

UNDERSTANDING TOURIST HABITS THROUGH THE DATA OBTAINED BY PUBLIC WI-FI NETWORKS

Dr. Bartomeu ALORDA-LADARIA⁽¹⁾, Dr. Vicente RAMOS-MIR⁽²⁾, Sr. Marcos RADO-LORT⁽²⁾,
Dr. Pere A. PALMER-RODRÍGUEZ⁽³⁾, Dr. Ricardo ALBERICH MARTÍ⁽⁴⁾

Smart Destination Working Group, University of the Balearic Islands

(1) Physics Dept. electronic System Group, tomeu.alorda@uib.eu

(2) Applied Economics Dept., Competitiveness, ecology and tourism, vicente.ramos@uib.es

(3) Mathematic Sciences and Informatics Dept. Multimedia information technologies, pere.palmer@uib.es

(4) Mathematic Sciences and Informatics Dept. Computational biology and bioinformatics, r.alberich@uib.es

1. INTRODUCTION

According to the World Tourism Organization, at least 1.184 billion people travelled internationally last year for tourism purposes (UNWTO, 2016). From those, about 68.1 million chose Spain as their vacation destination (Instituto Nacional de Estadística –INE-Frontur, 2016). Consequently, the tourism sector and its related activities combined is considered to account for approximately 11% of the Gross Domestic Product (GDP) of the country (INE-CST, 2016).

In the case of the Balearic Islands Exceltur (2014) estimated that in 2014 tourism represented 45.5% of its GDP, corresponding to 34.8% of Employment. The archipelago received last year 11.65 million visitors (INE-Frontur, 2016). Given its importance, it is not difficult to conclude that exists an interest in studies related with tourism.

Research has traditionally been carried out using complex and costly surveys. For example, 56915 tourists were surveyed in the span of three years to estimate the length of their vacations (Alegre and Pou, 2006).

Nevertheless, by taking advantage of the sensing capabilities of mobile phones, researchers are able to collect an unprecedented amount of fine-grained behavioral data from people (Phithakkitnukoon et al., 2014). Some examples can be found where it has been proved the feasibility of using Global Positioning System (GPS) data loggers to improve the quality or completely replace traditional surveys (Wolf et al., 2001). In fact, the use of specific applications to offer the touristic services available is used to collect behaviour data more efficiently. Obviously, the user must download and install the specific application. Therefore, this method to collect data requires an enough percentage of mobile phones with the specific application to estimate good results.

Furthermore, many studies outlining the new important role-played by the information and communication technologies (ICTs) at all levels of society. This role is independent of social or importance of the business sector in which the studies are focused. In this sense, the tourism sector is no stranger to this trend especially when "Travel and accommodation" is one of the topics most searched by Spanish Internet users in 2013 according to the Spanish National Institute of Statistics (59.4 % of users connecting daily for 37.9% of frequent users).

The deployment grade and the extension of ICT uses, and the capability to cover the needs of tourism expectative from previous phase to post travel phase is reported in (SEGITTUR, 2014). The study concludes indicating that the ICTs have a strong presence for age ranges between 18 and 45 years. The ICTs increase their importance when the travel distances increases and for international destinations. Finally, the Balearic Islands are reported as the tourism destination with a percentage of 50,9% travellers using ICTs in any moment during their stance.

Considering this ICTs trends and the main objective to construct a better tourism destination, this work focuses the research in using existing collection of data in the destination ICTs. Looking at the technical logs generated by the network infrastructure nodes, any event in the network is stored containing a unique device identifier in order to tracking the event. In the case of client devices, the MAC address is used as the unique identifier number and can be used to tracking the device (Chilipirea et al., 2015) throughout the area covered by the network. This capability of users' tracking is especially useful in analysis of the space used by tourists,

because despite the diversity of their travel history, humans follow simple reproducible patterns (González et al., 2008).

In this sense, the objective of this paper is to present and initial analysis of the possibilities of using data available in the public WIFI's network logs in order to provide private and public tourism managerial applications. At this stage, many of this data are stored for short periods just for technical maintenance operations of the network. But it can provide insights about the tourist behavior, which can be used to improve the tourism products at the destination.

The appropriate management of the destinations is a critical aspect, which is widely recognized as a determinant of destination competitiveness (Crouch and Ritchie, 2003). However, any managerial decision should be based on available data, and that is a well-known challenge in tourism, as data tends to be partial and incomplete. In this context, the possibility of extracting relevant touristic information from the digital traces related with the use of mobile devices can become a great opportunity for tourism public and private managers. One advantage of this source of data as compared with other conventional sources of information as surveys or public data is that it provides much more observations than any other dataset. However, the main advantage is the possibility of monitoring the destination use in real time, because the data is collected in real time. In this sense, with the appropriate protocols of reaction that can be defined once the information's uses are defined, the decision can be made with only minutes from the events

The main contribution of the paper is that it uses a tourism destination approach to study the lessons and insights that can be extracted from tourists digital traces as they move through the destination. Among other applications, that information can be used to:

- ✓ Understand congestion in time and space, public managers can take advantage of that information for traffic management, security, etc. Private managers can also benefit from a precise and in real time information of such points for marketing purposes, or location decisions.
- ✓ Identify patterns of movement. The combination of information about device, position and time can provide a huge amount of information about itineraries and correlations of given locations that individuals tend to make. That has obvious applications for both Public and private managers.

The remaining of the paper is as follows; next section describes the network from which the data is generated. Section 3 explains the procedure to classify the devices as touristic or non-touristic, and to identify a set of locations of interest at the destination. Section 4 presents some applications of the type of information that can be obtained and its applications. The paper ends with the concluding section and the identification of issues of further research

1. Description of network infrastructure and the data source

The Internet access service in a touristic environment has gained in importance in the last years. In the same trend as the ICTs have begun to change the relationship between the citizen and the city services (in the context of the SmartCity paradigm change). The wireless fidelity, wireless internet (WIFI) infrastructures offered in a touristic destination has been increased extending the area of coverage. Initially, the WIFI service, in the context of tourism, was offered in public buildings (mainly libraries) and in private buildings as a pay-for-service (hotels, restaurants, etc.). But, now days, the WIFI coverage has been extended to everywhere and the access has become a free service. Mainly motivated by the digital behaviour changes of the tourists: when travelling not only the physical person moves from one place to the other, the digital profile of each person must be maintained during the travel. Therefore, it requires maintain a good quality digital connection.

This work takes profit of the existence of public WIFI services and explores the capabilities of the maintenance data set to be useful to know more about the users. This research tries to obtain results useful for tourism sector; therefore, the first step will be the selection of an adequate ICT to be considered as a data source.

In 2012, the SmartDestination working group proposed to the Playa de Palma consortium the implementation of a public ICT able to be used for public, private and research activities. In 2013, the SmartWiFiPdP (Smart WIFI Playa de Palma) was born and installed by TELCO S.L. (MallorcaWIFI, a local WIFI provider). Figure 1 shows the network coverage located at a mature touristic zone of Playa de Palma.

Fig. 1: SmartWiFiPdP Network coverage map located in a mature touristic zone (“Playa de Palma”) in the Island of Mallorca, Spain



The SmartWIFIPdP covers 5 Km of grounded beach during touristic season and it rises to record marks just in the first year of use (Crespí Seguí, B. (2015)). From that year, the SmartWIFIPdP has increased the number of users per day and the number of services offered. The main statistic features of the SmartWIFIPdP network shown in Figure 2 reveal why the network has been selected as a data source: the network is located in a very crowded touristic area, and exist an specific app (“SmartApp”) offered as a complementary touristic service. Therefore, the statistic data reported in Figure 2 comes from two sources: the network and the app. This work centres the attention of the network as a source.

In terms of technologies implemented in the SmartWiFiPdP, it is important to note that the network devices (routers and antennas) has the capability to sense any device with the wireless port activate, although it is not effectively connected to the network. Therefore, it is not necessary register the device in the network to sense data from it. Table 1 list the data fields provided by the SmartWiFiPdP maintenance platform from every interaction between the network infrastructure and the user device.

Table 1: List of the data fields provided by SmartWIFIPdP

Data Field	Format	Example
Router MAC address	Text	dd:dd:dd:dd:dd:dd ^(*)
User device MAC address	Text	dd:dd:dd:dd:dd:dd ^(*)
SSID	Text	Welcome_to_Plajta_de_Palma
User device model	Text	Apple
Event date	Date	2016-05-15
Event Time	Time	21:58:45
Event latitude	Float	39.53079170431497
Event longitude	Float	2.727746944874525

^(*) hidden numbers for privacy issues

The data showed in table 1 obtained from the network is just a small portion of the huge quantity of information used by TELCO S.L. to maintenance the infrastructure. In this sense, Figure 2 reports the representation of the statistics offered by TELCO S.L. to the municipality of Palma in order to summarize the benefits of SmartWifiPdP on the touristic knowledge of Playa de Palma.

Fig 2: Main features of SmartWifiPdP



This work tries to analyse the network data using the SmartWifiPdP as a data source. The fact that SmartWifiPdP covers only an area (Playa de Palma) offers the possibility to obtain information just from the destination, using the data from the destination. Of course, the results can be combined with social network data sets or other sources, but this research focus the attention of the data available in the destination itself.

2. Categorization of the observations and identification of relevant locations

Because the SmartWifiPdP is locate on a crowded area, the number of events produced during the day is high. Each day, the network collects approximately 2 Gbytes of observations data; which are saved in the project server. The dimension of the data set depends on the number of devices available in the network. This work reports the analysis performed using the observations from the second half of May 2016, from 15th to 31th of May.

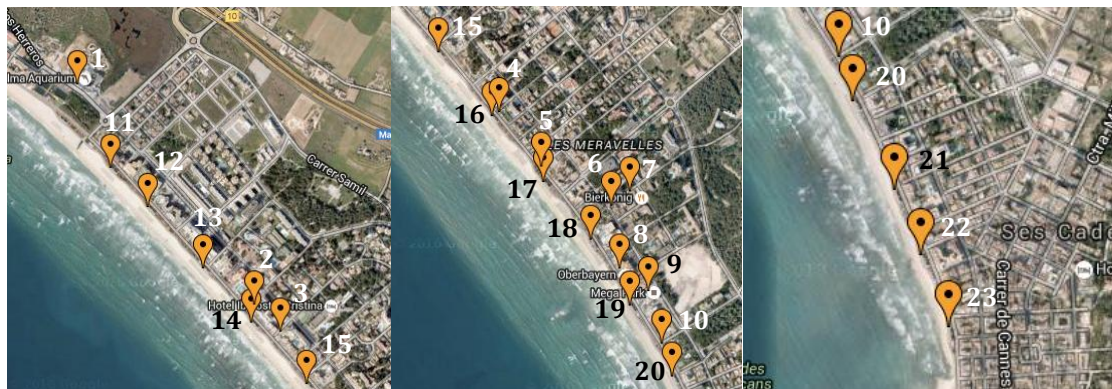
From the tourism perspective, the observation registers are relevant if they become from devices owned to a tourist. Therefore, it is necessary to apply a categorization process to analyse only the adequate observations. In this sense, it is necessary to distinguish between tourist's devices from the others (workers, residents, etc.). The categorization process has been based on the assumption that the regular touristic stance in Playa de Palma is between 5 and 7 consecutive days. Therefore, the observations from 15th to 21th of May has been used to obtain a set of user MAC addresses formed by tourist's devices and others. This initial set has been compared with the observations collected in 30th of May, and only different user MAC addresses were considered useful for this touristic analysis. This historical approach to categorize touristic users from others guarantees the quality of the observations from the point of view of tourism. In other way, not all observations can be analysed directly, it is necessary to adapt the data set to the specific research goals in order to clean the source data.

Another important issue to consider in this work is related to the geographical analysis. The SmartWifiPdP covers a big touristic area with different relevant locations from the point of view of tourism. In addition, the observation data includes geographical information of user device, as reports the list of data fields showed in Table 1. Therefore, one of the goals of this work is related to analyse the use of the space and try to do tracking of the tourist devices. In this sense, Table 2 identifies the relevant locations considered in this work. A total of 23 locations have been selected to explore all area of SmartWifiPdP coverage. Figure 3 represents on maps the relevant locations using the tag to identify each place.

Table 2: Relevant locations considered in this work

Relevant Location	Tag	Relevant Location	Tag
Aquarium	1	Balneario 11	13
Iberostar Bahia de Palma	2	Balneario 10	14
Iberostar Playa de Palma	3	Balneario 9	15
McDonalds	4	Balneario 8	16
Parque	5	Balneario 7	17
Grillmeister	6	Balneario 6	18
Bierkonig	7	Balneario 5	19
Burger King	8	Balneario 4	20
Mega Park	9	Balneario 3	21
Pabisa Beach Club	10	Balneario 2	22
Balneario 13	11	Balneario 1	23
Balneario 12	12		

Fig. 3: Geographical representation of relevant locations considered in this work using tags showed in Table 2



The relevant places selected are in the area of coverage of the SmartWifiPdP network. This is a hard requirement. Any place outside the coverage area can be included in the analysis. Although, this hard rule is limiting in terms of analysis area, it is extremely positive in terms of resolution, because the results obtained may be extremely precise from the point of view of destination behaviour analysis.

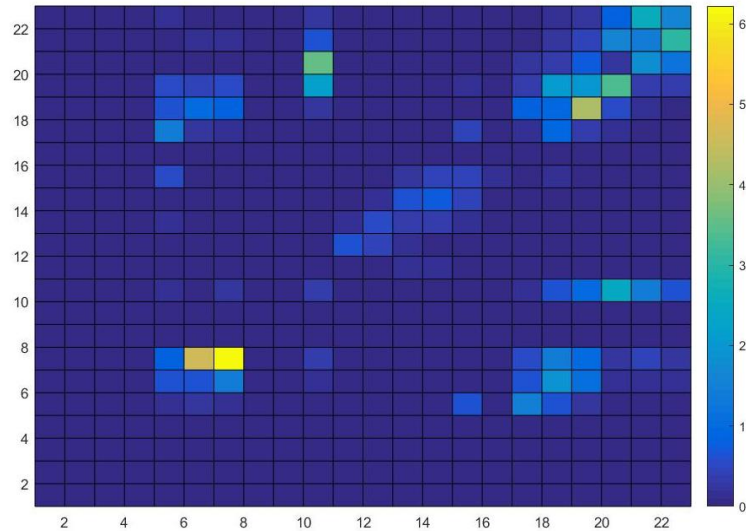
3. Data analysis for public and private destination intelligence

This section presents some examples of the type of analysis that can be provided with the available data and some uses to generate value from those analysis. As this is an initial exploratory study, it does not have a completeness perspective.

Figure 4 presents the dynamics of movement inside the covered area. The values in the diagonal capture those individual, which appear only in one location. In this sense, we can see from the graph that there is a relevant percentage of individual, which are present only at location 7 (Bierkonig). For the rest of the matrix, each column indicates which is the percentage distribution of previous locations. For example, column six captures individual that are at location 6 (Grillmeister) in a given point in time (t), and that have been monitored in a different location in their previous observed period (t-1). For each of those, each cell indicates the

percentage distribution of origins. In this case, we can see a relevant number of individuals captured at location 6 that come from location 7, followed by location 18 and 19. Destination planner to improve the routes among each location and their likely origins, for example, in relation with pedestrian movements, or security positions, could use this definition of likelihood of origin for each location. Private managers can also benefit from that information to define co-marketing strategies of other bundling options.

Fig. 4: Mobility analysis. Percentage of users at given location (x axis) at t-1 with respect to the next location detected, at t, (y axis)



The next set of figures (5, 6 and 7), identify the pattern of movement for different periods of the day. In a mass tourism destination as Playa de Palma, it is easy to relate those periods with differentiated types of touristic products as nightlife or sand and beach. The paper proposes the use of 3 periods during the day, as indicated in Table 3. The first period covers the start of the day and access to the beach. The second period is related with lunch and post lunchtime. Finally, the last period covers the usual time of leaving the beach, dinner and nightlife time

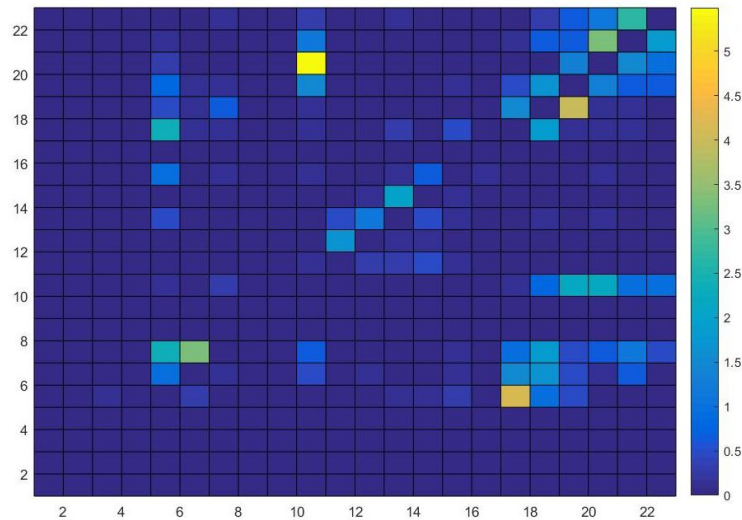
Table 3: List of relevant periods for the study

Period	Start time	End time
Morning	08:00:00	11:59:59
Evening	12:00:00	17:59:59
Night	18:00:00	07:59:59

Figures 5, 6 and 7 indicate the movement for each of the time periods defined in Table 3. Each of the Figures captures in the x axis the location at which each device is first capture at the corresponding time period. In this sense, the column indicates the percentage distribution of places visited afterwards during that period.

For the morning period, Figure 5, the matrix indicates relevant movements, as the ones indicated for individuals starting at location 11 (Balneario 13) that visit location 21 (Balneario 3), followed by the pattern from 17(Balneario 7) to 19 (Balneario 5) and from 19 (Balneario 5) to 18 (Balneario 6), etc.

Fig. 5: Mobility during the morning period. Percentage of users that are first observed at a given location (x axis) and the places where they go after during the morning period (y axis)



For the evening and night period equivalent analysis can be made. In this sense, in the evening period, it is possible to capture a similar movement as in the morning from location 11 (Balneario 13) that visit location 21 (Balneario 13) and from 19 (Balneario 5) to 18 (Balneario 6). For the night period represented in Figure 7 it is worth mentioning the relevance of the pairs from location 6 (Grillmeister) to location 7 (Bierkonig), followed by the pairs from 19 (Balneario 5) to 18 (Balneario 6) and 20 (Balneario 4) to 19 (Balneario 5)

More precise managerial applications as the ones described for the analysis of Figure 4 can be extracted from this time related movements. As an example, a good location to locate promotion after 17h related with the private activities at Bierkonig is Grillmeister. Obviously, the short geographical distance that can be seen at Figure 3 can explain that movement.

The same generic information without disaggregating the time differences is presented at Figure 8. In this sense, the application of that information could be related with other forms of generic promotion, as the one provided by the Wi-Fi network during the day

Fig. 6: Mobility during the evening period. Percentage of users located in on place (x axis) and the places where they have gone during the evening period (y axis)

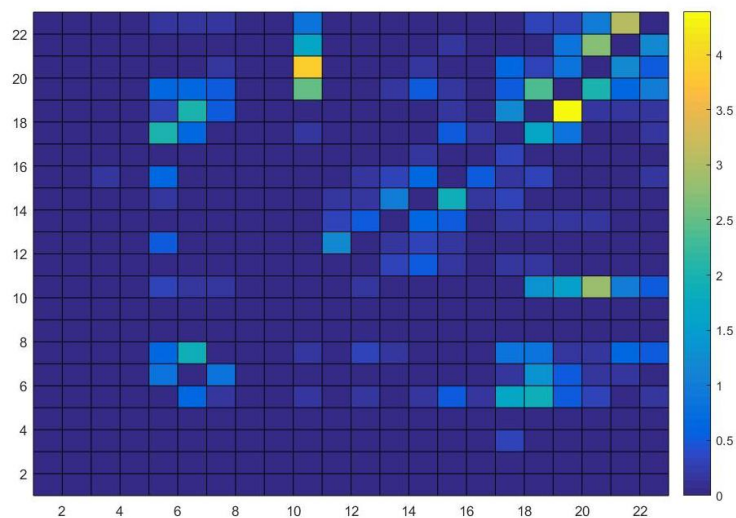


Fig. 7: Mobility during the night period. Percentage of users located in on place (x axis) and the places where they have gone during the night period (y axis)

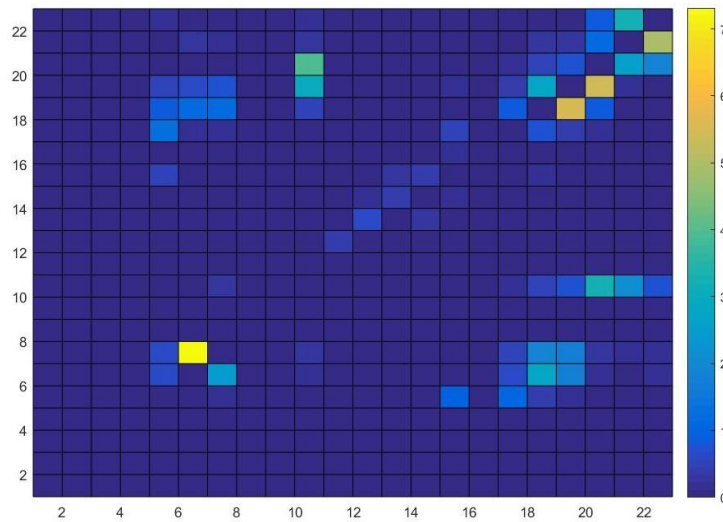


Fig. 8: Mobility during the day. Percentage of users located the first time in on place (y axis) and the places where they have gone during the day (x axis)

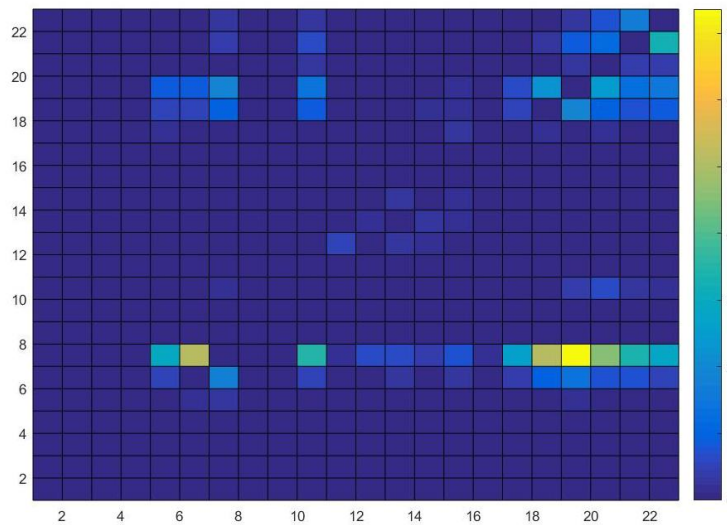


Figure 9 and 10 provide other type of information, which relates the geographical position and the temporal approach.

Figure 9 captures the duration for which the individuals remain at different locations. In this sense, Locations 7 (Bierkonig) and 19 (Balneario 5), are the ones, which present longer remaining times. Finally, Figure 10 captures the distribution of visit at the different hours of the day. This is a measure of the temporal aggregation of people at different locations. For instance, location 14 (Balneario 10 is related with early morning longer stays, while location 9 (Mega Park) is related with late high longer stays

Fig. 9: Percentage of users remaining the same time duration represented in periods of 5 minutes (x axis) in the relevant places (y axis)

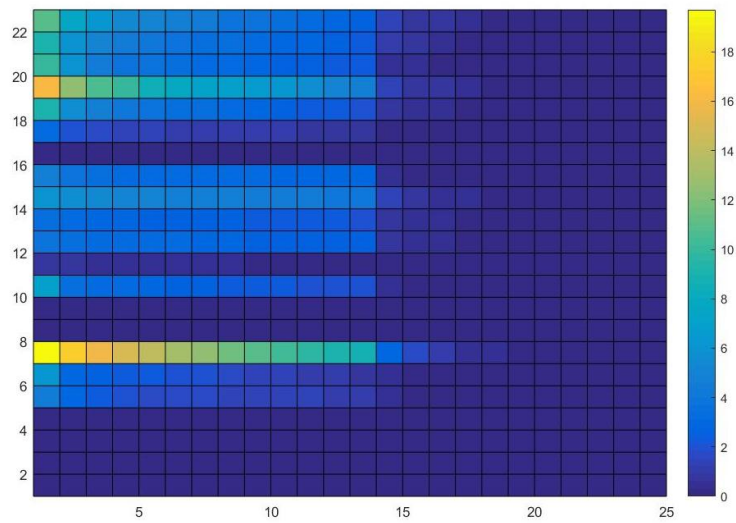
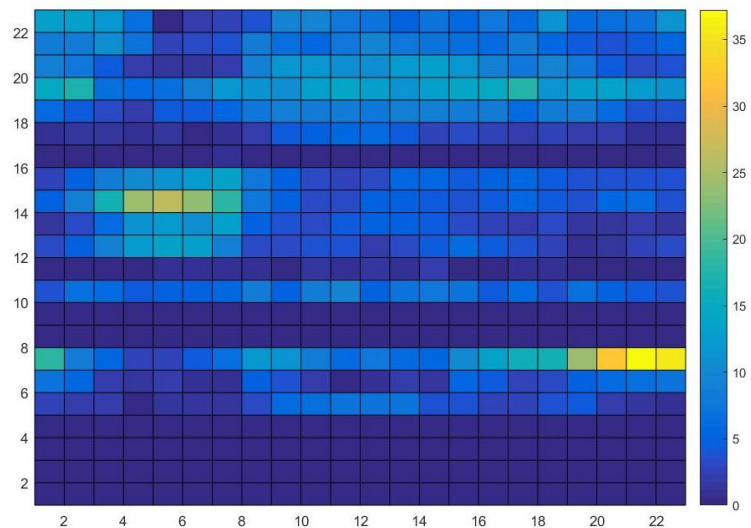


Fig. 10: Percentage of users in relevant location (y axis) depending on the time of day (x axis)



4. Conclusion and further research

The objective of this paper was to provide an initial exploratory analysis of the type of valuable information that can be obtained from a digital related database which was not exploited for tourism management before. The database was collected only for network technical reasons, but it contains information, which could add value for both private and public tourism managers.

The contribution of this paper is to provide a destination approach to the analysis of that database and to relate it with the potential private and public uses, which could help to generate value for the tourism industry at the destination. Particularly in a context of data scarcity, the possibility of extending our knowledge about tourists' movement across the geographical distribution of our destination could represent a remarkable improvement for tourism industry.

An initial proposal for data management identification is proposed, and an exploratory study of how to extract information with market intelligence applications is presented in section 3.

A lot of challenging issues remain in the sphere of further research issues as the researchers know-how on the treatment of the dataset improves. The extension of the time under analysis is

clearly the first in that list, which was constrained in this paper by the difficulties in managing the huge amount of data of the 24 hours database. Without being exhaustive, a more precise way of defining stops and locations, the identifications of itineraries and patterns of movement, and the creation of measures of congestion are among the ones under current development

REREFENCES

- Alegre, J., Pou, L. (2006). The length of stay in the demand for tourism, *Tourism Management*, 27(6), 1343–1355.
- Chilipirea, C., Petre, A., Dobre, C., van Steen, M. (2015). Filters for Wi-Fi Generated Crowd Movement Data, *10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*, 285-290. doi:10.1109/3PGCIC.2015.36
- Crouch, G. I., Ritchie, B. (2003). *The Competitive Destination: A Sustainable Tourism Perspective*. Cabi Publishing. ISBN: 978-1845930103
- Exceltur (2014). *Impactur Baleares (2014)*. <http://www.exceltur.org/wp-content/uploads/2015/10/IMPACTUR-Baleares-2014-informe-completo.pdf>
- González, M. C., Hidalgo, C. A., Barabási, A.-L. (2009). Understanding individual human mobility patterns, *Nature*, 458, 238–238. doi:10.1038/nature07850.
- Instituto Nacional de Estadística-Frontur. (2016) *Estadística de Movimientos Turísticos en Fronteras*. Retrieved July 9, 2016, from <http://www.ine.es/daco/daco42/frontur/frontur1215.pdf>
- Instituto Nacional de Estadística-CST. (2016) *Cuenta satélite del turismo de España. Base 2010*. Retrieved July 9, 2016, from <http://www.ine.es/prensa/np975.pdf>
- Phithakkitnukoon, S., Teerayut Horanont, T., Witayangkurn, A., Siri, R., Sekimoto, Y., Shibasaki, R. (2014). Understanding tourist behavior using large-scale mobile sensing approach: A case study of mobile phone users in Japan, *Pervasive and Mobile Computing*.
- UNWTO. (2016). *Annual Report 2015*. Retrieved July 9, 2016, from <http://www2.unwto.org/publication/unwto-annual-report-2015>
- Wolf, J., Guensler, R., and Bachman, W. (2001). Elimination of the travel diary: experiment to derive trip purpose from global positioning system travel data. *Transportation Research Record*, 1768, 125–134.
- SEGITTUR (2014). *Impacto de las nuevas tecnologías en los hábitos del nuevo turista en España*, Secretaria de estado de turismo, Instituto Tecnológico hotelero.
- Crespí Seguí, B. (2015), *Infraestructura SmartWifi Playa de Palma*. Libro de comunicaciones del I Congreso de Ciudades Inteligentes, ISBN: 9788460667551, <http://www.congreso-ciudades-inteligentes.es>