City Indicators for Mobility Data Mining

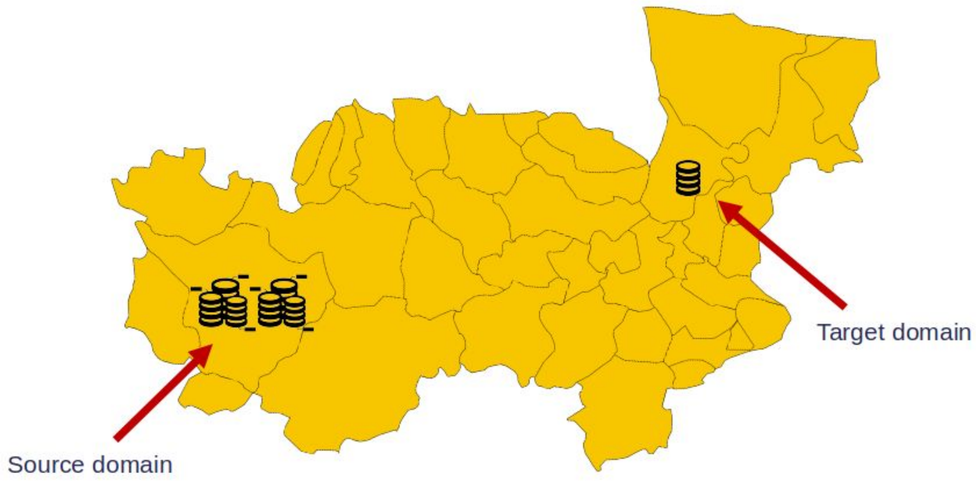
Mirco Nanni, Agnese Bonavita, Riccardo Guidotti

Introduction & Motivation

- **Characterization of a geographical area** through a set of quantitative measures is one of the most common tasks in mobility data analysis
- Difference display between cities, municipalities and other geographical units
- Application task: **Geographical Transfer Learning**
  - What if we want to predict the impact of an event on the urban mobility without having historical data on it?
  - Can knowledge be transferred from any city or are there some constraints?

Transfer Learning

Given some observations in a set of source geographical areas for some mobility-related tasks, and some observations about a target geographical area for some task, exploit the knowledge from the source domains to enable or improve models on the target domain.



Methodology

- Define a notion of **similarity** between areas
  - Based on several properties (geography, mobility, road network,etc)

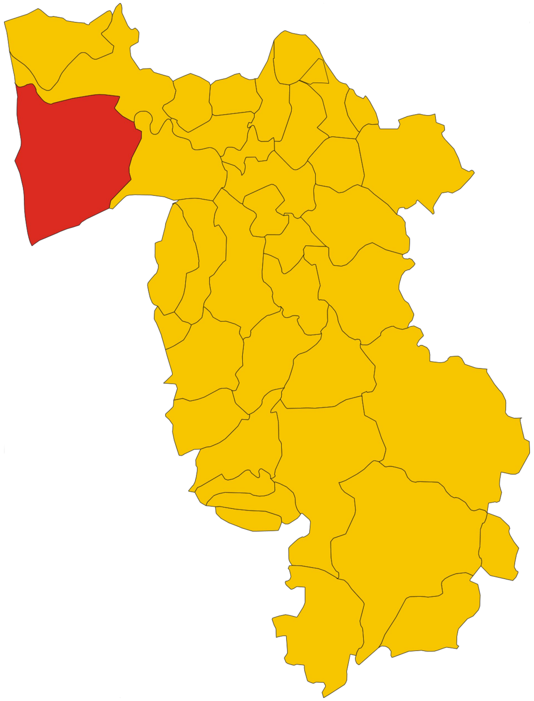
- Study the model **transferability** in a simple case
  - Task: Local, short term traffic prediction
  - Objective: test relation between transferability and similarity

Extracting of mobility related descriptive features of areas

- Objective: describe areas through multi-dimensional views

- Local features categories:
  - Spatial concentration
  - Intra-city Flows
  - Individual Mobility
  - Roads and traffic

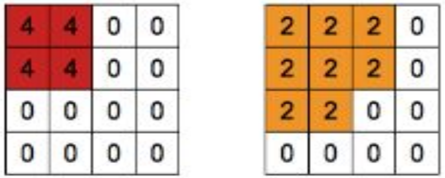
- Global features categories:
  - City Networks
  - Ego-Networks



Spatial Concentration

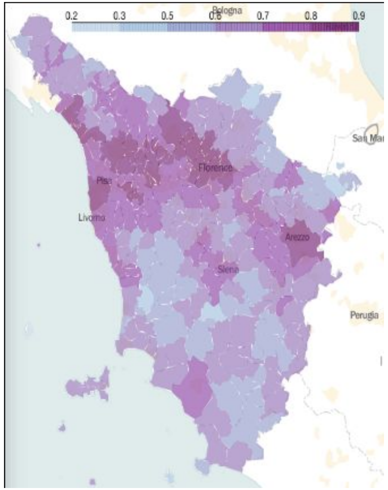
- Based on distribution of presence
  - Inferred from GPS traces
  - Based on a grid over the area

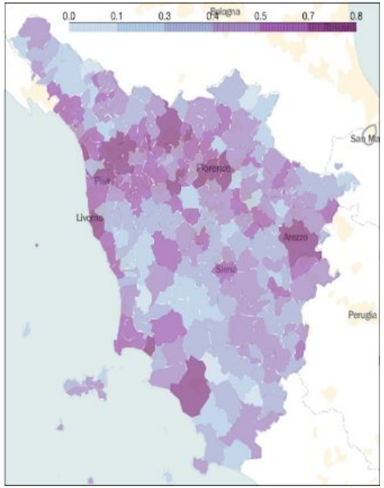
- Three values:
  - Entropy
  - Moran's I (spatial autocorrelation)
  - Average nearest neighbor distance



Entopy = 1.38
Norm. Entr. = 0.5
Moran's I = 0.61

Entopy = 2.08
Norm. Entr. = 0.75
Moran's I = 0.52





Flows in a Grid Network

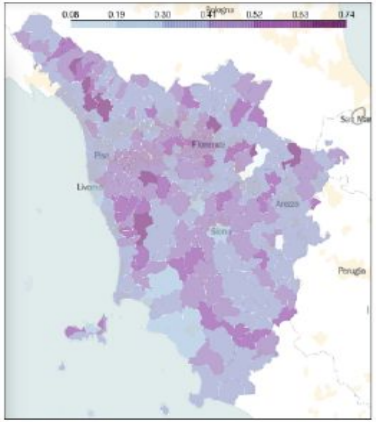
- Based on the Origin-destination flow matrix

- Network measures:
  - Average Node degree
  - Modularity of network communities (Louvain algorithm)

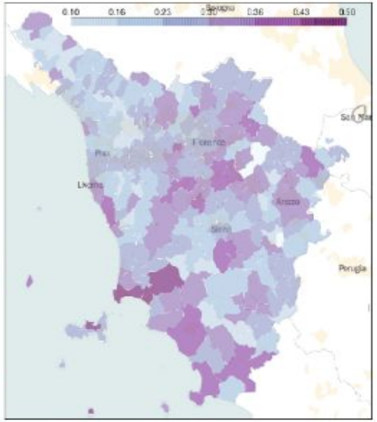
- Fitting with Physical mobility models:
  - Gravitation Model
  - Radiation Model

$$G_{ij} = A\frac{m_i^{\alpha} n_j^{\beta}}{r^{\gamma}}$$

$$T_{ij} = \frac{T_i}{1-\frac{m_i}{M}} \frac{m_i\, n_j}{(m_i+s_{ij})(m_i+n_j+s_{ij})}$$



Modularity (after Louvain partitioning)

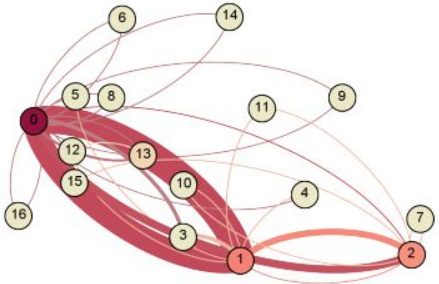


R2 fitting for Gravitational Model

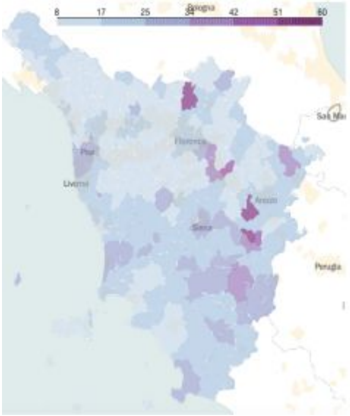
Individual Mobility

- Based on Individual Mobility Networks
  - Considers IMNs of users mostly moving in the area

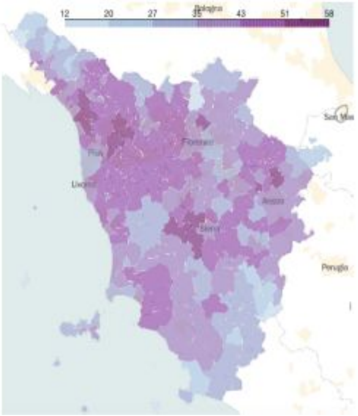
- Various mobility & network measures:
  - Average size of IMNs
  - Average km driven by users
  - Time-uncorrelated entropy of in-degree
  - Radius of gyration
  - Percentage of regular trips
  - Modularity of network (Louvain)



Individual Mobility Networks



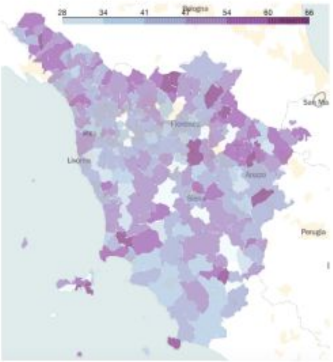
IMN Radius of gyration



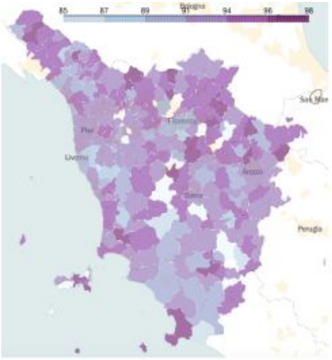
IMN number of nodes

Road and Traffic Networks

- Consider both the road network and how the traffic distributes on it
- Basic statistics on the road network
  - Density of road edges, intersections, intersection degree, total and average length
  - Network centrality of roads
- Traffic distribution
  - Concentration of traffic over the hottest network edges



Traffic % in top 10% roads



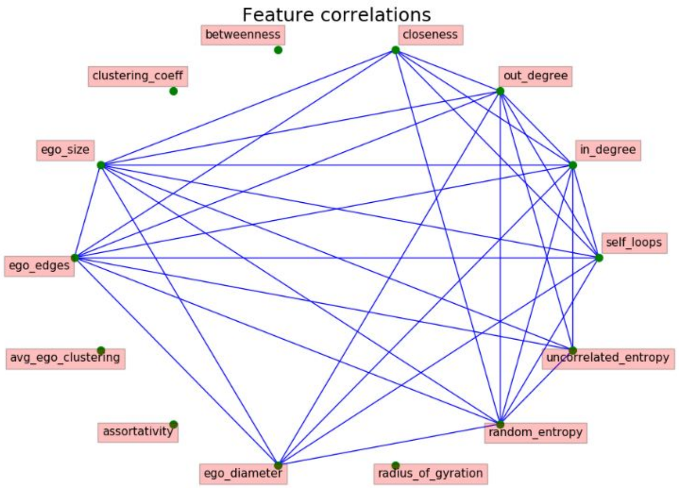
Traffic % in top 50% roads

Complete Network of Cities

- Network where each city is a node and edges are drawn based on trajectories between them

- Origin-Destination Matrix for the trips between two cities
- Extraction of attributes for the clustering step
  - self-loops
  - In/Out degrees
  - Closeness

- Ego-Network for each city
  - # of nodes
  - # of edges
  - average clustering coefficient

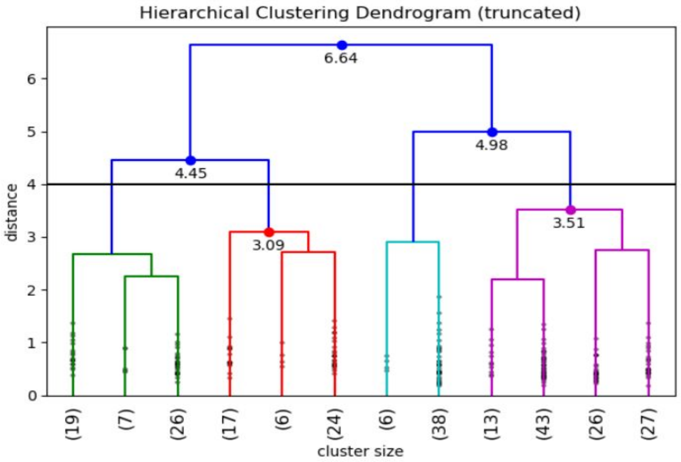


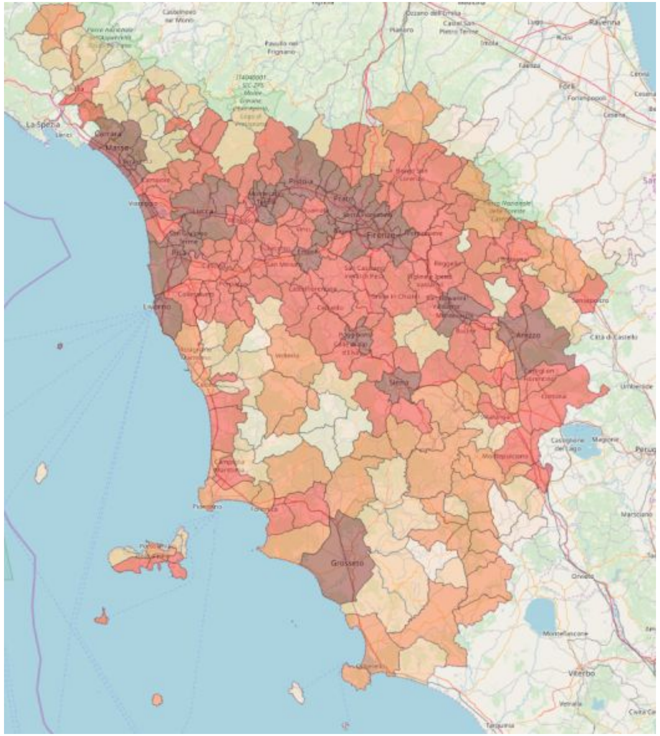
Case Study

- Dataset of GPS traces from private vehicles
- 18.9 million trajectories of 250 cars in Tuscany region
- Tuscany divided in 276 municipalities
- Areas of 10 x 10 km for each municipality

City Clustering

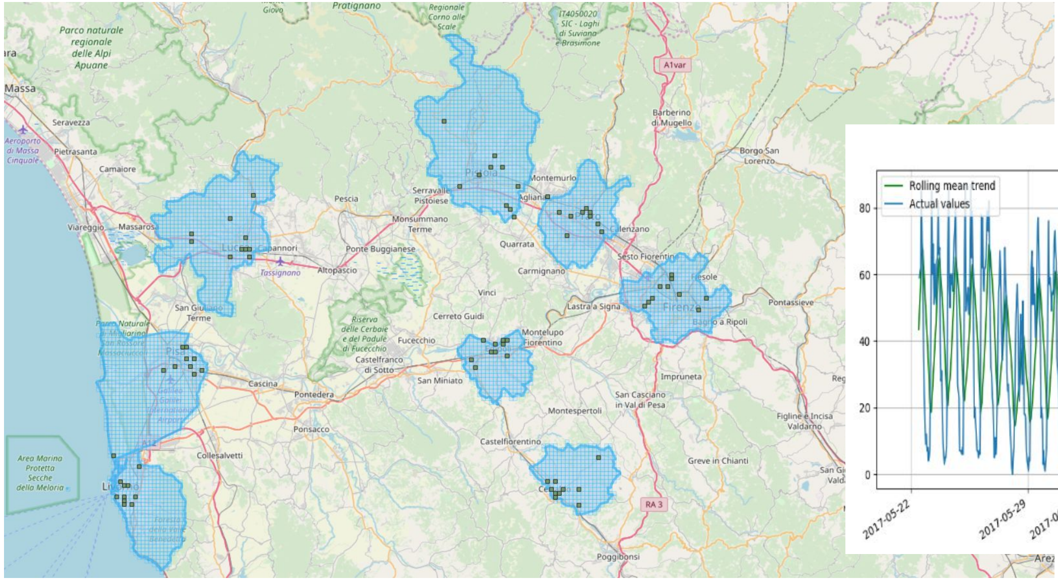
- Hierarchical cluster
- 5 different clusters

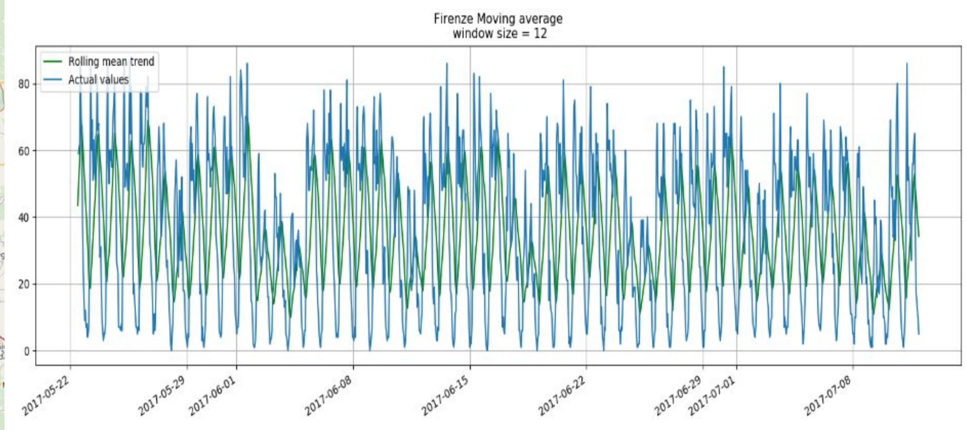




Similarity vs Transferability: Prediction Task

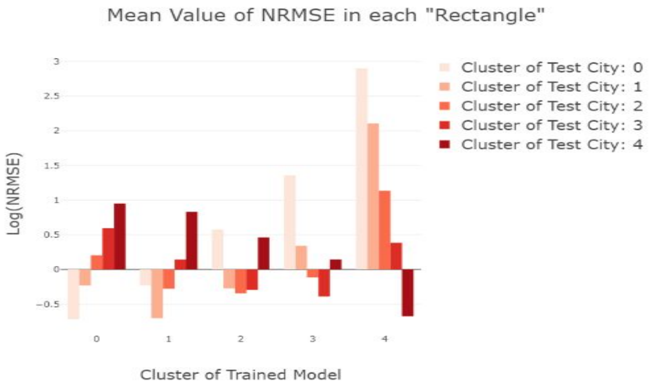
- Predict traffic volume in key portion of each city
- Prediction model in each city: XGBoost regression

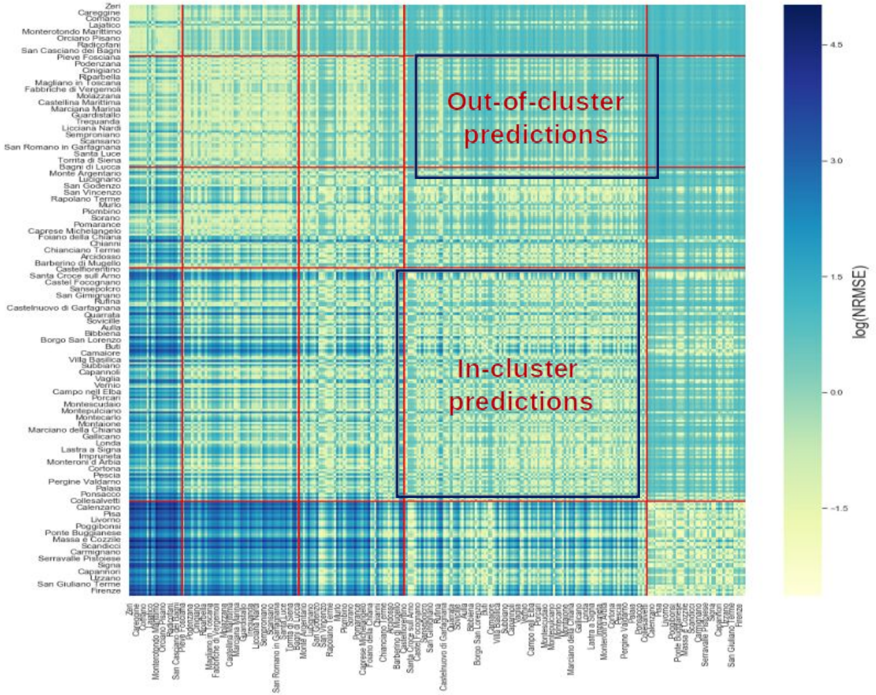




Testing Model transferability

- We want to test if similarities based on the city indicators are useful to identify areas with a better transferability
- The transfer is better between cities of the same cluster
- Matrix is not simmetric



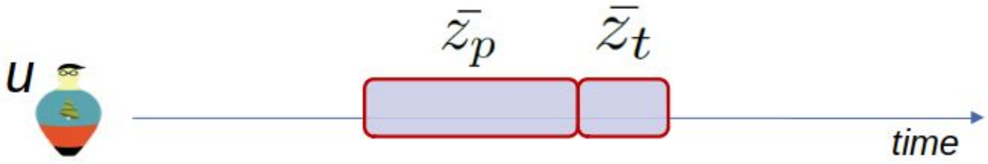


Conclusions

- Definition of a set of local and global city indicators
- Test on a real case study
- Transfer Learning goal
- Results show that models trained on a municipality performs better when transfered on other municipalities belonging to the same cluster

Future and Ongoing Works: Crash Prediction task

- Given the historical mobility data ***H*** of a user in a time period z_p



- Predict his crash probability in a future period z_t

$$p_{crash}(u) = P(u \ has \ crash \ in \ \bar{z}_t \mid H_u^{\bar{z}_p})$$