Visualizing a City Within a City – Mapping Mobility Within a University Campus

Dirk Ahlers$^{(\boxtimes)}$, Kristoffer Gebuhr Aulie, Jeppe Eriksen, and John Krogstie

NTNU – Norwegian University of Science and Technology, Trondheim, Norway
{dirk.ahlers,john.krogstie}@idi.ntnu.no

Abstract. Urban mobility analysis usually examines large cities or even regions. We take another angle and examine a university campus as a city within a city to focus on small-scale and hyperlocal characteristics. The campus mobility data exhibits a high spatial and temporal granularity that we use to drive analyses and visualizations towards the aim of campus analytics. We describe the abstraction approaches and visualizations used towards the development of our tool and share initial results of campus analytics.

Keywords: Visualization · Visual analytics · Mobility · WLAN localization · Object traces · Campus analytics

1 Introduction

The analysis of urban mobility can be used to understand behavior and improve services, transportation, and area use. In this paper, we present a system for visual analytics of people's movements on a large university campus. A campus can be understood as a miniature city embedded in a city. In our case, we model the campus as a combination of multiple large-scale places with indoor localization infrastructure.

The system serves as a proof-of-concept for several building-related analyses in a campus scenario. Current use cases and applications include improved awareness of building usage including use of labs, lecture halls, reading rooms; learning and possible improving routes on campus; improvement of services and their locations; energy savings for a sustainable campus; and research investigating the characteristics of human mobility on campus-like structures. We use the campus as a lab infrastructure to test approaches that we later consider to scale to a city level as part of the smart cities infrastructure. For the presented application, we work together with stakeholders from the university's building management and scheduling offices to finetune application scenarios of building use on campus. Our approach to smart city mobility analysis on campus is to support stakeholders in making sense of the huge amounts of positioning data by reducing and abstracting it in a way that facilitates visual analytics with a *campus analytics* approach.

In the growing interest areas of smart cities, urban computing [13], and mobility analysis, much work has been done on the basis of mobile phone call data records (CDR) or social networks. For example, [10] describe trajectory analysis and show the existence of regularity and similarity of individual's mobility patterns. [5] look at aggregate mobility and transportation patterns with a broad set of methods. These scenarios usually are using individual's traces to gain aggregate knowledge over cities or regions.

We are in a privileged position where we can have access to mobility data at both a high temporal and spatial resolution, thus delivering much finer grained analyses that only times of calls or tweets. Compared to CDRs, WLAN trilateration has a much higher spatial and temporal granularity, but usually a much lower coverage. This makes it viable to look at smaller populations or smaller areas, complementing city-level approaches and instead, aggregate on an intra-city level. In our case, we can only cover a limited area of the city, but in the form of a self-contained campus it is functionally closed and is a high density area of people's movements.

In related work, [2] describe visual analytics tools for transportation patterns on a regional regional scale and [4] follow a visual analytics approach on an urban scale for land use analysis. We are scaling this down to a campus, similar to [9] who analysed a year of mobility data. [12] examines how indoor location influences information needs, based on access to a WLAN network and also to Web traffic over the network inside a mall. We only look at mobility data without traffic content at the scale of a larger set of buildings.

We do not aim to track actual persons, therefore we use device movement as a proxy for people's movement and operate on anonymous data that we aggregate. The smartphone ownership in Norway is above 75 %, and we estimate it even higher in the student and staff population of the university. Additionally, some feature phones also are WLAN-enabled. The multi-device ownership (tablets, notebooks, etc.) rate is also high. This means that we are able to capture a majority of people in the device tracking. At the same time, for this reduction in sampling bias we also get a certain number of cases where we track multiple devices belonging to the same person, thus increasing bias towards multiple device owners by double counting. Because usually people take their devices with them if they move, the proxy assumption holds. Investigating these issues are part of future work combined with deeper accuracy evaluation.

In our approach we use a stacked campus movement abstraction. It comprises the steps of all device positions, extracted movement heatmap, inter-building movement matrix visualization, and specific use-case functionality. The logical steps we take are position gathering, movement extraction, building graph extraction, and visualization. The present paper includes results of two theses [3,7] done in collaboration with NTNU and Wireless Trondheim.

We start by describing the setup of data collection and our data set in Sect. 2, describe our visual analytics approach and data processing in Sect. 3, present results in Sect. 4, and discuss limitations and future work in Sect. 5 before closing in Sect. 6.

2   Data Set

NTNU together with Wireless Trondheim and Mazemap is employing a WLAN infrastructure in the city center and the campuses, which is set up as a living lab [1]. In this work, we are focusing on Gløshaugen, the biggest campus of NTNU. Existing work on this infrastructure for example enables campus wayfinding [6] or an estimation of available reading rooms for students [8].

The WLAN infrastructure at NTNU employs a Cisco MSE (mobility services engine) controller with a location engine that derives device positions by trilateration over the access points [11]. The coverage is based on the roughly 1800 access points on 350000 sqm of campus. The tracking is based on passive location sensing, i.e., any device with enabled WLAN can be tracked, by means of its probe requests. These are transmitted by the devices at varying intervals, which sets the sampling rate usually to less then 60 s. Thus stickiness is a lesser issue as even when the device is not yet connected to a closer access point, it will already be sensed. The system is setup to provide coverage mainly indoors, but also covers outdoor areas close to buildings. Therefore it captures most indoor but also some outdoor movement.

An abstraction and processing layer is deployed on top of this. The data we can gather is already processed/abstracted and anonymized, and we do not have access to raw data. Other data that would be available on the backend is stripped out or otherwise unavailable, for example the device type or model as well as any traffic over the WLAN network. The output we can access contains data points that consist of an anonymized device ID, coordinate pair of latitude and longitude, a hierarchical description of the position, a timestamp, accuracy measure, and a salt timestamp. The anonymization uses a hashing function on the MAC addresses of devices to generate a random ID. This changes regularly by changing the salt of the hash. It thus ensures that a single device cannot be tracked over an extended period of more than one day. For this reason the salt timestamp is included in the data record. As long as the timestamp has not changed, an ID continuously identifies the same device. For internal university use of the data, this process was approved by the The Norwegian Data Protection Authority. We also feel that privacy is important especially for such high-granularity data. Our approach thus uses not individual, but general patterns.

The hierarchy description contains three items: campus name, building name, and floor, for example *Gloshaugen > IT-Vest > 1. etasje* or *Gloshaugen > Sentralbygg II > 13. etasje*. Further details such as mapping to individual rooms are not included in the data, but can be map-matched later [6,8]. The accuracy gives the estimated radius of error for each measurement. The average indoor accuracy is about 5–10 m, outdoors this can be higher as positions are estimated close to the buildings. For the visualization, we already have tailored digital maps, building models, and positions available [6].

For initial data collection and partition, we use a fixed window of one day starting at midnight. Compared to other city scenarios, this is sufficient for the campus area, where there is usually very little activity at night. We are currently using a collection of historic data, but the system could easily be adapted to

handle live data. We define a trace as a sequential list of positions for a single device that contains at least two positions and is delimited either by the end of day or by a 1-h timeout with no further data points. There are about 43000 devices per day with 3.2 million position measurements available for the campus with a captured historical dataset of two months.

3   Approach for Visual Analytics

Our approach uses visual analytics of the position data of the devices on campus, with gradually refined data processing steps. We start with simple device positions, then animate them over time, then derive movement data and finally move to a graph-based analysis.

The first iteration is an animation of density heatmaps over all data. It shows changes in the device density, which could be interpreted as movement, but is difficult to analyze. Of course, areas filling up or emptying out can easily be identified. But for example, slight fluctuations between densities in neighboring buildings could mean few movements or larger reciprocal movements happening between these two buildings. This means there is a need to extract the movement as a derivative and possibly also make out the trajectories of movements. As we do not only have point clouds, but stable randomized IDs over a certain period, we can extract device traces.

In a heat map visualization of movement data, trajectories are abstracted in a non-directional heat map. In a graph-based approach, directional inter-building traffic can be analyzed. This manages to integrate two complementary geospatial dimensions. First, it uses plain geospatial data in the form of coordinates, second, it uses a conceptual view by abstracting and mapping positions to buildings and generate a building graph. The visualization of movement instead of raw position achieves two things. First, it obviously filters out those devices that do not move. Second, it allows to see where the movement takes place and at what magnitude (as an aggregated temporal view). The standard heatmap can make it difficult to distinguish between static and moving devices if they do not move far beyond the kernel bandwidth or radius. Thus, this shows the areas where movement takes place, and together with a temporal dimension can derive the direction. For more abstract analysis that is difficult to visualize as heatmaps, we chose an abstraction to buildings as functional areas with a graph-based approach.

3.1   Extraction of Movement

When at rest, devices will be located at similar positions because the trilateration is done via probe requests to any base station picking up the signal. Detected movement correlates highly with actual device movement. But on the individual level, there may be slight deviations during movement about the exact position. Yet, the indoor accuracy is usually within 5–10 m.

To abstract the data and filter out jitters in the positions, we apply a simple track smoothing with a distance filter and then consider a device in movement

Fig. 1. Heatmap of all device positions on campus

only if its position differs by a determined threshold of 10 m from its previous position. Data points matching these conditions are then considered as in movement and are visualized. This can be used to show the movement and intensity in different time periods.

3.2 Extraction of Inter-Building Movement

The next step is to go beyond positional movement data and look at trajectories on campus. We chose to abstract movement to a building level, or rather, to movement between buildings. To this end, we build an inter-building movement matrix. Movement between floors was not yet considered. Device traces between buildings are counted to build a building movement graph.

As we already have the building names in the position data, we do not need a map matching or clustering approach. Thus the building extraction is rather trivial. In the WLAN setup, a device does not actually have to enter a building to be registered or even to connect. This is of course desirable to deliver at least partial outdoor coverage. For example, if a person walks along the outside of a building, the device can still be registered by the access points within, thus registering the device to that building. We do no further mapping or snap-to-building, so we may capture some outdoor positions as well.

To model device movement between buildings, we have experimentally implemented two approaches. For the 'all-movement' approach, all movement in the location traces between buildings is considered. This means that whenever we see a device in a building A and later in its location trace in another building B, we count one movement from A to B in the building movement matrix. Each device's location trace is iterated to capture all these events. The most obvious limitation of this approach is that it will cut apart longer trajectories between buildings of there are other buildings on or along the way. Thus, a movement does not mean that the registered destination of a movement is actually the destination of the user moving around campus. It may just lie on the way and be either bypassed or passed through to the true destination. This is especially the case for centrally located building on campus that have a lot of through traffic.

To counteract this and filter out buildings that were only passed on the way, we attempt to identify movements with a defined origin and destination. The improved 'start-stop' approach delivers better results towards the extraction of the true user-intended destination of a movement. The approach does not currently consider movements outside the system boundary, for example, users entering or leaving the campus. Again, each location trace is iterated. While a group of three subsequent positions are at rest, a halt is assumed. After a halt, when three subsequent positions have a distance to each other of more than 10 m, they are considered in movement and a movement origin is registered in the current building. When a halted stated is detected in the further iteration of the trace, this is registered as the destination and a movement between the two buildings is generated.

4   Results

We share our insights into the use of these campus mobility visualizations. We are able to generate heatmaps with very high spatial resolution that allow to analyze movement even in small indoor structures. Mapping all devices as in Fig. 1 shows the full amount of position data for a morning. It can be zoomed in to enable drill-down, and the heat map parameters can be adapted to change the visualization. The mapped positions follow largely the buildings on campus and their expected traffic densities. In the static view, there is limited information to be gained, but we can animate this map over time to show changing densities, which makes this already a useful tool.

However, as noted above, it is difficult to make out movements in this mapping, as there is a lot of ambiguous fluctuation. In the following we take the next step of our approach and use only extracted movement.

If we use the visualization of extracted movement to move through a typical day, the first thing to note is that there is a reduction in density of the data points which allows a more focused view of only movement. As an example, we show the evolution of the movement on campus for one morning. We use only selected time slices to show the most distinctive visualizations. We can see daily patterns revolving around campus structures and lectures. From 8:00–9:00 there
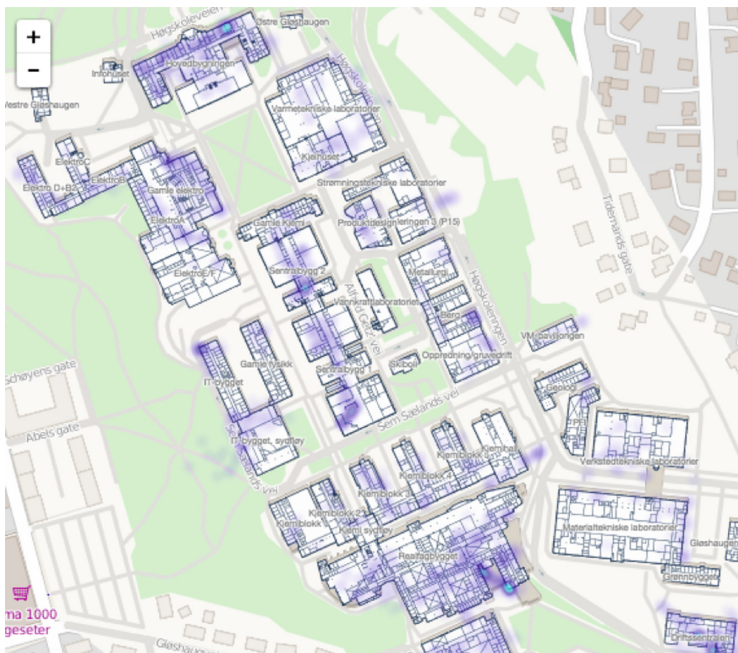
Fig. 2. Evolution of a morning: movements 8:00–9:00
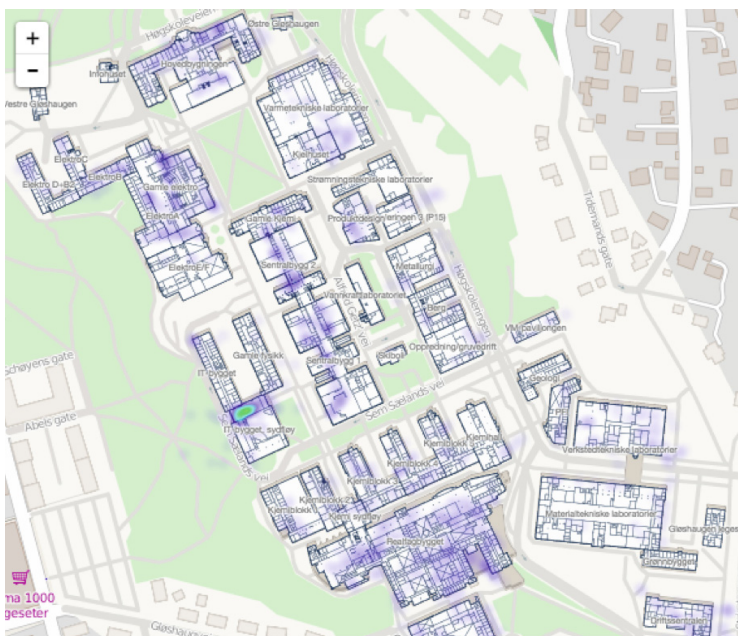
Fig. 3. Evolution of a morning: movements 9:45–10:00
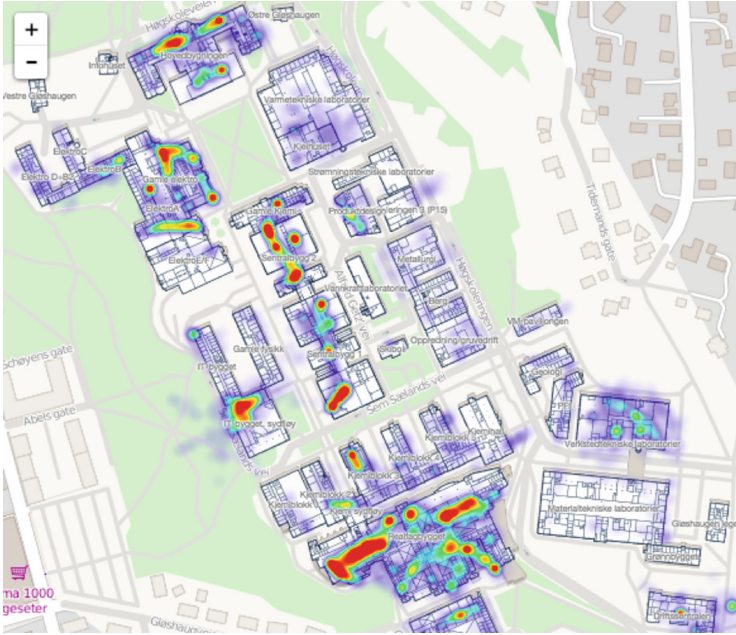
Fig. 4. Evolution of a morning: movements 11:00–12:00

is little movement overall when the first lectures of the day start (Fig. 2); at 9:45–10:00 there is a slight increase due to the end and the start of the second lectures (Fig. 3) and at 11:00–12:00 the campus really comes to life (Fig. 4). There is one obvious pattern that the early morning is not a preferred time for users to arrive on campus. Lunchtime sees highly increased movement as people move from lecture halls or labs to the canteens.

For the time being, this visualization cannot yet fully distinguish between a lot of local movement and longer trajectories, however, the former will halt once in a while while the latter continues stronger and thus delivers more movement density.

For a more detailed view into the longer trajectories, we can switch to the building graph view. In the default state, it shows the contribution of each building to the overall movement by the size of the overlaid bubble plot as in Fig. 5 and provides numerical data on hovering.

Furthermore, the underlying origin-destination building matrix can be explored for each building as shown in Fig. 6, which shows the movement from or to all other buildings, in this case limited to the top-5 connected nodes which again show their weight when hovering. This can also be viewed as the raw matrix and further processed in other tools. We provide a sample of the matrix for the top-k buildings regarding traffic count in Table 1. One result that can be derived for example is that movement is correlated with building size, i.e., there is more movement inside and to/from larger buildings. Another quite interesting

Fig. 5. Visualization of building traffic embedded into the application with a selection of interaction and analysis tools
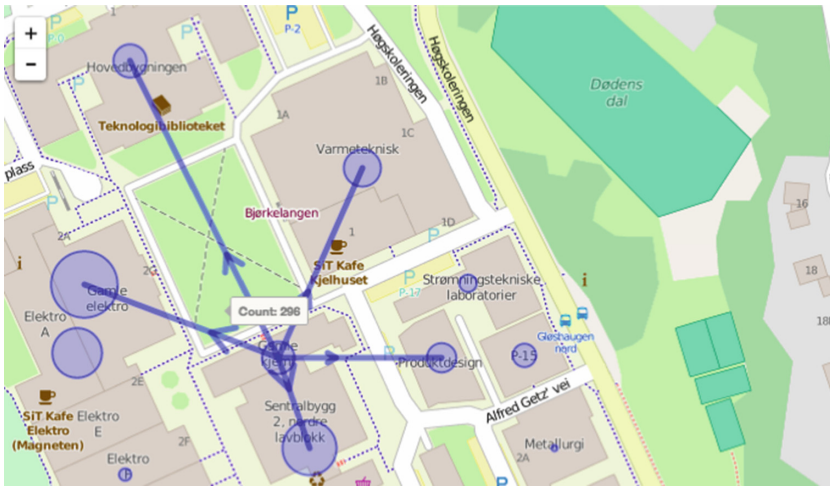
Fig. 6. Visualization of the origin-destination building matrix: top-5 destinations of traffic flow from the Gamle Kjemi building

finding is that there seems to be a lot of reciprocal movement between buildings. The matrix is not fully symmetric, but the numbers of ingoing and outgoing trajectories are strongly correlated over different periods during the day. This is surprising because it means that buildings have the same average usage over the day and that also the initial movement of people onto the campus (which we do not capture in the matrix) is well distributed over all buildings. What we of course also see is that some buildings act as hubs for movement, either by their central location or by housing a canteen.

Table 1. Selection from the building matrix of the top 8 building traffic counts for a 4 h period

From \ To	Gamle Elektro	Realfagbygget	Sentralbygg I	Sentralbygg II	Elektroblokk A	Kjemi 3	Gamle Fysikk	Driftssentralen
Gamle Elektro	0	36	33	153	1308	24	220	4
Realfagbygget	46	0	255	85	22	660	152	402
Sentralbygg I	35	165	0	1083	12	255	321	3
Sentralbygg II	83	80	1031	0	21	27	215	0
Elektroblokk A	1245	7	16	38	0	6	83	3
Kjemi 3	23	800	316	26	7	0	209	3
Gamle Fysikk	234	187	270	157	80	284	0	5
Driftssentralen	1	482	7	8	3	7	3	0

Switching between the two modes of inter-building detection, it is also possible to distinguish those buildings that are mostly passed by and whose WLAN contributes strongly to the outdoor coverage. These disproportionately lose traffic when switching to the start-stop method.

5 Limitations and Future Work

As discussed before, the access points are only indoors, so there is no full view of the campus, but outdoor is covered close to buildings. However, the accuracy of the localization is much higher indoors than outdoors and often outdoor positions get localised indoors. A classification of these cases is not directly obvious from the data, but may be explored in future work. An interesting aspect is that the system can be used to detect outdoor events based on the detected (indoor) positions. Figure 7 shows the movement during a career day for students. The event took place in a large industrial tent on the large square in the North of the campus. No positions were located on the square, but there is much higher movement in all neighboring buildings as compared to the normal day in Fig. 4.

Issues of data sparsity for outdoor areas and some erroneous attributions will need to be improved when looking at higher granularity estimates, for example at room level and to better separate functional areas. Finding the entrance points into buildings and including movements crossing the system boundary

Fig. 7. Influence of (outdoor) events on indoor detected mobility: movements 11:00–12:00 of a student event in a temporary industrial tent on the square in the Northeastern part of the campus

(when people enter or leave the campus) will help in a detailed understanding of movement and transportation. We also aim to more fully utilize the height information of floors inside buildings to improve the mobility models towards three dimensions.

Another point is the improvement of the movement extraction and classification and the use of more complex data mining approaches to drive the visualizations. For added accuracy of the origin-destination matrix, we plan several improvements on the movement extraction. For example, the assumed pattern underlying the approach does not always hold and may not always be representative of actual movements. It filters out well those buildings that are merely bypassed, but when users go through a building, they may not always exhibit a consistent movement with stopping at doors or other bottlenecks, reading bulletins or meeting friends. Incidentally, the identification of bottlenecks inside buildings is another application scenario that we are working on. Additionally, considering different timings or inconsistencies of probe requests and devices that go 'dark' intermediately should also improve the measure. Finally, we will develop a stronger focus on the temporal aspect by providing additional analysis beyond time selection or animation.

6   Conclusion

We have presented our approach of campus analytics to use WLAN localization to derive movement patterns on a university campus. We have shown that the localization is sufficiently accurate to build this type of application and employ smart city analyses also on this smaller scale. Having mobility data available at very high spatial and temporal granularity makes this a very rich data source. Results show that both heat map and building graph mobility visualization can provide value to stakeholders and that patterns of movement can be identified. Further development of the NTNU campus visualization application will improve the developed methods and integrate additional modes of analysis towards using the campus analytics as a contribution for smart cities.

References

1. Andresen, S.H., Krogstie, J., Jelle, T.: Lab and research activities at wireless trondheim. In: 4th ISWCS (2007)
2. Andrienko, G., Andrienko, N., Wrobel, S.: Visual analytics tools for analysis of movement data. SIGKDD Explor. Newsl. **9**(2), 38–46 (2007)
3. Aulie, K.G.: Human Mobility Patterns from Indoor Positioning Systems. Master's thesis, Norwegian University of Science and Technology, Trondheim, Norway (2015)
4. Bak, P., Omer, I., Schreck, T.: Visual analytics of urban environments using high-resolution geographic data. In: Painho, M., Santos, M.Y., Pundt, H. (eds.) Geospatial Thinking. LNGC, vol. 0, pp. 25–42. Springer, Heidelberg (2010)
5. Becker, R., Cáceres, R., Hanson, K., Isaacman, S., Loh, J.M., Martonosi, M., Rowland, J., Urbanek, S., Varshavsky, A., Volinsky, C.: Human mobility characterization from cellular network data. Commun. ACM **56**(1), 74–82 (2013)
6. Biczok, G., Diez Martinez, S., Jelle, T., Krogstie, J.: Navigating MazeMap: indoor human mobility, spatio-logical ties and future potential. In: PerMoby 2014 (2014)
7. Eriksen, J.B.: Visualization of Crowds from Indoor Positioning Data. Master's thesis, Norwegian University of Science and Technology, Trondheim, Norway (2015)
8. Gao, S., Krogstie, J., Thingstad, T., Tran, H.: A mobile service using anonymous location-based data: finding reading rooms. Int. J. Inf. Learn. Technol. **32**(1), 32–44 (2015)
9. Ghosh, J., Beal, M.J., Ngo, H.Q., Qiao, C.: On profiling mobility and predicting locations of wireless users. In: REALMAN 2006, pp. 55–62. ACM (2006)
10. Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.L.: Understanding individual human mobility patterns. Nature **453**(7196), 779–782 (2008)
11. Little, J., O'Brien, B.: A Technical Review of Cisco's Wi-Fi-Based Location Analytics. Technical report, Cisco (2013)
12. Ren, Y., Tomko, M., Ong, K., Bai, Y.B., Sanderson, M.: The influence of indoor spatial context on user information behaviours. In: Workshop on Information Access in Smart Cities. ECIR 2014 (2014)
13. Zheng, Y., Capra, L., Wolfson, O., Yang, H.: Urban computing: concepts, methodologies, and applications. ACM Trans. Intell. Syst. Technol. **5**(3), 38:1–38:55 (2014)