Revealed preference data from WiFi traces for pedestrian activity scheduling

Antonin Danalet
TRANSP-OR, EPFL

8th Workshop on Discrete Choice Models
May 31st, 2013

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Presentation outline

- Motivation
- Data requirement
- Methodology
- A case study on EPFL campus
- Conclusion
- Future work

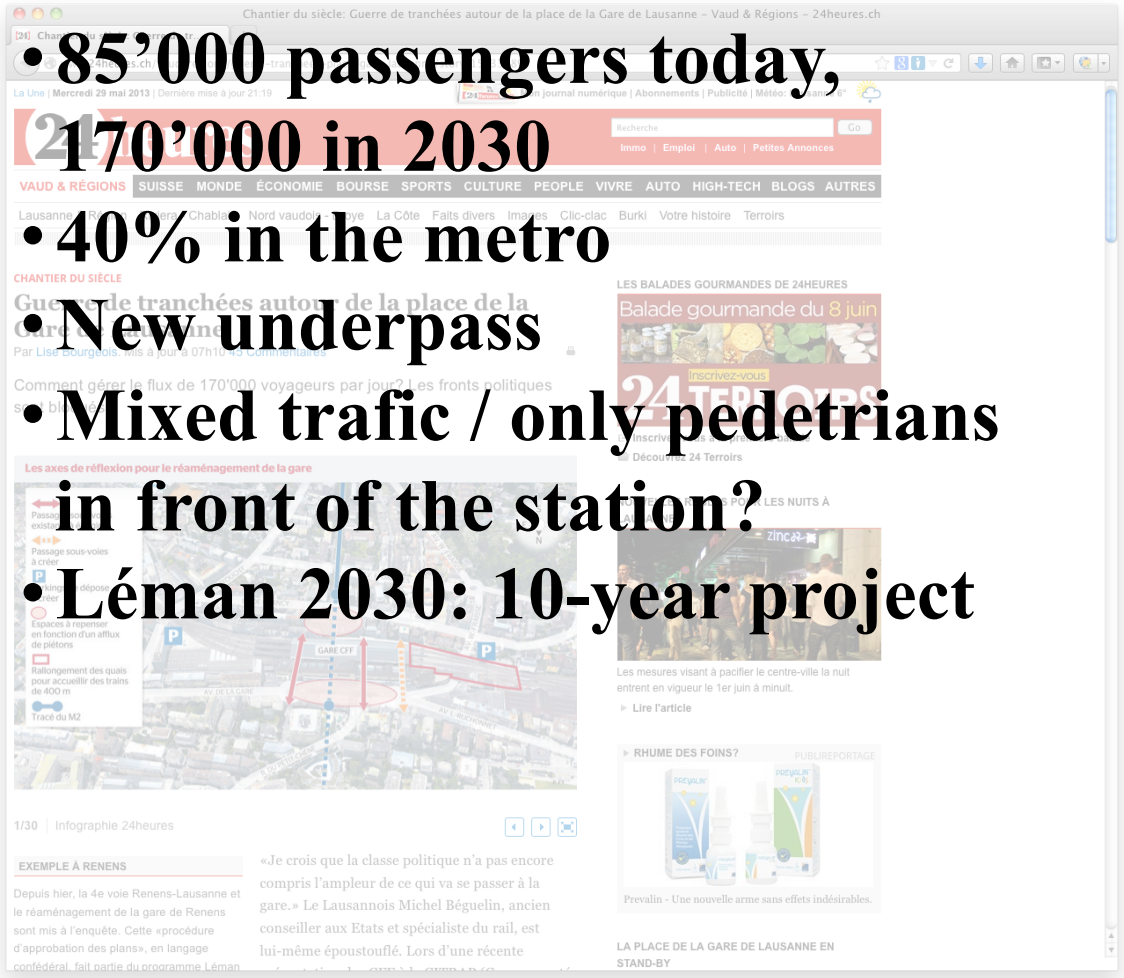
MOTIVATION

Lausanne railway station

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

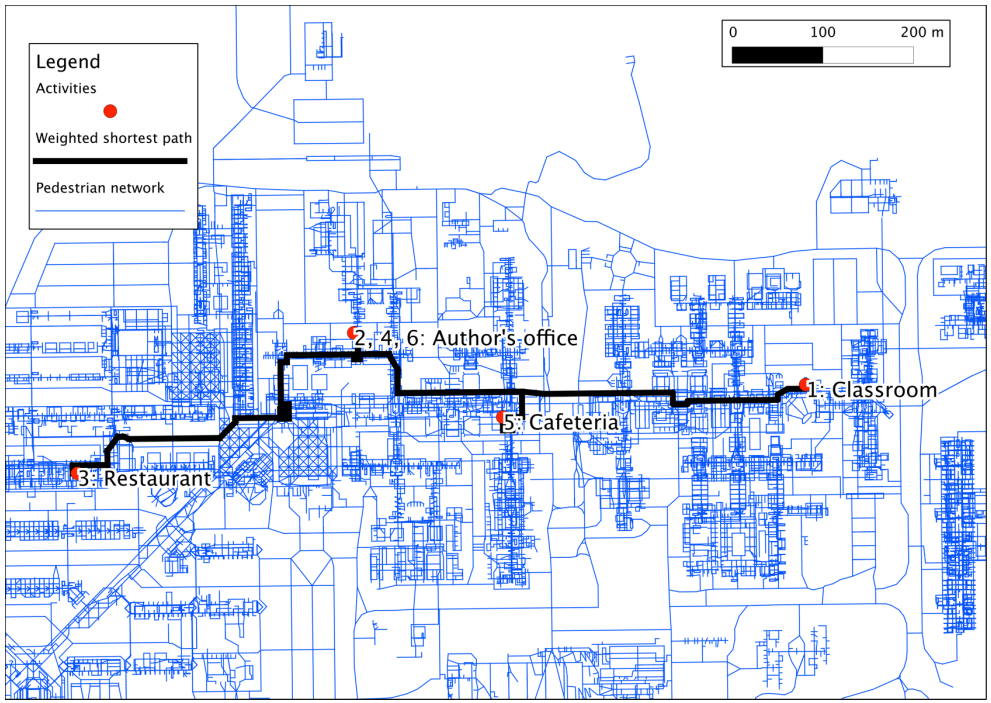
Lausanne railway station



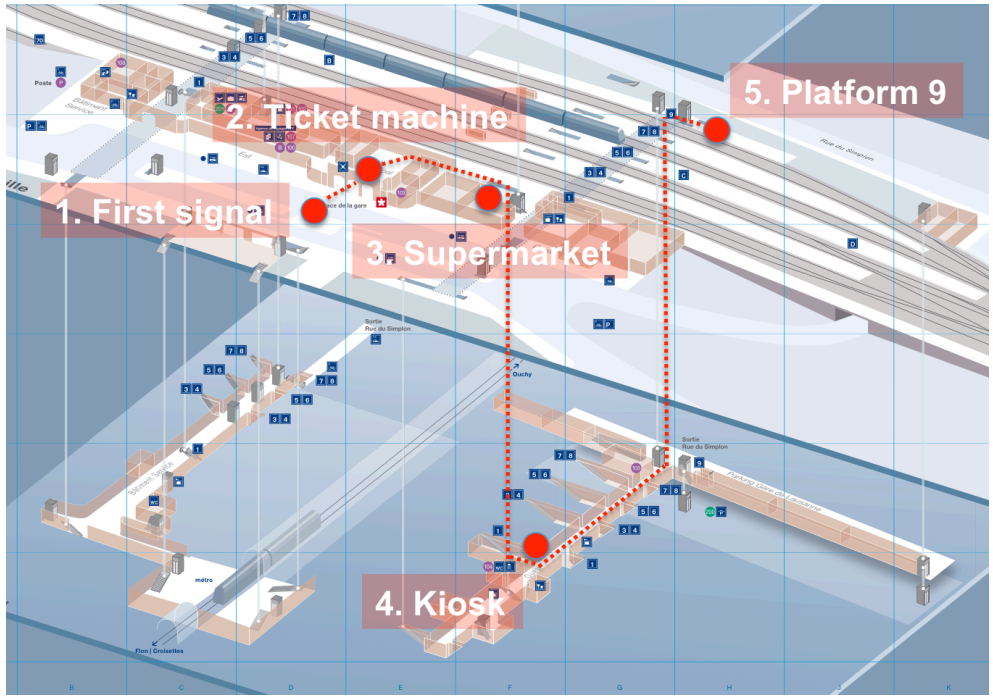
- 85'000 passengers today, 170'000 in 2030
- 40% in the metro
- New underpass
- Mixed trafic / only pedetrians in front of the station?
- Léman 2030: 10-year project

Understand pedestrian activities

What we are doing: Campus



What we want to do: Station



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

DATA REQUIREMENT

Data requirement

- Required
  - Localization data with full coverage of the facility
  - Pedestrian network

- Not really required but often available information
  - Prior aggregated occupancy

Data requirement: Localization

- Data from communication network infrastructure
  - GSM traces (Calabrese et al. 2011, Bekhor et al. 2011)
  - WiFi traces
- Post-processing is needed (Rieser-Schüsseler 2012)
  - Detection of stop points
  - Activity purpose detection through land-use information and spatial matching

Data requirement: Pedestrian network

- We need maps
  - With **points of interests** and destinations
  - With turn-by-turn **directions**
  - With **multi-floor** management
- More and more available in airports, malls, museums, campuses, hospitals
  - **Nokia**: 214 shopping malls in 2011, 4605 indoor maps in July 2012, 5100 in December 2012
  - **Microsoft**: 2700 indoor maps
  - **Google**: > 10'000 indoor maps
  - **Start-ups**: Wifarer, Meridian, Point Insider, ByteLight

Data requirement: Prior occupancy

- The *a priori* **number of people C(x,t)** who are performing an activity
  - Capacity (nb of seats), registered students, expected passengers, capacity of a scene, point-of-sale data, …
- At each **destination x**
  - Classroom, platform, scene, …
- At any **time t**
  - class schedules, train schedules, opening hours, …
- Example: 1500 passengers on platform 4 at 16h04
  32 students in GC B3 31 at 17h15

METHODOLOGY

Methodology

- **Goal:** extract the possible activity-episodes performed by pedestrians from digital traces from communication networks

Input

- Localization measurement
- Prior occupancy
- Pedestrian map

Output

- set of candidate activity-episodes sequences associated with the probability of being the true one

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Methodology

- Probabilistic measurement model:
  A Bayesian approach
  - Measurement equation
  - Prior
- Generation of activity-episode sequences
  - Episode location
  - Episode start and end times
- Intermediary signals
- Sequence elimination procedure

Definitions / Notations

- Measurement: $\hat{s} = (\hat{x}, \hat{t})$
- Activity-episode: $a = (x, t^-, t^+)$
- Episode location, start time and end time
- Activity-episode sequence: $(a_1, ..., a_m) = a_{1:m}$
- Activity: $A(a)$
- Activity pattern: $(A_1, ..., A_m) = A_{1:m}$

Probabilistic measurement model

Measurement likelihood

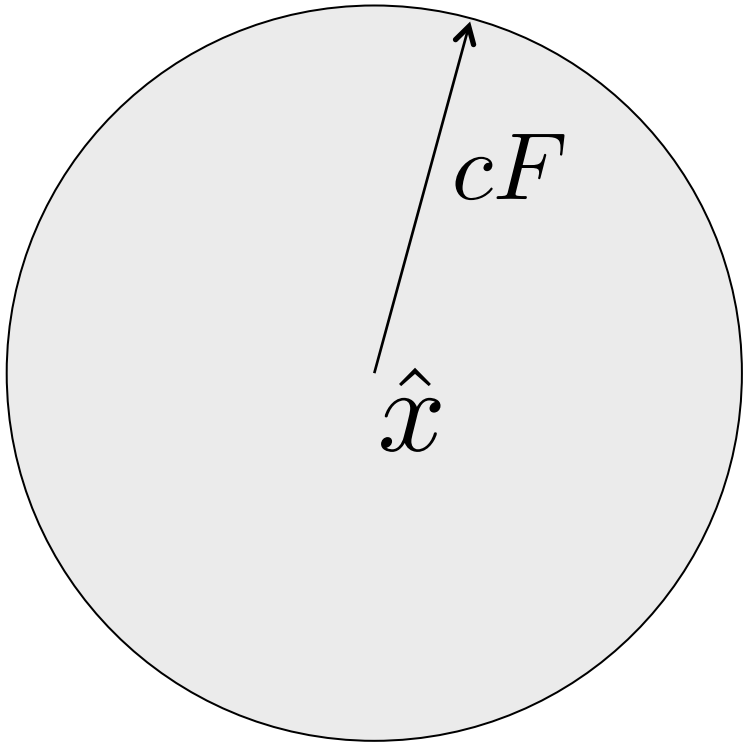
Prior

$$P(a_{1:m}|\hat{s}_{1:n}) \propto P(\hat{s}_{1:n}|a_{1:m}) \cdot P(a_{1:m})$$

Activity model

Probabilistic measurement model

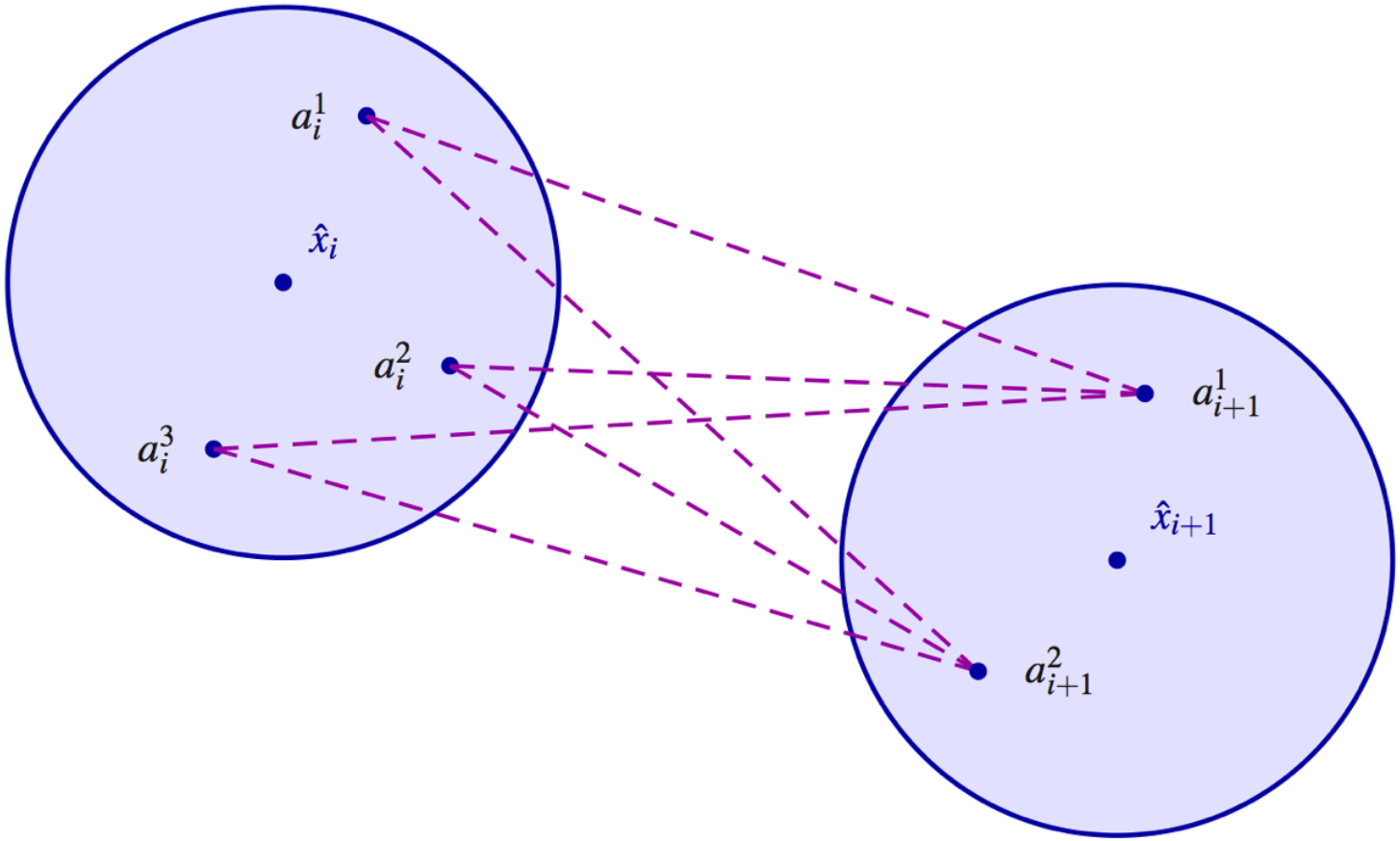
$$P(\hat{s}_{1:n}|a_{1:m}) = \prod_{j=1}^{m} P(\hat{s}_{i_{j-1}+1:i_j}|a_j)$$
$$= \prod_{j=1}^{m}\prod_{i=1}^{n} P(\hat{s}_{i_j}|a_j)$$
$$= \prod_{j=1}^{m}\prod_{i=1}^{n} P(\hat{x}_{i_j}|x_j)$$



Prior

$$
\begin{aligned}
P(a_{1:m}) &= \prod_{j=1}^{m} P(a_j) && (1)\\
&= \prod_{j=1}^{m} P(x_j, t_j^-, t_j^+) && (2)\\
&= \prod_{j=1}^{m} \frac{C_{x_j}(t_j^-, t_j^+)}{\sum_{x \in X} C_x(t_j^-, t_j^+)} && (3)\\
&= \prod_{j=1}^{m} \int_{t_j=t_j^-}^{t_j^+} \frac{C_{x_j}(t_j)}{\sum_{x \in X} C_x(t_j)} dt && (4)
\end{aligned}
$$

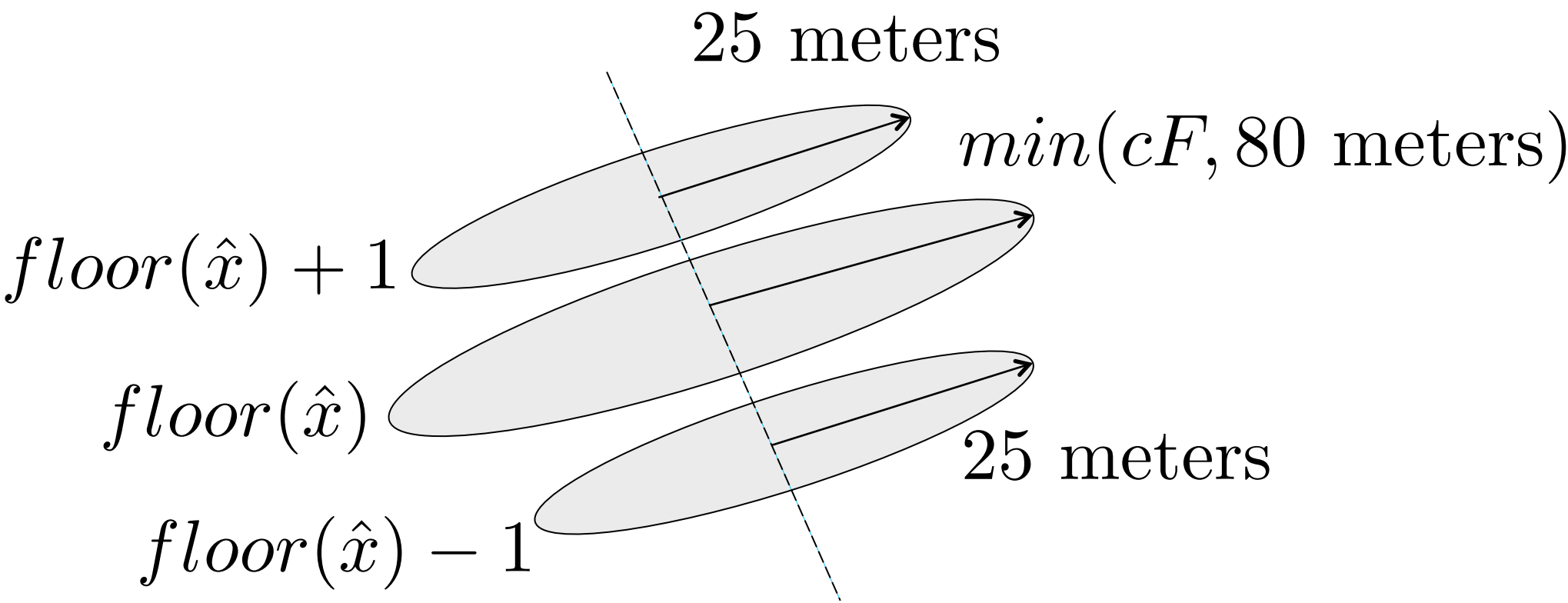
Generation of activity-episode sequences



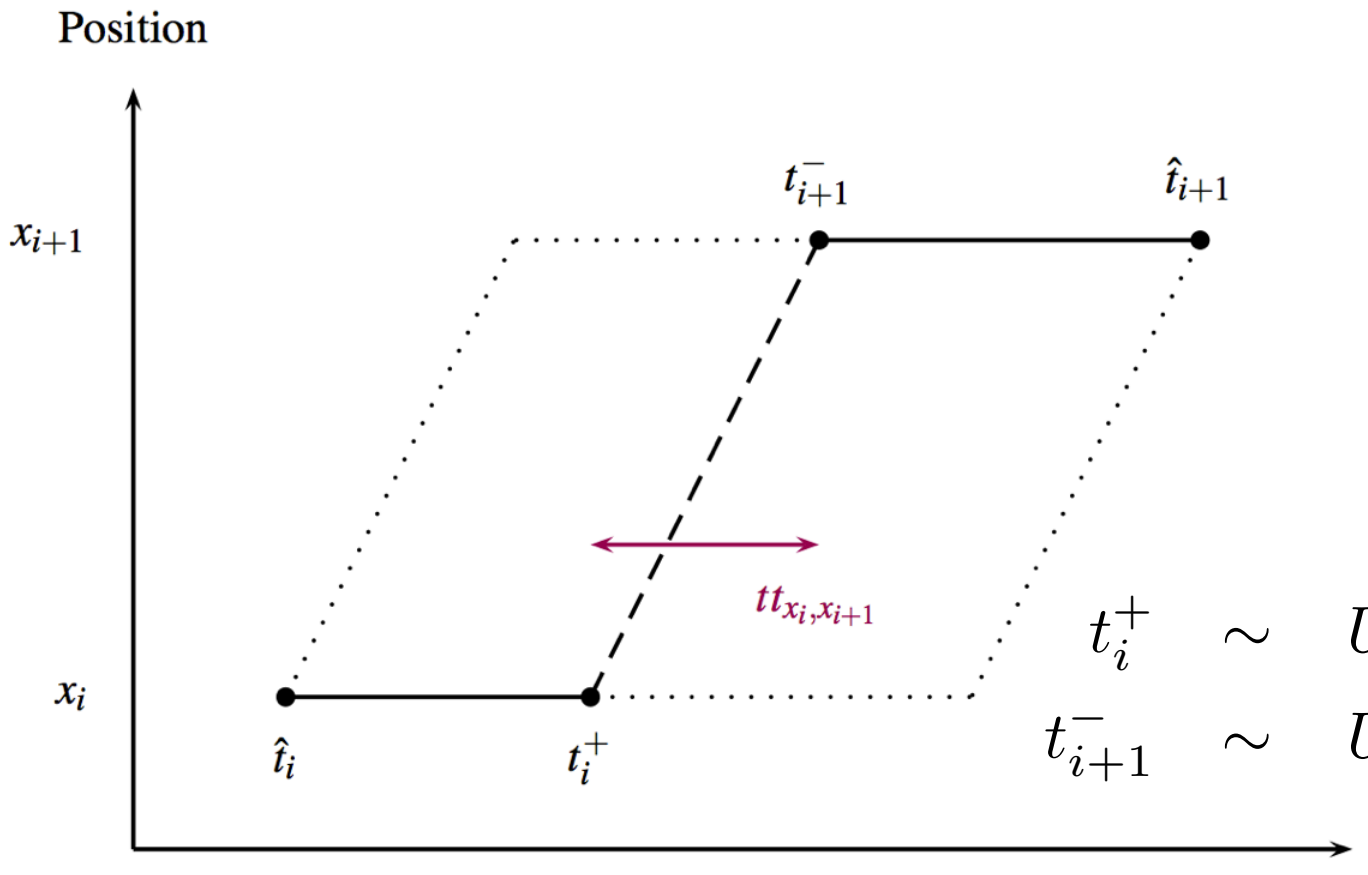
TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Domain of data relevance



Generation of activity-episode sequences



$$t_i^+ \sim U(\hat{t}_i, \hat{t}_{i+1} - tt_{x_i,x_{i+1}})$$

$$t_{i+1}^- \sim U(t_i^+ + tt_{x_i,x_{i+1}}, \hat{t}_{i+1})$$

Generation of activity-episode sequences

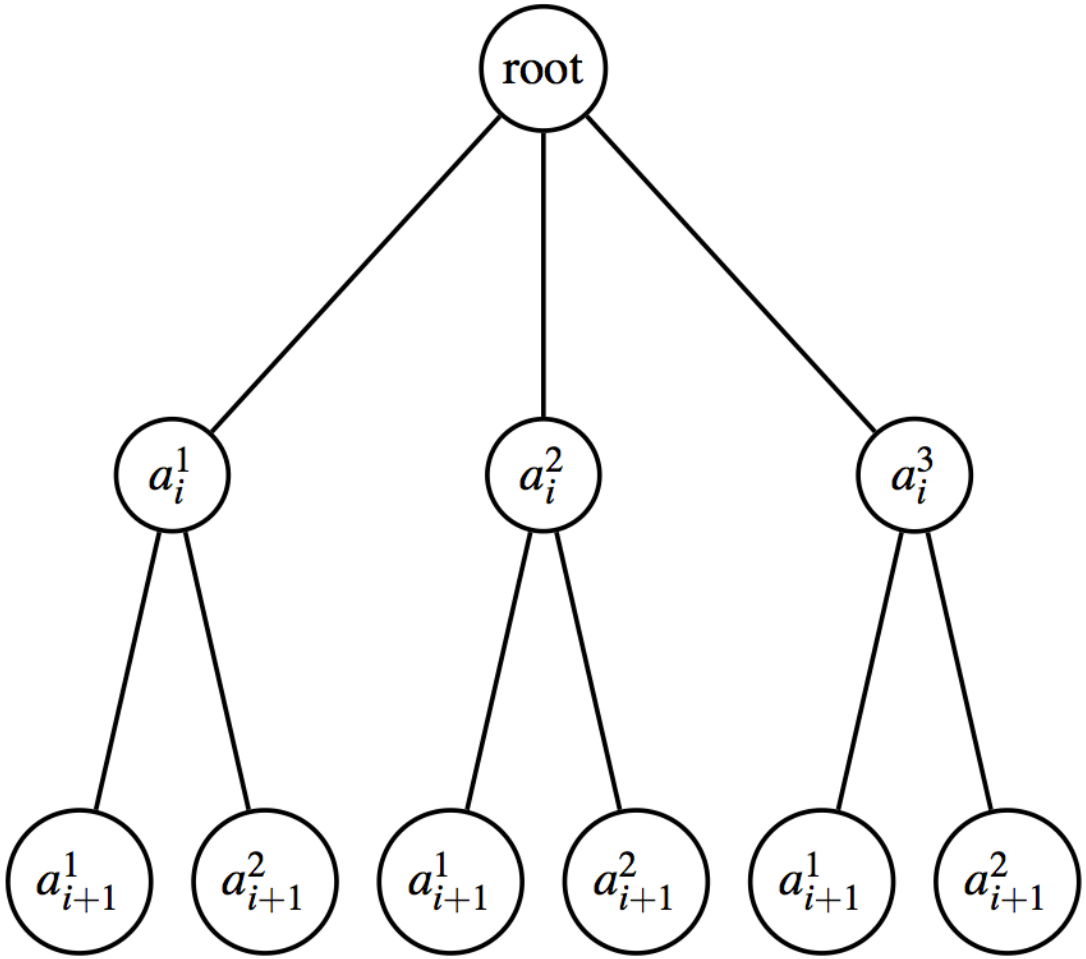
$$f(t_{i+1}^{-}) = \frac{1}{\hat{t}_{i+1} - tt_{x_i,x_{i+1}} - \hat{t}_i} \ln \frac{\hat{t}_{i+1} - tt_{x_i,x_{i+1}} - \hat{t}_i}{\hat{t}_{i+1} - t_{i+1}^{-}}$$

$$E(t_{i+1}^{-}) = \frac{\hat{t}_i + tt_{x_i,x_{i+1}}}{4} + \frac{3 \cdot \hat{t}_{i+1}}{4}$$

Generation of activity-episode sequences

- $$tt_{x_i,x_{i+1}} = \frac{dist(x_i, x_{i+1})}{v}$$
- Distance: shortest path in the pedestrian graph
- Speed: 1.34 meters/second (Buchmueller and Weidmann, 2006)

Generation of activity-episode sequences



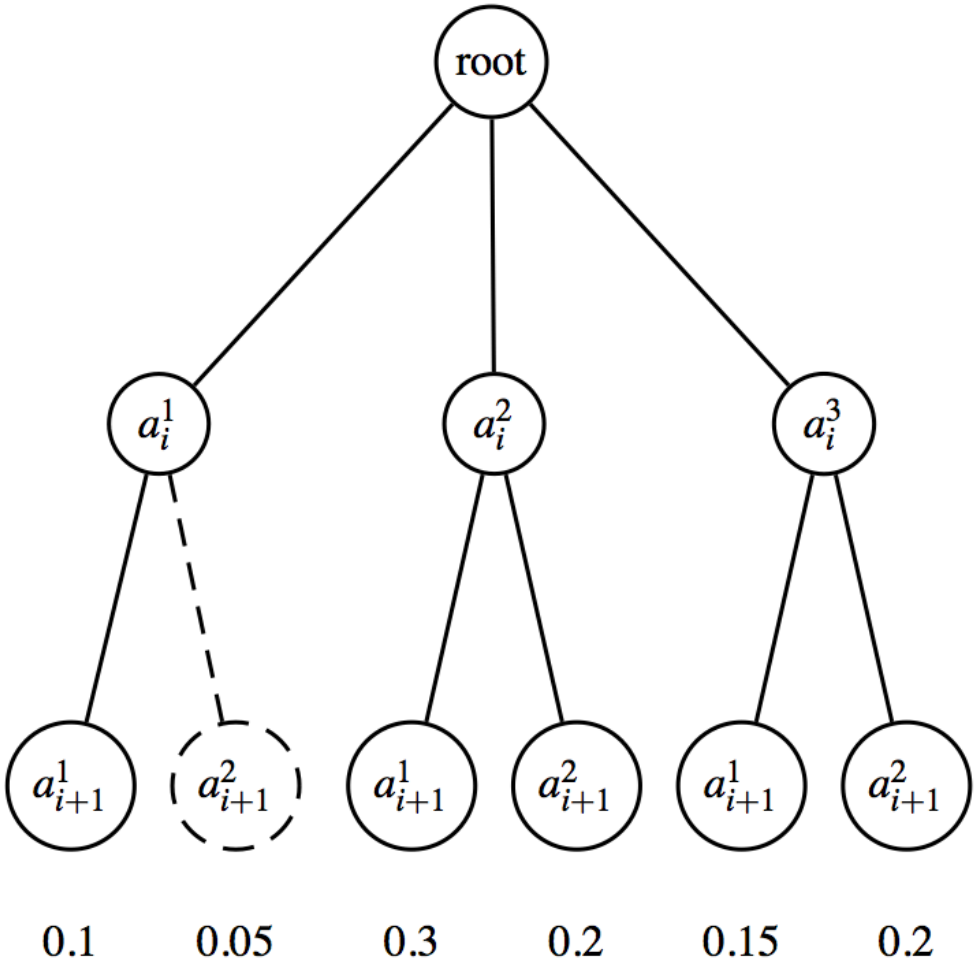
Intermediary signals

- Eliminate intermediary signal if

$$E(t^+) - E(t^-) < T_{min}$$

since we generate an activity episode at each signal.

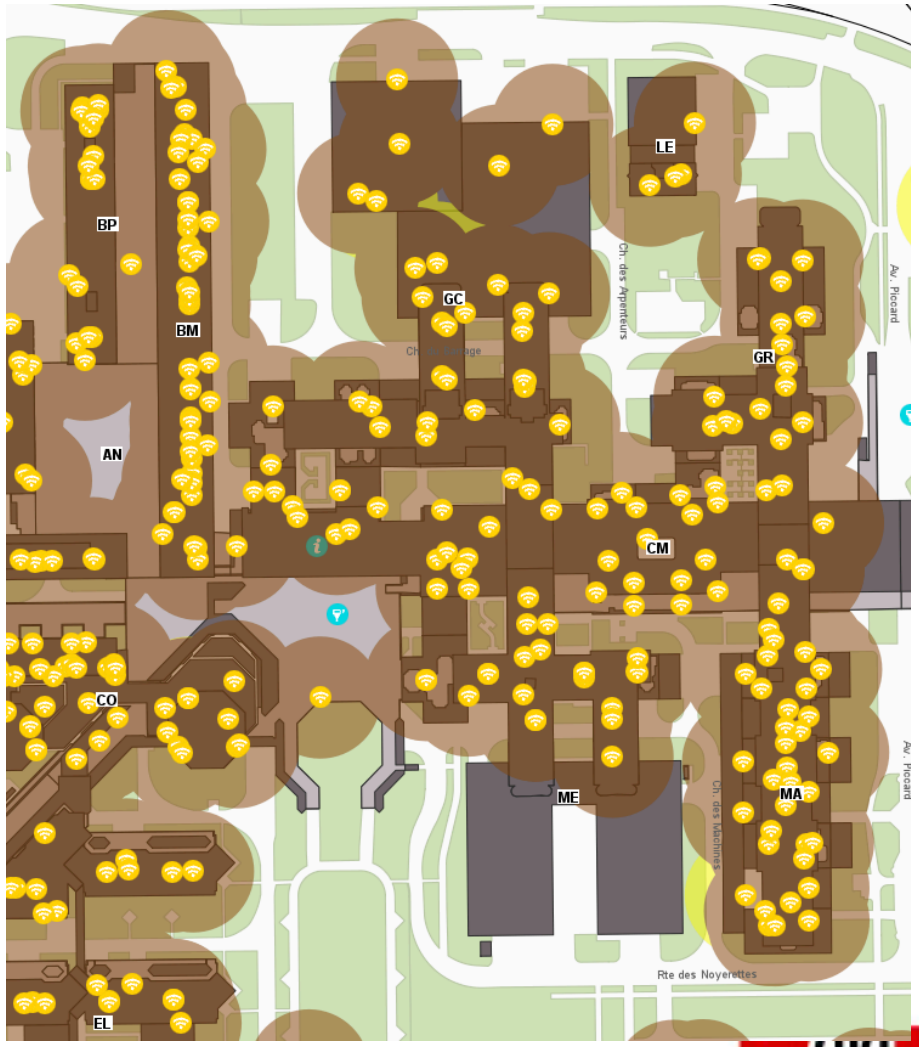
Sequence elimination



A CASE STUDY ON EPFL CAMPUS

EPFL data: Localization

- 8 participants for 2 months with known ID
- Non-participants: 46 days, but only 10 with courses
  - 200 students in 6 different classes
  - 317 employees
  - 700 students from University of Lausanne
- For 151 GC students, 152’598 observations

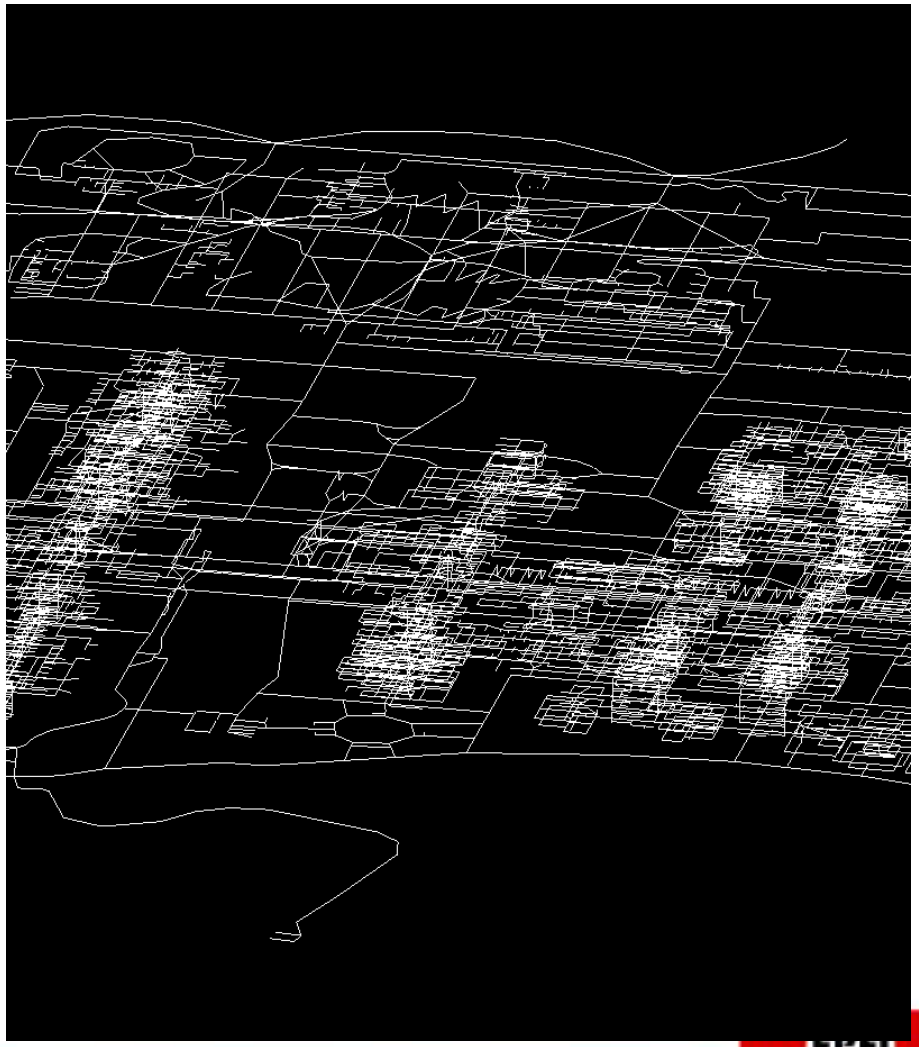


TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

EPFL data: Pedestrian network

- Source: map.epfl.ch
- 56’655 edges
- 4 different levels of path
  - Major (« highway »)
  - Inter-building
  - Intra-building
  - Access to offices
- Shortest path
- All offices, restaurants, classrooms and other points of interest are coded: X

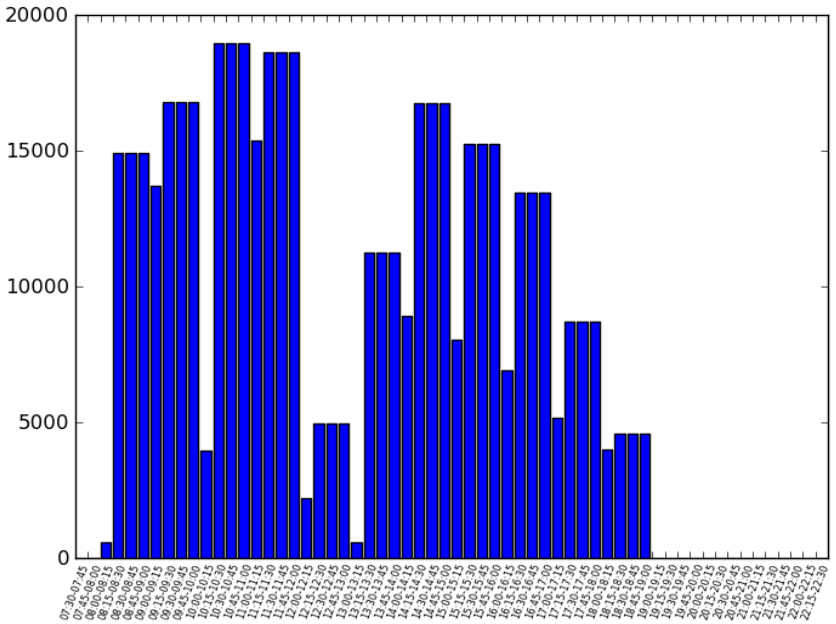


TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

EPFL data: Potential occupancy C(x,t)

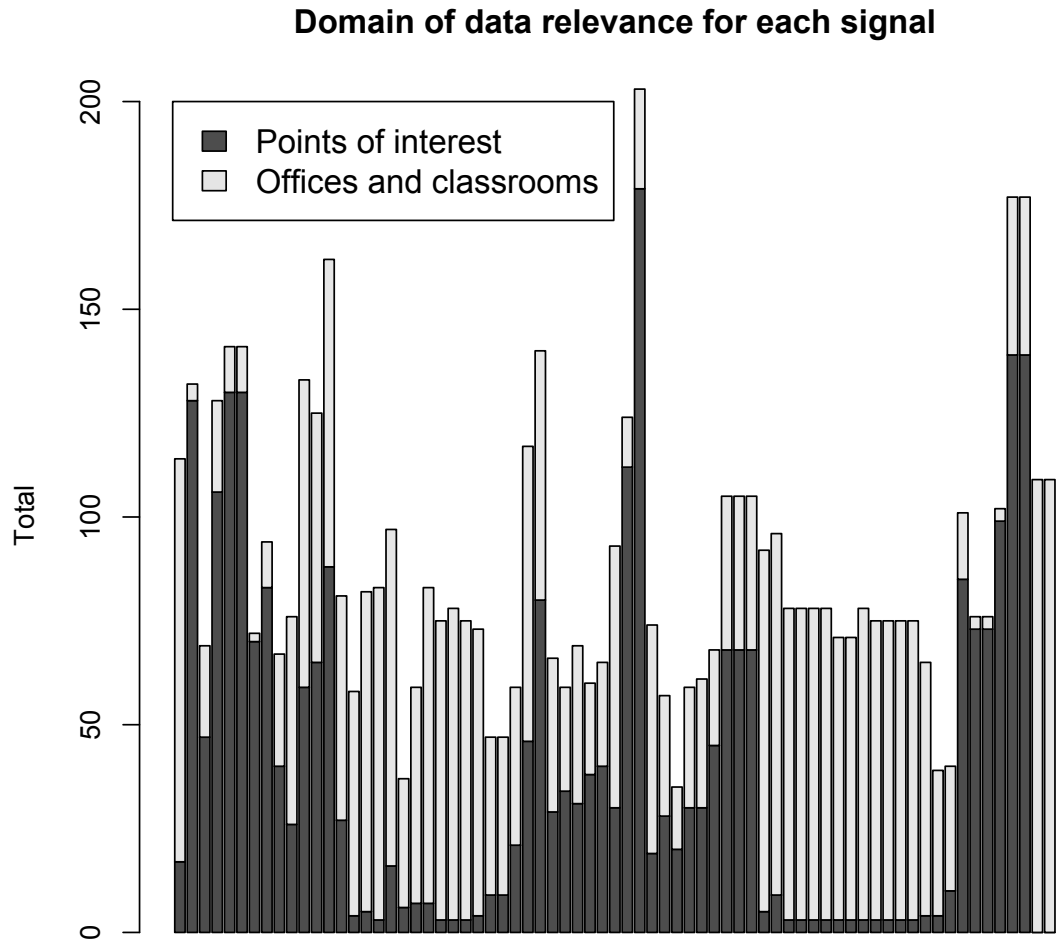
- Class schedules with
  - Number of students
  - Name of the classroom
- Number of employees per office
  - Name of the office
  - Sum of percent of work (e.g, 3 full times = 300%)
- Number of seats in restaurants
  - Localization
  - Opening hours
- Number of seats in library



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

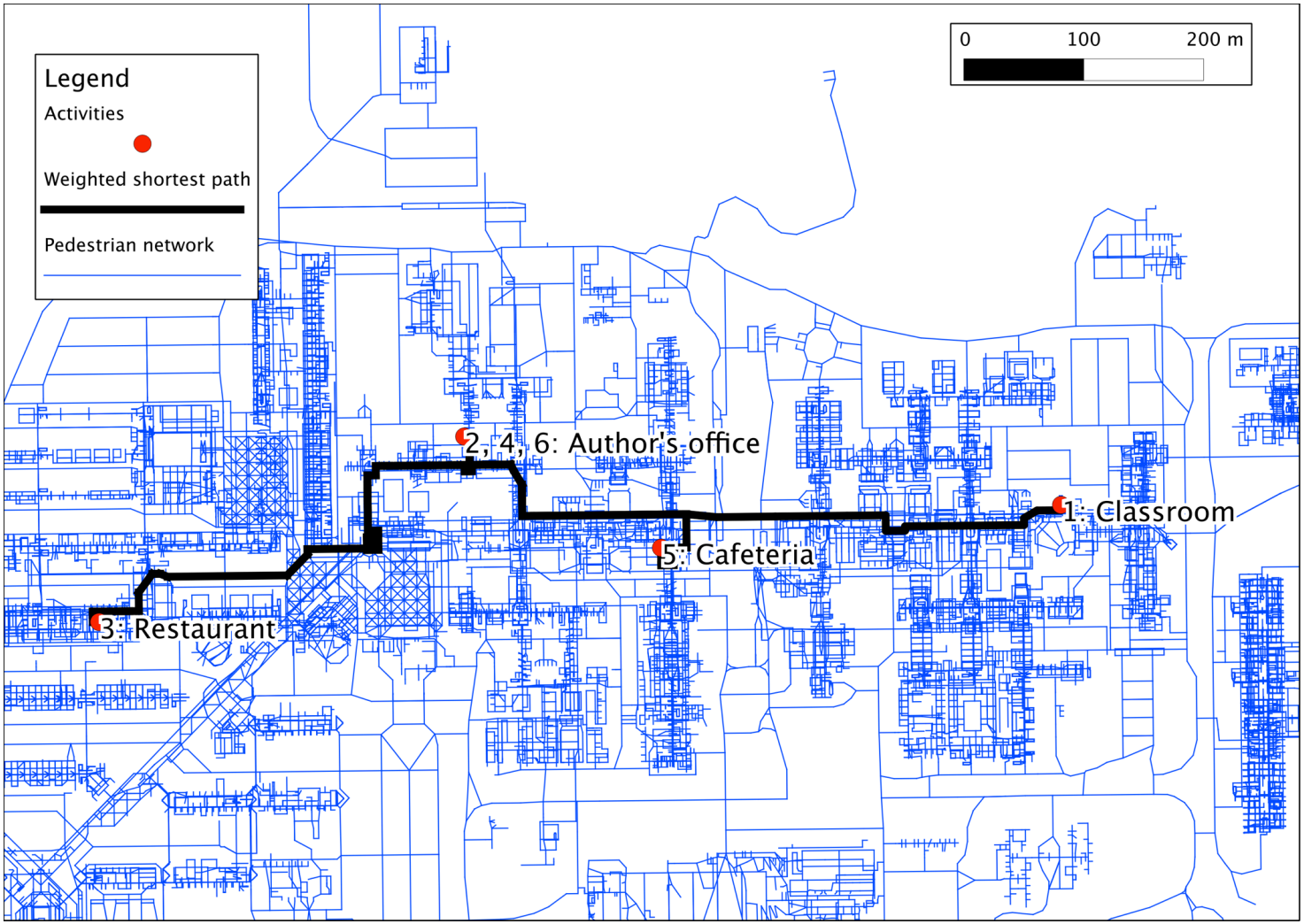
Generation of activity-episode sequences



Results

Model				Truth			Δx
Arrival time	*Departure time*	*Floor*	*Location*	*Time spent*	*Floor*	*Location*	*(in m.)*
8:33-8:33	10:38-10:38	1	Classroom	8.32am-10.30am	1	Classroom	0
10:40-10:40	11:51-11:51	3	Office	Until 11.47am	3	Author's office	7
11:54-11:54	12:47-12:53	1	Restaurant	From 11.55 am	1	Restaurant	0
12:51-12:58	13:03-13:44	3	Office	Around 1pm	3	Author's office	7
13:06-13:47	13:53-14:02	2	Cafeteria	Around 2pm	2	Cafeteria	0
13:55-14:04	19:45-19:45	3	Office	Until around 7.45pm	3	Author's office	7
19:47-19:47	19:52-19:52	3	Workshop	-	3	Metro stop	366

Results


Legend
Activities
Weighted shortest path
Pedestrian network
0
100
200 m
2, 4, 6: Author's office
1: Classroom
5: Cafeteria
3: Restaurant

CONCLUSION

Conclusion

- Prior is needed to **overcome low precision**: make use of non-localization aggregate data, with theoretical or no time dimension in it
- **Localization data brings dynamics** in the model: individual sequences with start and end times
- Our methodology is **merging** these two different types of data
- Robust for **low density data** as well

FUTURE WORK

Further work

- Binary choice model for class attendance
  - No use of the dynamic part of the sequence
- Analysis of the access to campus
  - First and last destination of the sequence
  - Arrival times on campus, departure times
- Based on class attendance and on available time budget: activity scheduling

References

- Bekhor, S., Cohen, Y. and Solomon, C. (2011), Evaluating long-distance travel patterns in Israel by tracking cellular phone positions. J. Adv. Transp..
  doi: 10.1002/atr.170
- Buchmüller, S. and Weidmann, U. (2006). Parameters of pedestrians, pedestrian traffic and walking facilities, IVT-report Nr. 132, Institut for Transport Planning and Systems (IVT), ETHZ.
- Calabrese, F.; Di Lorenzo, G.; Liang Liu; Ratti, C., "Estimating Origin-Destination Flows Using Mobile Phone Location Data," Pervasive Computing, IEEE , vol.10, no.4, pp.36,44, April 2011
  doi: 10.1109/MPRV.2011.41
- Rieser-Schüssler, N. (2012). Capitalising Modern Data Sources for Observing and Modelling Transport Behaviour, Transportation Letters: the International Journal of Transportation Research (2): 115–128.
  doi: 10.3328/TL.2012.04.02.115-128