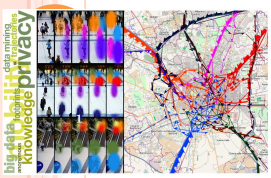
CIKM 2010 Tutorial on
Mobility Data: Modeling, Management, and Understanding

S. Spaccapietra, E. Zimányi, C. Renso

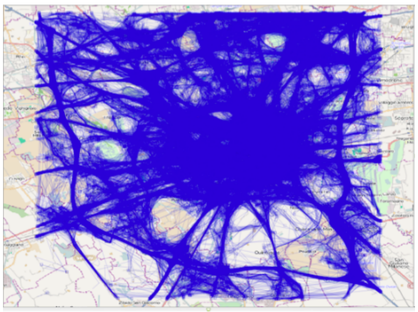
Mobility Data Understanding



Chiara Renso
KDDLab, ISTI-CNR,
Pisa, Italy
chiara.renso@isti.cnr.it

People move in many ways…….

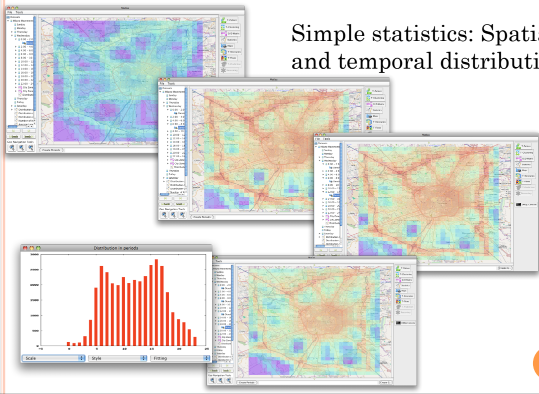
How do people move during the day?



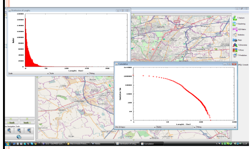
One day GPS tracks of 17.000 cars in the Milan (Italy) area

How do people move during the day?

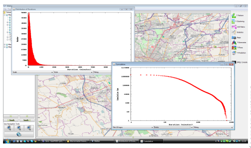
Simple statistics: Spatial and temporal distributions



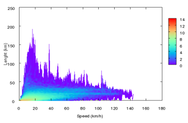
THE COMPLEXITY OF PEOPLE'S TRAVELS



How long: several short trips, few very long ones



How far: most trips last less than two hours, a few last up to 5 hours



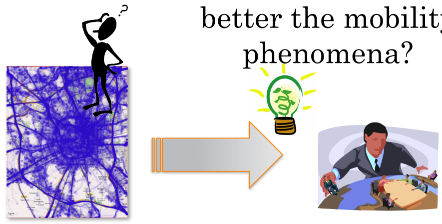
How fast (length & speed): long travelers go slow (only slow trips are very long up to 200 km), fast travelers go short trips

Simple statistics do not help!

The average behavior makes no sense

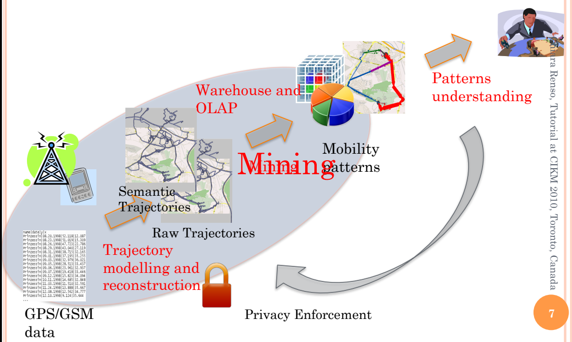
We need more complex techniques to fully understand mobility data!

How can we understand better the mobility phenomena?



DESIGNING A MOBILITY KNOWLEDGE DISCOVERY PROCESS!

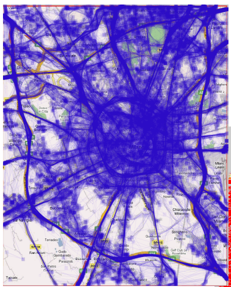
MOBILITY KNOWLEDGE DISCOVERY PROCESS

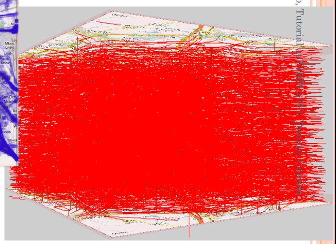


OUTLINE

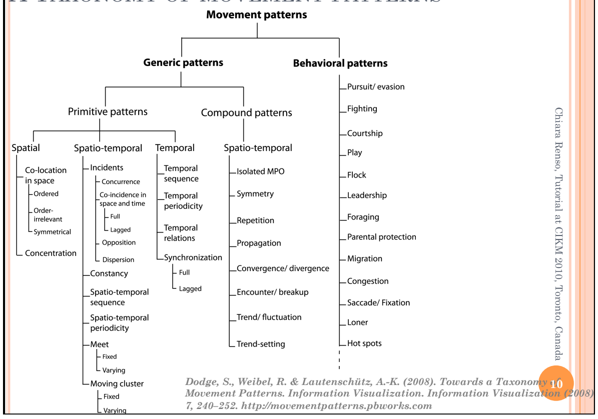
- Mobility Data Mining
- Semantic enrichment of mobility patterns
- Environments for the support of knowledge discovery process of mobility data
- What about Privacy?

What is a mobility pattern?





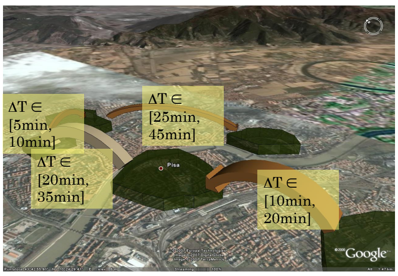
A Taxonomy of movement patterns



Dodge, S., Weibel, R. & Lautenschütz, A.-K. (2008). Towards a Taxonomy of Movement Patterns. Information Visualization. Information Visualization (2008) 7, 240–252. http://movementpatterns.pbworks.com

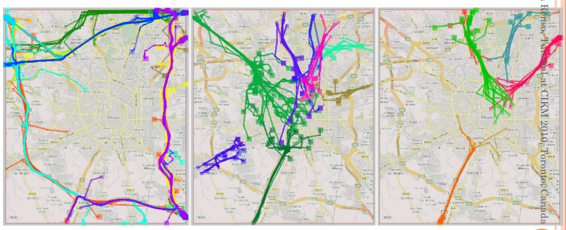
Trajectory patterns…

…Frequent sets/sequences of places visited?



Trajectory patterns…

… Groups of objects moving together?



Mobility Data Mining
Raw trajectories

Data Mining is the automatic extraction of useful, often previously unknown information from large datasets.

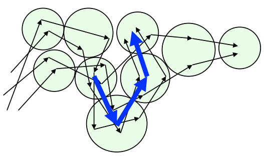
Frequent Sequences in trajectories: T-Pattern

Fosca Giannotti, Mirco Nanni, Fabio Pinelli, Dino Pedreschi: Trajectory pattern mining. KDD 2007: 330-339

Frequent pattern mining

Discover frequent routes, where frequency is measured by a parameter called *support*.

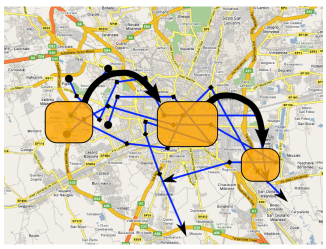
In trajectories the sequence should be preserved



A spatio-temporal sequential pattern

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

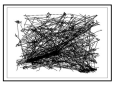
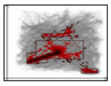
$$(x_0, y_0) \xrightarrow{\alpha_1} (x_1, y_1) \xrightarrow{\alpha_2} \cdots \xrightarrow{\alpha_k} (x_k, y_k)$$



T-Patterns for trajectories

- 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
- What does "matches" mean in space/time?
- Approximation matches
  - a notion of *spatial neighborhood*
  - a notion of *temporal tolerance*

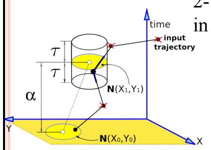
T-Pattern discovery

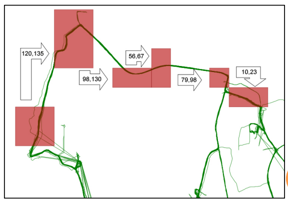



1. Find Regions of Interest

Given as a parameter or computed as dense regions

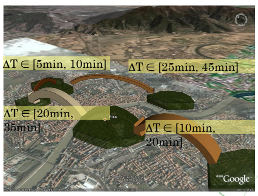
2- Find candidate patterns (similar Trajectory in space and time)

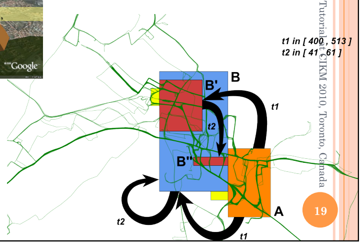




3- Extract frequent patterns with given support

Examples T-patterns





Trajectory Clustering: T-Clustering

Nanni, Pedreschi. Time-focused clustering of trajectories of moving objects. J. of Intelligent Information Systems, 2006

Andrienko, G., Andrienko, N., Rinzivillo, S., Nanni, M., Pedreschi, D.: A Visual Analytics Toolkit for Cluster-Based Classification of Mobility Data. In: SSTD, pp. 432–435 (2009)

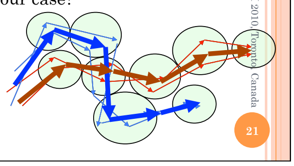
Rinzivillo, Pedreschi, Nanni, Giannotti, Andrienko, Andrienko. Visually-driven analysis of movement data by progressive clustering. J. of Information Visualization, 2008

Trajectory Clustering

Cluster Analysis: Find groups where objects in a group will be similar (or near) to one another and different from (or distant from) the objects in other groups

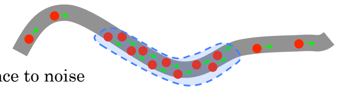
Key Questions:

- Which distance between trajectories?
- Which kind of clustering?
- What is a cluster 'mean' in our case?
  - A representative trajectory?



Trajectory Clustering

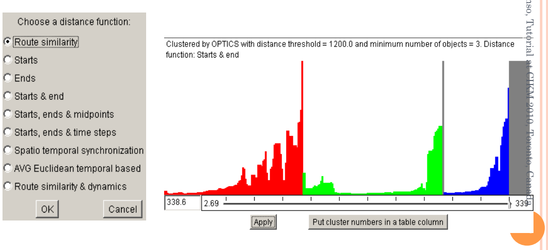
- General requirements:
  - Non-spherical clusters should be allowed
    - E.g.: A traffic jam along a road = "snake-shaped" cluster



  - Tolerance to noise
  - Low computational cost
- A suitable candidate: Density-based clustering
  - OPTICS (Ankerst et al., 1999)
    - ➔ **T**(rajectory)-**OPTICS**

Trajectory Clustering

Euclidean syncronized distance is too complex to compute for huge trajectory dataset:
The Distance function is parametric



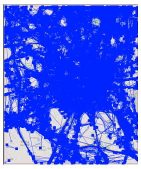
Visually-driven clustering

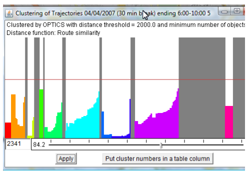
- Progressive refinement through visually-driven exploration
- Progressively complex similarity functions:
  - First, create a large clusters of trajectories using the "common ends" distance function,
  - Concentrate on the (big) cluster of inward trajectories (routes towards the city center)
  - Refine by creating subclusters using a more sophisticated distance function (route similarity)

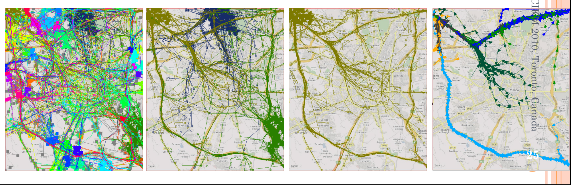
Andrienko, G., Andrienko, N., Rinzivillo, S., Nanni, M., Pedreschi, D.: A Visual Analytics Toolkit for Cluster-Based Classification of Mobility Data. In: SSTD, pp. 432–435 (2009)

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

Analytical effect of progressive clustering







Chiara Renso, Tutorial at CIKM 2010, Toronto, Canada

25

Semantic Enrichment of Mobility Patterns

26

Chiara Renso, Tutorial at CIKM 2010, Toronto, Canada

Incorporating semantics: a step towards the user

1. May the data mining tasks be more accurate if data are semantically enriched?
   - Data Mining on semantic trajectories
2. May we deduce more meaning from data and patterns?
   - Semantic Pattern interpretation
3. May data and patterns be re-combined and queried?
   - Semantich-Rich Environments for Mobility KDD

Chiara Renso, Tutorial at CIKM 2010, Toronto, Canada

27

Data Mining on Semantic Trajectories

28

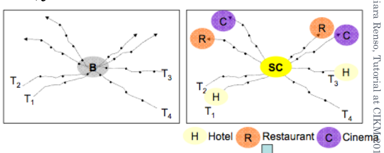
Bogorny, V.; Kuijpers, B.; Alvares, L.O. ST-DMQL: A Semantic Trajectory Data Mining Query Language. In: International Journal of Geographical Information Science. Taylor and Francis . pp.1245 – 1276, vol 23 (10)

Alvares, L.O, Oliveira, G., Heuser, C.A., Bogorny, V. A framework for Trajectory Preprocessing for Data Mining. In: SEKE (21st International Symposium on Software Engineering

Chiara Renso, Tutorial at CIKM 2010, Toronto, Canada

Geometric Patterns Vs Semantic Patterns

- Difficult to interpret from the user's point of view
- Do not discover semantic patterns, which can be independent of x,y coordinates



Semantic trajectory Pattern

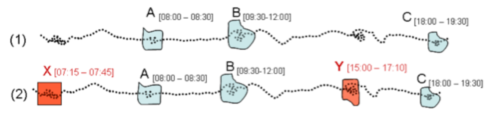
(a) Hotel to Restaurant, passing by SC
(b) (b) go to Cinema, passing by SC

Alvares, L. O.; Bogorny, V.; Kuijpers, B.; Macedo, J. A. F.; Moelans, B.; Vaisman, A.: A Model for Enriching Trajectories with Semantic Geographical Information.. In: Proc. of the ACM 15th International Symposium on Advances in Geographic Information Systems (ACM-GIS'07)

Semantic Trajectory Mining

Work of Bogorny et al. propose methods to compute stop and moves

1) SMoT (intersection-based)



2) CB-SMoT (clustering-based)

SMoT: intersection of the trajectory with *interesting places* (for the application) for a minimum amount of time
CB-SMoT: Cluster single trajectories based on speed: low speed → important place

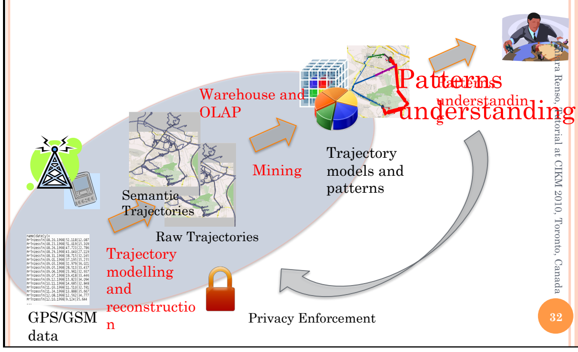
Frequent patterns on Semantic trajectories

Large Sequences of Length 2

(41803_ruas_tuesday,41803_ruas_tuesday)	Support: 9
(41803_ruas_tuesday,66655_ruas_tuesday)	Support: 5
(41803_ruas_monday,66655_ruas_monday)	Support: 5
(41803_ruas_monday,41803_ruas_monday)	Support: 11
(41803_ruas_monday,0_unknown_monday)	Support: 5
(41803_ruas_thursday,41803_ruas_thursday)	Support: 13
(41803_ruas_thursday,0_unknown_thursday)	Support: 6
(41803_ruas_wednesday,41803_ruas_wednesday)	Support: 7

Day of the week
gid
Spatial feature type (stop name)

Trajectory knowledge discovery



Athena: Semantic-rich movement analysis

Which are the home-work trajectories? And the common behaviors of them?

To answer these questions we need to define what is a **home-work trajectory** (or pattern)

The concept of the home-work trajectory can be **encoded** in a formal framework to automatically **infer** which trajectories are home-work

Chiara Renso, Tutorial at CIKM 2010, Toronto, Canada

33

The Athena tool

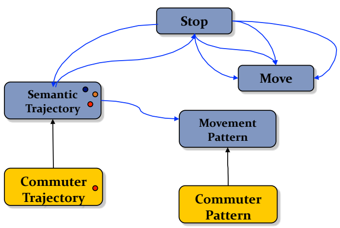
- Supports the post processing phase of the KDD process.
- Based on an ontology to represent domain knowledge and to infer the semantic types of the patterns (or trajectories).
- Classification of movement patters (trajectories) in domain concepts based on the semantic characteristics

Miriam Baglioni, José Antônio Fernandes de Macêdo, Chiara Renso, Roberto Trasarti, Monica Wachowicz: Towards Semantic Interpretation of Movement Behavior. AGILE Conf. 2009

Chiara Renso, Tutorial at CIKM 2010, Toronto, Canada

34

The Reasoning System of Athena



Trajectory data populates a domain ontology and the automatic reasoning engine classifies trajectories into the concept satisfiyng the defintion.

Commuter trajectory = a trajectory frequently starting outside the city, stopping inside the city for a long time and going back outside the city

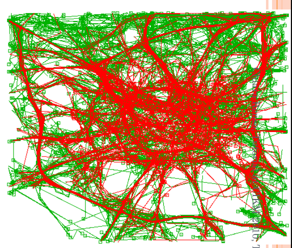
IS A relationship   Object property

Chiara Renso, Tutorial at CIKM 2010, Toronto, Canada

35

Running Example: Identifying the Commuters

We query the system to identify the trajectories whose "semantic" type is commuter, i.e. satisfying the ontology definition

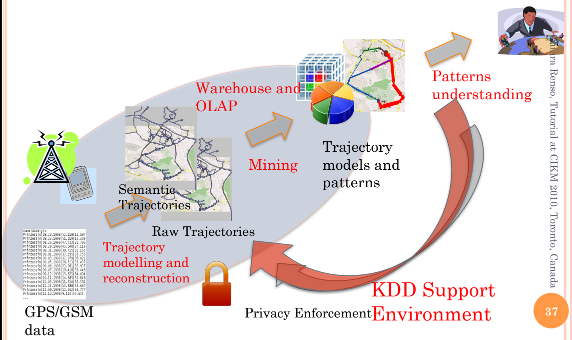


```
SELECT t.id, t.object
  FROM Milano_tr
  WHERE 'Commuter' in
SEMANTIC(t.object)
```

Chiara Renso, Tutorial at CIKM 2010, Toronto, Canada

36

Trajectory knowledge discovery



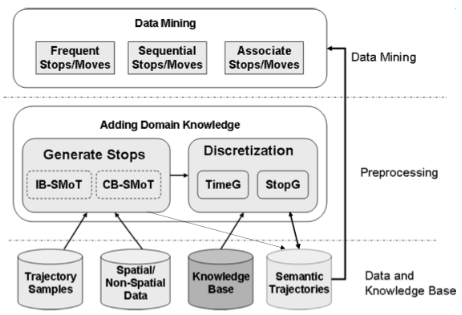
Environments for the support of mobility data KDD process

R. Trasarti et al. The DAEDALUS framework: progressive querying and mining of movement data. ACM-GIS 2008

Roberto Trasarti, Salvatore Rinzivillo, Fabio Pinelli, Mirco Nanni, Anna Monreale, Chiara Renso, Dino Pedreschi, Fosca Giannotti: Exploring Real Mobility Data with M-Atlas. ECML/PKDD 2010

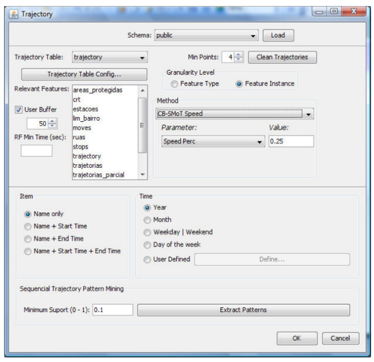
Bogorny, V.; Kuijpers, B.; Alvares, L.O. ST-DMQL: A Semantic Trajectory Data Mining Query Language. In: International Journal of Geographical Information Science. Taylor and Francis . pp.1245 – 1276, vol 23 (10)

Semantic Trajectories Data mining: Weka-STDM



Bogorny, V.; Kuijpers, B.; Alvares, L.O. ST-DMQL: A Semantic Trajectory Data Mining Query Language. In: International Journal of Geographical Information Science. Taylor and Francis. pp.1245 – 1276, vol 23 (10)

Semantic Trajectories Data mining: Weka-STDM – cont.



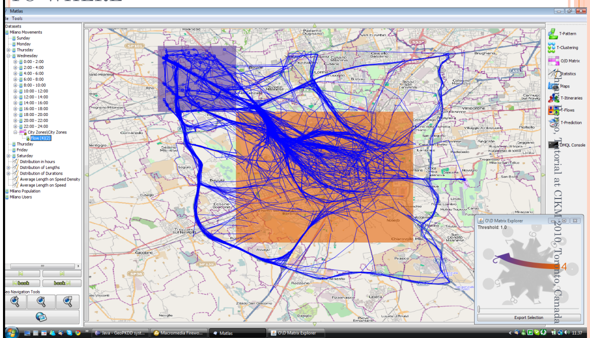
The M-Atlas tool

A knowledge discovery support environment for trajectory data:

- Data, Patterns and background knowledge need to be progressively combined
- T-patterns, T-clusters, etc. are the basic primitive within a DMQL – Data Mining Query Language, supporting the entire KDD process
- T-patterns, T-clusters, etc., once mined from a trajectory dataset, can be stored and later used for query, mining & interpreting in a progressive way

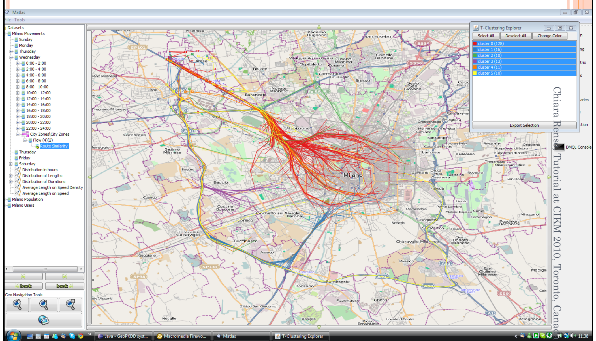
Roberto Trasarti, Salvatore Rinzivillo, Fabio Pinelli, Mirco Nanni, Anna Monreale, Chiara Renso, Dino Pedreschi, Fosca Giannotti: Exploring Real Mobility Data with M-Atlas. ECML/PKDD 2010

Looking for movement patterns: from where to where

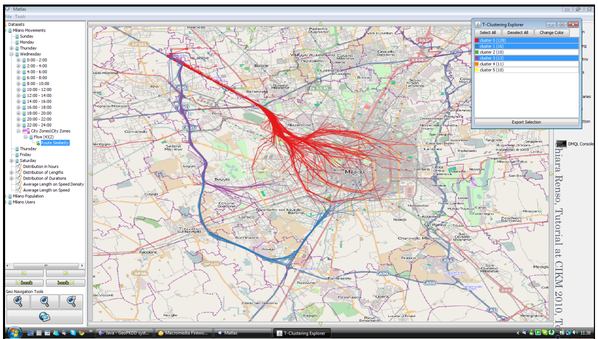


Select trajectories exiting the city from the centre towards North-West

Discovering typical routes

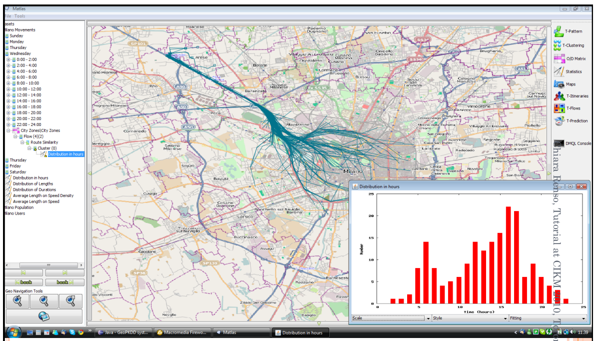


T-clustering divides trips based on route similarity



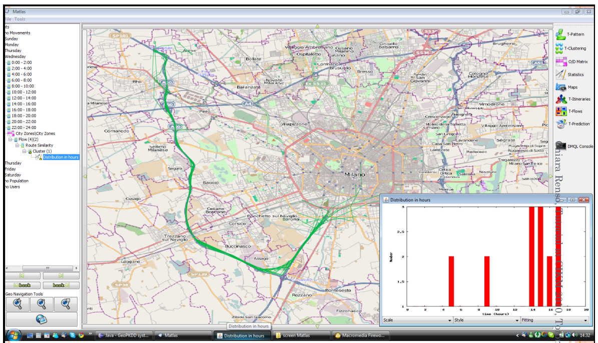
Focus on three larger clusters

One group (red) goes straight to NW, other follow alternative routes



Temporal analysis on each group

A small group of commuters in the morning and a larger one in the afternoon



Temporal analysis on each group

A group of commuters in the afternoon follows an alternative route: are they *smart*?

Privacy Aspects

Today, tracking is an everytime / everywhere process

- the places we visit, we live or work at
- the people we meet

are extremely sensitive:

Therefore, privacy-preservation is a must!

What is required:

Privacy-aware KDD process

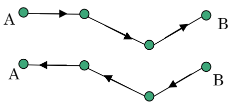
De-identify is not enough..!

"De-identified mobility data are enough to reconstruct aggregate movement behaviour, pertaining to groups of people".

Is de-indentified data really anonymous?

Unfortunately not! 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…

Chiara Renso, Tutorial at CIKM 2010, Toronto, Canada

49

49

A Possibile Solution: Privacy by Design

- Hiding personal identifiers may not be sufficient
- Need for new privacy-preserving DM techniques
  - **Privacy by Design**
- Natural trade-off between privacy quantification and data utility
  - Analysis results should not be altered significantly
  - Privacy has to be maximized

Chiara Renso, Tutorial at CIKM 2010, Toronto, Canada

50

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

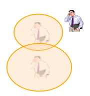
Chiara Renso, Tutorial at CIKM 2010, Toronto, Canada

51

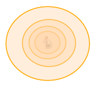
51

Concepts for Location Privacy - Camouflage

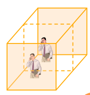
- Location Perturbation
  - The user location is represented with a fake value



- Spatial Cloaking – Generalization
  - The user exact location is represented as a region that includes the exact user location



- Spatio-temporal generalization
  - Generalize also the temporal dimension



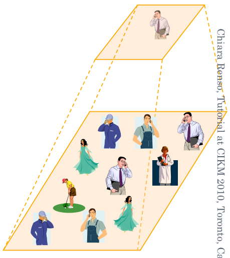
Chiara Renso, Tutorial at CIKM 2010, Toronto, Canada

52

52

Concepts for Location Privacy
Hiding in the crowd

- k-anonymity
  - User's position is generalized to a region containing at least k users
  - The user is indistinguishable among other k-1 users

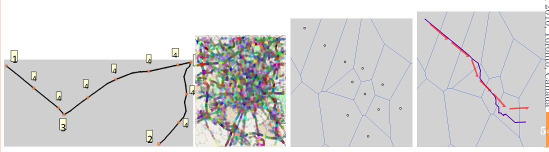


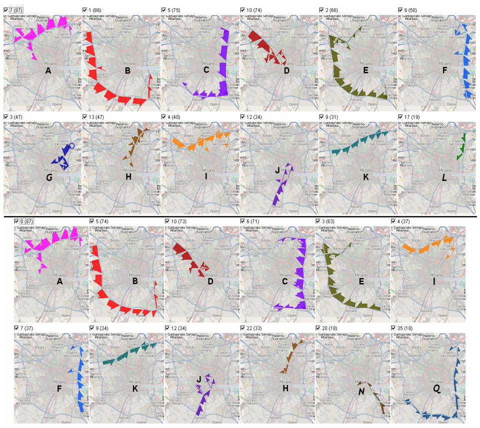
10-anonymity

Example of spatial generalization

Anonymization through generalization

- Extract characteristic points from the trajectories
- Group the extracted points in space
- Partition the territory into Voronoi cells around the centroids of the groups
- Divide the trajectories into segments that link Voronoi cells





Original Patterns

Anonymized Patterns

Acknowledgments

Part of the material has been inspired/extracted from:

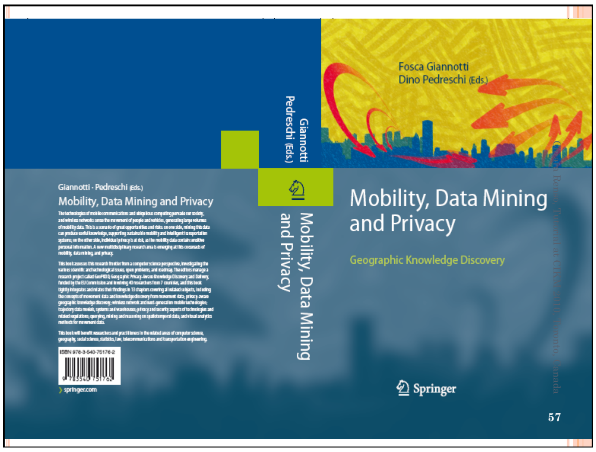
Fosca Giannotti, Dino Pedreschi, Yannis Theodoridis: Geographic Privacy-aware Knowledge Discovery and Delivery – Tutorial ad EDBT 2009

Dino Pedreschi & Fosca Giannotti: Mobility Data Miningat MODAP Summer School "I know where I'll be next summer", Rodhes, 2010

Vania Bogorny and Shashi Shekhar: Spatial and Spatio-temporal Data Mining, Four hours tutorial do be presented at ICDM 2010 (Sydney, Australia).

Special thanks to KDDLab group (http://www-kdd.isti.cnr.it) for the help!

Work partially supported by EU FET CA MODAP: http://www.modap.org



Chiara Renso, Tutorial at CIKM 2010, Toronto, Canada

57

Selected literature on …

- Mobility Pattern Querying & Mining

- Cao, H. et al. (2005) Mining frequent spatio-temporal sequential patterns. Proceedings of ICDM.
- Djafri, N. et al. (2002) Spatio-temporal evolution: querying patterns of change in databases. Proceedings of ACM-GIS.
- Giannotti, F. et al. (2006) Efficient Mining of Temporally Annotated Sequences. Proceedings of SDM.
- Giannotti, F. et al. (2007) Trajectory Pattern Mining. Proceedings of KDD.
- Hadjieleftheriou, M. et al. (2005) Complex Spatio-Temporal Pattern Queries. Proceedings of VLDB.
- Horvitz, E. et al. (2005) Prediction, expectation, and surprise: Methods, designs, and study of a deployed traffic forecasting service. Proceedings of Conference on Uncertainty in Artificial Intelligence.
- Kalnis, P. et al. (2005) On discovering moving clusters in spatio-temporal data. Proceedings of SSTD.
- van Kreveld, M. et al. (2007) Efficient Detection of Motion Patterns in Spatio-Temporal Data Sets. GeoInformatica, 11(2):195-215.
- Laube, P. et al. (2005) Discovering relative motion patterns in groups of moving point objects. Int. Journal of Geographical Information Science, 19(6): 639-668.
- Li, X. et al. (2007) Traffic density-based discovery of hot routes in road networks. Proceedings of SSTD.
- Liu, Y. et al. (2006) A scalable distributed stream mining system for highway traffic data. Proceedings of PKDD.
- Mamoulis, N. et al. (2004) Mining, indexing, and querying historical spatiotemporal data. Proceedings of KDD.
- du Mouza, C. and Rigaux, P. (2005) Mobility Patterns. GeoInformatica, 9(4): 297-319.

Chiara Renso, Tutorial at CIKM 2010, Toronto, Canada

58

Selected literature on …

- Mobility Pattern Querying & Mining (cont.)

- Nakata, T. and Takeuchi, J. (2004) Mining traffic data from probe-car system for travel time prediction. Proceedings of KDD.
- Qu, Y. et al. (2003) Supporting Movement Pattern Queries in User-Specified Scales. IEEE Transactions on Knowledge and Data Engineering, 15(1): 26-42.
- Shekhar, S. et al. (2001) Data mining and visualization of twin-cities traffic data. Technical Report, TR-01-015, University of Minnesota.
- Dodge, S., Weibel, R. & Lautenschütz, A.-K. (2008). Towards a Taxonomy of Movement Patterns. Information Visualization. Information Visualization (2008) 7, 240–252.

Chiara Renso, Tutorial at CIKM 2010, Toronto, Canada

59

Selected literature on …

- Mobility Data Clustering

- Mihael Ankerst et al. 1999: OPTICS: Ordering Points To Identify the Clustering Structure. SIGMOD Conference 1999: 49-60
- Gadez, I.V. et al. (2000) A General Probabilistic Framework for Clustering Individuals and Objects. Proceedings of KDD.
- Gaffney, S. and Smyth, P. (1999) Trajectory Clustering with Mixtures of Regression Models, Proceedings of KDD.
- Gaffney, S. et al. (2006) Probabilistic Clustering of Extratropical Cyclones Using Regression Mixture Models, Tech. Rep. UCI-ICS 06-02.
- Kalnis, P. et al. (2005) On Discovering Moving Clusters in Spatio-temporal Data. Proceedings of SSTD.
- Lee, J.-G. et al. (2007) Trajectory Clustering: A Partition-and-Group Framework, Proceedings of ACM SIGMOD.
- Nanni, M. and Pedreschi, D. (2006) Time-focused clustering of trajectories of moving objects. J. of Intelligent Information Systems, 27(3): 267-289.
- Rinzivillo, S. et al. (2008) Visually-driven analysis of movement data by progressive clustering. J. of Information Visualization, 7(3/4) 225-239.

Chiara Renso, Tutorial at CIKM 2010, Toronto, Canada

60

Selected literature on …

- Mobility KDD Environments

  - R. Trasarti et al.: The DAEDALUS framework: progressive querying and mining of movement data. ACM-GIS 2008
  - Roberto Trasarti, Salvatore Rinzivillo, Fabio Pinelli, Mirco Nanni, Anna Monreale, Chiara Renso, Dino Pedreschi, Fosca Giannotti: Exploring Real Mobility Data with M-Atlas. ECML/PKDD 2010
  - Mirco Nanni, Roberto Trasarti, Chiara Renso, Fosca Giannotti, Dino Pedreschi: Advanced knowledge discovery on movement data with the GeoPKDD system. Demo at EDBT 2010, ECML-PKDD 2010
  - Fosca Giannotti, Mirco Nanni, Dino Pedreschi, Fabio Pinelli, Chiara Renso, Salvatore Rinzivillo, Roberto Trasarti: M-Atlas: creating the atlas of urban mobility by mining massive GPS trajectory data. NetSci 2010
  - Roberto Trasarti, Miriam Baglioni, Chiara Renso: DAMSEL: A System for Progressive Querying and Reasoning on Movement Data. DEXA Workshops 2009: 452-4
  - Miriam Baglioni, José Antônio Fernandes de Macêdo, Chiara Renso, Roberto Trasarti, Monica Wachowicz: Towards Semantic Interpretation of Movement Behavior. AGILE Conf. 2009
  - Bogorny, V.; Kuijpers, B.; Alvares, L.O. ST-DMQL: A Semantic Trajectory Data Mining Query Language. In: International Journal of Geographical Information Science. Taylor and Francis . pp.1245 – 1276, vol 23 (10)

Chiara Renso, Tutorial at CIKM 2010, Toronto, Canada

61

Selected literature on …

- Mobility Data Privacy & Anonymity

  - Abul, O. et al. (2007a) Hiding Sensitive Trajectory Patterns. Proceedings of ICDM Workshops.
  - Abul, O. et al. (2007b) Hiding Sequences. Proceedings of ICDM Workshops.
  - Abul, O. et al. (2008) Never Walk Alone: Uncertainty for Anonymity in Moving Objects Databases. Proceedings of ICDE.
  - Gedik, B. and Liu, L. (2005) Location Privacy in Mobile Systems: A Personalized Anonymization Model. Proceedings of ICDCS.
  - Ghinita, G. et al. (2008) Private Queries in Location Based Services: Anonymizers are not Necessary, Proceedings of ACM SIGMOD.
  - Gkoulalas-Divanis, A., and Verykios, V.S. (2008a) A Privacy-Aware Trajectory Tracking Query Engine. SIGKDD Explorations, 10(1): 40-49.
  - Gkoulalas-Divanis, A., and Verykios, V.S. (2008b) A Free Terrain Model for Trajectory K-Anonymity. Proceedings of DEXA.
  - Gkoulalas-Divanis, A. et al. (2009a). A Network Aware Privacy Model for Online Requests in Trajectory Data. Data and Knowledge Engineering, to appear.
  - Gkoulalas-Divanis, A. et al. (2009b). PLOT: Privacy in Location-Based Services: an Open-Ended Toolbox. Proceedings of MDM.
  - Gkoulalas-Divanis, A. et al. (2009c). Identifying Unsafe Routes for Network-Based Trajectory Privacy. Proceedings of SDM.

Chiara Renso, Tutorial at CIKM 2010, Toronto, Canada

62

Selected literature on …

- Mobility Data Privacy & Anonymity (cont.)

  - Gruteser, M. and Grunwald, D. (2003) Anonymous Usage of Location-based Services through Spatial and Temporal Cloaking. Proceedings of MobiSys.
  - Gruteser, M. and Liu. X. (2004) Protecting Privacy in Continuous Location-tracking Applications. IEEE Security and Privacy, 2(2): 28-34.
  - Kalnis, P. et al. (2006) Preserving Anonymity in Location Based Services. Technical Report TRB6/06, Department of Computer Science, National University of Singapore.
  - Mokbel, M.F. et al. (2007) The New Casper: A Privacy-Aware Location-Based Database Server. Proceedings of ICDE.
  - Terrovitis, M. and Mamoulis, N. (2008) Privacy Preservation in the Publication of Trajectories. Proceedings of MDM.
  - Zacharouli, P. et al. (2007) A K-Anonymity Model for Spatiotemporal Data. Proceedings of STDM.

Chiara Renso, Tutorial at CIKM 2010, Toronto, Canada

63