Mobility, Data Mining and Privacy
Lessons from the GeoPKDD EU project

Fosca Giannotti and Dino Pedreschi

Pisa KDD Lab   www-kdd.isti.cnr.it

University of Pisa and ISTI-CNR, Italy

BiSS 09 – Bertinoro international Spring School

Mobile devices and services

- Large diffusion of mobile devices, mobile services and location-based services

Wireless networks as mobility data collectors

- Wireless networks infrastructures are the **nerves of our territory**
- besides offering their services, they gather highly informative **traces** about the human mobile activities
- UbiComp infrastructure will further push this phenomenon
- Miniaturization, wearability, pervasiveness will produce traces of increasing
  - positioning accuracy
  - semantic richness

Which mobility data?

- Location data from mobile phones, i.e. cell positions in the GSM/UMTS network.
- Location data from GPS-equipped devices – Galileo in the (near?) future
  - Next/current generation of Nokia mobile phones have on-board GPS receiver, and can transmit GPS tracks by SMS/MMS
- Location data from
  - peer-to-peer mobile networks
  - intelligent transportation environments – VANET
  - ad hoc sensor networks, RFIDs (radio-frequency ids)

Mobility, Data Mining and Privacy

- Towards an **archaeology of the present**?
- A scenario of great opportunities and risks:
  - mining mobility data can yield useful knowledge;
  - but, individual privacy is at risk.
- A new multidisciplinary research area is emerging at this crossroads, with potential for broad social and economic impact
  - F. Giannotti and D. Pedreschi (Eds.)
    Mobility, Data Mining and Privacy. Springer, 2008.

A paradigmatic project: ***GeoPKDD***

http://www.geopkdd.eu

A European FP6 project

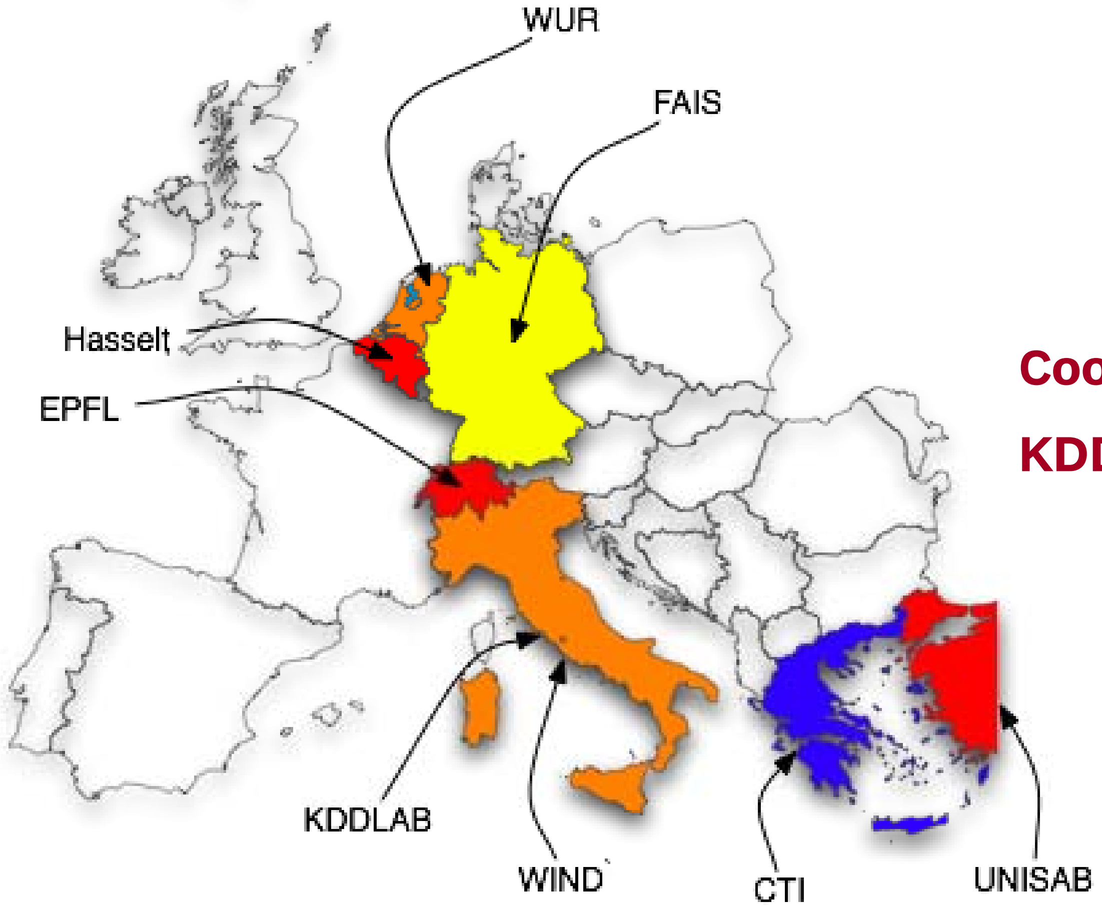
Geographic Privacy-aware

Knowledge Discovery and Delivery

www.geopkdd.eu

GeoPKDD

Geographic Privacy-aware Knowledge Discovery and Delivery



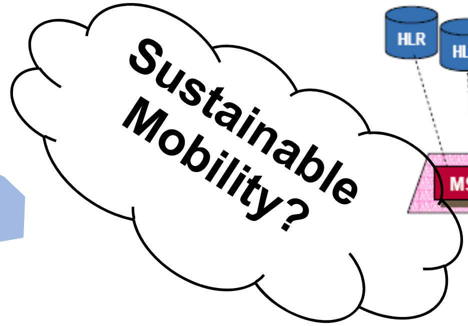
Coordinator:

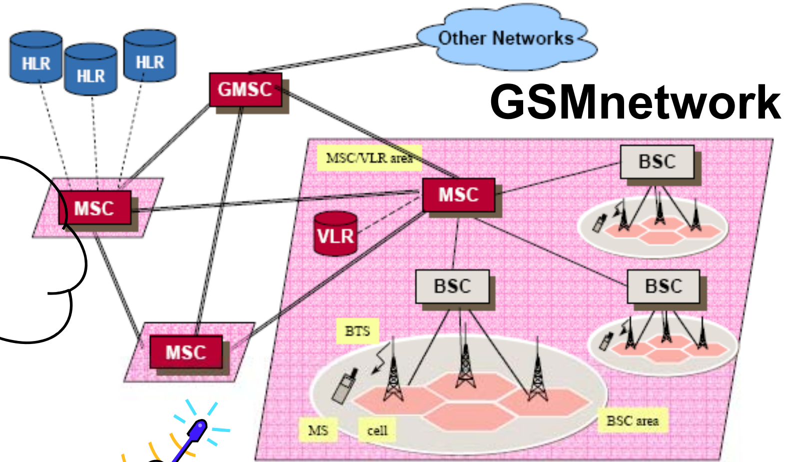
KDD-LAB Pisa, ISTI-CNR

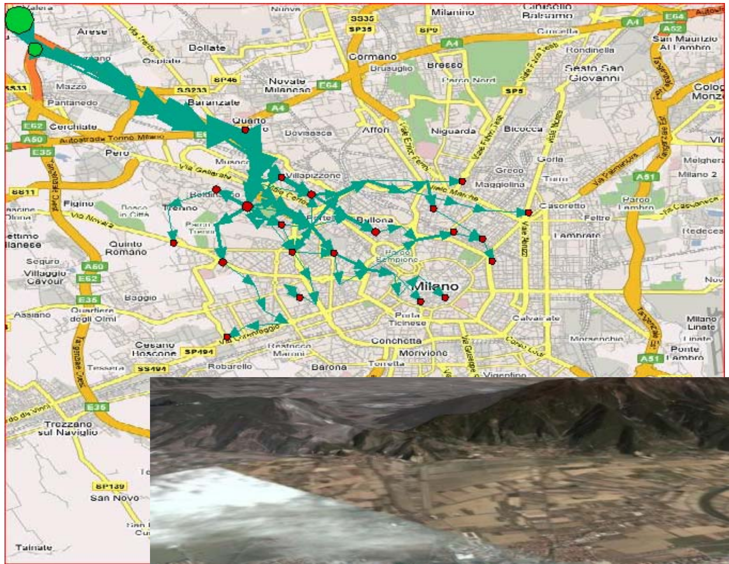
The GeoPKDD scenario

- **From the analysis of the traces of our mobile phones it is possible to reconstruct our mobile behaviour, the way we collectively move**
- **This knowledge may help us improving decision-making in many mobility-related issues:**
  - Planning traffic and public mobility systems in metropolitan areas;
  - Planning physical communication networks
  - Localizing new services in our towns
  - Forecasting traffic-related phenomena
  - Organizing logistics systems
  - Avoid repeating mistakes
  - Timely detecting changes.

Mobility Manager



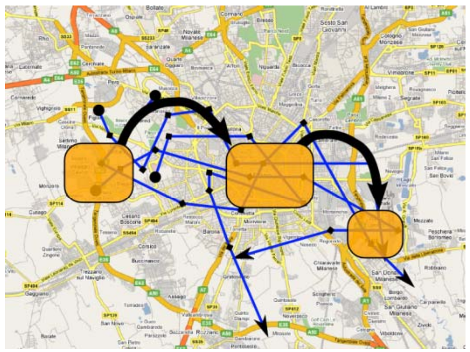




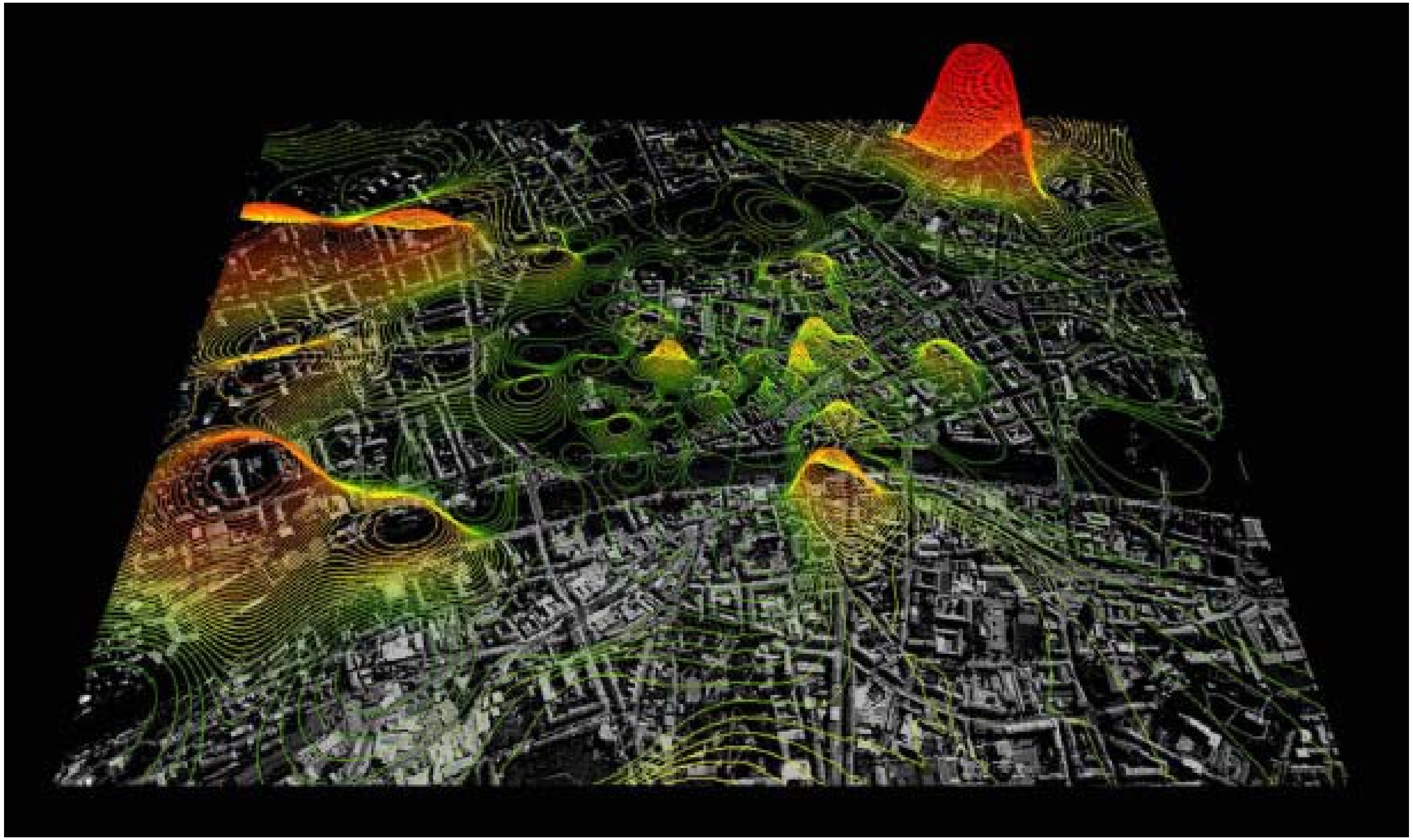

```
ate|y|x
in|08.20.1998|52.118|12.087
|08.23.1998|51.019|15.309
n|08.26.1998|47.723|22.786
in|08.29.1998|43.040|27.119
ssin|08.31.1998|38.715|32.165
essin|09.01.1998|37.195|35.255
zessin|09.03.1998|32.979|36.021
nzessin|09.05.1998|28.513|33.437
inzessin|09.06.1998|23.961|32.937
rinzessin|09.07.1998|19.418|33.446
Prinzessin|09.12.1998|15.823|34.094
Prinzessin|10.11.1998|14.685|32.848
Prinzessin|11.03.1998|11.510|32.591
Prinzessin|11.24.1998|13.888|35.667
Prinzessin|12.08.1998|12.562|34.777
Prinzessin|12.10.1998|9.124|35.644
...
```

Location data

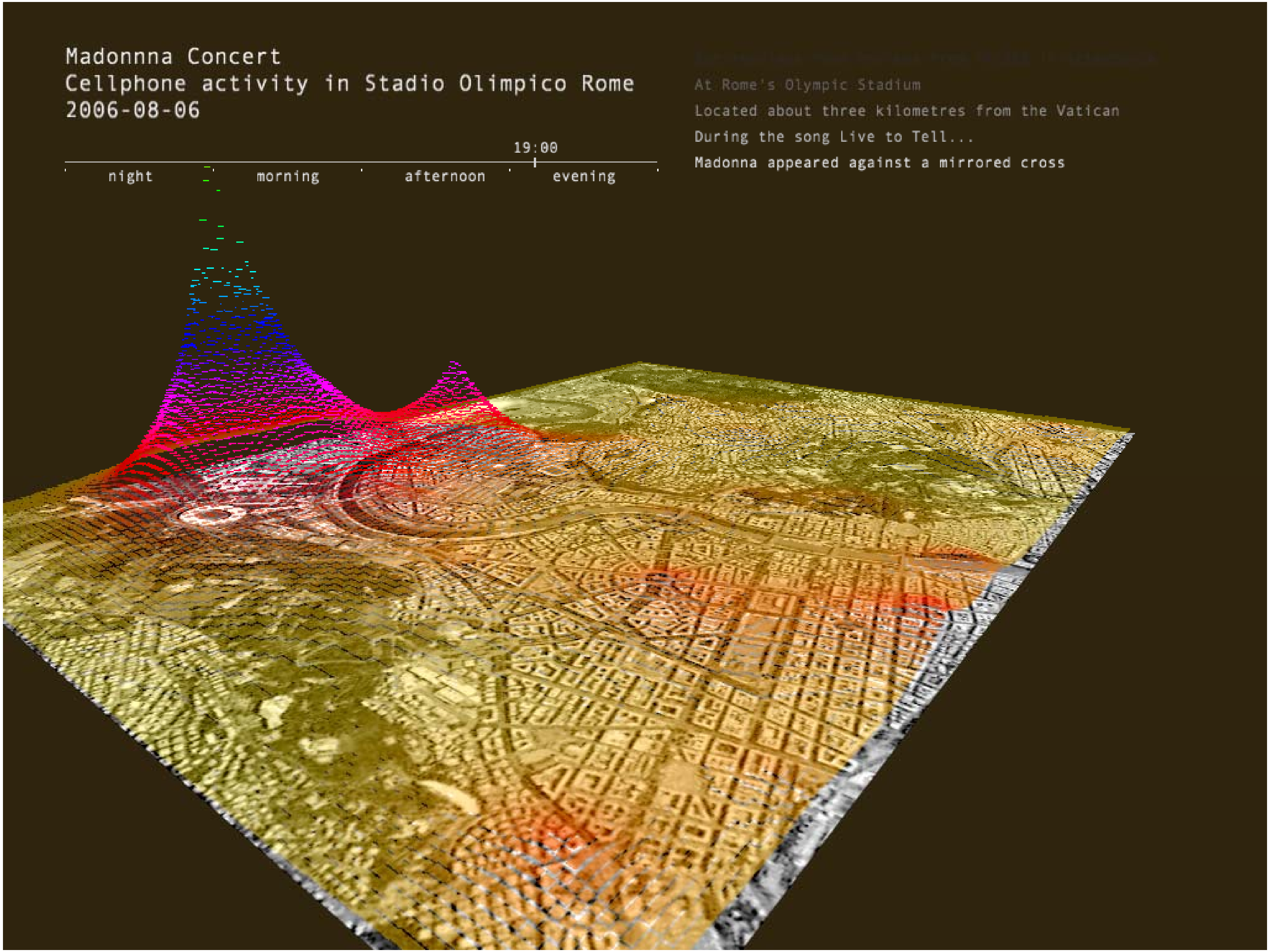
Mobility models



Real-time density estimation in urban areas



The senseable project: http://senseable.mit.edu/grazrealtime/

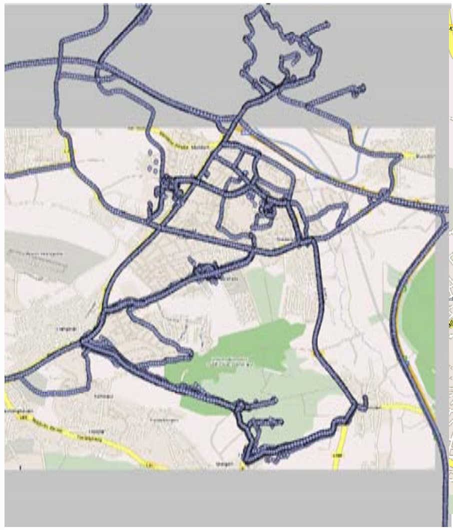

Madonnna Concert
Cellphone activity in Stadio Olimpico Rome
2006-08-06
19:00
night
morning
afternoon
evening
At Rome's Olympic Stadium
Located about three kilometres from the Vatican
During the song Live to Tell...
Madonna appeared against a mirrored cross

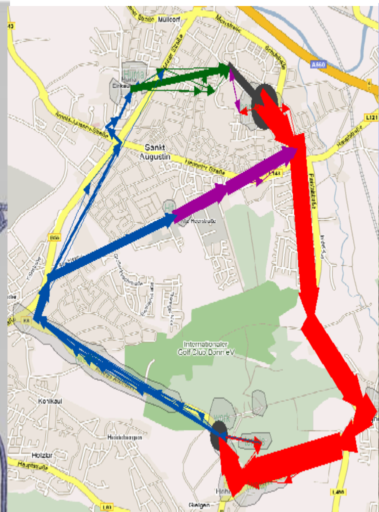
More ambitiously: mobility patterns

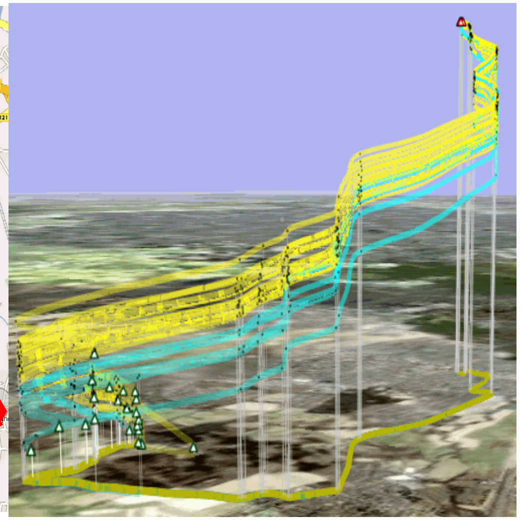
From mobility data to mobility patterns

From mobility data to mobility patterns

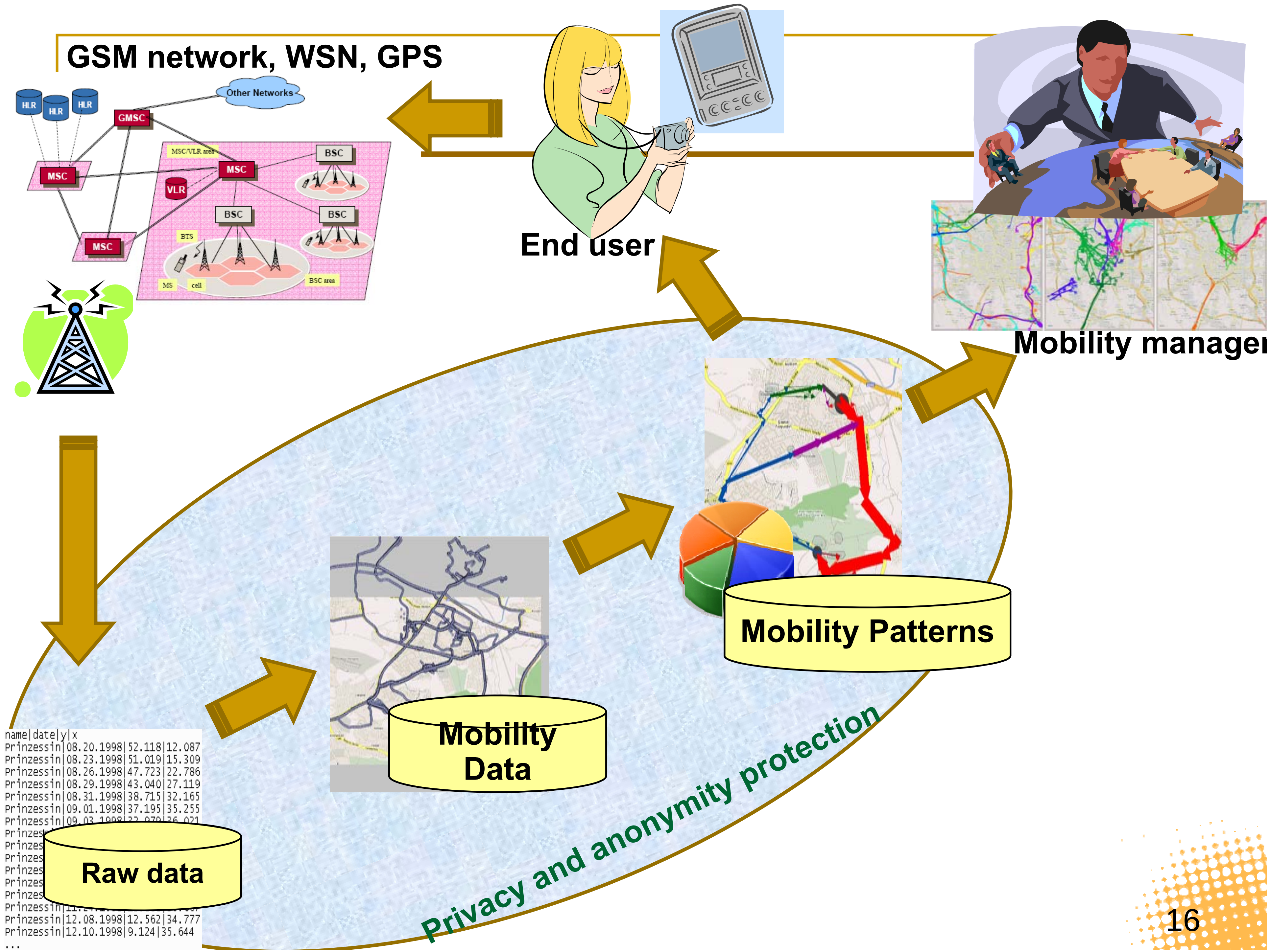
```
name|date|y|x
Prinzessin|08.20.1998|52.118|12.087
Prinzessin|08.23.1998|51.019|15.309
Prinzessin|08.26.1998|47.723|22.786
Prinzessin|08.29.1998|43.040|27.119
Prinzessin|08.31.1998|38.715|32.165
Prinzessin|09.01.1998|37.195|35.255
Prinzessin|09.03.1998|32.979|36.021
Prinzessin|09.05.1998|28.513|33.437
Prinzessin|09.06.1998|23.961|32.937
Prinzessin|09.07.1998|19.418|33.446
Prinzessin|09.12.1998|15.823|34.094
Prinzessin|10.11.1998|14.685|32.848
Prinzessin|11.03.1998|11.510|32.591
Prinzessin|11.24.1998|13.888|35.667
Prinzessin|12.08.1998|12.562|34.777
Prinzessin|12.10.1998|9.124|35.644
...
```

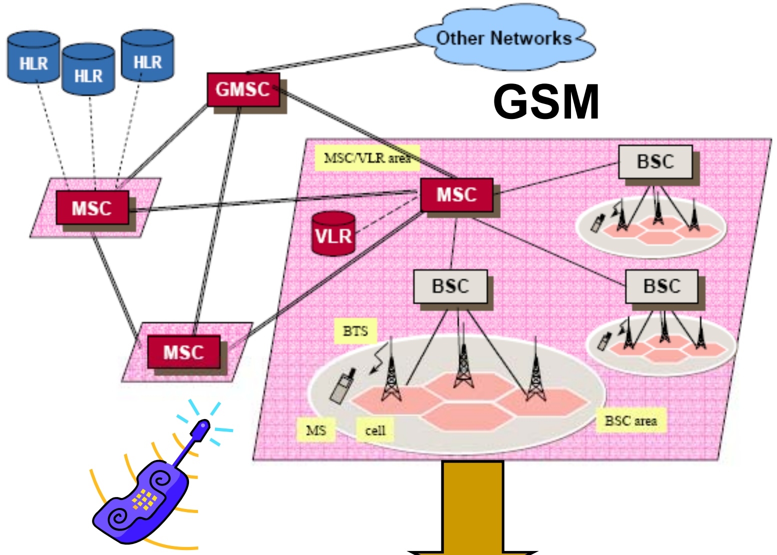




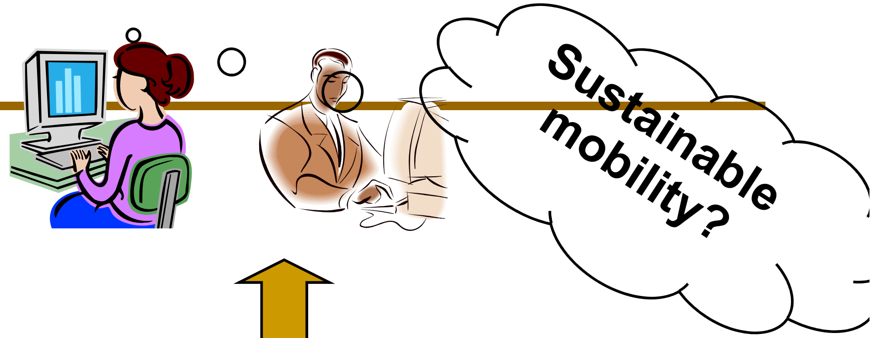


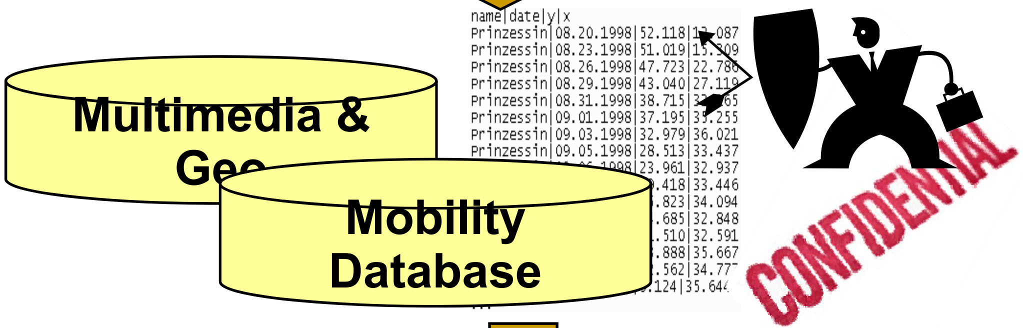
Mobility data mining and the Geographic Knowledge Discovery process

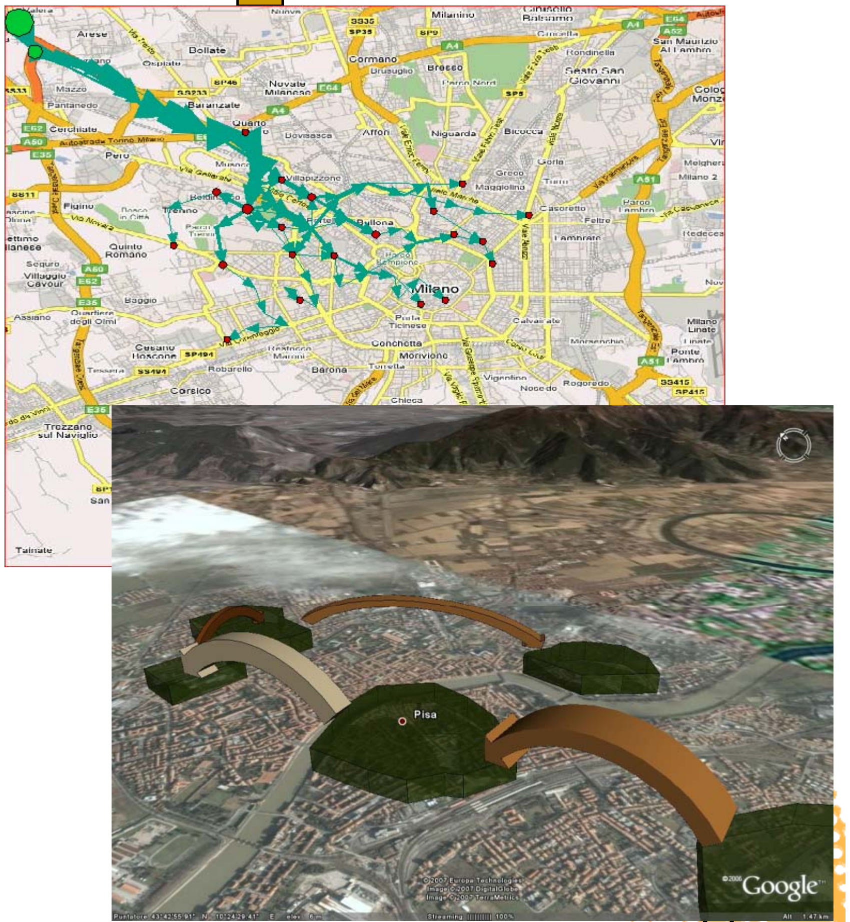

GSM network, WSN, GPS
Other Networks
HLR
HLR
HLR
GMSC
MSC
MSC
MSC
VLR
BSC
BSC
BSC
MSC/VLR area
BTS
MS
cell
BSC area
End user
Mobility manager
name|date|y|x
Prinzessin|08.20.1998|52.118|12.087
Prinzessin|08.23.1998|51.019|15.309
Prinzessin|08.26.1998|47.723|22.786
Prinzessin|08.29.1998|43.040|27.119
Prinzessin|08.31.1998|38.715|32.165
Prinzessin|09.01.1998|37.195|35.255
Prinzessin|12.08.1998|12.562|34.777
Prinzessin|12.10.1998|9.124|35.644
...
Raw data
Mobility Data
Mobility Patterns
Privacy and anonymity protection



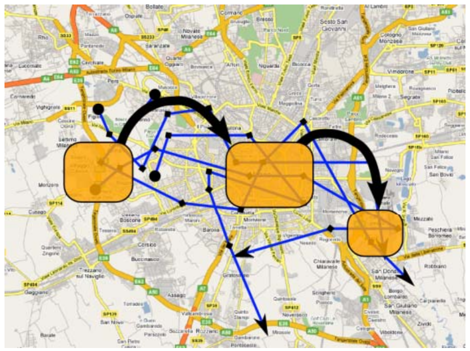
Mobility management

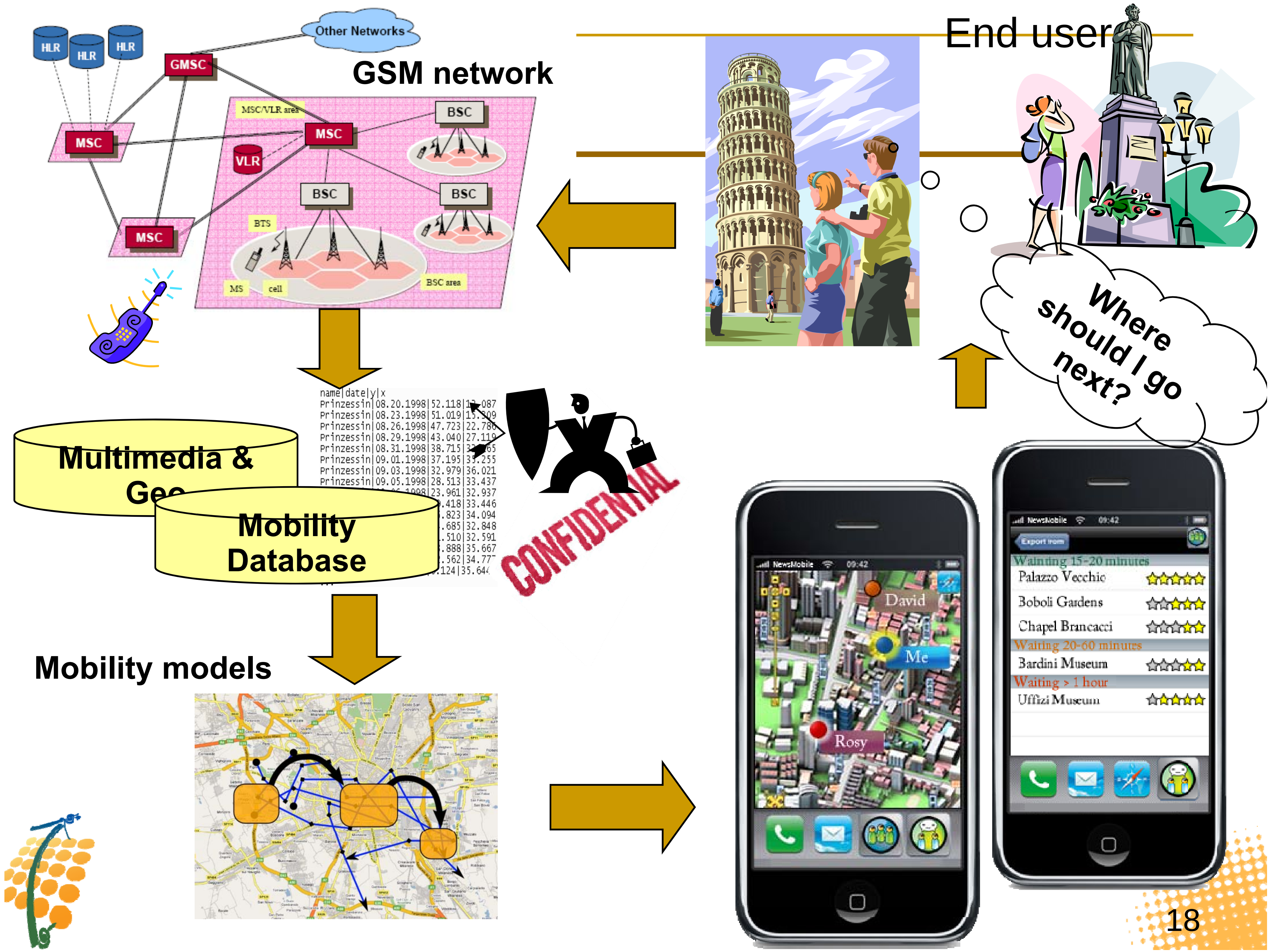






Mobility models





End user
Other Networks
GSM network
HLR
HLR
HLR
GMSC
MSC/VLR area
MSC
MSC
MSC
VLR
BSC
BSC
BSC
BSC
BTS
MS
cell
BSC area
Where should I go next?
name|date|y|x
Prinzessin|08.26.1998|47.723|22.786
Prinzessin|08.29.1998|43.040|27.119
Prinzessin|09.03.1998|32.979|36.021
Prinzessin|09.05.1998|28.513|33.437
CONFIDENTIAL
Multimedia &
Geo
Mobility
Database
Mobility models
NewsMobile 09:42
David
Me
Rosy
Export from
Waiting 15-20 minutes
Palazzo Vecchio
Boboli Gardens
Chapel Brancacci
Waiting 20-60 minutes
Bardini Museum
Waiting > 1 hour
Uffizi Museum

Key questions

- How to reconstruct a trajectory from raw logs, how to store and query trajectory data?
- How to classify trajectories according to means of transportation (pedestrian, private vehicle, public transportation vehicle, …)?
- Which spatio-temporal pattern and /models are useful abstractions of mobility data?
  - How to compute such patterns and models efficiently?
- Privacy protection and anonymity – how to make such concepts formally precise and measurable?
  - How to find an optimal trade-off between privacy protection and quality of the analysis?

GeoPKDD highlights

- Trajectory DB Management System and DW
  - Theodoridis and colleagues, Athens, Raffaetà and colleagues Venice
- A repertoire of mobility patterns and models
  - Nanni, Pedreschi and colleagues, Pisa
- A visual analytics environment for mobility data
  - Andrienko's, Fraunhofer – Rinzivillo, Pedreschi, Pisa
- A repertoire of PP analysis techniques
  - Saygin, Istanbul – Bonchi, Giannotti, Pedreschi, Pisa – Damiani, Milan
- A mobility data mining query language
  - Giannotti, Manco, Renso and colleagues, Pisa + Cosenza
- A reasoning framework for mobility data mining applications
  - Macedo, Spaccapietra, EPFL + Renso, Pisa + Wachowicz, Madrid

GeoPKDD highlights

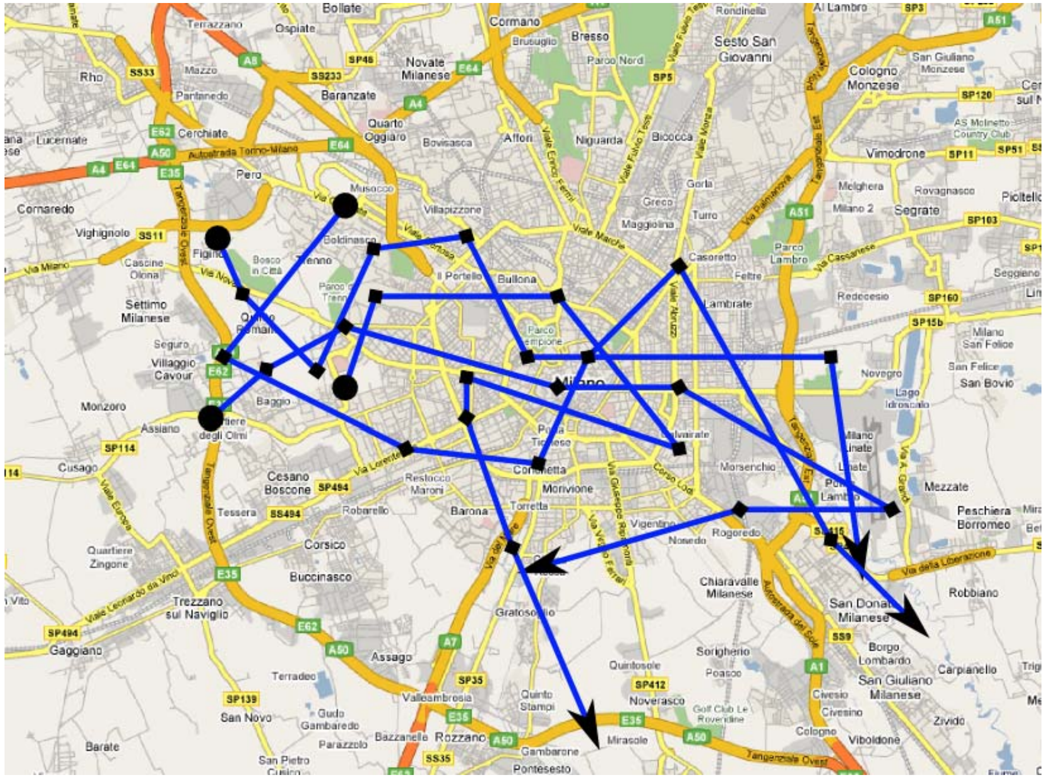
- Trajectory DB Management System and DW
  - Theodoridis and colleagues, Athens, Raffaetà and colleagues Venice
- A repertoire of mobility patterns and models
  - Nanni, Pedreschi and colleagues, Pisa
- A visual analytics environment for mobility data
  - Andrienko's, Fraunhofer – Rinzivillo, Pedreschi, Pisa
- A repertoire of PP analysis techniques
  - Saygin, Istanbul – Bonchi, Giannotti, Pedreschi, Pisa – Damiani, Milan
- A mobility data mining query language
  - Giannotti, Manco, Renso and colleagues, Pisa + Cosenza
- A reasoning framework for mobility data mining applications
  - Macedo, Spaccapietra, EPFL + Renso, Pisa + Wachowicz, Madrid

Spatio-temporal data mining

Trajectory Pattern Mining

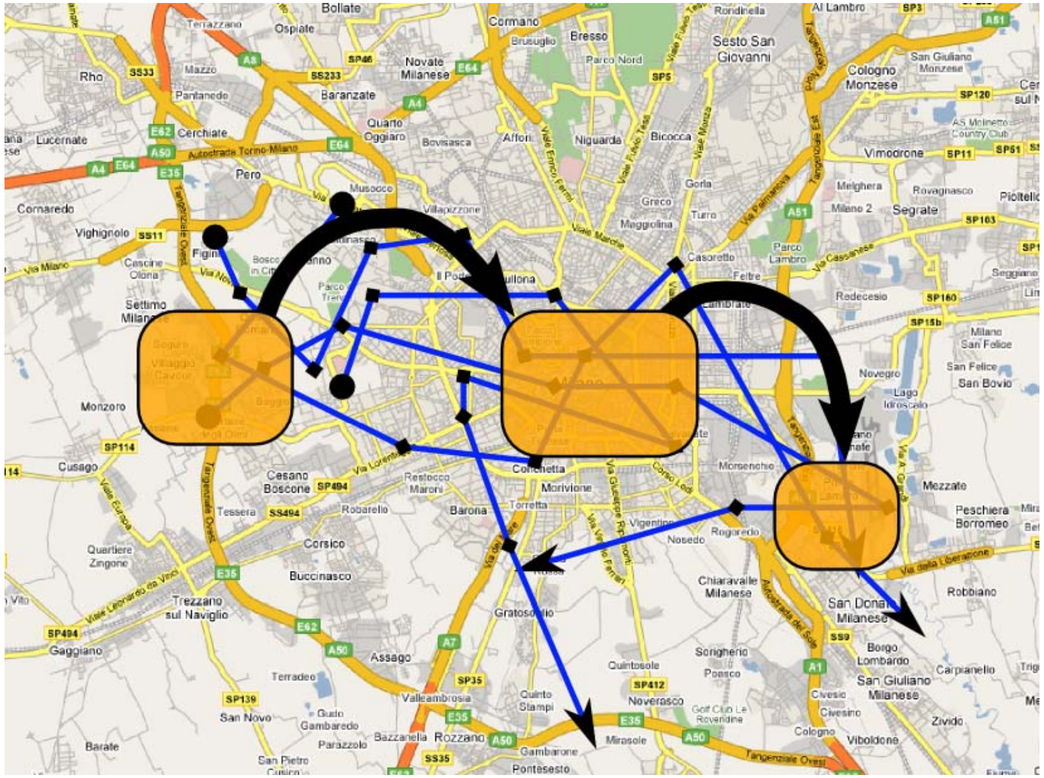
Trajectory Clustering

***Q**: What is a trajectory pattern?*



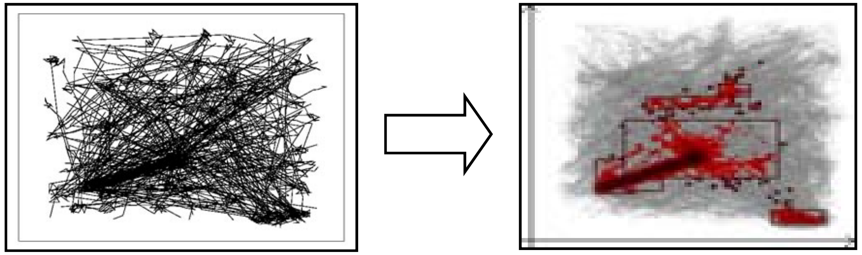
A: *A spatio-temporal sequential pattern*

- A sequence of visited regions, **frequently** visited in the **specified order** with **similar transition times**



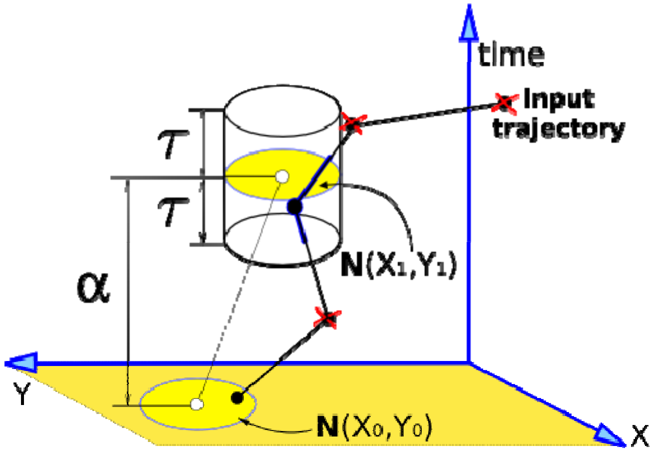
- Giannotti, Nanni, Pedreschi, Pinelli.
  Trajectory pattern mining. In Proc. ACM SIGKDD 2007

T-Pattern discovery

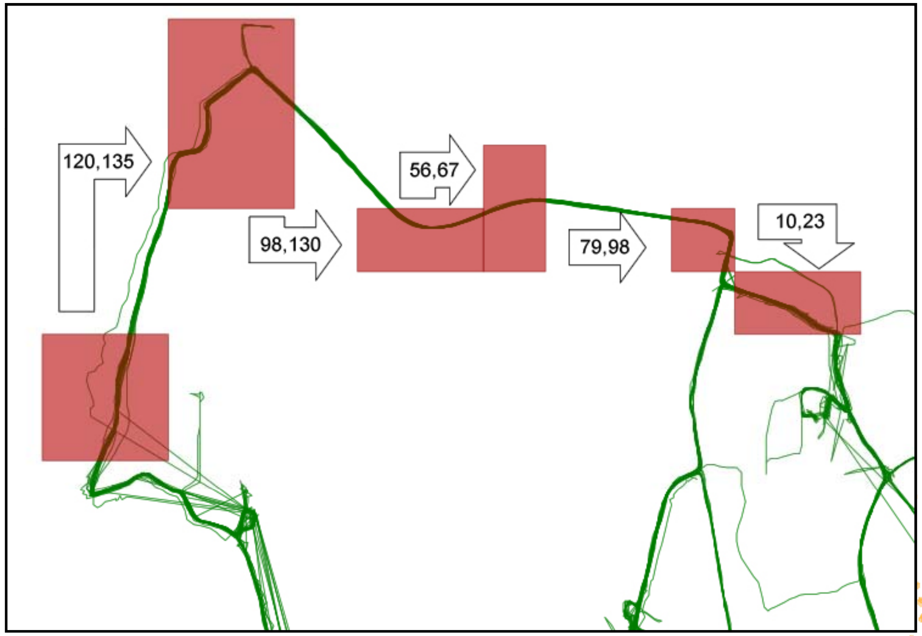


1- Find Regions of Interest

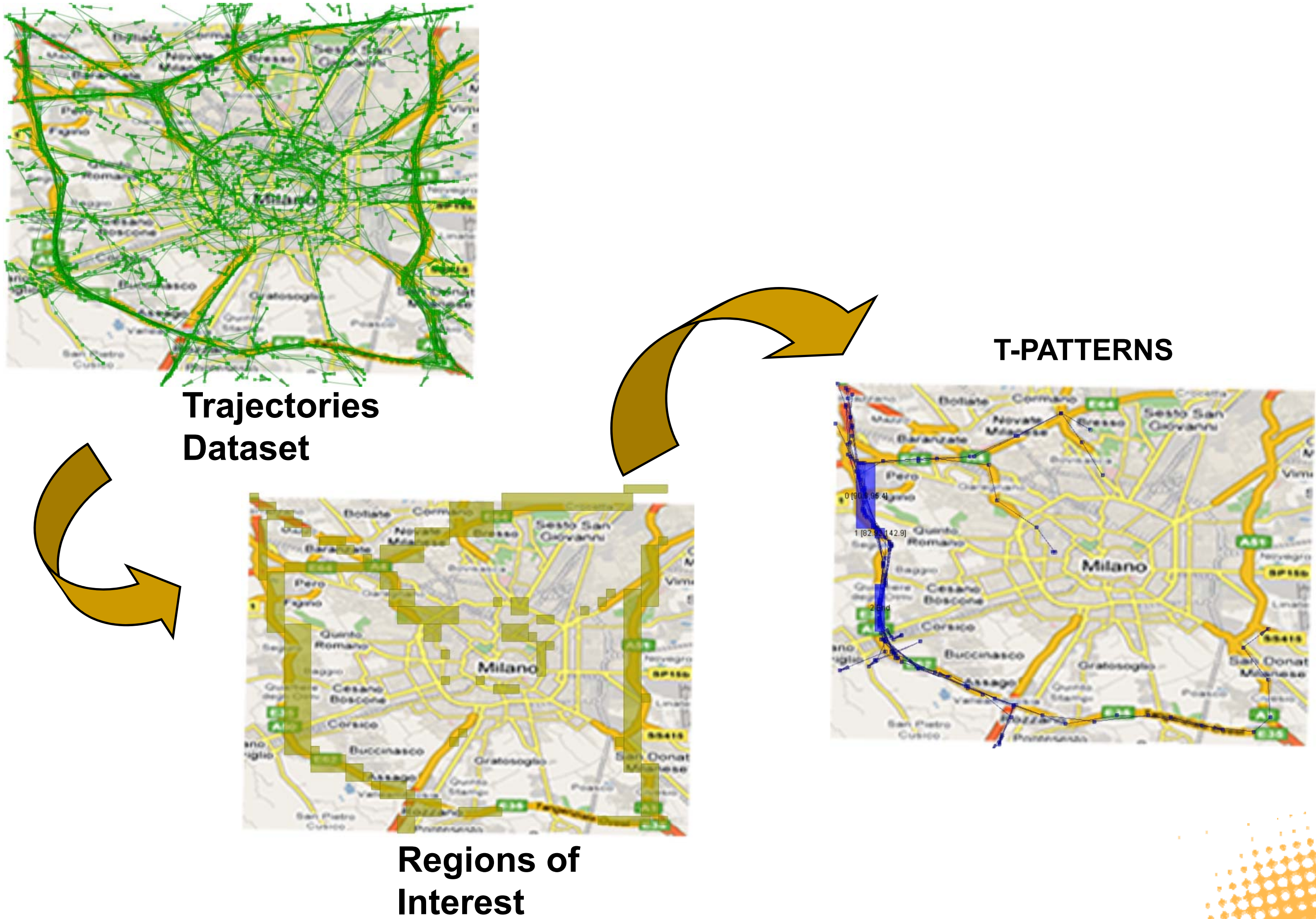
2- Find similar Trajectory in space and time



3- Extract patterns:



T-Pattern: Extraction Process



T-Patterns for trajectories

- A **Trajectory Pattern** (T-pattern) is a pair (**s**, α):
  - **s** = $<(x_0,y_0),\ldots,(x_k,y_k)>$ is a sequence of k+1 locations
  - $\alpha = <\alpha_1,\ldots,\alpha_k>$ are the transition times (*annotations*)

  also written as: $$(x_0, y_0) \xrightarrow{\alpha_1} (x_1, y_1) \xrightarrow{\alpha_2} \cdots \xrightarrow{\alpha_k} (x_k, y_k)$$

- A T-pattern T_p **occurs** in a trajectory if it contains a sub-sequence S such that:
  - each (x_i,y_i) in T_p matches a point (x_i',y_i') in S, and
  - the transition times in T_p are similar to those in S

Continuity issues (space & time)

- The same exact spatial location (x,y) usually never occurs twice

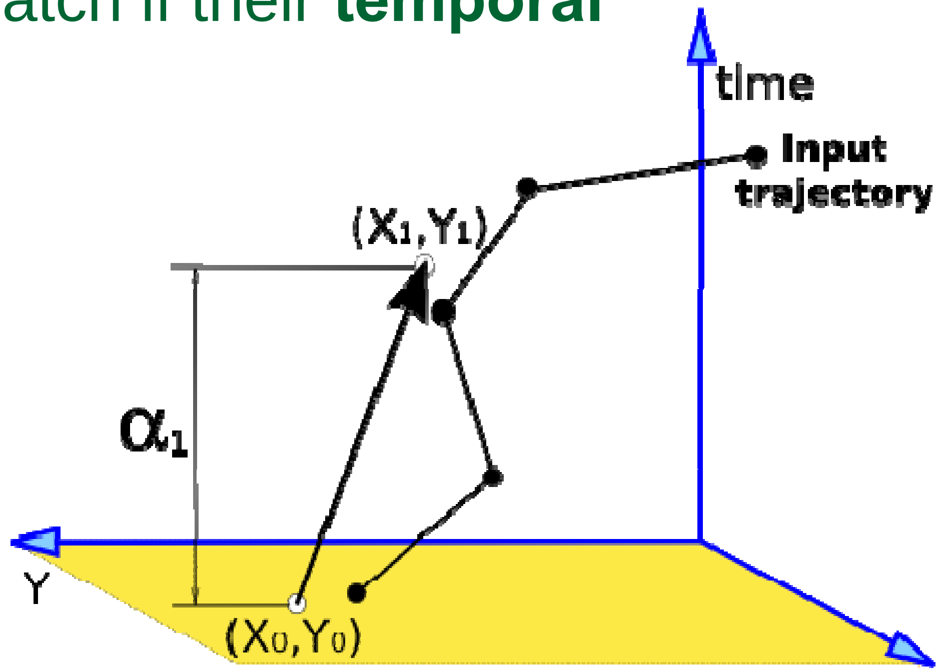
- The same exact transition times usually do not occur twice

- Solution: allow approximation
  - a notion of *spatial neighborhood*
  - a notion of *temporal tolerance*

T-Pattern: approximate *occurrence*

- Two points match if one falls within a **spatial neighborhood N()** of the other
- Two transition times match if their **temporal difference is ≤ τ**
- Example:

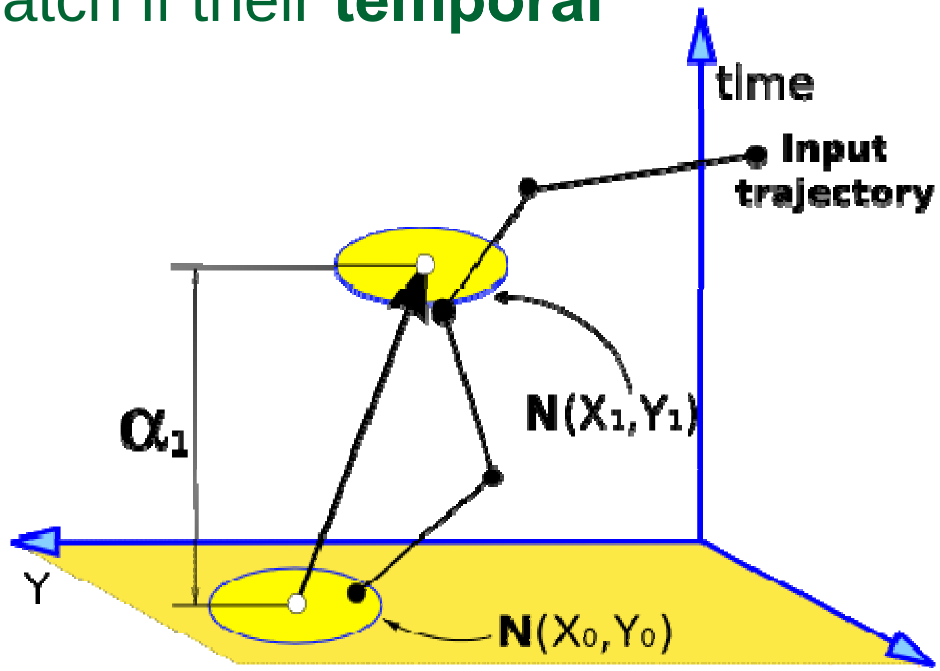
$$(x_0, y_0) \xrightarrow{\alpha_1} (x_1, y_1)$$



T-Pattern: approximate *occurrence*

- Two points match if one falls within a **spatial neighborhood N()** of the other
- Two transition times match if their **temporal difference is ≤ τ**
- Example:

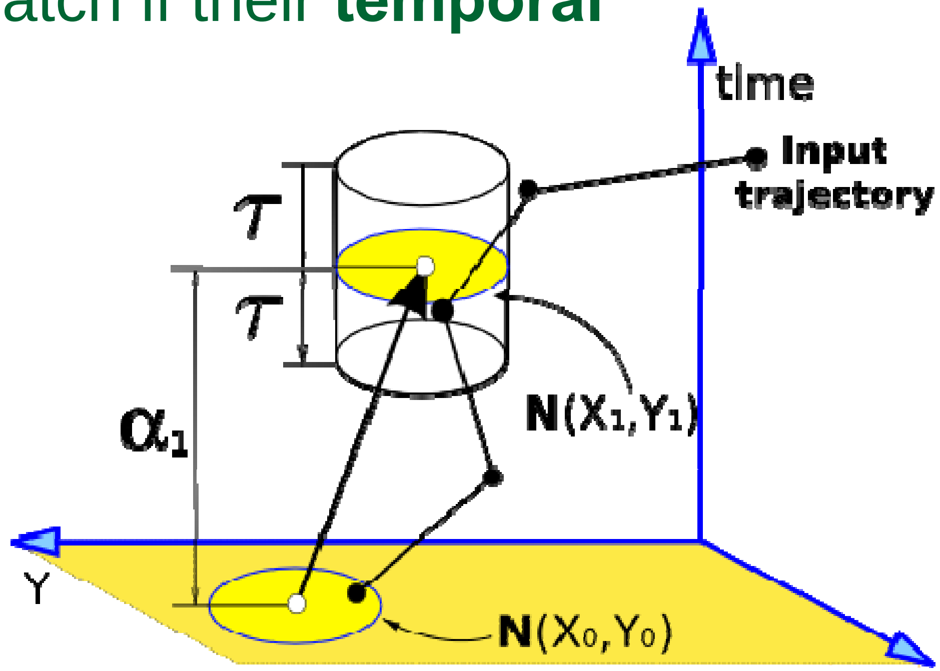
$$(x_0, y_0) \xrightarrow{\alpha_1} (x_1, y_1)$$



T-Pattern: approximate *occurrence*

- Two points match if one falls within a **spatial neighborhood N()** of the other
- Two transition times match if their **temporal difference is ≤ τ**
- Example:

$$(x_0, y_0) \xrightarrow{\alpha_1} (x_1, y_1)$$



Computing general T-Patterns

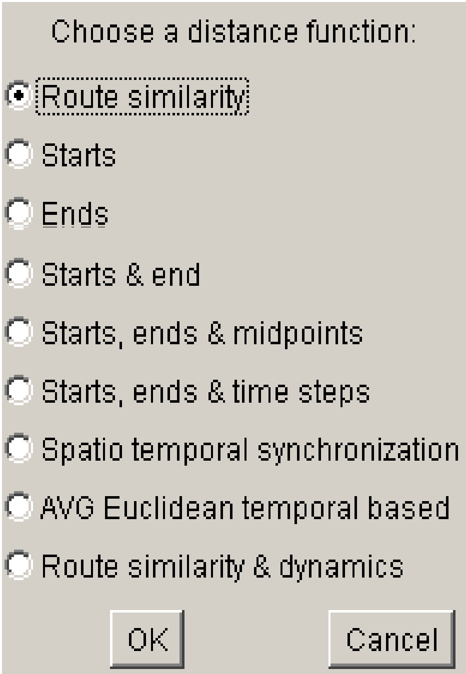
- T-pattern mining can be mapped to a density estimation problem over R^{3n-1}
  - 2 dimensions for each (x,y) in the pattern (2n)
  - 1 dimension for each transition (n-1)
- Density computed by
  - mapping each sub-sequence of n points of each input trajectory to R^{3n-1}
  - drawing an influence area for each point (composition of **N()** and **τ**)
- Too computationally expensive, heuristics needed!!!

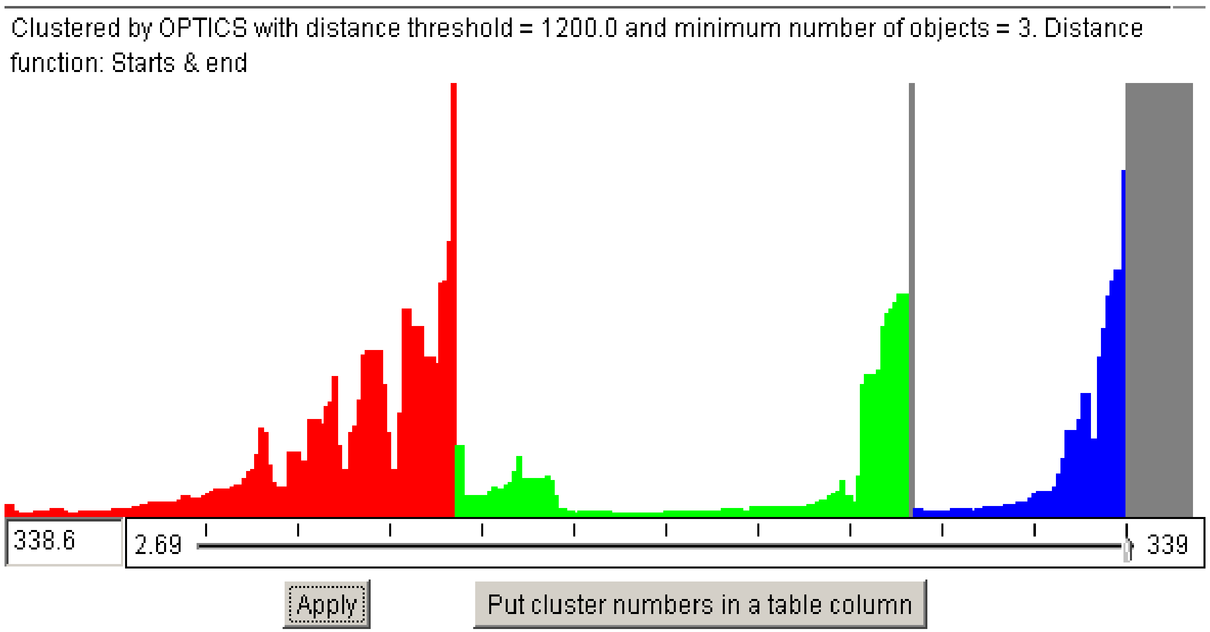
Spatio-temporal data mining

Trajectory Pattern Mining

Trajectory Clustering

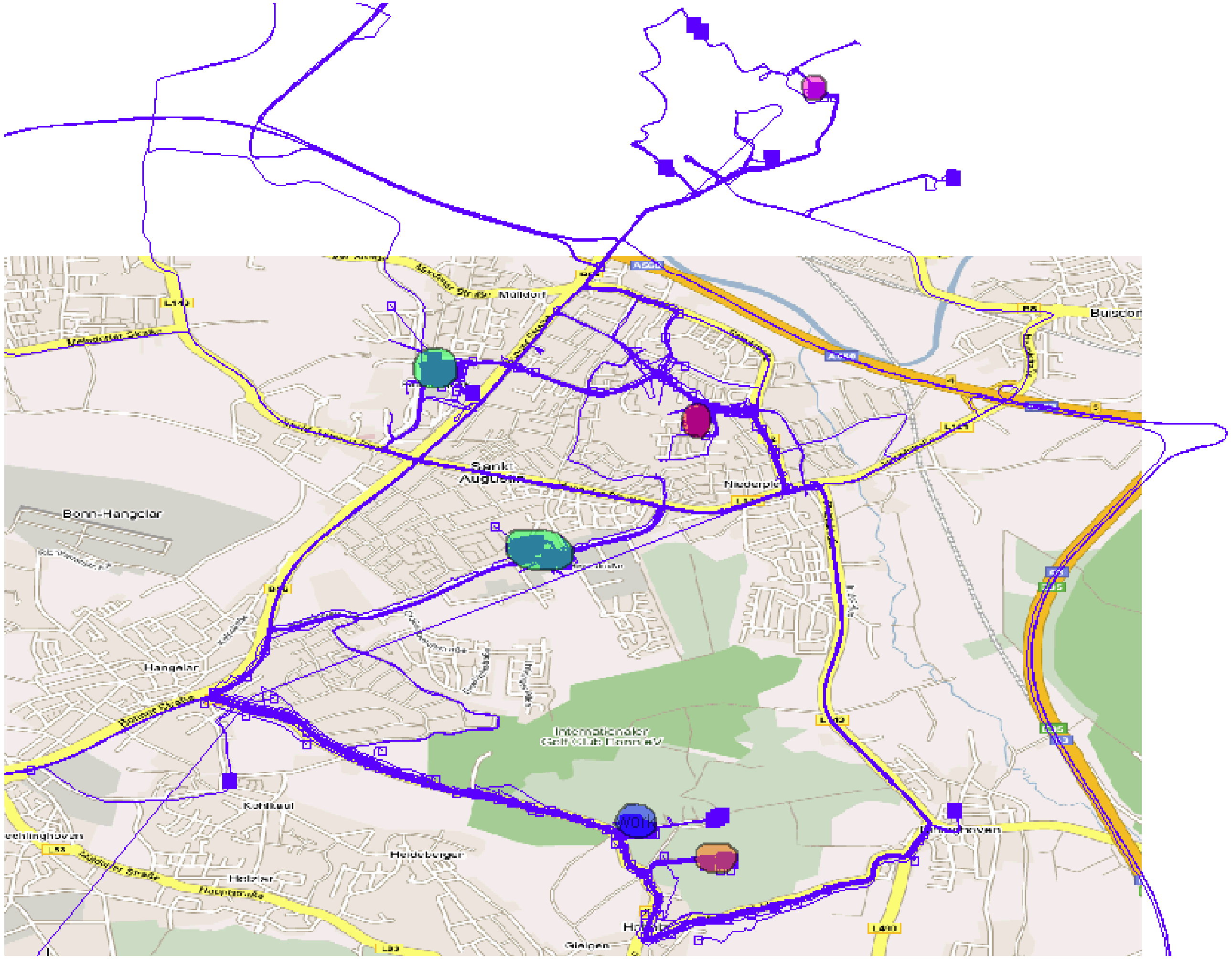
Interactive density-based trajectory clustering



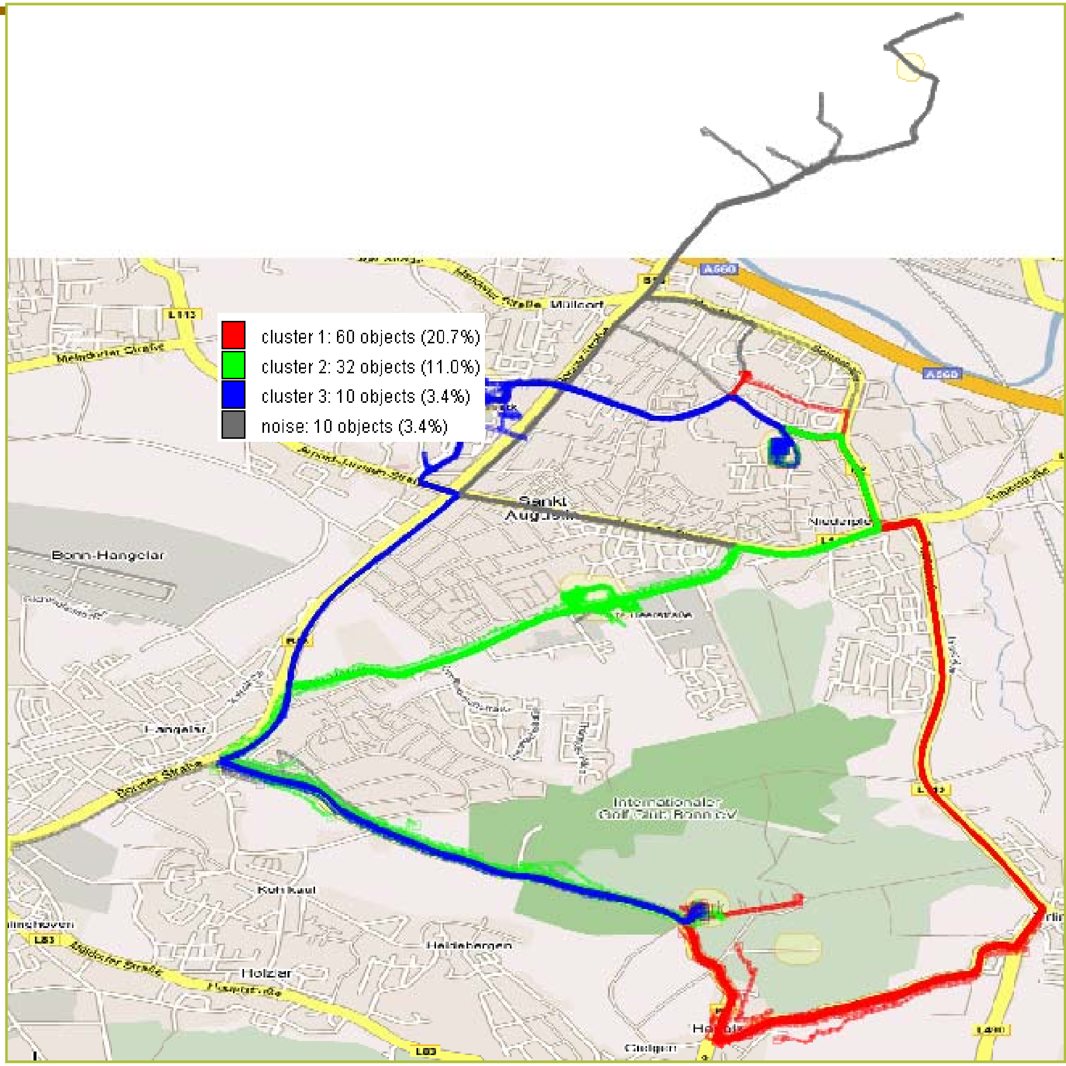


- Nanni, Pedreschi.
  Time-focused clustering of trajectories of moving objects.
  J. of Intelligent Information Systems, 2006
- Rinzivillo, Pedreschi, Nanni, Giannotti, Andrienko, Andrienko.
  Visually-driven analysis of movement data by progressive clustering. J. of Information Visualization, 2008

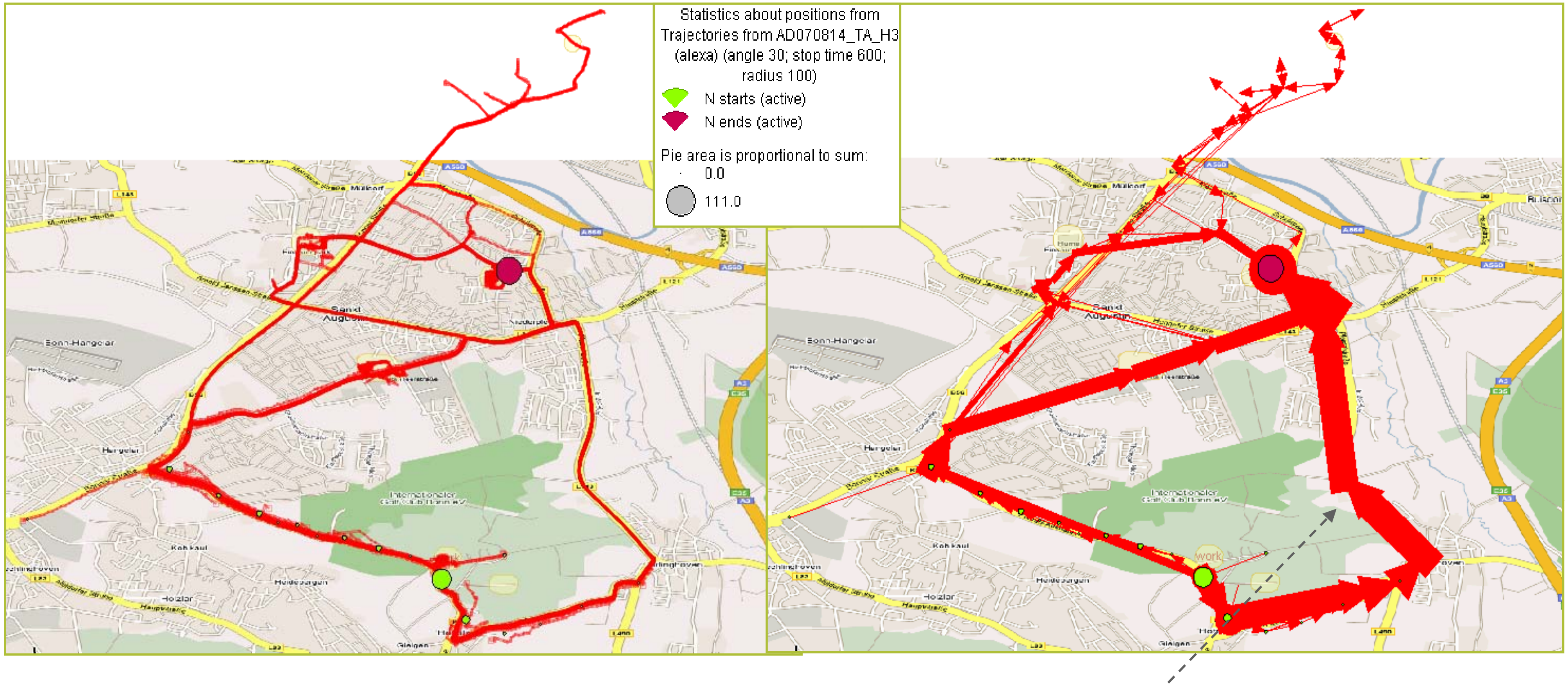
Looking for frequent stops & moves



Clusters of typical trips

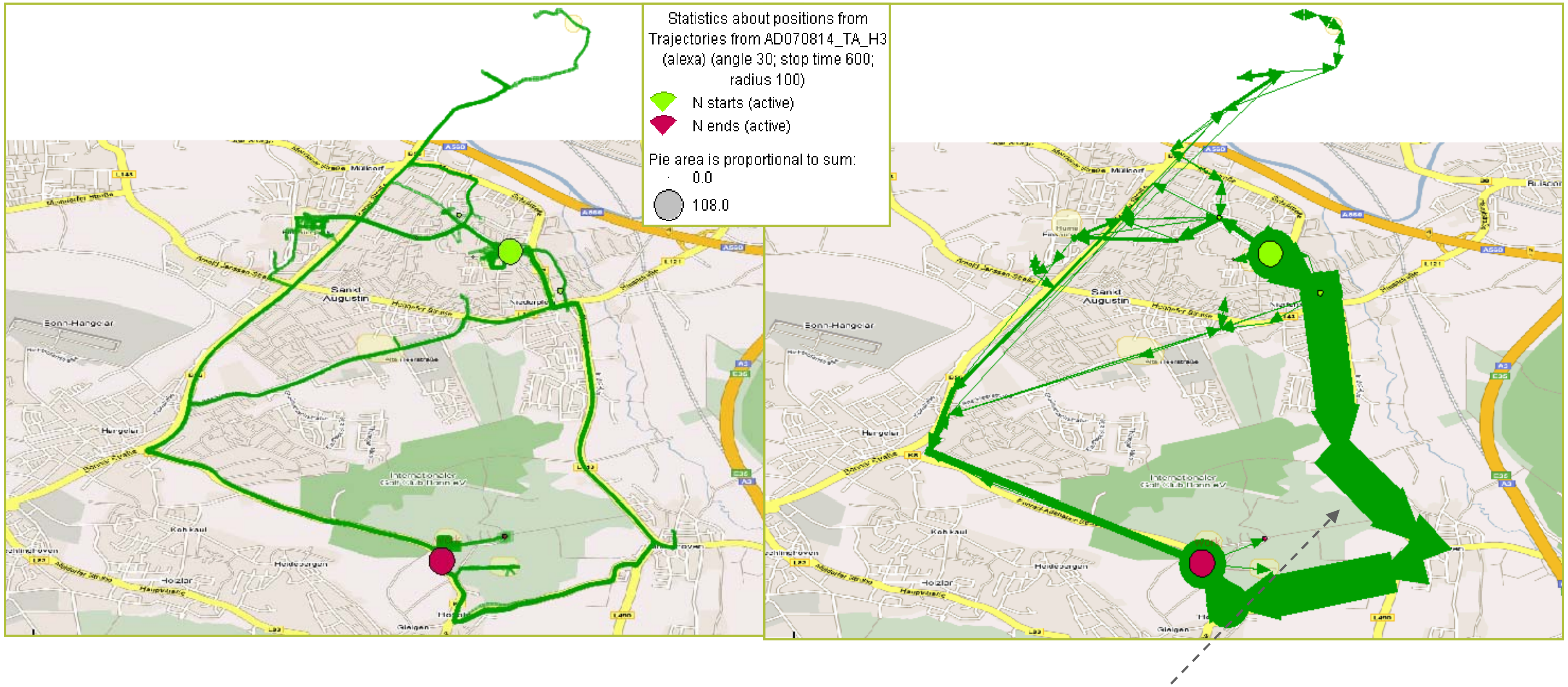


Cluster 1: from work to home



Observation: the eastern route is chosen more often

Cluster 2: from home to work

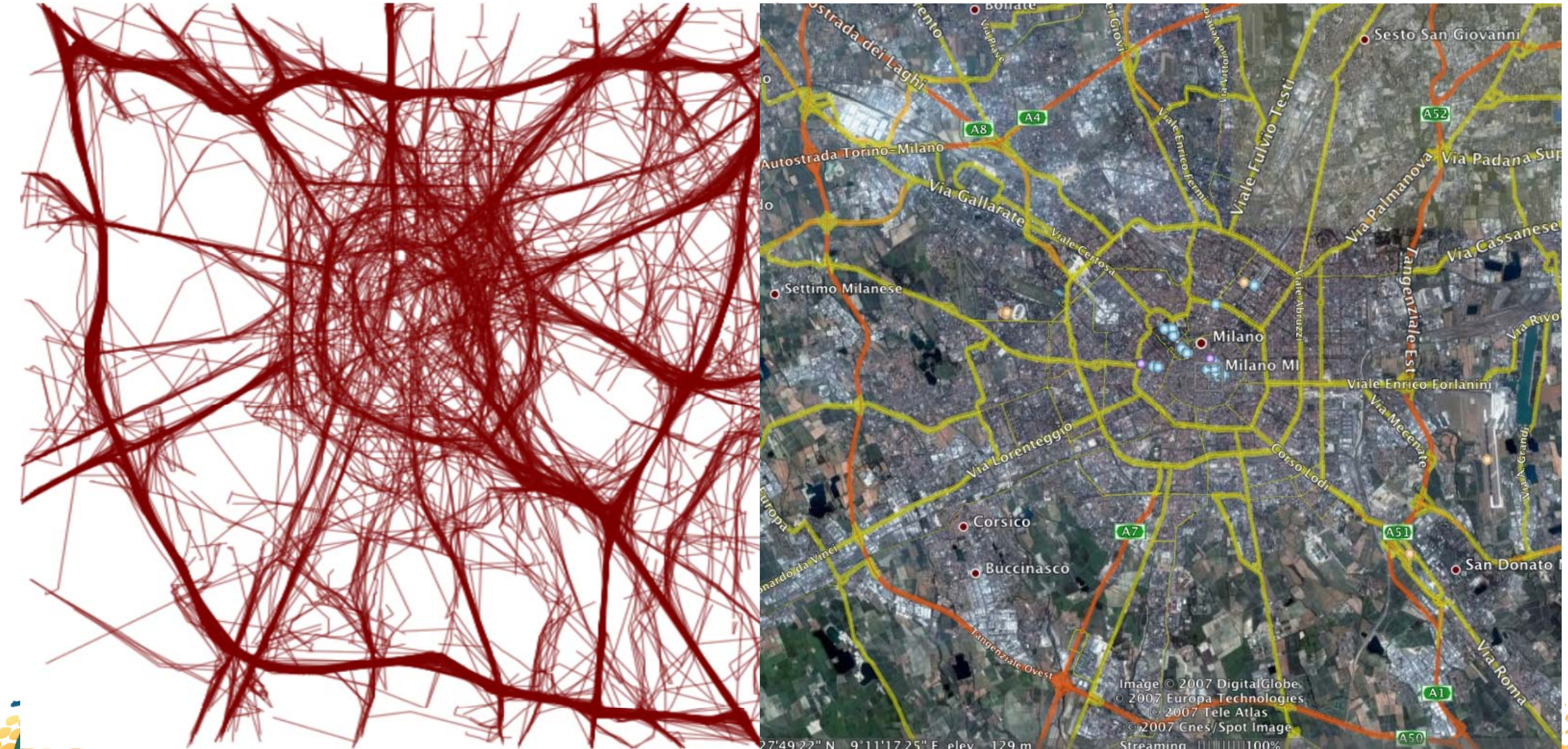


Observation: the eastern route is chosen much more often

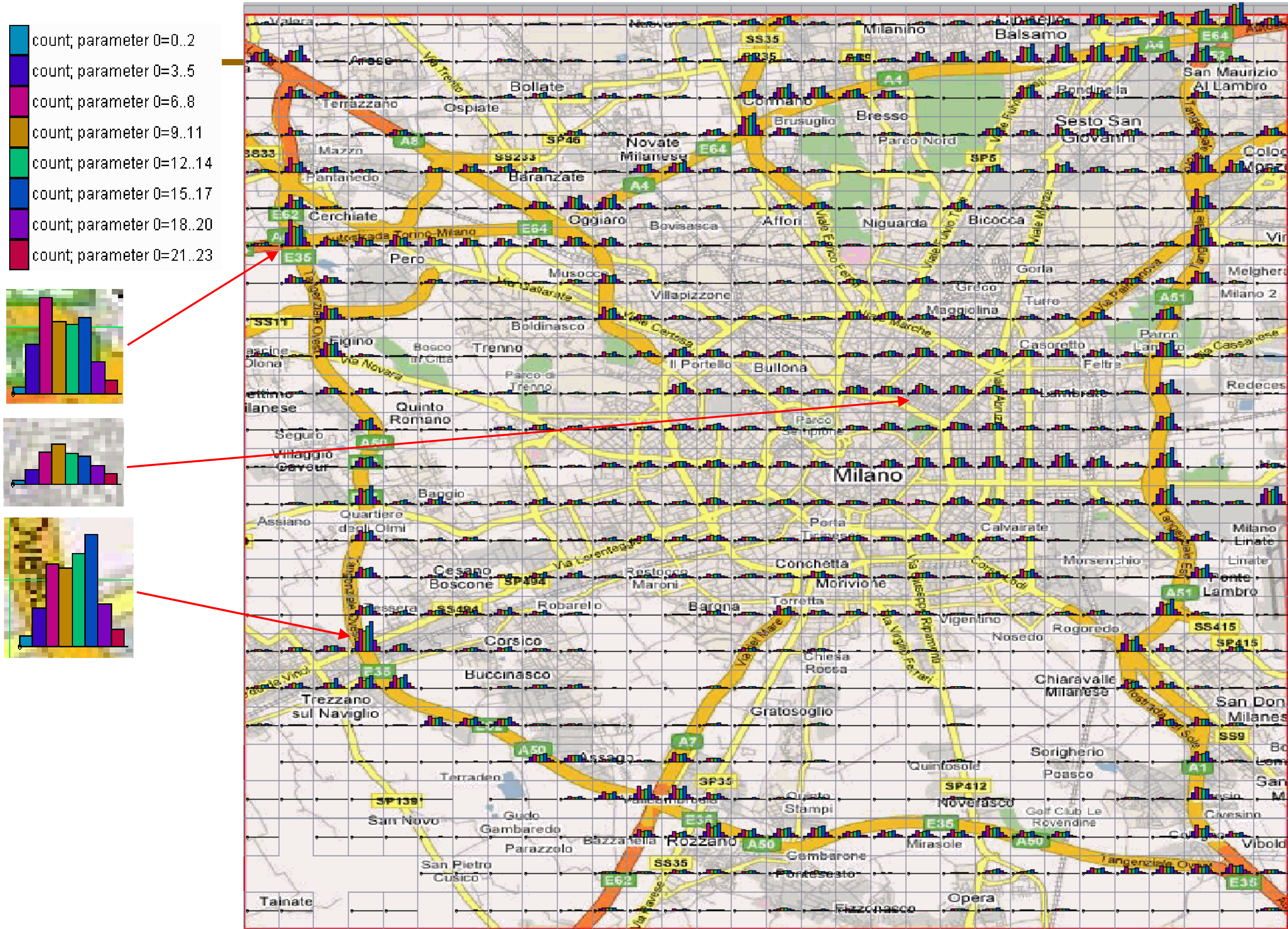
Mobility data analysis in Milano

- WIND Telecomunicazioni spa (major telecom provider, GeoPKDD partner)
  - GSM data (Handover data: aggregated flows between adjacent cells)
- Other collaborations:
  - Comune di Milano, Mobility Agency
  - Infoblu and OctoTelematics (GPS receivers on board of cars with special insurance contract)
- Experience on a a dataset of
  - 2 M positions,
  - 17 K vehicles,
  - 200 K trajectories

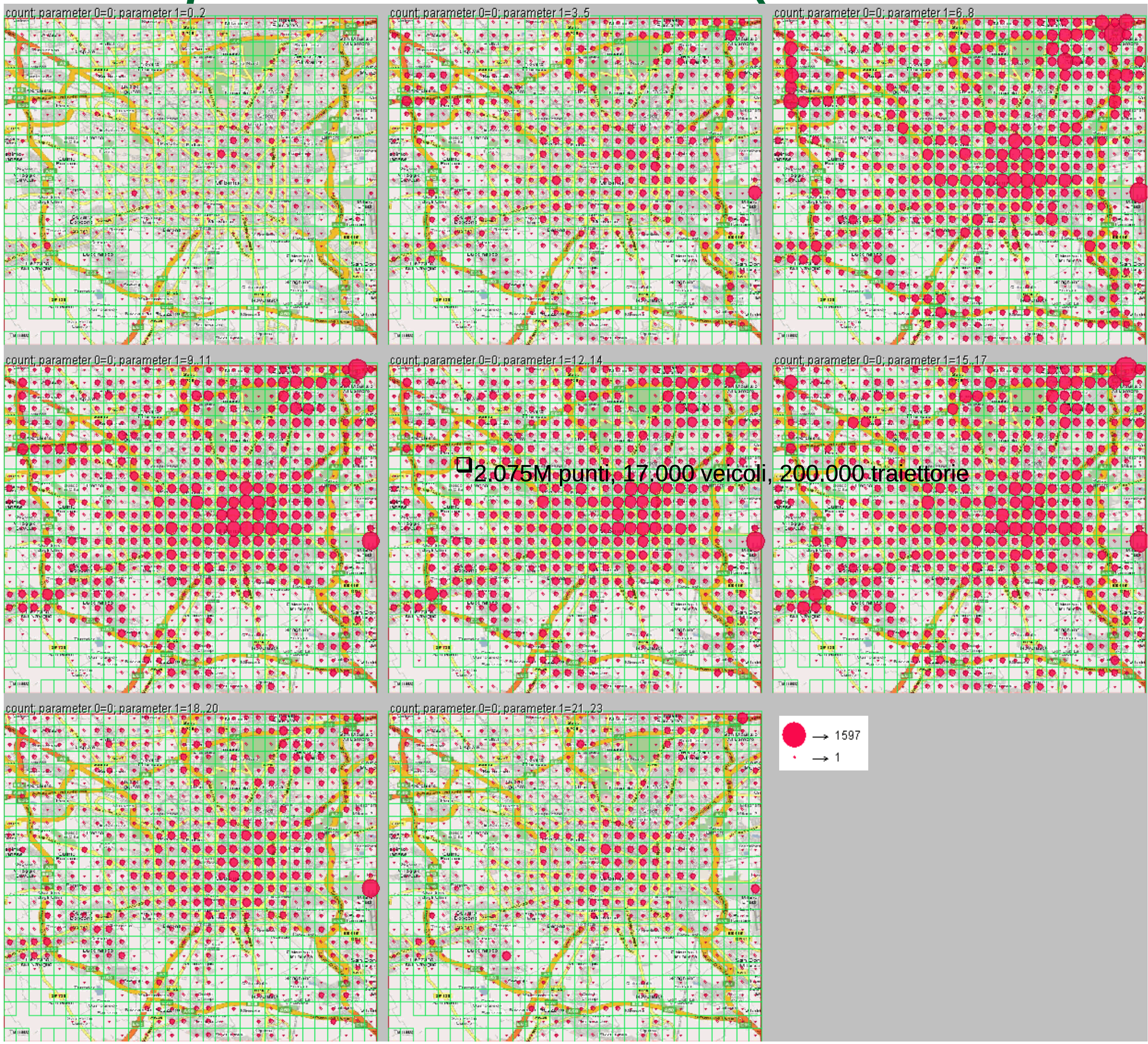
MILANO: data on the map



Traffic density patterns (spatio-temporal aggregation)



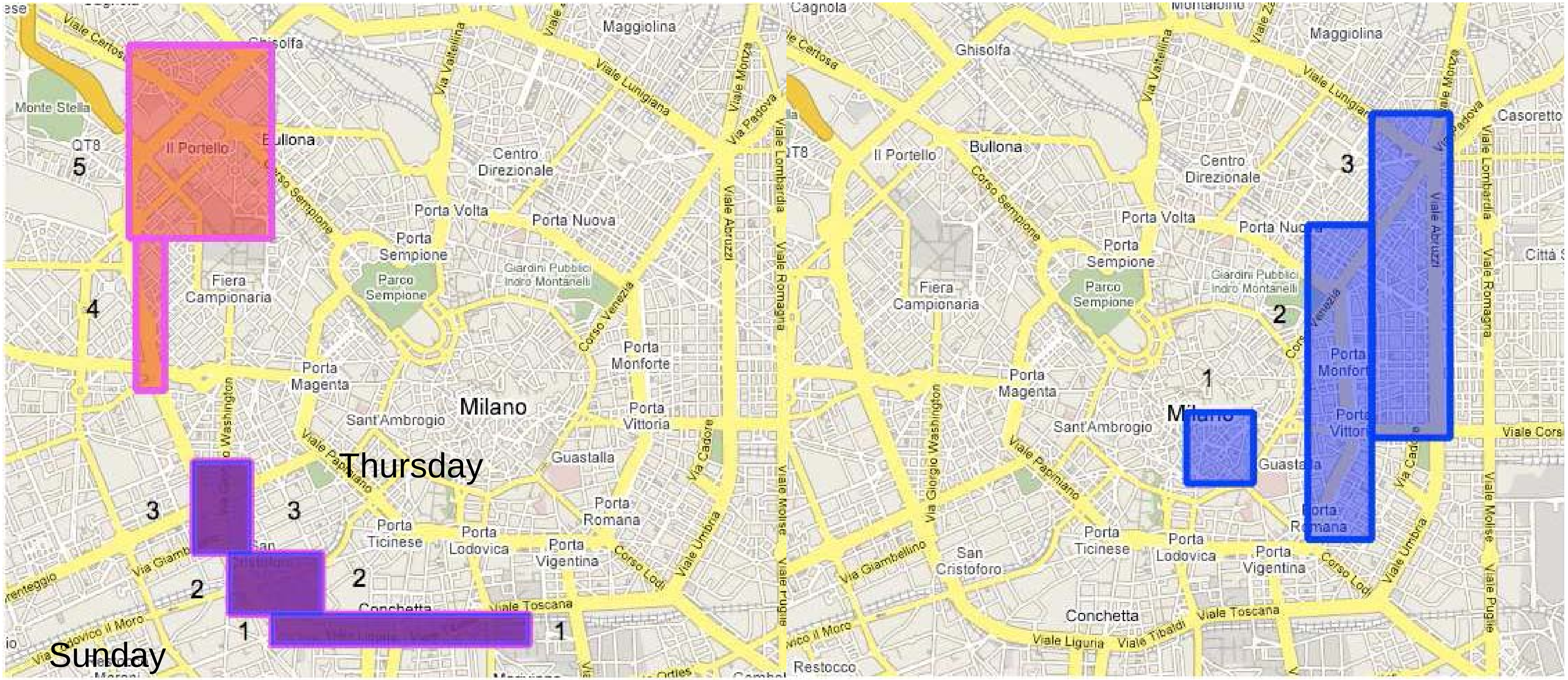
Low-speed movement (counts, 3h intervals)



- 2.075M punti, 17.000 veicoli, 200.000 traiettorie

→ 1597
→ 1

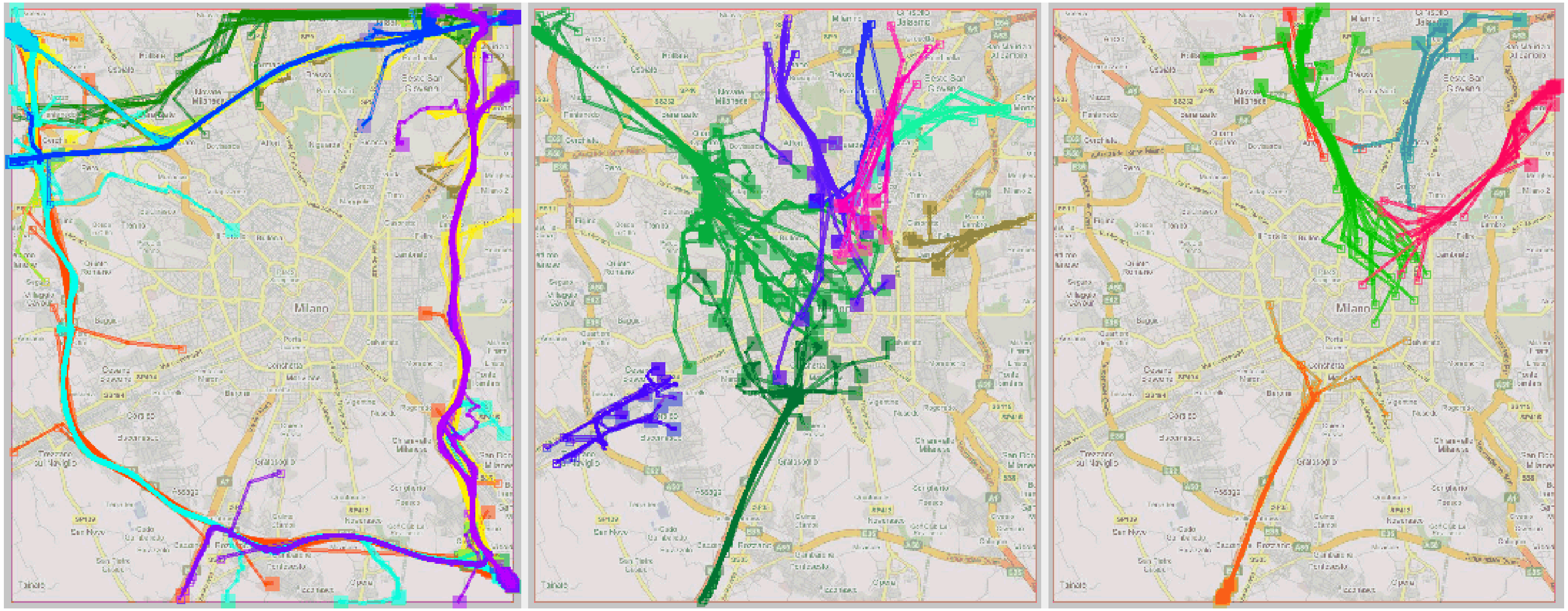
T-Patterns



Working days

Sunday

Clustering trajectories on “route similarity”



Left: peripheral routes; middle: inward routes; right: outward routes.

- Rinzivillo, Pedreschi, Nanni, Giannotti, Andrienko, Andrienko **Visually-driven analysis of movement data by progressive clustering**. J. of Information Visualization, 2008

Challenges of visually-driven clustering

- Progressive refinement through visually-driven exploration
  - Progressively complex similarity functions
- Scalability
  - Index structures to support efficient neighborhood queries for trajectory clustering (Nanni, Pedreschi, Pelekis, Theodoridis, 2008)
  - Progressive clustering by sampling
- Incremental clustering and concept drift

On ZDNet: Windows 7: The Anti-Vista? Log in | Sign up Why join?

c|net NEWS.com

Search: News Go Advanced search

Today on CNET Reviews **News** Downloads Tips & Tricks CNET TV Compare Prices Blogs

Business Tech | Cutting Edge | Green Tech | Access | Security | Media 2.0 | Markets | Personal Tech | News Blogs | Video | Extra | My News

Subscribe to News Blog

February 8, 2008 5:56 PM PST

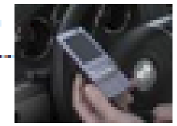
Nokia turns people into traffic sensors

Posted by Erica Ogg | 8 comments

UNION CITY, Calif.--On a cool, overcast morning in the parking lot of a Lowe's hardware store, 100 UC Berkeley students lined up in rows ready to jump into a bevy of idling vehicles.

With media and VIPs from companies like Nokia, Navteq, General Motors, BMW, and CalTrans looking on, wave after wave of students left the parking lot to drive a 10-mile stretch of the nearby 880 freeway as part of a large-scale experiment to test how cell phones can monitor and predict traffic.

The test, conducted all day Friday, was put on by the California Center for Innovative Transportation (CCIT) as a joint project between Nokia, CalTrans, and Berkeley's Department of Civil and Environmental Engineering.



Video: Using cell phones to track traffic

Each student car was issued a Nokia N95 phone with GPS and special traffic-monitoring software developed by Nokia's Palo Alto, Calif.-based research lab--plus a Bluetooth headset. As the students drove the freeway, the phone sent data about each car's speed and position back to the company's research facility. The data is compiled and used to predict traffic patterns and help drivers get where they need to be quickly. Nokia hopes that one day the system could be a significantly cheaper way to track traffic than the permanent sensors installed in roadways or next to them because it uses equipment most people already own: cell phones.

Alex Bayen, a professor of civil and environmental engineering and lead researcher on the project for Berkeley, called the experiment "a glimpse into the future of traffic information

About News Blog

Recent posts on technology, trends, and more.

Subscribe to this blog
Click this link to view this blog as XML.

Add this feed to your online news reader

- Add to Google
- Add to my Yahoo
- Add to MSN
- Add to Bloglines
- Add to Newsgator

An archaeology of the present

- The opportunity to discover, from the **digital traces** of human activity, the **knowledge** that makes us comprehend timely and precisely the way we live, the way we use our time and our land.

- **Mobility data mining**

From opportunities to threats

- Personal mobility data, as gathered by the wireless networks, are extremely sensitive
- Their disclosure may represent a brutal violation of the privacy protection rights, i.e., to keep confidential
  - the places we visit
  - the places we live or work at
  - the people we meet
  - …

The naive scientist's view

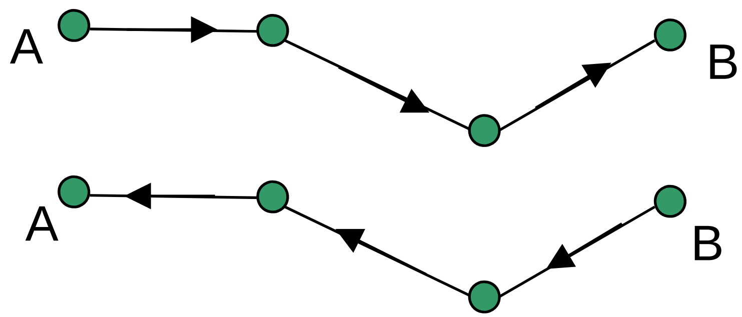
- Knowing the exact identity of individuals is not needed for **analytical** purposes
    - De-identified mobility data are enough to reconstruct aggregate movement behaviour, pertaining to groups of people.
- Reasoning coherent with European data protection laws: personal data, **once made anonymous**, are not subject to privacy law restrictions
- Is this reasoning correct?

Unfortunately not!

- Making data (reasonably) anonymous is not easy.
- Sometimes, it is possible to reconstruct the exact identities from the de-identified data.
- Many famous example of re-identification
  - Governor of Massachusetts' clinical records (Sweeney's experiment, 2001)
  - America On Line August 2006 crisis: user re-identified from search logs
- Two main sources of danger:
  - **Many observations** on the same "anonymous" subject
  - **Linking data**, after joining separate datasets

Spatio-temporal linkage in Mobility Data

Id:
34567



[almost every day mon-fri between 7:45 – 8:15]

[almost every day mon-fri between 17:45 – 18:15]

- By intersecting the phone directories of locations A and B we find that only one individual lives in A and works in B.
- Id:34567 = Prof. Smith
- Then you discover that on Saturday night Id:34567 usually drives to the city red lights district…

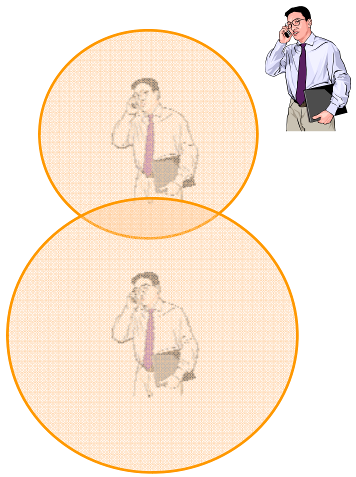
Basic ideas for anonymity preserving data analysis

How do people (try to) stay anonymous?

- either by **camouflage**
  - pretending to be someone else or somewhere else
- or by **hiding in the crowd**
  - becoming indistinguishable among many others

Concepts for Location Privacy
Location Perturbation – Randomization

- The user location is represented with a **fake** value

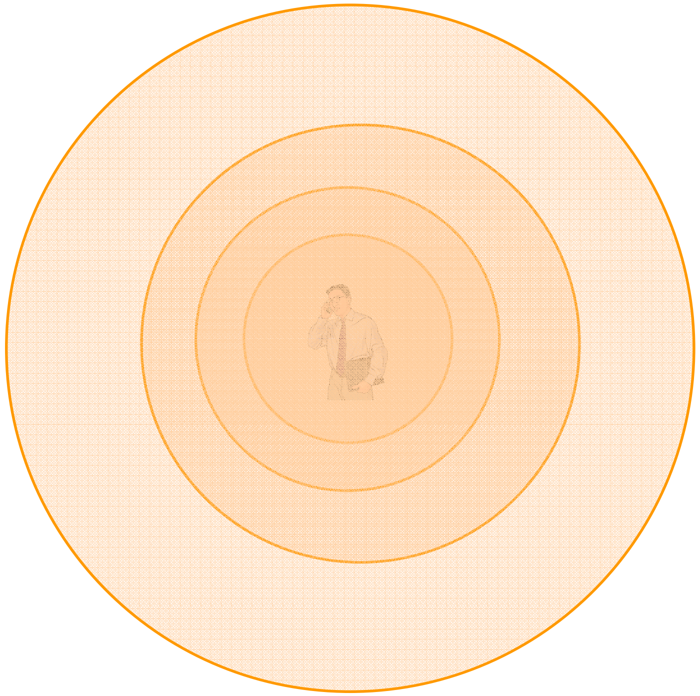


- Privacy protection is achieved from the fact that the reported location is false

- The accuracy and the amount of privacy mainly depends on how far is the reported location from the exact location

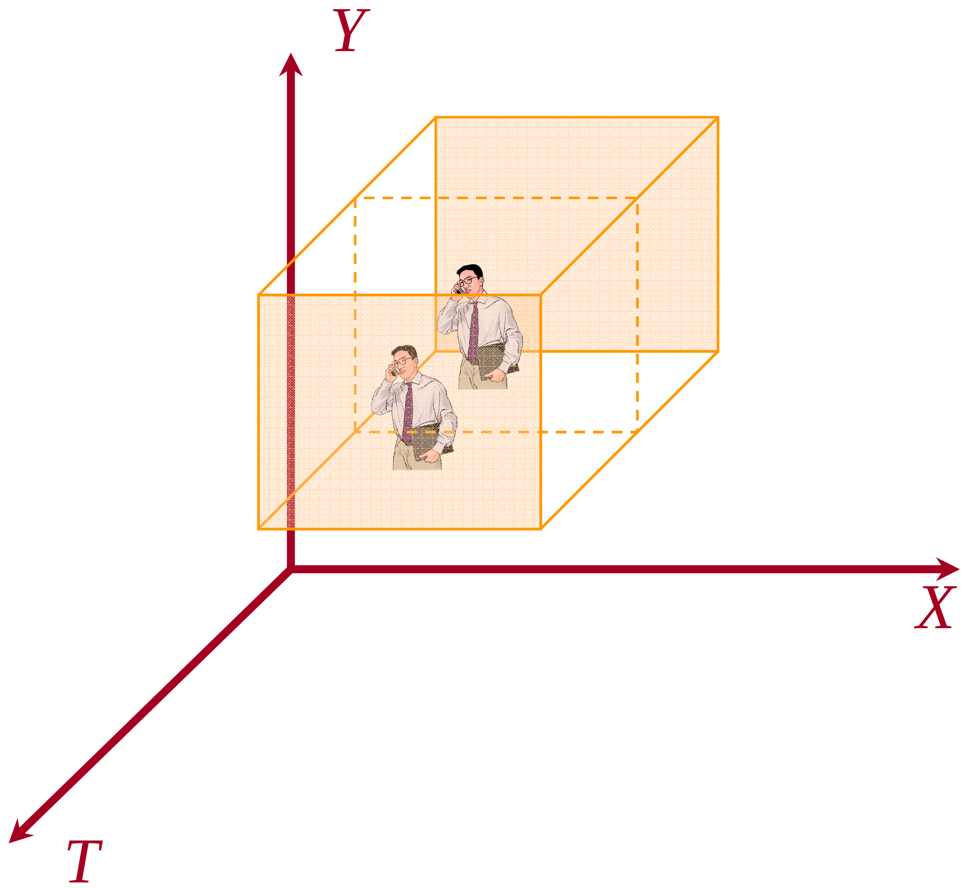
Concepts for Location Privacy
Spatial Cloaking – Generalization

- The user exact location is represented as a region that includes the exact user location
- An adversary does know that the user is located in the region, but has no clue where the user is exactly located
- The area of the region achieves a trade-off between user privacy and accuracy



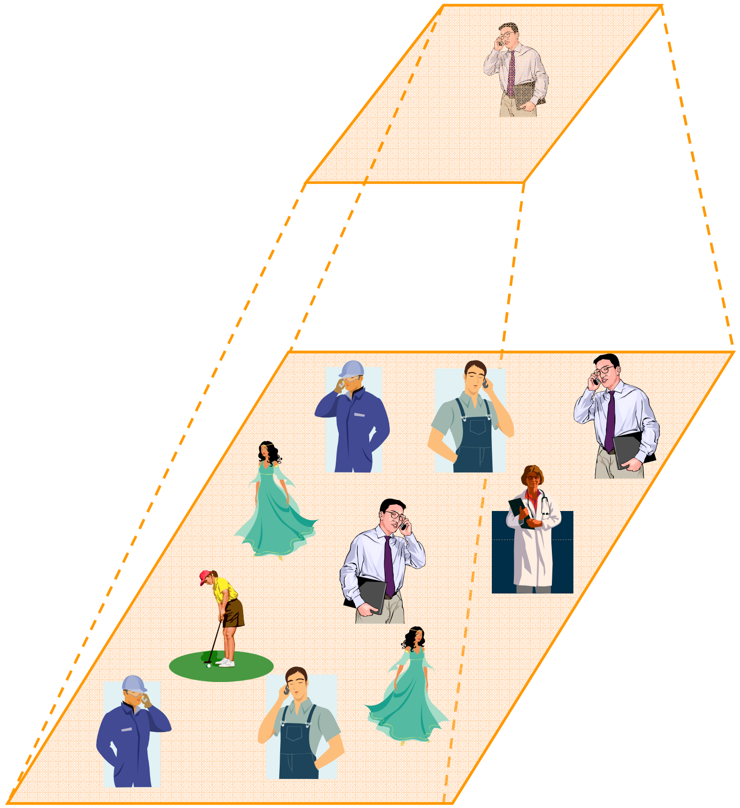
Concepts for Location Privacy
Spatio-temporal generalization

- In addition to the spatial dimension, generalize also the temporal dimension



Concepts for Location Privacy
k-anonymity

- User's position is generalized to a region containing **at least *k* users**
- The user is indistinguishable among other *k* users
- The area largely depends on the surrounding environment.
- A value of *k* =100 may result in a very small area downtown Hong Kong, or a very large area in the desert.



10-anonymity

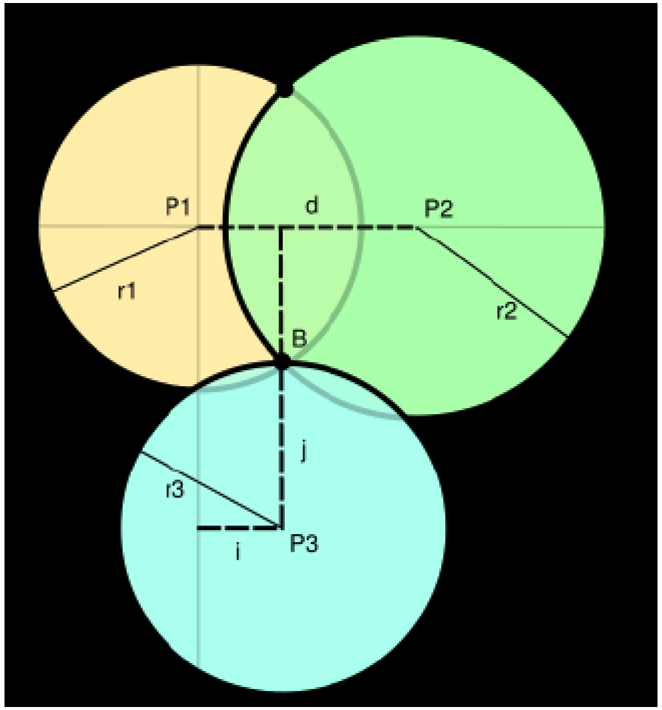
Privacy- preserving spatio-temporal data mining

Trajectory randomization is risky!

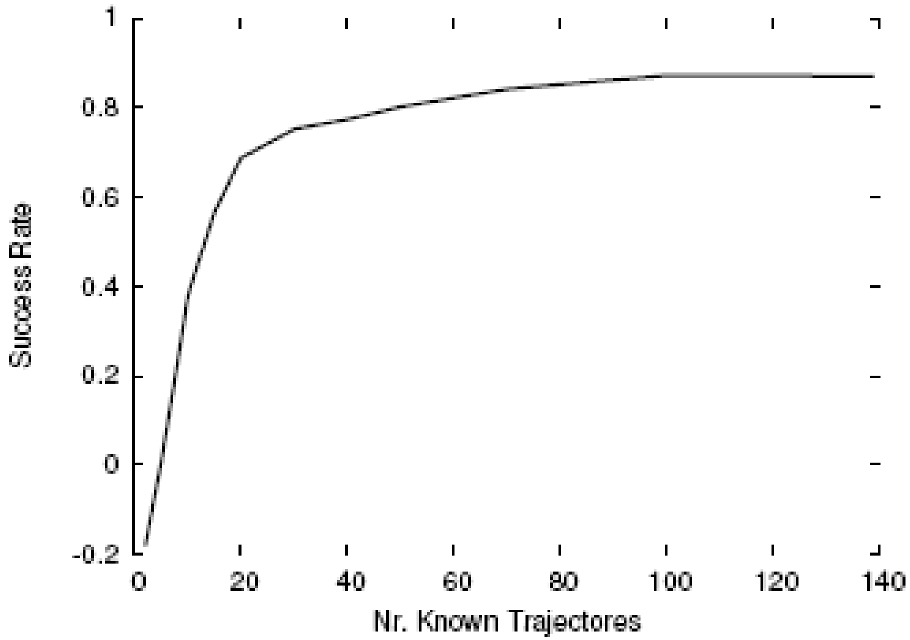
Trajectory anonymization

A subtle re-identification attack

- Disclosure Risks of Distance Preserving Data Transformations
  - Erkay Savas, Yucel Saygin, Emre Kaplan, and Thomas B. Pedersen (Sabanci Univ., Istanbul)
- What if the attacker knows:
  - Some trajectories
  - All mutual distances
- Hyper-lateration
  - Works in d dimensions given d + 1 points
  - If known trajectories are few, then approximate!



Red: true traj Blue: approx traj



(b) Success-rate vs. number of known trajectories (Each sample is the average of 60 experiments run for 50.000 iterations).

Privacy- preserving spatio-temporal data mining

Trajectory randomization is risky!

Trajectory anonymization

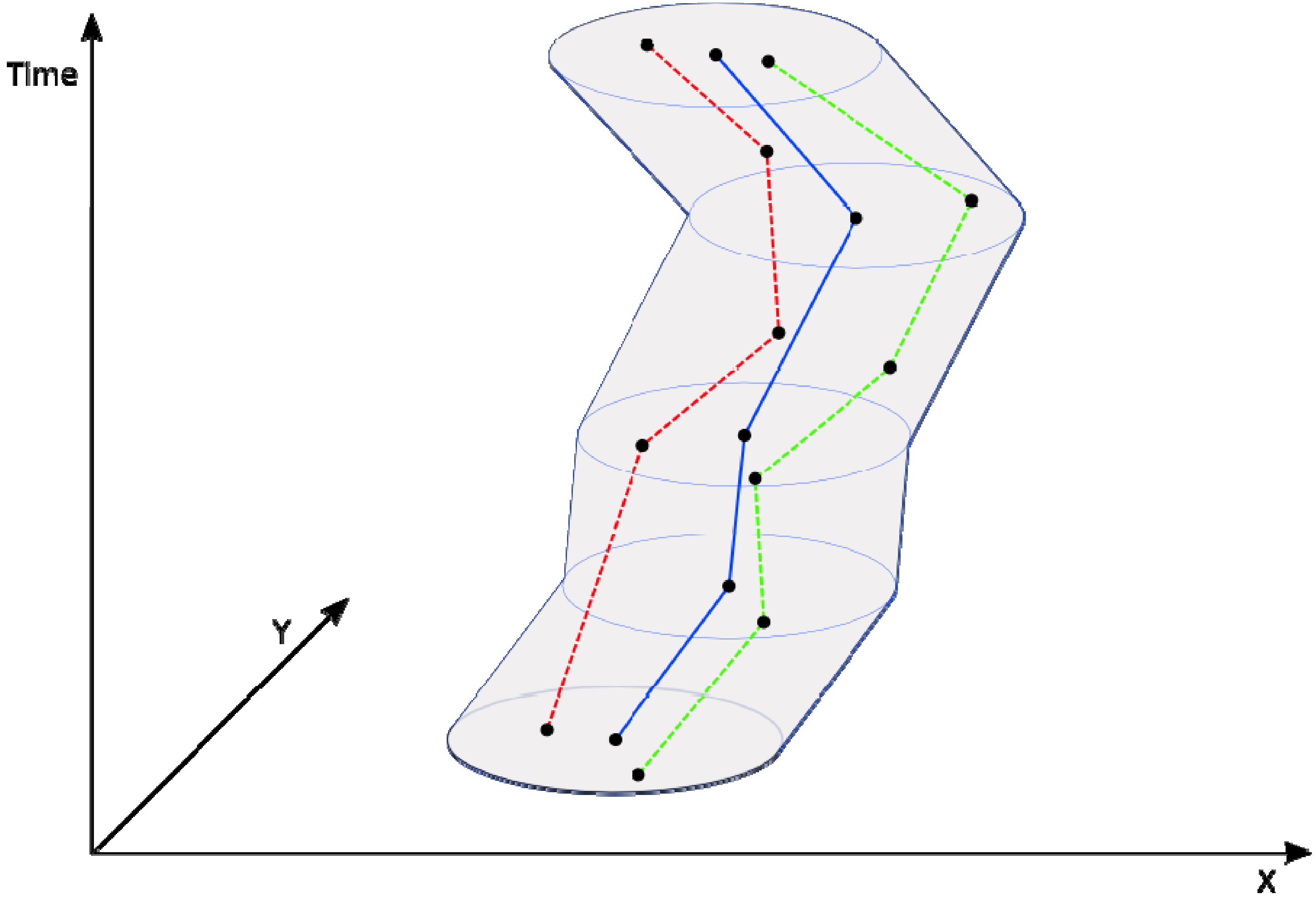
Trajectory anonymization

- Several variants developed in GeoPKDD:
  - Abul, Bonchi, Nanni (Pisa KDD LAB). Int. Conf. Data Engineering ICDE 2008
  - Nergiz, Atzori, Saygin (Sabanci Univ. + Pisa KDD LAB). 2007 (submitted)
  - Gkoulalas-Divanis, Verykios (Univ. Thessaly). 2007 (submitted)
  - Monreale, Giannotti, Pensa, Pedreschi, Pinelli (Pisa KDD LAB) 2008
- Common goal: construct an anonymized version of a trajectory dataset, preserving some target analytical properties
- Different techniques adopted

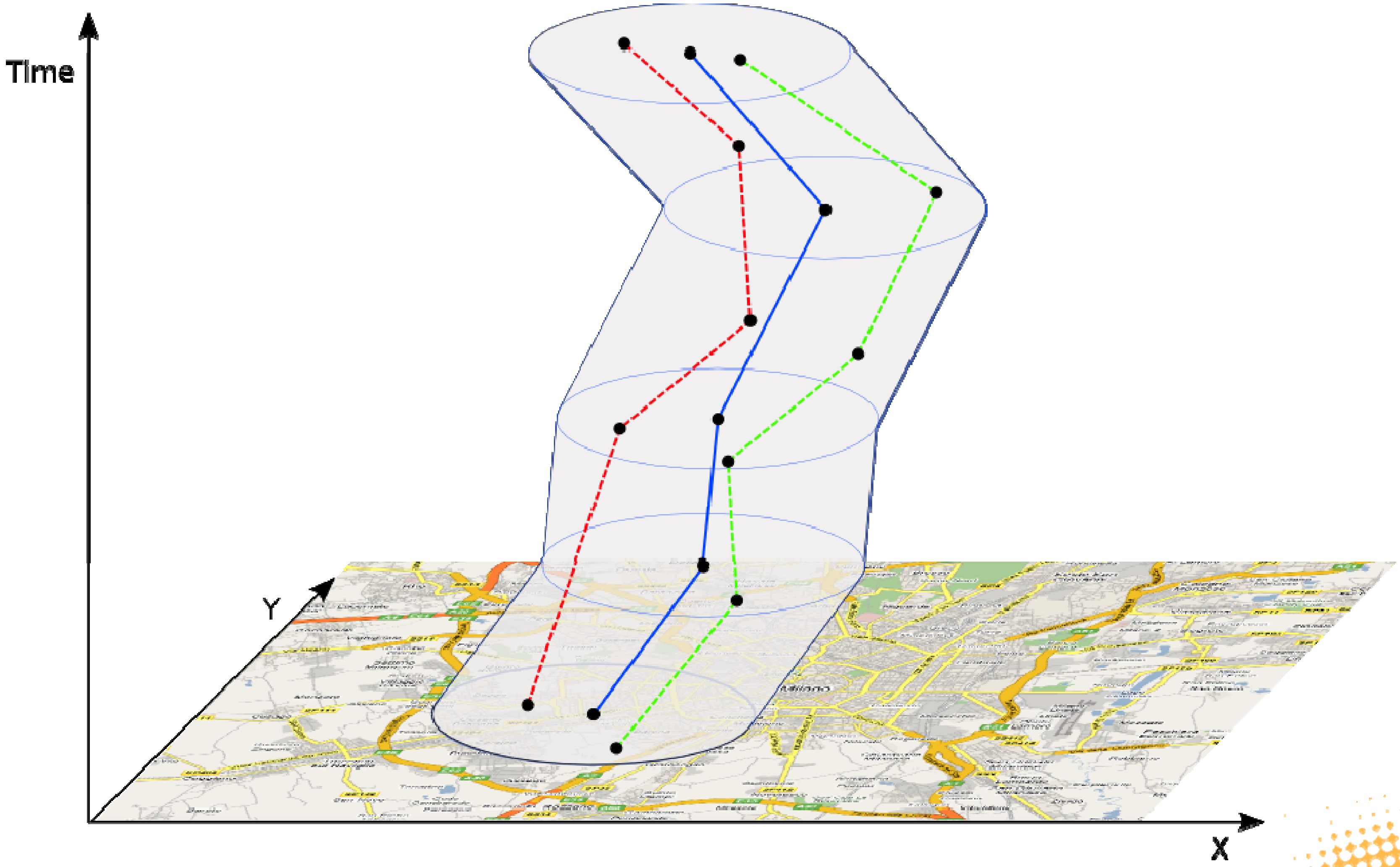
Example result: Never Walk Alone

- Bonchi, Abul, Nanni. *Never Walk Alone: Uncertainty for Anonymity in Moving Objects Databases*. ICDE 2008
- Basic ideas:
  - Trade uncertainty for anonymity: trajectories that are close up the uncertainty threshold are indistinguishable
  - Combine k-anonymity and perturbation
- Two steps:
  - Cluster trajectories into groups of k similar ones (removing outliers)
  - Perturb trajectories in a cluster so that each one is close to each other up to the uncertainty threshold

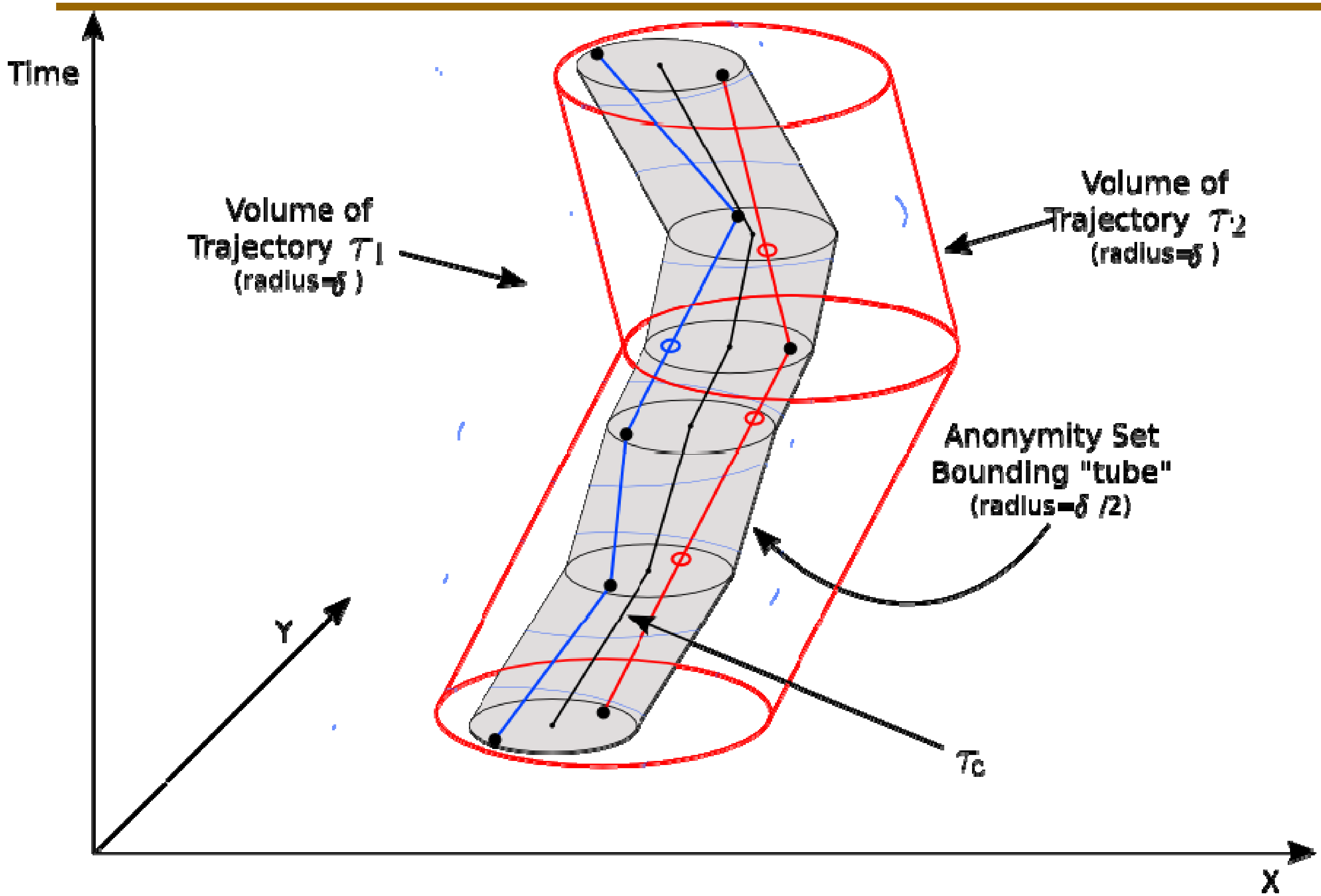
Trajectory cluster



Trajectory cluster



(K, δ) –anonymity set



- K = minimum number of trajectories in the set
- δ = uncertainty threshold (e.g., measurement error of GPS device)

Quality of anonymized datasets

- For reasonable values of K and δ, some interesting analytical properties of the original dataset are preserved by the anonymized trajectories :
  - **density** (aggregate count of mobile users in the spatio-temporal dimension)
  - **clustering**
  - **T-patterns**
- Prototype **trajectory anonymity toolkit** available

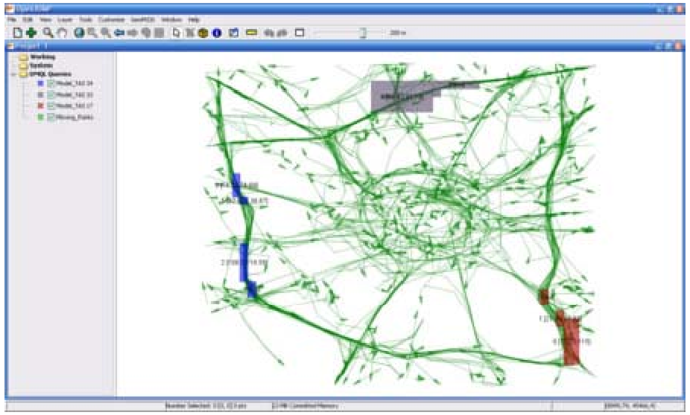
Key issues

- Define an acceptable formal measure of anonymity protection:
  - Probability of re-identification (in a given context)
  - A (technically supported) juridical issue!
- Sampling: a necessity **and** an opportunity!
  - Necessary for performance/feasibiliy of data mining from massive mobility datasets
  - Good for anonymity (re-identification probability decreases)

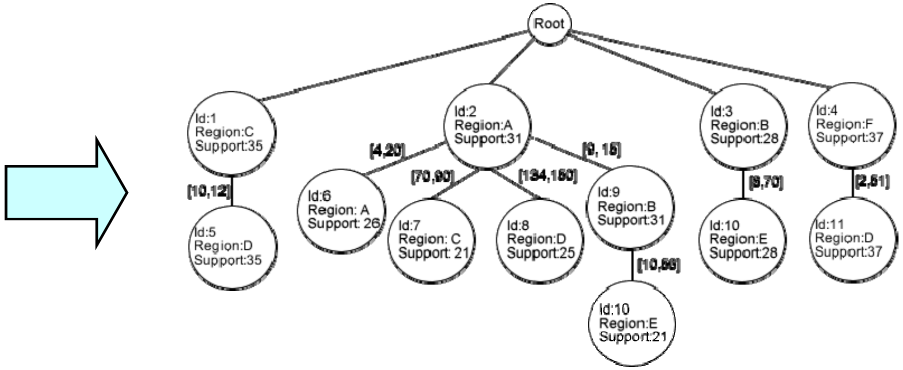
From local patterns to global models

From T-patterns to T-models

- Mobility raw data are huge and noisy
- To create good models, better first simplify data into patterns
  - … as in association rule based classification
- Location prediction based on T-patterns (Giannotti et al., Int. Workshop on Computational Transportation Science, 2008)



Extracting Trajectory Patterns from the data



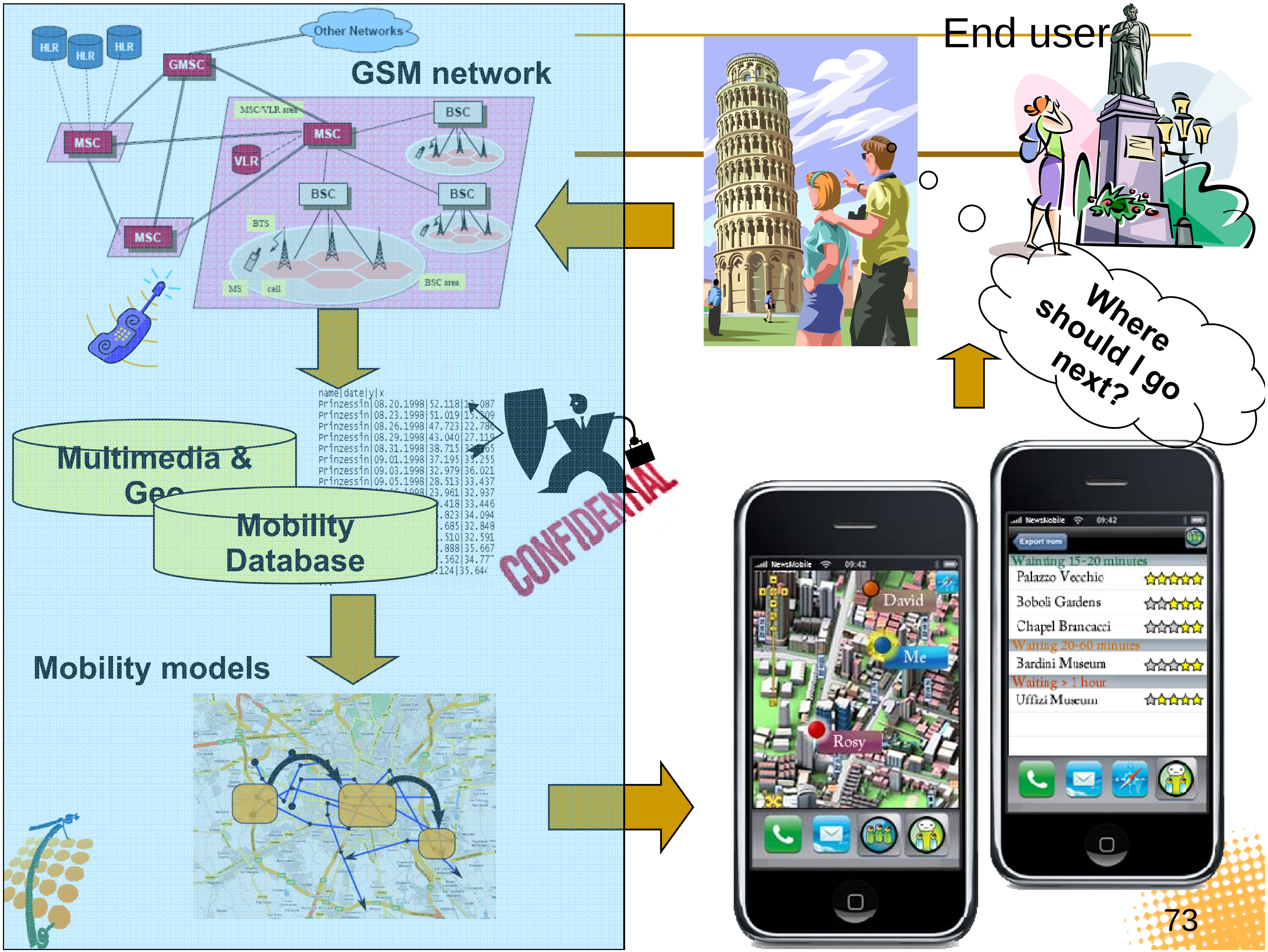
Prediction Tree

- T-anonymization based on T-patterns (Pedreschi et al., 2008)

A reasoning framework for mobility data mining applications

Building mobility data mining applications…

- requires reasoning on a richer form of knowledge about mobility
  - Geographic semantics
    - Landmarks and interesting places
    - Road network
    - Landscape
    - …
  - Movement sematics
    - stops and moves
    - Purposes of movement
    - means of transportation
    - …


Other Networks
HLR
HLR
HLR
GMSC
GSM network
MSC/VLR area
MSC
MSC
MSC
VLR
BSC
BSC
BSC
BTS
MS
cell
BSC area
name|date|y|x
Multimedia &
Geo
Mobility
Database
CONFIDENTIAL
Mobility models
End user
Where should I go next?
David
Me
Rosy
Walking 15-20 minutes
Palazzo Vecchio
Boboli Gardens
Chapel Brancacci
Waiting 20-60 minutes
Bardini Museum
Waiting > 1 hour
Uffizi Museum

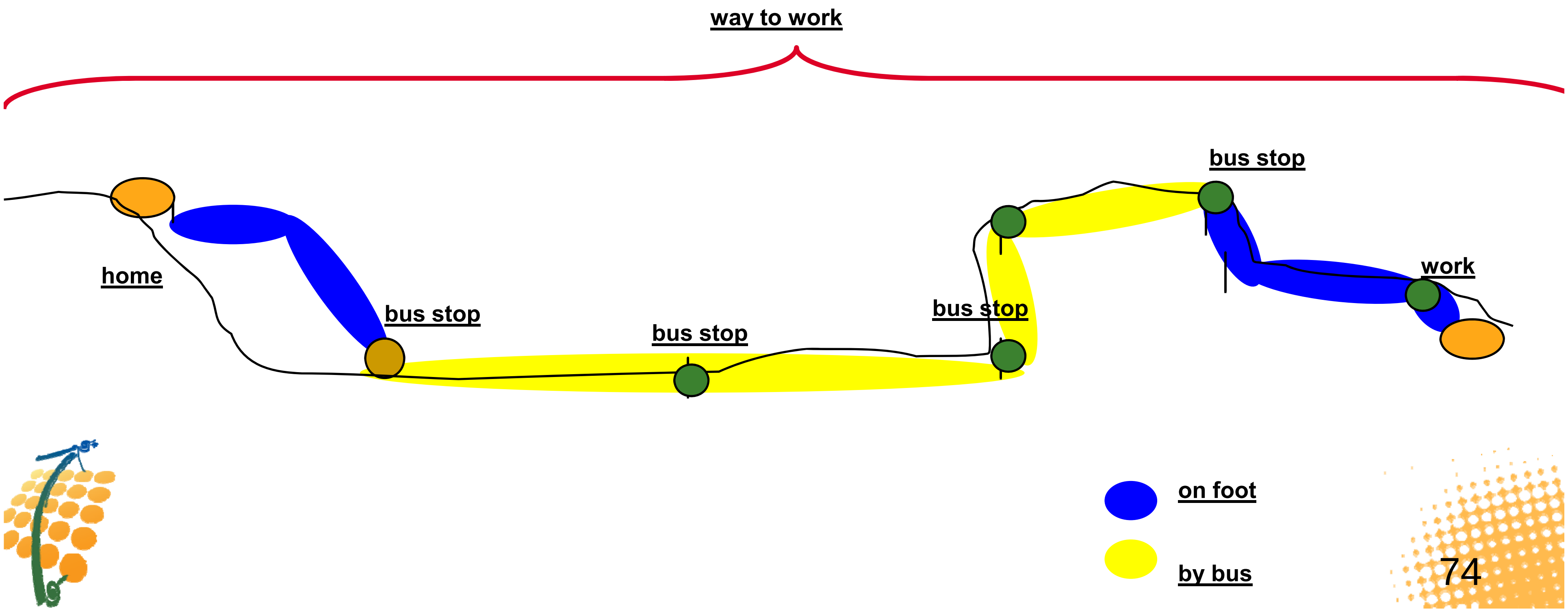
Semantic Trajectory Data

Physical Trajectory:

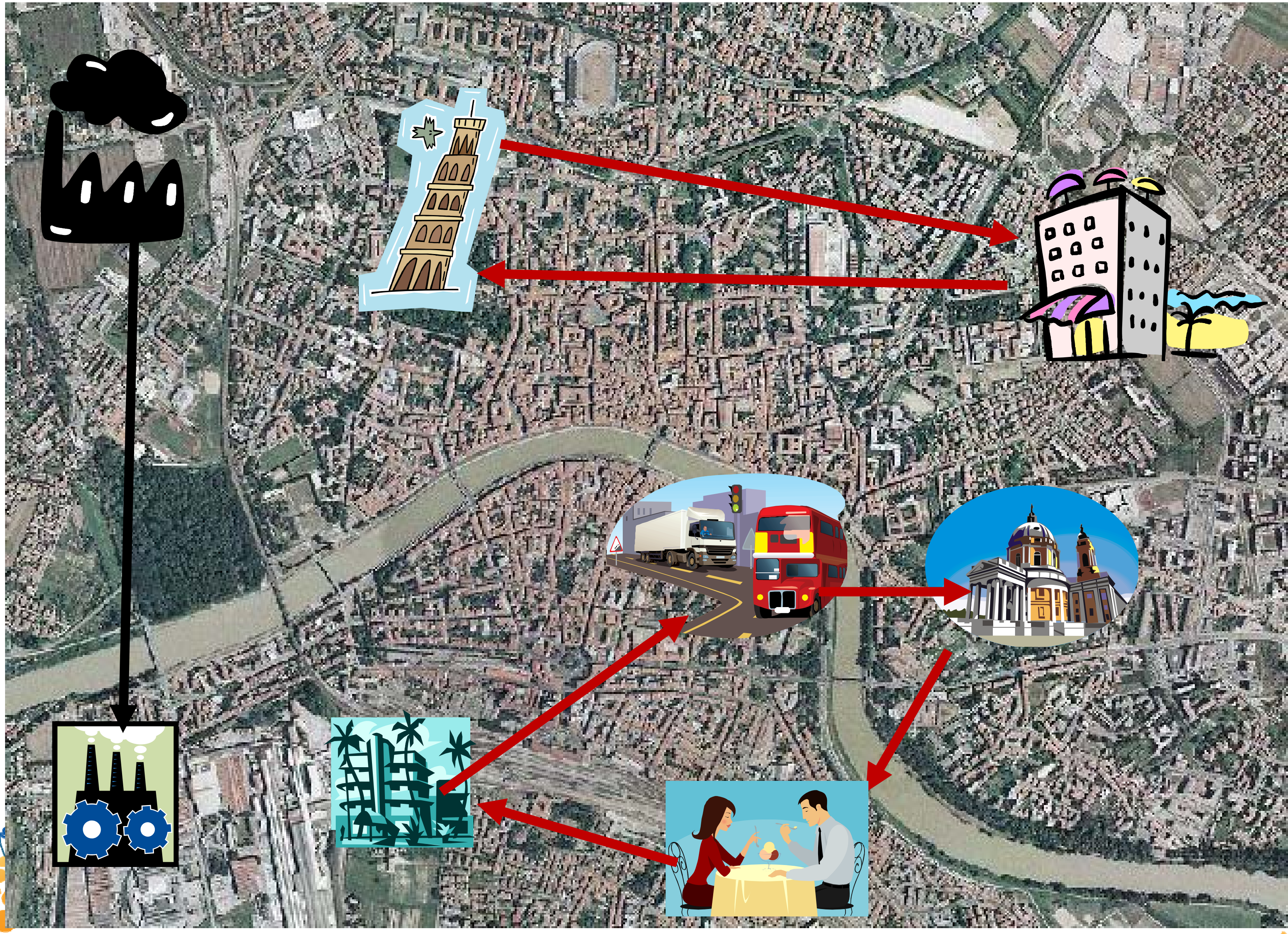
- e.g. GPS recording over some period of time

Semantic Trajectory:

- places where a person stayed
- means of transportation
- combination of above elements for higher-level description



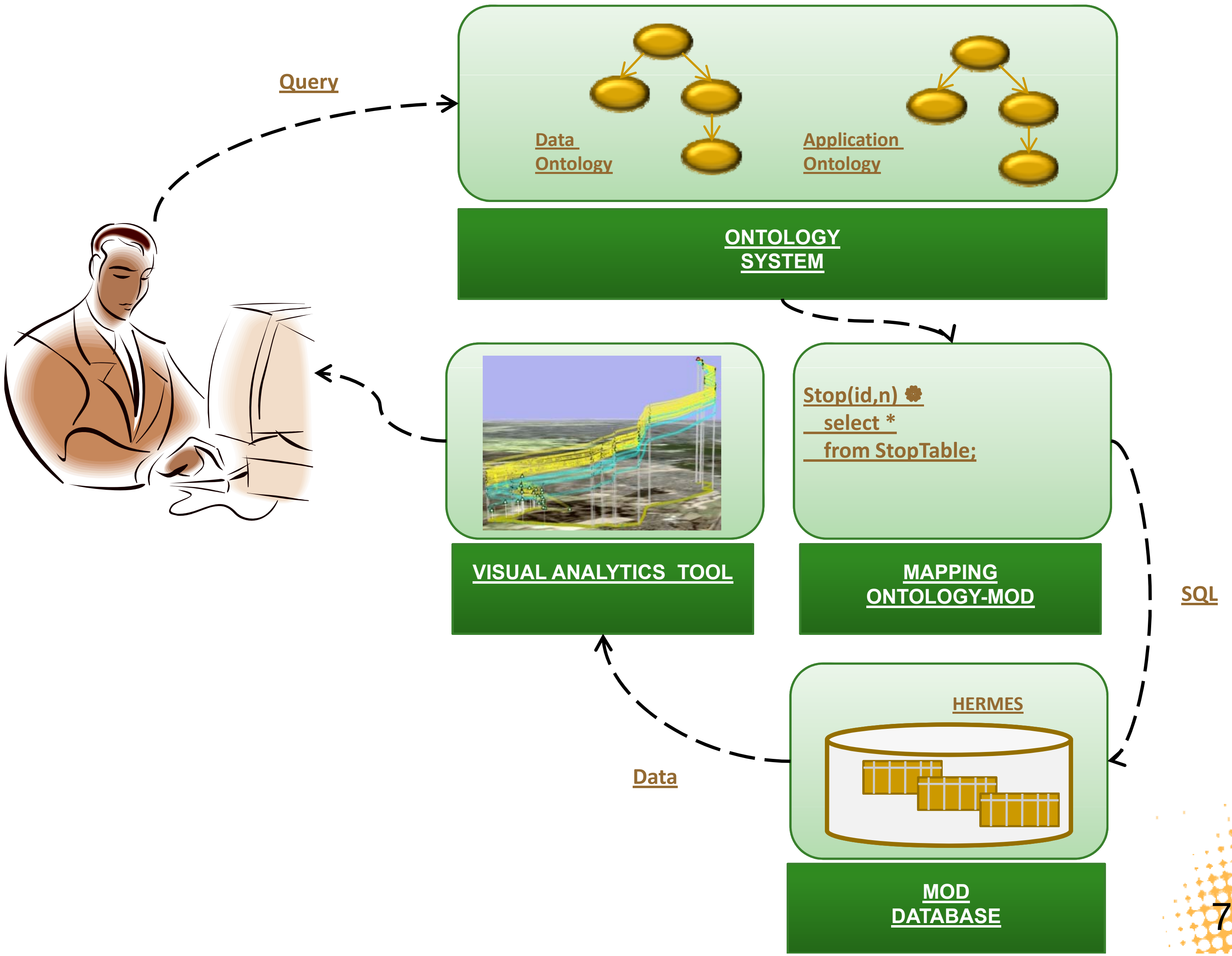
Semantic (frequent) patterns



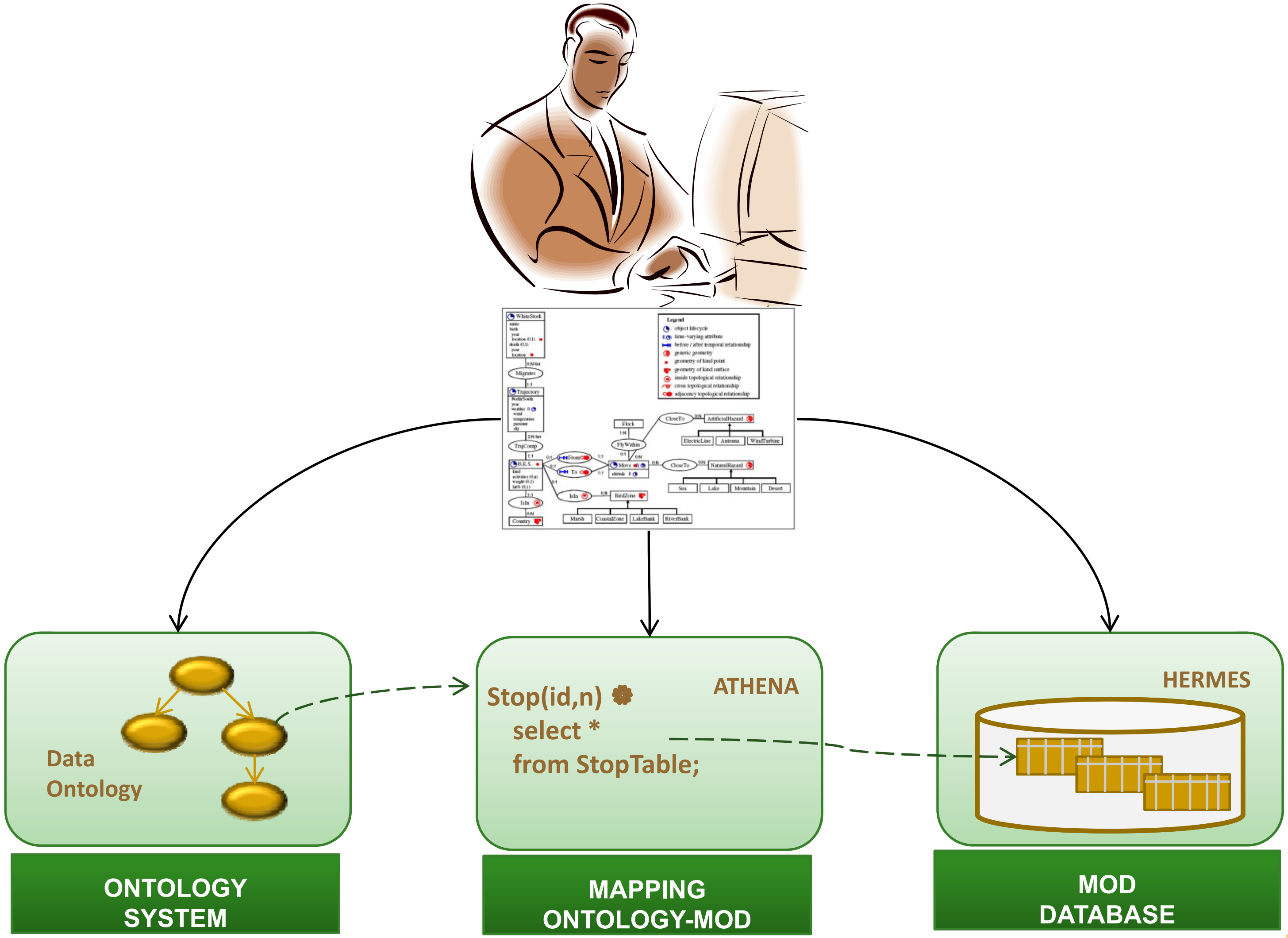
An ontological framework ..

- enables a progressive semantic enrichment of mobility data and patterns
  - From **raw mobility data** collected by devices
  - To **semantic trajectories** – from (x,y,t) to sequences of stops and moves
  - From **trajectory patterns**
  - To **behavioral motion patterns**

The GeoPKDD ontological framework

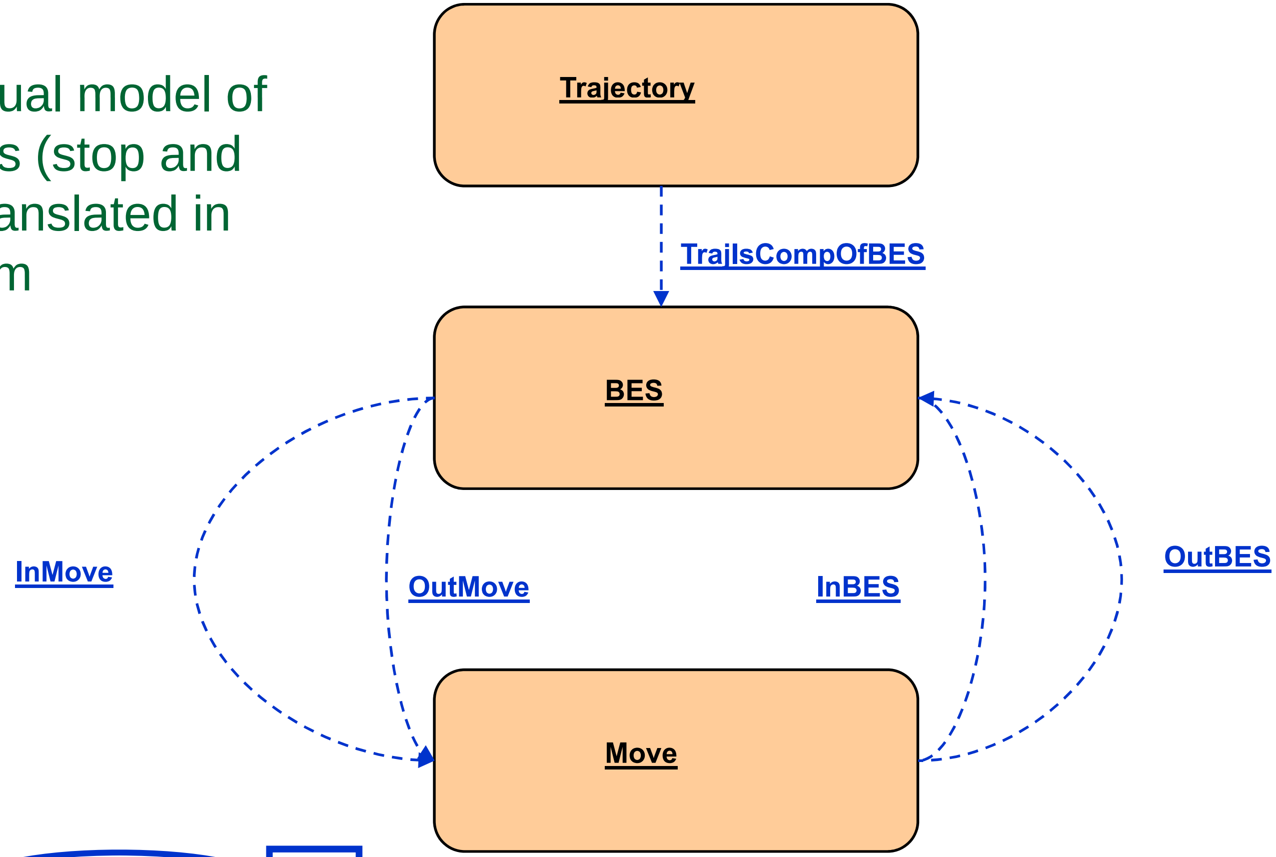


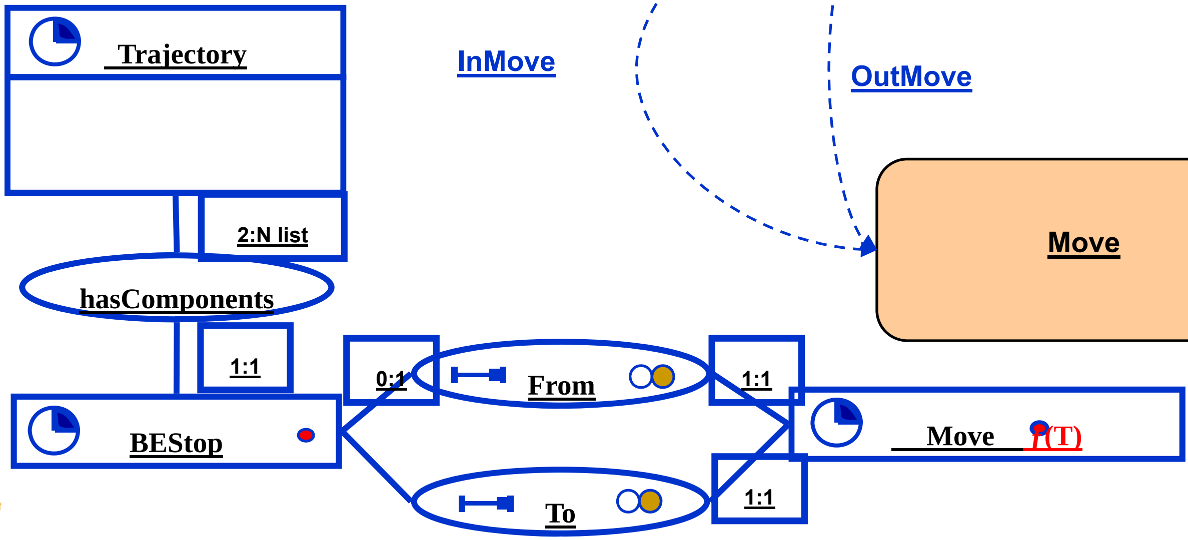
Mapping the data ontology to the trajectory DB



Conceptual Model of Semantic Trajectories

The MADS conceptual model of semantic trajectories (stop and moves) has been translated in ontological formalism

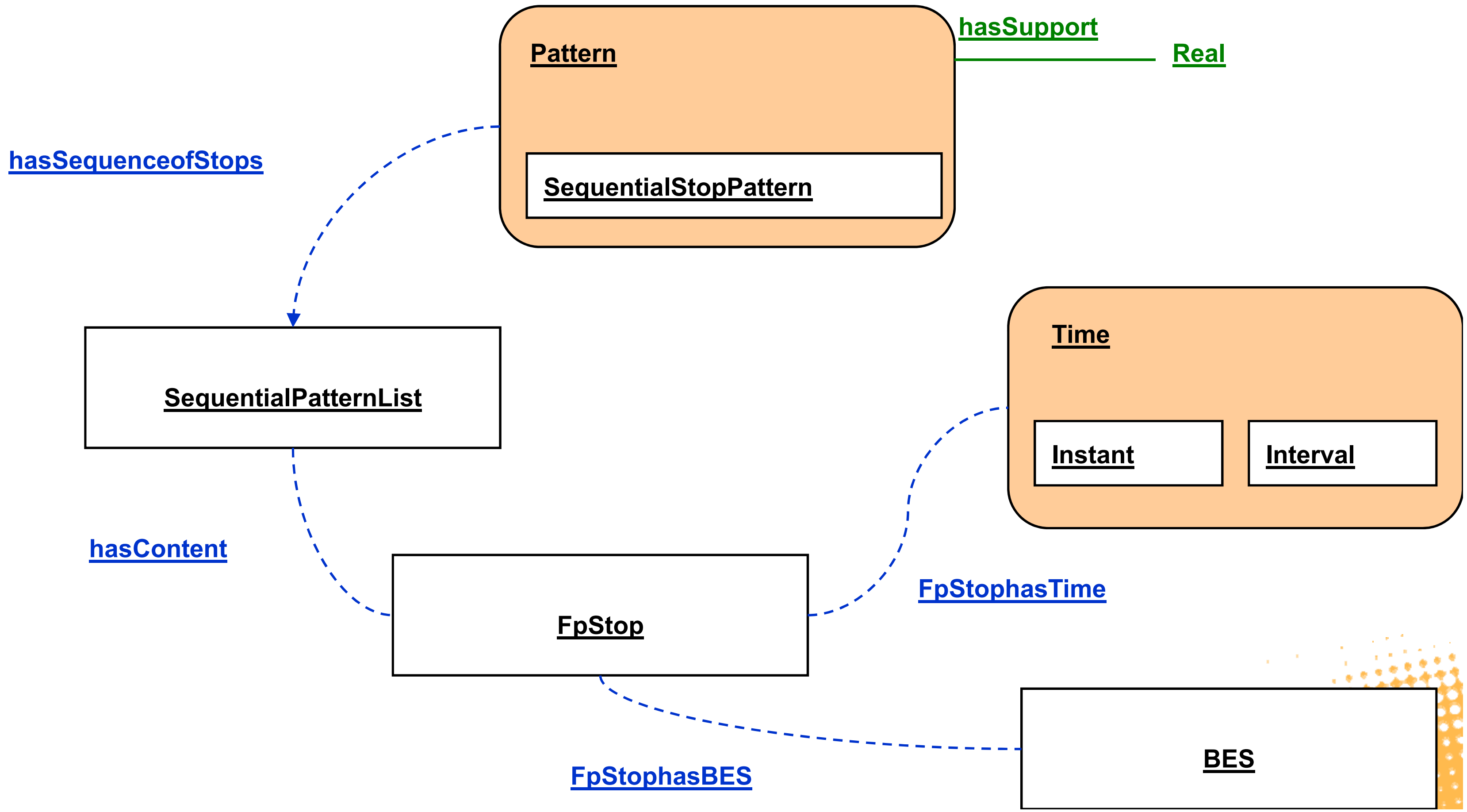




Conceptual Model of Motion Patterns

Applying sequential pattern on semantic trajectories:
4% of people stops at CentrumHotel in the morning and then
at Pisa Tower in the afternoon

<CentrumHotel[08:00-12:00], PisaTower[12:01-18:00] (s=0.04)>



Definition of Touristic Path

Data Ontology concepts are exploited in the application ontology to give new concepts definitons through axioms.

A touristic path is a sequential pattern that has some stops in an accomodation area in the morning, followed by some touristic area and then by stops in an accomodation area in the evening

TouristicSeqPath ≡ hasSequenceOfStops some (SequentialPatternList and (hasContent some (FPStop and (inside some AccomodationArea) and (fpStopHasTime has Morning))))) and (isFollowedBy some (SequentialPatternList and (hasContent some FPStop and (inside some TouristicArea)) and (isFollowedBy some (SequentialPatternList and (hasContent some (FPStop and (inside some AccomodationsArea) and (fpStopHasTime has Evening))))) and (hasNext some EmptyList))))))

Architecture features

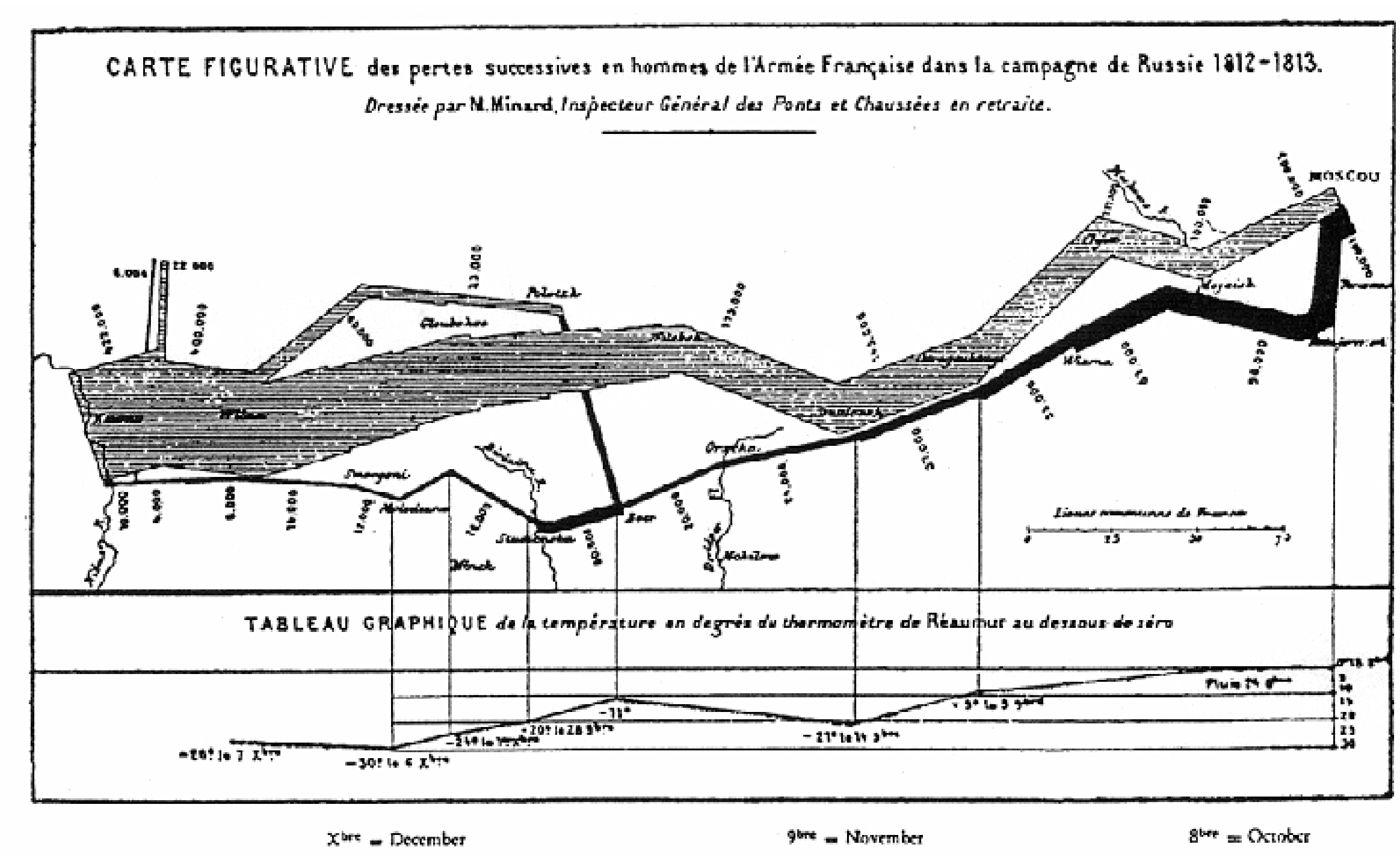
- Design
  - Trajectory Conceptual model based on MADS;
  - Mappings: TCM-Hermes, TCM-Data Ontology and Data Ontology-Hermes
  - Ontology language: OWL
- Querying
  - Conjunctive query over a DL knowledge base;
  - Reformulate query taking into account Tbox;
  - Translate reformulated query into SQL query;
- Integration with GeoPKDD components:
  - Visual analytics tool
  - DMQL

Conclusions

While mobility data flood us …

- … mobility data mining is a emerging as an exciting new field
- **GeoPKDD.eu** is in the mix, shaping up the area
- Challenge: Ubiquitous Computing will provide us with streams of semantic-rich mobility data (in a decentralized setting)
  - We have only begun to scratch the surface of this problem

… trying to accomplish a long-time dream



The representation of Napoleon's Russian campaign of 1812 produced by Charles Joseph Minard in 1861

Giannotti · Pedreschi (Eds.)

Mobility, Data Mining and Privacy

The technologies of mobile communications and ubiquitous computing pervade our society, and wireless networks sense the movement of people and vehicles, generating large volumes of mobility data. This is a scenario of great opportunities and risks: on one side, mining this data can produce useful knowledge, supporting sustainable mobility and intelligent transportation systems; on the other side, individual privacy is at risk, as the mobility data contain sensitive personal information. A new multidisciplinary research area is emerging at this crossroads of mobility, data mining, and privacy.

This book assesses this research frontier from a computer science perspective, investigating the various scientific and technological issues, open problems, and roadmap. The editors manage a research project called GeoPKDD, Geographic Privacy-Aware Knowledge Discovery and Delivery, funded by the EU Commission and involving 40 researchers from 7 countries, and this book tightly integrates and relates their findings in 13 chapters covering all related subjects, including the concepts of movement data and knowledge discovery from movement data; privacy-aware geographic knowledge discovery; wireless network and next-generation mobile technologies; trajectory data models, systems and warehouses; privacy and security aspects of technologies and related regulations; querying, mining and reasoning on spatiotemporal data; and visual analytics methods for movement data.

This book will benefit researchers and practitioners in the related areas of computer science, geography, social science, statistics, law, telecommunications and transportation engineering.

ISBN 978-3-540-75176-2

9 783540 751762

springer.com

Giannotti
Pedreschi (Eds.)

Mobility, Data Mining and Privacy

Fosca Giannotti
Dino Pedreschi (Eds.)

Mobility, Data Mining and Privacy

Geographic Knowledge Discovery

Springer