2012 International Conference on Indoor Positioning and Indoor Navigation, 13-15th November 2012

Large scale movement analysis from WiFi based location data

Filipe Meneses, Adriano Moreira
Mobile and Ubiquitous Systems research group
Centro Algoritmi, University of Minho
Guimarães, Portugal
{filipe.meneses, adriano.moreira}@algoritmi.uminho.pt

Abstract **— Understanding and modeling the way humans move in urban contexts is beneficial for many applications. The recent advances on positioning technologies, namely those based on the ubiquity of wireless networks, is facilitating the observation of people for human motion analysis. In this paper we present the result of a large scale work conducted to study the human mobility in a University's campuses. The study was conducted along several months, using data collected from thousands of users that freely moved inside the numerous buildings existent in two University campuses and a few other buildings in the city center. A Wi-Fi infrastructure of more than 550 access points provides Internet access to the academic community. We tracked the user movements by logging the devices connected to each access point. Based on that data, an analysis process that highlights the relationships between space features and human motion has been developed. In this paper we introduce the concepts of "place connectivity" and "flow across a boundary" to model these relationships. Results show the mobility patterns detected, which are the attraction places along the day, and what places are more strongly connected. This paper also includes an analysis of the short and long term movements between places. With this study we extended our understanding of the life in the campus, enabling us to feel the campus "pulse".**

Keywords - Human motion, WiFi networks, tracking, movement patterns.

I. INTRODUCTION

The study and understanding of the human motion is very important in numerous activities: it can be used in urban planning activities, to plan new roads based on the traffic, and to predict the spread of virus and diseases, just to name a few applications. To understand the human motion in large scale spaces, several studies were conducted based on data collected using GPS receivers, data from the usage of GSM networks, or data from other sources (e.g. money bills circulation) [1-6]. The work performed so far revealed that the spatio-temporal behavior of humans is a lot more predictable than we would expect (and would like it to be).

Studying the mobility of people in large buildings or groups of buildings is also very important, and raises new challenges on how to observe the human motion: how to collect the people's location inside buildings, how to observe a large number of persons, how to do it in the entire building (large spatial contexts), and how to collect data for long time periods (weeks to years). Collecting data for long time periods is fundamental to detect changes in the human motion behavior, as shown in [7], and to cope with seasonal events. Observing people on their daily lives is, therefore, a fundamental step on human mobility analysis. The choice of the positioning/location technology has a significant impact on the mobility analysis [3].

A. *Observing people for motion analysis*

There are numerous technologies being used to locate people and objects inside buildings, with different degrees of accuracy and precision. However some of the technologies cannot be easily deployed in large facilities or be used by a large number of users. One potential solution is to exploit the ubiquity of wireless networks.

In our university, access to the Internet is provided through a large scale Wi-Fi network, deployed across the several universities' facilities and covering all the indoor spaces and some of the outdoors space around buildings. This network is also integrated into a European-wide network of Wi-Fi infrastructures used in universities and research centers across Europe – the eduroam network [8]. A roaming protocol between all the participating institutions allows users from one university, students or staff members, to use the infrastructures in all other institutions without any costs. Access to the network is controlled through an authentication process, with the RADIUS protocol being part of the process. Every time one user accesses the network, the RADIUS protocol logs that event (Start event). A similar log is done whenever a connected device moves from one Wi-Fi Access Point (AP) to another (re-association; sequence of Stop and Start events), or when the users disconnect from the network (Stop event). Therefore, RADIUS logs provide a convenient proxy for observing the motion of people by observing the usage of the Wi-Fi network by their devices. Since the members of the academic community are using more and more mobile devices, portable computers and, increasingly, smartphones with Wi-Fi network interfaces, tracking the use of these devices ends up to be a good way of tracking the mobility of people.

We started to collect data about the use of our Wi-Fi network back in 2005, when the eduroam network was still in its infancy. Initially, we developed and deployed a network monitoring applications that, by resorting to the Simple Network Management Protocol (SNMP), collected data at regular intervals (5 minutes) about the devices associated with each one of the APs. Later we started to collect the RADIUS log files as they provide a more reliable and detailed view of the network usage. Moreover, the RADIUS logs also include records about users associated with APs while roaming in other institutions.

The data provided by the RADIUS service enables a very detailed view about the mobility of people over the space covered by the Wi-Fi infrastructure. There are, though, two major limitations in this approach: spatial extent and spatial resolution. The spatial extent limitation refers to the limited observation space – as users move away from the spaces served by the eduroam network, we can no longer track their movements. This limitation is being tackled on a current project using a collaborative sensing approach [9]. The problem with the spatial resolution derives from the fact that, through the RADIUS logs, we can only detect the position of devices (people) at the Basic Service Set (BSS) level. However, as the size of each Wi-Fi cell (BSS) is just a few tens of meters wide, the obtained spatial resolution is adequate to analyze the motion at a building level. There is a third potential problem in using the RADIUS logs as a proxy for observing human motion: each record on the log refers to a BSS by the MAC address of the AP, but do not include the physical location of that AP (coordinates). Therefore, additional information is required about the physical position of each AP, to enable analysis processes based on the Euclidean distance. In our university, a large number of APs is geo-referenced using latitude/longitude coordinates. However, no data is available about the APs of other institutions. Mapping the physical location of APs automatically, from data collected by the users, is a workaround for this problem that we are developing on an ongoing project [10]. An alternative, used in the work reported in this paper, is to make use of the names each network administrator assigns to each AP for management purposes. In our university, these names follow a hierarchical structure, starting by the institution name, then the city name, then the campus or zone, followed by the department or building name and, finally, the floor and AP name (e.g. `UMINHO-BRG-rstd-2a`). By using these names we have been able to perform mobility analysis at several different spatial scales, from institution level to AP level.

There are numerous techniques that can be used to acquire the user' position in indoor environments. In this paper we present the results achieved with a basic positioning technique for Wi-Fi networks. Our aim is not to improve a positioning technique directly but to use the positioning information to conduct a movement analysis process producing knowledge in terms of movement patterns. Later, this can eventually be integrated into the existing positioning techniques to provide higher positioning accuracy by taking advantage of the knowledge of users' movement patterns.

Our goal with this analysis is to go one step further on human mobility analysis by looking for relationships between space features and the way people move. In this analysis, we introduce the concepts of "place connectivity" and "flow across a boundary" to model these relationships. These concepts are described in the next section. In section III we present the dataset used in the study. Section IV discusses the obtained results, presenting a place connectivity analysis done at different levels and a time profile of the inbound and outbound flows. In Section V, we revise related work and highlight the new findings emerging from our work. Finally, in Section VI, we present our concluding remarks.

II. HUMAN MOVEMENT ANALYSIS

In [11], the authors proposed a set of concepts for the process of human mobility analysis, in order to build a homogeneous and generic reference model. In this paper we adopt those concepts.

The concept of *Observation* represents the presence of an artifact in a specific point of a spatio-temporal space, and is described as:

(*Id_Observation, Artifact, Location, Timestamp*)

where the *Location* can be represented in a geometric or symbolic space model. In our case, *Observations* are derived directly from the RADIUS logs as we can observe when a device enters (Start event) or leaves (Stop event) the eduroam network. In every observation, the *Location* is represented by the reference to an AP (MAC address and name). The *Artifact* is represented by the MAC address of the device connecting to the Wi-Fi network. As described in [11], an observation might include optional attributes. In our case, adding the type of event, Start or Stop, as an optional attribute makes it easy to transform *Observations* into *Stays* and *Space Leaps*.

The concept of *Stay* represents the permanence of an artifact on a single *Place* for a certain amount of time, and is described as:

(*Id_Stay, Artifact, Place, Timestamp_Initial, Timestamp_Final*)

where a *Place* represents the aggregation of two or more nearby *Locations*. In our case, a *Stay* might represent the permanence of a person while associated to an AP for some time, or the presence of a person within a region covered by a set of nearby APs. In [11], the authors distinguish *Stays* from *Time Leaps*, with *Time Leaps* representing the sequence of two *Observations* on the same *Place*, but with a time span between them that is too long for the analyst to assume that the artifact did not leave that *Place* during that time span. In our case, Start and Stop event clearly define the arrival and departure from a place, and provide evidence that the device did not leave the *Place* between the Initial and Final Timestamps even for long time spans. In the RADIUS terminology, *Stays* are the synonymous of Sessions.

The concepts of *Space Leap* and *Elementary Movement* represent the change in *Location* (moving state) of an *Artifact* over time. Both have similar representations but with different semantics: An *Elementary Movement* occurs when the time span is short enough for the analyst to assume, with reasonable confidence, that the *Artifact* moved along a straight line connecting the two *Positions*. Otherwise the movement is described as a *Space Leap*, and is described as:

(*Id_SpaceLeap, Artifact, Location _Start, Location _End, Timestamp_Initial, Timestamp_Final*)

The spatial reference model represents an important role in the understanding of the people movements. The two main types of spatial reference models are the geometric and the symbolic ones. In case of a geographic space model, like the WGS-84 datum, in some situations it is easier than in others to distinguish between an Elementary Movement and a Space Leap. Although one can easily calculate the geometric distance between two points, it is not as easy to calculate the time that is necessary to go from one point to another. For example, to travel between two points different, several alternative means of transportation can be used (car, bicycle or on foot), each one having different average travel speeds. Similarly, two places can be geographically very near but the time to travel between them can be high if, for example, those places are separated by a river.

On a symbolic spatial reference model to distinguish an Elementary Movement from a Space Leap is, in most cases, even more difficult. The lack of universal space model makes it nearly impossible to implement a universal solution. Postal codes, wireless networks cell-id, postal addresses are some examples of symbolic space models. The diversity of symbolic space models and the personal interpretation of some of those models make it impossible to achieve a universal solution.

In our case one should represent the movement as a set of *Space Leaps* as Locations are described in a symbolic space model – the set of APs' names and MAC addresses. In some cases, where the time span is very short, in the order of a few seconds, the movements could be described as *Elementary Movements* but, due to radio signal fluctuations, these movements might be apparent and not the real movement of a device from one position to another. For this reason, we refer to all the movements as *Space Leaps*. Extracting *Space Leaps* from our *Observations* is straightforward, since the sequence of a Stop event and a Start event, on a different AP, represent a change in *Position*.

Space features affect the human motion. The mobility of humans is also conditioned by a number of variables, such as the weather, the accessibility, the mean of transportation, and the location. Most people perform regular movements between their home and work place. The path used on every day movements is conditioned by the existing roads, the available buses or metropolitan lines, etc. Many real constraints like a car accident or a traffic jam can make the users change their usual path or take much more time to travel between two places. Many constraints also apply to short distance movements. For example, bad weather can make people to walk through a certain path to avoid rain.

On the other hand, the presence/movement of people also shapes the space. It is what many authors describe as the "carpet effect" – the constant movement of person along the same path can be detected by the wear out of the carpets. If a large number of persons travel directly between two places then we can say that those places are symbolically near or highly connected. We define the concept of *Place Connectivity* as a metric to measure the degree of linkage or connectivity existent between two places. *Place Connectivity* is computed as the number of *Space Leaps* observed between two *Places* within a given observation time period, as is described as:

(*Id_PlaceConnection, Place_Origin, Place_Destination, Number_Of_SpaceLeaps*)

As a directed metric, the *Place Connectivity* between two *Places* might or not be equal in both directions.

As shown in the following sections, some *Places* are connected with a lot of other Places, while others are not. We define the concept of *Hub Level* to refer to the number of connections between distinct places, that is, the number of places for which the *Place Connectivity* is larger than 1, and describe it as:

(*Id_HubLevel, Place, Number_Of_Connections*)

A *Place* is an abstract concept that can be used to refer to a small area, like an office inside a building, or to a wider area like a harbor or a district in a city. The flow of people across the boundary around a *Place* is a measure of the number of persons that enters and leaves a place. We define the concepts of *Inbound Flow* and *Outbound Flow* of a *Place* as the number of *Space Leaps* from external *Places* to and from a given Place, within a given observation time period, and describe it as:

(*Id_Flow, Place, Number_SpaceLeaps*)

As for the *Place Connectivity*, given a certain *Place*, the *Inbound* and *Outbound Flows* might not be equal. Even when similar, their time profiles might be quite different, as shown in the following sections.

III. Data Analysis

Each institution participating in the eduroam network runs a RADIUS server that is used to authenticate the users. In case of roaming, the RADIUS server contacts the user' home institution to verify the user credentials (login and password).

The University of Minho is spatially organized in two campuses and a few other buildings located in two different cities 20 km apart. The Wi-Fi network contains around 550 Access Points that provide complete indoor coverage in all the facilities.

A. The dataset

The dataset used for the mobility analysis presented in this paper was obtained from the RADIUS log file for the month of March 2011. Each record in the RADIUS log includes a large number of parameters. For this study we extracted the following parameters:

- timestamp;
- *Acct_Session_Id*: (sessionID);
- *Calling_Station_Id*: (client device's MAC address);
- *Called_Station_Id*: (Access Point's MAC address);
- *Acct_Status_Type*: (event type, "Start" or "Stop");
- *WISPr_Location_Name*: (the name of the AP).

This file includes around 1.437.504 records, about half of them being "Start" records and the other half being "Stop" records. These records refer to 1.138 different Access Points and 5.989 different client devices. The number of detected Access Points clearly exceeds the number of Access Points installed in our facilities, meaning that about half of them were used by home users while in roaming in other institutions.

The number of *Observations* per device varies significantly, with an average number of 240 *Observations* per device. Some devices have one single observation: a logout from a session previously started or the login on a session that was still running when the log file was generated. A few devices have a high number of *Observations*.

In case of roaming, the RADIUS server of the visited institution exchanges data with the corresponding server at user's home institution. Many institutions do not include the optional fields in the data exchanges between the RADIUS servers, thus, for many roaming records information like the name of access point is not available.

Observations can be converted to *Stays* by linking the corresponding Start and Stop events. Processing the Observations by chronological order we have been able to detect a total of 715.169 *Stays*. The large majority of these *Stays* (709.829, corresponding to 99,25% of all *Stays*) correspond to Wi-Fi sessions that took place within our institution network.

The *Space Leaps* can also be extracted directly from the *Observations* or, alternatively, from the *Stays*. In order to perform motion analysis at several different scales, we resorted to the hierarchical structure of the APs' names. These names are strings in the format University-city-building-department-APnumber. This hierarchical structure allows us to know the location of an access point up to department level. Some departments occupy an entire floor of a building or span across more than one floor. In a few cases, one department occupies rooms in different buildings, or even in different campuses.

B. Space Leaps

A sequence of a Stop and a Start RADIUS event for the same client device represents a *Space Leap* from one AP to another one. A total of 690.154 *Space Leaps* have been detected.

For various reasons, including operating systems faults and network load, sometimes the association between a client device and an Access Point fails and have to be reestablished, creating a new RADIUS session. If a device restarts the association using the same Access Point that it was using just a few seconds ago, then it is not a movement but clearly the reestablishment of a previous session.

The coverage area of most of the Access Points overlaps that of other Access Points, in order to provide a complete spatial coverage and enough network capacity. Quite often, due to signal level fluctuations, a device stops using an Access Point and a few seconds later starts to use another one. Clearly, these re-associations do not represent movement but just the reestablishment of an existing network session (now in another access point). This is known as the "*ping-pong*" effect.

We cleaned all these cases where the leap time is shorter than 10 seconds. We also filtered out all the *Space Leaps* occurring between similar APs in less than 60 seconds as these are not also representative of a device's movement. After filtering, we ended up with a total of 155.974 Space Leaps.

Each *Space Leap* is characterized by a departure time, an arrival time, and a time duration (time span). Figure 1 shows the distribution of the time span of all the Space Leaps.

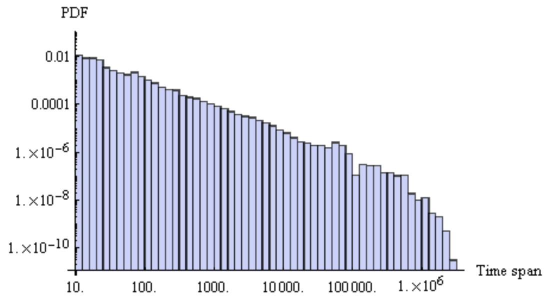

Figure 1. Probability Distribution Function of the time span of Space Leaps (in seconds).

Short duration *Space Leaps* are a lot more frequent than long duration *Space Leaps* (note the log scale in both axes). The graph in Figure 1 also shows that the time span of *Space Leaps* approximately follows a power law. In previous work described in [6] and [12], the authors use the concept of "jump size" as the distance in space, measured in meters or kilometers, between the two places where the user is located consecutively. Our results are based on time but follow a similar distribution to the results achieved by [6] using spatial distance.

C. Hierarchical Space Model

Place Connections between *Places*, the *Hub Level* of each *Place*, and the *Inbound* and *Outbound Flows* for each *Place* were computed at the several levels of the hierarchical space model.

For the APs from our university, 5 different levels were considered: institution; city, campus, department; and AP. For the other institutions only one level was considered (name "outside"), since the naming space used for the APs in other institutions is not uniform and, in many cases, the names of the APs is even missing in the RADIUS logs.

Computing the *Place Connectivity* between *Places*, as well as the other metrics, is a simple task. However, the large number of records demands, in some cases, a long computing period and impose some limitations on the visualization of the results. The next section describes some of the main results.

IV. RESULTS

In this section we describe the findings that emerged from the analysis of a dataset extracted from a RADIUS log file.

A. Place Connectivity

The analysis of the *Space Leaps* at the top level of the hierarchy (institutions) shows that most of the displacements occur within our university (99.8% of the total *Space Leaps*).

From those that are to or from other institutions, 54% are outbound displacements and 46% are inbound.

Looking inside our university, at the second level of the hierarchy, reveals the movement of people between the two cities, and also the displacements to and from other institutions. Figure 2 illustrates these movements as a graph, where nodes represent the *Places* - in this case the two cities, in red, and all other institutions, in green - and edges represent the number of *Space Leaps* between *Places* during the observation period.

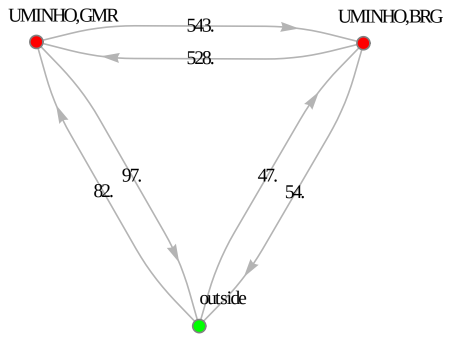

Figure 2. Number of Space Leaps between places, at level 2 of the hierarchy.

The edge weights in Figure 2 represent the number of *Space Leaps* between every two *Places*. Note the high level of symmetry on the number of *Space Leaps* in both directions. Some of these *Space Leaps* are of short duration (up to 3 hours) while others are longer. These longer *Space Leaps* cannot be seen as direct movements between nearby *Places*, as the corresponding persons might have taken longer routes between these *Places* but without using the eduroam network in the meantime. Figure 3 shows the time profile of the *Space Leaps* between the two cities where the university has its facilities. Although not that much frequent, the *Space Leaps* between the two cities occurs during the working hours, and mostly on the afternoon.

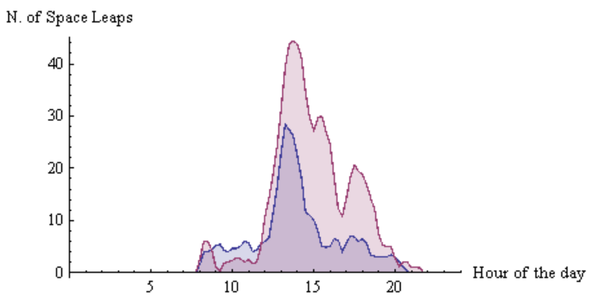

Figure 3. Time profile of the movements between the two major locations in our university.

The third level of the hierarchy provides details about the motion within each city, and from these cities to and from outside institutions. The graph in Figure 4 illustrates the *Place Connections* between the major university locations within each city.

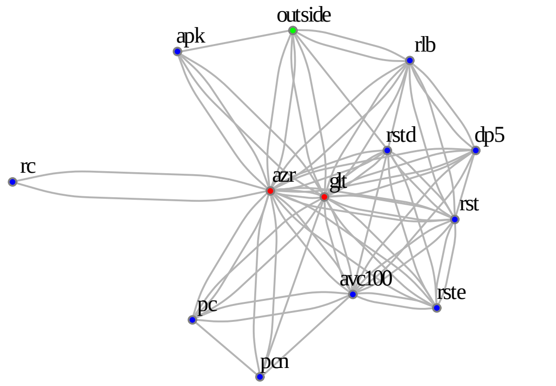

Figure 4. The graph representing the Place Connectivity at level 3 of the hierarchy.

What comes out of Figure 4 is that some *Places* act like central *Hubs* for the movement between the other *Places*. This is the case of the *Places* “azr” and “glt” (in red), that exhibit connections to most of the other *Places*. The *Place* “azr” is connected to all other *Places*, with *Space Leaps* in both directions, in a total of 24 connections. The *Place* “gtl” is also connected to most of the other *Places* (21 connections). These results are not surprising, and those two *Places* represent the two major campuses in the university, one in each city.

Some of the *Places* in Figure 4 are tightly connected, while between other pairs of *Places* the number of *Space Leaps* is much lower. The stronger links are shown in Figure 5.

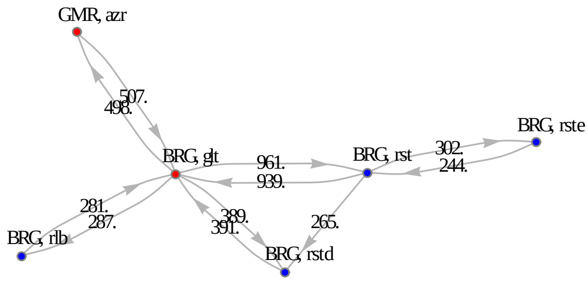

Figure 5. The strongest connections, at level 3 of the hierarchy.

At this level, there are two pairs of places tightly connected: “GMR, azr” ↔ “BRG, glt”, and “BRG, glt” ↔ “BRG, rst”. The first one represents the motion between the two campuses, while the second one represents the motion between the Gualtar (“glt”) campus and one of the students’ residences. Again, these were expected results. However, we were not expecting such a strong connectivity between the two major campuses, as the students studying in one campus do not have to move to the other campus for taking classes or getting involved in other activities. Again, note the high level of symmetry on the number of *Space Leaps* in both directions.

At level 4 of the hierarchy, 124 distinct *Places* are identified, corresponding to the several departments of the university. These 124 *Places* are connected through 3013 connections (unidirectional). Some of these connections are really strong, but most of them as weak connections, as illustrated in Figure 6 where the distribution of the number of Space Leaps per connection is shown.

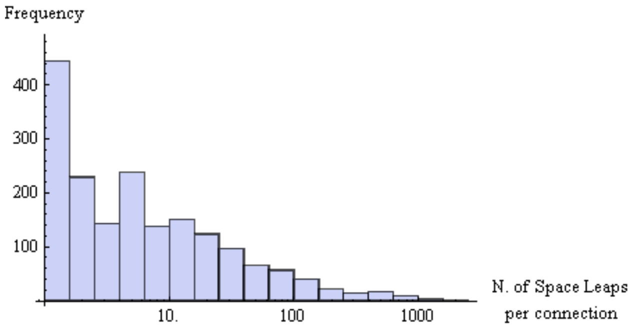

Figure 6. Distribution of the number of Space Leaps per connection, at level 4 of the hierarchy.

The top 5 connections have more than 1.000 *Space Leaps*. These stronger connections are shown in the graph of Figure 7.

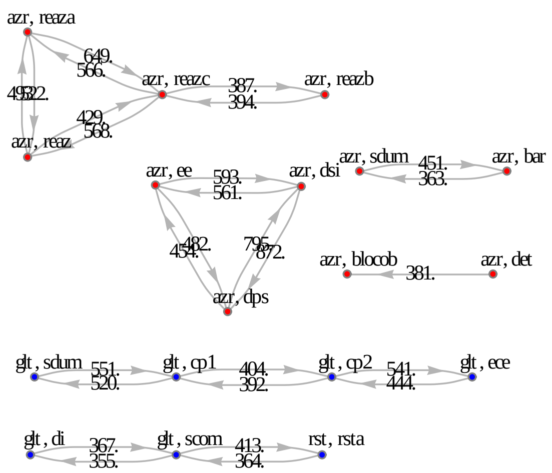

Figure 7. The strongest connections, at level 4 of the hierarchy.

The results in figure 7 show that the strongest connections are between nearby *Places*. In the Azurém campus (*Places* in red), the strongest connections are between *Places* in the same buildings or nearby buildings (e.g, "azr,reaz", "azr,reaza", "azr,reazb", and "azr,reazc", all represent *Places* in the students' residences). There are, also, strong connections between *Places* that are not near each other, such as the connections between *Places* "glt,di" and "glt,scom", revealing affinities between these *Places* that were, otherwise, not obvious. Some of these strongly connected *Places* are also central hubs in the motion within the campuses. Figure 8 shows the top 6 *Places* (in red) in number of connections with other *Places*.

Five of these hubs (*Places*) are in the Gualtar campus, and the other one is in the Azurém campus. The Gualtar campus is larger than the Azurém campus and, hosts many different schools, while the Azurém campus host only two schools. Therefore, people in the Gualtar campus concentrate around many central *Places*, while in Azurém people concentrate more around one single *Place*. These *Places*, though, are not related to the several schools but, instead, related to the major services: the library ("sdum") is top in each one of the campuses, and "cp1" and "cp2" are the major classroom buildings. In the Azurém campus, the classrooms are not so concentrated in a few buildings but, instead, spread throughout most of the buildings. As a consequence, we do not observe the emergence of these central hubs as in the Gualtar campus.

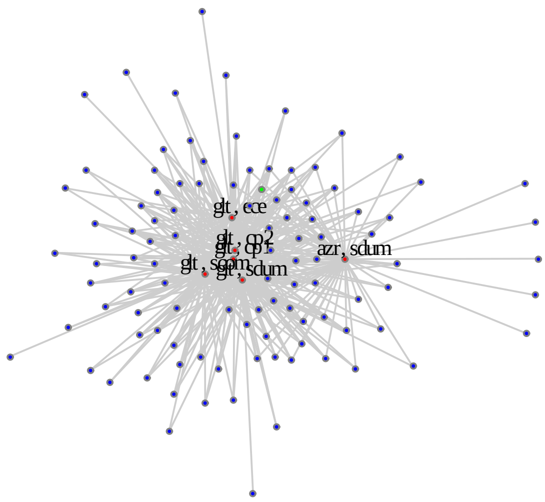

Figure 8. The top 6 places in number of connections with other places, at level 4 of the hierarchy.

There are also preferential places for going from one city to the other one. Figure 9 shows the strongest connections between the two cities.

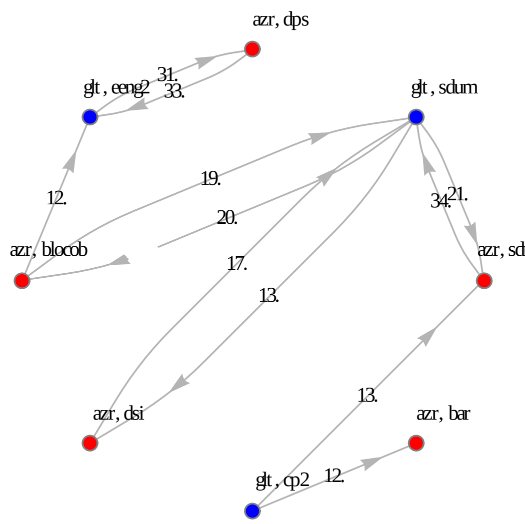

Figure 9. Strongest connections between the two cities, at level 4 of the hierarchy.

The strongest of these connections are between two pairs of *Places*: “azr,sdum” ↔ “glt,sdum”, connecting the libraries in the two campuses, and “azr,dps” ↔ “glt,eeng2”, connecting two *Places* used by the same department in the two campuses.

At level 5 of the hierarchy – the AP level – one can find which areas (APs) in each of the central *Places* are actually the central hubs within the campus. Figure 10 shows the graph of connections for the 6 top places in number of connections with other *Places*.

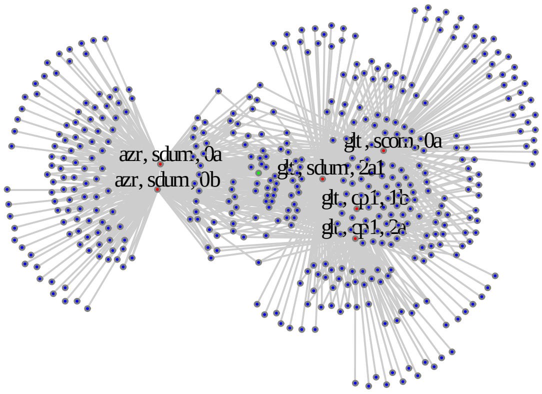

Figure 10. The top 6 places in number of connections with other places, at level 5 of the hierarchy.

What is interesting about the distribution of the number of connections per *Place* is that it is not a monotonic function of the number of connections, as shown in Figure 11. There are just a few *Places* with a lot of connections (the central hubs), but there are also not that much *Places* with just a few connections.

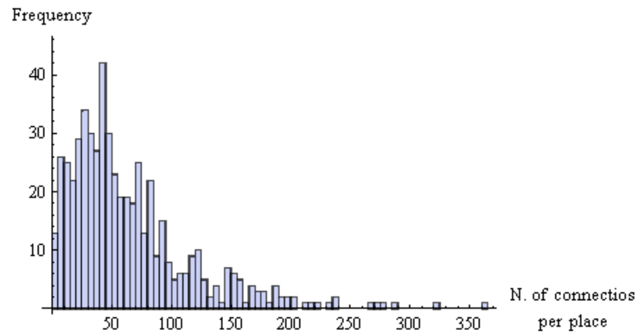

Figure 11. Distribution of the number of connections per place, at level 5 of the hierarchy.

B. *Flows across boundaries*

Figure 12 shows the time profile of the inbound and outbound flows for several places and levels. Figure 12 a) shows the general view of the inbound and outbound in the University (level 1). The number of artifacts increases a lot a few minutes before 9 am, decrease slightly during lunch time and increases again after lunch time. Between 5 pm and 8 pm the number of artifacts decreases significantly. Most of the remaining artifacts leave the University until 1 am. During night the number of artifacts is reduced.

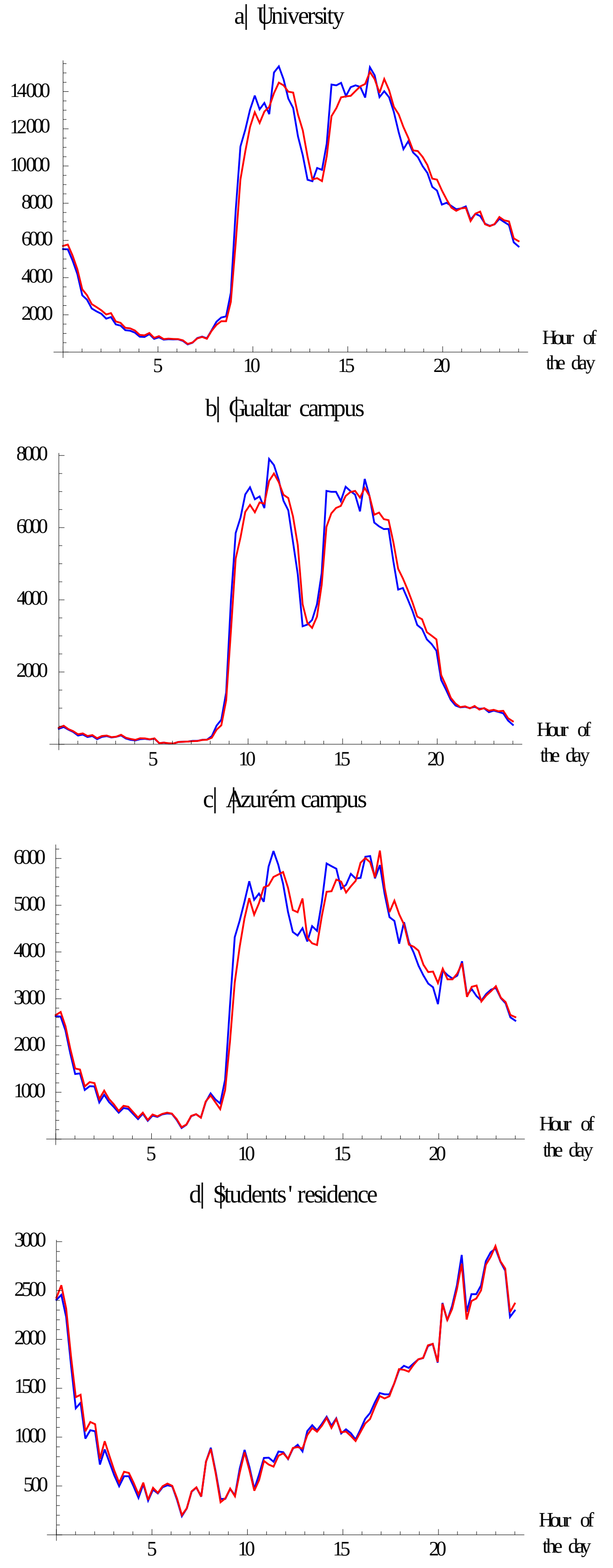

Figure 12. Time profile of the inbound (in blue) and outbound (in red) flows for the a) University, b) Gualtar campus, c) Azurém campus; and d) students’ residences, along the 24 hours of the day.

In figures 12 b) and c) both campuses follow the University general trend, with a higher number of artifacts during the day. The general trend reflects the standard working hours, with an increase in the number of artifacts in begin of the day and after lunch time.

The Gualtar campus is larger than the Azurém campus (more students and more departments) and thus the total number of artifacts is also higher. However, between 7 pm and midnight, the time profile show a higher number of inbound and outbound flows in Azurém campus when compared to the larger campus of Gualtar. This means that Azurém campus has more "life" during and after dinner time. Azurém students' residences profile (in figure 12 d)) is clearly different from the campus. Students start to use the Internet mostly in the afternoon and after dinner until late in the night.

Inbound and outbound flows are very similar in all graphs in figure 12, which denotes a high degree of mobility.

V. RELATED WORK

Modelling human motion has attracted the attention of the scientific community during the past few years [13, 14]. Significant progresses have been achieved, namely through the work described in [6] and [1]. Both these works identified fundamental, and eventually universal, characteristics of the human motion, and proposed mathematical models to describe it. Although these two works contributed significantly to the understanding of how an individual person moves, they also proved the importance of the data (about human motion) that must be available to assess the validity of the proposed models and to identify other universal characteristics.

In [6], the authors used a data set containing positioning records of around 100 000 users of a cellular network collected over a period of six months. Since mobile phones are personal devices, the trajectory of a mobile phone is highly correlated to that of his owner. Cellular networks are, therefore, a good proxy to observe the trajectories of humans. However, the majority of the data records in the above referenced data set correspond to position data collected only when a person initiates/receives a phone call or a SMS message. Therefore, the collected phone trajectories may represent the sum of a number of individual human trajectories. Another limitation of data collected through mobile cellular networks is accuracy. Since positioning is based on CellID, this data cannot be used to validate models at short scales.

In [13, 14], Ahas and his team used data collected by one cellular network operator to study the seasonal characteristics of tourism in Estonia. At a campus scale, the iSPOTS project [7] used data collected by a campus wide WiFi network to study the usage patterns of spaces and to assess the impact that the deployment of the WiFi network had on the academic life. We go a step further and use Wi-Fi log data to study the mobility between the spaces.

In other projects, data about the trajectories of people in urban environments have been explicitly collected through the use of GPS. An example is the Spatial Metro project [15], where GPS trajectories being used to help urban planners and architects to better adapt several city centres to pedestrians.

One common factor in all these projects is the importance of data and how it is collected. We have built a large infrastructure for data collection that provided access to data about indoor mobility in university context, at a scale much better than that provided by cellular networks. One advantage of our approach is the possibility to collect data at high resolution about human motion and space usage patterns in real environments.

The achieved results can be used by a wide range of different applications and scenarios. It can be used to rearrange a space, considering the usage pattern of that place, or to improve the physical safety and security of persons that use an area. Results can also be used to plan and increase the WiFi network capacity, to improve transportations networks, or simply to understand the need to increase flexibility in the opening and closing hours of the facilities in the campus.

VI. CONCLUSIONS

We have used Eduroam Wi-Fi network as a large sensor to observe people in the campuses. We processed the RADIUS log file converting *Observation* into *Space Leaps* and studied the artifacts movements in order to understand life in the campuses and gain knowledge about the relationships existent between the different places. Eduroam network can only be used by authenticated users, which ensures the completeness of the used dataset.

The coverage area of an access point can span across more than one floor or departments. This limits the fine grain of the analysis but it is enough to do a complete study at building level. The place connectivity analysis, done at different levels, allowed determine the "central" places, finding the places that are tightly connected to other locations. We found a strong connectivity between the two campuses and in a more detailed level results show a strong connectivity between places located in the same building or nearby buildings. Time profiles analysis, done at three different levels, show that inbound and outbound flows are very similar. However, the temporal profiles of the two campuses are slightly different.

Using Eduroam Wi-Fi network it is possible to study human mobility at different levels and look for relationships between space and way people move. In future, we expect to extend this work integrating data collected in other Eduroam participant institutions.

REFERENCES

[1] Brockmann, D., Theis, F.: Money Circulation, Trackable Items, and the Emergence of Universal Human Mobility Patterns. IEEE Pervasive Computing. 7, 28-35 (2008).

[2] Reades, J., Calabrese, F., Sevtsuk, A., Ratti, C.: Cellular census: Explorations in urban data collection. IEEE Pervasive Computing. 6, 30-38 (2007).

[3] Moreira, A., Santos, M., Wachowicz, M., Orellana, D.: The impact of data quality in the context of pedestrian movement analysis. In: Painho, M., Santos, M., and Pundt, H. (eds.) Geospatial Thinking. pp. 61-78. Springer-Verlag (2010).

[4] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. Nature, 439(7075):462–465, January 2006.

[5] K. Farrahi and D. Gatica-Perez. Learning and predicting multimodal daily life patterns from cell phones. In Proceedings of the 2009

international conference on Multimodal interfaces, pages 277–280, 2009.

[6] Marta C. Gonzalez, Cesar A. Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. Nature, 453(7196):779–782, June 2008.

[7] Sevtsuk, A., Ratti, C.: iSPOTS. How Wireless Technology is Changing Life on the MIT Campus. Proceedings of the 9th International Conference on Computers in Urban Planning and Urban Management CUPUM 2005 (2005).

[8] Eduroam website, available at http://www.eduroam.org/. 2012.

[9] Helena Rodrigues, Maria João Nicolau, Rui João José, Adriano Moreira, “Engaging participants for collaborative sensing of human mobility” UbiComp’12, September 5-8, 2012, Pittsburgh, USA

[10] Carlos Pérez-Penichet, Ângelo Conde, Adriano Moreira, “Human mobility analysis by collaborative radio landscape observation”, In. Jérôme Gensel, Didier Josselin and Danny Vandenbroucke (eds.), “Multidisciplinary Research on Geographical Information in Europe and Beyond”, Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, ISBN: 978-90-816960-0-5, pp. 153-158, 2012.

[11] João Peixoto, Adriano Moreira, “Dealing with Multiple Source Spatio-temporal Data in Urban Dynamics Analysis” in B. Murgante et al. (Eds.): ICCSA 2012, Part II, LNCS 7334, pp. 450–465, 2012.

[12] Karamshuk, D.; Boldrini, C.; Conti, M.; Passarella, A.; , "Human mobility models for opportunistic networks," Communications Magazine, IEEE , vol.49, no.12, pp.157-165, December 2011 doi: 10.1109/MCOM.2011.6094021

[13] Rein Ahas et al, “Seasonal tourism spaces in Estonia: case study with mobile positioning data”, Tourism Management 28(3), 2006, pp. 898-910

[14] Rein Ahas et al, “Mobile Positioning in Space – Time Behaviour Studies: Social Positioning Method Experiments in Estonia, Cartography and Geographic Information Science, Vol.34, No. 4, 2007, pp- 259-273

[15] Stefan van der Spek, “ Mapping Pedestrian Movement: Using Tracking Technologies in Koblenz”, Lecture Notes in Geoinformation and Cartography, 2008, pp. 95-118

ACKNOWLEDGMENT

This work was supported by the FEDER program through the COMPETE and the Portuguese Science and Technology Foundation (FCT), within the context of projects SUM – Sensing and Understanding human Motion dynamics (reference PTDC/EIA-EIA/113933/2009) and TICE.Mobilidade (COMPETE 13843).