open data feeds for human mobility

- Spanish Traffic Incident Data (https://infocar.dgt.es/etraffic/)

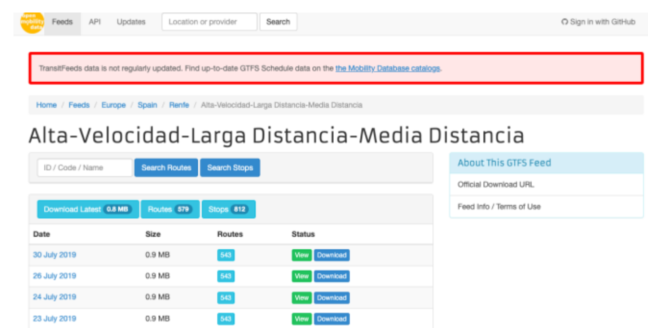


2. Open data feeds for human mobility

- Open Flights (https://openflights.org/)
  - Data with the flight connections across 10,000 worldwide airports.
    - The official dataset only converts to 2014.
    - *Hint: Go to the data supplier (the Airline Route Mapper app) donwload at http://arm.64hosts.com/ and unzip it an you get acess to a dataset updated until 2022.*



- TransitFeed (Mobility dataset)
  - Dataset with open data about public transport movement in different countries.
  - For example: Train connections and lines in Spain on a daily basis (departure and arrival time to each station) (https://transitfeeds.com/p/renfe/1018?p=38)



Tip: Once you find an interesting dataset, download it ASAP!!

2. Open data feeds for human mobility

- Are you familiar with open data sources in your country (related to mobility)?

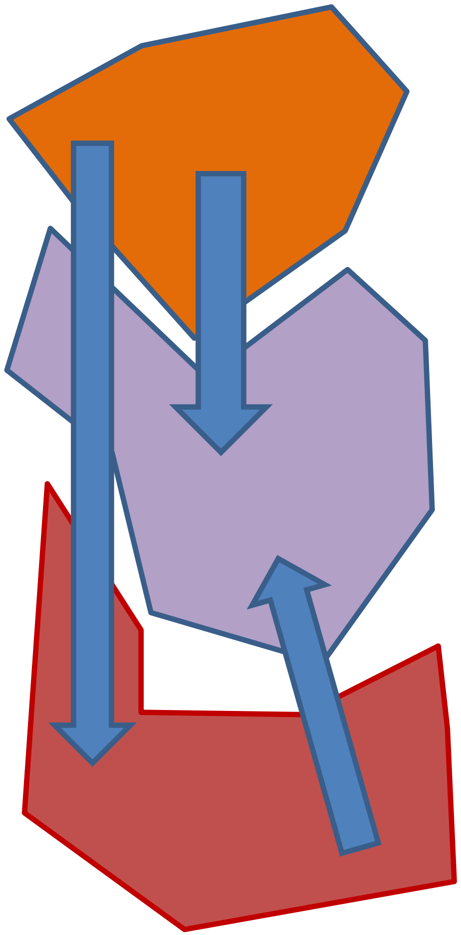
- In Greece:
  - Real-time Transit: http://gr-city.census.okfn.org/dataset/transport-realtime.html
  - Hellenic Institute of Transport: http://data.nap.gov.gr/dataset?organization=hit
    - Use of parking áreas
    - Travel times
    - Maritime transport
  - And then I found this ☹ https://lab.imedd.org/en/the-slow-death-of-open-data-in-greece/

Agenda

1. Introduction*.*
2. Where’s my (open) data?
3. ***Human mobility as a timeseries service***
4. Human mobility as a graph-based service
5. Real-time? Movement Data Analysis
6. Real-time, again? Traffic alerts
7. Mobility in the digital world.
8. Conclusion

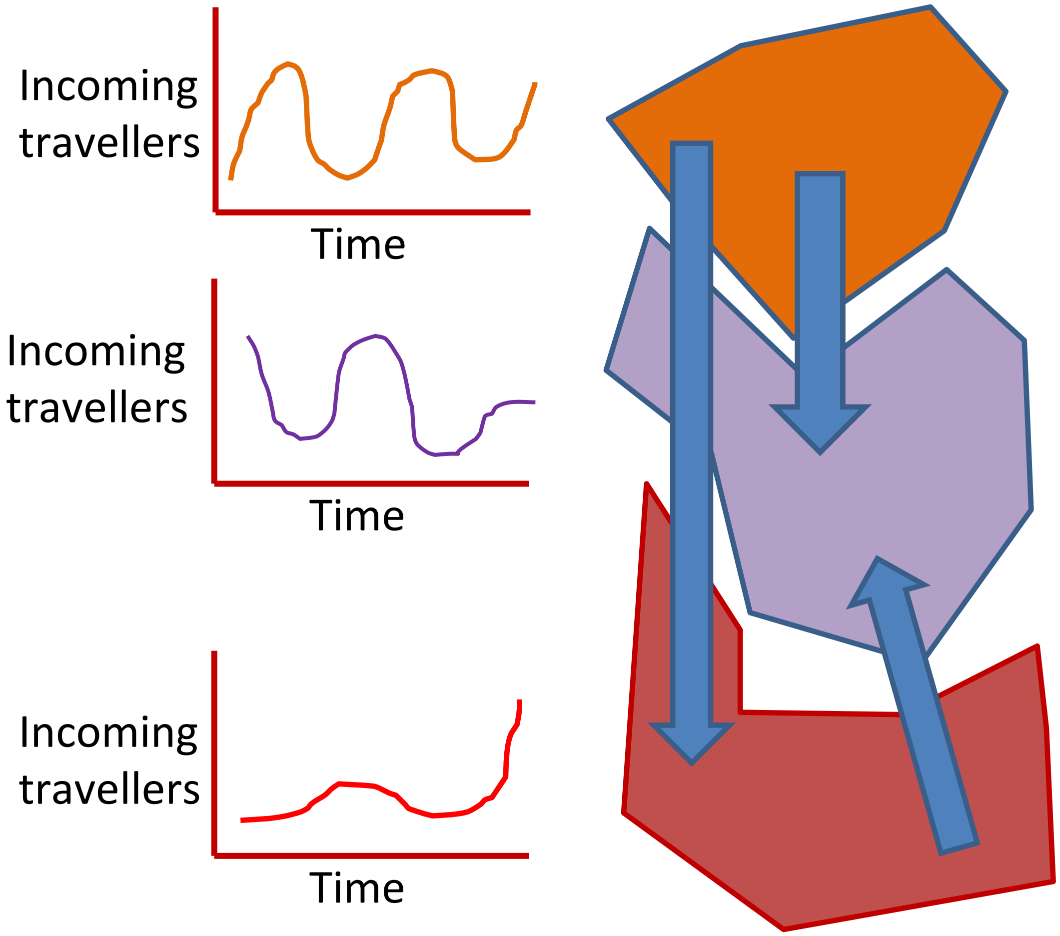
3. Human mobility as a timeseries problem.

- Using open-data mobility feeds allows us to define OD-matrices among spatial areas providing some type of aggregation.
- We can now analyze the incoming and outgoing flows of travellers for each area to other areas.



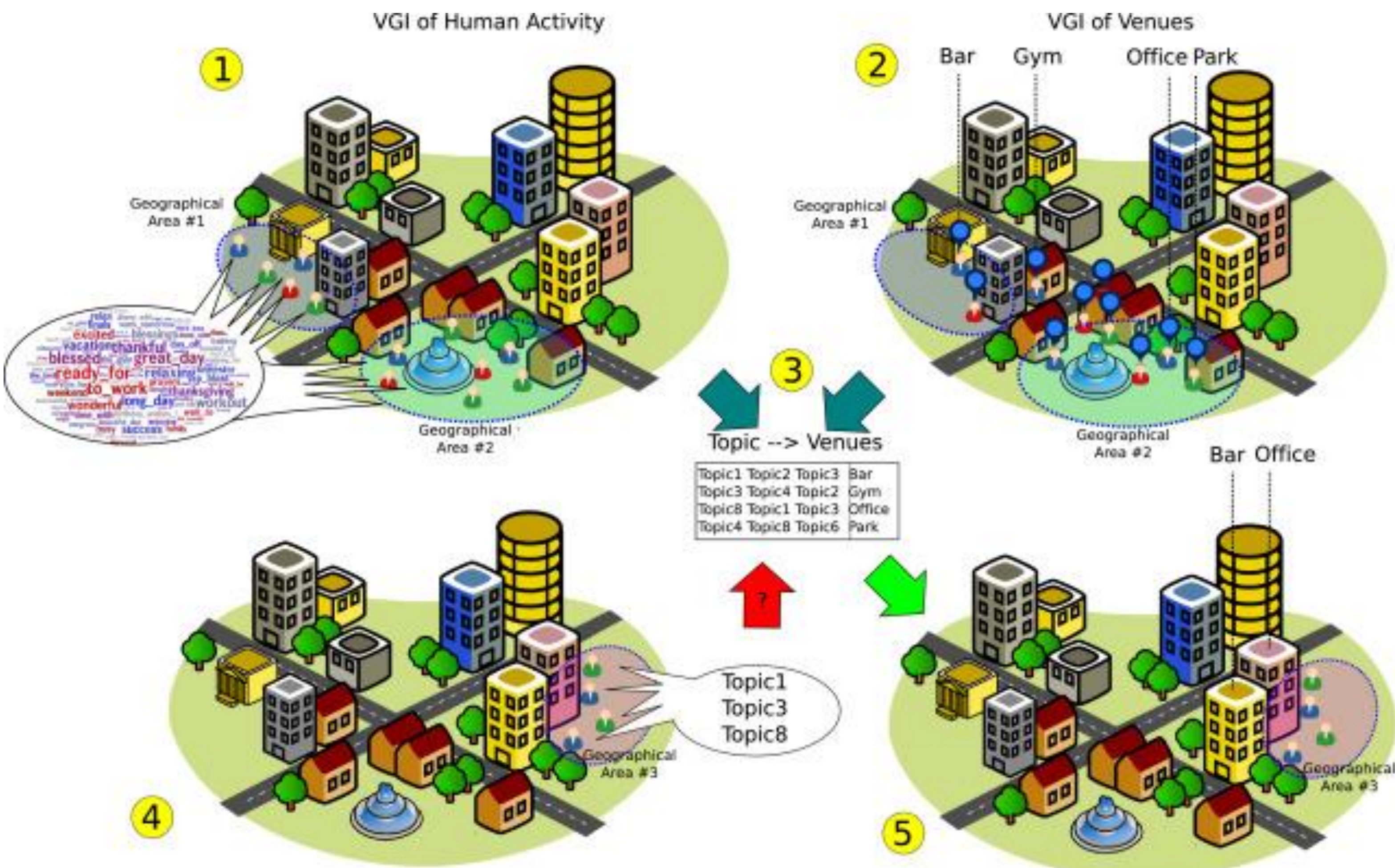
3. Human mobility as a timeseries problem.

- Such flows can be regarded as an time-evolving data → timeseries problem!!



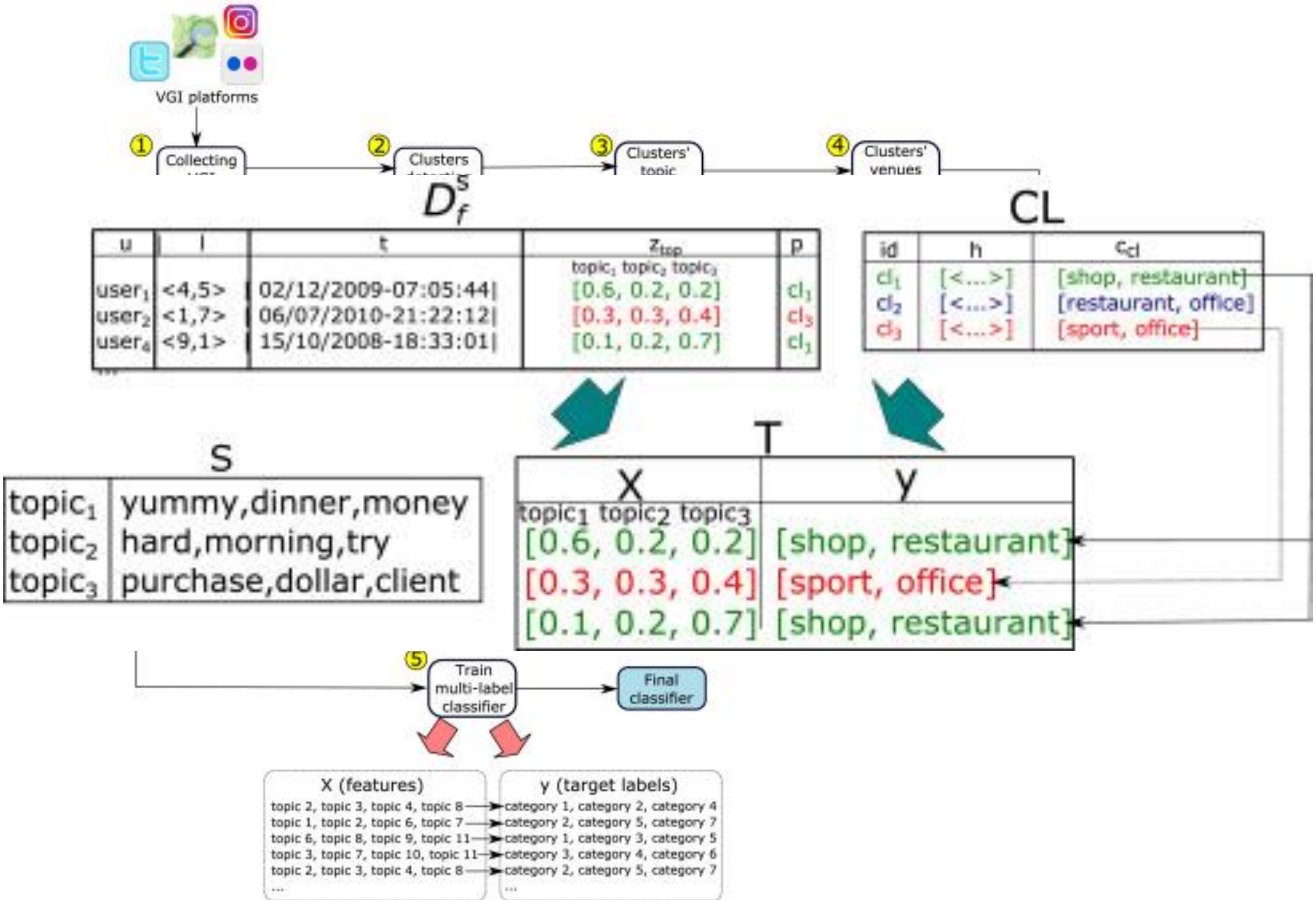
3. Human mobility as a timeseries problem.

- **<u>1st application: Land use discovery</u>**



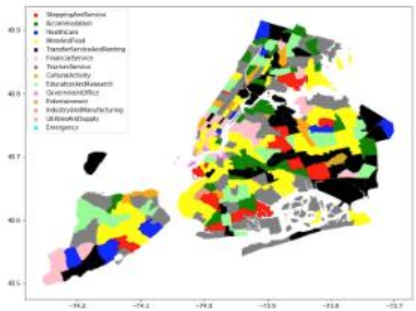
3. Human mobility as a timeseries problem.

- **<u>1st application: Land use discovery</u>**

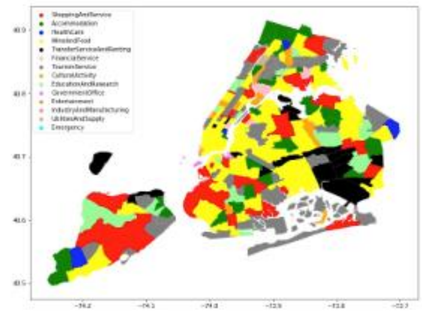


3. Human mobility as a timeseries problem.

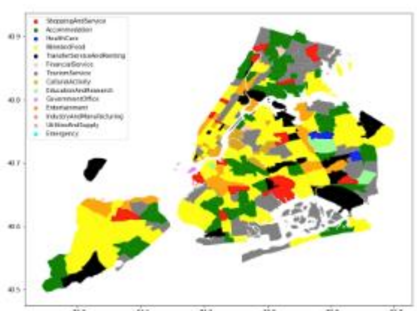
- **<u>1st application: Land use discovery</u>**



(a) Working mornings



(b) Working afternoons

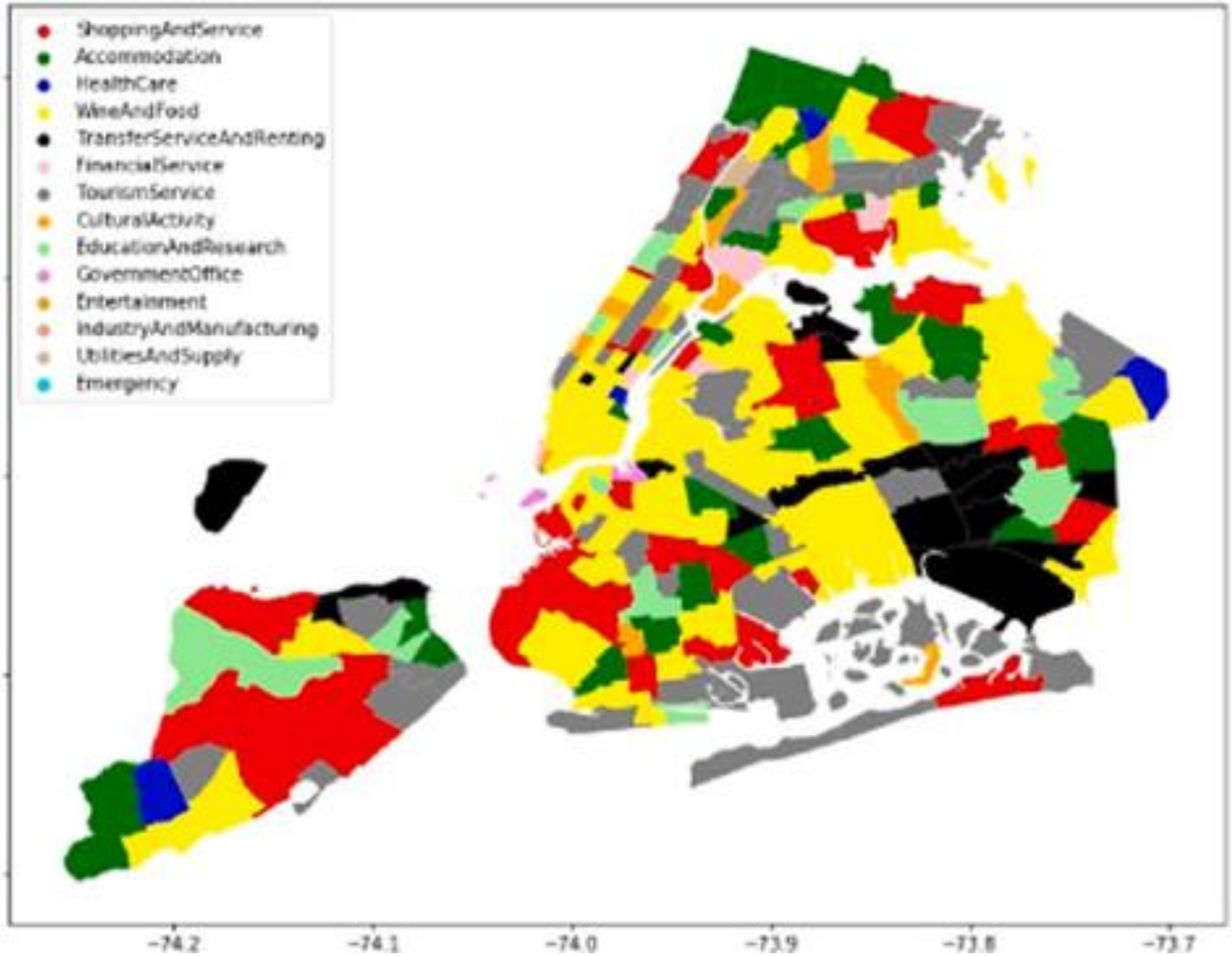


(c) Working evenings

(d) Working nights

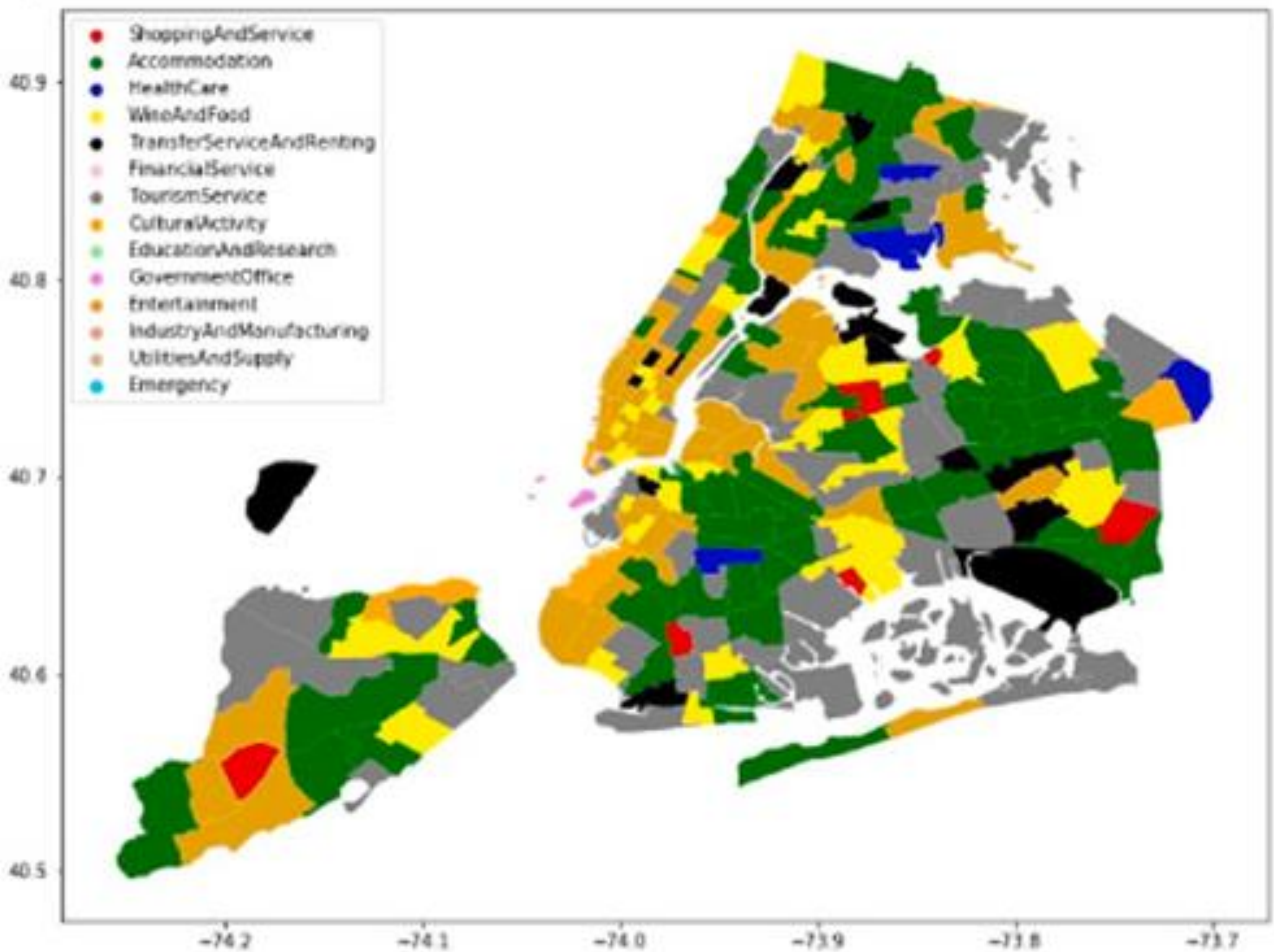
3. Human mobility as a timeseries problem.

(B)



3. Human mobility as a timeseries problem.

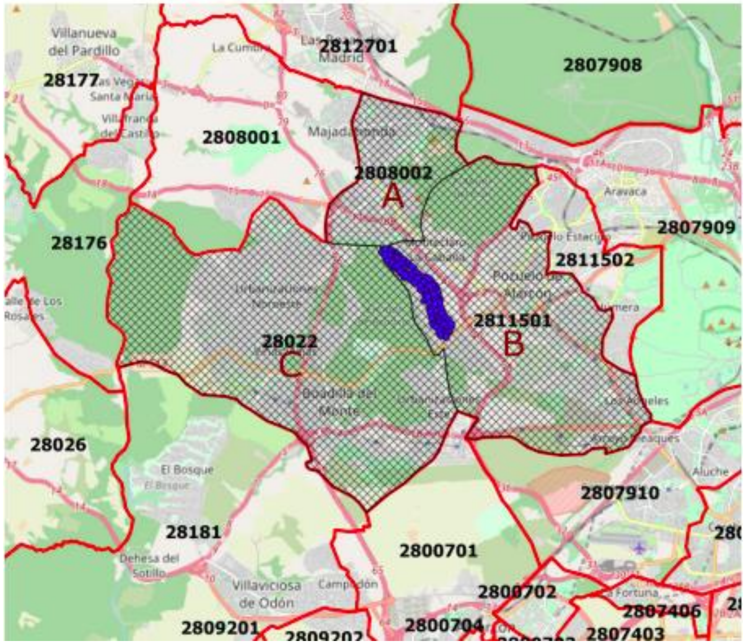
(D)



3. Human mobility as a timeseries problem.

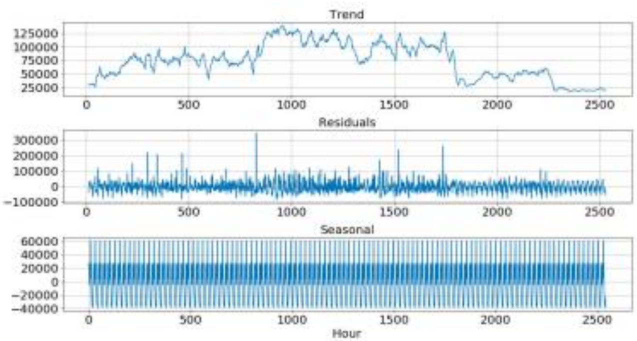
- **<u>2nd application: Mobility prediction</u>**

- We also apply this view for the prediction of the incoming home trips of three regions near the city of Madrid.
  - Movement data extracted from the *Open Data Movilidad* dataset.
- Novelty: We combine the movement data with household water consumption from different smart water meters.

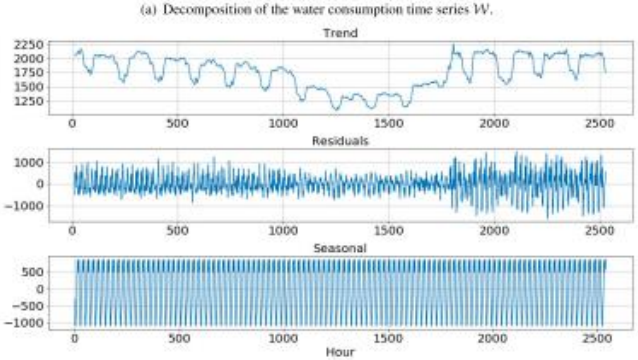


3. Human mobility as a timeseries problem.

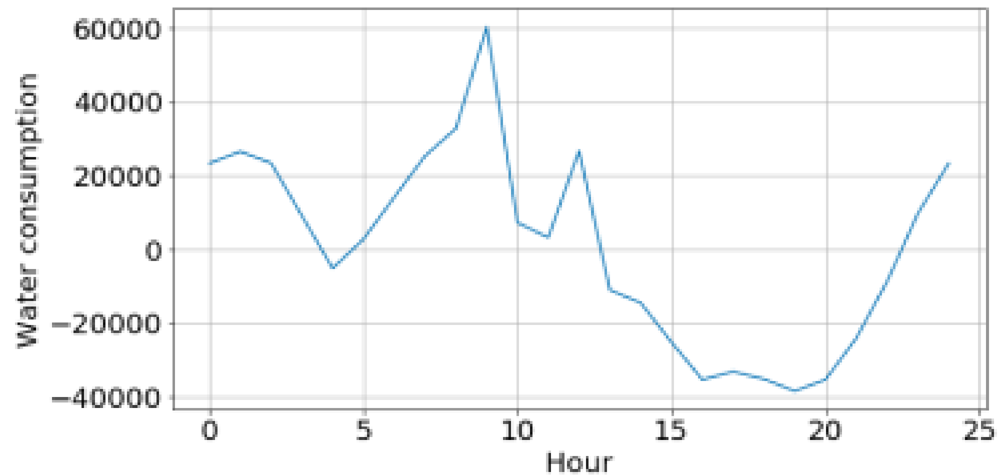
- We observed an interesting correlation between the seasonal components of the flow and water consumption timeseries.



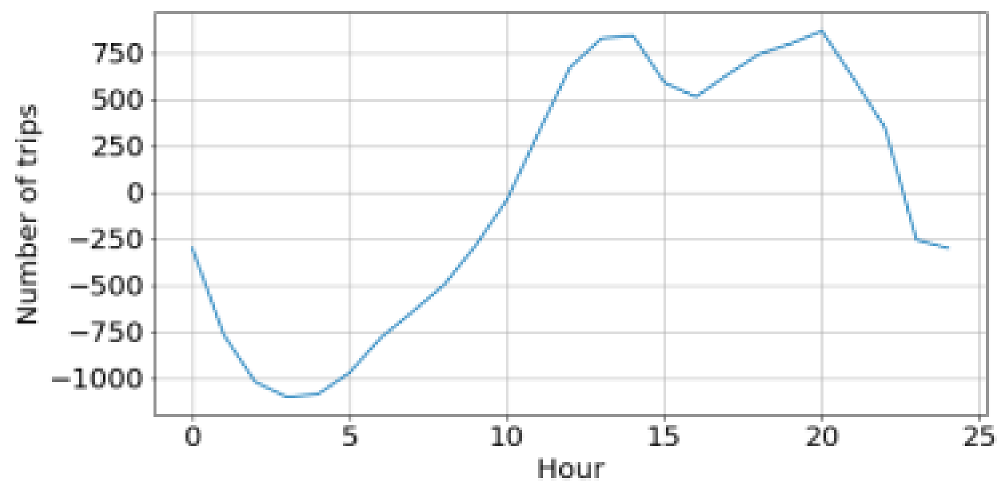
(a) Decomposition of the water consumption time series $\mathcal{W}$.



(b) Decomposition of the human flow time series $\mathcal{F}_{target}$.



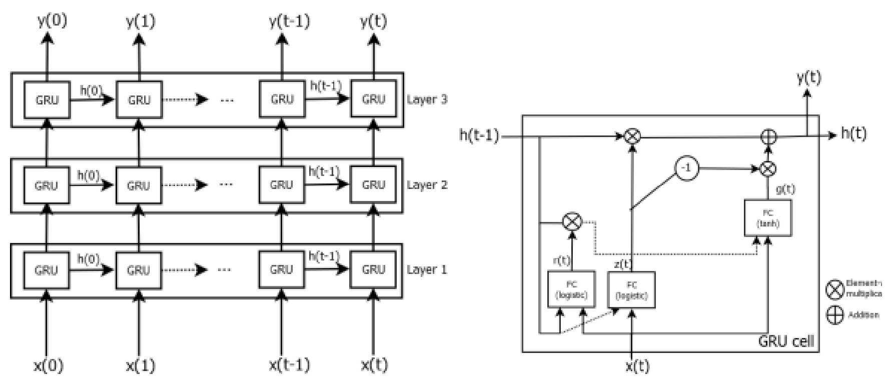
(a) Water-consumption series $\mathcal{W}$.



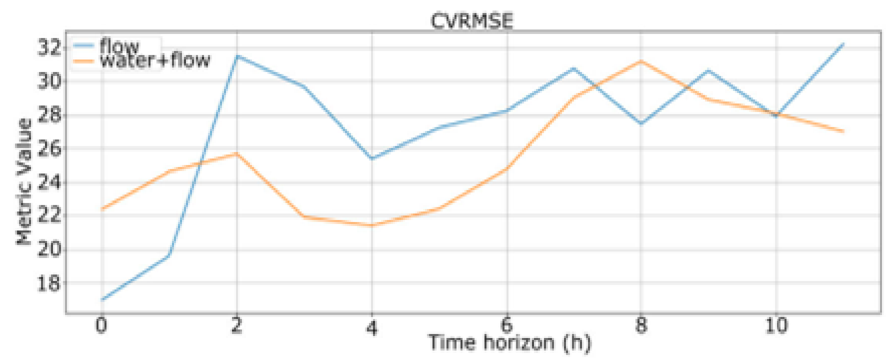
(b) Human-flow series $\mathcal{F}_{target}$.

3. Human mobility as a timeseries problem.

- We develop a DL predictor based on GRU cells.
- Results showed a clear improvement of the model enriched with the water consumption data.



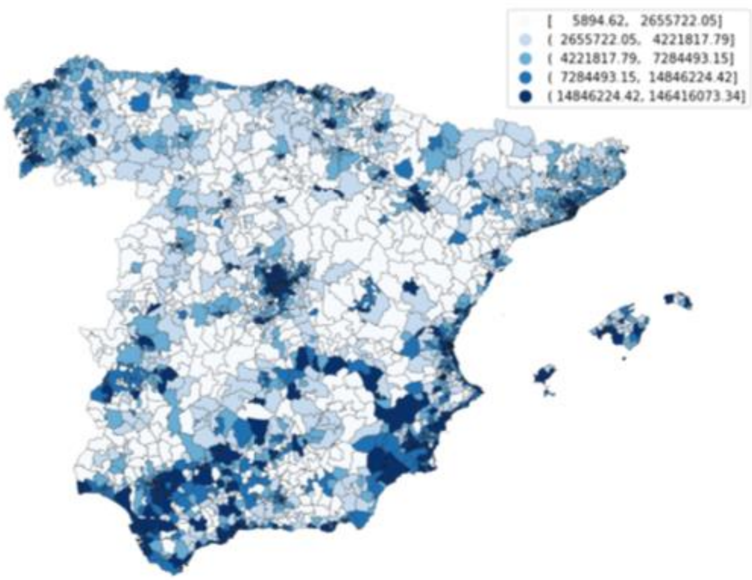
(a) Example of GRU architecture with three layers unrolled thought time. (b) Gated structure of a GRU cell. FC stands for Ful Connected.



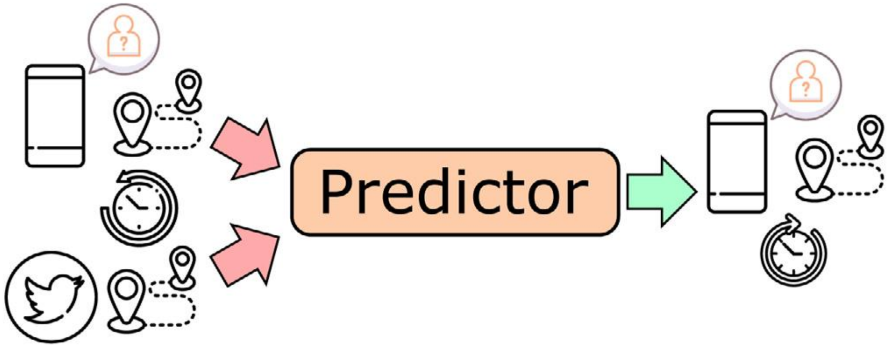
(b) CVMSE

3. Human mobility as a timeseries problem.

- We also develop a human mobility predictor to estimate the total number of displacements at nation level.
- We used the Open Data Movilidad to develop such a predictor.
- We also used a nation-wide Twitter dataset of human trips.

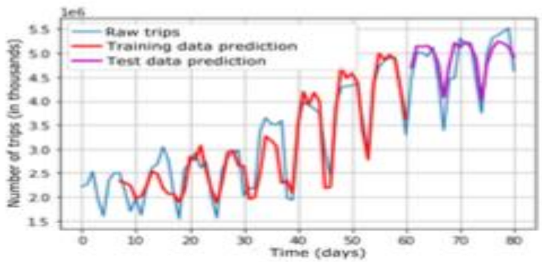


# of outgoing trips

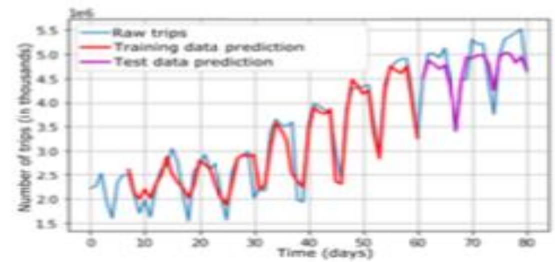


3. Human mobility as a timeseries problem.

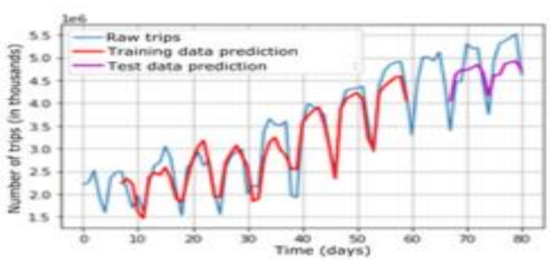
- We feed different types of ML models for timeseries forecasting after COVID-19 lockdown in Spain



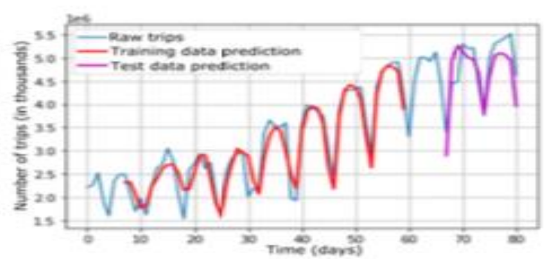
(a) MLP($\mathcal{T}_{SMT}$).



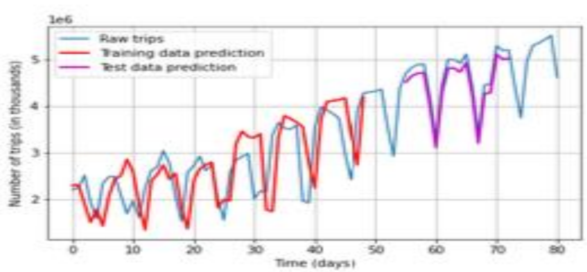
(b) MLP($\mathcal{T}_{SMT}$, $\mathcal{T}_{TWT}$).



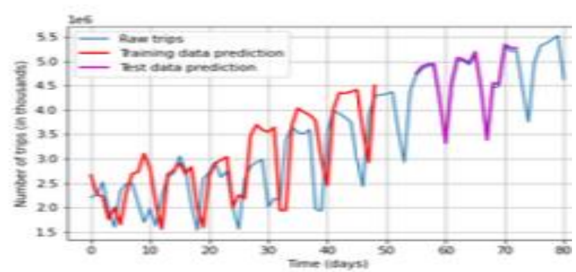
(c) LSTM($\mathcal{T}_{SMT}$).



(d) LSTM($\mathcal{T}_{SMT}$, $\mathcal{T}_{TWT}$).



(e) VARMAX($\mathcal{T}_{SMT}$, $\mathcal{T}_{TWT}$).



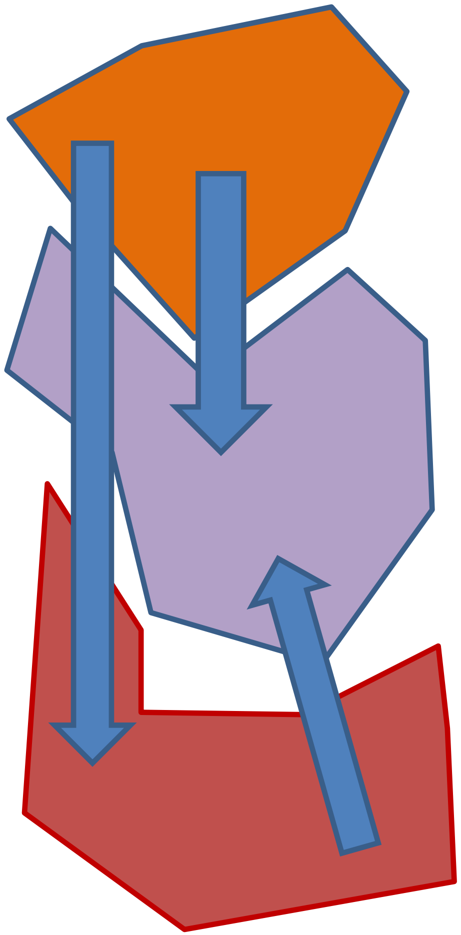
(f) VAR($\mathcal{T}_{SMT}$, $\mathcal{T}_{TWT}$).

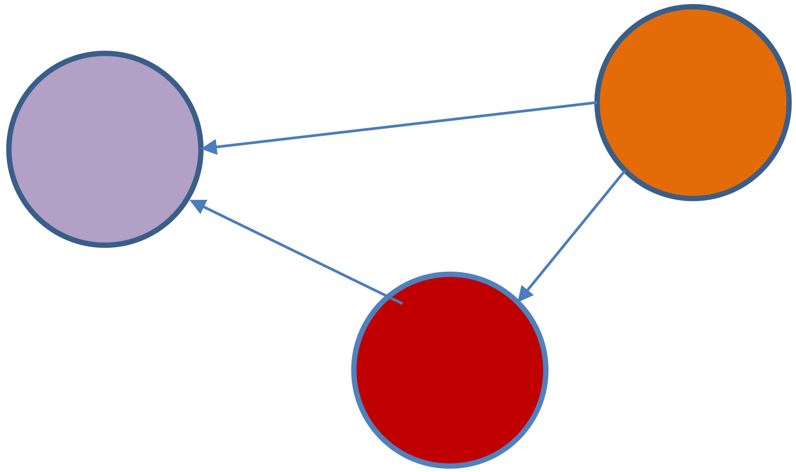
Agenda

1. Introduction*.*
2. Where's my (open) data?
3. Human mobility as a timeseries service
4. ***Human mobility as a graph-based service***
5. Real-time? Movement Data Analysis
6. Real-time, again? Traffic alerts
7. Mobility in the digital world.
8. Conclusion

4. Human mobility as a graph-based problem.

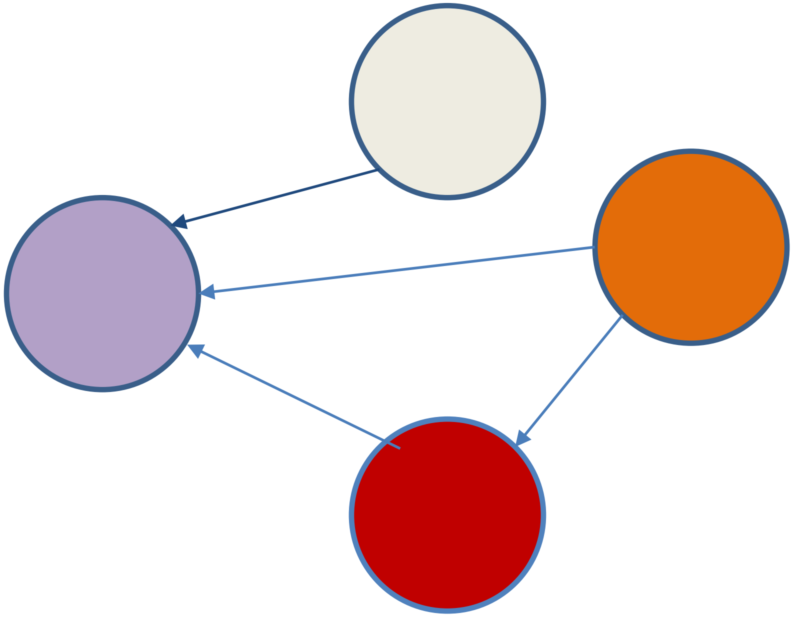
- The region-based flows derived from the open-data feeds allow us to represent the mobility of a spatial region as a graph.

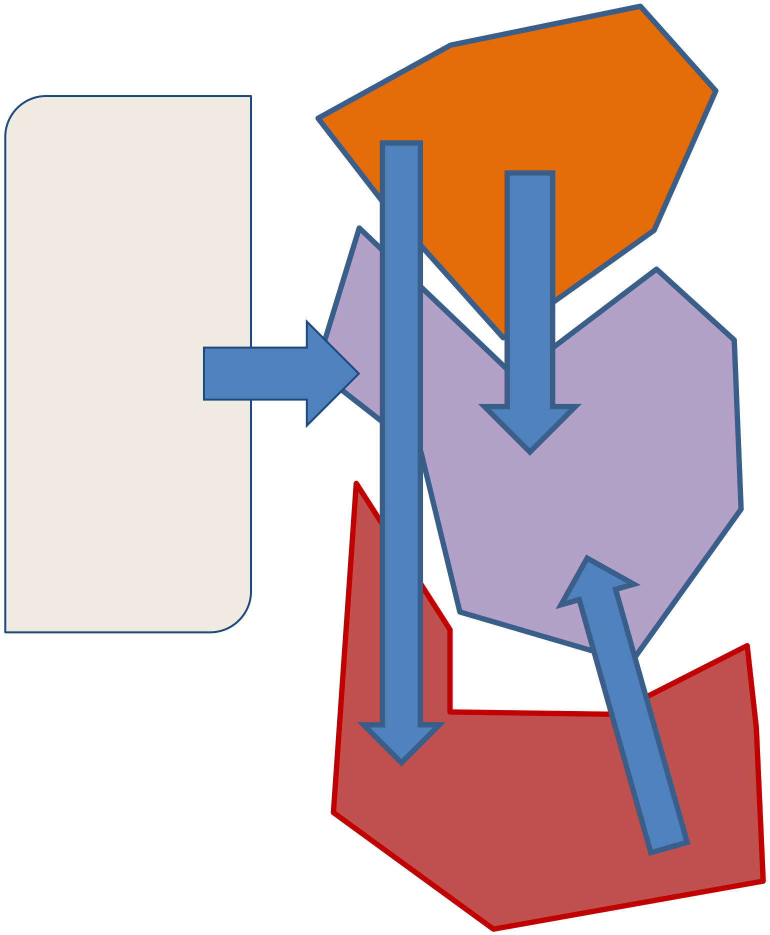




4. Human mobility as a graph-based problem.

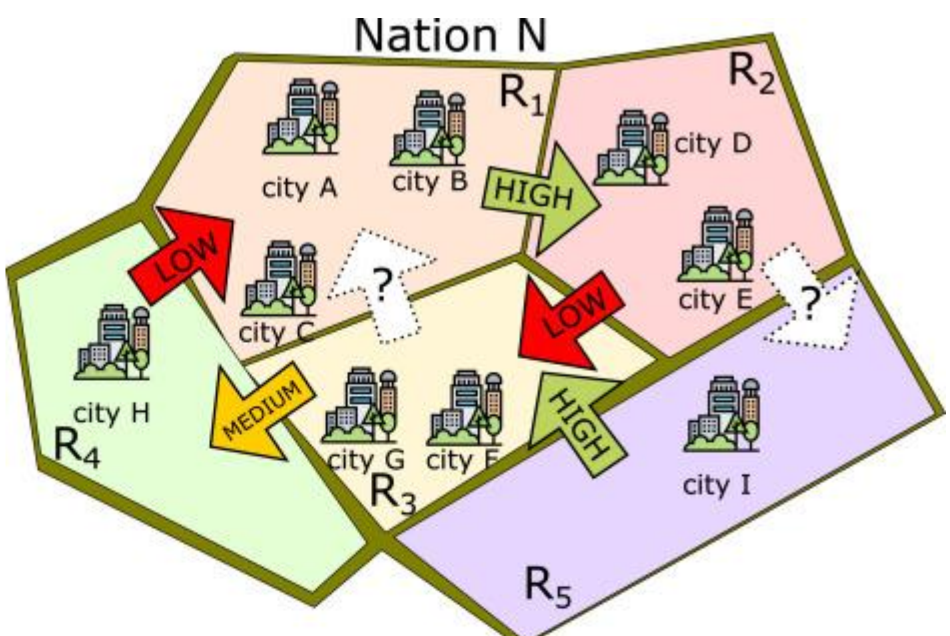
- At the same time, the DL field has witnessed the evolution of Graph Neural Networks (GNNs) as interesting solutions in fields can not be represented in a tabular or grilled format.





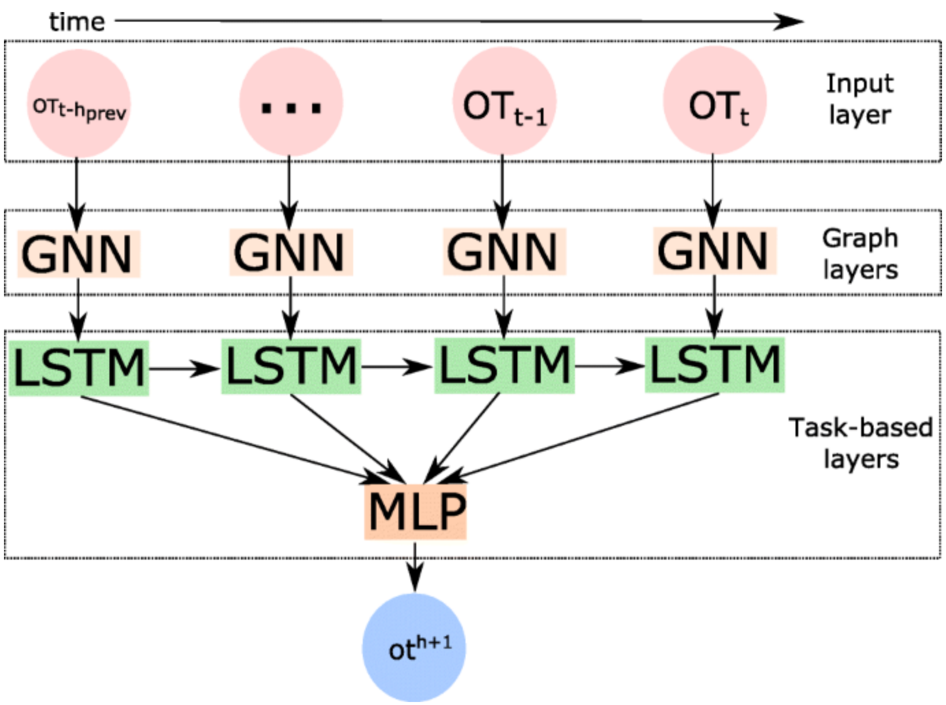
4. Human mobility as a graph-based problem.

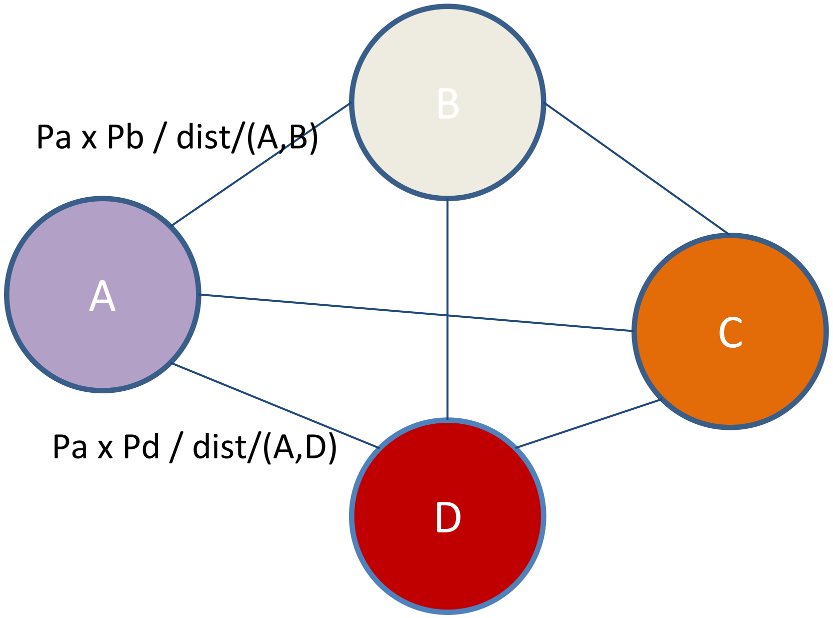
- We used the Open Data Movilidad dataset to predict nation-wide mobility.
- Target: Incoming types of flow for all TA (Touristic Areas).
- Combined with Twitter, weather & tourist infrastructure data, we aim to predict the demand of tourist accomodation
- We adopted a Graph approach → A single model!



4. Human mobility as a graph-based problem.

- Nodes were Spanish regions, edges based on Gravity model.

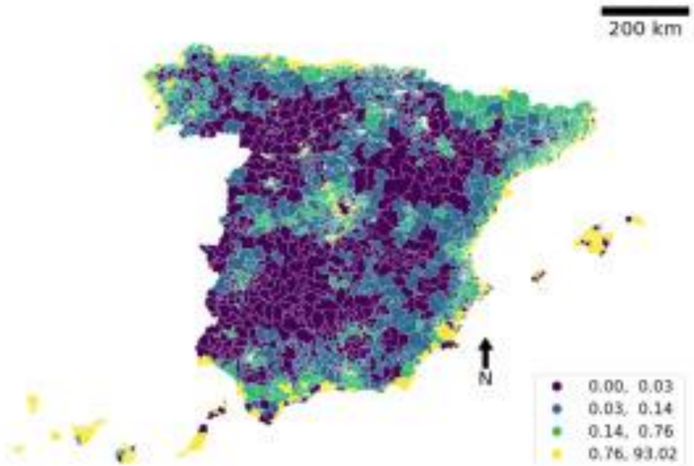




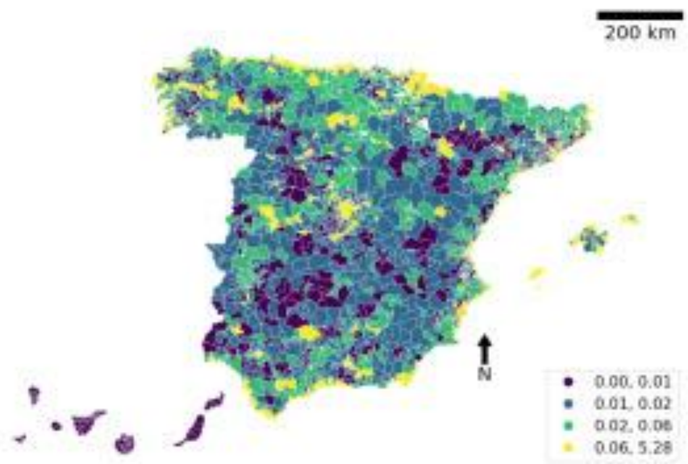
4. Human mobility as a graph-based problem.

- Weather data extracted from stations in Spanish airports (https://rp5.ru/)

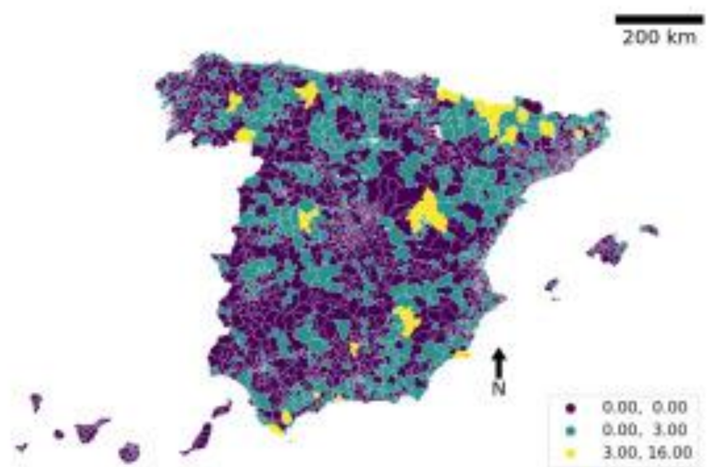
4. Human mobility as a graph-based problem.



(a) Distribution of touristic houses per TA.

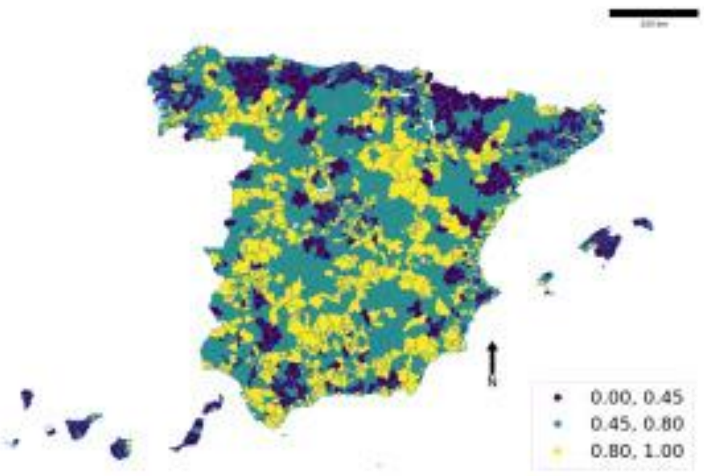


(b) Distribution of hotels per TA.



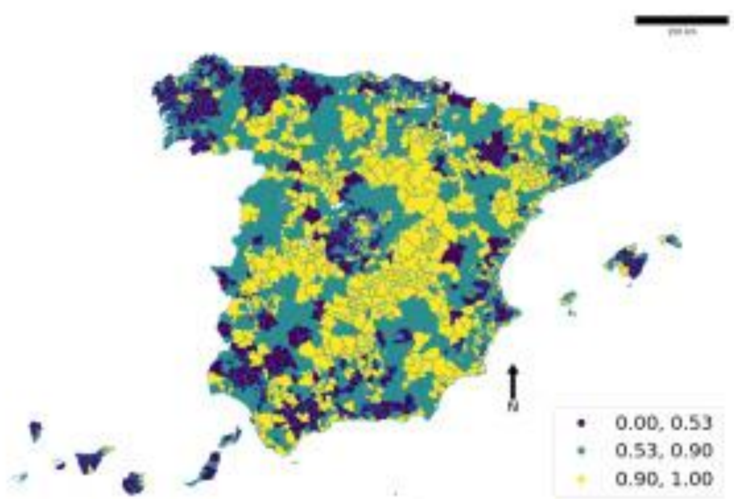
(c) Distribution of campsites per TA.

4. Human mobility as a graph-based problem.



(a) GNN_{low}.

(b) GNN_{medium}.



(c) GNN_{high}.

Agenda

1. Introduction*.*
2. Where's my (open) data?
3. Human mobility as a timeseries service
4. *Human mobility as a graph-based service*
5. ***Real-time? Movement Data Analysis***
6. Real-time, again? Traffic alerts
7. Mobility in the digital world.
8. Conclusion

5. ***Real-time? Movement Data Analysis***

- We have witnessed an increasing increment to develop real-time machine learning models.
- An endless discussion about online vs batch training of the models.

5. ***Real-time? Movement Data Analysis***

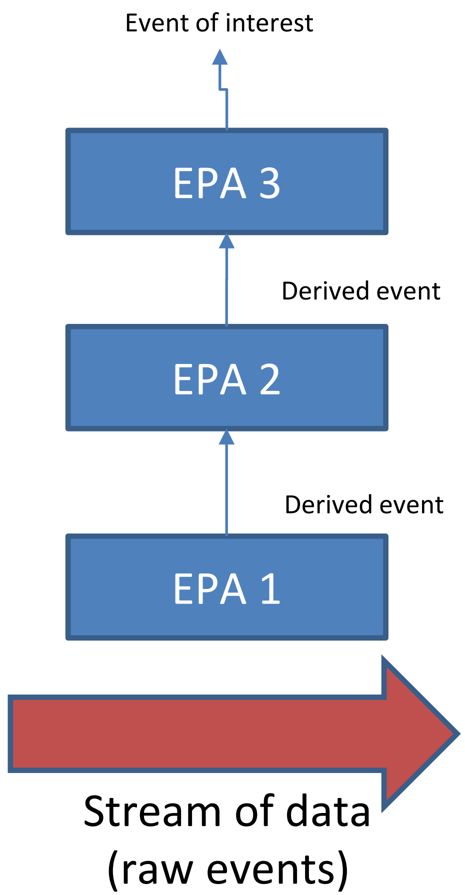
And what about rule-based models?

5. ***Real-time? Movement Data Analysis***

- There are some interesting platforms to develop event-driven applications that do not rely on ML models.
- We will briefly see one of them: Complex Event Processing (CEP) and how we can use it to process movement data.

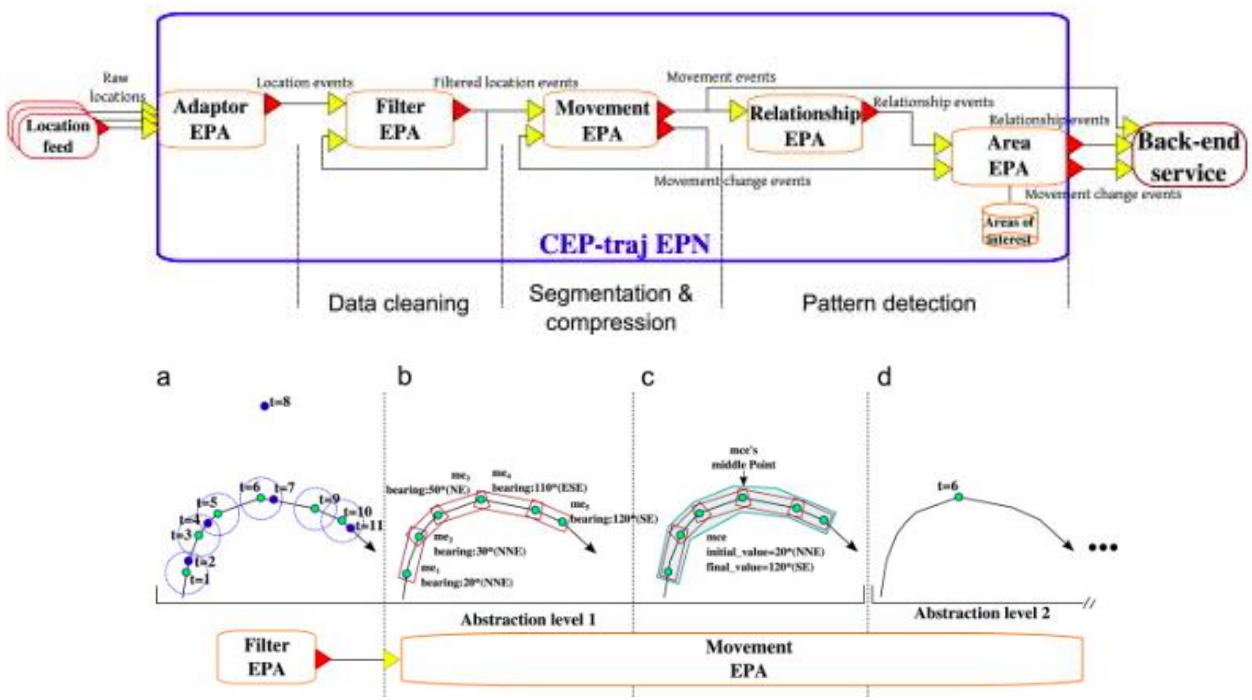
5. ***Real-time? Movement Data Analysis***

- But first, what is CEP?
  - CEP is a paradigm that focuses on timely processing unbounded flows of information items, so-called events.
  - A CEP system compounds a set of Event Processing Agents.
  - An EPA is a software module that consumes certain types of events (raw or derived) and creates new (derived) ones
  - They are based on event-condition-action rules.
    - Detect event → Fire rule → perform action



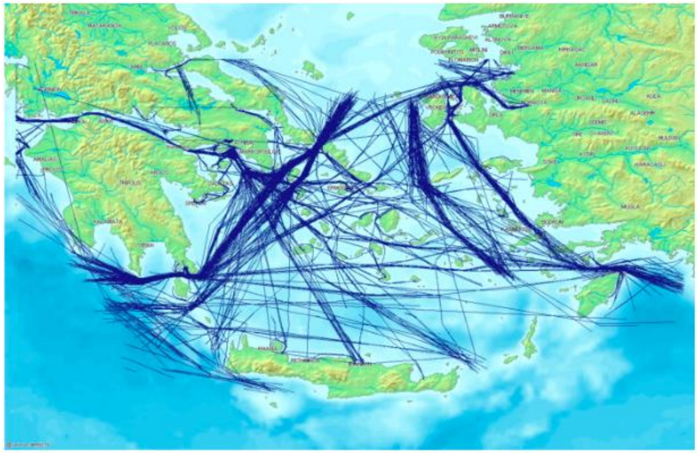
5. ***Real-time? Movement Data Analysis***

- Based on this architecture we defined *CEP-traj*. to filter and detect meaningful events in a spatio-temporal trajectories.



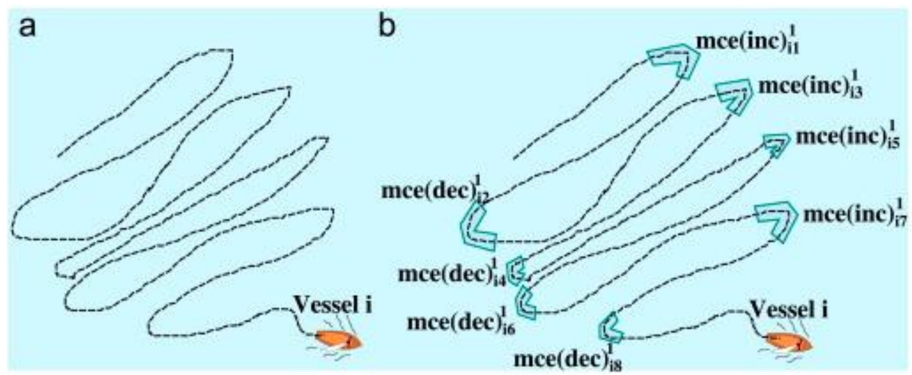
5. ***Real-time? Movement Data Analysis***

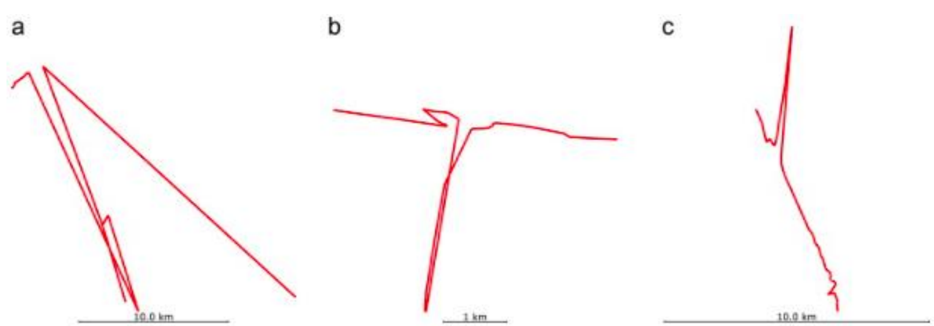
- We apply CEP-traj to a large dataset of vessel data.
  - 993 trajectories with 3 173 957 spatio-temporal points (3402 points per trajectory on average) covering a 3-day period



5. *Real-time? Movement Data Analysis*

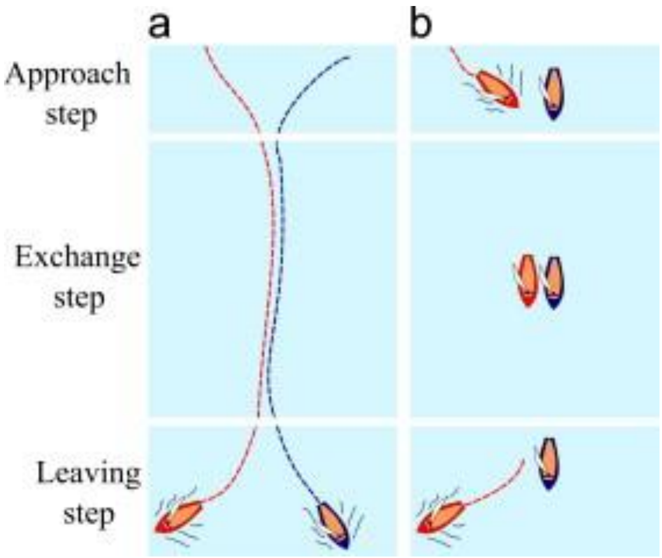
- One of the goals was to identify ilegal fishing behaviours.

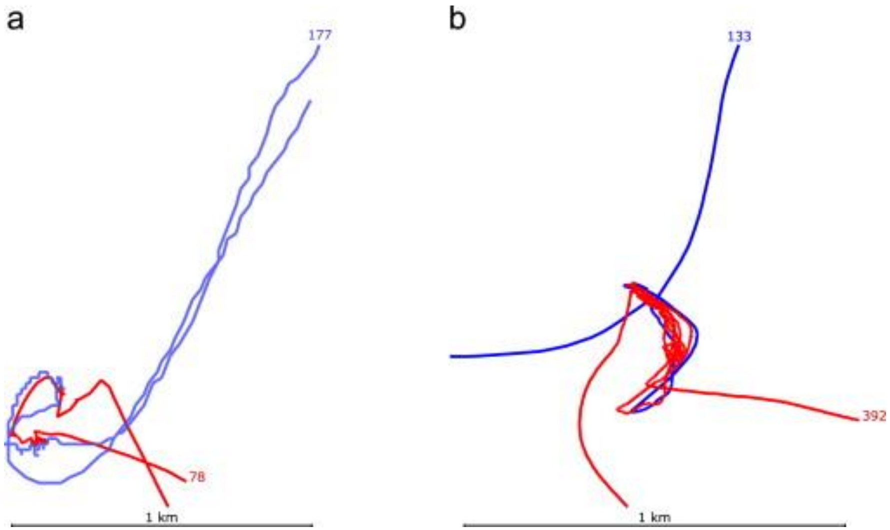




5. ***Real-time? Movement Data Analysis***

- Another appilication: Smuggling behaviours





Agenda

1. Introduction**.**
2. Where’s my (open) data?
3. Human mobility as a timeseries service
4. *Human mobility as a graph-based service*
5. Real-time? Movement Data Analysis
6. ***Real-time, again? Traffic alerts***
7. Mobility in the digital world.
8. Conclusion

6. ***Real-time, again ? Traffic alerts***

Introduction

Background

Traffic congestion is a global problem that especially affects urban areas.

Road operators utilize a variety of sensors to detect vehicles, including inductive loop detectors, microwave radar detectors, and **video cameras**.

Deep Learning (DL) techniques such as **Convolutional Neural Networks (CNN)** are increasingly used for **image recognition** tasks.

Challenge

We aim to deliver a reliable **traffic alert service to road operators**, whether public or private, by utilizing their **own data sources**. This allows them to avoid dependence on third-party data providers such as Google Maps or Waze.

The service needs to be tested **using real-world images** from different camera types, weather conditions, light intensity, camera position, etc.

6. ***Real-time, again ? Traffic alerts***

Introduction

Objectives

Design of a system for real-time traffic alerts based on the analysis of images captured from the network of cameras of the DGT in Spain

DGT
Dirección General de Tráfico

6. ***Real-time, again ? Traffic alerts***

FLUID

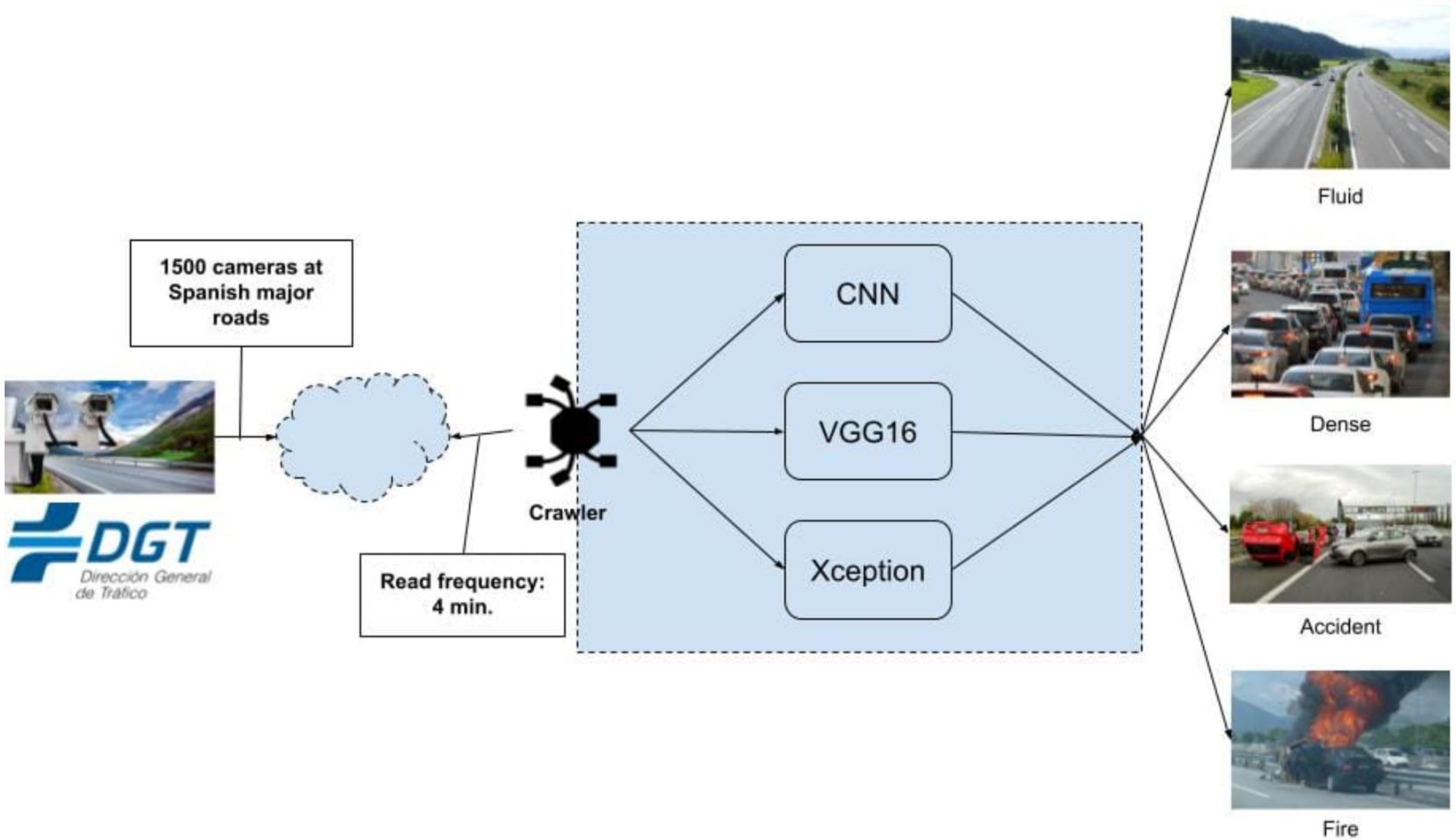
DENSE

ACCIDENT

FIRE

6. ***Real-time, again ? Traffic alerts***

System Architecture. Proposal



6. ***Real-time, again ? Traffic alerts***

System Architecture. Datasets

1

Traffic-Net2: Used for training and test the models

It contains 4,400 images covering 4 classes, namely *Accident*, *Dense traffic*, *Fluid traffic* and *Fire*.

There are 1,100 images for each category (900 images for training and 200 images for testing). As a result, we have a dataset of 3600 images for training and 800 for testing

2

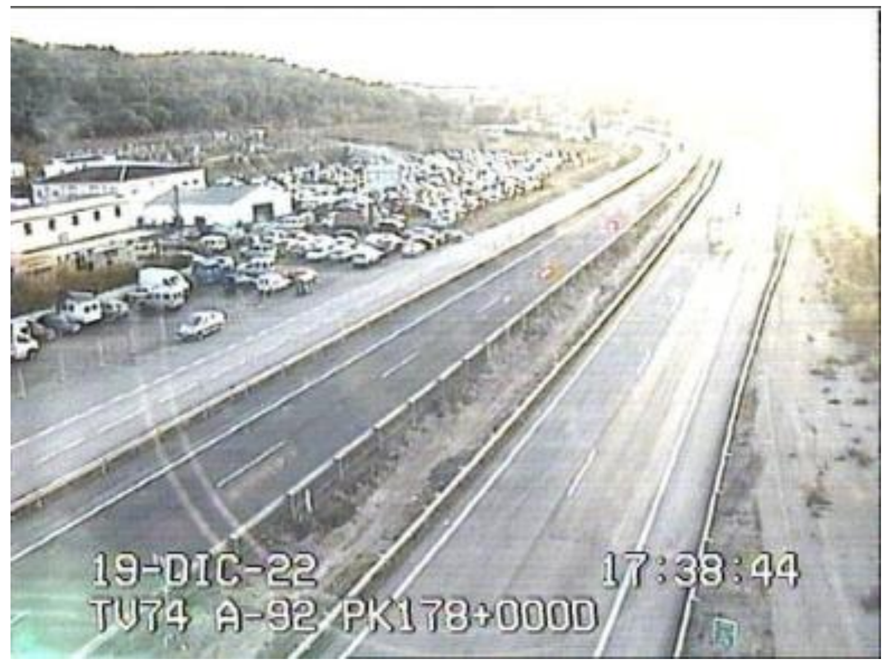
DGT dataset: Test dataset, manually labelled by capturing traffic images from the DGT cameras every 4 minutes during a week (14-12-2022 to 21-12-2022).

A total of 595 images, being 337 classified as *dense traffic* and 258 as *fluid traffic*.

Affected by weather phenomena, such as fog or rain

6. ***Real-time, again ? Traffic alerts***

System Architecture. Datasets



6. ***Real-time, again ? Traffic alerts***

System Architecture. Image Recognition Models

1

CNN model: Two convolutional layers, two max-pooling layers, and one fully connected layer.

ReLU activation function applied in convolutional and dense layer.

Output is a layer with four perceptrons and the softmax activation function.
Implemented in Python 3.10 with the Keras library

2

VGG16: Blocks composed of an incremental number of convolutional layers with filters of size 3×3.

Final part, two fully connected dense layers of 4096 neurons each and a final softmax output layer of 1000 neurons.

Pre-trained with the ImageNet image library

3

Xception: Based on a model called Inception, which uses wider and shallower layers

Introducing in the same layer several convolution operators where their results are finally concatenated.

Pre-trained with the ImageNet image library

6. ***Real-time, again ? Traffic alerts***

Results. Metrics

Precision: Proportion of classifications that were correct.

Recall: Proportion of real positives that were correctly identified

F1-score: Balance between accuracy and recall

Accuracy: Average value for each of the previous measures

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1\text{-}score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

6. ***Real-time, again ? Traffic alerts***

Results. Evaluation with Traffic-Net

TABLE I
EVALUATION ON TEST DATA FOR THE PROPOSED CNN MODEL

	Precision	Recall	F1-score
Accident	0.62	0.78	0.70
Dense	0.79	0.74	0.77
Fire	0.83	0.84	0.84
Fluid	0.83	0.68	0.74
Accuracy			0.76

TABLE II
EVALUATION ON TEST DATA FOR THE PROPOSED VGG16 MODEL

	Precision	Recall	F1-score
Accident	0.80	0.87	0.83
Dense	0.92	0.85	0.89
Fire	0.93	0.93	0.93
Fluid	0.85	0.85	0.85
Accuracy			0.88

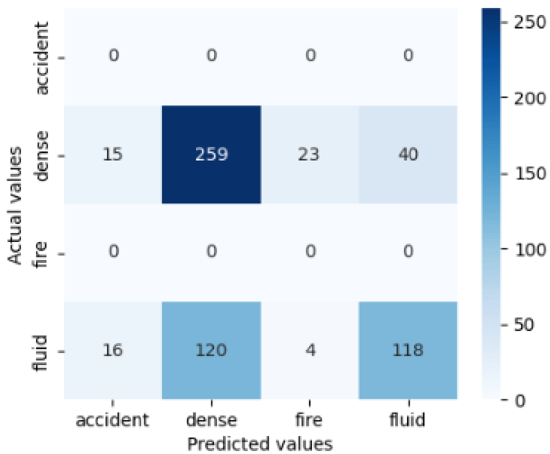
TABLE III
EVALUATION ON TEST DATA FOR THE PROPOSED XCEPTION MODEL

	Precision	Recall	F1-score
Accident	0.92	0.94	0.93
Dense	0.97	0.92	0.94
Fire	0.97	0.97	0.97
Fluid	0.90	0.93	0.91
Accuracy			0.94

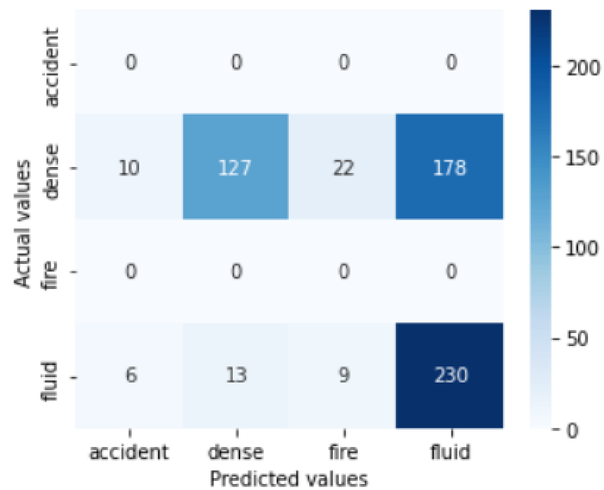
6. ***Real-time, again ? Traffic alerts***

Results. Evaluation with DGT dataset

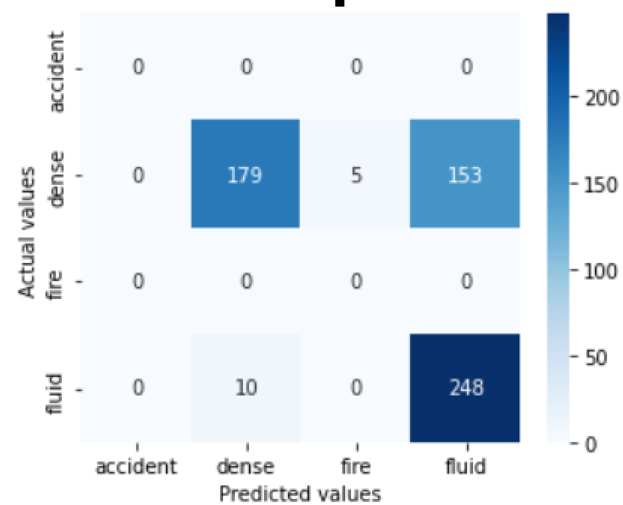
CNN



VGG16



Xception



6. ***Real-time, again ? Traffic alerts***

Results. Limitations

Dense, classifed as Fire

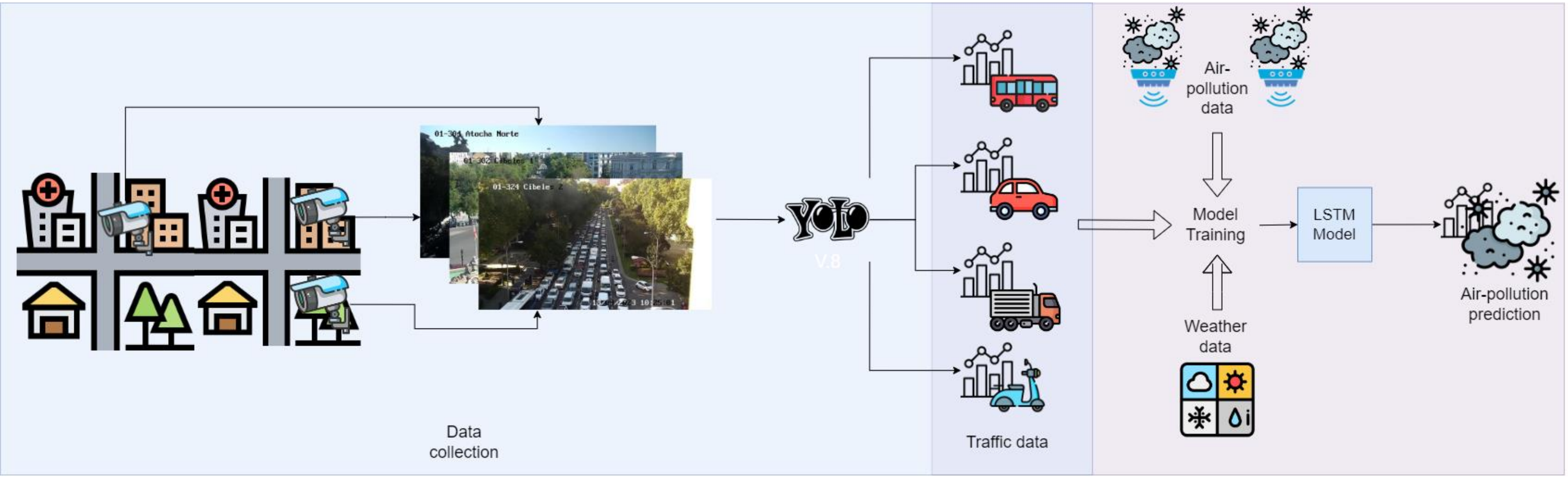
Fluid, classified as Dense

6. ***Real-time, again ? Traffic alerts***

Ongoing Work

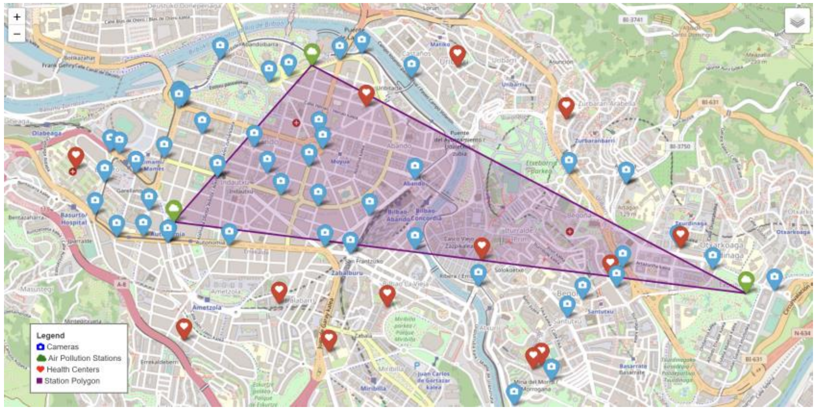
6. ***Real-time, again ? Traffic alerts***

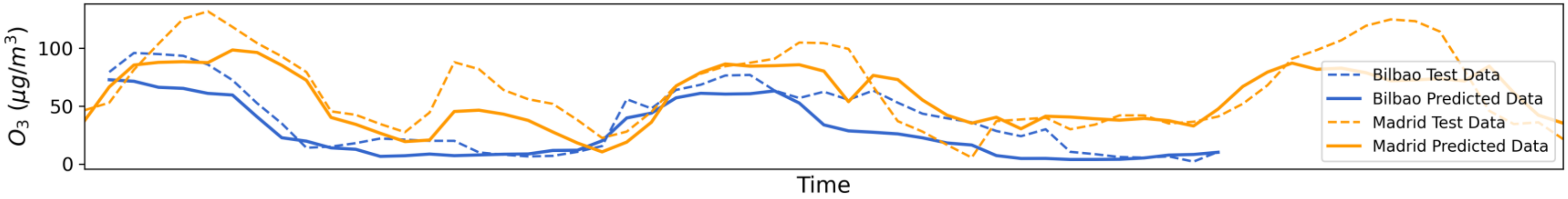
Ongoing Work. Air pollution prediction

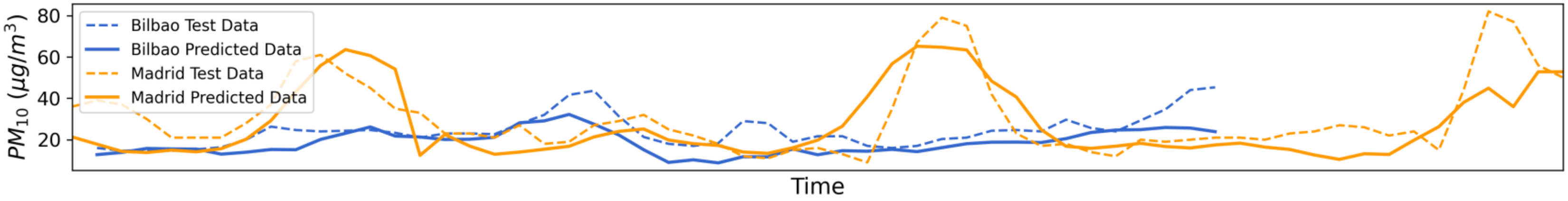


6. ***Real-time, again ? Traffic alerts***

Ongoing Work: Air-Pollution Prediction







Agenda

1. Introduction**.**
2. Where’s my (open) data?
3. Human mobility as a timeseries service
4. *Human mobility as a graph-based service*
5. Real-time? Movement Data Analysis
6. Real-time, again? Traffic alerts
7. ***Mobility in the digital world***
8. Conclusion

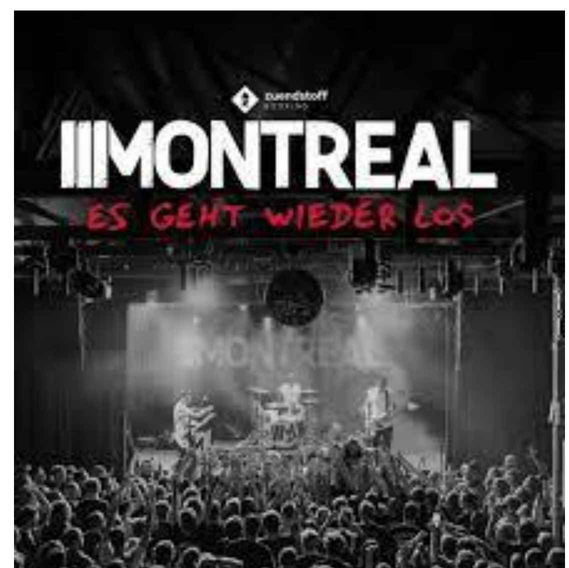
7. Mobility in the digital world.

- The research community has also studied the mobility patterns that arise in digital environments.
  - The patterns of users when they are surfing the net (places are websites and movements are clicks on links).
  - Other works have made comparative analyses between traces of World of Warcraft (WoW) players and GPS traces*.

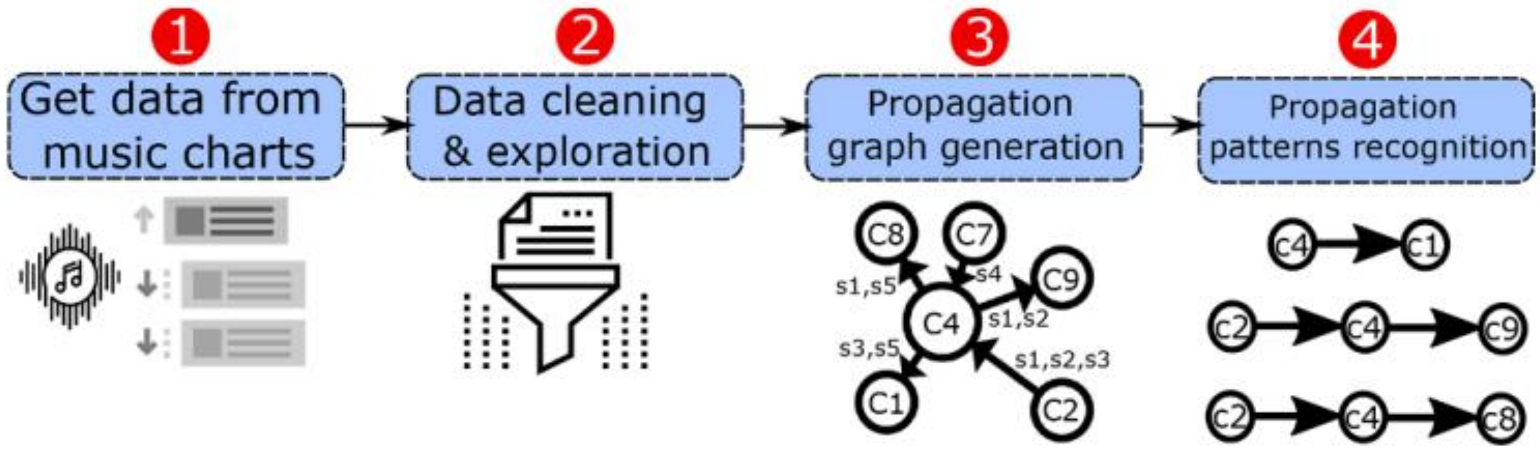
*S. Shen, N. Brouwers, A. Iosup, D. Epema, Characterization of human mobility in networked virtual environments, Proceedings of Network and Operating System Support on Digital Audio and Video Workshop (2014) 13:13—-13:18.

7. Mobility in the digital world.

- In our research team, we ask ourselves some questions:
  - *And how about the music sector?*
  - *Some songs propagate from one country to another until they become worldwide hits.*
  - *Could we regard songs as travellers that move from one country to another to become music hits?*
  - *Could we extract mobility patterns from such propagation?*

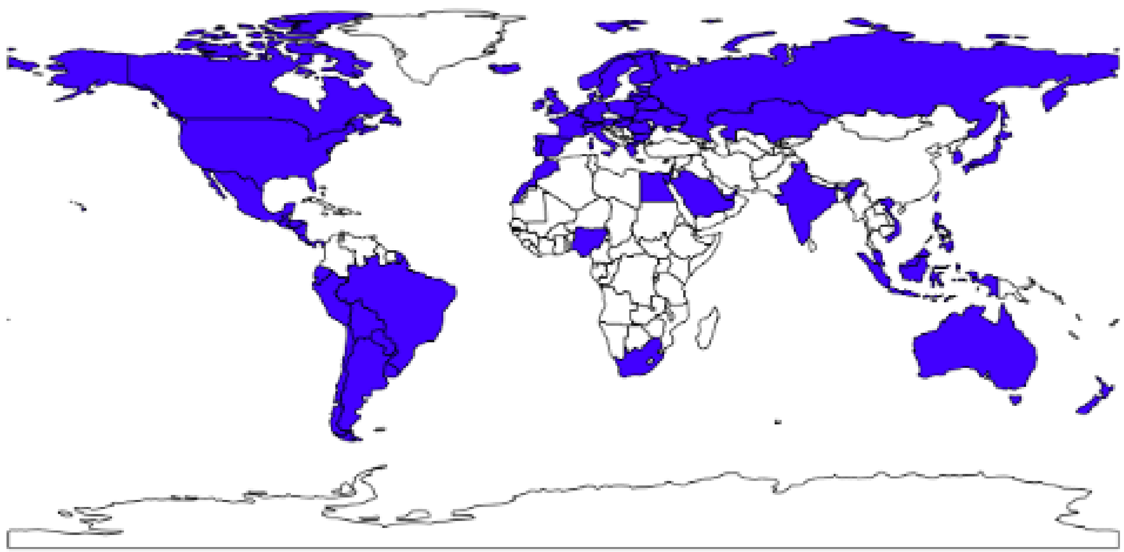


7. Mobility in the digital world.



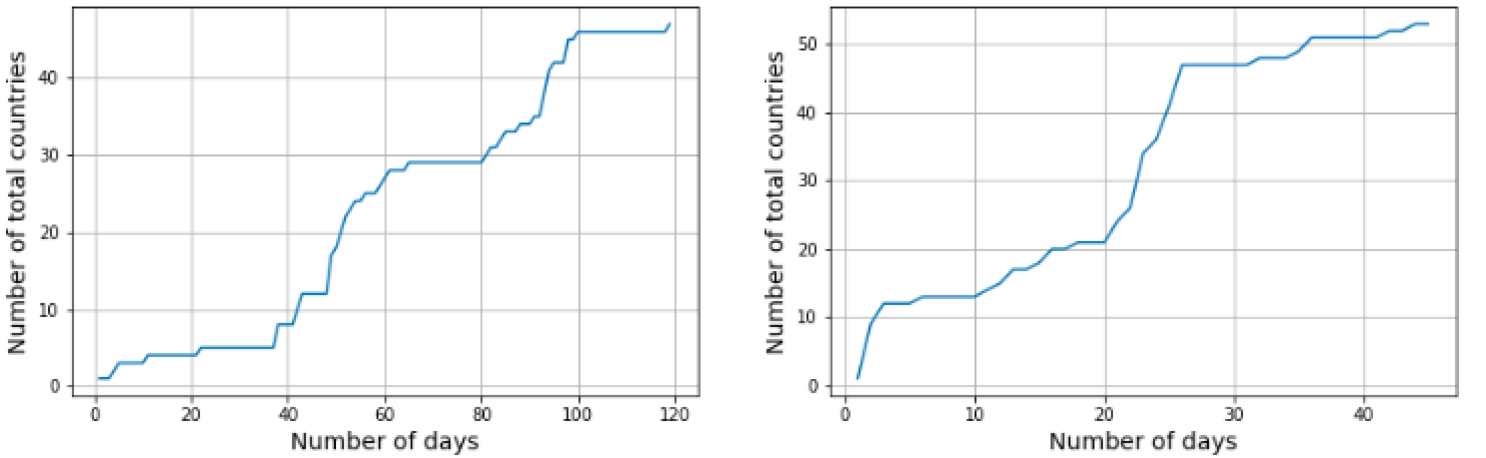
7. Mobility in the digital world.

We extract from Spotify Chart the top-200 list of 69 countries covering a 4-year period.



7. Mobility in the digital world.

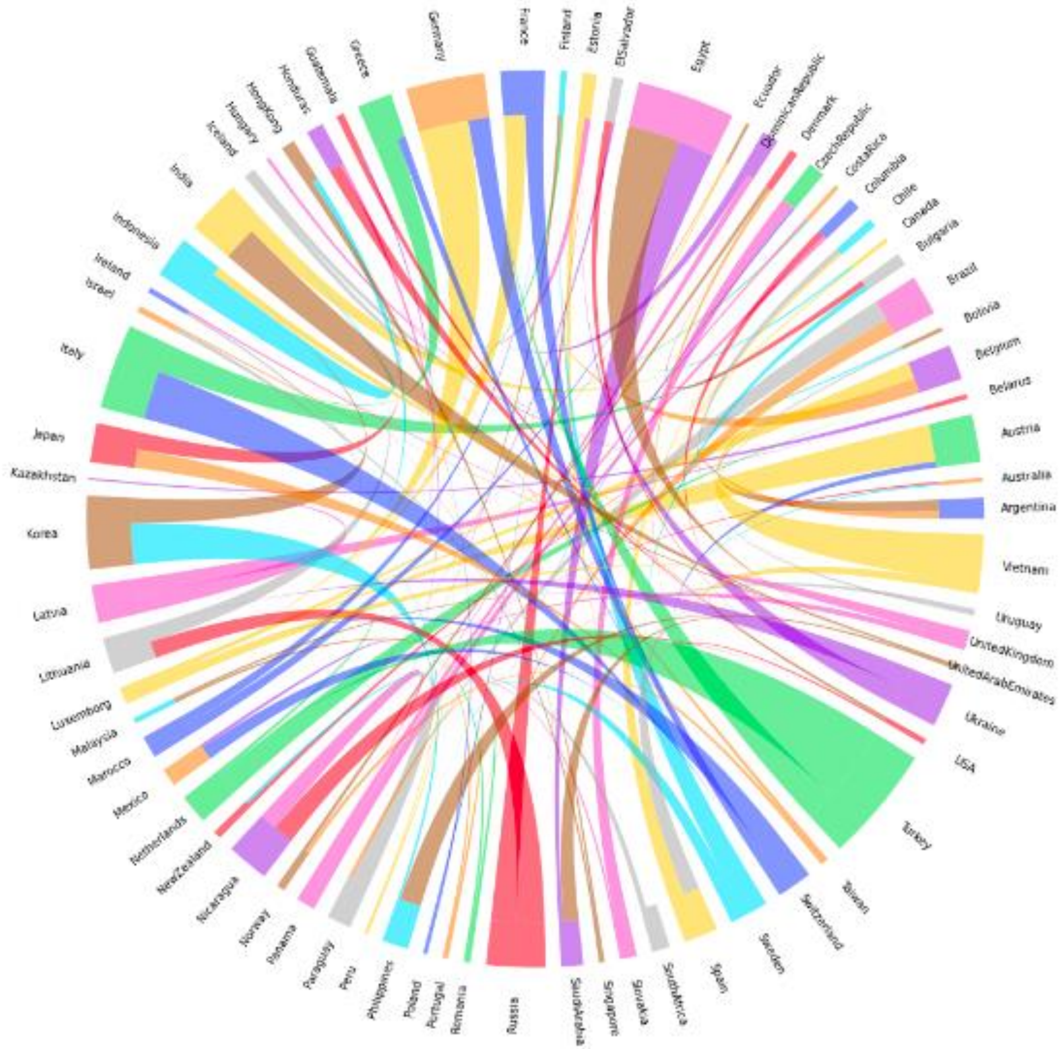
- We remove mainstream songs from very well-known singers or bands.
- We just kept with songs showing a snowball propagation.



(a) Dance Monkey. Artist: Tones and I. (b) Life Goes On. Artist: Oliver Tree.

7. Mobility in the digital world.

- How do these songs propagate across countries?



7. Mobility in the digital world.

Table 2. The five MMPs with the highest weight for lengths 3, 4 and 5. The frequency of each MMP is shown in brackets.

Pos	$\mathscr{M}_{(3,)}$	$\mathscr{M}_{(4,)}$	$\mathscr{M}_{(5,)}$
1	Lithuania → Norway → United Kingdom (77)	Switzerland → Belgium → Lithuania → Latvia (64)	USA → Hungary → Belgium → Lithuania → Latvia
2	Iceland → Lithuania → Belgium (70)	Netherlands → Finland → Slovakia → Hong Kong (60)	Norway → Denmark → Slovakia → Belgium → Estonia (64)
3	Iceland → Sweden → Belgium (65)	Norway → Denmark → Slovakia → Hong Kong (60)	Norway → Latvia → Switzerland → Slovakia → Hong Kong
4	Slovakia → Lithuania → Latvia (64)	UK → Australia → New Zealand → South Africa (60)	Norway → Denmark → Finland → Slovakia → Hong Kong (60)
5	USA → Hungary → Latvia (64)	UK → Australia → New Zealand → Romania (60)	Netherlands → Finland → Slovakia → Belgium → Estonia (60)

7. Mobility in the digital world.

- Prediction timeline of the pop song ”Toxic” (BoyWithUke)

Date	$\mathcal{P}_s^d$	$\mathcal{C}_s^d$	$\mathcal{O}_s^d$
2021-10-11	Indonesia → Lithuania	Austria	Estonia
2021-10-12	Indonesia→ Lithuania→ Austria	Slovakia, **Germany** , Malaysia	**Germany**
2021-10-13	Indonesia →Lithuania → Austria → Slovakia Indonesia →Lithuania → Austria → Malaysia Indonesia →Lithuania → Austria → Germany Indonesia →Lithuania → Austria → Slovakia Indonesia →Lithuania → Austria → Malaysia Indonesia →Lithuania → Austria → Germany Indonesia →Lithuania → Austria → Slovakia Indonesia →Lithuania → Austria → Malaysia	Latvia, Hungary, CzechRepublic	Slovakia, Austria, Portugal
2021-10-14	Indonesia →Lithuania → Austria → Germany → Latvia Indonesia →Lithuania → Austria → Germany → Hungary Indonesia →Lithuania → Austria → Germany → CzechRepublic Indonesia →Lithuania → Austria → Slovakia → Latvia Indonesia →Lithuania → Austria → Slovakia → Hungary Indonesia →Lithuania → Austria → Slovakia →Czech Republic Indonesia →Lithuania → Austria → Malaysia → Latvia Indonesia →Lithuania → Austria → Malaysia → Hungary Indonesia →Lithuania → Austria → Malaysia → Czech Republic	Switzerland, **Estonia**, Bolivia	Italy, **Estonia**, Netherlands, Ireland, Denmark, Australia
2021-10-18	Austria → Malaysia → Czech Republic → Bolivia	Costa Rica, Panama, **Ecuador**	**Ecuador**
2021-10-21	Austria → Malaysia → Czech Republic → Bolivia → Costa Rica Austria → Malaysia → Czech Republic → Bolivia → Panama Austria → Malaysia → Czech Republic → Bolivia → Ecuador	**Peru**	El Salvador, **Peru**

Agenda

1. Introduction**.**
2. Where's my (open) data?
3. Human mobility as a timeseries service
4. *Human mobility as a graph-based service*
5. Real-time? Movement Data Analysis
6. Real-time, again? Traffic alerts
7. Mobility in the digital world.
8. ***Conclusion***

8. Conclusions

- From the Covid-19 pandemic, there is a large interest in human mobility analysis.
- The heterogeneity of data sources is one of the key points to consider in this domain now.
- In open datasets where flows are defined as region-based links, timeseries or graph-based approaches are promising solutions.

8. Conclusions

- Future work will define new integration of datasources in graph-based models.

- The development of interactive dashboards to provide final services to citizens.

- Use of LLMs in analyzing mobility data?
  https://arxiv.org/abs/2308.15197
  https://arxiv.org/abs/2402.14744

Papers

Terroso-Sáenz, F., Muñoz, A., Fernández-Pedauye, J., & Cecilia, J. M. (2021). Human Mobility Prediction With Region-Based Flows and Water Consumption. *IEEE Access*, *9*, 88651-88663.

Terroso-Saenz, F., Flores, R., & Muñoz, A. (2022). Human mobility forecasting with region-based flows and geotagged Twitter data. *Expert Systems with Applications*, 117477.

Terroso-Sáenz, F., & Muñoz, A. (2022). Nation-wide human mobility prediction based on graph neural networks. *Applied Intelligence*, *52*(4), 4144-4160.

Terroso-Saenz, F., Valdés-Vela, M., den Breejen, E., Hanckmann, P., Dekker, R., & Skarmeta-Gomez, A. F. (2015). CEP-traj: An event-based solution to process trajectory data. *Information Systems*, *52*, 34-54.

Terroso-Saenz, F., Soto J., R., & Muñoz, A.. Music Mobility Patterns: How Songs Propagate Around The World Through Spotify. Pattern Recognition

Code: https://github.com/fterroso

Thank You for Your Attention!
Questions?

Acknowledgements

CREPES
CREACIÓN Y EVALUACIÓN EN ESCENARIOS AMPLIADOS DE APRENDIZAJE CON PRESERVACIÓN DE PRIVACIDAD
Project grant PID2020-115844RB-I00

PHADAS
Human-Centered Privacy for Sustainable Open Data Analytics
Project grant TED2021-132073B-I00

ALLEGRO
Smart multi-modal crowdsensing-based system as a service oriented to the prediction of social problems
Project grant PID2020-112827GB-I00

MINISTERIO DE CIENCIA E INNOVACIÓN

AGENCIA ESTATAL DE INVESTIGACIÓN

Funded by the European Union
NextGenerationEU