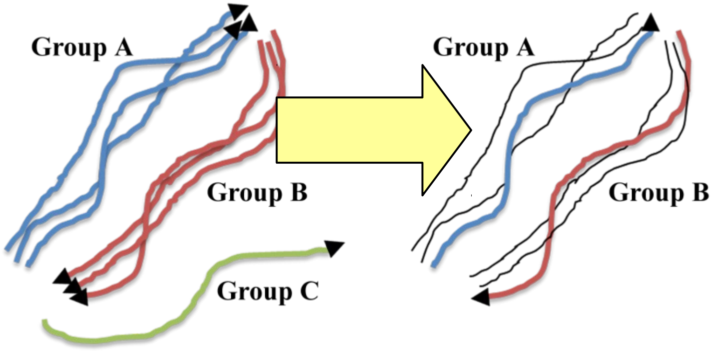
Mobility Data Mining

Understanding Human Mobility

Mobility Profiles

Derived patterns and models

- Combination & refinement of basic patterns and models



- Individual Mobility Profile: routines consistently followed by a single moving object

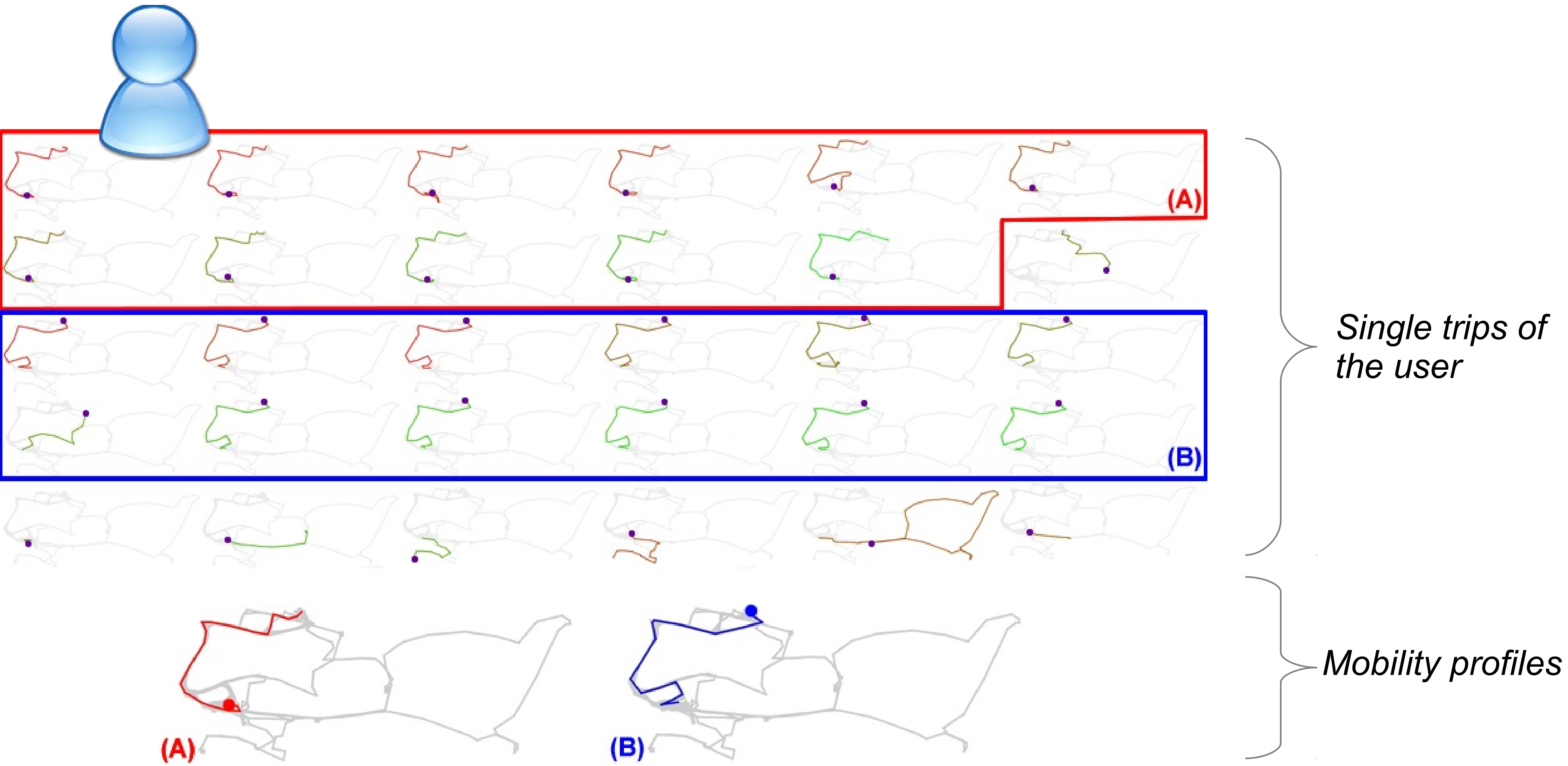
User’s Mobility Profile

Given the user history as an ordered sequence of spatio-temporal points, we want to extract a set of *routines* in order to create the his\her *mobility profile.*

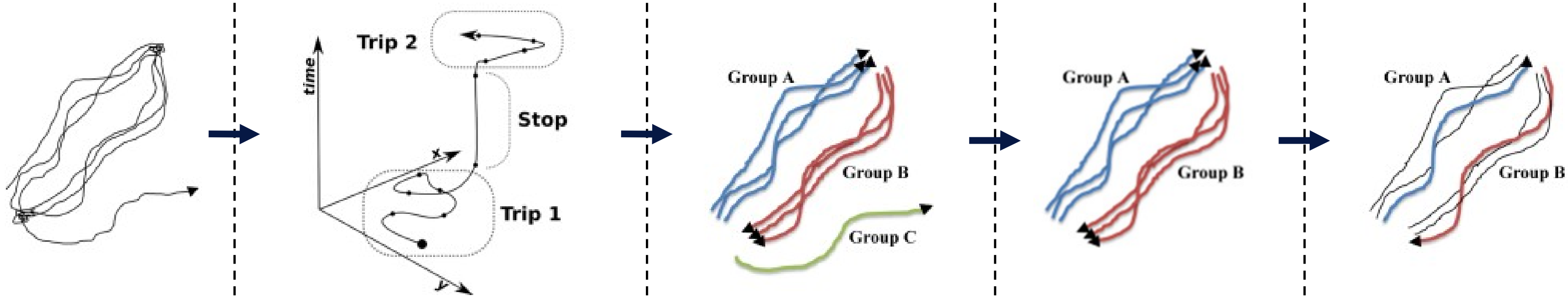
Where:

- A *Routine* is a typical local behavior of the user.
- A *Mobility profile* is the set of user’s routines

User's Mobility Profile



Derived patterns and models: mobility profiles



User history	Trips construction	Grouping	Pruning	Profile extraction
An ordered sequence of spatio-temporal points.	Cutting the user history when a **stop** is detected	Performing a density based **clustering** equipped with a **spatio temporal** distance function	Groups with a small Number of trips are Pruned	The **medoid** of each group becomes user's **routines** and the all set become the user's **mobility profile**
	Stops Spatial Threshold Stops Temporal Threshold	Spatial Tollerance Temporal Tollerance Spatio temporal distance	Support Threshold	

Trasarti, Pinelli, Nanni, Giannotti.
***Mining mobility user profiles for car pooling**. ACM SIGKDD 2011*

What kind of distance?

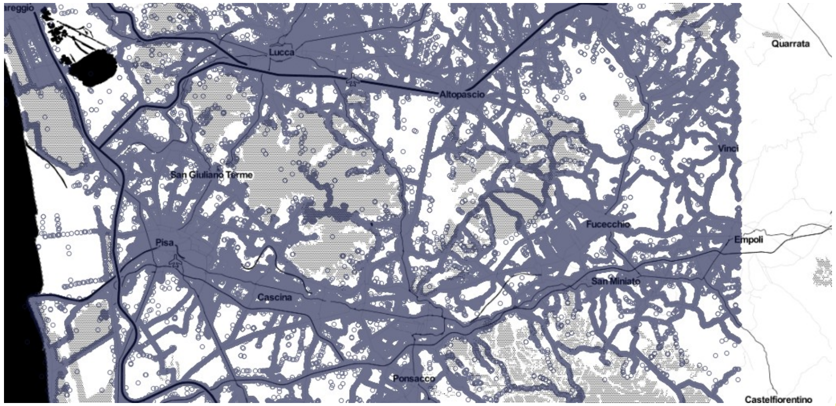
- Start + End
  - Look for origin-destination pairs
- Route similarity
  - Look for recurrent paths followed
- Temporal dimension
  - Include time (of the day) to distinguish temporal regularity

What kind of representative?

- Classical average centroid cannot be applied
  - What is the centroid trajectory? Could make no sense
- Two practical solutions
  - Medoid: most central element of the cluster, e.g. minimized the sum of distances
  - Random: good enough if the clustering parameters are tight

Map Matching

Objective

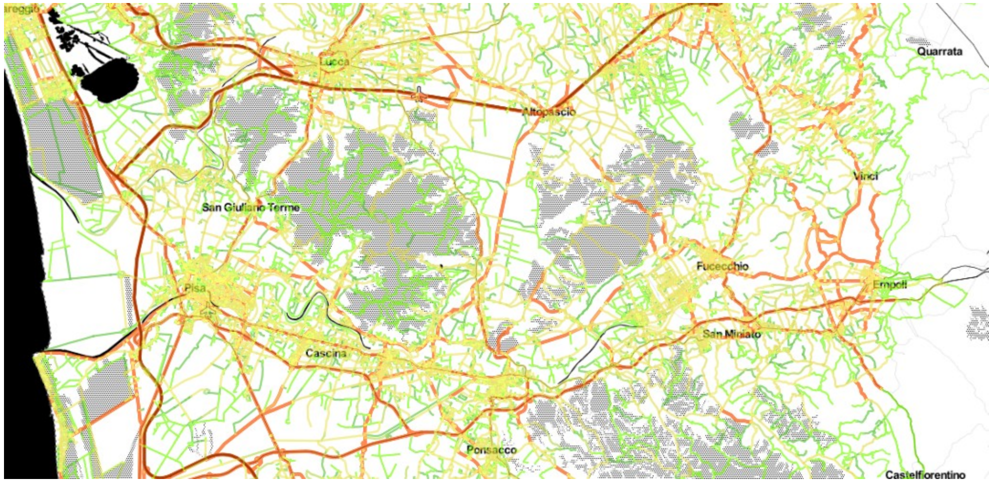


How to transform this...

Gps raw trajectories
Avg sampling rate 90 seconds
Affected by GPS positioning error

...into this?

Sequences of road segments crossed

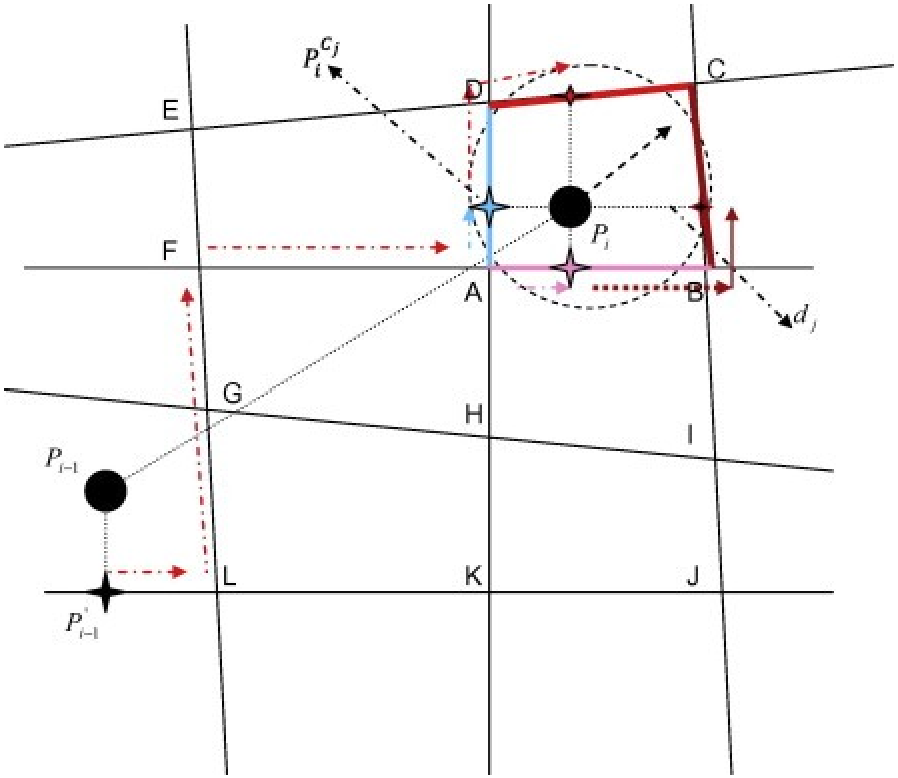


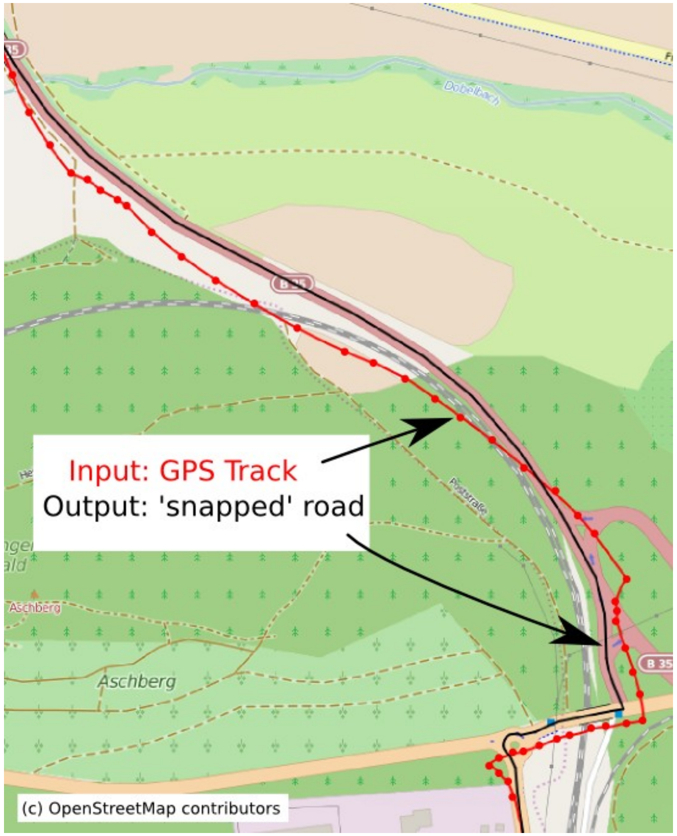
Objective

- Associate a sorted list of user positions to the road network on a digital map
- Two kinds of problems to solve
  - Map points to streets
  - Reconstruct path between points

Point mapping

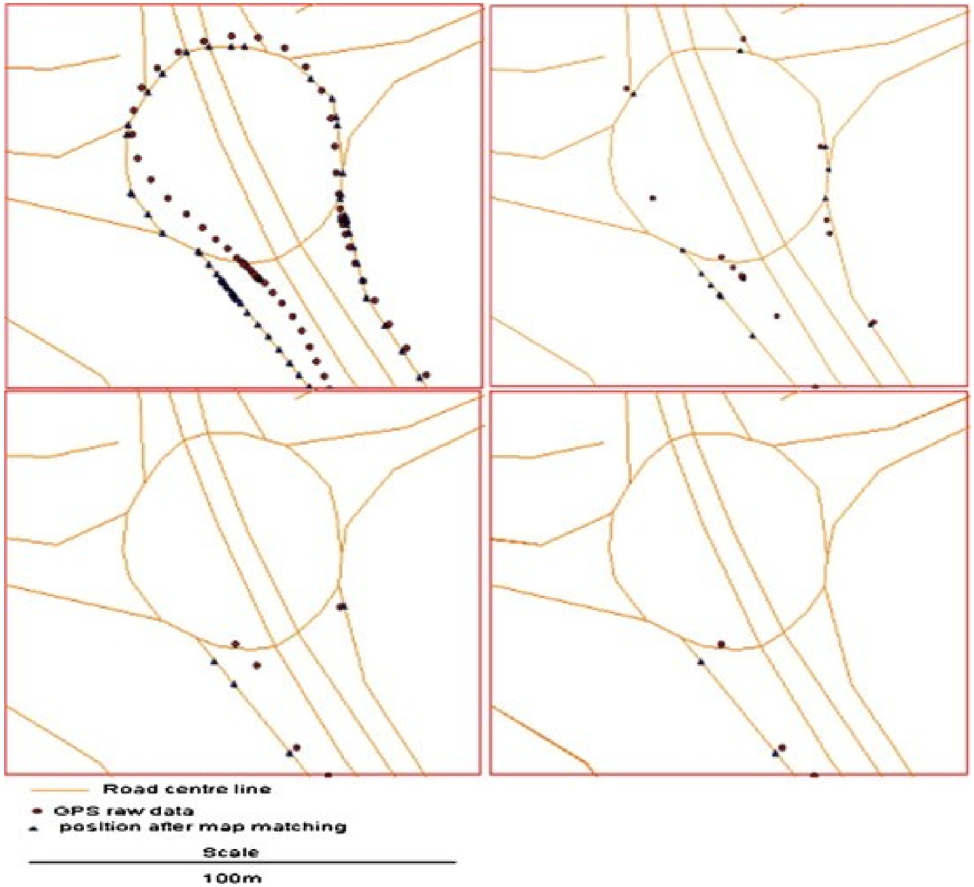
- Determine which road segment a point belongs to
- Choose position within the segment





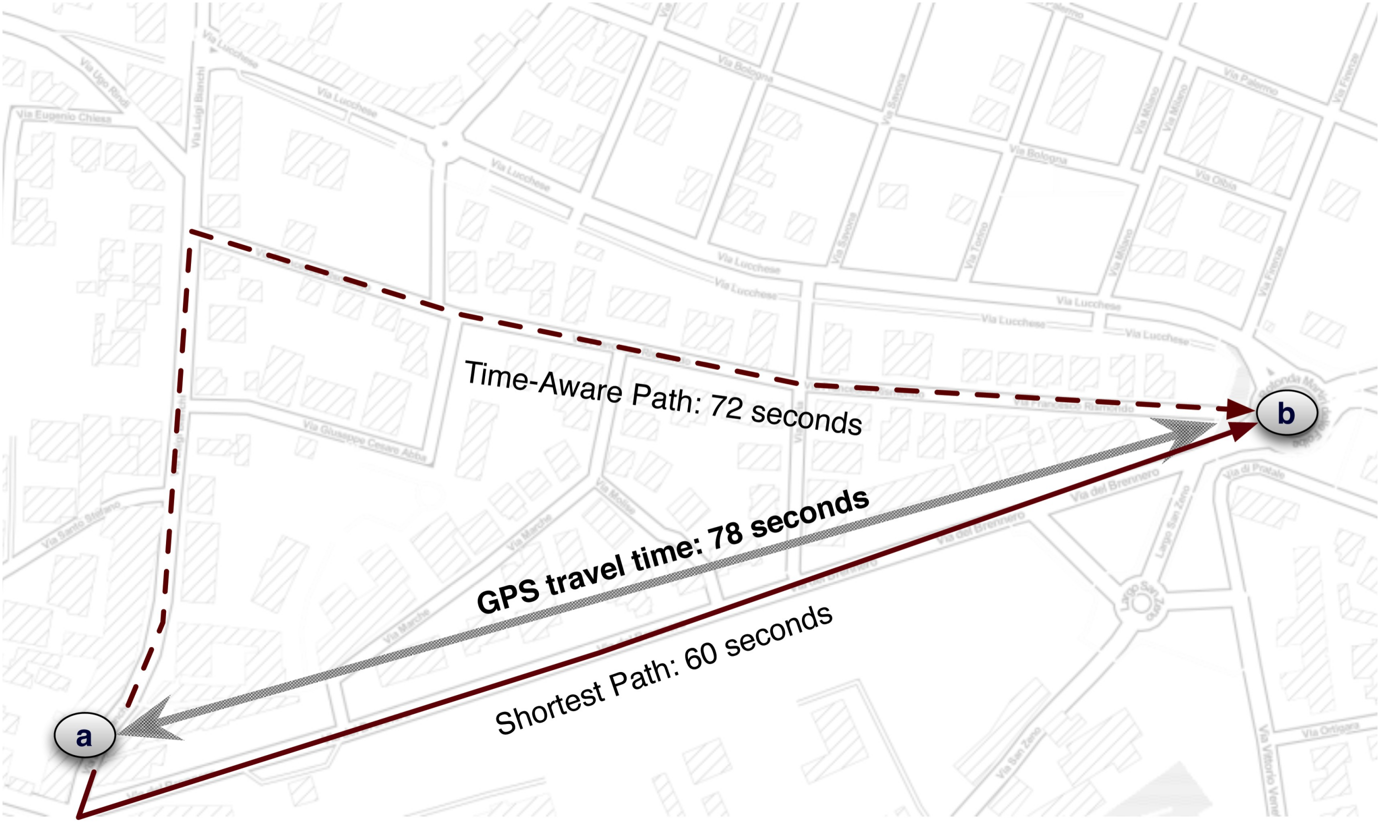
Objective

- Path reconstruction
  - Needed when gap between points is large



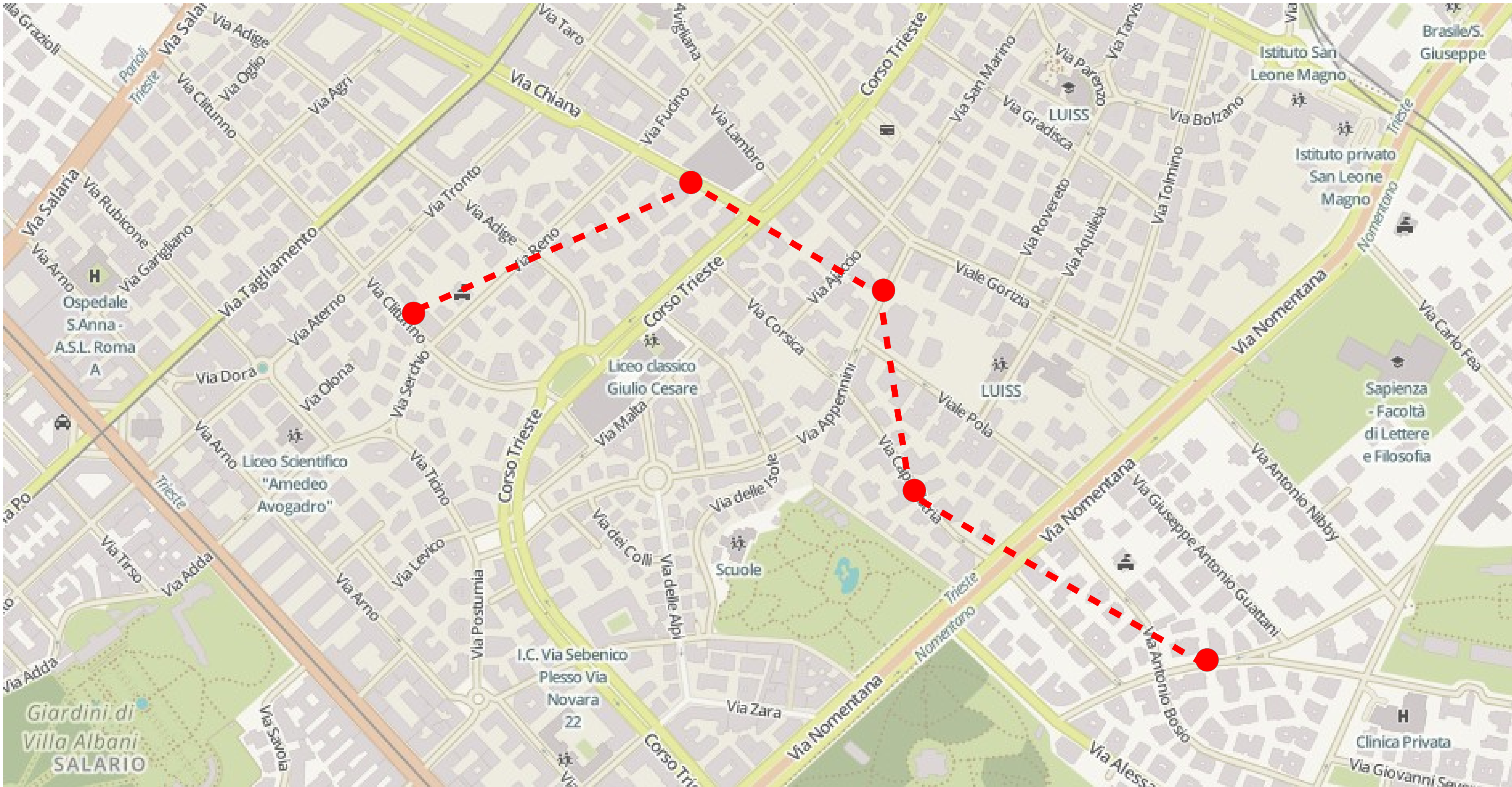
State of the art

- Map matching algorithms rely on shortest path between GPS points



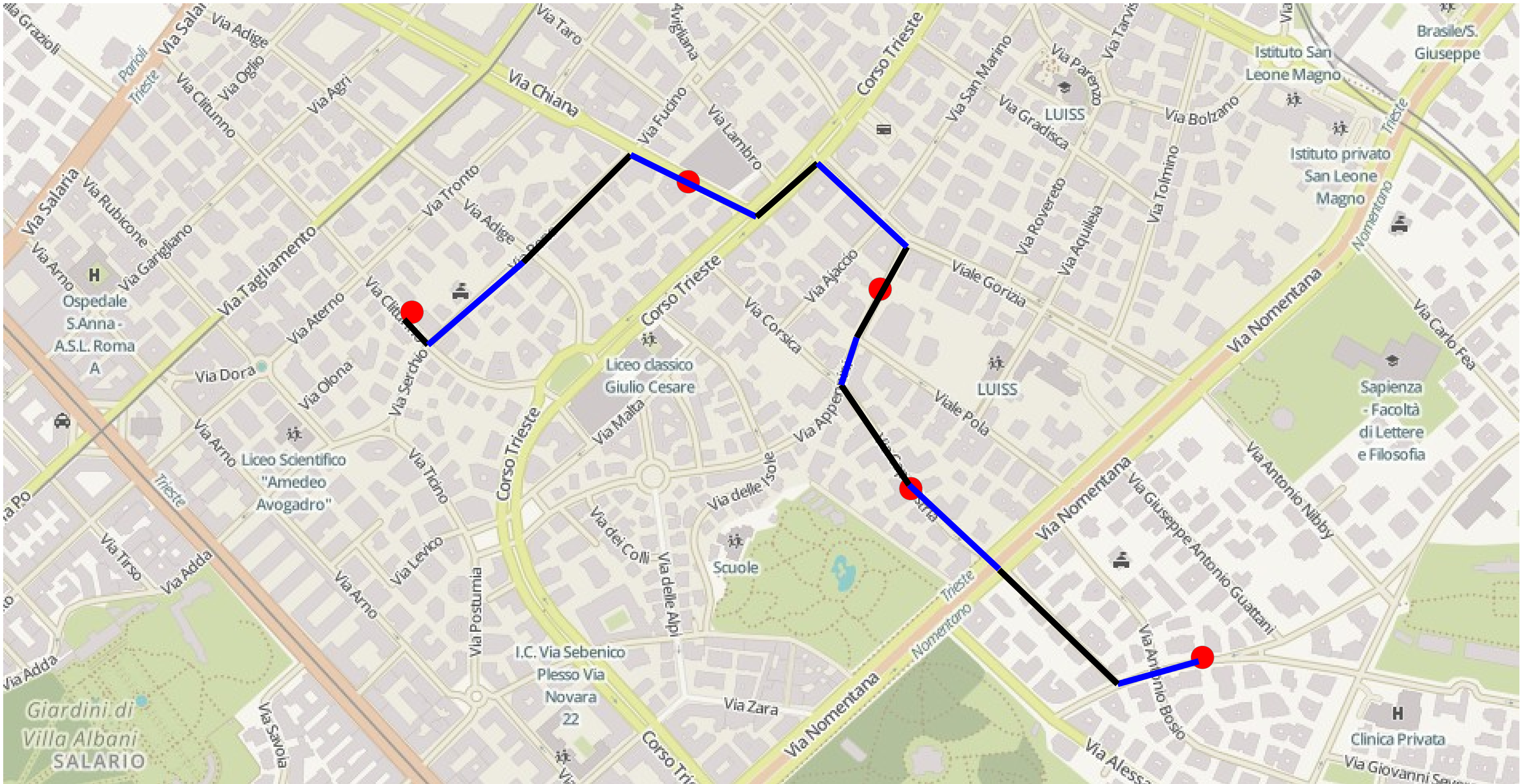
Result

- Trajectory → sequence of road segments

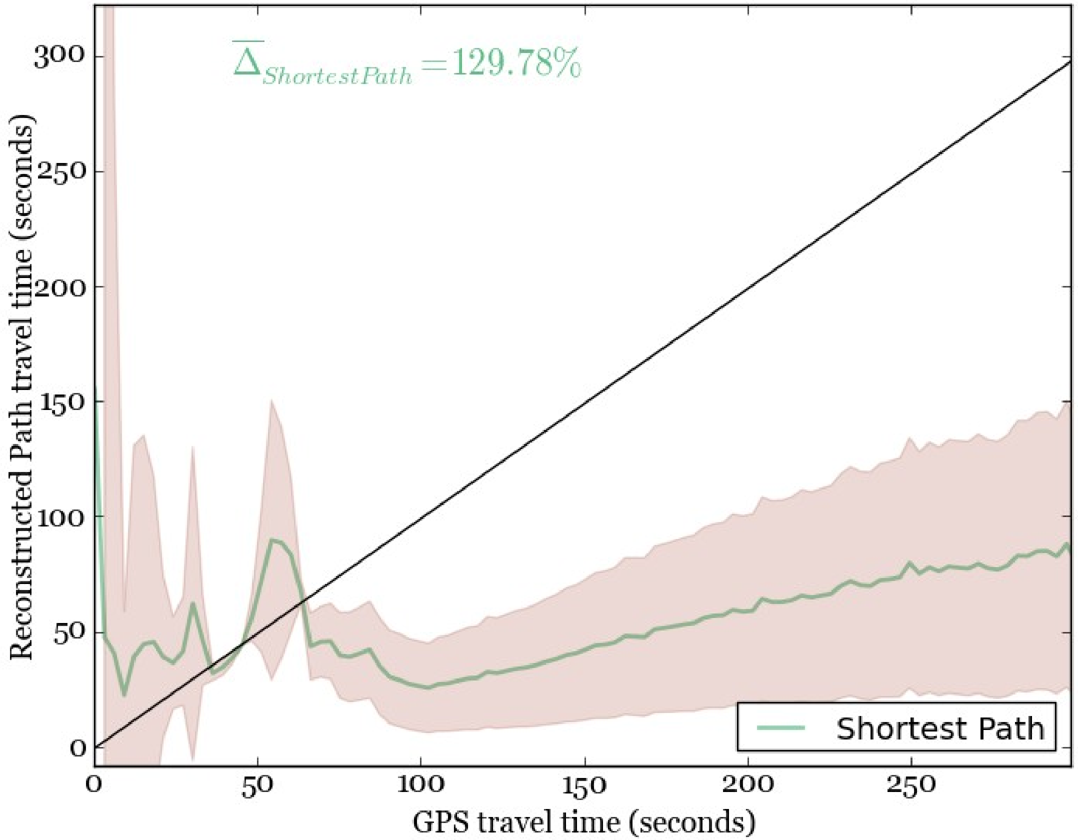


Result

- Trajectory → sequence of road segments



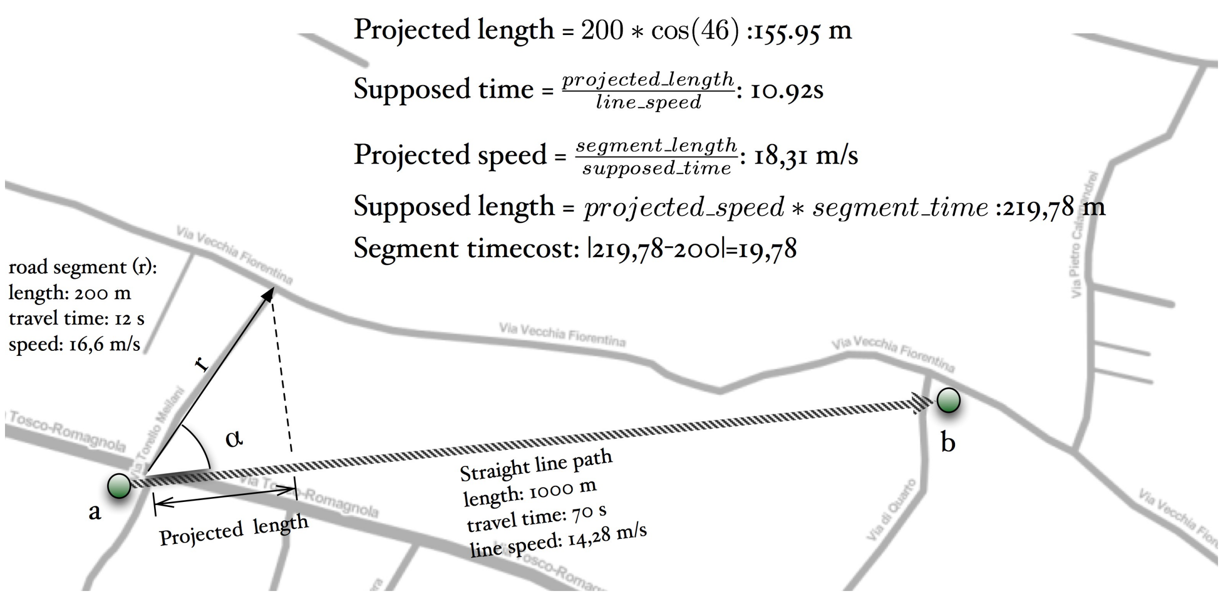
Shortest path vs real GPS time



Matching GPS data with shortest path leads to significant differences w.r.t. real GPS travel time between two points

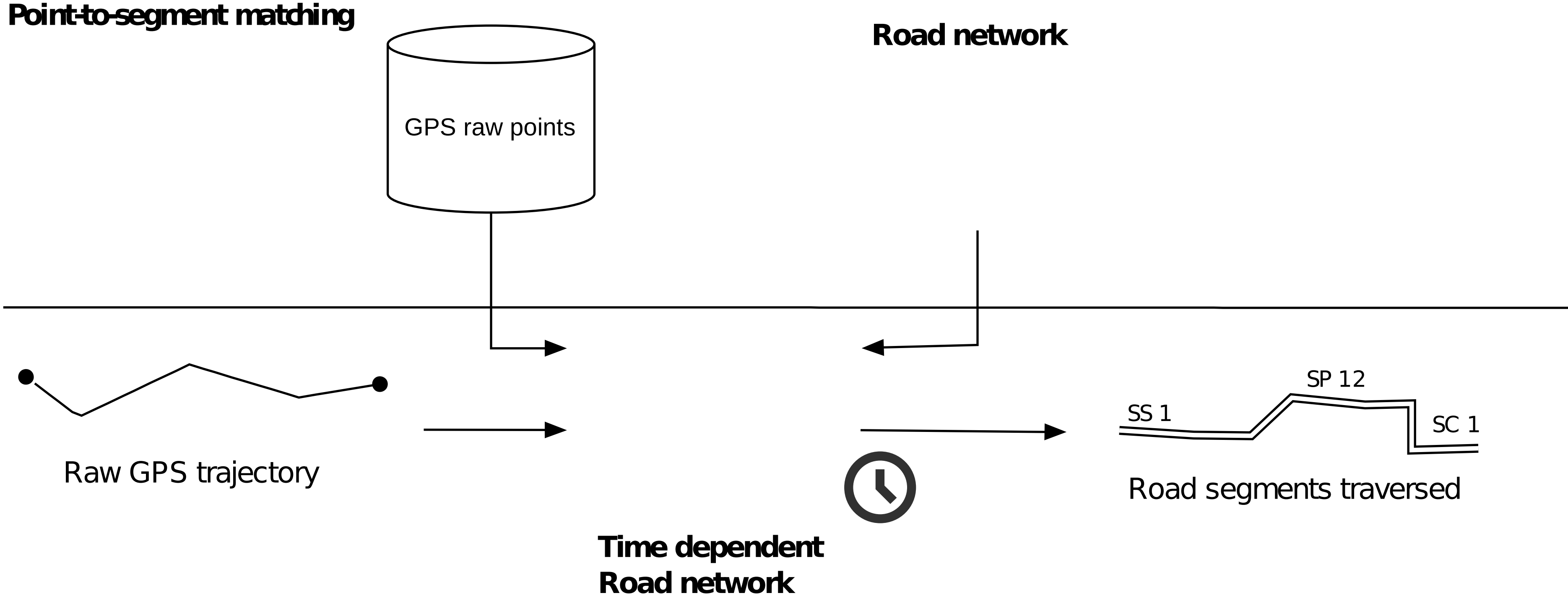
Alternative, Time-Aware approach

- Given a road network with travel times for each edge, find the path that best fits given total travel time
- Satisfy some basic constraints, e.g. no useless turnarounds



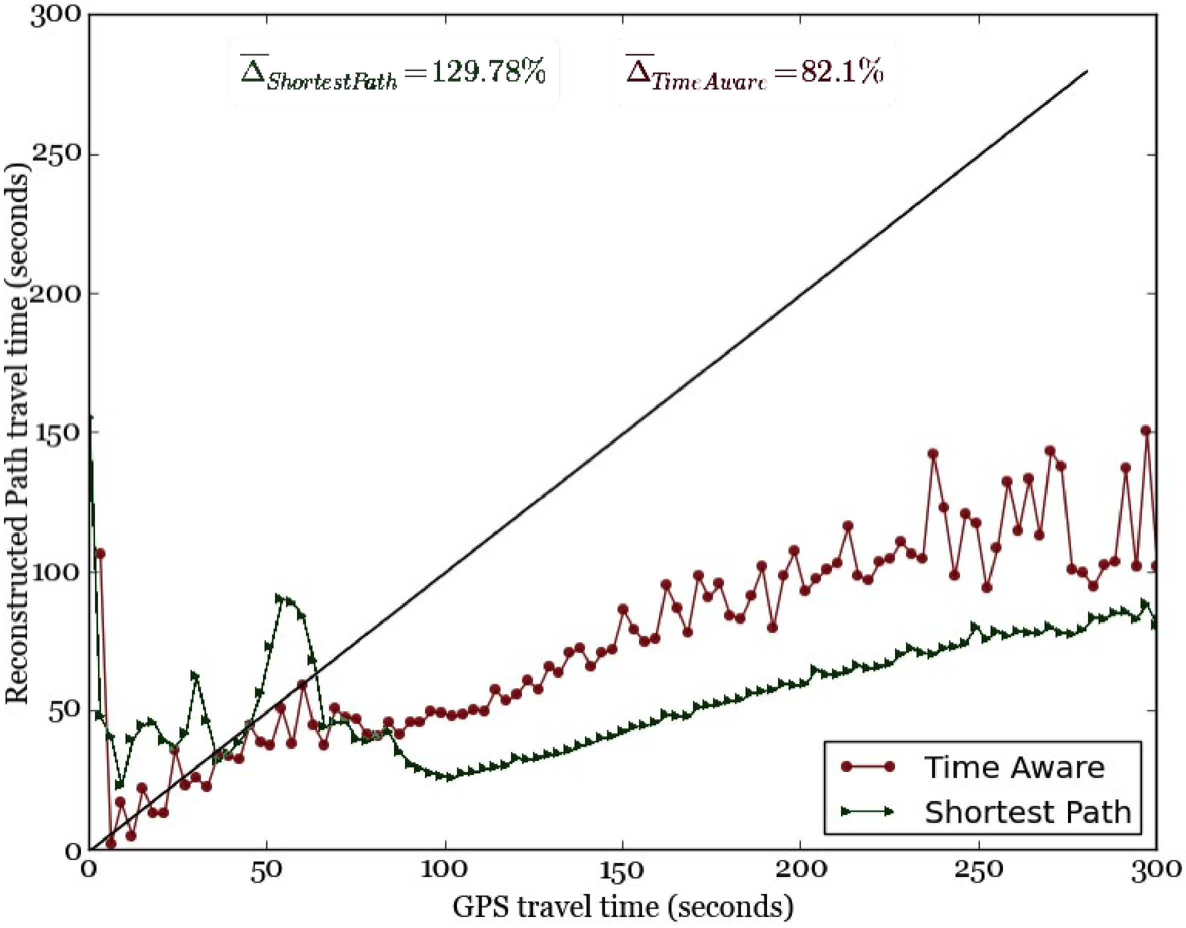
Workflow

Point-to-segment matching

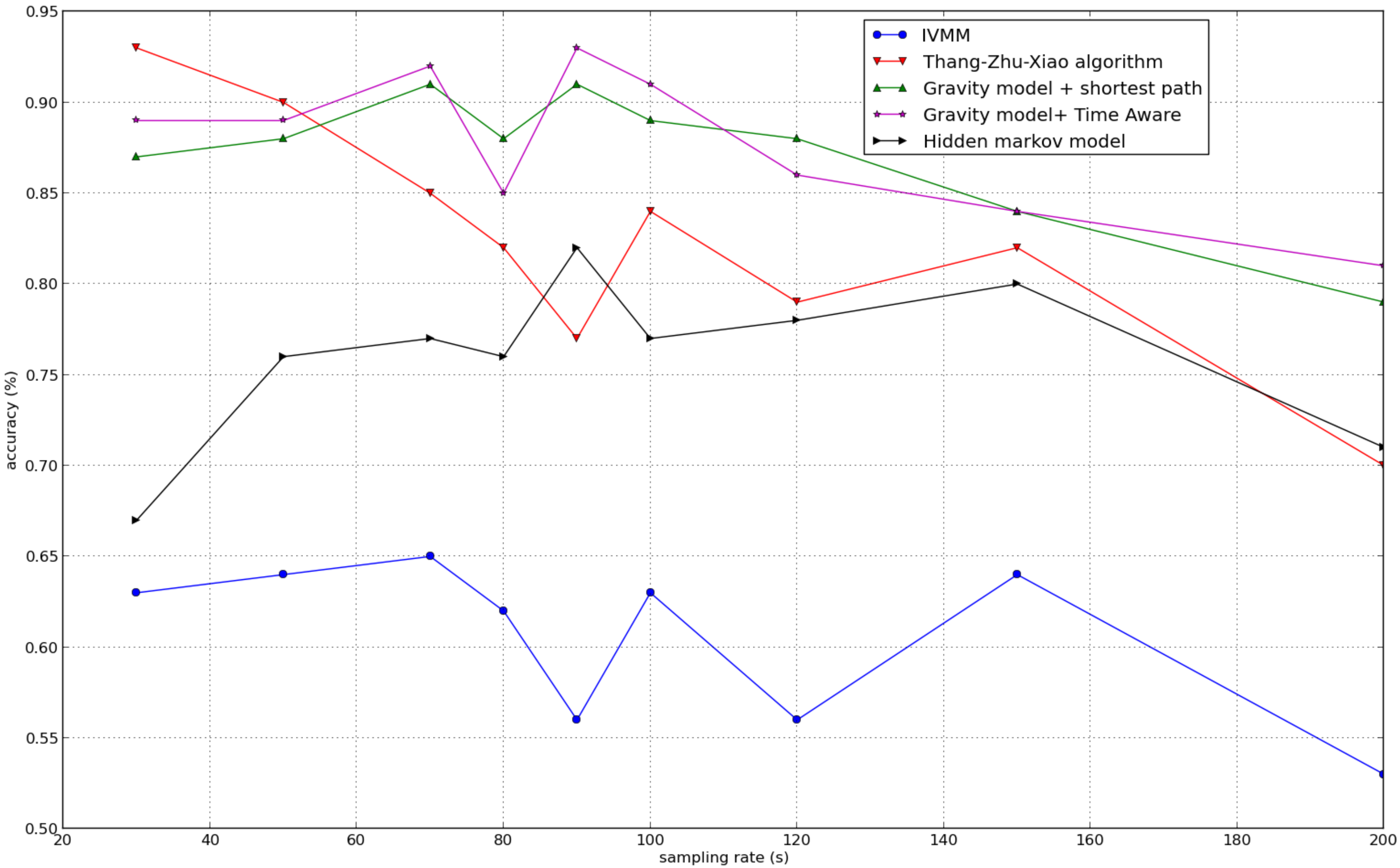


Time-Aware map matching

Finding the travel time



Effectiveness



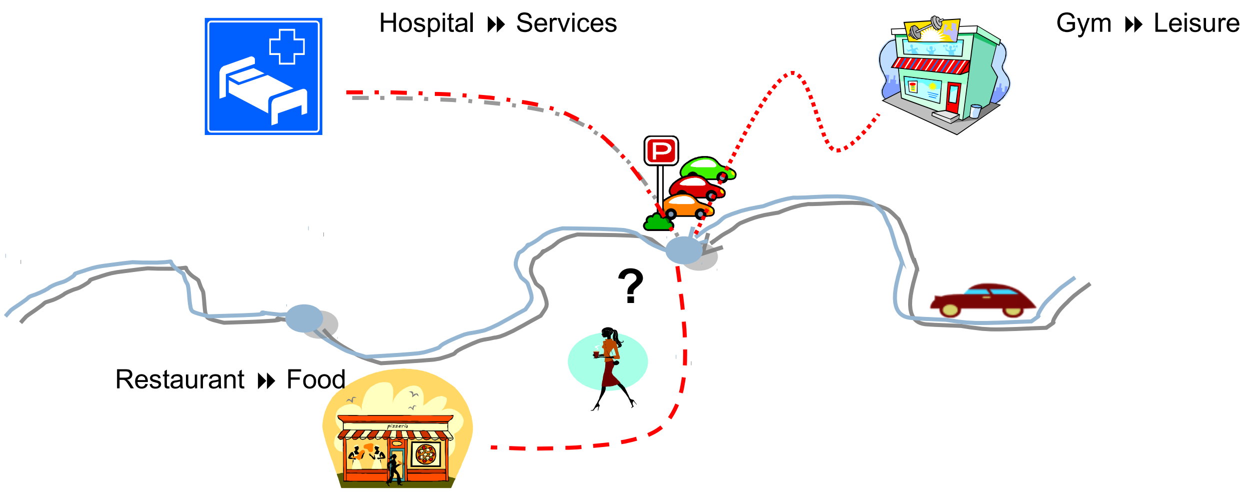
Activity Recognition

Objective

- Infer the purpose and/or activity performed of trips and locations
- Two approaches
  - Consider what kind of activities can be performed in that area
  - Consider how the user behaves (when and how he reaches the are, etc.)

Recognition through Points-of-Interest

Given a dataset of GPS tracks of private vehicles, annotate trajectories with the most probable activities performed by the user.



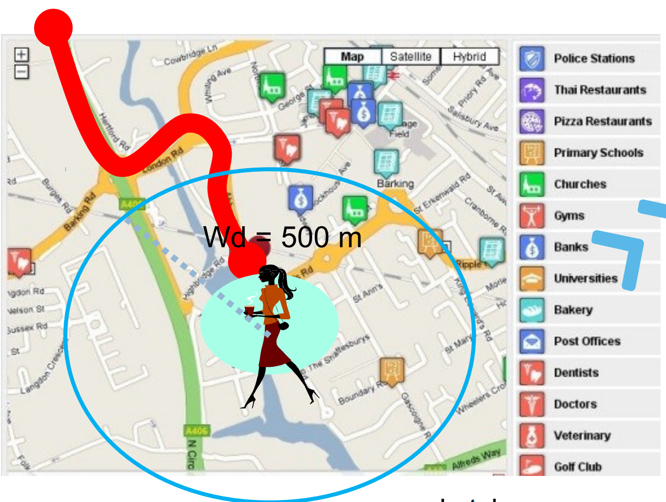
Associates the list of possible <u>POIs</u> (with corresponding probabilities) visited by a user moving by car when he stops.

A mapping between POIs categories and Transportation Engineering activities is necessary.

The enrichment process

- **POI collection**: Collected in an automatic way, e.g. from Google Places.
- **Association POI – Activity**: Each POI is associated to a ``activity''. For example Restaurant → Eating/Food, Library → Education, etc.
- **Basic elements/characteristics:**
  - C(POI) = {category, opening hour, location}
  - C(Trajectory) = {stop duration, stop location, time of the day}
  - C(User) = {max walking distance}
- **Computation of the probability to visit a POI/ to make an activity:** For each POI, the probability of ``being visited'' is a function of the POI, the trajectory and the user features.
- **Annotated trajectory**: The list of possible activities is then associated to a Stop based on the corresponding probability of visiting POIs

Input & Output



Lat; Lon
TimeStamp: Sun 10:55 – 12:05

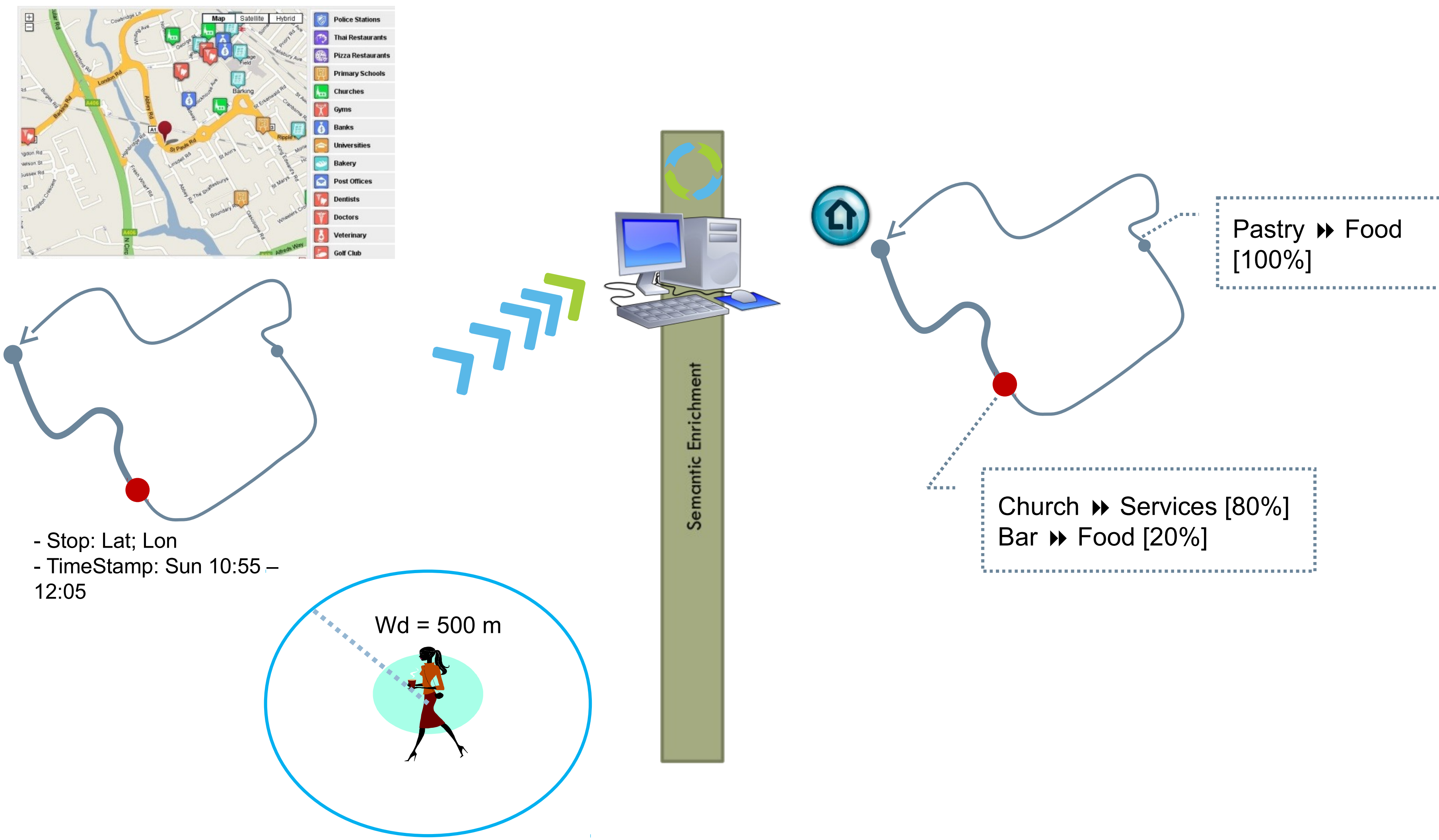
Bank: Mon – Fri [8:00 – 15:30]

Dentist: Mon – Sat [9:00 – 13:00] [15:30 – 18:00]

Church: Mon – Sat [18:00 – 19:00]
Sun [11:00 – 12:00]

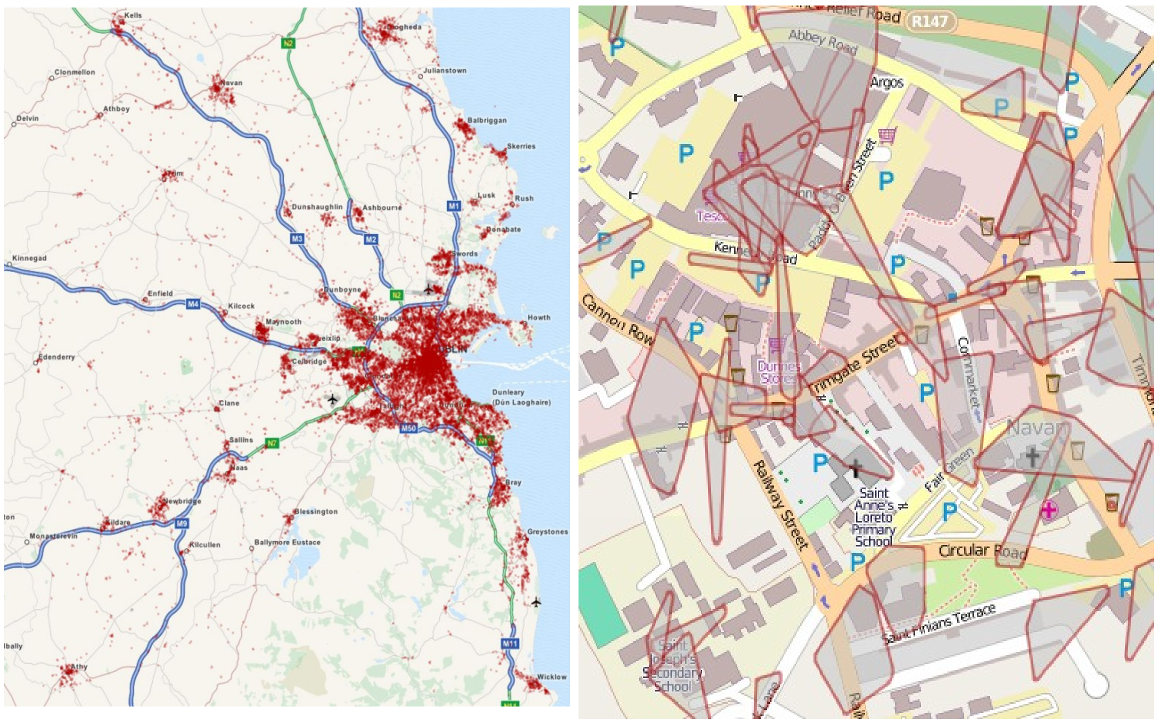
Primary School: Mon – Sat [8:00 – 13:00]

Input & Output



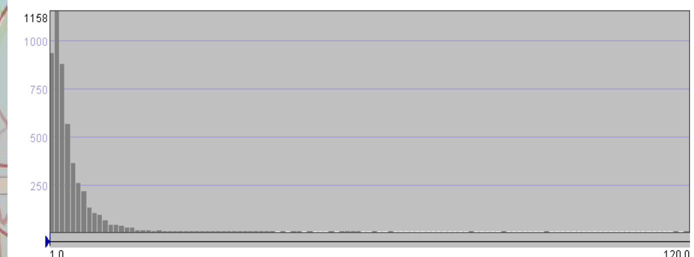
Extraction of personal places from Twitter trajectories in Dublin area

The points of each trajectory taken separately were grouped into spatial clusters of maximal radius 150m. For groups with at least 5 points, convex hulls have been built and spatial buffers of small width (5m) around them.
1,461,582 points belong to the clusters (89% of 1,637,346); 24,935 personal places have been extracted.



Examples of extracted places

N	N?	min	q1	med	q3	max	ave	stdd
5180	0	1.0	2.0	3.0	5.0	120.0	4.8	6.3



Statistical distribution of the number of places per person

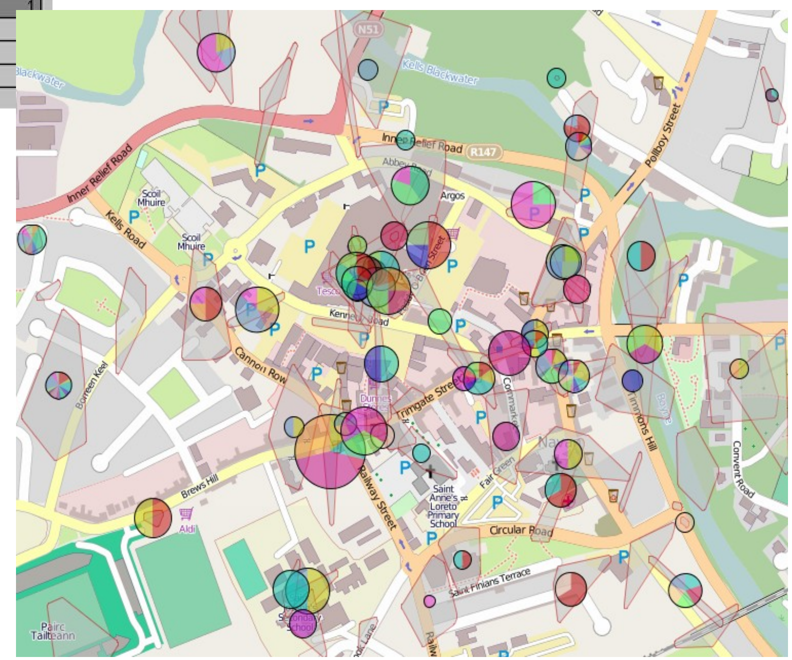
Recognition of the message topics, generation of topical feature vectors, and summarization by the personal places

Topics have been assigned to 208,391 messages (14.3% of the 1,461,582 points belonging to the personal places)

Message	Features	topic=family: Occurrences of topic	topic=home: Occurrences of topic	topic=education: Occurrences of topic	topic=work: Occurrences of topic
@joe_lennon I usually	education	0	0	1	0
@joe_lennon together	education	0	0	1	0
@jas_103 deadly; don	work	0	0	0	1
Just got home and see	home	0	1	0	0
So excited about my ne	sweets	0	0	0	0
@lamtcdizzy I haven't b	shopping	0	0	0	0
Get in from my night ou	family;home;work	1	1	0	1
Home again at 6pm! N	home	0	1	0	
Bussing it home for th			1	0	
Ah shite. It's been a p			0	0	
@ronanhutchinson be	education	0	0	1	

Get in from my night out; my dad gets home from work two minutes later. Great timing :)

...

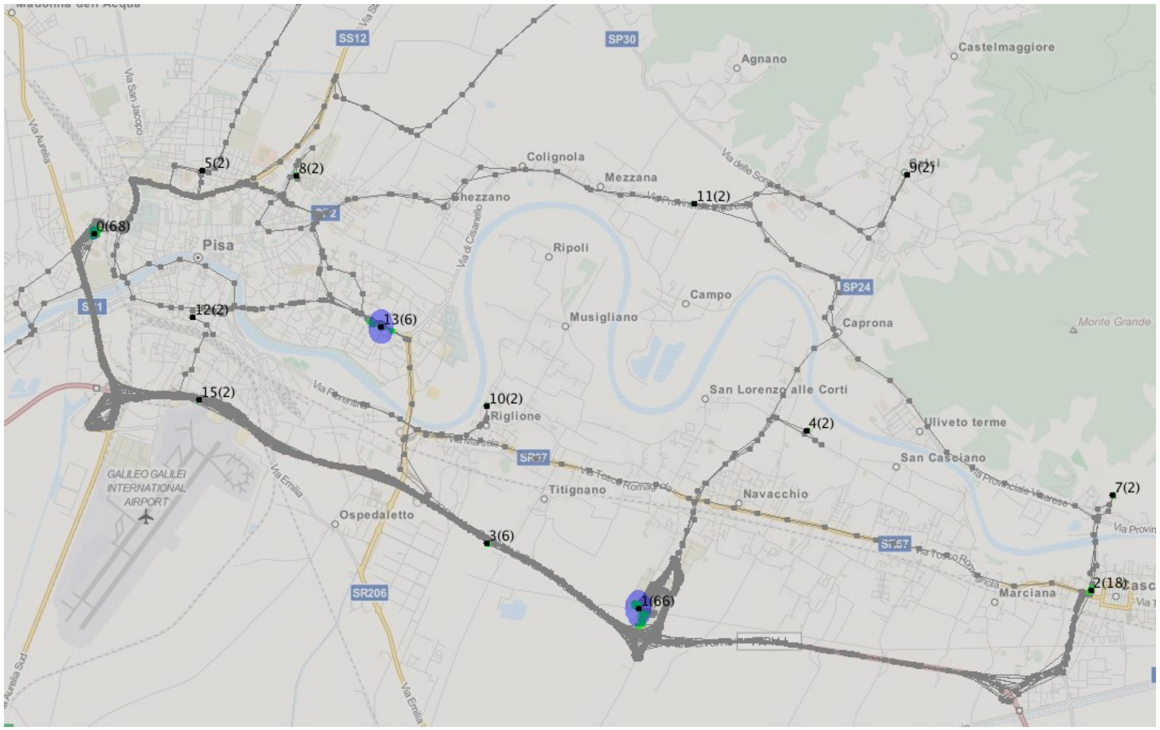


1) Some places did not get topic summaries (about 20% of the places)
2) In many places the topics are very much mixed
3) The topics are not necessarily representative of the place type (e.g., topics near a supermarket: family, education, work, cafe, shopping, services, health care, friends, game, private event, food, sweets, coffee)

Activity Recognition

Individual Mobility Networks

How to synthesize Individual Mobility?



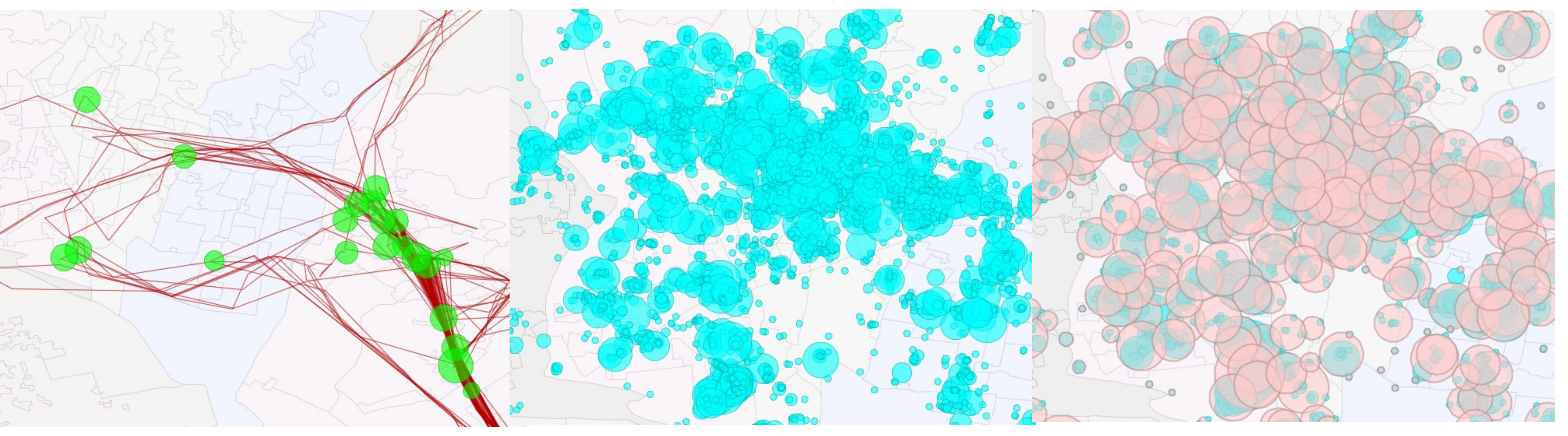
Mobility Data Mining methods automatically extract relevant episodes: locations and movements.

How to synthesize Individual Mobility?

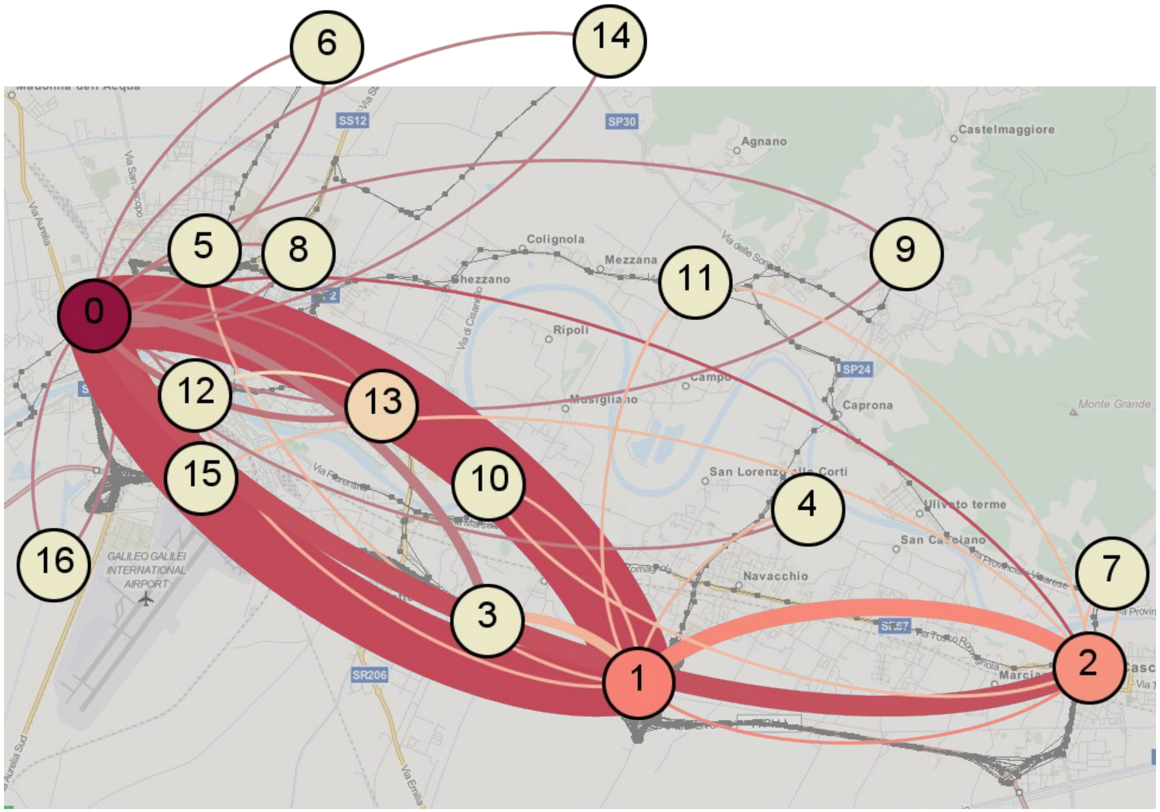
- Basic approach: compute movement features of each trip
  - Length
  - Average speed or Duration
  - Bee-line length
  - Time of the day
  - ...

How to synthesize Individual Mobility?

- More advanced approach: consider overall mobility of the user
- First step: rank individual preferred locations

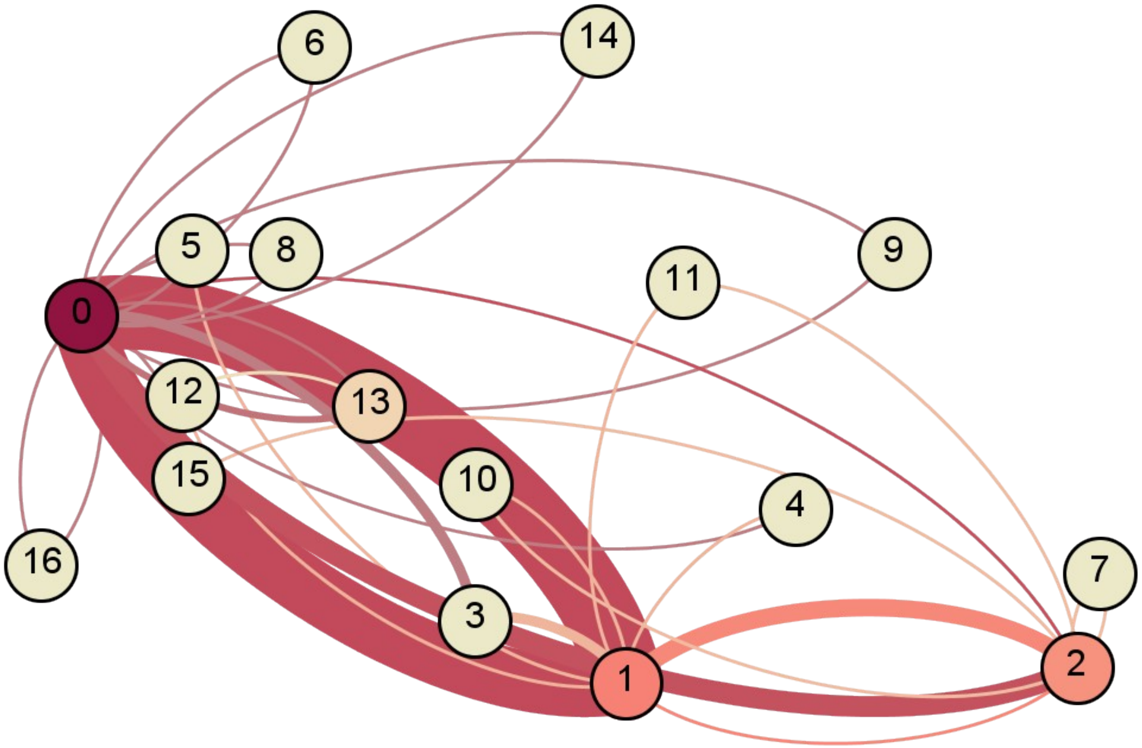


How to synthesize Individual Mobility?



Graph abstraction based on locations (nodes) and movements (edges)

How to synthesize Individual Mobility?

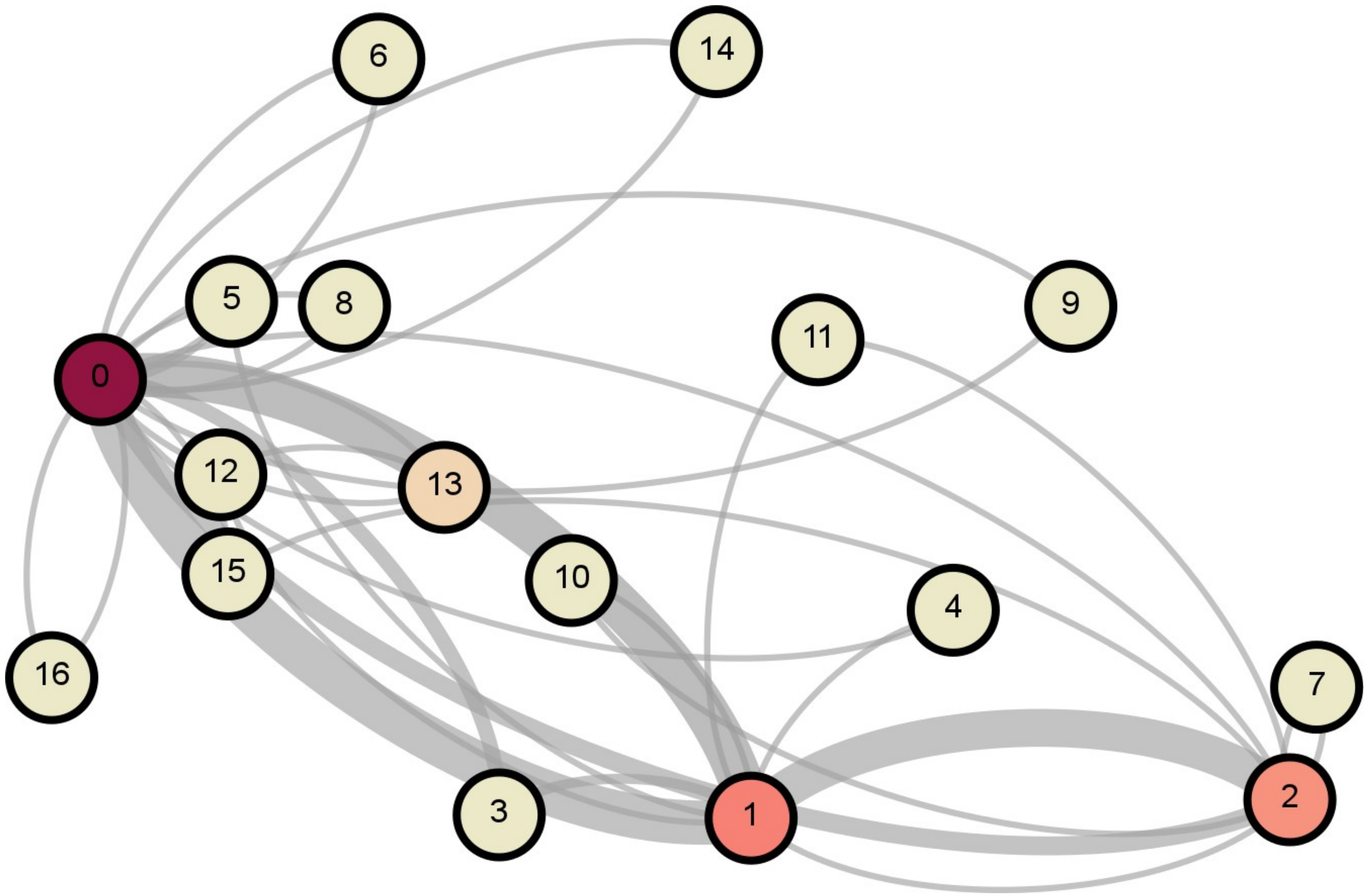


High level representation

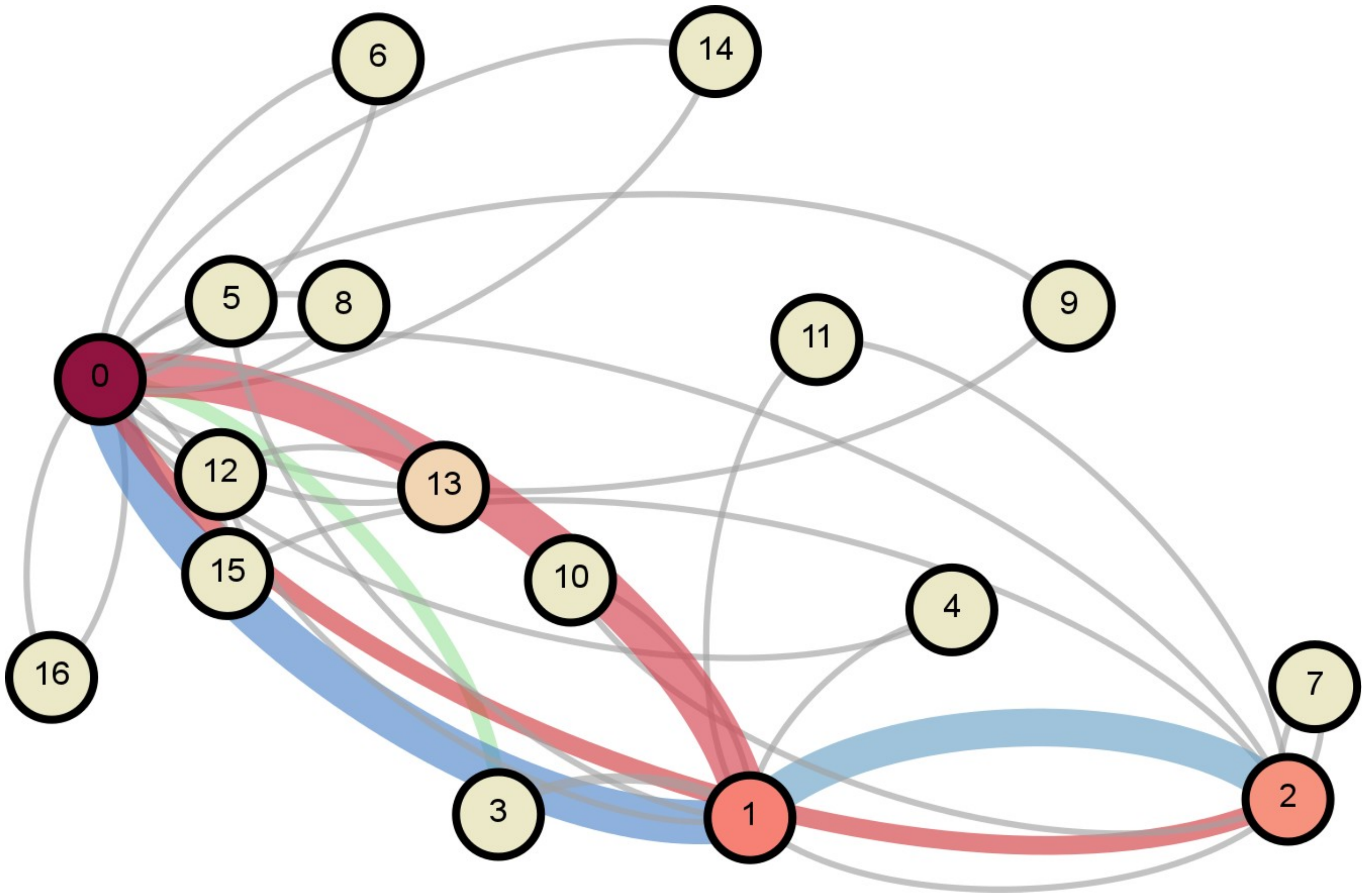
Aggregation of sensitive data

Abstraction from real geography

From raw movement…



… to annotated data



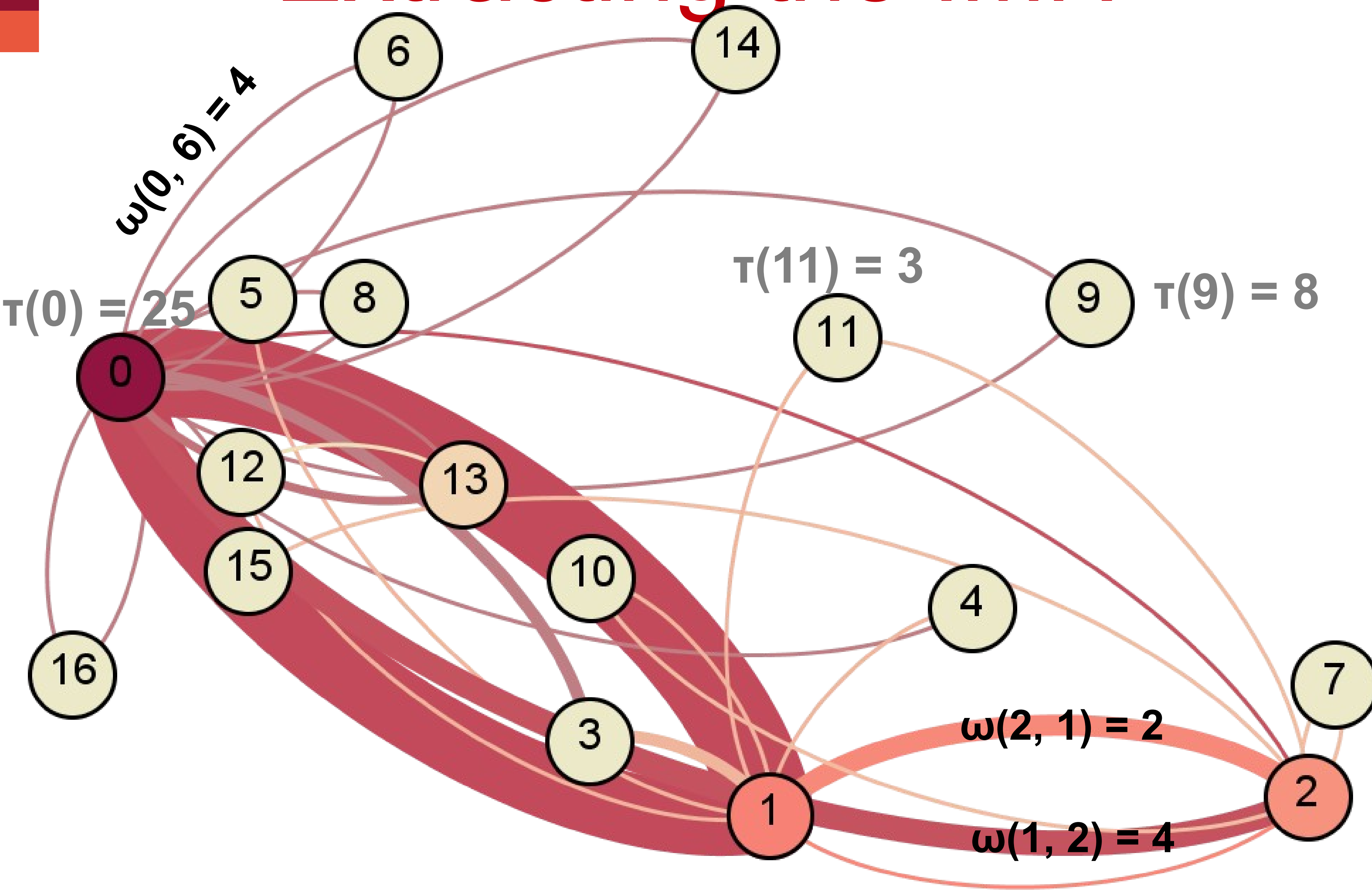
The ABC classifier

1) Build from data an
   Individual Mobility Network (IMN)

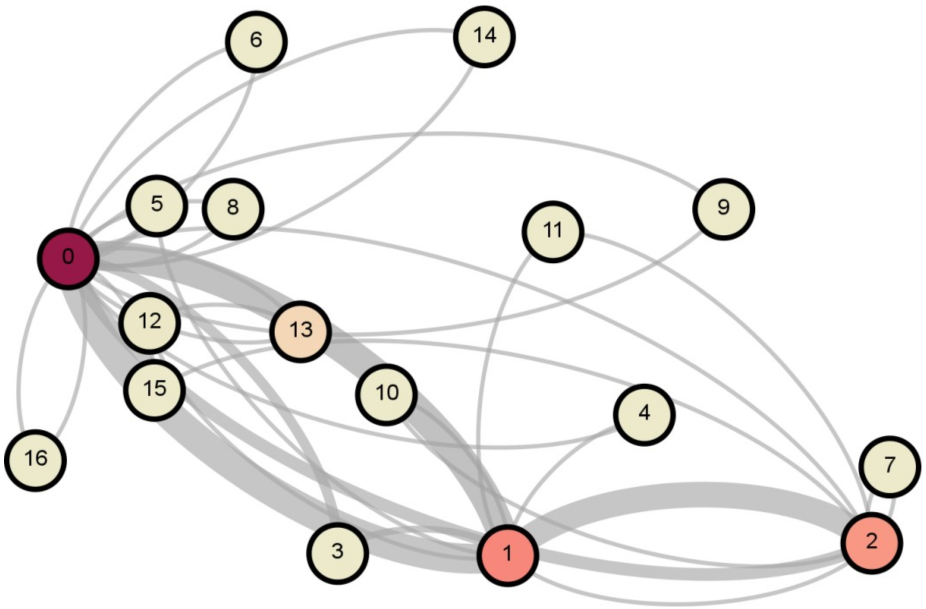
2) Extract structural features from the IMN

3) Use a cascading classification with label propagation (ABC classifier)

Extracting the IMN



Extracting the IMN



Trip Features

Length

Duration

Time Interval

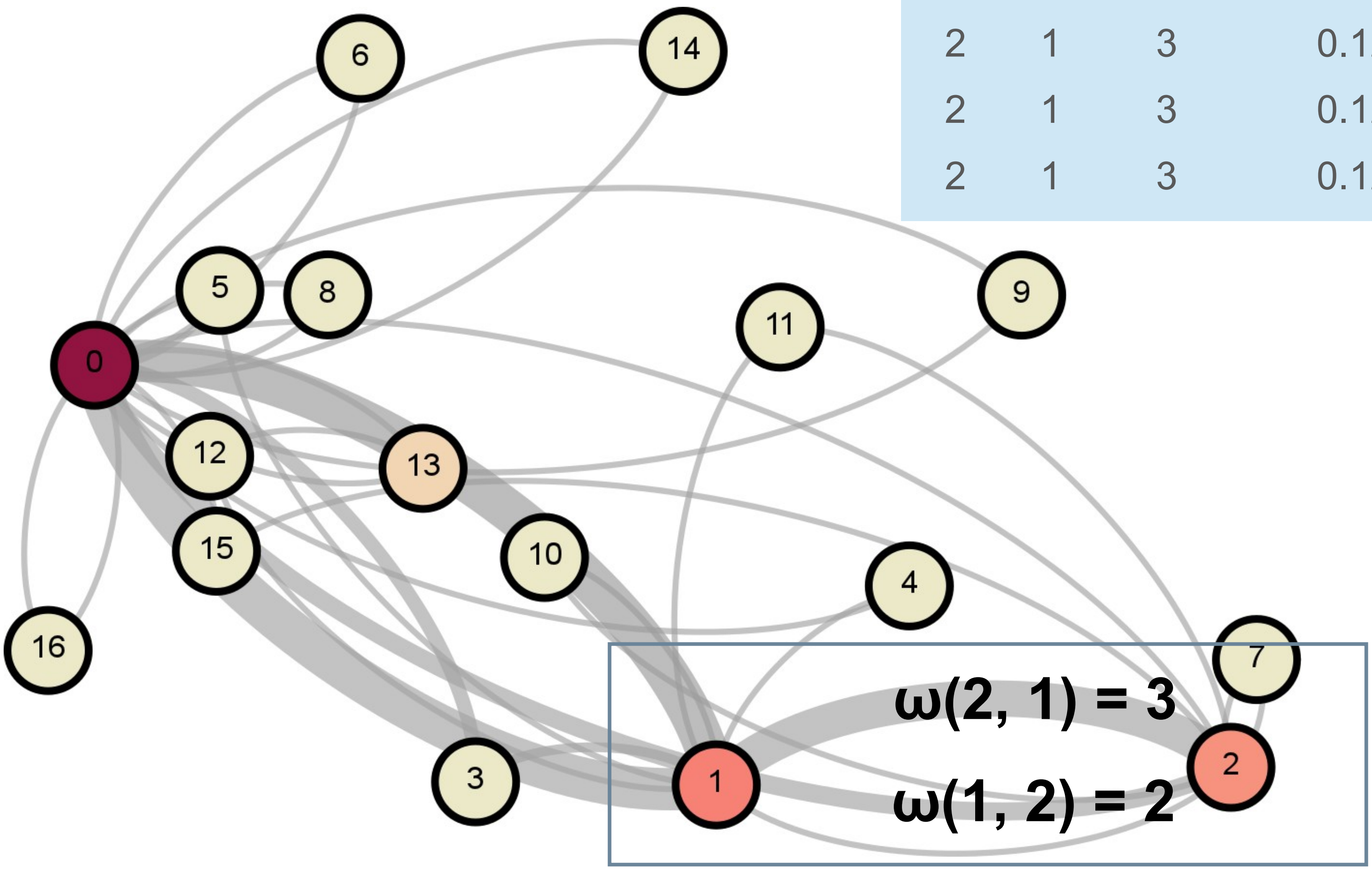
Average Speed

Network Features

centrality	clustering coefficient average path length
predictability	entropy
hubbiness	degree betweenness
volume	edge weight flow per location

Extracting the IMN

from	to	weight	ccFrom	ccTo	duration
1	2	2	0.22	0.12	10 min
1	2	2	0.22	0.12	5 min
2	1	3	0.12	0.22	4 min
2	1	3	0.12	0.22	6 min
2	1	3	0.12	0.22	4 min



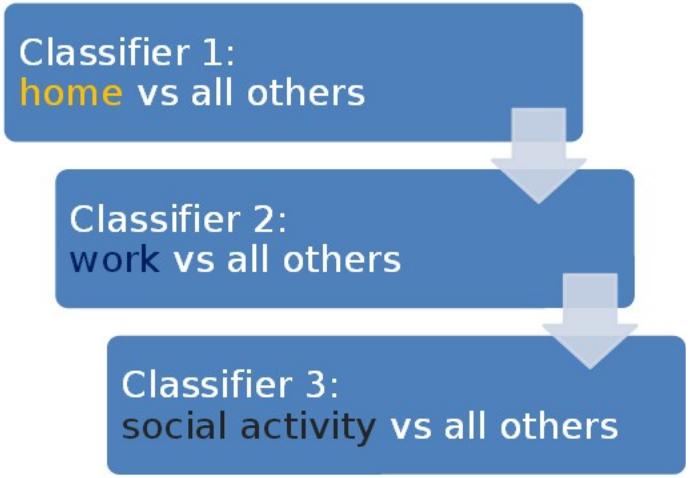
ABC Classifier

- Principles:
  - The activities of a user should be predicted as a whole, not separately
  - Some activities are easy to classify
  - Other activities might benefit from contextual information obtained from previous predictions
- E.g.: a place frequently visited after work might be more likely to be leisure / shopping

ABC Classifier

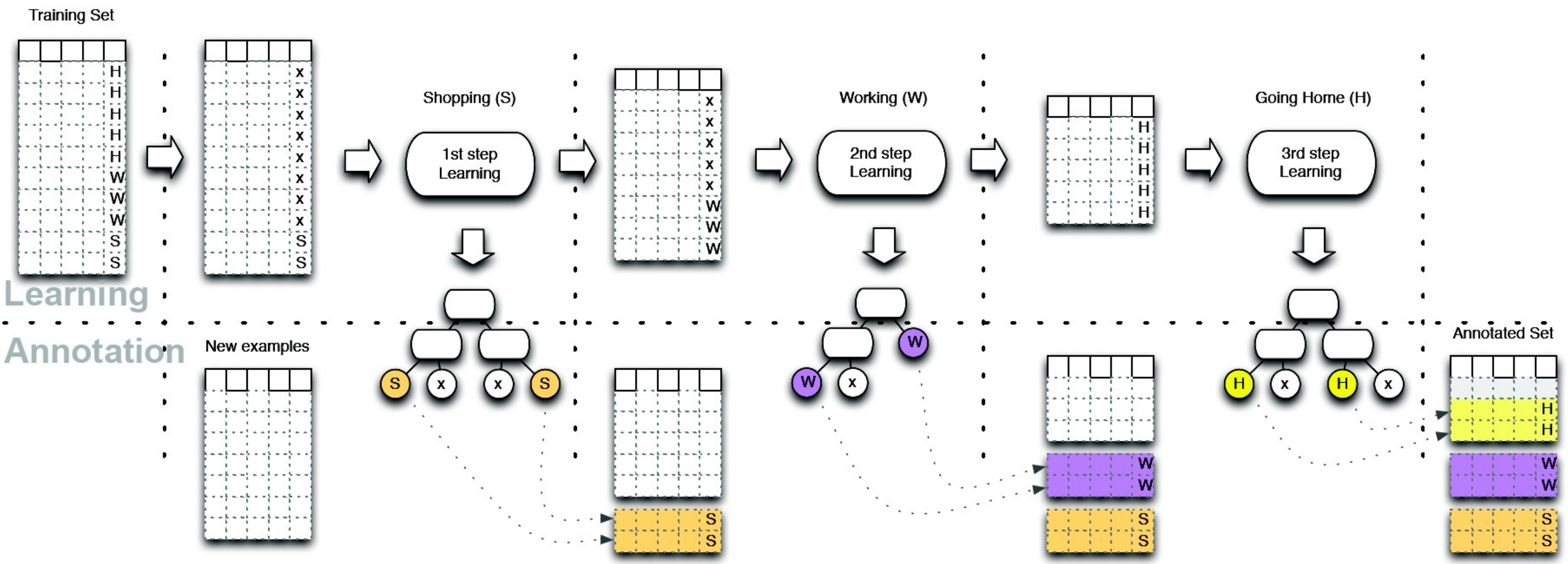
- Reduce the multi-class problem into several binary problems
- The binary classifiers are learnt in cascade
- The classification results of each step are used as source for later classifications

Example



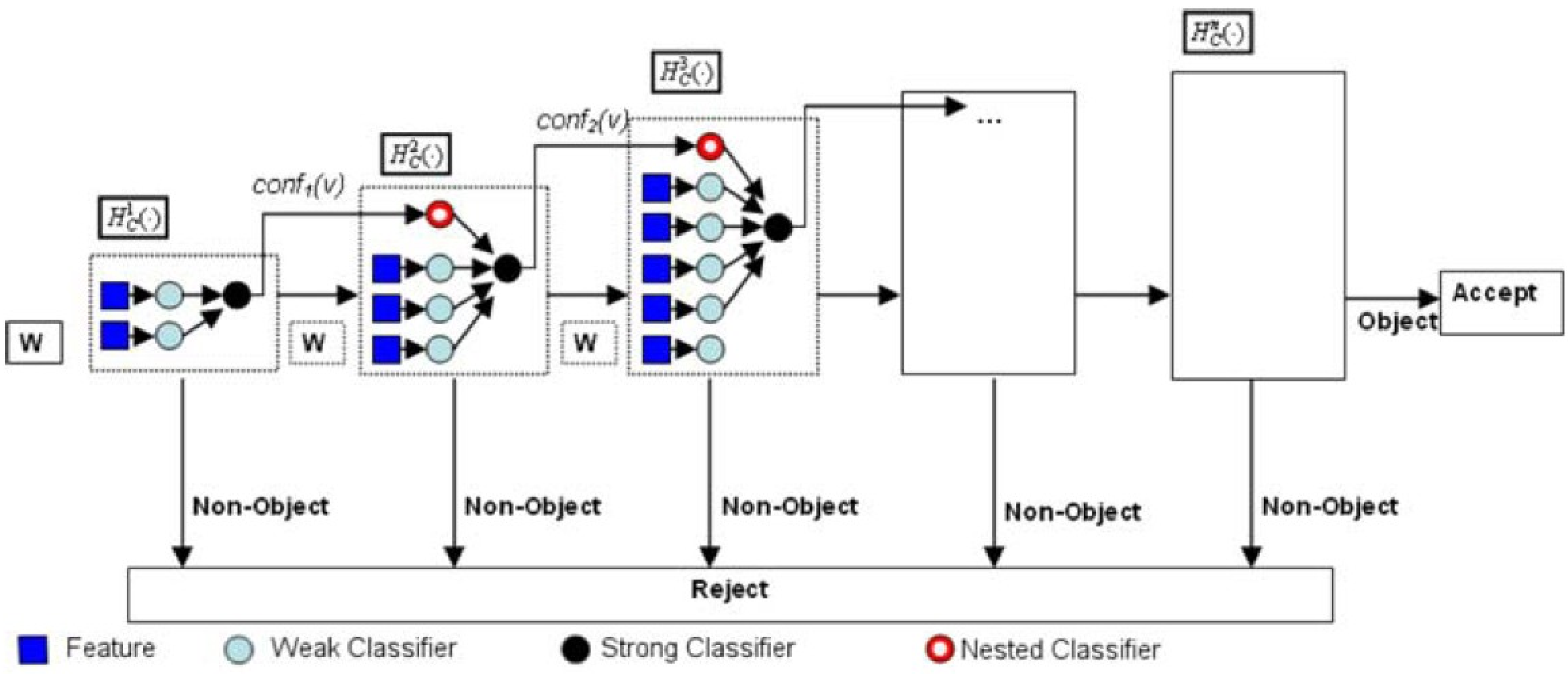
ABC Classifier

- Inspired by Nested Cascade Classification



ABC Classifier

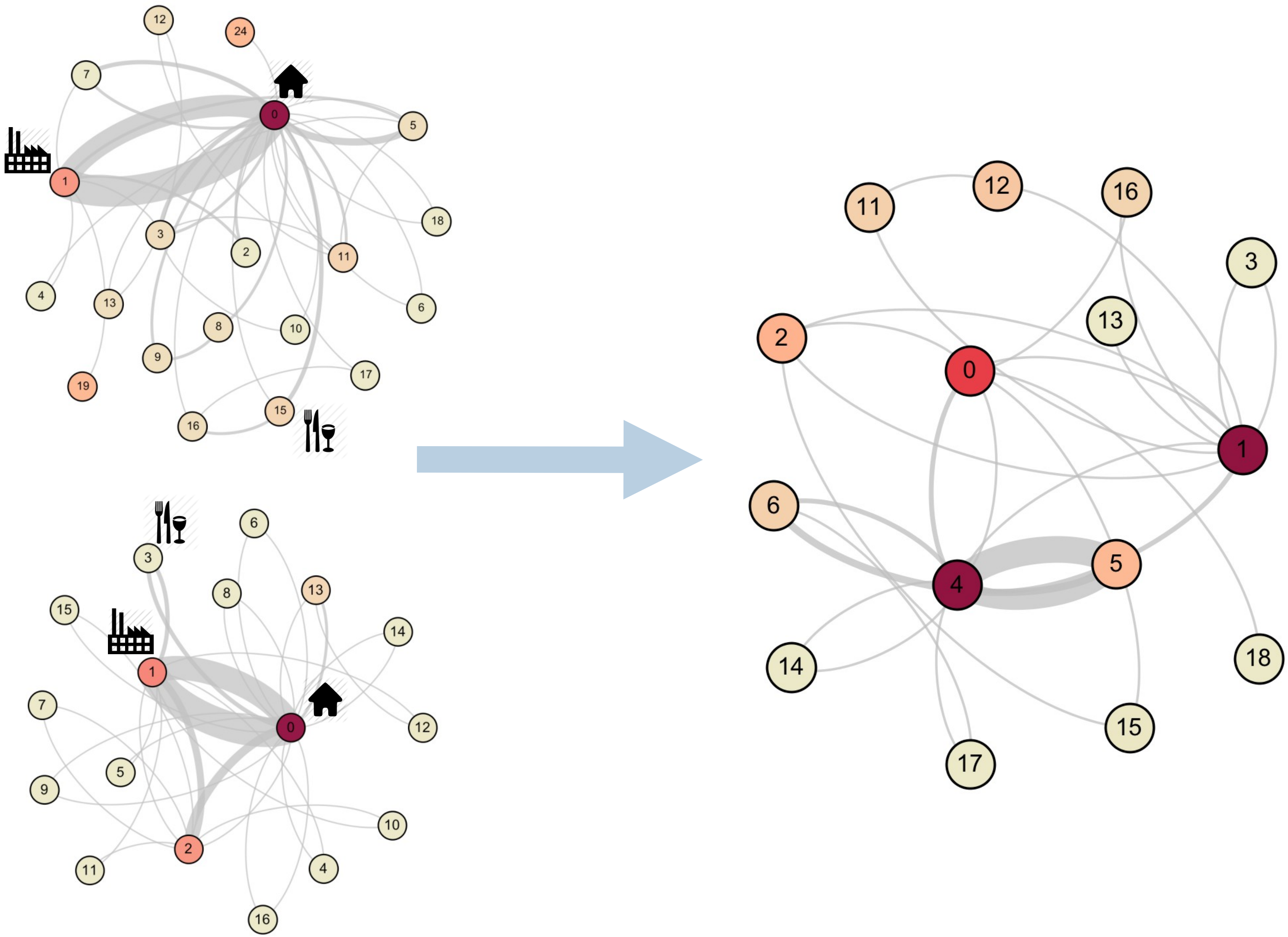
- Inspired by Nested Cascade Classification



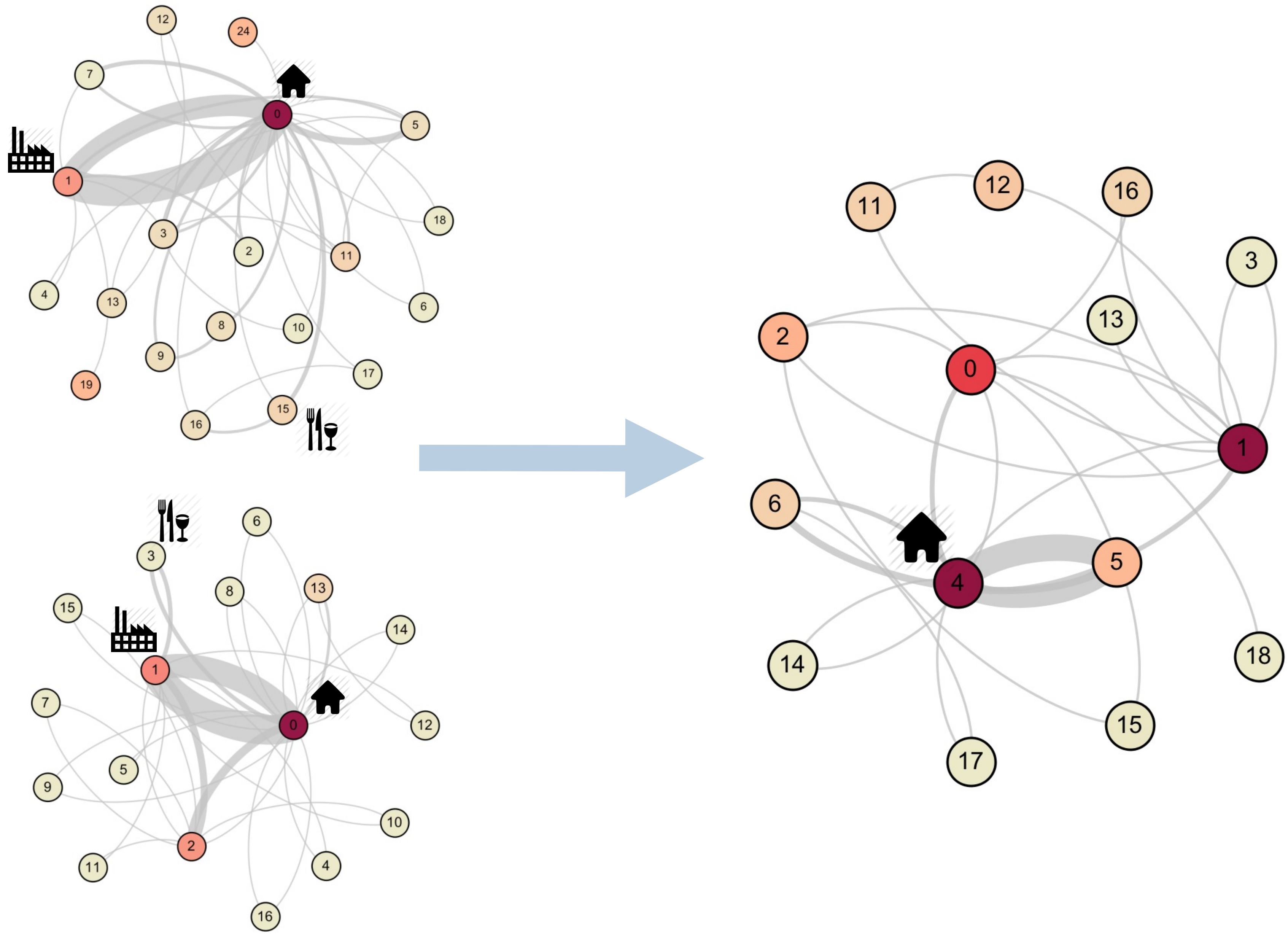
ABC Classifier

- After recognizing an activity (e.g. work), we use this information to enrich the features of the yet-unclassified trips
- E.g. add a feature describing whether the remaining trips are adjacent to the previous activity
  - Are there direct trips from work to the new place?

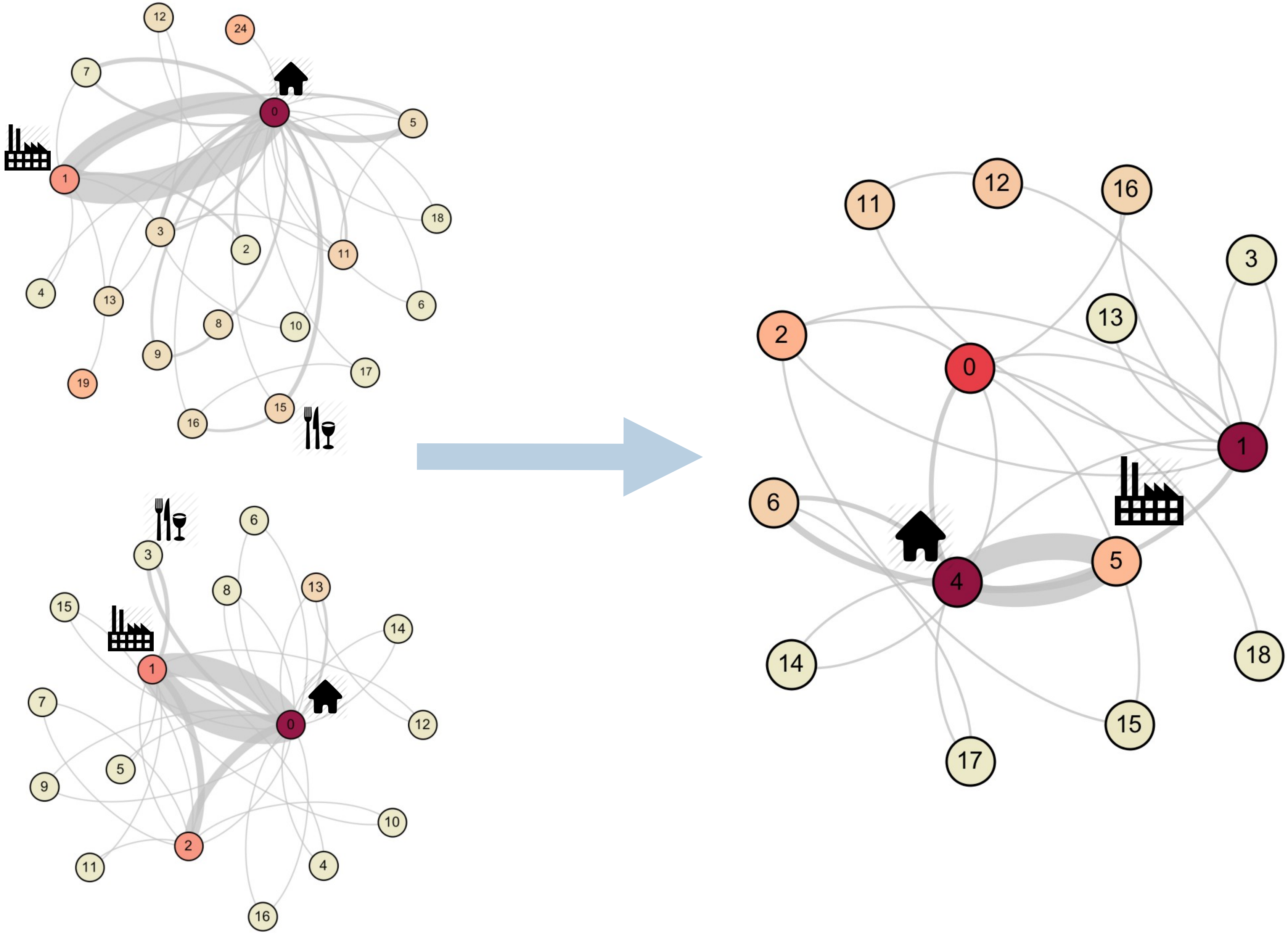
The ABC classifier



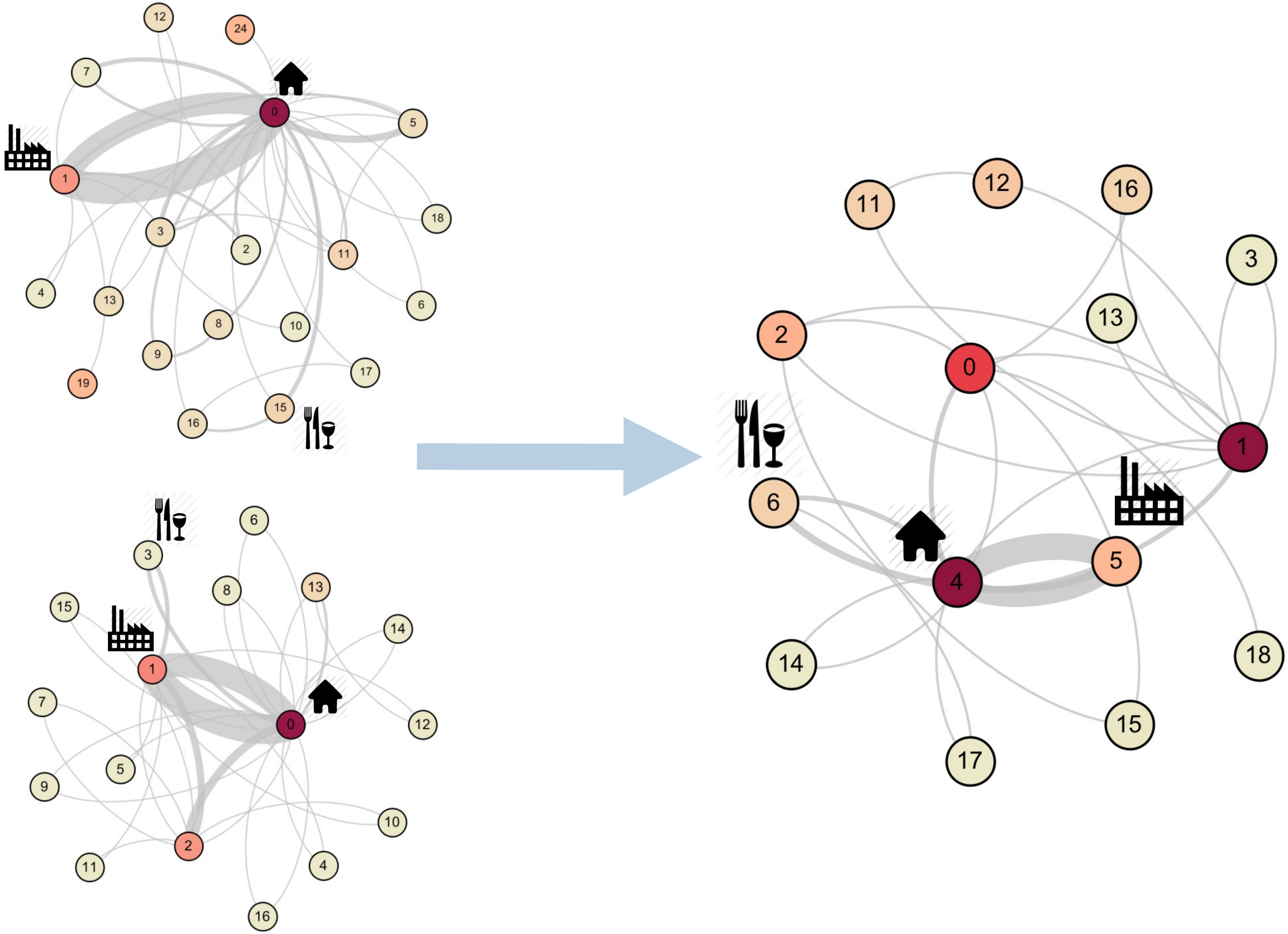
The ABC classifier



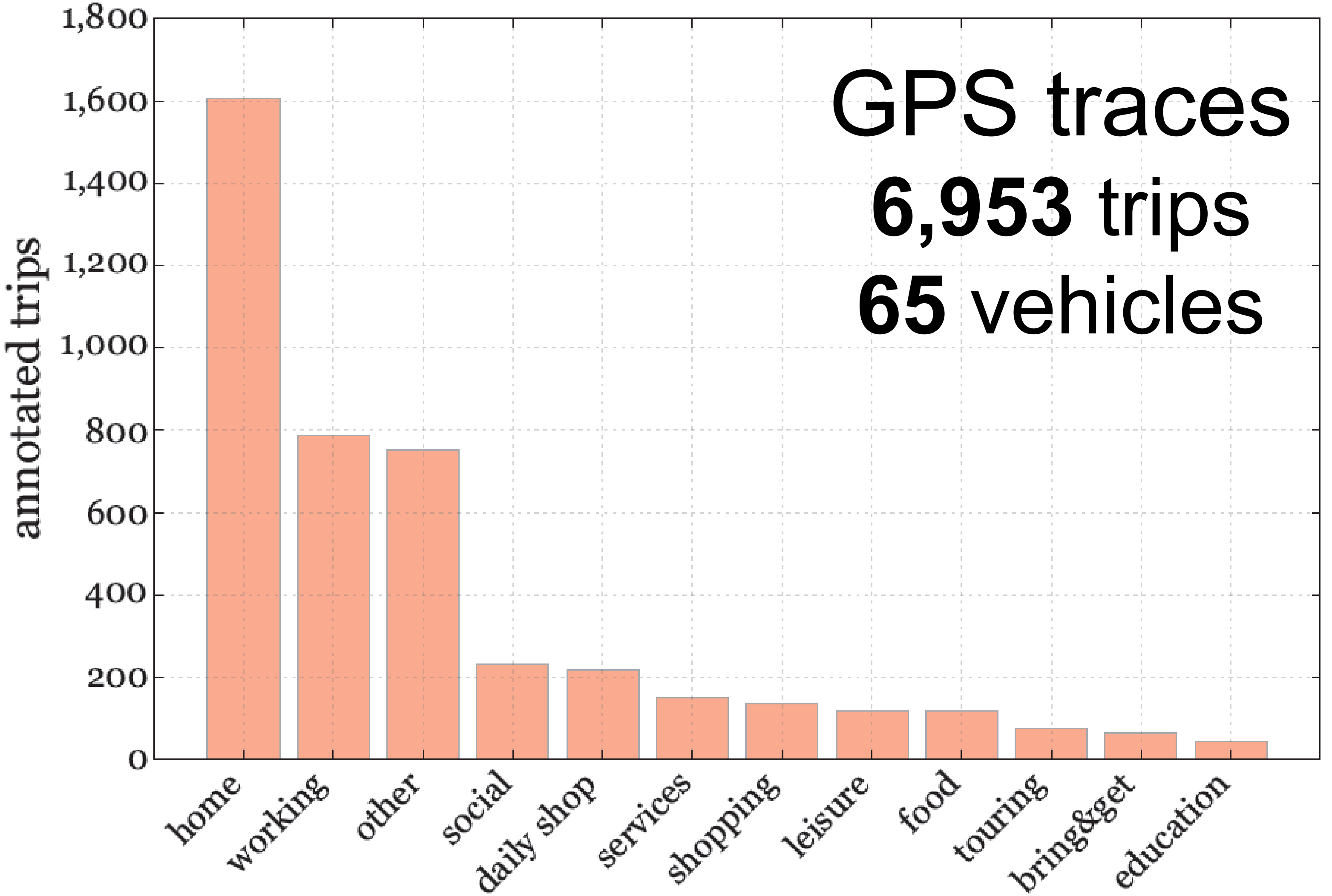
The ABC classifier



The ABC classifier



Experiments



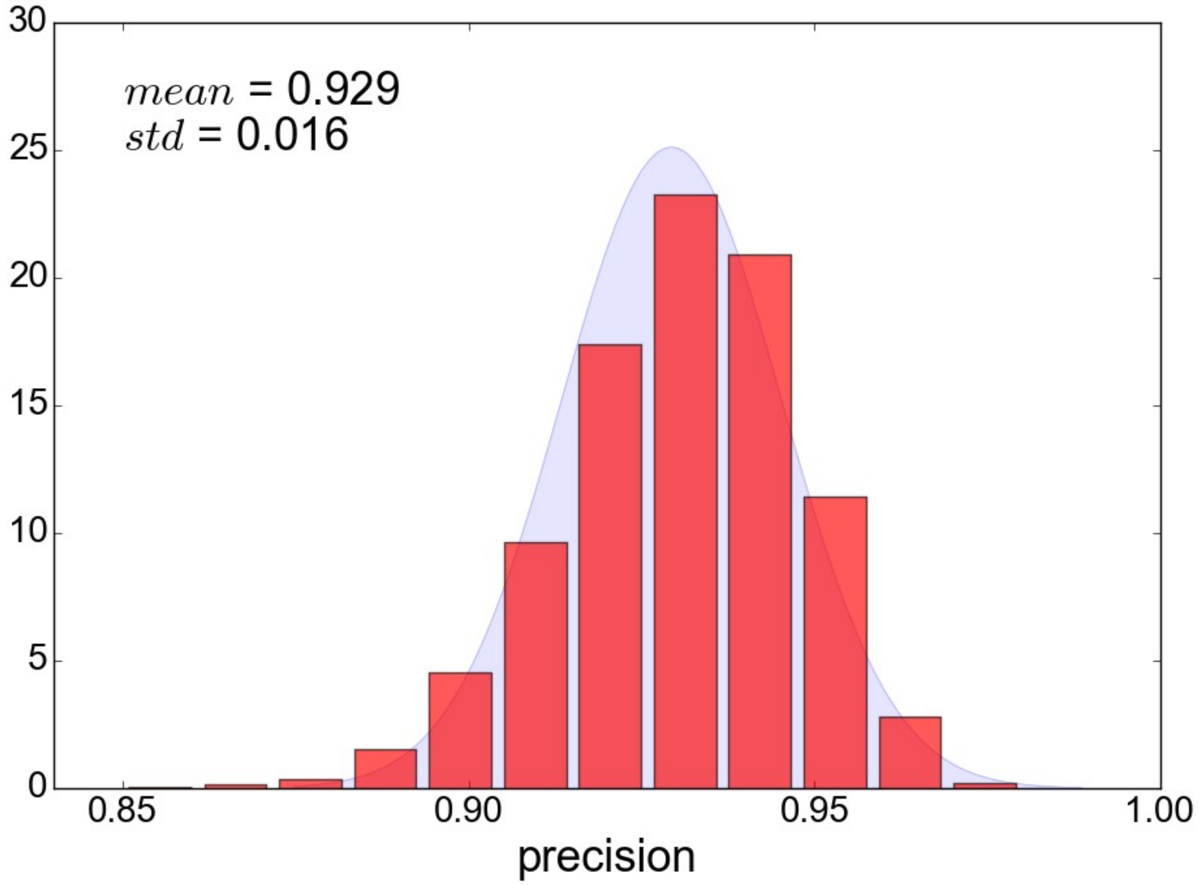
Experiments

RF classifier			activity	ABC classifier			
precision	recall	f1-score		precision	recall	f1-score	support
0.91	0.93	0.92	Going Home	0.99	0.99	0.99	930
0.00	0.00	0.00	Bring and get	1.00	0.68	0.81	22
0.00	0.00	0.00	Education	1.00	0.25	0.40	4
0.15	0.14	0.14	Daily shopping	0.97	0.85	0.91	68
0.50	0.37	0.43	Working	0.93	0.96	0.94	258
0.34	0.6	0.43	Other	0.90	0.98	0.94	384
0.09	0.05	0.06	Shopping	0.87	0.85	0.86	39
0.13	0.09	0.11	Leisure	0.84	0.84	0.84	51
0.24	0.11	0.15	Services	0.83	0.71	0.77	42
0.04	0.01	0.02	Touring	0.77	0.83	0.80	12
0.06	0.06	0.06	Food	0.76	0.49	0.60	53
0.08	0.04	0.05	Social activities	0.65	0.69	0.67	49
0,54	**0,54**	**0,54**	avg / total	**0.94**	**0.94**	**0.94**	1912

0.54 **0.94**

Experiments

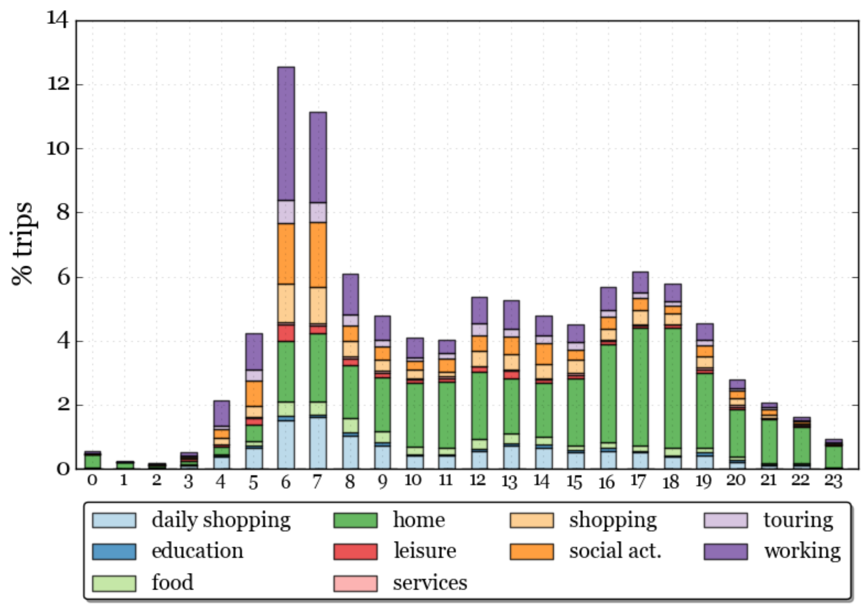
- Is the order of activities in the learning relevant?



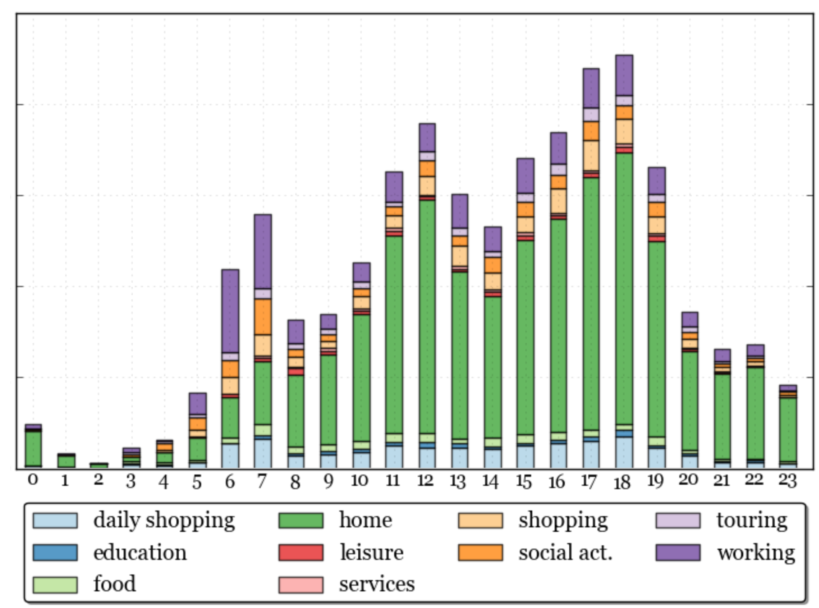
Semantic Mobility Analytics

Temporal Analysis

- Pisa traffic



In

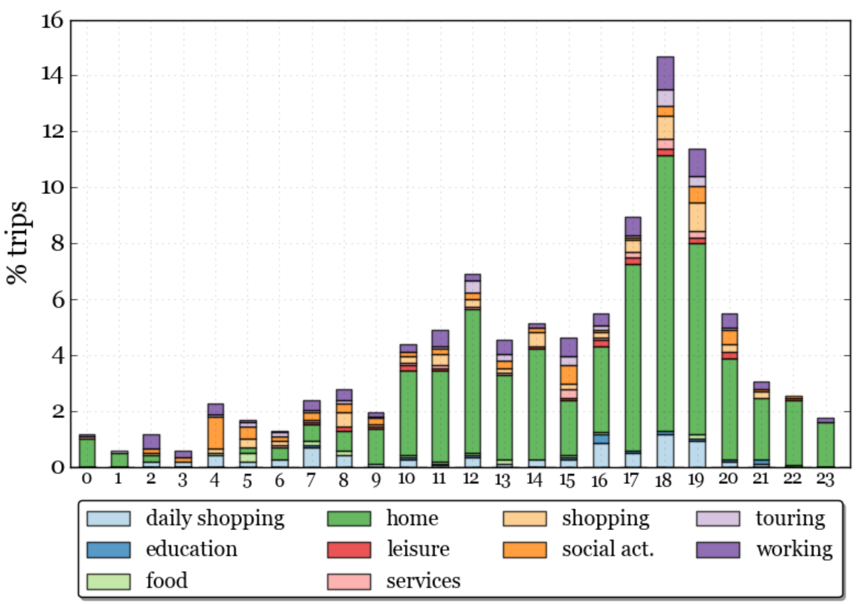


Out

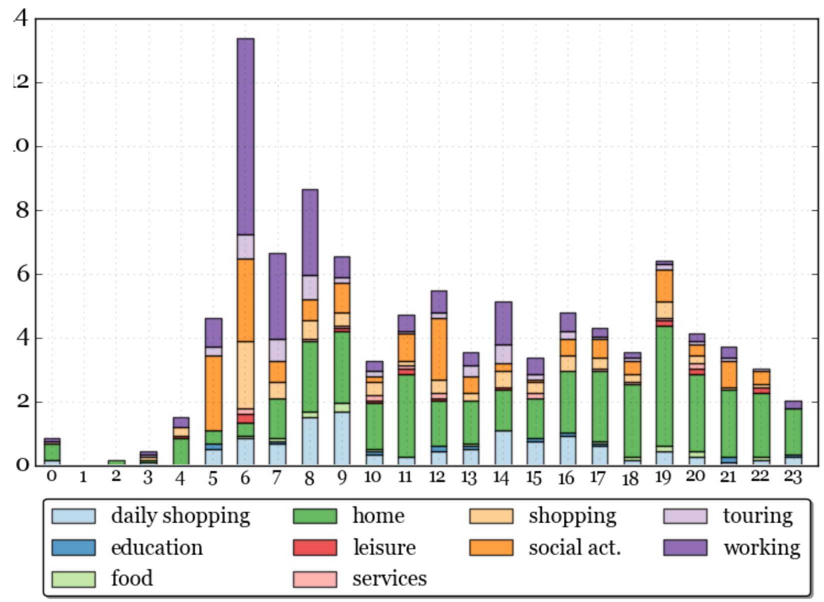
Semantic Mobility Analytics

Temporal Analysis

- Calci traffic



In



Out

Semantic Mobility Analytics

Temporal Analysis