



Mobility, Data Mining and Privacy

www.geopkdd.eu

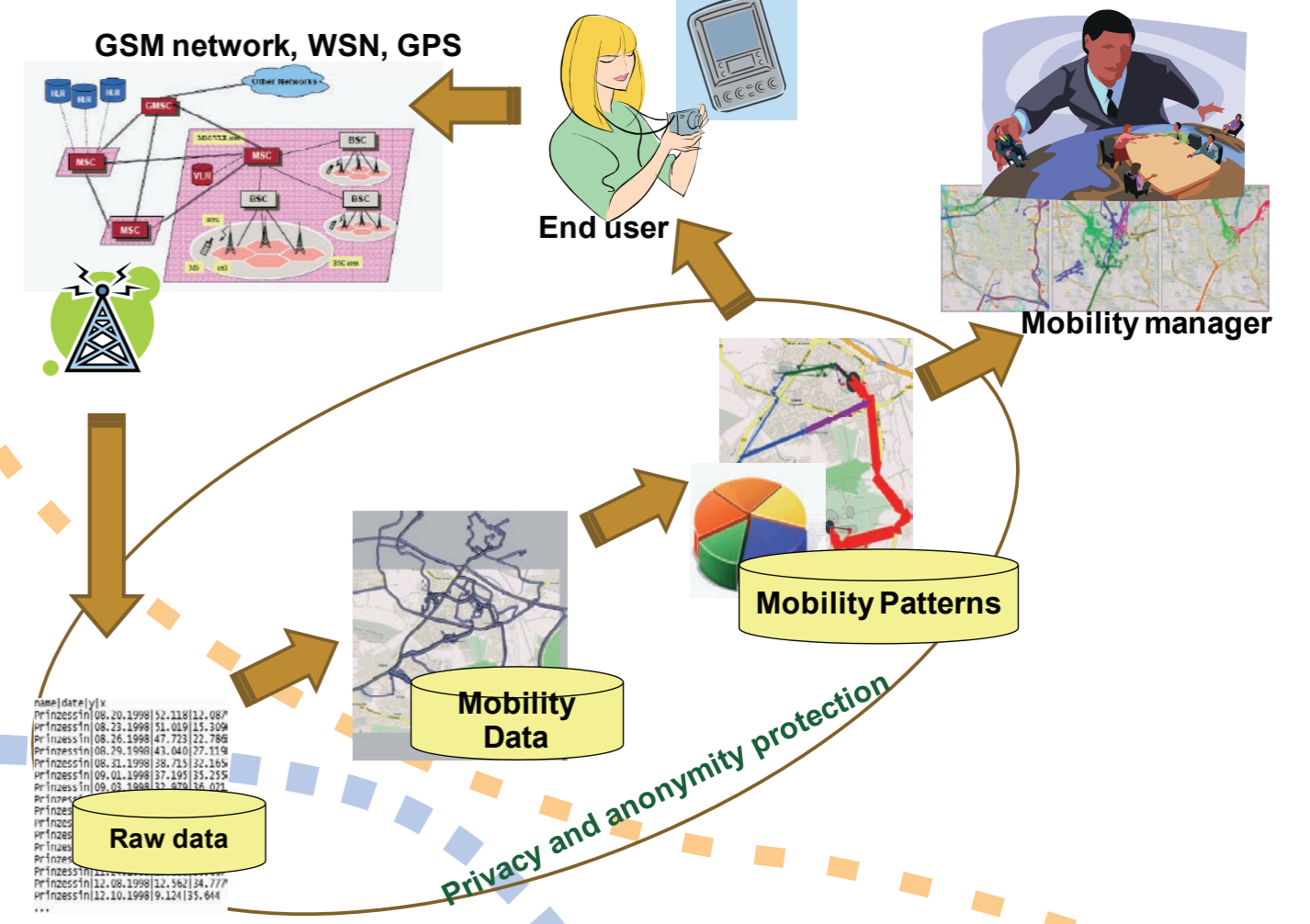
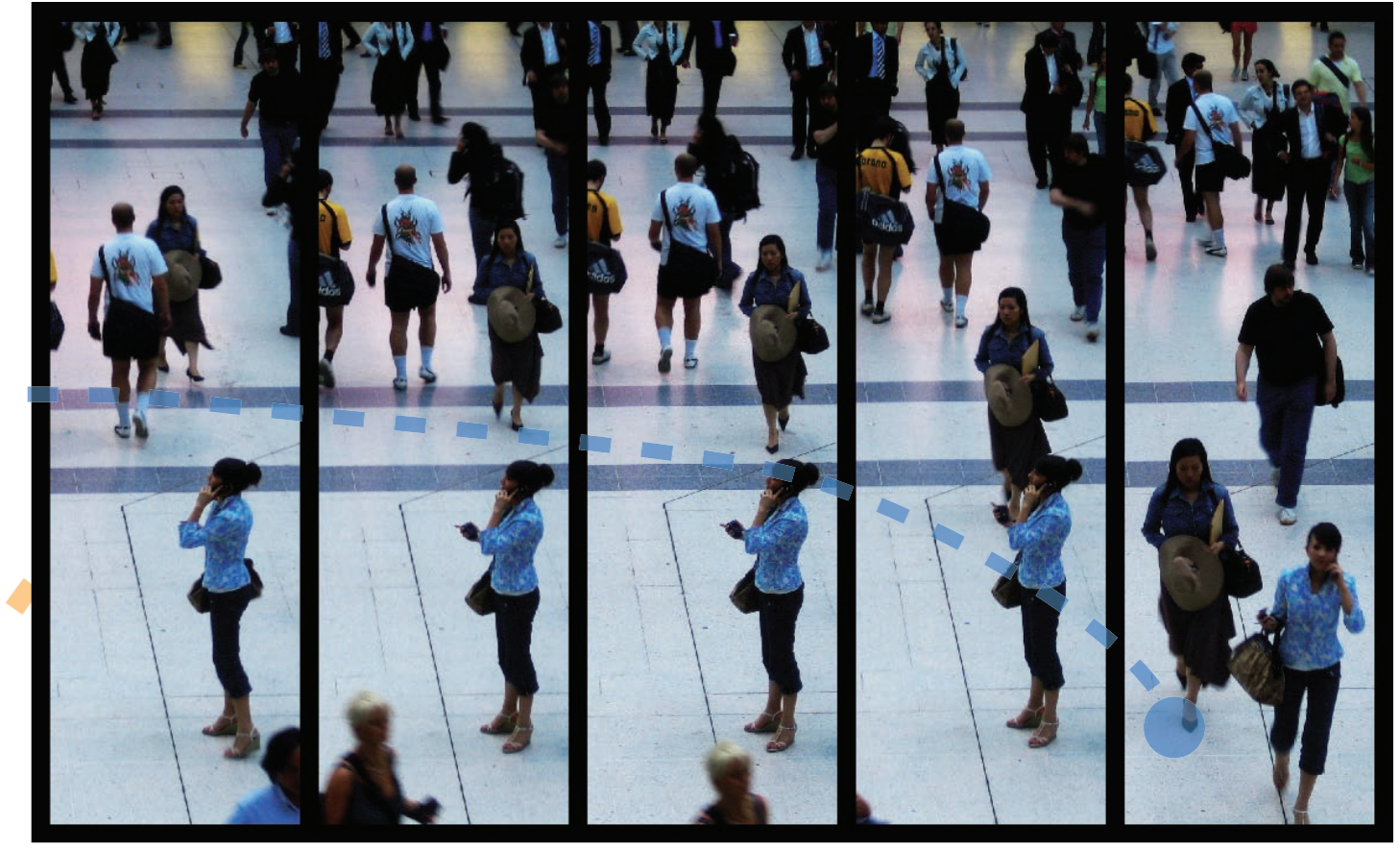
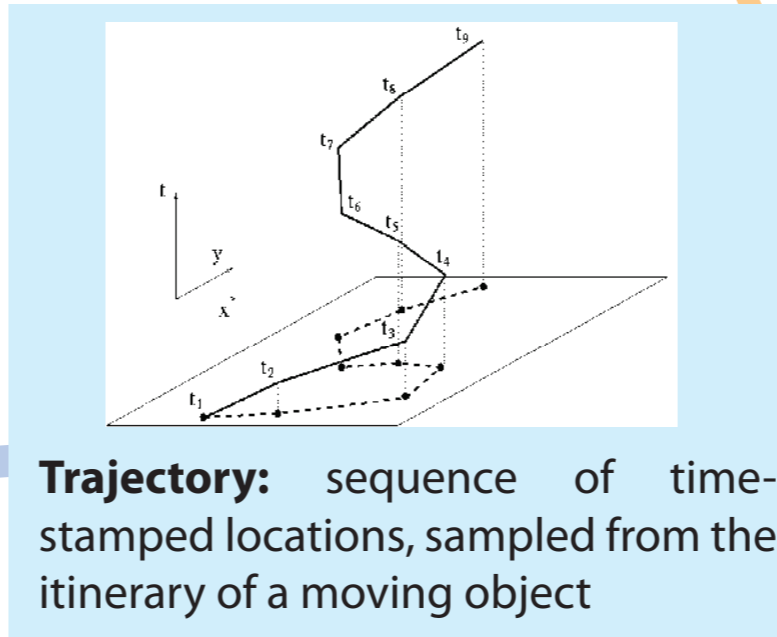
The technologies of mobile communications and ubiquitous computing pervade our society, and wireless networks sense the movement of people and vehicles, generating large volumes of mobility data. Miniaturization, wearability, pervasiveness is producing traces of our mobile activity, with increasing positioning accuracy and semantic richness:

- Location data from mobile phones: GSM cell positions
- GPS tracks from mobile devices receiving geo-positions from satellites

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.

GeoPKDD objective: How to discover useful knowledge about human movement behaviour from mobility data, while preserving the privacy of the people under observation? GeoPKDD has started a new exciting multidisciplinary research area, at this crossroads of **mobility, data mining, and privacy.**

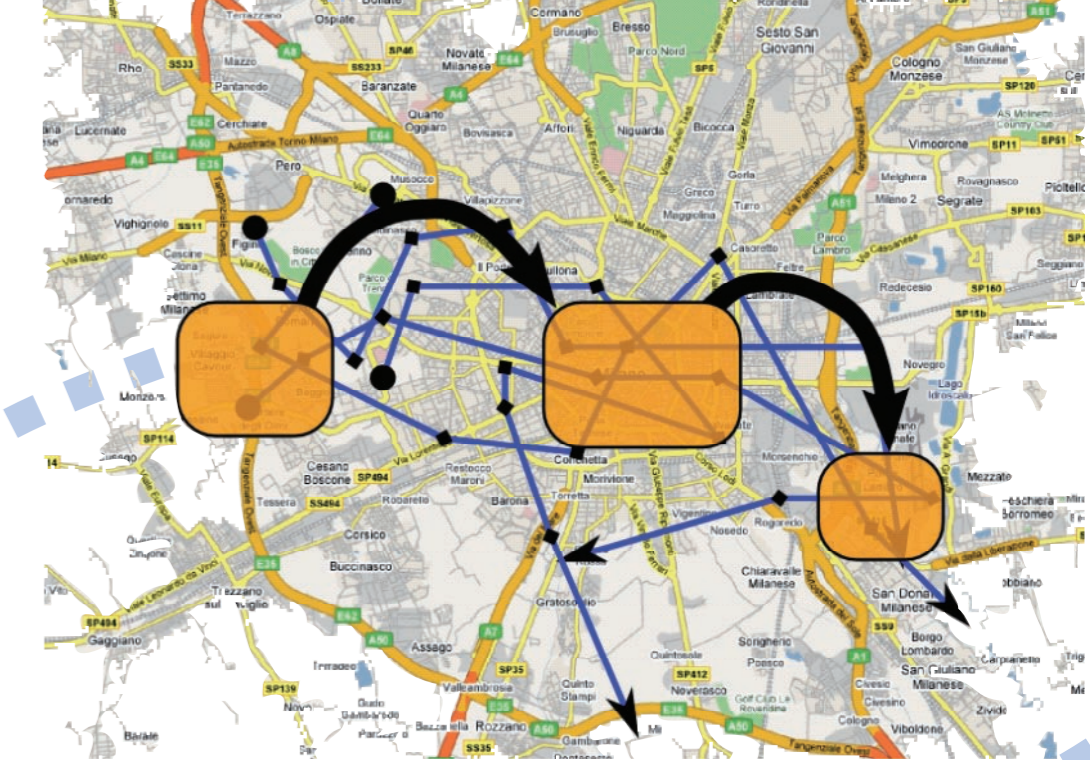
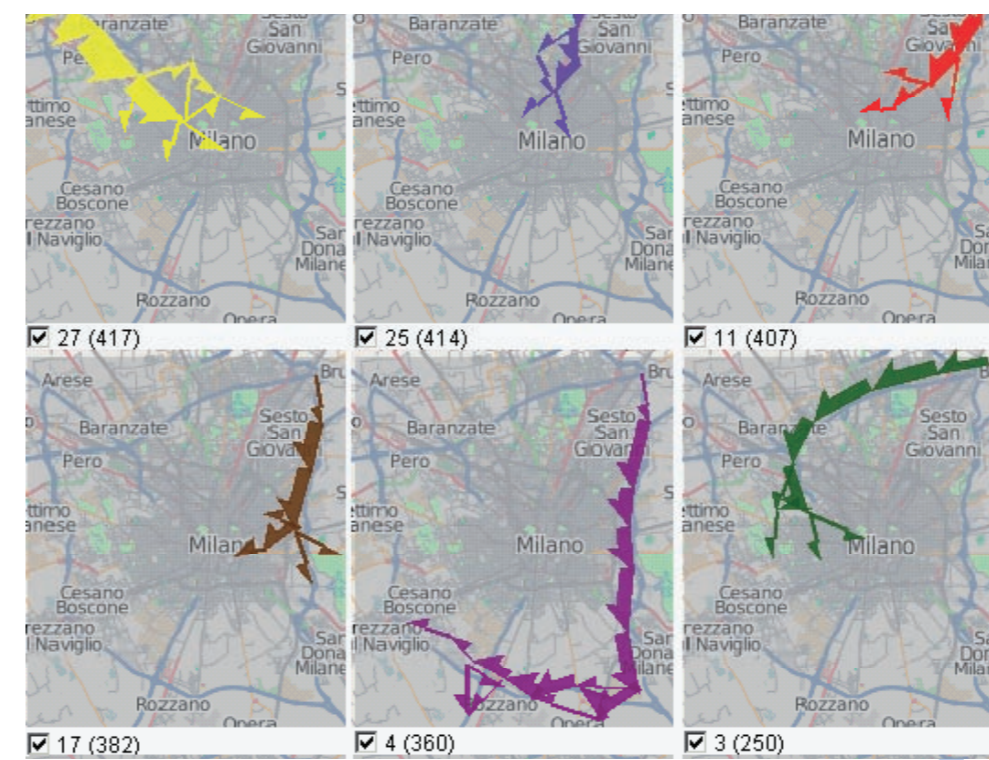
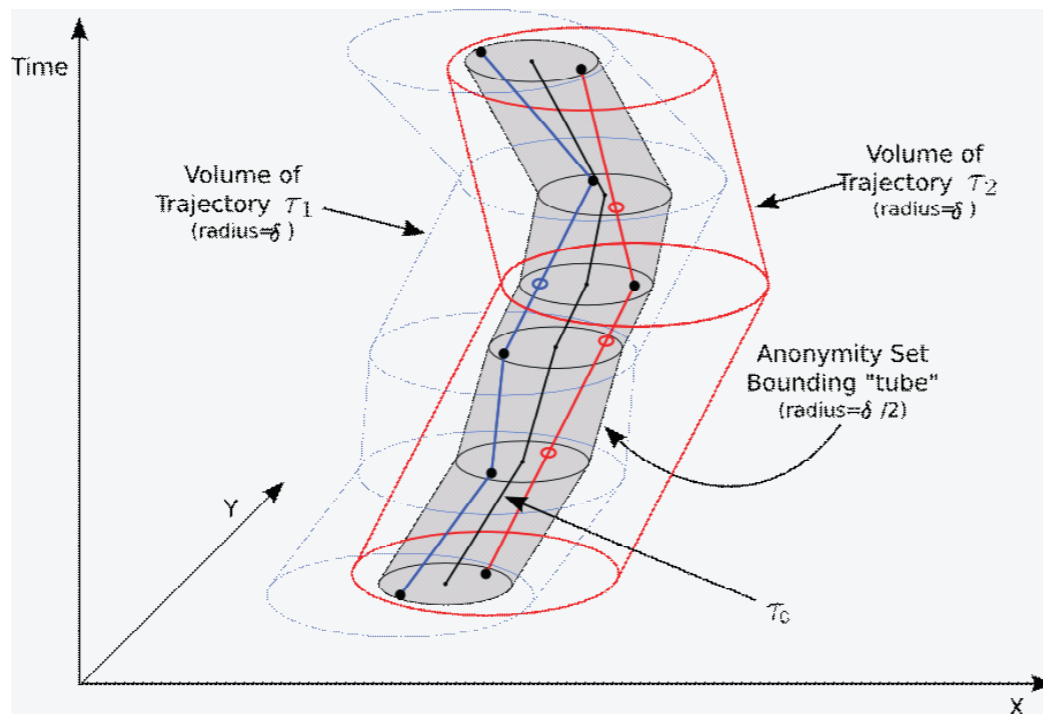
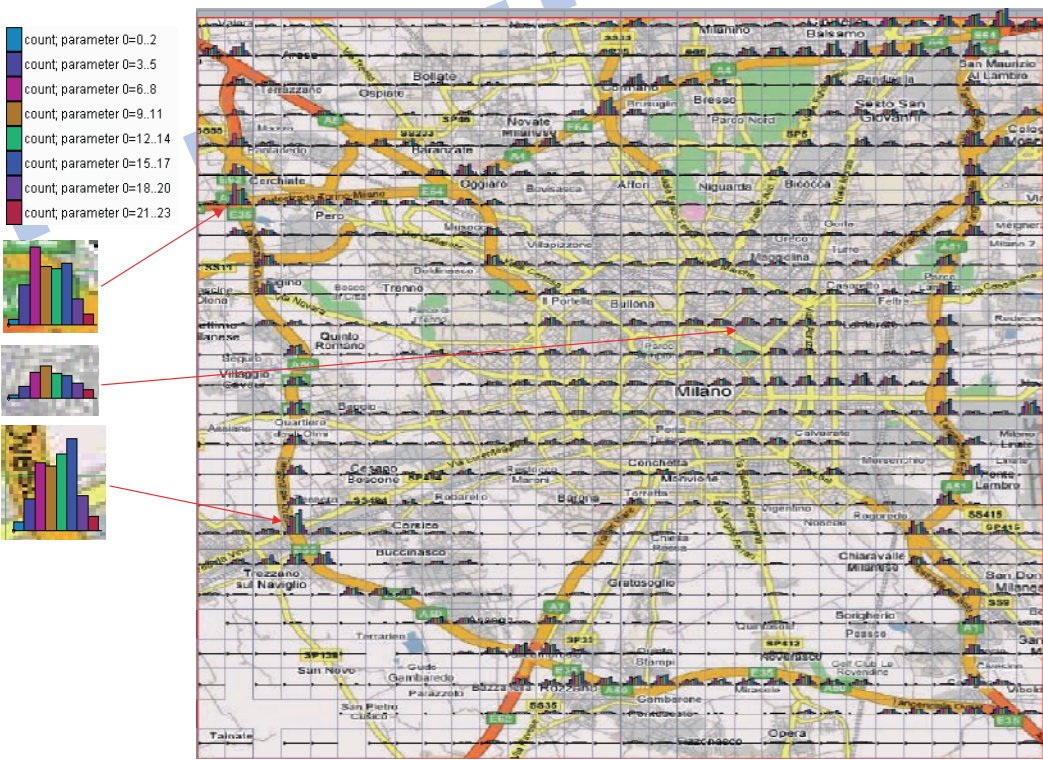


Project Number IST-6FP-014915
Project Start 1 December 2005
Project End 31 March 2009
Coordinator KDDLAB, Knowledge Discovery and Delivery Laboratory, ISTI-CNR and University of Pisa

GeoPKDD is a project in the Future and Emerging Technologies programme of the Sixth Framework Programme for Research of the European Commission, FET-Open contract n: 014915



Breakthrough analytical methods for mining from massive trajectory datasets



T-Warehouse

Problem: How to explore the aggregated properties of moving objects along the spatio-temporal dimensions, such as the presence or speed of vehicles in the various locations at varying times?
Idea: A T-warehouse is a spatio-temporal data cube representing various aggregated measures of the moving objects.
Method: our T-OLAP engine allows exploratory analysis, drilling up and down the spatial and temporal dimensions.

T-Anonymity

Problem: Mobility data may reveal personal habits, preferences, etc. How to avoid that a malicious attacker reconstructs the exact identity of a person associated to a de-identified trajectory?
Idea: A k-anonymous trajectory dataset is one where the itinerary of each person is indistinguishable from that of other k-1 persons (anonymity = hiding in the crowd).
Method: our T-anonymity methods transform a trajectory dataset into a new, k-anonymous dataset, such that the key analytical properties are preserved.

T-Clustering

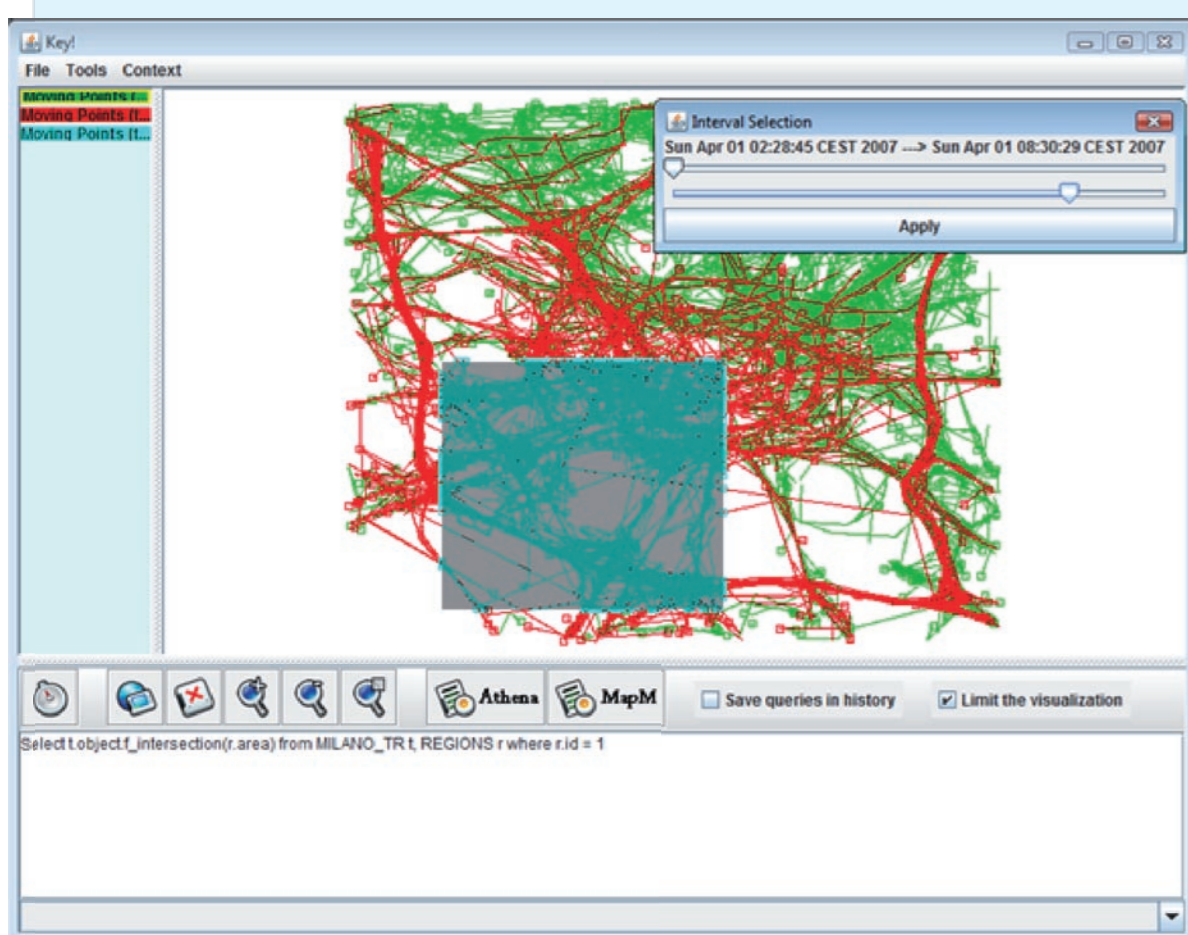
Problem: How to discover groups of objects with similar movement behaviour in massive trajectory datasets, such as a systematic home-work-home commuting behaviour?
Idea: A T-cluster is a set of similar trajectories, according to a repertoire of trajectory similarity functions
Method: our density-based T-clustering algorithm discovers clusters in trajectory data

T-Patterns

Problem: How to discover frequent itineraries in massive trajectory datasets?
Idea: A T-pattern is a sequence of locations that are frequently visited in the specified order with similar transition times
Method: our T-pattern mining algorithm automatically discovers frequent T-patterns in trajectory data.

Mastering the Complexity of the GeoPKDD Process

Semantic-based query and reasoning system



Spatio-temporal query primitives

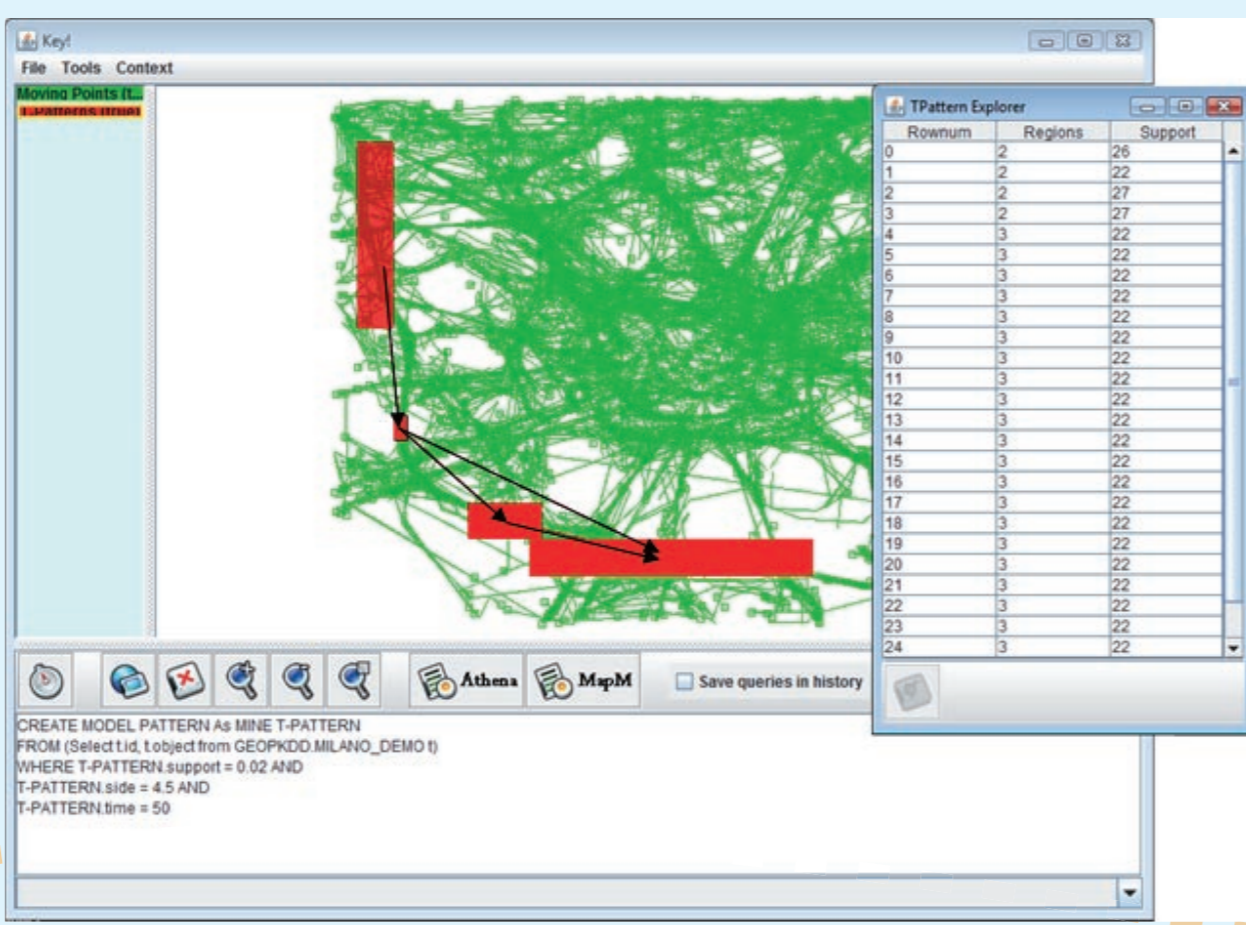
- to select and pre-process trajectory data w.r.t. geographic background knowledge
- trajectory reconstruction and anonymization primitives

Trajectory mining primitives

- to extract various mobility patterns and models
- a **Data Mining Query Language**, to allow the user drive the entire analytical process

A reasoning component

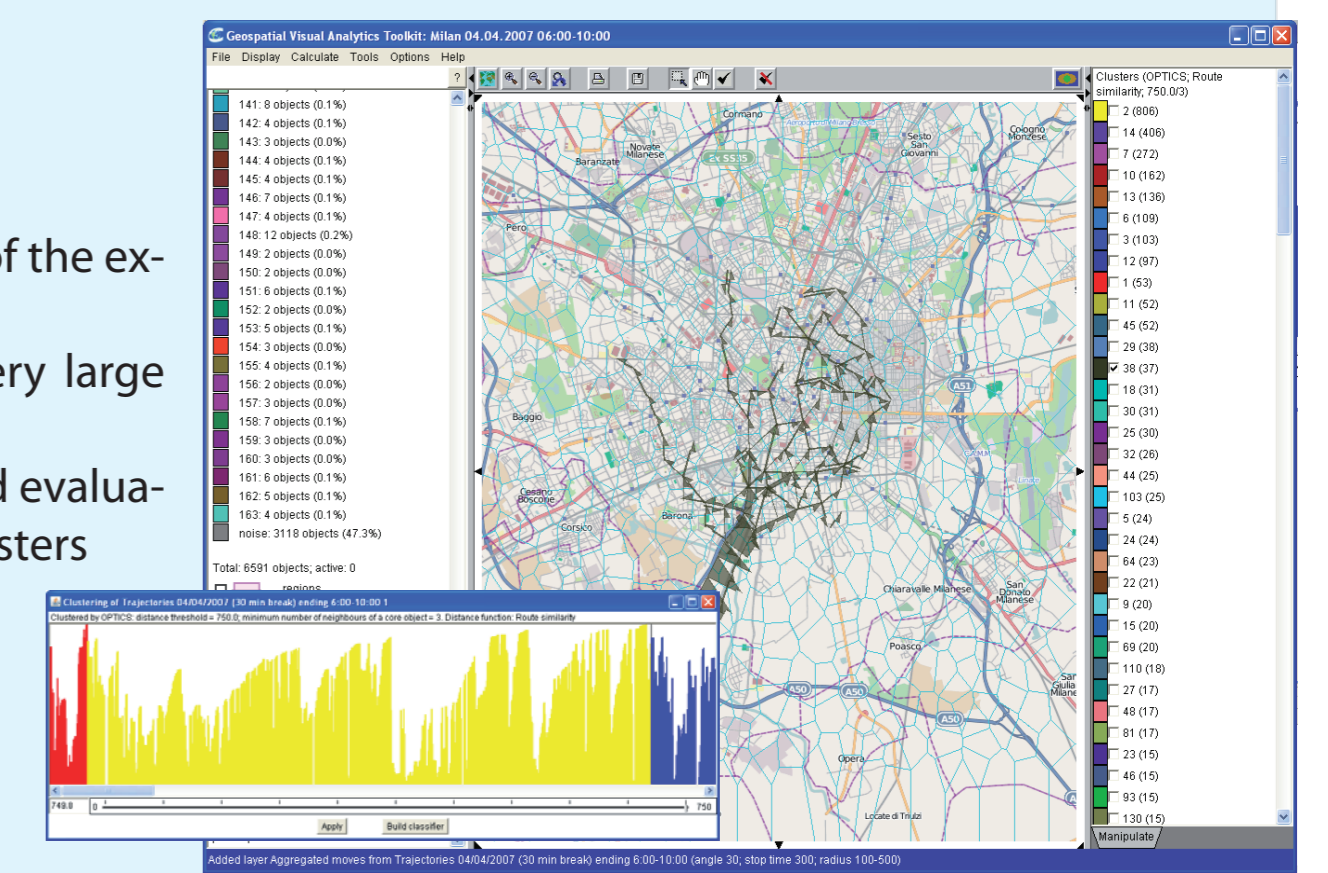
- to specify domain-driven ontologies
- to infer types of trajectories and patterns



Visual Analytics environment

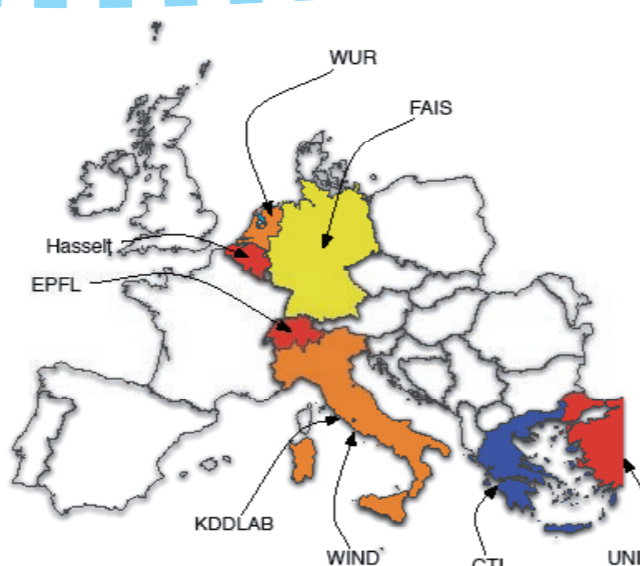
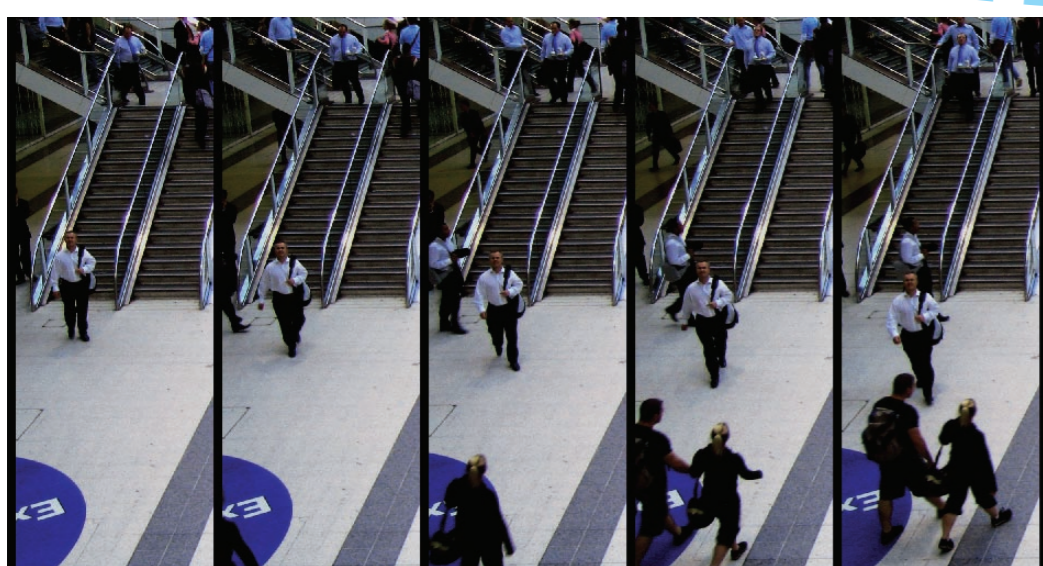
Visualization of T-Patterns

- to support the navigation of the extracted patterns
- interactive clustering of very large trajectory datasets
- user-driven exploration and evaluation of the discovered T-clusters

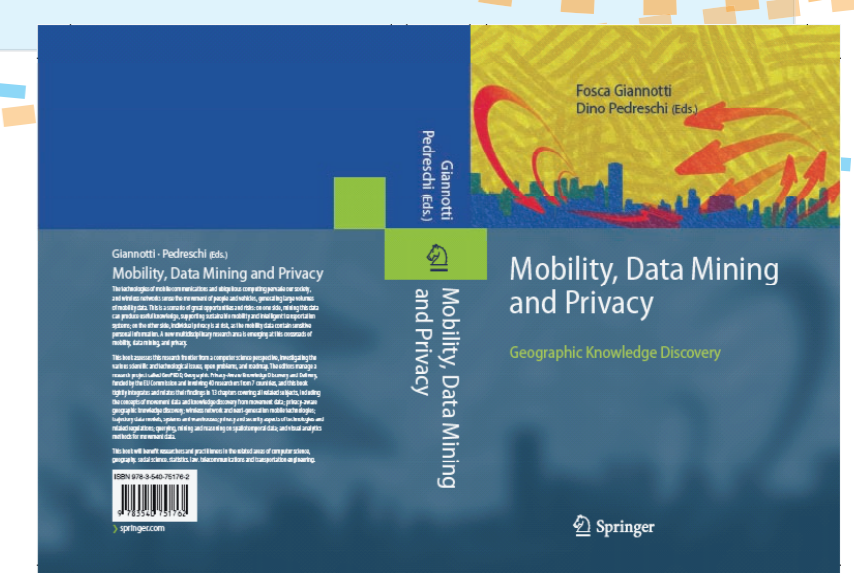


Visual exploration of the T-Warehouse

- visual T-OLAP operations to browse aggregated measures, such as aggregated presence and speed of moving objects (triangle base = presence, triangle height = speed)



- (I) **KDDLab** Knowledge Discovery and Delivery Laboratory, ISTI-CNR, Istituto di Scienza e Tecnologie dell'Informazione, jointly with Univ. Pisa <http://www.isti.cnr.it/> — <http://www.di.unipi.it/> Fosca Giannotti (fosca.giannotti@isti.cnr.it)
- (B) **Hasselt University** Theoretical Computer Science Group <http://www.luc.ac.be/theocomp> Bart Kuijpers (bart.kuijpers@uhasselt.be)
- (CH) **EPFL** <http://lbdwww.epfl.ch/> Stefano Spaccapietra (stefano.spaccapietra@epfl.ch)
- (D) **Fraunhofer IAIS** Institute for Autonomous Intelligent Systems, Sankt Augustin. <http://www.ais.fraunhofer.de/> Michael May (michael.may@ais.fraunhofer.de)
- (GR) **CTI** Research Academic Computer Technology Institute, Research and Development Division — jointly with Univ. Piraeus, Dept. of Informatics <http://www.cti.gr/> — <http://www.unipi.gr/> Yannis Theodoridis (lytheod@unipi.gr)
- (NL) **Wageningen UR** Centre for Geoinformation <http://cgi.girs.wageningen-ur.nl/>
- (ES) **Universidad Politécnica de Madrid** Monica Wachowicz (Monica.Wachowicz@wur.nl) Ecole Polytechnique Fédérale de Lausanne. <http://www.sabanciuuniv.edu/>
- (TK) **Sabancı University** Faculty of Engineering and Natural Sciences. <http://www.sabanciuuniv.edu/> Yucel Saygin (ysaygin@sabanciuuniv.edu)
- (I) **WIND** Telecomunicazioni SPA, Direzione Reti Wind Progetti Finanziati & Technology Scouting Riccardo Mazza (Riccardo.Mazza@Mail.wind.it)



F. Giannotti and D. Pedreschi (Eds.)
Mobility, Data Mining and Privacy. Springer, 2008