

Semantically enriching mobile subscribers' trajectories

Elis Kõivumägi

University of Tartu

Outline

- Introduction
- Data overview
- Algorithm description
- Results
- Conclusion and next steps
- Acknowledgments

Introduction

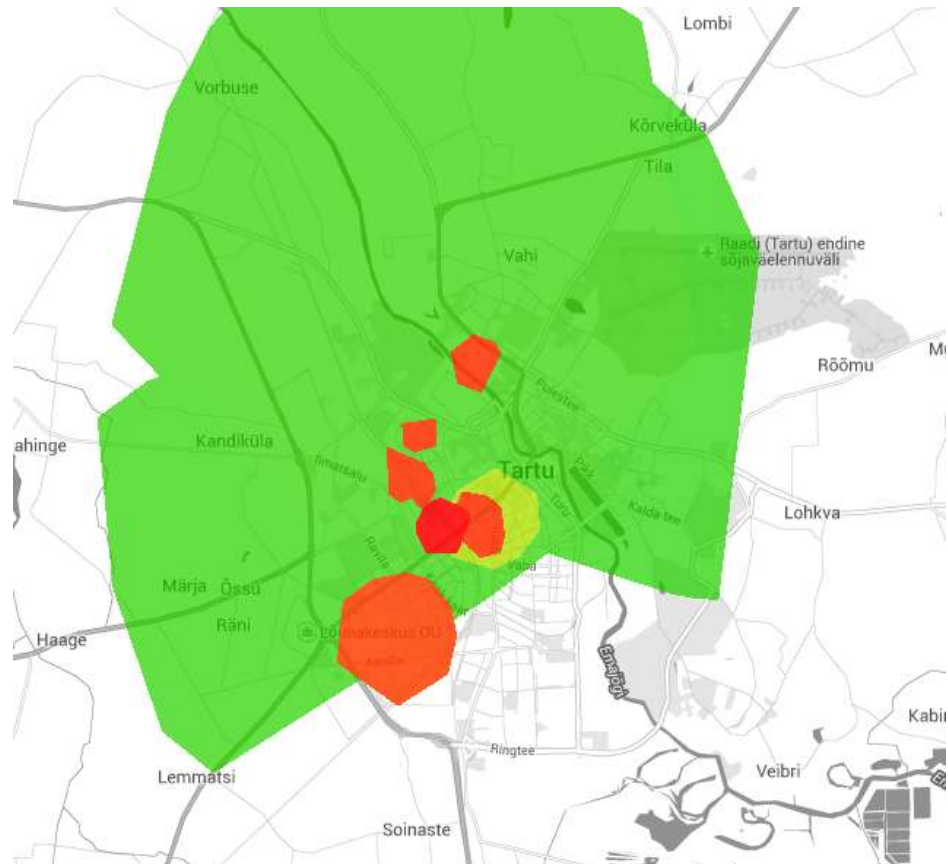
- I've been working in co-operation with mobile operators and location data for several years
- Published a paper about framework for home/work detection and real time labeling
- PhD thesis is about profiling mobile subscribers

Data (1)

- 20 people
 - 12 months
- Location events from
 - Real time event stream (calls, SMS messages)
 - Raw data events (sessions)
 - Cell level accuracy
 - Average cells: 10-11(in a day)
 - Average number of events: 60-70 (in a day)
 - No lat,lon coordinates at GPS accuracy

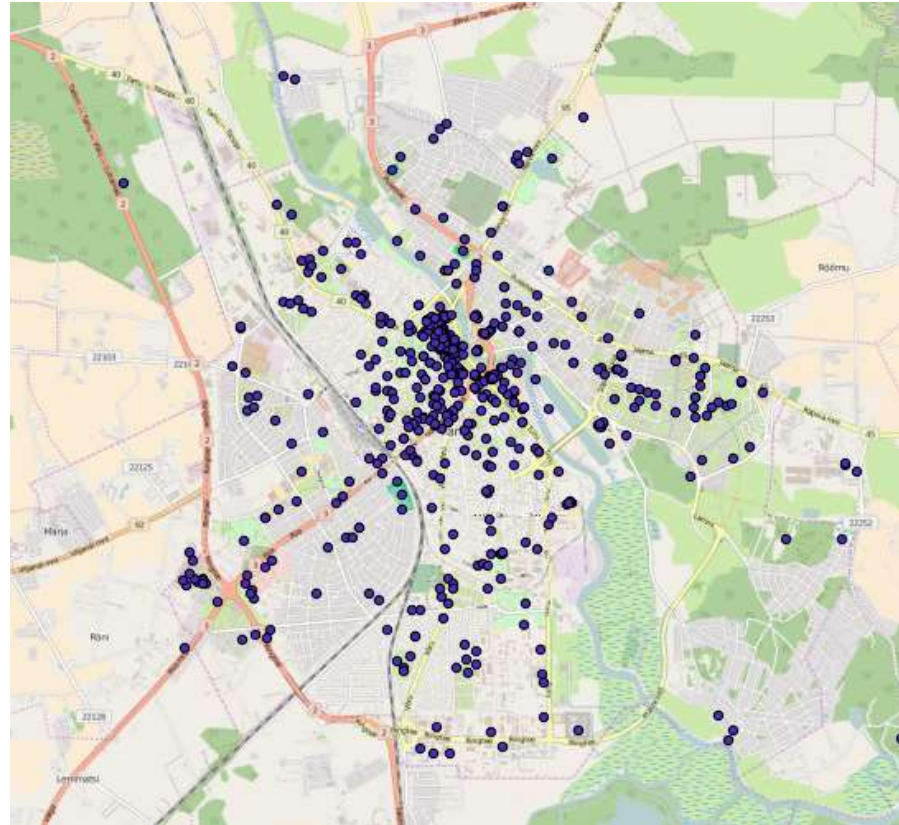
Data (2)

- Cellplan
 - Coverage areas for -90dbm
 - Required for mapping POIs to events -> from events to subscribers



Data (3)

- POI data
 - Nr of objects ~8000
 - Categories
 - Education
 - Sport
 - Food
 - Shopping
 - Entertainment



Algorithm (1)

- Read input location data to cell by hour matrix
 - Reverting to hourly accuracy
- Configurable
 - By hours
 - By days
 - How many days
- Read input cell data
 - Can be configured to ignore cells larger than X km²
 - Generate cell intersections array
 - Cell areas overlap

Algorithm (2)

- Detect episodes using cell intersection array
 - All episode cells must overlap with each-other
 - Each cell can only be part of one episode
 - Disregard large cells by eliminating them from the input
- Episodes will consist of:
 - Cell IDs
 - Hours for each cell
 - Event counts for each hour

Algorithm (3)

- Find POIs for each cell
 - Lat, lon inside of cell coverage area
 - Cell-POI dictionary
- Configure POI visiting hours

POI type	Mon-Fri	Sat-Sun
Sport	7:00-9:00 or 18:00-21:00	14:00-18:00
Entertainment	19:00-02:00	18:00-02:00
Shopping	17:00-20:00	10:00-18:00
Food	12:00-14:00	18:00-20:00
Education	9:00-15:00	-

Algorithm (4)

- Find semantics for episodes based on cell-POI dictionary
- Only select those POIs to episodes that have matches to POI visiting hours
- Weights for POIs will be calculated using event counts for episode cells only for valid hours
- The higher the episode hours match rate to POI hours the better
- Also supports adjusting weights to POI frequencies per cells but currently disabled

Example (1)

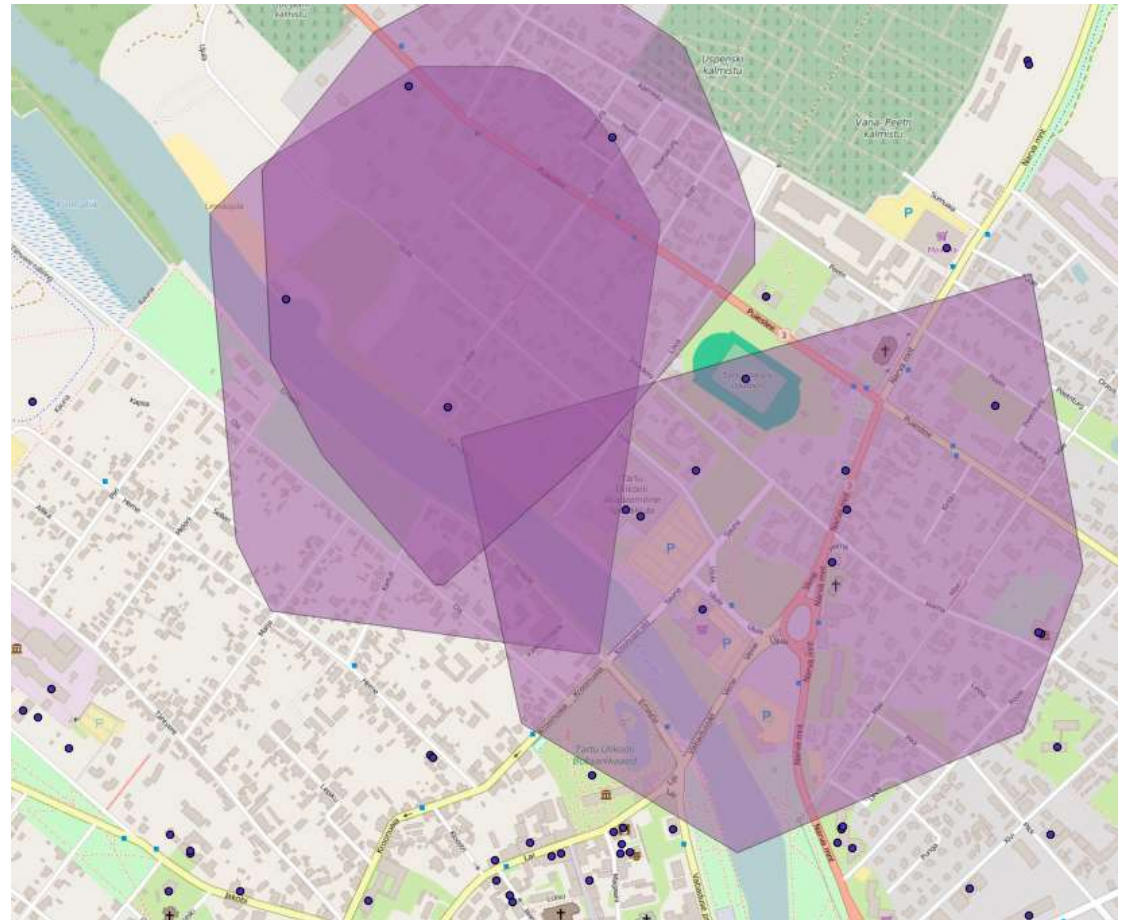
- Episode = {C1, C2, C3, C1}
- Hours = {18, 18, 20, 19}
 - Sport = {7,8,9,18,19,20,21}
 - Education = {9,10,11,12,13,14,15}
 - Food = {12,13,14} //lunch
 - Shopping = {17,18,19,20}
- Events = {25, 18, 4, 20}
 - Sum(Events) = 67
- POIs
 - C1 = {sport, education, shopping}
 - C2 = {sport, food}
 - C3 = {food, shopping}
- Weights
 - Sport = 0.94
 - Shopping = 0.73

Example (2)

- Adjust weights depending on matching hours
- In this example weights remain the same because all episode hours are within both Sport and Shopping category hours
- Final weights:
 - Sport = 0.94
 - Shopping = 0.73

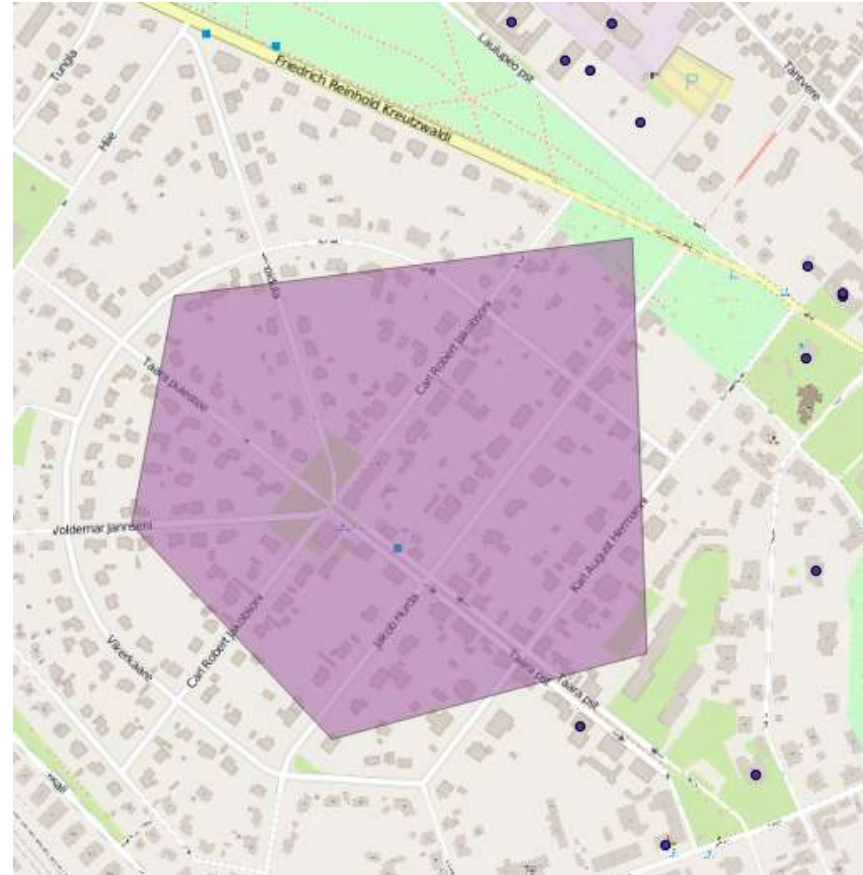
Results 1

- POIs:
 - Sport: 0.32
 - Shopping: 0.32
 - Entertainment: 0.03
- Hours
 - 12, 18, 18, 19
- Before adjustment
 - Sport: 0.42
 - Shopping: 0.42
 - Entertainment: 0.14



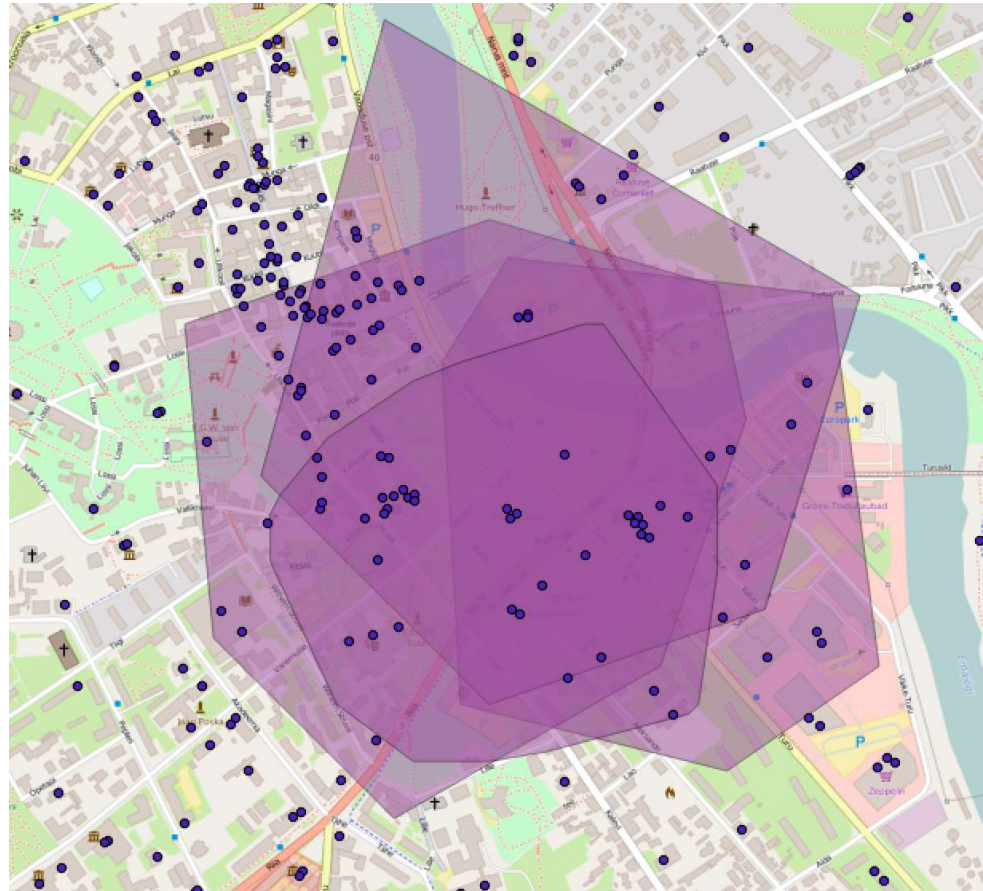
Results 2

- POIs
 - -
- Hours
 - 8, 19



Results 3

- POIs
 - Education: 0.37
 - Food: 0.21
- Hours
 - 13, 14, 15
- Before adjustment
 - Education: 0.46
 - Food: 0.53



Conclusions/challenges

- Good way to find where people have lunch, train, shop/buy groceries
- Difficult to identify:
 - Other repetitive activities – drop off children to school/kindergarten
 - Trajectory density issue
- Inaccurate cell coverage areas reduces the accuracy drastically
 - -90dbm seems too pessimistic
 - On the other hand – increasing the area will have a reverse effect
- Use home/work locations to decouple episodes
- Blacklist cells with no POIs or find nearest POI(s)?

Acknowledgments

- University of Tartu and Distributed System Group
- STACC
 - Software and Technology Competence Center
- Reach-U



STACC

Reach  U



European Union
Regional Development Fund



Investing in your future

Distributed Systems Group
University of Tartu

 positium