Pedestrian activity pattern mining in WiFi-network connection data

Guilhem Poucin*
Master's student
Civil Engineering Department
École Polytechnique de Montréal
2500, Chemin de Polytechnique, Montreal
Tel: (514) 340-4711
E-mail: guilhem.poucin@gmail.com

Bilal Farooq, Ph.D. Ing. Jr.
Assistant Professor
École Polytechnique de Montréal
2500, Chemin de Polytechnique, Montreal
Tel: (514) 340-4711 ext. 4802
E-mail: bilal.farooq@polymtl.ca

Zachary Patterson, Ph.D.
Associate Professor
Department of Geography, Planning and the Environment, Concordia University
1255-15, Hall Building, 1455 De Maisonneuve W.,
Montreal, Quebec, Canada
Tel: (514) 848-2424 ext 3492
Fax: (514) 848 2032
zachary.patterson@concordia.ca

Word count: **5,209** words text + 9 tables/figures x 250 words (each) = **7,459** words

* Corresponding author

Submission date: 1 August 2015

Submitted for presentation to the 95th Annual Meeting of the Transportation Research Board and publication in Transportation Research Record: Journal of the Transportation Research Board.

Abstract

This article proposes a methodology to mine valuable information about pedestrian use of a facility based only on the WiFi network connection history data. Data are collected in Concordia University, Montreal, Canada. Working with a limited set of information, we tried to characterize the different pedestrian activity patterns in an analytic way without the prior knowledge of the different locations covered by the WiFi connection data. The goal of this research is to develop an analytical tool that is spatially transferable to different facilities. Moreover it is able to distinguish the main pedestrian activity patterns by looking at the WiFi network logs only. The methodology is based on the identification and generation of pertinent variables for data clustering and time-space activity identification. A K-means clustering algorithm is then used for the construction of a set of 6 activity patterns associated with activities in a campus context. We then discuss a few potential additional applications by analysing the inter-access point behaviour that WiFi connection data offer, as well as the challenges caused by space-time inaccuracies.

Keywords: Information mining; Activity pattern recognition; WiFi network; Pedestrian movement; Big Data

I - Introduction

Nowadays, the vast majority of collected data concerning mobility are obtained through surveys or studies based on samples of known populations. At the same time, traditional methods of data collection are expensive in terms of direct costs as well as time and may not represent actual behaviour of users due to sampling biases and even the reliability of data reported by respondents.

The increasing availability of passively collected Big Data and the spread of smartphones are opening interesting opportunities for automatic data collection, processing and behavioural inference. Even if using ubiquitous data from wireless network information, GSM traces, smart-card data is challenging due to the lack of information concerning users, caused by the common necessity of anonymizing the data, they still present non negligible assets. These new sources have large quantities of ubiquitous data, providing us with the opportunity to capture population level behaviour; contribute to the parametrization of models; and far reduced cost options than traditional methods.

A considerable amount of work using wireless network data has already been done, especially with respect to network optimization and the computation of pedestrian traces and the prediction of pedestrian behaviour (see section II). The work presented here, while using individual wireless connections, focuses not so much on the individuals connecting to the access points, but rather on the access points themselves and what can be learned about the activities taking place around them. The reason for this is that when beginning this project, while we were able to obtain wireless access point connection data, we weren't granted access to any complementary information (such as access point location). Thus, the work began identifying what could be done using only the data available.

As a result, the first part of this paper is dedicated to the study of intra-access point relationships. We studied connection patterns at access points in order to understand and categorize the activities taking place in proximity to them. This part of the work has many applications including analysis of activities in multi-purpose environments such as festivals or other public events. In the second part, we focus on inter-access point data discussing the possibilities offered in terms of visualization and analysis, as well as the limits of the use of such data.

The rest of the paper is structured as follows: a literature of current literature on the use of WiFi data is followed by a description of the case study location and the dataset used in the analysis. This leads to a section describing the methodology adopted in the analysis and the results related to the classification of access points in terms of their surrounding activities as well as an analysis of activity transitions. This section finishes with a description of the opportunities for using of this same data in the calculation of travel times between access points, and thereby locations in the buildings in which they are found. The last section provides a summary of conclusions from the work and a description of fruitful areas of future research.

II - Literature review

WiFi signal treatment in the field transportation can be classified between device- and network-centric approaches. Device-centric (i.e. methods collecting data directly from mobile phones, through for example the use of apps methods) have the advantage of being able to associate user characteristics with data that are collected, but typically require the development or tailoring of applications, rely on users to install and keep applications running, and samples tend to be small and of questionable representativity. Network-centric approaches (i.e. methods using data collected from wireless and mobile networks) have the disadvantage of having limited information about the characteristics of “respondents,” but allow the collection of exhaustive and potentially comprehensive sets of data thanks to their ubiquity. As a result, some have argued that network-centric approach seem better suited to pedestrian facility analysis [1]. In this article, we focus on network based approaches and as a result concentrate on that literature in this review. The interested reader can, however, refer to the following literature concerning device centric approach for more information [2][3][4][5].

In the literature concerned with network-based data collection approaches and mobility, few challenges related to the processing of network-based data before it can be used successfully for analysis are highlighted. A first challenge relates to the importance of, and issues surrounding, the protection of individual identity in the use of such data. While the importance of assuring anonymity has been raised, being able to identify users' characteristics can be a complex task [6]. Another important issue network-centred data processing encounter is the so-called Ping Pong effect. This arises when the user is connected and disconnected to different access points (APs) when the user is not mobile [1]. When this happens, non-existent trips can be created within data sets. A number of solutions have been proposed to deal with this issue [7][8].

A second aspect of literature on network-centred data collection approaches relates to applications to which these data have been applied. The first type of application has been to use such data to optimize the placement of WiFi network infrastructure [9]. A second application has been the use of this data to improve the connectivity of the network by predicting the next Access Point (AP) of attachment using pedestrian activity based choice models [10][12]. A third related application is the computation and treatment of the spatial localisation of users obtained through various procedures (see [8]) and that serve as the basis for the development of mobility models more generally. A good review of such models can be found in [11].

At a larger scale, studies have been made on campuses and offices using WiFi connection data to analyse global mobility behaviour between different APs. Meneses & Moreira [13], for example, use connection rates to compute the most important corridors.

From a theoretical perspective, there have been attempts to move away from a strict geographical concept of location (i.e. coordinates in space) to include in the notion of locations, the activities that are conducted there and as a result, to emphasize the concept of “place” in the description of a wireless environment [14]. From this perspective, it is though that users are more interested by the type of activities taking place in a location rather than its spatial location. Such considerations are taken into account by some mobility models (e.g. [1]) through a function associating activities to places in an enriched graph, but those functions are obtained through a previous knowledge of the services furnished within the places.

Related to this emphasis on activities surrounding APs, Calabrese et al. [15] introduce the possibility of identifying the type of activities taking place around WiFi APs, adopting an eigenvector analysis of signals in connection patterns. The use of signal theory methods, such as signal decomposition and unsupervised machine learning on WiFi connection data allows them to classify APs into 5 clusters associated with different types of activity.

As such literature on network-centred data collection has been concerned primarily with the processing of such data before it can be used, how the data can be used to better plan wireless infrastructure, as well as observing and modelling individual movements through such networks? A relatively unexplored aspect of research in this field concerns not so much individual users, but rather the characterization of activities associated with APs, based on network-centred data. In this article, we extend this burgeoning research by concentrating on APs, and not on users. As such we propose (and take) an AP-centred approach as opposed to a user-centred approach. Using this approach we explore some aspects of how WiFi network connection data can be used to associate APs with different types of activities, as well as the potential to use such data in analysing inter-AP connectivity.

III - Concordia University Access Point Log Dataset

The data used in the research were provided by Concordia University in Montreal, Canada. Concordia Concordia University is made up of 57 buildings across two campuses (one in downtown, and the other in suburban, Montreal). The university counts of 47,000 students, faculty and staff with 36,000 undergraduate students. In this study, we analize connection data from all APs of the campus during one week (02/02/15-08/02/15). The data identifies each connected individual through a unique code associated to the device's Media Access Control (MAC) address in a confidential database to provide users anonymity. Each connection record includes an access point ID, user device ID, connection and disconnection time.

We expect that this sample of device connections will be representative of the activity on the Concordia campus despite the fact that non connected individuals are invisible and over connected individuals (users using more than one connected device e.g. smartphone/laptop/tablet) are over represented. Past studies and our experience in previous usage of similar data in public spaces (e.g. outdoor street festival, train station) suggest that the sample from WiFi connection data represents 20% to 30% of the population. As this case study concerns a university campus, we expect the percentage to be even higher thus giving us a greater confidence in terms of representativity [16].

Summary results for the data in question allows some basic observations. We observe over the course of the week there were almost 1.7 million connections by 60,500 devices to over 1,048 APs throughout the campus. Maximum connection time recorded was 5 days and 20 minutes, surely a fixed devices such as desktop computer with a have very long connection time.

As mentioned above, there are many buildings on the two campuses of the university. Instead of considering all APs across all buildings, we concentrated on the APs of only one building. As such, in this article we focus on the set of data concerning the largest building (John Molson building) during a Wednesday, considered as the most representative day of the week (04/02/15).

While we were provided with AP log data, we were not provided initially with the locations of APs throughout the campus. At the same time the AP IDs allowed us to determine the building and floor where the APs were located. While accurate location information would have been ideal, we decided to explore what we could do with this non-localized data. As a result, we decided to focus our study on the inference of activities surrounding APs only using device connection information available through the logs without knowledge of the buildings, activities, or AP locations.

IV - Intra-AP Activity analysis and pattern clustering

This section describes both the methodology adopted using information on wireless AP connections to derive activities in locations near to these APs, as well as the results of this clustering process.

1 - Pattern definition and pertinent clustering variables

To characterize the connectivity of an AP, we generate a vector containing the number of connections by time at intervals of 5 minutes. This provides us an intuitive representation of the use of an AP over the day, as shown Figure 1. We refer to this type of a graph as a "connection profile." The first step of our analysis was to analyze the connection profile of each AP individually as an isolated system. That is, we generate time independent variables to characterize the observed signal.

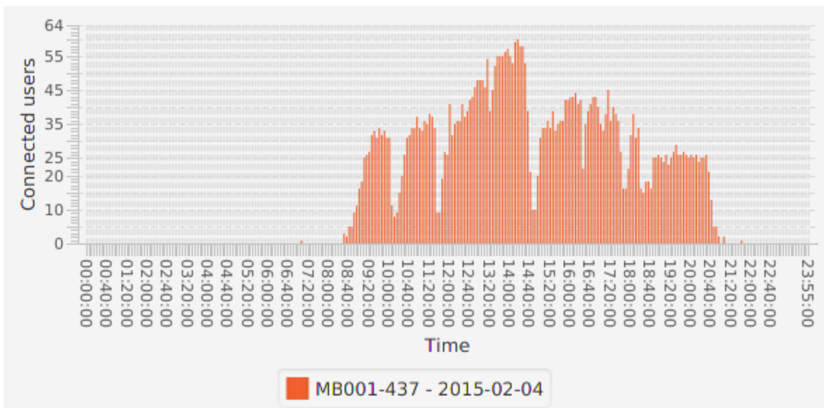

Fig. 1. Connection profile of an AP during one day.

While signal decomposition through a Fourier series would have been a possible avenue to describe the main characteristics of the signal observable Figure 1, we chose not to use it in order to be able to incorporate other dimensions to which we had access in characterizing APs. In particular, while the connection profile provides information about the number of connections to an AP, a number of other characteristics of the connections were available to us from the AP logs. After a great deal of examination and comparison of different connection profiles, we chose three variables concerning the user connections and four concerning signal characterization.

As duration is a fundamental dimension of activity, we chose to study average connection time at each access point and its standard deviation. The third variable is the ratio between the number of connections and the number of different MAC addresses registered at these locations during the day. This was included to inform us on the potential repetitivity of users in the set of data staying independent from the connections number.

The characterization of the signal was made through the computation of the two first derivatives of the connexions number as instantaneous rates of changes through time. We didn't use the number of connections during the day because we didn't believe that it had an influence on the activities associated with APs. We then extracted the amplitude and standard deviations (equivalent here to the quadratic mean, the standard mean being null). While the standard deviation was intended to inform us on the global quantity of variation, the amplitude provided us information on their interval of variation. We are conscious that such a characterization could be considered as poor for a signal, but we wanted to keep a balance between location and user perspective for our clustering.

2 - Clustering and Pattern Identification

Many clustering algorithms exist and as suggested by [15], unsupervised machine learning algorithms such as the k-means algorithm, help to avoid the transposition of expectations on clustering results. However, the k-means algorithm requires the specification of the number of clusters as a prerequisite. The appropriate number of clusters can be extrapolated from previous knowledge of the number of classes expected or through mathematical methods.

We first compute the optimal within cluster sum of squares for each number of clusters and 1000 different initial configurations (computed automatically by R). Such results allow us to visualize the potential impact of the cluster number variation on the cluster fit.

Figure 2 obtained from this analysis shows that the variation of the value that we want to minimize decreases slowly after 4 clusters. As such, we considered 4, to be the minimum number of clusters that would be appropriate.

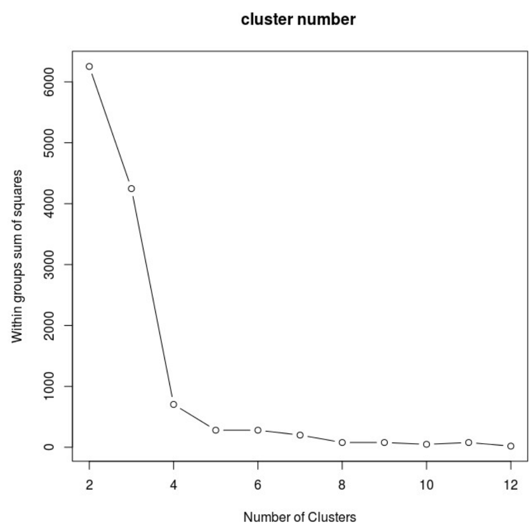

Fig. 2. Within sum of square of the optimal solution found for each number of cluster.

We then manually analysed the different connection profiles and created a subset of “representative patterns” classified in four families: *corridors, shared places, classrooms* and *offices and laboratories*. We used this subset of locations as a calibration tool for the machine learning process, modifying the parameters of the algorithm (number of clusters and initial centroids) until we found a satisfying clustering. We obtained a 6 cluster partition of the subset of profiles, 2 of the initial clusters being each split in 2 more accurate sub groups.

We then used the centroids obtained during the calibration step as the initial conditions for the entire population clustering. Doing so rather than fixing these centroids can lead the clustering to a non-optimal solution "because of the probability that usage at some outlying APs deviates so far from the norm that it skews the clustering process toward solutions in which most clusters contain just a few extreme APs [15]. We chose to proceed in this way because of the partial character of our calibration subset of signal and its potential non representativity of the global behaviour of access points.

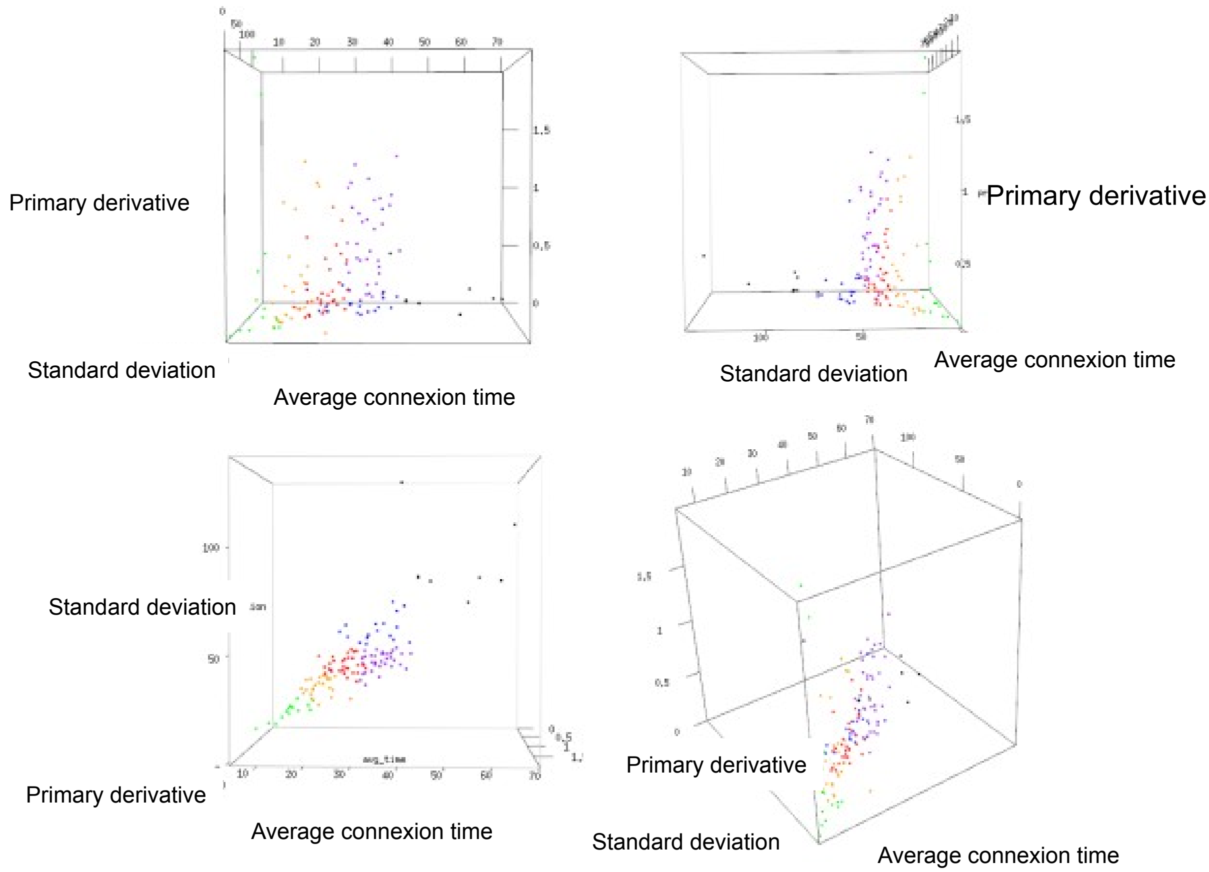

Fig. 3. Representation of the 6 cluster function of three of the clustering variables

Figure 3 shows the different APs, classified by the obtained clusters as function of the average connection time, the standard deviation of the connection time and the standard deviation of the primary derivative of the connection number through time. At first glance, one can observe that there are some extreme observations, but they are captured in non-negligible clusters (the smallest cluster contains 8 access points or 5.7% of the total population against 26% for the biggest). As well, the proximity of some clusters shows a limit of the approach: even if the locations show some characteristics and patterns allowing them to be classified into a “main activity,”, the gap between different types of APs can be small. Indeed, we considered in our study, that each place/access point would be associated to a unique activity which is a simplification of reality

Table 1 shows the 5 minute aggregation rate clustering variables for each cluster. We associated each cluster to a characteristic type of activity and campus place (discussed further in the article). These data can be compared with the connection profiles of representative places from each cluster, proposed in Figure 4.

Table 1. Characteristics of the 6 clusters

Cluster	1	2	3	4	5	6
Name	Shared places	Corridors	Laboratories/ offices	Offices	Auditorium	Classrooms
Number of access point	32	14	17	8	36	23
Average number of connexions during day	246.4	592.4	129.5	86.6	494.9	302.8
Ratio number of device/number of devices	0.54	0.76	0.48	0.52	0.55	0.61
Average time (mn)	24.66	10.71	32.92	53.3	34.66	18.33
Average time standard deviation	37.17	11.1	57.32	95.57	41.11	25.06
Primary derivative of connection number standard deviation	0.317	0.425	0.198	0.158	0.636	0.406
Primary derivative of connection number amplitude	3.32	4.31	2.07	1.35	7.37	4.44
Secondary derivative of connection number standard deviation	0.094	0.124	0.06	0.043	0.17	0.113
Secondary derivative of connection number amplitude	0.892	1.211	0.595	0.37	1.879	1.233

Cluster 1: Shared places

These locations are public places shared by all users of the building such as cafeterias or common working spaces. They often have an increase in connection number during lunch times (an average increase of 200% of the number of connections between 12:00 and 15:00 in the case of the Figure 4.A). The number of connections tend to vary in a progressive way as those places fill and empty progressively.

Cluster 2: Corridors

Corridor access points are passing places, they are characterized by short connection times and a fluctuating number of connections (corresponding to the noise observable Figure 4.B). The low standard deviation show that the activity times don't vary much.

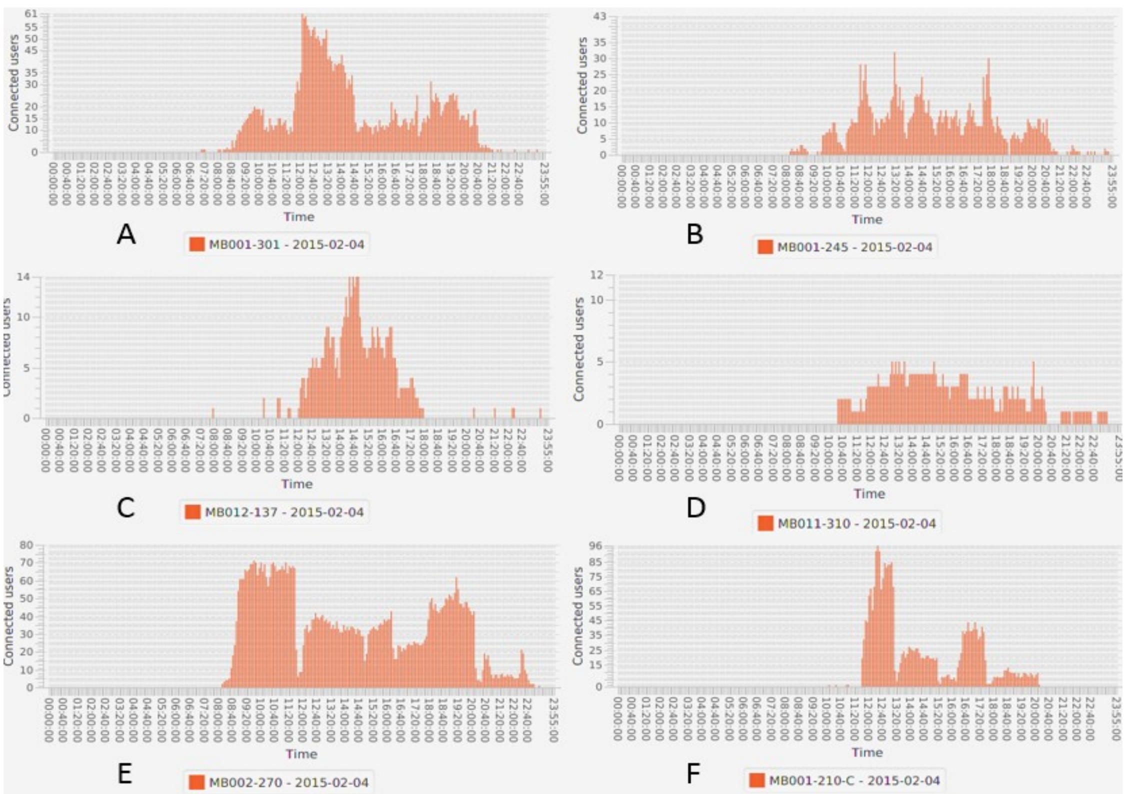

Fig. 4. A-F. Representative connection profile of each cluster

Cluster 3: Laboratories/private places

These locations have a low number of connections and a strong variation within this small number of connections, depending on the hour of the day (Figure 4.C).

Cluster 4: Offices

Offices have a very low number of connection but are very regular in their behaviour with low variation in number of connections. The average connection time and its standard deviation are the highest of the 6 clusters (Figure 4.D).

Cluster 5: Classroom/auditorium

The classroom clusters are characterized by very strong and punctual variations (explaining the high standard deviation of the derivative). At the same time, the connection number stays relatively constant during the peaks: number of users vary before and after the class but not during the activity time interval (Figure 4.E). This cluster shows a large number of connected people and average connection time than the cluster 6: we supposed than the auditoriums hosting various kind of event could show this behaviour.

Cluster 6: Classroom

This cluster exhibits similar behaviour as cluster 5, but with a lower usage (Figure 4.F). We hypothesize that cluster 5 may include undergraduate classes with high attendance, while cluster 6 may include graduate classes that are usually in the afternoon and where the number of attendees is lower.

3 - Visualization of the activity

As mentioned previously, we didn’t get access to the maps of the university preventing us from proceeding to an accurate spatial analysis or representation of the data. However, the Access Point ID allowed us to extrapolate the floors of the locations. Figure 5 represent the distribution of different access point activity clusters for each floor of John Molson building on the downtown campus of the University. The low number of access points on floors 7 and 13 and the missing 9 floor come from the data's structure and are not a voluntary filtering. Such data give information about the use that is made from the facilities by the users.

We observe that shared place APs are present on most of the floors, which is not surprising considering the fact that the waiting/working/eating activities are often performed close to the other activity locations. Two main corridors can be observed: on the first floor corresponding to the entrance of the building, and the second on the 10th floor which seems to be a transition floor. Indeed, while most of the classroom activity APs are located under the 10 floor, most of the labs and offices are located above.

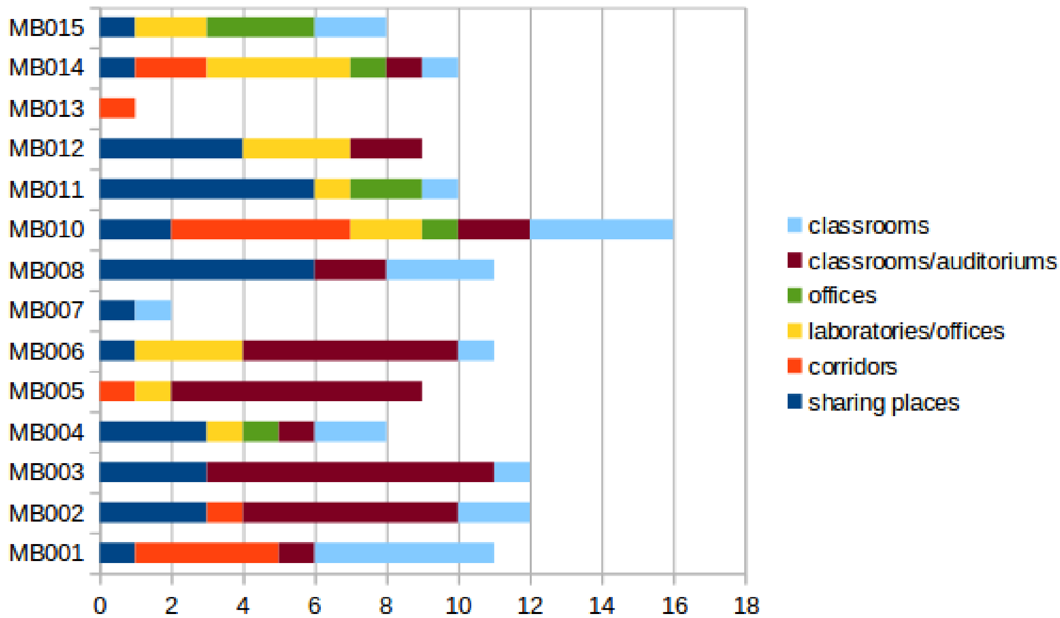

Fig. 5 Distribution of activities within floors

To conclude, in this section, we presented a methodology to extract general information about facility use only from WiFi Access Points logs. Such an approach is applicable in environments in which the places purposes are not well defined in advance e.g. public places for events or festivals. Then it would be possible to use that kind of work as environment definition for mobility analysis such the one realized in [1]. If the use of a clustering algorithm on a set of objects with close characteristics brings sensitivity, it is acceptable because this analysis studies patterns and similarities in wireless network usage: pattern clusters' semantic meaning is still is up to the human judgement.

V - Inter-AP behaviour analysis

Till now, we have been focusing on the evolution of the number of connections through time and the average connection time for each location to infer the type of activities performed in a given location. In the following section, we will focus on the service provided in terms of accessibility. To do so, we analyse the trips realized by the users during Wednesday, 04-02-2015 and want to explore the possibilities offered by wireless connection data concerning time analysis.

1 - Visualization of inter-AP trips

Trips in the set of connection data are associated to transitions for a user from one access point to another. While the generation of the set of data is not complicated, the definition and computation of travel time is a challenge that will be discussed in section V-3. We are first interested in the distribution of the trip origins and destinations within the MB building access point subset.

Figure 6 shows the transition distributions observed on Wednesday the 4th of February, while removing trips with the same origin and destination. The access point IDs allowed us to order them by floor resulting in the pattern observed in Figure 6.B. Here, we observe the presence of square blocks located on the diagonal of the graph corresponding to the inner floor trips. Parallel to those trips, we can observe vertical and horizontal lines corresponding to alternative paths to all floors successively which are probably elevators. The Figure 6.B is an interesting visualization tool to analyze the main path through the facility if (as in our case) we are forced to work on a symbolic space model (model without geometric localization of the objects).

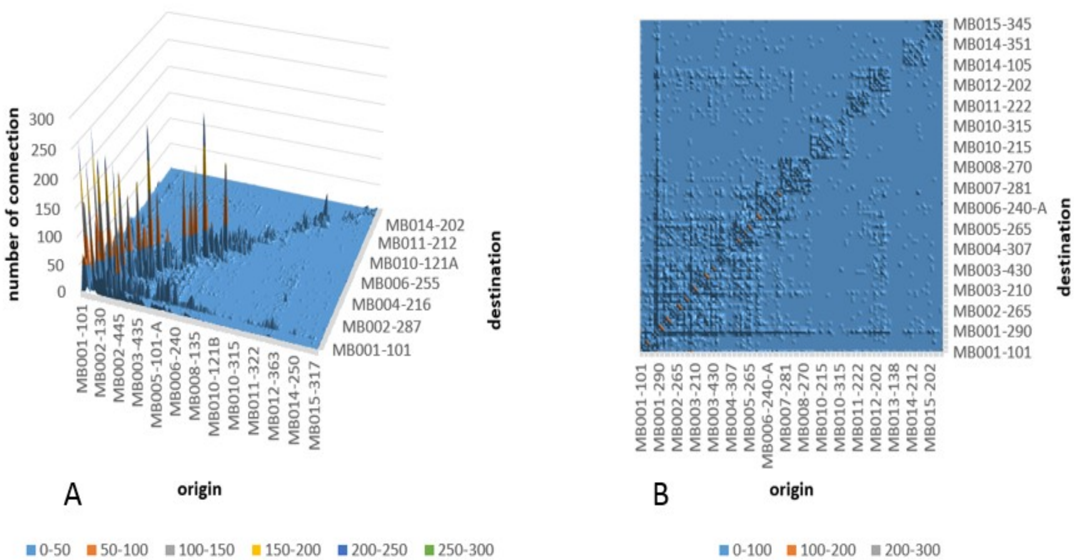

Fig. 6. Visualization of the distribution of transitions

2 - Activity transition

After considering the level of connection of the different places, it's interesting to associate the concept of activity to these locations. Indeed, as raised in [14], the trips realized by users are more influenced by the type of activities performed at the destination rather than by its spatial localization.

Table 2 presents the distribution of transitions between activities types of locations based on the clusters defined in section II. Such data is interesting to understand the global behavior of users through the network: it could help in the design of locations in the network.

We first observe that classrooms always represent at least 50% of destinations, and sharing places plus corridors 30%: they represent the largest proportion users' trips. It needs to be recognized that in this subset of the data, we have only considered internal. As a result, the first trip of user when they enter the building is not taken into consideration. Including it would have considerably increased the proportion of corridors as a destination.

Then the high number of transitions starting and finishing in auditoriums or classrooms in table 2 could seem surprising, but it could be explain by the elementary character of the trips represented here. Indeed, these transitions are the basic blocks for constructing trip chains which are more representative of the behavioural character of trips. In this context, we could speculate that classrooms/auditorium access points are grouped in the same area creating intermediate destination transitions.

Table 2. Distribution of activity transitions

Transitions O/D (%)	Sharing places	Corridors	Laboratories	Offices	Auditorium	Classrooms
Sharing places	21.53	17.67	3.55	2.36	40.60	14.24
Corridors	16.21	24.08	4.55	1.38	32.84	20.91
Laboratories	15.25	13.42	6.68	2.25	50.74	11.67
Offices	22.51	13.16	8.70	1.91	28.66	25.05
Auditorium	19.09	15.68	6.68	1.24	46.36	10.92
Classrooms	15.99	24.25	3.31	2.20	26.56	27.69

Other observation could be made linking these data with the real world context, however the point is here to show that connections data enable observations to be made about user behavior within the network.

3 - Travel time and limitations

In this last section, we discuss the temporal aspect of trips which appears to be one of the main limitations of this connections data set. Indeed, while the use of signal strength to localize device allows us to follow the user through the network and to localize trips in the time dimension, pure connection data do not seem well suited to do this.

The main reason comes from symbolic space used to locate the user: access points are associated to areas to which the user could be connected. A trip in such a space corresponds to a transition from one area to another. Working with data with dense AP networks, which is our case, becomes a burden because time transitions between APs are mostly instantaneous, as shown in Figure 7. In this figure we see that in the vast majority of transitions between APs, transition time is 0. As a result, trips between distant access points representing totally distinct portions of space are decomposed into successive elementary trips between close access points, thus making total travel time the sum of a series of instantaneous transitions, and thereby null.

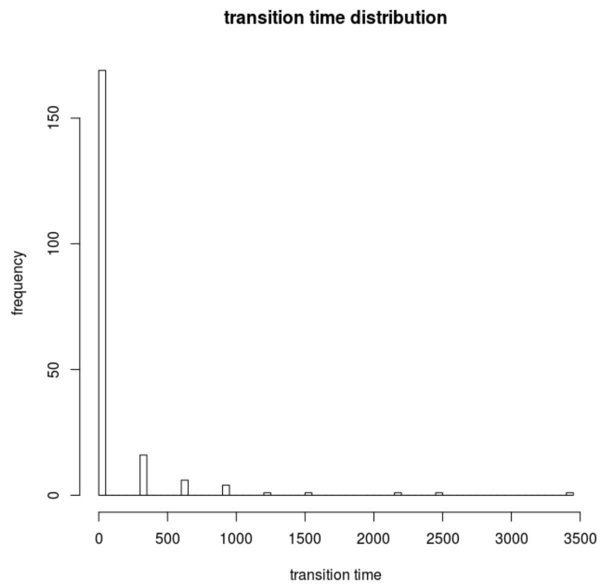

Fig. 7. Distribution of the transition time between to access points in the 1st floor

By focusing on the user behavior, we reach one of the limits of the network connection analysis. Indeed, if strength of these data lay in their potential to continuously furnish global information about the whole population of entire facilities as shown section II, they appear to be weak in characterizing precisely individual mobility behaviours if not enhanced with localization data such as signal strength.

VI - Conclusions

We extensively explored the possibilities offered independently by wireless connection data. In the first part, we studied number of connections to the access points through time and average connection number, to infer the main activity associated to the location. We obtained 6 clusters of activity each one associated to a practical situation derived from the university context. We were then able to visualize the use of the building, based on those activities. In the second part, we computed the trips from connections data and presented graphical object to analyse the service of the facility in term of connectivity. We finally discussed the limitation created by the binary nature of the connection information and the additional data required to overpass the limitation. It appears that connections data allow to characterize the demand from users in a facility describing the use of its different locations. However, the quantification of the service provided, in term of accessibility seems to need additional data.

The results obtained through the treatment of wireless connection data could appear as limited. Indeed, they can be found by the knowledge of the environment or users surveys. However, the exhaustive knowledge of the possibilities and limitations of a set of data is essential in the perspective of the Big Data treatment. Indeed, the development of the data collecting tools and processes open new perspectives in term of data fusion. The anonymity, essential for privacy protection, make that merging process complex and demanding in term of work. In that context, the knowledge of the data's available characteristics and their potential way to fit together is the direction to improve our capacity to analyse mobility through a new perspective.

Acknowledgements

This research would not have been possible without the help of a number of people and institutions. First, the research FR-QNT. Mike Babin, Myung Hong and Lichuan Yang of IITS. Special thanks to Adam Taka, a Ph.D. student at Polytechnique for developing the MAC anonymization script for the preprocessing of WiFi network logs. This research has been funded by FRQNT-New Researcher.

References

[1] Danalet, A., Farooq, B., and Bierlaire, M. (2014) A Bayesian Approach to Detect Pedestrian Destination-Sequences from WiFi Signatures. Transportation Research Part C: Emerging Technologies 44: 146-170. doi:10.1016/j.trc.2014.03.015

[2] Naini, F. M., Dousse, O., Thiran, P. and Vetterli, M. (2011). Population size estimation using a few individuals as agents, 2011 IEEE International Symposium on Information 45 Theory Proceedings, IEEE, pp. 2499–2503.

[3] Boldrini, C., Conti, M., Passarella, A. (2008) User-Centric Mobility Models for Opportunistic Networking, Bio-Inspired Computing and Communication pp 255-267

[4] Su, J., Chin, A., Popivanova, A., Goel, A., de Lara, E.: User Mobility for Opportunistic Ad-Hoc Networking. In: Proc. of IEEE WMCSA (2004)

[5] Su, J., Goel, A., de Lara, E.: An Empirical Evaluation of the Student-Net Delay Tolerant Network. In: Proc. of Mobiquitous (2006)

[6] Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira, J. and Ratti, C. (2013). Understanding individual mobility patterns from urban sensing data : A mobile phone trace example, Transportation Research Part C 26(0) : 301–313.

[7] Yoon, J., Noble, B. D. and Liu, M. (2006). Building realistic mobility models from coarse-grained traces, in Proc. MobiSys, ACM Press, pp. 936–5983.

[8] Aschenbruck, N., Munjal, A. and Camp, T. (2011). Trace-based mobility modeling for multi-hop wireless networks, Computer Communications 34(6) : 704–714.

[9] Conti, M. and Giordano, S. (2007). Multihop Ad Hoc Networking : The Theory, IEEE Communications Magazine 45(4) : 78–86.

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

[11] Bettstetter, C., (2001) Mobility modeling in wireless networks: categorization, smooth movement, and border effects, ACM SIGMOBILE Mobile Computing and Communications Review p 55-61

[12] Thor S. Prentow a, ∗ , Antonio J. Ruiz-Ruiz b , Henrik Blunck a , Allan Stisen a , Mikkel B. Kjærgaard a (2015) Spatio-temporal facility utilization analysis from exhaustive WiFi monitoring, Pervasive and Mobile Computing 16 (2015) 305–316

[13] Meneses, F. and Moreira, A. (2012). Large scale movement analysis from WiFi based location data, 2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN), IEEE, pp. 1–9.

[14] Kang, J. H., Welbourne, W., Stewart, B. and Borriello, G. (2004). Extracting places from traces of locations, Proceedings of the 2nd ACM international workshop on Wireless mobile applications and services on WLAN hotspots - WMASH ' ACM Press, New York, New York, USA, p. 110.

[15] Calabrese, F., Reades, J. and Ratti, C. (2010). Eigenplaces : Segmenting Space through Digital Signatures, IEEE Pervasive Computing 9(1): 78–84.

[16] Farooq, B., Bierlaire, M., Hurtubia, R., and Flötteröd, G. (2013). Simulation Based Population Synthesis. Transportation Research Part B: Methodological 58: 243-263.