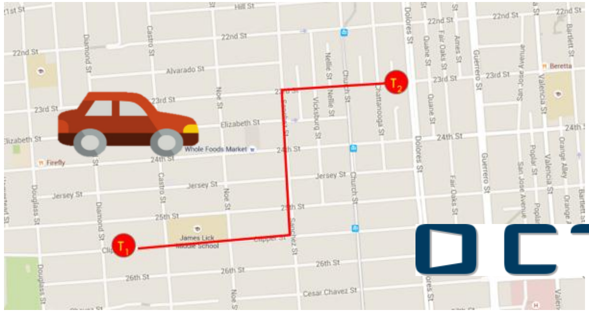
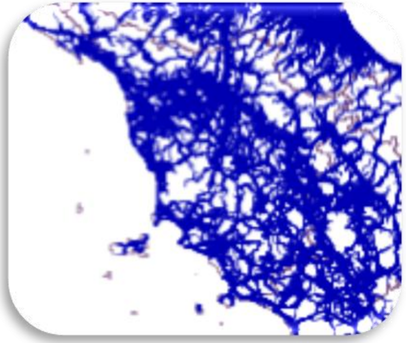
Strumenti di studio data-driven per la mobilità toscana

Mirco Nanni
mirco.nanni@isti.cnr.it

Sorgenti informative

- **Tracce GPS di mobilità veicolare**





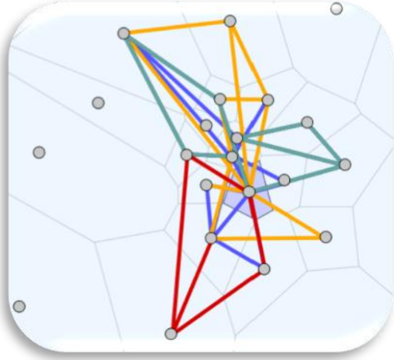
Feb-Mar 2014
Tempo

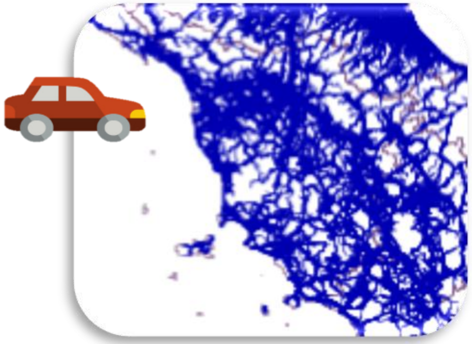
150K
Numero di veicoli

12 mln
Viaggi

- **Dati di telefonia (CDR)**

d	Time start	Cell start	Cell end	Duration
294595	"2014-02-20 14:24:58"	"PI010U2"	"PI010U1"	48
10294595	"2014-02-20 18:50:22"	"PI002G1"	"PI010U2"	78
10294595	"2014-02-21 09:19:51"	"PI080G1"	"PI016G1"	357





1. MFAD

Mobility Functional Area Discovery

Problem definition

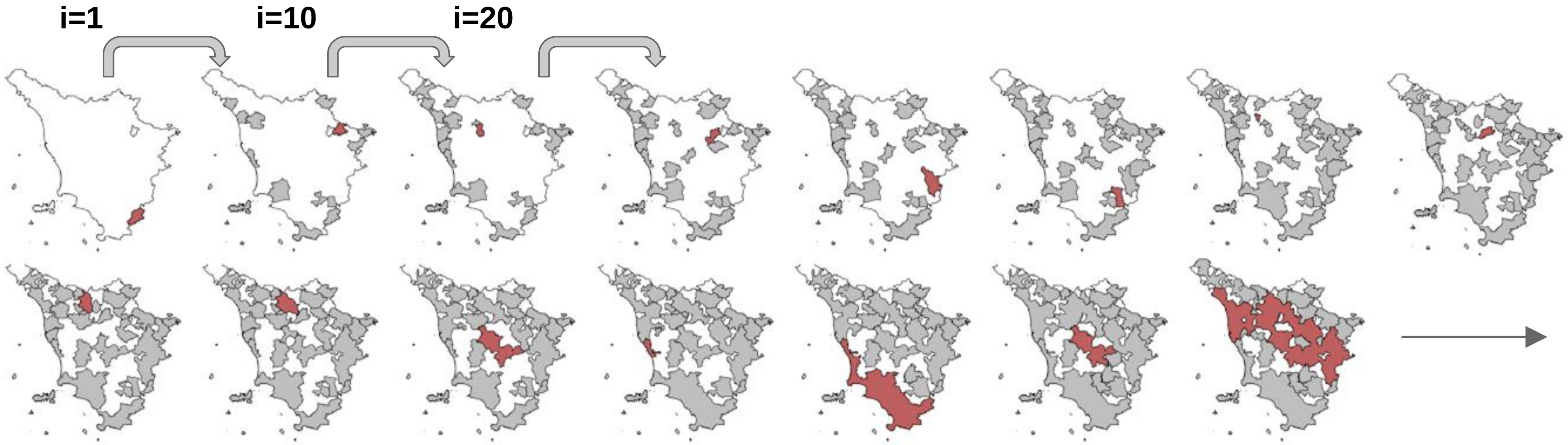
- **Q1**: I dati di mobilità possono suggerirci quali aree funzionali esistono sul territorio?
- **Q2**: Possono aiutarci a comprendere quale tipo di struttura (monocentrica vs. policentrica) hanno?

Proposta

- **Identificare i confini urbani “naturali” che emergono dai dati**
  - Criterio: i comuni di un’area hanno molto più traffico interno (all’area) che da/verso l’esterno

- **Studiare la struttura interna di ogni area**
  - Criterio: identificare i comuni cardine che definiscono l’area

Proposta: algoritmo agglomerativo



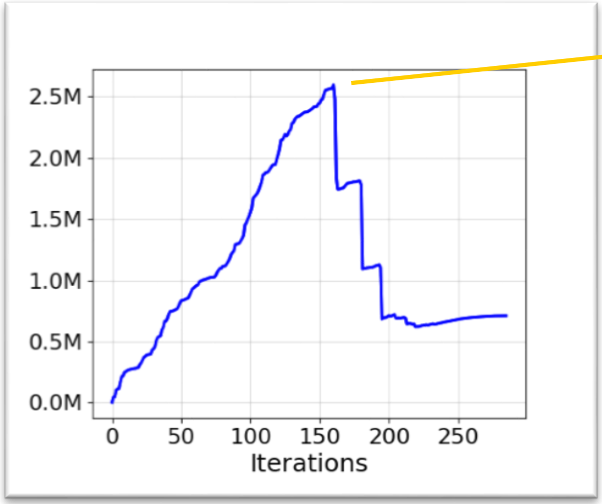
$$localQ(a,b,G) = \frac{F(a,b) + F(b,a)}{\sum_{(x,y)\in E \wedge \{a,b\}\cap\{x,y\}\neq\emptyset} F(x,y)}$$

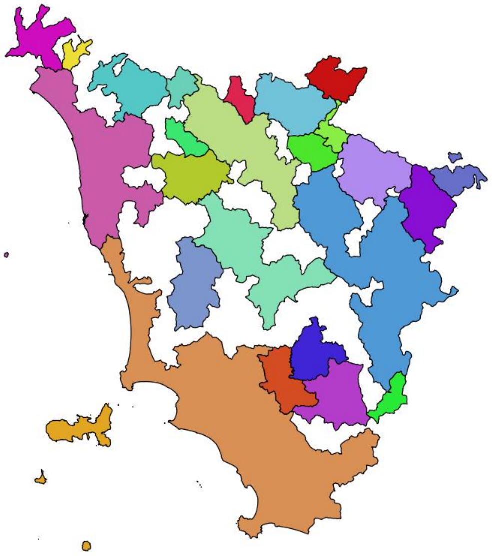
Red Comunità/aree create nell'ultima iterazione

Grey Comunità/aree create nelle precedenti iterazioni

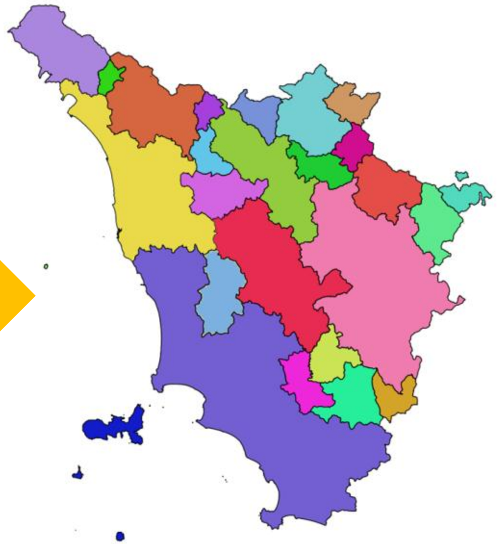
Città policentrica: RISULTATI

$$globalQ(G) = \sum_{(i,j)\in E} F(i,j) - F(i\rightarrow) * \frac{F(\rightarrow j)}{K}$$





25 Comunità scoperte

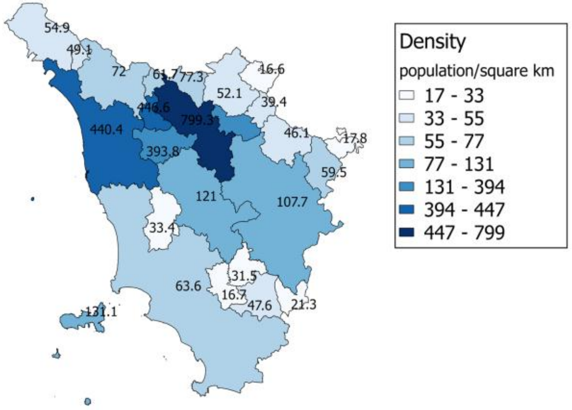


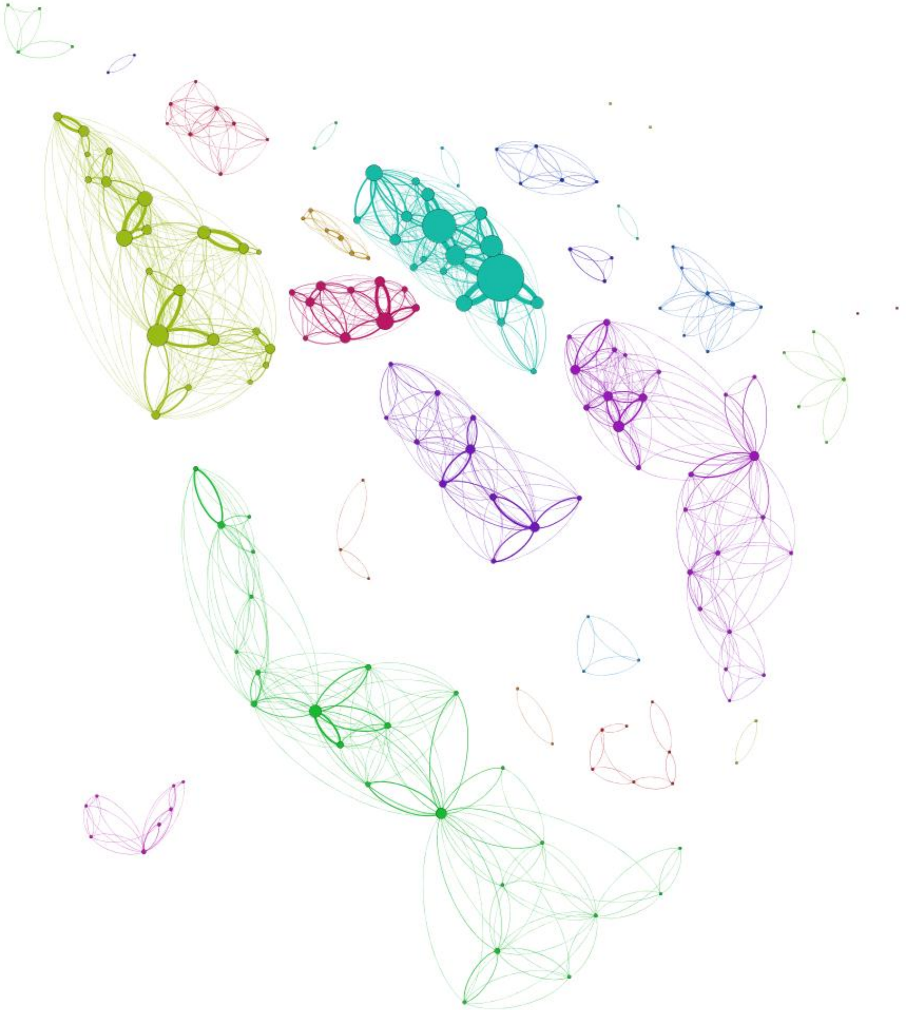
Saturazione

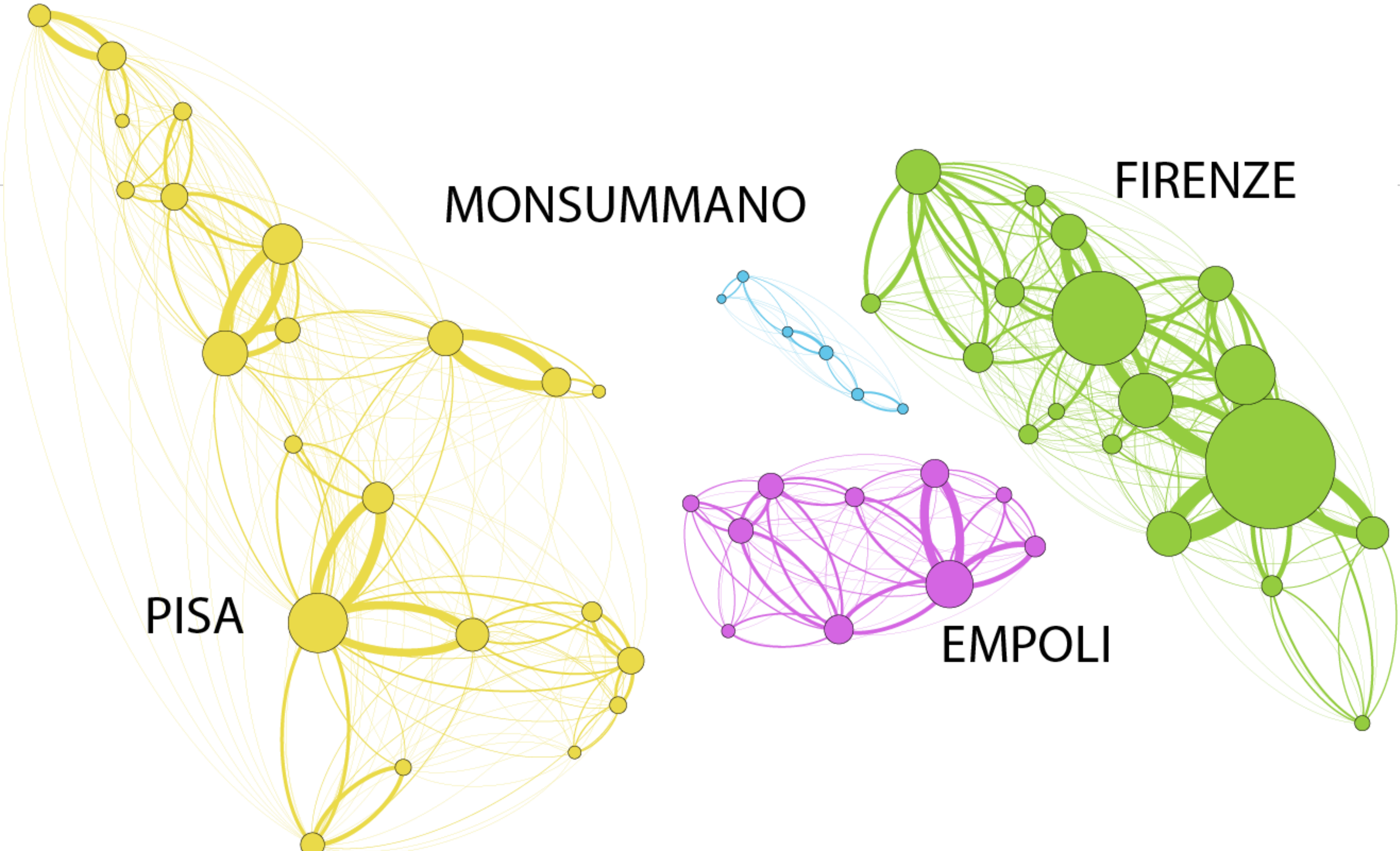
Assegnazione comuni isolati

Sub network

24 communities

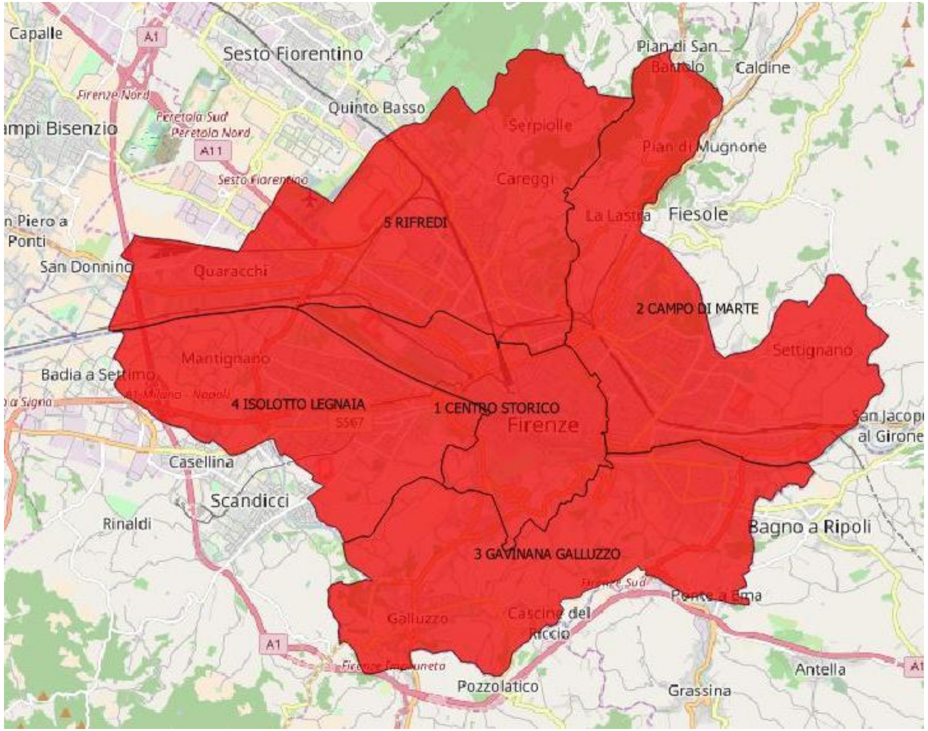




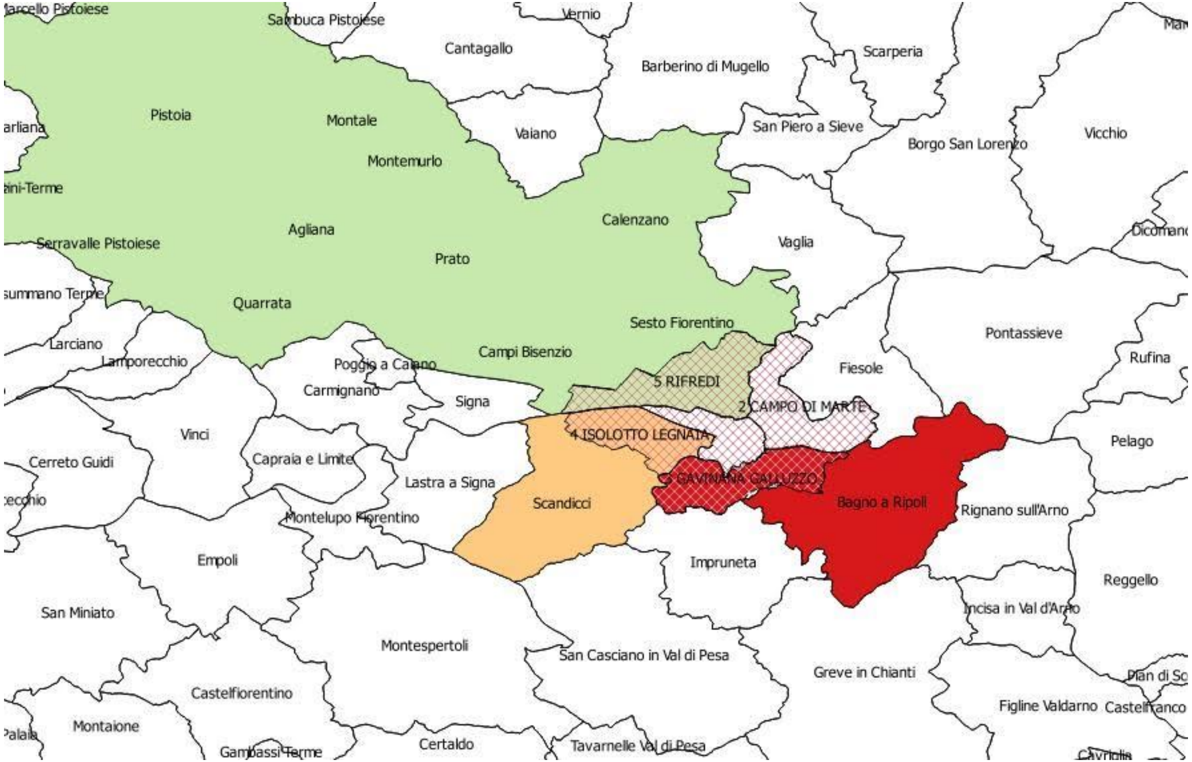

MONSUMMANO
FIRENZE
PISA
EMPOLI

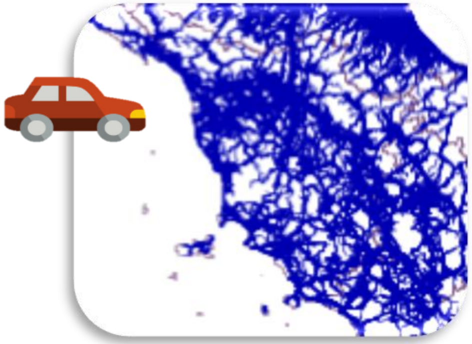
Focus: i quartieri di Firenze

Quartieri	Traiettorie	Incidenza (%)
Rifredi	212.651	31,9
Campo di Marte	126.046	18,9
Centro	103.041	15,6
Gavinana	88.825	13,4
Isolotto	134.171	20,2



Focus: i quartieri di Firenze





2. Analisi di attrattori

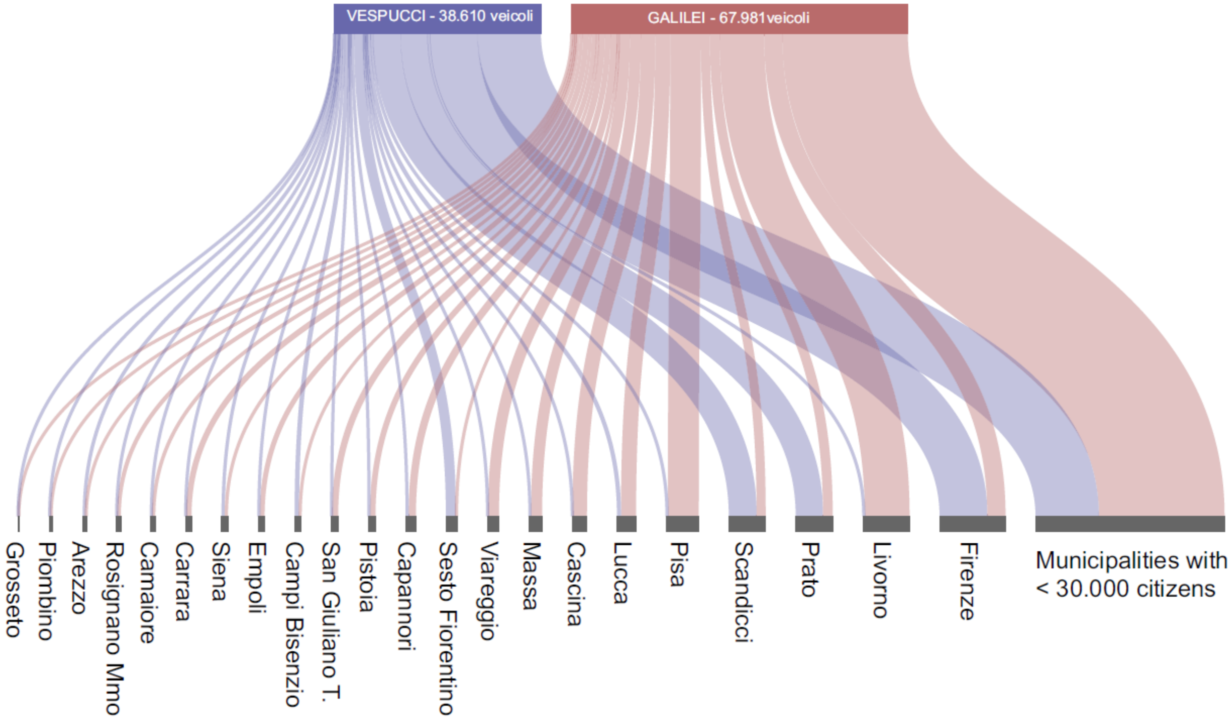
Analisi degli attrattori: aeroporti

Analizzare l'influenza dei grandi attrattori sulla mobilità dei territori circostanti

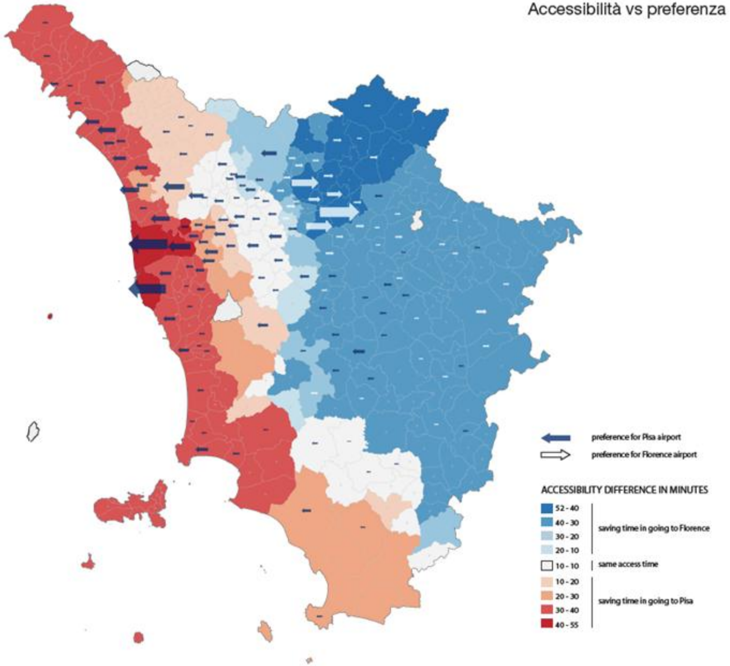
CASO STUDIO:

gli aeroporti di Firenze e Pisa e la propensione dei residenti toscani all'uso delle due infrastrutture.

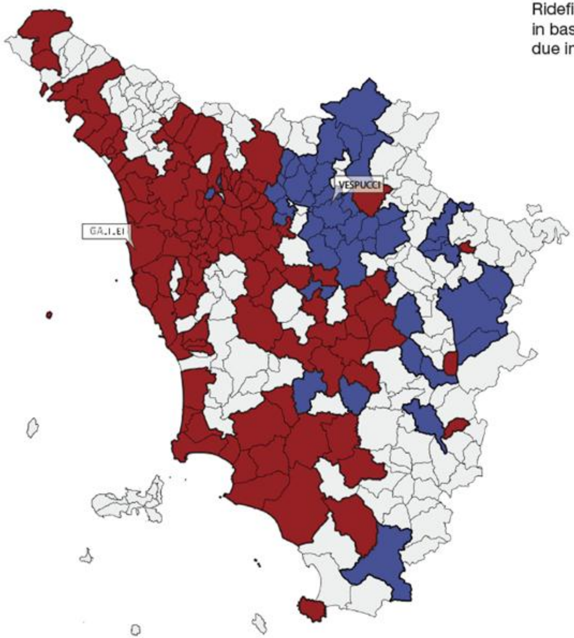
Analisi degli attrattori: aeroporti

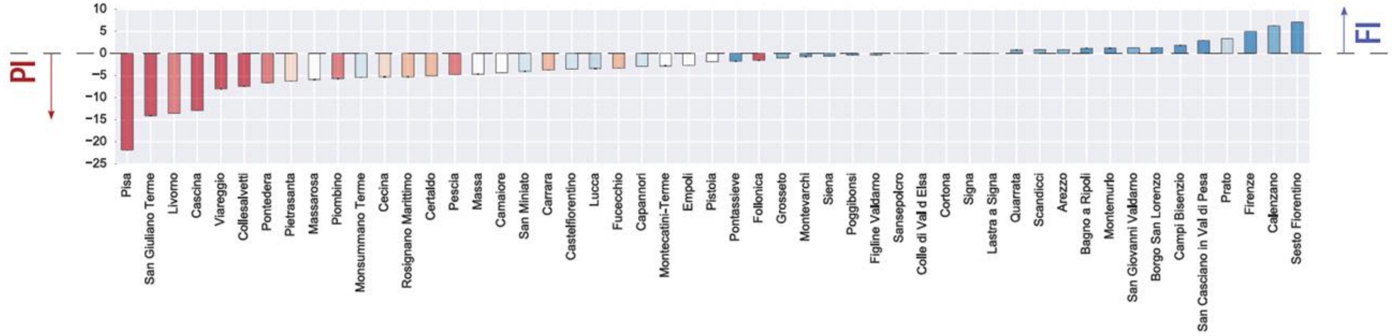


Accessibilità vs preferenza



Ridefinizione del territorio
in base all'attrazione delle
due infrastrutture





Modelli di investimento vs. attrattività

Modeling Investments and Attractiveness on Tuscan Airports.
Ioanna Miliou, Diana Knipl, Salvatore Rinzivillo, Seven R. Bishop, Dino Pedreschi.

An intertwined system based on investment and attractiveness

$$\frac{d}{dt}A = s(mF - (k+e)A),$$
$$\frac{d}{dt}F = -rF + re\frac{bA}{1+bhA}.$$

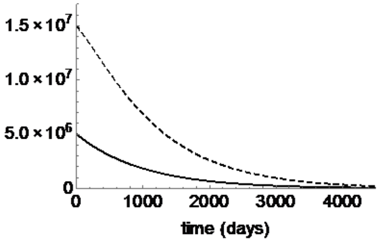
A - > Attractiveness of airport

F -> Number of passengers served

Attractiveness is proportional to the cost of operating the airport (k) and the extra investments (e)

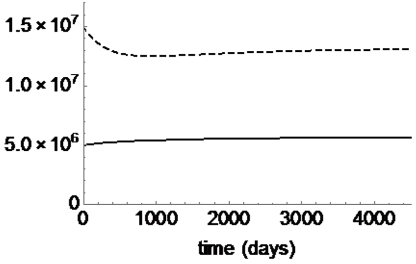
Simple case: non spatial model

No extra investments (e=0)



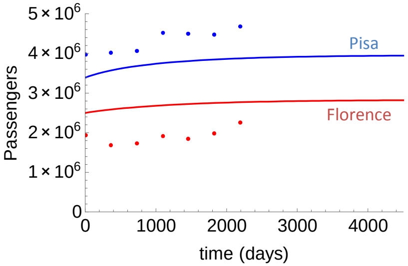
(a) $e = 0$

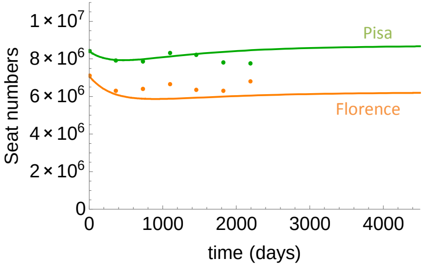
Structural investments (e=0.05)



(b) $e = 0.05$

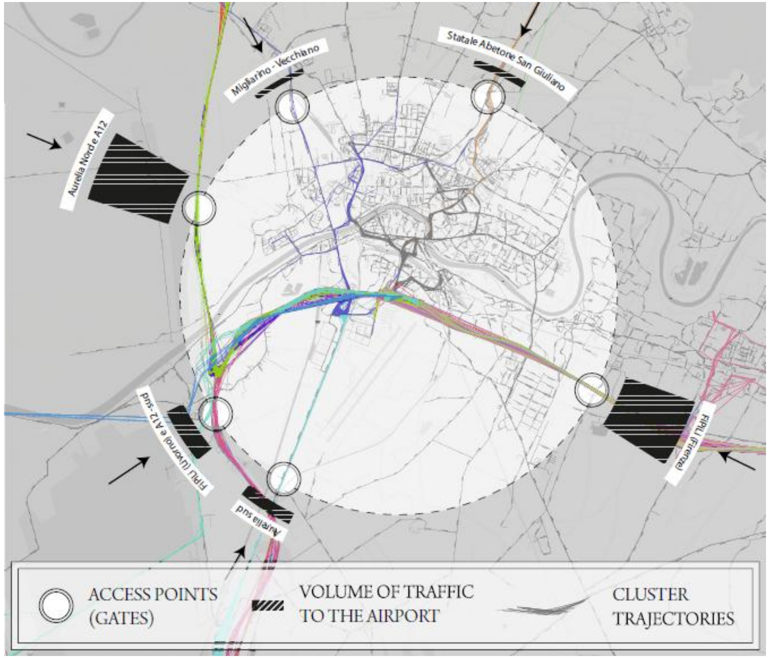
Spatial model: two airports, two populations



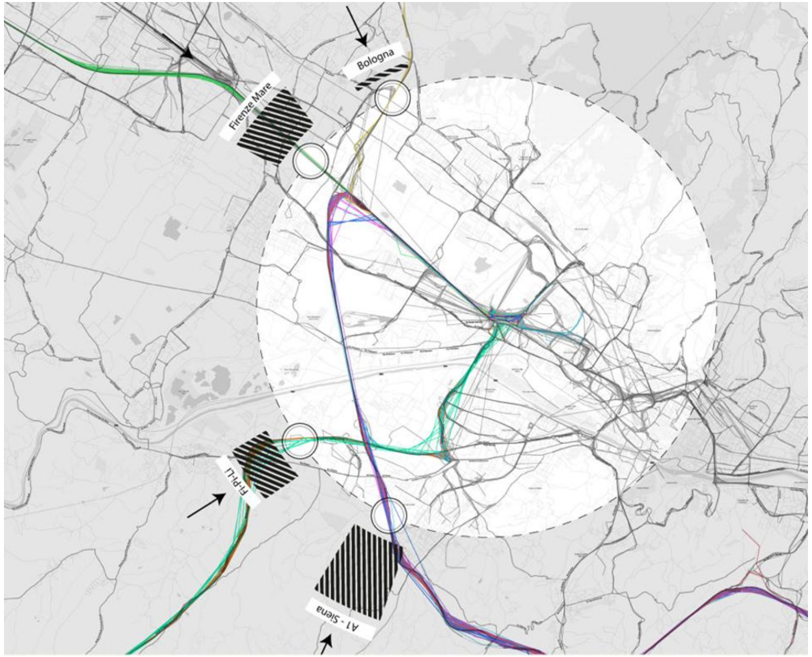


The two airports reach an equilibrium: neither of the two is overwhelming the other

Percorsi di accesso

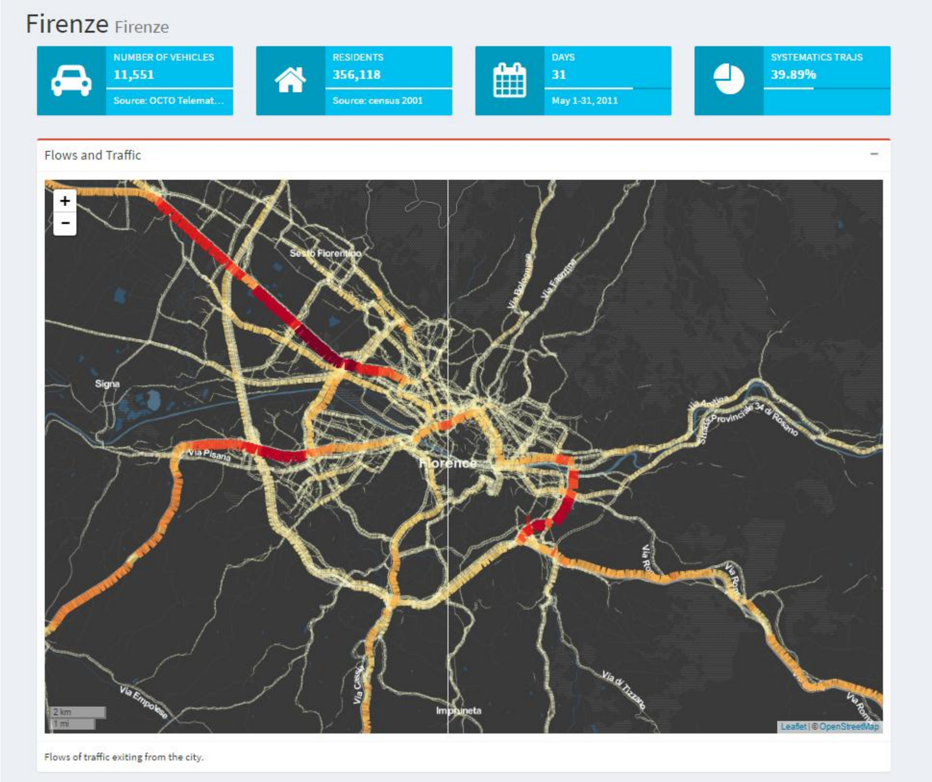


Pisa



Firenze

Urban Mobility Atlas (UMA)

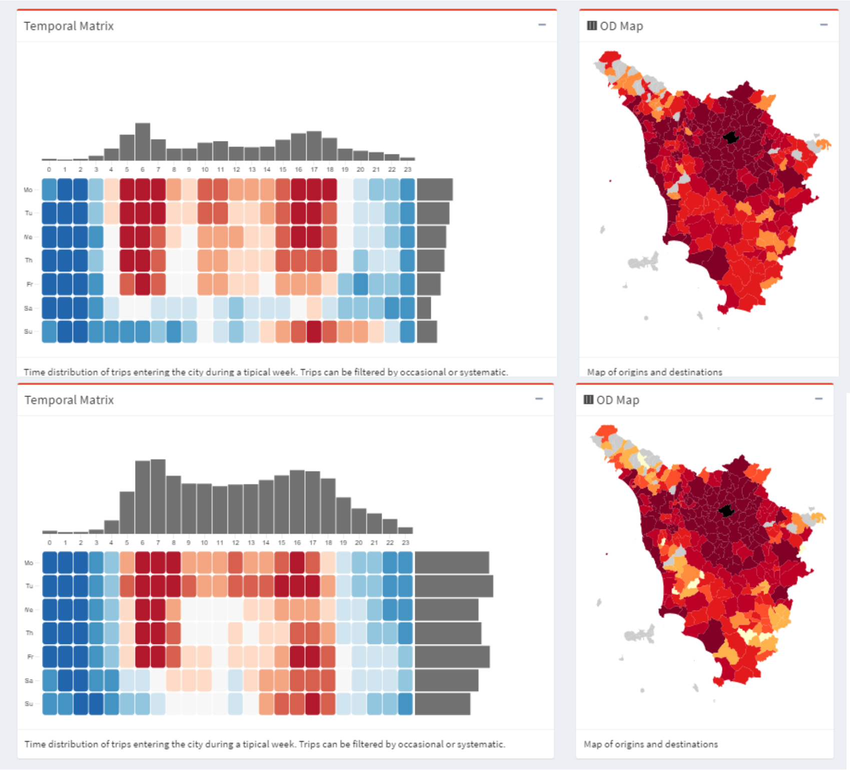


http://kdd.isti.cnr.it/uma2/

Urban Mobility Atlas (UMA)

Sistematico

Occasionale



http://kdd.isti.cnr.it/uma2/

3. Analisi di eventi

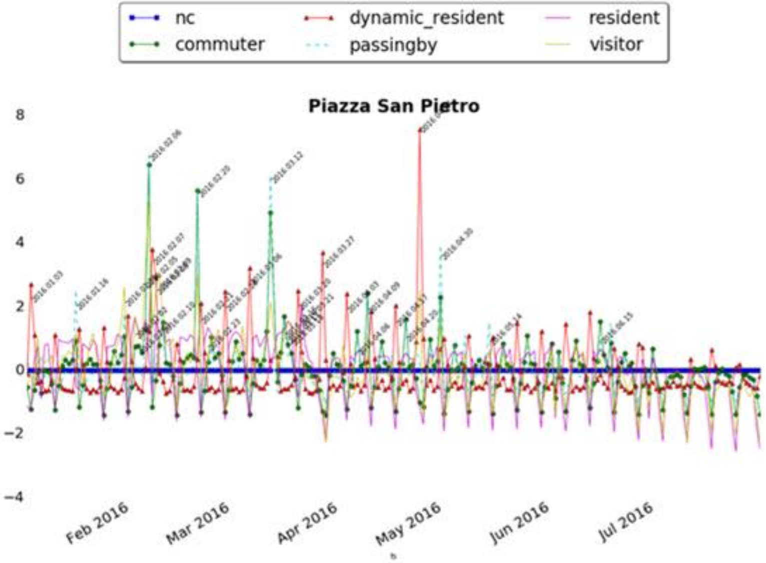
St. Peter's Square
(Piazza San Pietro)

Olympic Stadium
(Stadio Olimpico)

Circus Maximus
(Circo Massimo)

Rilevazione e misurazione eventi

Peak detection



Piazza San pietro

Misuriamo le presenze distinguendo I diversi *City User*:

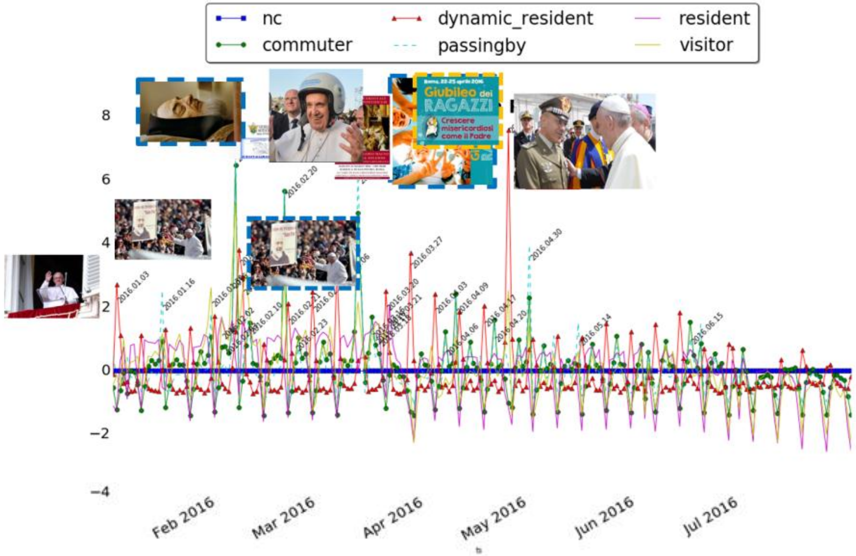
- Residenti
- Pendolari / Visitatori costanti
- Turisti / Visitatori occasionali

Identifichiamo presenze anomale

- in un period specific
- rispetto a presenze passate

Rilevazione e misurazione eventi

Peak detection



Piazza San pietro

Misuriamo le presenze distinguendo I diversi *City User*:

- Residenti
- Pendolari / Visitatori costanti
- Turisti / Visitatori occasionali

Identifichiamo presenze anomale

- in un period specific
- rispetto a presenze passate

SoBigData Research Infrastructure

Social Mining & Big Data Analytics

Social Mining & Big Data Ecosystem

H2020 - www.sobigdata.eu

September 2015– August 2019

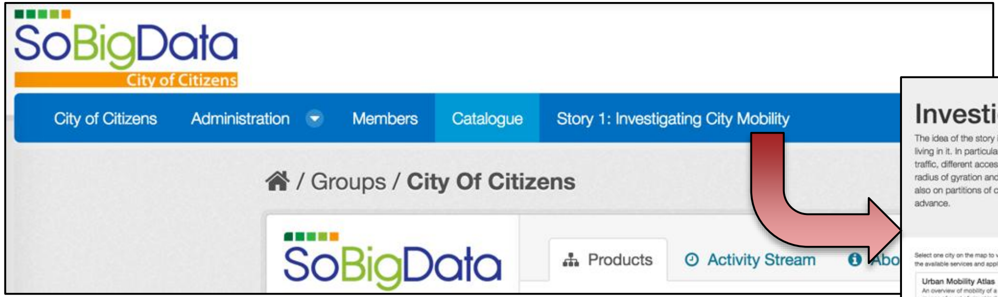
@SoBigData (https://twitter.com/SoBigData)
https://www.facebook.com/SoBigData

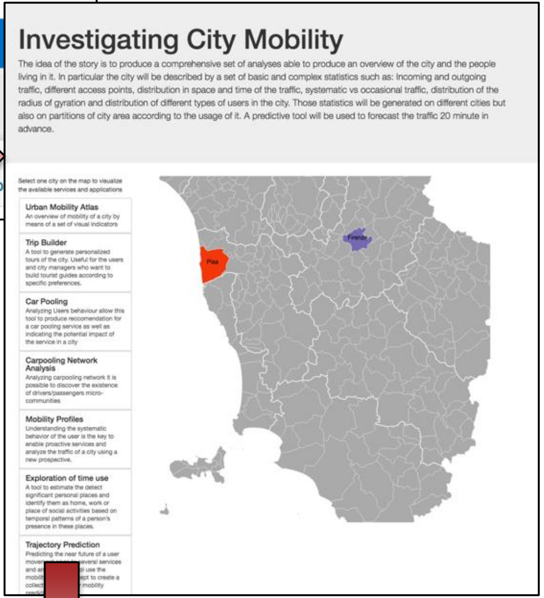
Objectives and Stakeholders

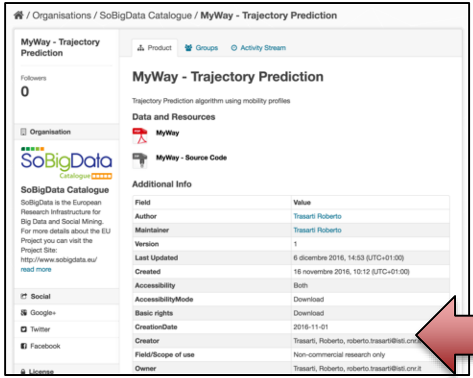
A **Multidisciplinary** European **Infrastructure** for **Big Data** and **Social Data Mining** providing an integrated ecosystem for **ethically sensitive scientific discoveries** and advanced applications of social data mining on the various dimensions of social life, as recorded by “big data”.

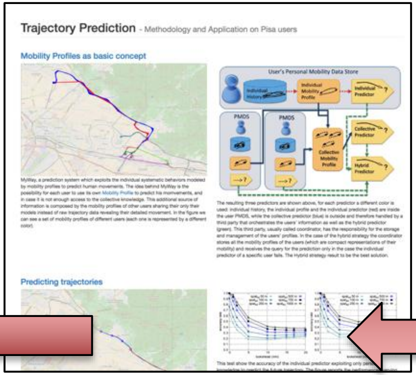
- Big data analysts and social informatics researchers
- Economists, social science and humanities researchers
- journalists, policy and law makers
- Researchers in related communities
- Industrial innovators & startuppers
- The public as a whole

Virtual Research Enviroments





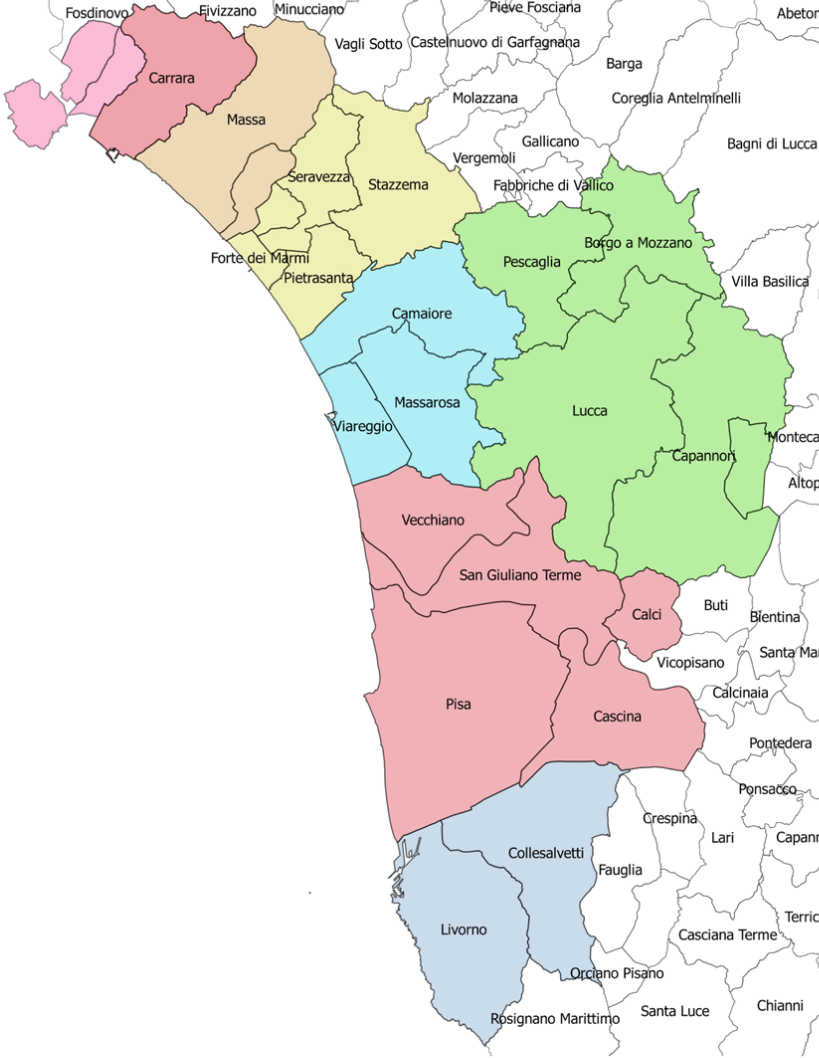




Training and Industry

Grazie!

domande ?



BACKUP Slides

Competitors (state of the art)

Network Based

- Louvain
  MODULARITY BASED
- Demon
  EGO NETWORK BASED
- Infohiermap
  CONDUCTANCE BASED

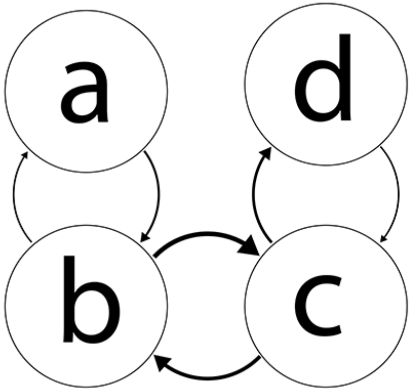
Cluster Based

- DBSCAN
- KMedoid

NETWORK approach

In order to *compare our approach* with the state of art we observe those measures:
internal density conductance modularity

Quality score locale: indice di autoconteminmento



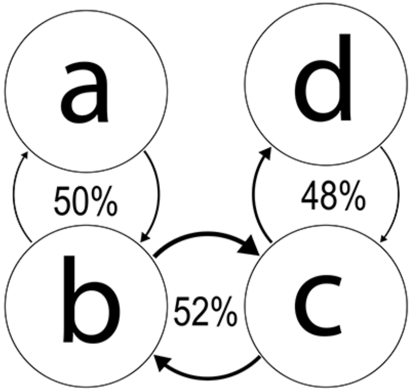
Iteration *n*

Local Function	Inner traffic
localQ(a,b)	50%
localQ(a,d)	0%
localQ(a,c)	0%
localQ(b,c)	**52%**
localQ(d,c)	48%

Evaluation$_n$ (t,z)= tUz/(t*+z*)

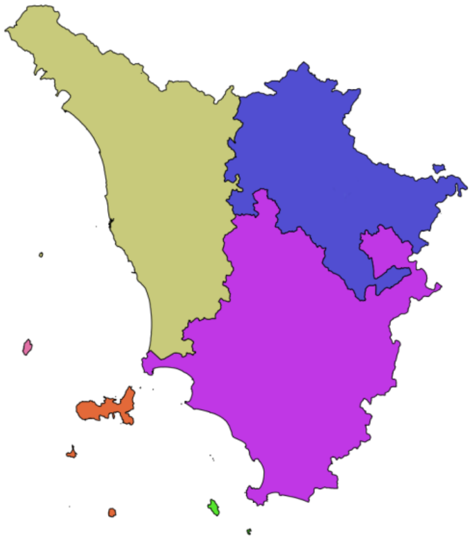
a**U**b = 50% b**U**c = 52%

c**U**d = 48% d**U**e = 40%



Iteration *n+1*

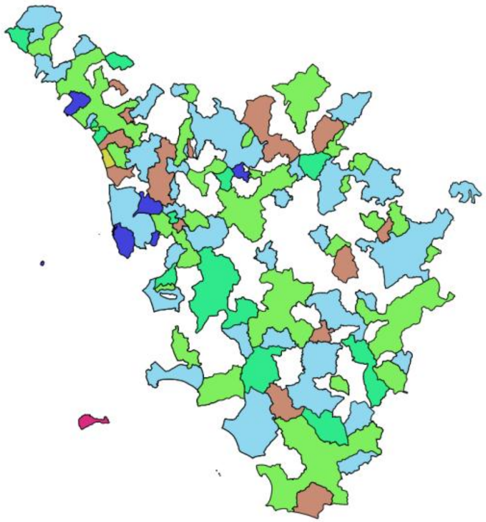
Network approach LOUVAIN



Measure min/max/avg/std	Louvain	Policentrometer
Internal Edge Density	0.15/0.32/0.21/0.07	0.27/0.75/**0.49**/0.20
Conductance	0.014/0.58/0.38/0.27	0.014/0.97/**0.88**/0.19
Modularity	0.16	-0.06

Drawback too few and too big communities

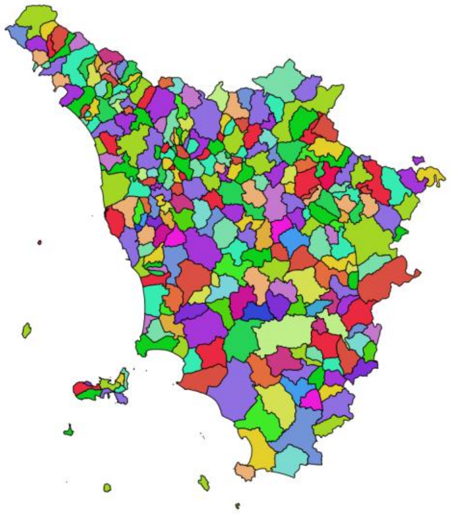
Network approach DEMON



Measure min/max/avg/std	Demon	Policentrometer
Internal Edge Density	0.12/0.50/0.28/0.18	0.27/0.75/**0.49**/0.20
Conductance	0.37/0.90/0.50/0.17	0.014/0.97/**0.88**/0.19
Modularity	-0.38	-0.06

Drawback too overlapping communities

Network approach INFOHIERMAP



Measure min/max/avg/std	INFOHIERMAP	Policentrometer
Internal Edge Density	0.09/0.50/0.18/0.10	0.27/0.75/**0.49**/0.20
Conductance	0.90/0.98/**0.95**/0.24	0.014/0.97/**0.88**/0.19
Modularity	0.006	-0.06

Drawback non contiguous communities

Lesson Learned

Comparison with network algorithms:

- Too few communities
- Too big
- Not contiguos

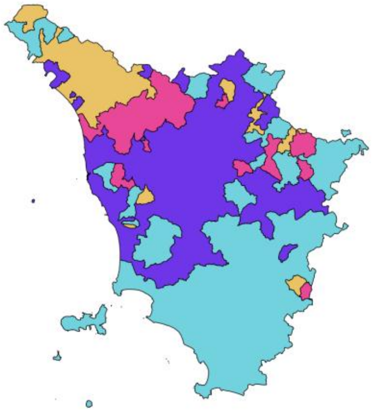
CLUSTERING approach

partitions obtained with K-medoid
and DBSCAN cluster methods

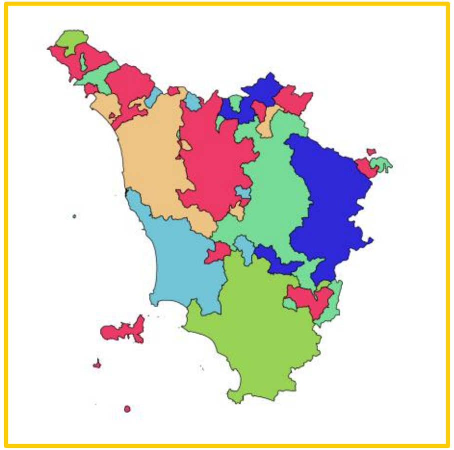
K MEDOID

Drawback non contiguous communities

k=2



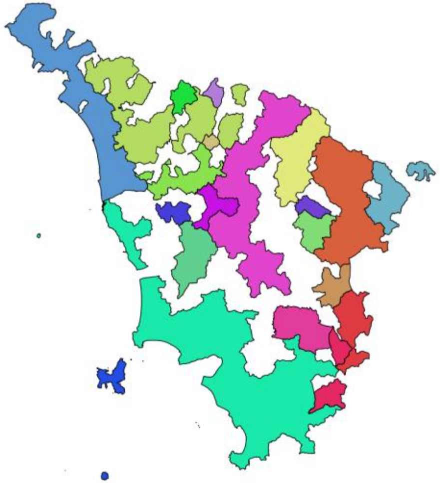
k=4



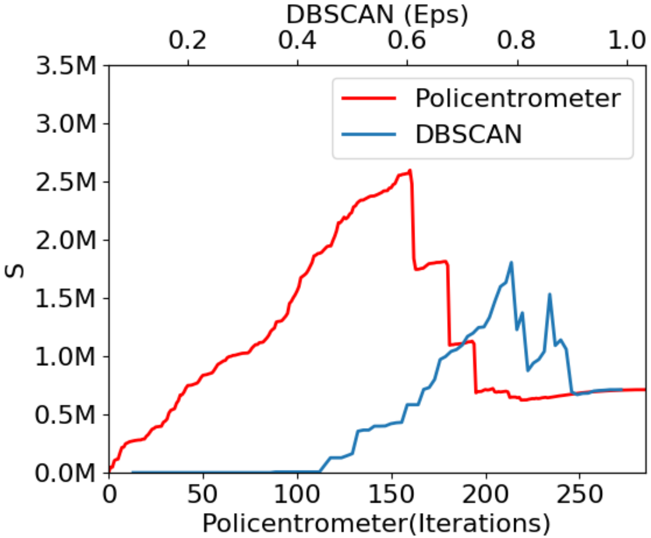
k=6

DBSCAN

Drawback: Policentrometer get a higher global score than DBSCAN



dbscan communities

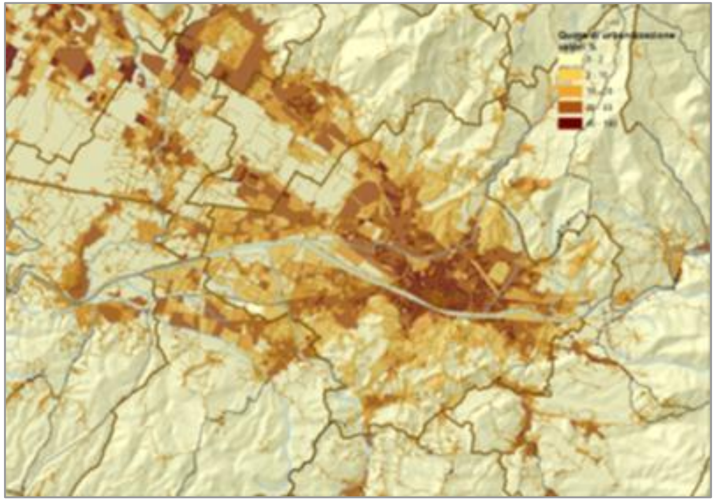


optimal choice for the two methods

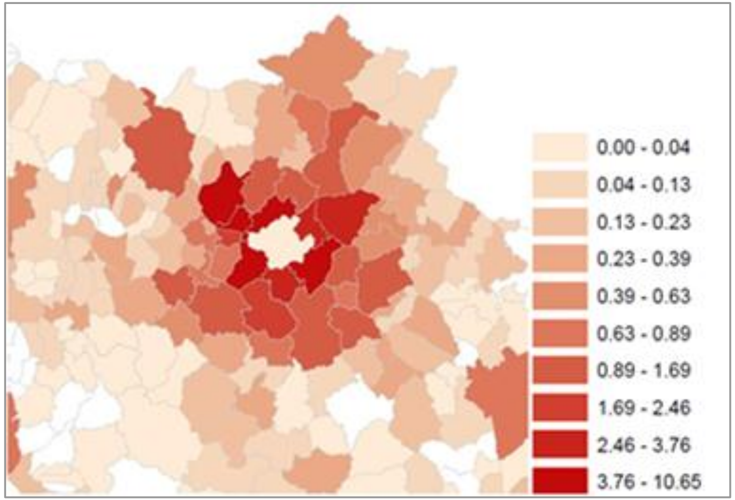
Conclusion

- Flow inclusion based problem definition
- Ad hoc algorithm that outperforms state-of-the-art methods
- Preliminary evaluation of results

Stato dell’arte



Analisi Morfologica



Analisi funzionale