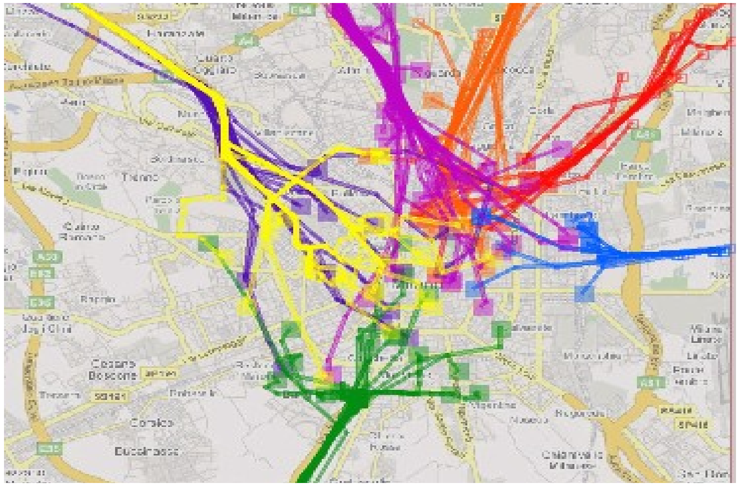
Mobility Data Mining

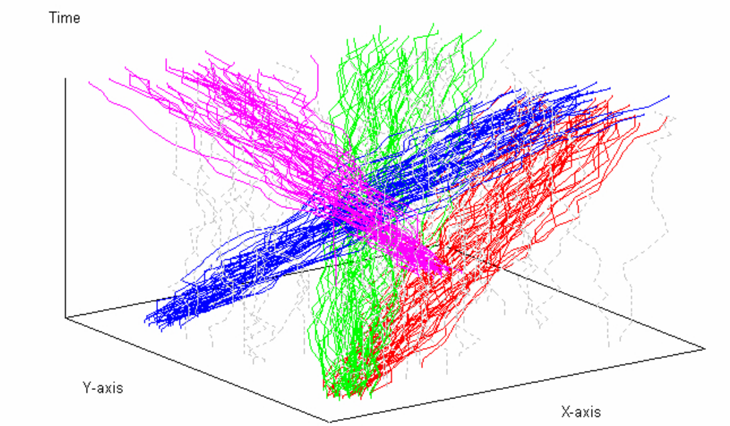
Mobility data Analysis Foundations

Trajectory Clustering

T-clustering

- Trajectories are grouped based on similarity
- Several possible notions of similarity
  - Start/End points
  - Shape of trajectory
  - Shape & time
  - Etc.



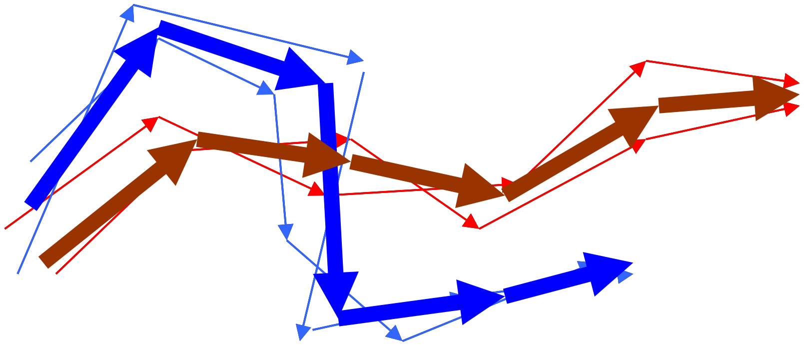


Nanni, Pedreschi. **Time-focused clustering of trajectories of moving objects**. J. of Intelligent Information Systems, 2006.

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

Trajectory Clustering

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



Which distance?

- Average Euclidean distance (Spatio-temporal distance)

$$D(\tau_1,\tau_2)|_T = \frac{\int_T \boxed{d(\tau_1(t),\tau_2(t))}\,dt}{|T|}$$

distance between moving objects $\tau 1$ and $\tau 2$ at time t

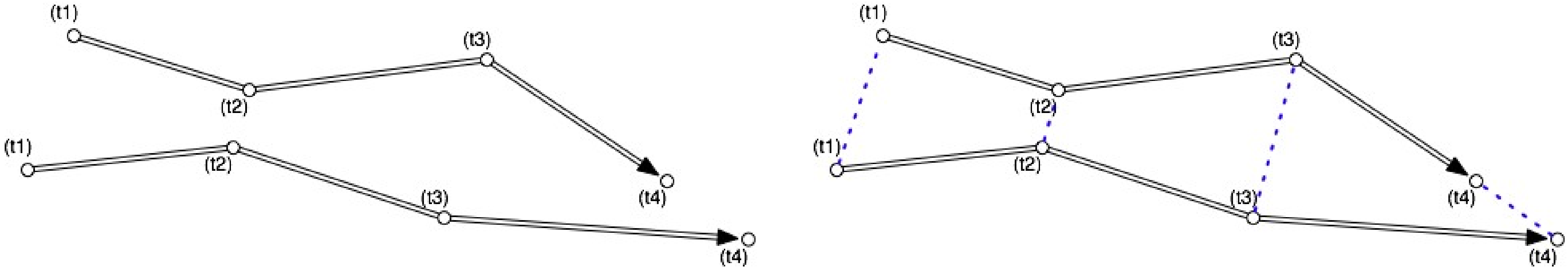
- “Synchronized” behaviour distance
  - Similar objects = almost always in the same place at the same time
- Computed on the whole trajectory

Average Euclidean Distance Sincronized

- Align point temporally

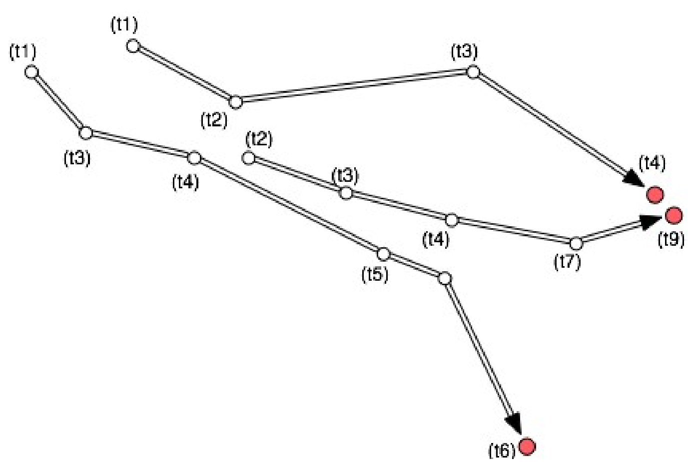
$$D(\tau_1,\tau_2)|_T = \frac{\int_T d(\tau_1(t),\tau_2(t))dt}{|T|}$$

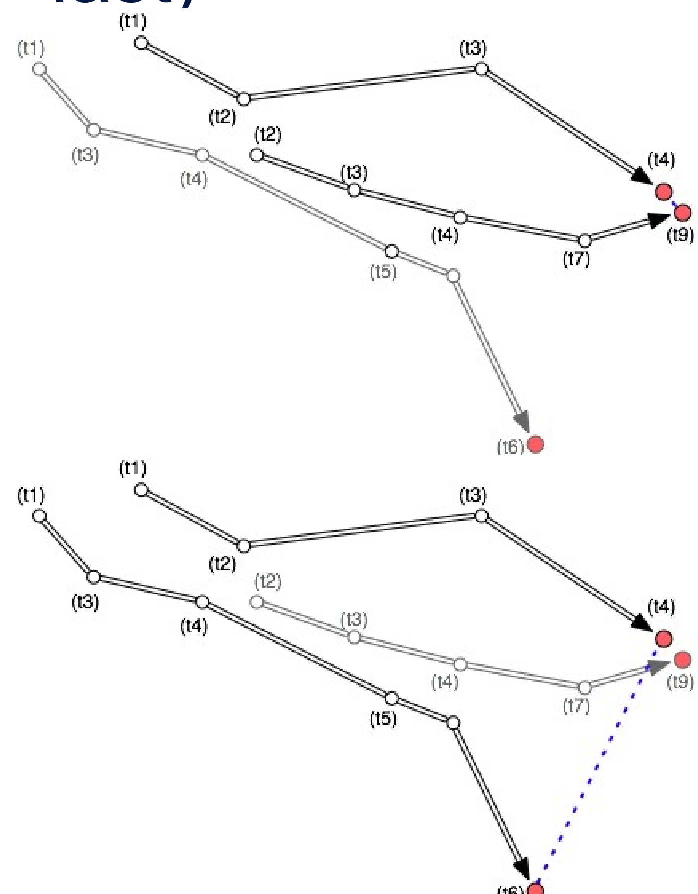
- Eventually assign penalties to non matching points



Common Destination

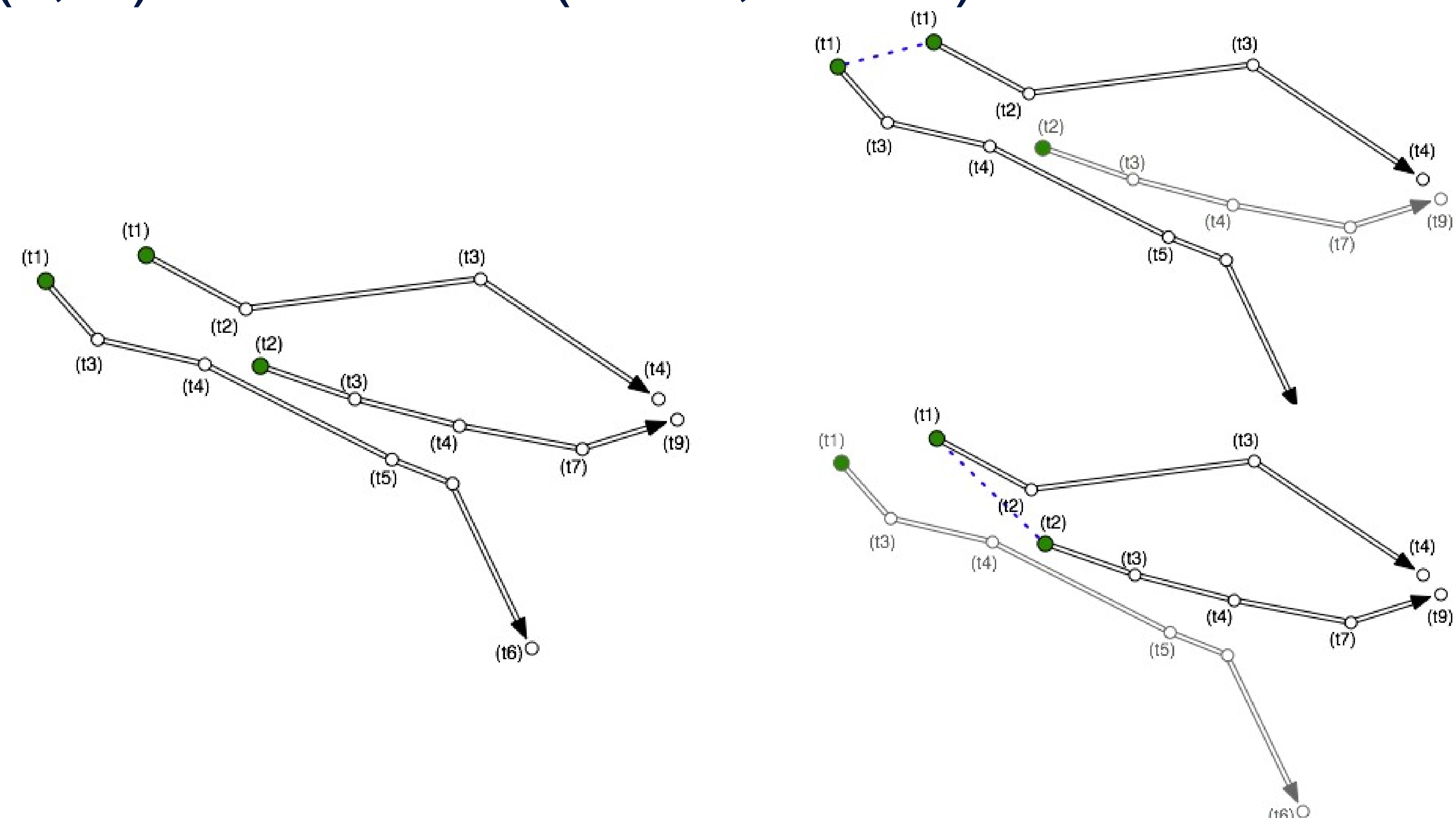
- Select last point *Plast* for each trajectory
- D(T,T’) = Euclidean(Plast, P’last)





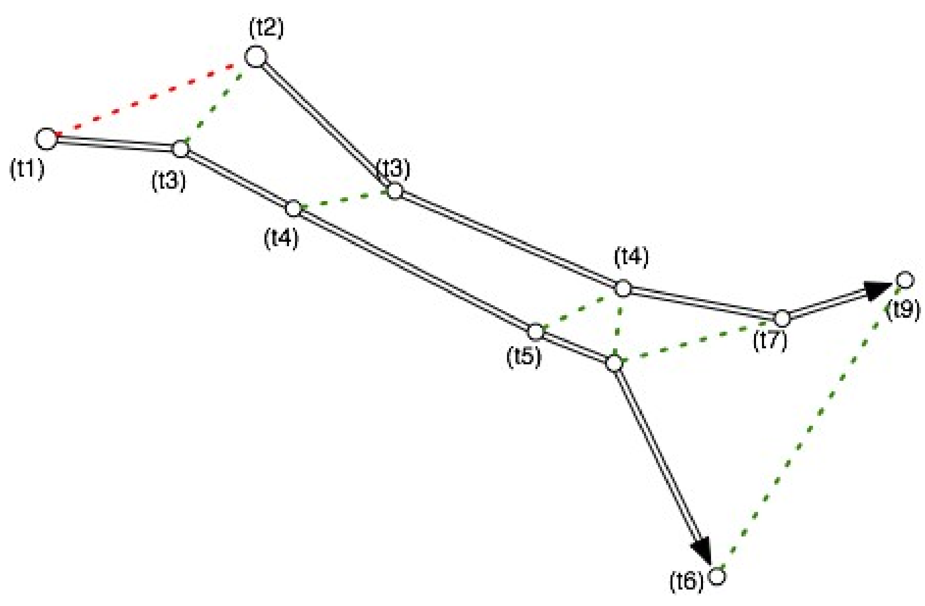
Common Origins

- Select first point *Pfirst* for each trajectory
- D(T,T') = Euclidean(Pfirst, P'first)



Route Similarity

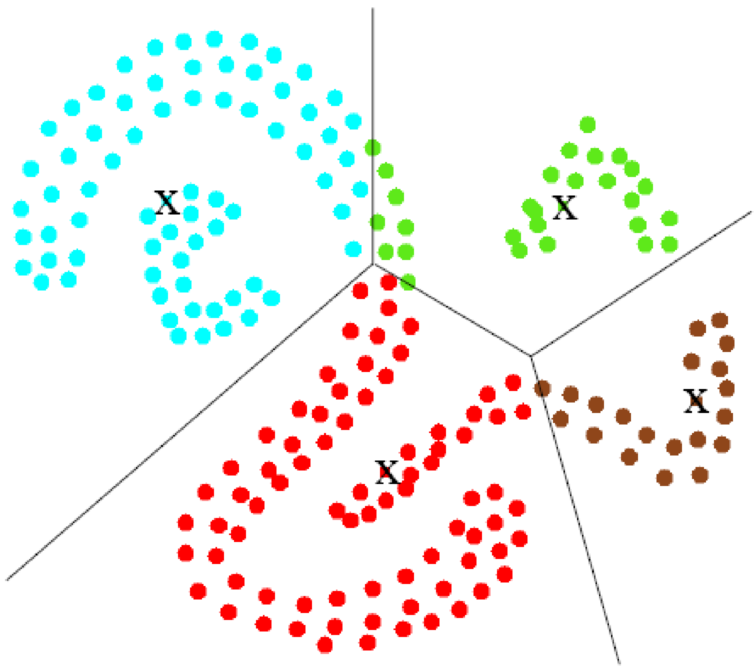
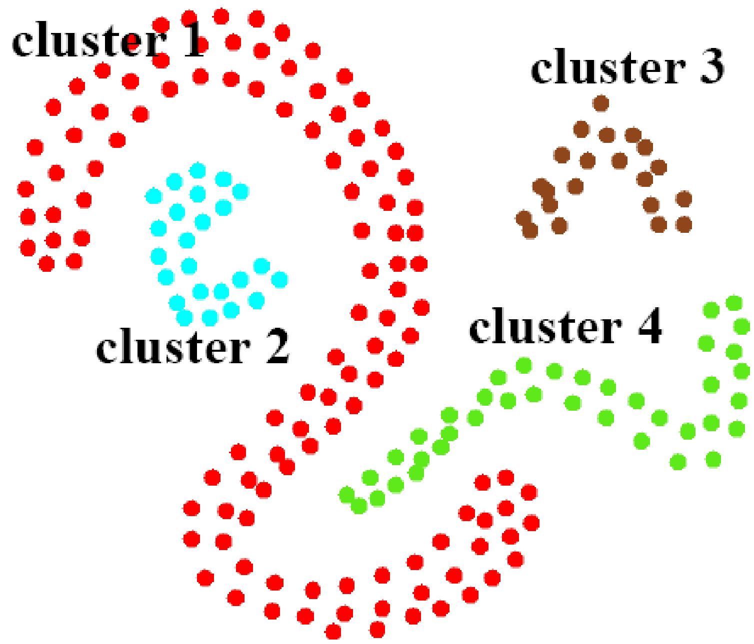
- Alignment of points, multiple matches
- Average Euclidean Distance
- Penalties for non matching initial points (no penalties for destinations)



Which kind of clustering?

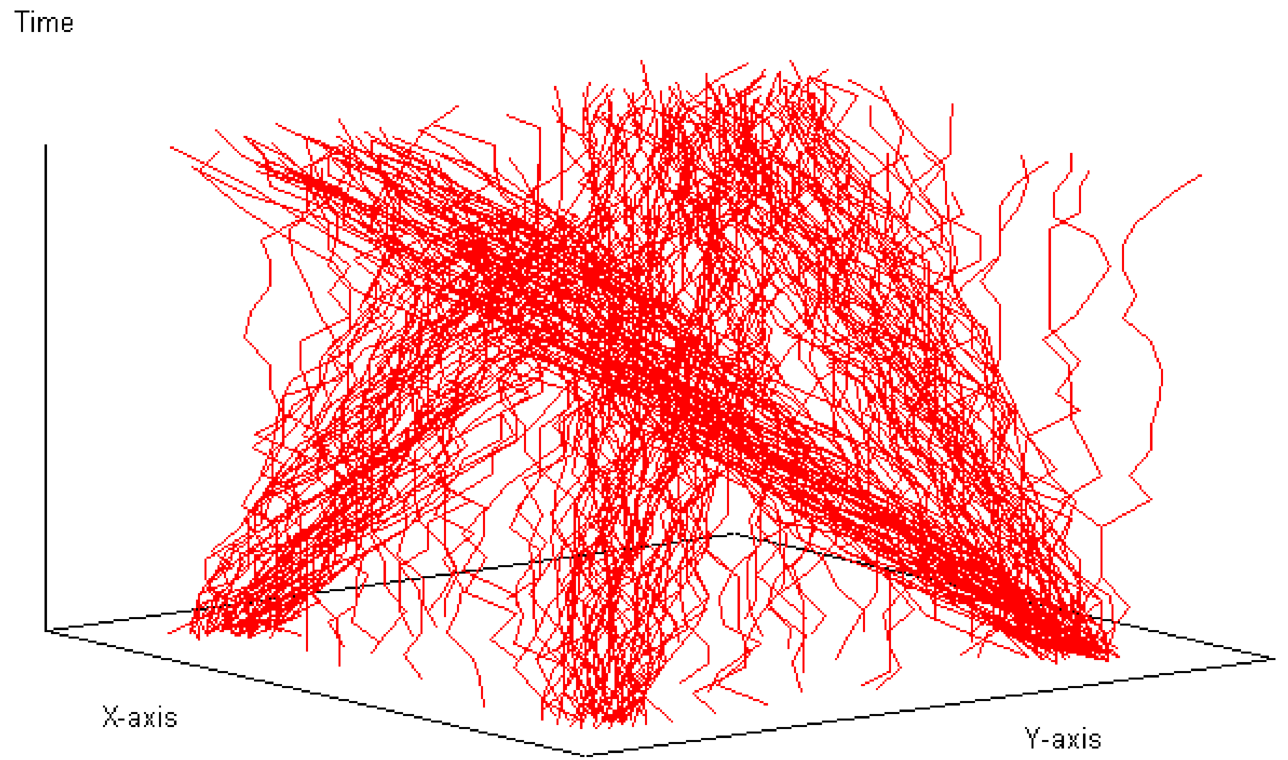
- General requirements:
  - Non-spherical clusters should be allowed
    - E.g.: A traffic jam along a road = “snake-shaped” cluster
  - Tolerance to noise
  - Low computational cost
  - Applicability to complex, possibly non-vectorial data
- A suitable candidate: Density-based clustering
  - OPTICS (Ankerst et al., 1999) → **T**(rajectory)**-OPTICS**
  - Evolution of basic DBSCAN

Density Based Clustering

K-means	*Density-based*
	

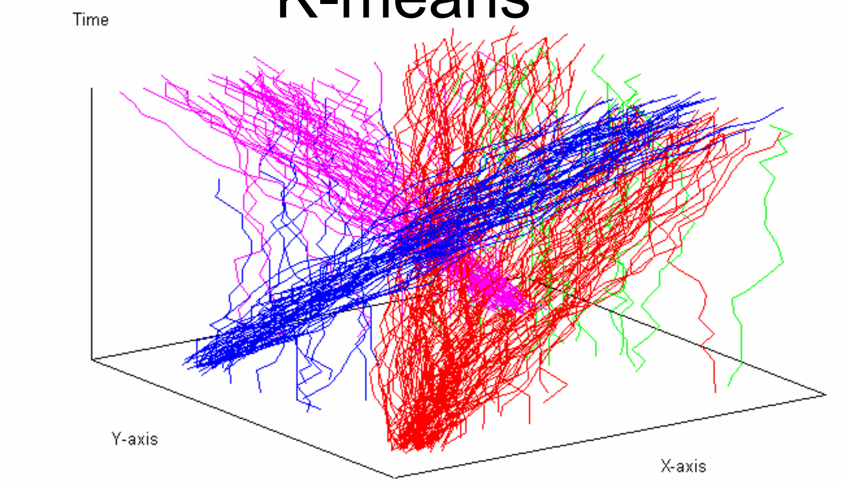
A sample dataset

- A set of trajectories forming 4 clusters + noise (synthetic)

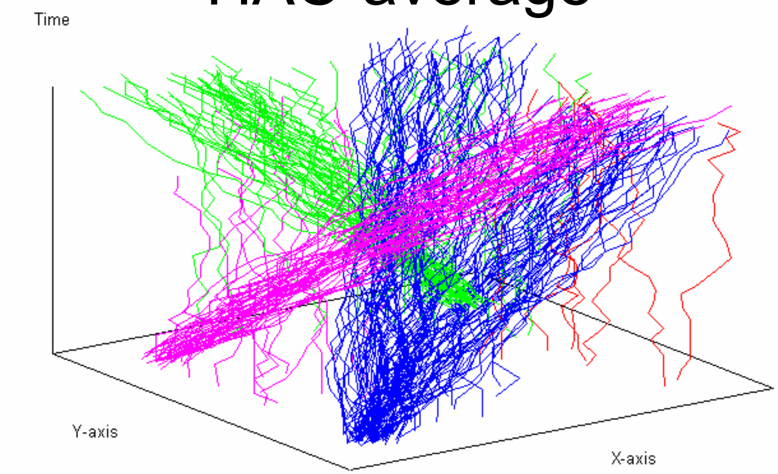


T-OPTICS vs. HAC & K-means

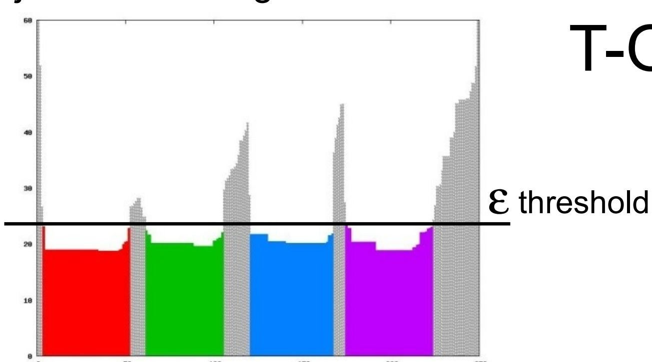
K-means



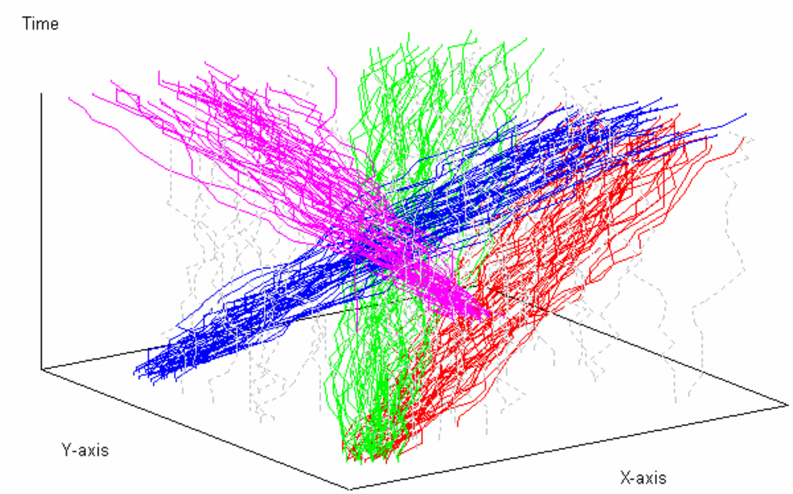
HAC-average



Reachability plot
(= objects reordering for distance distribution)



T-OPTICS



Ad-hoc distance functions

- Colocation –
  - Link prediction,
  - Semantic behaviors,
  - GSM data
- Spatio-temporal Colocation –
  - Link prediction,
  - Semantic behaviors,
  - GSM data
- *Start and End inclusion*
  - Car Pooling Matching
- *Align to end –*
  - Incoming flows
- *Align to start –*
  - Outcoming flows

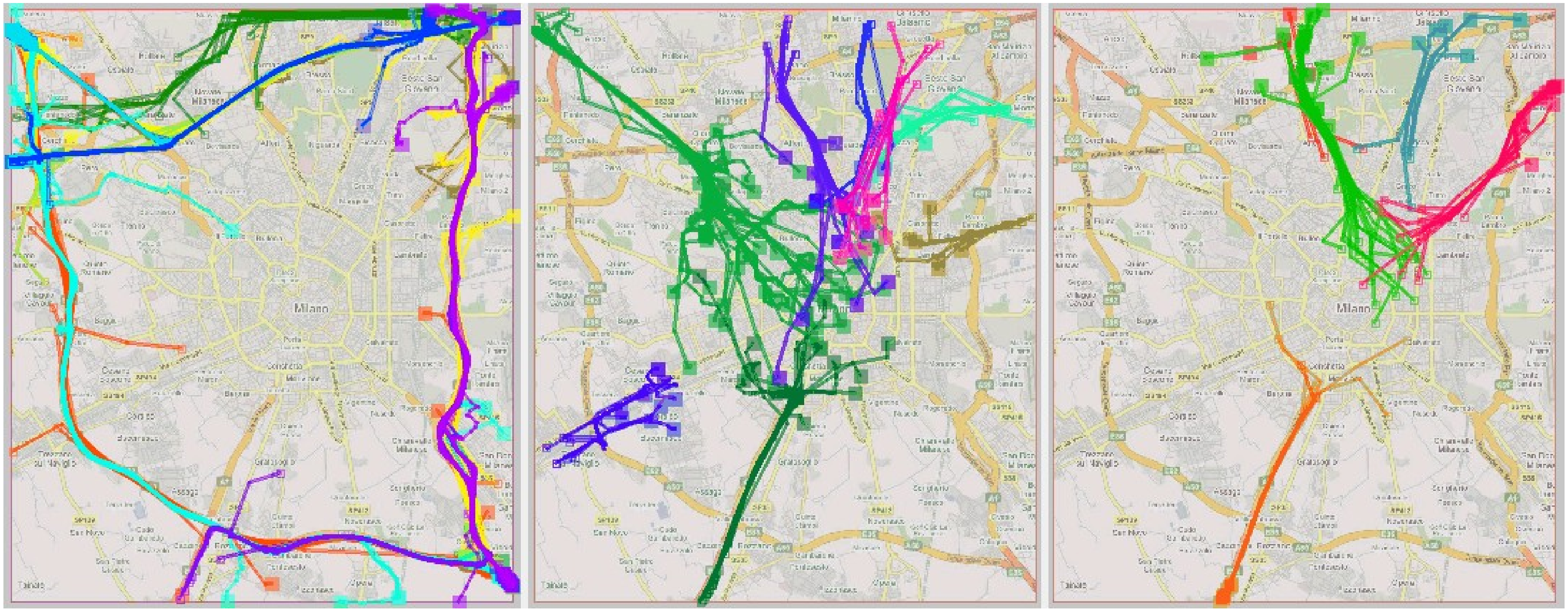
Progressive clustering

- *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)*

Clustering Data
(Common Destination)

Select a Cluster

Clustering Data
(route similarity)

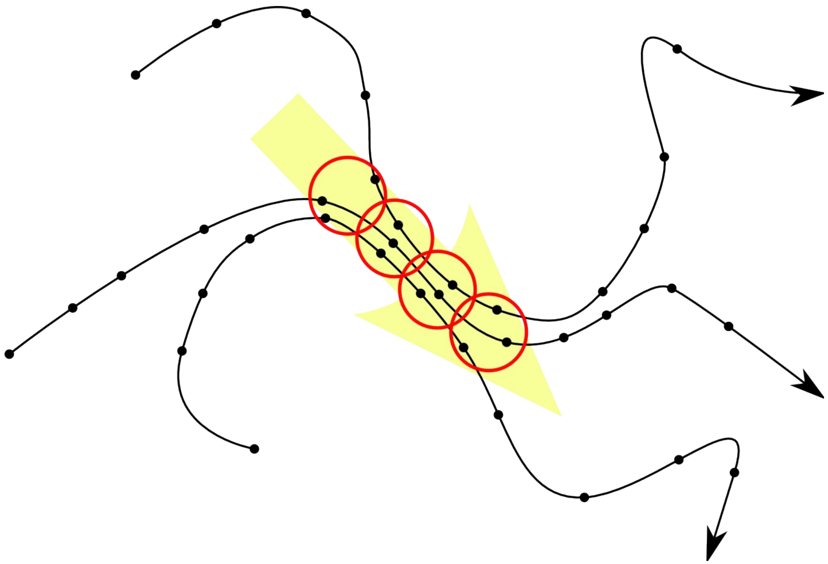


Left: peripheral routes; middle: inward routes; right: outward routes.

Trajectory patterns

Are there groups of objects that move together for some time or in a similar way?

Moving Trajectory Flocks



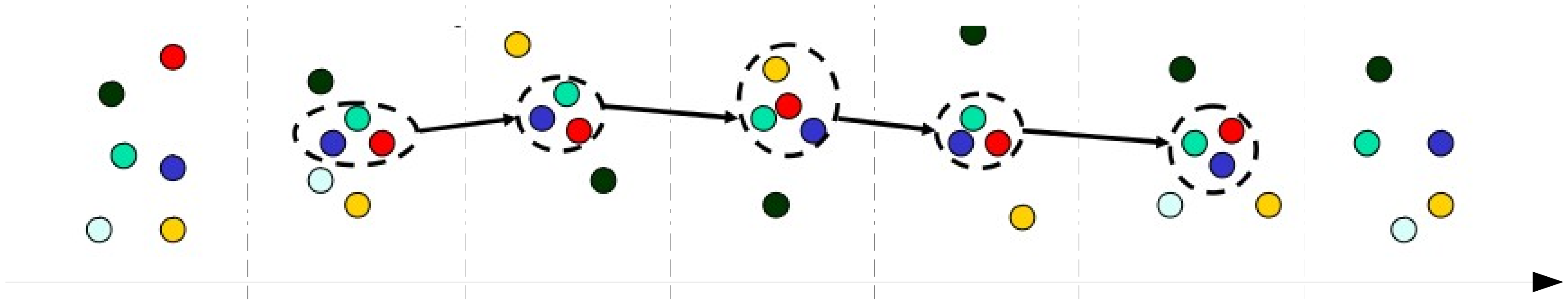
- Group of objects that move together (close to each other) for a time interval

- Discover all possible:
  - sets of objects O, with |O| > min_size and
  - time intervals T, with |T| > min_duration
- such that for all timestamps $t \in T$ the points in O|t are contained in a circle of radius *r*

M. Wachowicz, R. Ong, C. Renso, M. Nanni: ***Finding moving flock patterns among pedestrians through collective coherence.*** *IJGIS 25(11): 1849-1864 (2011)*

Moving Clusters

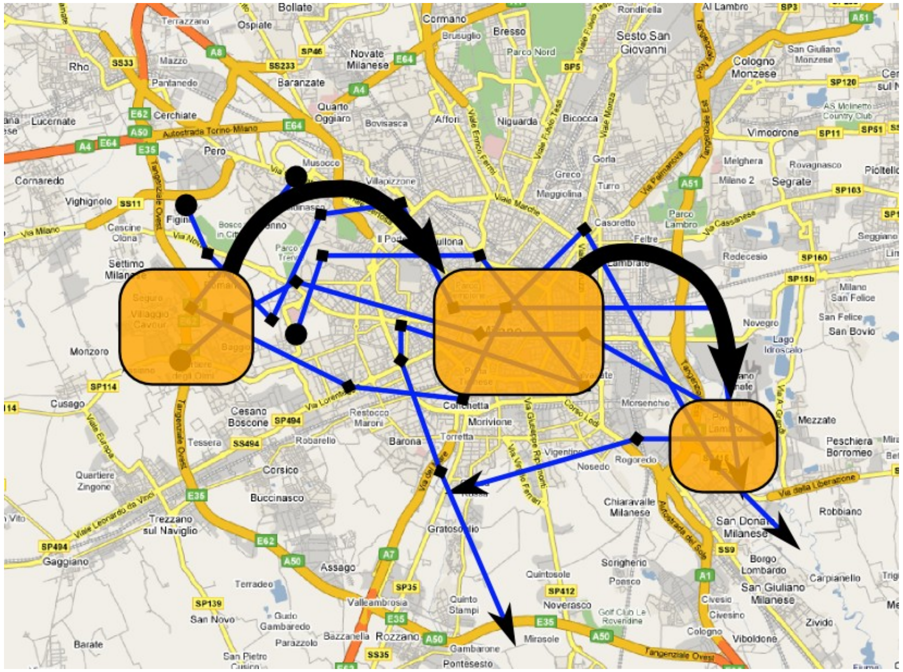
- A ***moving cluster*** is a set of objects that move close to each other for a long time interval



- Formal Definition [Kalnis et al., SSTD'05]:
  - A ***moving cluster*** is a sequence of (snapshot) clusters ***c1, c2, …, ck*** such that for each timestamp i ($1 \le i < k$), $|c_i \cap c_{i+1}| / |c_i \cup c_{i+1}| \ge \theta$ $(0 < \theta \le 1)$

T-Patterns

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

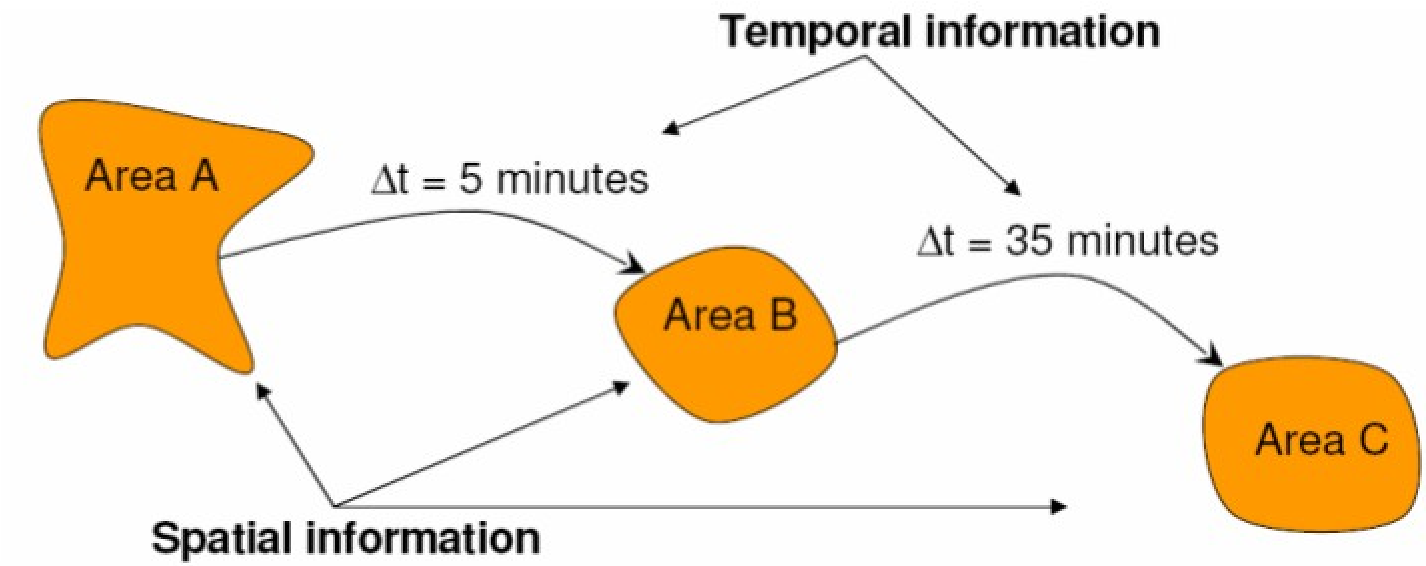


Giannotti, Nanni, Pedreschi, Pinelli.
Trajectory pattern mining. Proc. ACM SIGKDD 2007

T-Patterns

$$A_0 \xrightarrow{t_1} A_1 \xrightarrow{t_2} \ldots \quad A_{n-1} \xrightarrow{t_n} A_n$$

- $\boldsymbol{t_i}$ = transition time, $\boldsymbol{A_i}$ = spatial region



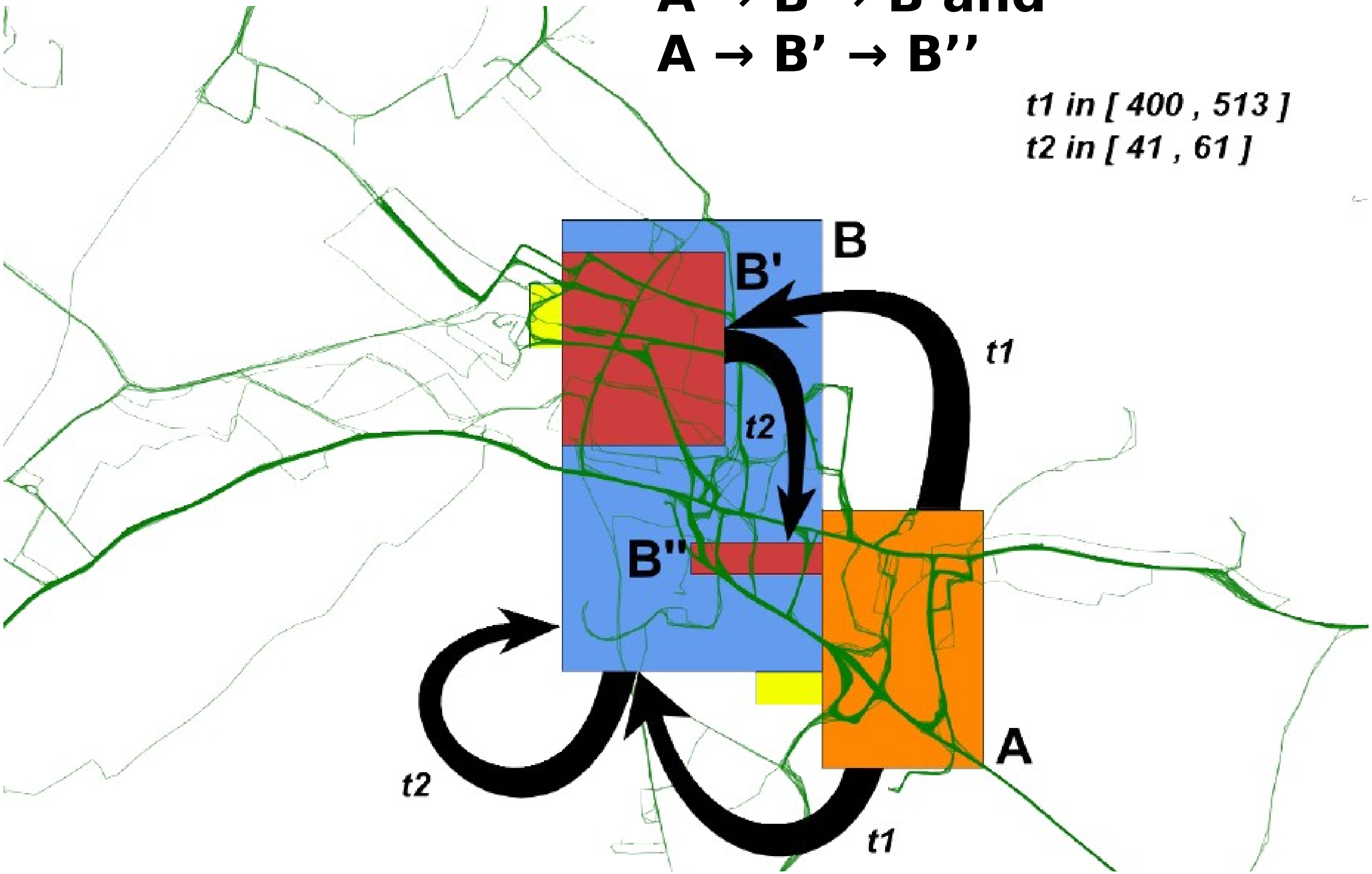
$$Station \xrightarrow{20\text{min.}} Castle \xrightarrow{65\text{min.}} Museum$$

Sample Trajectory Pattern

Data Source: Trucks in Athens – 273 trajectories)

A → B → B and
A → B’ → B’’

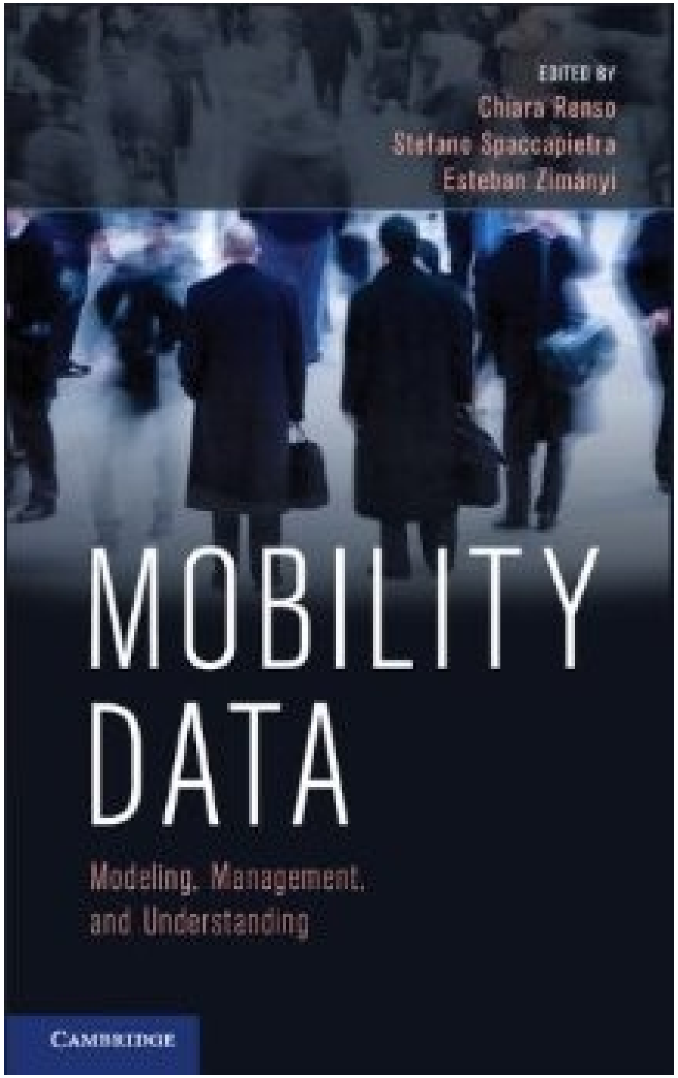
t1 in [400 , 513]
t2 in [41 , 61]



Short bibliography

- G. Andrienko, N. Andrienko, S. Rinzivillo, M. Nanni, D. Pedreschi, and F. Giannotti. Interactive visual clustering of large collections of trajectories. In Visual Analytics Science and Technology, 2009. VAST2009. IEEE Symposium on , pages 3{10, 2009.
- F. Giannotti, M. Nanni, D. Pedreschi, F. Pinelli, C. Renso, S. Rinzivillo, and R. Trasarti. Unveiling the complexity of human mobility by querying and mining massive trajectory data. VLDB J. , 20(5):695{719, 2011.
- Fosca Giannotti, Mirco Nanni, Fabio Pinelli, and Dino Pedreschi. Trajectory pattern mining. In KDD , 2007.
- M.Nanni, R.Trasarti, G.Rossetti, and D.Pedreschi. Ecient distributed computation of human mobility aggregates through user mobility proles. In UrbComp12 , 2012.
- Roberto Trasarti, Fabio Pinelli, Mirco Nanni, and Fosca Giannotti. Mining mobility user proles for car pooling. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining , pages 1190{1198. ACM, 2011.
- Monreale, A., Pinelli, F., Trasarti, R., & Giannotti, F. (2009). WhereNext: a location predictor on trajectory pattern mining. In KDD '09 (p. 637-646).
- Mobility Data. Cambridge Press:
- Mobility Data Mining and Privacy

Mobility data: Modeling, Managing and understanding, Cambridge press.



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