Pattern mining in tourist attraction visits through association rule learning on Bluetooth tracking data: a case study of Ghent, Belgium

Abstract

The rapid evolution of information and positioning technologies, and their increasing adoption in tourism management practices allows for new and challenging research avenues. This paper presents an empirical case study on the mining of association rules in tourist attraction visits, registered for 15 days by the Bluetooth tracking methodology. This way, this paper aims to be a methodological contribution to the field of spatiotemporal tourism behavior research by demonstrating the potential of ad-hoc sensing networks in the non-participatory measurement of small-scale movements. An extensive filtering procedure is followed by an exploratory analysis, analyzing the discovered associations for different visitor segments and additionally visualizing them in 'visit pattern maps'. Despite the limited duration of the tracking period, we were able to discover interesting associations and further identified a tendency of visitors to rarely combine visits in the center with visits outside of the city center. We conclude by discussing both the potential of the employed methodology as well as its further issues.

1. Introduction

Movement represents a key aspect of tourism, both in order to reach a certain destination from an individual's habitual environment and to move around within that tourist destination. As a consequence, many research efforts have focused on the spatiotemporal behavior of tourists in order to *inter alia* optimize tourist infrastructure, for marketing incentives, and to better manage the impacts of tourist mobility on the environment (Shoval & Isaacson, 2009). Due to the complex nature of tourism (McKercher, 1999), there is a growing need for empirical movement data to accompany theoretical models. Yet, empirical studies into tourist mobility have traditionally been rather scarce due to the labor-intensive and often expensive nature of traditional methods such as direct observation (Hartmann, 1988) or personal interviews (Kemperman, Borgers, & Timmermans, 2009). Space-time diaries (Connell & Page, 2008; Janelle, Goodchild, & Klinkenberg, 1988; Lau & McKercher, 2006) shift some of the weight away from the researchers but are often characterized by a low reliability as respondents tend to forget or neglect to register certain activities. Recently, however, tracking technologies offer a more scalable and objective way to capture spatiotemporal behavior in a detailed way (Shoval & Isaacson, 2009). The use of global navigation satellite systems – such as GPS – is currently the dominant approach and its adoption in tourism research through the distribution of logging devices is well-documented (Shoval & Isaacson, 2007a, 2007b; Shoval, McKercher, Ng, & Birenboim, 2011; Tchetchik, Fleischer, & Shoval, 2009). An alternative approach is to track the movement of mobile phones through a cell tower network without the direct participation of the phone's owner (González, Hidalgo, & Barabási, 2008; Ratti, Pulselli, Williams, & Frenchman, 2006). Particularly in Estonia, this method has already been extensively used for studying regional movement patterns of tourists (Ahas, Aasa, Mark, Pae, & Kull, 2007; Ahas, Aasa, Roose, Mark, & Silm, 2008).

Despite the undeniably important contribution of both tracking methodologies to the research field, we argue that both approaches have certain limits. The distribution of logging devices necessitates the direct collaboration of the tracked individual. This makes it hard to scale up the methodology to large groups of individuals. Additionally, any participatory methodology presents a risk for self-selection bias where individuals with certain characteristics would show a higher degree of cooperation and thus be overrepresented in the sample. While the use of smartphone apps for tracking tourist movements – e.g. through shared user-generated content such as pictures (Jankowski, Andrienko, Andrienko, & Kisilevich, 2010) – may decrease some of the intrusive nature in comparison with the use of logging devices, it still represents a participatory methodology with an inherent risk for bias and data sparseness in some locations. Finally, GPS technology is not applicable to indoor contexts. Cell phone tracking, on the other hand, encompasses other limitations. First, the

spatial accuracy of the method is limited by the density of cell towers over the study area. In Estonia, for example, around 50% of measurements were correct to within only 400 meters in urban areas and only 2600 meters in rural areas (Ahas, Laineste, Aasa, & Mark, 2007). While this does not hinder the study of regional movements, it does pose a problem when studying movement within a certain tourist destination (e.g. a city). Second, these datasets are property of mobile operators and – as such – not freely available. In summary, it seems that small-scale spatiotemporal behavior cannot be measured without the direct involvement of the individual to be tracked. This hinders studying larger groups of individuals.

A recent alternative in the non-participatory tracking of mobile phones is the use of ad-hoc sensor networks distributed over a study area. Bluetooth technology, for example, has already been employed for studying pedestrian flows at mass events (Delafontaine, Versichele, Neutens, & Van de Weghe, 2012; Stange, Liebig, Hecker, Andrienko, & Andrienko, 2011; Versichele, Neutens, Delafontaine, & Van de Weghe, 2012; Versichele, Neutens, Goudeseune, Van Bossche, & Van de Weghe, 2012) and in social studies (Eagle & Pentland, 2005). WiFi (Bonné, Barzan, Quax, & Lamotte, 2013) and RFID (Öztayşi, Baysan, & Akpinar, 2009) technology provide similar possibilities. Due to the limited coverage of each sensor, a careful deployment of sensors may thus provide movement records with a granularity that is much smaller than the accuracy level of cell phone tracking data. By deploying sensors with these wireless technologies at a set of pre-defined tourist locations, one is able to study the spatiotemporal behavior at and between these locations. Despite this potential in the non-participatory registration of small-scale movements, we have as yet no indication of the application of the methodology for tourism management purposes.

This paper aims to address this issue by presenting a case study where visitors to tourist attractions in Ghent, Belgium were registered through an ad-hoc Bluetooth sensor network. Due to the novelty of Bluetooth technology – and the use of ad-hoc sensing networks in general for that matter – we will not only elaborate extensively on the working principle of the methodology, but also on the analytical potential of such tracking data. Ad-hoc sensor network data lack the typical socio-demographic or psychographic variables used as explanatory factors in various studies related to tourism behavior. In contrast with hypothesis testing procedures, sensor network data often need to be investigated without any a priori assumptions. The collection of such methods that can be used to discover (non-trivial) patterns and knowledge from large data sets is called data mining (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). Several data mining techniques have already been frequently applied to tourism data, including *regression* techniques (Song & Li, 2008; Witt & Witt, 1995), *clustering* (Bloom, 2005; Cini, Leone, & Passafaro, 2010; Dolničar & Leisch, 2003; Dolničar, 2004; Tchetchik et al., 2009), *sequential pattern mining* (Orellana, Bregt, Ligtenberg, & Wachowicz, 2012; Shoval & Isaacson, 2007a), and *classification* (Law & Au, 2000; Law, Bauer, Weber, & Tse, 2006). *Association rule learning* is concerned with discovering associations between variables without fixing the output variable, as is the case in classification. In comparison with the other techniques, implementations of association rule learning in tourism research are rather scarce. Documented applications found in literature include tourism product development (Al-Salim, 2008; Liao, Chen, & Deng, 2010), domestic tourist profiling (Emel, Taskin, & Akat, 2007), sharers and browsers of touristic websites (Rong, Vu, Law, & Li, 2012), and change and trend identification in Hong Kong outbound tourism (Law, Rong, Vu, Li, & Lee, 2011).

This paper aims to be a methodological contribution to the field of spatiotemporal tourism behaviour research by demonstrating the potential of ad-hoc sensing networks in the non-participatory measurement of small-scale movements. We describe a case study where visitors to 14 tourist attractions were registered through Bluetooth technology sensing the mobile devices they were carrying around. In an attempt to investigate the analytical potential of the resulting data, we employ an association rule learning algorithm to mine for 'interesting' patterns in the combinations of visits to different attractions (in the sense that they represent potentially valuable information which would be hard to discover in another more straightforward way). As the tracking data are completely anonymous, it is impossible to directly distinguish between local visitors and actual tourists as defined by the World Tourism Organization: people "traveling to and staying in places outside their usual environment for not more than one consecutive year for leisure, business and other purposes" (World Tourism Organization, 1995). By deploying sensors in 14 hotels, however, some visitors will be identified as hotel guests therefore giving a strong suggestion that they are indeed tourists. Extra context is added by tracking visitors at the tourist inquiry desk as well. Combining the tracking data with these contextual assumptions, we will investigate patterns for different visitor *segments* (e.g. those that were only detected on one day, those that were identified as hotel guests, etc.). For the sake of clarity, we will always use the term *visitors* instead of further labeling them as tourists.

The remainder of the paper is organized as follows. In section 2, we first discuss the Bluetooth tracking methodology and its specific implementation in the case study (2.1). Next, we describe association rule learning in more detail (2.2) and how the information it generates can be summarized in *visit pattern maps* (2.3). Section 3 outlines the filtering of the raw tracking data in detail, and section 4 presents a first data exploration. The actual association rule mining is performed for the different visitor segments in section 5. We finish with a discussion and conclusion (6).

2. Methods and data

2.1. Bluetooth tracking

For this study, scanners with Bluetooth sensors were deployed at 29 locations in and around the historical center and the 'arts quarter' of Ghent (Belgium) for 15 days in May of 2012. Ghent was chosen as the study area because of its rather unique touristic character: despite its wealth of attractions and historical significance, it was once described as "Belgium's best kept secret" (Lonely Planet, 2011) due to the nearby presence of better known destinations such as Bruges. As a result, it attracts a more diverse (and probably less predictable) audience, which makes it an intriguing yet challenging test bed for the suggested approach. Additionally, some of its attractions are located at considerable distances from the historical center and the tourism department was very receptive to any methodology which could offer additional insights in the visiting behavior over the entire city.

An overview of the study area and the sensor locations is given in Figure 1. The full names of the different venues are shown in Figure 4. The locations consist of fourteen hotels (a-n), three open (1-3) and eleven of the most visited closed (4-14) tourist attractions, and the inquiry desk for tourists. We make the distinction between open and closed attractions based on the need for visitors to either buy a ticket or register. The open (i.e. no registration required) attractions consisted of a cathedral, a church and an indoor market. All closed attractions were museums, covering a wide range of interests such as classic/modern arts, history, textiles, and the former 'Castle of the Counts'. In 2012,

4

these fourteen selected attractions were responsible for around 76% of the total number of visits to all attractions in Ghent. The hotels comprised the entire range of common classes and price-ranges: one hotel without stars (a), one *- (b), two **- (c-d), four ***- (e-h), four ****-hotels (i-l), and two hostels (m-n). Together, these fourteen hotels contained 67% of the total number of available beds in the city.

The Bluetooth scanners continuously searched for discoverable Bluetooth devices within their detection range, and registered the MAC address and COD (class-of-device) code of each detected device together with the detection timestamp. The MAC address acts as a unique identifier of the detected device, and can be used to link different detections (at different locations) to the same device and thus generate trajectories. The COD code can be used to deduce the type of device (phone, car kit, mp3-player, etc.). More details on the Bluetooth tracking methodology and the deployed hardware can be found in a previous study (Versichele, Neutens, Delafontaine, et al., 2012). All Bluetooth sensors used were Class 2 devices, which have a theoretical communication range of around 10 meters according to the official Bluetooth specifications. The actual detection range of Bluetooth sensors, however, largely depends on the environment and the presence/absence of a line-of-sight between the sensor and the detected device. As such, the exact location of each sensor was chosen by balancing the need for an optimal position (in order to only detect devices inside the attraction or hotel) and the sake of convenience (i.e. the presence of a power supply). All Bluetooth scanners were connected to the internet (18 wired, 11 wireless) in order to facilitate the remote monitoring of their correct operation.

As a first generalization of the detection data, each scanner concurrently generated a compressed log format where successive detections of the same device within 10.24 s of each other were compressed into detection intervals. This duration corresponds to the standard Bluetooth inquiry time (Peterson, Baldwin, & Kharoufeh, 2006). The difference between the detections and the resulting detection intervals is illustrated in Figure 2. The resulting dataset, hereafter referred to as the raw data, consisted of 17,496 Bluetooth devices being detected over 235,597 time intervals over all locations. As tracked individuals were not approached, no additional socio-demographic or other variables were present in the dataset. The owners of the detected devices thus remain completely anonymous, and were in fact not aware of being part of a scientific study. In previous experiments, we observed that around 8% of a general public is traceable through a detectable Bluetooth device with the class 'phone' by comparing manual head counts with tracking data at the same location over periods of five to fifteen minutes. We will use this figure to provide a rough approximation of the number of detected individuals based on the number of detected phones. We continue this chapter by describing the association rule learning method.

2.2. Association rule learning

Association rule learning represents a popular data mining method for discovering interesting relationships between variables in large databases. Adhering to the original definition (Agrawal, Imieliński, & Swami, 1993), an association rule can be defined as $X \Longrightarrow Y$ with $X, Y \subseteq I$ and $X \cap Y \neq \emptyset$. The itemsets X and Y are called antecedent and consequent respectively. The total itemset I in this study consists of the fourteen attractions: $I = \{1, 2, \ldots, 14\}$. The database of transactions can be formalized as $D = \{t_1, t_2, \ldots, t_n\}$ with each transaction $t_i \subseteq I$. Note that the general notion of a transaction is borrowed from the domain of market basket analysis (Chen, Tang, Shen, & Hu, 2005), but consists of an unordered set of locations visited by a Bluetooth device i. The

time ordering of visits is thus ignored, which distinguishes the method from sequential pattern mining. The rules are generated by the *Apriori* algorithm (Agrawal & Srikant, 1994) with the *arules* package (Hahsler, Chelluboina, Hornik, & Buchta, 2011) in the R (2.14.0) statistical environment.

Three measures are used to compare rules: support, confidence and lift. The *support* of a rule is a measure of the share of tracked individuals to which the rule applies: $s(X \Longrightarrow Y) = s(X \cup Y)$, with the support of an itemset Z (in this case $Z = X \cup Y$) defined as the proportion of transactions in the dataset which contain the itemset: $s(Z) = \#\{t_i \in D : Z \subseteq t_i\}/\#D$. The *confidence* of a rule is a measure of the probability of its consequent given its antecedent: $c(X \Longrightarrow Y) = s(X \cup Y)/s(X)$. The *lift* of a rule is a measure of its support compared with the support that can be expected if X and Y were independent: $l(X \Longrightarrow Y) = s(X \cup Y)/(s(X) \times s(Y))$. As this measure indicates whether a rule's support is lower, similar or higher than would be expected if X and Y are assumed independent, it is often used as the primary measure for the interestingness of a rule. In other words, rules with a higher lift indicate a stronger association between antecedent and consequent than what could be predicted based on the frequency of the items separately and are thus potentially more informative and valuable. In order to further clarify these concepts, Table 1 shows an illustrative example of transactions (visitors) being constituted of visits to four of the main tourist attractions in Paris. Say, the following rule is generated: $\{Louvre\} \Longrightarrow \{Arc\ de\ Triomphe, Notre\ Dame\}$. The support of this rule would be the share of visitors that visited all three attractions: $s = 2/5 = 0.4$. Its confidence would be its support divided by the support of its antecedent: $c = (2/5)/(3/5) = 2/3 \cong 0.67$. The lift would be calculated as follows: $l = (2/5)/((3/5) \times (3/5)) \cong 1.11$. To limit the number of generated rules and enhance interpretability, the following constraints were used for the *Apriori* algorithm: $s \geq 5/\#D$, $c \geq 0.05$ and $\#X + \#Y \geq 1$ (in order to filter out rules with an empty antecedent). This way, rules needed to be supported by at least 5 tracked individuals or roughly 60 individuals taking the detection ratio of 8% into account. We finish this chapter with more details on how we will present the information extracted from this data mining method.

2.3. Visit pattern maps

A correct dissemination of patterns or knowledge discovered through data mining is essential. The output of this specific case study should be tailored for all stakeholders involved in the tourism management of the study area. Association rule learning methods are known to generate large amounts of rules, and the selection of those rules with a higher relevance to the research question is a non-trivial task. Several approaches in visualizing association rules, in contrast with the classical tabular representation, have already been documented. These include the use of scatter plots and matrix-based visualizations (Hahsler & Chelluboina, 2010), graph-based representations (Appice & Buono, 2005), parallel coordinate plots (Bruzzese & Davino, 2003; Yang, 2005), 3D volumes (Compieta, Di Martino, Bertolotto, Ferrucci, & Kechadi, 2007), or others (Techapichetvanich & Datta, 2005). As a way of summarizing the gathered knowledge on tourist attraction visits of a specific segment of individuals, we introduce an alternative approach called a 'visit pattern map'. This map is a geographical depiction combining two types of information: the spatial distribution of visits over the study area, and the association (combination) of visits to different attractions. The spatial distribution of visits is visualized by proportionally sized circles showing the share of tracked individuals that visited each attraction. The association between the different attractions is visualized by means of lines connecting different attractions. We believe that a geographical depiction of association rules will enhance the interpretability, in contrast with the traditional tabular fashion of

representation. In order to avoid cluttering, we only visualize rules with single-item antecedents. This way, rules are only associated with one item in both the antecedent and consequent. Rules can now be represented by a single line connecting two attractions at their geographical location. Although it could be possible to include all three indicators in the visualization, we opt to neglect the confidence as it is known to be biased by frequent items in the consequent (Tan, Steinbach, & Kumar, 2005). Including it would additionally clutter the visualization as it is the only measure which is not symmetrical for the case of two-item rules ($s(a \rightarrow b) = s(b \rightarrow a), l(a \rightarrow b) = l(b \rightarrow a), c(a \rightarrow b) \neq c(b \rightarrow a)$). The support of a rule is linked to the width of the line, the lift is represented by a discontinuous color scale. Rules with higher lift values are plotted after (above) rules with lower lift values, making the former easier to identify. The visit pattern maps generated for this dataset are shown in Figure 7 and Figure 8.

3. Filtering

As mentioned above, nothing was known on the individuals carrying the detected Bluetooth devices. Before analyzing the dataset for associations between the different attractions, we needed to make sure we know whether devices detected at a certain attraction represented actual visitors or individuals that merely passed the location (either because of a sensor with a detection range that was too large, because these individuals physically approached a registration desk but only for information purposes, or in the case of staff). Analogously, we needed to distinguish between actual hotel guests and detected individuals with other purposes (hotel staff, restaurant guests, convention attendees, browsers, etc.). In order to make this distinction, we applied a progressive filtering process on the set of detected Bluetooth devices at each location. The filtering was based on a combination of three parameters: the type of device (accessible through the COD code), the *duration of visit* (d_v) to a location and the *duration of presence* (d_p) at a location. By taking the device type into account, we could filter on 'phones' and remove all other classes that do not represent a moving individual including carkits, printers, etc. The duration of visit was calculated as the time difference between the very first and very last detection of a device at the corresponding location. In contrast, the duration of presence is the duration that a device was actually detected by the sensor of that location (after merging of co-located detection intervals, i.e. intervals that were less than one minute apart). This way, a device that was detected at a location from 20:35 until 20:42 and from 08:32 until 08:35 the next morning would have a duration of visit d_v of 12 hours and a duration of presence d_p of 10 minutes. The concepts of d_v and d_p are further illustrated in Figure 2.

For most locations, we received actual visitor/guest counts for the 15-day tracking period (n_v). By taking the detection ratio of 8% (ρ) into account, we could estimate the number of visitors that should have been tracked at each location as such: $n_v' = n_v \times (\rho/100)$. Where available, we could compare these figures with the actual number of detected devices (n_d) that remained after each successive filtering step. The ratio $\delta = n_d/n_v'$ could then act as a reference for stating when the filtering process had reached an appropriate end point ($\delta \rightarrow 1$). The limits imposed on d_v and d_p were chosen by heuristic common sense linked to the type of venue. As such, we describe the filtering for the group of hotels, open attractions and inquiry desk, and closed attractions separately. Figure 3 represents a graphical overview of the filtering process. In the remainder of the paper, all visitor counts will refer to detected visitors (n_d) unless otherwise stated.

As can be seen in Figure 3a, the number of unfiltered devices detected at most hotels significantly exceeded estimates based on guest counts (with hotels e, i and k being the extremes with $\delta \cong 20$). Only hotels a and d seemed to represent a set of devices with an acceptable size without any filtering. A constraint on phones caused a moderate decrease in δ for all hotels, but a significantly larger decrease in hotel c. Further investigation indicated this sensor's range overlapping with a nearby parking lot, causing an overrepresentation of devices associated with vehicles (28% vs. 7±4% for all other hotels). The subsequent filtering steps were based on d_v. Setting the lower limit to one hour clearly removed the largest share of noise from the dataset, but common sense dictates that a visit to a hotel should range from at least 8 hours (a guest checking in at night, and checking out early next morning), up to a reasonable maximum number of days (in order to filter out hotel staff, subcontractors, etc.). The upper limit was fixed at 5 days, which is still rather conservative compared to the average duration of a visit in Flanders of 2,43 days (Toerisme Vlaanderen, 2012). After filtering on the visit duration, all hotels except hotels e, i and k were now associated with sets of devices that corresponded with or were slightly smaller than the estimations. For these three hotels, a further filtering on the actual duration of presence yielded acceptable sets when the lower limit was set at one minute. The detection of individuals frequently passing these hotels or their registration desks over several days is the most probable reason for the necessity of this extra filtering step. We opted to apply this additional filtering solely on these three hotels, because it caused a significant additional decrease in four other hotels.

For the three open attractions and the tourist inquiry desk (Figure 3b), no visitor counts were available and the filtering was based on a conservative minimum visit duration of 10 seconds. This choice may seem arbitrary, but was made on the notion that some of these attractions are known to serve as passageways for general movements throughout the center. A further distinction between individuals merely glancing at the attraction, and purposeful visits will not only necessitate further data but also entails a semantic discussion on how to define a 'visit' to such a location.

The filtering for the closed attractions was again based on a combination of the constraint to phones, duration of visit, and the duration of presence. The heuristic lower and upper limits of the visit duration were now set to 15 minutes and 5 hours respectively, thus filtering out inquirers and museum staff. Figure 3c shows that all but four attractions reached an acceptable δ value after filtering on d_v alone. As with the hotels, a further filtering on the duration of presence (minimum of 1 minute) was necessary for the remaining venues. Only attraction 9 was associated with a device set that was still somewhat larger than would be expected after this filtering ($\delta = 1.4$). In absolute numbers, the difference seems less pronounced ($n_d = 59$ vs. $n_v' = 42$). As the lower limit on d_p would need to be set at nearly 5 minutes, we chose to stop the filtering and accept one device set that was slightly larger than expected. The nearby presence of a bar associated with the museum (the visitors of which could be tracked but are not included in the visitor counts) might have caused this anomaly.

4. Data exploration

As a summary of the progressive filtering process, the size of each filtered device set associated with all covered locations is depicted in Figure 4. The filtered sets can now be aggregated into three different sets of tracked individuals: visitors (symbolized as V, part of at least one of the filtered device sets at attractions 1-14), hotel guests (symbolized as H, part of at least one of the filtered

device sets at hotels a-n) and inquirers (symbolized as I, part of the filtered device set at the tourist inquiry desk). The total filtered population of tracked individuals $P = H \cup V \cup I$ contains 7,326 devices, which represents a 58% reduction by the filtering process. Looking at the number of hotel guests and comparing with the map of the study area depicted in Figure 1, we can generally distinguish between cheaper hotels further from the center with lower guest numbers, larger and more expensive hotels in the center, and two hostels accommodating a very small share of guests. The open attractions are associated with significantly larger numbers of visitors than the closed attractions. Based on this finding and the previously mentioned different characteristics of a visit/visitor between the two types of attractions, we will further distinguish between visitors *sensu lato* ($V = V_o \cup V_c$) and visitors *sensu stricto* (V_c), with V_o representing all visitors to at least one of the open attractions and V_c all visitors to at least one of the closed attractions.

As already stated in section 3, there is a significant semantic difference between visiting an open and a closed attraction as visits to open attractions can be very short and sometimes coincidental in nature due to the free entrance and lack of a registration procedure. We therefore suspect that both types of visits are generally performed by different individuals. In order to explore this hypothesis, we start by defining 5 visitor segments based on (different combinations of) the sets V, V_o and V_c. These five segments are: V (visited at least one open/closed attraction), V_o (visited at least one open attraction), V_c (visited at least one closed attraction), $V_o \setminus V_c$ (visited at least one open attraction but none of the closed attractions), $V_c \setminus V_o$ (visited at least one closed attraction but none of the open attractions). The similarities between these segments are depicted in Table 2, where Jaccard-indices (size of the intersection divided by size of the union) were calculated as measures for the similarity. As expected, the number of open attraction visitors V_o (80% of V) clearly exceeds the number of closed attraction visitors V_c (36%). Additionally, the overlap is quite small: 80% of the open attraction visitors never visited any of the closed attractions, 56% of the closed attraction visitors never visited any of the open attractions, only 16% combined both types of attractions. Next, we investigated the share of hotel guests, inquirers and one-day/several-day visitors (calendar days) for the different visitor segments. It appears that especially individuals that only visited one or more closed attractions ($V_c \setminus V_o$) show a deviating (lower) hotel (4 vs. 8%) and inquiry desk (4 vs. 13%) use, and contain a slightly higher frequency of one-day visitors (83 vs. 78%) compared to V. It would be reasonable to assume that this is caused by the geographical distance between most of the closed attractions and the historical center where most tourists (identified hotel guests) stay, making that this group has a higher representation of local (one-day) visitors.

Almost two thirds of all visitors (V) only visited one or more of the open attractions ($V_o \setminus V_c$). Due to the low number of open attractions (3), this would make a large part of the dataset be composed of a rather homogeneous and less informative subpopulation. As such, we continue exploring the set V_c (i.e. the individuals having visited at least one closed attraction) as a more heterogeneous and better candidate set for a data mining method. We continue the investigation by distinguishing between one-day and several-day visitors, leading to two extra visitor segments: $V_c \cap P_{1d}$ and $V_c \cap P_{>1d}$ respectively. As Table 2 shows, one in four visitors in V_c is present in the database over more than one (calendar) day. Around 37% of the several-day visitors were identified as hotel guests, and 23% went to the inquiry desk. Please note that the hotel usage will be an underestimation of the real figure as only 67% of the total hotel capacity was tracked in the experiment, so the other 63% of several-day visitors will be a combination of hotel guests outside of the tracked sample and visitors that performed visits over several days that may or may not be contiguous (e.g. over a first visit in

the first weekend, and a second visit in the following weekend). As a last segment, we additionally constrained to identified hotel guests ($V_c \cap P_{>1d} \cap H$). This segment, which comes as close to the definition of tourists as explained in the introduction, shows an even slightly larger use of the inquiry desk.

In order to explore the possible effect of hotel choice on visiting patterns, we further also distinguish between the guests of hotels located far from the center (*a*, *c*, *g* or *j*; H_{far}), four-star hotels (*i-l*, H_{4*}) and hostels (*m*, *n*; H_{hostel}). Guests of a hotel far from the center clearly less often visit an open or closed attraction, or the inquiry desk. The four-star hotel guests seem to follow the pattern of the more general set of hotel guests. Hostel guests hardly visit the inquiry desk, but seem to visit tourist attractions slightly more often than the average hotel guest.

We also investigated the total duration covered in the tracking data and the number of visited attractions in the different visitor segments. Figure 5 shows the resulting distributions of the number of calendar days, number of attractions and number of closed attractions for the ten visitor segments which will be further investigated in section 5. Around 80% of the visitor population V is tracked over only one calendar day, and practically none over more than five days. Visitors to at least one closed attraction V_c and those that did not visit any open attraction $V_c \setminus V_o$ do not seem to deviate significantly from this distribution. Several-day visitors that were identified as hotel guests seem to cover a slightly higher number of calendar days than those that were not. Please recall that we cannot state with certainty whether an individual did not stay at a hotel because only a subset of hotels was covered by a Bluetooth sensor, so it is difficult to explain this difference directly. Concerning the hotel-based visitor segments, guests of the remote hotels cover slightly less calendar days, and hostel guests slightly more than on average. More than 60% of the visitors *sensu lato* only visit one attraction. Those that visited at least one closed attraction are more distributed towards a higher number of attractions. Looking at the number of visited *closed* attractions, however, the share of visitors that only visited one closed attraction is even higher (over 80%). Several-day visitors (both those identified as hotel guests and those that were not) visit a larger number of attractions when the open attractions are included, but do not deviate significantly from the general trend that most visitors only visited one closed attraction. Remote hotel guests visit slightly less closed attractions, hostel guests slightly more.

As a first approach to investigating the degree of association between the different venues, we also examined the degree of overlap in between the sets of visitors/guests of the different locations as listed in Figure 4. These overlaps are again calculated by the Jaccard index and are depicted in Figure 6 for both the unfiltered and filtered sets. As expected, the overlaps between the hotels is very low: tourists usually only stay at one hotel during a visit. The remaining overlaps after filtering are mainly caused by geographical proximity (e.g. hotels *h* and *k* nearly face each other across the same street) and guests of one hotel being mistakenly classified as guests of the other. The open attractions, and closed attractions 5 (Belfry) and 7 (Castle of the Counts) all show significant mutual overlaps. As can be seen in Figure 1, proximity is probably the most important cause besides similarity in characteristics (all are historical buildings). The same effect also explains the higher overlaps between the tourist inquiry desk and the open attractions in the center, and the moderate overlaps with attractions 5-7. Outside of the center, attractions 10, 12 and 14 in the 'arts quarter' also show significant mutual overlaps. A deeper understanding of the associations, however, will be mined for by the association rule learning method. In the next section, we will outline its results.

5. Visit pattern mining

The degree of overlap between the different locations and their visitors discussed in the previous section offers a first insight into the degree of association between the different attractions. To obtain a deeper understanding, however, a more thorough analysis is needed. In this section, we will mine for association rules between the fourteen (open and closed) attractions for the previously identified visitor segments. The mining process was described in section 2.2 (minimum support of 5 devices, minimum confidence of 5%). The subset of rules with only one item in the antecedent is visualized in their geographical context, together with the share of visitors in the segment that visited each attraction, in a 'visit pattern map' as described in section 2.3. For each segment, the top-20 of *all* association rules is additionally listed in a tabular fashion. The rest of this section is divided into two parts. In section 5.1, 'visitor segment exploration', we describe the patterns and their differences found going from the general set of visitors *sensu lato* V to the most specific segment of hotel guests that visited at least one closed attraction and have a duration of visit of at least 1 day ($V_c \cap P_{>1d} \cap H$). In section 5.2, we investigate the potential differences in patterns for remote hotel guests, four-star hotel guests, and hostel guests.

5.1. Visitor segment exploration

Figure 7 shows the visit pattern maps for the five visitor segments going from visitors *sensu lato* V, over V_c, $V_c \cap P_{1d}$, and $V_c \cap P_{>1d}$ to $V_c \cap P_{>1d} \cap H$. Table 3 lists the top-20 of all rules (including those with more than one item in the antecedent) for all these visitor segments. Looking at the share of visitors visiting each attraction (proportionally-sized circles) on the visit pattern map for visitors *sensu lato* (V), we clearly observe a concentration of visits in the city center and its open attractions (attractions 1 and 2 each attract nearly 50% of the tracked population). Of the closed attractions, the Castle of the Counts (7) attracts the largest share of visitors (12%), followed by the Belfry (5) with 6%. All other closed attractions, both in the center and more remote, attract significantly smaller shares – the largest being the Museum of Contemporary Art (SMAK, 12) which attracts around 4% of the population. Concerning the rules for the visitors *sensu lato* (V) segment, both the map and the top-20 show that all association rules with a high lift have a very low support (the rule with the highest lift in the map is supported by only 1% of the population, rules with more items in the top-20 have even lower supports). As previously mentioned, visitors *sensu lato* represent a heterogeneous group of individuals – many of which only visited one or more open attractions (Figure 5) on relatively short trips. Despite low supports, the strong associations between the SMAK museum (12) and the Museum of Fine Arts (10) and especially between the SMAK museum and the Saint Peter's Abbey (14) are noteworthy. Longer-distance associations have a low lift and support, indicating that visitors seem to rarely combine in-center and out-of-center visits. The associations between the open attractions logically have a larger support, but show lift values close to 1 suggesting that the associations are not significantly stronger than expected.

With regard to the more specific visitor segments and starting by constraining to closed attraction visitors (V_c), the associations between the three art museums in the south are no longer the strongest in the set (although they still have lift values higher than 1). Instead, the highest lift values are now found in the city center, especially between Saint Nicholas' Church (3), and the Belfry (5) and Saint Bavo's Cathedral (2): for visitors that made at least one conscious choice in visiting a closed attraction we find a higher than expected association between the attractions in the city center (lift > 1).

Next, for visitors that visited none of the open attractions ($V_c \setminus V_o$) very few associations are found, only with low lift values: only the combinations 12-10 and 12-14 remain, together with an association between the Castle of the Counts (7) and the House of Alijn (8).

Visitors *sensu stricto* that were only tracked on one calendar day ($V_c \cap P_{1d}$) show nearly no association between attractions in the center and those outside of the center, revealing they either visit one or the other.

Those that were tracked over several days ($V_c \cap P_{>1d}$) do show these associations as they probably have more free time to cover these distances. Comparing these visitors to the visitors *sensu stricto* in general (V_c), there are no clear differences in both patterns for the city center.

Finally, comparing identified hotel guests ($V_c \cap P_{>1d} \cap H$) with the general several-day visitors ($V_c \cap P_{>1d}$) some differences between the patterns become visible. Saint Peter's Abbey looses its significance, together with the associations between the art museums in the south. In contrast, both the House of Alijn (8) and the City Museum of Ghent (STAM, 13) appear to gain in importance: both now show significant (lift>1) and considerable (support of almost 10%) associations. The lift of the association between the Museum of Industrial Archeology and Textiles (MIAT, 9) and attraction 1 also seems to have risen, but the support of the rule is very low (3,5%, 7 tracked individuals).

5.2. Effect of hotel location and type

The potential effect of the hotel location and type is investigated by comparing the visit patterns of four distinct segments of hotel guests: the general set of hotel guests that visited at least one open/closed attraction ($V \cap H$), those that stayed at one of the four most remote hotels (*a*, *c*, *g* or *j*, $V \cap H_{far}$), those that stayed at one of the four 4-star hotels ($V \cap H_{4*}$), and finally those that stayed in one of both hostels ($V \cap H_{hostel}$). The four corresponding visit pattern maps are shown in Figure 8. The top-20 of rules for each segment is listed in Table 4. Remote hotel guests, who were already shown to significantly more often engage in one or more visits (Table 2), clearly show a preference for the open attractions and the Castle of the Counts (7) in the center. They rarely visit any of the closed attractions further from the center, and associations besides those between open attractions are rare. The 4-star hotel guests show a visit pattern that is almost identical to the general visit pattern for hotel guests, which makes sense as they form the largest share of hotel guests. Hostel guests, finally, show a pattern that bears some similarities to the remote hotel guests. The Museum of Contemporary Art (SMAK, 12) seems to attract a significantly larger share of these visitors (24% vs. 4% of the general hotel guest segment). It should be noted, however, that the size of this last segment has become rather small, possibly limiting the representativeness of the pattern it exhibits.

6. Discussion and conclusion

In this section, we will first further interpret the filtering and mining processes and the results that were generated from the dataset. Subsequently, we will discuss the current and future potential of the presented methodology (Bluetooth tracking + association rule learning) for tourism management purposes. We conclude with further avenues for future research.

6.1. Further interpretation of filtering, mining and results

As in any knowledge discovery process, data mining techniques form only part of a chain of subprocesses going from data to knowledge (Fayyad et al., 1996). As described in section 3, the

Bluetooth tracking data needed to undergo significant preprocessing before being able to function as an input for association rule learning. There are two main reasons for the need for such an extensive filtering procedure. First, Bluetooth is a very popular and widely distributed technology available on a large variety of devices. Whereas phones can be assumed to be linked to one individual each, other classes, such as carkits, cannot. Second, the proximity of an individual to a place of interest cannot be directly associated with a certain activity related to that location. This becomes clear in Figure 4, where most sensors at both hotels and attractions clearly detected significantly more Bluetooth devices than predicted according to the actual guest/visitor counts and the detection ratio of 8%. As can be seen in Figure 3, most hotels are associated with very short detections indicated by the large decrease in filtered devices by choosing d_v to a minimum of one hour. A further constraint of this parameter to the heuristic interval between 8 hours and 5 days leads to acceptable filtered sets, except for three hotels where a further constraint had to be placed on d_p. A similar trend, but with different constraints on d_v, is visible for the closed attractions. While it is impossible to statistically verify the accuracy of the applied temporal filtering, Figure 6 at least gives an indication that most significant overlaps between hotels (where there should not be any) have decreased significantly. The overlaps that remain after filtering are mainly due to proximity (e.g. hotels *h* and *k*). The uncertainty over whether a tracked individual actually visited an attraction and/or stayed at a hotel could be addressed in future work by making the scanning of the device part of the registration process. This shift in a participatory nature would make the methodology significantly more labor-intensive, though, thereby potentially nullifying its main advantage.

Before we started mining for association rules, we performed an extensive data exploration. Two hypotheses received special attention. First, the combination of open and closed attractions was suspected to result in a heterogeneous group of 'visitors' because of the semantic difference between both types. Second, we needed to investigate the tendency of individuals to visit more than one attraction over the 15-day tracking period before looking at the specific associations. As was listed in Table 2, visitors *sensu lato* show little overlap between visitors *sensu stricto*: only 36% of the former group also visit at least one closed attraction, and only 16% combine both types. Figure 5 showed that only a small fraction of visitors combined more than one closed attraction, which would certainly influence the mining for association rules. Readers should bear in mind that 15 days is quite a short period for capturing combination preferences between museums for visits that are not part of a chained trip of visits (e.g. for local residents), and that the tracking period would ideally be composed of the entire touristic season. Where the mined associations for these individuals would certainly be an underestimation and thus possibly not representative, identified hotel guests can be assumed to perform visits that form part of a trip. As such, this group ($V_c \cap P_{>1d} \cap H$) received special attention in the analyses.

Subsequently, the associations between the different open and closed attractions were investigated through the application of the *Apriori* algorithm and the interpretation and visualization of the extracted rules. Combining the exploratory findings, we gradually restricted the set of visitors *sensu lato* to individuals that visited at least one closed attraction (as this indicates at least one conscious choice in their itinerary) and additionally to identified hotel guests. Figure 7 shows the resulting series of visit pattern maps. Taking all visitors *sensu lato* as input, large but predictable associations in the center appear next to very small but more interesting associations between the art-oriented attractions in the south of the study area. Constricting to visitors *sensu stricto* decreased this conceptual difference due to the lower frequency (higher lift) of the center-based attractions and the

higher frequency (lower lift) of the art attractions. Several-day visitors clearly show associations (though most of them are quite small) between the center and more remote attractions, whereas one-day visitors rarely combine both areas. The constrained set of identified hotel guests, which could be argued to be the only representative segment in the series, shows some interesting differences with the general several-day visitor set (which also includes individuals that did not stay overnight). The strong associations between the art museums disappear together with the significance of Saint Peter's Abbey individually, but two new locations appear in considerable associations with the open attractions in the center: the City Museum of Ghent (STAM), and the House of Alijn'. Subsequently, we investigated the additional differences in visit patterns between remote hotel guests, four-star hotel guests and hostel guests. Remote hotel guests show a very large preference in the open attractions in the center and rarely venture further away, whereas hostel guests show a similar pattern with the exception of a higher representation of the Museum of Contemporary Art (SMAK). Four-star hotel guests followed the general pattern.

6.2. Potential of the employed methodology for tourism management

The potential of Bluetooth tracking in tourism management practices was illustrated by applying an association rule learning scheme on the visits to different attractions in an urban environment. In this specific case study, we focused on discovering *interesting* associations as indicated by attractions appearing together more often than predicted by an independent choice model. Such associations or the lack thereof can be used for a wide array of purposes. Focusing on the attractions, existing associations could either be strengthened or non-existent or weak associations could be created by applying specific promotional advertisements at each attraction, thereby urging visitors to visit other attractions as well. In contrast, the focus could also lie on the tourist and discovered patterns could be used in recommendation systems based on collaborative filtering. These recommendations could be disseminated through the use of smartphones. In a long-term strategic context, pattern maps such as those in Figure 7 could also be used by urban planners for optimizing tourist accessibility and facilities. The three art museums in the south of the study area (promoted as the 'arts quarter') can serve as an illustrative example. While they do seem associated with each other mutually, our analysis has also pointed out that there is little association with the attractions in the center. Planners could tackle this issue by developing the necessary tourist facilities (e.g. hotels) in the 'arts quarter' itself, improving the (visibility of) public transport options between both areas, or by designing a corridor to minimize the perceived distance between the center and the 'arts quarter' (e.g. by improving pedestrian accessibility, creating more green and open spaces, etc.). The effectiveness of certain actions could also be investigated by tracking during a period both before and after the action was taken.

6.3. Further issues surrounding the methodology

Three main issues can be identified with regard to the used methodology: (i) errors introduced when using presence detections to deduce activities, (ii) potential bias in the tracked sample introduced by using Bluetooth technology, and (iii) lack of any metadata on tracked visitors due to the non-participatory nature of the methodology. We will shortly reflect on all of these issues.

Deducing activities from presence detections through sensors is a process which can essentially be affected by two types of errors. The first error could be labeled as false presence detections, where a sensor detects devices that do not physically enter a building or approach a desk. Filtering procedures were able to remove most of this sensor noise but further efforts are certainly warranted

in order to calibrate the sensors thereby minimizing false presence detections. The second type of error is due to actual presences not automatically implying certain activities. In the context of this study, these could be caused by staff members or inquirers. Filtering on common-sense thresholds for the duration of visit and duration of presence, we were able to deduce acceptable device sets. However, questions still remain on why additional filtering steps were necessary for some of the locations. In the end, a higher accuracy in deducing activities will most probably imply contacting individuals and registering the MAC address of their device(s) in contrast with the use of an unsupervised filtering procedure.

Second, it is possible that certain age segments, or one gender might have a higher usage rate of devices with a discoverable Bluetooth interface. Tourists might, for example, preferentially opt to turn off their mobile phones in order not to be disturbed and thus be under-represented. Despite the undeniable importance of these potential influencing factors on any tourism management incentive, it falls outside of the scope of this research to fully investigate this issue. As in any other tracking study, however, it is of vital importance that more attention should be devoted to this question.

The lack of metadata, finally, can be interpreted as the downside of using a non-participatory methodology. Instead of dealing participatory and non-participatory methodologies as complete opposites, we argue that both methodologies might be combined. Alternatively, ad-hoc sensing networks could be made semi-participatory by approaching and interviewing part of the tracked population. In our scenario, for example, individuals could be contacted in the hotel where they are staying. Again, this fell outside of the current scope but could aid in further strengthening assumptions made about visitors.

6.4. Future research

Due to the limited number of sensors at our disposal, this study only included primary touristic locations. It is well known that a significant share of tourists may wander from these locations and engage in other related activities such as shopping, visiting bars, etc. Future studies may include some of these secondary locations in order to provide a more holistic image of tourism. While we applied an association rule learning technique in this case study, other data mining tasks could be used or combined for answering other or similar research questions as those put forward in this paper. If the order of visits were important for example, sequential pattern mining techniques could be used. Clustering methods could segment tourists based on their associations of visits. In addition, a second and longer period of tracking could serve as a point of comparison with the patterns found during the 15-day tracking period of this study. The visit pattern maps summarize a considerable amount of information in one map, which possibly makes them challenging to interpret. User studies could clarify the way in which people read the map (e.g. by eye-tracking procedures), and the provided insights could be used to further fine-tune the visualization. The selective visualization of 2-item-rules causes some of the (potentially valuable) information contained within rules with more items to be lost. How to efficiently visualize these complex rules in a legible way is saved for future work.

Bibliography

Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining Association Rules between Sets of Items in Large Databases. In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data - SIGMOD '93* (Vol. 22, pp. 207–216). New York, USA: ACM Press.

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. In *Proceedings of the 20th International Conference on Very Large Data Bases, VLDB* (pp. 487–499). Santiago, Chile.

Ahas, R., Aasa, A., Mark, Ü., Pae, T., & Kull, A. (2007). Seasonal tourism spaces in Estonia: Case study with mobile positioning data. *Tourism Management*, *28*(3), 898–910.

Ahas, R., Aasa, A., Roose, A., Mark, Ü., & Silm, S. (2008). Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. *Tourism Management*, *29*(3), 469–486.

Ahas, R., Laineste, J., Aasa, A., & Mark, Ü. (2007). The Spatial Accuracy of Mobile Positioning: Some experiences with Geographical Studies in Estonia. In G. Gartner, W. Cartwright, & M. P. Peterson (Eds.), *Location Based Services and TeleCartography* (pp. 445–460). Berlin: Springer.

Al-Salim, B. (2008). Mass customization of travel packages: data mining approach. *International Journal of Flexible Manufacturing Systems*, *19*(4), 612–624.

Appice, A., & Buono, P. (2005). Analyzing Multi-level Spatial Association Rules Through a Graph-Based Visualization. In A. Moonis & F. Esposito (Eds.), *Innovations in Applied Artificial Intelligence* (Vol. 3533, pp. 448–458). Springer.

Bloom, J. Z. (2005). Market Segmentation: A Neural Network Application. *Annals of Tourism Research*, *32*(1), 93–111.

Bonné, B., Barzan, A., Quax, P., & Lamotte, W. (2013). WiFiPi: Involuntary tracking of visitors at mass events. In *2013 IEEE 14th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*. Madrid, Spain.

Bruzzese, D., & Davino, C. (2003). Visual post-analysis of association rules. *Journal of Visual Languages & Computing*, *14*(6), 621–635.

Chen, Y.-L., Tang, K., Shen, R.-J., & Hu, Y.-H. (2005). Market basket analysis in a multiple store environment. *Decision Support Systems*, *40*(2), 339–354.

Cini, F., Leone, L., & Passafaro, P. (2010). Promoting Ecotourism Among Young People: A Segmentation Strategy. *Environment and Behavior*, *44*(1), 87–106.

Compieta, P., Di Martino, S., Bertolotto, M., Ferrucci, F., & Kechadi, T. (2007). Exploratory spatio-temporal data mining and visualization. *Journal of Visual Languages & Computing*, *18*(3), 255–279.

Connell, J., & Page, S. J. (2008). Exploring the spatial patterns of car-based tourist travel in Loch Lomond and Trossachs National Park, Scotland. *Tourism Management*, *29*(3), 561–580.

Delafontaine, M., Versichele, M., Neutens, T., & Van de Weghe, N. (2012). Analysing spatiotemporal sequences in Bluetooth tracking data. *Applied Geography*, *34*, 659–668.

Dolničar, S. (2004). Beyond "Commonsense Segmentation": A Systematics of Segmentation Approaches in Tourism. *Journal of Travel Research*, *42*(3), 244–250.

Dolničar, S., & Leisch, F. (2003). Winter tourist segments in Austria - Identifying stable vacation styles using bagged clustering techniques. *Journal of Travel Research*, *41*(3), 281–292.

Eagle, N., & Pentland, A. (2005). Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, *10*(4), 255–268.

Emel, G. G., Taskin, Ç., & Akat, Ö. (2007). Profiling a Domestic Tourism Market By Means Of Association Rule Mining : A Case Study. *Anatolia*, *18*(2), 1–12.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI magazine*, *17*(3), 37–54.

González, M. C., Hidalgo, C. A., & Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, *453*(7196), 779–782.

Hahsler, M., & Chelluboina, S. (2010). Visualizing Association Rules : Introduction to the R-extension Package arulesViz. *R project module*.

Hahsler, M., Chelluboina, S., Hornik, K., & Buchta, C. (2011). The arules R-Package Ecosystem: Analyzing Interesting Patterns from Large Transaction Data Sets. *Journal of Machine Learning*, *12*, 2021–2025.

Hartmann, R. (1988). Combining field methods in tourism research. *Annals of Tourism Research*, *15*, 88–105.

Janelle, D., Goodchild, M., & Klinkenberg, B. (1988). Space-time diaries and travel characteristics for different levels of respondent aggregation. *Environment and Planning A*, *20*, 891–906.

Jankowski, P., Andrienko, N., Andrienko, G., & Kisilevich, S. (2010). Discovering Landmark Preferences and Movement Patterns from Photo Postings. *Transactions in GIS*, *14*(6), 833–852.

Kemperman, A., Borgers, A., & Timmermans, H. (2009). Tourist shopping behavior in a historic downtown area. *Tourism Management*, *30*(2), 208–218.

Lau, G., & McKercher, B. (2006). Understanding tourist movement patterns in a destination: A GIS approach. *Tourism and Hospitality Research*, *7*(1), 39–49.

Law, R., & Au, N. (2000). Relationship modeling in tourism shopping: a decision rules induction approach. *Tourism Management*, *21*(3), 241–249.

Law, R., Bauer, T., Weber, K., & Tse, T. (2006). Towards a Rough Classification of Business Travelers. In *Advanced Data Mining and Applications* (Vol. 4093, pp. 135–142).

Law, R., Rong, J., Vu, H. Q., Li, G., & Lee, H. A. (2011). Identifying changes and trends in Hong Kong outbound tourism. *Tourism Management*, *32*(5), 1106–1114.

Liao, S., Chen, Y.-J., & Deng, M. (2010). Mining customer knowledge for tourism new product development and customer relationship management. *Expert Systems with Applications*, *37*(6), 4212–4223.

Lonely Planet. (2011). Ghent: Belgium's best kept secret.

McKercher, B. (1999). A chaos approach to tourism. *Tourism Management*, *20*, 425–434.

Orellana, D., Bregt, A. K., Ligtenberg, A., & Wachowicz, M. (2012). Exploring visitor movement patterns in natural recreational areas. *Tourism Management*, *33*(3), 672–682.

Öztayşi, B., Baysan, S., & Akpinar, F. (2009). Radio frequency identification (RFID) in hospitality. *Technovation*, *29*(9), 618–624.

Peterson, B. S., Baldwin, R. O., & Kharoufeh, J. P. (2006). Bluetooth inquiry time characterization and selection. *IEEE Transactions on Mobile Computing*, *5*(9), 1173–1187.

Ratti, C., Pulselli, R. M., Williams, S., & Frenchman, D. (2006). Mobile Landscapes: using location data from cell phones for urban analysis. *Environment and Planning B: Planning and Design*, *33*(5), 727–748.

Rong, J., Vu, H. Q., Law, R., & Li, G. (2012). A behavioral analysis of web sharers and browsers in Hong Kong using targeted association rule mining. *Tourism Management*, *33*(4), 731–740.

Shoval, N., & Isaacson, M. (2007a). Sequence Alignment as a Method for Human Activity Analysis in Space and Time. *Annals of the Association of American Geographers*, *97*(2), 282–297.

Shoval, N., & Isaacson, M. (2007b). Tracking tourists in the digital age. *Annals of Tourism Research*, *34*(1), 141–159.

Shoval, N., & Isaacson, M. (2009). *Tourist mobility and advanced tracking technologies*. (S. Page, Ed.) (Vol. 19, p. 207). New York, London: Routledge.

Shoval, N., McKercher, B., Ng, E., & Birenboim, A. (2011). Hotel location and tourist activity in cities. *Annals of Tourism Research*, *38*(4), 1594–1612.

Song, H., & Li, G. (2008). Tourism demand modelling and forecasting—A review of recent research. *Tourism Management*, *29*(2), 203–220.

Stange, H., Liebig, T., Hecker, D., Andrienko, G., & Andrienko, N. (2011). Analytical Workflow of Monitoring Human Mobility in Big Event Settings using Bluetooth. In *Third ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness* (pp. 51–58). Chicago, IL, USA: ACM.

Tan, P.-N., Steinbach, M., & Kumar, V. (2005). Association Analysis: Basic Concepts and Algorithms. In *Introduction to Data Mining* (p. 769). Boston, MA, USA: Addison-Wesley Longman Publishing.

Tchetchik, A., Fleischer, A., & Shoval, N. (2009). Segmentation of Visitors to a Heritage Site Using High-resolution Time-space Data. *Journal of Travel Research*, *48*(2), 216–229.

Techapichetvanich, K., & Datta, A. (2005). VisAR : A New Technique for Visualizing Mined Association Rules. In L. Xue, S. Wang, & D. Zhao Yang (Eds.), *Advanced Data Mining and Applications* (Vol. 3584, pp. 88–95).

Toerisme Vlaanderen. (2012). *Toerisme in cijfers 2011: de Belgische markt in Vlaanderen* (p. 17).

Versichele, M., Neutens, T., Delafontaine, M., & Van de Weghe, N. (2012). The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent Festivities. *Applied Geography*, *32*(2), 208–220.

Versichele, M., Neutens, T., Goudeseune, S., Van Bossche, F., & Van de Weghe, N. (2012). Mobile Mapping of Sporting Event Spectators Using Bluetooth Sensors: Tour of Flanders 2011. *Sensors*, *12*(10), 14196–14213.

Witt, S. F., & Witt, C. a. (1995). Forecasting tourism demand: A review of empirical research. *International Journal of Forecasting*, *11*(3), 447–475.

World Tourism Organization. (1995). *Collection of Tourism Expenditure Statistics* (p. 14).

Yang, L. (2005). Pruning and visualizing generalized association rules in parallel coordinates. *IEEE Transactions on Knowledge and Data Engineering*, *17*(1), 60–70.
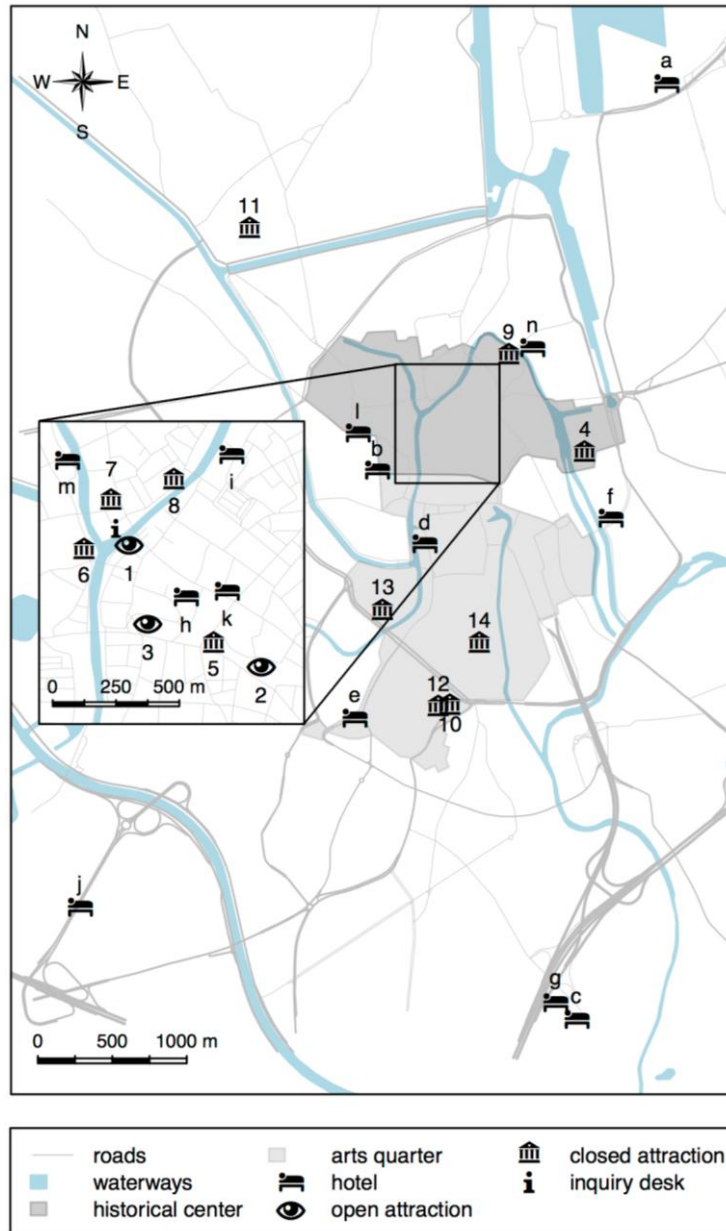
Figure 1. Overview of the Bluetooth sensor placement in Ghent, Belgium. The inset map shows a more detailed view of the city center.
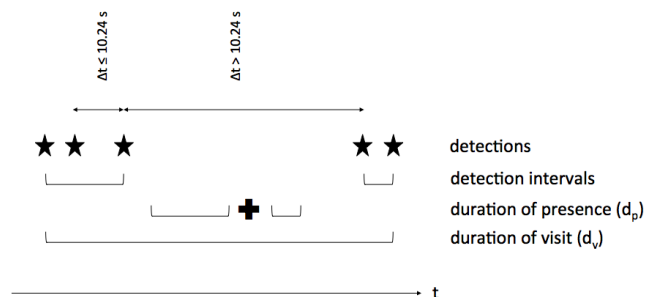
Figure 2. Schematical representation of detections, detection intervals, the duration of presence (d_p), and the duration of visit (d_v).

Table 1. Illustrative example of a transaction database in the context of tourist attractions in Paris (1: visited, 0: not visited).

Transaction (=visitor)	Eiffel Tower	Louvre Museum	Arc de Triomphe	Notre Dame
1	1	1	0	1
2	0	0	1	0
3	1	0	1	1
4	1	1	1	1
5	0	1	1	1
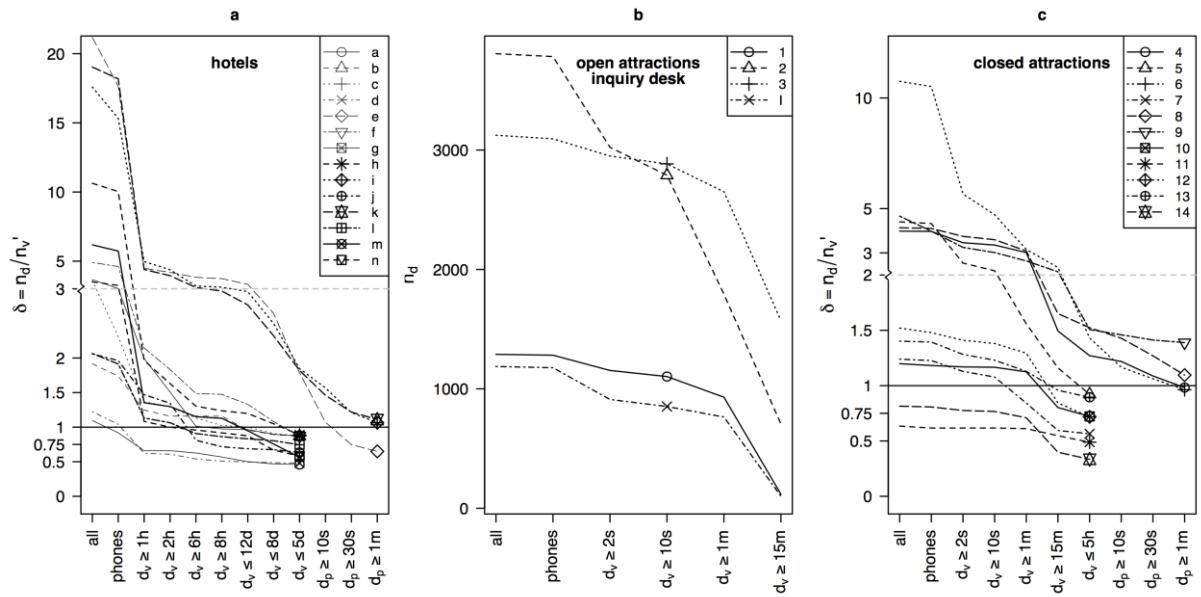
Figure 3. Progressive filtering process on the detected Bluetooth devices for the hotels (a), open attractions and tourist inquiry desk (b) and closed attractions (c). The point symbols indicate the filtering end points for each location. For the hotels and closed attractions, the filtering is based on $\delta = n_d/n_v'$ with n_d representing the number of devices and n_v' the estimated number of tracked visitors/guests based on a detection ratio ρ of 8%. For the open attractions and the inquiry desk, no visitor counts were available and only n_d was taken into account. Note the breaks in scale depicted as the dashed horizontal line on the y-axes in (a) and (c).
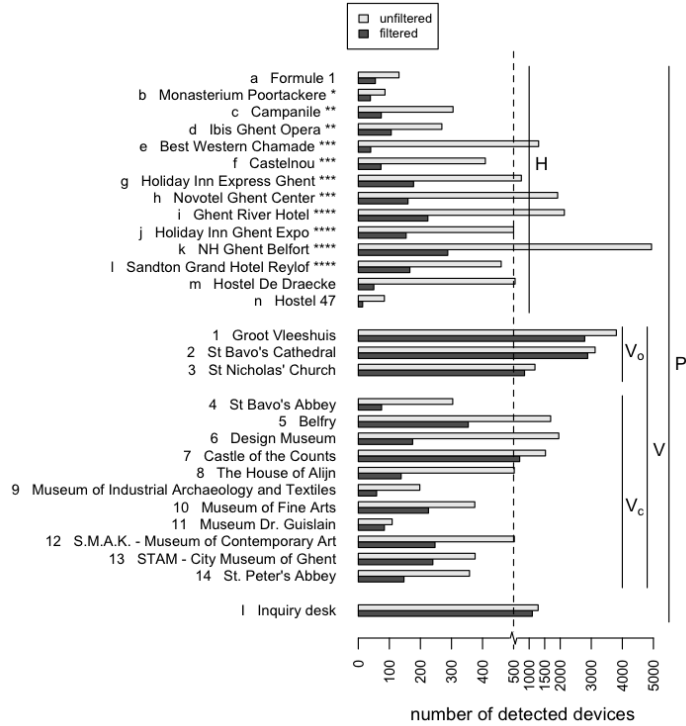
Figure 4. Preprocessing summary showing the number of detected devices before and after filtering at each location, and the aggregation of the filtered devices into the sets of hotel guests H, visitors V (*sensu lato*), information seekers I, open attraction visitors V_o, closed attraction visitors V_c, and the entire population of tracked individuals P. Note the break in scale on the x-axis depicted by the dashed vertical line.

Table 2. Sizes of, similarities between different visitor segments (Jaccard-index), and the corresponding share of hotel guests (H), inquirers (I), one-day (P_{1d}) and several-day visitors ($P_{>1d}$) for each visitor segment. The five visitor segments are: V (all visitors), V_o (open attraction visitors), V_c (closed attraction visitors), $V_o \setminus V_c$ (only visited one or more open attractions but none of the closed attractions), $V_c \setminus V_o$ (only visited one or more closed attractions but none of the open attractions).

		Jaccard-index: $\#(A \cap B)/\#(A \cup B)$					$\#(A \cap B)/\#A$						
$A \downarrow \ B \rightarrow$	#A	V	V_o	V_c	$V_o \setminus V_c$	$V_c \setminus V_o$	H	I	P_{1d}	$P_{>1d}$	V	V_o	V_c
V	5,891	1	0.80	0.36	0.64	0.20	0.08	0.13	0.78	0.22			
V_o	4,726	0.80	1	0.16	0.80	0	0.09	0.15	0.77	0.23			
V_c	2,095	0.36	0.16	1	0	0.56	0.10	0.14	0.75	0.25			
$V_o \setminus V_c$	3,796	0.64	0.80	0	1	0	0.07	0.12	0.80	0.20			
$V_c \setminus V_o$	1,165	0.20	0	0.56	0	1	0.04	0.04	0.83	0.17			
$V_c \cap P_{1d}$	1,564						0.00	0.11	1	0			0.75
$V_c \cap P_{\geq 1d}$	531						0.37	0.23	0	1			0.25
$V_c \cap P_{>1d} \cap H$	196						1	0.27	0	1			0.09
H	1,581							0.08			0.31	0.28	0.13
H_{far}	456							0.05			0.16	0.15	0.03
H_{4*}	675							0.09			0.37	0.33	0.16
H_{hostel}	64							0.02			0.39	0.33	0.19

22
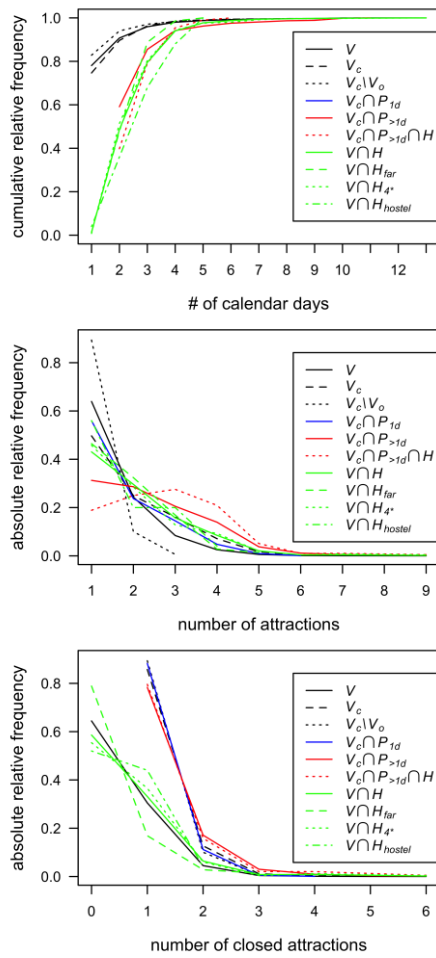
Figure 5. Cumulative relative frequency of the number of tracked calendar days (top), and absolute relative frequency of the number of visited attractions (middle) and closed attractions (bottom).

hotels — unfiltered

	a	b	c	d	e	f	g	h	i	j	k	l	m	n
a		0	0	1	0	0	0	0	0	0	0	0	0	1
b	0		1	0	0	0	0	0	0	0	0	0	0	0
c	0	1		0	0	0	4	0	0	0	1	0	0	0
d	1	0	0		1	0	0	1	1	0	1	1	1	0
e	0	0	0	1		1	1	2	2	0	2	1	1	0
f	0	0	0	0	1		1	1	1	0	1	1	1	1
g	0	0	4	0	1	1		1	1	2	1	1	0	0
h	0	0	0	1	2	1	1		3	0	17	2	2	0
i	0	0	0	1	2	1	1	3		0	5	1	2	1
j	0	0	0	0	0	0	2	0	0		1	0	0	0
k	0	0	1	1	2	1	1	17	5	1		2	2	0
l	0	0	1	1	1	1	1	2	1	0	2		2	0
m	0	0	0	1	1	1	0	2	2	0	2	2		1
n	1	0	0	0	0	1	0	0	1	0	0	0	1	

hotels — filtered

|	a	b	c	d	e	f	g	h	i	j	k	l	m	n
a		0	0	1	0	0	0	0	0	0	0	0	0	0
b	0		0	0	0	0	0	0	0	0	0	0	0	0
c	0	0		0	0	0	2	0	0	0	0	0	0	0
d	1	0	0		0	0	0	0	0	0	0	0	0	0
e	0	0	0	0		0	0	1	0	0	1	0	0	0
f	0	0	0	0	0		0	0	0	0	0	0	0	0
g	0	0	2	0	0	0		0	0	0	0	0	0	0
h	0	0	0	0	1	0	0		0	0	4	0	0	0
i	0	0	0	0	0	0	0	0		0	0	0	0	1
j	0	0	0	0	0	0	0	0	0		0	0	0	0
k	0	0	0	0	1	0	0	4	0	0		0	0	1
l	0	0	0	0	0	0	0	0	0	0	0		0	0
m	0	0	0	0	0	0	0	0	0	0	0	0		0
n	0	0	0	0	0	0	0	0	1	0	1	0	0	

open and closed attractions, inquiry desk — unfiltered

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	I
1		24	12	1	13	14	16	5	1	0	2	2	1		14
2	24		22	1	22	14	22	5	1	2	0	1	1	1	14
3	12	22		1	17	10	16	5	1	2	0	1	1	1	12
4	1	1	1		1	1	1	4	1	2	0	6	1	6	1
5	13	22	17	1		11	14	4	1	1	0	1	1	1	10
6	14	14	10	1	11		12	5	1	2	0	2	1	2	13
7	16	22	16	1	14	12		6	1	2	0	1	1	1	15
8	5	5	5	4	4	5	6		2	2	0	2	1	2	8
9	1	1	1	1	1	1	1	2		2	0	2	1	1	1
10	1	2	2	2	1	2	2	2	1		21	2	11		2
11	0	0	0	0	0	0	0	0	1	0		0	1	1	0
12	2	1	1	6	1	2	1	2	2	21	0		3	20	2
13	2	1	1	1	1	1	1	1	1	2	1	3		3	2
14	1	1	1	6	1	2	1	2	1	11	1	20	3		1
I	14	14	12	1	10	13	15	8	1	2	0	2	2	1	

open and closed attractions, inquiry desk — filtered

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	I
1		25	11	0	5	2	9	2	0	1	0	1	1	0	14
2	25		19	0	8	3	12	2	0	2	0	1	1	0	14
3	11	19		0	10	2	9	2	1	2	0	1	1	1	12
4	0	0	0		0	0	0	2	2	0	0	1	1	0	1
5	5	8	10	0		3	9	3	1	1	0	1	0	1	6
6	2	3	2	0	3		3	2	0	2	0	1	1	1	4
7	9	12	9	0	9	3		4	1	1	0	1	1	1	9
8	2	2	2	2	3	2	4		3	1	1	1	1	1	2
9	0	0	1	2	1	0	1	3		2	0	0	1	0	1
10	1	2	2	0	1	2	1	1	2		1	8	2	3	2
11	0	0	0	0	0	0	0	1	0	1		0	1	0	0
12	1	1	1	1	1	1	1	1	0	8	0		1	10	1
13	1	1	1	0	1	1	1	1	1	2	1	1		2	1
14	0	0	1	0	1	1	1	1	0	3	0	10	2		1
I	14	14	12	1	6	4	9	2	1	2	0	1	1	1	

Figure 6. Overlap between unfiltered (left) and filtered (right) device sets at hotels (a-n, top), and the open and closed attractions and inquiry desk (1-14 + I, bottom). The numbers in the grid are the Jaccard indices of each combination of device sets, and represent the degree of overlap (0: less than one % overlap, 100: completely identical).
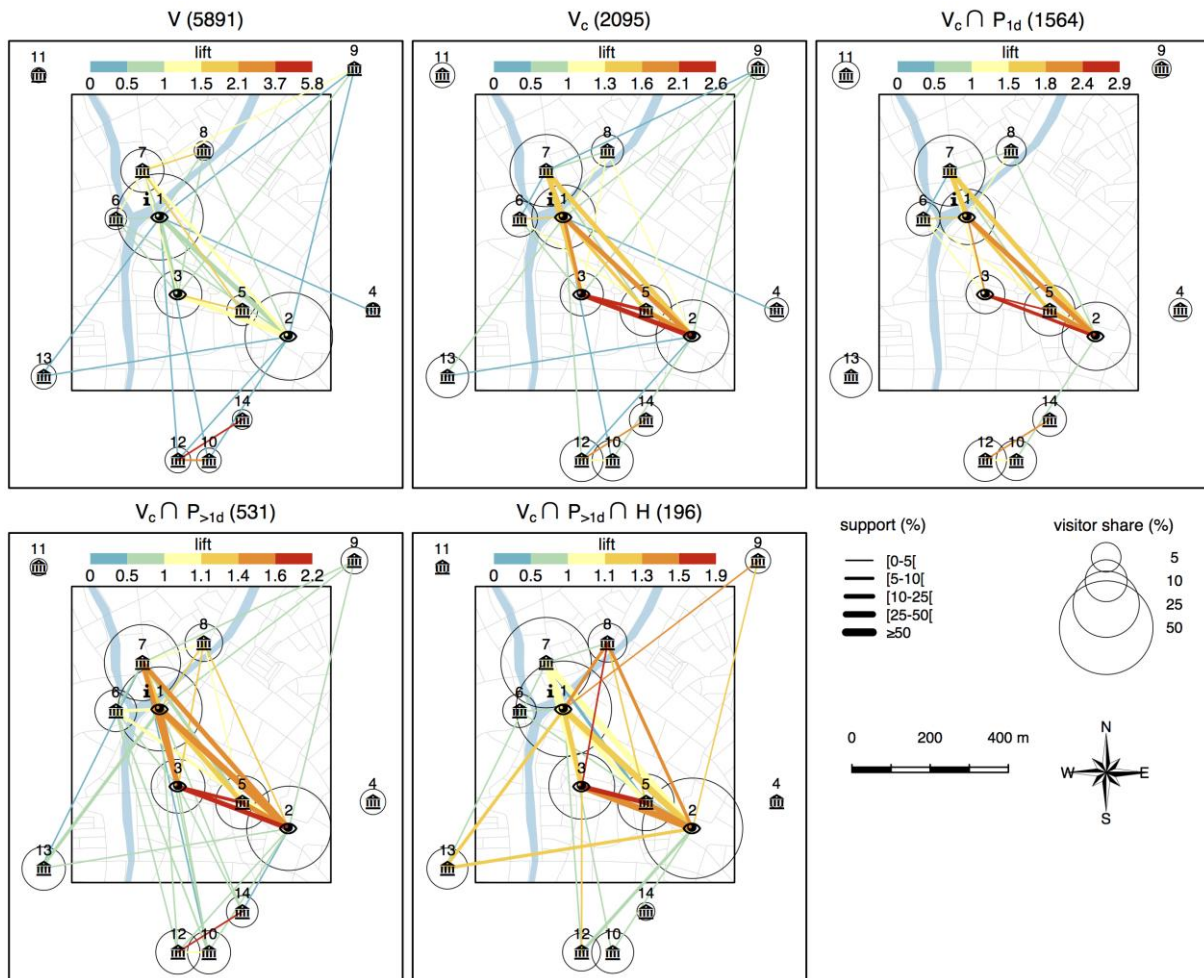
Figure 7. Visit pattern maps for visitor segments $V, V_c, V_c \cap P_{1d}, V_c \cap P_{>1d}$, and $V_c \cap P_{>1d} \cap H$. The size of each segment is given between brackets. The spatial distribution of visits is represented by proportionally sized circles symbolizing the share of visitors in the segment visiting each attraction. Two-element association rules are visualized as lines connecting the locations in the antecedent and consequent. The support of a rule is symbolized by the width of the line, the lift by its color (as shown by the bars above each map, classification was equal-range for lifts below 1 and according to Jenks natural breaks above 1). Attractions far from the city center (depicted by the rectangle) are not depicted on their actual geographical location in order to increase the legibility of the visualization.

Figure 8. Visit pattern maps for visitor segments $V \cap H$, $V \cap H_{far}$, $V \cap H_{4*}$, **and** $V \cap H_{hostel}$. **The size of each segment is given between brackets. The spatial distribution of visits is represented by proportionally sized circles symbolizing the share of visitors in the segment visiting each attraction. Two-element association rules are visualized as lines connecting the locations in the antecedent and consequent. The support of a rule is symbolized by the width of the line, the lift by its color (as shown by the bars above each map, classification was equal-range for lifts below 1 and according to Jenks natural breaks above 1). Attractions far from the city center (depicted by the rectangle) are not depicted on their actual geographical location in order to increase the legibility of the visualization.**

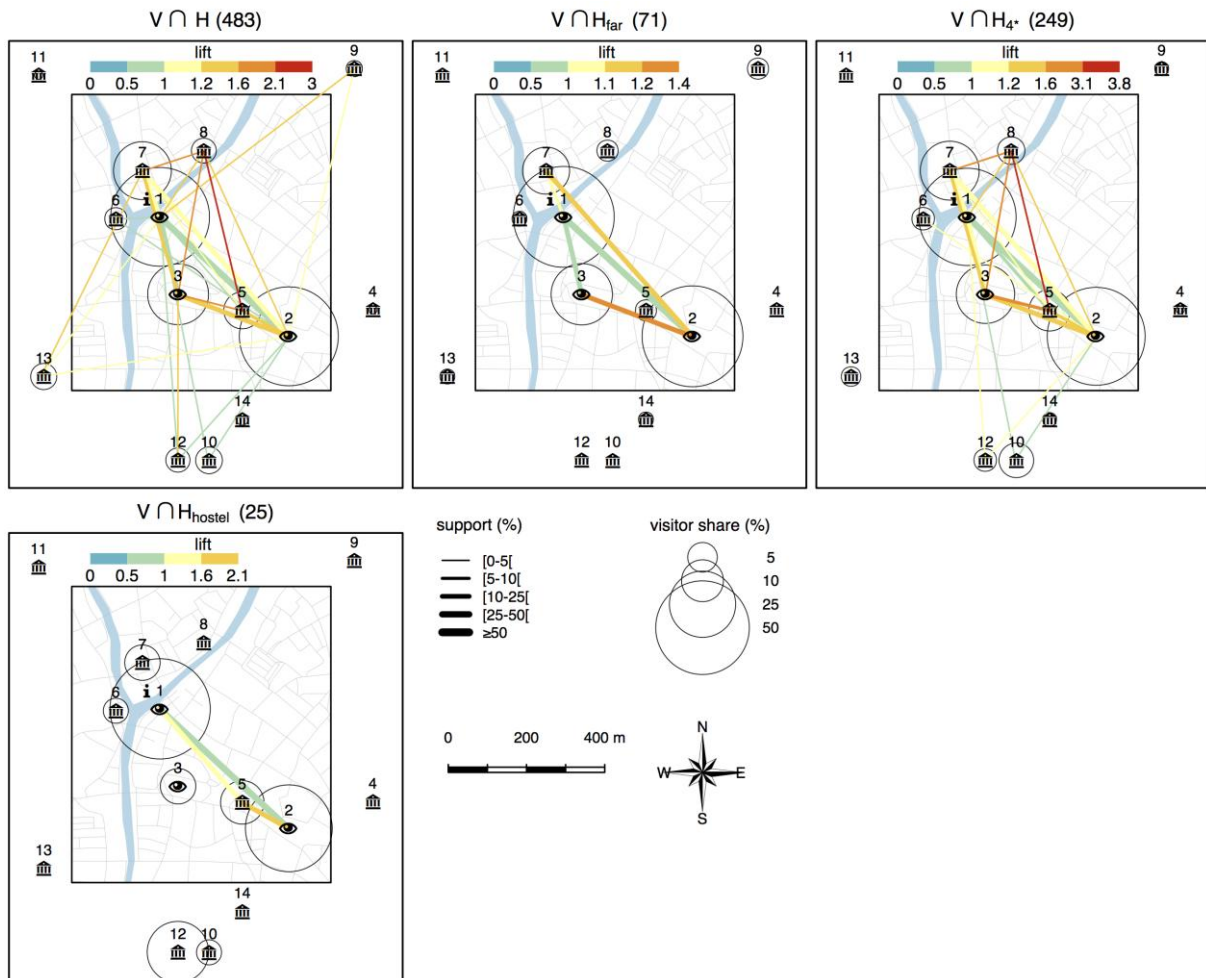**Table 3. Top-20 (where applicable) of association rules $X \Rightarrow Y$ for visitor segments $V, V_c, V_c \cap P_{1d}, V_c \cap P_{>1d}$, and $V_c \cap P_{>1d} \cap H$. Rules were generated based on constraints on both support ($s_a \geq 5$) and confidence ($c \geq 0.05$), and are sorted on their lift (l) value. The relative (s_r) and absolute (s_a) support, as well as the confidence (c) are also shown. Rules with only item in the antecedent (marked by a grey shading) are also visible in**

	V (5,891 devices)						V_c (2,095 devices)						$V_c \cap P_{1d}$ (1,564 devices)					
	X	Y	s_r	s_a	c	l	X	Y	s_r	s_a	c	l	X	Y	s_r	s_a	c	l
1	3,7,8	5	8.5E-04	5	0.63	10.40	2,5,7,8	3	2.4E-03	5	0.71	6.01	12,2	10	3.2E-03	5	0.50	4.80
2	2,3,7,8	5	8.5E-04	5	0.63	10.40	2,5,8	3	3.3E-03	7	0.64	5.35	1,5,7	3	5.8E-03	9	0.39	4.05
3	10,2,7	5	8.5E-04	5	0.56	9.25	10,5	3	2.4E-03	5	0.63	5.26	5,6	1	3.8E-03	6	0.75	3.83
4	10,7	5	8.5E-04	5	0.45	7.56	10,2,5	3	2.4E-03	5	0.63	5.26	1,2,5,7	3	5.1E-03	8	0.36	3.77
5	1,2,7,8	5	8.5E-04	5	0.45	7.56	1,2,5,8	3	2.4E-03	5	0.63	5.26	2,5,6	1	3.2E-03	5	0.71	3.65
6	12,2	10	1.4E-03	8	0.29	7.45	5,7,8	3	2.4E-03	5	0.56	4.67	10,2	3	5.8E-03	9	0.35	3.59
7	2,3,8	5	1.2E-03	7	0.44	7.28	1,5,8	3	2.4E-03	5	0.56	4.67	6,7	1	5.8E-03	9	0.69	3.54
8	1,7,8	5	1.0E-03	6	0.43	7.13	10,2,7	3	2.4E-03	5	0.56	4.67	1,2,5	3	1.5E-02	24	0.34	3.50
9	2,7,8	5	1.2E-03	7	0.41	6.85	1,2,7,8	3	2.9E-03	6	0.55	4.59	2,5	3	4.0E-02	62	0.34	3.49
10	3,8	5	1.2E-03	7	0.39	6.47	5,8	3	3.3E-03	7	0.54	4.53	2,6,7	1	3.8E-03	6	0.67	3.41
11	2,3,5,7	8	8.5E-04	5	0.15	6.47	1,2,5,7	3	1.1E-02	22	0.52	4.41	1,3,6	2	3.2E-03	5	1.00	3.37
12	1,3,8	5	8.5E-04	5	0.38	6.40	1,5,7	3	1.1E-02	23	0.52	4.40	1,5	3	1.8E-02	28	0.32	3.33
13	1,2,3,8	5	8.5E-04	5	0.38	6.40	2,7,8	3	3.8E-03	8	0.47	3.96	5,7	3	1.1E-02	17	0.31	3.26
14	3,5,8	7	8.5E-04	5	0.71	6.05	10,7	3	2.4E-03	5	0.45	3.82	5,7	2	3.3E-02	52	0.96	3.25
15	2,3,5,8	7	8.5E-04	5	0.71	6.05	3,7,8	5	2.4E-03	5	0.63	3.70	1,5,7	2	1.4E-02	22	0.96	3.22
16	3,5,7	8	8.5E-04	5	0.14	5.93	2,3,7,8	5	2.4E-03	5	0.63	3.70	3,5	2	4.0E-02	62	0.93	3.12
17	5,8	7	1.5E-03	9	0.69	5.87	5,7	3	1.7E-02	36	0.43	3.65	2,8	1	9.0E-03	14	0.61	3.11
18	14	12	6.1E-03	36	0.24	5.84	1,7,8	3	2.9E-03	6	0.43	3.61	1,2	3	3.5E-02	54	0.30	3.06
19	12	14	6.1E-03	36	0.15	5.84	2,5,7	3	1.6E-02	33	0.42	3.56	1,2,7	3	1.9E-02	29	0.29	3.03
20	1,5,7	8	1.0E-03	6	0.14	5.82	1,2,5	3	2.4E-02	50	0.42	3.51	10,3	2	5.8E-03	9	0.90	3.03

	$V_c \cap P_{>1d}$ (531 devices)						$V_c \cap P_{>1d} \cap H$ (196 devices)					
	X	Y	s_r	s_a	c	l	X	Y	s_r	s_a	c	l
1	1,2,7,8	3	1.1E-02	6	0.75	4.06	1,7,8	3	2.6E-02	5	0.83	3.02
2	2,5,8	3	9.4E-03	5	0.71	3.87	1,5,7	3	2.6E-02	5	0.83	3.02
3	2,7,8	3	1.3E-02	7	0.70	3.79	1,2,7,8	3	2.6E-02	5	0.83	3.02
4	1,2,5,7	3	2.6E-02	14	0.70	3.79	1,2,5,7	3	2.6E-02	5	0.83	3.02
5	2,5,7	3	3.4E-02	18	0.69	3.75	5,7	3	4.1E-02	8	0.80	2.90
6	1,5,7	3	2.6E-02	14	0.67	3.61	2,5,7	3	4.1E-02	8	0.80	2.90
7	5,7	3	3.6E-02	19	0.66	3.55	7,8	3	3.1E-02	6	0.75	2.72
8	1,7,8	3	1.1E-02	6	0.60	3.25	2,7,8	3	3.1E-02	6	0.75	2.72
9	5,8	3	9.4E-03	5	0.56	3.01	12,2	3	3.1E-02	6	0.60	2.18
10	1,2,5	3	4.9E-02	26	0.53	2.88	2,5	3	1.1E-01	21	0.60	2.18
11	2,5	3	6.8E-02	36	0.53	2.87	1,3,7	8	2.6E-02	5	0.21	2.15
12	3,8	5	9.4E-03	5	0.45	2.51	1,2,3,7	8	2.6E-02	5	0.21	2.15
13	2,3,8	5	9.4E-03	5	0.45	2.51	2,3,7	8	3.1E-02	6	0.21	2.13
14	1,2,7	3	7.5E-02	40	0.45	2.46	2,3	5	1.1E-01	21	0.42	2.06
15	1,2,8	3	1.7E-02	9	0.45	2.44	1,2,3	8	3.6E-02	7	0.19	2.01
16	1,5	3	4.9E-02	26	0.45	2.43	3,7	8	3.1E-02	6	0.19	2.00
17	2,3	5	6.8E-02	36	0.42	2.34	1,3	8	3.6E-02	7	0.19	1.95
18	1,2,3	5	4.9E-02	26	0.42	2.32	5	3	1.1E-01	21	0.53	1.91
19	2,8	3	2.1E-02	11	0.42	2.29	3	5	1.1E-01	21	0.39	1.91
20	5,7	8	1.1E-02	6	0.21	2.29	1,2,5	3	6.1E-02	12	0.52	1.89

Table 4. Top-20 (where applicable) of association rules $X \Longrightarrow Y$ for visitor segments $V \cap H$, $V \cap H_{far}$, $V \cap H_{4*}$, and $V \cap H_{hostel}$. Rules were generated based on constraints on both support ($s_a \geq 5$) and confidence ($c \geq 0.05$), and are sorted on their lift (l) value. The relative (s_r) and absolute (s_a) support, as well as the confidence (c) are also shown. Rules with only item in the antecedent (marked by a grey shading) are also visible in Figure 8.

	$V \cap H$ (483 devices)						$V \cap H_{far}$ (71 devices)						$V \cap H_{4*}$ (249 devices)					
	X	Y	s_r	s_a	c	l	X	Y	s_r	s_a	c	l	X	Y	s_r	s_a	c	l
1	1,3,7	8	1.0E-02	5	0.21	5.30	1,2	7	7.0E-02	5	0.21	1.48	1,3,7	8	2.0E-02	5	0.36	7.41
2	1,2,3,7	8	1.0E-02	5	0.21	5.30	3	2	2.1E-01	15	0.88	1.36	1,2,3,7	8	2.0E-02	5	0.36	7.41
3	2,3,7	8	1.2E-02	6	0.21	5.26	2	3	2.1E-01	15	0.33	1.36	2,3,7	8	2.4E-02	6	0.33	6.92
4	3,7	8	1.2E-02	6	0.19	4.92	1,3	2	9.9E-02	7	0.88	1.35	3,7	8	2.4E-02	6	0.32	6.55
5	2,5	8	1.0E-02	5	0.14	3.53	7	2	1.1E-01	8	0.80	1.23	1,2,7	8	2.0E-02	5	0.23	4.72
6	1,7,8	3	1.0E-02	5	0.83	3.53	2	7	1.1E-01	8	0.17	1.23	7,8	3	2.4E-02	6	1.00	4.61
7	1,5,7	3	1.0E-02	5	0.83	3.53	1,2	3	9.9E-02	7	0.29	1.22	1,7,8	3	2.0E-02	5	1.00	4.61
8	1,2,7,8	3	1.0E-02	5	0.83	3.53	1,7	2	7.0E-02	5	0.71	1.10	2,7,8	3	2.4E-02	6	1.00	4.61
9	1,2,5,7	3	1.0E-02	5	0.83	3.53	7	1	9.9E-02	7	0.70	1.08	1,2,7,8	3	2.0E-02	5	1.00	4.61
10	5,7	3	1.7E-02	8	0.80	3.39	1	7	9.9E-02	7	0.15	1.08	2,5	8	2.0E-02	5	0.22	4.51
11	2,5,7	3	1.7E-02	8	0.80	3.39	2,7	1	7.0E-02	5	0.63	0.96	1,2,3	8	2.8E-02	7	0.21	4.40
12	1,3,8	7	1.0E-02	5	0.71	3.35	1	2	3.4E-01	24	0.52	0.81	2,8	5	2.0E-02	5	0.45	4.19
13	1,2,3,8	7	1.0E-02	5	0.71	3.35	2	1	3.4E-01	24	0.52	0.81	1,3	8	2.8E-02	7	0.20	4.15
14	1,2,7	8	1.2E-02	6	0.13	3.25	3	1	1.1E-01	8	0.47	0.73	5,7	3	2.4E-02	6	0.86	3.95
15	2,8	5	1.0E-02	5	0.28	3.19	1	3	1.1E-01	8	0.17	0.73	2,5,7	3	2.4E-02	6	0.86	3.95
16	7,8	3	1.2E-02	6	0.75	3.18	2,3	1	9.9E-02	7	0.47	0.72	8	5	2.0E-02	5	0.42	3.84
17	2,7,8	3	1.2E-02	6	0.75	3.18							5	8	2.0E-02	5	0.19	3.84
18	2,3,7	5	1.7E-02	8	0.28	3.17							2,3	8	3.2E-02	8	0.17	3.61
19	3,8	7	1.2E-02	6	0.67	3.13							2,7	8	2.4E-02	6	0.17	3.56
20	2,3,8	7	1.2E-02	6	0.67	3.13							3,8	7	2.4E-02	6	0.75	3.40

	$V \cap H_{hostel}$ (25 devices)					
	X	Y	s_r	s_a	c	l
1	1,2	5	1.2E-01	3	0.43	3.57
2	5	2	1.2E-01	3	1.00	2.08
3	2	5	1.2E-01	3	0.25	2.08
4	1,5	2	1.2E-01	3	1.00	2.08
5	5	1	1.2E-01	3	1.00	1.56
6	1	5	1.2E-01	3	0.19	1.56
7	2,5	1	1.2E-01	3	1.00	1.56
8	2	1	2.8E-01	7	0.58	0.91
9	1	2	2.8E-01	7	0.44	0.91