PhD thesis

Activity choice modeling for pedestrian facilities

Antonin Danalet

Outline

Outline

Motivation: Understanding pedestrian demand

Detecting activity-episode sequences

A path choice approach to activity modeling

Location choice with panel effect

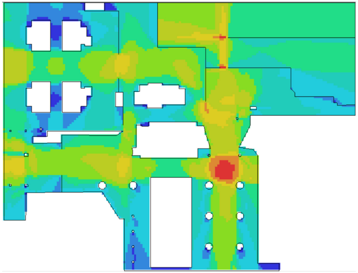
Conclusion and future work

New interest in pedestrian modeling

- Urban growth and its pressure on pedestrian facilities
- Availability of new tracking data

In airports...

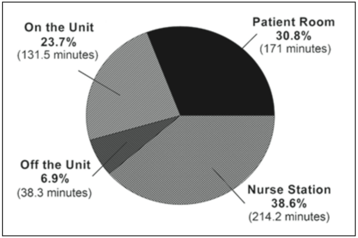
- +38 % air passengers (2008-2013)
- Surveying [LUS14], space syntax [KBM14]



[KMM15]

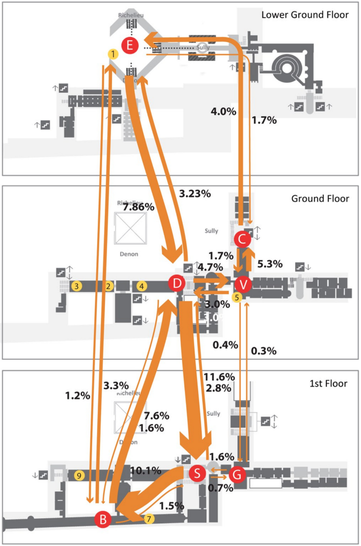
In hospitals...

- US: Hospital-building and -renovation boom [HCSL08]
- Time use of nurses using RFID [HCSL08]



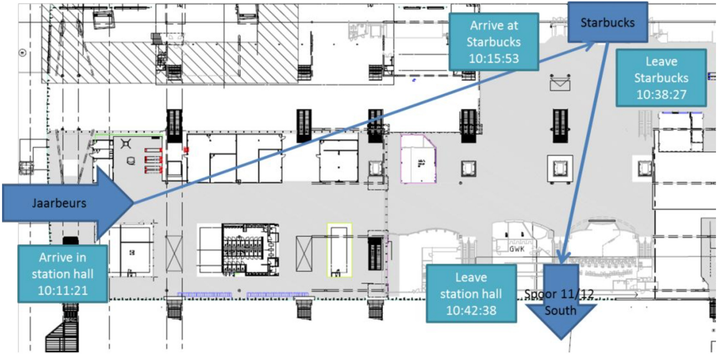
In museums...

- Louvre: +35 % visitors (2004-2014)
- Understanding congestion using Bluetooth [YSR+14]



In train stations...

- Utrecht Central Station: +14 % visitors by 2020
- Activity location choice using WiFi and Bluetooth [Ton14]



Challenges of pedestrian facilities

- Knowing the number of visitors
- Determining the source of congestion
- Localizing points of interest
- Modifying/building new facilities
- Defining timetables

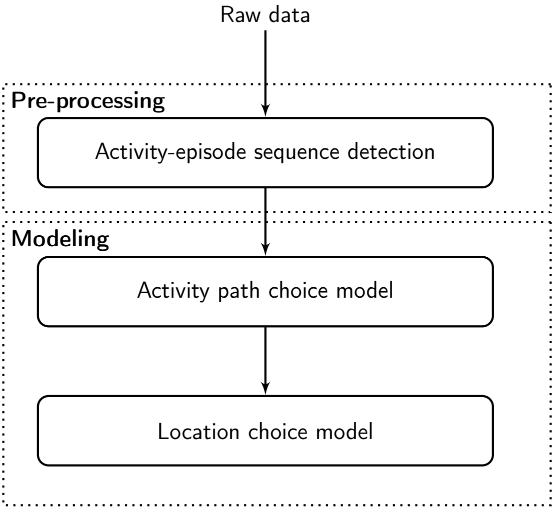
Data from communication antennas

+	−
• Large sample size	• No socioeconomics
• Low cost	• Not representative
• Low privacy risk	• Privacy risk
• No recall bias	• Low frequency
• No need to distribute devices	• Low precision
• Tracking non-travelers	• No stops
• Full coverage of the facility	• No activity purpose

Goal: Understanding pedestrian demand

- **Where, when and for how long** do pedestrians perform activities in pedestrian facilities?
- Based on **communication network traces** from **existing antennas**

Activity path approach



Activity-episode sequence detection

- Explicit modeling of the imprecision in the measure
- Usage of prior knowledge of the infrastructure
- Avoidance of the pingpong effect

Activity-path choice model

- No tours, no priorities
- Managing large choice sets
- Unique utility for activity type, time-of-day and duration choices

Location choice model

- Including panel data
- Correcting for serial correlation

More details

- Introduction: Chapter 1 in [Dan15b]
- Literature review: Chapter 2 in [Dan15b]

Outline

Motivation: Understanding pedestrian demand

Detecting activity-episode sequences

A path choice approach to activity modeling

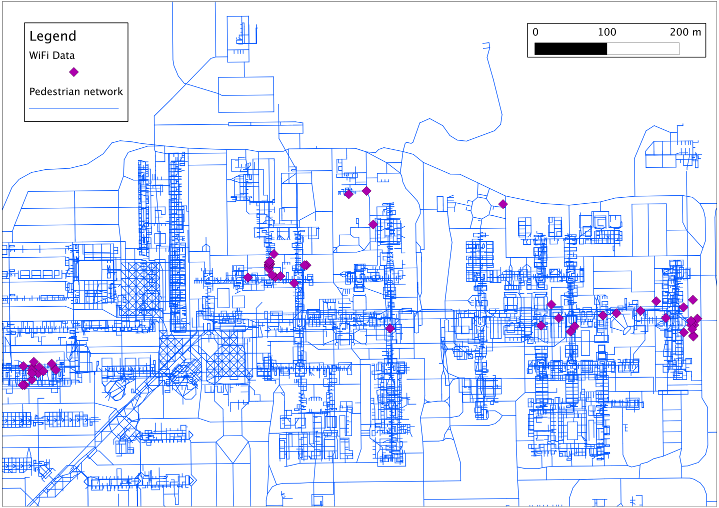
Location choice with panel effect

Conclusion and future work

Data requirement

- Required
  - Localization data with full coverage of the facility
  - Semantically-enriched routing graph for pedestrians
- Not required but often available information
  - Potential attractivity measure

Data requirement: Localization



Data requirement: Map (POI + network)

Potential attractivity measure

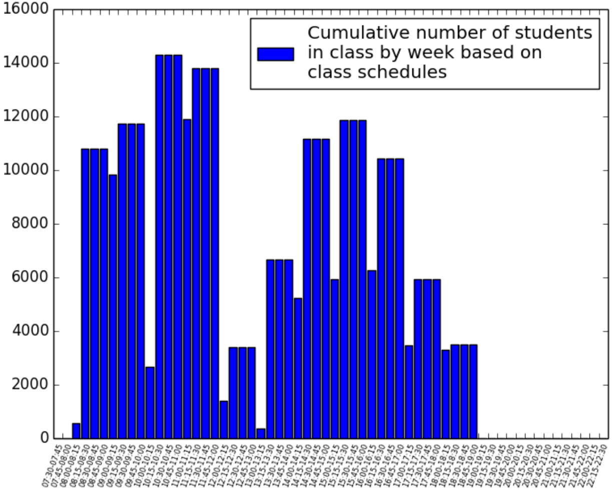
For individual n, point of interest x, start and end times t^- and t^+:

$$S_{x,n}(t^-, t^+) = \int_{t=t^-}^{t^+} \delta_{x,n}(t) \cdot att_n(x, t) dt$$

with

- Time constraints $\delta_{x,n}$
  (e.g., train or class schedules, opening hours)
- Destination attractivity $att_n(x, t)$
  (e.g., classroom, platform, scene aggregate occupancy)

Data requirement: Potential attractivity



Methodology

Input

- Localization measurement
- Semantically-enriched routing graph
- Potential attractivity measure

Output

- Set of candidate activity-episode sequences associated with the likelihood to be the true one

Probabilistic measurement model: a Bayesian approach

Measurement likelihood

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

Activity probability

Prior

with

- measurement $\hat{m} = (\hat{x}, \hat{t})$, $(\hat{m}_1, \hat{m}_2, ..., \hat{m}_j, ..., \hat{m}_J) = \hat{m}_{1:J}$
- activity episode $a = (x, t^-, t^+)$, $(a_1, a_2, ..., a_\psi, ..., a_\Psi) = a_{1:\Psi}$

Measurement likelihood

$$P(\hat{m}_{1:J} | a_{1:\Psi}) = \prod_{\psi=1}^{\Psi} P(\hat{m}_{1:J}^{\psi} | a_{\psi}) \quad \Leftrightarrow \quad \text{Independence between activities}$$

$$= \prod_{\psi=1}^{\Psi} \prod_{j=1}^{J} P(\hat{m}_{j}^{\psi} | a_{\psi}) \quad \Leftrightarrow \quad \text{Independence between measurements}$$

$$= \prod_{\psi=1}^{\Psi} \prod_{j=1}^{J} P(\hat{x}_{j}^{\psi} | x_{\psi}) \quad \Leftrightarrow \quad \text{No time measurement error}$$

Prior: Potential attractivity measure

$$
\begin{aligned}
P(a_{1:\Psi}) &= \prod_{\psi=1}^{\Psi} P(a_\psi) \\
&= \prod_{\psi=1}^{\Psi} P(x_\psi, t_\psi^-, t_\psi^+) \\
&= \prod_{\psi=1}^{\Psi} \frac{S_{x_\psi,n}(t_\psi^-, t_\psi^+)}{\sum_{x \in POI} S_{x,n}(t_\psi^-, t_\psi^+)}
\end{aligned}
$$

Probabilistic measurement model: a Bayesian approach

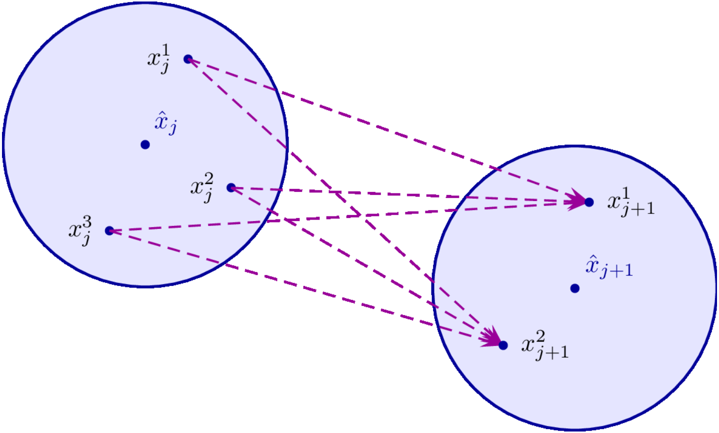
Measurement likelihood

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

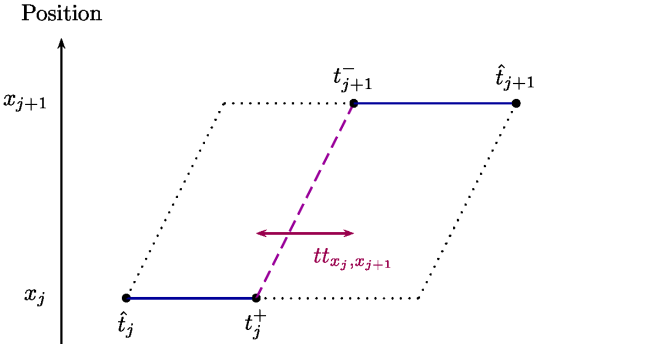
Activity probability

Prior

Generation of activity-episode sequences

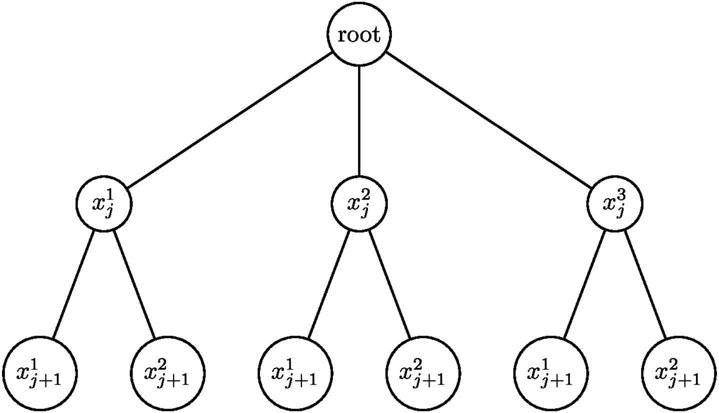


Generation of activity-episode sequences



with $tt_{x_j,x_{j+1}}$ the travel time from x_j to x_{j+1}

Generation of activity-episode sequences



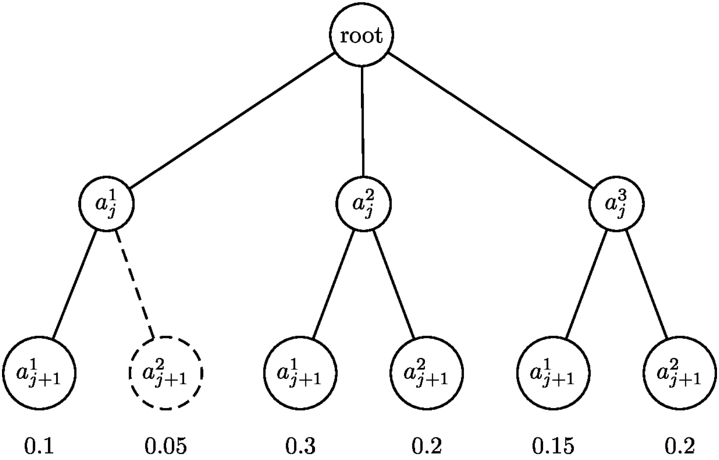
Intermediary measurements

Eliminate intermediary measurements if

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

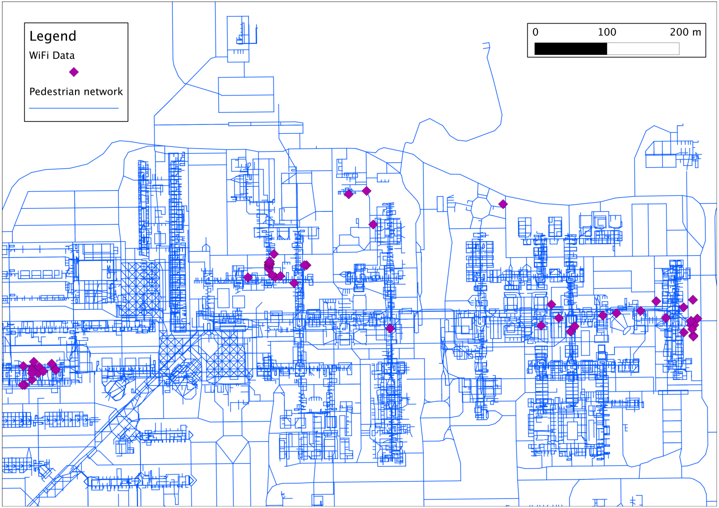
since we generate an activity episode at each measurement.

Sequence elimination

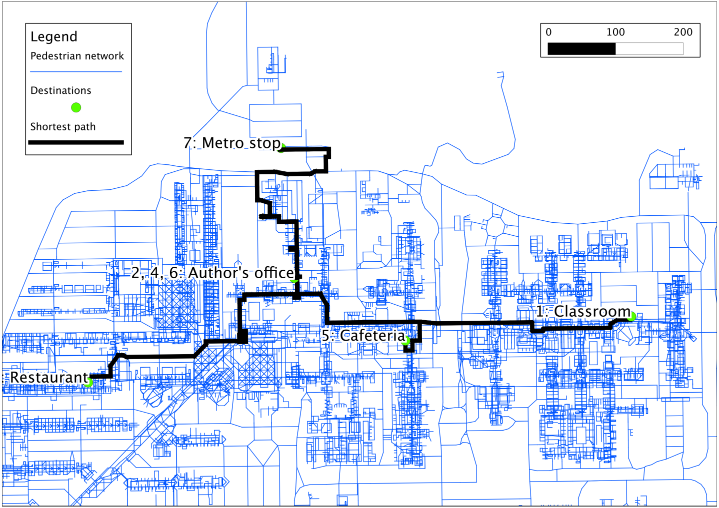


We keep L (here, $L = 5$) most likely activity-episode sequences

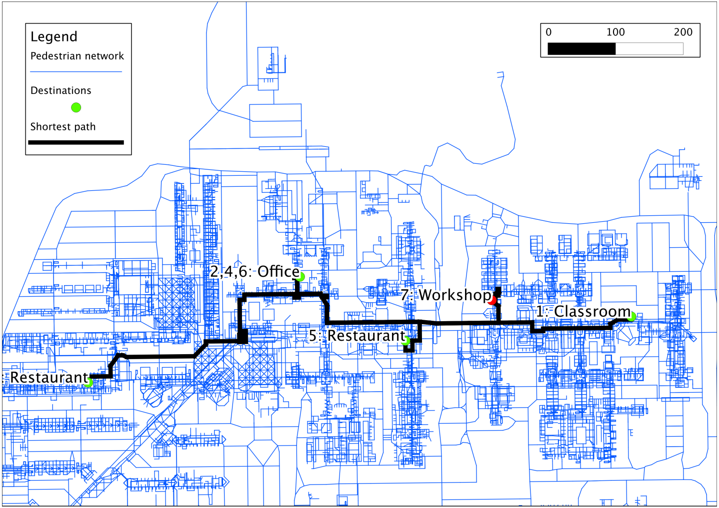
Results: me on EPFL campus, raw data



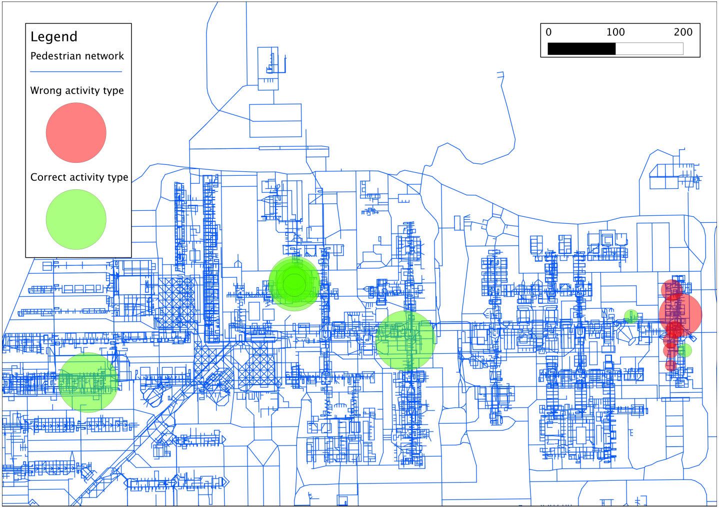
Results: me on EPFL campus, truth



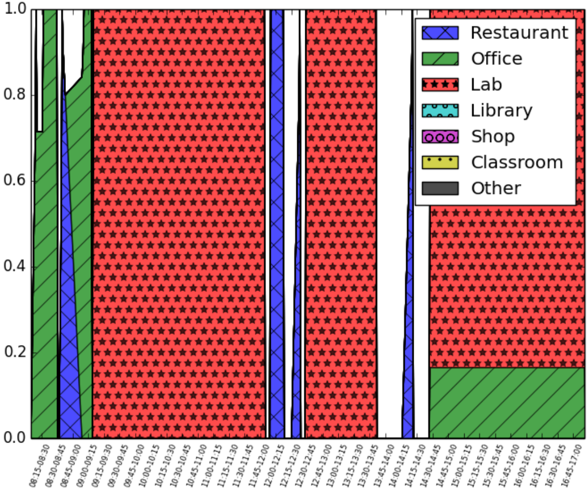
Results: me on EPFL campus, model, $L = 1$



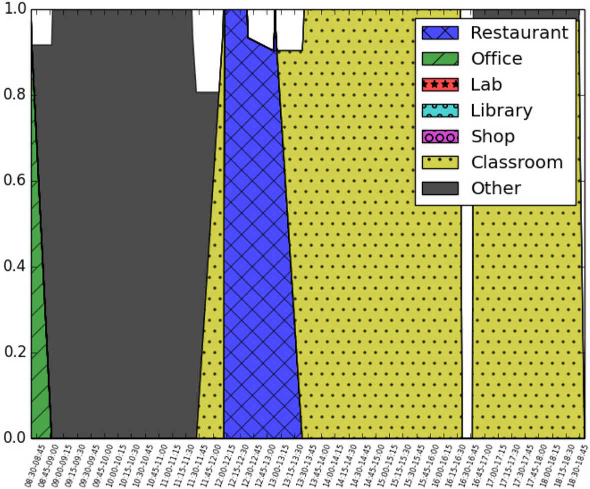
Results: me on EPFL campus, model, $L = 100$



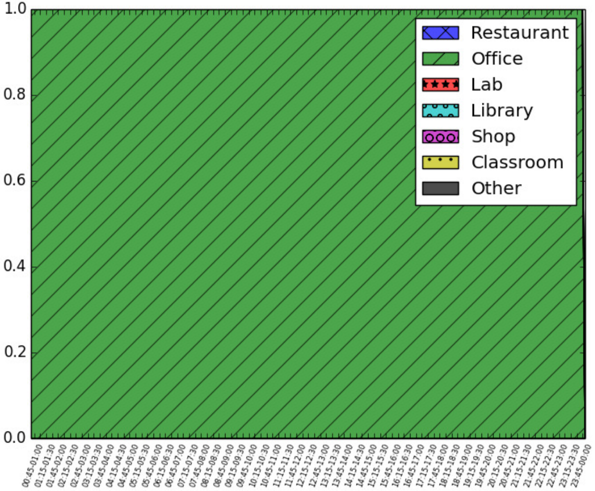
Results: an employee on EPFL campus, $L = 20$



Results: an computer science student, $L = 20$

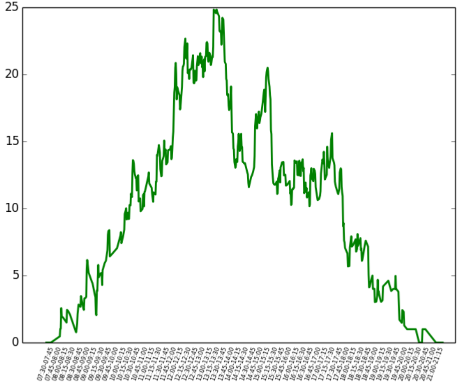


Results: an employees?, $L = 20$



Detection: Results for full population

- 3 activity episodes on average
- 1h37 on each activity
- Devices detected in restaurant during lunch break (see figure)



More details

- Article: [DFB14]
- Chapter 3 in [Dan15b]

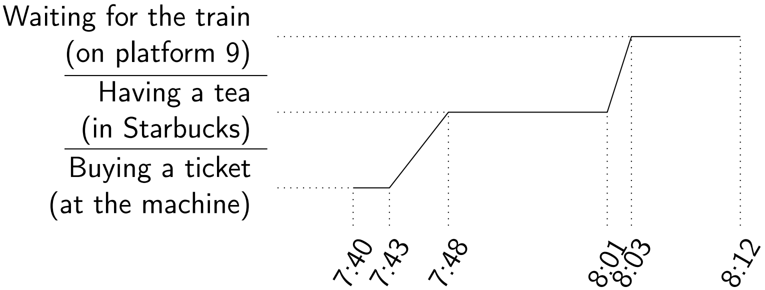
Outline

Modeling assumption

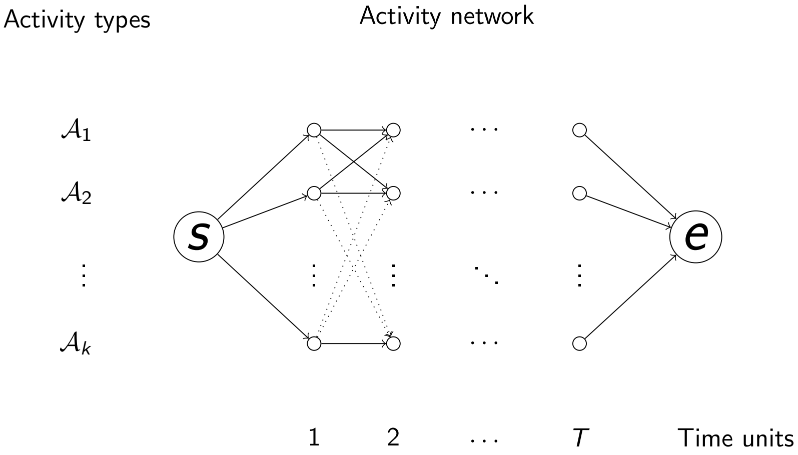
- Sequential choice:
  1. activity type, sequence, time of day and duration
  2. destination choice conditional on 1.
- Motivations:
  - Behavioral: precedence of activity choice over destination choice [BBA01, AT04, HB04, AZBA12, KR13]
  - Dimensional: destinations $\times$ time $\times$ position in the sequence is not tractable

Observations: activity patterns in a transport hub

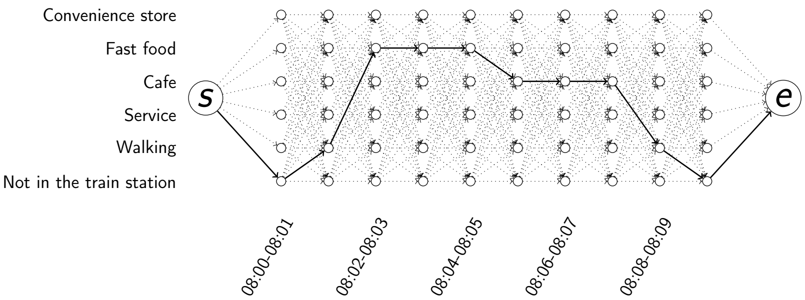
Activity types



Activity network



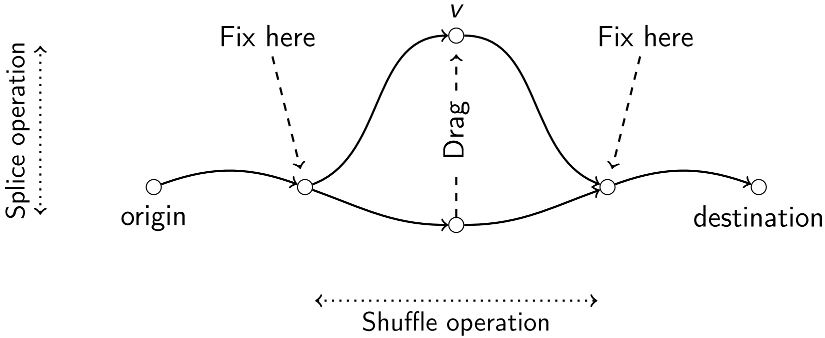
Activity path



Sampling strategies for choice set generation

- Simple random sampling (SRS)
- Importance sampling using Metropolis-Hastings algorithm [FB13] and strategic sampling [LK12]

Metropolis-Hastings sampling of paths



[FB13]

Metropolis-Hastings sampling of paths

- Sample paths from given distribution, without full enumeration
- To be defined:
  - Target weight: Also with non-node-additive utility
  - Proposal distribution:

$$P_{\text{insert}} = \frac{e^{-\tilde{\mu}\delta_{SP}(origin,v)+\delta_{SP}(v,destination)}}{\sum_w e^{-\tilde{\mu}\delta_{SP}(origin,w)+\delta_{SP}(w,destination)}}$$

  Relies on shortest paths, node-additive cost.

Strategic sampling

- Target weight:
  previously estimated model
- Proposal distribution:
  previously estimated model using only time-of-day preferences (node-additive)

Utility structure

- Utility of activity pattern:
  - Node utility $V(\mathcal{A}_{k,t})$
    - time-of-day preferences
  - Activity-episode utility $V(a)$
    - satiation effects: decreasing marginal utility, $\eta \ln(\mathit{duration})$
    - scheduling constraints: schedule delay
  - Activity path utility $V(\Gamma)$
    - primary activity
    - number of episodes
- Sampling correction

$$\mu\left(\sum_{k=1}^{K}\sum_{\tau=1}^{T} V(\mathcal{A}_{k,\tau}) + \sum_{a\in\mathcal{A}_{1:T}} V(a) + V(\Gamma)\right) + \ln\frac{k_{\Gamma n}}{b(\Gamma)}$$

Case study: pedestrians on EPFL campus

- 13'000 people per day
- 8 activity types:
  - classrooms,
  - shops,
  - offices,
  - restaurant,
  - library,
  - lab,
  - other and
  - not being detected
- 12 time units in the activity network, from 7am to 7pm

Proposal distribution (using simple random sampling)

Description	Coeff. estimate	Robust Asympt. std. error	t-stat
$\beta_{\textbf{NA, 17-19, employees}}$	0.263	0.0302	8.70
$\beta_{\textbf{NA, 14-17, students}}$	-0.222	0.191	-1.16
$\beta_{\textbf{NA, 7-8, students}}$	0.349	0.0281	12.44
$\beta_{\textbf{NA, 7-9, employees}}$	0.326	0.0262	12.43
$\beta_{\textbf{NA, 17-19, students}}$	1.14	0.187	6.09
$\beta_{\textbf{classroom, 12-14, students}}$	-0.336	0.337	-1.00
$\beta_{\textbf{classroom, 7-12, employees}}$	-0.723	0.397	-1.82
$\beta_{\textbf{classroom, 7-12, students}}$	0.598	0.262	2.28
$\beta_{\textbf{library, 14-19, employees}}$	-0.624	0.553	-1.13
$\beta_{\textbf{library, 12-14, employees}}$	-0.575	0.481	-1.20
$\beta_{\textbf{library, 7-12, employees}}$	-1.57	0.508	-3.09
$\beta_{\textbf{office, 14-19, employees}}$	1.41	0.246	5.73
$\beta_{\textbf{office, 7-12, employees}}$	1.12	0.228	4.92
$\beta_{\textbf{restaurant, 14-19, students}}$	-0.410	0.185	-2.21
$\beta_{\textbf{restaurant, 12-14, employees}}$	0.136	0.0259	5.26
$\beta_{\textbf{restaurant, 12-14, students}}$	0.665	0.286	2.32
	...		

Number of observations = 1087

Number of estimated parameters = 43

$$\mathcal{L}(\beta_{\mathbf{0}}) = -5016.636$$

$$\mathcal{L}(\hat{\beta}) = -453.225$$

$$\rho^2 = 0.910$$

$$\bar{\rho}^2 = 0.901$$

Target weight (using simple random sampling)

Description	Coeff. estimate	Robust Asympt. std. error	t-stat
$\beta_{\text{library 7-12, employees}}$	-2.08	0.422	-4.93
$\beta_{\text{office 7-12, 14-19, employees}}$	1.69	0.393	4.30
$\beta_{\text{restaurant 12-14, employees}}$	1.22	0.502	2.43
$\beta_{\text{shop 12-14, students}}$	-7.36	1.24	-5.92
$\beta_{\text{shop 7-12, 14-19, students}}$	-1.16	0.538	-2.16
$\beta_{\text{NA 7-8, students}}$	4.27	0.995	4.29
$\beta_{\text{NA 8-12, students}}$	1.40	0.498	2.82
$\beta_{\text{NA 17-19, students}}$	1.75	0.568	3.08
$\beta_{\text{NA 9-17, employees}}$	1.43	0.296	4.84
$\beta_{\text{NA 7-9, 17-19, employees}}$	3.34	0.554	6.02
$\eta_{\text{Office, Lab, Classroom}}$	5.22	0.764	6.83
$\eta_{\text{Restaurant, Library, Other}}$	7.85	1.11	7.10
η_{Shop}	7.33	0.894	8.20
η_{NA}	2.75	0.393	7.00
$\beta_{\text{3+ lab episodes}}$	-5.03	0.952	-5.28
$\beta_{\text{3+ resto episodes}}$	-2.50	0.759	-3.29
	...		

Number of observations = 1087
Number of estimated parameters = 22

$$\mathcal{L}(\beta_0) = -5016.636$$
$$\mathcal{L}(\hat{\beta}) = -47.218$$
$$\rho^2 = 0.991$$
$$\bar{\rho}^2 = 0.986$$

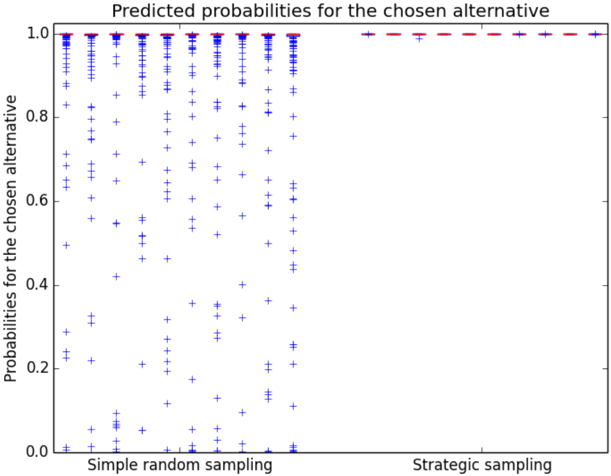
Model using strategic sampling

Description	Coeff. estimate	Robust Asympt. std. error	t-stat
$\beta_{\textbf{classroom 7-12, students}}$	0.478	0.238	2.01
$\beta_{\textbf{restaurant 12, students}}$	2.69	0.527	5.10
$\beta_{\textbf{shop 14-19, students}}$	1.46	0.343	4.27
$\beta_{\textbf{NA 7-12, students}}$	2.33	0.285	8.17
$\beta_{\textbf{NA 17-19, students}}$	2.83	0.343	8.24
$\beta_{\textbf{NA 7-9, 17-19, employees}}$	2.91	0.303	9.60
$\eta_{\textbf{office, lab, classroom}}$	-6.85	0.379	-18.09
$\eta_{\textbf{restaurant, library, other}}$	-6.58	0.360	-18.31
$\eta_{\textbf{shop}}$	-3.72	0.278	-13.40
$\eta_{\textbf{NA}}$	-7.63	0.541	-14.12
$\beta_{\textbf{0 restaurant episode}}$	4.11	0.365	11.28
$\beta_{\textbf{0 classroom episodes, employees}}$	10.3	0.887	11.65
$\beta_{\textbf{1 shop episodes}}$	-3.87	0.573	-6.76
$\beta_{\textbf{2+ shop episodes}}$	-3.49	1.08	-3.24
$\beta_{\textbf{0 library episode, employees}}$	2.72	0.335	8.10
$\beta_{\textbf{0 library episode, students}}$	4.77	0.495	9.64
...			

Number of observations = 1087
Number of estimated parameters = 39

$$\mathcal{L}(\beta_0) = -5016.636$$
$$\mathcal{L}(\hat{\beta}) = -400.633$$
$$\rho^2 = 0.920$$
$$\bar{\rho}^2 = 0.912$$

Validation



Predicted probabilities for the chosen alternative
Probabilities for the chosen alternative
1.0
0.8
0.6
0.4
0.2
0.0
Simple random sampling
Strategic sampling

More details

- Conference proceeding: [DB15]
- Chapter 4 in [Dan15b]

Outline

Goal

- Model location choice conditional on an activity type
- Adapted to panel data

Static model

$$U_{int} = V_{int} + \varepsilon_{int}$$

Ignores two aspects:

- Dynamics
- Serial correlation

Dynamic model without agent effect

$$U_{int} = V_{int} + \rho y_{in(t-1)} + \varepsilon_{int}$$

Assumes

- Dynamic process of order one
- Location-specific dependence
- Previous choice $y_{in(t-1)}$ independent of error term ε_{int}

Relaxing the independence assumption of error terms

- Agent effect α_{in}: time-invariant factor (*"between" individuals variability*)
- Unobserved heterogeneity ε'_{int}: short-term variation of probabilities (*"within" an individual variability*)

$$U_{int} = V_{int} + \rho y_{in(t-1)} + \alpha_{in} + \varepsilon'_{int}$$

Endogeneity issue:

- $y_{in(t-1)}$ and α_{in} are correlated

An approach by Wooldridge [Woo05]

For activity location i, individual n, at time t:

Lagged variable

Agent effect

$$U_{int} = V_{int} + \rho y_{in(t-1)} + \alpha_{in} + \varepsilon'_{int}$$

$$\alpha_{in} = a + by_{in0} + c'\bar{x}_n + \xi_{in}$$

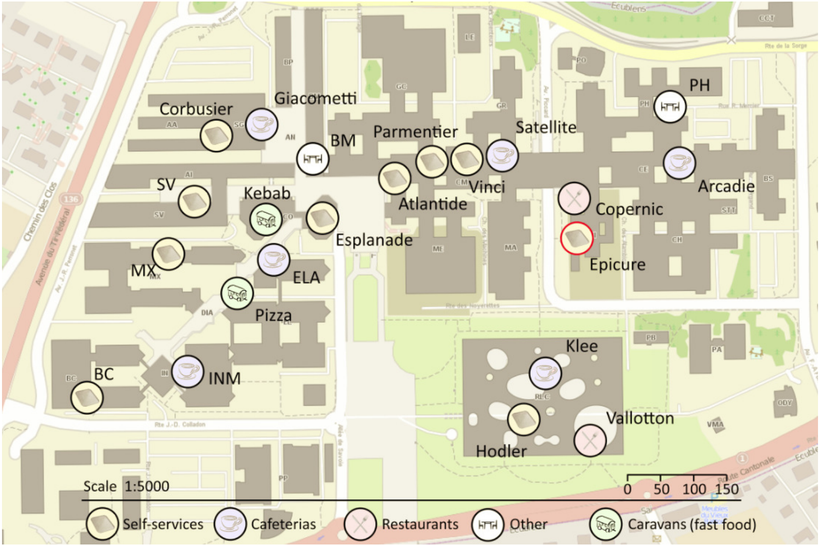
$$\sim N(0; \Sigma_\alpha)$$

Endogeneity issue solved [Woo05]

3 different models

Static model	Dynamic model without agent effect	Dynamic model with agent effect
$\rho = 0$ $a, b, c, \sigma_\alpha^2 = 0$	$\rho \neq 0$ $a, b, c, \sigma_\alpha^2 = 0$	$\rho \neq 0$ $a, b, c, \sigma_\alpha^2 \neq 0$

Case study: EPFL catering locations



Two specifications of the agent effect

- First choice

$$\alpha_{in} = a + by_{in0} + \xi_n$$

- First choice and frequency

$$\alpha_{in} = a + by_{in0} + cy_{int}^{\text{count}} + \xi_n$$

$$\sum_{t'=1}^{t-1} I(y_{int'})$$

4 models estimated

Static model	Dynamic model without agent effect	Dynamic model with agent effect correction	
		First choice	First choice and frequency
$\rho = 0$	$\rho \neq 0$	$\rho \neq 0$	$\rho \neq 0$
$a = 0$	$a = 0$	$a \neq 0$	$a \neq 0$
$b = 0$	$b = 0$	$b \neq 0$	$b \neq 0$
$c = 0$	$c = 0$	$c = 0$	$c \neq 0$
$\sigma_\alpha^2 = 0$	$\sigma_\alpha^2 = 0$	$\sigma_\alpha^2 \neq 0$	$\sigma_\alpha^2 \neq 0$

Estimation results

- **Distance** has a negative impact
- **Yearly evaluation** has a positive impact
- **Beer after 14:00** has a positive impact
- **Cost** has a negative impact
- **Dinner** has a positive impact
- **Capacity** has a positive impact

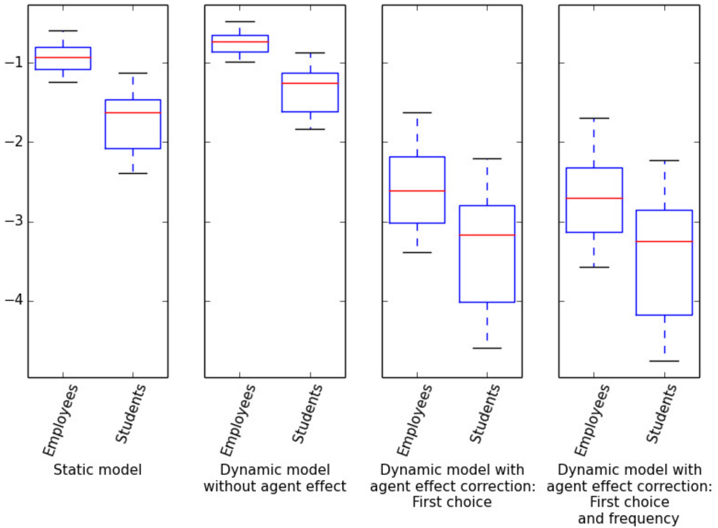
Likelihood ratio tests

Static model	**Dynamic model without agent effect**	**Dynamic model with agent effect correction**	
		First choice	First choice and frequency
	354.003 (> 5.99)	920.354 (> 58.12)	16.172 (> 5.99)

Validation

	Predicting last observations based on past observations			
	Static model	Dynamic model without agent effect	Dynamic model with agent effect correction	
			First choice	First choice and frequency
Sum of the squares of the errors	232.95	204.01	184.16	173.85

Elasticities to price

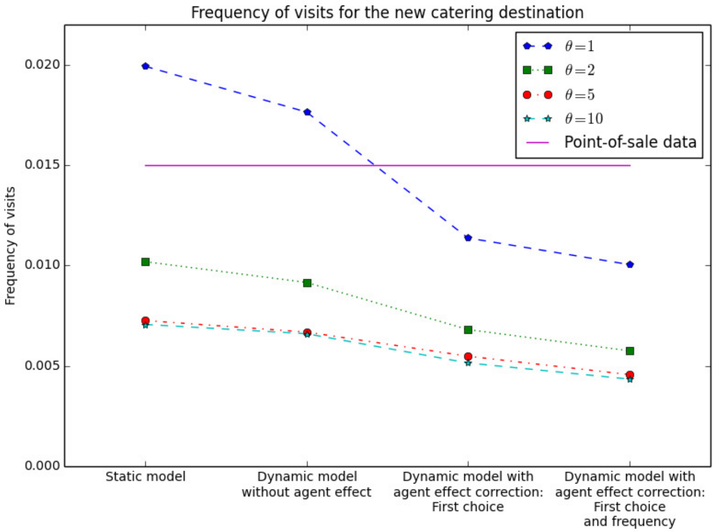


Forecasting: opening a new catering location

Nesting structure with the most similar alternative

- Nesting parameter $\theta = 1$: logit model, independent error terms
- Nesting parameter $\theta \to \infty$: perfectly correlated error terms

Forecasting: opening a new catering location



Outline

Motivation: Understanding pedestrian demand

Detecting activity-episode sequences

A path choice approach to activity modeling

Location choice with panel effect

Conclusion and future work

Activity-episode sequence detection

- Explicit modeling of the imprecision in the measure
- Usage of prior knowledge of the infrastructure
- Avoidance of the pingpong effect

Activity-path choice model

- No tours, no priorities
- Managing large choice sets
- Unique utility for activity type, time-of-day and duration choices

Location choice model

- Including panel data
- Correcting for serial correlation

Limitations

- Activity purpose is extracted from map data
- No mode detection
- No congestion

Future work

- Congested case study
- Include the uncertainty from detection in modeling
- Metropolis-Hastings algorithm for the sampling of activity paths
- More complex correlation structure for the choice of an activity path
- Include other sources of endogeneity (group, queue)

Thank you

PhD thesis:

Activity choice modeling
for pedestrian facilities

Antonin Danalet

– antonin.danalet@epfl.ch

Bibliography I

Theo A Arentze and Harry J.P Timmermans.
A learning-based transportation oriented simulation system.
Transportation Research Part B, 38(7):613–633, aug 2004.

Maya Abou-Zeid and Moshe Ben-Akiva.
Well-being and activity-based models.
Transportation, 39(6):1189–1207, jan 2012.

John L Bowman and Moshe Ben-Akiva.
Activity-based disaggregate travel demand model system with activity schedules.
Transportation Research Part A, 35(1):1–28, jan 2001.

Bibliography II

Antonin Danalet.
A Bayesian Approach to Detect Pedestrian Destination-Sequences from WiFi Signatures: Data (Transp. Res. Part C, 2014), 2015.

Antonin Danalet.
Activity choice modeling for pedestrian facilities.
PhD thesis, EPFL, 2015.

Antonin Danalet and Michel Bierlaire.
Importance sampling for activity path choice.
In *15th Swiss Transport Research Conference (STRC)*, Monte Verità, Ascona, Switzerland, 2015.

Bibliography III

Antonin Danalet, Bilal Farooq, and Michel Bierlaire.
A Bayesian approach to detect pedestrian destination-sequences from WiFi signatures.
Transportation Research Part C, 44:146–170, 2014.

Gunnar Flötteröd and Michel Bierlaire.
Metropolis-Hastings sampling of paths.
Transportation Research Part B, 48:53–66, feb 2013.

Serge P Hoogendoorn and Piet H L Bovy.
Pedestrian route-choice and activity scheduling theory and models.
Transportation Research Part B, 38(2):169–190, 2004.

Bibliography IV

Ann Hendrich, Marilyn P Chow, Boguslaw a Skierczynski, and Zhenqiang Lu.
A 36-hospital time and motion study: how do medical-surgical nurses spend their time?
The Permanente journal, 12(3):25–34, 2008.

Sofia Kalakou, Michel Bierlaire, and Filipe Moura.
Effects of terminal planning on passenger choices.
In *14th Swiss Transport Research Conference (STRC)*, Monte Verità, Ascona, Switzerland, 2014.

Bibliography V

Sofia Kalakou, Filipe Moura, and Valério Medeiros.
Analysis of airport configuration and passenger behaviour.
In *Proceedings of the 10th International Space Syntax Symposium (SSS10)*, London, 2015.

Jee Eun Kang and Will Recker.
The location selection problem for the household activity pattern problem.
Transportation Research Part B, 55:75–97, sep 2013.

Bibliography VI

Jason D. Lemp and Kara M. Kockelman.
Strategic sampling for large choice sets in estimation and application.
Transportation Research Part A: Policy and Practice, 46(3):602–613, mar 2012.

Xuan Liu, John M. Usher, and Lesley Strawderman.
An analysis of activity scheduling behavior of airport travelers.
Computers and Industrial Engineering, 74(1):208–218, 2014.

Danique Ton.
NAVISTATION: a study into the route and activity location choice behaviour of departing pedestrians in train stations.
Master thesis, Delft University of Technology, 2014.

Bibliography VII

Jeffrey M. Wooldridge.
Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity.
Journal of Applied Econometrics, 20(1):39–54, 2005.

Yuji Yoshimura, Stanislav Sobolevsky, Carlo Ratti, Fabien Girardin, Juan Pablo Carrascal, Josep Blat, and Roberta Sinatra.
An analysis of visitors' behavior in The Louvre Museum: a study using Bluetooth data.
Environment and Planning B: Planning and Design, 41(6):1113–1131, 2014.

Privacy issues in this thesis

- EPFL ethics committee:
  - “No personal identifier when sharing data”
- In practice:
  - We have no access to MAC addresses in our dataset
  - The dataset is public [Dan15a]