I-CiTies 2019

Pisa, 18 September 2019

Vehicle mobility data analysis and Individual Mobility Networks for crash prediction

Mirco Nanni

Leonardo Longhi

Sistematica

Work funded by:

Track & Know

Big Data for Mobility Tracking Knowledge Extraction in Urban Areas

EU H2020 G.A. 780754

Crash risk prediction

- **Context:** insurance companies
  - Base their business on evaluating the expected costs (for the company) of a client
  - Cost is a function of client's risk of suffering an accident/loss
  - Specific to car insurance: cost ~ crash risk in the next term

- **Problem definition**: given the recent mobility history of a user predict whether he/she will have a crash in the next future
  - E.g. predict crashes in the next month based on data of the last 4

State of the Art

Mainly based on

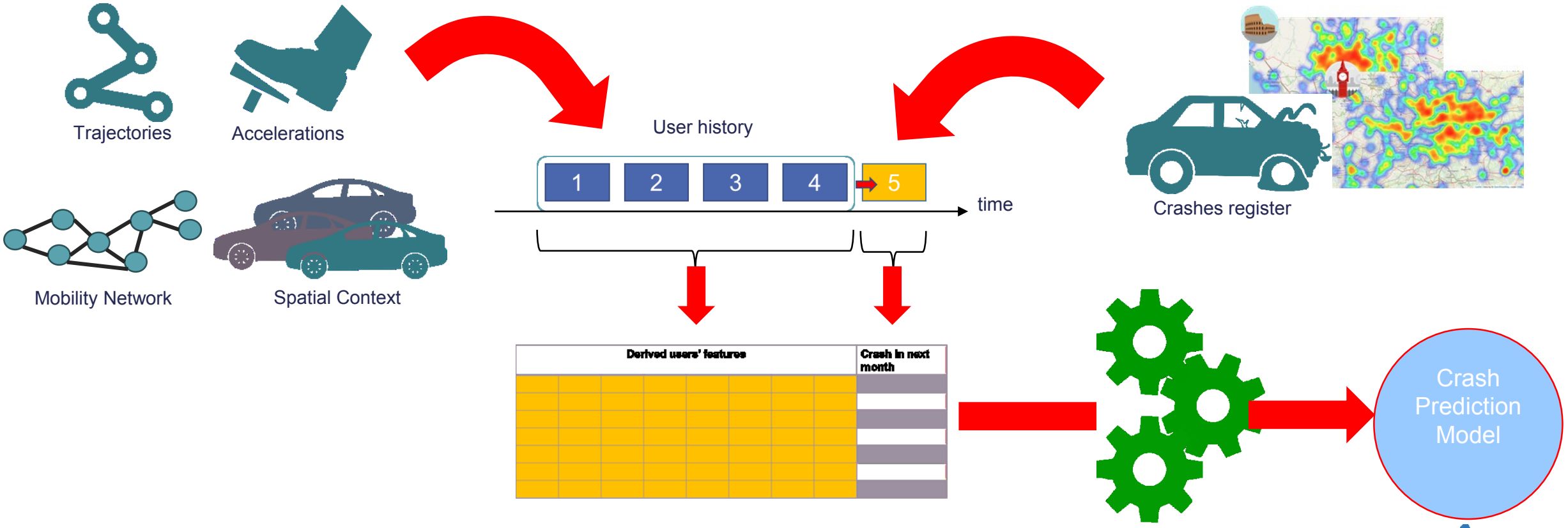
- Real Time Crash Prediction
- Contextual conditions (weather, traffic intensity)

Our approach:

- Individual, medium-term prediction (e.g. 1 month)
- Based on several features extracted from the history of the individual

The approach

- Machine Learning approach: directly learn from past crashes



Individual Trajectory Features

Trips count

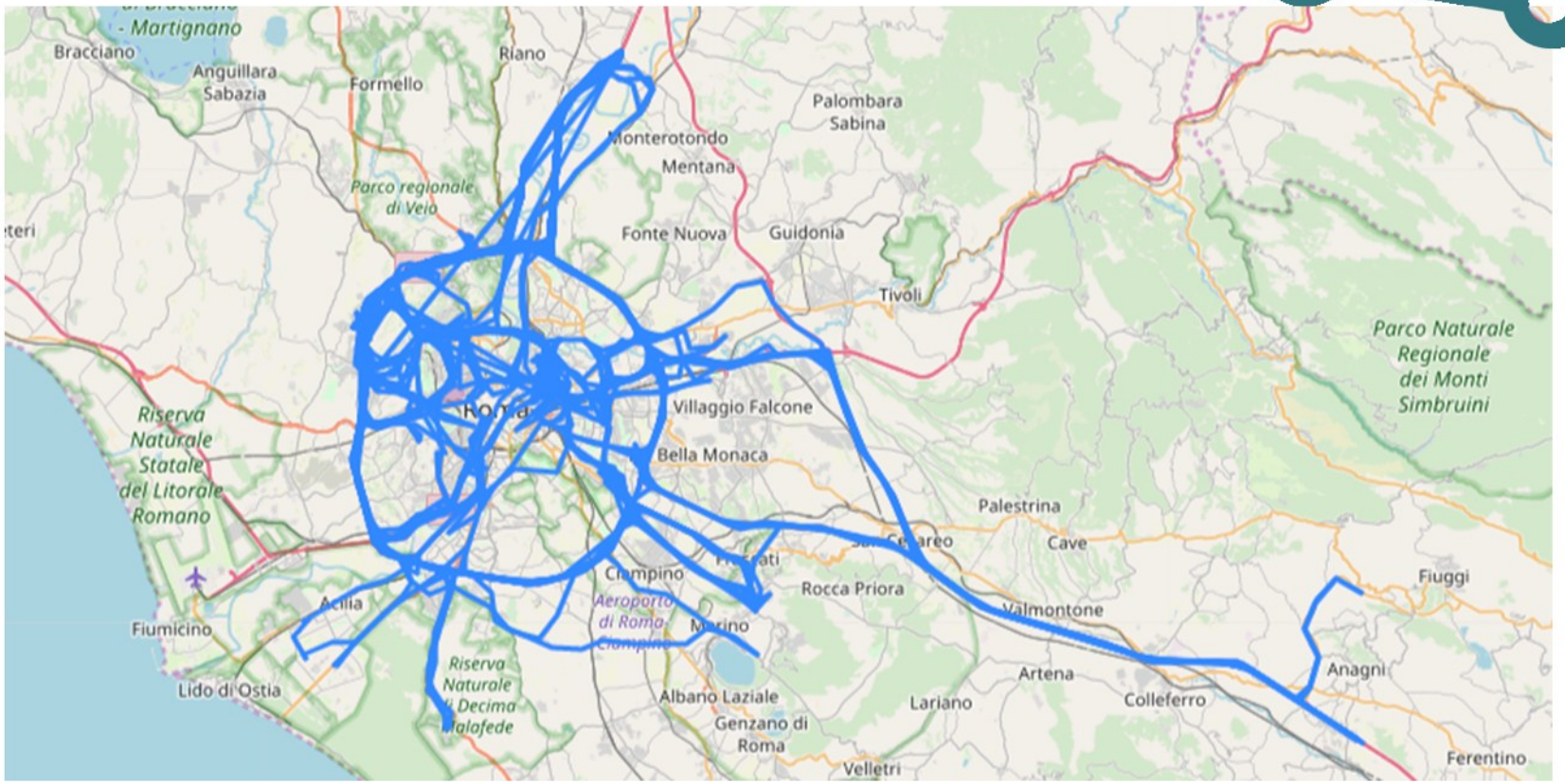
Lengths

Durations

Time of day

Avg speed

…



Events Features

Accelerations

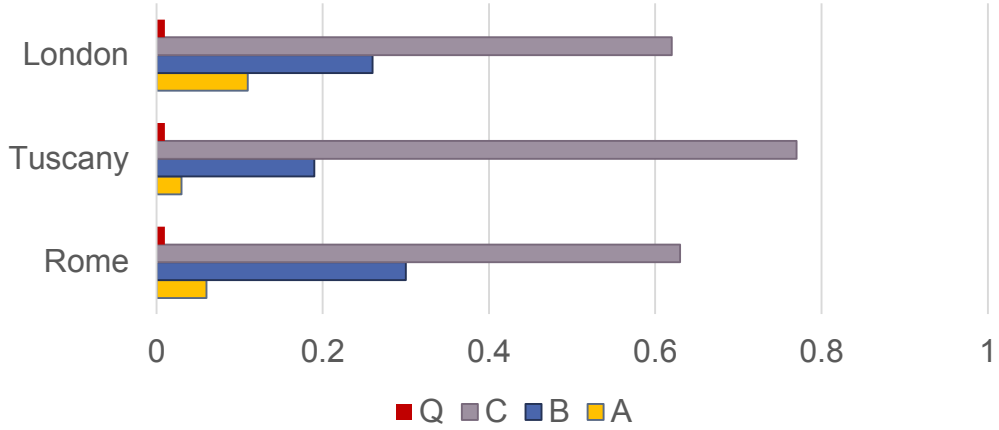
Breaks

Cornering

Zig-zag

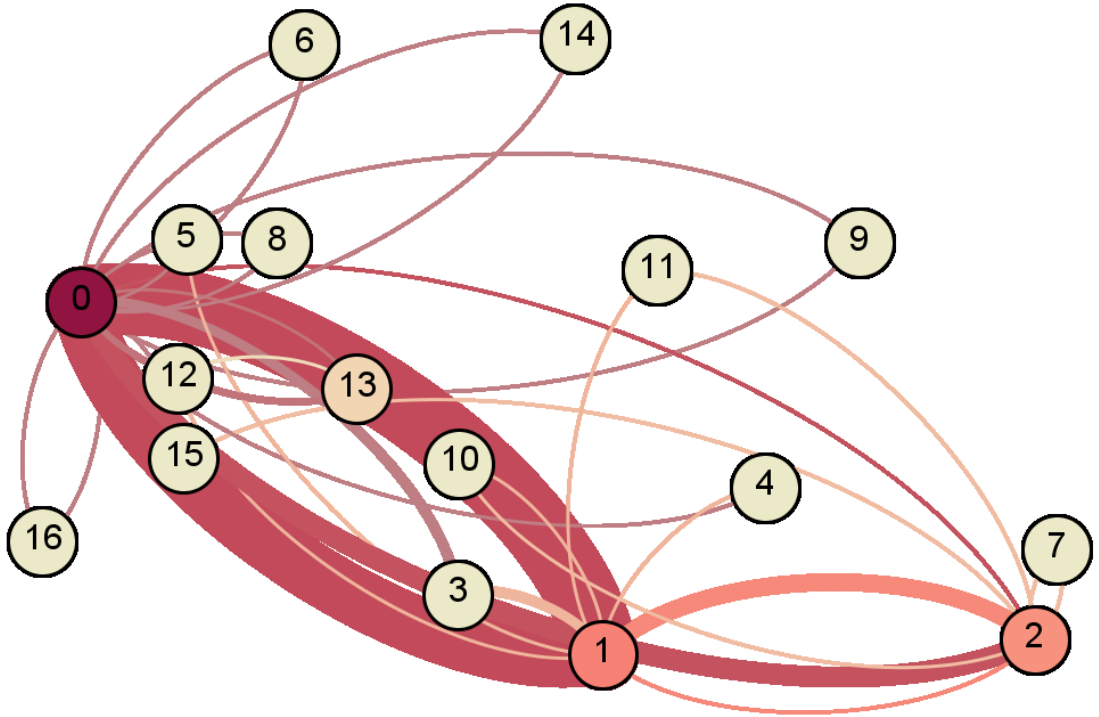
Mini-stops

- Short engine switch-offs



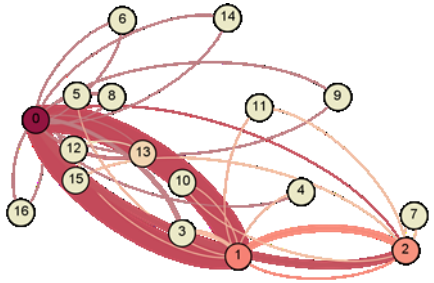
Individual Mobility Network

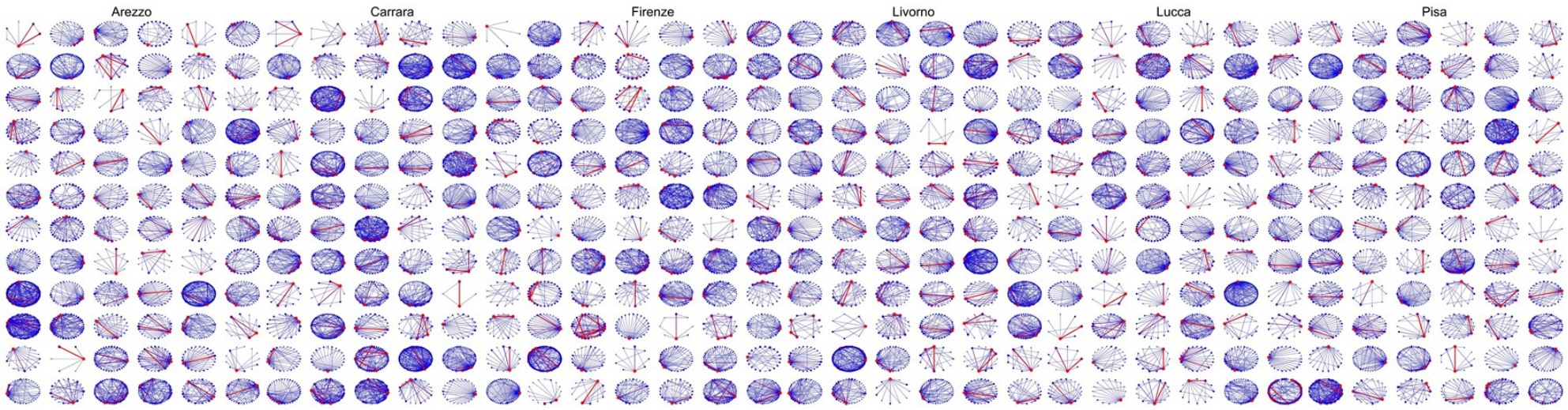
Graph abstraction of the overall mobility of the individual based on locations (nodes) and movements (edges).



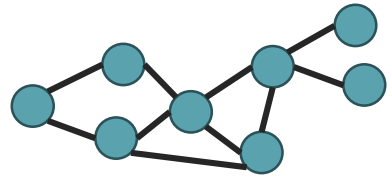
Individual Mobility Network

Graph abstraction of the overall mobility of the individual based on locations (nodes) and movements (edges).





IMN Features: network measures



Nodes

Edges

Node frequency

Edge frequency

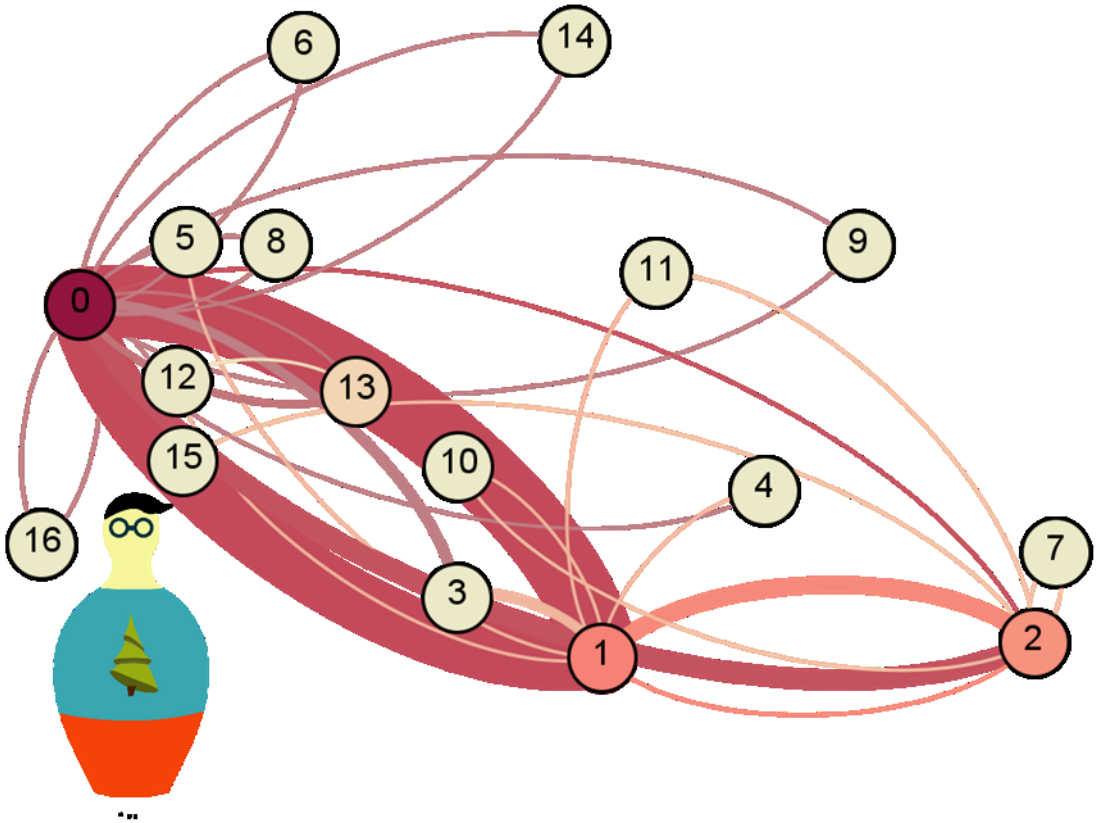
Indegree

Outdegree

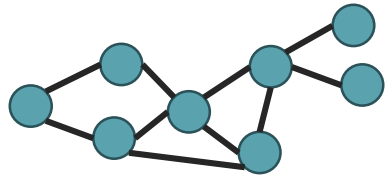
Clustering Coefficient

Entropy

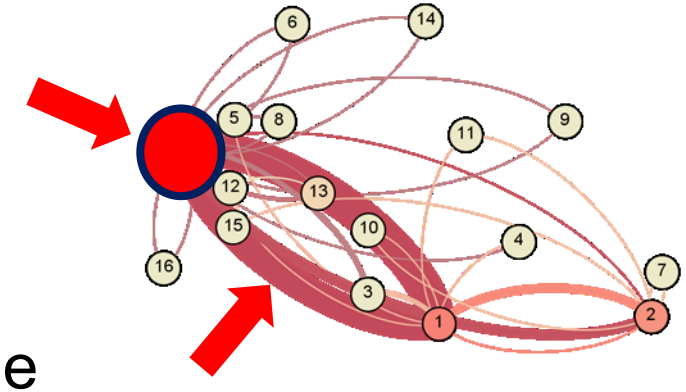
…



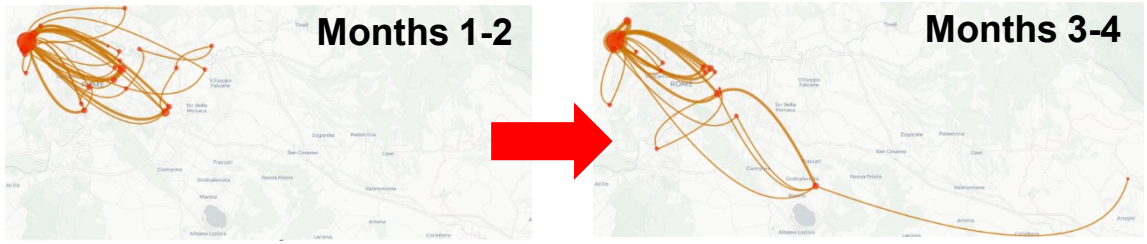
IMN Features: ON GOING WORK



- Compute features focused on parts of IMN
  - on frequent locations (in-flow)
  - along frequent movements



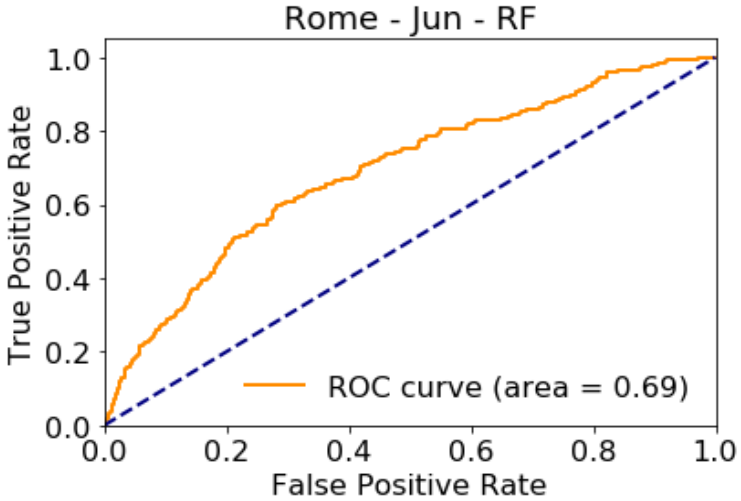
- Measure how much the IMN changes in time
  - Might reflect changes in habits or mobility needs



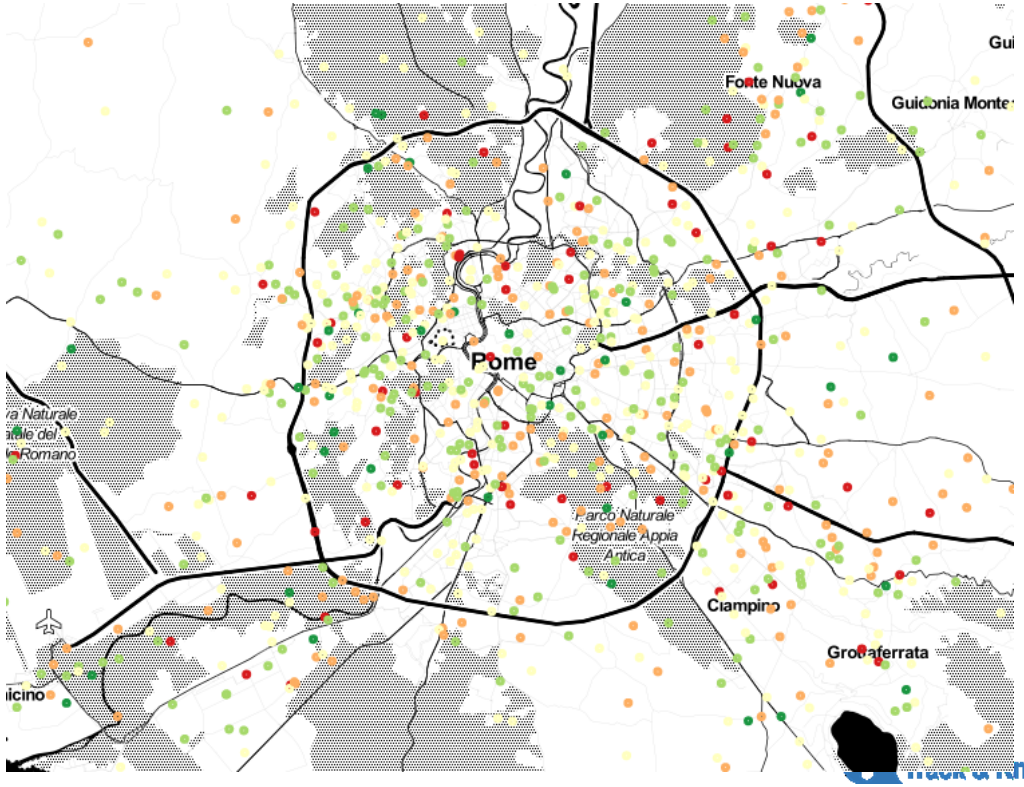
- Characterize what kind of areas I cross/stay in
  - Measure through mobility information on the areas aggregated over all the vehicles (e.g. traffic density)

Preliminary performance tests

- Various tree-based classification algorithms tested
- Promising yet to improve results
  - Moderately good precision (up to 48%)
  - Low recall (up to 6%)
  - Low F1-score (up to 0.098)
  - Moderately good AUC(ROC)



Spatial distribution of prediction (period = June)



Next steps on crash prediction

- Features & model refinements to improve performances, especially recall
  - Work on developing richer IMNs
  - Integrate information from other sources (e.g. weather and POI enrichment)
- Test model geographical transferability at various scales
  - Problem: how to adapt a model built in area A in order to work well in area B?
- From predictions to prescriptive rules
  - Explainable AI !

THANK YOU!

Mirco Nanni
mirco.nanni@isti.cnr.it
ISTI-CNR, Italy