Data Mining II

Mobility Data Mining

F. Giannotti& M. Nanni KDD Lab – ISTI – CNR
Pisa, Italy

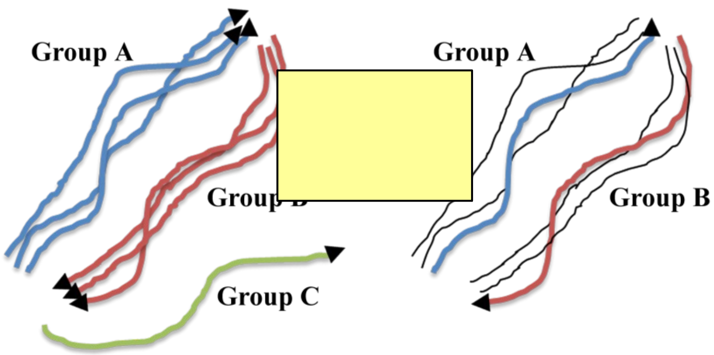
Outline Mobility Data Mining

- Introduction
- MDM methods
  - Clustering
  - Trajectory Pattern Mining
  - Prediction
- MDM methods at work. Understanding Human Mobility
  - Dimensions of mobility analytics
  - Models of human mobility
  - The Mobility Atlas

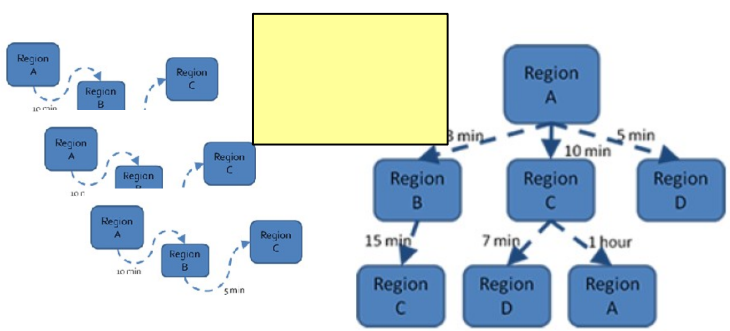
- Module 3 Case studies
  - OD Matrix, D4D, Sociometer,
  - Network& Mobility

Derived patterns and models

- Combination & refinement of basic patterns and models



- Individual Mobility Profile: routines consistently followed by a single moving object



- T-PTree: predictive tree built by combining T-Patterns

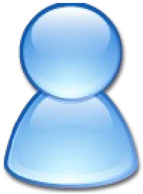
User's Mobility Profile

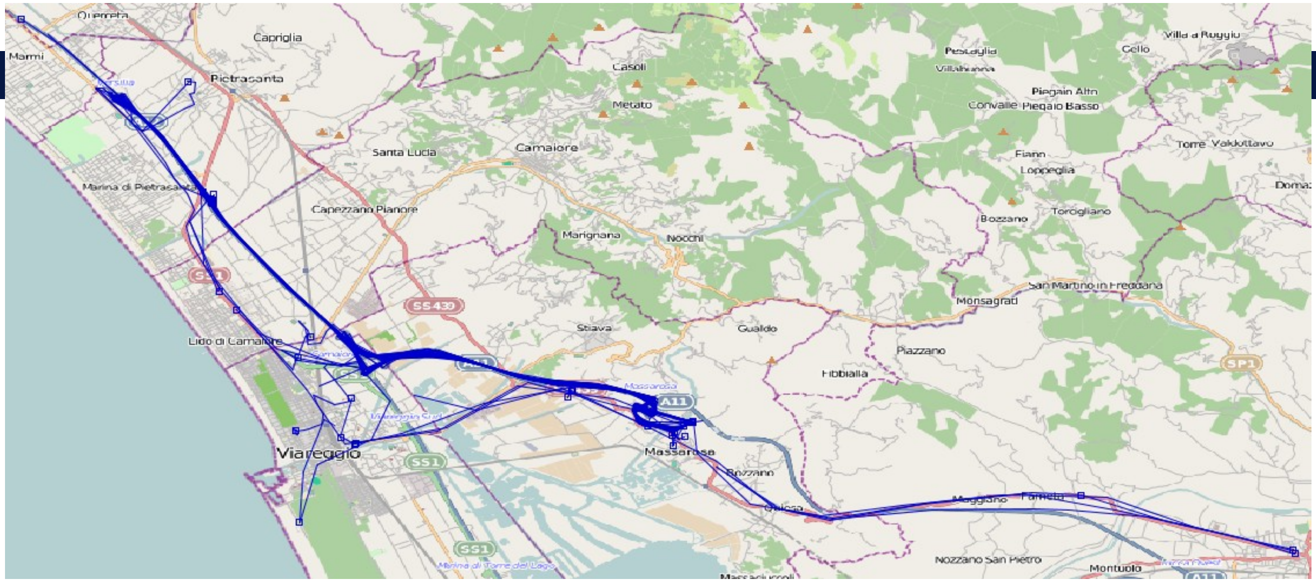
Given the user history as an ordered sequence of spatio-temporal points, we want to extract a set of *routines* in order to create the his\her *mobility profile.*

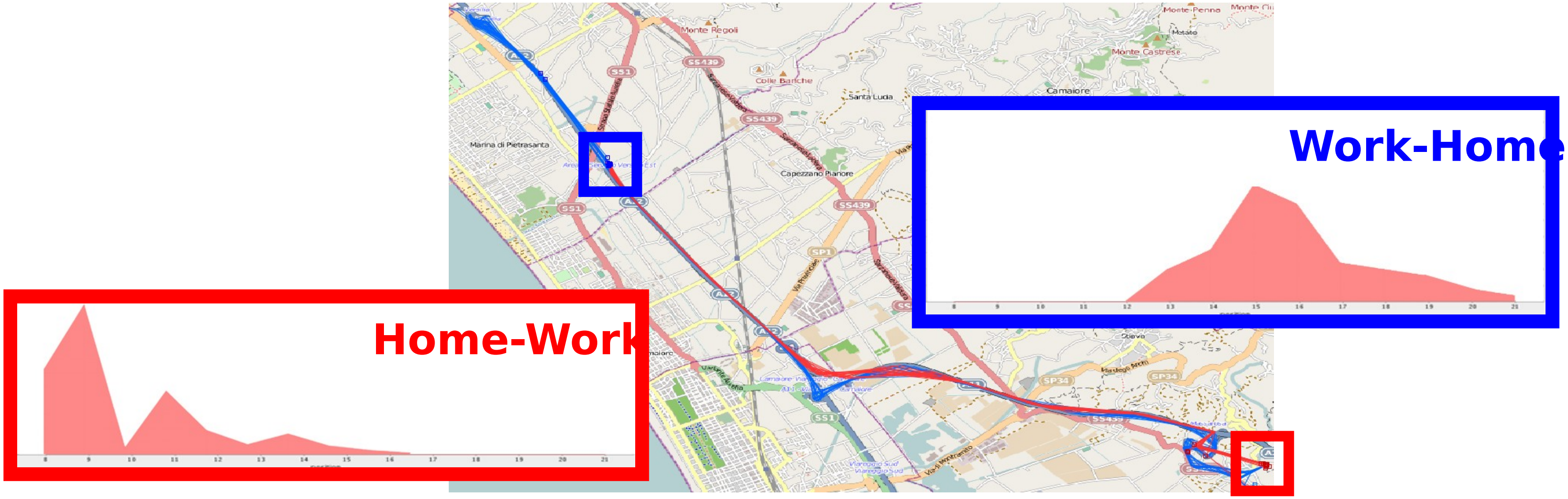
Where:

- A *Routine* is a typical local behavior of the user.
- A *Mobility profile* is the set of user's routines

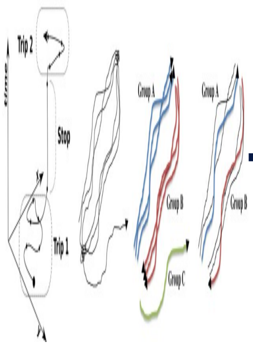
Discovering individual systematic movements





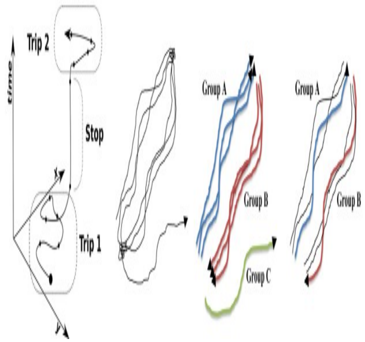


Derived patterns and models: mobility profiles



User history

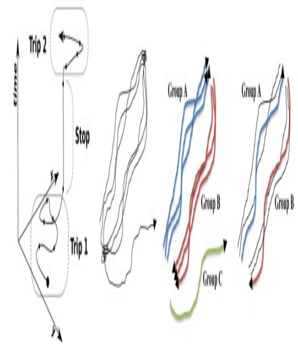
An ordered sequence of spatio-temporal points.



Trips construction

Cutting the user history when a **stop** is detected

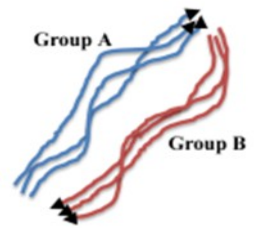
Stops Spatial Threshold
Stops Temporal Threshold



Grouping

Performing a density based **clustering** equipped with a **spatio temporal** distance function

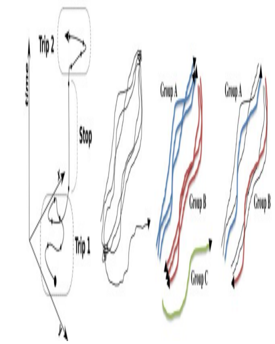
Spatial Tollerance
Temporal Tollerance
Spatio temporal distance



Pruning

Groups with a small Number of trips are Pruned

Support Threshold



Profile extraction

The **medoid** of each group becomes user's **routines** and the all set become the user's **mobility profile**

Trasarti, Pinelli, Nanni, Giannotti. **Mining mobility user profiles for car pooling**. ACM SIGKDD 2011

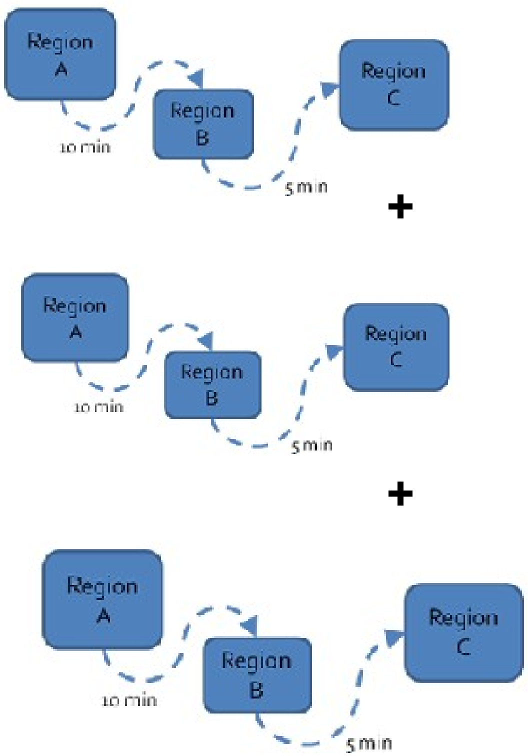
Derived patterns and models: T-Prediction Tree

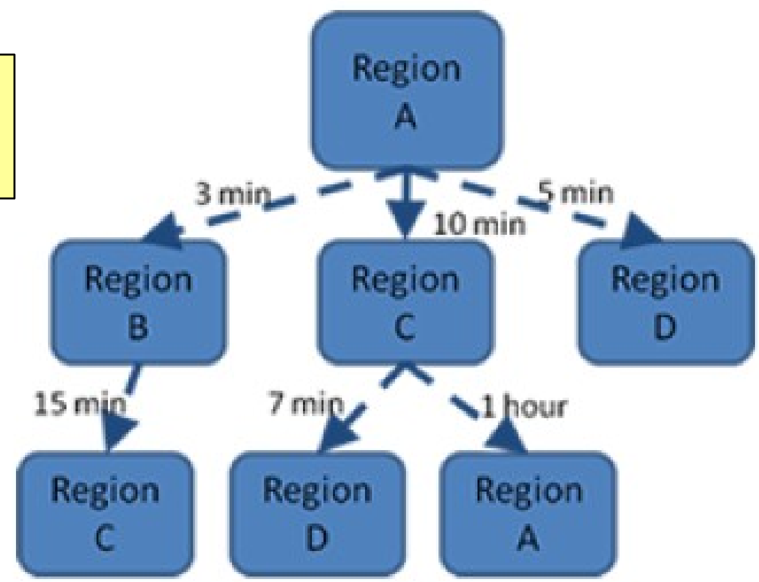
- Rule-based prediction model
- Each T-Pattern is

e”

ation /

f a set





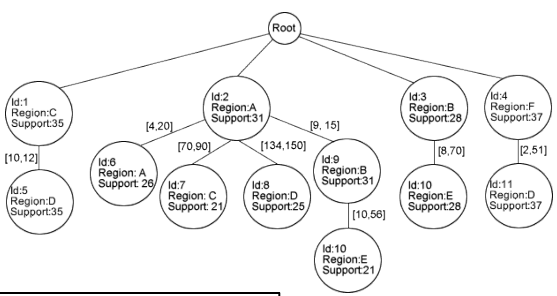
Monreale, Pinelli, Trasarti, Giannotti.
Where Next: a predictor on Trajectory pattern mining. Proc. ACM SIGKDD 2009

Basic Idea: People move as the crowd moves

How to realize this idea:

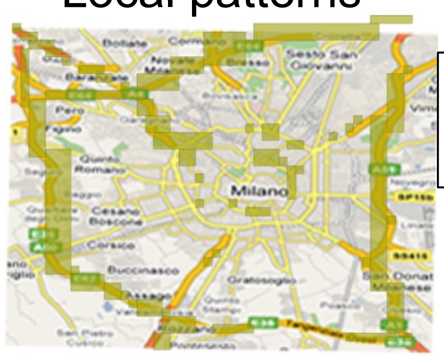
- Extract patterns from **all the available movements** in a certain area instead of on the individual history of an object;
- Using these **Local movement patterns** as predictive rules.
- Build a prediction tree as global model.

Prediction Tree



Trajectories dataset

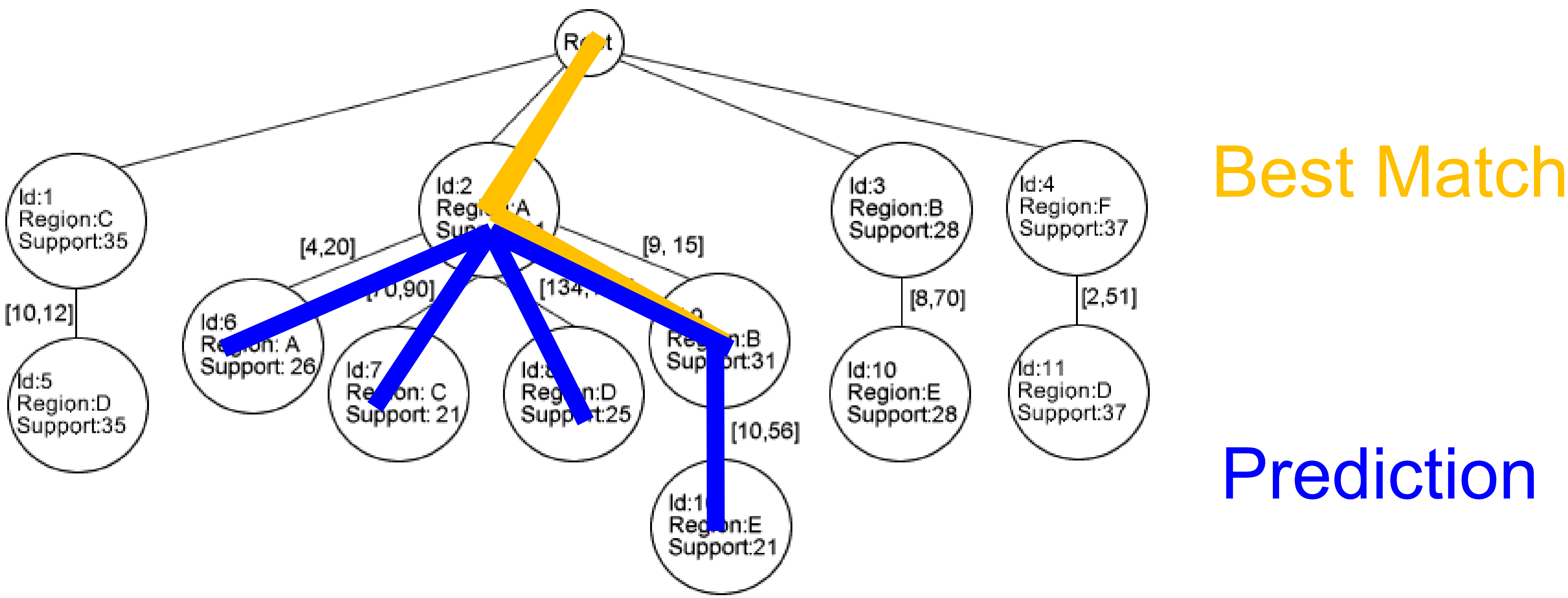
Local patterns



Predict by means of T-Pattern tree

Given a new trajectory:

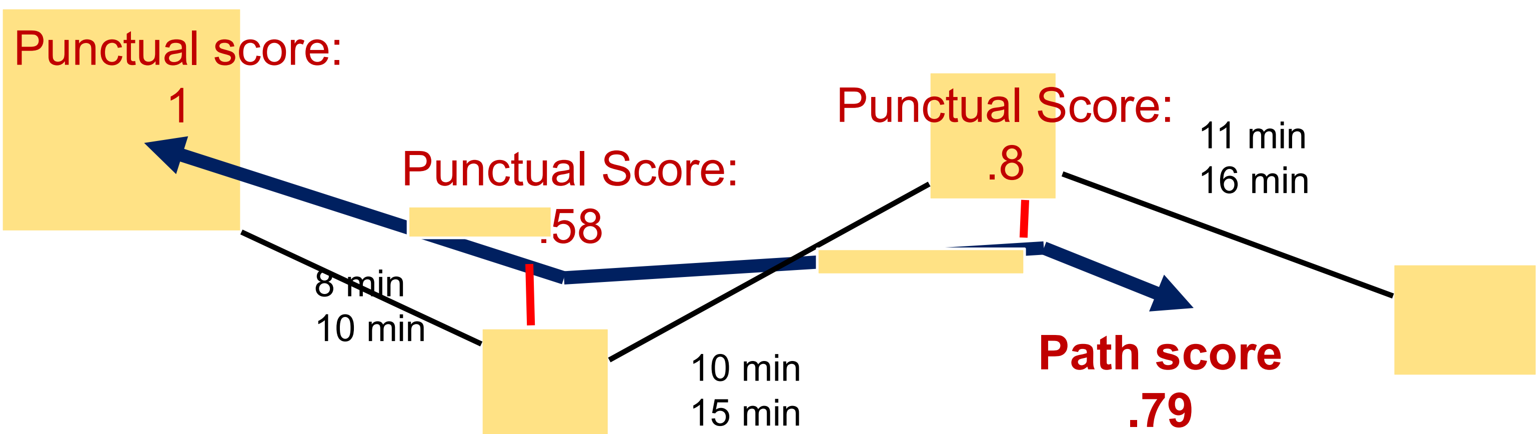
1. Search for best match
2. Candidate generation
3. Make predictions



How to compute the Best Match?

Computing the path score

The path score is the aggregation of all punctual scores along a path.

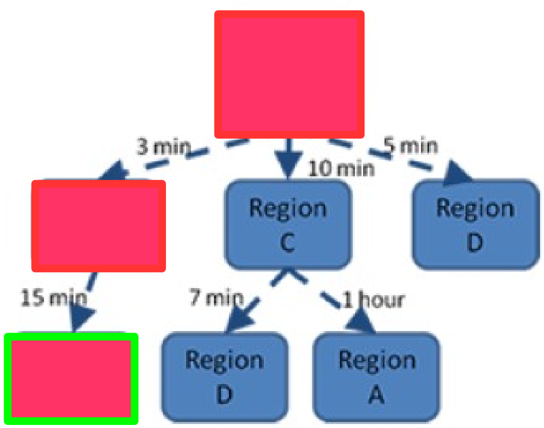


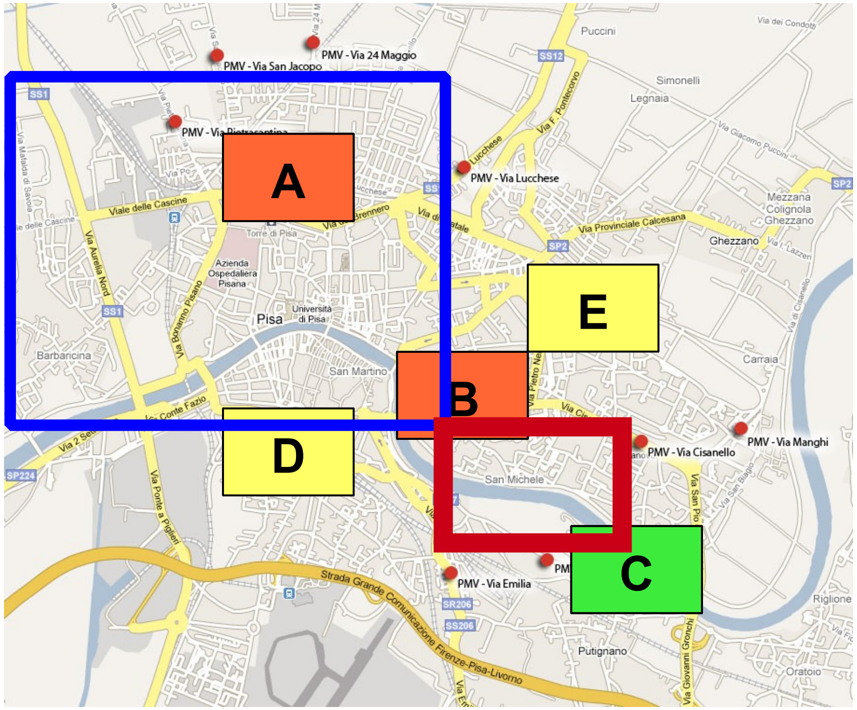
The **Best Match** is the path having:

- ✓ the maximum path score;
- ✓ at least one admissible prediction.

Derived patterns and models: T-PTree

- Example: Compare actual trajectory against the T-PTree
  - Spatial and temporal similarity used to choose best "rule"



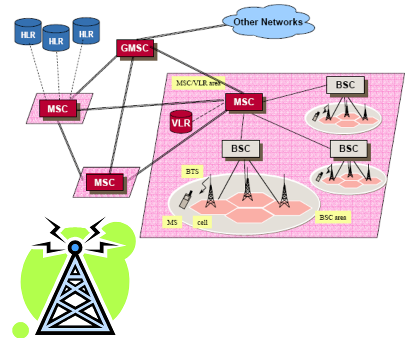


M-Atlas system

Download from: http://m-atlas.eu

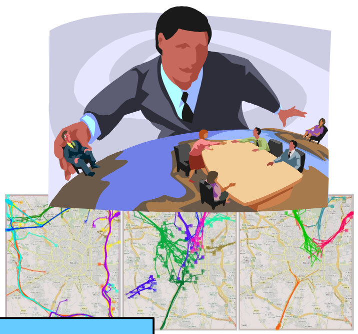
The (GeoP)KDD process

Mobile phone data, GPS tracks



[redacted]

End user

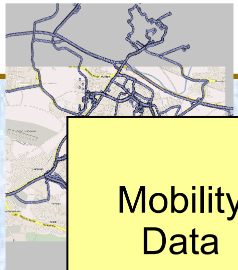


manager

[redacted]

[redacted]

Mobility Data Mining



[redacted]

```
name|date|y|x
Prinzessin|08.20.1998|52.118|12.087
Prinzessin|08.23.1998|51.019|15.309
Prinzessin|08.26.1998|47.723|22.786
Prinzessin|08.29.1998|43.040|27.119
Prinzessin|08.31.1998|38.715|32.165
Prinzes
Prinzes
Prinzes
Prinzes
Prinzes
Prinzes
Prinzes
Prinzes
Prinzes
Prinzes
Prinzessin|12.10.1998|9.124|35.644
...
```

Raw data

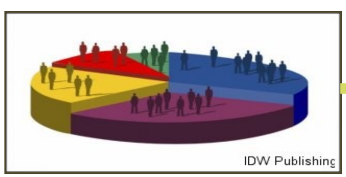
Mobility Data

Privacy and anonymity protection

M-Atlas input

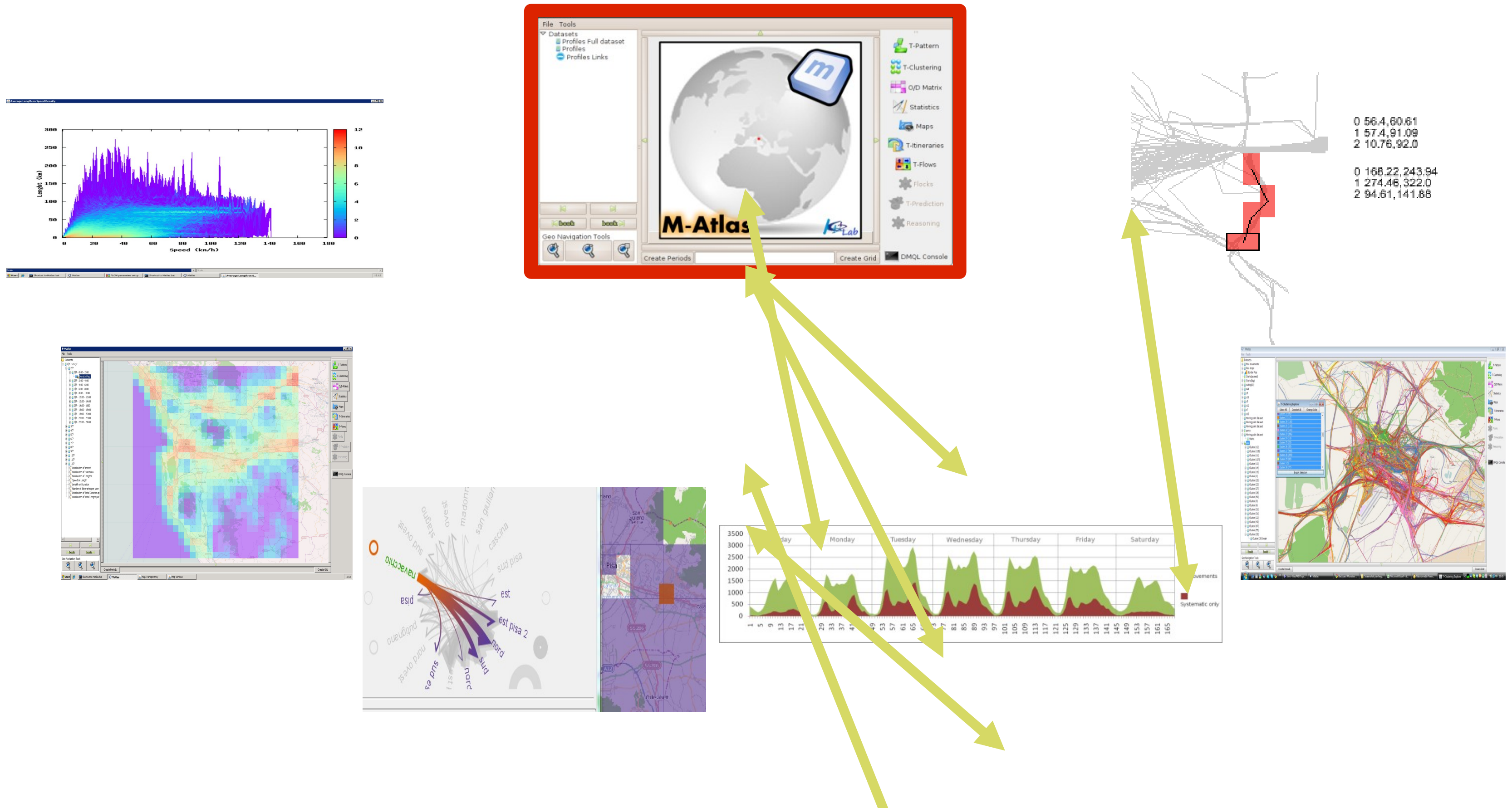
- M-Atlas: An atlas for “urban mobility behaviors”. A framework to query, analyze and navigate the results on mobility data



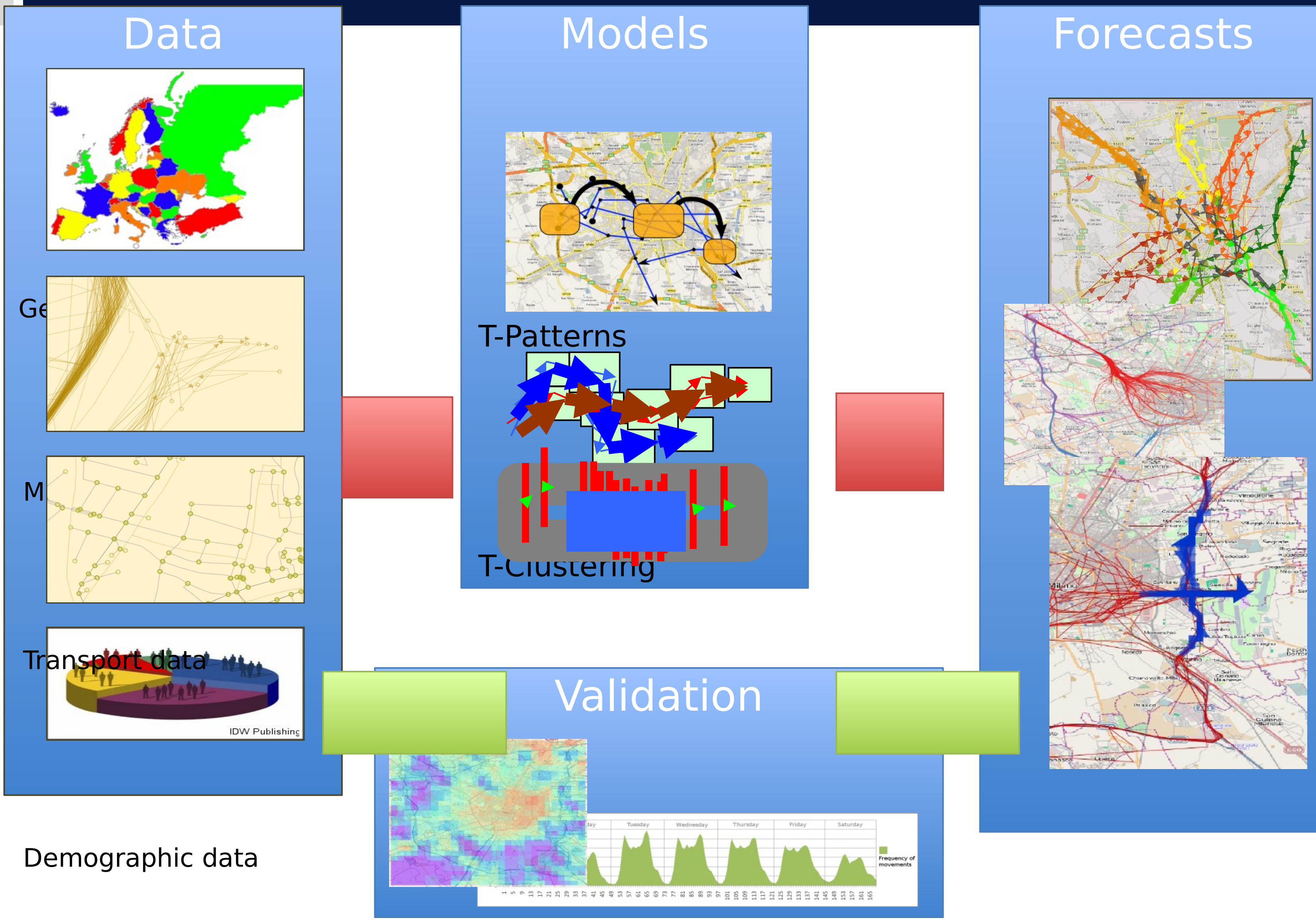



M-Atlas platform

- A tool kit to extract, store, combine different kinds of models to build mobility knowledge discovery processes.



From DATA to KNOWLEDGE



outline

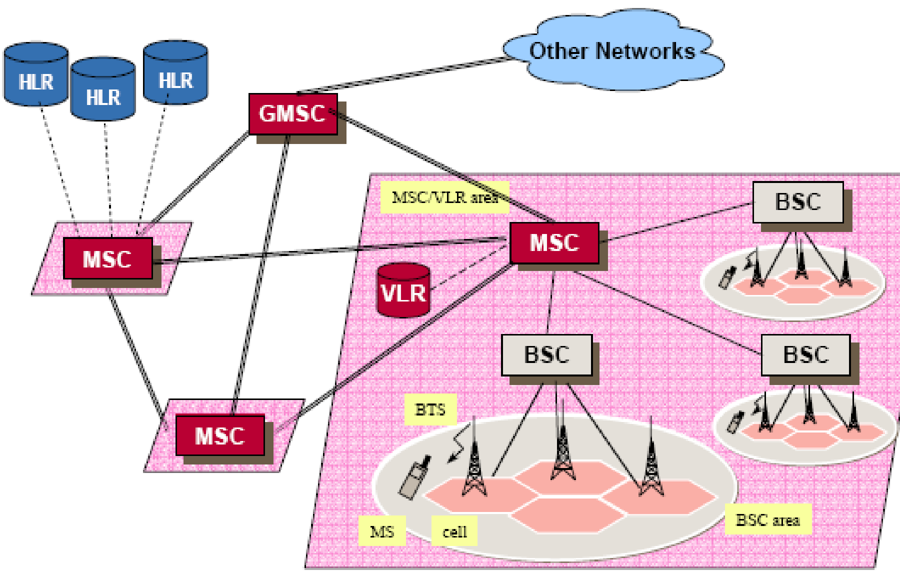
- Introduction
- MDM methods
  - Clustering
  - Trajectory Pattern Mining
  - Prediction
  - Semantic enrichment
- **MDM methods at work. Understanding Human Mobility**
  - Dimensions of mobility analytics
  - Models of human mobility
  - The Mobility Atlas

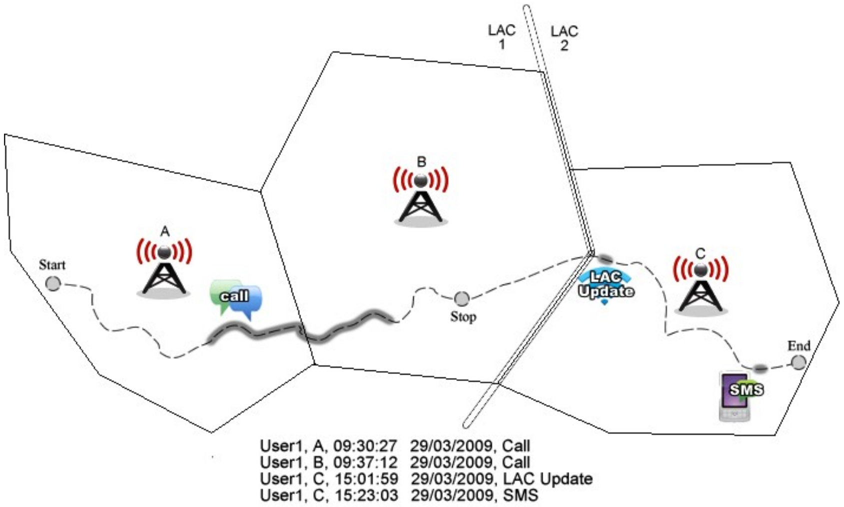
Sensing the movement

Several datasources avaiable

GSM data

- Mobile Cellular Networks handle information about the positioning of mobile terminals
  - CDR Call Data Records: call logs (tower position, time, duration,..)
  - Handover data: time of tower transition
- More sophisticated Network



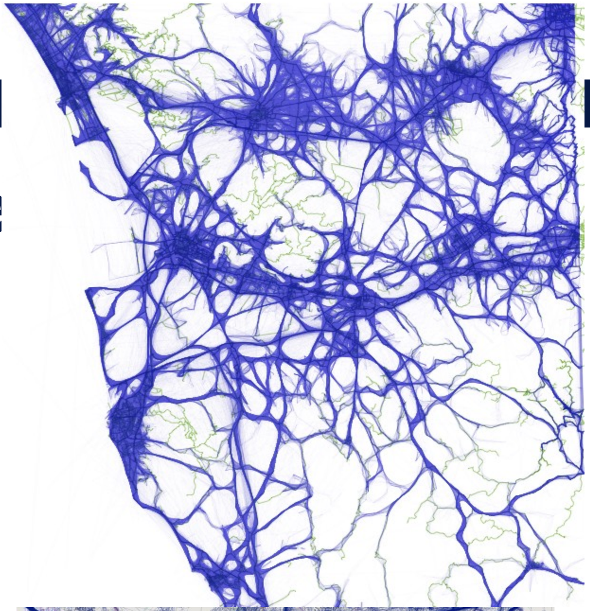


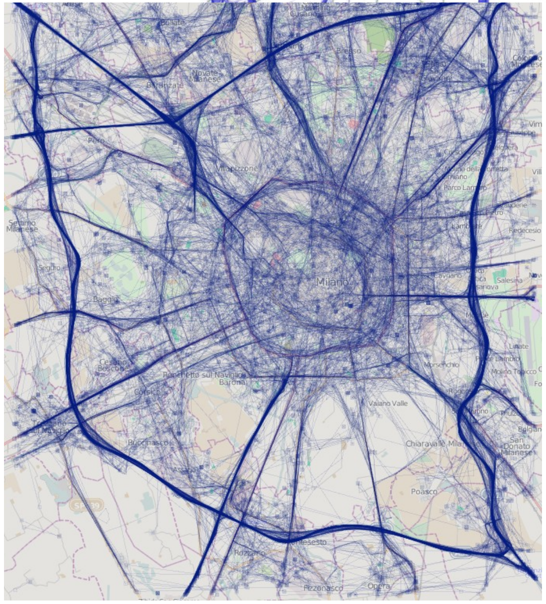
GPS tracks

- Onboard navigation devices se
  to central servers

```
Ide;Time;Lat;Lon;Height;Course;Speed;PDOP;State;NSat
...
8;22/03/07 08:51:52;50.777132;7.205580; 67.6;345.4;21.817;3.8;1808;4
8;22/03/07 08:51:56;50.777352;7.205435; 68.4;35.6;14.223;3.8;1808;4
8;22/03/07 08:51:59;50.777415;7.205543; 68.3;112.7;25.298;3.8;1808;4
8;22/03/07 08:52:03;50.777317;7.205877; 68.8;119.8;32.447;3.8;1808;4
8;22/03/07 08:52:06;50.777185;7.206202; 68.1;124.1;30.058;3.8;1808;4
8;22/03/07 08:52:09;50.777057;7.206522; 67.9;117.7;34.003;3.8;1808;4
8;22/03/07 08:52:12;50.776925;7.206858; 66.9;117.5;37.151;3.8;1808;4
8;22/03/07 08:52:15;50.776813;7.207263; 67.0;99.2;39.188;3.8;1808;4
8;22/03/07 08:52:18;50.776780;7.207745; 68.8;90.6;41.170;3.8;1808;4
8;22/03/07 08:52:21;50.776803;7.208262; 71.1;82.0;35.058;3.8;1808;4
8;22/03/07 08:52:24;50.776832;7.208682; 68.6;117.1;11.371;3.8;1808;4
...
```

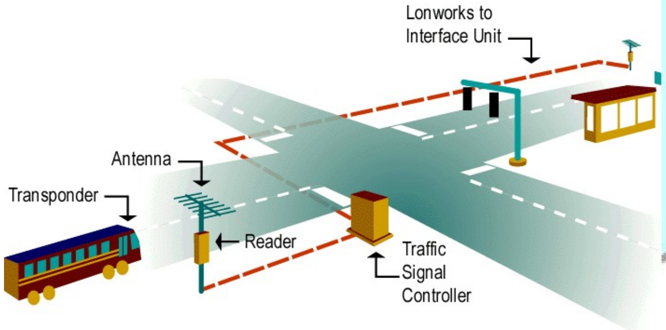
- Sampling rate ~30 secs
- Spatial precision ~ 10 m

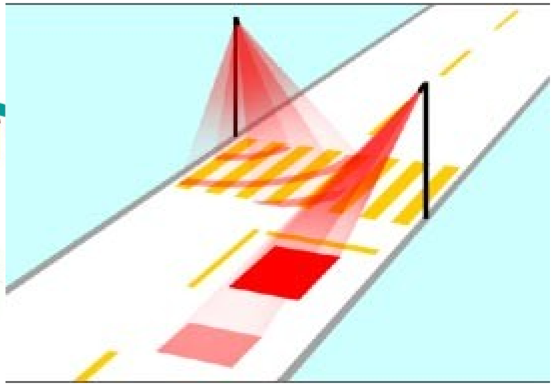


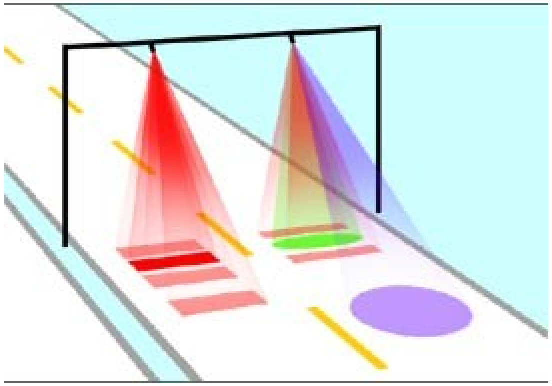


Road side sensors

- Measure the flow of a specific road arc
  - Laser-based sensors
  - Inductive loops
  - Traffic cameras







Other data sources

- Social web services
  - Flickr
  - Foursquare
  - Gowalla
  - Twitter

- *Presence estimation*
  - Hotel statistics
  - Airport departures and arrivals
  - Bus and public transportation
  - Park usage
  - Weather conditions

Dimensions to explore

Space

Dimensions

Individual

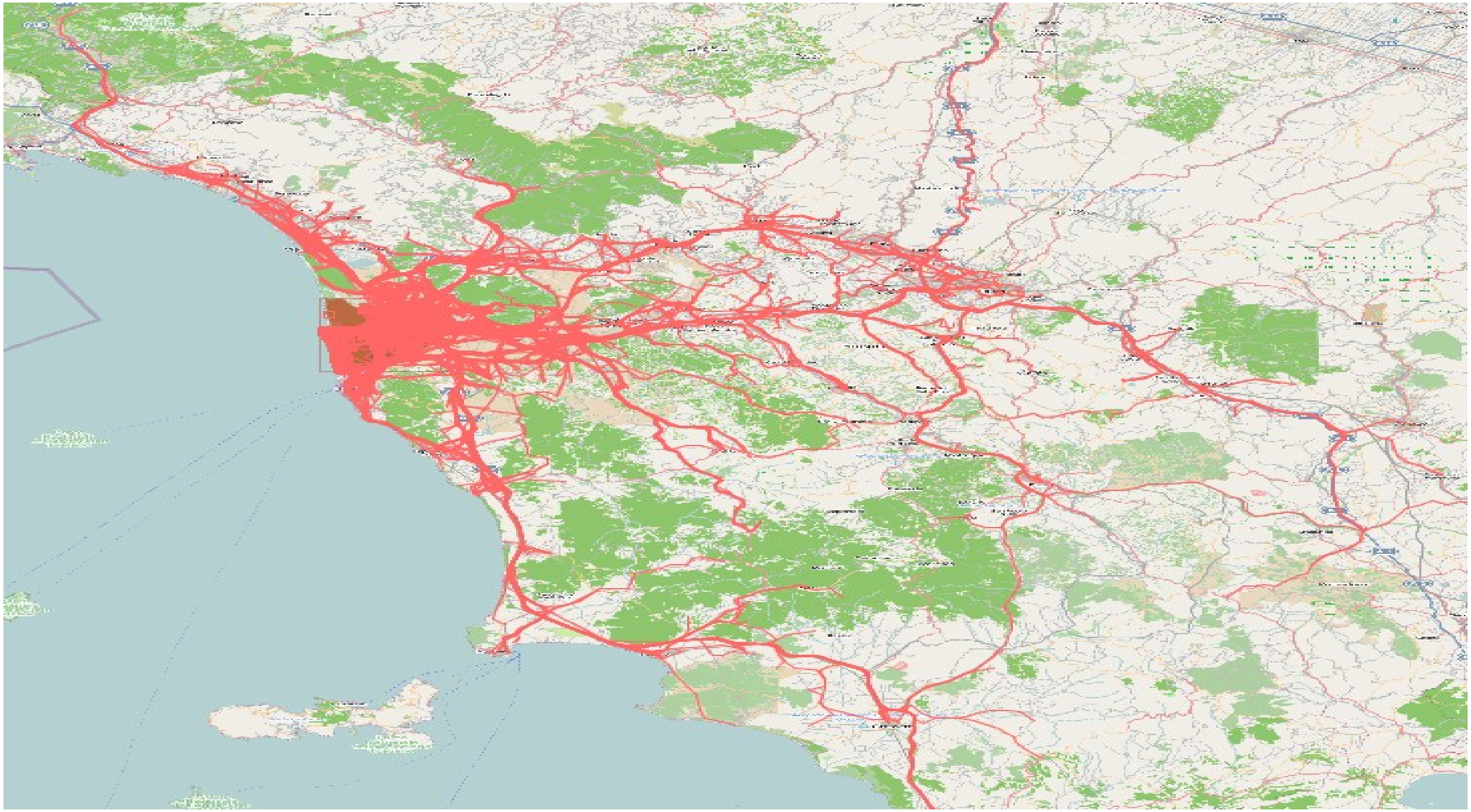
Time

- Space
  - Administrative borders
    - E.g.: city
  - Distance travelled
    - How much a person is travelling

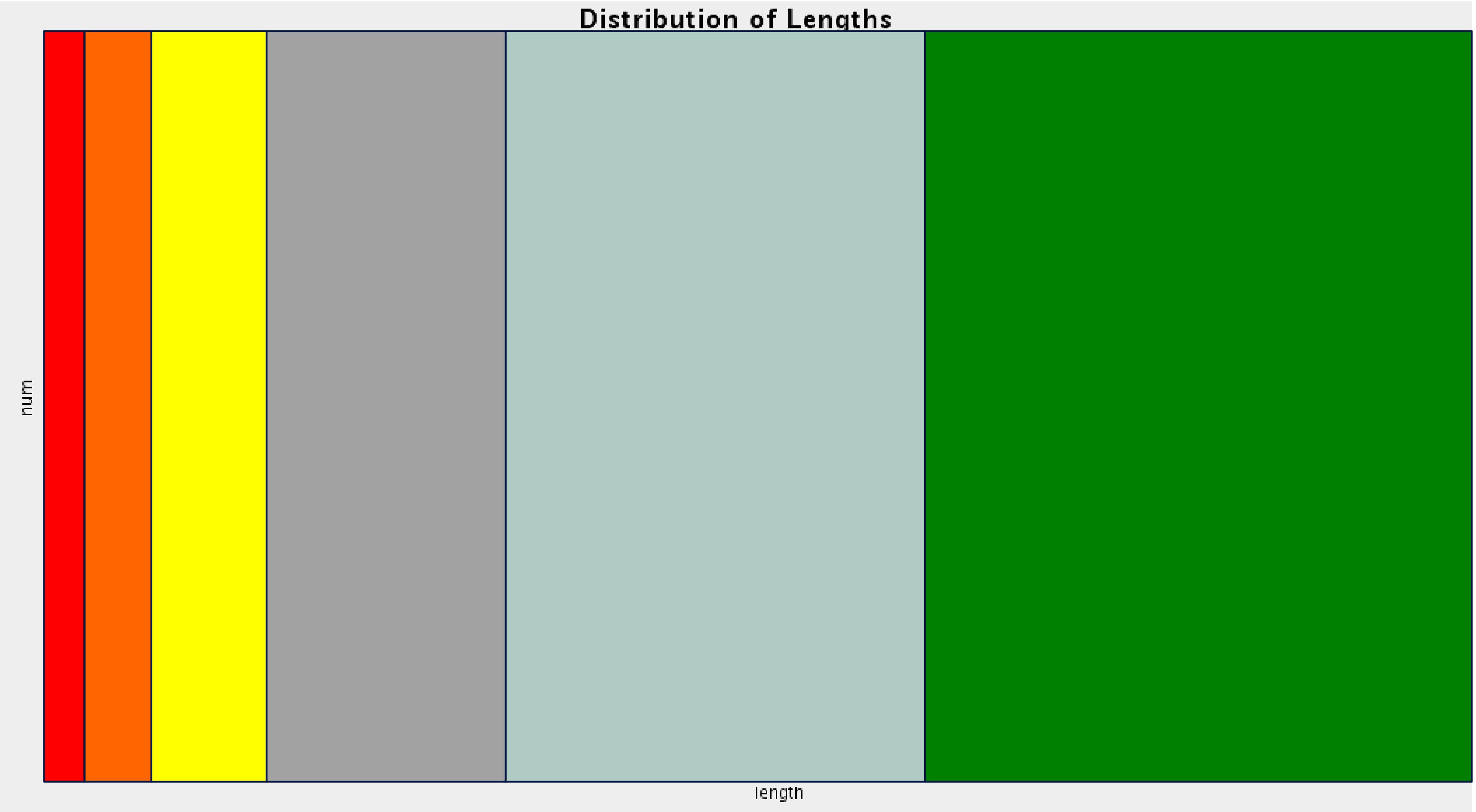
- Individual
  - Preferred locations
  - EigenMobility

- *Time*
  - Hour of day
  - Day of week
  - Weekdays/weekends

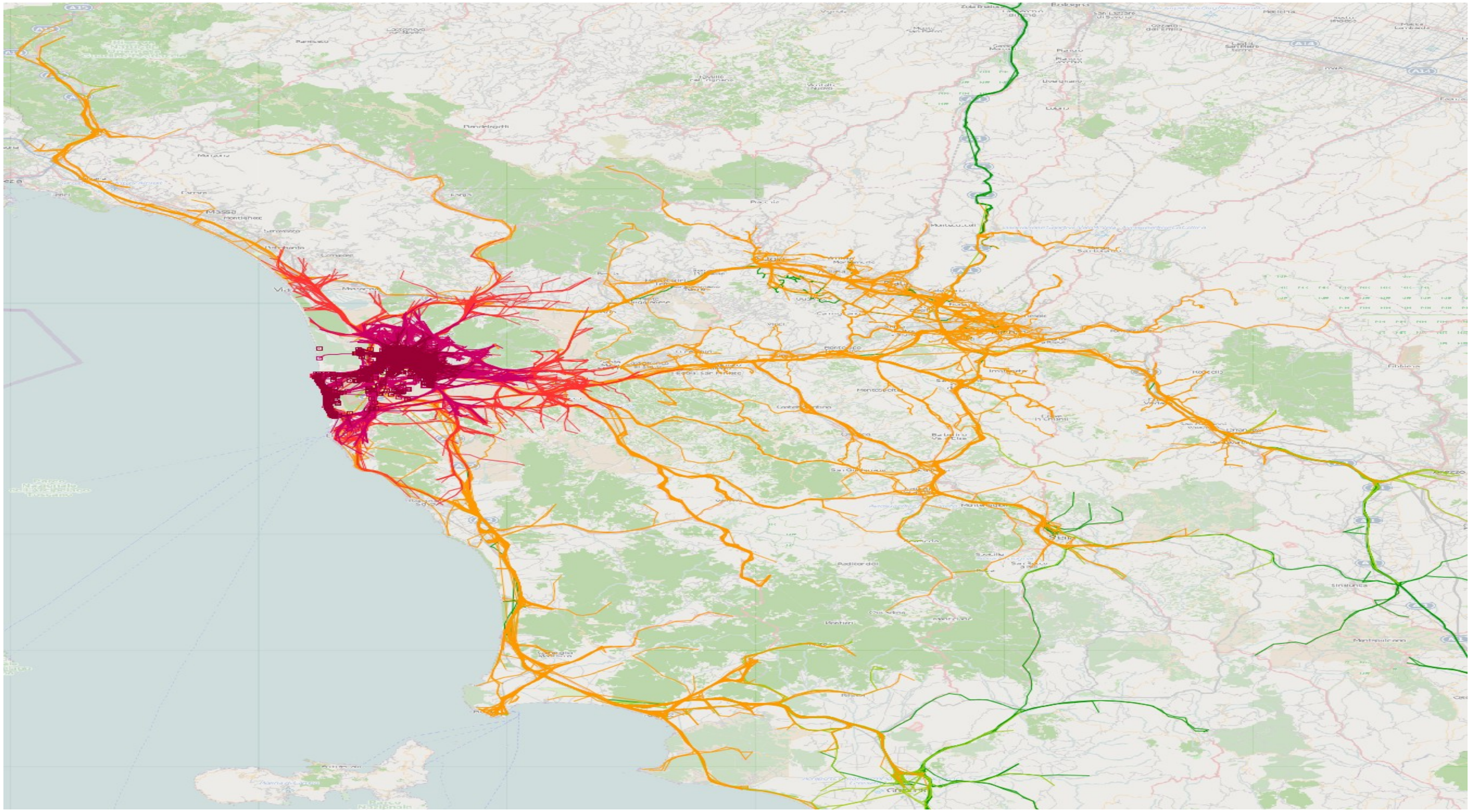
A small city: Pisa



First dimension: space
Travel length distribution



Travel length on the map

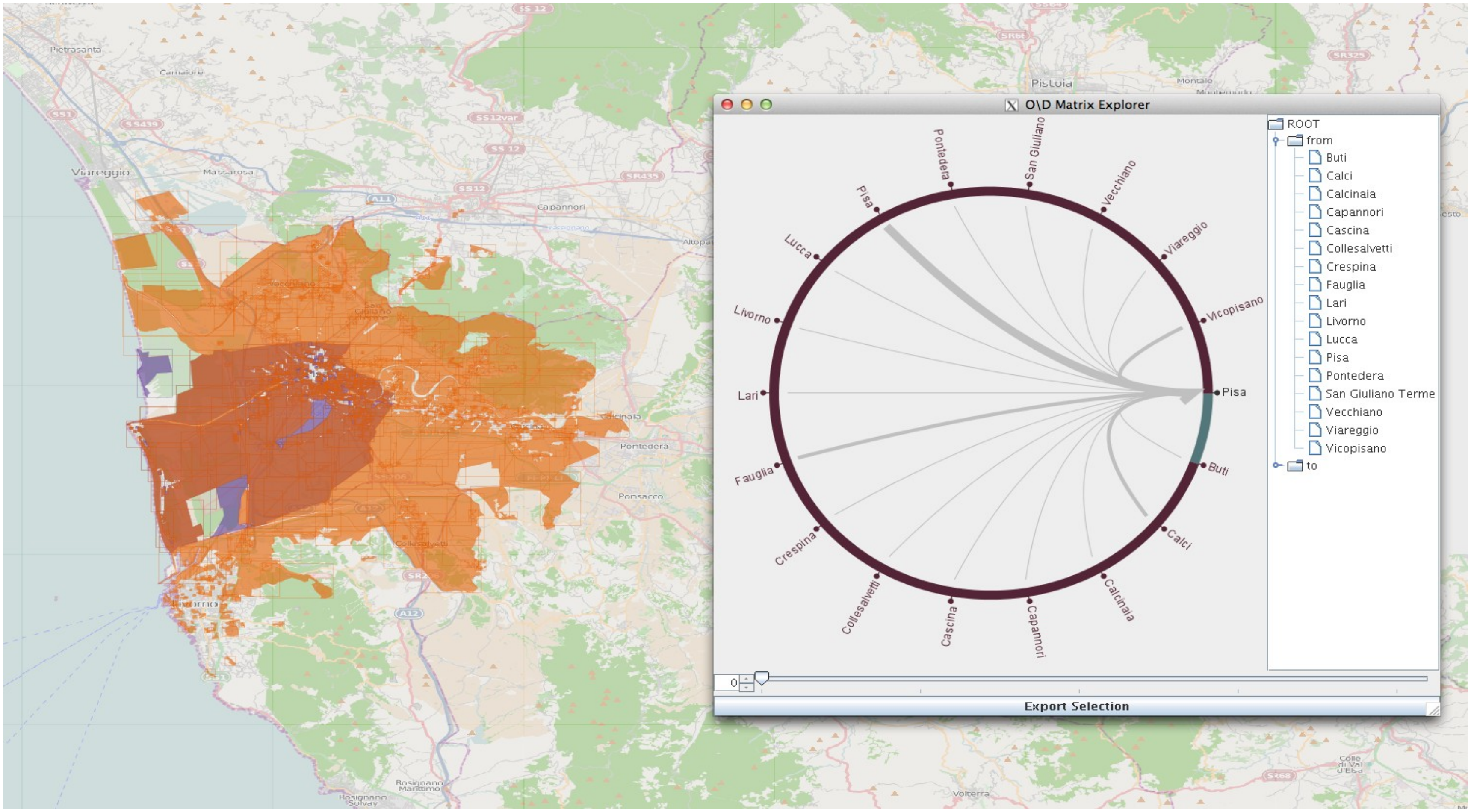


Exploring Origin and Destinations

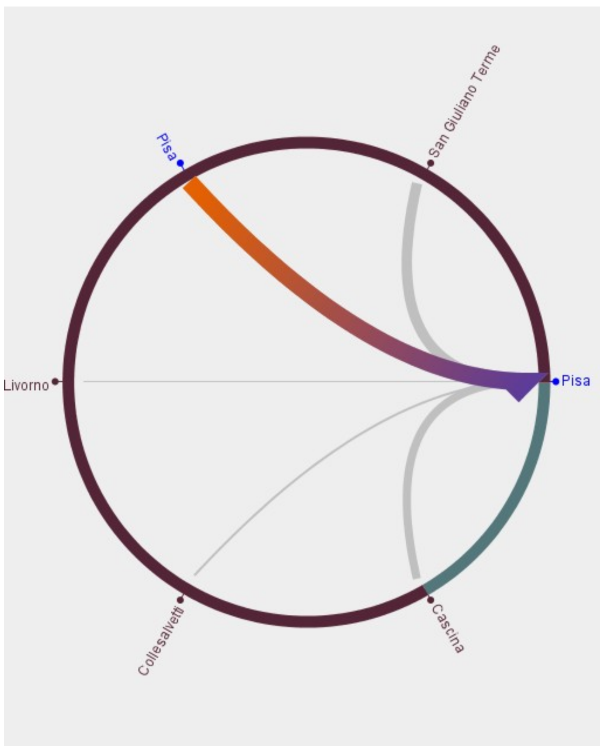
The general process

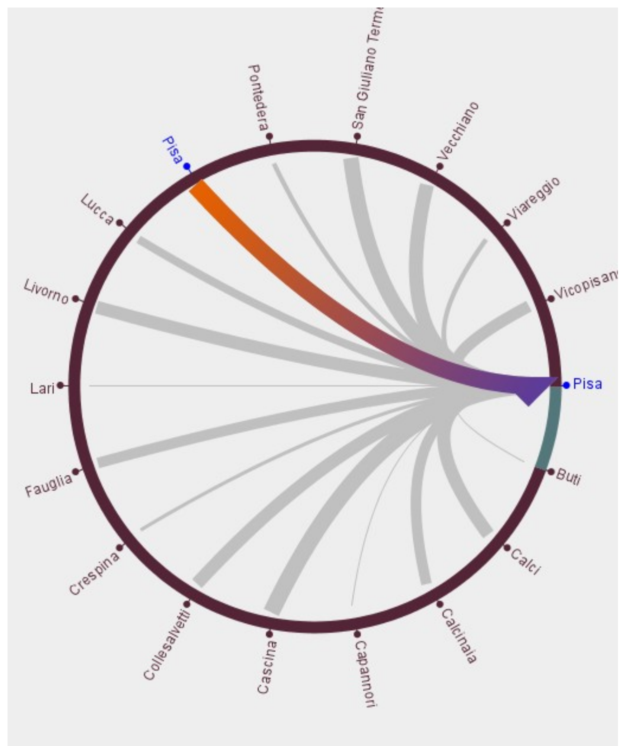
- Browse DWH at high spatial level (provinces)
- Identify interesting flows and drill down to specific flow
- Navigate the cube at finer level (cities)
- From flow to trajectories: entails original data
- Do specific analysis on the real trips

Exploring Origins and Destinations

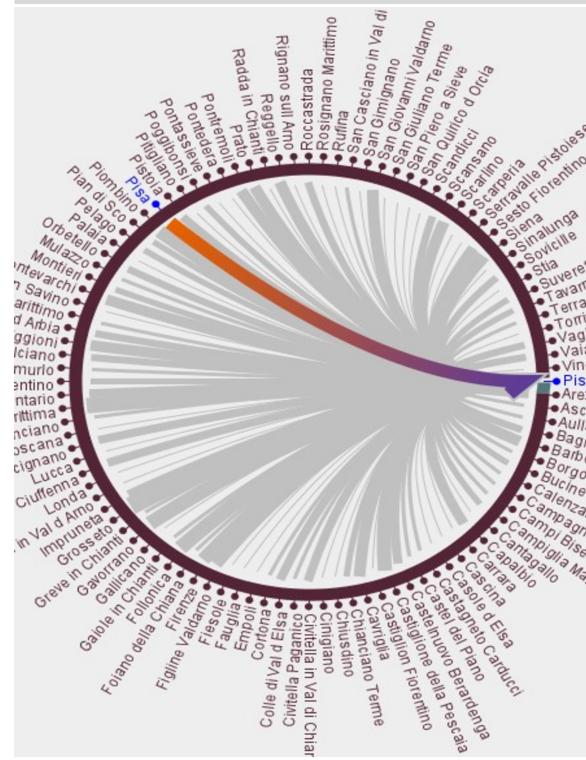


Exploring the origins of trips



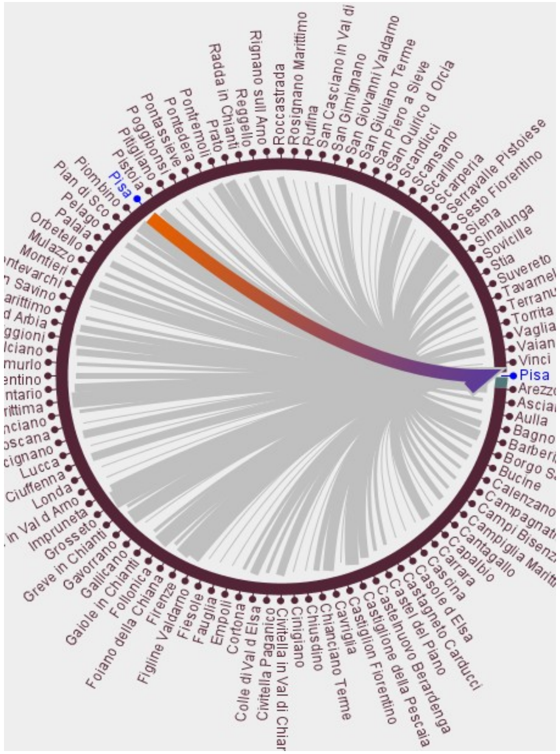


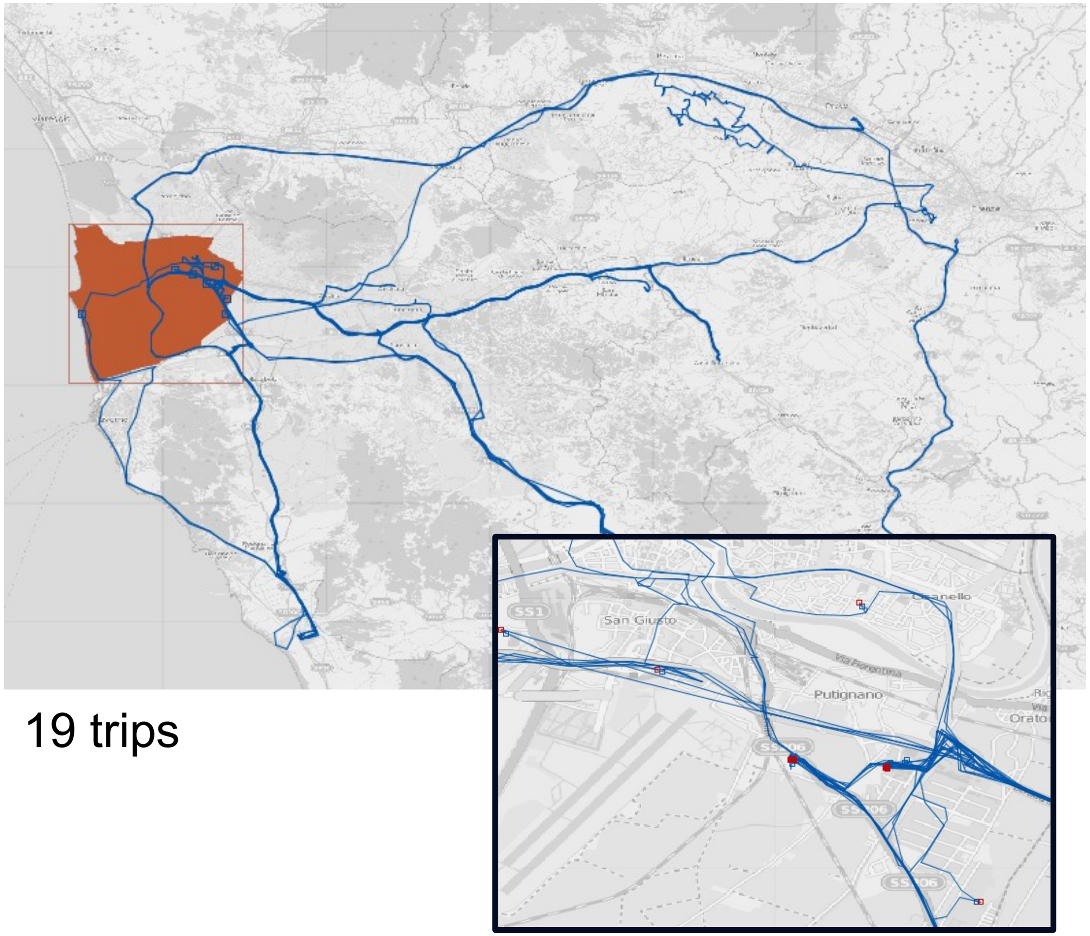
> 150km



Exploring origins of trips

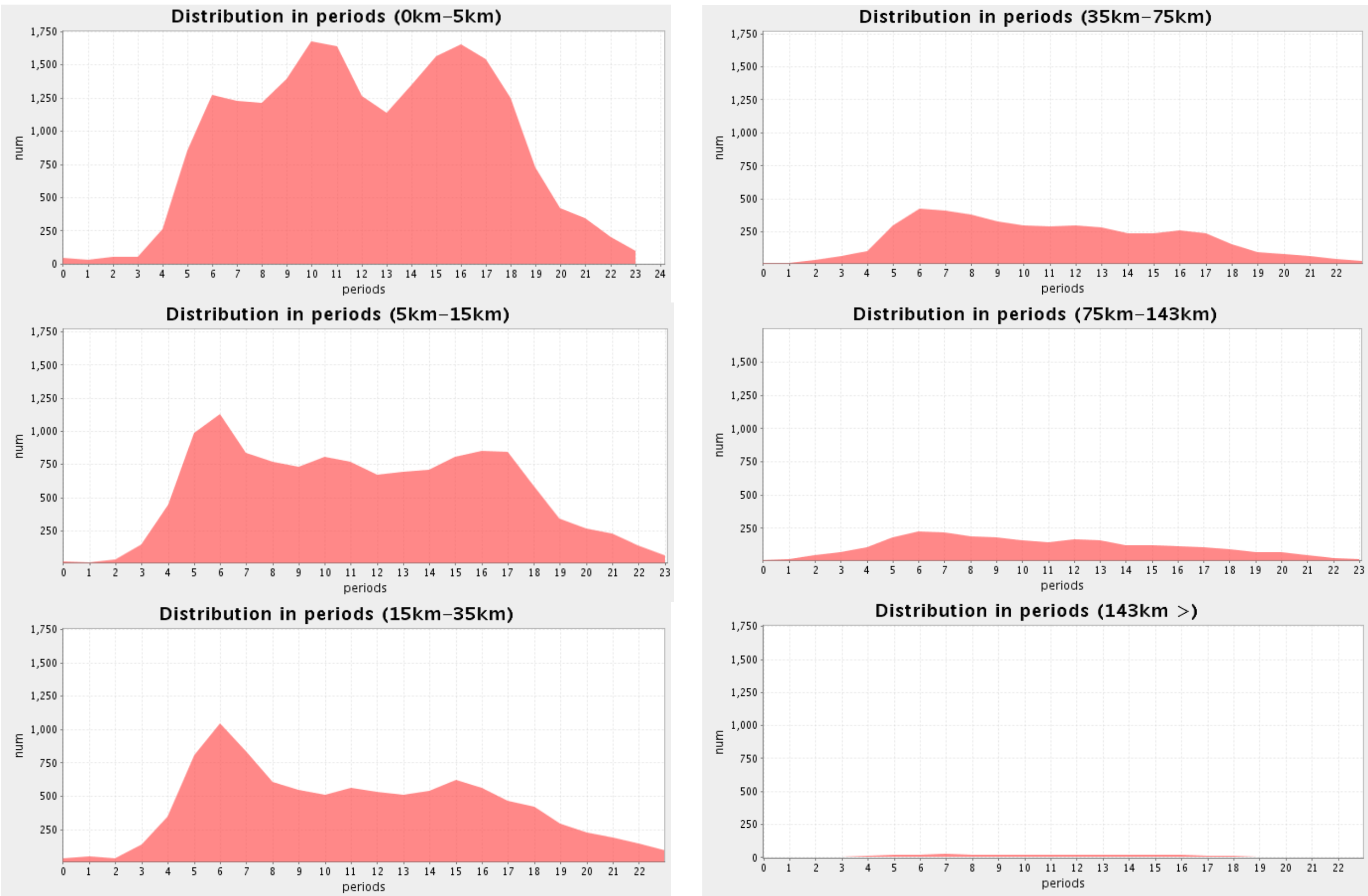
> 150km

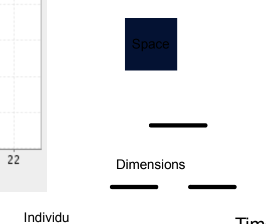




19 trips

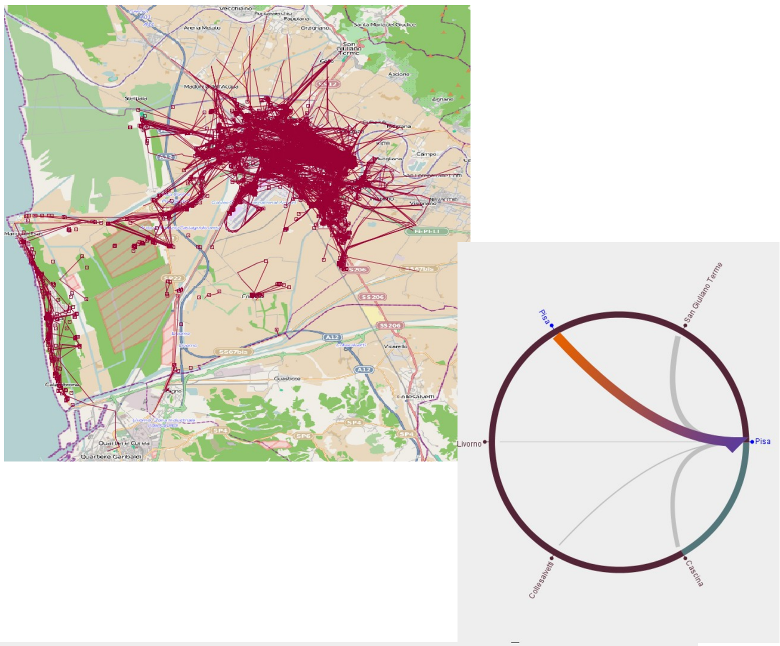
Second dimension: time
When people move to Pisa?

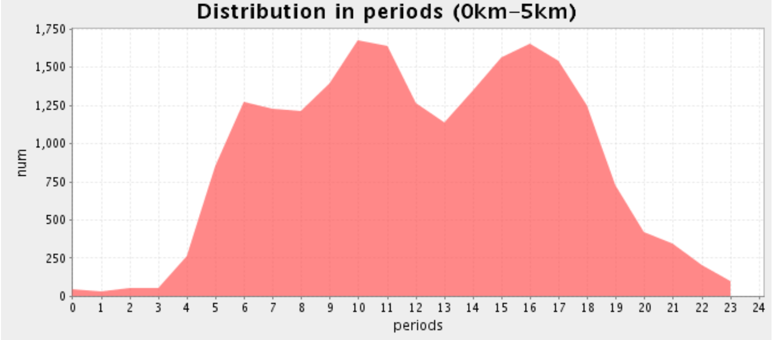




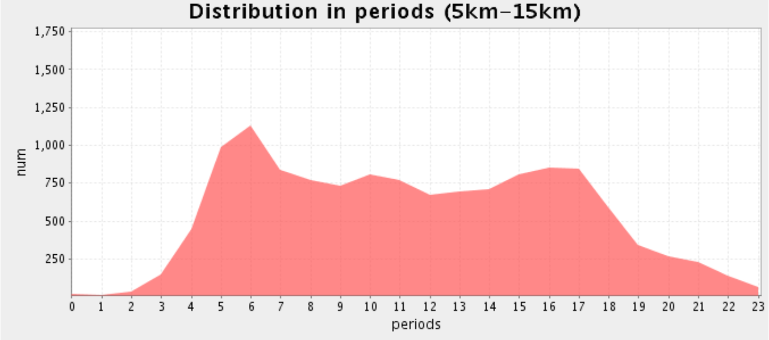
Let’s focus at city level

0km – 5Km

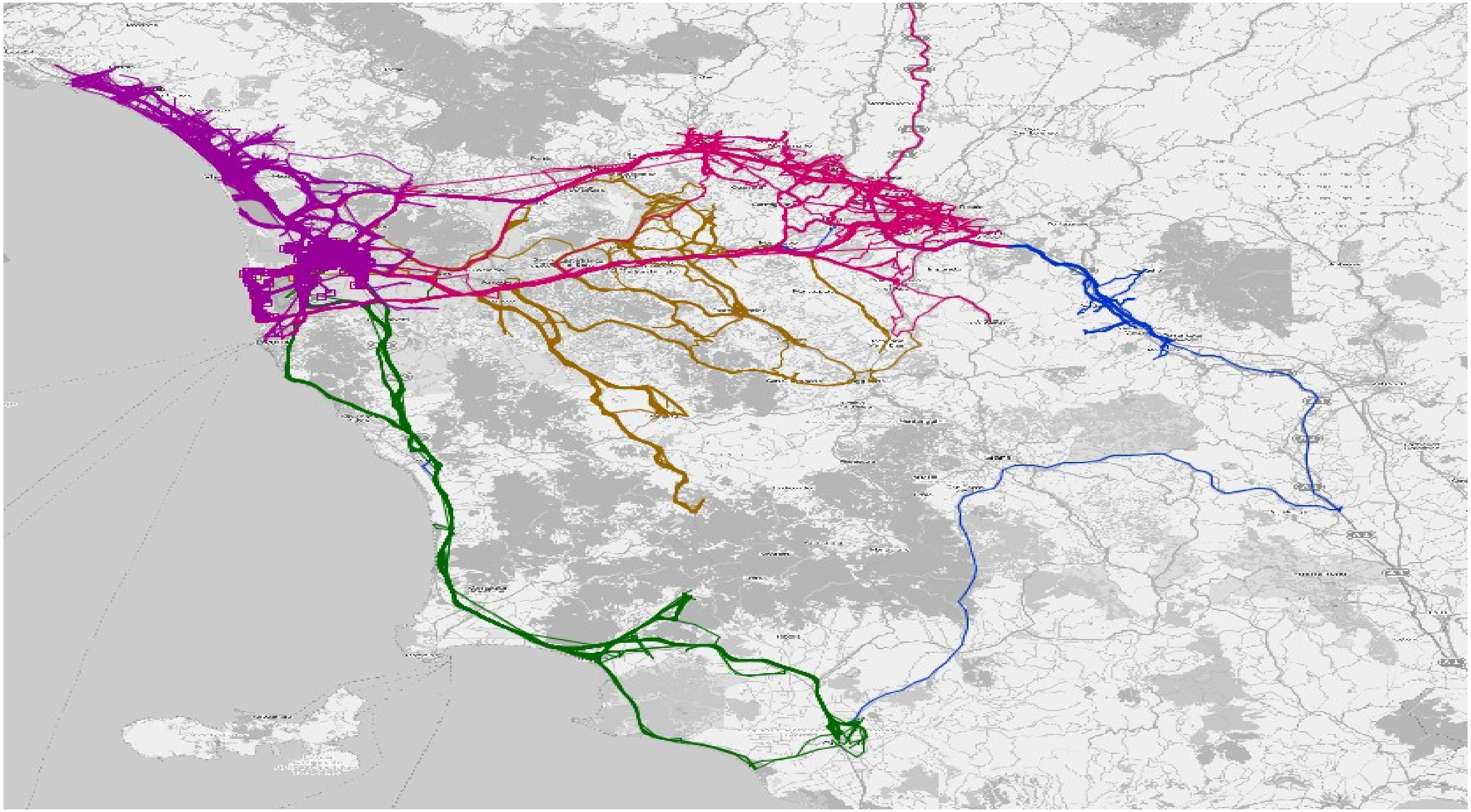




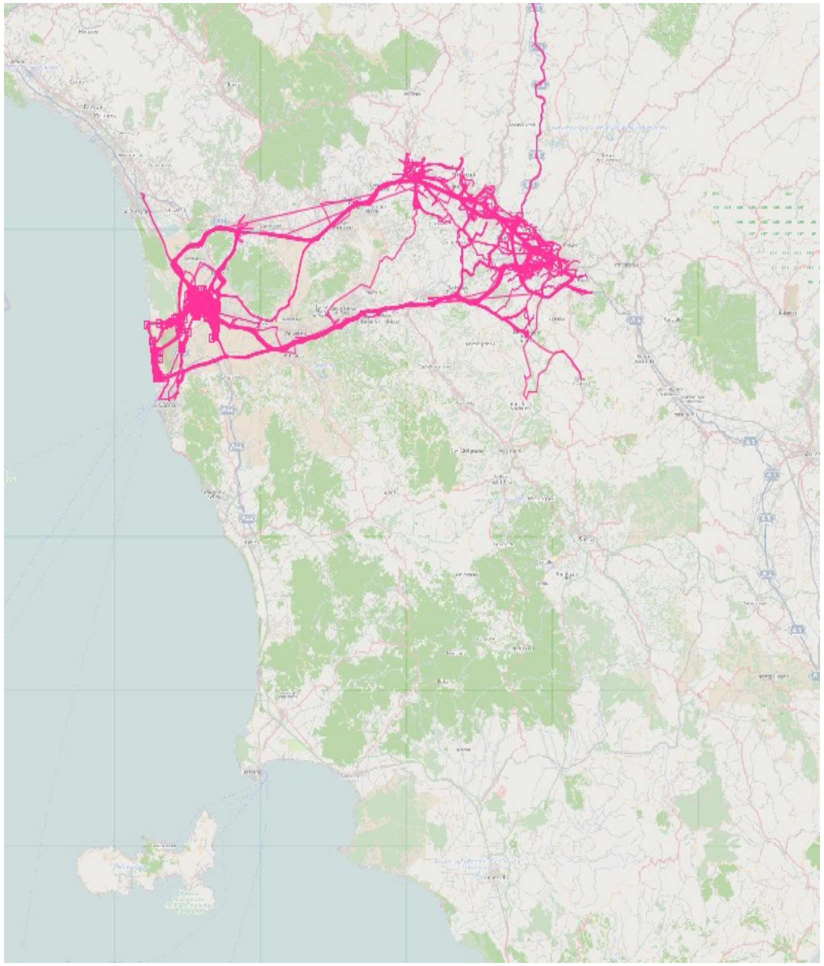
5km – 15Km

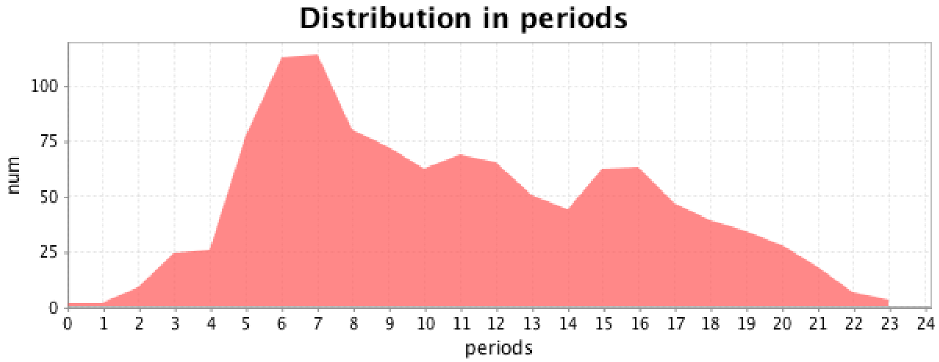


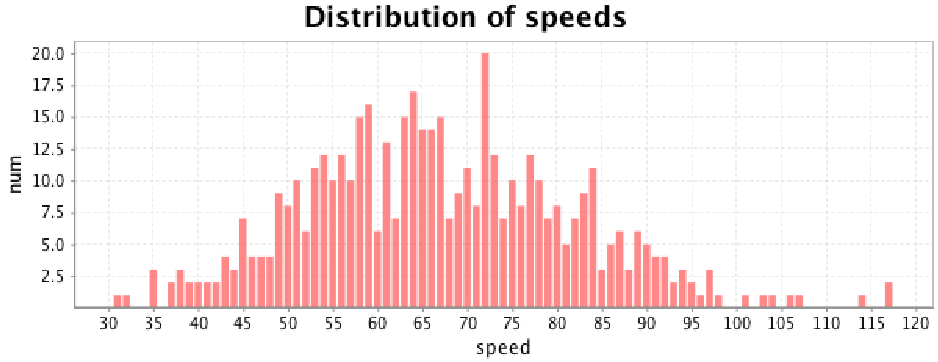
Trips segmented by similarity



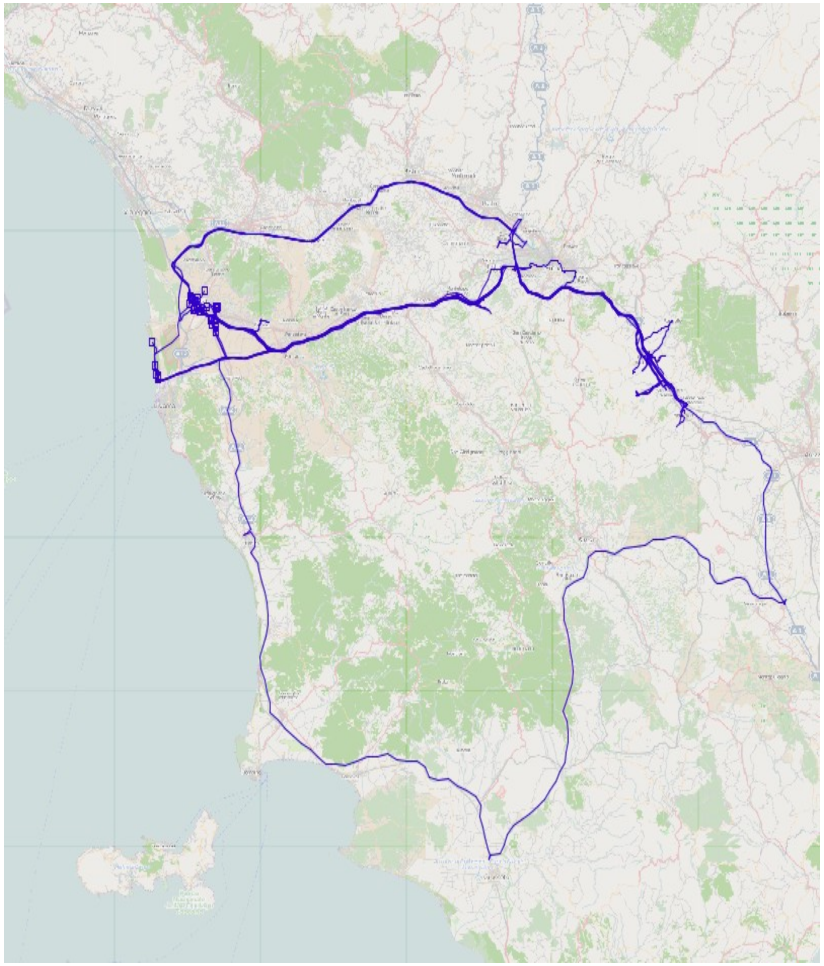
Explore clusters: Florence

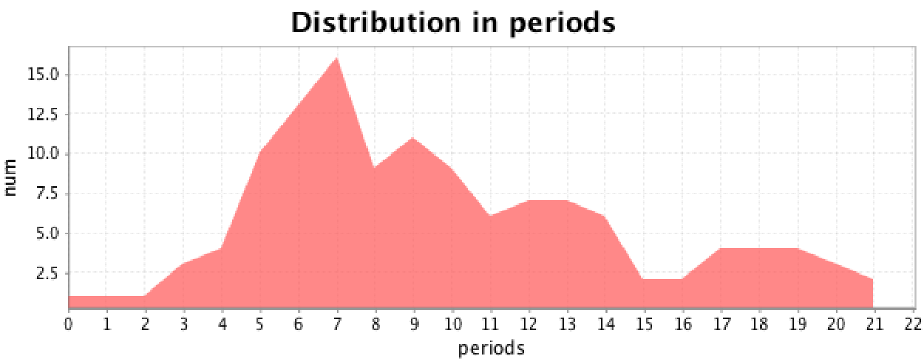


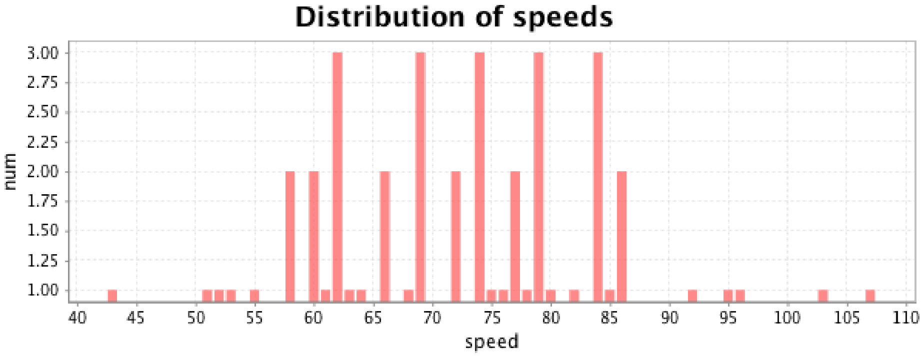




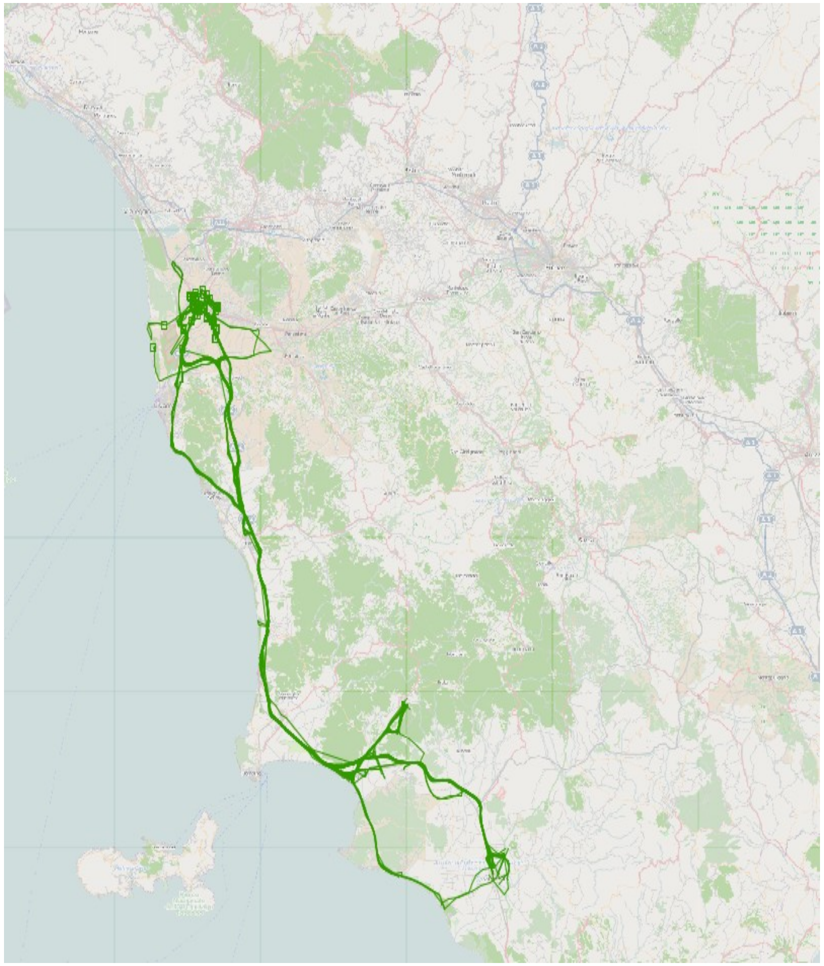
Explore clusters: A1

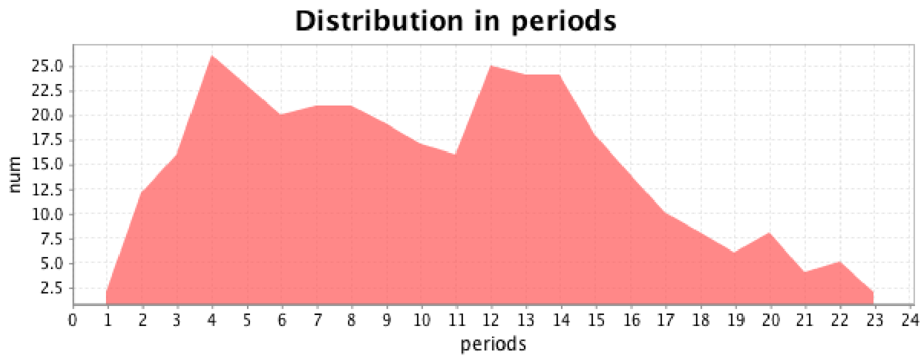


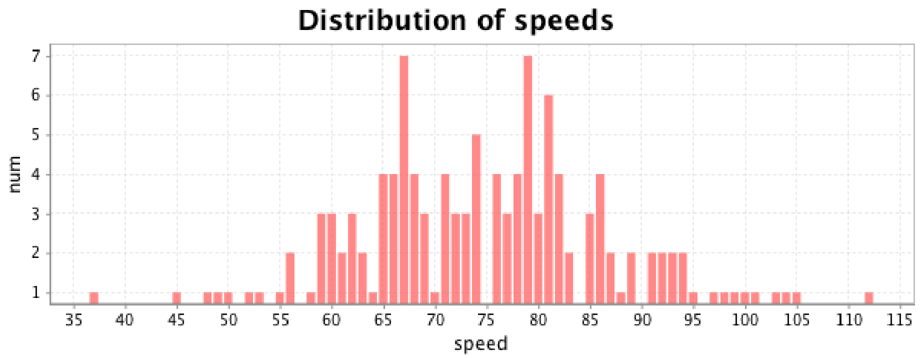




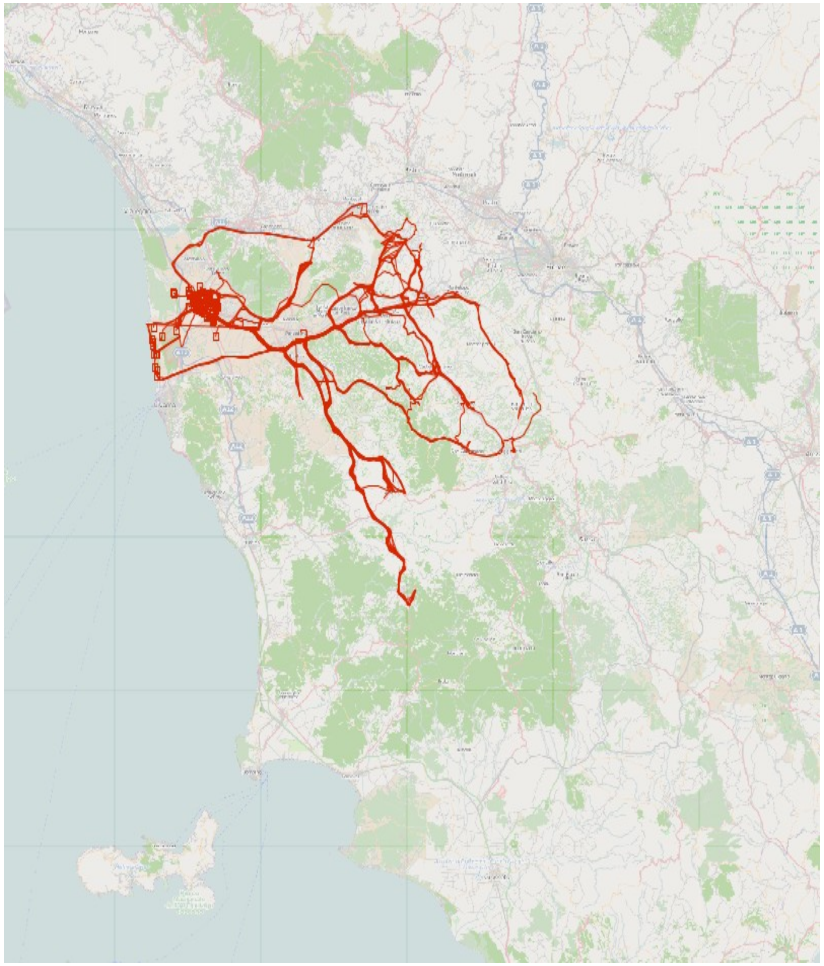
Explore clusters: A12

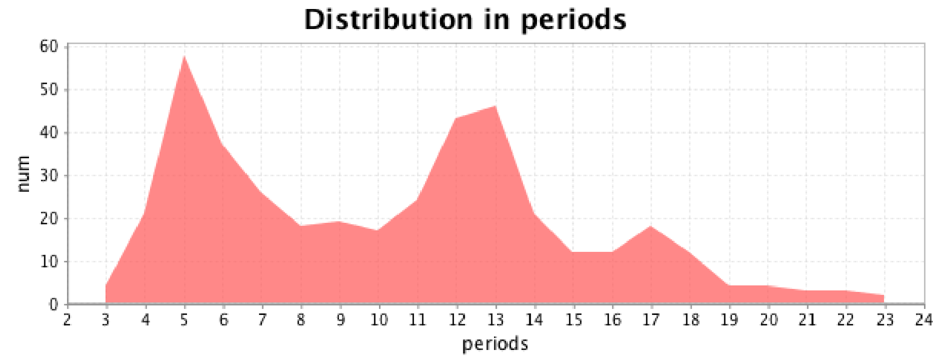


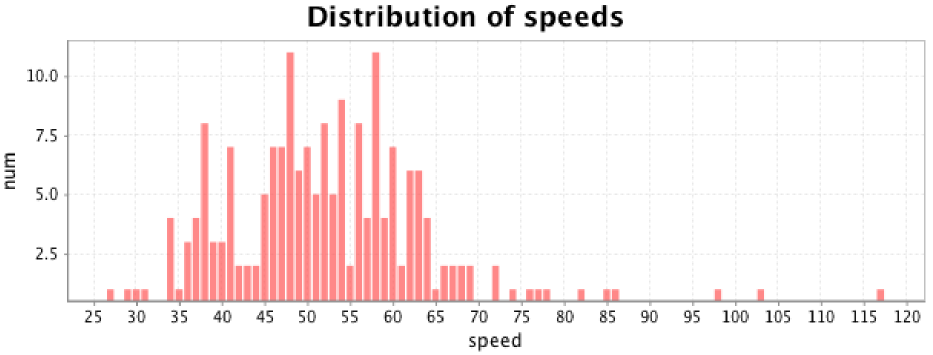




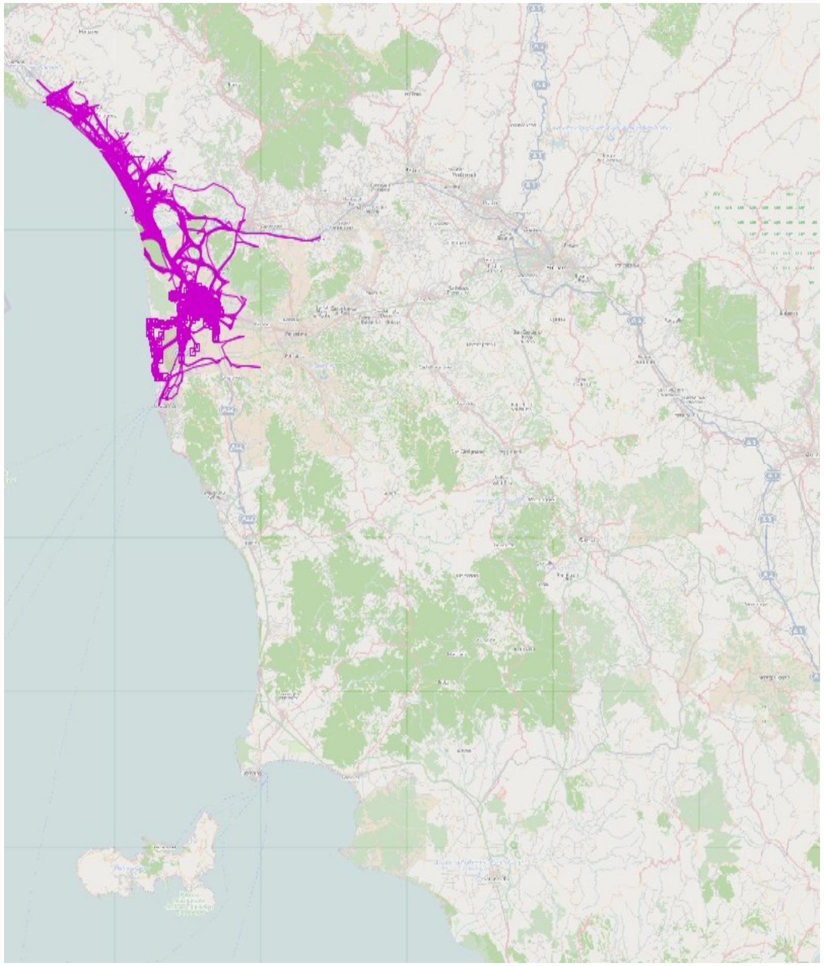
Explore Clusters: Valdera

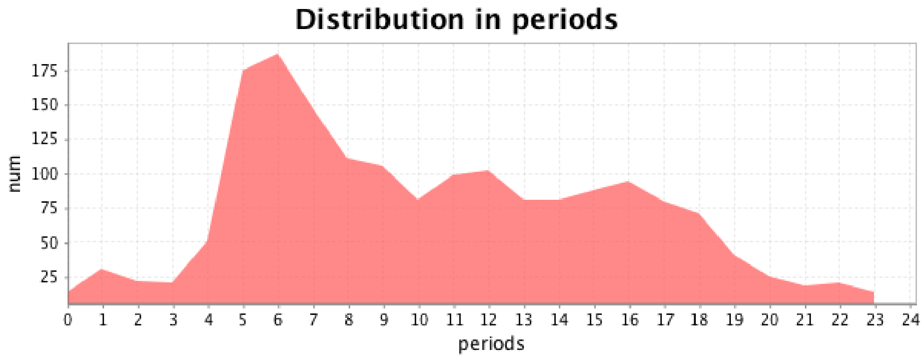


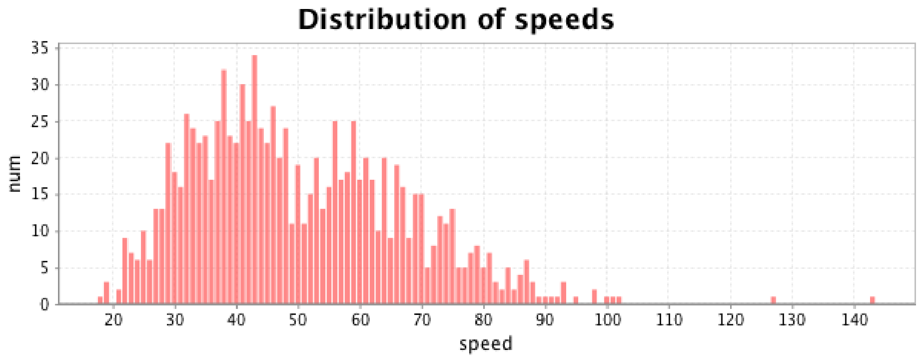




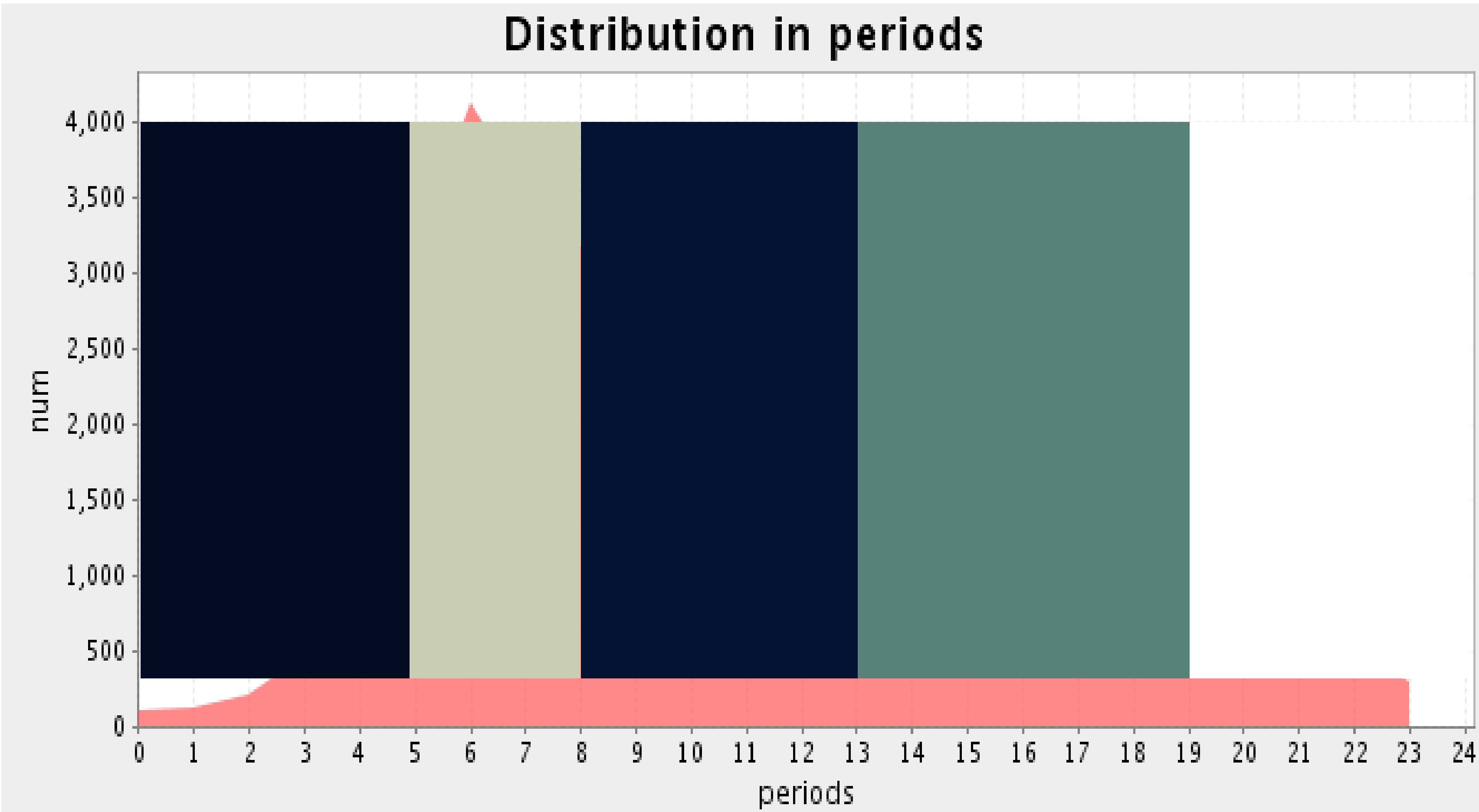
Explore clusters: Versilia



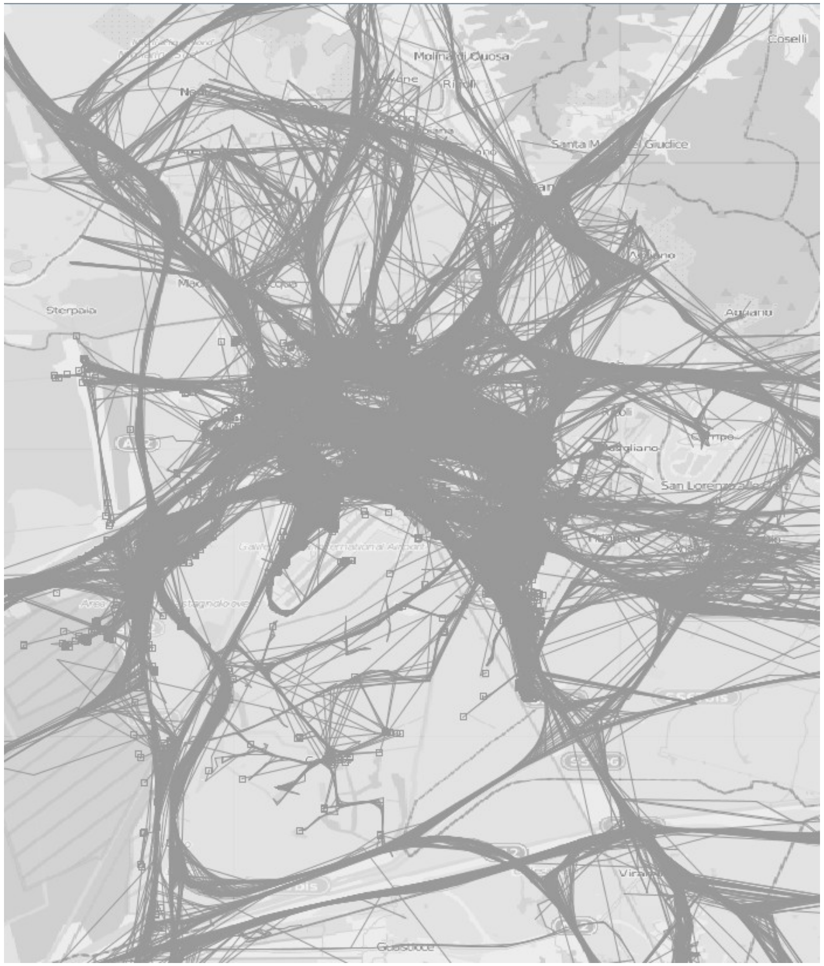


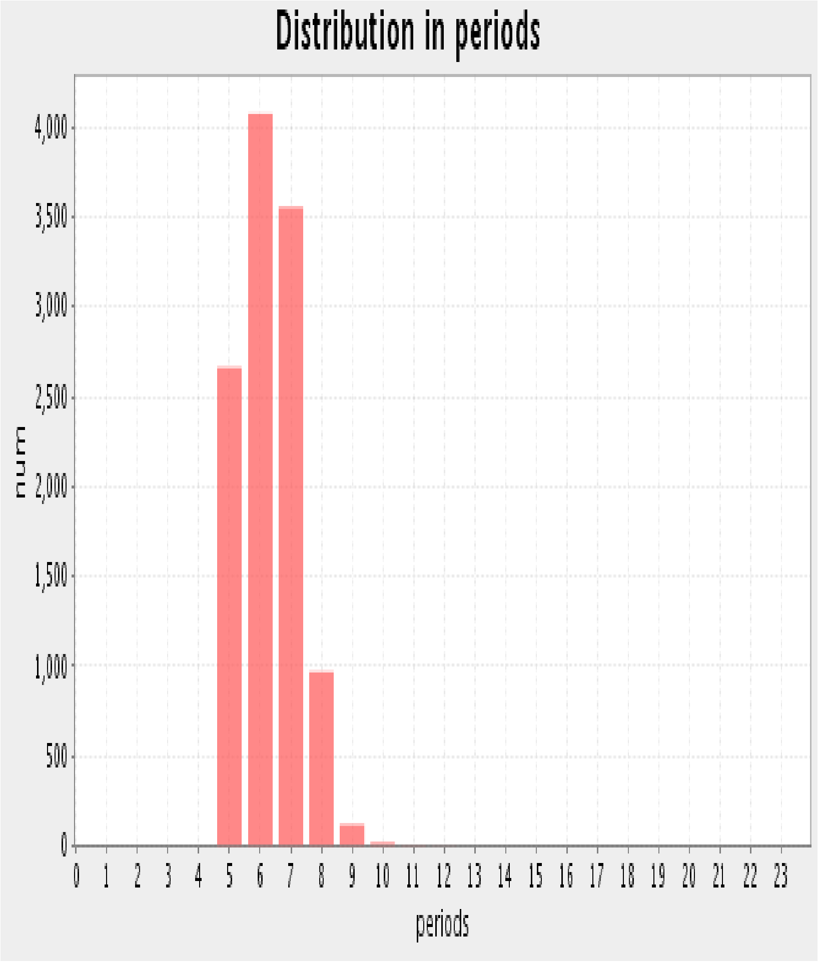


Trip segmentation by time

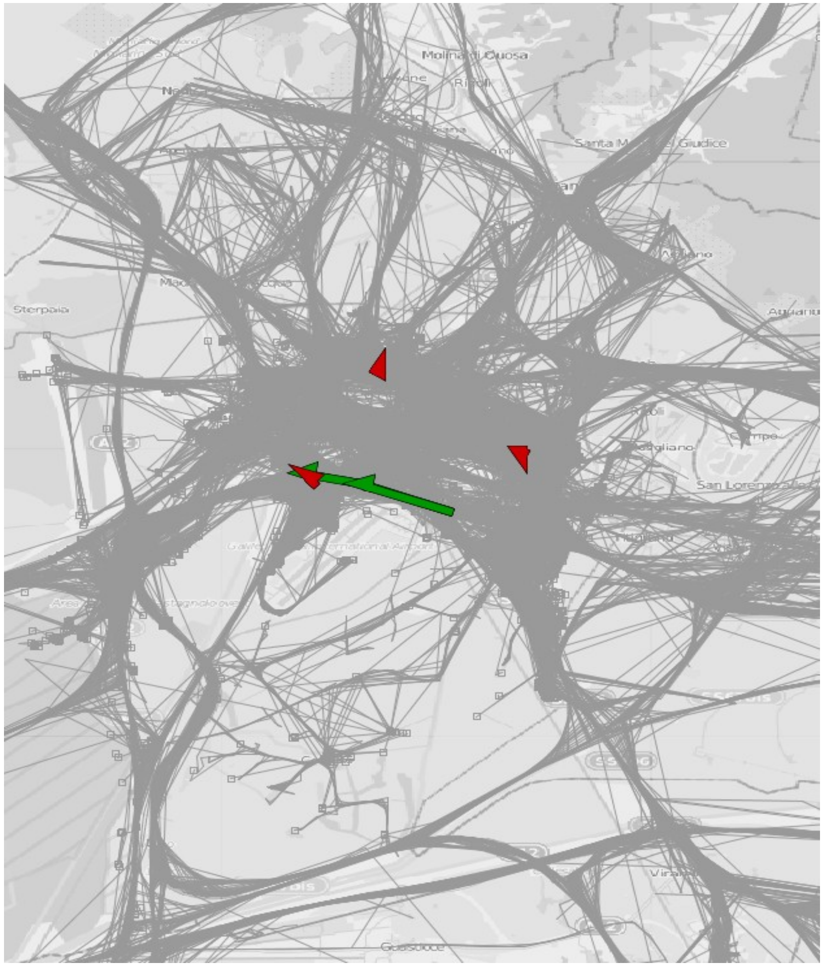


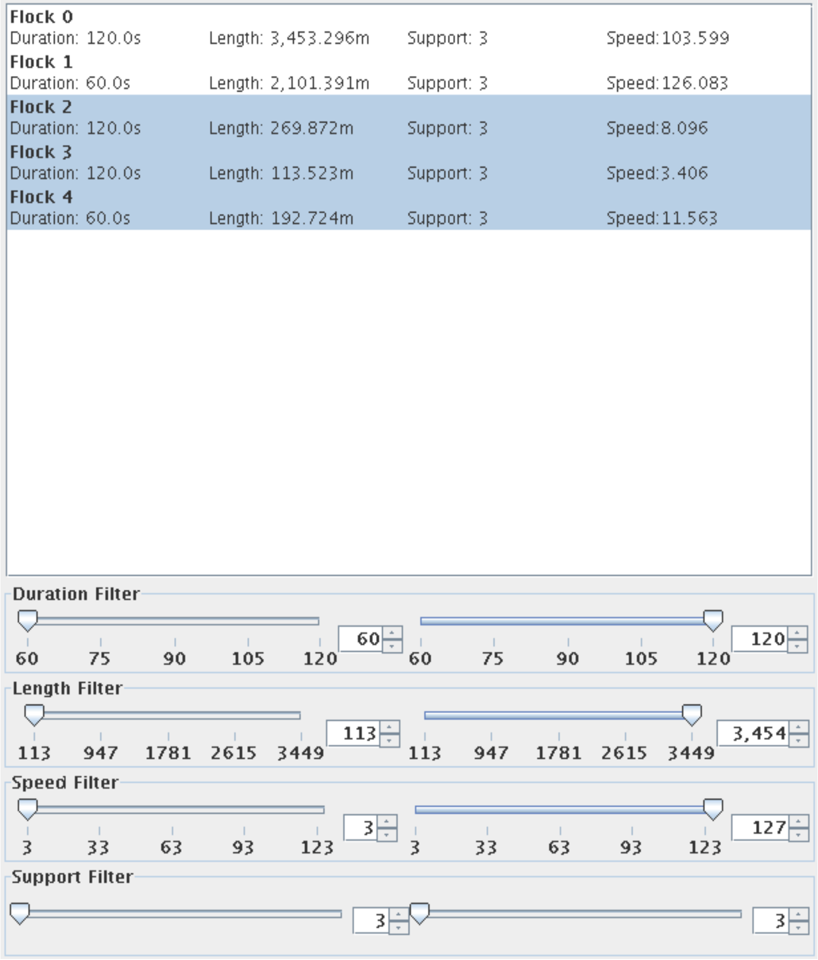
Trips Segmented by Time: from 5 to 8





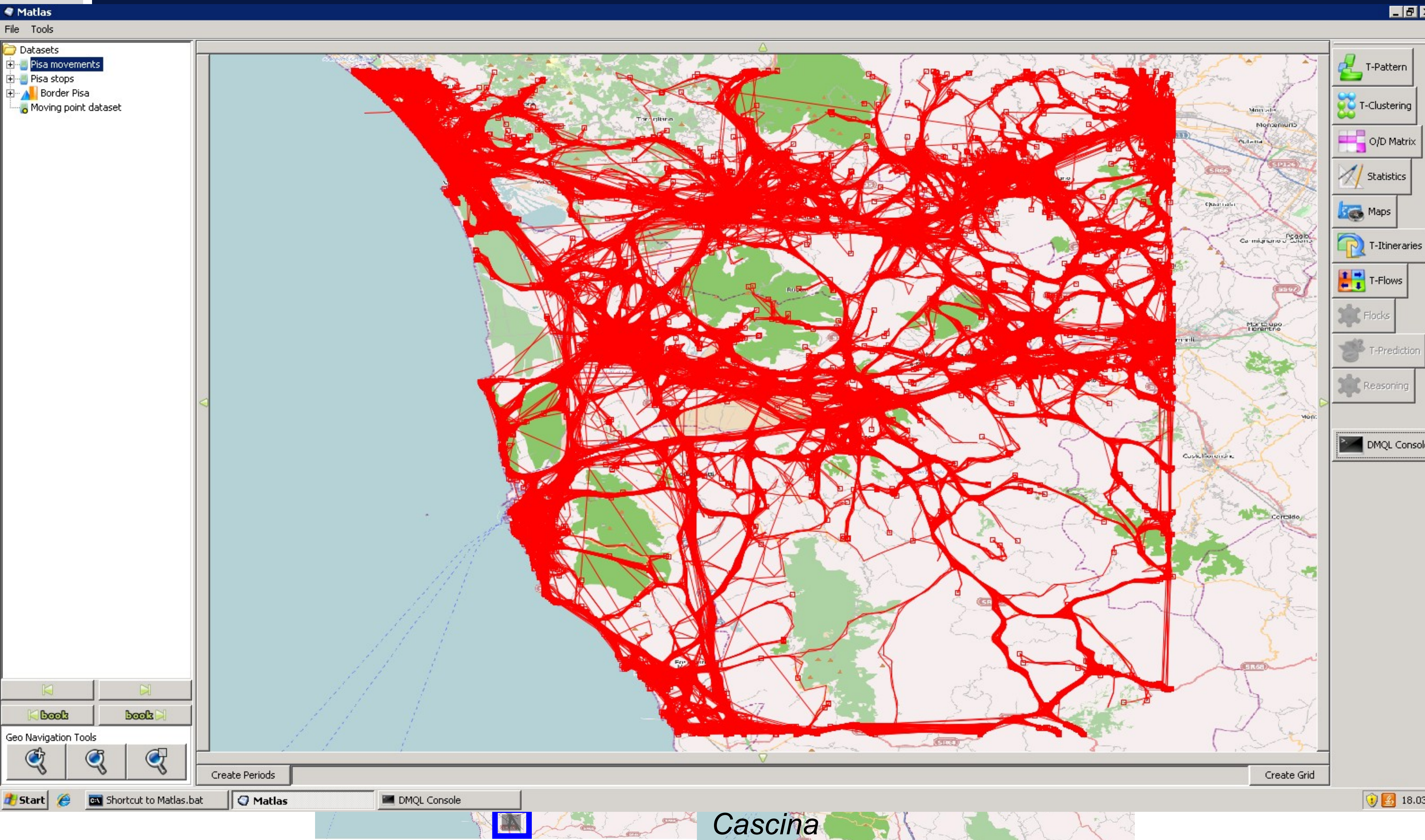
Discover traffic jams



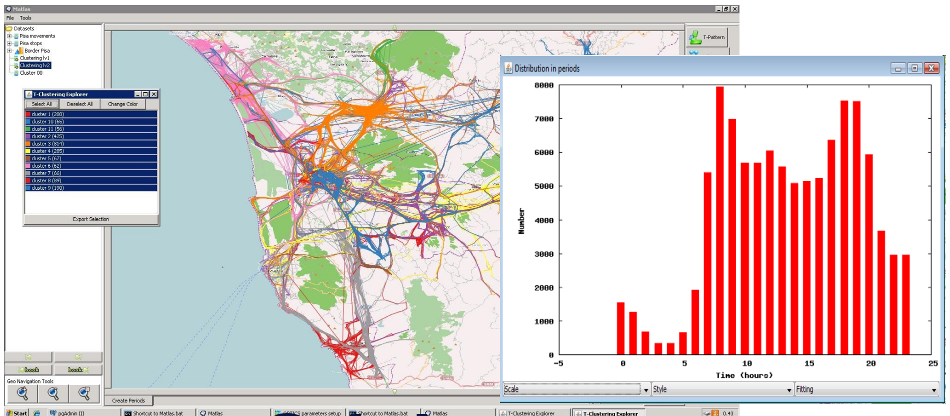


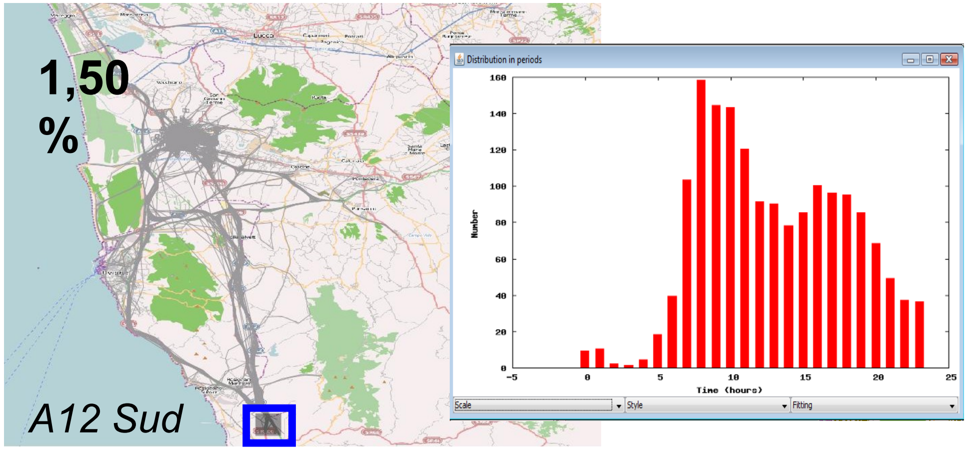
Discovering access patterns to Pisa with GPS tracks data

Access patterns using T-clustering

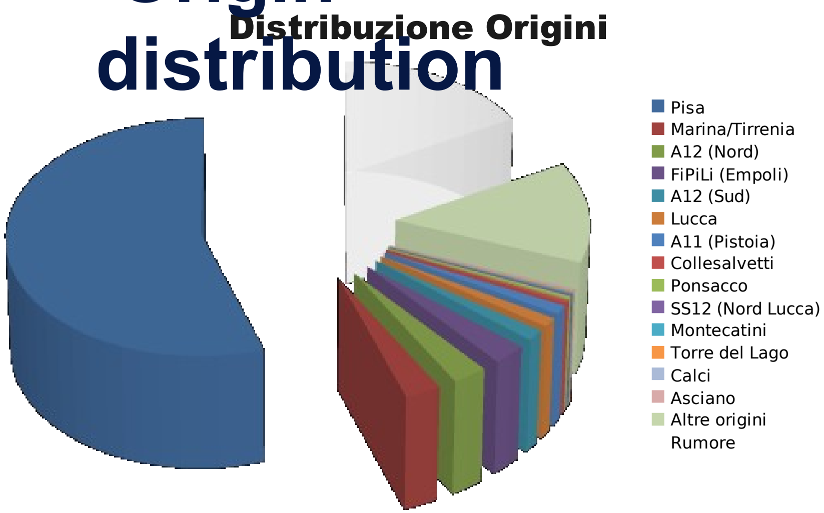


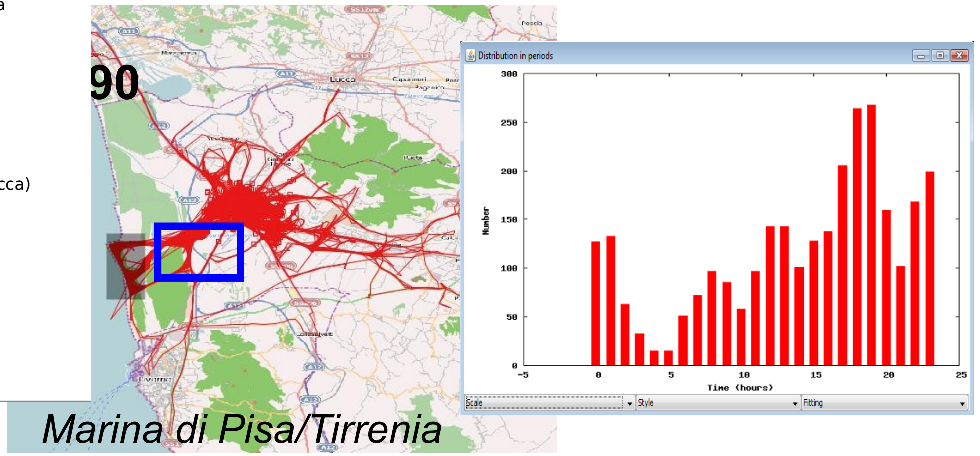
Characterizing the access patterns: origin & time



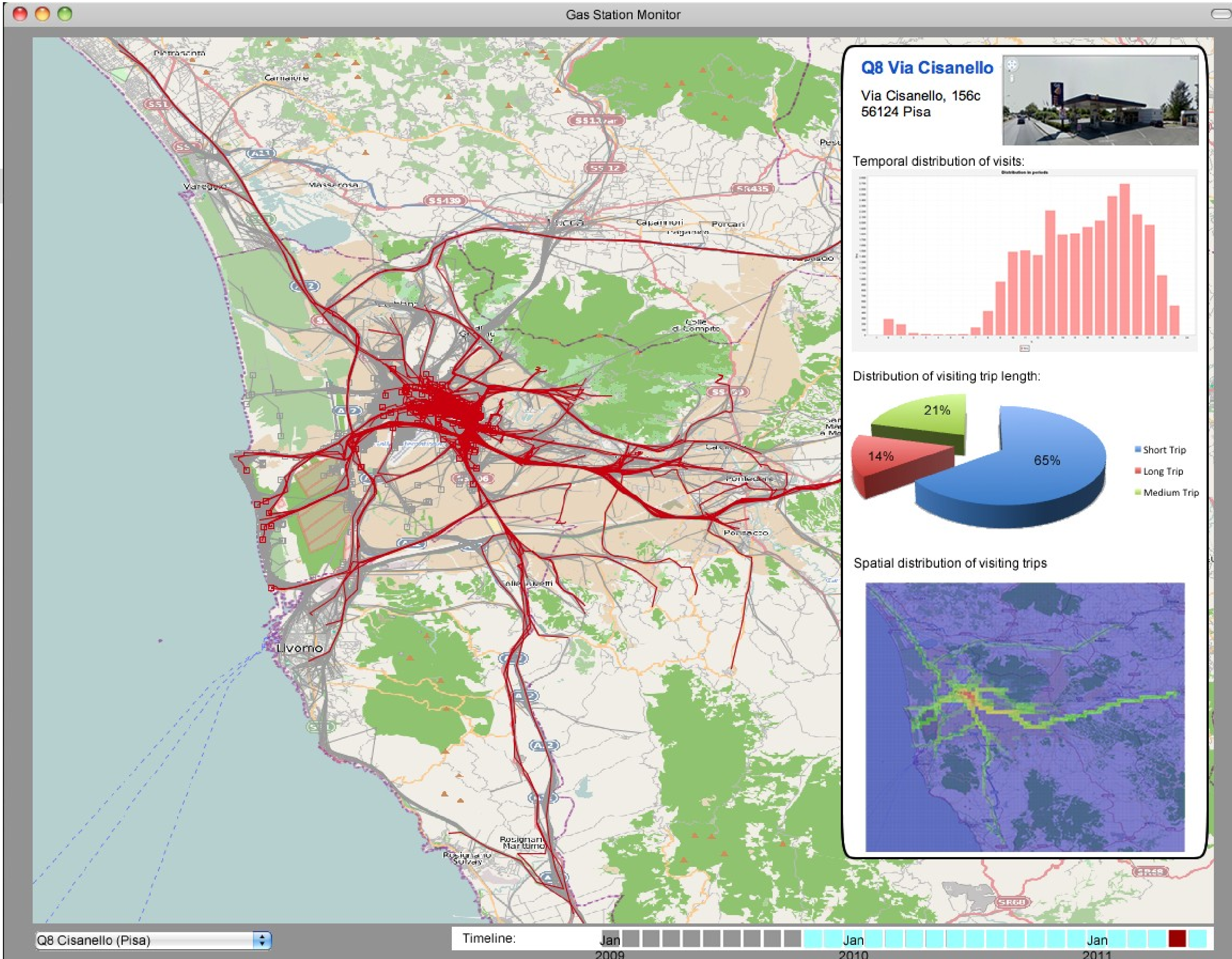


Origin distribution

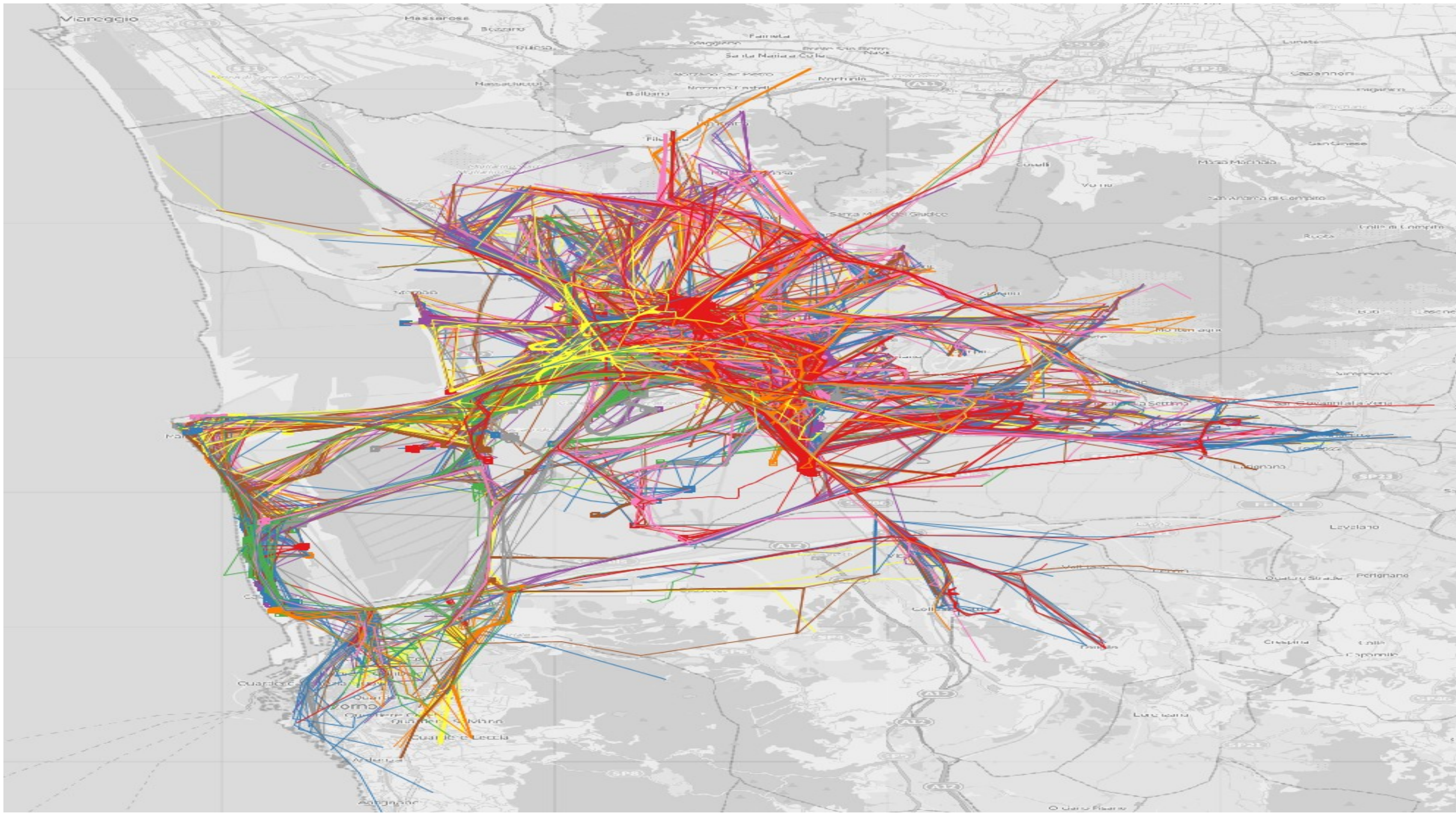




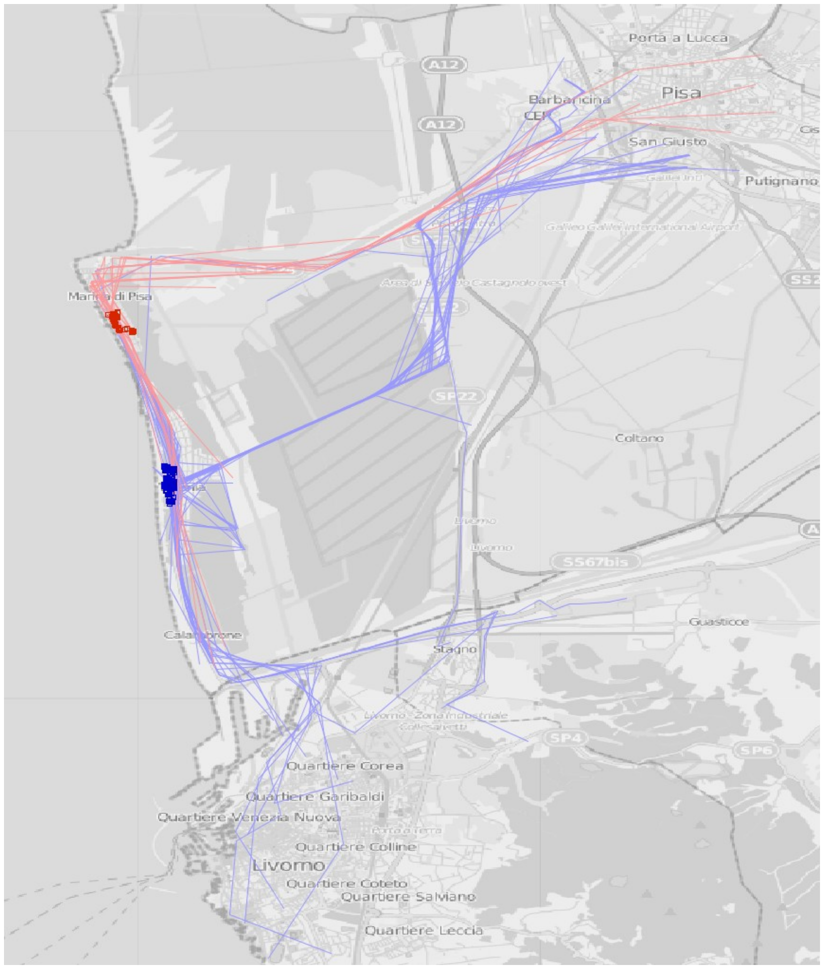
Studying the attractiveness/efficiency of a service with GPS tracks

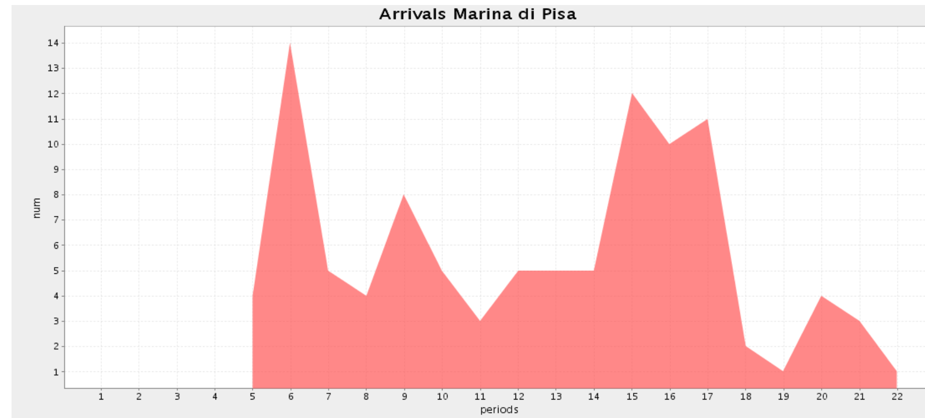

Gas Station Monitor
Q8 Via Cisanello
Via Cisanello, 156c
56124 Pisa
Temporal distribution of visits:
Distribution of visiting trip length:
21%
14%
65%
Short Trip
Long Trip
Medium Trip
Spatial distribution of visiting trips
Livorno
Q8 Cisanello (Pisa)
Timeline:
Jan
2009
Jan
2010
Jan
2011

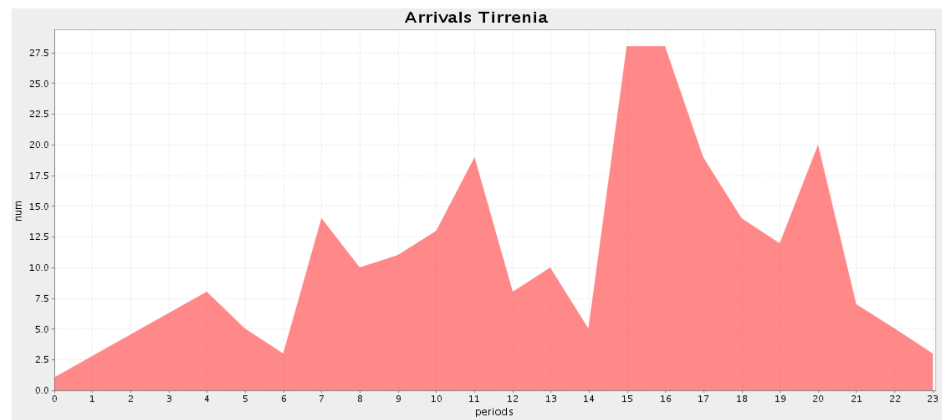
Aggregate trips by common destinations



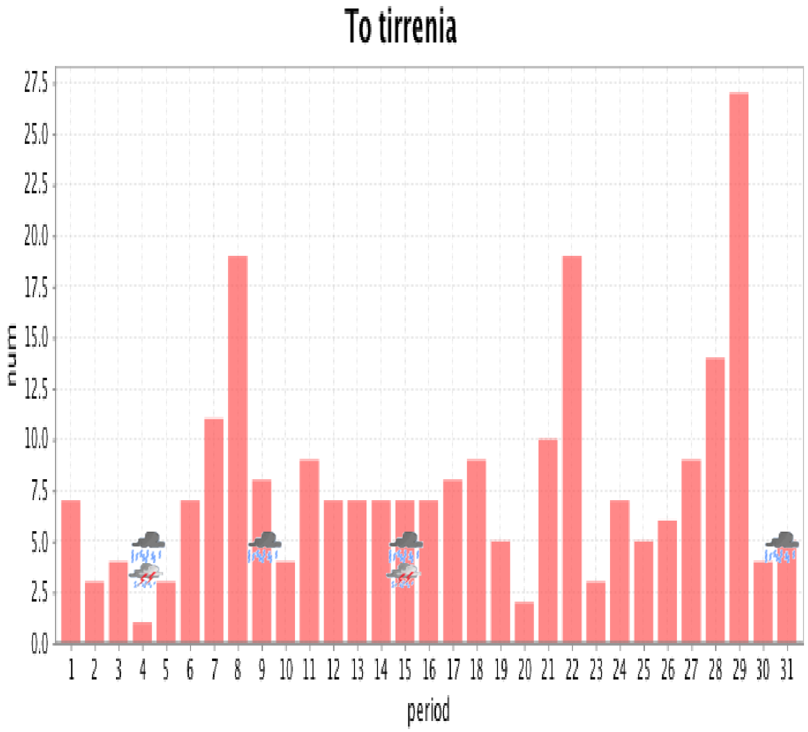
Seaside: Tirrenia and Marina di Pisa

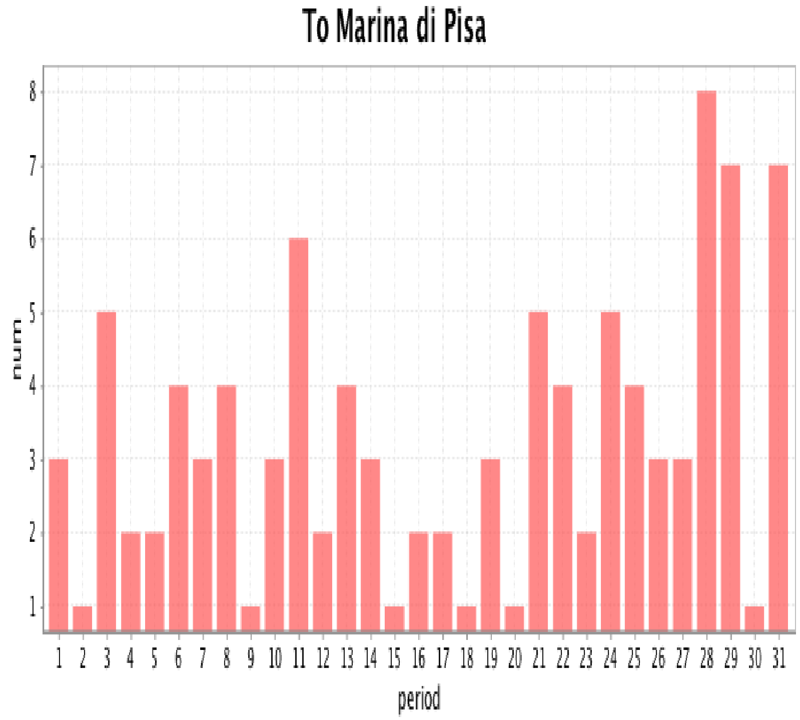






Seaside: Tirrena and Marina di Pisa

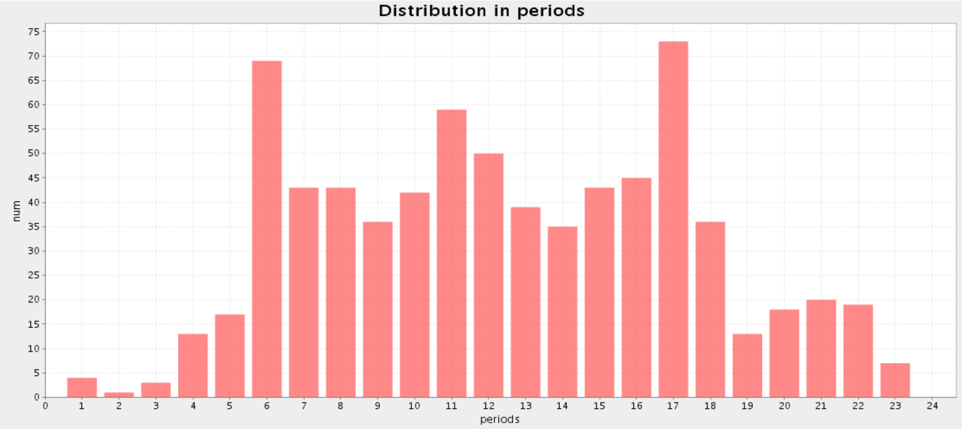




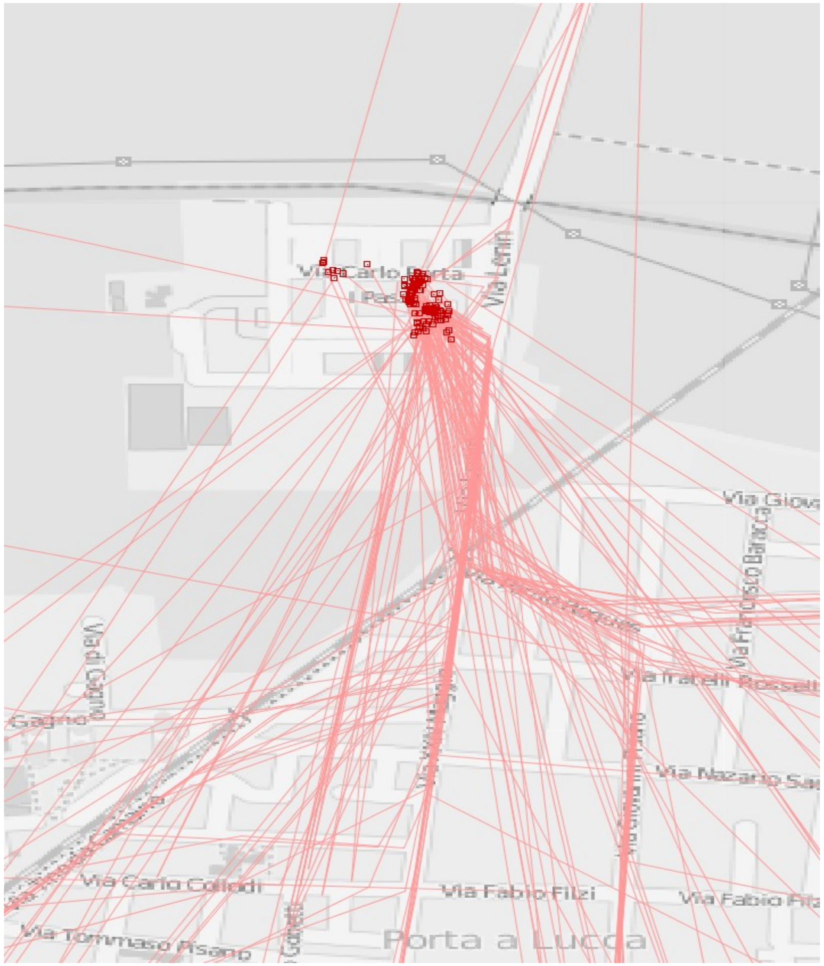
http://www.ilmeteo.it/portale/archivio-meteo/Tirrenia/2011/Maggio

Industry: Saint Gobain

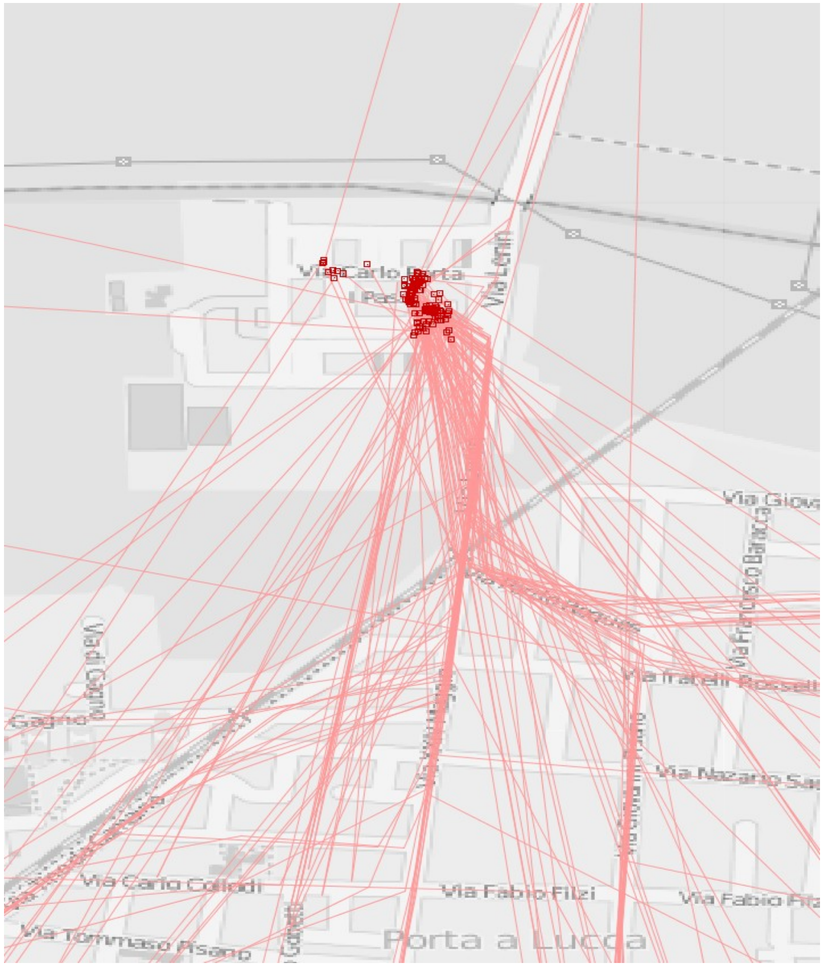
Industry: Saint Gobain

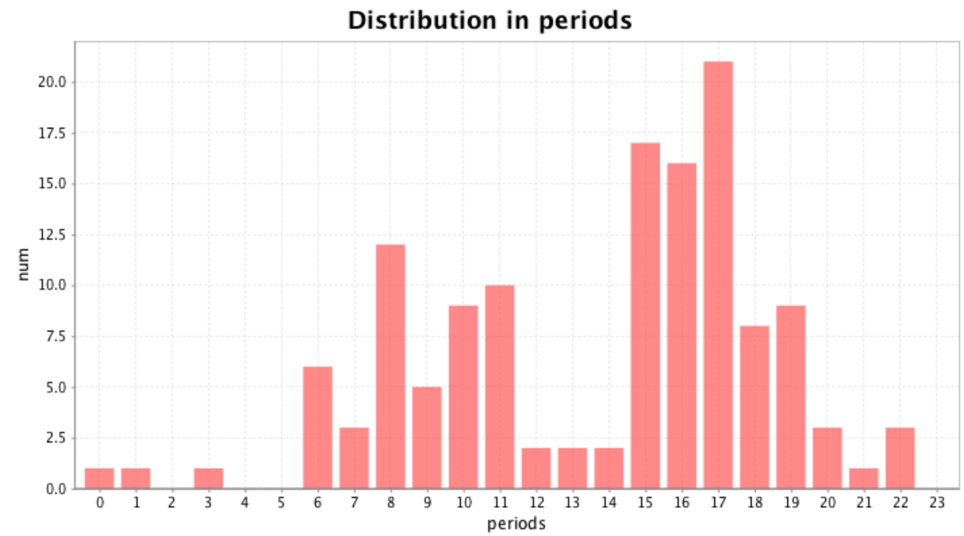


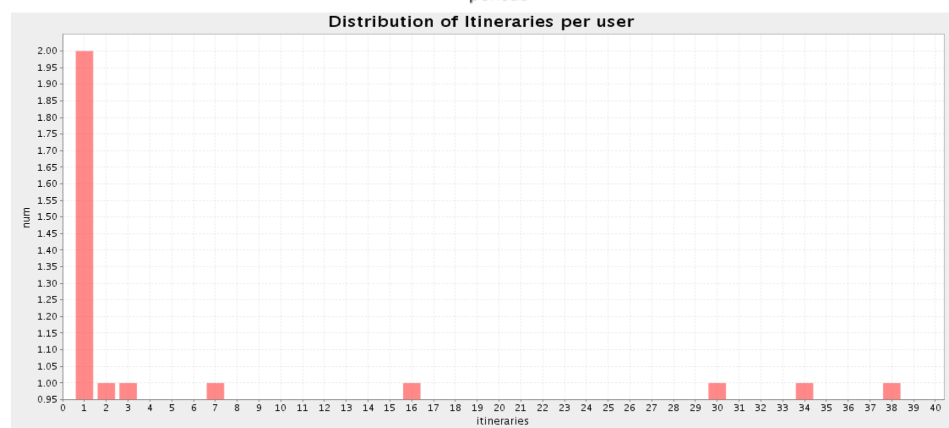
Residential Area: I Passi



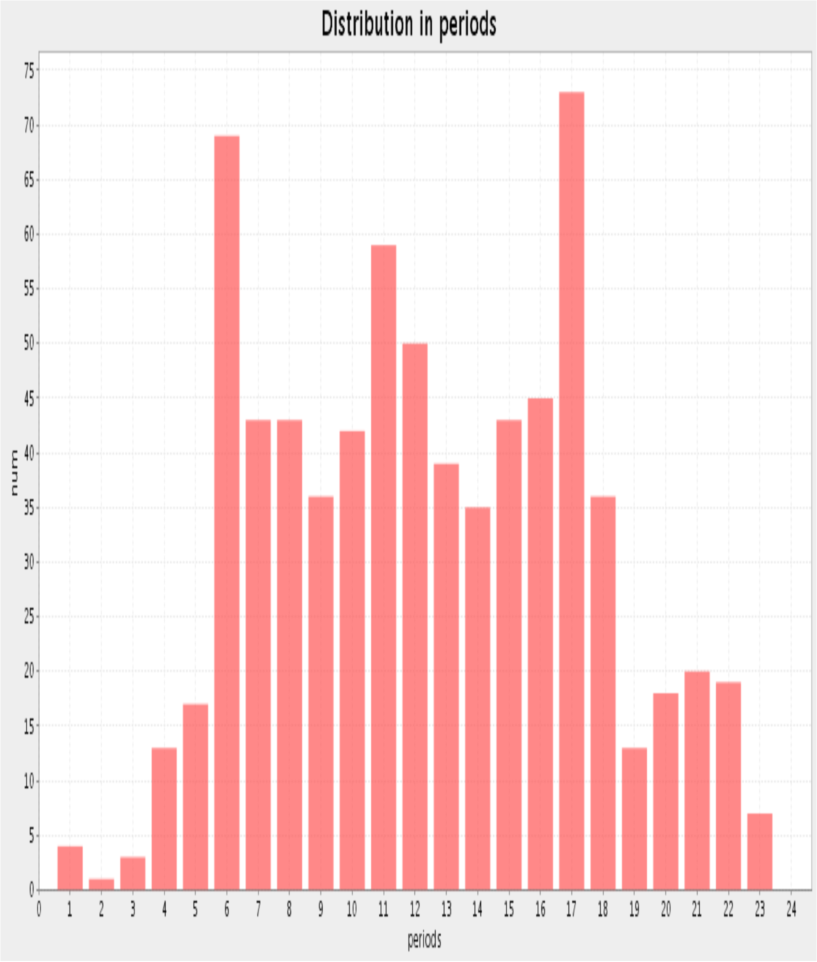
Residential Area: I Passi

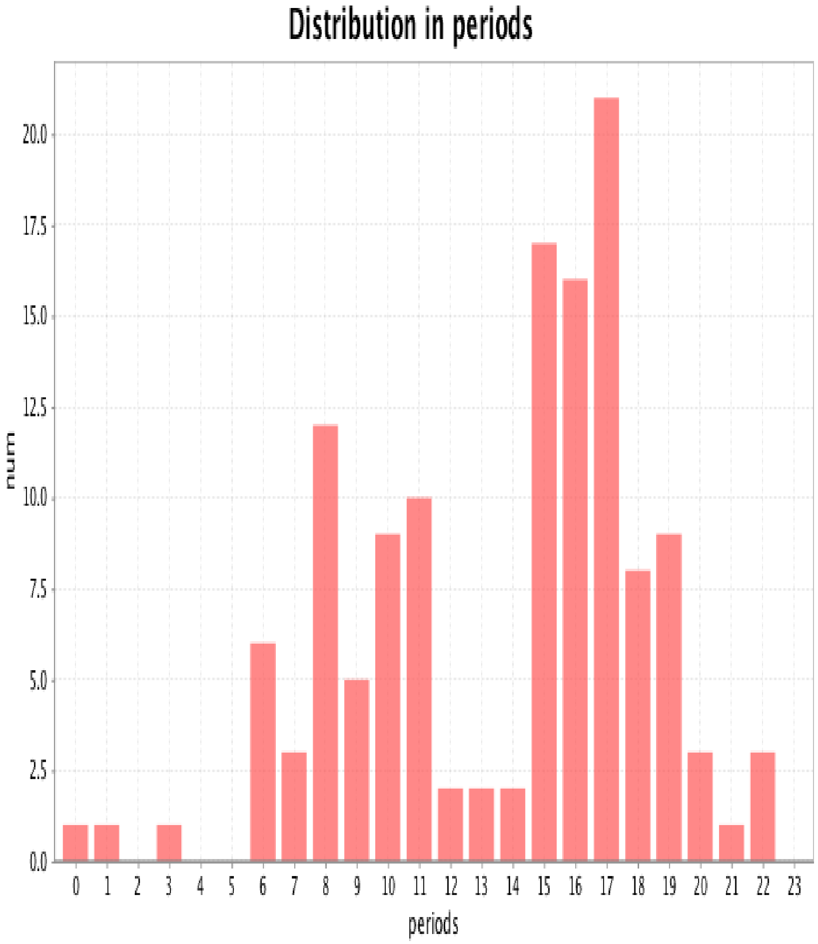




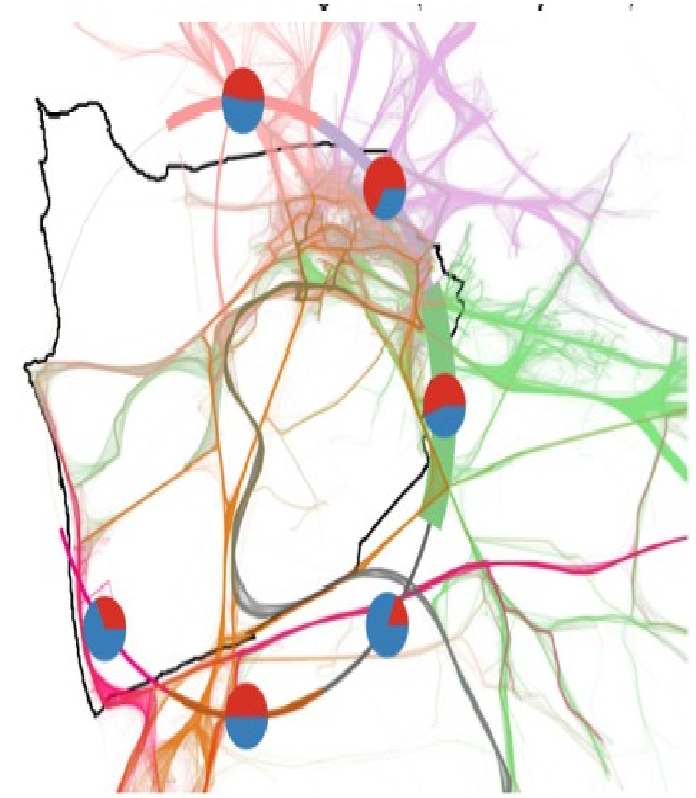


Residential vs Industrial



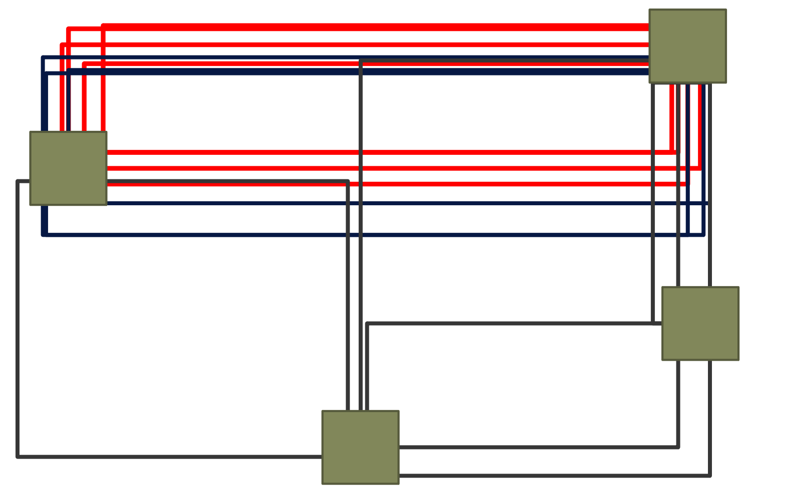


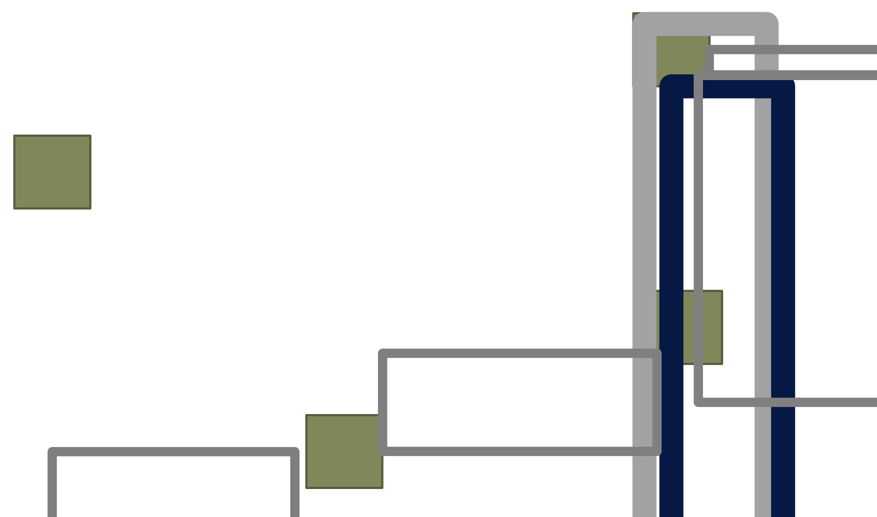
Atlas of Urban Mobility



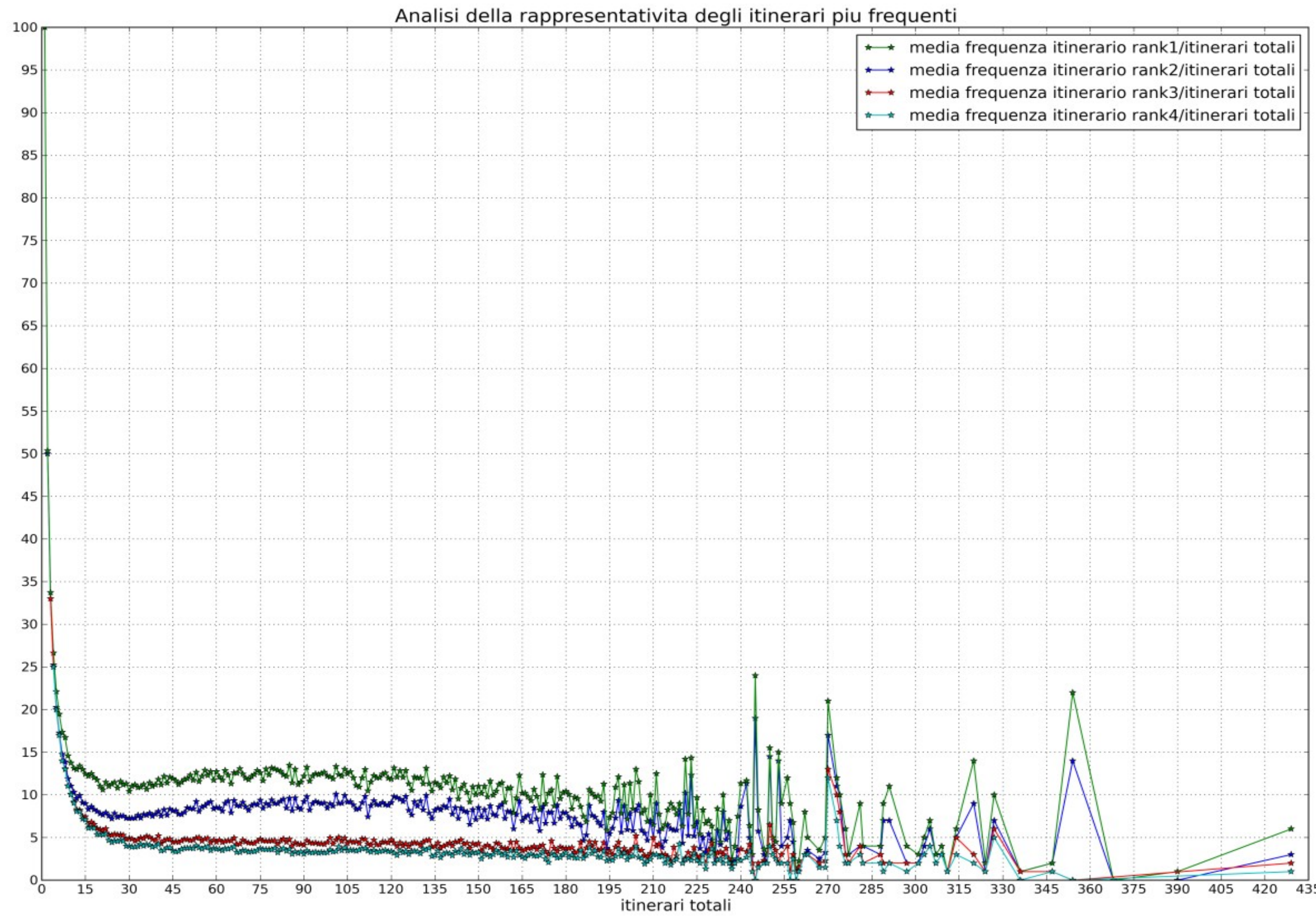
From Profiles to Systematicity Indicator

- Each routine of a profile is associated with a measure of frequency
- Routines are sorted according to their frequency: rank 1, rank 2, rank 3, …
- A minimum frequency threshold allow to distinguish a **systematic** trip from

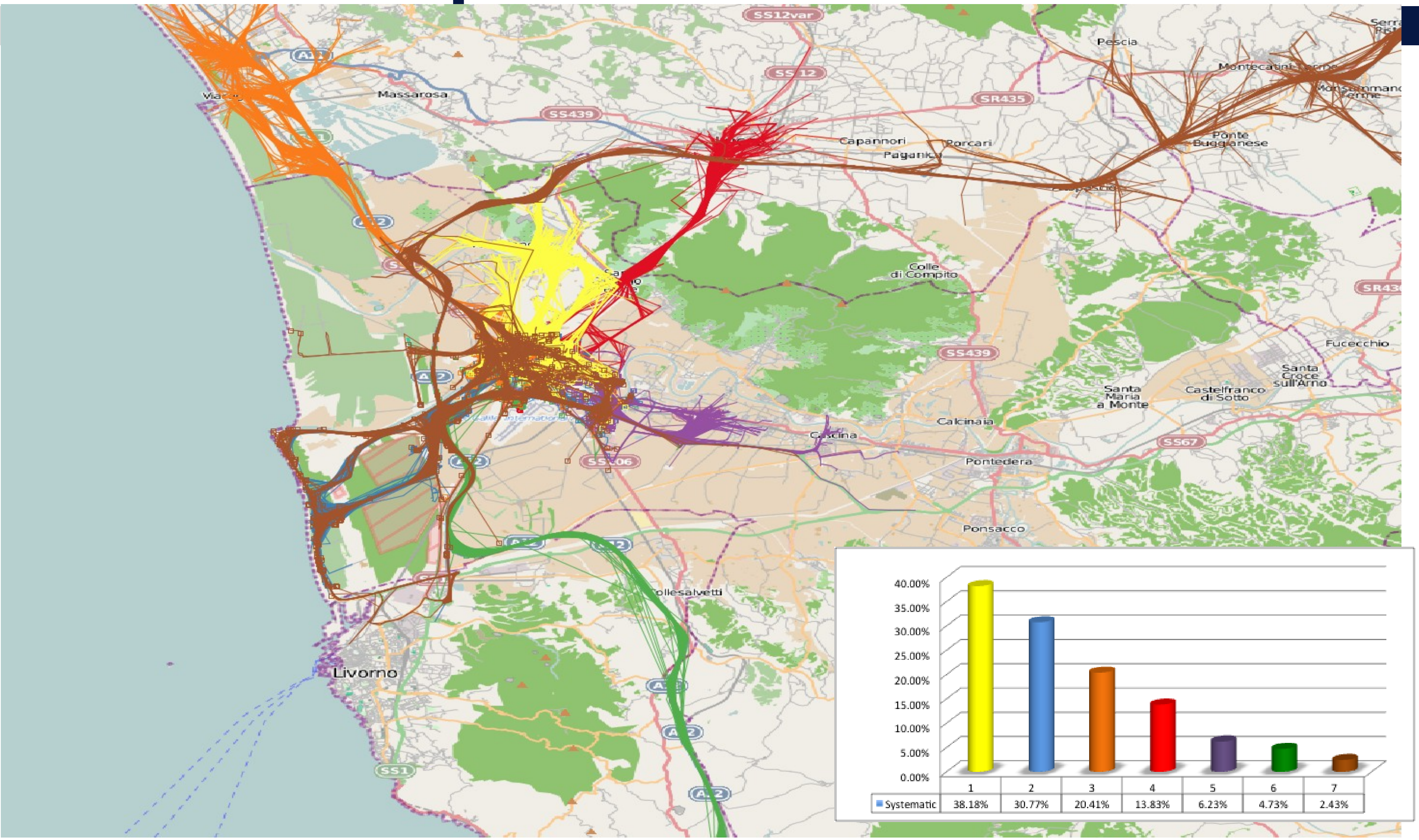




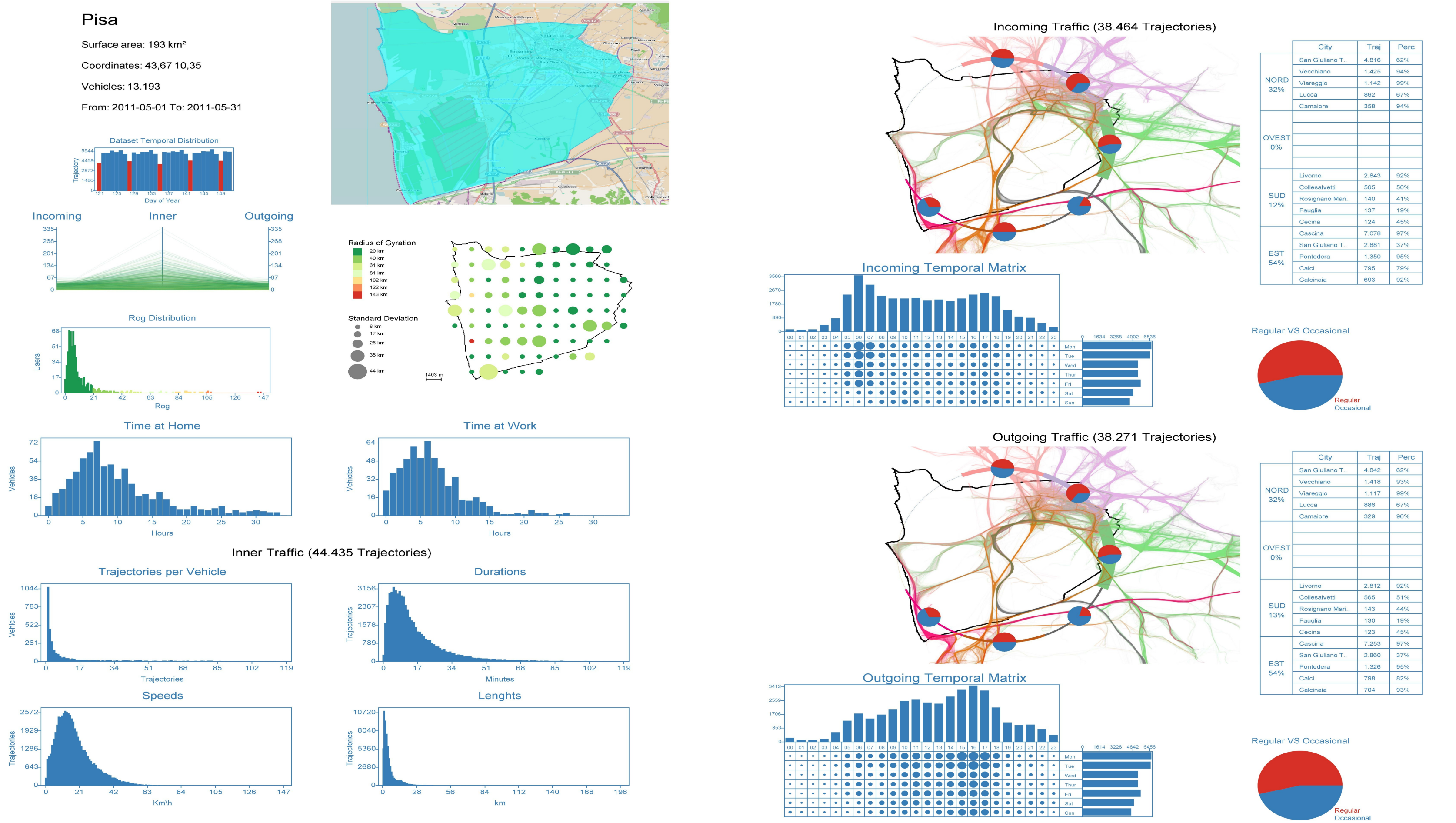
Rapporto Sistematici/Occasionali



Impact of systematic mobility on access patterns

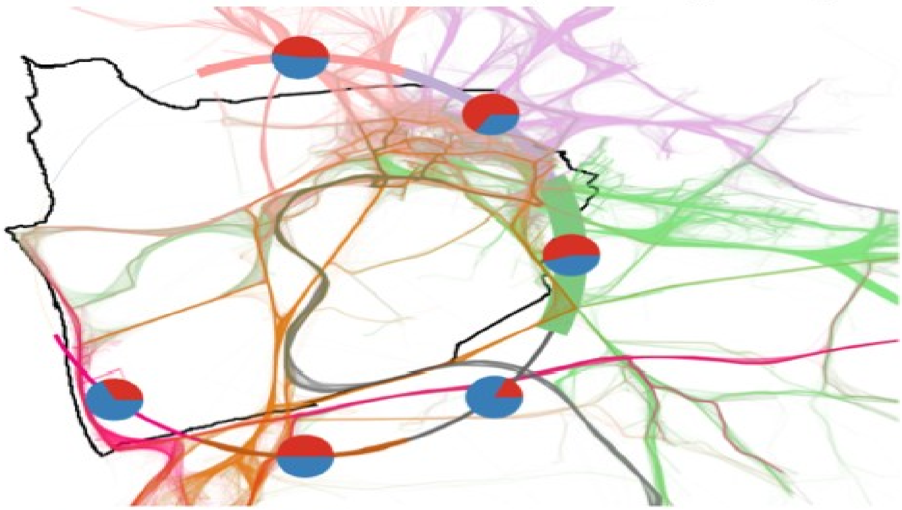


Atlas of Urban Mobility



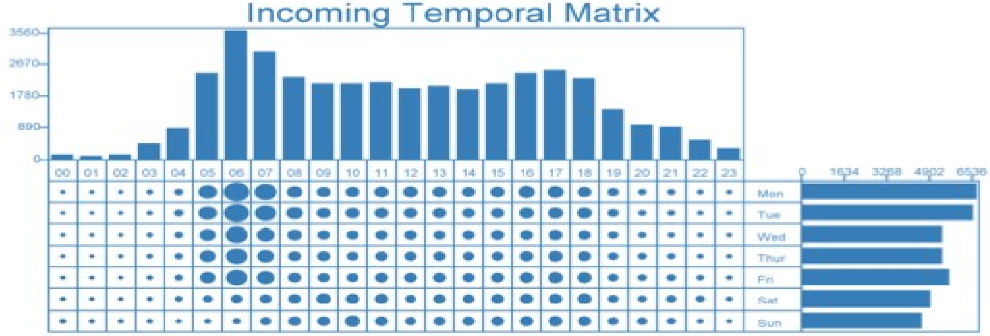
Pisa – Traffico in Ingresso

Incoming Traffic (38.464 Trajectories)

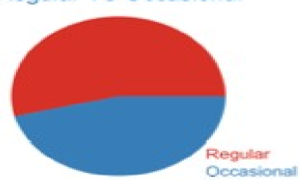


	City	Traj	Perc
NORD 32%	San Giuliano T..	4.816	62%
	Vecchiano	1.425	94%
	Viareggio	1.142	99%
	Lucca	862	67%
	Camaiore	358	94%
OVEST 0%			
SUD 12%	Livorno	2.843	92%
	Collesalvetti	565	50%
	Rosignano Mari..	140	41%
	Fauglia	137	19%
	Cecina	124	45%
EST 54%	Cascina	7.078	97%
	San Giuliano T..	2.881	37%
	Pontedera	1.350	95%
	Calci	795	79%
	Calcinaia	693	92%

Incoming Temporal Matrix

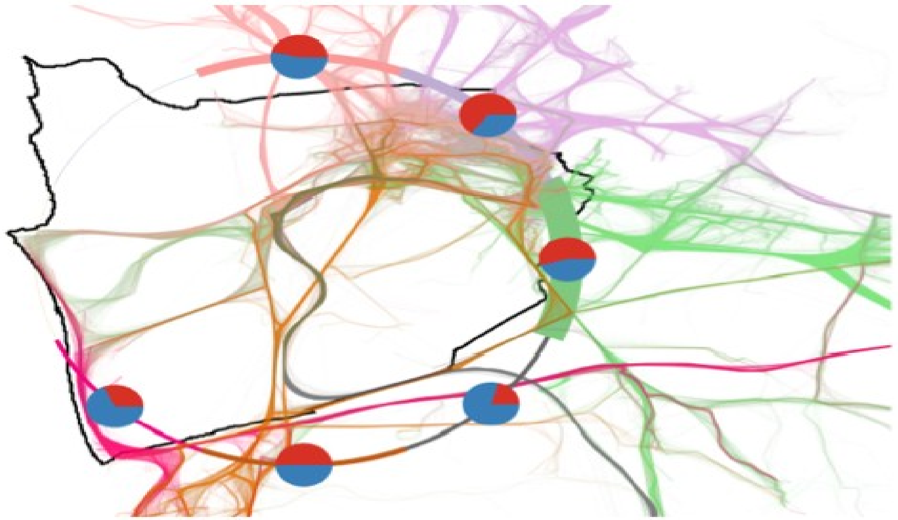


Regular VS Occasional



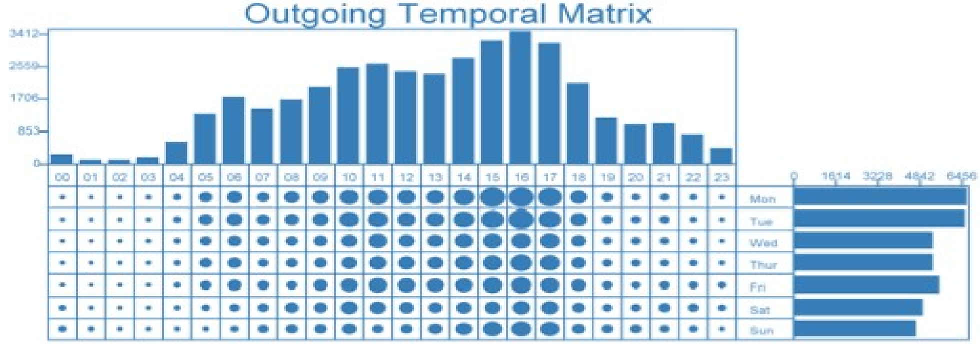
Pisa – Incoming Traffic

Outgoing Traffic (38.271 Trajectories)



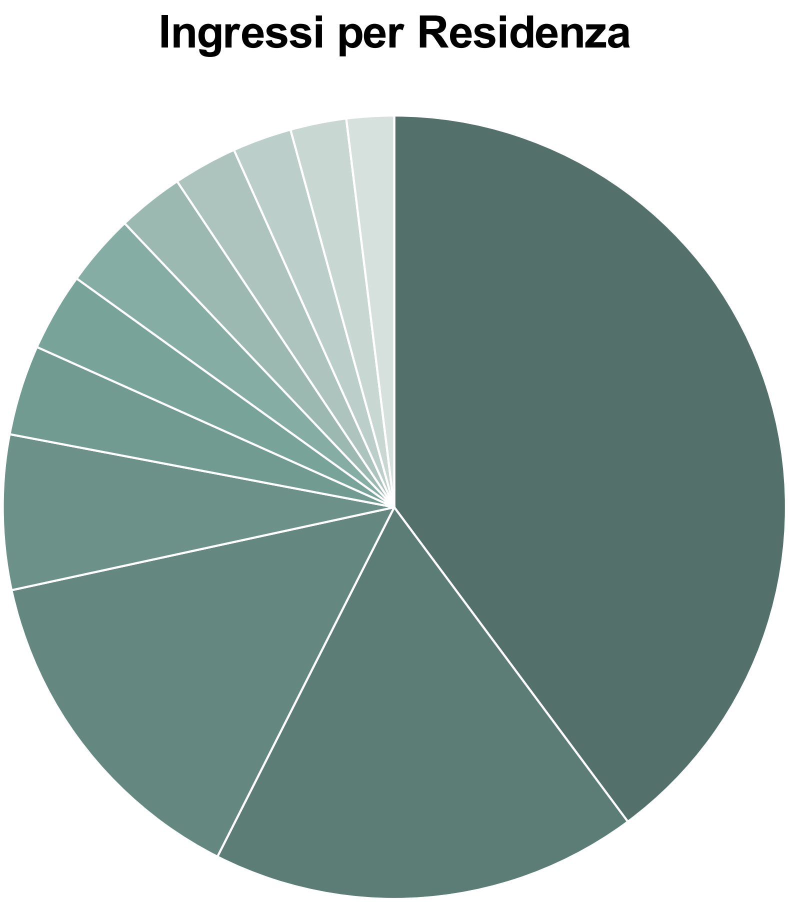
	City	Traj	Perc
NORD 32%	San Giuliano T...	4.842	62%
	Vecchiano	1.418	93%
	Viareggio	1.117	99%
	Lucca	886	67%
	Camaiore	329	96%
OVEST 0%			
SUD 13%	Livorno	2.812	92%
	Collesalvetti	565	51%
	Rosignano Mari...	143	44%
	Fauglia	130	19%
	Cecina	123	45%
EST 54%	Cascina	7.253	97%
	San Giuliano T...	2.860	37%
	Pontedera	1.326	95%
	Calci	798	82%
	Calcinaia	704	93%

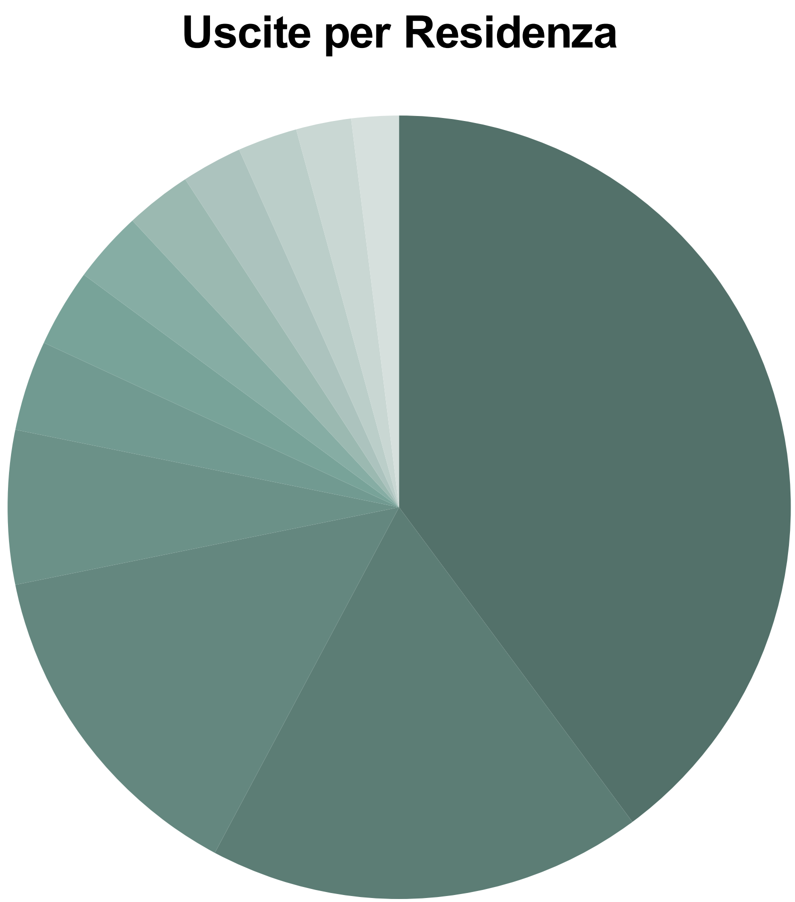
Outgoing Temporal Matrix



Regular VS Occasional

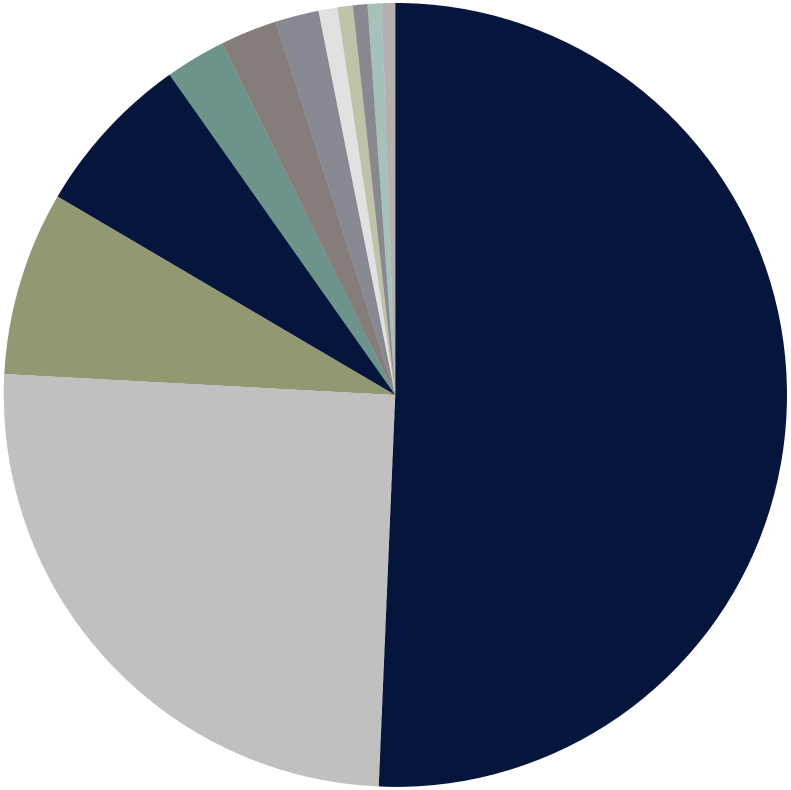
Trajectories by residence



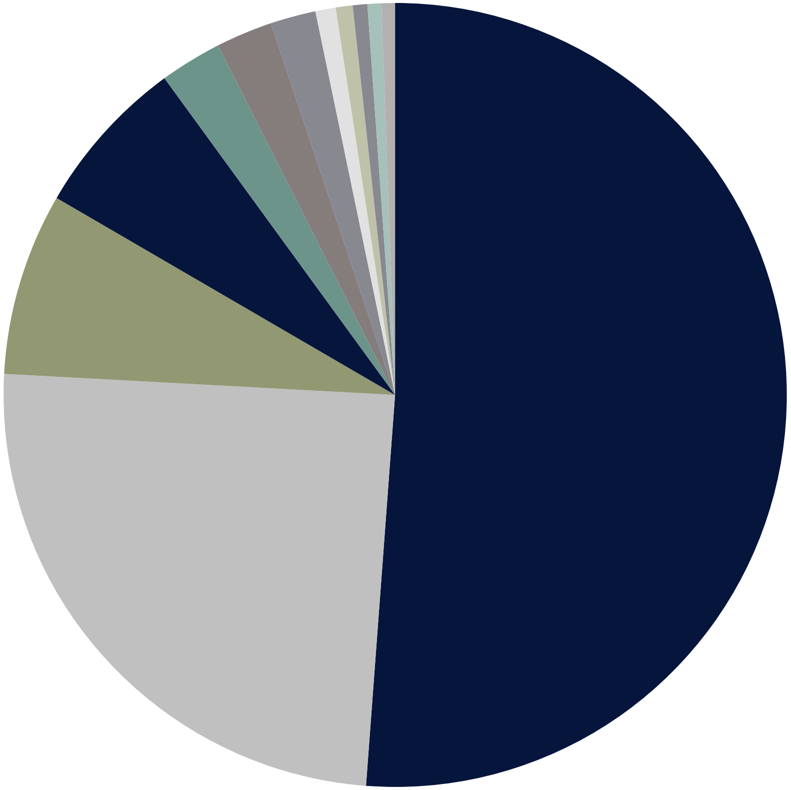


Cosa succede a San Giuliano Terme?

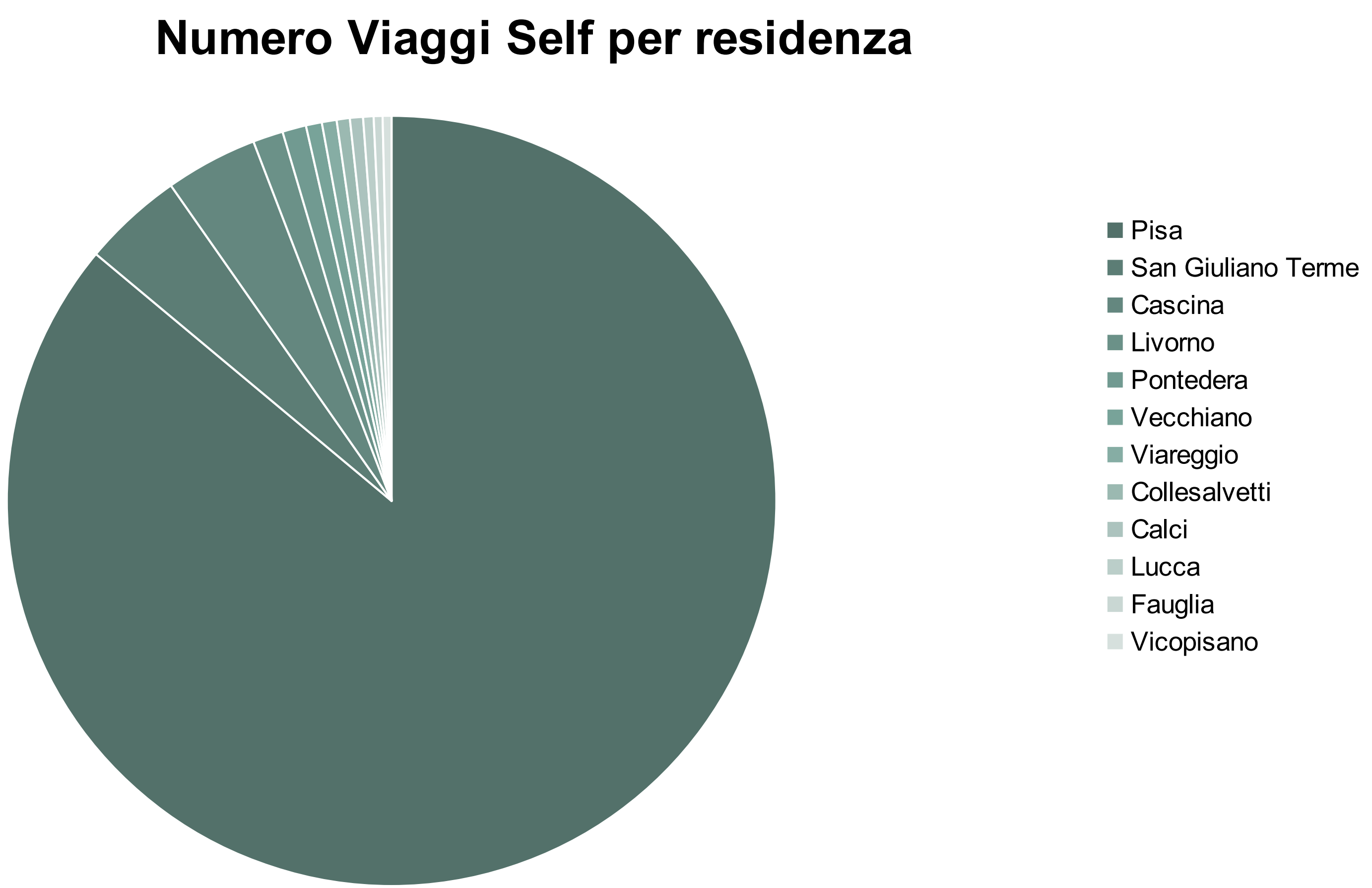
Ingressi per Residenza



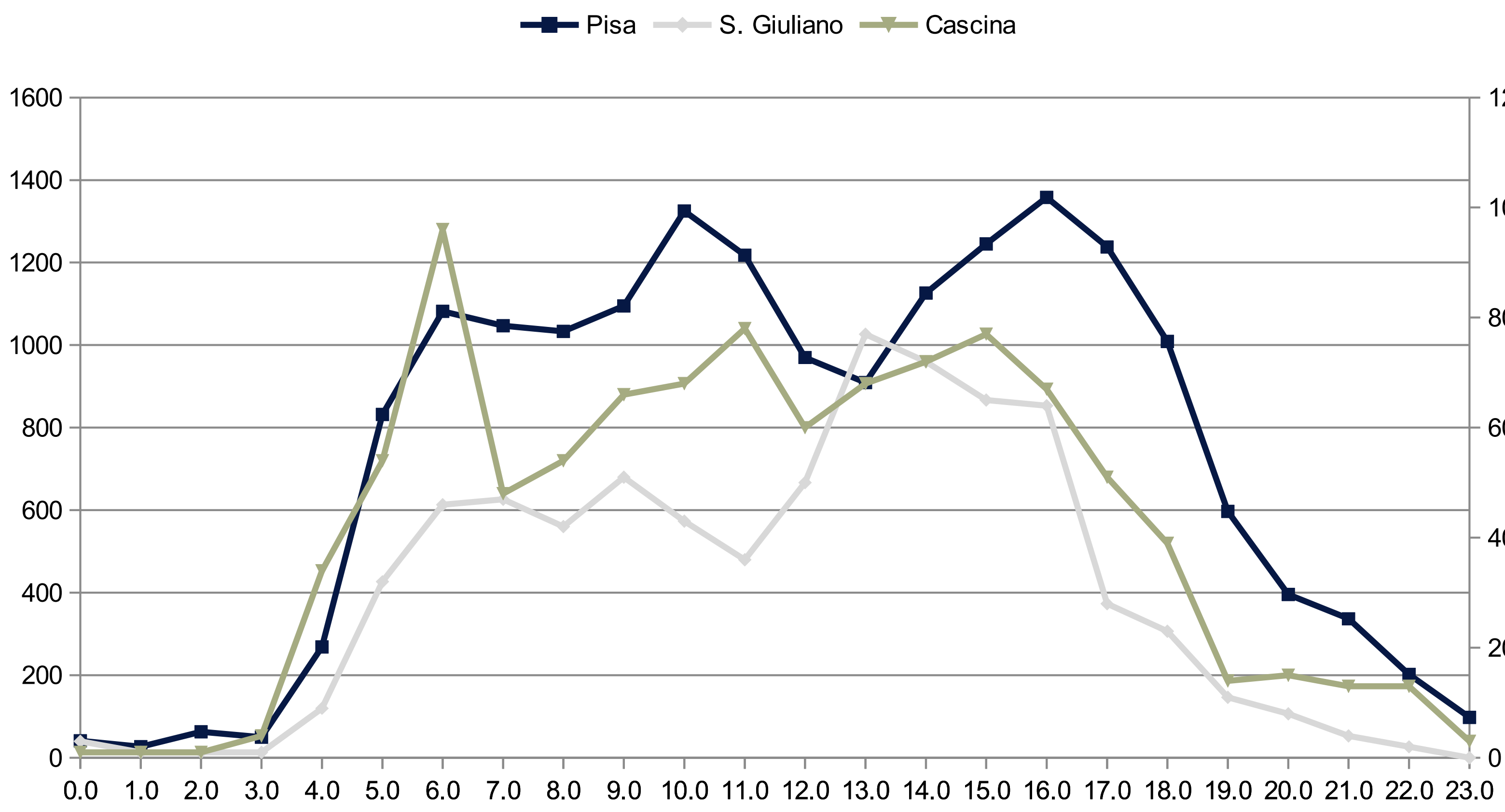
Uscite per Residenza



Internal trajectories in Pisa



Trip distribution per day

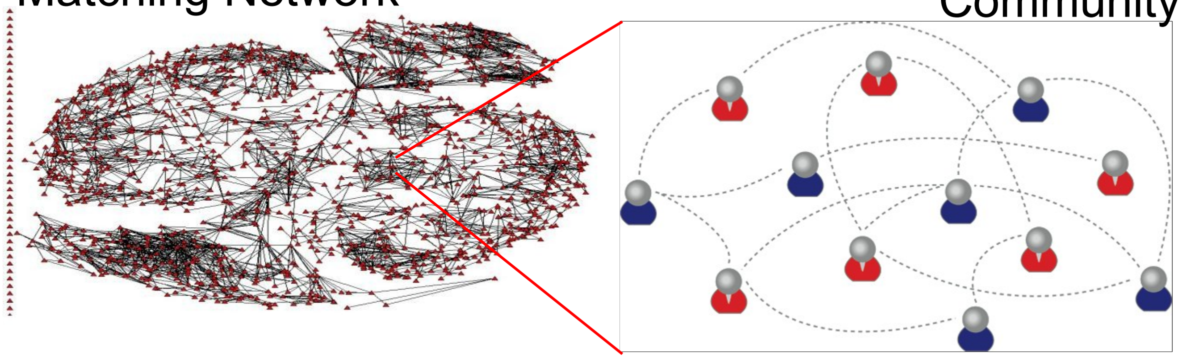


What-if scenarios

Studying proactive car pooling

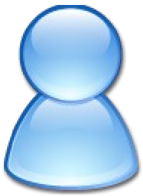
Matching Network

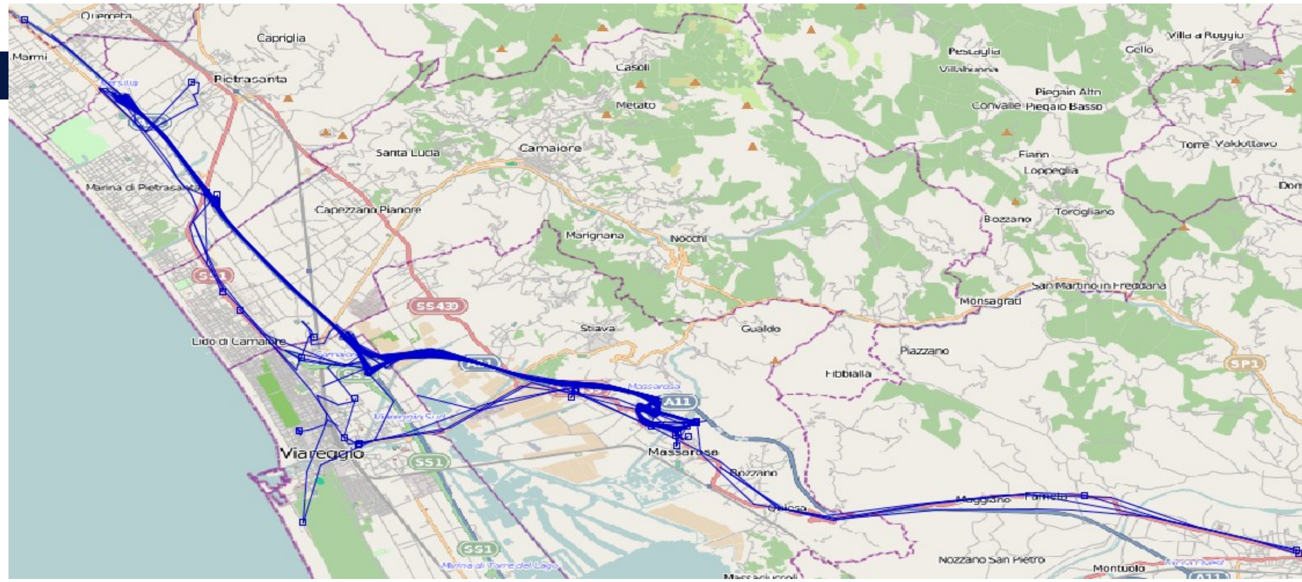
Community

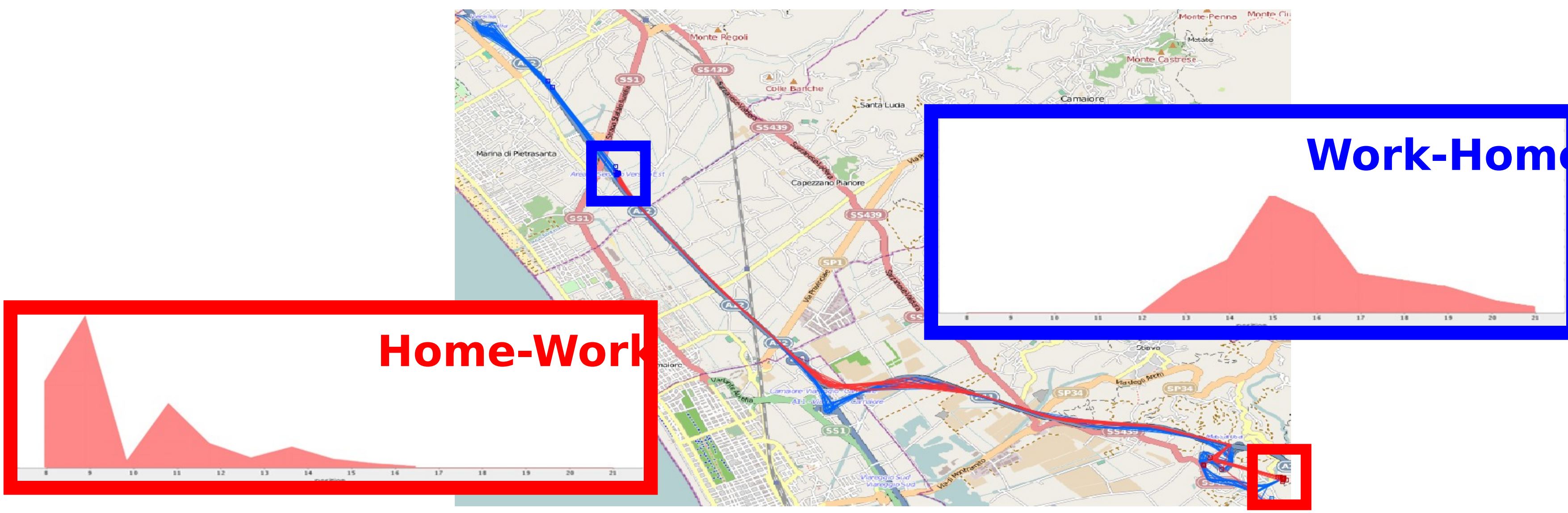


Suggestion

Discovering individual systematic movements







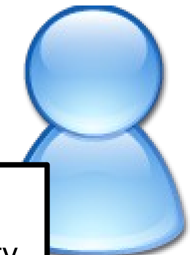
Mobility profile matching

User A (as driver)

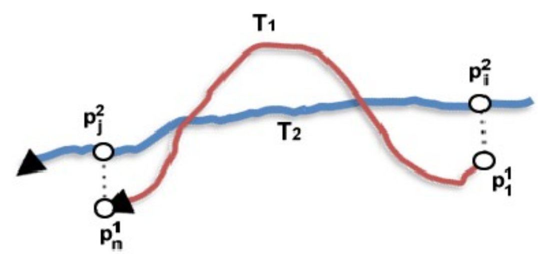
Mobility Profile

Spatio Temporal Routing matches

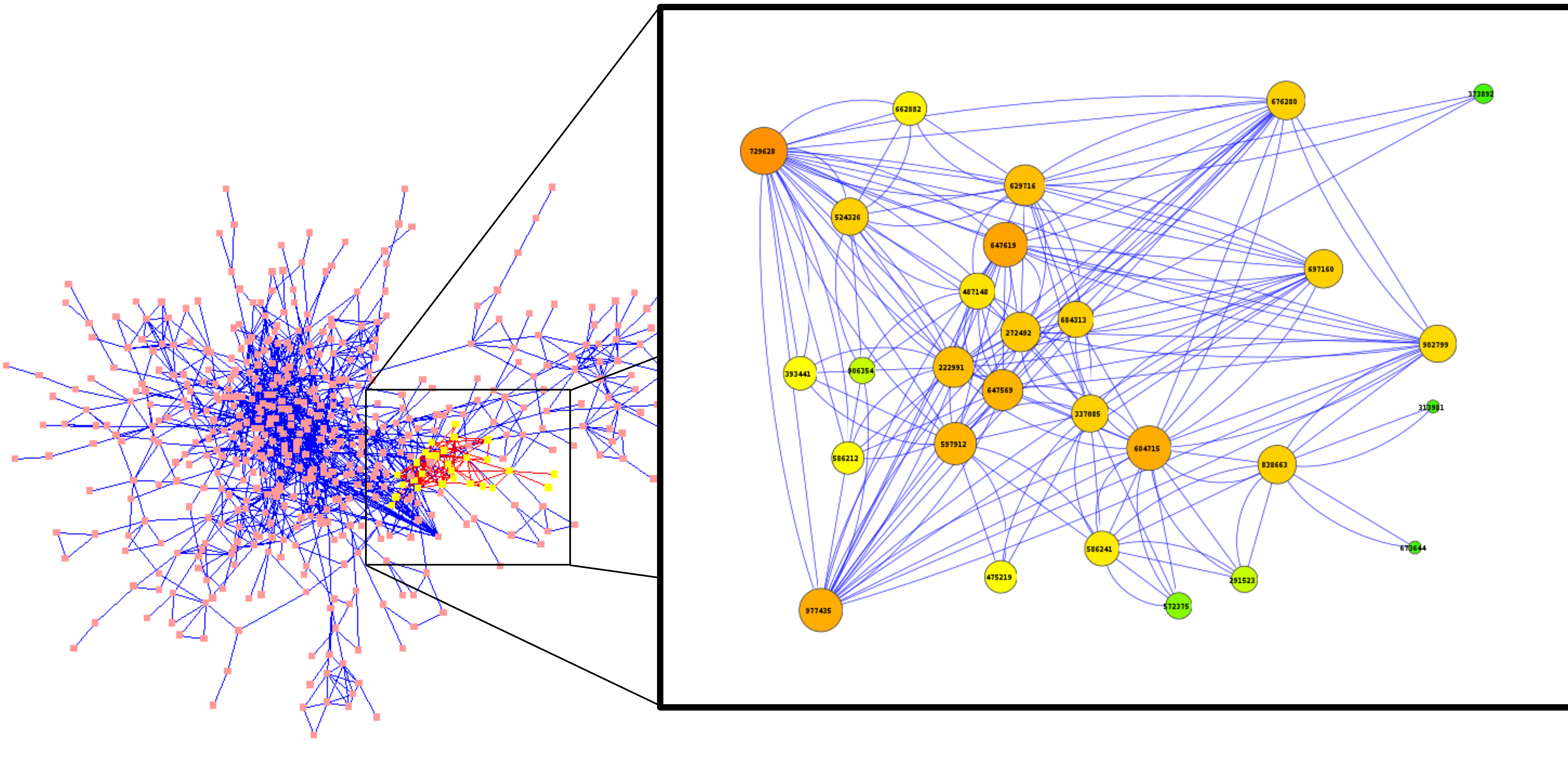
User B (as passenger)



Mobility Profile

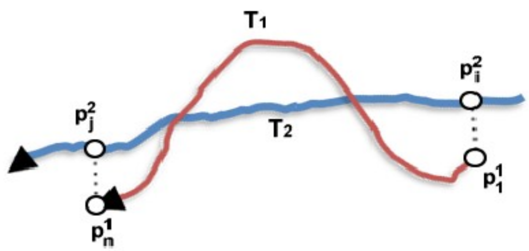


A can serve most of the routines of B the match is suggested.

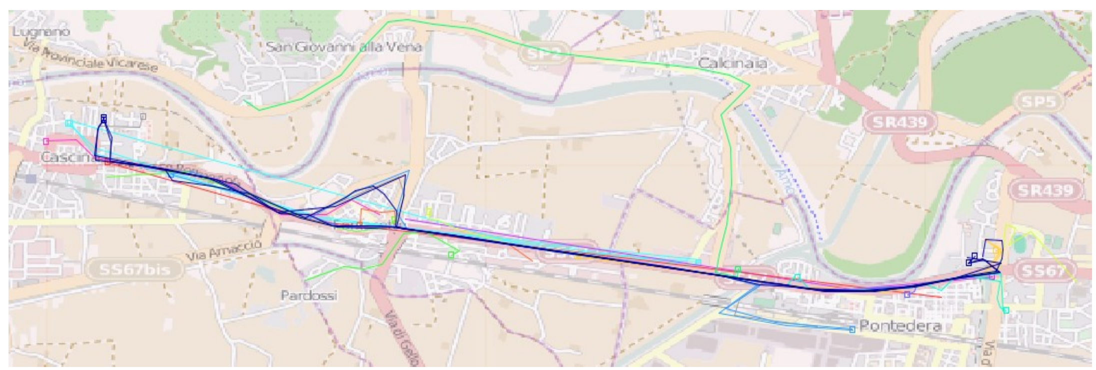


662882
676280
729628
629716
524326
647619
697160
487148
684313
272492
902799
393441
906354
222991
647569
337085
604715
838663
597912
586212
586241
475219
291523
572375
977435

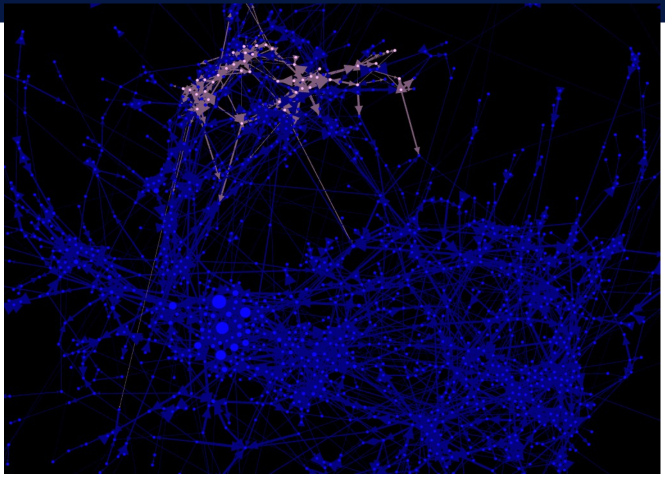
Carpooling Network

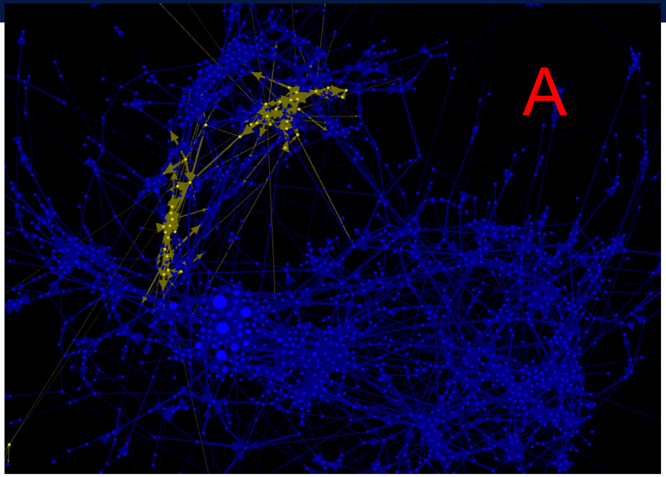


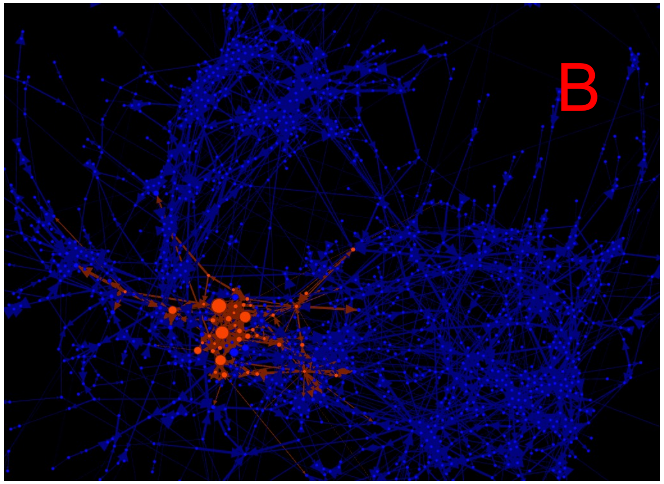
Th space = 1800s
Th time = 1000m
U ---w--> V
If U could take lifts from V

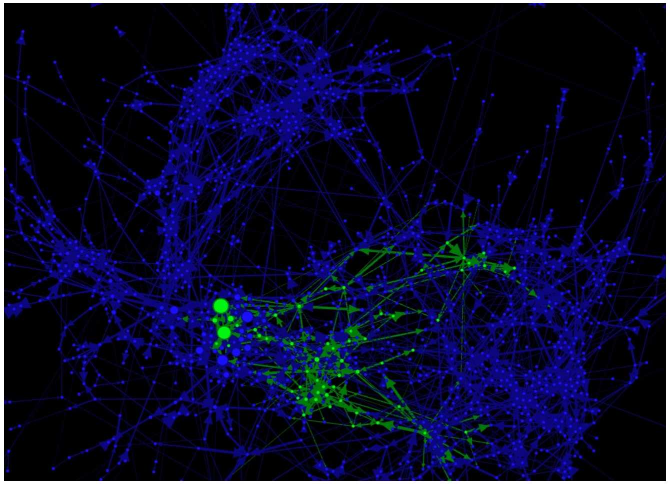


Carpooling Network Pisa - Communities

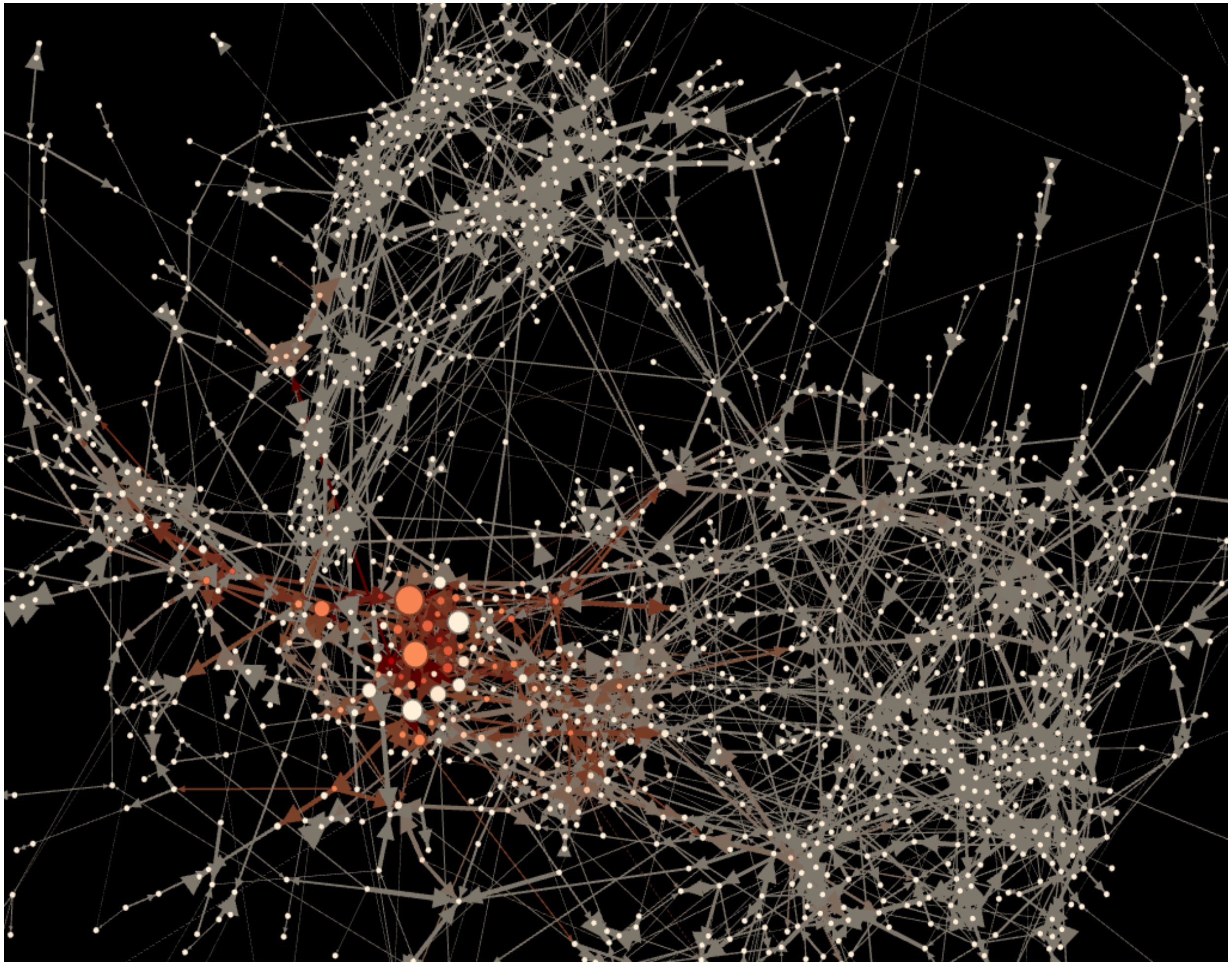




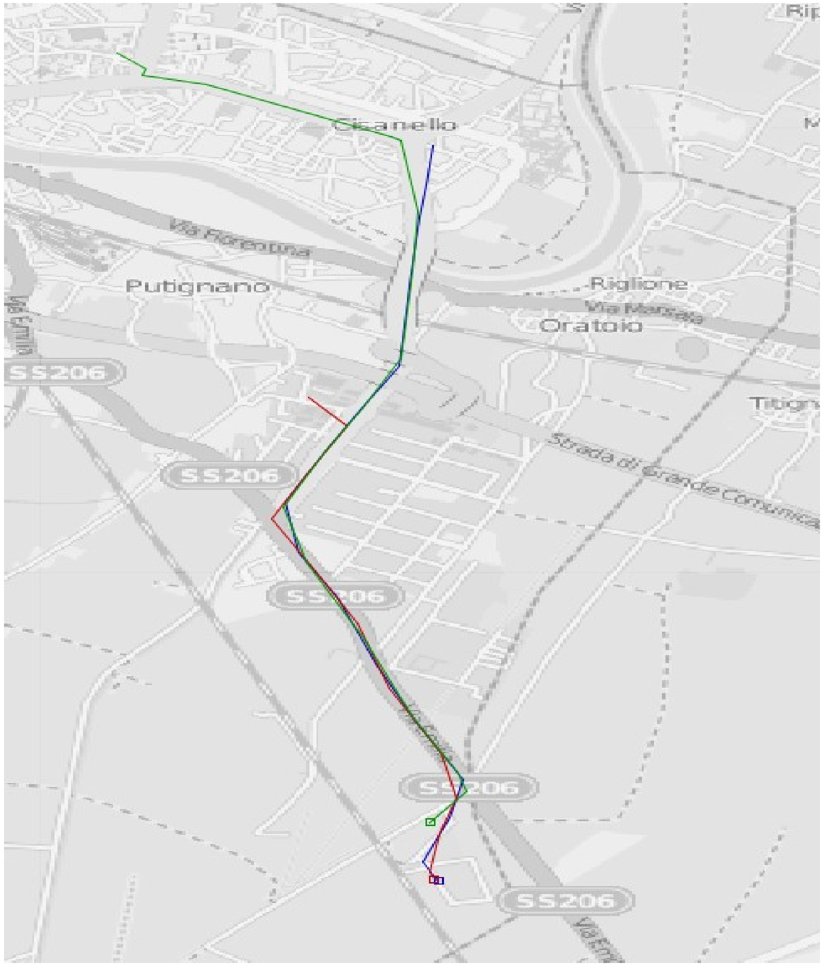




Carpooling Network Pisa



Service: Montacchiello (Car Pooling?)



- *Traj Blu*
  - DT: 06:46:53
- *Traj Red*
  - DT: 11:52:06
- *Traj Green*
  - DT: 06:51:41
- *Blu can give a ride to Green*

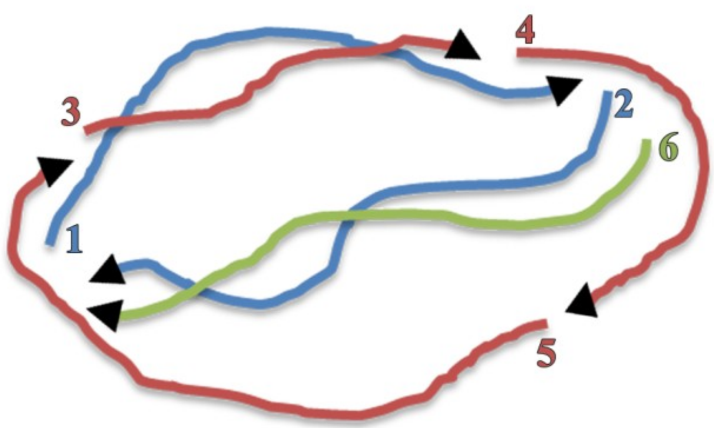
Application: Car pooling

Pro-active suggestions of sharing rides opportunities without the need for the user to explicitly specify the trips of interest.

Matching two routines:

$$
\begin{aligned}
contained(T_1, T_2, th^{walking}_{distance}, th^{wasting}_{time}) \equiv\ & \exists i, j \in \mathcal{N} \mid \\
& 0 < i \leq j \leq m \wedge \\
& Dist(p^1_1, p^2_i) + Dist(p^1_n, p^2_j) \leq th^{walking}_{distance} \wedge \\
& Dur(p^1_1, p^2_i) + Dur(p^1_n, p^2_j) \leq th^{wasting}_{time}
\end{aligned}
$$

mobility profile share ability:



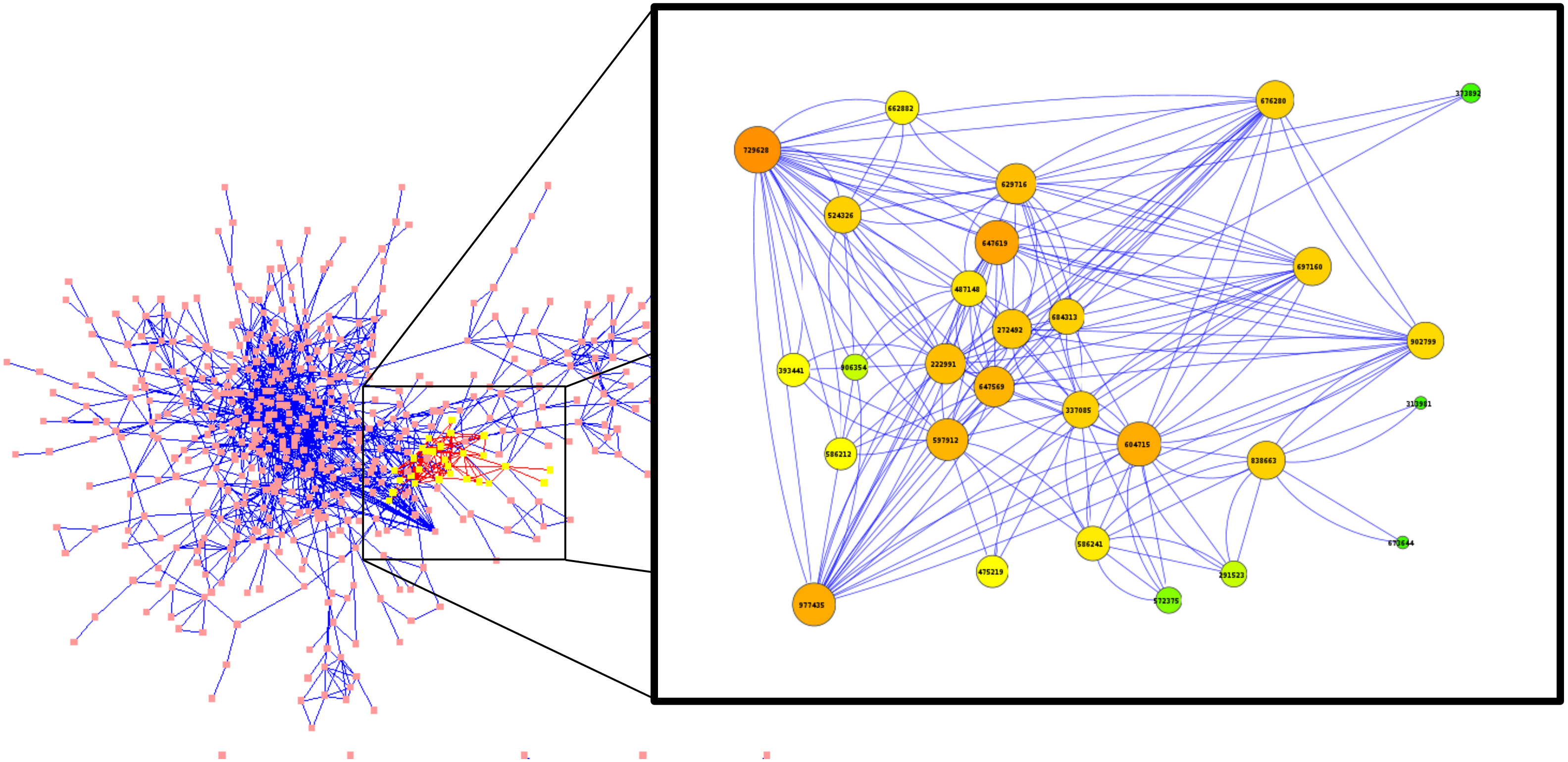
	1	2	3	4	5	6
1	-	-	F	F	F	F
2	-	-	F	F	F	**T**
3	**T**	F	-	-	-	F
4	F	F	-	-	-	F
5	F	F	-	-	-	F
6	F	**T**	F	F	F	-

	-	0	1/2
	1/3	-	0
	1	0	-

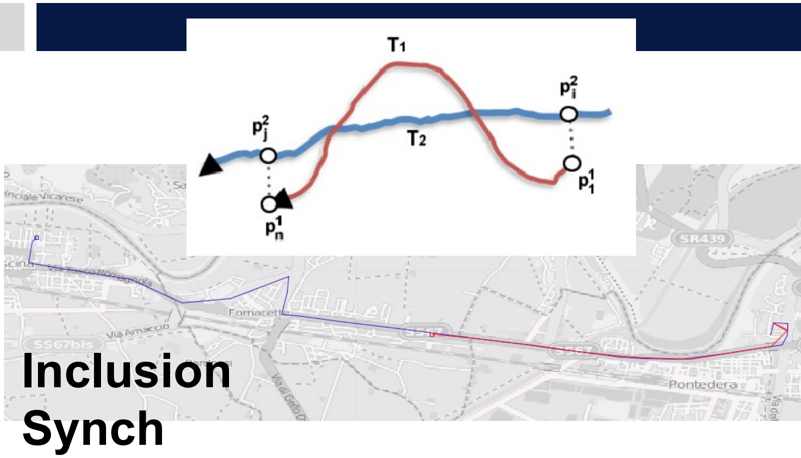
mobility profiles $\tilde{T}_1$ *and* $\tilde{T}_2$

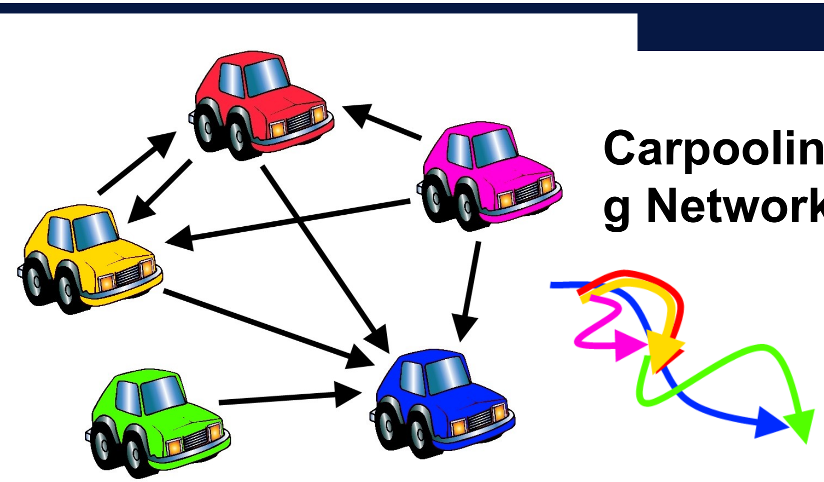
$$
profileShare(\tilde{T}_1, \tilde{T}_2, th^{walking}_{distance}, th^{wasting}_{time}) = \frac{\left| \left\{ p \in \tilde{T}_1 \mid \exists q \in \tilde{T}_2 . Share(p, q, th^{walking}_{distance}, th^{wasting}_{time}) \right\} \right|}{|\ \tilde{T}_1\ |}
$$

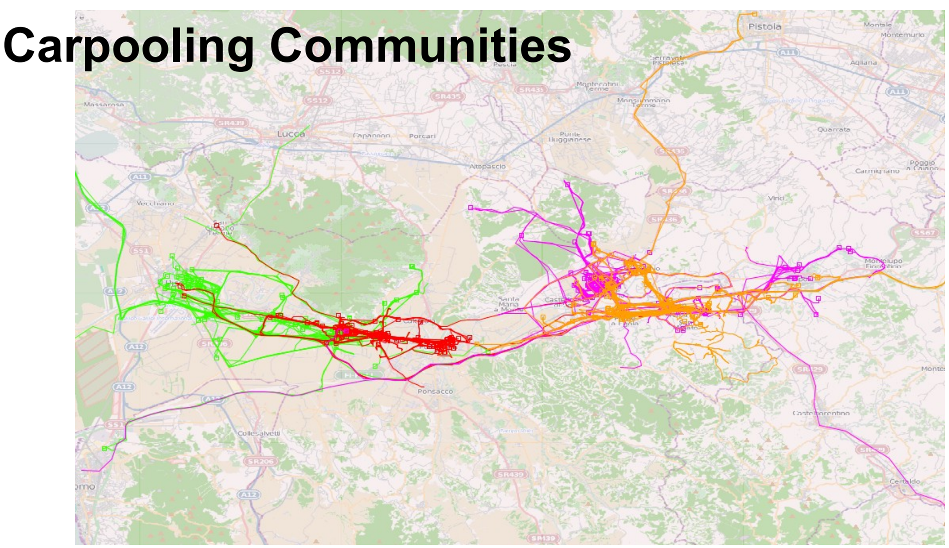
Communities of users

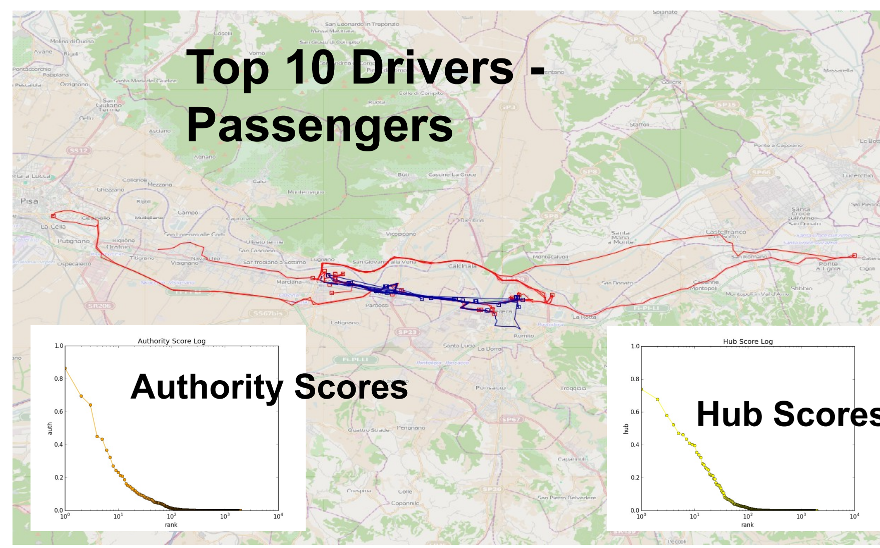


Car Pooling


T1
T2
Inclusion Synch

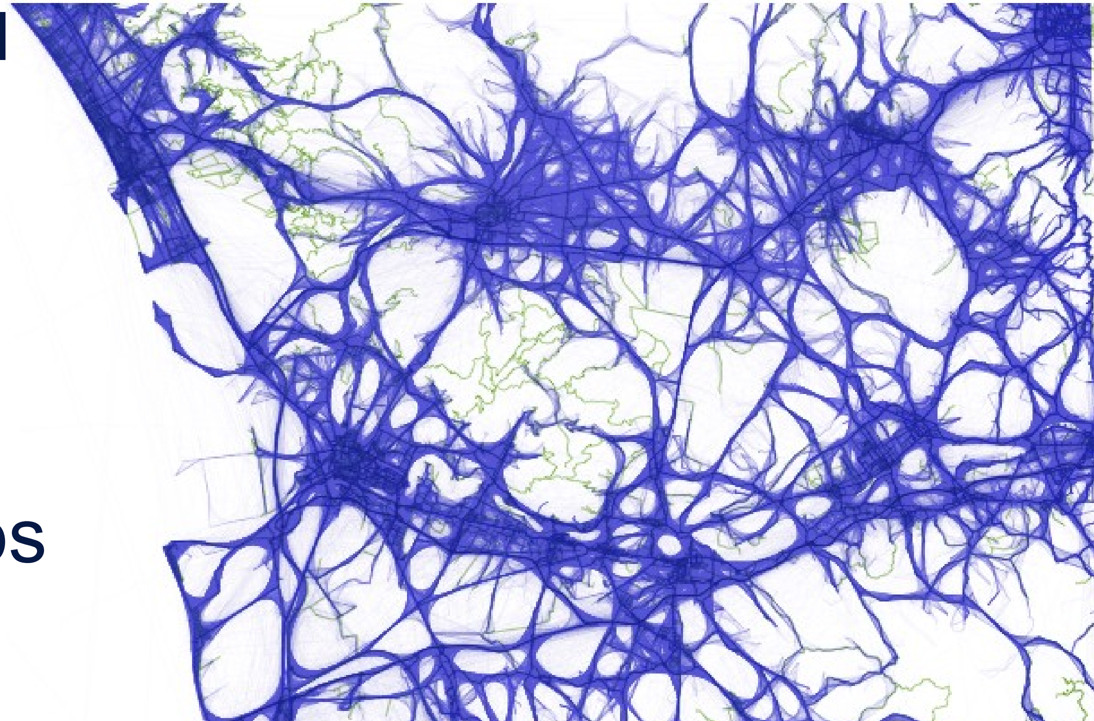

Carpooling Network

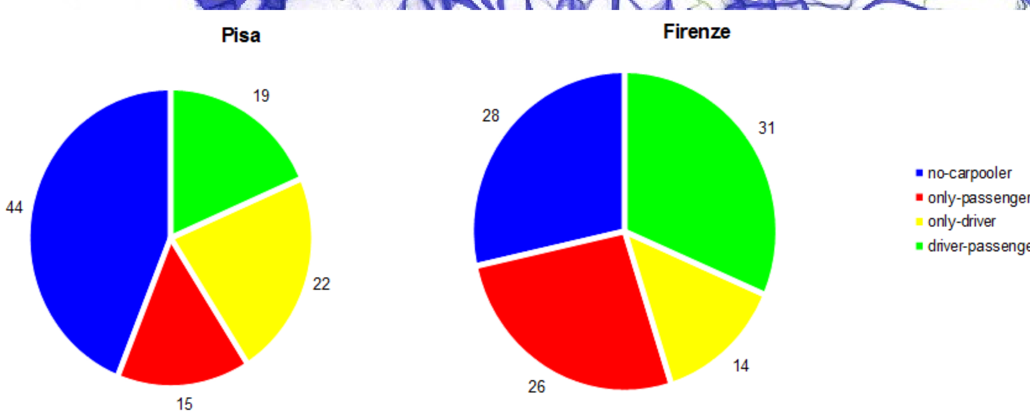

Carpooling Communities


Top 10 Drivers - Passengers
Authority Scores
Hub Scores

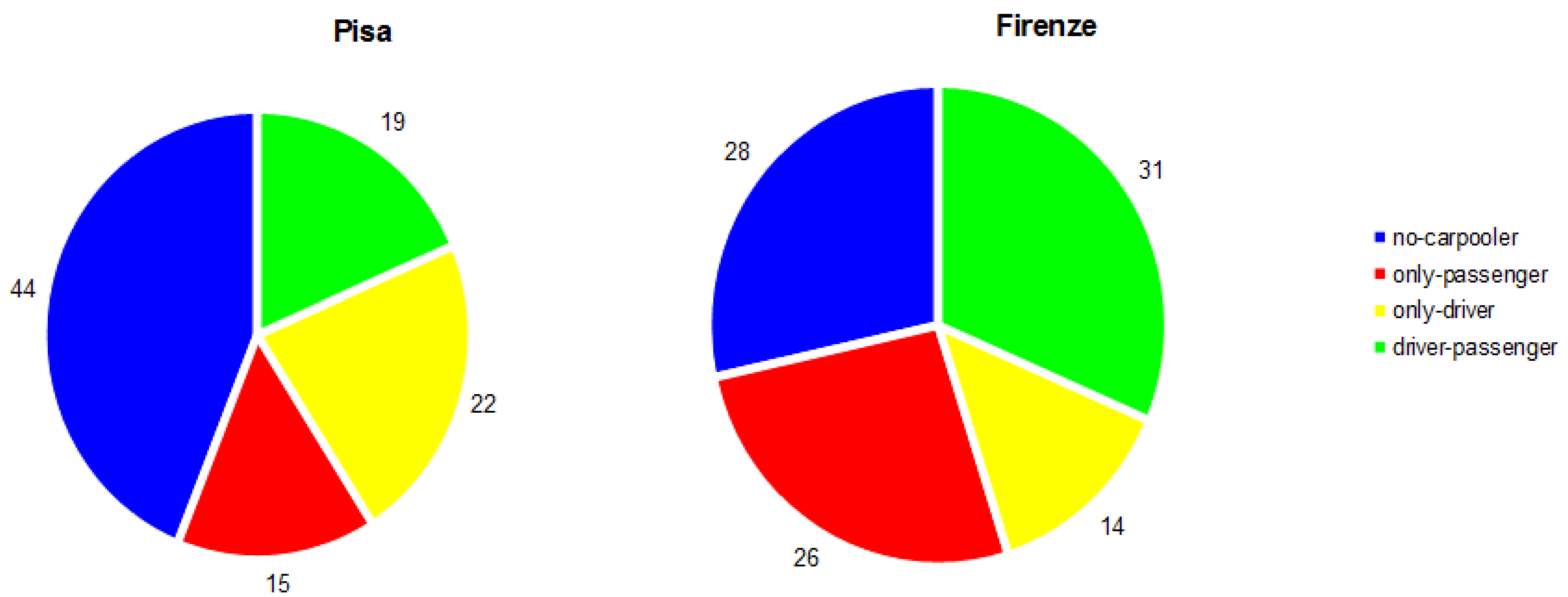
OctoPisa Carpooling potential

- 38416 vehicles
- 1,449,258 trips
- % di systematic trips
- % di matching trips
- % highly successful matching trips
- Saved trips
- Saved Kms





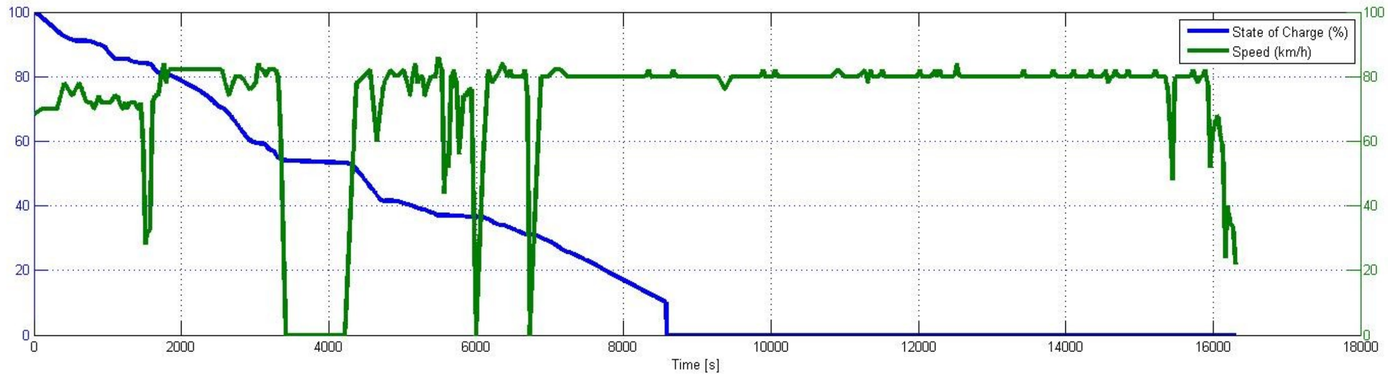
Carpooling propensity of Pisa & Florence



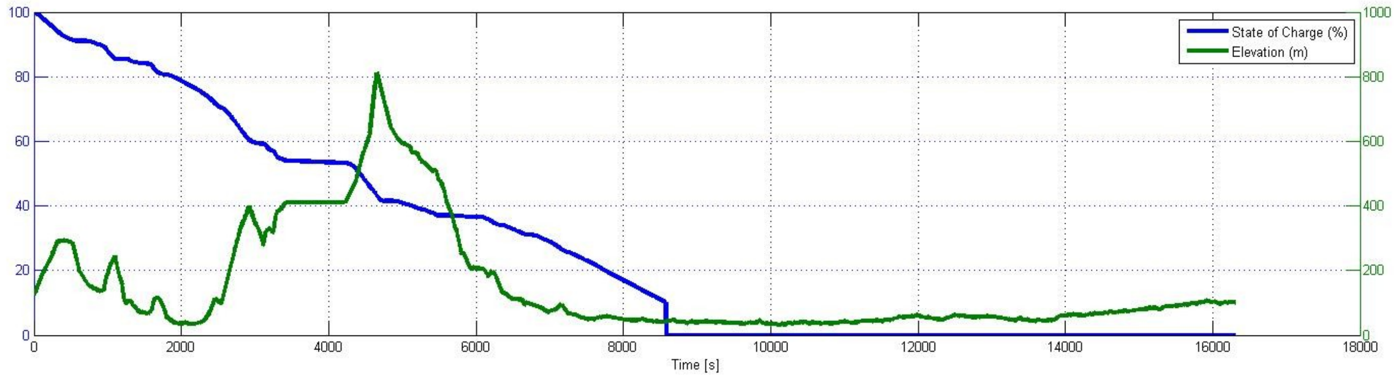
Car pooling potential

Electrificability

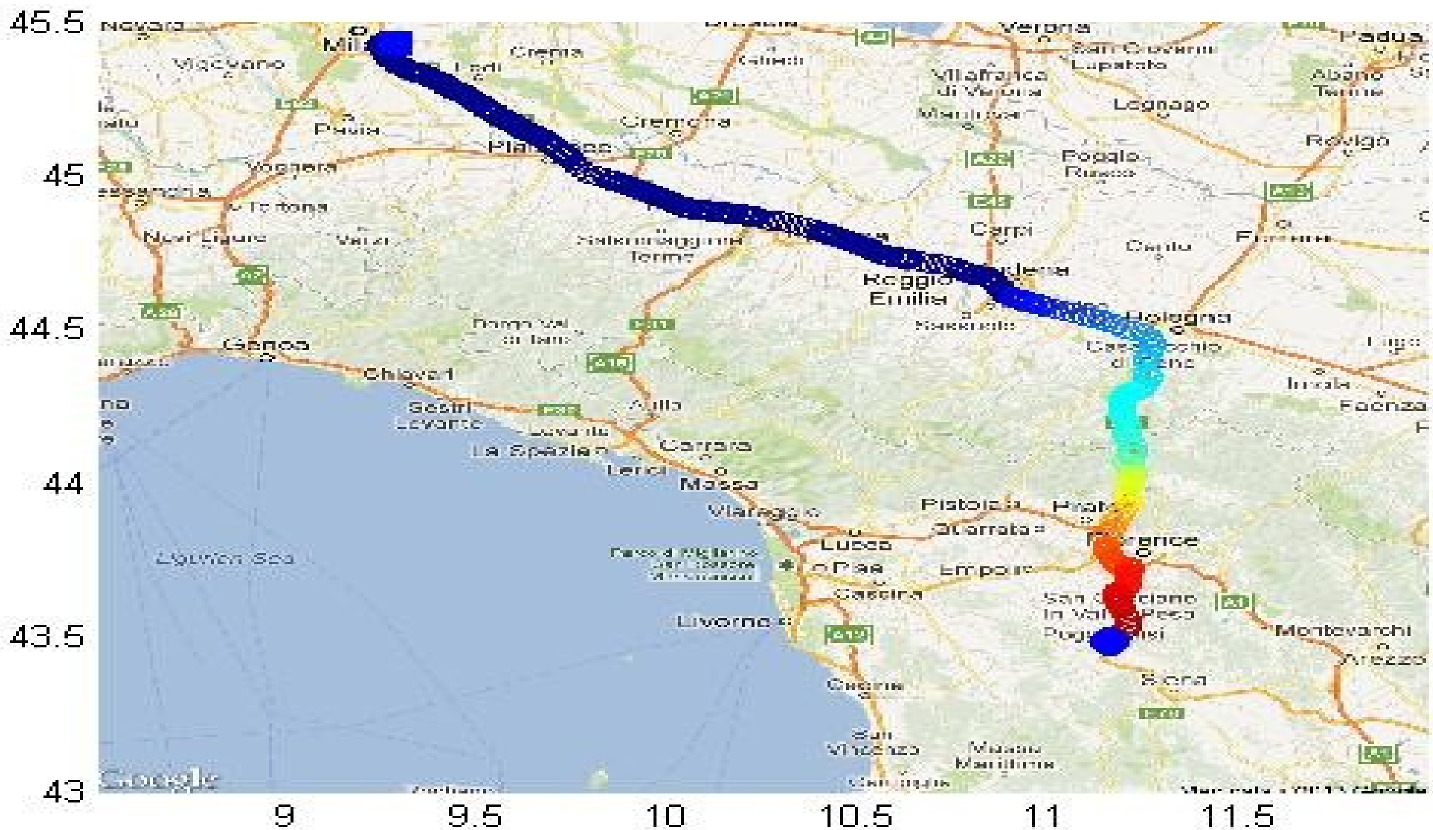
Joint work with UPM



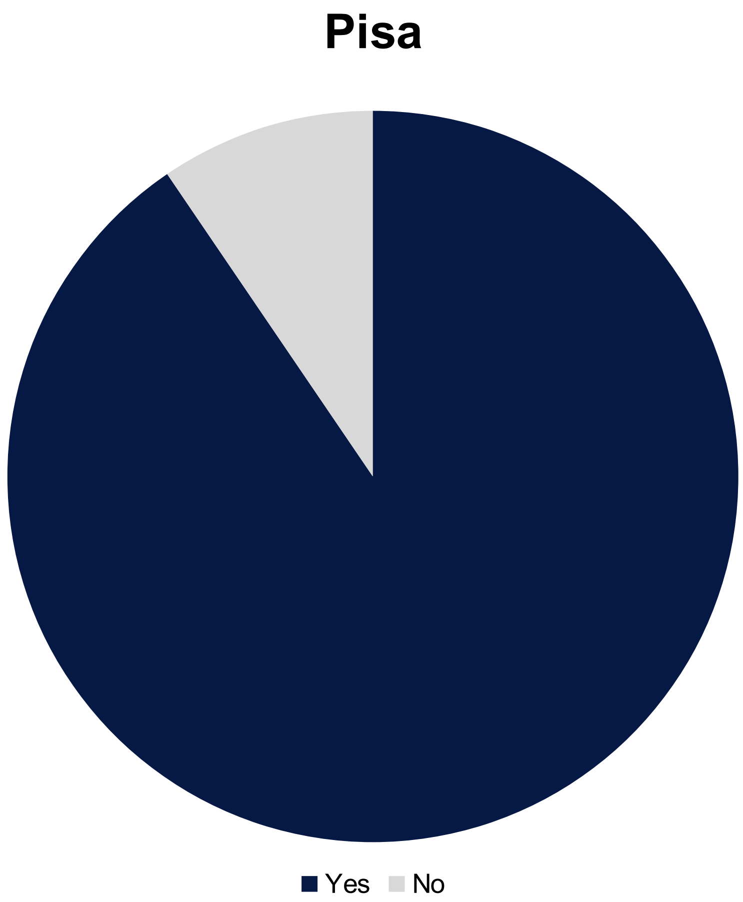
State of Charge (%)
Speed (km/h)
Time [s]

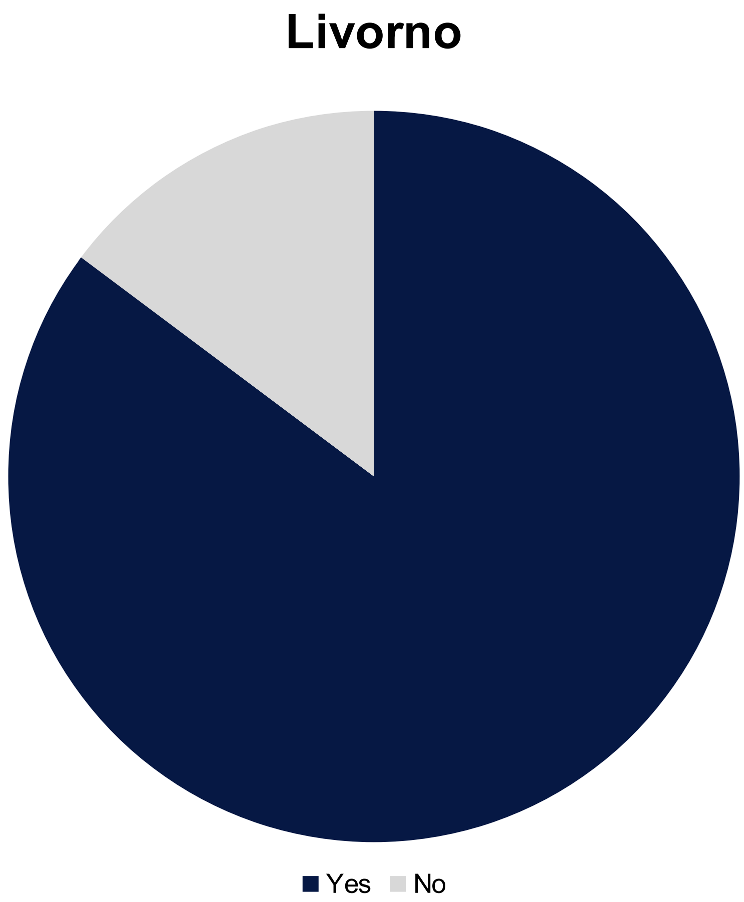


State of Charge (%)
Elevation (m)
Time [s]


45.5
45
44.5
44
43.5
43
9
9.5
10
10.5
11
11.5
Verona
Padua
Vigevano
Lodi
Cremona
Legnago
Pavia
Rovigo
Voghera
Carpi
Reggio
Emilia
Bologna
Genoa
Sestri
Levante
La Spezia
Carrara
Massa
Viareggio
Pistoia
Prato
Lucca
Empoli
Cascina
Livorno
Montevarchi
Faenza
Siena
Ligurian Sea
Google

Electrifiability index





Electrifiability index