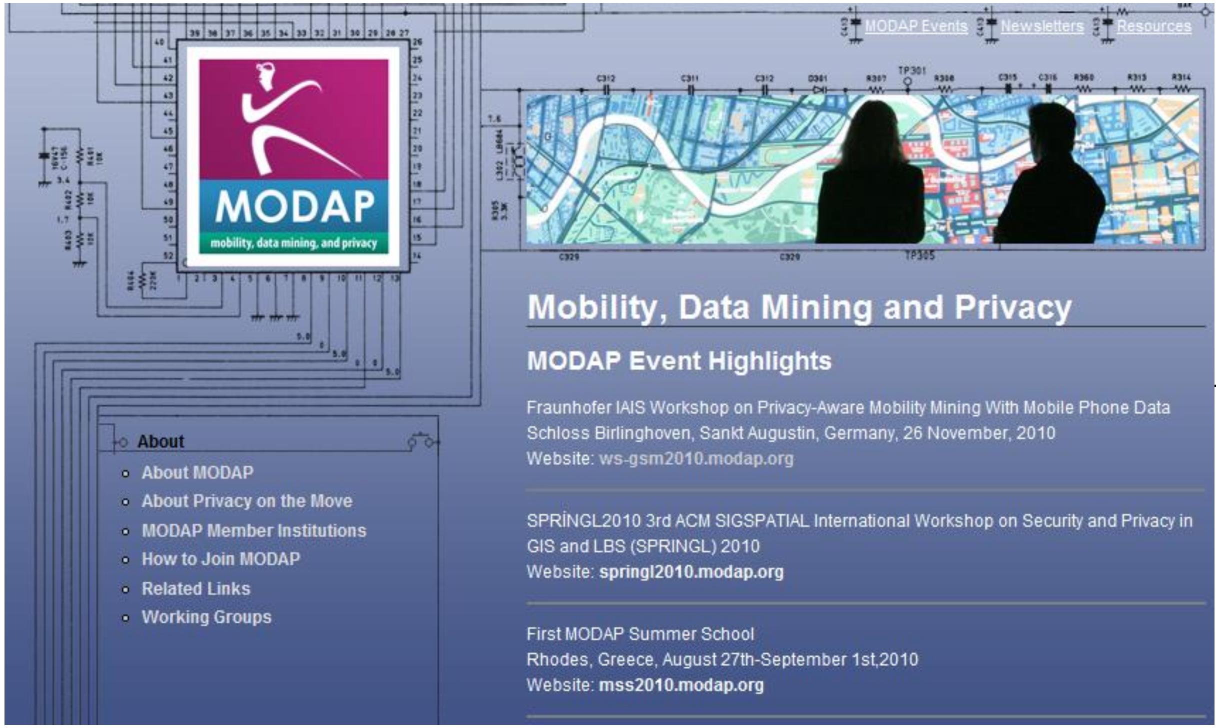
Privacy Issues in Geospatial Visual Analytics



Gennady Andrienko
Natalia Andrienko

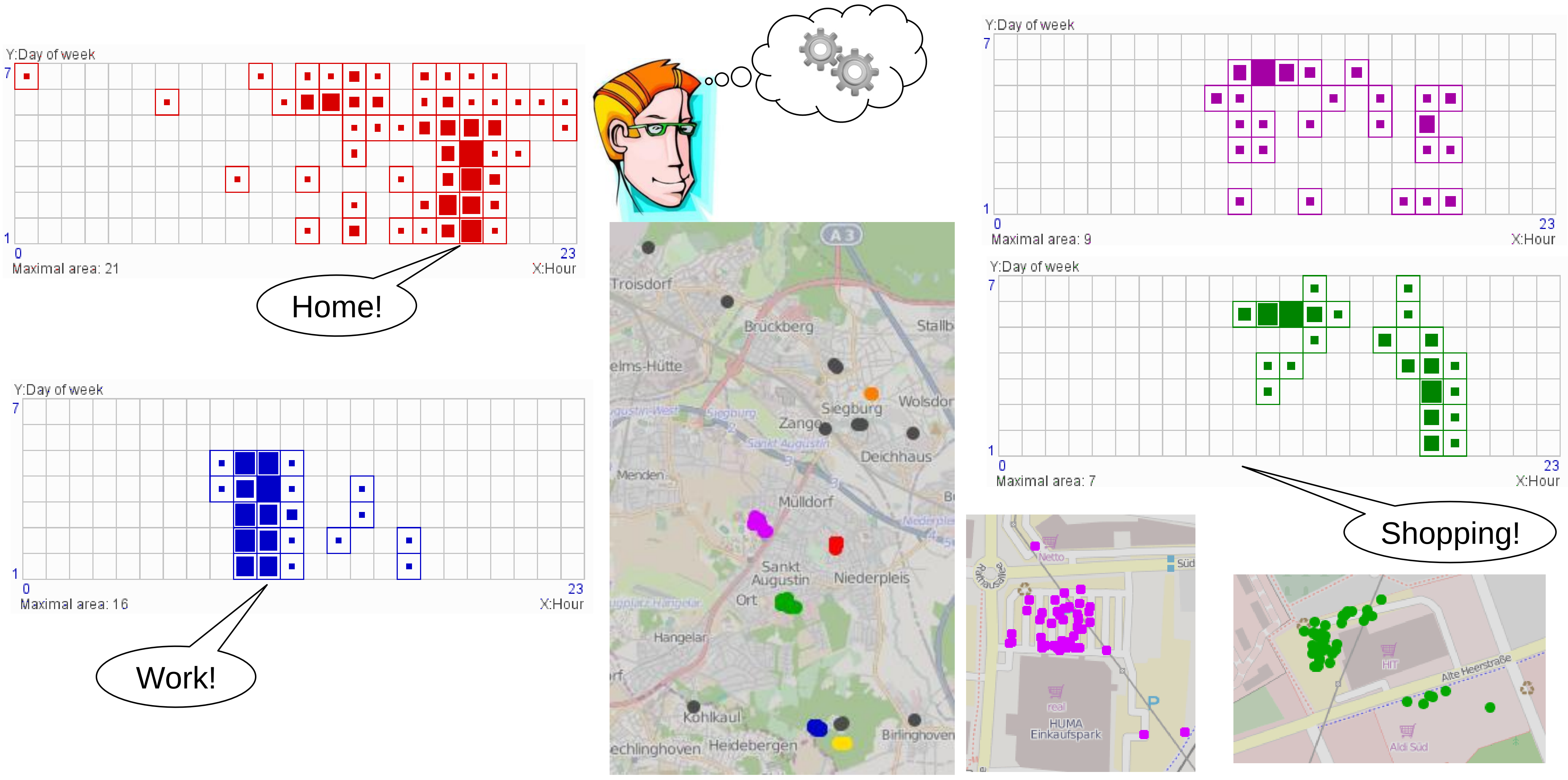
http://geoanalytics.net/and

Privacy issues

- Lots of personal data containing locations can be (automatically) collected and are actually collected:
  - Mobile phone use, car tracking, video observation, use of bank cards, …
- People can also make their locations known to others through Web2.0:
  - Georeferenced photos in flickr and Panoramio, Twitter messages, …
- These data can be (and often should be or need to be) analyzed and are actually analyzed
- The analyses may disclose potentially sensitive personal information
- Visual analytics approaches can aggravate the problem
  - Enables human-computer collaboration where each side applies its unique capabilities
  - Humans use their knowledge, experience, common sense; can easily relate pieces of information; do not require formal representations

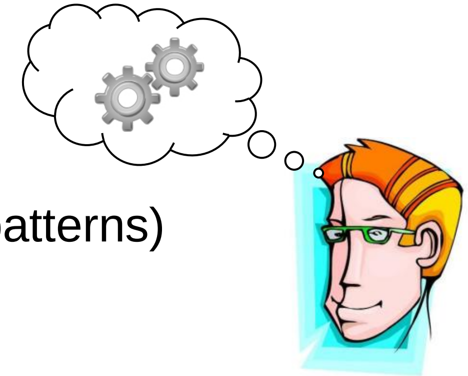
Example 1: GPS tracks of a personal car

Spatial clusters of stops for >=30 minutes



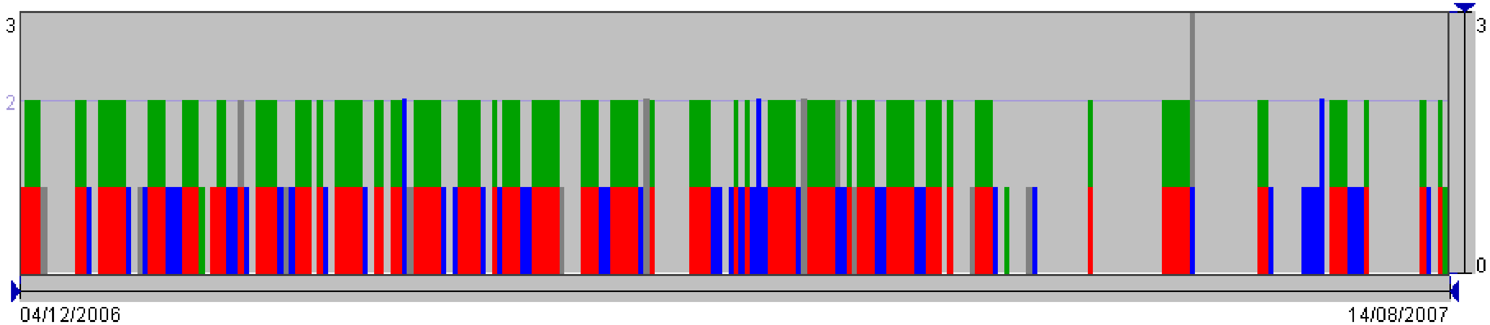
What was learned about the car owner:

- The places where the person lives, works, and shops and the places the person sometimes visits
- The durations of the stops; the times spent in the shopping areas
- The usual times of the trips and stops
- The typical routes and their distribution over time (daily and weekly patterns)



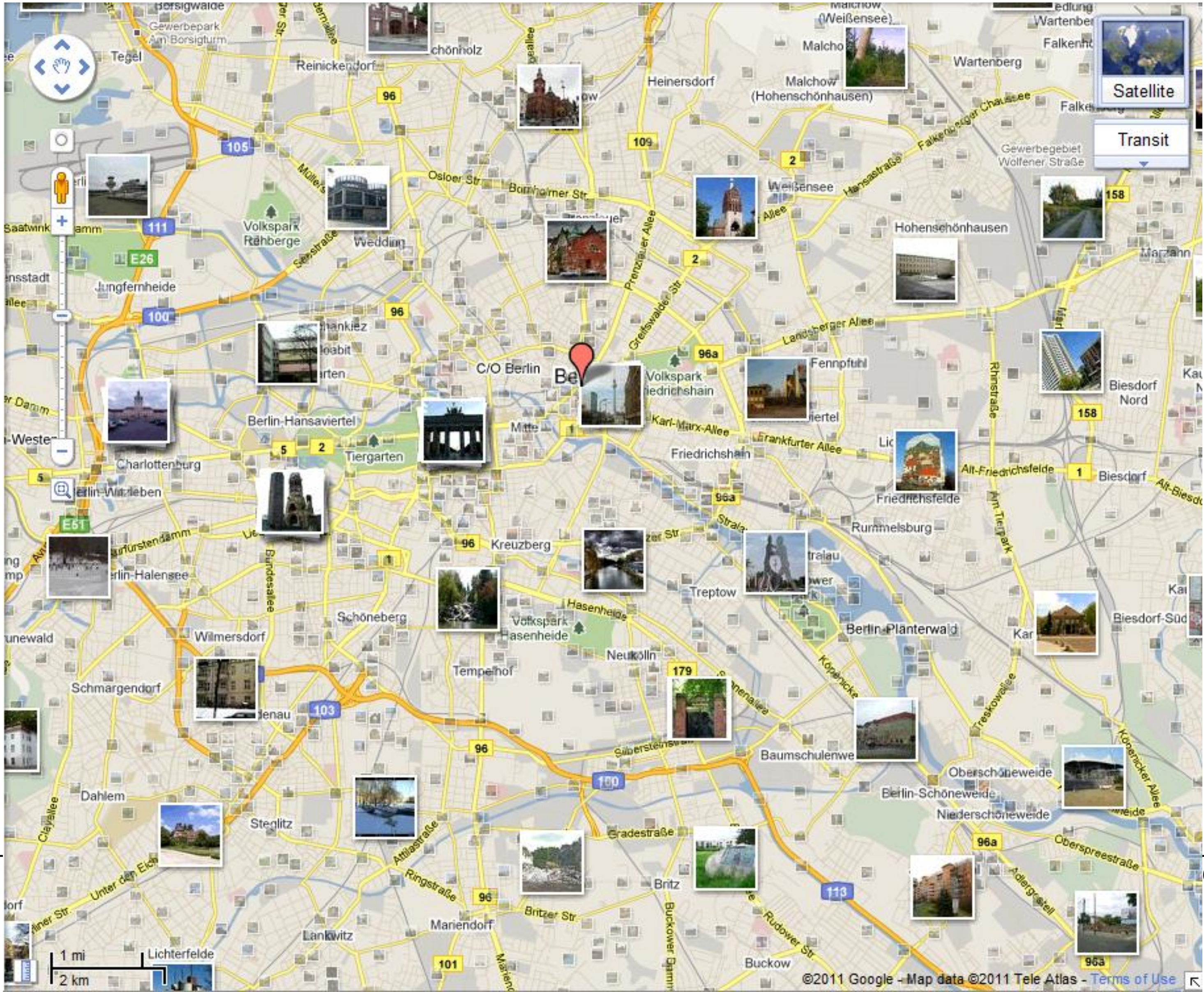
Other inferred personal information:

- The person has a flexible work schedule, has no small children, is often away or sick



green: home to work
red: work to home
blue: home – home

Example 2: Photos in flickr or Panoramio



Repeated meetings of people

identifiers	Overall frequency	Frequency after filtering
311327;953473	67	67
311327;1367968	50	50
953473;1367968	26	26
1592718;1620799	24	24
311327;2038072	8	8
953473;2038072	5	5
1367968;2038072	3	3
899174;953473	2	2
899174;1367968	2	2
1690654;2038072	2	2
1592718;1962862	2	2
1038820;2044833	2	2
986555;1592718	1	1
953473;1690654	1	1
953473;1258045	1	1
953473;1178121	1	1
687113;1620799	1	1

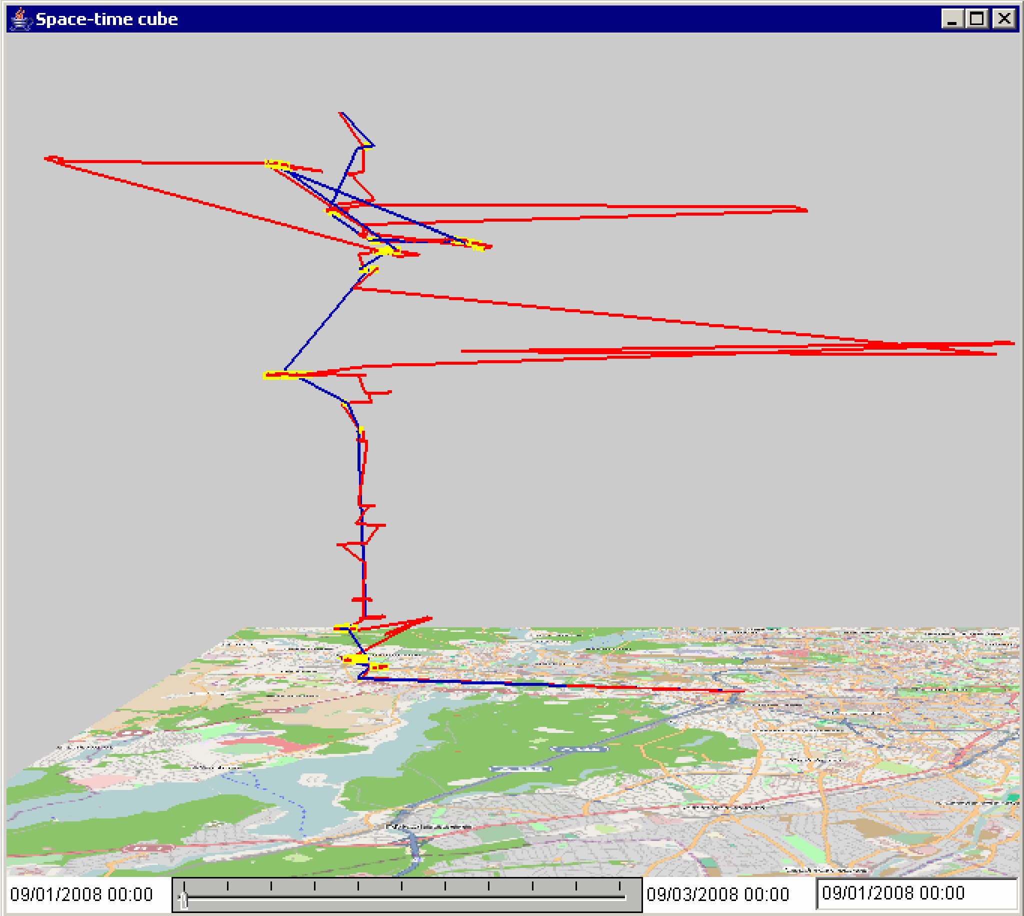
Pairwise interactions from Trajectories (time <...

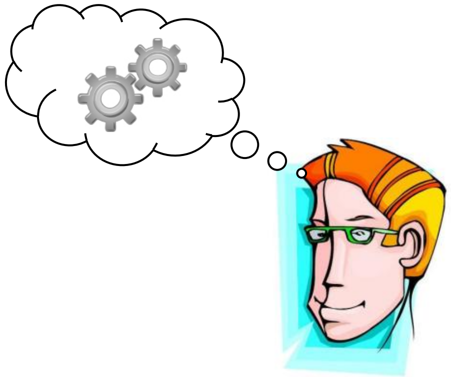
129: [11/06/2008 12:42 - 11/06/2008 14:48]

Id	Name	Earliest time	Latest time
311327	311327	11/06/2008 12:42	11/06/2008 14:48
953473	953473	11/06/2008 12:55	11/06/2008 14:05

171: [28/06/2008 12:44 - 28/06/2008 14:16]

Id	Name	Earliest time	Latest time
311327	311327	28/06/2008 12:44	28/06/2008 14:16
953473	953473	28/06/2008 13:37	28/06/2008 14:08





Users 311327, 953473, and 1367968 frequently meet each other in Berlin and sometimes meet also user 2038072.

Users 1592718 and 1620799 frequently meet.

…

Example 3: the iPhone case

- Availability of important but challenging spatio-temporal data sets (geospatial imagery, sensors, GPS and movement tracking, geo-tagging, flickr, wiki, …)
  - Sometimes {unintentionally} breaking personal privacy

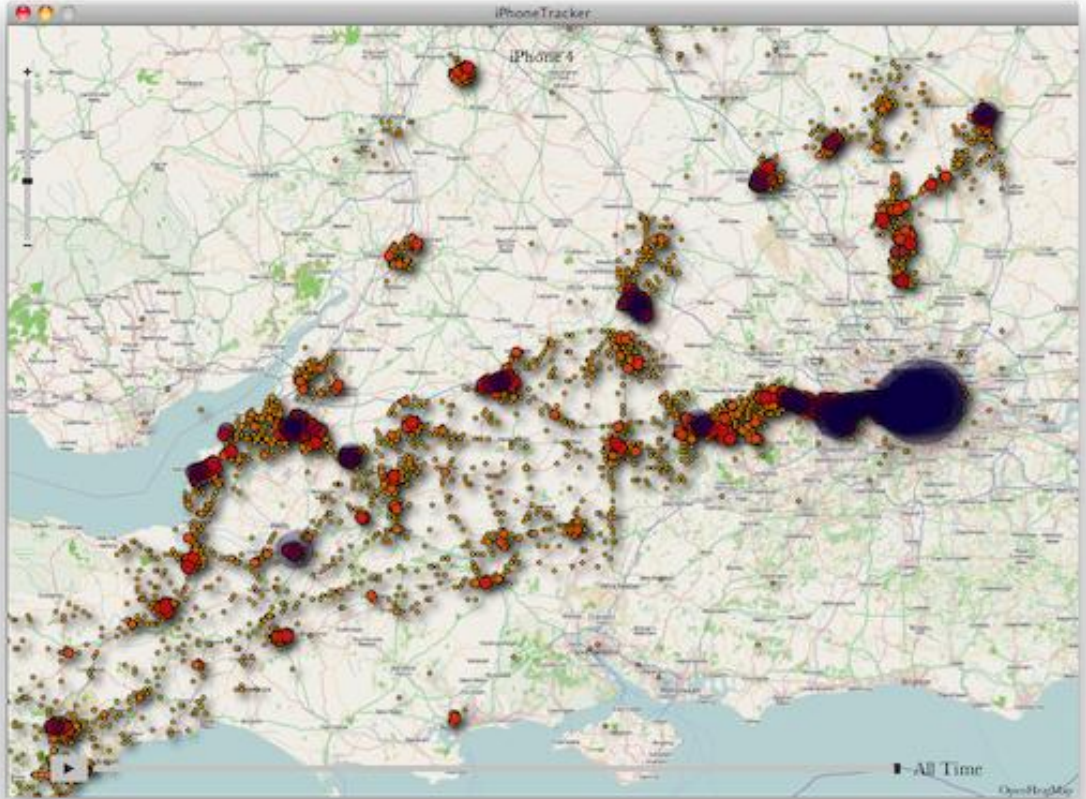
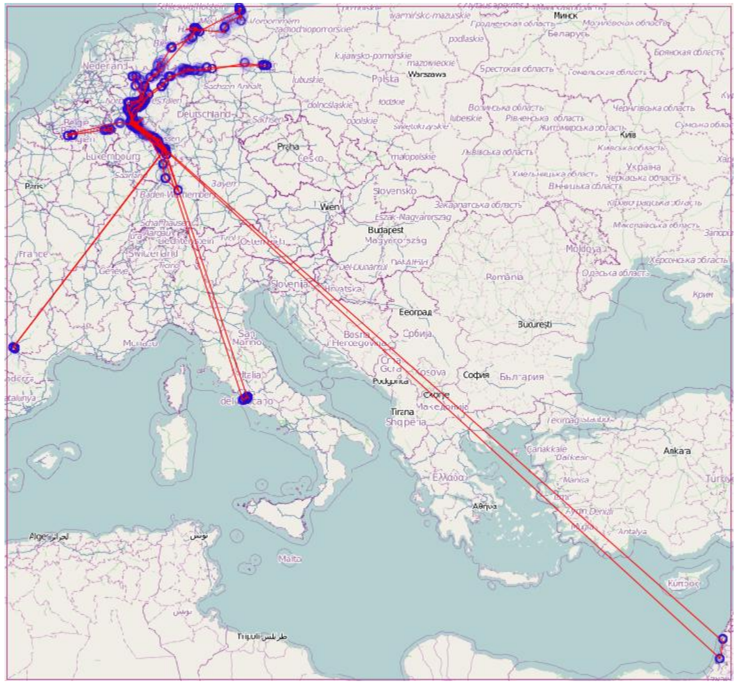
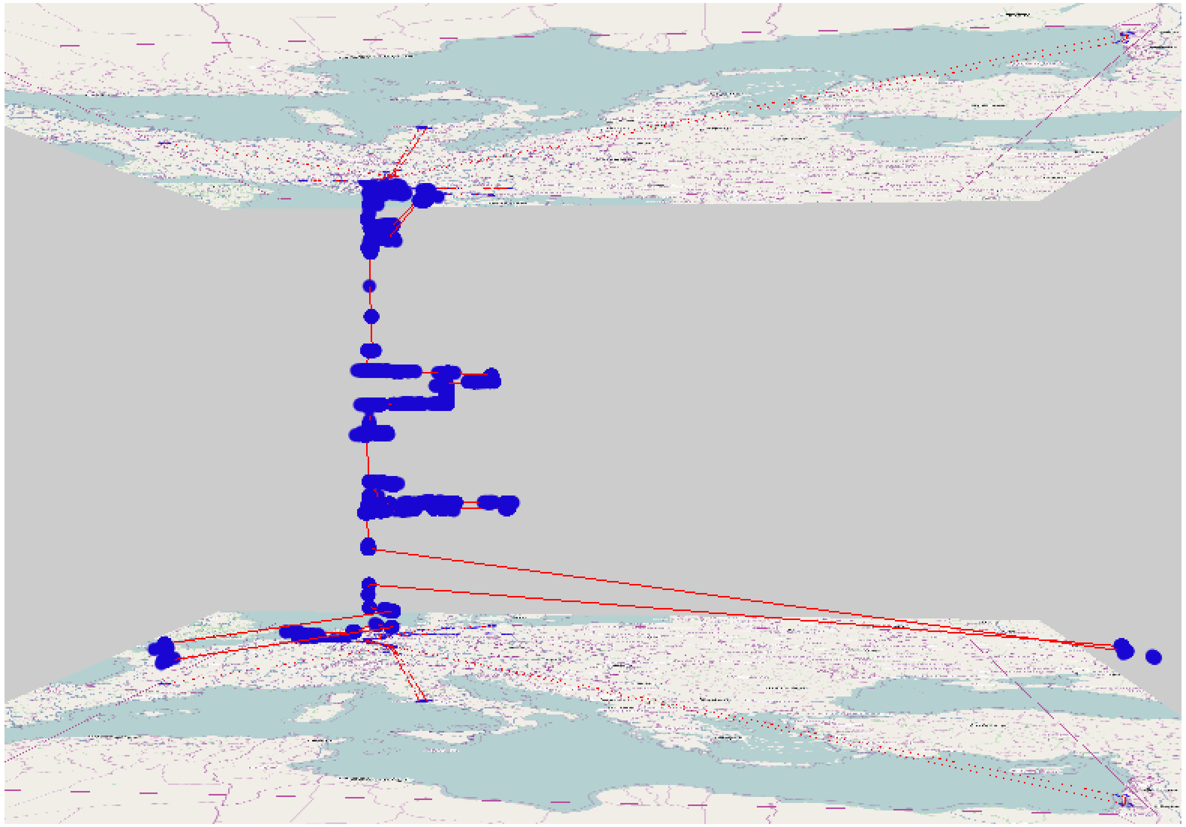


Image courtesy of http://petewarden.github.com/iPhoneTracker/

Are iPhone data really dangerous for {my} personal privacy?

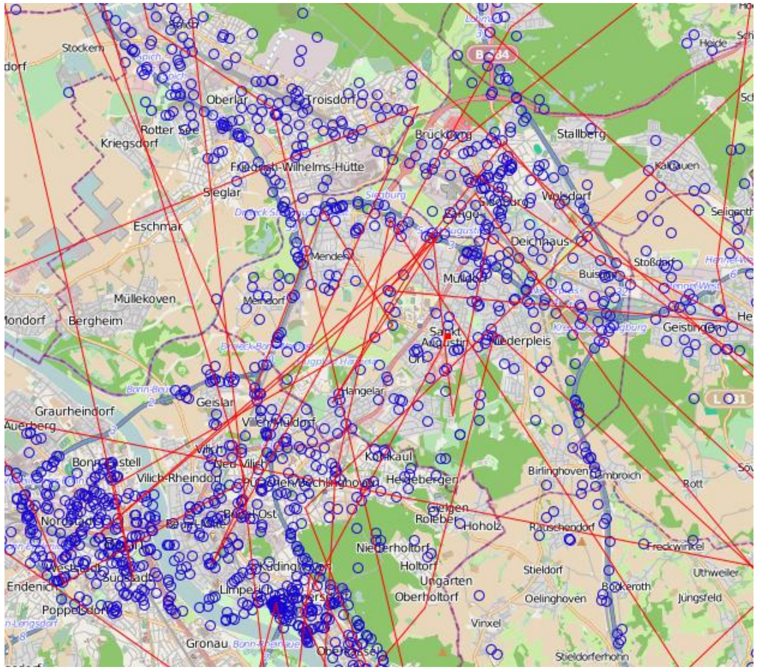
- At large scale: definitely YES – show which cities have been visited and when

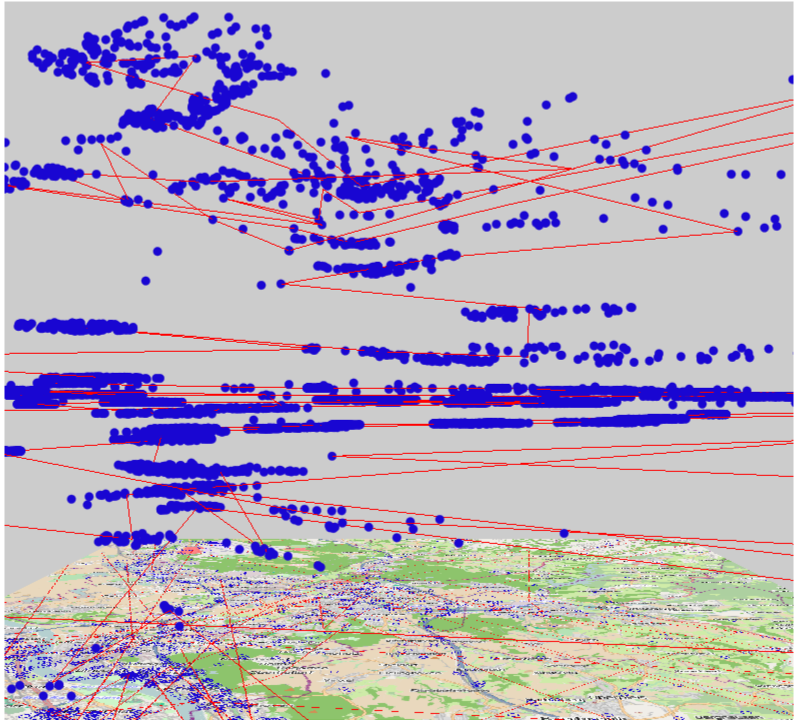




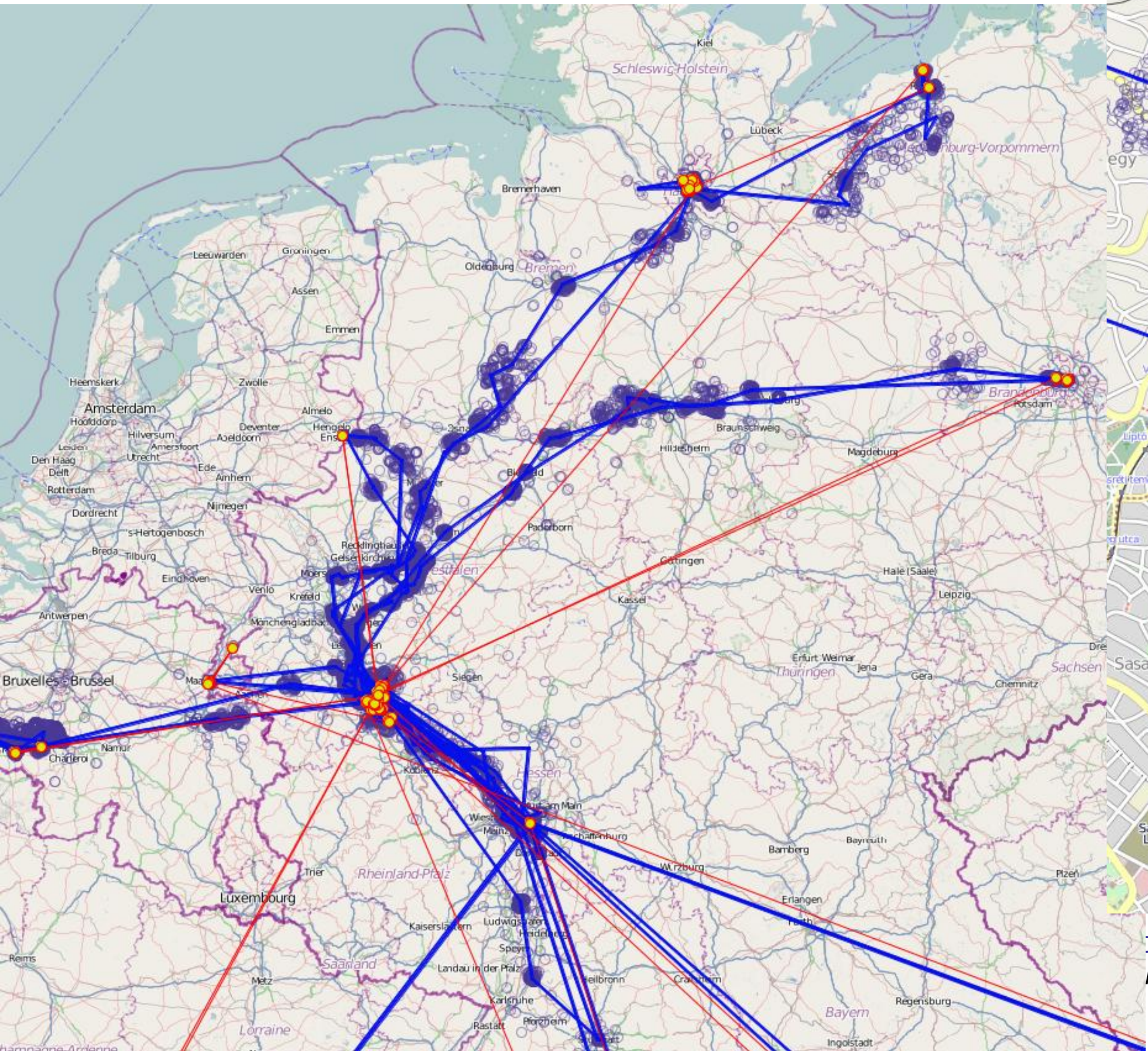
Are iPhone data really dangerous for {my} personal privacy?

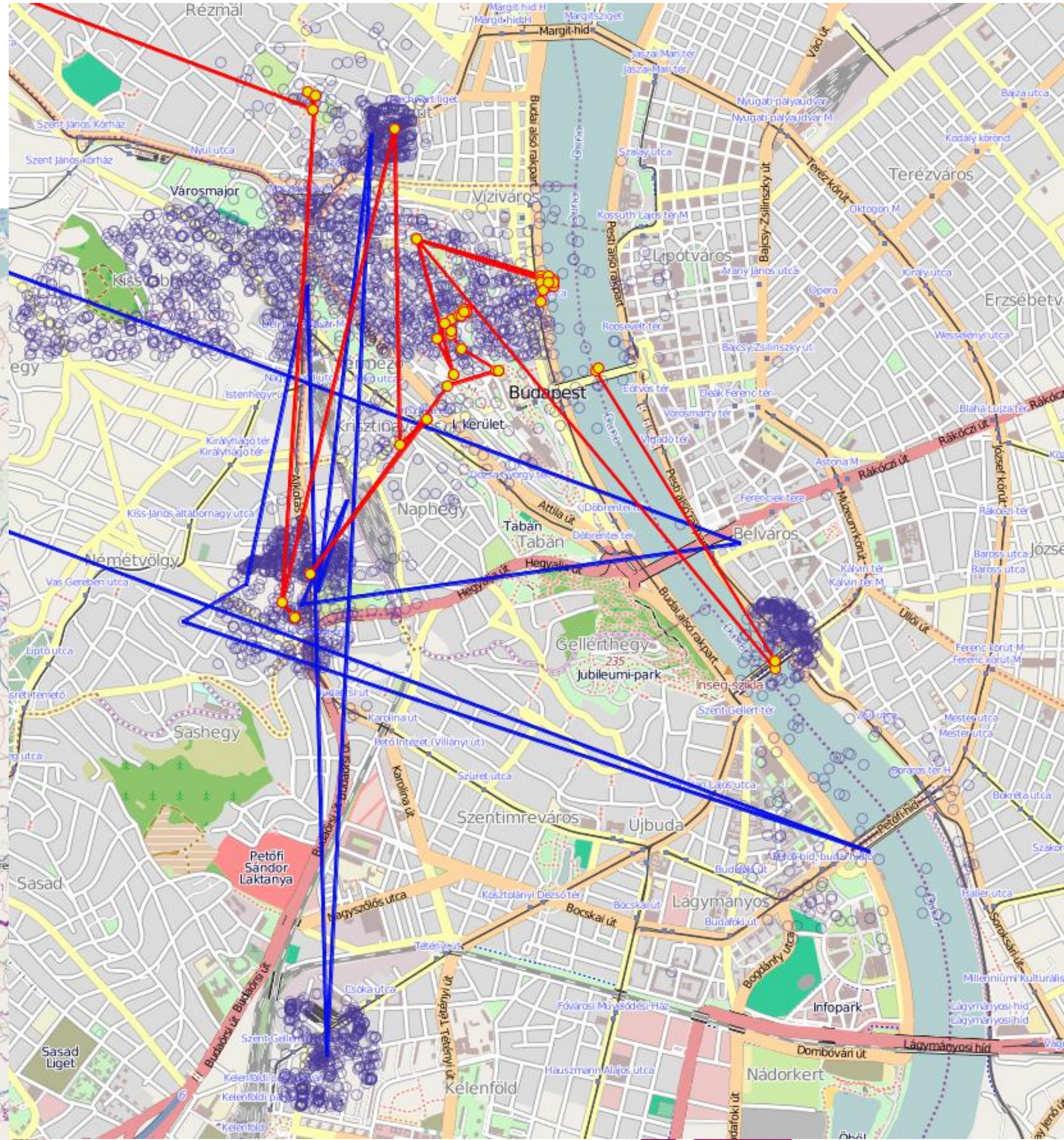
- At small scale: probably NO – real mobility patterns are not visible





iPhone data: cross-checking with other data sources (positions of photos)





What is special about human's involvement in analysis?

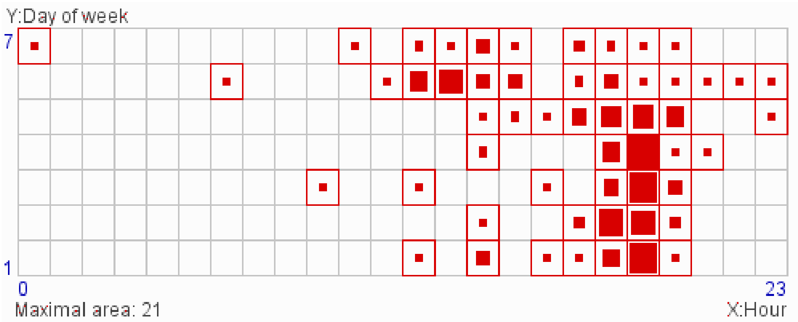
- A human flexibly links data and extracted patterns to *context*
  - *Spatial (geographic) context*: properties of places, spatial objects and their properties, spatial relationships
  - *Temporal context*: properties of time moments and intervals, temporal objects (events and processes) and their properties, temporal relationships
  - *Conceptual context*: general and domain-specific concepts and their relationships
- The context does not need to be formally or even explicitly represented
- A human flexibly infers new information by linking pieces of known information

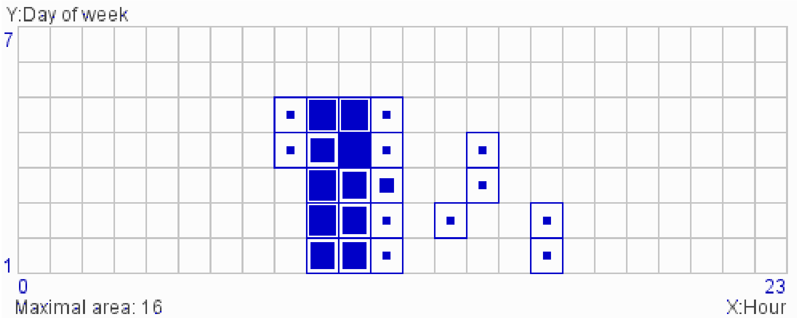
Hence, how can (geo)visual analytics contribute to privacy protection?

- Try to describe the context
- Try to describe the possible inferences

to inform and direct the privacy protection research

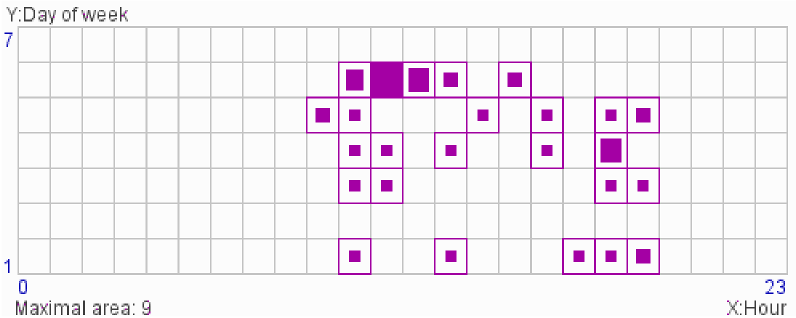
Examples of describing context and inferences

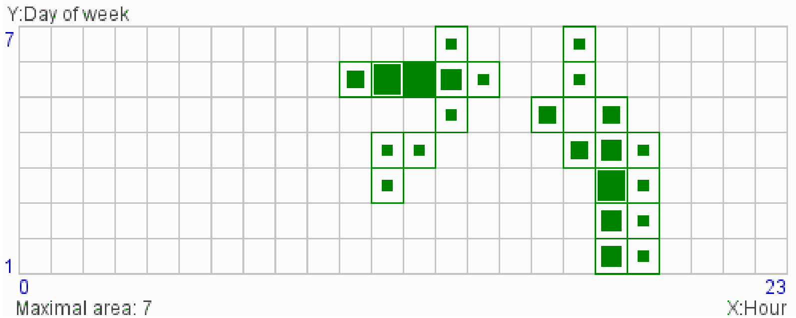




Temporal context:
- days of the week
- times of the day

Conceptual context:
- activities of people
- typical days and times of the activities
- typical durations
- typical places



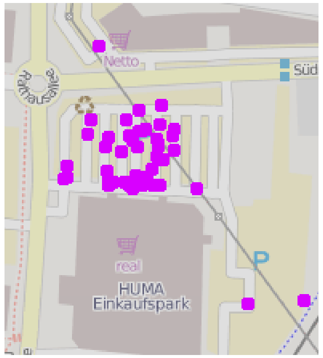


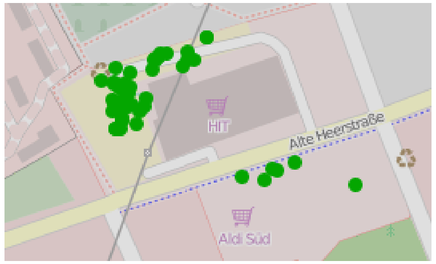
Long stops in the evenings of the working days and at any time in the weekends ⇒ Home!

Long stops in the mornings of the working days and no stops in the weekends ⇒ Work!

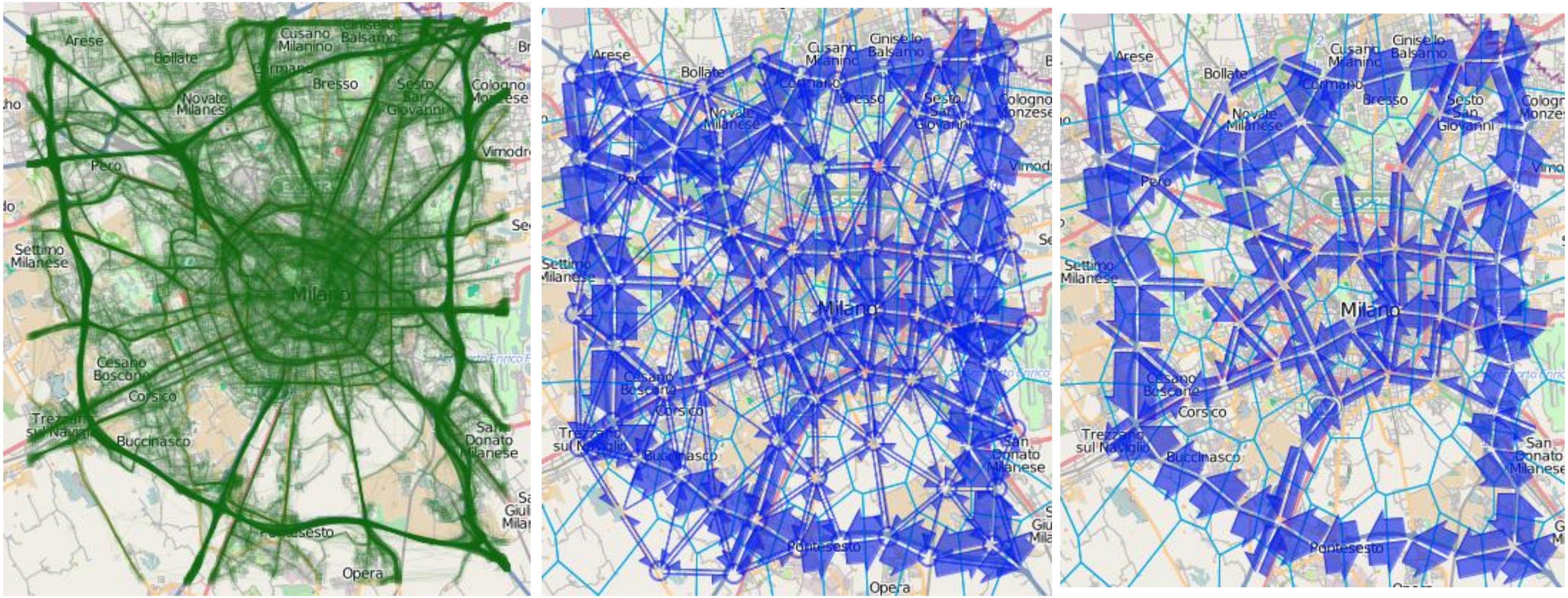
Stops >= 30 minutes in shopping areas ⇒ Shopping!

Spatial context:
- shopping areas



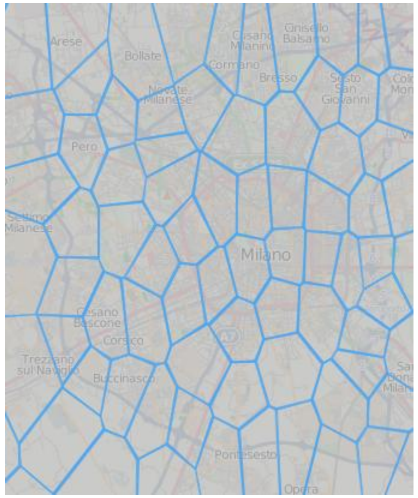


Another possible contribution: generalization

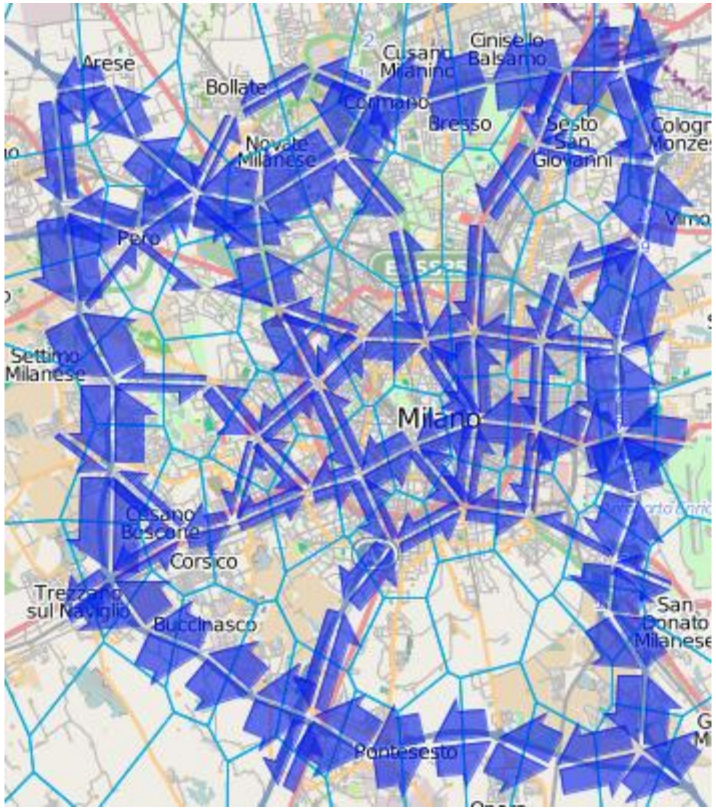


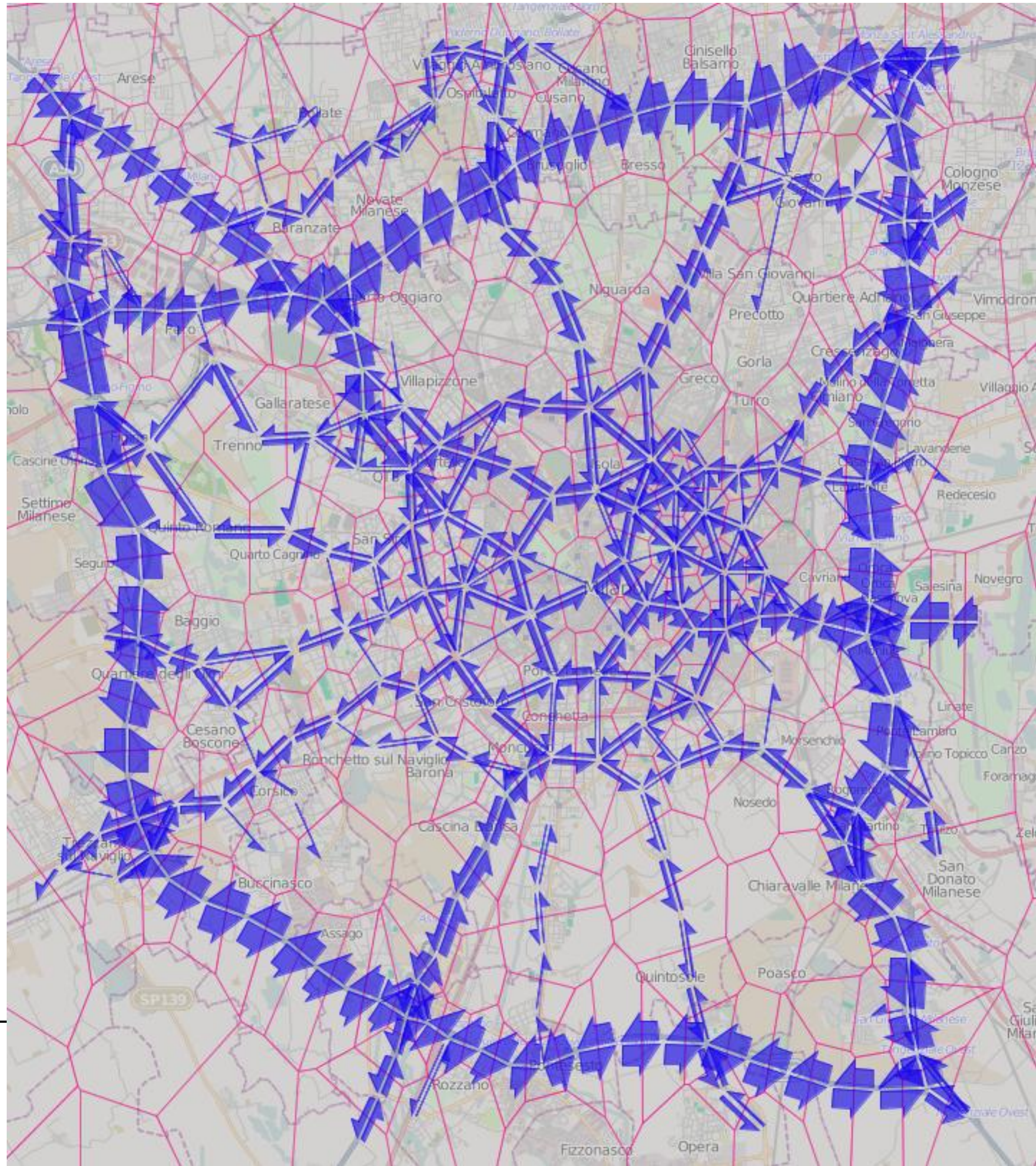
Generalization and aggregation are used to visualize large amounts of data and avoid display clutter.

Generalization as an approach to privacy protection



Variable generalization level depending on data density





Inter-disciplinary collaboration needed

Geovisual analytics + privacy protection research

Positive example:

A.Monreale, G.Andrienko, N.Andrienko, F.Giannotti, D.Pedreschi, S.Rinzivillo, S.Wrobel
Movement Data Anonymity through Generalization
Transactions on Data Privacy, 2010, v.**3** (3), pp. 91-121
http://www.tdp.cat/issues/abs.a045a10.php

Conclusion

- Space- and time-referenced data may allow extraction of sensitive personal information
- Geovisual analytics enables humans to establish links and make inferences
- Humans can do this more flexibly than computers
  - Do not require formal representation, can deal with incomplete information, can use previous knowledge and experience, …
- Researchers on privacy protection focus on automated analysis methods and may overlook the capabilities of human analysts
- Geovisual analytics should collaborate with privacy research
  - to identify potential risks to personal privacy from involving humans in analysis
  - to use visual analytics approaches for privacy protection

http://www.modap.org