



Practical simulation of virtual crowds using points of interest



Matti Pouke*, Jorge Goncalves, Denzil Ferreira, Vassilis Kostakos

Center for Ubiquitous Computing, University of Oulu, Finland

ARTICLE INFO

Article history:

Received 23 May 2015

Received in revised form 4 February 2016

Accepted 11 February 2016

Available online xxx

Keywords:

Smart cities

Virtual environments

Pedestrian modelling

ABSTRACT

We present a computational method that exploits points of interest (POIs) to generate realistic virtual pedestrians for a city model, i.e., a simulated crowd. Our method is validated using mobility traces collected longitudinally from a city-wide free and open Wi-Fi network in downtown Oulu, Finland. Analysing this data, we first construct a time-varying Origin–Destination matrix that describes how individual pedestrians in our city move at different times and places. We compare this ground-truth against a random pedestrian model to investigate how the latter underestimates or overestimates movement at various locations or times of day. By identifying these deviations, we can calibrate a weighted model that uses POIs from OpenStreetMap to adjust the simulated crowd. Our results show a significant accuracy improvement over the random model, while at the same time our work is readily applicable to simulating crowds in other cities (real and virtual) as long as POI can be defined spatially.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Understanding how humans navigate in urban spaces has been the interest of scientific disciplines ranging from psychology through civil engineering to computer science. The specifics of the navigation depend on how an individual understands the physical setting in which the navigation takes place. How the physical setting is specifically understood always depends on each individual. However, environments such as cities contain properties which are more or less common to all individuals navigating them. Kevin Lynch called “the public image” of a city to its collective understanding (Lynch, 1960). This public image is the overlap of the individual images, and according to the study data of Lynch et al., this can be described consisting of the following properties: *paths, edges, districts, nodes and landmarks*. Paths are the channels through which individuals move. Edges divide areas physically or metaphorically. Districts are geographical medium-to-large areas of the city. Nodes are strategic points that can be street junctions or other types of crossings or convergence of paths. Landmarks are distinctive physical points, which can be observed visually such as churches, parks, skyscrapers or mountaintops (Lynch, 1960).

The acquisition of spatial knowledge, both metric and qualitative, begins as soon as an individual starts navigating in a new environment (Montello, 1998). Through increased exposure and familiarity, the quality of the spatial knowledge increases, and with enough exposure, an individual can connect separately learned places in a larger spatial

understanding (Montello, 1998). It is suggested that individuals organise spatial knowledge according to *anchor points*, salient locations that form the cognitive map that the individual uses to navigate. Besides geographical points, such as landmarks, anchor points can be path segments, nodes or even distinctive areas, similar to city properties categorized by Lynch (Lynch, 1960), (Golledge & Gärling, 2002). Anchor points of an individual usually consist of home, work place (or similar) as well as other locations that are somehow meaningful for the navigation of said individual. Especially striking locations, such as famous landmarks might be common to almost every visitor of the city while some anchor points are shared between various demographics (Couclelis, Golledge, Gale, & Tobler, 1987; Golledge & Gärling, 2002; Golledge & Spector, 1978).

As individuals navigate between places, they usually associate locations according to their closest anchor points (Couclelis et al., 1987). These associations form regions connected to anchor points and individuals tend to displace these regions within the direction of the anchor points, causing metric distortions in their mental models (Couclelis et al., 1987). This is related to the findings of human tendency to store spatial information according to hierarchical categorization (Couclelis et al., 1987). A largely cited example is the experiment of Stevens and Coupe (Stevens & Coupe, 1978), where US citizens systematically considered the city of Reno, Nevada to be east of San Diego, California because the state of Nevada is associated to being east from California (Stevens & Coupe, 1978).

The empirical macro-scale analysis performed by Manley et al. (Manley, Addison, & Cheng, 2015) supports the theory that anchor points play a dominant role in urban navigation. Their findings from observing 700,000 minicab routes within London suggest that urban anchor points are more suitable for basis of urban travelling models

* Corresponding author at: Center for Ubiquitous Computing, Erkki Koiso-Kanttilan katu 3, door E, P.O. Box 4500, FI-90014, University of Oulu, Finland.

E-mail addresses: matti.pouke@ee.oulu.fi (M. Pouke), jgoncalv@ee.oulu.fi (J. Goncalves), denzil.ferreira@ee.oulu.fi (D. Ferreira), vassilis@ee.oulu.fi (V. Kostakos).

instead of cost-minimisation by shortest routes. Accordingly, the minicab drivers rarely followed route choices given by shortest path routing methodologies but instead repeatedly established common anchors in route choices (Manley, Addison and Cheng, 2015).

As stated in (Golledge & Spector, 1978), while anchor points are always subjective, groups of people might share some anchor points. In a city scale this means that, college students might share their campus as a common anchor point as well as people living in the same neighbourhood might share anchor points related to that neighbourhood. However, the work of (Bromley, Tallon, & Thomas, 2003) points out that different demographics can also be segregated according to time, at least when examining inner city use. Daytime visitors of inner city visit different locations than night time visitors and often belong to different demographics.

In this work, we examine the use of crowdsourced points of interest (POIs) as a means to generate anchor points into city microsimulations. The rationale for exploiting POIs as an alternative for such a purpose, is the ease of which they can be acquired. Our attempt is not to provide a complete microsimulation system that would provide alternatives to complete activity-based microsimulation models, such as the social force model (Helbing & Molnar, 1995). Instead, we study POIs in isolation as time varying inner city anchor points. This can help, for example, in identifying after-hour crime hotspots (Nelson, Bromley, & Thomas, 2001). Here, temporal data is acquired from a time-varying Origin–Destination (OD) matrix, a result from analysing municipal Wi-Fi network user data. In addition to examining crowdsourced POIs as anchor point data, this work contributes to pedestrian microsimulation research by demonstrating how a municipal Wi-Fi network can be used to simulate granular pedestrian mobility.

2. Related work

2.1. A cursory overview of travel behaviour studies

Simulation of the total flow of pedestrians across a city has been repeatedly addressed in research concerning topological studies of street networks. Space Syntax has been used to describe how and why people move through certain routes within a city (Hillier & Hanson, 1984). Space Syntax literature suggests that roughly 70% of pedestrian volume at particular street segments can be predicted from the closeness connectivity metric of the street network, while Jiang and Jia (2011) claim that a weighted version of PageRank is a more suitable metric. They also argue that the underlying street network is the most influential factor in guiding pedestrian movements, and therefore randomly moving agents and real pedestrians move essentially in a same way through the same street network. A space syntax study by Lerman and Omer (2016) combined land use, physical properties of road sections as well as demographics information with street connectivity to study the relative contribution of each of the aforementioned dimensions to pedestrian movement. Their findings substantiate the claims of previous studies (Hillier & Hanson, 1984; Jiang & Jia, 2011); street connectivity was the most significant contributor to pedestrian movement and can in itself cause changes in the physical properties of road sections and land use (Lerman & Omer, 2016).

However, city topology alone is not sufficient to fully describe all travel behaviour aspects, and therefore travel behaviour has been studied across multiple disciplines such as geography, urban planning, transportation and even computer science research.

A land-use approach has been often used in transportation research since the early 20th century. It describes the characteristics of travel behaviour between different types of land use, such as the traffic between residential zones and industrial zones. Alan M. Voorhees (2013) described how travel between different types of origins and destinations roughly follows gravitational laws, with different types of destinations generating certain types of “pull” towards the origins. The amount of the pull in all types of destinations depends on the size of the origin

and destination, as well as the travel time between them. However, how the theoretical pull is calculated, depends on the type of the travel, i.e., the types of the origin and the destination (Voorhees, 2013). This gravitational law is still often revisited in research, such as in the work of Simini, González, Maritan, and Barabási (2012); their radiation model can predict mobility patterns using population density as the only input data, eliminating the need for parameter adjustments. Land use has a different effect on various aspects of travel behaviour, such as trip generation, distance travelled and choice of mode (M. G. Boarnet, Joh, Siembab, Fulton, & Nguyen, 2011). Criticism towards the effect of land use to travel behaviour has been presented by for example Boarnet and Sarmiento (1998); Stead (2001) and Ewing, Deanna, and Li (1996). In any case, transportation planning and land use continue to meet heavily in research, as seen in (Waddell, Ulfarsson, Franklin, & Lobb, 2007).

Besides land use, it is possible to also consider smaller-scale segregation of locations into urban research, while time also plays an important role. Land use research usually focuses on travel behaviour between, for example, residential zones, industrial zones and shopping districts. However, there are also significant differences in the use of inner city locations according to the time of the day and week (Bromley et al., 2003).

Activity based models have often been used to estimate Travel Behaviour since the 1990s. Such models rely on the fact that people travel because they have needs, activities to which they must tend. How these activities are scheduled, given various conditions, such as household characteristics, properties of potential destinations and the state of the transportation network, is what activity based approaches seek to answer. However, activity based approaches have received criticism for their complexity and intense data requirements (Ettema & Timmermans, 1997). Activity-based models rely on realistic modelling on schedules of individuals and households which is not a simple task; properties such as speed of spatial knowledge acquisition and its role in scheduling decisions, assigning activities to utilities and interaction between household members are difficult to model as observed by (Kay W. Axhausen & Gärling, 1992). Since this observation, studies have emerged to tackle these problems. For example, the work of (Arentze & Timmermans, 2005) works in simulating spatial knowledge acquisition and (Zhang, Timmermans, & Borgers, 2005) model household interaction. However, intense data requirements of activity-based models are still a problem. It has been speculated whether it is even possible to gather exhaustive dataset for a truly precise activity based model (K W. Axhausen, 1998). While large-scale data collection efforts have been made, it is difficult to find a representative set of participants willing to commit to a long-term data gathering effort (K. W. Axhausen, Zimmermann, Schönfelder, Rindsfuser, & Haupt, 2002).

Since the 1990s, activity based approaches have been common in travel behaviour studies. In a 2001 survey, Timmermans categorized these approaches into: constraints based models, utility-maximizing models, computational process models and microsimulation models. Microsimulation models – such as ours – attempt to simulate individual activity patterns according to probability conditions, while the other approaches infer rules and parameters from empirical data (Timmermans, Arentze, & Joh, 2002).

Microsimulations can either simulate all aspects of activity based approaches or concentrate on certain properties. RAMBLAS (Veldhuisen, Timmermans, & Kapoen, 2000) and TRANSIMS (Nagel & Rickert, 2001) are examples of microsimulation models that replicate city-wide traffic according to multiple parameters. There are, however microscopic simulations that are not activity-based approaches, but simulate pedestrian activity at the micro-level movement of individual pedestrians concentrating only on detailed aspects of pedestrian flow. These models can effectively analyse and simulate pedestrian flows and interactions through narrow spaces with varying sets of rules such as lanes and pathways, estimate concepts from vehicular traffic such as level of service and estimate effects of various types of pedestrians, such as the obese or the elderly, on pedestrian flow (Galiza & Ferreira, 2012; Guo, Wong,

Huang, Zhang, & Lam, 2010; Helbing, Molnár, Farkas, & Bolay, 2001; Helbing & Molnar, 1995; K Teknomo, 2002; Kardi Teknomo, 2006). The recent study conducted by Ronald, Arentze, and Timmermans (2016) addressed the issue of validating complex transport models, such as agent-based activity microsimulations. While the authors experimented with one model only, they argue that their suggested approach to sensitivity testing should be adopted to the validation of complex transport models in general (Ronald et al., 2016).

2.2. Empirical observation

Modern computer science offers various methods for observing pedestrian traffic. Broadly speaking, the different data collection methods can be classified as *Direct Observation or Investigation*, *Scene Analysis*, *Proximity Sensing*, *Continuous Localization Systems*, and *Sensor Networks* where the latter method can be a combination of any of the previous methods (Bandini, Federici, & Manzoni, 2007). *Direct Observation* refers to observing individuals directly. In *Scene Analysis* individual pedestrians are not tracked, but the behaviour of a crowd is estimated by observing crowd densities and flows from image or video footage. *Proximity Sensing* refers to detecting passing entities using fixed or mobile sensors and *Localization Systems* in tracking individual entities' locations continuously using techniques such as GPS.

Computer vision allows for rather accurate tracking and analysis of crowd behaviour. Jacques Junior et al. (Jacques Junior, Raupp Musse, Jung, & Junior, 2010) classify computer vision crowd tracking methods into *object-based approaches* and *holistic approaches*. *Object-based approaches* focus on tracking individual pedestrians while *holistic approaches* treat the crowd as a single entity following a top-down methodology (Jacques Junior et al., 2010). Although machine vision techniques are very powerful, their obvious limitation is the restriction of space. The number and location of cameras limit the spaces that can be studied.

The popularity of mobile phones has provided the research community with various alternatives for collecting traffic data (pedestrian and otherwise). A popular way for tracking an individual with a cell phone would be using the phones built-in GPS. Work and Tossavainen successfully transformed GPS traces into a velocity field describing highway traffic (Work & Tossavainen, 2008). GPS data was also successfully utilized by Xiao, Juan, and Zhang (2015), who used Bayesian Networks to distinguish between travel modes from GPS data. Tracking a large number of pedestrians using GPS is not a straightforward task because large numbers of people have to specifically volunteer. However, Wirz, Franke, Roggen, Mitleton-Kelly, Lukowicz, and Tröster (2012, 2013) successfully estimated pedestrian movement and crowd densities at mass events using a subset of event attendees as probes who voluntarily shared their location using a mobile phone application (Wirz et al., 2012, 2013). The key to success was to convince people to install the required app by offering benefits such as real-time event related information (Wirz et al., 2012, 2013). However, this also suggests that tracking subsets of a crowd may provide enough information to reconstruct the movement of the whole crowd.

Other mobile phone related data such as in-network localization have been harnessed for tracking pedestrian movement at large. Calabrese et al. have estimated city-wide traffic by recording network bandwidth usage from signalling events (Calabrese, Colonna, Lovisolo, Parata, & Ratti, 2011; Calabrese, Pereira, & Lorenzo, 2010). The limitation of in-network localization is the typically large distance between cell towers ranging from average distance of 250 m to 500 m (Calabrese et al., 2011, 2010). Another work has investigated using online behaviour as a proxy for studying urban mobility by correlating urban traffic patterns with online search trends (Kostakos, Juntunen, Goncalves, Hosio, & Ojala, 2013).

Reflecting on the theoretical background and related research presented earlier, our contribution is as follows. We show that crowdsourced POI data can act as a simple data source for estimating

inner city pedestrian traffic. POIs lend themselves easily to the emulation of anchor points. They are salient locations that are known by at least part of the city population in one way or another. While POIs primarily refer to geographic points, POI frameworks such as the entity-component model described in (Heikkinen, Okkonen, Karhu, & Koskela, 2014) allow richer geographic description to POIs such as lines or polygonal areas, as well as the support for nested POIs. While it would be possible to model POIs that attempt to emulate anchor points as closely as possible, for example according to some or all categories described Lynch (Lynch, 1960), in this work we wish to specifically focus on crowdsourced POIs. Using crowdsourced POIs we wish to address the problem of data requirements in activity based models and provide a simple option to consider in activity based modelling or other similar travel behaviour simulations.

Our study is a microsimulation model of pedestrian movement similarly to agent navigation in virtual environments. While the POI data provides a geographical dataset, its pinpoint localisation is more precise than in land-use studies that rely on zoning. The category information of each POI acts as a source of activities that the simulated pedestrians pursue within inner city. We complement the activity data with temporal information obtained from the analysis of the municipal Wi-Fi logs of our city. This provides a heuristic model for simulating spatio-temporal inner city activities with minimal data collection effort. While the proposed model is unable to compete with more specific activity based models, we demonstrate that it can provide a more realistic pedestrian mobility that is explained with only street connectivity and should be considered as a supplement to data sources in activity-based models.

3. Modelling crowd movement

The topic of this work is the usage of crowdsourced POIs for travel behaviour simulation; for this purpose, we present a model called "The POI model" that simulates pedestrian mobility within the inner city. However, before we can define the POI model, we must first define two preliminary models, which are called "The ground truth model" and "The random model." These two models are used to generate temporal weights for POIs as well as for validation. We begin by describing our pedestrian simulation environment that defines the city geography as well as basic pedestrian movement, which is common to all models we present.

3.1. Pedestrian simulation environment

Our virtual city model describes the city centre of Oulu and covers approximately 1.5 km × 1.5 km area. This area was chosen for two reasons. Firstly, the area contains the inner city of Oulu with its shopping and nightlife districts, similarly to what was the focus of the works of (Bromley et al., 2003) and (Nelson et al., 2001). Secondly, the area has the densest distribution of Wi-Fi hotspots, allowing us to use dense empirical observations in the basis of our ground truth model. Borrowing the concept from virtual environments, the geography of the city model, in which pedestrian agents navigate, is defined as a navigation mesh (Demyen & Buro, 2006). In our city model, the navigation mesh is projected to a 2D plane where a single coordinate unit equals one metre, the negative Y axis points towards north and the positive X axis points towards east. The navigation mesh defines all the streets and pedestrian pathways within the modelled area as a triangle mesh. Although there are tools for defining navigation meshes automatically in virtual environments, in our case, since the simulated area is relatively small, the mesh was modelled manually with the Blender 3D modelling software. The modelling was carried out using the actual map of Oulu as a reference. Scaling the map so that 1 m in the map corresponds to 1 Blender coordinate unit, the navigation mesh was modelled as a horizontal plane going through all roads, clearings as well as minor pedestrian pathways that exist in the simulated area. The mesh was

modelled using as small number of triangles as possible so that the triangles could be efficiently used as nodes with pathfinding algorithms such as the A* (Hart, Nilsson, & Raphael, 1968). Fig. 1 displays a screenshot of the modelling process.

The movement of the agents takes place by calculating the path between origin and destination triangles, (nodes from now on) with the A* algorithm (Hart et al., 1968). The path is initially calculated through node centrepoints, after which it is smoothed with a funnel algorithm (Demyen & Buro, 2006). The funnel algorithm finds the shortest routes through adjacent nodes, which makes the resulting paths shorter and more likely to correspond to natural movement on a micro-level as the agents move through and within nodes. While the navigation mesh acts as a platform for moving agents across city, we also use its nodes for quantifying street usage and crowdedness, as described later on.

The autonomous agents act as virtual pedestrians in our ground-truth model. When a pedestrian agent is created, it is assigned a preferred walking speed chosen randomly between 1.25 m/s and 1.5 m/s (LaPlante & Kaeser, 2007). Unless interfered by other pedestrian agents, the agent continues to use this walking speed until removed from the simulation. Agent location is updated once per simulated second. Since the population of our simulated city is relatively low, we do not simulate the effect of crowdedness to walking speed at micro-level but estimate it as follows: Each node of the navigation mesh upkeeps their pedestrian density as $D = N/A$, where N is the number of agents within a node and A is the area of the node. To simulate the effect of street crowdedness, we use the pedestrian density-velocity relationship as described in (Seyfried, Steffen, Klingsch, & Boltes, 2005; Weidmann, 1993) to estimate agent walking speed while walking in proximity of other agents. The density-velocity relationship diagram identifies four density-domains with varying changes in the decrease of velocity; according to (Seyfried et al., 2005) and (Weidmann, 1993), changes in pedestrian dynamics can qualitatively observed between these domains. The slope of the density-dependent velocity decrease changes at the domain edges (as observed by Weidmann (1993)), which are $D = 0.7$, $D = 2.3$ and $D = 4.7$. The diagram with the domains partitioned has been reproduced in Fig. 2. We use the domains to estimate the agent walking speed by treating each domain as a line with a slope $m = y_2 - y_1 / x_2 - x_1$. We then find the corresponding walking speed with the linear equation $y = mx + b$ where x is the density within the node and m the slope of the density-domain in which x resides in. Walking speeds for each pedestrian agent within that node are then adjusted accordingly.

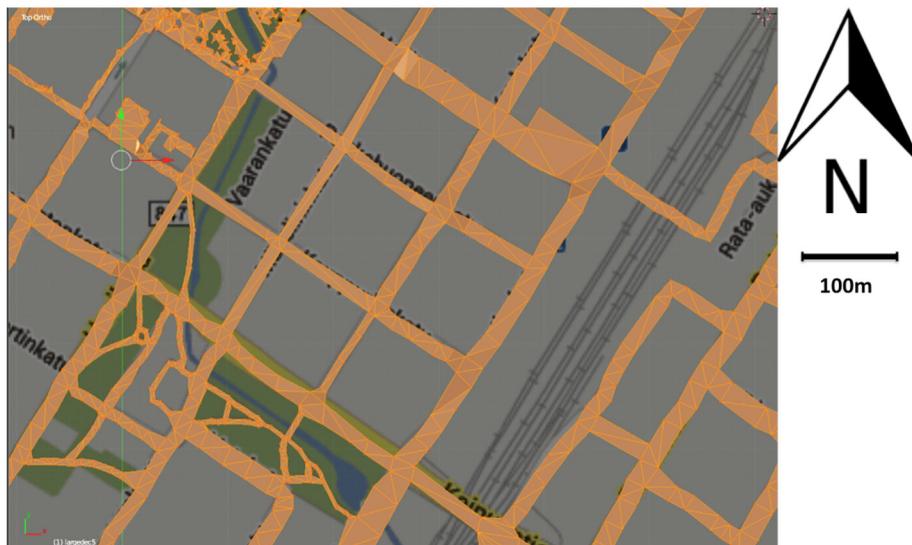


Fig. 1. Screenshot from the Blender modelling software coupled with a scale bar and a north arrow exemplifying the navigation mesh construction process. The inside of each triangle represents a node that defines a walkable area for pedestrian agents. Map data by Google.

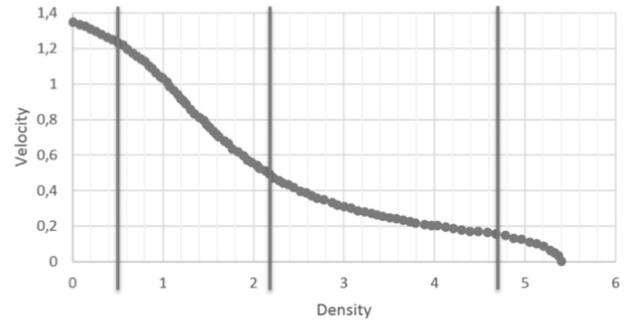


Fig. 2. The relation between density and velocity according to Weidmann (1993).

Although the model can successfully simulate the movement of thousands of agents in real-time, some optimizations were introduced to speed up the simulation for one month's worth of pedestrian traffic at a time. The density calculation $D = N/A$ is computed only when there is more than one pedestrian agent in a node. To find the number of pedestrians' N per node, the closest node centre for each agent is determined. The node centres are found using a kd-tree search (Maneewongvatana & Mount, 1999). The node of an agent is then determined by the closest centrepoint. This speeds up the simulation significantly in comparison to a point-in-triangle check for each pedestrian agent during every update. According to our analysis, this optimization does not significantly reduce accuracy as long as the triangles of the underlying navigation mesh do not contain extremely small angles, i.e. are closer to equilateral and right shapes than oblique. For this reason, our simulation ignores triangles with an area less than 1 m to prevent the agents from getting caught in small oblique "invisible" triangles that can sometimes result from the triangulation process. Finally, we multiply N by four to take into account the fact that the observations in the OD matrix reflect roughly one quarter of the total population. While we admit this is a simplified manner compensate for a smaller number of agents, we claim that this is sufficient for the purposes of this study, especially considering the small size and population of our city.

3.2. The ground-truth model

The ground truth model is based on the existence of a rich and detailed Wi-Fi mobility dataset that granularly describes the movement of individuals across our city. The dataset was generated by quantifying

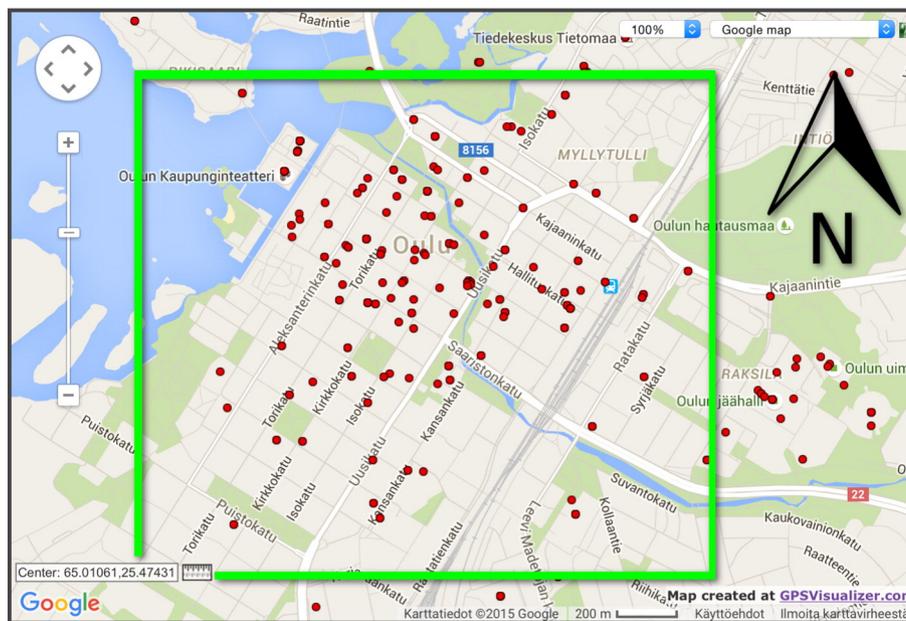


Fig. 3. The study area and its immediate surroundings. The exact study area is shown with a green rectangle. Each red dot marks the location of a Wi-Fi hotspot. Map data by Google. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the hourly connectivity and horizontal handover data of municipal Wi-Fi hotspots. In our case (Oulu, Finland), the network has approximately 1300 access points in 25 km². This basic analysis resulted in a time-varying origin–destination (OD) matrix, which describes the volume of pedestrians transitioning from any given location to any other given location at 1-h intervals over a large number of days. The Wi-Fi hotspots are not uniformly distributed across the city. In this paper we base our analysis on two one-month periods: May 2011 and May 2012, with each month having approximately 40,000 unique devices, which correspond to approximately 25% of the city population. While the OD matrix does not distinguish choice of travel mode, considering the simulation takes place in the inner city, we consider all trips to be made on foot. This is a simple generalisation we had to make because the mode of travel cannot be distinguished from the used dataset. A Google Maps visualisation of the study area including the Wi-Fi hotspots is in Fig. 3.

Our OD matrix does not contain any qualitative data about location types, but a very dense trip volume dataset between geographical points. Our virtual inner city model, which is a 1.5 km times 1.5 km, portion of the actual city, is populated with 249 virtual Wi-Fi access points (“hotspots” from now on) to act as targets for pedestrian navigation. The location of these virtual hotspots corresponds to the real-life ones found in our city, as listed in the OD matrix. The hotspots were mapped into the virtual model by simply transforming their GPS coordinates into the coordinate system of the virtual model. Because many real-life hotspots exist outside the scope of our virtual model, we filter the OD matrix data as follows. First we consider origin–destination pairs of hotspots that are both within the scope of our virtual city model as observations of how people move within the city. Next, origin–destination pairs where only one of the hotspots is within the scope of our virtual model are used to describe how agents enter or exit the city. Given the geographic landscape of our city, we define four major entry/exit directions: Northwest, Northeast, Southeast and Southwest. Effectively, hotspots within the virtual city model act as targets for agents, whereas the hotspots outside the scope of the model, so called *entry/exit hotspots*, are used for adding or removing pedestrians from the model.

Since we are modelling the ground truth model on exact empirical observations only, the agents have no assumed schedules similarly to activity-based models. We instead mirror the observations of the inner city mobility into pedestrian agents. Using the OD matrix and the virtual city model, we are able to model a type of “pull” between

the geographical points within our city. While the theory of Voorhees (2013) described a gravitational pull between district types, we model a linear pull between exact inner city locations. However, in our case, the pull is essentially a probability of the hotspot to be chosen as a destination. It is defined by adding weights to each hotspot based on the OD matrix data. The weights are based on trip volume, i.e. how many times each origin–destination pair was observed within any given 1-h slot in our dataset. The weights are updated during our simulation according to the simulated day and time. At the beginning of each simulated hour, each hotspot gets a list of destination hotspots as a property. Each destination hotspot is assigned a weight: the number of trips from the origin, as described in the OD matrix at the corresponding day and time.

Subsequently, the agents move between the hotspots, picking their next destination according to their current hotspot and its destination weights. The next destination is picked with a *weighted choice*, a random choice in which destination's probability to be chosen equals its trip volume's ratio to the sum of trip volumes of all destinations of the current hotspot. More specifically $p(A) = \frac{T_A}{T_{All}}$ where A is a destination and T refers to trip volume. The agents are not targeting the exact geographical position of the hotspot but pick the destination from 10 random nodes closest to the target from navigation mesh. The route is calculated with the A* pathfinding algorithm and further smoothed with a funnelling algorithm as previously described. Currently, all the agents are considered to move individually as we are not identifying group movement from our data. While previous research has shown that shortest route algorithms are not optimal for urban traffic modelling (Manley, Addison and Cheng, 2015; Manley, Orr, & Cheng, 2015), it should be noted that in our case, the trip lengths are quite short meaning that there are rarely multiple routing options for reaching the next destination.

The entry/exit hotspots are a special case of hotspots that add and remove agents from the simulation. The entry/exit hotspots are combinations of actual geographical hotspots that reside outside the virtual model limits in each ordinal direction. Computationally, they are treated as single hotspots with trip volumes equal to the sum of the trip volumes of the actual hotspots. An agent can pick entry/exit nodes similarly to any other nodes, as a result of a weighted choice. However since entry/exit nodes don't have exact geographical locations, the agents pick their geographical destination according to the closest road/sidewalk leading outside the virtual city model. An agent can exit the model at any time,

however new agents can enter the model only when the number of simultaneous agents is smaller than the current hourly preferred number of agents. The preferred number of agents is defined as the sum of all OD matrix observations for each particular 1-h slot. The ordinal direction for the entry of a new agent is picked by a weighted choice according to the trip volumes from each entry/exit hotspot. The exact geographical location of an entry is randomly picked among the roads and sidewalks leading to the city model from the ordinal direction that is the result of the weighted choice.

3.3. Using a random model to analyse spatio-temporal location attractiveness

To investigate the temporal characteristics of pedestrian flows, we compared the trips generated by our ground truth model to randomly moving agents at different time segments. Comparison with a random model allowed us to segregate pedestrian traffic typical to specific inner city location types in our city from traffic that is generated solely from street network structure. In this analysis, we used an OD matrix dataset containing trip volumes from May 2012.

The agent navigation in our random model is simple compared to the ground truth model. The random agent movement resembles the *purposive movement* in agent simulation by Jiang and Jia (2011). The agents randomly pick their destinations and navigate to them among a shortest route generated by A* and funnelling algorithms. Each time an agent reaches its destination, it has a chance of 1 to 10 to pick an entry/exit node and leave the simulated model. Otherwise the agent picks the next destination randomly among all the nodes of the navigation mesh. On entering the virtual model, the agents pick their entry location randomly among roads and sidewalks leading to the virtual model. Beyond this, agents have no assumed schedules. Because the random model has no data to guide the pedestrian agents, the agent mobility is essentially guided by street connectivity only (Jiang & Jia, 2011). For being able to properly compare the traffic patterns, the random model upkeeps the same number of simultaneous pedestrians as the ground truth model.

To compare the pedestrian traffic generated by the two models, a way was needed to quantify the pedestrians' movement across the virtual model. Two common ways to aggregate traffic flows are *gate counting* and *footprint counting*; the former meaning, for example, counting each time a pedestrian enters a street segment while the latter refers to counting footprints at fixed time intervals (Jiang & Jia, 2011). In our footprint analysis, we use the navigation mesh to quantify pedestrian flows. The navigation mesh contains thousands of nodes, with area typically between 10m² and 40m². Each time a pedestrian's route goes through a node, the footprint count of the node is incremented by one. Our first footprint analysis compared one month's (May 2012) worth of traffic as generated by the ground truth model and the random model.

In search for temporal patterns in the pedestrian traffic, we further analysed the agent movement for different time periods. We split the week into three periods: *week* (Monday to Thursday), *end of week* (Friday and Saturday) and *Sunday*. This segmentation is culturally derived, since Fridays and Saturdays are usually busier than other days of the week, while Sundays are typically quieter than other days. This is due to the fact that 5-day working weeks in Finland are the most common, making Friday and Saturday popular days for shopping and other pastime ((OSF) Official Statistics of Finland, 2009). Furthermore, we split each day into four periods: *morning* (8:00–10:00), *working hours* (11:00–17:00), *evening* (18:00–22:00) and *night* (23:00–7:00). Again this segmentation is culturally derived and differs slightly from the convention typical in other studies. When analysing the pedestrian traffic with this segmentation, we reset the footprint counts in the beginning of each time period and finally average the results of each time period. This results in total of 12 time segments. Currently, seasonal variation

is not considered in the segmentation. All our experiments focus on simulating the month of May.

Fig. 4a and b use a heatmap visualization to describe footprint analyses for weekend daytime time segment. This period has especially high traffic in the centre area and highlights well the differences between the models. Fig. 4a shows the analysis for the ground-truth model and Fig. 4b the random model. Brighter streets have more traffic. Our comparison shows that certain densely populated streets have high traffic in both models.

The average number of pedestrian agents per hour of each time period can be seen in Fig. 5. This volume is averaged from the hourly observations within the OD matrix. As can be seen, the Sunday Night time segment contains more traffic than Sunday Morning while the morning is more crowded elsewhere in the segmentation. This is most likely due to Saturday nightlife traffic.

To further quantify the differences between the ground-truth and random models, we averaged the footprint analyses within each time period and subtracted one model from the other: i.e. we subtracted the footprints in the random model from the footprints in the ground-truth model. This helps us identify the areas where the random model generates too much or too little traffic. We found that the extent of deviation of the random model is not constant but fluctuates across different time periods. We were also able to identify certain types of locations that according to these differences seem to attract more traffic than others. Fig. 6a and b show the comparison between the two models for weekday morning and weekday night hours respectively. The colour scheme is adopted from www.colorbrewer.org (Harrower & Brewer, 2011). Streets in red indicate the random model generates too little traffic, green areas are populated too heavily in the random model, while bright white areas (i.e., closer to the background) matches the ground-truth model.

The busiest region in the ground truth model is the shopping/nightlife district located in the very centre of the city. Other attractors seemed to be schools and libraries, as well as other points of interest. Suburbs had the least traffic and the random model typically generates too much traffic in those areas while underestimating the centre area.

4. Introducing the weighted POI model

Utilizing crowdsourcing, and building on our footprint analysis between the random model and the ground-truth model, we constructed a "weighted POI model". In this model, we substitute the Wi-Fi hotspots of the ground truth model with the crowdsourced POIs collected from the OpenStreetMap database (www.openstreetmap.org). As in the previous models, the agents have no realistically assumed schedules; in this case, the mobility is defined only by the location of POIs and their temporally varying attractiveness. The locations of the POIs act as geographical targets for pedestrian agents similarly to the Wi-Fi hotspots in the ground truth model. Unlike Wi-Fi hotspots, however, individual POIs do not possess weights, but their categories do. POI's typically have a "category" datafield which offers a general one-word description of the location type such as "restaurant, museum, and park". In our model, these categories act as activities for pedestrian agents that have a different probability to be chosen according to the simulated day and time. Our model contains a total number of 134 POIs that belong in 21 categories. Again, the agents are using shortest routes to navigate between POIs. However, we consider the POIs to act similarly to anchor points, providing realistic routing in inner city mobility.

Unlike the hotspot weights, the choice of next destination is not affected by the previous place visited; all POIs belonging to the same category have an equal chance of being chosen. Similar to the Random model, the agents have a 1 to 10 chance to exit the virtual model. In addition, the agents also have a 1 to 10 chance to pick a completely random destination. This is to generate traffic into areas, which are not close to any POIs. Another big difference between the ground truth model and the POI model is the frequency of weight changes. In the

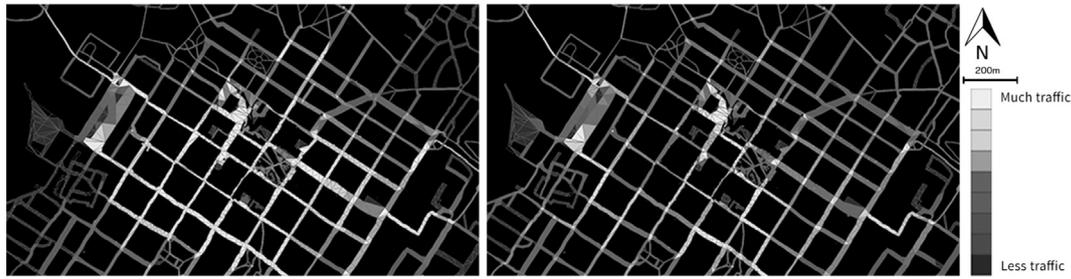


Fig. 4. Pedestrian footprint analysis with the ground-truth (a) and the random model (b). Brighter streets have more traffic.

ground truth model the each simulated hour has a unique set of weights while the category changes in the weighted POI model change according to the 12 time segments as defined in the footprint analysis. The next subsection explains in detail how the category weights of the weighted POI model are defined. An overview of all models can be seen in Table 1.

4.1. Modelling POI category weights

As previously stated, in footprint analysis we derived the over- and underestimations of the random model by subtracting the footprints produced by the random model from the footprints produced by the ground truth model for each of the 12 time segments. The random model gave us the attractiveness of locations according to street connectivity only; comparison with the ground truth model in turn gave us the temporal variations that show as over- and underestimations of the random model. Using the comparison values we estimate temporal attractiveness of various location types while attempting to negate the effect of street connectivity and location centrality. Therefore, for each node of the navigation mesh, we receive the comparison footprint value $N_c = N_g - N_r$, where N_g is the number of footprints of a node produced by the Ground Truth model and N_r the number of footprints produced by the Random model. The average comparison value P_c for each POI is acquired by averaging all comparison footprint values N_c of nodes inside the 75-m radius of the POI and taking the additive inverse of the result. Each POI comparison value P_c is then scaled between 0 and 10 according to the following formula where P_{all} is a list of all comparison values P_c :

$$P_w = \frac{(P_c - \min(P_{all})) * 10}{\max(P_{all}) - \min(P_{all})}$$

Finally the category weights C_w are acquired by averaging the individual scaled POI comparison values P_w that belong to the same

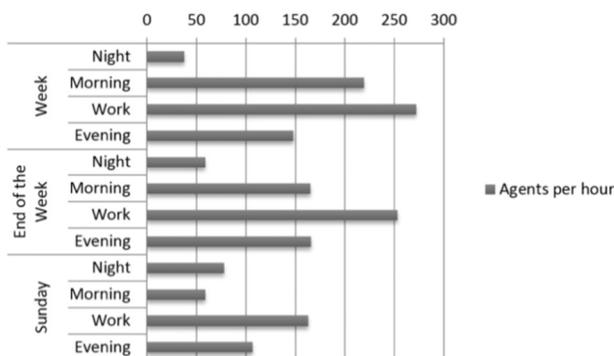


Fig. 5. Average number of pedestrian agents per hour within each time period in the ground-truth model.

category. The category weights obtained from analysing the May 2012 dataset are shown in Table 2.

4.2. Evaluating the POI model and the random model against the ground truth model

Similar to the analysis conducted for the Random model, the POI model was also analysed by performing a footprint analysis by quantifying the agents' passage through navigation mesh nodes. Again, we were essentially interested in the over- and underestimations compared against ground truth data. However, since we used May 2012 data to train the POI model, we compared the footprints against a ground truth model using data from May 2011 dataset instead. This was to gain insight to the generalisability of the POI model. Furthermore, we also conducted a comparison analysis between the random model and the ground truth model with the May 2011 dataset to assess the performance of the POI model against the random model. As stated previously, the random model has no POI or Wi-Fi data to guide pedestrian agents; the agents instead pick random locations from the virtual model and pursue those locations according to shortest possible route. Thus, the footprints the random model generates are essentially determined solely by the street connectivity of the inner city (Jiang & Jia, 2011). After letting each model simulate one month's worth of pedestrian mobility according to simulated date and time, we analysed the over- and underestimations produced by the models to evaluate how much the generated footprints deviate from the ground truth overall.

We calculated the deviations as a percentage difference from the footprints produced by the ground truth model. The percentage error of a single node is calculated as $error = \frac{|fp_g - fp_m|}{fp_g}$ where fp_g is the number of footprints produced by the ground truth model and fp_m is the number of footprints produced by the model under investigation. The resulting percentage errors were then averaged across regions of interest. Our first observation was that the weighted POI model's footprints are closer to ground truth especially at the POI locations. The total error produced by the POI model was 34% whereas the Random model produced a 51% error when measured from inside the radii of the POI hotspots. Outside the POIs the difference between the models was smaller with the POI model producing a 47% error and the Random model a 51% error. This is due to the fact that areas where no POIs are defined have no data to guide pedestrian movement. Thus, the footprints that are generated into these areas are either caused by the occasional pedestrian with a random target, pedestrians travelling between POIs, or a pedestrian entering or exiting the virtual model. Averaged throughout the entire city model, the POI model produced a 41% error and the Random model a 50% error. Scatterplots of the entire model's footprints per triangle averaged through all time segments can be seen in Figs. 7 a (ground-truth model vs. Random model), b (ground-truth model vs. POI) and c (POI model vs random model). Investigating the figures as well as the R^2 values contained within, it can be seen that the POI model is statistically closer to the ground truth when averaged through all time segments. The POI



Fig. 6. Colour maps visualizing the differences in traffic between the ground-truth and random models in the Weekday morning time segment a) and the Weekday night time segment b). The random model does not generate enough traffic at the red areas whereas it creates too much traffic at the green areas. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

model and the Random model, however are quite different from each other.

We also inspected the deviations at individual time segments that were introduced in Section 3.3. The POI model outperformed the random model in almost every time segment with the *end of the week morning* and *Sunday morning* being outliers inside POIs and *week night* time segment outside. An overview of the comparison between the two models can be seen in Table 3. We split the evaluation into categories “inside POIs”, “outside POIs” as well as “Total”. While including POI areas in the evaluation might be favourable for the performance of the POI model, we can gain some insights in the functioning and further improvement of the model this way, as can be later seen when examining individual POI categories.

Examining two outlier time segments, *end of the week morning* and *Sunday morning*, we identify that the POI model tends to overestimate the pedestrian traffic whereas the random model is usually more prone to underestimations. Clear examples are the *Casino* and *Bureau de change* POIs where the overestimations of the weighted POI model heavily outweigh the underestimations of the Random model (see Supplementary Table 1 and Supplementary Table 2 for comparisons of individual POI categories at every time segment). Generally, however the

underestimations of the Random model are larger than the POI model's overestimations making the POI model perform better when averaged through all time segments as can be seen in Table 4.

When inspecting the individual POI deviations at every time segment (Supplementary Tables 1 and 2), we found that the POI model overestimates the *Casino* and the *Bureau de change* locations heavily in almost every time segment. Similarly, these locations are underestimated by the random model. However, not all locations heavily underestimated by the random model are in turn overestimated by the POI model. The *Bar* and the *Fast Food* POIs are such examples.

It is also noteworthy that all three *Night* time segments are heavily underestimated in both models. These segments are outliers within the POI model which is not as prone to underestimations as the Random model. Deviations exceeding 100 negative footprints exist only within the *Night* time segments in the POI model whereas they are commonplace in the random model (see Supplementary Tables 1 and 2).

The over- and underestimations of individual POI locations as total can be seen in Table 4. Individual POI locations deviations at all time segments can be seen at Supplementary Table 1 and Supplementary Table 2 for weighted POI model and random model respectively.

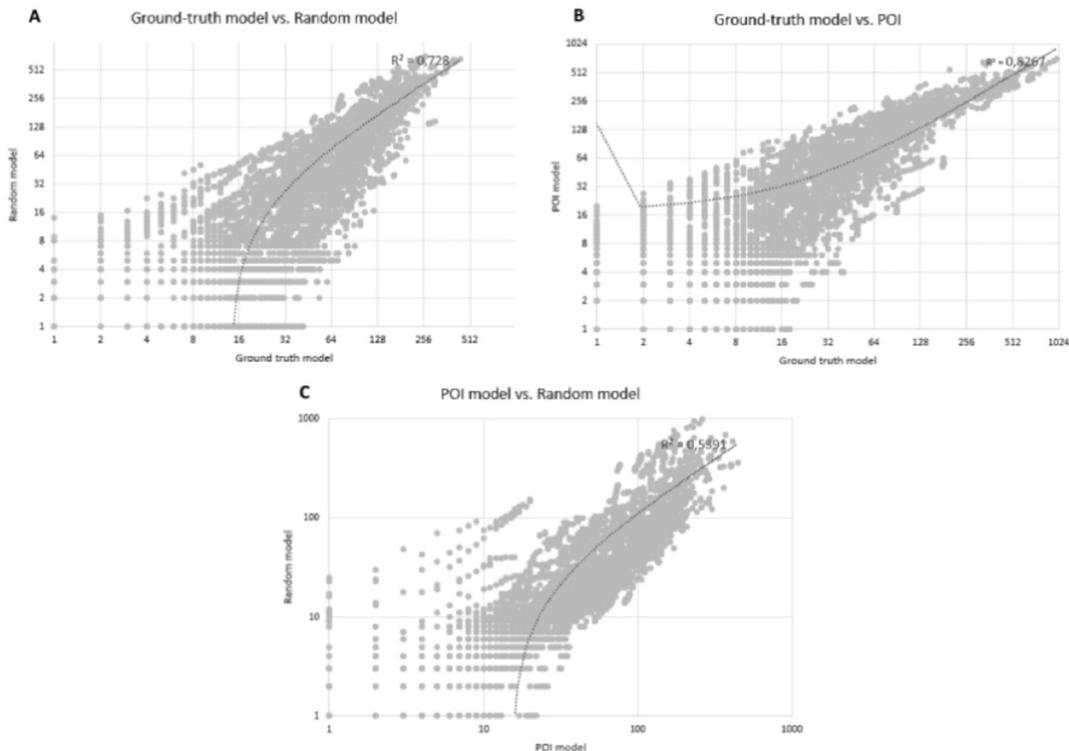


Fig. 7. Scatterplots with trendlines comparing entire city model's footprints averaged through all time segments. A: ground-truth model vs. random model ($R^2 = 0.7228$), B: ground-truth model vs. POI ($R^2 = 0.8267$), C: POI model vs. random model ($R^2 = 0.5391$).

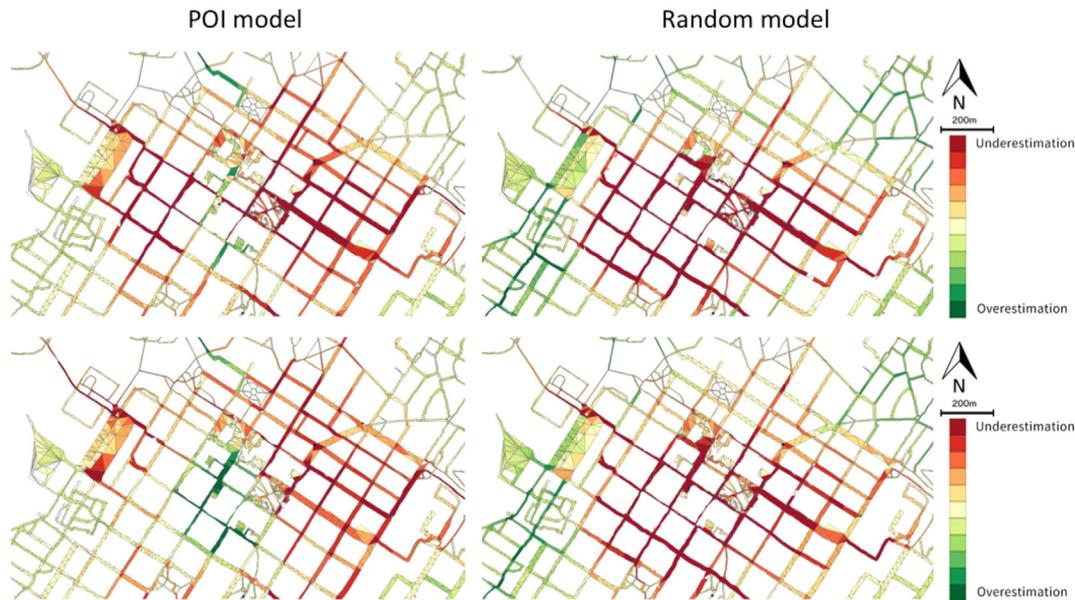


Fig. 8. Comparison maps showing examples of over- and underestimations made by the POI model (left) and the random model (right). The upper half shows the “Sunday Night” time segment while the lower half shows the “Week Working Hours” time segment.

Fig. 8a, b, c and d show example heatmaps of the overestimation and underestimation charts made by the POI model (leftmost figures), as well as the Random model (rightmost figures) in the centremost district of the simulated area. The upper half displays the time segment “Week Working Hours” while the lower half displays the “Sunday Night” time segment. It can be seen that the POI model brings the very centre of the shopping district as well as the areas outside of the shopping district closer to ground truth. However, there are still quite many streets that are underestimated. The most significant difference to the Random model is that while the footprints produced by the Random model gradually turn from heavy underestimation to overestimation while moving away from the centre, the POI model is able to somewhat level the deviations according to most distinct location types. An interesting observation is that when examining the maps visually, it seems that the lower half, “Week working hours,” is better treated by the POI model than the upper half “Sunday Night,” when the truth is actually the opposite. When examining the deviations from ground truth as a whole (see Table 3), the POI model is much closer to the Random model in the “Week working hours” segment than in the “Sunday Night” segment. This is probably due to the POI model heavily overestimating the centre part of the district in the former segment.

5. Discussion & conclusion

Our work presents a method to simulate temporal variations of inner city pedestrian traffic with minimum data collection effort.

Using the weighted POI model we have developed, the varying attractiveness of various POI categories at various time of day can be used to calibrate the fluctuations of pedestrian traffic in a random model. There are various sources for obtaining POIs, or alternatively for simulated cities POIs can simply be defined. The POI categories and their time-dependent attractiveness can be used to examine the characteristics of different parts of cities.

Of course, the functions of inner city locations change according to time of the day and different functions attract people from various social classes; these city usage characteristics have implications for planning and safety (Bromley et al., 2003). Currently, we use POI data categories exactly as they are reported and curated by a community of volunteers. For example, there are separate categories for *bar* and *pub* which could probably be combined. Obviously, we cannot guarantee that the POI dataset we extracted from OpenStreetMap is optimal for pedestrian simulation. While the results of the weighted POI model are promising in replicating the ground truth model's pedestrian flows, it lacks in performance at certain instances. Most notable shortcomings are the occasional overestimations especially of certain category types. Currently our model relies solely on analysing the pedestrian footprint numbers and attempting to replicate these footprints through unmodified POI data. However, the categories can be easily modified to be more suitable for various analytical purposes. For example, the large number of POI categories could be abstracted into “Shopping,” “Cafes,” “Restaurants,” “Pubs,” “Nightclubs,” “Cinemas,” and “Theatres” similarly to inner city classifications in (Bromley et al., 2003). Additionally, the seminal work

Table 1
The differences between the ground truth model, the random model and the weighted POI model.

Model	Ground truth	Random	POI
Dataset	OD-matrix derived from Wi-Fi network analysis	None	Crowdsourced points of interest (POIs)
Agent attractors	249 Wi-Fi hotspots	Random locations	134 POIs
Choice of next destination in virtual model	According to destination weights of last visited hotspot	Random choice among all the nodes of the navigation mesh	Each POI category has probability weights. A category is chosen with a weighted choice. The destination is then chosen randomly among the POIs of the chosen category. In addition, agents have 1 in 10 chance of picking a random location.
Weight change frequency	Hourly	None	12 time segments
Exiting city model	According to destination weights of last visited hotspot	On picking the next destination, agents have a 1 to 10 chance of choosing to exit the virtual model.	On picking the next destination, agents have a 1 to 10 chance of choosing to exit the virtual model.

Table 2
POI category weights.

Category	Week				End of the week				Sunday			
	Morning	Work	Evening	Night	Morning	Work	Evening	Night	Morning	Work	Evening	Night
Fast food	4	5	6	6	5	5	6	5	4	5	5	5
Place of worship	5	6	5	0	6	6	6	0	5	6	6	1
Cafe	3	4	4	4	4	4	4	3	3	4	3	3
ATM	5	6	5	1	6	6	6	0	5	5	6	2
Hospice	5	6	5	0	6	6	6	0	5	6	6	2
Theatre	6	7	6	2	6	7	6	2	6	6	6	3
Restaurant	5	6	6	5	6	6	6	4	5	6	6	4
Gym	7	7	6	1	7	7	7	1	7	7	6	2
Cinema	4	5	6	5	5	6	6	3	4	5	5	4
Car rental	5	6	5	1	5	6	6	0	5	5	6	2
Townhall	5	6	6	1	6	6	6	1	5	6	5	2
Pub	4	5	5	4	5	5	5	4	4	5	5	4
Nightclub	5	6	6	7	5	6	6	5	5	6	6	5
Library	5	6	6	1	6	6	7	1	5	7	6	2
Parking	5	5	5	4	5	6	6	3	4	5	5	4
Bar	4	5	5	7	5	5	6	8	3	5	4	7
Casino	0	1	1	6	1	1	1	2	0	1	1	0
Pharmacy	4	4	4	4	4	5	4	2	4	4	4	2
Hospital	5	6	6	4	6	6	6	2	6	6	6	3
Bank	3	4	4	4	4	4	4	3	3	4	4	3
Bureau de change	0	0	0	5	0	0	0	3	0	0	0	2

of Lynch (1960) offers a categorization that can be utilized if POIs are not defined literally only as geographical points but as areas as well. In this case, the POI categories could follow Lynch's list of city elements: "paths," "edges," "districts," "nodes" and "landmarks." As for crowdsourced data, contemporary social media as well as specialized sites offer numerous sources for POI data besides OpenStreetMap.

While we made an attempt to negate the effect of street connectivity using the random model and ground truth model comparison values for POI weights, the results suggest that some of the POI model shortcomings can be attributed to location centrality. A good example is, the outlying category type *bureau de change*, which consists of a single POI residing at very central location within the Oulu city as well as close proximity to other popular POI's. The large number of footprints the general model produces within the category is probably due to the location than the attractiveness of the category itself. The effect of POIs with categories at very central locations could be thus reinterpreted. Perhaps, previous findings in macroscopic studies (Hillier & Hanson, 1984; Jiang & Jia, 2011) could be leveraged further to automatically classify locations that are naturally attractive due to the street structure of the city. This information could be used to negate the effect of overestimation at central locations more effectively and thus further improve the weighted POI model.

5.1. Future work

Our work presents opportunities for further studies. The most critical drawback is that it is difficult to validate our model against the ground truth model in other cities. This is due to the fact that large open municipal Wi-Fi networks are rare at this time. Even though the weights we presented might not yield optimum results for all cities as

they are, we believe that our findings can be used as rough heuristics for inner city usage analysis. Estimating the required changes to the POI weights for different cities is an interesting topic for future research.

Our work acts as an example on how modern societies contain abundant, sometimes unexpected data sources. It is interesting to further investigate whether the POI model can be combined with other data sources or into more complete activity-based models. For instance, do POIs generate somewhat realistic traffic independently, and can more realistic results be obtained by incorporating our POI model into a complete activity-based system.

There are possibilities to improve the POI model in isolation as well. The current model does not consider distance when choosing the specific POI among the chosen category. However, favouring closer POIs over further ones might increase the realism of the model. It would also be interesting to model POIs according to the anchor point theory, for example, using path nodes as POIs as described in Lynch (Lynch, 1960). A comparison between modelled POIs vs crowdsourced POIs would bring new knowledge.

Currently, the OD matrix consists solely of data collected from municipal Wi-Fi usage. Even though the density of the Wi-Fi network in our city allows for rather accurate tracking of pedestrians, combining other tracking methods might further increase the realism of the ground truth model.

Finally, in our current model pedestrians simply use the shortest routes to move between locations. Incorporating simulated human behaviour at the micro level could further increase the realism of both models. Similarly, the addition of simulated public transport would increase the realism of the model in larger cities. Our model currently simulates an area that is comparatively small and independent of public transportation. The OD matrix however, does contain long distance

Table 3

Average deviation (percentage difference from the footprints produced by the ground truth model) from the ground-truth model when compared to the other models. The results of the better model are in bold.

Location	Model	Week				End of the week				Sunday				Total
		Morning	Work	Evening	Night	Morning	Work	Evening	Night	Morning	Work	Evening	Night	
Inside POIs	Random	0.41	0.43	0.52	0.75	0.36	0.43	0.48	0.69	0.35	0.38	0.44	0.65	0.51
	POI	0.39	0.37	0.28	0.52	0.41	0.33	0.30	0.39	0.38	0.36	0.36	0.30	0.34
Outside POIS	Random	0.48	0.51	0.52	0.58	0.44	0.49	0.55	0.61	0.46	0.45	0.53	0.64	0.52
	POI	0.42	0.45	0.50	0.63	0.40	0.42	0.46	0.56	0.44	0.45	0.47	0.49	0.47
Total	Random	0.45	0.47	0.51	0.66	0.39	0.45	0.50	0.64	0.42	0.41	0.48	0.62	0.50
	POI	0.40	0.40	0.40	0.58	0.38	0.36	0.36	0.48	0.41	0.40	0.41	0.39	0.41

Table 4
Weighted POI model and Random model over- and underestimations at Individual POI locations averaged through all time segments. The results of the better performing model are in bold.

Model	Fast food	Place of worship	Cafe	ATM	Hospice	Theatre	Restaurant	Gym	Cinema	Car rental	Townhall
Random	-132	45	-100	7	17	-38	-90	-2	-79	2	-24
POI	-10	11	37	9	3	-22	-14	-31	1	12	-2
Model	Pub	Nightclub	Library	Parking	Bar	Casino	Pharmacy	Hospital	Bank	Bureau de change	
Random	-79	-108	-19	-57	-186	-202	-117	-26	-128	-242	
POI	-12	-45	-31	-8	-8	139	40	-9	48	153	

origin–destination data as well. Such information can be used for the simulation of public transportation in future research.

Acknowledgements

This research was funded partly by the European Commission (Grants 6Aika, AKAI, PCIG11-GA-2012-322138 and 645706-GRAGE), and the Academy of Finland (Grants 276786-AWARE, 285062-iCYCLE, 286386-CPDSS, 285459-iSCIENCE).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.compenvurbsys.2016.02.004>.

References

- (OSF) Official Statistics of Finland (2009). *Time use survey [e-publication]*. (Retrieved February 6, 2015, from) http://www.stat.fi/til/akay/akay_2014-12-01_men_001_en.html
- Arentze, T. A., & Timmermans, H. J. P. (2005). Representing mental maps and cognitive learning in micro-simulation models of activity-travel choice dynamics. *Transportation*, 32(4), 321–340.
- Axhausen, K. W. (1998). Can we ever obtain the data we would like to have? *Theoretical foundations of travel choice modeling* (pp. 305–323).
- Axhausen, K. W., & Gärling, T. (1992). Activity-based approaches to travel analysis: Conceptual frameworks, models, and research problems. *Transport Reviews*, 12(4), 323–341 (<http://doi.org/10.1080/01441649208716826>).
- Axhausen, K. W., Zimmermann, A., Schönfelder, S., Rindsfuser, G., & Haupt, T. (2002). Observing the rhythms of daily life: A six-week travel diary. *Transportation*, 29(2), 95–124 (<http://doi.org/10.1023/A:1014247822322>).
- Bandini, S., Federici, M. L., & Manzoni, S. (2007). A qualitative evaluation of technologies and techniques for data collection on pedestrians and crowded situations. *Proceedings of the 2007 Summer Computer Simulation Conference* (pp. 1057–1064) (Retrieved from <http://dl.acm.org/citation.cfm?id=1358075>).
- Boarnet, M., & Sarmiento, S. (1998). Can land-use policy really affect travel behaviour? A study of the link between non-work travel and land-use characteristics. *Urban Studies*, 35(7), 1155–1169.
- Boarnet, M. G., Joh, K., Siembab, W., Fulton, W., & Nguyen, M. T. (2011). Retrofitting the suburbs to increase walking: Evidence from a land-use-travel study. *Urban Studies*, 48(1), 129–159.
- Bromley, R. D. F., Tallon, A. R., & Thomas, C. J. (2003). Disaggregating the space–time layers of city-centre activities and their users. *Environment & Planning A*, 35(10), 1831–1851.
- Calabrese, F., Colonna, M., Lovisolo, P., Parata, D., & Ratti, C. (2011). Real-time urban monitoring using cell phones: A case study in Rome. *IEEE Transactions on Intelligent Transportation Systems*, 12(1), 141–151 (<http://doi.org/10.1109/ITITS.2010.2074196>).
- Calabrese, F., Pereira, F. C., & Lorenzo, G. Di. (2010). The geography of taste: analyzing cell-phone mobility and social events. *Pervasive Computing*, 6030, 22–37.
- Couclelis, H., Golledge, R. G., Gale, N., & Tobler, W. (1987). Exploring the anchor-point hypothesis of spatial cognition. *Journal of Environmental Psychology*, 7(2), 99–122.
- Demyen, D., & Buro, M. (2006). Efficient triangulation-based pathfinding. *AAAI'06 Proceedings of the 21st National Conference on Artificial Intelligence*. Vol. 1. (pp. 942–947). AAAI Press (Retrieved from <http://dl.acm.org/citation.cfm?id=1597538.1597687>).
- Ettema, D., & Timmermans, H. J. P. (1997). *Activity-based approaches to travel analysis*. Emerald Group Publishing Limited.
- Ewing, R., Deanna, M., & Li, S. -C. (1996). Land use impacts on trip generation rates. *Transportation Research Record*, 1518(1), 1–6 (<http://doi.org/10.3141/1518-01>).
- Galiza, R. J., & Ferreira, L. (2012). Standard pedestrian equivalent factors: New approach to analyzing pedestrian flow. *Journal of Transportation Engineering*, 139(2), 208–215 (Retrieved from [http://ascelibrary.org/doi/abs/10.1061/\(ASCE\)TE.1943-5436.0000441](http://ascelibrary.org/doi/abs/10.1061/(ASCE)TE.1943-5436.0000441)).
- Golledge, R. G., & Gärling, T. (2002). *Cognitive maps and urban travel*. University of California Transportation Center.
- Golledge, R. G., & Spector, A. N. (1978). Comprehending the Urban Environment: Theory and practice. *Geographical Analysis*, 10(4), 403–426 (<http://doi.org/10.1111/j.1538-4632.1978.tb00667.x>).
- Guo, R. Y., Wong, S. C., Huang, H. J., Zhang, P., & Lam, W. H. K. (2010). A microscopic pedestrian-simulation model and its application to intersecting flows. *Physica A: Statistical Mechanics and its Applications*, 389(3), 515–526.
- Harrower, M., & Brewer, C. A. (2011). ColorBrewer.org: An online tool for selecting colour schemes for maps. *The map reader: Theories of mapping practice and cartographic representation* (pp. 261–268). John Wiley and Sons.
- Hart, P. E., Nilsson, N. J., & Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, 4(2), 100–107 (<http://doi.org/10.1109/TSSC.1968.300136>).
- Heikkinen, A., Okkonen, A., Karhu, A., & Koskela, T. (2014). A distributed POI data model based on the entity-component approach. *IEEE Symposium on Computers and Communications (ISCC)* (pp. 1–6). IEEE (<http://doi.org/10.1109/ISCC.2014.6912570>).
- Helbing, D., & Molnar, P. (1995). Social force model for pedestrian dynamics. *Physical Review E* (Retrieved from http://pre.aps.org/abstract/PRE/v51/i5/p4282_1).
- Helbing, D., Molnar, P., Farkas, I. J., & Bolay, K. (2001). Self-organizing pedestrian movement. *Environment and Planning B: Planning and Design*, 28(3), 361–383 (<http://doi.org/10.1068/b2697>).
- Hillier, B., & Hanson, J. (1984). *The social logic of space*. Cambridge University Press.
- Jacques Junior, J. C. S., Raupp Musse, S., Jung, C. R., & Junior, J. J. (2010). Crowd analysis using computer vision techniques. *IEEE Signal Processing Magazine*, 27(5), 66–77 (<http://doi.org/10.1109/MSP.2010.937394>).
- Jiang, B., & Jia, T. (2011). Agent-based simulation of human movement shaped by the underlying street structure. *International Journal of Geographical Information Science*, 25(1), 51–64 (Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/13658811003712864>).
- Kostakos, V., Juntunen, T., Goncalves, J., Hosio, S., & Ojala, T. (2013). Where Am I? Location archetype keyword extraction from urban mobility patterns. *PLoS One*, 8(5), e63980 (<http://doi.org/10.1371/journal.pone.0063980>).
- LaPlante, J., & Kaeser, T. P. (2007). A history of pedestrian signal walking speed assumptions. *3rd urban street symposium: Uptown, downtown, or small town: Designing urban streets that work* (pp. 1–8).
- Lerman, Y., & Omer, I. (2016). Urban area types and spatial distribution of pedestrians: Lessons from Tel Aviv. *Computers, Environment and Urban Systems*, 55, 11–23 (<http://doi.org/10.1016/j.compenvurbsys.2015.09.010>).
- Lynch, K. (1960). *The image of the City*. The M.I.T. Press.
- Maneewongvatana, S., & Mount, D. (1999). It's okay to be skinny, if your friends are fat. *Center for Geometric Computing 4th Annual Workshop on Computational Geometry* (pp. 1–8) (Retrieved from <https://myplan.cs.umd.edu/~mount/Papers/cgc99-smpack.pdf>).
- Manley, E. J., Addison, J. D., & Cheng, T. (2015a). Shortest path or anchor-based route choice: A large-scale empirical analysis of minicab routing in London. *Journal of Transport Geography*, 43, 123–139 (<http://doi.org/10.1016/j.jtrangeo.2015.01.006>).
- Manley, E. J., Orr, S. W., & Cheng, T. (2015b). A heuristic model of bounded route choice in urban areas. *Transportation Research Part C: Emerging Technologies*, 56, 195–209 (<http://doi.org/10.1016/j.trc.2015.03.020>).
- Montello, D. R. (1998). A new framework for understanding the acquisition of spatial knowledge in large-scale environments. *Spatial and temporal reasoning in geographic information systems* (pp. 143–154).
- Nagel, K., & Rickert, M. (2001). Parallel implementation of the TRANSIMS micro-simulation. *Parallel Computing*, 27(12), 1611–1639 ([http://doi.org/10.1016/S0167-8191\(01\)00106-5](http://doi.org/10.1016/S0167-8191(01)00106-5)).
- Nelson, A. L., Bromley, R. D. F., & Thomas, C. J. (2001). Identifying micro-spatial and temporal patterns of violent crime and disorder in the British city centre. *Applied Geography*, 21(3), 249–274.
- Ronald, N., Arentze, T., & Timmermans, H. (2016). Towards process validation for complex transport models: A sensitivity analysis of a social network-enhanced activity-travel model. *Computers, Environment and Urban Systems*, 55, 24–32 (<http://doi.org/10.1016/j.compenvurbsys.2015.09.005>).
- Seyfried, A., Steffen, B., Klingsch, W., & Boltes, M. (2005). The fundamental diagram of pedestrian movement revisited. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(10), P10002 (Physics and Society; Statistical Mechanics. <http://doi.org/10.1088/1742-5468/2005/10/P10002>).
- Simini, F., González, M. C., Maritan, A., & Barabási, A. -L. (2012). A universal model for mobility and migration patterns. *Nature*, 484(7392), 96–100 (<http://doi.org/10.1038/nature10856>).
- Stead, D. (2001). Relationships between land use, socioeconomic factors, and travel patterns in Britain. *Environment and Planning B: Planning and Design*, 28(4), 499–528 (Retrieved from <http://www.envplan.com/abstract.cgi?id=b2677>).

- Stevens, A., & Coupe, P. (1978). Distortions in judged spatial relations. *Cognitive Psychology*, 10(4), 422–437.
- Teknomo, K. (2002). *Microscopic pedestrian flow characteristics: Development of an image processing data collection and simulation model*. (PhD thesis) Japan, Sendai: Tohoku University (Retrieved from <http://people.revoledu.com/kardi/publication/Dissertation.pdf>).
- Teknomo, K. (2006). Application of microscopic pedestrian simulation model. *Transportation Research Part F: Traffic Psychology and Behaviour*, 9(1), 15–27.
- Timmermans, H., Arentze, T., & Joh, C. -H. (2002). Analysing space–time behaviour: New approaches to old problems. *Progress in Human Geography*, 26(2), 175–190 (<http://doi.org/10.1191/0309132502ph363ra>).
- Veldhuisen, J., Timmermans, H., & Kapoen, L. (2000). RAMBLAS: A regional planning model based on the microsimulation of daily activity travel patterns. *Environment & Planning A*, 32(3), 427–443 (Retrieved from <http://www.envplan.com/abstract.cgi?id=a325>).
- Voorhees, A. M. (2013). A general theory of traffic movement. *Transportation*, 40(6), 1105–1116 (<http://doi.org/10.1007/s11116-013-9487-0>).
- Waddell, P., Ulfarsson, G. F., Franklin, J. P., & Lobb, J. (2007). Incorporating land use in metropolitan transportation planning. *Transportation Research Part A: Policy and Practice*, 41(5), 382–410.
- Weidmann, U. (1993). *Transporttechnik der Fussgänger - Transporttechnische Eigenschaften des Fussgängerverkehrs (Literaturstudie)* (Schriftenreihe des IVT). In IVT (Ed.), *Transporttechnische Eigenschaften des Fussgängerverkehrs (Literaturauswertung)*. CH-8093 Zürich: ETH-Hönggerberg (Retrieved from weidmann-1993.pdf).
- Wirz, M., Franke, T., Roggen, D., Mitleton-kelly, E., Lukowicz, P., & Tr, G. (2013). Probing crowd density through mobile phones in city-scale mass gatherings. *EPJ Data Science*, 2(1), 1–24.
- Wirz, M., Franke, T., Roggen, D., Mitleton-Kelly, E., Lukowicz, P., & Troster, G. (2012). Inferring crowd conditions from pedestrians' location traces for real-time crowd monitoring during city-scale mass gatherings. *2012 IEEE 21st International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises* (pp. 367–372). IEEE (<http://doi.org/10.1109/WETICE.2012.26>).
- Work, D. B., & Tossavainen, O. P. (2008). An ensemble Kalman filtering approach to highway traffic estimation using GPS enabled mobile devices. *47th IEEE Conference on Decision and Control, 2008. CDC 2008* (pp. 5062–5068) (Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4739016).
- Xiao, G., Juan, Z., & Zhang, C. (2015). Travel mode detection based on GPS track data and Bayesian networks. *Computers, Environment and Urban Systems*, 54, 14–22 (<http://doi.org/10.1016/j.compenvurbysys.2015.05.005>).
- Zhang, J., Timmermans, H. J. P., & Borgers, A. (2005). A model of household task allocation and time use. *Transportation Research Part B: Methodological*, 39(1), 81–95.