Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

Human mobility analysis by collaborative radio landscape observation

Carlos Pérez-Penichet
Centro Algoritmi
University of Minho
Guimarães, Portugal
cpp@dsi.uminho.pt

Ângelo Conde
Centro Algoritmi
University of Minho
Guimarães, Portugal
conde@dsi.uminho.pt

Adriano Moreira
Centro Algoritmi
University of Minho
Guimarães, Portugal
conde@dsi.uminho.pt

Abstract

A new method to analyze the spatio-temporal activities of humans based on the symbolic information that can be extracted from a set of observations of mobile networks taken through smart phones is presented. Specifically, GSM and WiFi network observations collected by several users are gathered to collaboratively build a symbolic base map of the logical structure of the geography. At the same time a map of the mobility of each individual is also created from the same set of observations. The Proximity Map is then used to provide some spatial context to the Individual Mobility Maps. This information is intended to be used for the analysis of transportation efficiency.
Keywords: Human Mobility, Spatiotemporal Data Analysis, Urban Dynamics, Symbolic Modeling, Collaborative Sensing, Mobile Networks

1 Introduction

The mobility patterns of humans are of big importance to many areas, in particular to the analysis of the efficiency in transportation systems. The TICE.Mobility project aims to help users create personal reports about their mobility in order to assist them in making a more efficient and sustainable use of the transportation systems in urban areas. Within the context of the project the information about WiFi and GSM networks, obtained using smartphones, is used as a proxy to study the mobility patterns of users. For instance, if the most usual displacements the user does can be identified, more environment friendly, cheaper or faster alternatives could be automatically suggested to him. Here we propose a new form of integration of different sources of geographic clues based on the communication infrastructure to infer the symbolic structure of space and the patterns of mobility of humans.

Several studies show an increasing popularity on the use of wireless networks [2, 5, 12, 15]. At the same time, works like [11] focus on individual mobility patterns, obtaining their information from mobile networks and other sources.

Studies conducted using the Reality Mining dataset [9, 10], which contains data related to users lives, show that it is possible to discover daily-life routines applying probabilistic topic models. The authors' main goal is to characterize, in terms of location patterns, both individual and group behaviors of human routines. Using the same dataset, Eagle and Pentland [8] applied Principal Component Analysis (PCA) to identify the main features which structure daily human behavior, such as leaving home in the morning, and returning home in the evening.

Other important studies about human mobility have been conducted. Gonzalez *et al.* [11] used a dataset of 100.000 anonymized mobile phone users to study the mobility patterns of humans. They found that individual travel patterns can collapse into a single spatial probability distribution showing that humans follow simple, reproducible patterns. Using a similar dataset, Song *et al.* [25] conclude that despite our desire for change and spontaneity, the human mobility is characterized by a deep-rooted regularity.

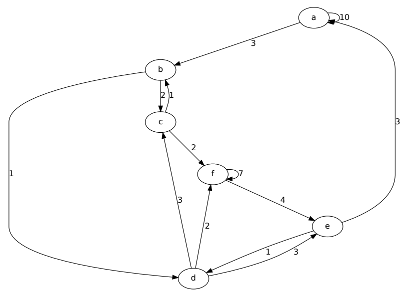
In addition to data from mobile cellular networks, other studies have been focused on data collected from GPS sensors [22], WiFi access points [14] and tracking of individual bank notes [4], to analyze and discover the fundamental statistical features of human mobility from real traces. These studies report the statistical behaviors of human mobility over various scales of time and space.

The analysis of human mobility necessarily starts by collecting position data. Many systems have been developed to detect user locations, maybe the best known of them is the Global Positioning System (GPS) but it suffers from a number of limitations, especially limited coverage in indoor environments [19] where most individuals spend around 87% of their time [17]. Furthermore, the power consumption of today's GPS modules in smartphones make it impractical to implement any real world solution based on GPS alone, since 24/7 collection of position data would considerably reduce the energy autonomy of smartphones.

A different type of localization system is the one developed by LaMarca *et. al.* known as PlaceLab [19, 24]. PlaceLab is aimed at providing localization services based solely on the presence of beacons of different networks (GSM, WiFi and Bluetooth) and extracting the associated location from a prepopulated database.

Location plays a very important role when determining a user's context, however in order to extend the possibilities of context aware applications, location alone is of limited usefulness [23]. For this reason, applications would highly benefit from a means of gathering higher level information about the surroundings of users. Many research efforts have been made in this direction, creating different sorts of abstractions related to users' location and its relation to context. For instance, in [21] a method for detecting the means of transportation and the possible future routes of users is presented. Authors are able to automat-

Figure 1: A personal mobility map represented as a graph.



ically classify users depending if they are traveling by foot, car or bus. The system works based solely on positions supplied by a GPS receiver and the prior knowledge of bus routes.

In [13] an algorithm to extract semantically meaningful places from location information is developed. Authors exploit location information from PlaceLab [19] to infer the places that are of meaning to the users based on the amount of time spent on each location. The algorithm presented is able to extract clusters of location information that represent frequented places but it is unable to label them in any way, which is a needed step in order for this information to be useful. An improvement to this idea is presented in [20], here the authors present a probabilistic model to detect candidate places for the "workplace" and the "home" of users.

The system presented in [6, 7] may be considered an integration of several of these ideas. The authors developed a smartphone application that can recognize a user's activity and is able to track his position using a combination of dead reckoning, WiFi fingerprinting and GPS [16]. From all this information the application is able to automatically recognize logical places and assign a location estimate to them. Then the application presents a historical summary of the mobility pattern of the user.

Mobile networks can be exploited in order to determine the locations of their users [1, 18, 19]. The goal here is to go one step further and use GSM and WiFi networks to infer the mobility patterns in a symbolic space, without necessarily knowing the geographic coordinates of places. The logical space structure is also inferred from the radio landscape and is used to provide some spatial context to the mobility patterns. This information is intended to be used for the analysis of transportation efficiency.

2 Personal Mobility Maps

The main objective of our work is to create a Personal Mobility Map of the user's spatio-temporal activities based on the symbolic information that can be extracted from a list of observations of mobile networks using smartphones. Here we refer to "observations" as samples of the radio landscape (WiFi and GSM) taken by specific sensory software running on each person's smartphone. This map should be represented by a graph $G_m = \{N_m, E_m\}$ which nodes N_m would be used to represent the places the user visits and its edges E_m would represent displacements among places (Figure 1).

Each observation includes the identification of the moving entity (a person) mID, a timestamp t and a list BSs of the Access Points (WiFi) or Base Stations (GSM/UMTS) visible to the user at that given moment. The list of all observations is represented as:

$$O = \{(mID_1, t_1, BSs_1), \cdots (mID_{Ns}, t_{Ns}, BSs_{Ns})\} \quad (1)$$

We assume there are Ns observations and that in the i'th record Nbs_i was the number of base stations detected. The lists BSs_i have the form:

$$BSs_i = \left\{ \left(bsID_{i,1}, RSSI_{i,1}\right), \cdots \left(bsID_{i,Nbs_i}, RSSI_{i,Nbs_i}\right) \right\} \quad (2)$$

where $bsID_{i,j}$ is the identification of the j'th detected station in the i'th observation and $RSSI_{i,j}$ is the received signal strength of the j'th detected base station in the i'th observation.

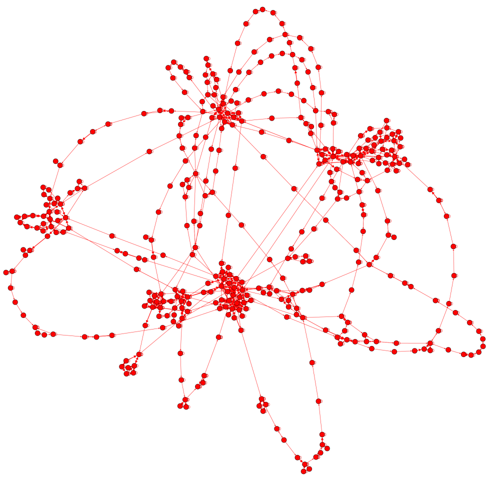
Ideally every node of the Personal Mobility Map should represent a meaningful location such as "The University Library" or "Steve's House". By observing the number and pattern of appearance of each of this locations in the observations form a single user it is possible to characterize each of them. For instance, it is expected that the locations with a larger number of occurrences are the more relevant to that particular user such as his or her home and workplace. A detailed temporal analysis could lead to better insight about the users utilization pattern of every place he visits such as how long he stays in each place and during which period of the day or the week. Since for every location there are potentially many visible APs, a way of grouping all the APs corresponding to the same location should be used. Because APs are not geo-referenced, grouping cannot be based on traditional Euclidean distance applied to pairs of coordinates. While a suitable solution to this difficulty is implemented, the Personal Mobility Maps are built simply using the nearest (the one with strongest signal) base station (WiFi or cellular) to represent the users presence in a place for every observation, thus the nodes in the Personal Mobility Map represent base stations.

On the other hand, by counting the number of displacements between pairs of nodes, it is possible to obtain the most frequent trips the user does. This information characterizes the displacement of the users and is represented in the Personal Mobility Graph as the weights of the edges. A detailed temporal analysis of the trips could also help further characterize the users' mobility patterns.

In Figure 2, the Personal Mobility Map of a single real user is depicted. The map was obtained from a set of 93.629 WiFi observations obtained during a period of 22 days. The resulting graph is formed by 404 nodes and 1.183 edges.

It must be noted that the obtained graph successfully shows the aggregated movements of the user and the trajectories he follows, however it lacks spatial context, that is: there is no relation between the relative position of the nodes in the graph of a given user and the actual relative position of the places they represent in space. A possible solution to these difficulties is to gather symbolic information about the topology of the underlying infrastructure in another graph, called the Proximity Graph.

Figure 2: Example of a real personal mobility map.



3 Proximity Graphs

The Proximity Graph $G_p = \{N_p, E_p\}$ is a map of the proximity of network access points, represented by the nodes N_p, based on their visibility to, possibly many, users through their smartphones. The main assumption is that if two access points are simultaneously visible to a given user, then they must be close to one another. The nodes of the proximity graph represent all APs as reported by one or more user observations:

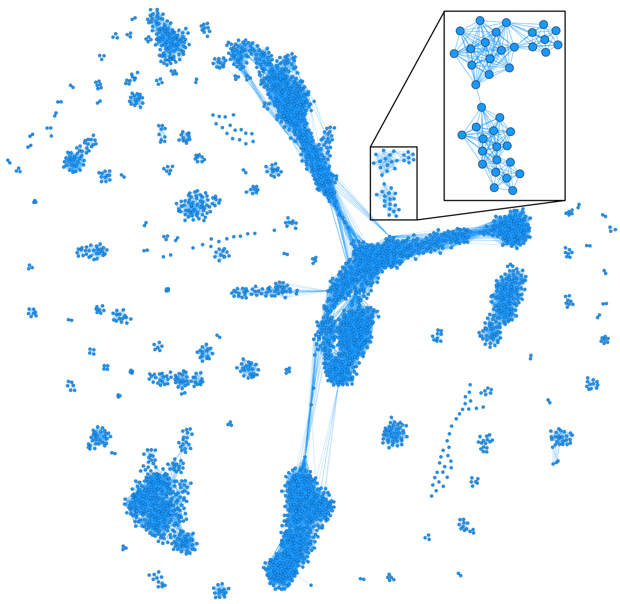
$$N_p = \bigcup_{\substack{i=1\\j=1}}^{\substack{Nbs_i\\N_s}} bsID_{i,j} \tag{3}$$

The edges of the proximity graph E_p represent proximity between pairs of nodes. An edge will exist between any pair of nodes if and only if those two nodes are detected simultaneously by a user, meaning the nodes must not be too far apart:

$$E_p = \left\{ \bigcup_i (bsID_{i,j}, bsID_{i,m}) : j \neq m \right\} \tag{4}$$

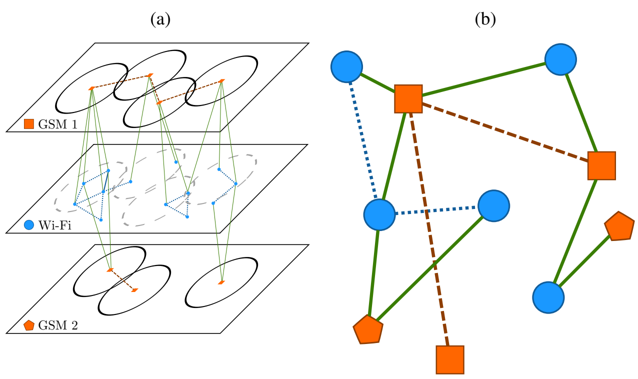
Figure 3 shows an example of a Proximity Map with 2.607 nodes and 41.470 edges computed from the observations gathered by two different users. The *ForceAtlas 2* layout algorithm as implemented in Gephi [3] was used to draw the graph. On the figure it can be seen that a number of disconnected subgraphs exists. These subgraphs represent places that are so far from each other that no pair of APs in either of them can be seen simultaneously. This lack of relationship among subgraphs is an unwanted feature since no spatial relation exists among those subgraphs. It can also be seen on Figure 3 that several nodes, representing APs, are strongly connected forming groups, while there exist relatively few connections among some of those groups. The strongly connected groups may represent locations with a dense availability of APs such as different buildings in a university campus.

Figure 3: Example of a real WiFi proximity map computed from a set of 93.629 WiFi observations collected over a period of 22 days by two users.



To overcome the issue of the disconnected subgraphs, another layer of spatially contextualized information may be added. One possible source of such layer is the GSM network that is ubiquitous nowadays. By analyzing GSM base station visibility in a way analog to what is done for WiFi, a so called GSM Proximity Map could be created. Such a map would probably be similar to the ones WiFi offers, except that it will have a much larger scale and probably a less spotty coverage of the geographic space. This new map would bring a new "dimension" to the WiFi data set. Since the characteristics of these maps should mimic some features of the same real world space, but probably at two different geometric scales, they should complement each other to bring a much more accurate representation of the geographic space (Figure 4). In fact, because the smartphone API will only

Figure 4: The matching relationship that exists between the WiFi and GSM layers.



Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

report the presence of GSM base stations of the associated service provider, and different users might have different providers, there will be more than just one single GSM layer.

From the outcome of the GSM map it should be expected that some of the disjoint clusters that are present in the WiFi map could be joined through the links between GSM towers on the GSM "plane" in a way similar to what is described in Figure 4. In order to link the WiFi and GSM layers from the observations of a particular user, a "time alignment" method has been used. If a certain WiFi access point was observed within a certain time interval Δt of the observation of a given GSM base station, then it is assumed that they are close enough to be observed simultaneously and they are linked together.

4 Results

Figure 5 depicts the Proximity Map obtained by merging one GSM layer (one provider) with observations made by both users to the WiFi layer shown in Figure 3. The GSM layer (computed from 346.150 observations using $\Delta t = 15s$) contains 353 nodes and 644 edges among GSM stations and 6.289 to WiFi APs. WiFi nodes are displayed in blue as are the edges among them. GSM nodes and the edges among them are drawn in orange. Green edges represent the connections among nodes in the GSM layer to nodes in the WiFi layer. The size of the nodes is proportional to the number of observations where they appear.

The motivation for building Proximity Maps was the construction of maps that could be used to provide spatial context to the abstract Personal Mobility Maps. Figure 6 shows the Personal Mobility Map of Figure 2 drawn over the Proximity Map obtained from the WiFi and GSM observations (Figure 5). Here all nodes and edges from the Proximity Graph are drawn in gray, then any node also present in the user's Personal Mobility Map is colored in red and finally the edges from the the Personal Mobility Map of the user are added and also colored in red to represent the user's displacements.

From this example it can be seen that the combined analysis of mobility and proximity maps can bring a much larger amount of information than what can be obtained with the mobility alone. Specifically, by putting spatial context into the Personal Mobility Map, it is possible to better understand the displacements of users and the characteristics of their daily routines in terms of mobility. With this information and the temporal patterns mentioned earlier it should be possible to perform a detailed analysis of many aspects of the mobility patterns of the users. It is expected that by combining the Proximity Maps of multiple users, a bigger and more detailed Proximity Map can be built and that then individual users can benefit from it to provide context to their Personal Mobility Maps.

Furthermore, it has been shown that the inclusion of multiple spatial proxies, such as WiFi and GSM beacons, can complement each other to greatly improve the quality and accuracy of the spatial representation they offer. In spite of these accomplishments there are still several issues that must be dealt with.

5 Research Challenges

The practice of setting up a WiFi access point in smartphones to be used for tethering is increasingly common. Consequently, the number of APs that appear in different locations at different times is increasing. This is an issue to be dealt with in the construction of the WiFi Maps because one implicit assumption is that APs do not move an thus can be used as a proxy for physical locations. A way to identify mobile APs should be found in order to discard them or maybe to use them in an advantageous manner. One possible way of telling mobile APs apart from static APs may be to analyze the time profile of neighbors for the APs. The mobile APs should have a much larger variability in neighbors than static ones.

Because the layout is needed to represent the inferred proximity of the nodes in space, a suitable metric for the edges of the Proximity Graph is still to be found. Pairs of nodes joined by edges with high value of this metric should be drawn closer to each other while edges with lower value must force their associated nodes to be placed further apart. Another issue is finding a graph layout algorithm that can, based on the aforementioned edge weight metric, draw a graph resembling the actual geographic distribution of places as closely as possible.

Besides dividing the graph into connected components as suggested in section 2, it is possible to gain a better resolution by running a clustering algorithm to detect the logical places visited by users in order to gain more resolution when building the mobility maps of individual users. The clustering process depends on finding the edge weighting metric mentioned above. This metric is also expected to be a key factor in validating and evaluating the obtained Proximity and Personal Mobility Maps.

Finally, dealing with the mobility data of multiple users raises obvious privacy concerns, data must be handled appropriately in order to avoid mistreatment of the information; however merging data into an aggregated Proximity Map contributes to hide individual patterns.

6 Conclusions

Despite the open issues, it was possible to build Personal Mobility Maps from which relevant information can be successfully extracted, showing the validity of our new approach. The presented maps show that it is possible to extract a lot of information about the mobility of smartphone users by the observation of the radio landscape surrounding them in their daily life. The detail and usefulness of this information is enhanced by the collaborative construction of WiFi and GSM proximity maps.

Adding the multiuser data about the GSM network allowed to connect the disconnected subgraphs present in the WiFi map and to improve the contextualization offered to Mobility Maps. With the merged data the relative position of the visited places became much more apparent in the graph, improving the obtained representation of space. This can only be achieved by aggregating the data from several users, showing how the obtained representation of space is enriched beyond what is achieved with single users.

These maps were built using the most simple of the available solutions to the presented issues and thus there may be a lot of room for improvement. The level of detail obtained can be expected to improve with the development of the proper techniques to handle the open issues.

Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

Figure 5: Proximity Map merging WiFi and GSM graphs. WiFi nodes appear in blue and GSM nodes are drawn in orange. Green edges represent connections between the WiFi and the GSM layers. The size of the nodes is proportional to the number of observations where they appear.

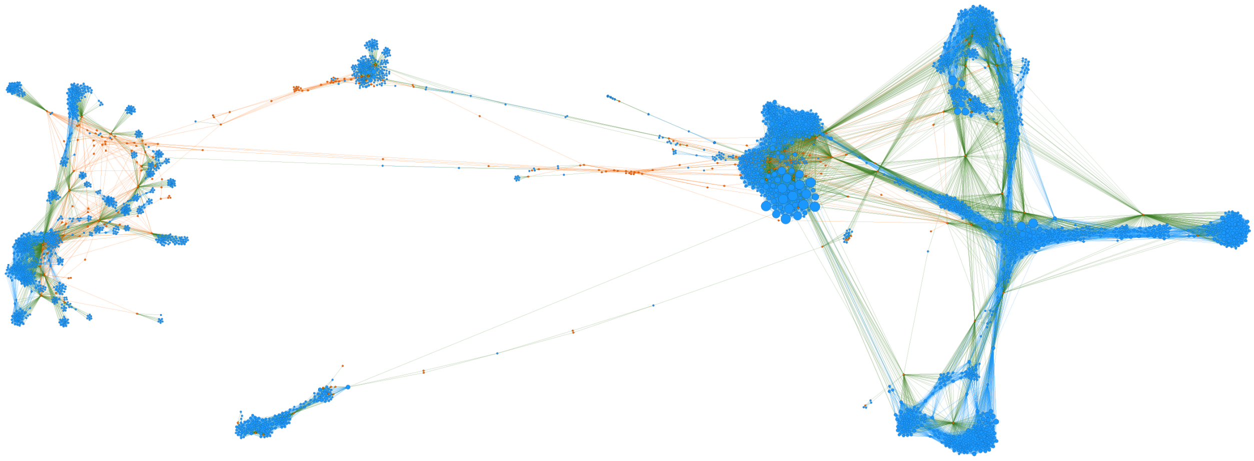
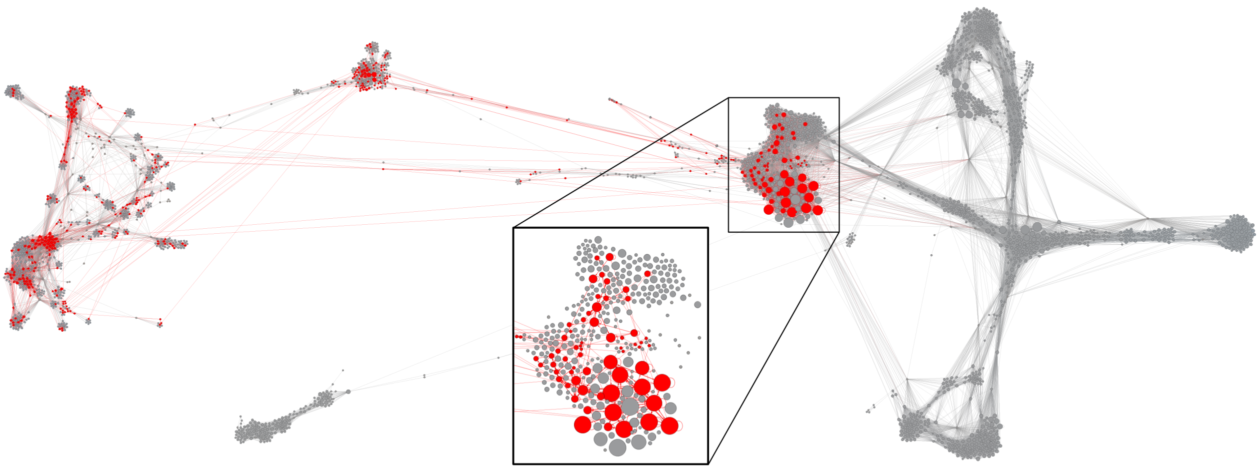


Figure 6: Personal Map overlaid on the GSM-WiFi proximity map. Nodes and Edges present in the Personal Mobility Map are colored in red.



References

[1] P. Bahl and V. N. Padmanabhan. RADAR: an in-building RF-based user location and tracking system. In *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 775–784, 2000.

[2] A. Balachandran, G. M. Voelker, P. Bahl, and P. V. Rangan. Characterizing user behavior and network performance in a public wireless LAN. In *ACM SIGMETRICS Performance Evaluation Review*, volume 30, pages 195–205, 2002.

[3] M. Bastian, S. Heymann, and M. Jacomy. Gephi: An open source software for exploring and manipulating networks. In *International AAAI Conference on Weblogs and Social Media*, page 361–362, 2009.

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

[5] F. Chinchilla, M. Lindsey, and M. Papadopouli. Analysis of wireless information locality and association patterns in a campus. In *INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 2, pages 906–917, 2004.

[6] J. Chon and Hojung Cha. LifeMap: a Smartphone-Based context provider for Location-Based services. *IEEE Pervasive Computing*, 10(2):58–67, February 2011.

[7] Y. Chon, E. Talipov, and H. Cha. Autonomous management of everyday places for a personalized location provider. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, PP(99):1–14, April 2011.

[8] Nathan Eagle and Alex Sandy Pentland. Eigenbehaviors: identifying structure in routine. *Behavioral Ecology and Sociobiology*, 63(7):1057–1066, April 2009.

[9] K. Farrahi and D. Gatica-Perez. What did you do today?: discovering daily routines from large-scale mobile data. In *Proceeding of the 16th ACM international conference on Multimedia*, pages 849–852, 2008.

[10] 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.

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

[12] Tristan Henderson, David Kotz, and Ilya Abyzov. The changing usage of a mature campus-wide wireless network. pages 187–201. ACM Press, 2004.

[13] J. H. Kang, W. Welbourne, B. Stewart, and G. Borriello. Extracting places from traces of locations. In *Proceedings of the 2nd ACM international workshop on Wireless mobile applications and services on WLAN hotspots*, page 110–118, 2004.

[14] Minkyong Kim and David Kotz. Extracting a mobility model from real user traces. *In Proceedings of IEEE INFOCOM*, 2006.

[15] Minkyong Kim and David Kotz. Periodic properties of user mobility and access-point popularity. *Personal and Ubiquitous Computing*, 11:465–479, October 2006.

[16] Y. Kim, Y. Chon, and H. Cha. Smartphone-Based collaborative and autonomous radio fingerprinting. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 42(1):112–122, December 2010.

[17] N. E. Klepeis, W. C. Nelson, W. R. Ott, J. P. Robinson, A. M. Tsang, P. Switzer, J. V. Behar, S. C. Hern, W. H. Engelmann, et al. The national human activity pattern survey (NHAPS): a resource for assessing exposure to environmental pollutants. *Journal of Exposure Analysis and Environmental Epidemiology*, 11(3):231–252, 2001.

[18] J. Krumm and E. Horvitz. Locadio: Inferring motion and location from wi-fi signal strengths. In *First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services (Mobiquitous 2004)*, pages 4–13, 2004.

[19] A. LaMarca, Y. Chawathe, S. Consolvo, J. Hightower, I. Smith, J. Scott, T. Sohn, J. Howard, J. Hughes, F. Potter, et al. Place lab: Device positioning using radio beacons in the wild. *Pervasive Computing*, pages 301–306, 2005.

[20] Adriano Moreira and Maribel Yasmina Santos. Enhancing a user context by Real-Time clustering mobile trajectories. In *Information Technology: Coding and Computing, International Conference on*, volume 2, pages 836–844, Los Alamitos, CA, USA, 2005. IEEE Computer Society.

[21] D. Patterson, L. Liao, D. Fox, and H. Kautz. Inferring high-level behavior from low-level sensors. In *UbiComp 2003: Ubiquitous Computing*, pages 73–89, 2003.

[22] Injong Rhee, Minsu Shin, Seongik Hong, Kyunghan Lee, Seong Joon Kim, and Song Chong. On the levy-walk nature of human mobility. *IEEE/ACM Trans. Netw.*, 19(3):630–643, June 2011.

[23] A. Schmidt, M. Beigl, and H. W. Gellersen. There is more to context than location. *Computers & Graphics*, 23(6):893–901, 1999.

[24] T. Sohn, W. G. Griswold, J. Scott, A. LaMarca, Y. Chawathe, and I. Smith. Place Lab—An open architecture for Location-Based computing. In *ESEC/FSE*, 2005.

[25] C. Song, Z. Qu, N. Blumm, and A.-L. Barabasi. Limits of predictability in human mobility. *Science*, 327:1018–1021, February 2010.

Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke