Mobility Data Management & Exploration

Ch. 09. Semantic Aspects on Mobility Data

Nikos Pelekis & Yannis Theodoridis

InfoLab | University of Piraeus | Greece
infolab.cs.unipi.gr

v.2014.05

“The map in not the territory.” Alfred Korzybski

Chapter outline

9.1. From Raw to Semantic Trajectories

9.2. The Semantic Enrichment Process of Raw Trajectories

9.3. Semantic Trajectory Data Management

9.4. Semantic Trajectory Data Exploration

9.5. Semantic Aspects of Privacy

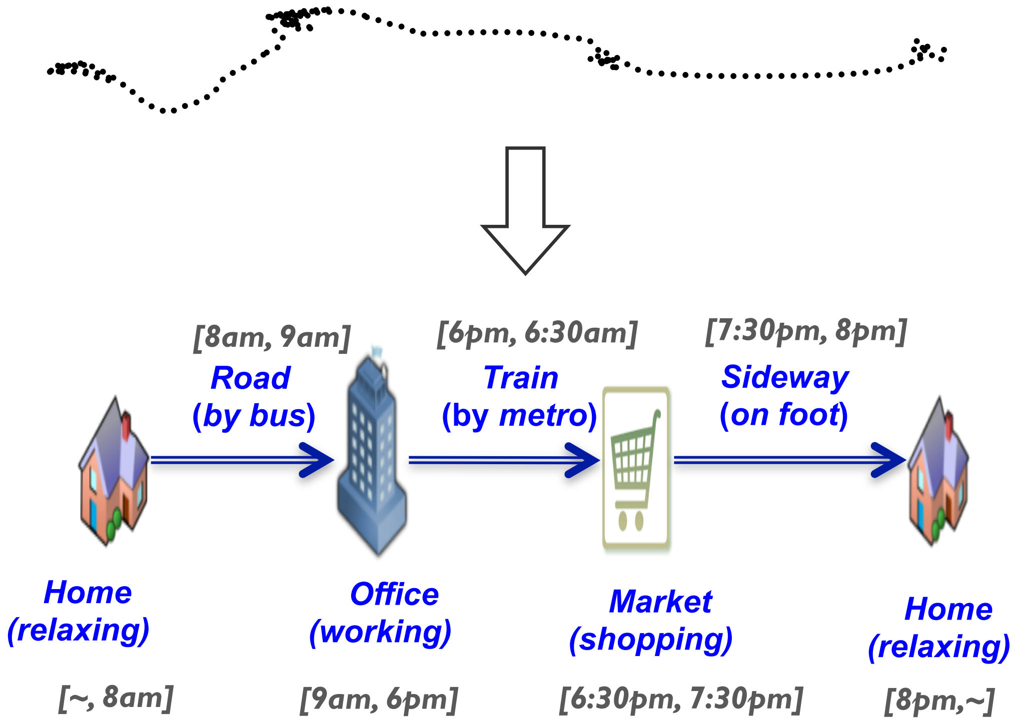
9.6. Summary

9.1.

From raw to semantic trajectories

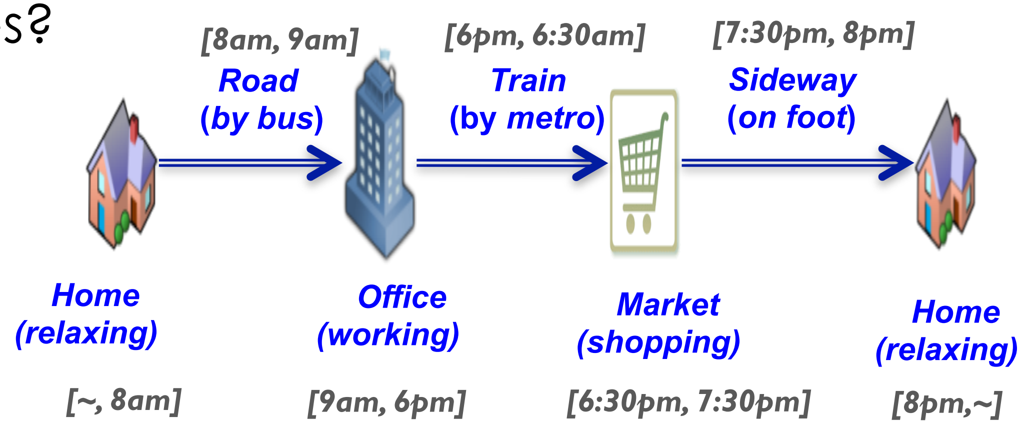
From “raw” to “semantic” trajectories

- From raw …
  - sequences of (x,y,t) points, e.g., GPS feeds
- … to meaningful mobility tuples <where, when, what/how/why>
  - Semantic Trajectory: $T = \{e_{first},\ldots,e_{last}\}$
  - Episode: e_i = (STOP | MOVE, t_{from}, t_{to}, *place*, *tag*)



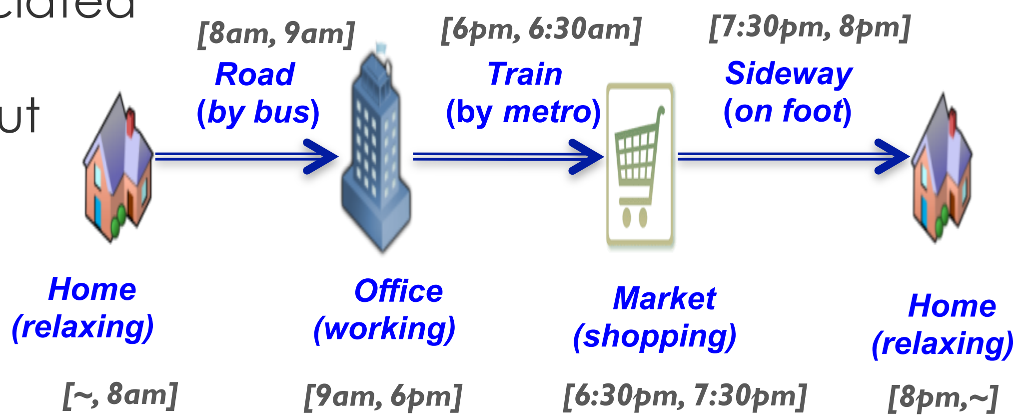
Semantic trajectory

- An alternative (semantically-annotated) representation of the motion path of a moving object
- Detection of homogenous fractions of movement,
  - A trajectory is reconstructed as a **sequence of episodes (stops/ moves)** along with appropriate **tags**
- Why semantic trajectories?
  - Not only a matter of downscaling the size of the database
  - Mainly, towards enriched movement analysis and understanding



Semantic trajectory (cont.)

- A trajectory is reconstructed as a **sequence of episodes (stops/ moves)** along with appropriate **tags**
  - **Stops** are the parts of the object's trajectory during which the object stays "static" at a place
  - **Moves** are the parts of the object's trajectory in between two Stops, i.e. where the object is "moving"
  - **Tags** are meta-data associated with Stops and Moves. Actually, information about
    - **when?**
    - **where?**
    - **how?**
    - **what?**
    - **why?**

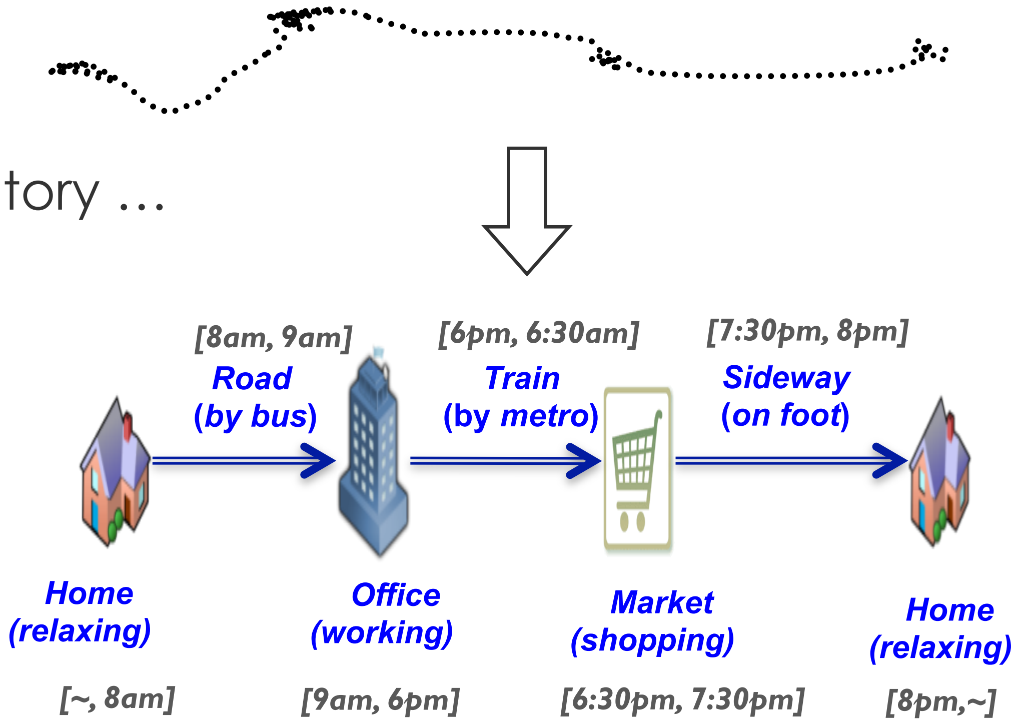


9.2.

The semantic enrichment process of raw trajectories

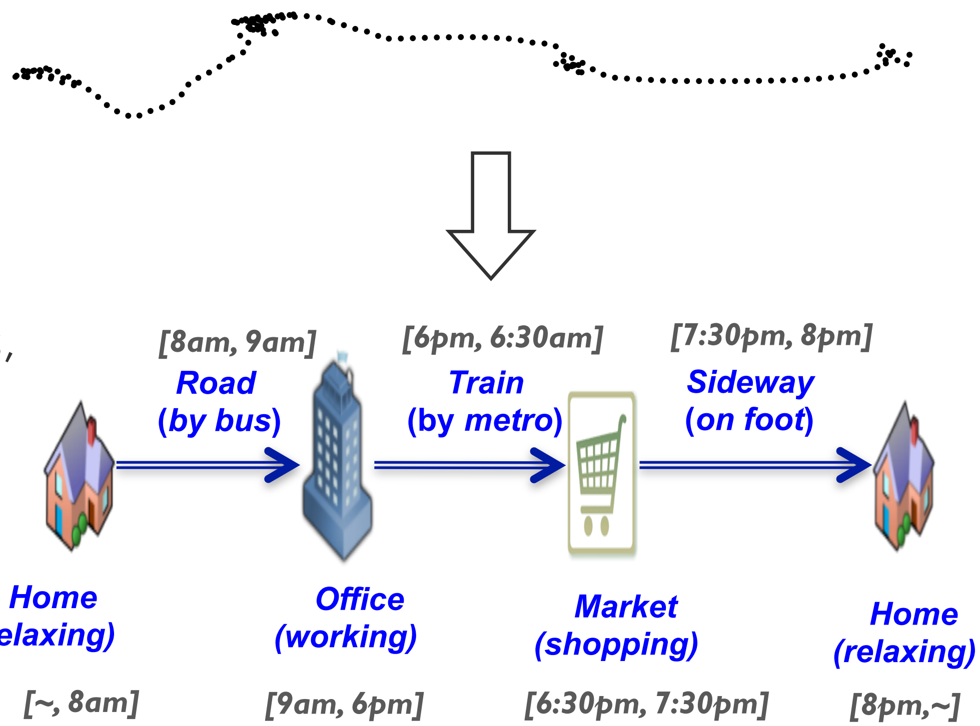
Semantic trajectory enrichment

- The process of adding application-oriented contextual information to raw trajectories
- Input: a 'valid' raw trajectory …
  - Recall Chap. 3
- … plus a contextual data repository
- Output: a semantic trajectory



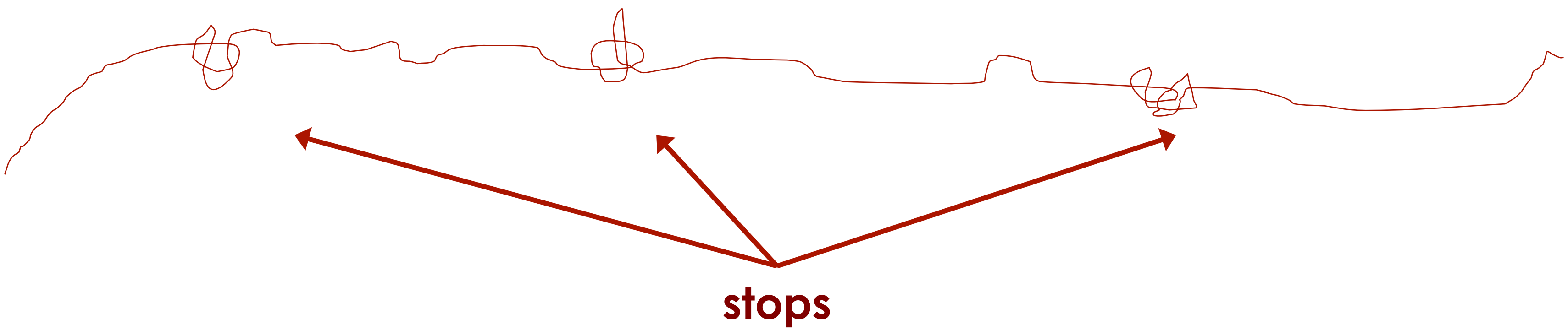
Semantic trajectory enrichment (cont.)

- Necessary intermediate step: **trajectory segmentation**
  - partition the trajectory into sub-trajectories that correspond to a specific behavior or activity
- Examples:
  - Detecting stops and moves,
  - Detecting changes in movement pattern,
  - etc.



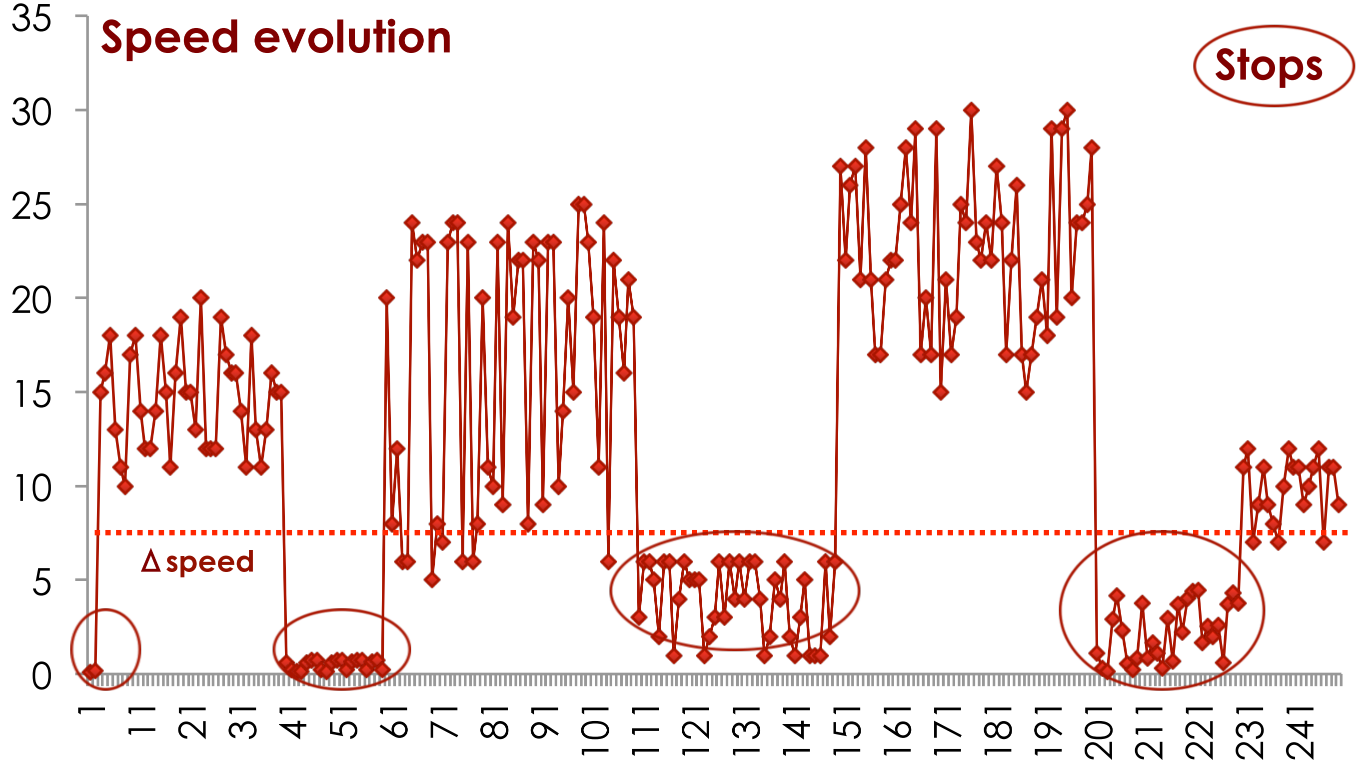
Stop discovery

- Issue: How can Stop be detected in a raw trajectory?
- Solutions:
  - when the trajectory intersects the geometry of a POI and the duration of intersection is above a given temporal duration threshold: **SMoT technique** (2007)
  - when dense areas of the trajectory points are detected, using e.g. a density-based clustering algorithm, and those areas are mapped to a POI: **CB-SMoT technique** (2008)



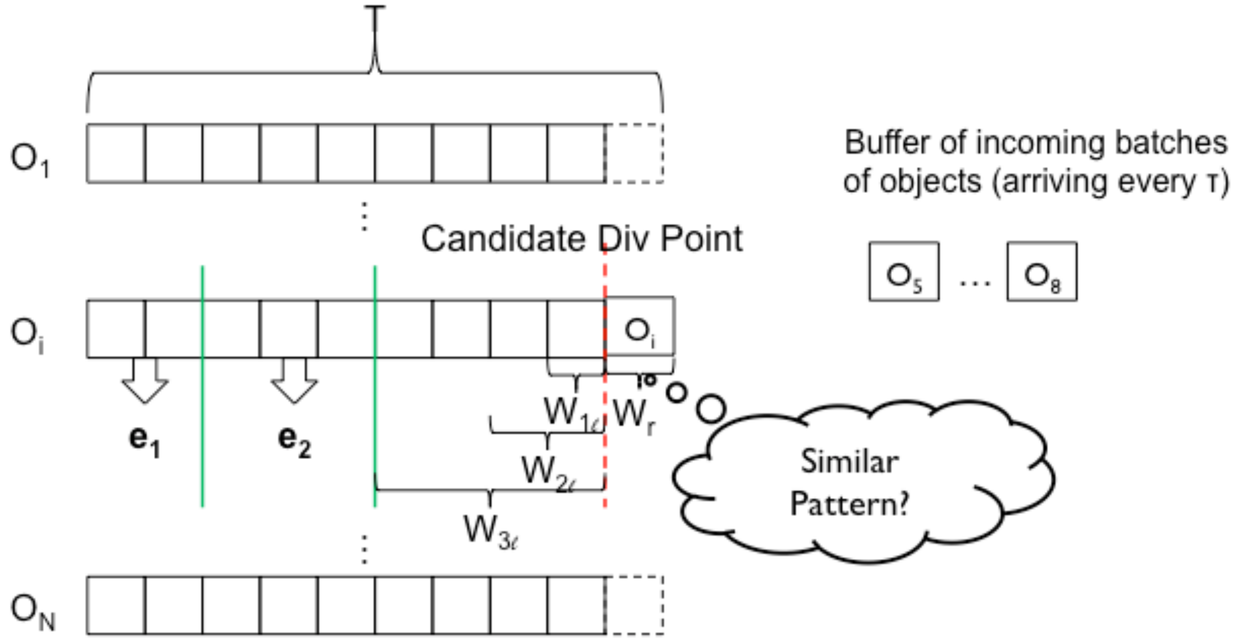
Stop discovery (cont.)

- Alternative: velocity-based stop identification



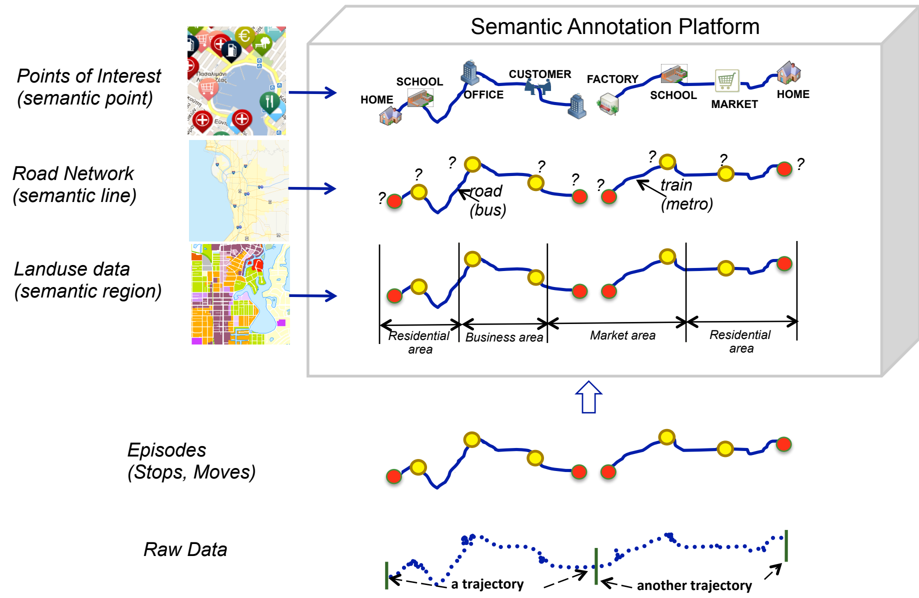
Online trajectory segmentation

- Quite useful in traffic monitoring scenarios
- An approach: **SeTraStream** (2011). Works in 3 steps:
  1. Cleaning and smoothing the incoming batch of status updates
     - status updates are described by Movement Feature Vectors (MVF)
  2. Compressing the batch by considering MVF characteristics
  3. Segmenting the batch by using a MVF matrix

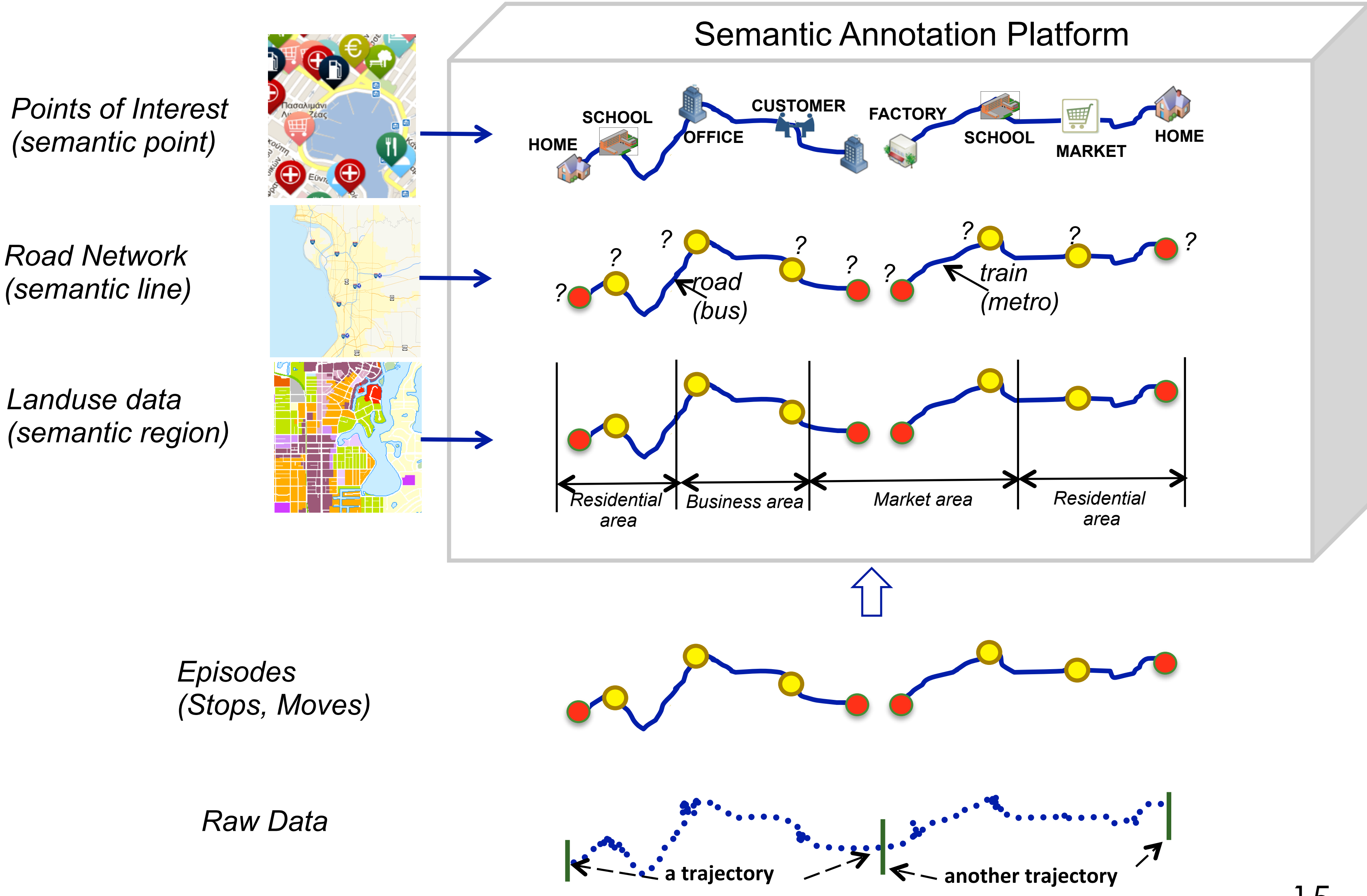


Semantic annotation of episodes

- … a complete framework: **SeMiTri** (2011-12)
- Three layers:
  1. **Semantic regions**: annotate trajectories with geographic regions of interest - ROIs (using OpenStreetMaps)
  2. **Semantic lines**: annotate trajectories with e.g. road network
  3. **Semantic points**: annotate Stops with POI types instead of POIs (using Hidden Markov Model)



The SeMiTri framework

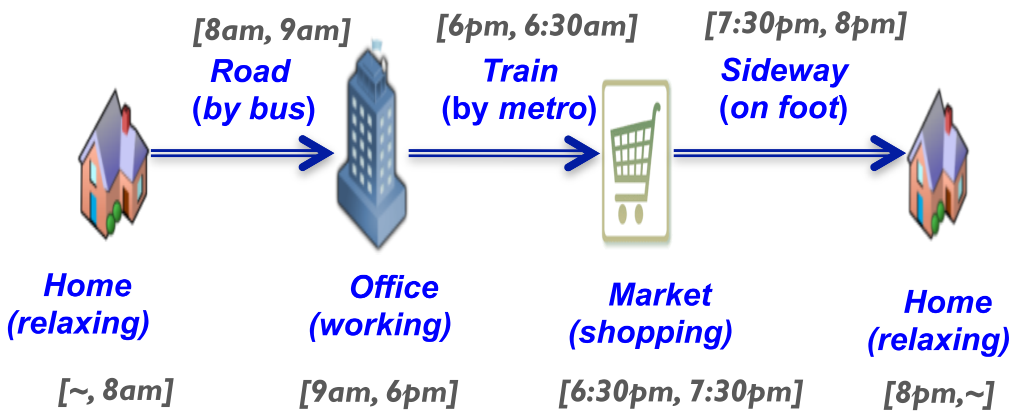


9.4.

Semantic trajectory data management

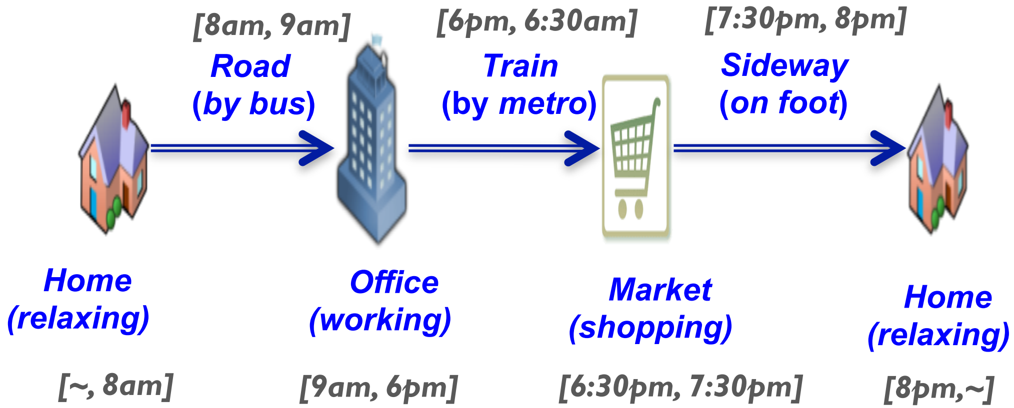
STD management requirements

- Efficient support of both raw and semantic trajectory databases (TD and STD, resp.)
- Indicative queries:
  - "Search for people who follow the home – office – home pattern every weekday"
  - "Search for people who cross the city center on their way from office back to home"
  - "Search for people who make long trips (e.g. more than 20 km) on their way from home to office without including intermediate stops"

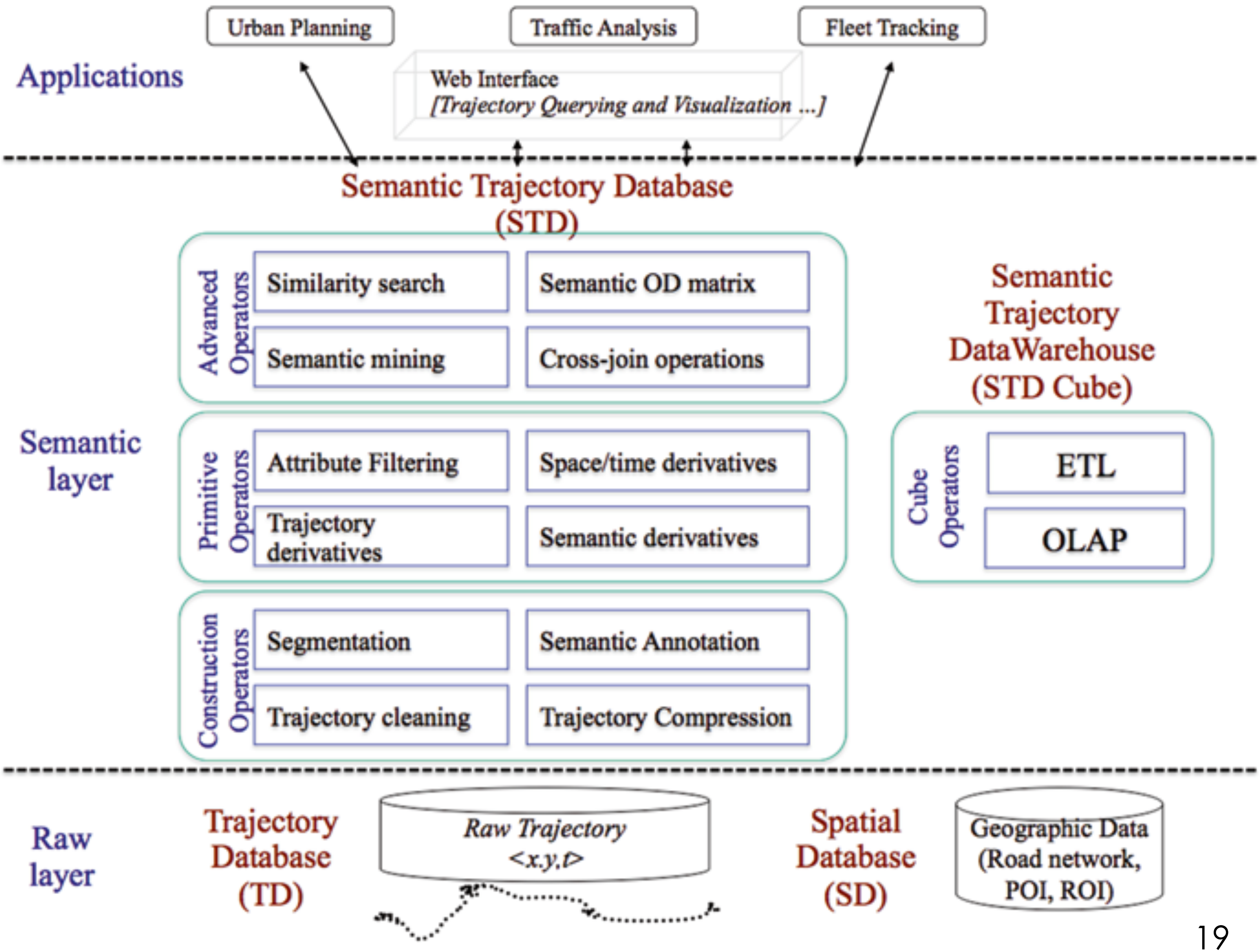


STD management requirements (cont.)

- Such queries are **innovative** and cannot be handled effectively and efficiently by existing approaches and corresponding MOD engines
- Moreover, by maintaining a semantic trajectory data cube, we are able to support analysis of type:
  - "when, where and why moving objects of a specific profile stop?"
  - "when, how and where from/to moving objects of a specific profile move?"

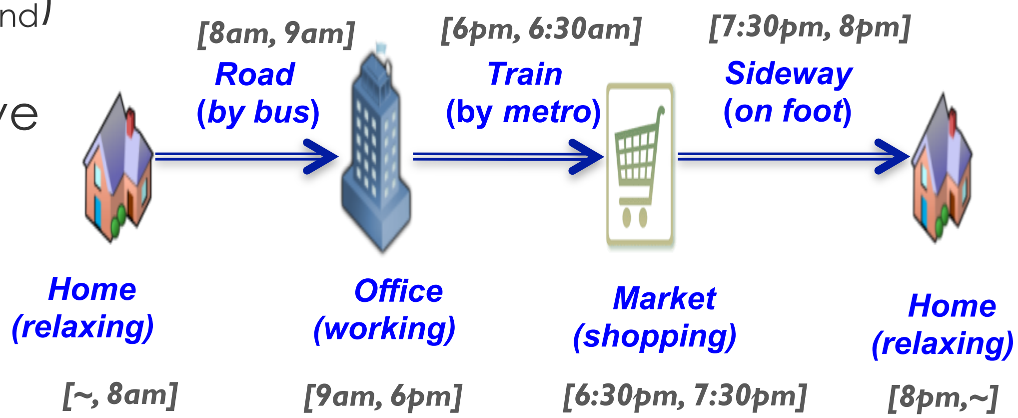


- The 'big picture' of a semantic trajectory DB/DW infrastructure …


Urban Planning
Traffic Analysis
Fleet Tracking
Applications
Web Interface
[Trajectory Querying and Visualization ...]
Semantic Trajectory Database
(STD)
Advanced Operators
Similarity search
Semantic OD matrix
Semantic mining
Cross-join operations
Semantic Trajectory DataWarehouse (STD Cube)
Semantic layer
Primitive Operators
Attribute Filtering
Space/time derivatives
Trajectory derivatives
Semantic derivatives
Cube Operators
ETL
OLAP
Construction Operators
Segmentation
Semantic Annotation
Trajectory cleaning
Trajectory Compression
Raw layer
Trajectory Database (TD)
Raw Trajectory
<x,y,t>
Spatial Database (SD)
Geographic Data
(Road network,
POI, ROI)

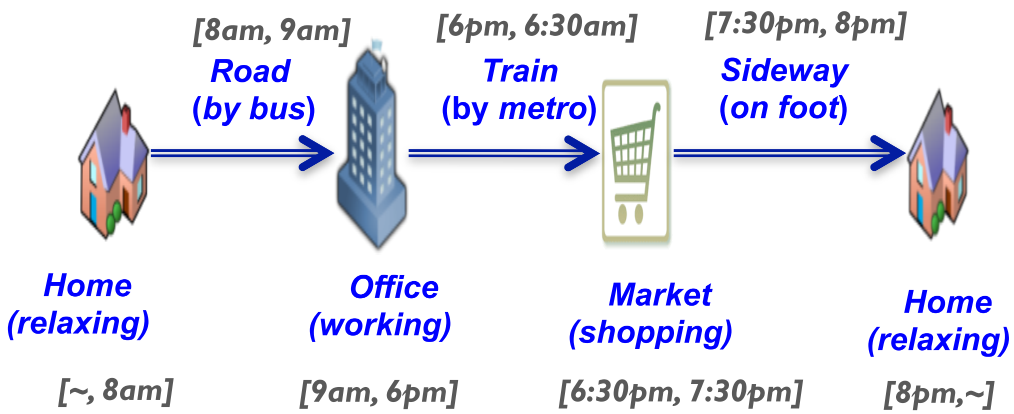
A data type system for STD

- Two novel datatypes, 'episode' and 'semantic trajectory'
  - corresponding to raw sub-trajectory and raw trajectory, resp.
- **episode**: a tuple (defineTag, MBB, episodeTag, activityTag, T-link), where:
  - defineTag: a flag in {Stop, Move}
  - MBB: a tuple (MBR, t_{start}, t_{end}) corresponding to the 3D coverage of the respective raw sub-trajectory
  - episodeTag and activityTag: semantic information
  - T-link: link to the respective raw sub-trajectory



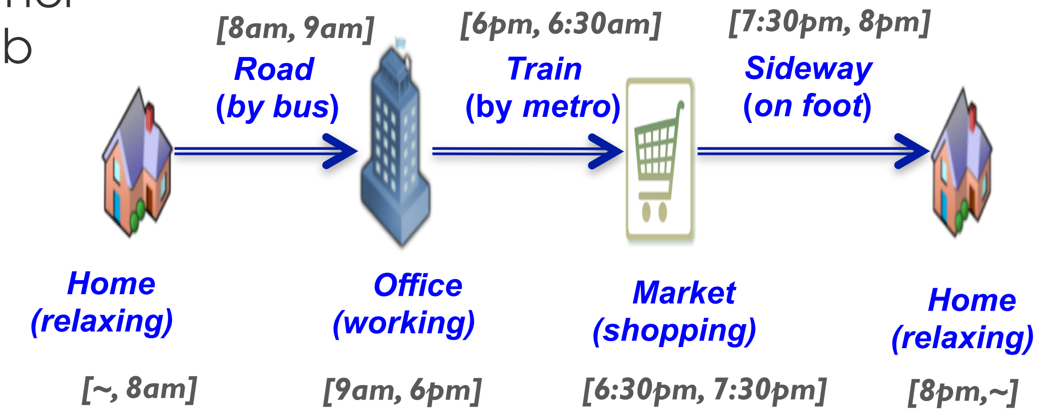
A data type system for STD (cont.)

- Two novel datatypes, 'episode' and 'semantic trajectory':
  - corresponding to raw sub-trajectory and raw trajectory, resp.
- **semantic trajectory**: a tuple (o-id, semtraj-id, T_{sem}), where:
  - o-id: the moving object identifier
  - sem-traj-id: the semantic trajectory identifier
  - T_{sem}: a sequence of episodes, $\{e_1, \ldots, e_n\}$, ordered in time



A data type system for STD (cont.)

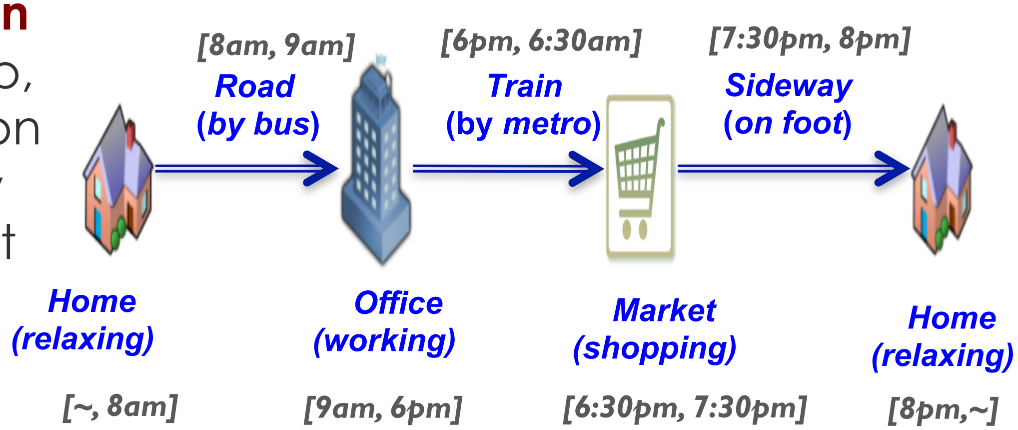
- Primitive methods and operators on 'episode' datatype:
  - number **duration** (), number **length** (), number **avg-speed** (): returns the duration, length, average speed, resp., of the episode
  - geometry **PAA** (): returns the Potential Area of Activity (PAA) of the episode
  - boolean **intersects** (MBB b): returns true or false, whether episode's MBB intersects b or not



A data type system for STD (cont.)

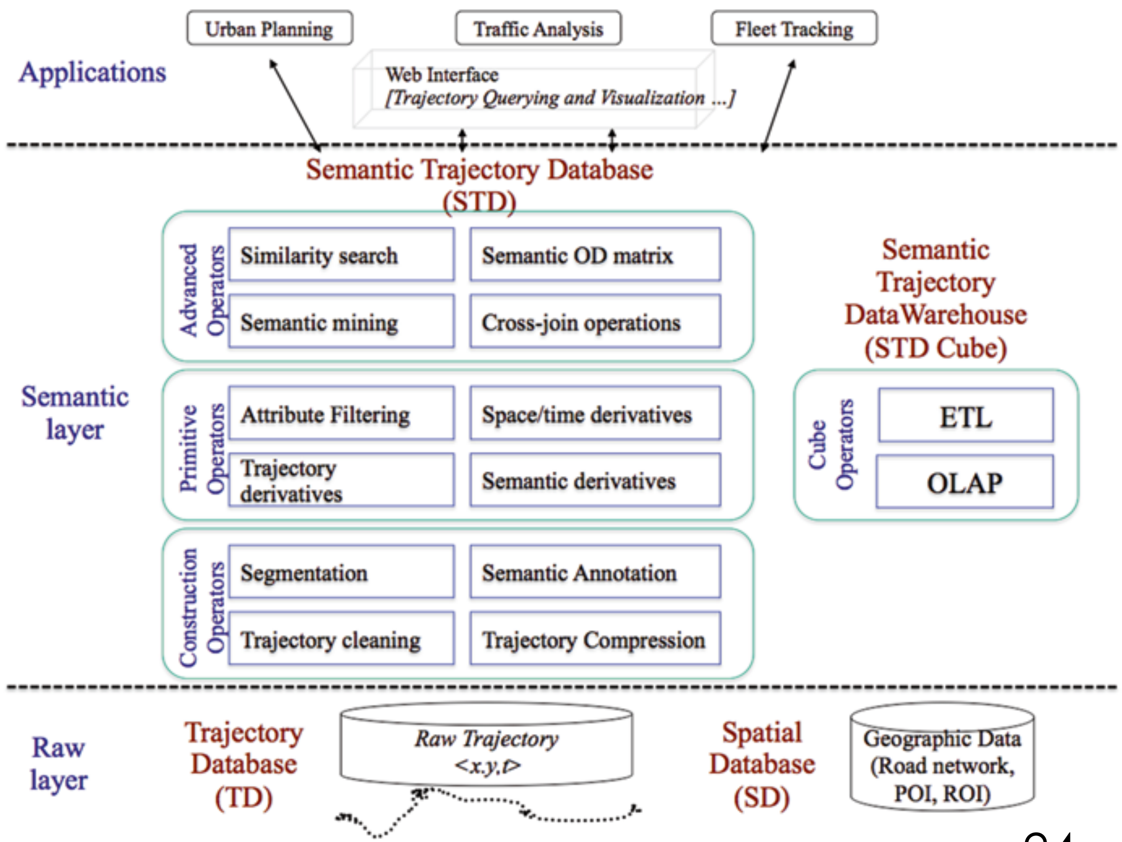
- Primitive methods and operators on 'semantic trajectory' datatype:
  - number **num_of_episodes** (string tag),

    set[episode] **episodes_with** (string tag): returns the number of episodes or the episodes themselves, resp., of the semantic trajectory that include tags LIKE tag
  - sem_trajectory **confined_in** (geometry g, timeperiod p, string tag): returns a portion of the semantic trajectory consisting of episodes that
    - spatially overlap with g,
    - temporally intersect with p, and
    - textually match with tag



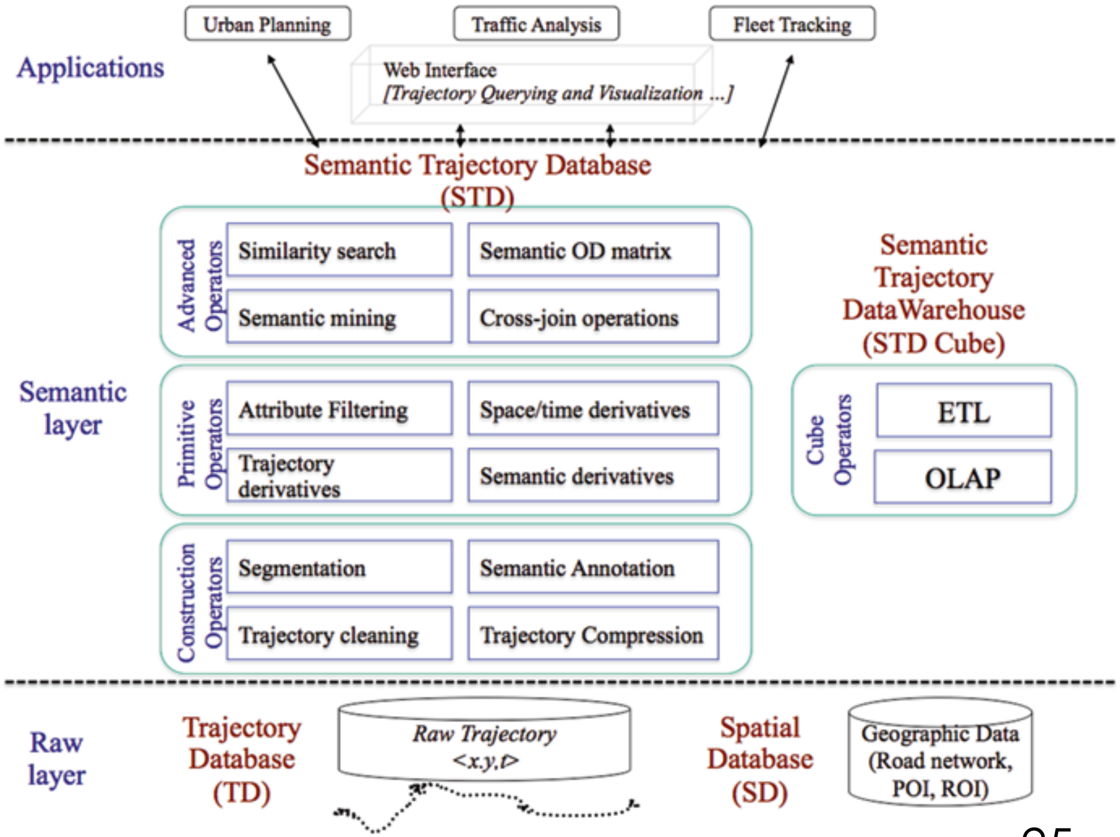
Querying STD

- Having such methods and operators in our hands, several types of queries can be defined:
  - **Q1 type**: raw trajectory queries (involving TD)
    - Example: "Search for people who crossed park X at night"
  - **Q2 type**: semantic trajectory queries (involving STD)
    - Example: "Search for people who follow the pattern home – office – home every weekday"
  - …



Querying STD (cont.)

- Having such methods and operators in our hands, several types of queries can be defined:
  - …
  - **Q3 type**: cross-over semantic trajectory queries (involving both TD and STD)
    - Example: “Search for people who cross the city center on their way from office back to home”
    - Technically supported through T-link (recall the def. of episode datatype)

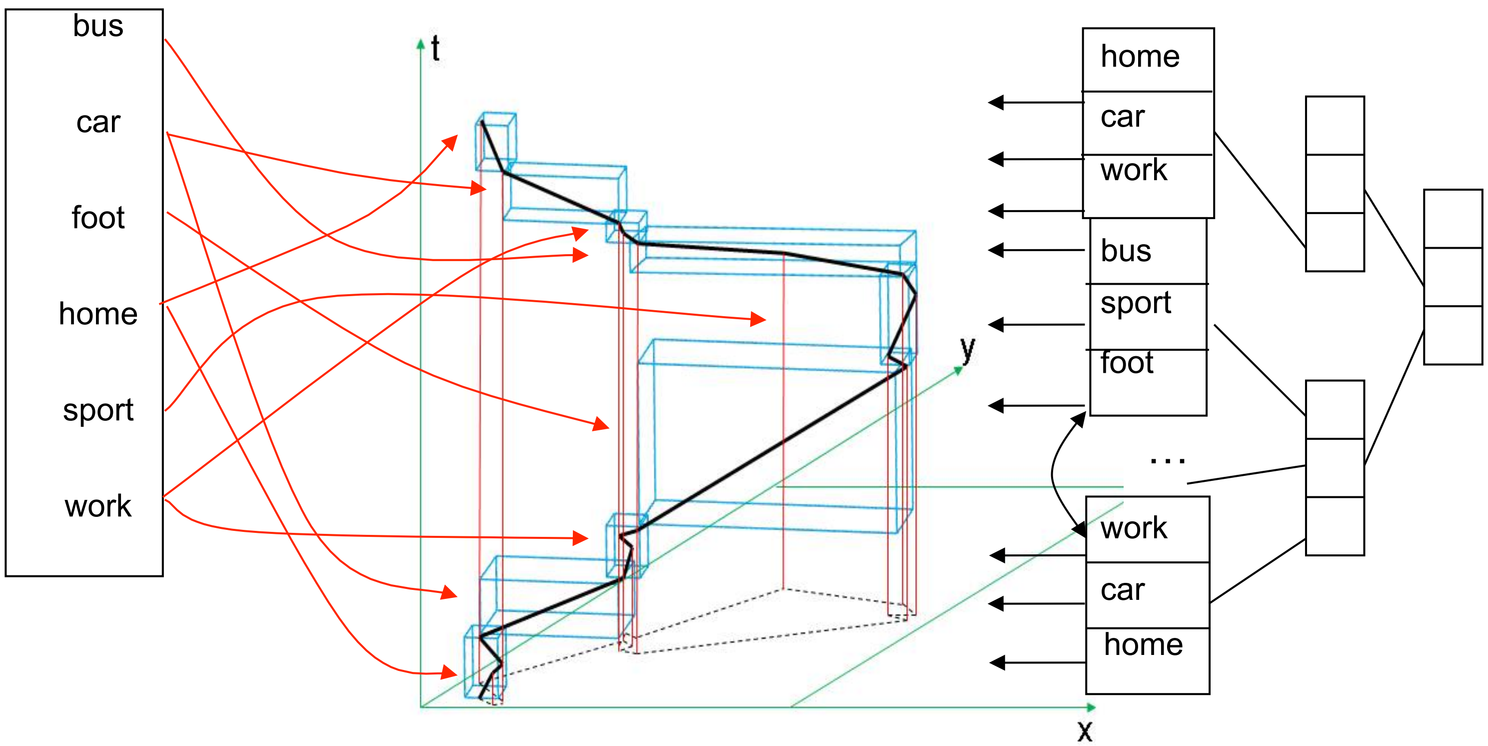


Indexing STD

- Issue: equally manage the spatio-temporal and the semantic (textual) component of semantic trajectories
- Baseline solutions:
  - Spatio-textual indexing structures (e.g. R-trees enhanced with textual search capabilities)
  - Trajectory data structures (e.g. TB-trees) along with text indexes (e.g. inverted files)
- A proposal: Semantic Trajectory Bundle (STB) tree
  - Builds upon spatio-temporal information of episodes (TB-tree) …
  - … also maintaining their textual information (inverted file)

Indexing STD (cont.)

- **STB-tree** (2013)
  - a TB-tree enhanced with textual information (right) …
  - … along with an inverted file on tags (left)

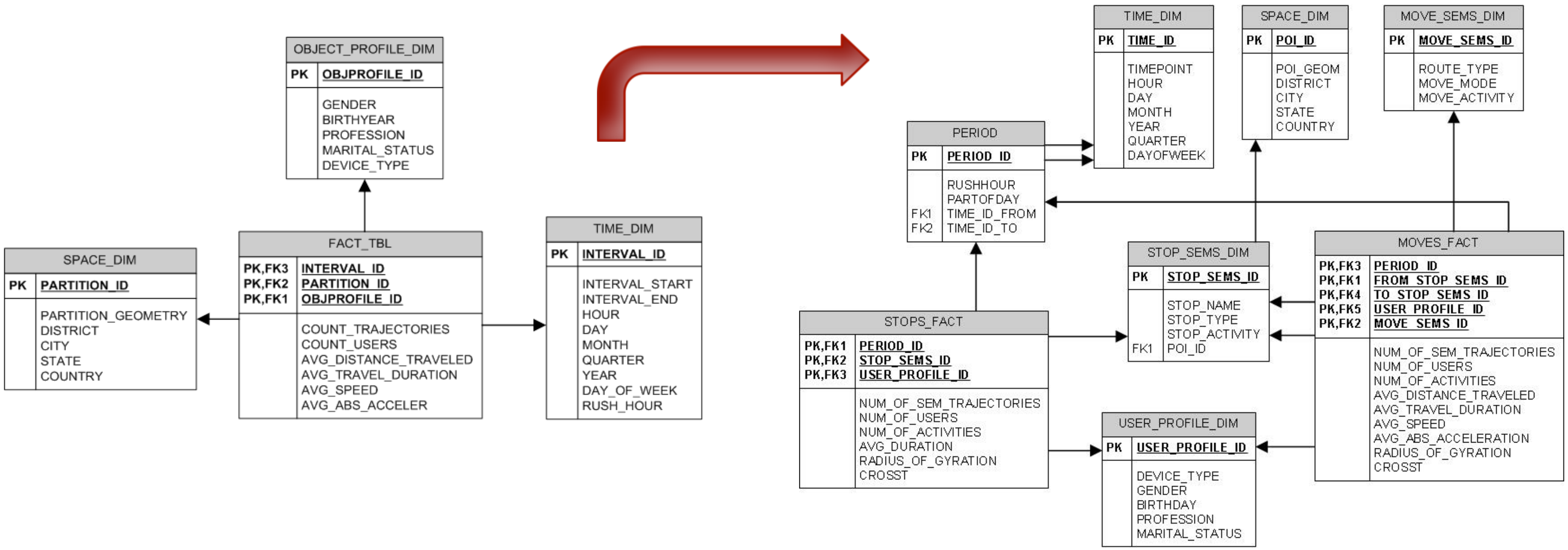


9.4.

Semantic trajectory data exploration

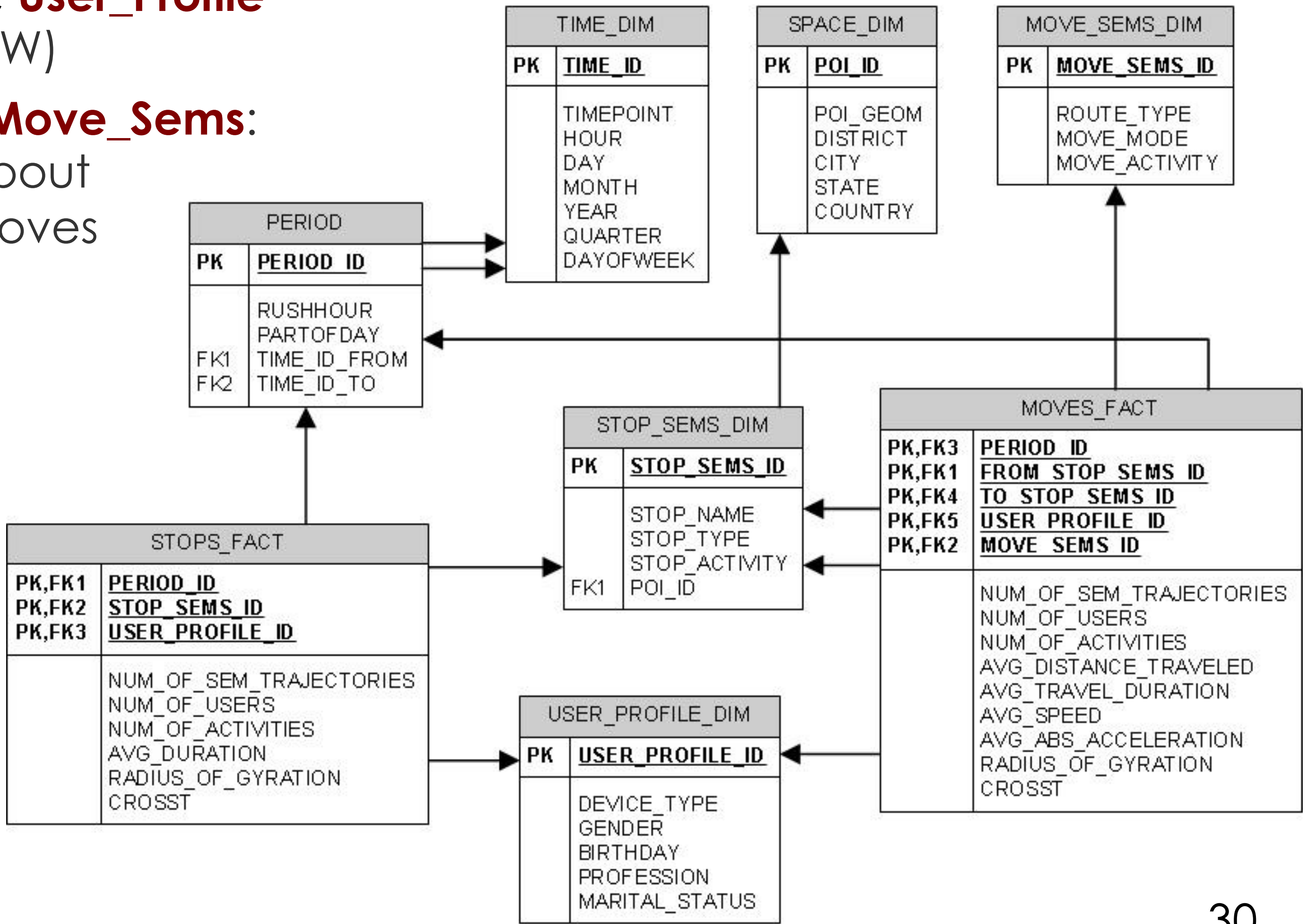
From raw to semantic TDW

- Extend TDW (recall Chap. 6) with dimensions and facts about:
  - **STOPS**: who made a stop? when and where? What did she do during her stop?
  - **MOVES**: who made a movement? when and where from/to? How did she move and what did she do during her motion?



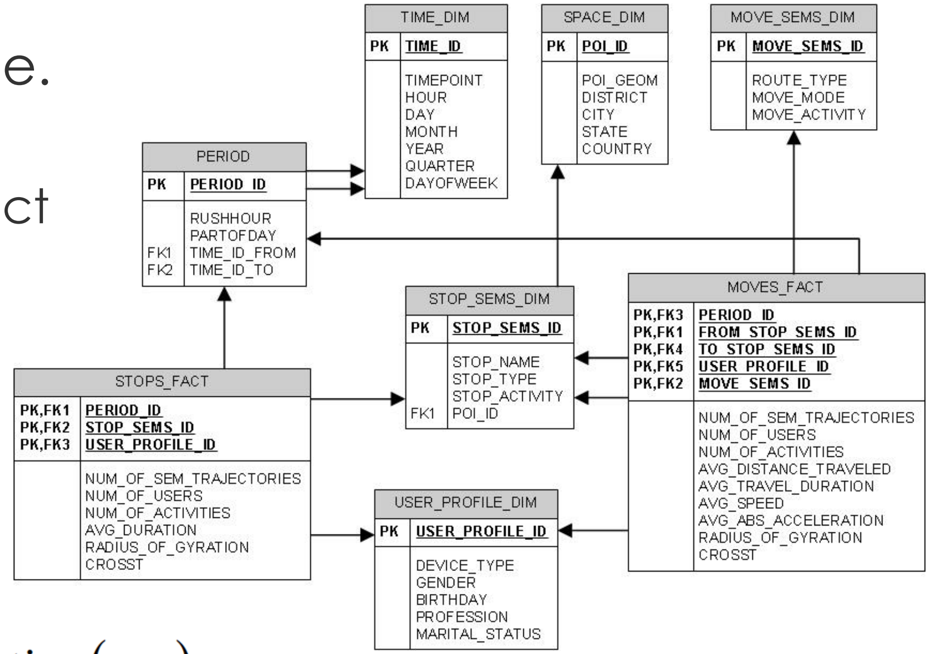
A semantic trajectory data cube

- 5 dimensions:
  - **Space**, **Time**, **User_Profile** (as in raw TDW)
  - **Stop_Sems**, **Move_Sems**: semantics about Stops and Moves
- 2 fact tables:
  - **STOPS_Fact**
  - **MOVES_Fact**



ETL in semantic trajectory data cubes

- (as usual) efficient ETL is an issue. For instance:
  - **num-of-sem-trajectories** is subject to the distinct count problem (recall Chap. 6)
  - **avg-duration** and **avg-distance-traveled** are Q3 type queries



$$avg\text{-}duration\left(bc_{stop}\right) = \frac{\sum_{ep_i \in bc_{stop}} duration\left(ep_i\right)}{num\text{-}of\text{-}sem\text{-}trajectories\left(bc_{stop}\right)}$$

$$avg\text{-}distance\text{-}traveled\left(bc_{move}\right) = \frac{\sum_{ep_i \in bc_{move}} length\left(ep_i\right)}{num\text{-}of\text{-}sem\text{-}trajectories\left(bc_{move}\right)}$$

Mining STD

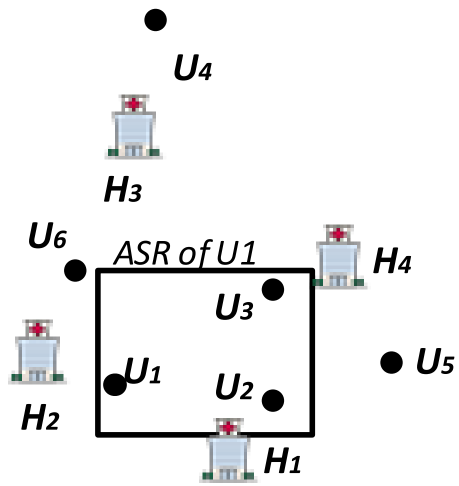
- Semantic trajectory mining is an infant, very challenging field, for analysis purposes
- Example (from clustering):
  - Try to discover a pattern like "this group of students (coming from different origins) stopped at piazza P participating in a protest for an hour; then walked to café C where they stayed for half an hour; and finally they went by tram T to bar B where they partied until midnight"
  - Constraints of types who? when? where? what? how?
- Critical technical objective: **semantic trajectory similarity function**. Alternative roadmaps:
  - Map semantic trajectories into vectors in a feature space (features may include spatial, temporal, and textual aspects)
  - Define an aggregate function between dissimilarity in spatio-temporal and dissimilarity in semantic (textual) space.

9.5.

Semantic aspects of privacy

LBS for sensitive semantic locations

- What if an adversary is aware of the semantic whereabouts of a territory? May be able to extract sensitive personal information
  - Hence, **sensitive locations** should be somehow protected
- An idea: **semantic location cloaking along with l-diversity** (2011)
  - Location is generalized to a **cloaked region** (CR) that includes at least K LBS users (recall Chap. 8) and at least l different POIs (recall Chap. 2)
  - However, we cannot avoid **semantic location identification attacks**



- Another idea: maintain a set of **strongly CRs** (2012)
  - Supports personalized privacy profile:
    - sensitive vs. non-sensitive POIs
    - the maximum probability of linking a user with a sensitive place

Privacy in STD

- Applying K-anonymity in semantic trajectories is not enough:
  - Example: all K trajectories stop at a clinic (sensitive place)
- An idea: **C-safe Anonymization of Semantic Trajectories** (CAST, 2011)
  - Produces a sanitized (c-safe) version of a semantic trajectory …
    - by generalizing at the semantic level (place taxonomy)
  - … under the assumption that an adversary is aware of
    - the utilized anonymization process
    - the place taxonomy
    - the presence of a user in the dataset
    - the quasi-identifier sequence of visited places

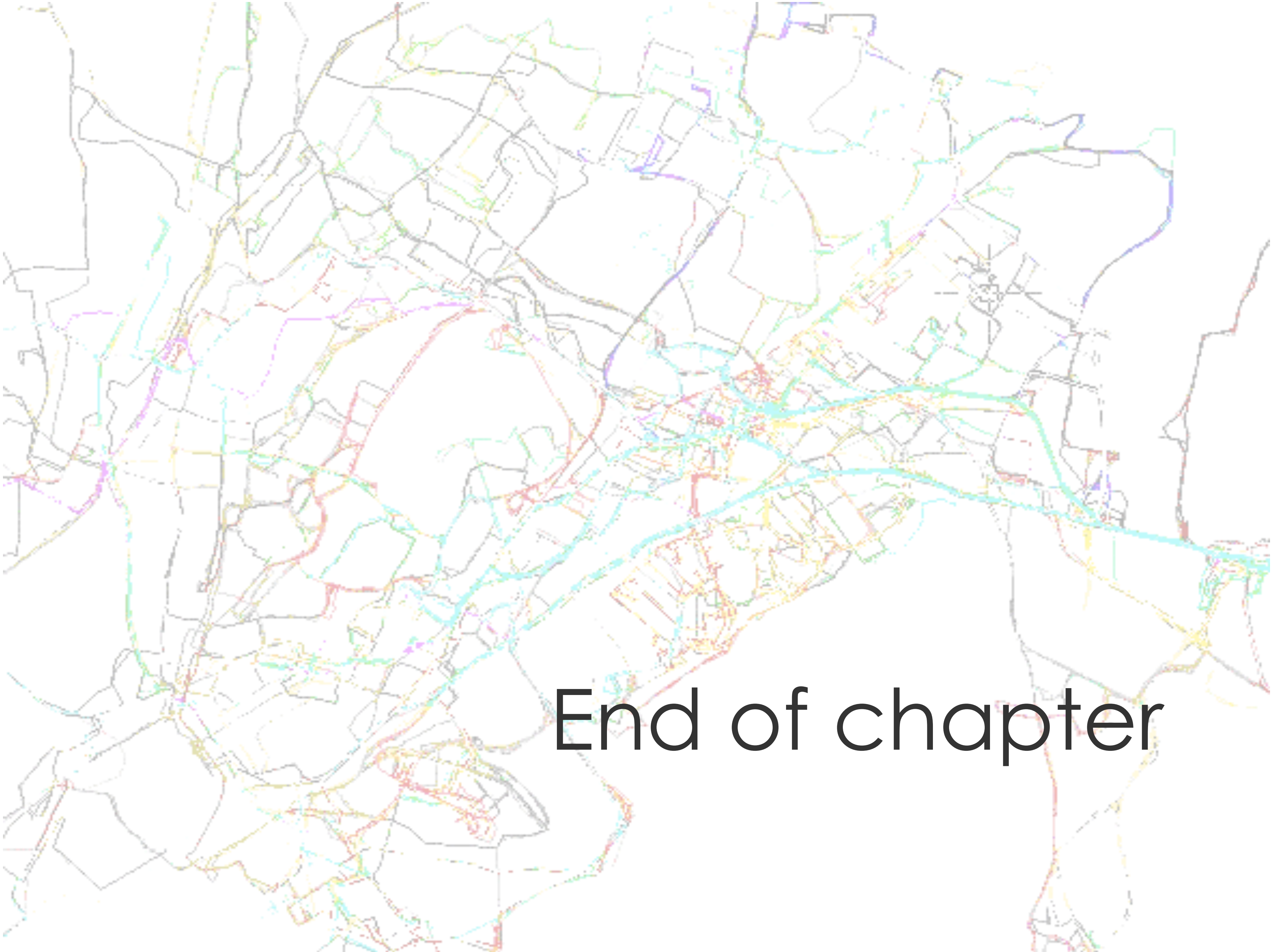
9.6.

Summary

Summarizing …

- (as for their raw counterparts) semantic aspects of mobility data ask for effective and efficient modeling, management, and knowledge discovery

- In this chapter, we presented:
  - How to model a semantic trajectory, from the abstract concept to the datatype level
  - The types of queries that are of interest in a STD
    - and how to index an STD for efficient query processing
  - The emerging challenges for STD mining and privacy



End of chapter