Probabilistic Pedestrian OD Matrix based on Digital Footprints

Antonin Danalet *

Transport and Mobility Laboratory,

School of Architecture, Civil and Environmental Engineering,

Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Michel Bierlaire

Transport and Mobility Laboratory,

School of Architecture, Civil and Environmental Engineering,

Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Bilal Farooq

Transport and Mobility Laboratory,

School of Architecture, Civil and Environmental Engineering,

Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

*Email: antonin.danalet@epfl.ch

1 Pedestrian Data Collections

Gathering data about pedestrian origins, destinations and routes is difficult, in particular indoor and on a large scale. Different sensors could be used, such as Bluetooth or GPS signals from smartphones or cameras. Such data are important for route choice modeling, description of congestion, efficient design of new facilities, and travel guidance and information systems.

According to a report by Cisco [1], there will be more than 7 billions smartphones and tablets on Earth end of 2012. In this context, research has been done on digital footprints on the smartphone for route choice [2] or real-time traffic monitoring [5]. Most data collections are device-centric. In this paper, we focus on the network infrastructure and the resulting digital footprints that exists in WiFi access points (APs) and cell towers.

Both WiFi and GSM traces allow for the localization of mobile devices, but they are scarce and fuzzy. We propose a methodology to use these traces to generate pedestrian OD matrices.

Using data from communication infrastructure has already been done, both with WiFi and GSM traces. Calabrese et al. [4] propose an OD matrix for the Boston metropolitan area based on data from the cell tower the mobile phone was connected to. Origins and destinations are regions, they compare they results with existing statistics and they don't consider the mode nor the underlying network. A large literature exists about WiFi as well, defining mobility models not as a tool to understand mobility itself, but in order to improve the quality of the WiFi (minimizing the handoff delay incurred scanning for available APs). The goal is to predict the next point-of-attachment of the user. Very often, field studies are done on campuses, but a recent paper [7] also applied its methodology to Municipal Wireless networks. In [7] and [8], the underlying mobility infrastructure is considered. They consider as destinations either buildings [8] or APs [7].

2 Destination Generation Based on WiFi Traces

We propose to extend this work with a focus on mobility behaviors of pedestrians. Our goal is not to improve the Wireless network, but to gather information on pedestrians from these data. We use the underlying pedestrian network, but also information about latent classes in the population, and information about potential destinations.

APs are usually not compliant with the behavior we would like to model. Thus, instead of considering a group of APs or a single AP as a destination ([4], [7], [8]), we consider the different points of interest around each AP as potential destinations, using the concept of "domain of data relevance" (DDR) [3].

We use EPFL campus as a field study. Two different datasets are available. The first one is based on the AP to which the device is connected. We have a signal every time the device creates a connection with the AP (see Figure 3). The second dataset uses triangulation from the signals received by all APs around the device. These data have been collected in collaboration with Cisco, using their Context Aware Mobility Service and are unique for this purpose to our knowledge. Both data collections are still running.

Based on the methodology developed in [6], we create both a structural and a measure-

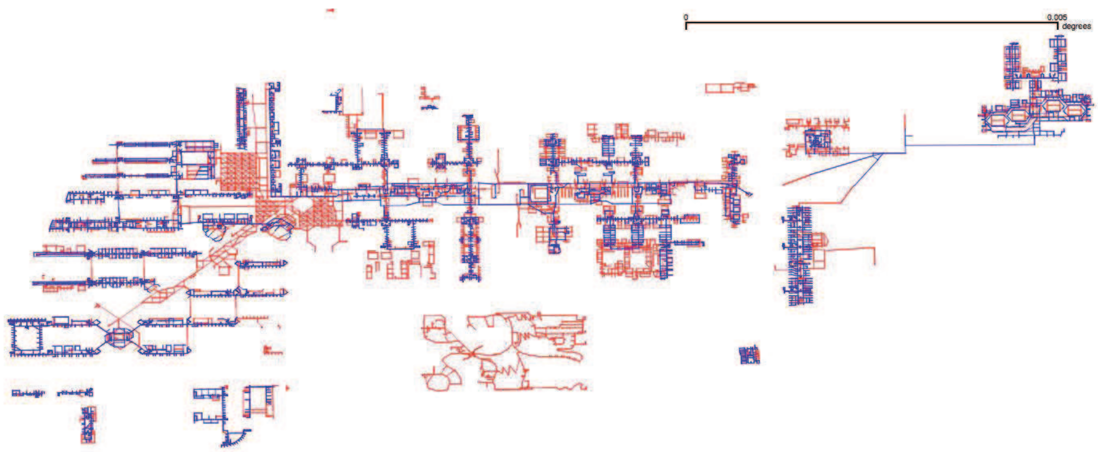

Figure 1: EPFL Pedestrian network. Red links are 1st floor, blue links are 2nd floor. We can see the shape of the buildings of the campus.

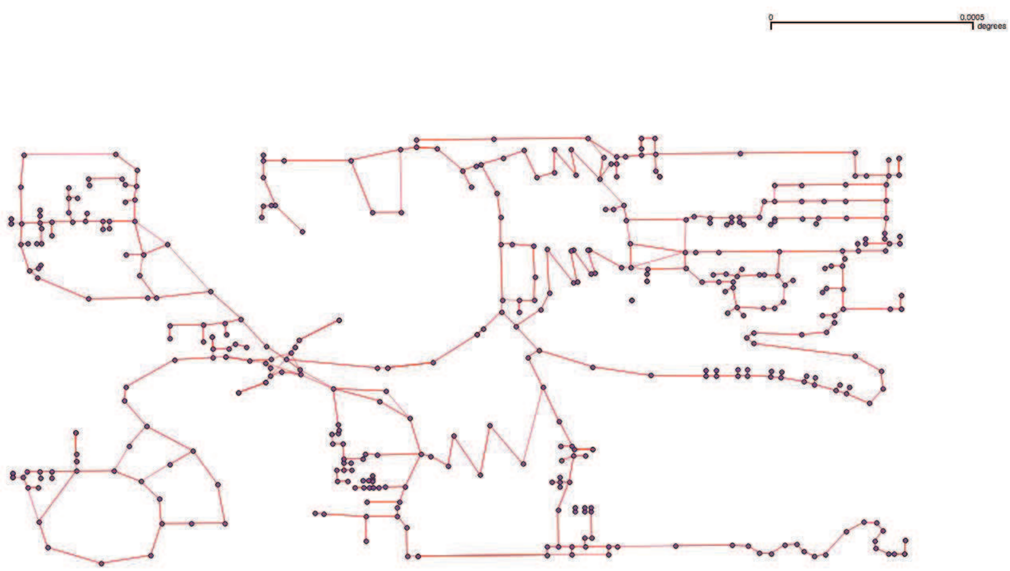

Figure 2: The Rolex Learning Center pedestrian network at EPFL. Contrarily to other buildings, the structure of the network is not a grid.

ment model. The structural model defines the next-destination prediction, based on time of the day, type of activities at destination and latent classes. It is independent from the data. More precisely, a destination model predicts the next destination of the device and for each destination both the arrival and departure time at this destination.

The measurement model gives a probability that a given list of destinations generates the data. This probability is decomposed recursively by destination. Spanning paths between destinations, we can compute the distance to the AP and thus the likelihood this AP generated a signal.

Since the generation of lists of destinations is facing a combinatorial effect, we use the domain of data relevance (DDR) to limit this generation to a tractable number of possible states. For each signal, we define the DDR and connect all destinations in this DDR to the previous destination. This process leads to the definition of intermediary destinations, i.e., signals while moving, that we need to filter. Between intermediary destinations, we assume the device to follow a weighted shortest path.

This framework gives the opportunity to use other data in the model. We can assume that people with an office don't have the same mobility patterns compared to students: the first ones are always coming back to their office, while the second one are not. We also know the proportion of students and employees on the campus, and thus we can use latent classes (since we don't know which signal correspond to which class). In the same way, we can use data about the access to campus. A survey was done about the transport mode used to reach the campus, and it gives an idea about the first destination of the users.

As a conclusion, we present a methodology for digital footprints of pedestrians, based on behavioral foundations with real destinations, and on a unique data collection. It allows us to explore the generation of OD matrices and route choices for pedestrians. These advances will benefit city centers, transportation hubs or commercial centers dealing with congestion, proposing travel guidance or planning a new pedestrian facility.

References

[1] Cisco visual networking index: Global mobile data traffic forecast update, 2011–2016, February 2012.

[2] Michel Bierlaire, Jingmin Chen, and Jeffrey P Newman. Using location observations to observe routing for choice models. In *Proceedings of the 89th Transportation Research Board Annual Meeting*, Washington D.C., US, 2010.

[3] Michel Bierlaire and Emma Frejinger. Route choice modeling with network-free data. *Transportation Research Part C: Emerging Technologies*, 16(2):187–198, 2008.

[4] F. Calabrese, G. Di Lorenzo, Liang Liu, and C. Ratti. Estimating origin-destination flows using mobile phone location data. *Pervasive Computing, IEEE*, 10(4):36 –44, april 2011.

[5] Juan C. Herrera, Daniel B. Work, Ryan Herring, Xuegang (Jeff) Ban, Quinn Jacobson, and Alexandre M. Bayen. Evaluation of traffic data obtained via gps-enabled mobile phones: The mobile century field experiment. *Transportation Research Part C: Emerging Technologies*, 18(4):568 – 583, 2010.

[6] Jeffrey P Newman, Jingmin Chen, and Michel Bierlaire. Generating probabilistic path observation from gps data for route choice modeling. In *Proceedings of the European Transport Conference*, Leiden Leeuwenhorst Conference Centre , Netherlands, 2009.

[7] Weetit Wanalertlak, Ben Lee, Chansu Yu, Myungchul Kim, Seung-Min Park, and Won-Tae Kim. Behavior-based mobility prediction for seamless handoffs in mobile wireless networks. *Wireless Networks*, 17:645–658, 2011. 10.1007/s11276-010-0303-x.

[8] Jungkeun Yoon, Brian D. Noble, and Mingyan Liu. Building realistic mobility models from coarse-grained traces. In *in Proc. MobiSys*, pages 936–5983. ACM Press, 2006.

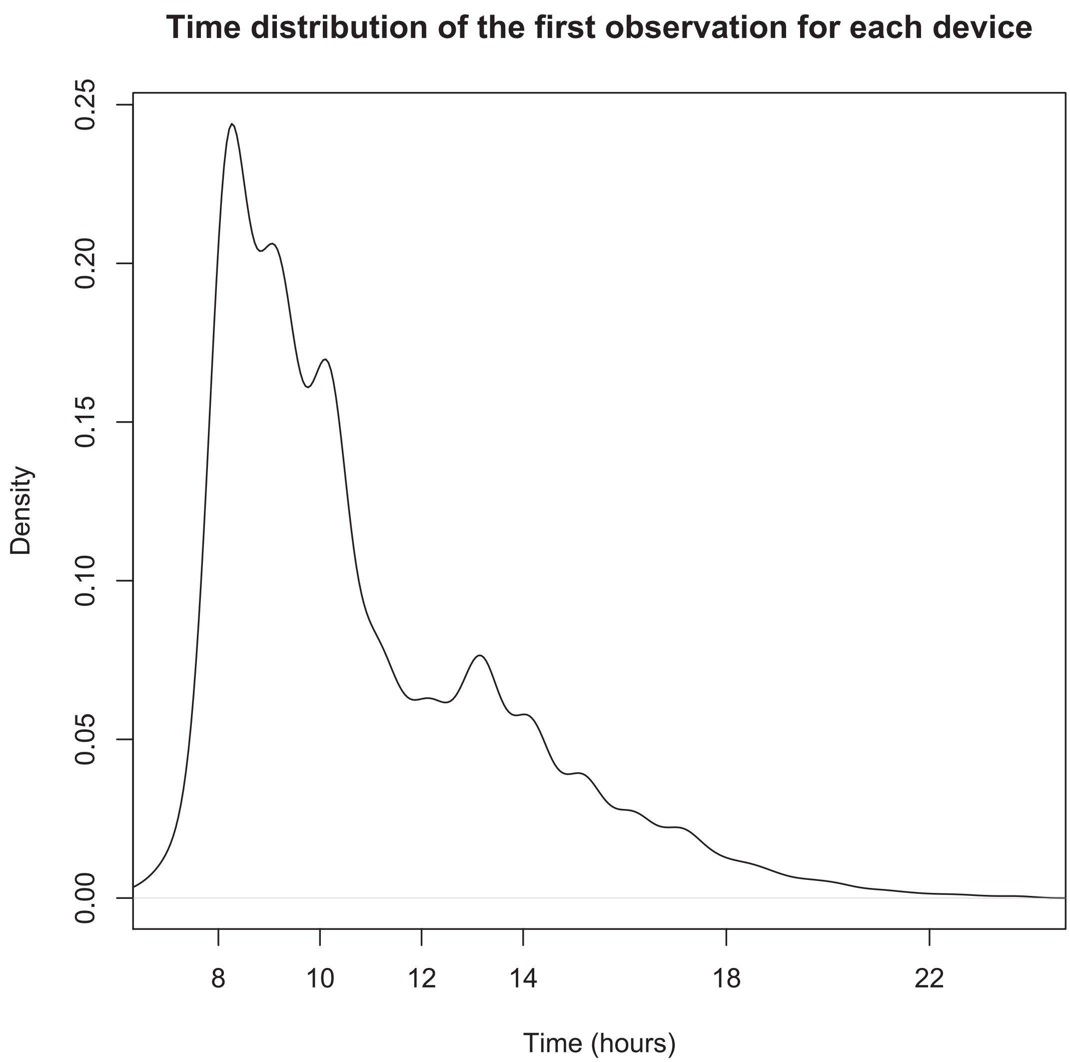

Figure 3: Arrival times based on WiFi traces (first dataset, 10 consecutive working days).