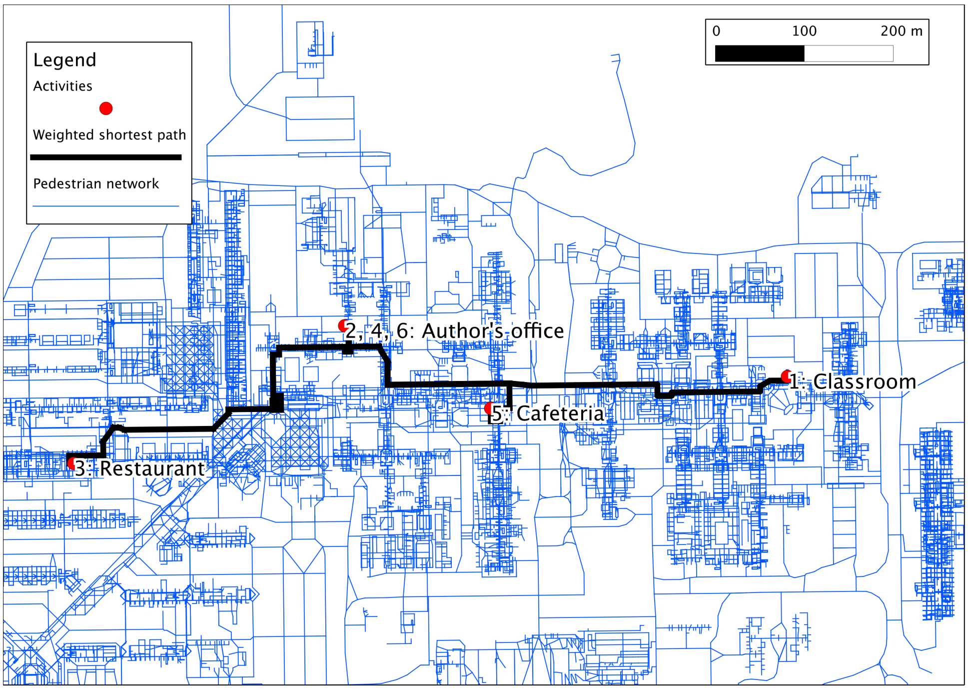

Towards an activity-based model for pedestrian facilities

Antonin Danalet

Bilal Farooq

Michel Bierlaire

EPFL

April 2013

STRC

13th Swiss Transport Research Conference

Monte Verità / Ascona, April 24 – 26, 2013

EPFL

Towards an activity-based model for pedestrian facilities

Antonin Danalet, Bilal Farooq, Michel Bierlaire
Transport and Mobility Laboratory (TRANSP-OR)
School of Architecture, Civil and Environmental Engineering (ENAC)
Ecole Polytechnique Fédérale de Lausanne (EPFL)
phone: +41-21-693 25 32
fax: +41-21-693 80 60
{antonin.danalet,bilal.farooq,michel.bieraire}@epfl.ch

April 2013

Abstract

This paper develops a framework for understanding pedestrian mobility pattern from WiFi traces and other data sources. It can be used to forecast demand for pedestrian facilities such as railway stations, music festivals, campus, airports, supermarkets or even pedestrian area in city centers. Scenarios regarding the walkable infrastructure and connectors, the scheduling (trains in stations, classes on campus, concerts in festivals) or the proposed services in the facility may then be evaluated.

It is inspired by activity-based approach. We assume that pedestrian demand is driven by a willingness to perform activities. Activity scheduling decision is explicitly taken into account. Activity-based approach for urban areas is adapted for pedestrian facilities, with similarities (scheduling behavior) and differences (no "home" in pedestrian facilities, thus no tours). This is a first attempt to define a integrated system of choice models in the context of pedestrian facilities.

Version 2

Keywords

Pedestrians, walk, pedestrian facilities, activity-based modeling, pedestrian demand, train stations, campus, music festival, WiFi traces, courses attendance

1 Introduction

Communication networks like WiFi or cellular networks have become ubiquitous in recent years. They offer opportunities for researchers to better understand mobility and improve urban policies. They can be used at the scale of a city, or even bigger, and are also available in pedestrian facilities, such as train stations. They can help in designing new infrastructures and optimizing existing ones. With urbanization, increase of the population and of the usage of public transportation, walking is important: it is not only green and healthy, but it is the key of an efficient multimodal transport system.

We will focus here on the modeling of pedestrian demand at a disaggregate level. Understanding and forecasting the demand for pedestrian infrastructure is important in many aspects. Mathematical models of pedestrian demand for activities are needed for problems like: (a) congestion of pedestrians; (b) efficient design of new facilities (such as transportation hubs, campuses or shopping malls) and daily operations of these facilities (public transit timetables); (c) large events gathering a high number of people; (d) travel guidance and information systems aim at helping the pedestrian in implementing her journey.

Pedestrians are highly complex to model and to apprehend from an analyst viewpoint. Contrary to other modes, they don't use a vehicle, they don't follow strict constraints (mostly legal and security ones), and the underlying infrastructure is highly heterogeneous (sidewalks, crossings, buildings, shopping malls, squares, etc.). The complexity of modeling pedestrian calls for a specific methodology. Surprisingly, the literature about demand for pedestrians is small.

This lack in research is probably not only related to the difficulties in modeling but also in collecting data. Detecting pedestrians at a large scale and indoor is challenging. The large literature about operational models of microscopic interactions between pedestrians in a crowd is related to the existence of the well known social force model, but also to the increasing availability of tracking of pedestrians using cameras. But at a large scale, using camera seems unadapted. We focus here on sensors with full coverage of pedestrian facilities, in order to estimate the overall demand. Such data exist, but with low precision (e.g., traces from WiFi infrastructures). They provide a large sample size for a low cost, but also have significant drawbacks: no socio-economic attributes are available and only smartphone users (with WiFi available) are tracked.

This paper aims at proposing a methodology for analyzing demand in pedestrian infrastructure. Our methodology merges data from sensors that can locate pedestrians, but also other information about activities during the journey, in order to model the demand for the infrastructure, understand what drives this demand, and finally forecast the effect of changes in policy. The main case studies are EPFL campus and Paléo music festival, but our methodology is general enough to be easily adapted for transport hubs, such as airports and train stations, any large event, shopping malls, pedestrian touristic attractions, museum or even pedestrian-only city centers.

We proposed a methodology to collect activity-episodes sequences from scarce data, directly modeling the imprecision in the measure in Danalet *et al.* (2012). It generates several candidate lists of activity-episodes sequences associated with a corresponding likelihood. In this paper, our goal consists in defining a location choice model from this output. The challenges are the choice set generation of possible activity locations, the measurement equations for observations, the dynamic of the system and the specification of the model.

In this paper, we present available data, we propose a modeling framework for a location choice model in pedestrian facilities and we discuss further data collection and how to reach an activity-based model for pedestrian infrastructure.

This research represents a step towards building a methodology to use WiFi traces for pedestrian research and to understand spatial and temporal pattern within walking areas.

2 Literature Review

Research on pedestrians is mostly focusing at the microscopic level of behavior, i.e., walking itineraries: how do pedestrian behave in front of an obstacle? On the other hand, practitioners are asking different questions: how many persons are going through an airport, how much time do they wait at the security gate, what is the demand for connecting the different platforms in a railway station, what is the demand for toilets in a music festival? They are mostly interested in macroscopic demand for a given or future infrastructure.

This paper focuses on generating macroscopic demand using individual activity-based models. The following literature review is divided in three parts, corresponding to the three challenges we meet in pedestrian demand modeling: data collection, representation of space and modeling itself.

2.1 Digital footprints: from network efficiency to human mobility

Gathering data about pedestrian origin, destination and route is difficult, particularly indoor and on a large scale. These data are important for route choice modeling, description of congestion, and flow estimation. Several data collection techniques are device-centric, based on data from smartphone. We focus here on data from communication network infrastructure ("infrastructure traces"). At the end of this section, we briefly mention research using data coming directly from the smartphone.

Device-centric data and infrastructure traces are using the fact that a majority of people are carrying a smartphone. Using traces from communication network infrastructure has several advantages on data from the smartphone. First, full coverage of the facility is usually cheap and allows for an estimation of the overall demand. The communication infrastructure sometimes already exists, and increasing its density has a positive side effect. Smartphone users don't need to install anything on their device, and so the access to sensitive information such as emails or address book is limited for the analyst, which ensure privacy for the users. Finally, traces from communication network infrastructure are related to the infrastructure and not to the individual: we are tracking all individual smartphones going through a facility and not all places visited by the same individuals. It allows the analyst to focus on the pedestrian facility covered by the communication network.

There are few drawbacks to infrastructure traces as well. Socio-economic and demographic attributes are difficult to collect due to both privacy concerns (if data already exist) and difficulty to survey the tracked person from the infrastructure side (if data don't exist). Plus, as mentioned by Calabrese *et al.* (2013), smartphones' users might not represent a random sample of the population.

Several applications using data from communication infrastructure, both with WiFi and GSM traces, has been developed to study mobility behavior. Calabrese *et al.* (2011) propose an OD matrix for the Boston metropolitan area based on data generated each time a mobile phone connects to the cellular network. In their paper, origins and destinations are regions. Some of the salient features are the comparison of their results with existing statistics, the absence of the underlying transportation network and of a particular mode, but the focus on global mobility in this area at a large scale.

A large literature exists about WiFi traces from a computer communication point of view. A complete review can be found in Aschenbruck *et al.* (2011). All references in this paper defines mobility trace-based models not as a tool to understand mobility itself, but to improve the quality of the WiFi, e.g., minimizing the handoff delay incurred when scanning for available access points (APs), or attempt to detect intrusion in the network (Balajinath and Raghavan, 2001). The goal is to predict the next point-of-attachment of the user (Wanalertlak *et al.,* 2011). This body of work studies mostly pedestrians, since the scale of the problem is related to offices or campuses. Yoon *et al.* (2006) study mobility models for WiFi infrastructure and try to make them representative of real mobility, and mention the possible applications to urban planning, socially-based games, or augmented reality.

Field studies have been done on campuses (Meneses and Moreira, 2012, Sevtsuk *et al.,* 2009, Wanalertlak *et al.,* 2011, Yoon *et al.,* 2006, Zola and Barcelo-Arroyo, 2011), but also with municipal wireless networks (Wanalertlak *et al.,* 2011), corporate networks (Balazinska and Castro, 2003), or conference rooms (Balachandran *et al.,* 2002). The first data collection occurred in 1999 (Tang and Baker, 2000) and recent papers are still published on this topic, with Zola and Barcelo-Arroyo (2011) being the first study with European data. The scale of the data collections range from a group of buildings on a campus in Zola and Barcelo-Arroyo (2011), hallways and atrium with 6 APs in a campus and 40 APs in the center of Portland, Oregon in Wanalertlak *et al.* (2011), a whole campus in Yoon *et al.* (2006).

The main results are the prediction of changes in APs. Wanalertlak *et al.* (2011) is reaching an accuracy of 23 to 43% in next-cell prediction with a reduction of the average handoff delay to 24~25 ms. Some descriptive statistics are also presented in Zola and Barcelo-Arroyo (2011) and Yoon *et al.* (2006). We learn that the activity is higher in the afternoon and lower in the weekends (Zola and Barcelo-Arroyo, 2011). The main conclusion in Zola and Barcelo-Arroyo (2011) consists in observing that almost 40% of the observed devices never connect to more than one AP during the day, based on 3 buildings representing a total of 37 APs. Yoon *et al.* (2006) are going further, defining transition probabilities between buildings on a whole campus. The amount of time a user spends at a location is supposed to follow a power-law distribution with small exponents (Tuduce and Gross, 2005), or a log-normal distribution (Kim and Kotz, 2006), similar to distribution of speed in the same article. Meneses and Moreira (2012) analyze the connectivity between the different visited locations at building level and between the two campuses of their university.

The main problem encountered in these articles is the *ping pong* effect, when the device has similar signal strengths from different APs and changes regularly from one to another. This is a problem from a network viewpoint, but also for modeling pedestrian origins and destinations. Yoon *et al.* (2006) propose to use a moving average weighted by time spent at destination to remove the extra signals. A general solution presented in Aschenbruck *et al.* (2011) consists in aggregation of data over different APs.

In most cases in previously cited article, the goal is to predict the next AP, not the next destination in term of users' choices. Most studies about WiFi are focusing on network performance and management and not on human mobility. In Yoon *et al.* (2006), contrarily to all other papers cited here, an OD matrix is estimated at the building level in Dartmouth college, but variations in time/day are not considered, as Aschenbruck *et al.* (2011) noticed.

In the literature about mobility models for WiFi infrastructure from a computer communication point of view, the most common model, Random Waypoint model (RWP), is often criticized as not representing real human mobility (Conti and Giordano, 2007). One of the problems with RWP consists in using straight lines between two signals in different APs, even if this path is not possible. This is the main reason why trace-based mobility models were developed in this domain of research. A key challenge in building a realistic model is to define a pedestrian network and the corresponding possible paths the user with a device can follow. This process of constructing and using the pedestrian network in order to improve the mobility model is not explicitly presented in Aschenbruck *et al.* (2011) in their large review of trace-based mobility models. The need for a more complex approach than a direct path is emphasized in Rojas *et al.* (2005).

The differentiation between symbolic and geometric map allows to analyze and understand human behavior with more depth. Kang *et al.* (2004) underline that users are more interested in places rather than locations, defining a place as a "locale that is important to an individual user and carriers important semantic meanings such as being a place where one works, lives, plays, meets socially with others, etc.". They propose an algorithm for extracting significant places from a WiFi trace of coordinates. Their goal is to automatically translate measured location into places of interest, as measurements of the same location are different due to errors. They use a time-based clustering algorithm to eliminate intermediate locations and determine the number of clusters as new location measurements come in. The algorithm also checks if a new cluster corresponds to a previously existing one. The authors mention that the automatically "extracted places need to be labeled to have semantic meanings". Future work consists for them in predicting a user's destination and to detect arrival and leaving time.

Calabrese *et al.* (2010) propose a way to generate labels for WiFi access point (AP) on a campus. Eigendecomposition is used to compress data in four eigenvectors, representing the daily office cycle and residential usage for the two first ones. This description, called "eigenplace", is used to cluster APs. The main conclusion of their work is that APs cover many different uses and it is difficult to label them: one AP serves different kind of locations surrounding it.

Using data from communication infrastructure has advantages. The data collection is related to a specific area (a campus, a railway station or an airport), and the mode is known in these context. In this way, there is no need to detect the mode of the user generating the signal. This is the main reason why we focused on non device-centric data collection techniques.

However, device-centric techniques could be useful and similar to traces from communication infrastructure in their quality. Movahedi Naini *et al.* (2011) present an experiment estimating the population size in a music festival using ten participants tracking bluetooth signals around them. Matching the IDs of the different detected devices, it is possible to track spectators. The precision of the data and the frequency of matches give information about the destination and the general patterns of visitors. In this context, the 10 different smartphones used to collect the surrounding Bluetooth signals are mimicking the role of access points in WiFi data.

Both WiFi and Bluetooth localization are very useful for giving geographical information. If we want to model not only the position but also the choice of destination and more generally of activities in a time period (see Section 2.3), there is a need for data about attributes of activities and the process of building chains of activities. The implementation of activity-based models is problematic relatively to data collection even if there exists a well-developed theoretical literature about activity modeling. It may be necessary to ask questions to the people we want to model. The process of building a schedule of activities has been observed through surveying tools, such as MAGIC, an interactive computer experiment for assessing scheduling rules. Ettema *et al.* (1994) asked respondents to develop activity schedules for the next day on a computer and found heuristics people used to complete the task. There are three problems with this technique. First, the data were collected at one point in time for a single person. Second, there could be a difference between what people plan and answer to the survey and their real behavior in the next day. And finally, the vast amount of information needed to understand the pattern of a day could create response burden.

CHASE (Computerized Household Activity Scheduling Elicitor) has been implemented on mobile devices Rindfüser *et al.* (2003) and is similar to some extent to MAGIC. Asking people to answer questions directly on their mobile device is a way to avoid the cognitive incongruence with actual behavior. Roorda (2005) suggests panel surveys as a way to avoid response burden, contacting respondents at multiple points over time.

Axhausen (1998) suggests that it is impossible to collect enough data about a single person to model the complete choice of activities, and that research should develop tools which generate schedules consistent with real behavior from more than one survey instrument, and methods to establish the consistency between generated and observed activity schedules. Using different instruments is also a way to evaluate bias associated to each of them.

2.2 Pedestrian Network

The transportation network depends on the scale of the study area and on the definition of destinations. All papers we presented in the previous section consider as destinations either buildings (Yoon *et al.,* 2006) or APs (Wanalertlak *et al.,* 2011, Tuduce and Gross, 2005).

In Yoon *et al.* (2006), the authors converted a map of the Dartmouth college to a graph between buildings and limited themselves to major roads. Wanalertlak *et al.* (2011) define explicitly the underlying mobility infrastructure in Portland, Oregon, and one floor of a building on campus as "a graphical representation of all possible paths". Sevtsuk *et al.* (2009) have access to a coded map of campus with rooms and their respective usage obtained from the MIT Department of Facilities.

As Kasemsuppakorn and Karimi (2013) mentioned very recently, "pedestrian networks are not as widely available as road networks are for many areas". They propose to build it from GPS traces, which obviously does not work indoor. Kang *et al.* (2004) and Calabrese *et al.* (2010) are using WiFi to cluster places of interest and label them, as mentioned in Section 2.1.

No indoor network of a pedestrian facility allowing for computation of the shortest path between two destinations is known to the authors and is publicly available.

2.3 Activity-based modeling: From Driving to Walking

Before 1950, transportation studies didn't have predictive power and mostly described the current state of traffic (Weiner, 1999). In 1954, Mitchell and Rapkin (1954) theorized models for travel patterns and behavior based on trips and susceptible to change depending on attributes such as land use or socioeconomic characteristics. They already mentioned at the time the limitations of a trip-based approach to really understand motivations of travel.

In practice, travel demand analysis has been decomposed in four sequential steps in the Urban Transportation Modeling System (UTMS): trip generation, trip distribution, modal split and traffic assignment. The fundamental unit of these analysis is trip. Trips are generated not from a behavior-based demand but from a physical analogy with gravity: from trip production location to trip attraction location. In 1955, the Chicago Area Transportation Study (CATS) used this analytical decomposition.

In the 1970s, a shift was observed from the traditional trip-based approach to a demand-oriented approach. Trips started to be considered as a derived demand. Scheduling decisions were surveyed, e.g., in the Household Activity Travel Simulator (HATS) (Heggie and Jones, 1978, Jones, 1979). At the same time, models started to be more of a disaggregate microsimulation approach. This change in paradigm is the origin of activity-based travel demand modeling. It is related to the rapid development of random utility-based individual choice models for route, mode, frequency and destination choice (McFadden, 1974, Stopher and Meyburg, 1975).

The premise of activity-based approach consists in considering activity as a choice and trips as a way to complete the chosen activity. In other words, modeling the daily activity patterns allows the development of behavioral travel demand models that are sensitive to changes in policy.

For Hägerstrand (1970) the choice set of activities is constrained in three ways:

Capability constraints are biological or technological constraints of the individual. Some are fundamentals and structure mostly the time when activities are performed, mainly sleeping at night and eating at some time of the day. Others are distance oriented and express the time-space constraints on movement[1]. People need to perform a trip to reach a destination before being able to perform an activity.

Coupling constraints define the requirement for some activities of other people or resources (salesman and customer in a shop, students and teacher in a class, meetings at work[2].

Authority constraints are protecting resources and limit accessibility to certain persons. They can be related to payment, invitation, power or custom: a seat in a theater is a temporary authority constraint, while traffic rules are permanent.

In practice, there are traditionally two main components in activity-based models: activity generation (dealing with the basic needs for self-realizations) and activity scheduling (spatio-temporal constraints and opportunities related to the actual activities). The interactions between these two components are in both directions.

Activity generation is facing the problem of generating the choice set for activity patterns and different techniques have been proposed. Using discrete choice models, Adler and Ben-Akiva (1979) proposed one of the earliest examples of these models. The patterns are chosen on attributes such as modes, number of destinations, purposes, time spent at destination, travel

[1] It is interesting to read Hägerstrand (1970) with today's eyes: "I am sure that we are still far from understanding the locational implications of the next enlargement of the range of this tube (i.e., telecommunications), which have entirely broken up this once so narrow spatial boundaries. One hears the most divergent opinions about future possibilities of having television screens substitute for face-to-face meetings around a table." (tubes are accessibility rings). So basically he was expecting Skype meetings. A bit further: "People need to have some kind of home base, if only temporary, at which they can rest at regular intervals, keep personal belongings and be reached for receiving messages." It seems he was not expecting Skype to be on smartphones.

[2] Again, Hägerstrand (1970) with today's eyes: "Today, hoards of administrators and secretaries spend their regular working hours trying to get other people together for future meetings" foresees Doodle (`http://www.doodle.ch`).

times, etc. Hybrid simulations have been proposed where the choice does not consist in an overall activity pattern, but patterns are sequentially built following a decision tree. Hybrid simulations focus on search heuristics for the choice set generation and create small choice sets. They are supposed to overcome the weaknesses of a discrete choice approach (Ettema *et al.,* 1996). Bowman and Ben-akiva (1996) criticized these models and the absence of dependencies in activity choices across the day. Their model proposes a hierarchical approach, with a decomposition into primary and secondary tours, assuming some activities are more important in structuring the travel patterns. For both types of tours, they model the activity pattern, the choice of tour and the choice of destination itself.

Many refined models have been developed later, based on the same assumptions, mainly that people are rational utility maximizers. Another stream of research exists, assuming people do not always take optimal decision. Based on the works by Simon, Tversky and Kahneman, sub-optimal decisions are related to a satisficing approach or a limitation in information acquisition and treatment. Habib (2007) mentions few problems of this kind of models. They are good for modeling short-term policy analyses but not for long-term demand forecasting and they need specific rules to be able to response to changes in policy. Inputs are generated empirically and lacks theoretical foundation. Handling of preference heterogeneity is a difficult task in these models.

Hybrid models have also been created, mixing the two approaches to gain flexibility. In PCATS (Prism-Constrained Travel Simulator) a module creates the choice set using behavioral rules, and gives the choice set as an input for a second module based on utility maximization (Kitamura *et al.,* 1997). It divides the day in two periods. Blocked periods represent fixed commitments, while open periods represent windows of opportunities. More generally, CEMDAP (Comprehensive Econometric Model of Daily Activity Pattern) uses 22 sequential and nested decision models.

A complete and general review not related to pedestrians can be found in Roorda (2005), Habib (2007) and Feil (2010).

With respect to pedestrians, Borgers and Timmermans (1986) develop a destination choice model as part of a system of models to predict the total demand for retail facilities within inner-city shopping areas. Timmermans *et al.* (1992) provide a review of models existing in 1992 and of a few applications to urban and transportation planning in The Netherlands. Zhu and Timmermans (2005) focus on shopping decision processes, using bio-inspired heuristics to mimic the decision process. Eash (1999) has developed models for non motorized destination choice and vehicle versus non motorized mode choice, with application to the Chicago Area.

Influenced by traditional practice in travel demand analysis, several models are derived from origin-destination matrices (Nagel and Barrett, 1997, Antonini *et al.*, 2006), where the set of potential origins and destinations is predefined, and flows between origins and destinations is estimated. In a disaggregate context, the choice of the destination can be modeled conditional to a given activity, or as a joint choice of an activity and a destination. In both cases, the choice set is typically large and difficult to characterize (Bierlaire and Robin, 2009).

In some circumstances, it is possible that no destination is explicitly chosen by a pedestrian. It is typical when walking is the activity as such, or in shopping and touristic activities. In these cases, an itinerary is undertaken without a known target, trying to maximize the chances to reach attractive places along the way (Borst *et al.*, 2001). This type of behavior is clearly difficult to formalize, but relates to behavior that highly characterizes pedestrian movements.

The current state of the literature consists on the one hand in very detailed conventional activity-based models for multimodal urban areas and on the other end in few specific models related to pedestrians. There is a big gap between these two worlds. Obviously, differences are important between them. A pedestrian facility such as a transportation hub is not as complex as a city but still share some characteristics with it. By drawing our inspiration from conventional models, we plan to reduce this gap and to develop a general model for activities in pedestrian facilities.

3 Available data in pedestrian facilities

Several data sources are available in pedestrian facilities. Localization data are probably the most crucial ones in order to build a demand model. Some systems exist to track pedestrians. We focus here on localization data from communication infrastructure, such as WiFi traces or Bluetooth traces.

A pedestrian map of the facility is also very important, in order to locate the possible walking paths as well as the potential destinations and categories of activities available. We present here two examples: a campus and a music festival.

Other data are needed: schedules, weather measurements, surveys where pedestrians are directly asked to answer on paper. A special kind of surveys concerns the access to the pedestrian facility. Since there is no “home” destination where all daily tours are starting and ending, we need to know how pedestrian access the facility, where and when. Mode choice survey to access the pedestrian facility may be available. Finally, the availability of socioeconomic data about pedestrians in a given location such as a music festival, a campus or a railway station is an open question.

3.1 Localization data

3.1.1 WiFi traces

WiFi network is not available in all pedestrian facilities but can be found in lots of railway stations, airports and campuses. Other pedestrian facilities, such as a music festival or a supermarket, can easily install a network for tracking people, providing internet to their customers, or both.

Two different kinds of data using traces from WiFi infrastructures exist: raw data from the access points, giving the identity of the access point the user is connected to, and localization data using triangulation and signal strength. In both cases, the identity of the user is uniquely defined through the MAC address.

There are 789 access points on campus (see Figure 1 and `http://map.epfl.ch` for more details, using the "Points of interest" tab and then the list of "Utilities"), covering the whole set of buildings. We used this infrastructure for two data collections.

Raw data from Radius The first data collection technique consists in the localization of the access point to which a user is connected. Users connected to the WiFi through WPA are authenticated and accounting data are collected by the Radius server (Rigney *et al.,* 2000). Radius log record allows to associate a user with an access point (Koo *et al.,* 2003). All access points on campus are located on the map. When moving, even if the user doesn't need to authenticate again, he generates new accounting data with the name of the new access point the user is associated with.

The access points are localized on a map by x-y coordinates and floor. Thus, the collected data are time-stamped and have the ID of the access point, plus its localization. Data are collected for all users of EPFL network. An unique and anonymous identification is generated daily for each device. In this way, a same device can be tracked for one day but not more and anonymity is ensured.

Localization using triangulation Tracking of WiFi devices positioning is also possible using triangulation based on signal strength. Moreover calibration information can be used to improve location accuracy.

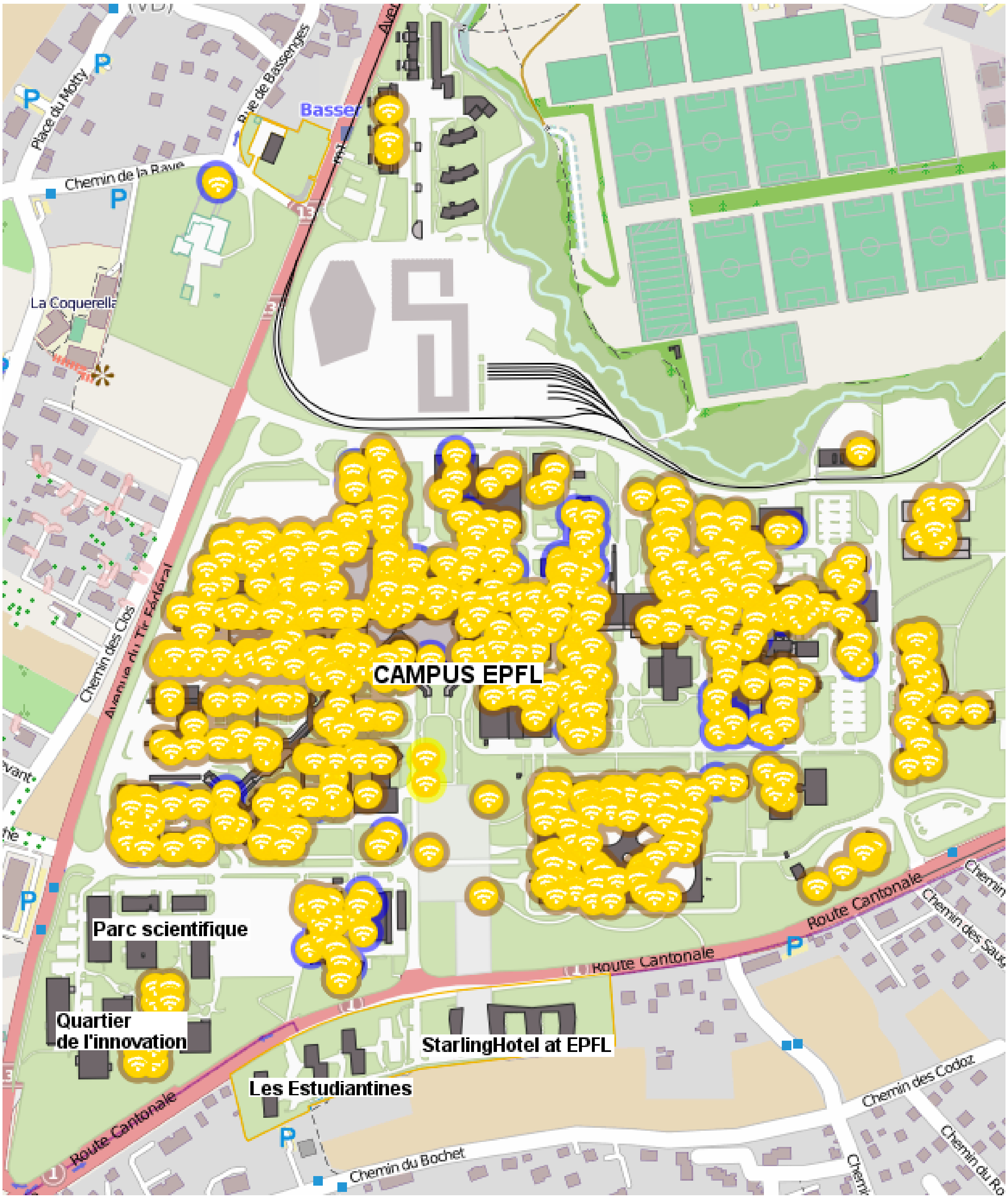

Figure 1: A map of EPFL campus with WiFi access points (source: `http://map.epfl.ch`)

The collected data has a time stamp, x-y coordinates, a floor and, possibly, a confidence area for the measurement.

On campus, we use Cisco Context Aware Mobility API with the Cisco Mobility Services Engine (MSE). A confidence factor cF provides a 95% confidence square around each x-y coordinates and is unique for both directions. It has a mean of 187 meters in our dataset.

We collected localization data with 12 users who accepted to participate in this experiment for two months. Moreover, 200 students from 6 different classes and 300 employees randomly chosen have been tracked for one day. Their identity has been anonymized.

3.1.2 Bluetooth tracking

Movahedi Naini *et al.* (2011) present a methodology to evaluate the number of people in a music festival. Ten attendees acted as Bluetooth probes using a smartphone. They sensed Bluetooth devices around them and captured the MAC addresses every 80 seconds. Each smartphone is localized through GPS, thus it gives an approximative location for Bluetooth devices nearby. Two mobile phones were installed at the entrances of the festival for validation purpose. In this music festival, in 2010, 8.2% of attendees had visible Bluetooth devices. The 10 agents detected 79.3% of these devices. Kondratieva *et al.* (2013) are evaluating the opportunity to use such data for the detection of activity patterns in a festival.

Versichele *et al.* (2012) are also tracking Bluetooth signals but from fixed detectors at the Ghent Festival, in Belgium. The full coverage of the festival is impossible. 22 sensors where located in the main spots of the festival. 81'000 visitors generated 153'000 traces. They estimate that 11% of visitors had visible Bluetooth devices. The traces allow to evaluate how many days people are coming, how long they are staying, the mode they used to access the festival and which part of the festival they visit and in which order.

3.2 Map data

Having access to a good pedestrian map is very important to understand pedestrian behavior. We present here two maps related to the datasets presented above, in in creflocdata: EPFL campus and Paléo music festival.

3.2.1 EPFL campus

EPFL website proposes an orientation tool for the campus, `http://map.epfl.ch`. It allows localization of someone's office or any place in the university, as well as itineraries between these places. Moreover, it offers thematic information, like restaurants.

Pedestrian network with 4 different levels of path The orientation tool is based on a pedestrian network created by Camptocamp, a company with offices on campus and developing Open Source softwares. It consists of a graph in three dimensions, containing 50'131 vertices. A vertex is defined by a x-y coordinate and a floor level. All edges are straight lines connecting two vertices. Each of the 56'655 edges contains a type of inter-floor connectors (stairs or lift), a type of access (free or locked) and a hierarchical factor.

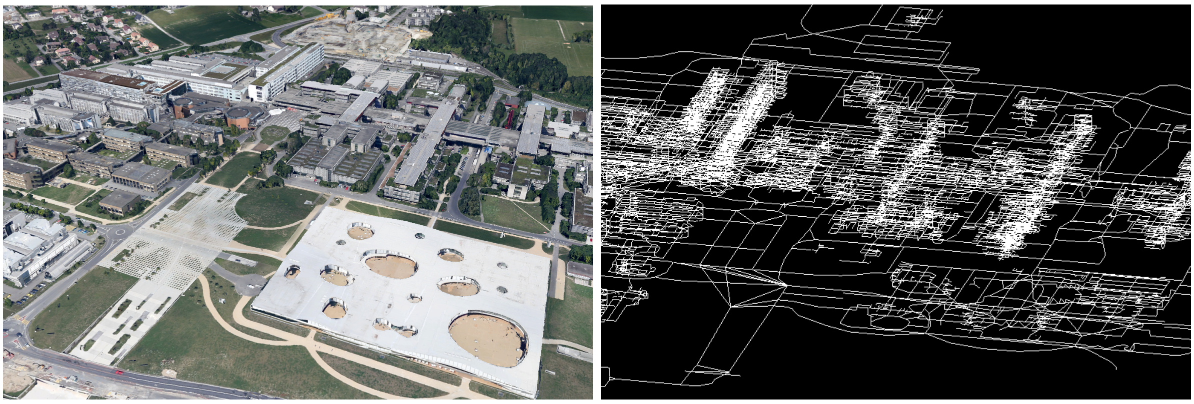

Figure 2: EPFL campus (source: GoogleEarth) and its underlying pedestrian network (source: Lopez-Montenegro Ramil *et al.* (2013))

Hierarchical factor Similarly to road networks for car driving, the pedestrian network is decomposed in four different levels. The *major* routes are "highway" for pedestrians, the main axes to walk through the campus. At a lower level, some routes are connecting the different buildings of EPFL. *Inter-building* routes are important, but are used by specific people and are not common knowledge. Then, some routes are axes for mobility inside a building; we call these routes *intra-building*. Finally, the last link connecting the intra-building routes to offices and classrooms are called *secondary* routes.

Description of activity locations The database contains 17'502 "points of interest" that corresponds to the list on `http://map.epfl.ch`. Most of them are locations of ethernet cables, access control readers, or WiFi access points, i.e., not a proper activity location for a pedestrian. All points of interest are separated in categories, such as "Infrastructures", "Pictures", or "Access". We selected some of these "points of interests" as potential activity locations for pedestrians (see Table 1).

In addition to public "points of interest", a list of all rooms on campus exists. It contains 13'783 rooms, in all categories. Only offices and classrooms where selected in our model, adding up to a total of 5387 on campus. All the rooms are localized on the pedestrian network.

Table 1: Points of interest of `http://map.epfl.ch`

<table>
<tr><th>Categories</th><th colspan="2">Activity locations</th><th>Count</th><th>In the model</th></tr>
<tr><td rowspan="13">Access</td><td colspan="2">Information Desk</td><td>1</td><td>Yes</td></tr>
<tr><td colspan="2">Entrances</td><td>42</td><td>No</td></tr>
<tr><td rowspan="2">Car</td><td>Parking lots</td><td>55</td><td>Yes</td></tr>
<tr><td>Electric plug</td><td>1</td><td>No</td></tr>
<tr><td colspan="2">Public transportations (bus stops, ticket machines)</td><td>9</td><td>Yes</td></tr>
<tr><td rowspan="3">Bike</td><td>Parkings</td><td>93</td><td>Yes</td></tr>
<tr><td>Electric plugs</td><td>1</td><td>No</td></tr>
<tr><td>Center</td><td>1</td><td>No</td></tr>
<tr><td colspan="2">Road accesses</td><td>5</td><td>No</td></tr>
<tr><td colspan="2">Delivery Point</td><td>1</td><td>Yes</td></tr>
<tr><td colspan="2">Disabled (obstacles, WC, automatic doors)</td><td>20</td><td>No</td></tr>
<tr><td colspan="2">Meeting points</td><td>5</td><td>No</td></tr>
<tr><td rowspan="3">Education</td><td colspan="2">Schools</td><td>9</td><td>No</td></tr>
<tr><td colspan="2">Auditorium</td><td>23</td><td>Yes</td></tr>
<tr><td colspan="2">Librairies</td><td>2</td><td>Yes</td></tr>
<tr><td rowspan="11">Services</td><td colspan="2">Restaurants</td><td>21</td><td>Yes</td></tr>
<tr><td colspan="2">Snack bars</td><td>17</td><td>Yes</td></tr>
<tr><td colspan="2">Shops</td><td>7</td><td>Yes</td></tr>
<tr><td colspan="2">Caretakers</td><td>11</td><td>Yes</td></tr>
<tr><td colspan="2">Associations</td><td>9</td><td>Yes</td></tr>
<tr><td colspan="2">Museums, exhibitions</td><td>2</td><td>Yes</td></tr>
<tr><td colspan="2">Services (child care, travel agency, language center, ...)</td><td>19</td><td>Yes</td></tr>
<tr><td colspan="2">Secretariats</td><td>12</td><td>Yes</td></tr>
<tr><td colspan="2">Workshops</td><td>29</td><td>Yes</td></tr>
<tr><td colspan="2">Hotels</td><td>2</td><td>Yes</td></tr>
<tr><td colspan="2">Student housing</td><td>4</td><td>Yes</td></tr>
<tr><td rowspan="10">Utilities</td><td colspan="2">Post stations</td><td>22</td><td>Yes</td></tr>
<tr><td colspan="2">Bornes camipro</td><td>9</td><td>Yes</td></tr>
<tr><td colspan="2">Electronic money chargers</td><td>3</td><td>Yes</td></tr>
<tr><td colspan="2">Cash</td><td>2</td><td>Yes</td></tr>
<tr><td colspan="2">Xerox machines</td><td>12</td><td>Yes</td></tr>
<tr><td colspan="2">Public printers</td><td>77</td><td>Yes</td></tr>
<tr><td colspan="2">Phones</td><td>32</td><td>No</td></tr>
<tr><td colspan="2">WiFi</td><td>791</td><td>No</td></tr>
<tr><td colspan="2">Showers</td><td>42</td><td>No</td></tr>
<tr><td colspan="2">Recycling</td><td>175</td><td>No</td></tr>
<tr><td rowspan="3">Infrastructure</td><td colspan="2">Network plugs (802.1x, DHCP, ...)</td><td>15024</td><td>No</td></tr>
<tr><td colspan="2">Camipro</td><td>870</td><td>No</td></tr>
<tr><td colspan="2">Fire safety</td><td>11</td><td>No</td></tr>
<tr><td>Pictures</td><td colspan="2">Panoramas</td><td>20</td><td>No</td></tr>
</table>

Weighted shortest path Finding a route on campus can be formulated as a shortest path problem and we can use traditional algorithm such as Dijkstra and A*. However, both in the case of cars and pedestrians, this is not necessarily consistent with real behavior. People usually tend to walk or drive on major roads, despite the fact that some secondary roads may result in a shorter path, since knowledge of the network is imperfect. The user usually knows the main routes, but only few secondary routes related to his daily routine. This bounded knowledge of the pedestrian network also occurs on EPFL campus.

A solution to balance between the shortest path and the simplest path is to give each edge of the pedestrian network a weight. It represents the aversion to floor changes and less important walkways.

In our example, weights were defined in the routing tool of EPFL website as shown in Algorithm 1. The higher the weight is, the less likely the link is to be selected for the shortest path. The length between floors is defined as being 0.5 meters in our pedestrian network. It explains why the floor factor is quite high compared to the hierarchical factor. In the case of elevators, the length is 0 and this is why we add 40 in the final weight formula.

3.2.2 Paléo music festival

Paléo music festival is one of the major open-air music festivals in Europe, taking place each year for six days in July in Nyon, Switzerland. Its area is around 120'000 square meters and it attracts almost 40'000 spectators per day.

A map is proposed to the spectators (Fig. 3) containing the main facilities of the festival. As a picture, it is not possible to us it for analysis. The festival also generates a GIS database every year for daily management. It is used by the staff and contains ... (Fig. 4). Kondratieva *et al.* (2013) has cleaned the map and selected potential destinations for spectators (Fig. 5): stages, bars and restaurants, stalls, kids activities, toilets and first-aid.

3.3 Capacity and schedule information

Lots of pedestrian facilities have schedules: class schedules on campus, train schedules in a station, flight schedules in an airport, concert schedules in a music festival. These schedules have a very strong impact on pedestrian demand. Moreover, schedules are often known in advance for near future and thus they can be used for forecasting.

Algorithm 1: Weight definition procedure for each edge in the pedestrian network

```
if door = closed then
    weight = ∞;
else
    if Major Route then
        hierarchical factor = 1;
    else if Inter-building Route then
        hierarchical factor = 1.2;
    else if Intra-building Route then
        hierarchical factor = 1.5;
    else if Access to Offices then
        hierarchical factor = 2.0;

    floor factor = 1;
    if Up then
        if Ramp then
            floor factor = 3 ;
        if Stairs then
            floor factor = 15;
    if Down then
        if Ramp then
            floor factor = 2 ;
        if Stairs then
            floor factor = 12;

    lift factor = 0;
    if Elevator then
        elevator factor = 40;

    weight = length · hierarchical factor · floor factor + elevator factor;
```

Similarly, capacity is known in some pedestrian facilities, mostly for security and evacuation reasons. It can be used as an attribute to explain people's choices. There are two kinds of capacities: the maximum capacity, e.g., for security reasons on a train platform or in a restaurant in the opening hours, and the expected capacity, often related to schedules, e.g. a classroom has a theoretical maximum capacity but also has defined number of students registered to classes during class schedules.

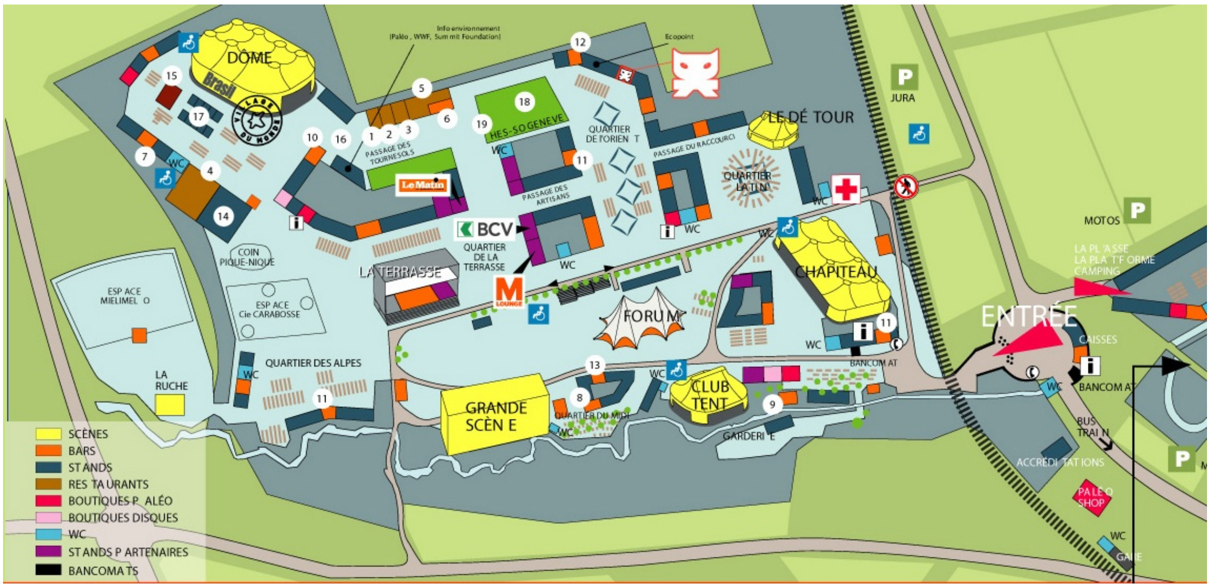

Figure 3: Paléo music festival: map for spectators.

Figure 4: Paléo music festival: Geographically coded map.

4 An activity-based model for pedestrian facilities

The objective is to adapt activity-based approach from urban areas to pedestrian facilities. It will allow to test hypotheses of how pedestrians' travel behavior reacts to changes in different attributes such as distances and pedestrian network, creation of new destinations or modification in schedules.

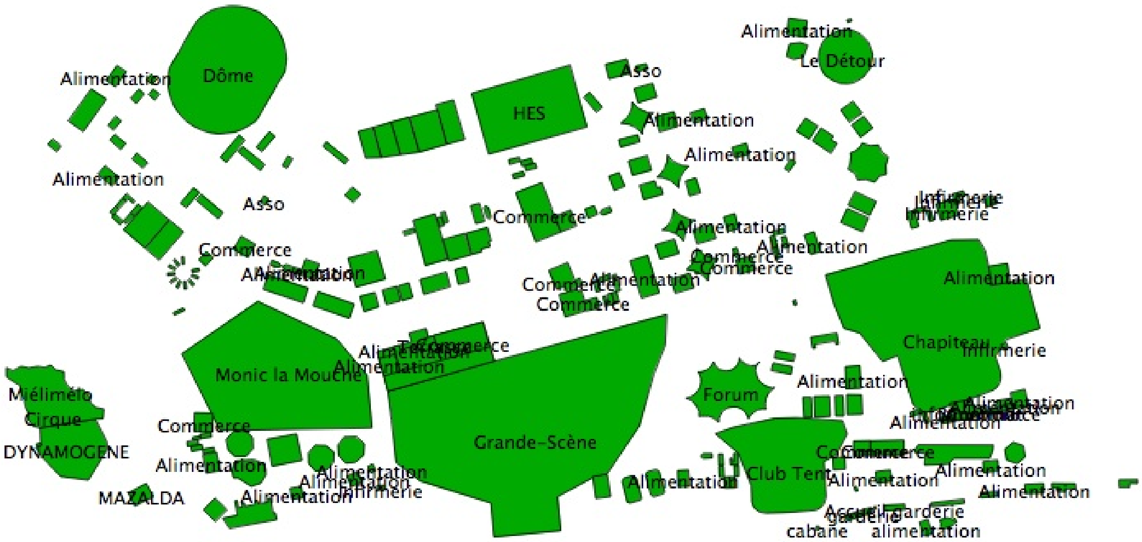

Figure 5: Paléo music festival: Potential pedestrian destinations.

4.1 Activity-based philosophy applied to pedestrians

Inspiration comes from existing activity-based models for urban areas and theoretical developments. However, there are fundamental differences when it comes to pedestrians. First, there is no “home”. If in some specific examples like work, there is a base camp being one’s office, there are usually no equivalent in pedestrian infrastructure. Second, the model is not covering a life, a week or a day, defined mostly by natural time units, but our model is covering a visit, where the beginning and the end are not clearly defined. Third, there is no mode choice. All studied individuals are walking. Fourth, there is no cost involved in moving like for cars or public transport. The closest equivalent to cost function is walking distance. Finally, the studied area is not an urban area but a pedestrian facility, such as an airport, a train station, a supermarket, a music festival, a stadium or a campus.

To paraphrase Hägerstrand (1970), we can ask “what about pedestrians in transportation hubs?”. In his system view and his attempt to unify human and physical geography, Hägerstrawnd is interested in human habitat and the life path of people in time and space. In transportation hubs, like airports or train stations, the same questions arise. The design of these facilities is nowadays mostly defined by security requirement. Crowd science often considers pedestrian as physical particles. But transportation hubs are lively places. The Swiss Railways calls its main stations “Rail Cities”, emphasizing the usage of stations not only as platforms and their accesses but as lively places with shops, tourism offices, post offices, restaurants, and even cultural events. Fig. 6, as a representation of activity-based thinking from Lenntorp (1978), needs to be adapted

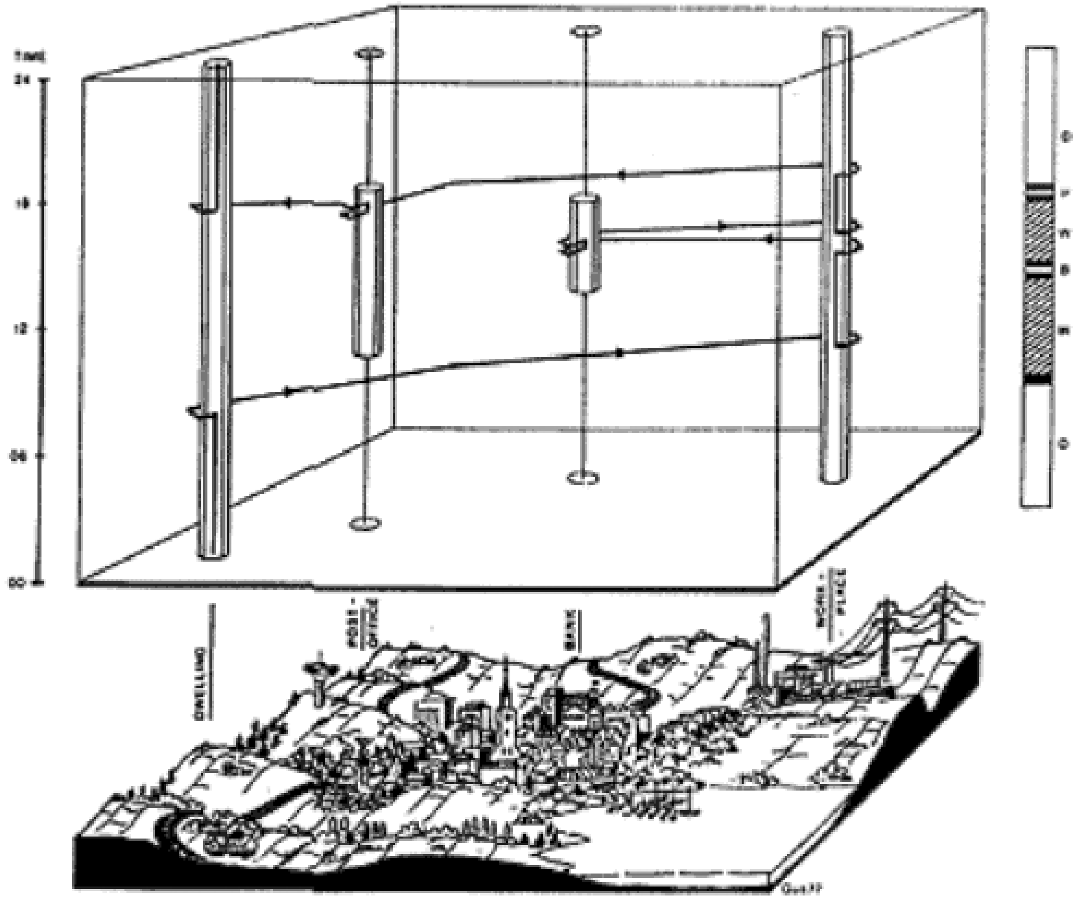

Figure 6: An individual's path in space-time cube. The vertical axis represents both the elevation of the map (below) and the time (top). Daily path is represented in space and time. The constraint related to home and the daily need for sleep can be seen on the left of the picture, where the daily path starts and ends. Source: Lenntorp (1978).

for transportation hubs, with a station instead of a village as the 3D ground of the picture (see Fig. 7).

As a capability constraint (see Section 2.3 for references on the different constraints), even if the necessity for sleeping at night is not present in most pedestrian facilities, a student spending his day on campus must generally eat for lunch. About coupling constraints, Hägerstrand (1970) mentions "compulsory timetables". They are present in lots of pedestrian facilities: class schedules on campus, train schedules in stations, concert schedule in music festivals, opening hours in shops. An authority constraint could be staff access in station.

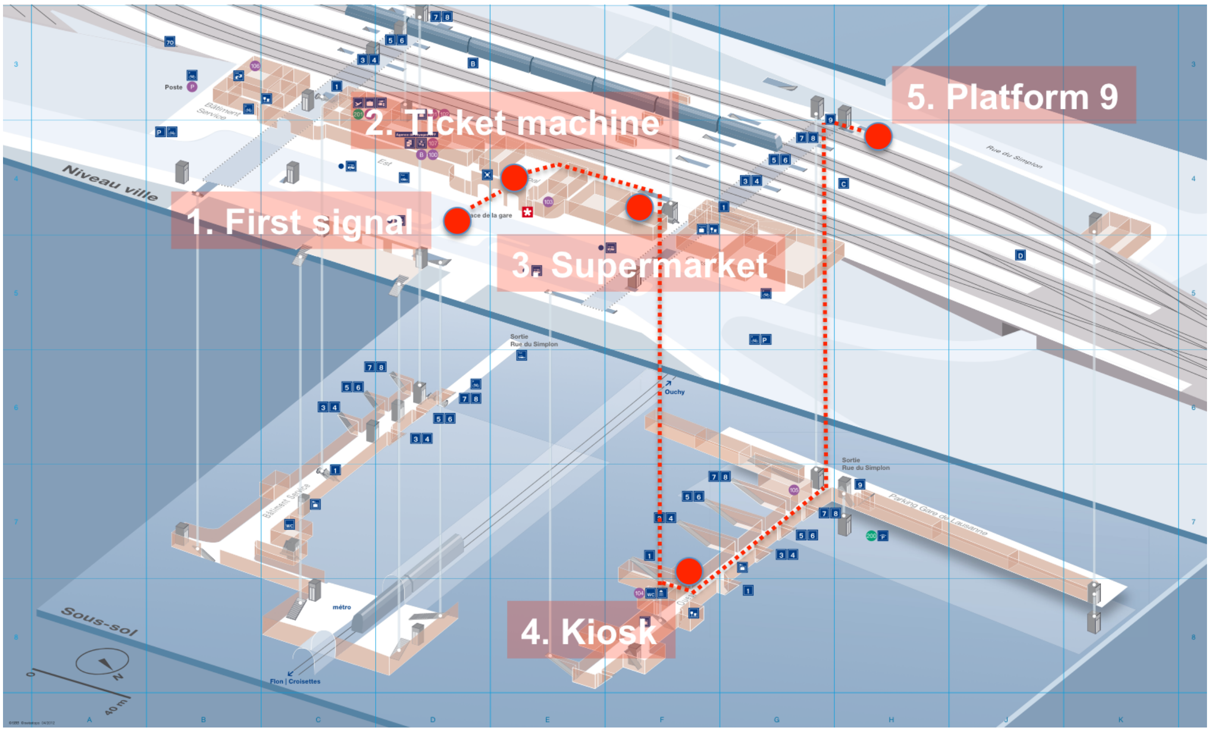

Figure 7: An individual's path in a railway station.

4.2 The modeling system

There are two modeling approaches. The first approach consists in modeling destination choices sequentially: a pedestrian in a facility is facing all possible destinations or a subset depending on some individual limitations. Once the destination is chosen, the duration of the activity is defined by the end time and the choice of the next destination. The literature review in Section 2 and the broad existence of schedules in pedestrian facilities suggest taking into account activity scheduling decision in our model. This is why we don't use the first modeling approach we have just described but a model including activity scheduling decision.

Instead of rule-based simulations, predicting sequential decision process outcomes from decision rules, we focus in this paper on an integrated system of choice models. Discrete choice models are in used for years in transportation modeling, as well as in marketing. They have a strong theoretical basis in consumer theory (Bierlaire, 1998).

4.2.1 Model structure

Inspired by Bowman (1998), two levels of decisions are taken into account. Primary activities are activities that are constrained by schedules or needs and that are planned in advance, before entering the pedestrian facility. They are decisions about the basic agenda and define the activity pattern of the day. Then, secondary activities represent the windows of opportunities that appear during the path in the pedestrian facilities. These opportunistic destinations are probably more common while walking than while driving since there is no need for parking. Moreover, transportation hubs are typical facilities where free time appears, since passengers may come in advance compared to their schedule. On campus, a typical example of primary activities would be "going to class", "going for lunch" or "going to office".

The concept of tours cannot be directly adapted for pedestrian facilities, since there is no "home"[3]. Instead, once the primary activity pattern is defined, the second choice consists in choosing the start and end time of the different activities. For the campus example, "going to class" define a schedule and the choice consists in the departure time from the previous activity-episode in order to be on time, while "going for lunch" define a softer schedule, meaning that the person may choose to adapt the start time of this episode depending on queue and previous experiences.

Finally, once the activity type and the start time are chosen, the destination can be chosen, both for primary activities and secondary ones.

4.2.2 Choice set

The question is to generate a choice set for activity patterns, i.e. alternative activity patterns than the observed one. Adler and Ben-Akiva (1979) showed that the choice set for 0estimation cannot be generated directly from a listing of all possibilities in the traditional urban case. Since the complete choice set is not necessary to estimate a logit model, they proposed to use as alternatives the chosen travel patterns of other households (in our case: of other pedestrians) in the same district.

4.2.3 Choice attribute

We also need to define what are the attributes of the choice of patterns. Adler and Ben-Akiva (1979) proposed a list of variables for households daily activity choices. It needs to be adapted for pedestrians. It consists broadly of schedule convenience of the pattern (number of destinations per tour, number of tours, purpose of tours and time of the day), transport levels-of-service of the pattern (out-of-vehicle travel time, travel distance, total travel time, out-of-pocket cost in relation with income), mode specific variables, destination attraction (total land area of the

[3]There is an exception on campus with the employee's office being the employee's "base camp". Then, utility of the activity pattern depends on office based tours. People may try to optimize their tours from their office.

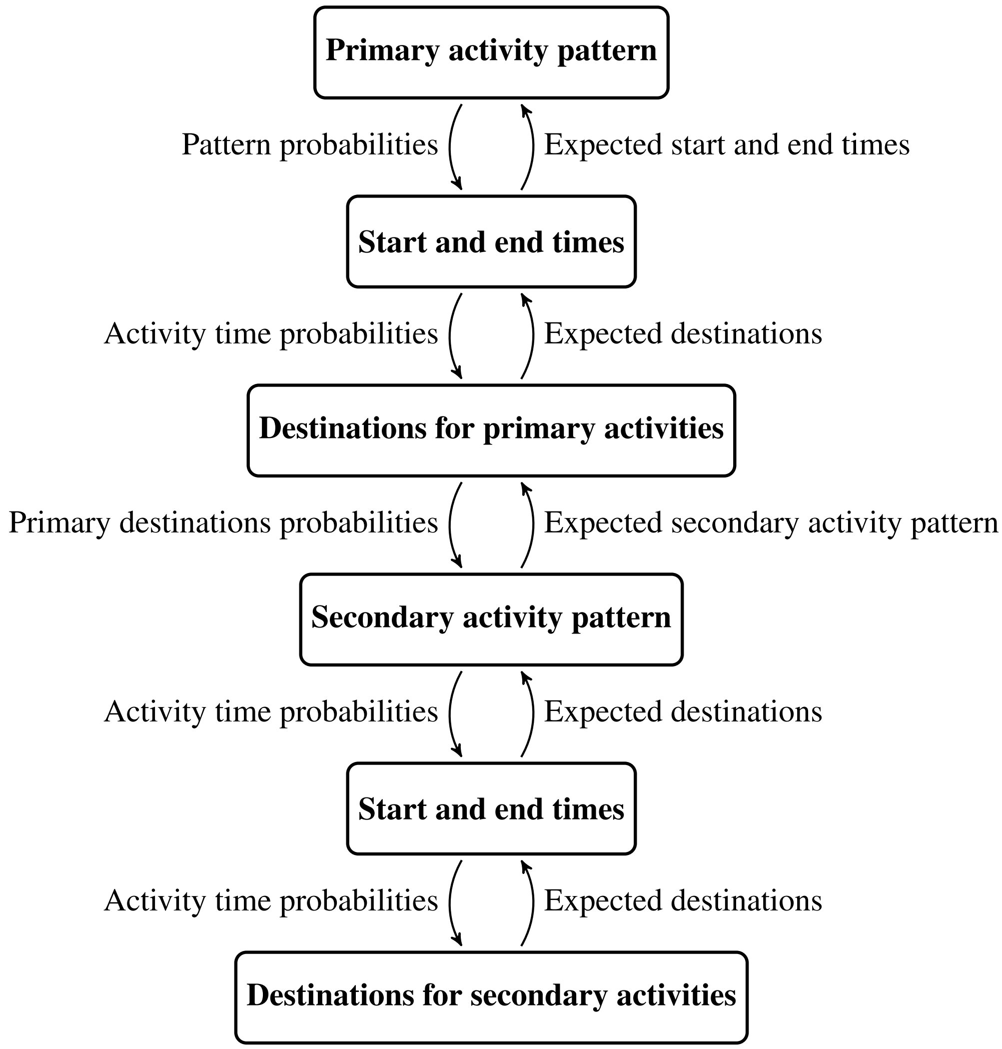

Figure 8: Pedestrian activity model

destination zones, retail employment density, service employment density, tourism areas in the zone), and finally data about the household (income, number of non-workers, information about the location of the household).

With the large variety of pedestrian facilities, attributes will change depending on the context. Cost and modes attributes are of course not applicable in this case. However, there are four families of attributes in pedestrian destination choice: distance, quality of the destinations (intrinsic but also crowdedness and occupancy), socioeconomic and demographic data, and finally schedules shared by many pedestrian facilities.

For pedestrians on campus, the choice of pattern would be strongly related to two base structures. Students are mostly following courses schedules, constrained in both time of day and duration. All other trips are less constrained. The employees are mostly constrained by their office, and it is necessary to detect what is their office, that would correspond to some extent to "household" in traditional activity-based models. Distance would prevent some activities (i.e., on campus people choose a close restaurant for lunch). We collected data using the class membership of students and employees. Thus we know from which class students are coming from, and we also differentiate students from employees. Attractiveness of a destination can be defined by evaluation of courses or restaurants or by the cost of the meal. Also, the number of other activities that may potentially be performed at a destination will influence the choice, as illustrated by the attractiveness of commercial centers or leisure parks in other contexts. DIstances can be computed from the pedestrian graph, as corridors define pedestrian roads.

In a music festival, there are few obstacles but not a network of roads. Distances are defined in an open space. Quality of the destination can be evaluated by queues in toilets, number of other people listening to the same concert, quality of the food (evaluated by experts in Paléo). Socioeconomic and demographic are unknown. Schedules are also present in this example.

5 A simple example

On campus, in our example, we assume there are three types of primary activities, i.e., three types of activities people are planning in advance, in order to simplify the model: going to course and following class schedules, going for lunch, and going in your office if you're an employee. Other activities, such as going for a coffee in the afternoon may be planned in advance in some cases, but we have no information about it.

In Danalet *et al.* (2012) and Danalet *et al.* (2013), we present a bayesian approach to build activity-episode sequences from WiFi traces using capacity as a prior. Here is an example of a daily activity-episode sequence.

The true sequence of activity-episodes has been recorded by the tracked author. He first went in a classroom from 8.32 to 10.30am, then in his office until 11.47am. For lunch break, he arrived in a restaurant on campus at 11.55am. He came back to his office around 1pm and went for a coffee around 2pm. Finally he came back in his office until the end of his working day, around 7.45pm. It is important to notice that the first destination of the day is a doctoral course followed by the author. Table 2 presents the truth and the output from the methodology we developed.

Based on the output from WiFi traces, our pedestrian activity model would suggest that "classroom", "office" and "restaurant" are primary activities and have been planed. Their order, i.e., classroom-office-restaurant-office, have been chosen compared to other activity patterns mostly based on class schedules and expected start and end times, available to the analyst. Conditionally

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

Table 2: Comparison between the output from the bayesian approach from WiFi traces and the truth as reported by one author. Δx represents the walking distance (in meters) between the true location and the output of the model.

on this pattern, the start and end time are defined by the choice maker probably depending on class schedules and expected queue in the restaurant. Finally, once primary activity pattern and start and end times are defined, the primary and secondary destinations are chosen. In some cases, like "classroom between 8.33am and 10.38am", the choice set contains only one alternative: the classroom defined in the schedules (or the classrooms if the course is divided in different rooms). In other cases, the choice set contains the list of all restaurants on campus. Its choice will depend on its intrinsic quality and on the distance between the previous destination and the restaurant.

6 Future works

We plan to implement this model on EPFL campus. There is a risk of low quality of the attributes of activities, i.e., only a little amount of the variance in the choice of patterns and destinations could be explained. Also, the usage of categories of destinations as activity may be misleading. In fact, observing a pedestrian in a cafeteria on campus does not mean the pedestrian is eating. He could be working for courses on one of the tables. Using categories of destinations as activity is a strong assumption. It is still interesting to model activity without asking explicitly pedestrians what they are doing, since it makes data collection much easier.

7 Conclusion

Developments in activity-based modeling is considered as an answer to more complex policies such as road pricing compared to simply enlarge the road network Jovicic (2001). Pedestrian infrastructures are facing the same challenges nowadays. Transport hubs can not be continuously enlarged, particularly in the city centers, and a smarter management of pedestrian demand needs to be implemented. There is a need to optimize the current and new infrastructures that will be developed, in order to answer to the increasing number of passengers, but also the limitations in budget and space. Policy makers and planners need to make optimal decisions on how to use scarce resources, but decision-aid models are missing in this area. These policies won't be related to monetary costs in most cases, but to its closest equivalent in walking trips: distance. The location of different facilities impacts how the demand is distributed in the facility.

Induced demand will appear in transport hubs. Traveling is not the unique activity in transport hubs. By increasing the number of shops in a railway station, more travelers and non-travelers will come in the facility for shopping. By improving the accessibility of different neighborhoods, more inhabitants of the surroundings will just cross the pedestrian facility. To predict this induced demand, our proposed activity-based model for pedestrian facilities must be connected with a larger model of the city in the future.

Our experiment on campus must be seen as a proof-of-concept for transportation hubs. In particular, a parallel is created between the schedules of classes on a campus and the schedules of trains in a railway station or the schedules of flights in an airport. This analogy could even be extended to music festivals and the schedules of concerts. In all theses examples, schedules are a prior information but are not followed by all pedestrians: people are using railway station as supermarket or just as a "bridge" between two parts of the city, students are not following all courses, and spectators in music festivals are also enjoying food and meeting friends.

Apart from being a interesting testbed for transport hubs, our campus study also gives an opportunity to know more about class attendance. The choice of attending or not a class can be related to the course evaluation. In a similar way, the relationship between class attendance and results in the exam can be explored at an aggregate level.

We propose a methodology to capture, understand and forecast pedestrian demand in such infrastructure at a large scale and from innovative and available data. This methodology is general enough to be adapted in the different contexts of pedestrian dynamics. We test it on real data on campus. It fills important gaps in the research community at the strategic level of pedestrian dynamics (and give small insight about students' behavior on campus).

Acknowledgement

The authors would like to thank Elisaveta Kondratieva and Javier Lopez-Montenegro Ramil for their work during their semester projects. Elisaveta provided the pictures from Paléo music festival in this report and Javier the 3D visualization of EPFL pedestrian network.

8 References

Adler, T. and M. Ben-Akiva (1979) A theoretical and empirical model of trip chaining behavior, *Transportation Research Part B: Methodological*, **13** (3) 243–257, ISSN 0191-2615.

Antonini, G., M. Bierlaire and M. Weber (2006) Discrete choice models of pedestrian walking behavior, *Transportation Research Part B: Methodological*, **40** (8) 667–687.

Aschenbruck, N., A. Munjal and T. Camp (2011) Trace-based mobility modeling for multi-hop wireless networks, *Comput. Commun.*, **34** (6) 704–714, May 2011, ISSN 0140-3664.

Axhausen, K. W. (1998) Can We Ever Obtain the Data We Would Like to Have?, *Theoretical Foundations of Travel Choice Modeling*, 234–305.

Balachandran, A., G. M. Voelker, P. Bahl and P. V. Rangan (2002) Characterizing user behavior and network performance in a public wireless LAN, *SIGMETRICS Perform. Eval. Rev.*, **30** (1) 195–205, June 2002, ISSN 0163-5999.

Balajinath, B. and S. V. Raghavan (2001) Intrusion detection through learning behavior model, *Computer Communications*, **24** (12) 1202–1212, ISSN 0140-3664.

Balazinska, M. and P. Castro (2003) Characterizing mobility and network usage in a corporate wireless local-area network, paper presented at the *Proceedings of the 1st international conference on Mobile systems, applications and services*, MobiSys '03, 303–316, New York, NY, USA.

Bierlaire, M. (1998) Discrete choice models, in M. Labbé, G. Laporte, K. Tanczos and P. Toint (eds.) *Operations Research and Decision Aid Methodologies in Traffic and Transportation Management*, vol. 166, 203–227, Springer, ISBN 3540646523.

Bierlaire, M. and T. Robin (2009) Pedestrians Choices, paper presented at the *Pedestrian Behavior. Models, Data Collection and Applications*, 1–26, ISBN 978-1-84855-750-5.

Borgers, A. and H. Timmermans (1986) A model of pedestrian route choice and demand for retail facilities within inner-city shopping areas, *Geographical analysis*, **18** (2) 115–128, April 1986.

Borst, H. C., H. M. E. Miedema, S. I. de Vries, J. M. A. Graham and J. E. F. van Dongen (2001) Relationships between street characteristics and perceived attractiveness for walking reported by elderly people, *Journal of Environmental Psychology*.

Bowman, J. L. (1998) The Day Activity Schedule Approach to Travel Demand Analysis, Ph.D. Thesis, Massachusetts Institute of Technology, May 1998.

Bowman, J. L. and M. Ben-akiva (1996) Activity based travel forecasting, paper presented at the *Conference of activity based travel forecasting (Transcript of a tutorial on activity based travel forecasting), Travel Model Improvement Program, US Department of Transportation and Environmental Protection*, 1–32, New Orleans, Louisiana.

Calabrese, F., G. Di Lorenzo, L. Liu and C. Ratti (2011) Estimating Origin-Destination Flows Using Mobile Phone Location Data, *Pervasive Computing, IEEE*, **10** (4) 36–44, April 2011, ISSN 1536-1268.

Calabrese, F., M. Diao, G. D. Lorenzo, J. F. Jr. and C. Ratti (2013) Understanding individual mobility patterns from urban sensing data: A mobile phone trace example, *Transportation Research Part C: Emerging Technologies*, **26** (0) 301–313, ISSN 0968-090X.

Calabrese, F., J. Reades and C. Ratti (2010) Eigenplaces: Segmenting Space through Digital Signatures, *Pervasive Computing, IEEE*, **9** (1) 78–84, ISSN 1536-1268.

Conti, M. and S. Giordano (2007) Multihop Ad Hoc Networking: The Theory, *Communications Magazine, IEEE*, **45** (4) 78–86, April 2007, ISSN 0163-6804.

Danalet, A., M. Bierlaire and B. Farooq (2012) Estimating Pedestrian Destinations using Traces from WiFi Infrastructures, paper presented at the *Pedestrian and Evacuation Dynamics 2012*.

Danalet, A., M. Bierlaire and B. Farooq (2013) Estimating Pedestrian Activity-Episodes Sequences using Sensors Data.

Eash, R. (1999) Destination and Mode Choice Models for Nonmotorized Travel, *Transportation Research Record: Journal of the Transportation Research Board*, **1674** (-1) 1–8.

Ettema, D., A. Borgers and H. Timmermans (1996) SMASH (Simulation model of activity scheduling heuristics): Some simulations, *Transportation research record*, **1551** (1) 88–94.

Ettema, D., A. Borgers and H. J. P. Timmermans (1994) Using interacting computer experiments for identifying activity scheduling heuristics, paper presented at the *7th IATBR Conference*.

Feil, M. (2010) Choosing the Daily Schedule: Expanding Activity-Based Travel Demand Modelling, Ph.D. Thesis, ETH.

Habib, K. M. N. (2007) Modelling activity generation processes, Ph.D. Thesis, University of Toronto.

Hägerstrand, T. (1970) What about people in regional science?, *Papers in regional science*.

Heggie, I. and P. Jones (1978) Defining domains for models of travel demand, *Transportation*, **7** (2), June 1978, ISSN 0049-4488.

Jones, P. M. (1979) HATS: a technique for investigating household decisions, *Environment and Planning A*, **11** (1) 59–70.

Jovicic, G. (2001) Activity based travel demand modelling, *Note*, **8**, Danmarks TransportForskning.

Kang, J. H., W. Welbourne, B. Stewart and G. Borriello (2004) Extracting places from traces of locations, paper presented at the *Proceedings of the 2nd ACM international workshop on Wireless mobile applications and services on WLAN hotspots*, WMASH '04, 110–118, New York, NY, USA, ISBN 1-58113-877-6.

Kasemsuppakorn, P. and H. A. Karimi (2013) A pedestrian network construction algorithm based on multiple GPS traces, *Transportation Research Part C: Emerging Technologies*, **26** (0) 285–300, ISSN 0968-090X.

Kim, M. and D. Kotz (2006) Extracting a mobility model from real user traces, paper presented at the *In Proceedings of IEEE INFOCOM*.

Kitamura, R., and Ryuichi Kitamura and S. Fujii (1997) Two Computational Process Models Of Activity-Travel Behavior.

Kondratieva, E., A. Danalet and M. Bierlaire (2013) Pedestrian Destinations in Paléo Music Festival from Bluetooth Traces.

Koo, S. G. M., C. Rosenberg, H.-H. Chan and Y. C. Lee (2003) Location discovery in enterprised-based wirless networks: case studies and applications, *Annales des T{é}l{é}communications (Annals of Telecommunications)*, **58** (3-4) 531–552.

Lenntorp, B. (1978) A Time-Geographic Simulation Model of Individual Activity Programmes, in T. Carlstein, D. Parkes and N. Thrift (eds.) *Timing space and spacing time, vol. 2; Human activity and time geography*, 286, Edward Arn, London, ISBN 0-7131-5989-8.

Lopez-Montenegro Ramil, J., A. Danalet and M. Bierlaire (2013) Visualization of pedestrian demand in a 3D graph.

McFadden, D. (1974) The measurement of urban travel demand, *Journal of Public Economics*, **3** (4) 303–328, November 1974, ISSN 00472727.

Meneses, F. and A. Moreira (2012) scale movement analysis from WiFi based location data, paper presented at the *2012 International Conference on Indoor Positioning and Indoor Navigation*.

Mitchell, R. and C. Rapkin (1954) Urban Traffic–A Function of Land Use.

Movahedi Naini, F., O. Dousse, P. Thiran and M. Vetterli (2011) Population Size Estimation Using a Few Individuals as Agents, paper presented at the *2011 {IEEE} International Symposium on Information Theory Proceedings*, 2499–2503.

Nagel, K. and C. L. Barrett (1997) Using Microsimulation Feedback For Trip Adaptation For Realistic Traffic In Dallas, *International Journal of Modern Physics C*, **8** (3) 505–525.

Rigney, C., S. Willens, A. Rubens and W. Simpson (2000) *Remote Authentification Dial in User Service (RADIUS)*, RFC 2865, June 2000.

Rindfüser, G., H. Mühlhans, S. T. Doherty and K. J. Beackmann (2003) Tracing the planning and execution of activities and their attributes: Design and application of a hand-held scheduling process survey, paper presented at the *10th IATBR Conference*.

Rojas, A., P. Branch and G. Armitage (2005) Experimental validation of the random waypoint mobility model through a real world mobility trace for large geographical areas, paper presented at the *Proceedings of the 8th ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems*, MSWiM '05, 174–177, New York, NY, USA, ISBN 1-59593-188-0.

Roorda, M. J. (2005) Activity-based modelling of household travel, Ph.D. Thesis, University of Toronto.

Sevtsuk, A., S. Huang, F. Calabrese and C. Ratti (2009) *Mapping the MIT campus in real time using WiFi*, IGI Global, Hershey, PA.

Stopher, P. and A. Meyburg (1975) *Urban transportation modeling and planning*, Lexington Books, Lexington, Mass., 1975.

Tang, D. and M. Baker (2000) Analysis of a local-area wireless network, paper presented at the *Proceedings of the 6th annual international conference on Mobile computing and networking*, MobiCom '00, 1–10, New York, NY, USA, ISBN 1-58113-197-6.

Timmermans, H., X. V. der Hagen and A. Borgers (1992) Transportation systems, retail environments and pedestrian trip chaining behaviour: modelling issues and applications, *Transportation Research Part B: Methodological*, **26B** (1) 45–59.

Tuduce, C. and T. Gross (2005) A mobility model based on WLAN traces and its validation, paper presented at the *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, vol. 1, 664 – 674 vol. 1, March 2005, ISSN 0743-166X.

Versichele, M., T. Neutens, M. Delafontaine and N. Van De Weghe (2012) The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent Festivities, *Applied Geography*, **32** (2) 208–220, ISSN 01436228.

Wanalertlak, W., B. Lee, C. Yu, M. Kim, S.-M. Park and W.-T. Kim (2011) Behavior-based mobility prediction for seamless handoffs in mobile wireless networks, *Wireless Networks*, **17** (3) 645–658, ISSN 1022-0038.

Weiner, E. (1999) *Urban Transportation Planning in the United States: An Historical Overview*, Greenwood Publishing Group, ISBN 0275963292.

Yoon, J., B. D. Noble and M. Liu (2006) Building realistic mobility models from coarse-grained traces, paper presented at the *in Proc. MobiSys*, 936–5983.

Zhu, W. and H. Timmermans (2005) Exploring Pedestrian Shopping Decision Processes and Application of Gene Expression Programming, in N. Waldau, P. Gatterman, H. Knoflacher and M. Schreckenberg (eds.) *Pedestrian and Evacuation Dynamics*, 145–154, Springer.

Zola, E. and F. Barcelo-Arroyo (2011) A comparative analysis of the user behavior in academic WiFi networks, paper presented at the *Proceedings of the 6th ACM workshop on Performance monitoring and measurement of heterogeneous wireless and wired networks*, PM2HW2N '11, 59–66, New York, NY, USA, ISBN 978-1-4503-0902-8.