Geographic Knowledge Discovery: towards Mobility Data Analysis

Fosca Giannotti,

KDD Lab Pisa, Italy

Knowledge Discovery & Data Mining Lab

ISTI/CNR – Dipartimento di Informatica/ University of Pisa

http://www-kdd.isti.cnr.it

Information Society Technologies

GeoPKDD

http://www.geopkdd.eu

A European FP7 project

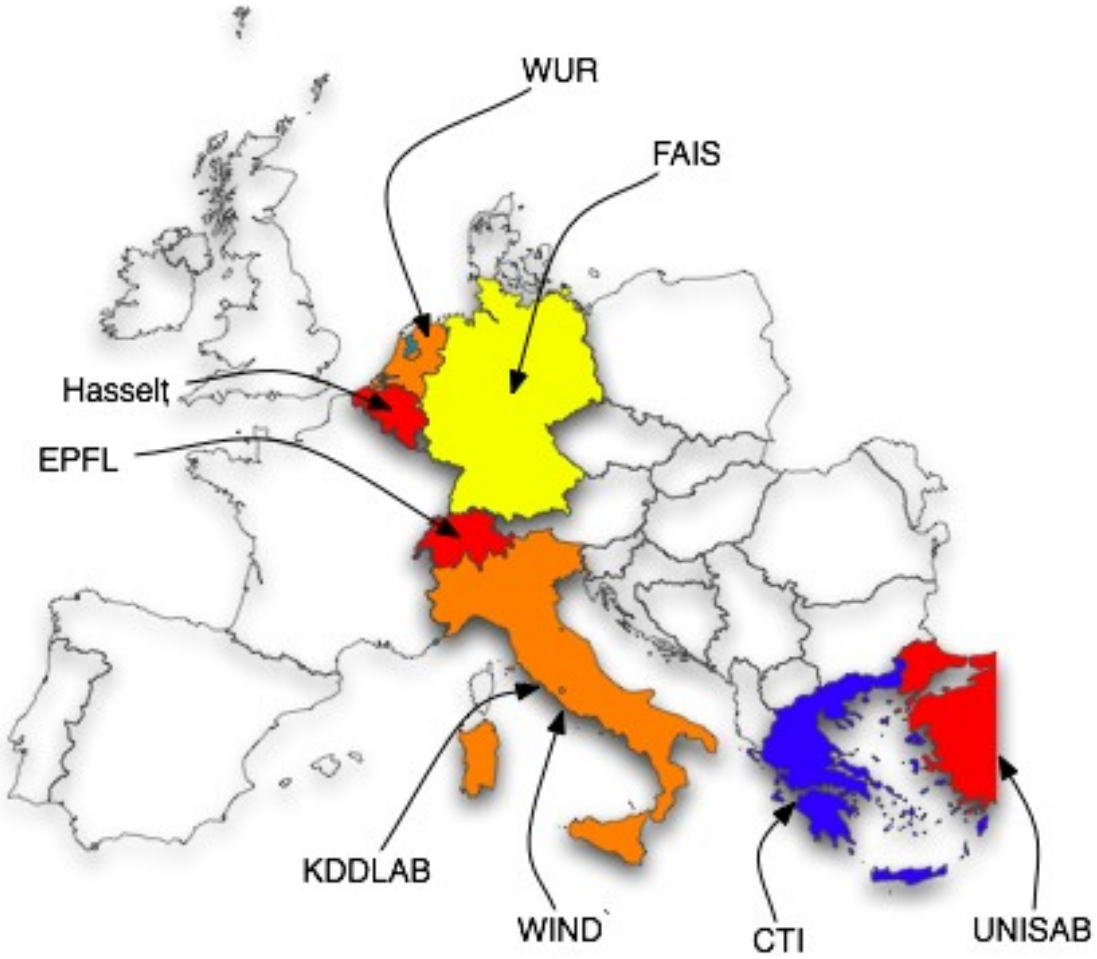
Geographic Privacy-aware Knowledge Discovery and Delivery

Information Society Technologies

www.geopkdd.eu

GeoPKDD

Geographic Privacy-aware Knowledge Discovery and Delivery



The Wireless Network

- The pervasiveness of mobile and ubiquitous technologies is increasing day after day
  - GSM wireless phone networks
    - 1.5 billions in 2005, still increasing at a high speed
    - Italy: # mobile phones ≈ # inhabitants
  - GPS and Galileo positioning systems
  - Wi-Fi and Wi-Max wireless networks
  - RFID's and sensor networks
- miniaturization
- positioning accuracy
  - location technologies capable of providing increasingly better estimate of user location

Which new opportunities?

- Location based services:
  - A certain service that is offered to the users based on their locations
- Mobility data analysis:
  - Discovering knowledge from the digital traces of our mobile activity to support decision making in mobility related issues.

Location-based Services: Then

Information Society Technologies

- **Limited to fixed traffic signs**

Location-based Services: Now

- Location-based traffic reports:
  - ***Range query:*** How many cars in the free way
  - ***Shortest path query***: What is the estimated time travel to reach my destination

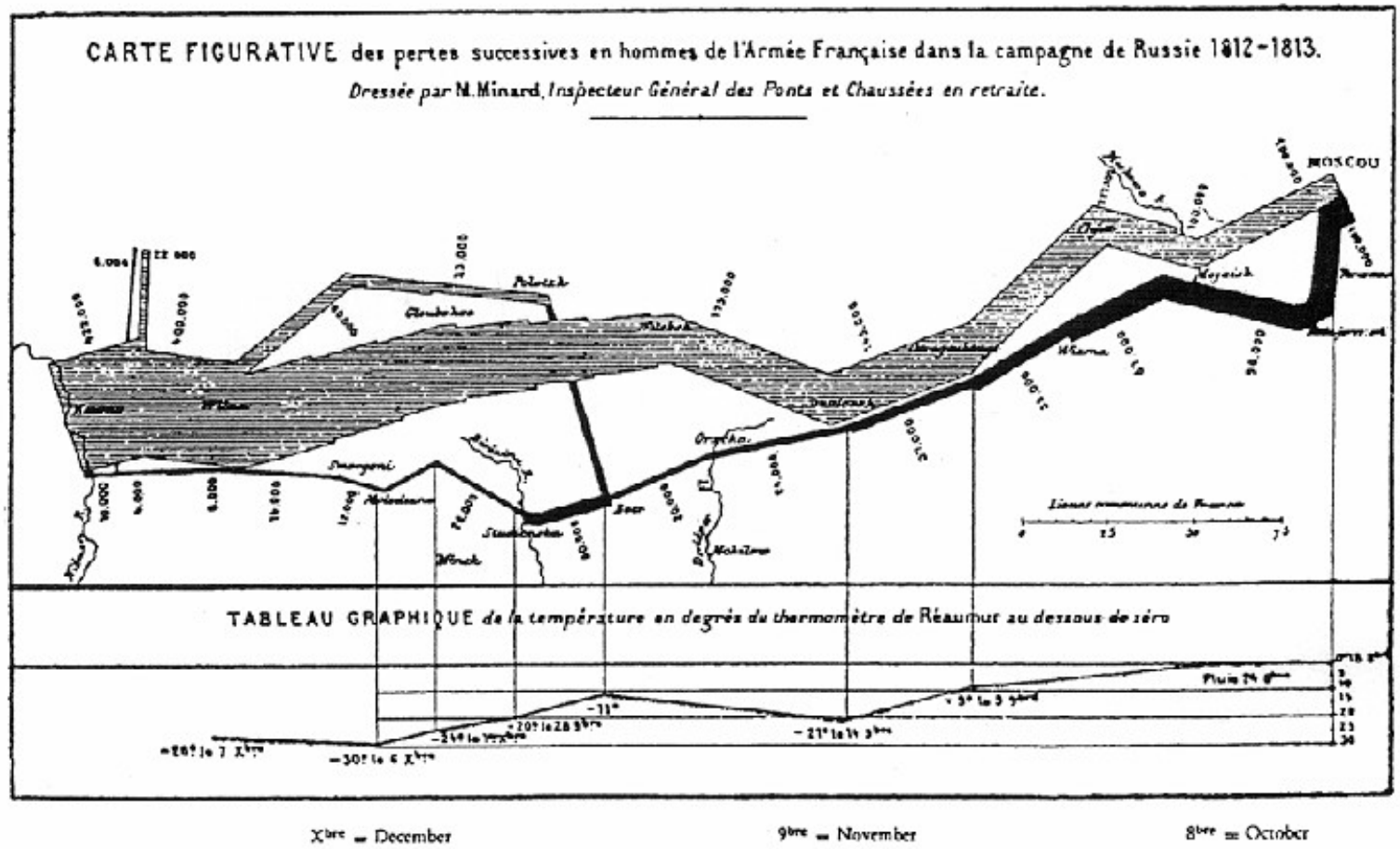
- Location-based store finder:
  - ***Range query:*** What are the restaurants within five miles of my location
  - ***Nearest-neighbor query***: Where is my nearest fast (junk) food restaurant

- Location-based advertisement:
  - ***Range query:*** Send E-coupons to all customers within five miles of my store

Mobility data analysis

- How people move around in the town
  - During the day, during the week, etc.
- Are there typical movement behaviours?
- Are there typical movement behaviours in a certain area at a certain time?
- How frequently people access the network?
- How are people movement habits changing in this area in last decade-year-month-day?
- Are there relations between movements of two areas?
- Are there periodic movements?

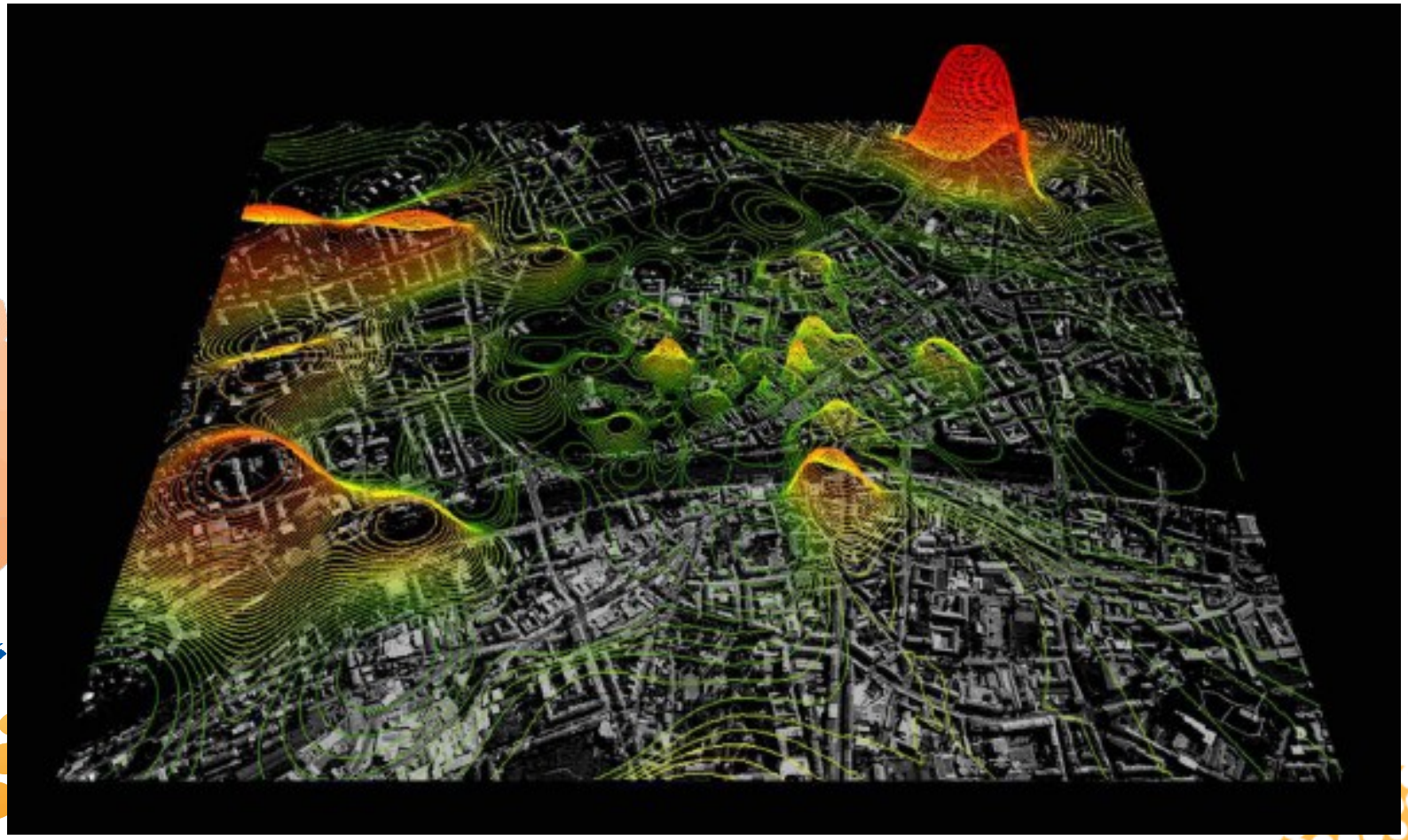
From *time-geography*



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

to interactive (recent/real) *time-geography*

Real-time density estimation in urban areas

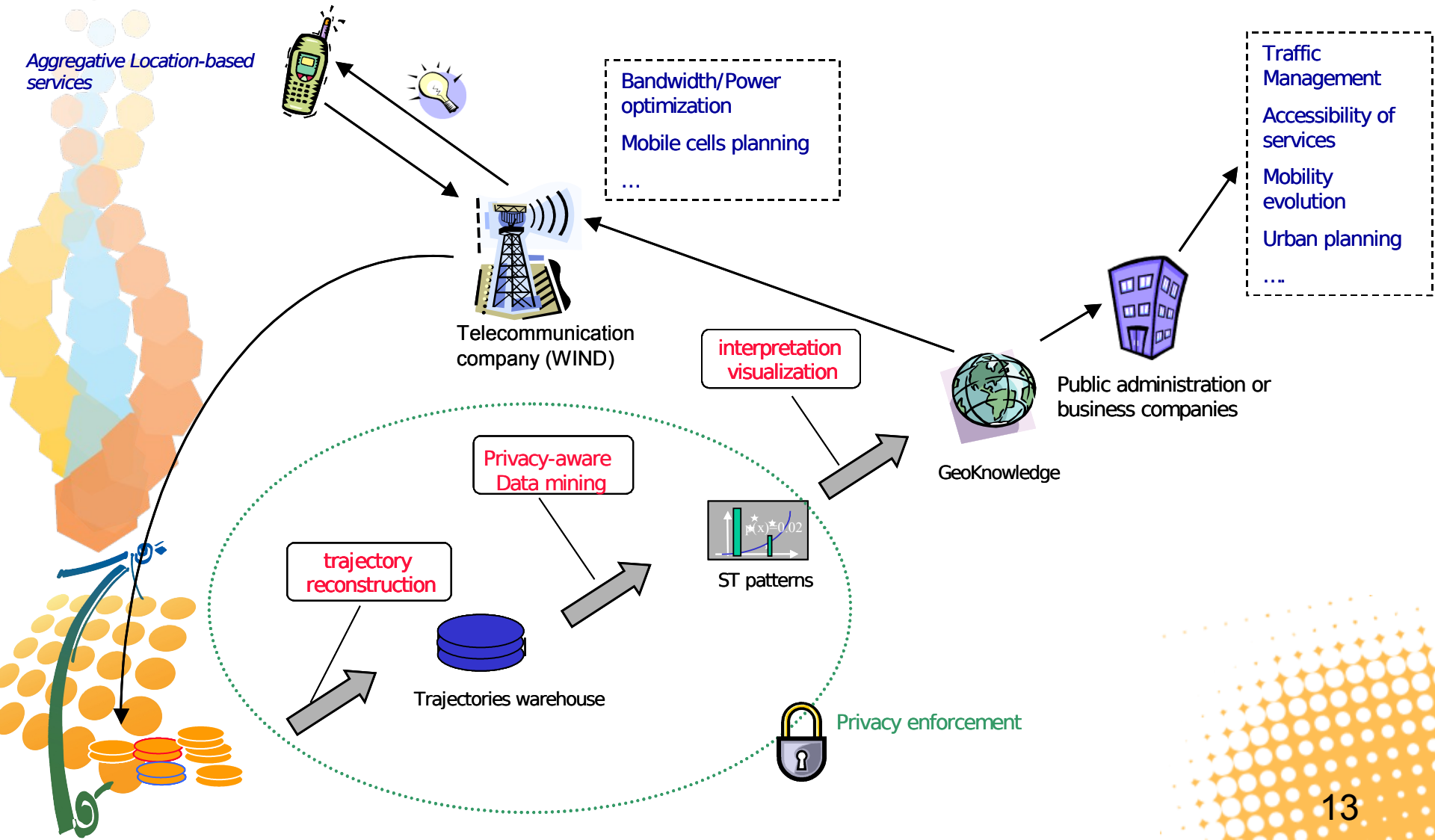


The senseable project: http://senseable.mit.edu/grazrealtime/

GeoPKDD general goal

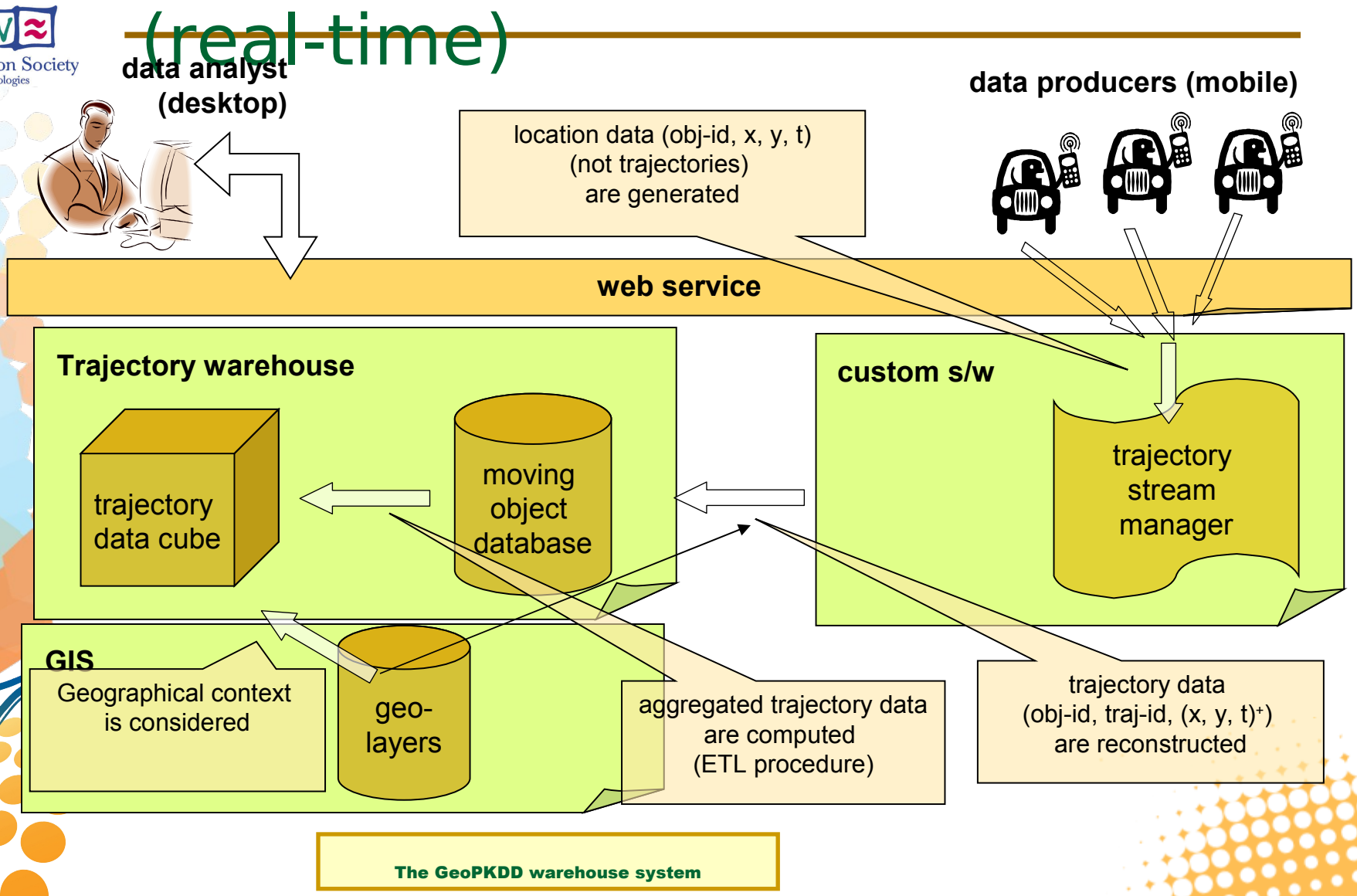
- is to develop
  - theory, techniques and systems for *geographic knowledge discovery and delivery*,
  - based on new automated *privacy-preserving* methods for extracting user-consumable forms of knowledge from large amounts of raw data referenced in *space* and in *time*.

Geographic privacy-aware Knowledge Discovery process



From movement data to movement patterns

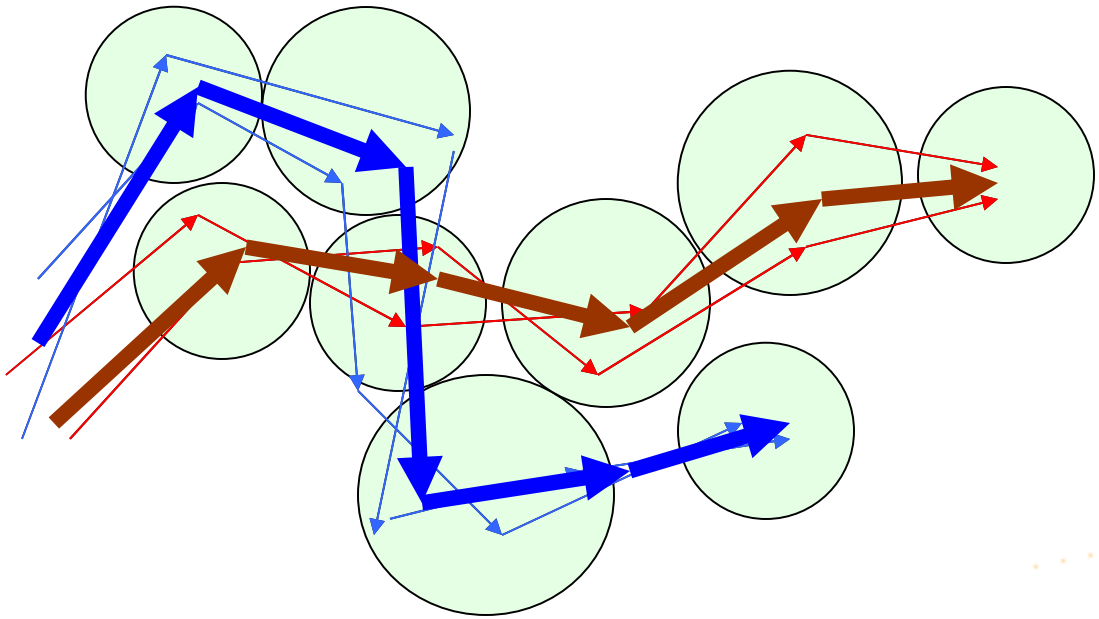
The GeoPKDD system architecture (real-time)



The GeoPKDD warehouse system

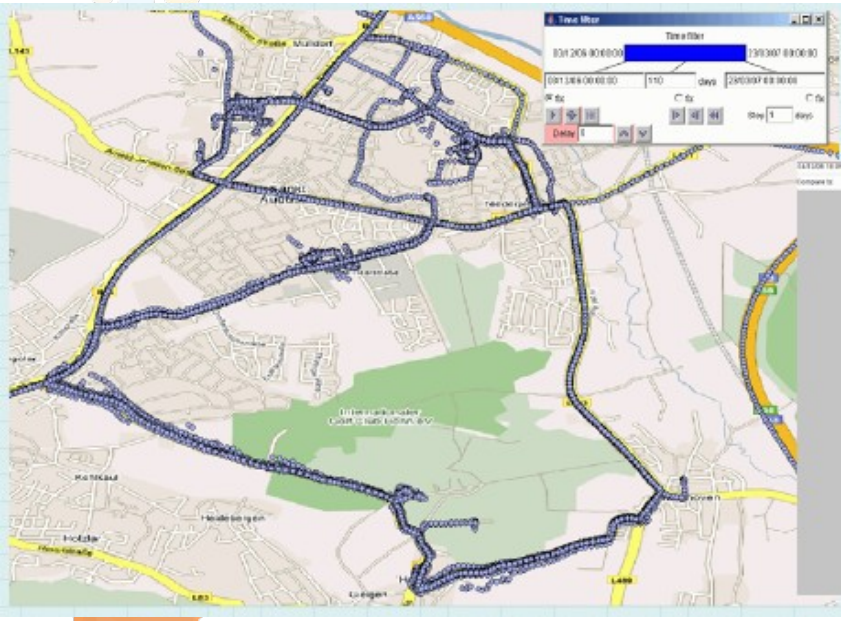
Mining Trajectories: Clustering

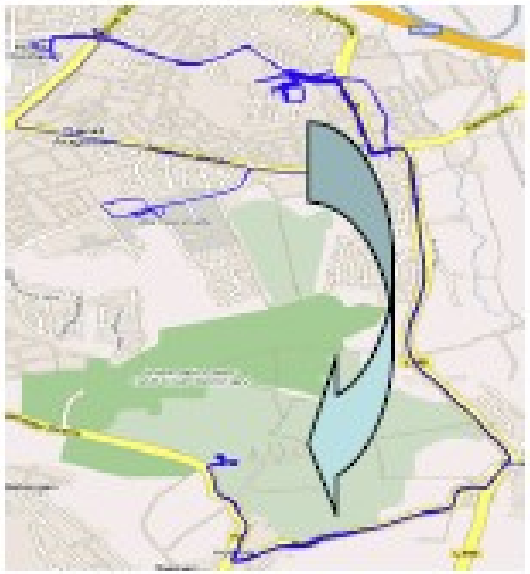
- Group together similar trajectories
- For each group produce a summary



= cell

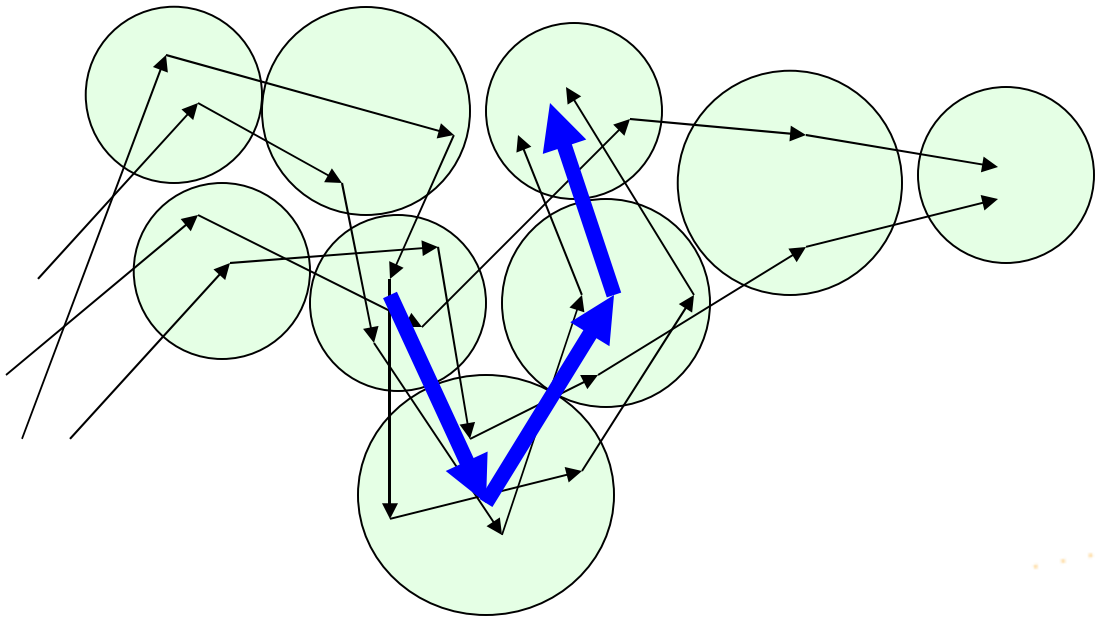
Trajectory Clustering





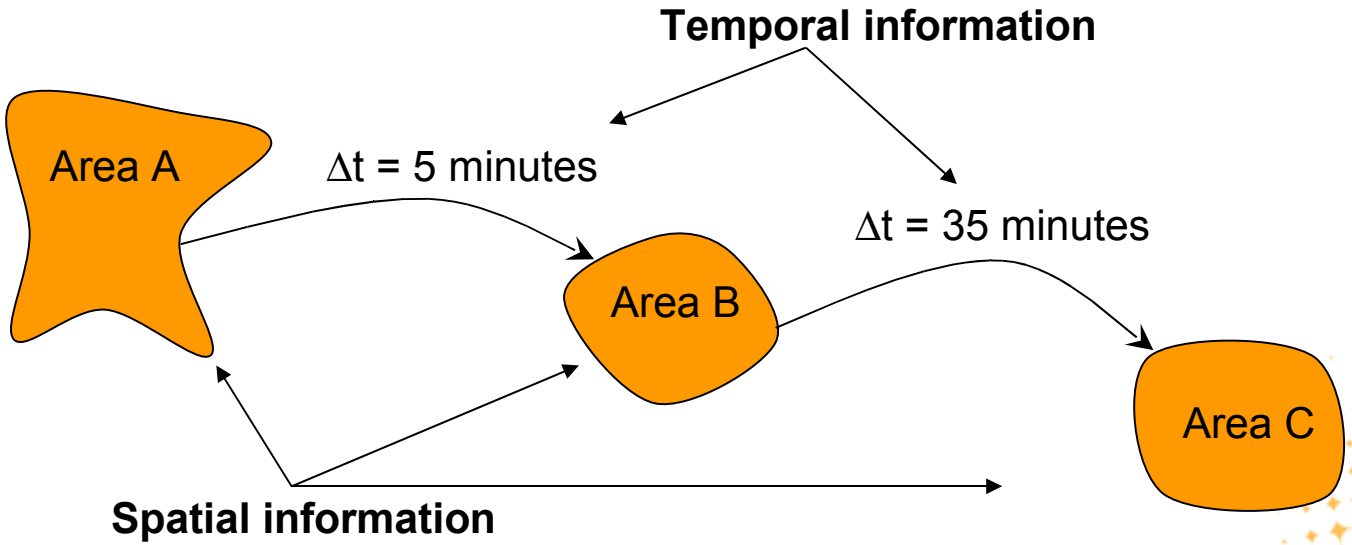
Mining Trajectories : Frequent patterns

- Discover frequently followed (sub)paths



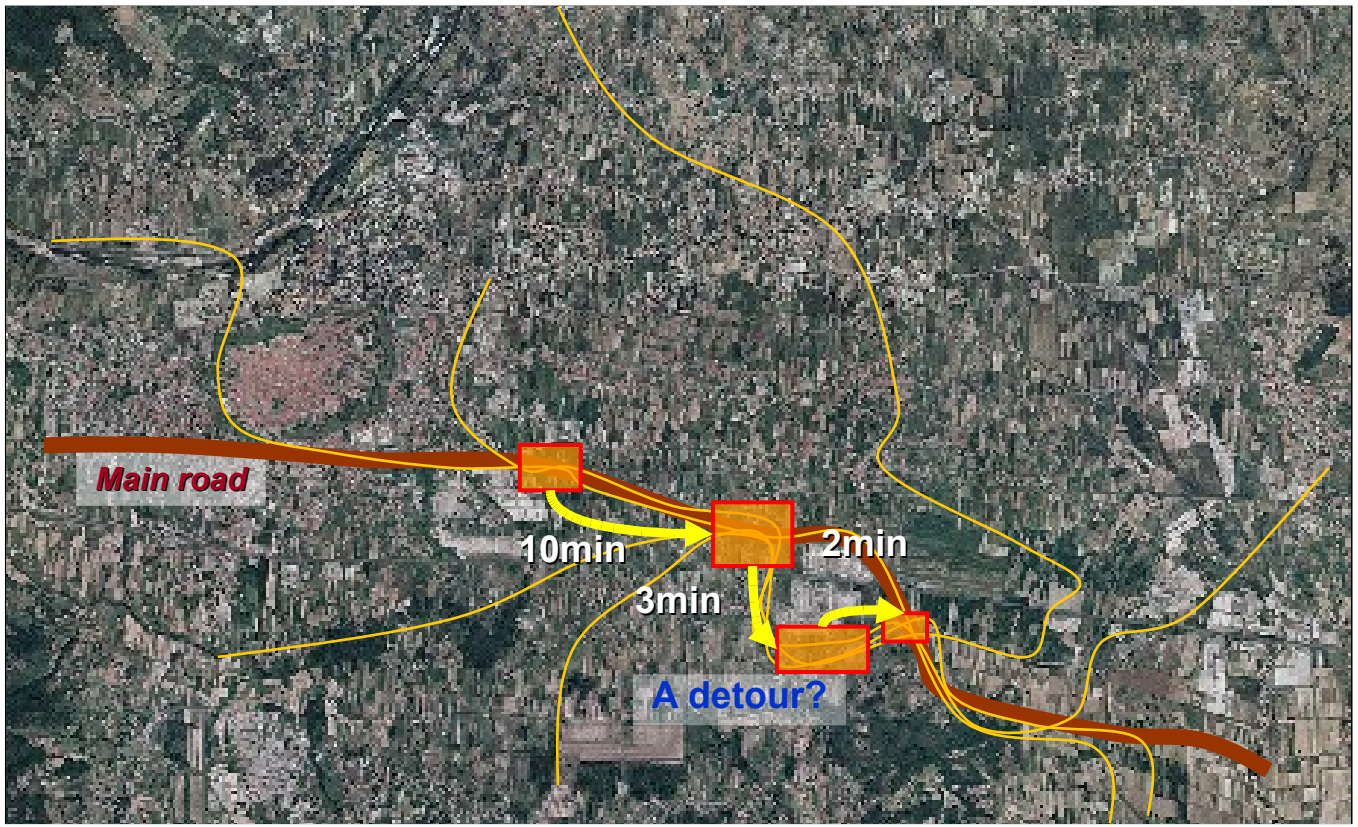
What is a trajectory pattern?

- A trajectory pattern is a **sequence of spatial regions** that, on the basis of the source trajectory data, emerge as frequently visited in the order specified by the sequence;
- in addition, the transition between two consecutive regions in such a sequence is annotated with a **typical travel time** that, again, emerges from the input trajectories.



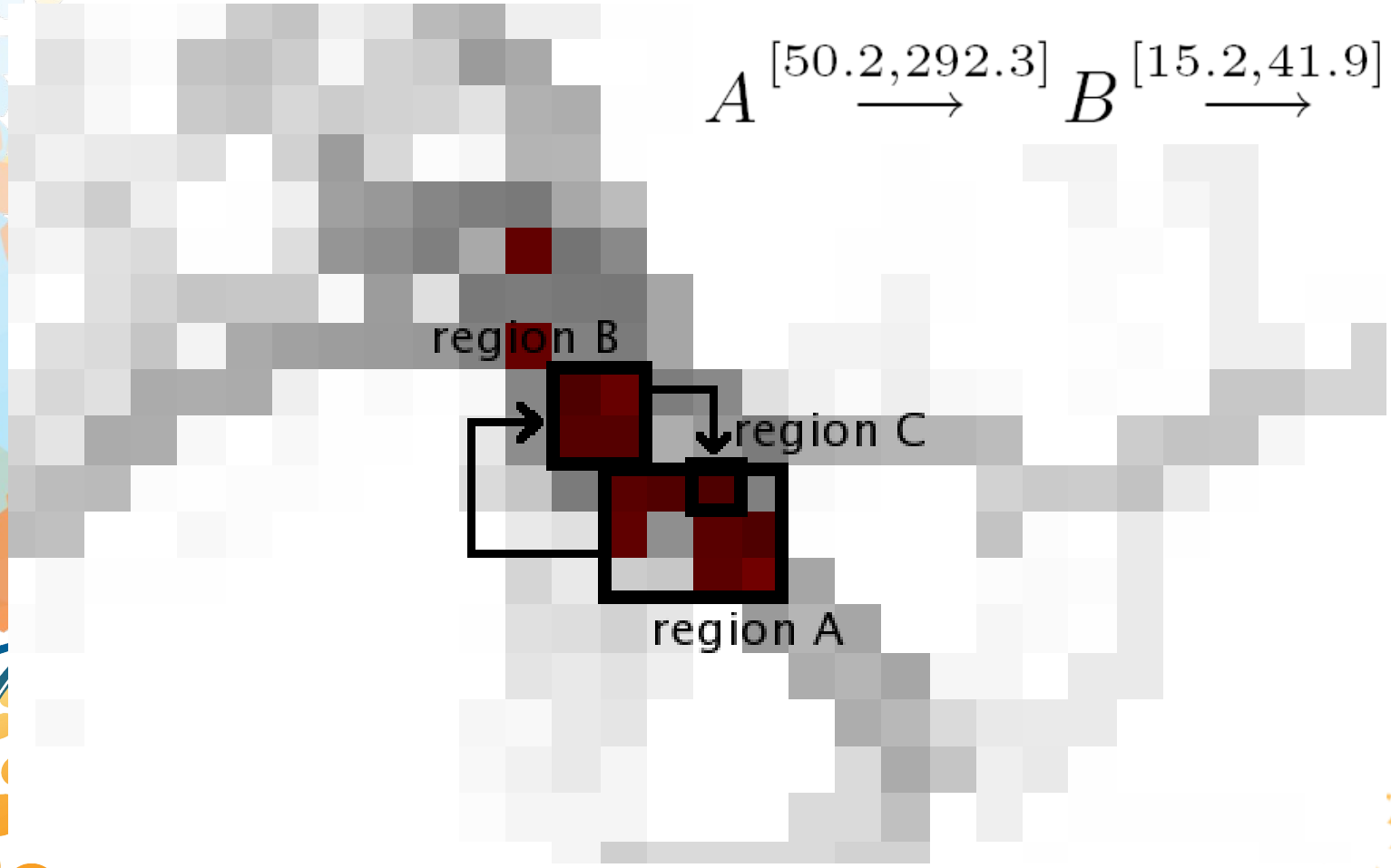
Visualizing T-patterns

Example



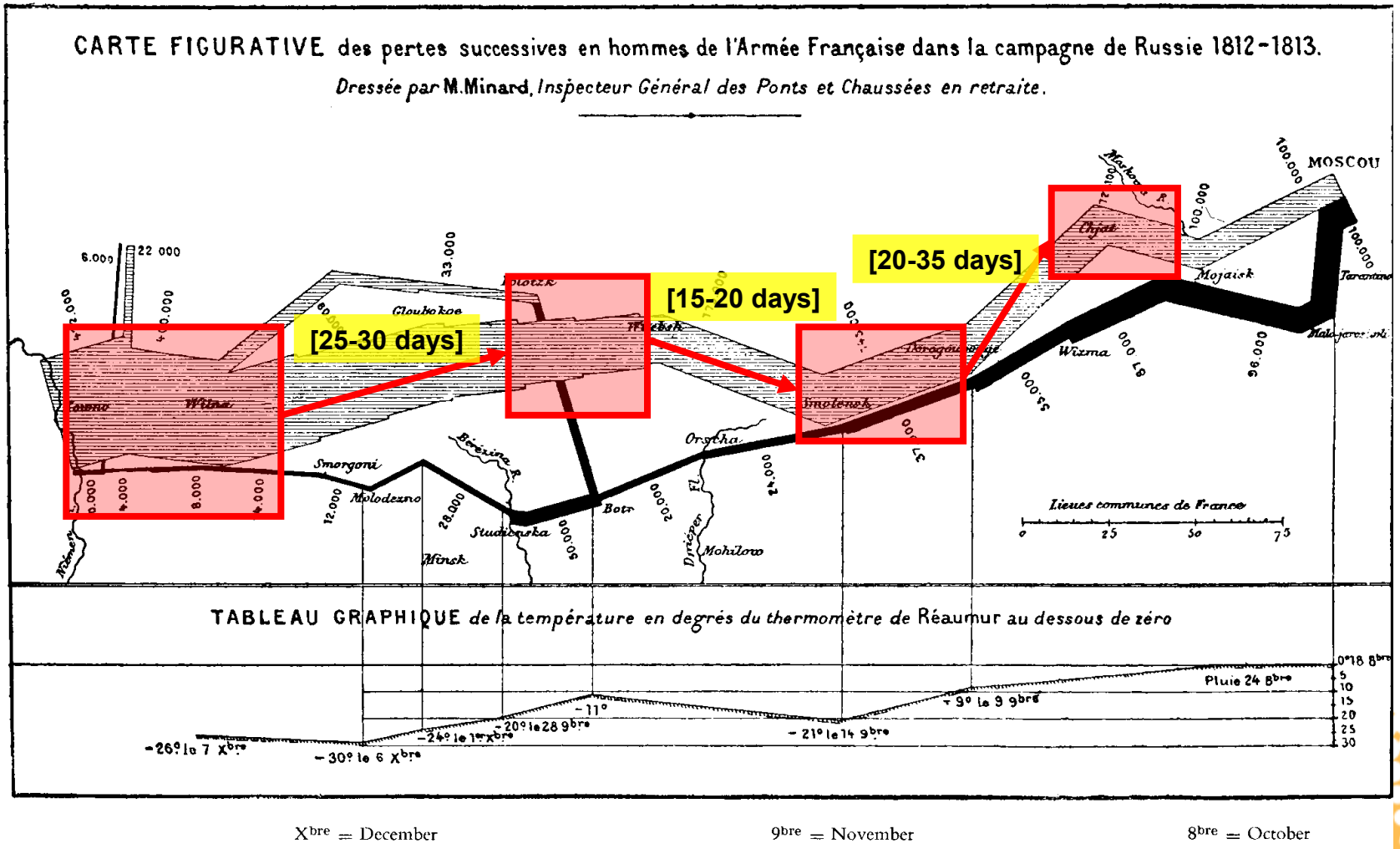
Visualizing T-patterns
Sample pattern (Athens trucks)

$$A \overset{[50.2,292.3]}{\longrightarrow} B \overset{[15.2,41.9]}{\longrightarrow} C$$



Visualizing T-patterns

Minard's map

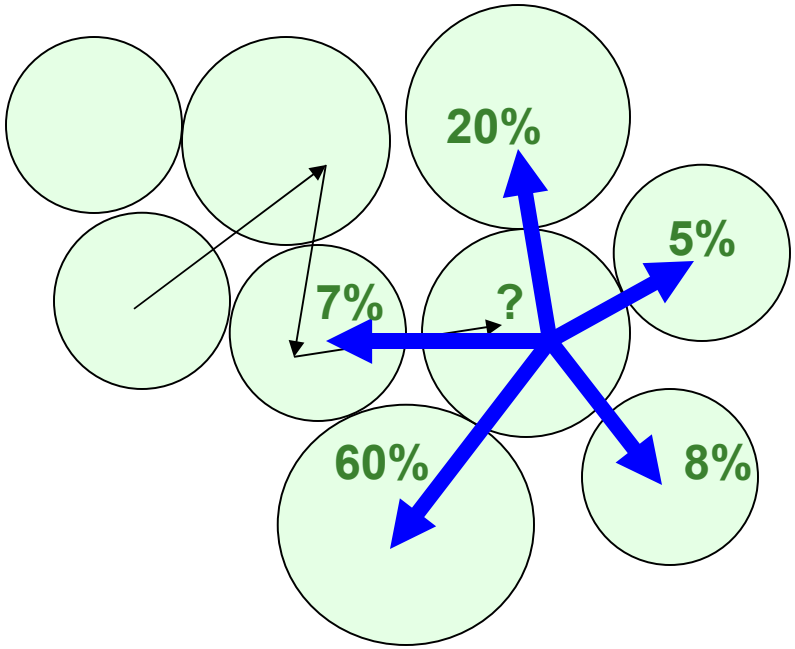


Visualizing T-patterns

Example

Mining Trajectories: classification models

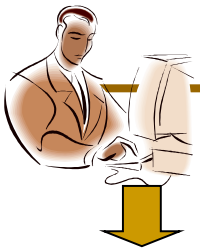
- Extract behaviour rules from history
- Use them to predict behaviour of future users

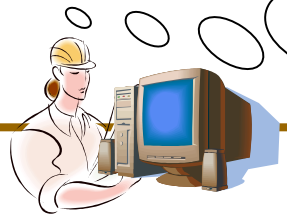


The GeoPKDD impact

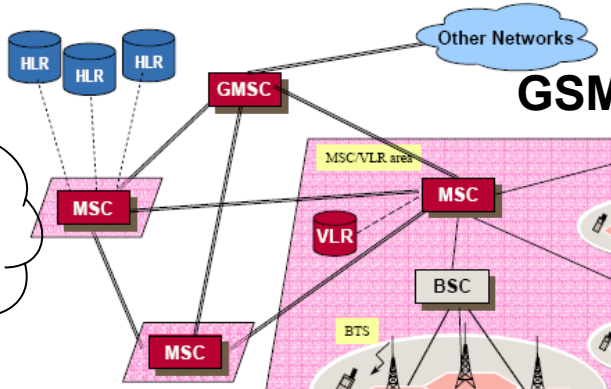
- Improving decision-making in mobility-related issues:
  - Planning traffic and public mobility systems in metropolitan areas;
  - Planning physical communication networks
  - Localizing new services in our towns
  - Forecasting traffic-related phenomena
  - Organizing logistics systems
  - Avoid repeating mistakes
  - Timely detecting changes.

Mobility Manager Office

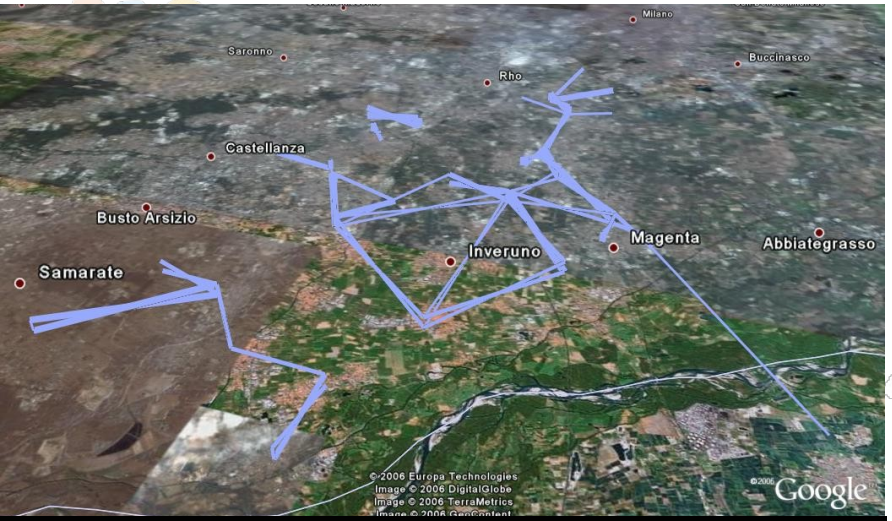


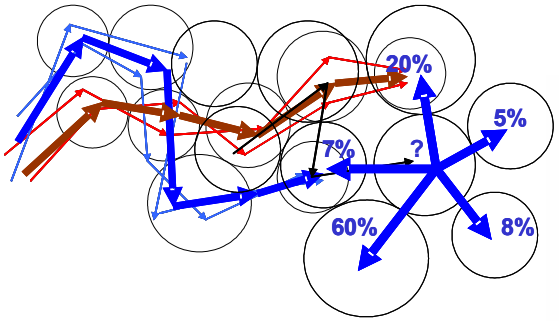


Sustainable Mobility?

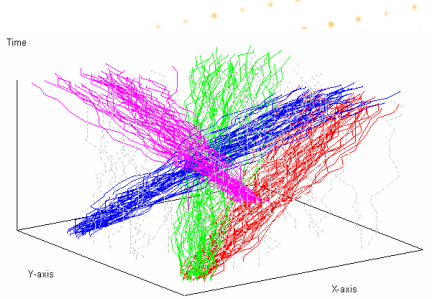


GSM network





Mobility Models



We are looking for real case to experiment on:

Information Society Technologies

Application Demonstrators

Amusement park Amsterdam

Brasilian Governamental Trasportation Management

GPS data from trucks in Athens

Comune di Milano ?

We are looking for case studies:

- **data sets, cartography and problems. The project will apply the mehods for free!**
- **Is there any public administration interested in experimenting on their data?**

We have a telecom. operator within the consortium, others may be involved.

Other experiments may be interesting: GPS data from car, or trucks

Why emphasis on privacy?

- More, better data are gathered, more vulnerability from linkage
- On the other hand, more and new data bring new opportunities
  - Public utility, new markets/paradigms, new services
- Need to maintain privacy without giving up opportunities
- Need to obtain social acceptance through demonstrably trustworthy solutions

Privacy in Mobility Data and Services

- Trusted/secure storage/Management of Mobility Data
- Privacy in Location Based Services:
  - the right of a user to receive a service without revealing his/her identity
  - Trade-off between quality of service and privacy protection
- Privacy and Anonymity in Mobility Data Analysis
  - Trade-off between privacy protection and analysis opportunities

Privacy in GeoPKDD

- How to design Data Analysis methods that, **by construction**, meet the the privacy constraints?
- How to develop trustable data mining technology capable of producing
  - ***provably/measurably*** privacy-preserving patterns
  - which may be safely distributed

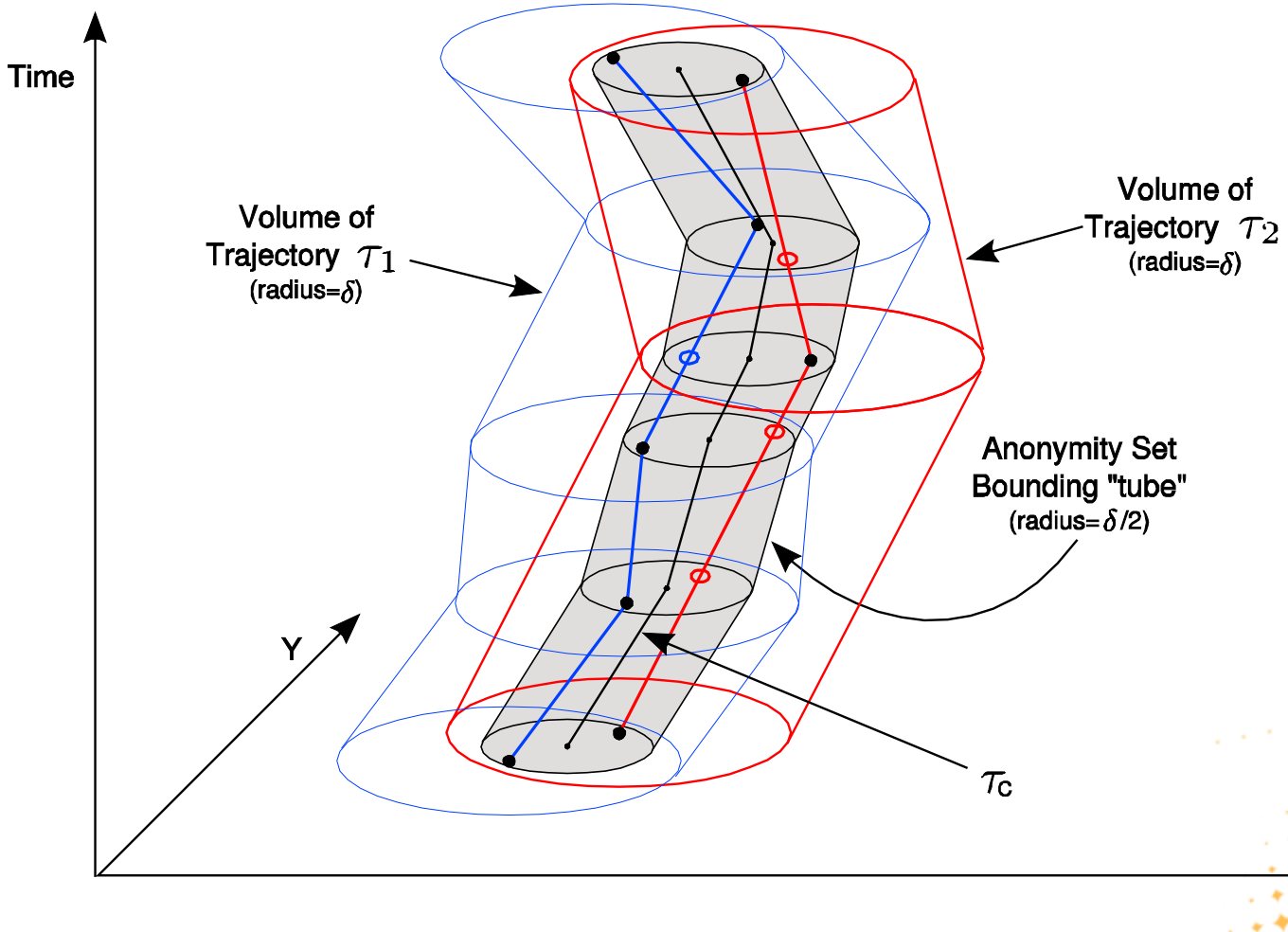
Scientific Privacy Issues in GeoPKDD

- Is there any specific challenge/risk/opportunity in the context of ST data?
  - New threats from traces analysis: learning who you are from where you have been (Malin et al 2003)
  - Space and Time in a trajectory can act as quasi-identifiers (Bettini and Jajodia 2005)
- How to formalize privacy measures over Spatio-Temporal data and Spatio-Temporal patterns?
  - E.g., anonimity threshold on clusters of individual trajectories

GeoPKDD Privacy Observatory

- Privacy cannot be achieved by technology alone
  - it's a combined social, ethical, legal and technological matter.
- The GeoPKDD Observatory interacts with stakeholders in privacy issues. Activities:
  - create and maintain relationships with European and national authorities for data protection and other privacy related organizations,
  - implement regulations into KDD methods and tools,
  - provide ideas for revisions of regulations themselves by means of novel privacy preserving technologies.
- http://www.geopkdd.eu/pro
- Tight collaboration with KDubiq's WG5

Anonymity and Trajectories



Announcements

Information Society
Technologies

The GeoPKDD boo

Information Society Technologies

Fosca Giannotti and Dino Pedreschi (Eds.)

Mobility, Privacy, and Data Mining.

Springer, Dec. 2007.