Daily and Weekly Patterns in Human Mobility

Eelco Herder and Patrick Siehndel

L3S Research Center, Leibniz University Hannover, Germany
{herder,siehndel}@L3S.de

Abstract. Most location-based services provide recommendations based on a user's current location or a given route or destination. Even though there are indications that human movement is highly predictable, daily and weekly routines of individual users constitute a largely unexplored and unexploited area. In this paper, we show and discuss such patterns, based on a large dataset of GPS trajectories. In particular, we show what information can be derived from the data of one exemplary user. These observations pave the way for personalization techniques that are based on the user's 'personal vicinity', as well as timely or delayed recommendations.

Keywords: GPS, Geolocation, Mobility Patterns, Personalization

1 Introduction

Location-based services have become a standard functionality on cell phones and other mobile devices [4]. These services may be offered on request, such as a list of nearest hospitals, restaurants, shopping malls or gas stations, or these services are automatically delivered when a certain event occurs. The latter category includes location-based tourist information, a notification that a friend is in the neighborhood, but also location-based advertising.

Human location tracking has further potential as well. The response time of police, emergency ambulances, fire brigades and road assistance can be greatly improved if they know where they need to go to. On a larger scale, human mobility tracking can be used for reducing traffic congestion, improving urban planning or limiting the spread of a disease [10]. There are negative aspects to location tracking as well: not all children will be equally amused with parental control, and the possibility that law enforcement agencies might follow your whereabouts or that insurance companies track how often you visit fast-food restaurants probably doesn't appeal to most people either. Still, it is likely that most people are willing to sacrifice some of their privacy in order to enjoy the benefits of location-based services [10].

The most common use of location-based services is to help users find points of interest in the direct neighborhood or along a route. As argued by Mokbel et al [8], distance is often the only criterion for selecting a destination: user preferences are typically not taken into account. Moreover, location-based services are usually 'stateless': they only take the current location or destination into account - daily or weekly patterns and routines are largely ignored.

Several studies confirmed the intuition that human mobility is highly predictable [6, 9], centered around a small number of base locations. This opens a wide range of opportunities for more intelligent recommendations and support of routine activities. Still, empirical studies on individual mobility patterns are scarce. In this paper, we aim to provide researchers with some new insights and inspiration. For this purpose, we analyze, visualize and discuss patterns in a large dataset of GPS trajectories, and discuss implications and opportunities for personalization.

The remainder of this paper is structured as follows. In the next section we discuss background and related work. Then we describe the dataset that we used, and present the results in Section 4. The result section is divided in two parts: first, we describe regularities found for the overall user population, and then we focus on 'a week in the life' of one exemplary user. The paper ends with a discussion of implications and opportunities for personalization.

2 Background and Related Work

Until recently, insight in human mobility patterns was limited, due to the lack of tools to monitor the movements of individuals. In particular the growing use of GPS technology has changed the situation. González et al [6] studied people movements, based on a sample of 100,000 randomly selected individuals, covering a six-month time period. The results show that human mobility patterns do not follow random-walk patterns: instead, trajectories show a high degree of spatial and temporal regularity. Further, individuals typically return to a few highly frequently locations and most travel trajectories are rather short in terms of distance and travel time. In a follow-up study, Song et al [9] found that 93% of human mobility is predictable; how predictable an individual's movements is, depends on the entropy of his patterns. For predictability it did not make a difference whether an individual's life was constrained to a 10-km neighborhood or whether he travels hundreds of kilometers on a regular basis. The researchers did not attempt to make actual mobility predictions, but indicated that 'appropriate data-mining algorithms' could be used for this purpose.

Zheng et al [13] used GPS data for mining interesting locations and 'classical sequences', based on the number of visits and the individual visitors' location interests. The outcomes are reported to be useful for tourists, who can easily discover landmarks and popular routes. At the same time, the authors note that the aggregated data may not be that useful for everyday location recommendations in a user's home city. Zheng et al [12] also investigated methods for mining correlation between locations, to serve as input for collaborative-filtering based location recommendations.

The above-mentioned studies focused on general human mobility patterns, with some insights on the predictability of individual mobility patterns. However, location-based services particularly depend on the locations and movements of individual users. The studies also did not investigate temporal dynamics in human mobility: can daily and weekly patterns be observed for the overall population and for individual people? And what can we learn from these patterns?

Apart from GPS data, a popular source for the analysis of human mobility is social media data - which usually is based on GPS data. However, social media data is reported to be sparse: most Twitter users only mention a very generic home location and less than 1% of tweets contains metadata on the location where it stems from [3]. Similarly, data from foursquare [1], a popular location-based social networking tool for mobile devices, is incomplete as well: foursquare does not automatically track the locations of users and only registers the users' location when they 'check in' at some place.

Dhar et al [4] discuss applications and business models for location-based services, which are separated in four categories: *information services*, such as yellow-page information about nearest hospitals, parking lots, restaurants, gas stations and other locations of interest; *tracking and navigation services*, which includes voice-enabled route description, locating friends in a particular area, and parental tracking of children; *emergency services* like roadside assistance, police and fire response; and *location-based advertising*, such as wireless coupons, marketing promotions and alerts, and customer notification.

Location-based information services are typically provided as recommendations [2] or as contextualized search results [11]. Several surveys show that restaurants and stores are the most popular locations that users search for, followed by local attractions and locations associated with leisure time [2, 11]. As noted before, these services usually provide suggestions based on the user's current location. In a recent study, Teevan et al. [1] showed the benefits of suggestions that are based on the user's predicted destination.

It goes without surprise that there are serious privacy issues associated with human location tracking. Among others, Krumm [7] showed that it is relatively easy to identify users' home locations within a trajectory log, making use of fairly straightforward heuristics. Golle and Partridge [5] showed that the approximate locations of an individual's home and workplace are sufficient to deduce her identity from anonymized trajectory logs. In both papers, several privacy-preserving techniques are discussed. From an end-user's perspective, location-based advertising - in particular unsolicited push advertisements - are generally considered intrusive. Soper [10] urges wireless service providers, mobile app developers, and others involved in GPS-based human mobility tracking to take steps towards self-regulation. However, at the end it is the user who *chooses* to allow this information to be tracked and used.

3 Dataset and Tools Used

As a basis for our analysis, we used the GeoLife GPS Trajectory Dataset [2] [13, 12], which contains a total of 17,621 trajectories from 178 users. The trajectories

[1] `https://foursquare.com/`

[2] `http://research.microsoft.com/en-us/downloads/b16d359d-d164-469e-9fd4-daa38f2b2e13/`

cover a total distance of about 1.2 million kilometers and a total duration of more than 48,000 hours. Locations are expressed in longitude and latitude.

Several preprocessing steps have been carried out before analysis. As we are interested in the start and end locations and the durations of the trajectories, we extracted the first and last entry of each trajectory in the dataset; this data was stored as a single entry in the database, representing a trajectory with a start point and an end point - the duration is the difference between the two time stamps. This reduced the data size from about 1 GB text data to a MySQL database of only a couple of megabytes. Subsequently, the different longitudes and latitudes were merged into (numbered) locations, by comparing the distance of each new start point or end point with the person's previously stored locations. After experimentation with different thresholds (starting with 20 meter, which is reported to be the current precision of GPS [3]), we finally chose a fairly large threshold of 300 meter. As a consequence, some locations that with only about 20-30 houses between one another may have been merged; we decided to consider this distance to be 'in the direct neighborhood'. In total, 14% of all trajectories were roundtrips, with durations varying from less than one minute to over 5 hours.

In this paper, we use 'User 160' as a running example. User 160 was ranked 9 of the people that provided most trajectories. His data (we do not know the gender of user 160, but we refer to the user as 'his' for convenience) consists of 401 trajectories, collected on 280 distinct days.

4 Results

4.1 Overall Travel Activity

As can be expected from human behavior, several aspects of the trajectories follow a power-law distribution: there are only a few locations that are start or end point of most trajectories, there is only a small number of trajectories that users follow most of the time; there are only few trajectories with a very long duration and many with a very short duration.

As we are interested in daily and weekly patterns, we visualized the number of trips that started at a specific hour (e.g. between 8am and 9am) on a certain day or group of days (week, weekend). Figure 1 shows the fluctuation of trips during the day for each weekday. The thick black line is the average of the five weekdays (Monday till Friday) and the thick grey line averages the weekend days (Saturday and Sunday).

Some strong regularities can be observed. First, there is very little traffic between about midnight and 7am. Further, on weekdays, the morning rush hour has a strong peak at 8am; the evening rush hour is more spread between 5pm and 9pm. Between both rush hours, traffic is moderate.

During weekends, the picture looks completely different: traffic starts between 7am and 8am and then remains relatively stable throughout the day, with a slight increase of traffic just before dinnertime.

[3] http://en.wikipedia.org/wiki/Global_Positioning_System

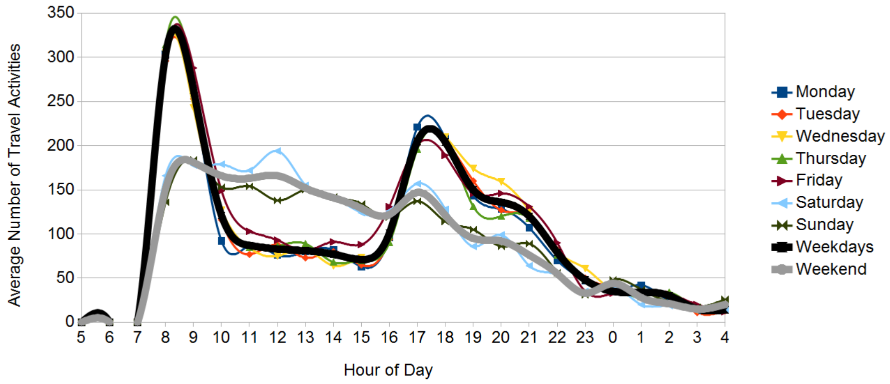

Fig. 1. Daily travel activity during the week and in weekends.

The differences between weekdays and weekends can obviously be explained by the fact that many people spend their weekdays at work; weekends are used for shopping, family visits and other sparetime activities throughout the day.

These observations may not come as a surprise, but the consistency and the strength of these regularities were larger than we expected. Naturally, these tendencies are averages of the 178 users represented in the dataset and individual differences do exist, as displayed in Figure 2.

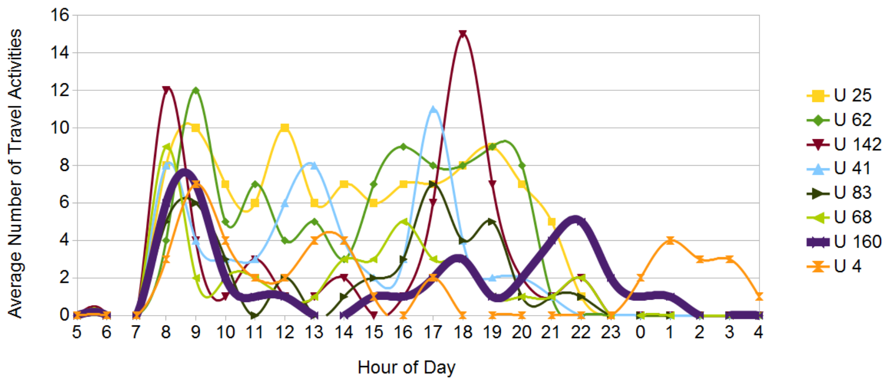

Fig. 2. Travel activities on weekdays of the top users in the dataset.

The morning rush hour is still visible, but the different peaks at 8 and 9 show that people start their day at different moments. User 142 (dark red, with a sharp peak between 6pm and 7pm) appears to have a job without much traveling and with strict office hours. User 41 (light blue) goes home relatively early and seems to have lunch outside of his office.

The daily activities of user 160 are displayed in bold. From the curve it appears that he starts his working days between eight and nine, and then returns home at either 6pm or 10pm. A rather average pattern, which we will further investigate in the next subsection.

4.2 A Week in the Life of User 160

In this section we further explore the daily movements of user 160. This user was ranked 9 of the people that provided most trajectories and, as discussed before, his behavior is quite representative for the average user. We do not aim to develop or evaluate prediction algorithms, but rather aim to 'get to know user 160 better'. We could have taken any other user with sufficient data - of course, this would have led to slightly different results due to individual differences.

Figures 3 and 4 show the movements of user 160 that started from the 5 most visited locations. On weekdays, he typically leaves 'location 1' (blue) between 8am and 9am. He usually leaves 'location 0' (red) at about 10pm. Further, he visits and leaves 'location 2' (yellow) at several points during the week, but particularly during the weekend, in the morning. 'Location 13' is visited in evening hours during the week; in the weekends, location 13 is usually visited during the day, in particular late afternoon.

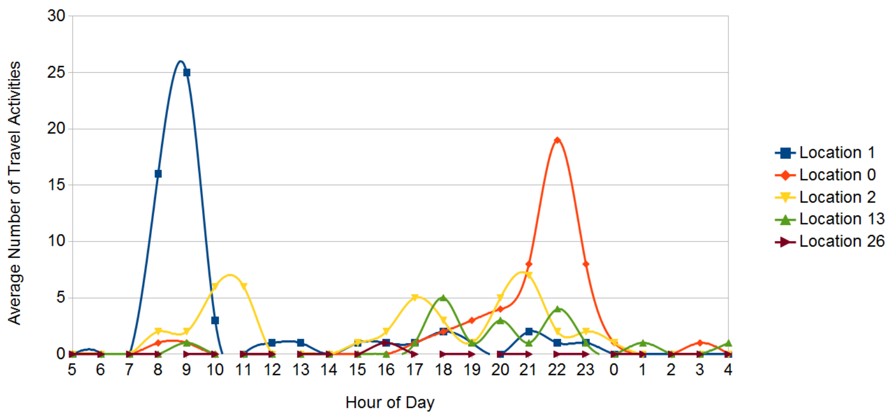

Fig. 3. Daily travel activities of user 160 on weekdays.

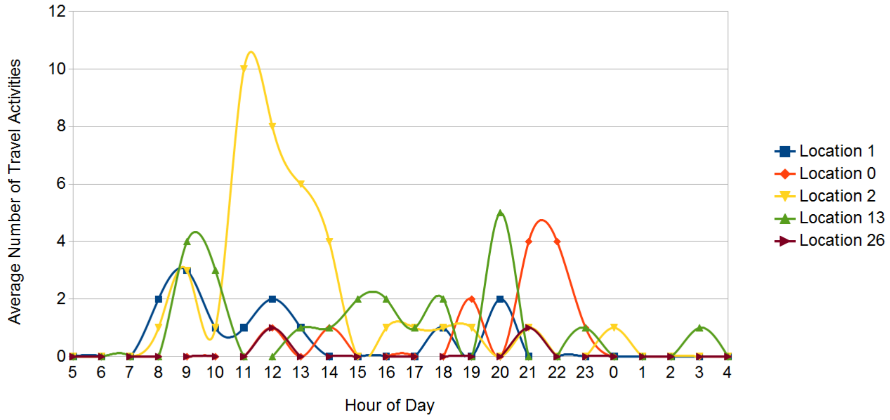

Fig. 4. Daily travel activities of user 160 during the weekend.

Without more context information, it seems likely that location 1 is user 160's Home. Location 0 is probably his Office - he seems to work late and sometimes also in the weekend during evening hours. Location 2 could be a shopping mall, and location 13 might be user 160's (sport) club.

Even though the exact purpose of these locations is just a more or less educated guess, from the figures it becomes clear that these are the four locations where user 160 spends most of his time and where one could find him at a certain point of day - during the week or during the weekend. The fifth most visited location, 'location 26' is visited far less often already and there is a long tail of locations (137 in total) that are visited only a couple of times.

Another interesting aspect is how these locations - the top 4 and the remaining locations at the tail - are related with one another. In order to find this out, we visualized the locations and the trajectories between them using the open source graph visualization toolkit Gephi [4], see Figure 5. The size and color of the nodes reflect the number of times that user 160 traveled to or from this location (large and red is very often, small and blue is only a few times). The width and color of the edges reflects the number of times that user 160 traveled between two locations. The graph layout is force-directed, with the longitude and latitude of the locations as a basis.

Figure 5 provides evidence for our assumption that user 160 has location 1 as his Home, and location 0 as his Office. Location 13 ('Sport club') is often visited from home, but also from his office. Location 26, Location 3 and various other locations are only visited when user 160 happens to be at Location 2 ('Shopping mall').

[4] `https://gephi.org/`

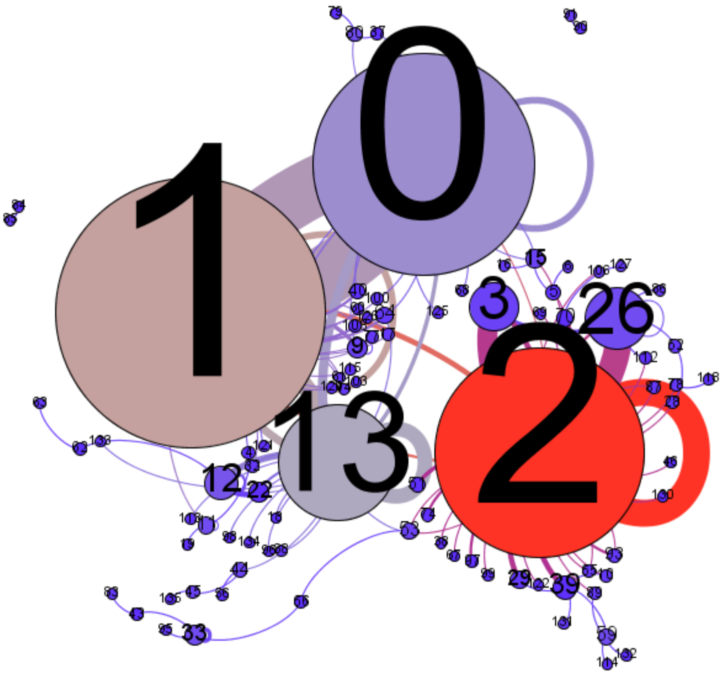

Fig. 5. Connections between user 160's locations - the thicker the edge, the more often user 160 traveled between these two locations.

In general, it is striking that many locations are connected to only one main location or are shared by two locations (the cluster of small dots between Home and Office probably represents places that user 160 typically visits on his commute).

It may have occurred to you that thus far we haven't discussed or mentioned the *real* locations at all. Indeed, the above discussion shows that many observations on a user's whereabouts can be interpreted without considering the exact geographical locations. In order to check to what extent our assumptions most likely hold, we now plot the top five locations on a map [5] - see Figure 6.

User 160 happens to live in Beijing. Location 1 ('Home') and location 0 ('Office') are in the University district - top left. This may also explain why he often seems to work until late in the evening. Location 13 ('Sport club') is located below Chaoyang Park. A closer look via Google Maps indicated that this is Palm Springs Campus, with among others an international club and sports facilities. Location 2 ('Shopping mall') - top right - turns out to be Jingkelong Convenience Store in what appears to be a shopping district.

[5] The map is generated using `http://open.mapquestapi.com/staticmap/`

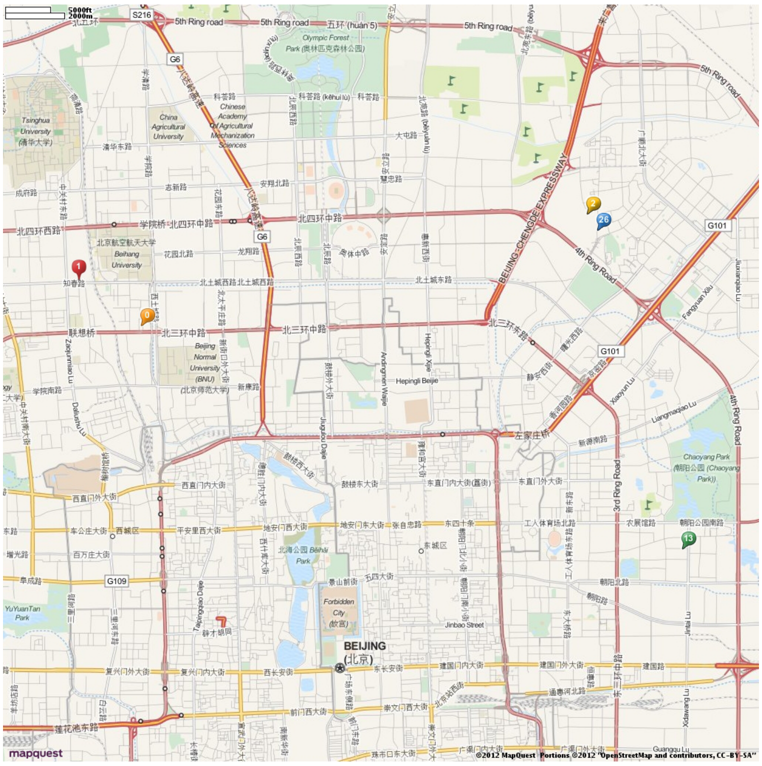

Fig. 6. Map of Beijing with User 160's top-5 locations.

5 Opportunities for Personalization and User Modeling

In the previous section we discussed regularities in mobility, based on GPS data of 178 people. First we focused on daily and weekly patterns for the overall population, and then focused on the specific trajectories of one exemplary user.

In line with [6], we observed that people typically spend most of their time on a small number of locations; most movements are between these locations. In addition we found that these popular locations and most-followed trajectories (e.g. the daily commute) also serve as the starting point for visits to several other locations that form the long tail of the person's whereabouts. The fact that most locations can be connected to one 'base location' or one trajectory can be exploited in various ways, varying from suggestions to navigate to regular stops on the way back home to targeted advertisements at the moment that a user embarks on a Saturday-morning shopping trip. In general, the 'personal vicinity' of locations (which does not necessarily have to match with geographical vicinity, as we have seen for user 160) provides huge potential to improve location-based recommendations by taking into account a user's (or a user group's) habits.

We also investigated the temporal character of human behavior, an aspect that seems not to have received much attention thus far. We found that most people have a relatively regular schedule of moments when they travel from one location to another (e.g. a daily commute on weekdays, fixed activities during the weekends). In this paper, we showed this by focusing on one particular user, but for other users that we investigated, the results were largely similar in character. As we have seen, the daily patterns (during the week and in weekends) and the connections between the locations provide sufficient indicators for an educated guess about the purpose of a particular visit to a location, even without knowing its coordinates. On the one hand, this implies that privacy-preserving location-based personalization techniques can be constructed that do not need exact geographical coordinates as an input. On the other hand, this also implies that when the geographical coordinates are available, it is fairly straightforward to find out where a person lives, works and shops - with modern GPS technology the addresses can be traced down with an accuracy of about 20 meter or less [6].

As discussed before, most location-based services are stateless: recommendations or suggestions are based on the user's current location and do not take a user's daily or weekly patterns into account. A potential that still needs to be explored, is the concept of timely (or delayed) targeted messages or suggestions. This may include suggestions to leave work a bit earlier in order to buy groceries or to pay a visit to a friend, or targeted advertisements and promotions at moments that a user is likely to appreciate them.

In this paper, we did not make actual predictions, but rather showed various promising directions that have thus far remained largely unexploited. To repeat the remark of [9], 'appropriate data-mining algorithms' combined with open map and directory services [7] could be used for this purpose.

References

1. Amini, S., Brush, A., Krumm, J., Teevan, J., Karlson, A.: Trajectory-aware mobile search. In: Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems. pp. 2561–2564. CHI '12, ACM, New York, NY, USA (2012)
2. Bellotti, V., Begole, B., Chi, E., Ducheneaut, N., Fang, J., Isaacs, E., King, T., Newman, M., Partridge, K., Price, B., et al.: Activity-based serendipitous recommendations with the magitti mobile leisure guide. In: Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems. pp. 1157–1166. ACM (2008)
3. Cheng, Z., Caverlee, J., Lee, K.: You are where you tweet: a content-based approach to geo-locating twitter users. In: Proceedings of the 19th ACM international conference on Information and knowledge management. pp. 759–768. CIKM '10, ACM, New York, NY, USA (2010), http://doi.acm.org/10.1145/1871437.1871535
4. Dhar, S., Varshney, U.: Challenges and business models for mobile location-based services and advertising. Commun. ACM 54(5), 121–128 (May 2011)

[6] http://en.wikipedia.org/wiki/Global_Positioning_System
[7] e.g. http://open.mapquestapi.com/nominatim/

5. Golle, P., Partridge, K.: On the anonymity of home/work location pairs. In: Pervasive Computing. pp. 390–397. Springer-Verlag (2009)
6. Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.L.: Understanding individual human mobility patterns. Nature 453(7196), 779–782 (June 2008)
7. Krumm, J.: Inference attacks on location tracks. In: Pervasive Computing. pp. 127–143. PERVASIVE'07, Springer-Verlag, Berlin, Heidelberg (2007)
8. Mokbel, Mohamed, F., Bao, J., Eldawy, A., Levandoski, J.J., Sarwat, M.: Personalization, socialization, and recommendations in location-based services 2.0. In: Proceedings of the PersDB 2001 Workshop (2011)
9. Song, C., Qu, Z., Blumm, N., Barabsi, A.L.: Limits of predictability in human mobility. Science 327(5968), 1018–1021 (2010)
10. Soper, D.: Is human mobility tracking a good idea? Commun. ACM 55(4), 35–37 (Apr 2012), `http://doi.acm.org/10.1145/2133806.2133819`
11. Teevan, J., Karlson, A., Amini, S., Brush, A.J.B., Krumm, J.: Understanding the importance of location, time, and people in mobile local search behavior. In: Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services. pp. 77–80. MobileHCI '11, ACM, New York, NY, USA (2011)
12. Zheng, Y., Zhang, L., Xie, X., Ma, W.Y.: Mining correlation between locations using human location history. In: Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. pp. 472–475. GIS '09, ACM, New York, NY, USA (2009)
13. Zheng, Y., Zhang, L., Xie, X., Ma, W.Y.: Mining interesting locations and travel sequences from gps trajectories. In: Proceedings of the 18th international conference on World wide web. pp. 791–800. WWW '09, ACM, New York, NY, USA (2009)