



# *senseable* city lab:...

---

## An analysis of visitors' behavior in The Louvre Museum: a study using Bluetooth data

---

**Yuji Yoshimura, Stanislav Sobolevsky, Carlo Ratti**

SENSEable City Laboratory, Massachusetts Institute of Technology,  
77 Massachusetts Avenue, Cambridge, MA 02139, USA;  
e-mail: yyoshi@mit.edu, stanly@mit.edu, ratti@mit.edu

**Fabien Girardin**

Near Future Laboratory, CP242, 3960 Sierre, Switzerland;  
e-mail: fabien@nearfuturelaboratory.com

**Juan Pablo Carrascal, Josep Blat**

Information and Communication Technologies Department, Universitat Pompeu Fabra,  
Roc Boronat, 138, Tanger Building 08018 Barcelona, Spain;  
e-mail: jp.carrascal@upf.edu, josep.blat@upf.edu

**Roberta Sinatra**

Center for Complex Network Research and Department of Physics, Northeastern  
University, 110 Forsyth Street, Boston, MA 02115, USA; e-mail: r.sinatra@neu.edu  
Received 7 April 2013; in revised form 13 August 2013; published online 31 July 2014

---

**Abstract.** Museums often suffer from so-called 'hypercongestion', wherein the number of visitors exceeds the capacity of the physical space of the museum. This can potentially be detrimental to the quality of visitors' experiences, through disturbance by the behavior and presence of other visitors. Although this situation can be mitigated by managing visitors' flow between spaces, a detailed analysis of visitor movement is required to realize fully and apply a proper solution to the problem. In this paper we analyze visitors' sequential movements, the spatial layout, and the relationship between them in a large-scale art museum—The Louvre Museum—using anonymized data collected through noninvasive Bluetooth sensors. This enables us to unveil some features of visitor behavior and spatial impact that shed some light on the mechanisms of museum overcrowding. The analysis reveals that the visiting styles of short-stay and long-stay visitors are not as significantly different as one might expect. Both types of visitors tend to visit a similar number of key locations in the museum while the longer-stay visitors just tend to do so more time extensively. In addition, we reveal that some ways of exploring the museum appear frequently for both types of visitors, although long-stay visitors might be expected to diversify much more, given the greater time spent in the museum. We suggest that these similarities and dissimilarities make for an uneven distribution of the number of visitors in the museum space. The findings increase the understanding of the unknown behaviors of visitors, which is key to improving the museum's environment and visitor experience.

**Keywords:** Bluetooth tracking, visitor behavior, museum studies, human mobility, building morphology

### 1 Mesoscopic research of visitors' sequential movement in an art museum

Falk and Dierking argue that "a major problem at many museums is crowding, and crowds are not always easy to control" (1992, page 145). Museums and their exhibits, along with their own spectacular architecture, become some of the most popular destinations for the tourists, thus triggering 'hypercongestion' (Krebs et al, 2007), as the number of visitors often exceeds the capacity of spaces, which results in the museum becoming overcrowded.

---

Congestion in museums shows, on the one hand, high attractiveness and vitality, resulting in positive economic impact. On the other hand, the increased number of visitors implies potential negative effects which are detrimental to the quality of visiting conditions and the visitors' experience can be disturbed by the behavior and presence of other visitors (Maddison and Foster, 2003, pages 173–174). In an age when museums play an important role in mass cultural consumption and with urban regeneration and the promotion of the image of cities (Hamnett and Shoval, 2003), museums are expected to achieve seemingly contradictory objectives at the same time; that is, to increase the number of visitors and also enhance the quality of their experience by achieving comfortable visiting conditions through management of the flow of visitors.

Visitors' movement and circulation patterns in museums are recognized as an important topic for research (Bitgood, 2006, page 463). However, most of these studies conducted in art museums have been done for only two extreme cases: (a) visitor patterns at the macroscale to investigate the basic demographic composition of the museum's visitors (Schuster, 1995), along with psychographic factors which influence visit motives and barriers (Hood, 1983); and (b) at the microscale to research visitor circulation in the individual exhibition rooms, limited galleries, or other areas. This often results in revealing that: (1) the visitor's attributive features from a sociocultural point of view (ie, highly educated people and wealthy upper- or middle-class people tend to visit more frequently than people from the lower social classes) (Hein, 1998, page 115–116); and (2) there is a local interaction between the layout of the exhibits displayed in the galleries and the visitors' behavior in those spaces (Klein, 1993; Melton, 1935; Parsons and Loomis, 1973; Weiss and Boutourline, 1963). This polarized research resulted in a shortage of mesoscopic empirical analysis of visitors in large-scale art museums, which have different research targets compared with a single exhibition, small or medium-sized museums (Serrel, 1998; Tröndle et al, 2012), or other types of museum (Kanda et al, 2007; Laetsch et al, 1980; Sparacino, 2002).

Space syntax (Hillier, 1996; Hillier and Hanson, 1984) applies a different approach to analyzing the influences of the spatial layout and design of buildings using visitors movement and behavior by describing the overall configuration of the museum setting (for a review, see Hillier and Tzortzi, 2006). This type of knowledge is key to producing patterns of exploration and interaction of visitors, and the copresence and coawareness that exists between visitors in the museum environment as a whole (Choi, 1999).

Yet all these studies rely on a spatially and temporally limited dataset, which often results in providing just a snapshot of a limited area in the built environment. Even a simulation-based analysis uses a simplification of human behavior to estimate visitors' behavior rather than revealing actual patterns of movement with real-world empirical data.

In this paper we analyze the sequential movement of visitors, the spatial layout, and the relationship between them in order to clarify the behavioral features of visitors in a large-scale art museum—The Louvre Museum. We focus on visitors' circulation from the entrance to an exit as a whole mobility network rather than their movement in particular individual rooms. The way of visiting exhibits is analyzed by means of the visitors' length of stay and the sequences in which they make their visits, because these determine the visitors' perceptions and attentions that shape their visiting experience (Bitgood, 2006). The length of stay might be thought to be the key factor that determines the number of places visited and the sequence in which they are visited, which results in a variety of different routes; the more time you are given, the more opportunities you have, and vice versa. The question to be asked is whether this hypothesis is actually true, and by its extension, how the length of stay and the sequence of the places visited make visitors' mobility style different, and how this dissimilarity is seen

---

in the museum. This understanding might be the key to improving the museum environment, as well as to enhancing visitors' experiences.

We employ a systematic observation method relying on Bluetooth proximity detection, which makes it possible to produce large-scale datasets representing visitors' sequential movement with low spatial resolution. 'Large-scale datasets' refers to the sample size we used being much larger than those collected in art museums for previous studies [eg, almost 2000 in Melton (1935); 689 in Serrell (1998); 576 in Tröndle et al (2012); and 50 in Sparacino (2002)], although each of them contains different types of information with sufficient resolution for their particular objectives and as good as human-based observation, GPS, RFID, or ultra-wideband technology can achieve. In our work we explore the global patterns of visitors' behaviors by increasing the quantity of the data, because "when we increase the scale of the data that we work with, we can do new things that weren't possible when we just worked with smaller amounts" (Mayer-Schönberger and Cukier, 2013, page 10).

Thus, we limit our research to dealing with visitors' physical presence in and between places, without questioning the introspective aspects (eg, learning process, making meaning from the experience of the museum), which the previous studies tried to answer by small-scale sampling [see Kirchberg and Tröndle (2012) for a review]. However, the superimposition of large amounts of data about individuals' movements over time allows some patterns to appear to be self-organizing in a bottom-up way from seemingly chaotic, disordered, and crowded movement. These results could shed light on the quality of visit conditions derived from overcrowding, not only around the spots where the iconic art works are placed, but also the spaces in the network between them that have dynamic visitor flow. A better understanding of visiting features would help in designing more adequate spatial arrangements and give insights to practitioners on how to manage visitor flow in a more efficient and dynamic way.

## **2 Visitor's sequential movement and analysis framework**

The use of large-scale datasets enables us to discover and analyze frequent patterns in human activities. Such analyses have been conducted in the specific spatiotemporal limitations derived from the limited measurement of mobile objects (Miller, 2005), in different contexts and at various scales. These analyses have shed light on unknown aspects of human behavior to discover patterns in human mobility (González et al, 2008; Hoteit et al, 2014; Kung et al, 2014), communication (Ratti et al, 2010; Sobolevsky et al, 2013), and urban activities (Grauwin et al, 2014; Ratti et al, 2006; Pei et al, 2014) by studying cell-phone usage at the regional scale. Other data like social media (Hawelka et al, 2014) or bank card transactions (Sobolevsky et al, 2014) have also been used. In particular, the sequential patterns of tourists at the local scale has been studied by looking at the number of locations visited, their order, and the length of stay, obtained from GPS data (Shoval et al, 2013), and, for instance, some aspects of customers' purchasing behavior in a grocery store have been disclosed by analyzing the customer's path, length of stay, and the categories of products purchased through RFID data (Hui et al, 2009).

Previous research (Yoshimura et al, 2012) proposed a Bluetooth-based data-collection technique in a large-scale art museum at the mesoscopic scale in order to classify visitors' behavior by their most-used paths and their relationship with the length of stay. Bluetooth data collection is based on systematic observation which detects Bluetooth-activated mobile devices, in the framework of 'unobtrusive measures', making use of the *digital footprint* unconsciously left by visitors. A considerable number of studies have employed this method but not in the context of large-scale art museums. Examples include measuring the relationship between peoples' social networks (Eagle and Pentland, 2005; Paulos and Goodman, 2004), analyzing mobility of pedestrians (Delafontaine et al, 2012; Kostakos et al, 2010; Versichele et al, 2012), and estimating travel times (Barceló et al, 2010).

---

A Bluetooth proximity-detection approach to the analysis of visitor behavior in museums has many advantages. Contrary to the granular mobile-phone tracking (Ratti et al, 2006), the detecting scale using Bluetooth is much more fine grained. In addition, in contrast to RFID tags (Hui et al, 2009; Kanda et al, 2007) and active mobile-phone tracking with or without GPS (Asakura and Iryob, 2007), with Bluetooth previous registration is not required and it is not necessary to attach any devices or tags. The fact that no prior participation or registration is required enables a mass participation of subjects and the collection of an enormous amount of data in the long term, unlike time constrained cases (McKercher et al, 2012; Shoval et al, 2013). Also, the unobtrusive nature of Bluetooth removes bias in the data, which could be created if a subject is conscious of being tracked. Furthermore, Bluetooth proximity detection succeeds inside buildings or in the proximity of tall structures, where GPS connectivity is limited. All these advantages make this method adequate for detecting visitors' sequential movement between key places, without specifying their activities, attributes, or inner thoughts, in a consistent way at the mesoscopic scale in a large-scale art museum.

We identify a visitor's length of stay at a particular location as the indicator for measuring their interest level at that exhibit by merely accounting for their presence without questioning their inner thoughts. We estimate visitors' routes between sensors and time at the place from the collected data.

As our analysis and interpretation of the data were conducted within a specific spatiotemporal framework, our approach has some limitations. Firstly the concept of trajectory used in this paper is different from the one usually available when working with data collected by GPS systems. This is because a Bluetooth proximity sensor just provides the time-stamped sequence of individual transitions of a mobile device between nodes (eg, sequence of A–B–D), while a GPS system can track all the movements of a device. However, the network of rooms derived from the spatial layout of the museum determines the feasible routes, and this enhances estimation of the paths used by visitors between sensors without observing their exact trajectories and orientations per room (Delafontaine et al, 2012). Secondly, we cannot deal directly with visitors' introspective factors, their expectations, experiences, and satisfactions (Pekarik et al, 1999). This results in excluding from our study research questions about 'wayfinding', which refers to a visitor's ability to find his or her way within a setting, and 'orientation', which indicates an available knowledge in a setting through the use of the hand-held maps and direction signs, because they consist of the complex interaction between environmental cognition and the orientation devices. In addition, a visitor's presence at a specific place is not necessarily related to their time engaging with the exhibits, although previous studies used this to measure visitor interest (Melton, 1935; Robinson, 1928). Finally, our sample is possibly biased in two ways. First, the sample composition is affected by the segments of the mobile-device holders and their decision to activate or not activate the Bluetooth function. Although the latter requires calculating the sample representativeness and is typically conducted by using a short-term manual counting method (Versichele et al, 2012), we employed a long-term (one-month) systematic comparison of the number of devices detected at the entrance with the official museum head count and ticket sales. This method provided us with more comprehensive information compared with previous research.

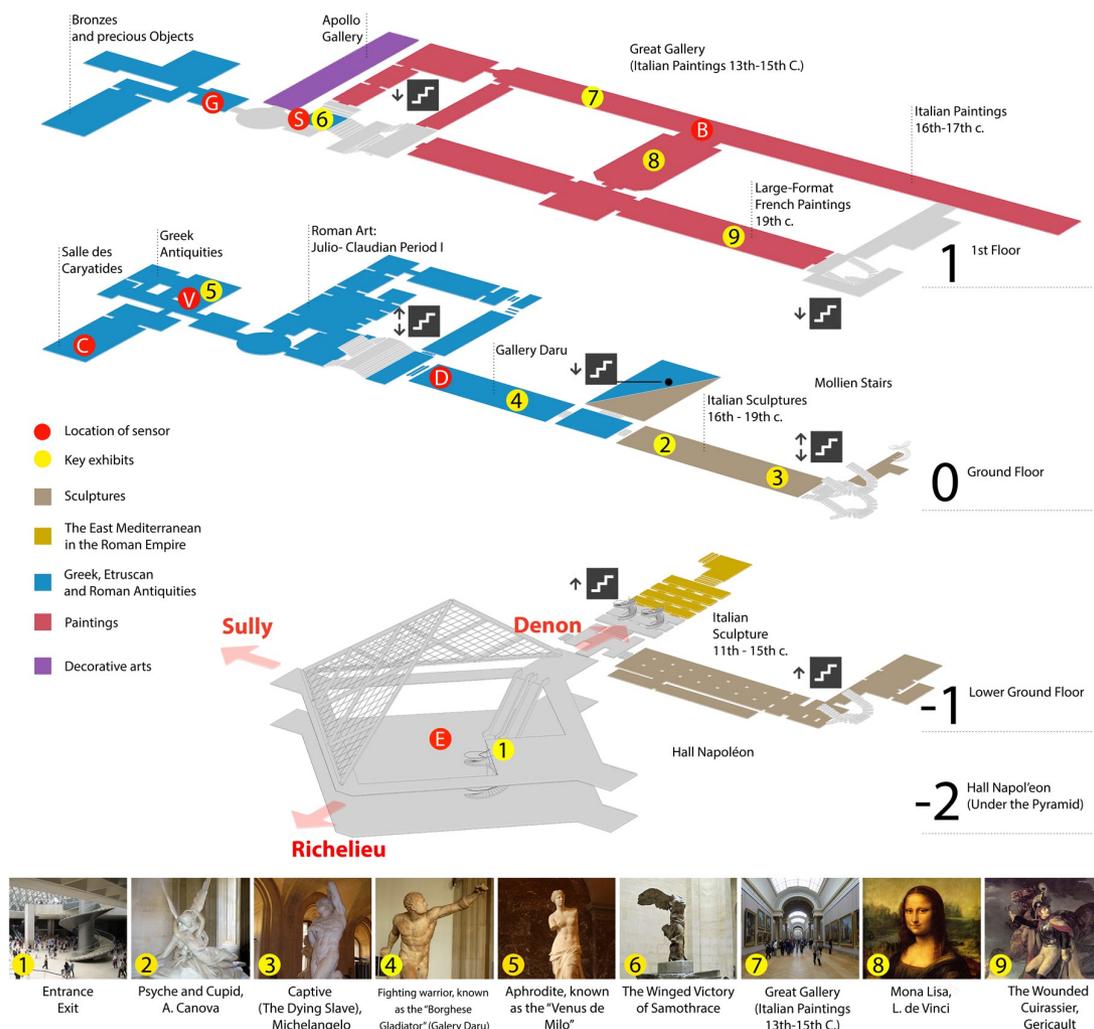
### **3 Concept definitions and data settings**

In this section we define the locations of sensors used and the components of the dataset in order to explore our method and data consistency. We collected our dataset during a specific period and processed it into a specific form required for the analysis.

### 3.1 Sensors settings in the museum and definition of node

Figure 1 shows the location of seven sensors, deployed throughout the museum, covering key places for detecting visitors. They are situated in one of the busiest trails, identified by The Louvre Museum authorities, which lead visitors from the entrance to the Venus de Milo; Entrance Hall (E), Gallery Daru (D), Venus de Milo (V), Salle des Caryatides (C), Great Gallery (B), Victory of Samothrace (S), and Salle des Verres (G).

Each sensor defined a detection area, identified as a *node*, approximately 20 m long and 7 m wide. The area varied in size, depending on the museum settings and the location of the sensor (eg, inside functional wooden boxes, desks, or in open space). However, all sensors covered targeted areas along the paths to key iconic art works. Once a Bluetooth-activated mobile device enters a detection area, the sensor receives the signal emitted by the mobile device and the detection continues until the device leaves the area. The sensor registers the time at which the signal from the mobile device first appears, called the check-in time, and when the signal disappears, called the check-out time; the time difference between each mobile device's check-in and check-out times can be calculated to define the length of stay at the node. Similarly, by looking at the first check-in time and the last check-out time for a mobile device over all nodes, provided that the first and last nodes correspond to an entry



**Figure 1.** [In color online.] Location of seven sensors E, D, V, C, B, S, and G, indicating their approximate sensing range.

point and an exit from the museum, respectively, it is possible to calculate how long a visitor stays in the museum. The series of check-in and check-out times registered for a mobile device by all the sensors makes it possible to construct a visitor's trajectory through the museum. In addition to the length of the stay, the sensors time stamps allow calculation of the travel time between nodes. The synchronization of all sensors makes it possible to perform fine-grained time-series analysis. All this information can be achieved without invading visitor privacy, because the SHA algorithm (Stallings, 2011, pages 342–361) is applied to each sensor where the MAC ID is converted to a unique identifier (Sanfeliu et al, 2010).

### 3.2 Collected sample

We collected data over 24 days; from 30 April to 9 May 2010, 30 June to 8 July 2010, and 7 August to 18 August 2010. We selected data starting and finishing at node E in order to measure the length of stay in the museum. Consequently, 24 452 unique devices were chosen to be analyzed for this study. On average, 8.2% of visitors activated Bluetooth on their mobile device while in The Louvre Museum (Yoshimura et al, 2012).

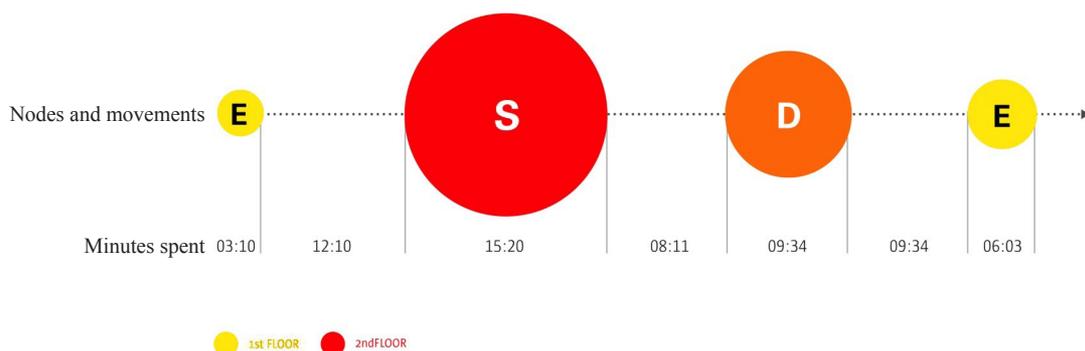
#### 3.2.1 Data clean up

The data collection was performed at different periods by a different number of sensors. We checked for possible synchronization issues arising from a lack of calibration, then adjusted the data to remove any inconsistencies. Finally, we only used data from visitors who started from node E and finished at node E in order to measure the complete length of the visit to the museum—such entries indicate that the visitor was correctly registered when he (or she) entered, moved around inside, and left the museum.

#### 3.2.2 Data processing

Figure 2 graphically shows the features of the logged data. It displays all entries in the database for a visitor for one day. Each lettered circle symbolizes detection at the corresponding node. It shows that this particular visitor made a sequential movement, E–S–D–E, and stayed at node E for 3 min 10 seconds, node S for 15 min 20 seconds, node D for 9 min 34 seconds and, again, node E for 6 min 3 seconds. The travel times between corresponding nodes were: 12 min 23 seconds for E–S, 8 min 11 seconds for S–D, and 9 min 34 seconds for D–E.

We built a database and designed a query engine to extract and transform the data for the different stages of the analysis. Table 1 shows an example of the components of the transformed dataset. There is one entry per visitor, and it includes the date of the visit, the path followed through the museum, the time of entry (check-in), time of exit (check-out), and the total length of the visit to museum.



**Figure 2.** [In color online.] Visualization of a relationship between the sequential movement and the time of stay of a visitor.

**Table 1.** Example of the dataset.

Reference	Date	Path	Check-in time	Check-out time	Length of visit
Unique ID	30 April 2010	E-S-D-E	09:04:35	11:07:52	02:03:17

### 3.3 Partitioning of visitors

In order to find the characteristics, the typical patterns of visits, and other determinant features of visitor behavior, we examined two extreme groups. Firstly, we sorted all the visits of our sample (24 452 visits) by their total time spent in the museum. By binning them into deciles, we obtained equal-sized clusters of approximately 2446 visits. Referencing all visits, those found in the first decile are called 'short visits' and we refer to these visitors as 'short-stay visitors'. Similarly, we refer to the visits in the tenth decile as 'long visits' and to these visitors as 'long-stay visitors'.

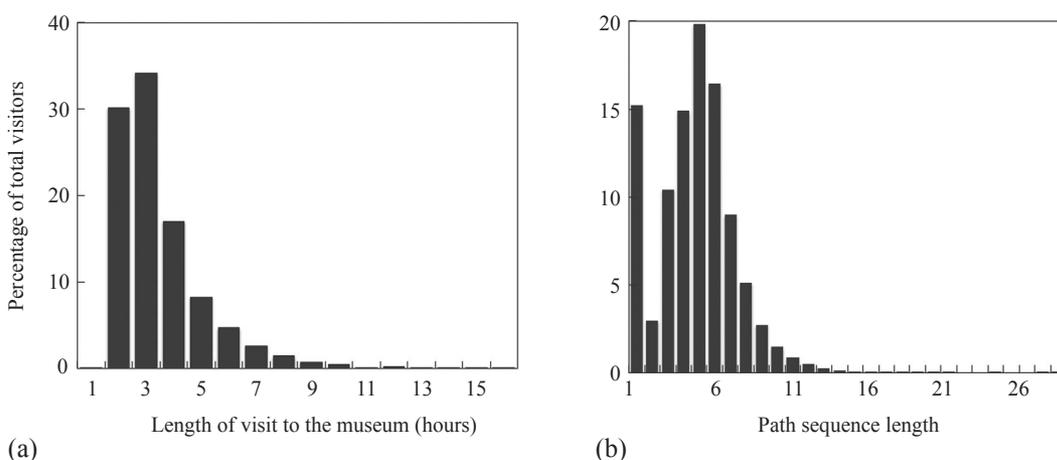
## 4 Results

In the following subsections, we present an overview of the statistical analysis built around the previously described dataset. We discuss the path sequence length, which is the number of nodes visited, including multiple visits executed without returning to E, the length of the visitors' paths, and the frequency of the appearance of each path. The distribution of the path sequence length is also presented and analyzed. We reveal visiting patterns, and the similarity and dissimilarity of the behaviors of the long-stay visitors and the short-stay visitors.

### 4.1 Basic statistics of visitors' behavior

We analyzed all visitor data to capture the features of their behavior, focusing on the path sequence length and its relationship with the length of stay in the museum.

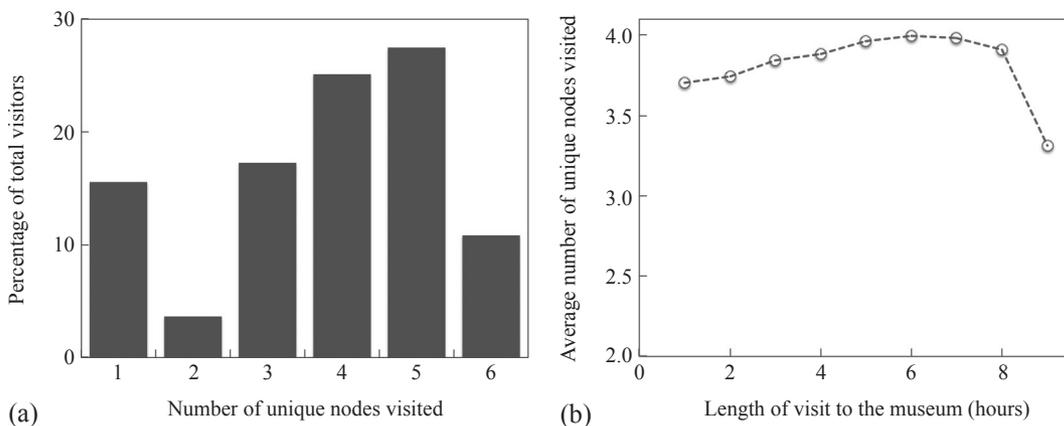
Figure 3(a) shows the distribution of the number of visits ( $y$  axis) to the length of stay in the museum binned for each hour ( $x$  axis). Although the maximum length of stay is more than 15 hours, only 410 visitors stayed for more than 8 hours, which corresponds to 1.6% of the total. Conversely, the minimum length of stay of less than 1 hour was for only one visitor, while more than 30% of visitors stayed for 1–2 hours. These facts indicate that the extreme visitors, whose length of stay is more than 8 hours or less than 1 hour, can be aggregated for the statistical reliability without substantially affecting the time-sensitive behavioral analysis. The distribution of the length of stay is positively skewed, with the majority of the visitors staying for 4–6 hours.



**Figure 3.** (a) The distribution of visits against the length of stay in the museum. (b) The distribution of the path sequence length.

Next, we look at the distribution of the path sequence length (number of nodes) [see figure 3(b)]. Although the maximum length of the path sequence length is thirty nodes, the percentage of visitors who visited more than fifteen nodes was only 0.5%. In general, this plot shows a distribution slightly skewed to the right, but visitors who visited only one node appear quite frequently, covering 15.2% of the total. Very few people visited two nodes (2.9%). However, the length of the sequence by itself does not necessarily reveal the size of the visitor mobility area, because a visitor could easily move between nearby nodes frequently without radially expanding throughout the museum.

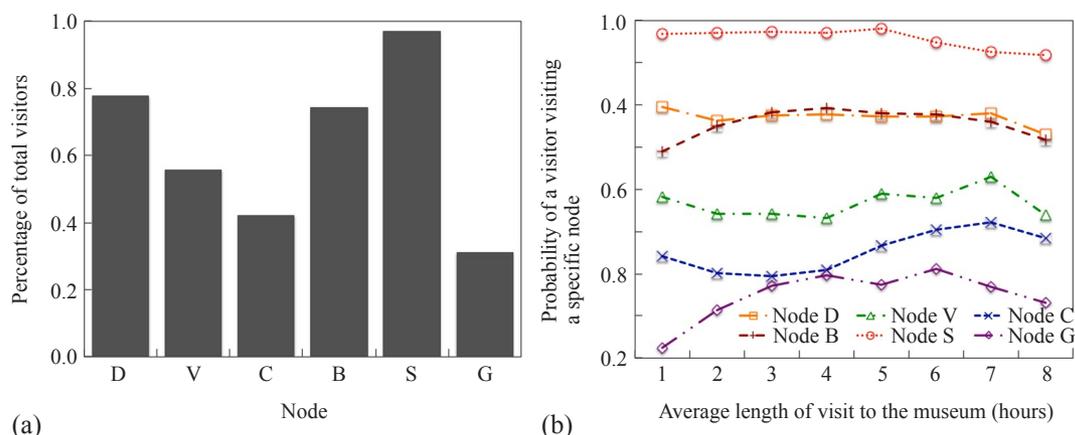
Figure 4(a) represents the number of unique nodes visitors passed during their stay in the museum. We can observe that visiting two nodes rarely happened, while visiting one or three nodes have almost the same frequency. The most frequent number of unique nodes visited is four or five nodes, while visiting all six nodes rarely happens. This indicates that in most of the cases some factors prevent the exploration of all the nodes, while all nodes but one could be explored much more often. In addition, figure 4(b) reveals that the average number of unique nodes visited against the duration of the visit is almost constant. The correlation coefficient between these two variables (Spearman's correlation = 0.072,  $p$ -value  $< 2.2 \times 10^{-16}$ ) indicates that the unique number of visited nodes is independent of the duration of the visit to the museum, and vice versa. Surprisingly, the long-stay visitors usually visit even fewer nodes than the short-stay ones.



**Figure 4.** (a) Distribution of the number of unique nodes visited other than node E. (b) The average number of unique nodes visited against the duration of the visit.

Figure 5(a) shows the frequency of visits for each node: 97% of all visitors passed node S. Nodes D and B are visited frequently (nearly 80% for each). On the other hand, node G is the most rarely visited, with just 30% of all visitors. Figure 5(b) presents the attractivity of the nodes depending on the duration of the visit. As we can see, for most nodes the probability of visiting does not depend on the length of stay in the museum as the probability is nearly constant for all nodes. Node G behaves differently from the others, as its probability of attracting visitors increases with the visitors' length of stay in the museum. This shows that short-stay visitors show a lower tendency to visit node G, while long-stay visitors seem more attracted to visit this node (perhaps having more time to explore this part of the museum), although its frequency does not surpass 40%, regardless of the visitor type.

We can observe the difference in the transition rates (probability of moving to the given destination node right after visiting the given origin) from any node to node G for the two types of visitors (see table 2). All the transition rates increase as the visitor's length of stay increases; nodes D, C, and S show substantial increases (shown in bold in table 2).



**Figure 5.** [In color online.] (a) The frequency of visits each node receives. (b) The frequency of visiting different nodes at least once against the duration of stay.

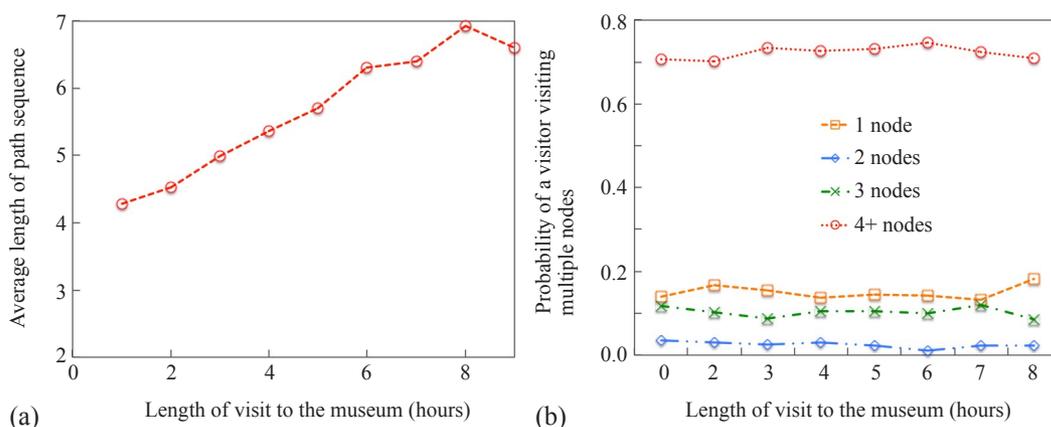
**Table 2.** Two types of visitors' transition rate from previous nodes to node G expressed as a percentage. Bold type indicates a substantial increase.

Node before node G	Short-stay visitors (%)	Long-stay visitors (%)	Difference (%)
D	4.00	7.17	<b>3.17</b>
V	1.38	3.17	1.79
C	4.86	9.91	<b>5.05</b>
B	5.60	6.30	0.70
S	2.53	5.69	<b>3.16</b>

**4.2 Similarity of visitor behaviors**

By looking at the path length of the visitors of different stay time we find another surprising effect. Although the path length increases slightly with increased length of visit, the path length of long-stay visitors is not substantially longer than that of the short-stay visitors. In addition, the number of nodes that make up a visit is very similar.

Figure 6(a) reveals that, while visitors tend to visit, on average, 4.3 nodes when visiting the museum for 1–2 hours, they are likely to visit only 5.5 nodes when they stay for 3–7 hours.



**Figure 6.** [In color online.] (a) The average length of path sequence (y axis) against the average length of stay in the museum (x axis). (b) The probability of a visitor's path length being 1, 2, 3, or more nodes by their length of stay in the museum (x axis).

The longer length of stay is three times the shorter, but it results in an increase of only 28% in the sequence length. In addition, the long-stay visitors (ie, 9–10 hours) visited 6.6 nodes on average, which is even less than the 8–9 hour visitors. The path sequence length increases as the duration of the visit increases, but the rate of change is not substantial (see table 3) especially if compared with the increases in visit times. Figure 6(b) presents the probability of visitors having a certain path length versus their length of stay in the museum. The probability of visiting 1, 2, 3, or more nodes against the length of stay aggregated by each hour appears almost flat, suggesting it is independent of the duration of the visit to the museum. We can also observe this tendency by examining the frequently appearing paths of the short-stay and long-stay visitors (table 4).

**Table 3.** The average length of path sequence (number of nodes visited) per hour and its percentage of increase.

Length of visit (hours)	Path sequence length (nodes)	Percentage increase
1–2	4.28	5.84
2–3	4.53	9.93
3–4	4.98	7.63
4–5	5.36	6.34
5–6	5.70	10.53
6–7	6.30	1.43
7–8	6.39	8.29
8–9	6.92	–4.62

**Table 4.** The probability of a visitor having a path length of 1, 2, 3, or more by the length of their stay in the museum.

Length of visit (hours)	Probability of path length			
	1 node	2 nodes	3 nodes	>3 nodes
1–2	0.14	0.03	0.11	0.70
2–3	0.16	0.03	0.10	0.70
3–4	0.15	0.02	0.08	0.73
4–5	0.13	0.03	0.10	0.72
5–6	0.14	0.02	0.10	0.73
6–7	0.14	0.01	0.10	0.74
7–8	0.13	0.02	0.12	0.72
8–9	0.18	0.02	0.08	0.70
Average	0.14	0.02	0.10	0.72
Standard deviation	0.01	0.00	0.01	0.01

Table 5 presents the top five most frequently appearing paths for both the short-stay and long-stay visitors. We counted the number of paths which appear in both groups (visited at least four nodes or visited less than four nodes) and divided by the total number of visitors in the group (ie, 2445), in order to obtain the frequency of a path appearing. This reveals that both groups have similar frequent path length; the short-stay paths are just slightly shorter compared with the long-stay paths. For both groups, the first and second most frequently

**Table 5.** Top five of the frequently appearing paths for paths of four nodes or more and for paths less than four nodes for the long-stay and short-stay visitors.

Path of long-stay visitors	Frequency (%)	Path of short-stay visitors	Frequency (%)
<i>Visitors whose length of path is 4 nodes or more</i>			
E-D-S-B-D-V-C-E	2.23	E-D-S-B-D-V-C-E	8.26
E-D-S-B-D-E	1.84	E-D-S-B-D-E	6.89
E-D-S-B-D-V-E	1.50	E-D-S-B-V-E	5.57
E-D-S-B-D-S-E	0.89	E-S-D-V-C-E	4.67
E-D-S-B-D-G-E	0.78	E-C-V-D-S-B-E	3.71
<i>Visitors whose length of path is less than 4 nodes</i>			
E-S-E	45.10	E-S-E	36.34
E-D-S-B-E	12.38	E-D-S-B-E	16.62
E-S-E-S-E	7.64	E-V-D-S-E	4.51
E-B-E	5.19	E-G-S-B-E	4.12
E-G-S-B-E	4.28	E-S-B-E	3.35

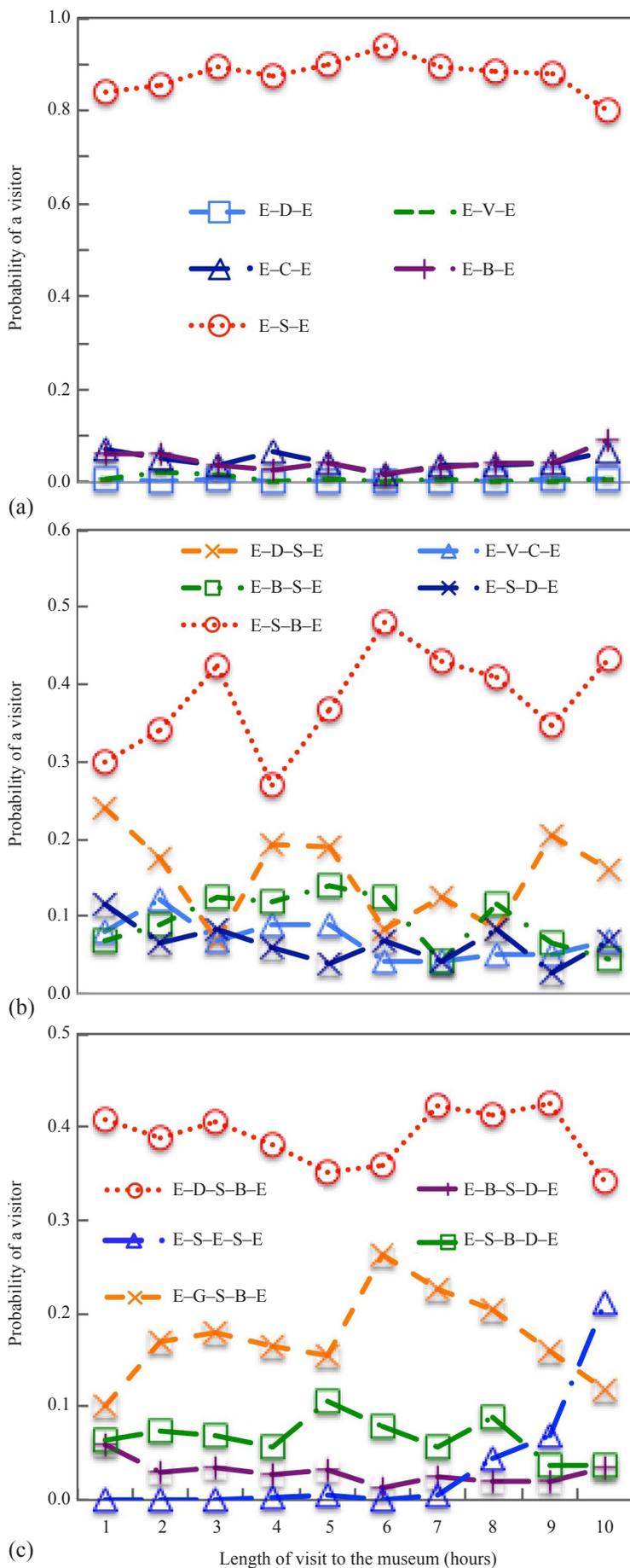
appearing paths for the long-stay and short-stay visitors are very similar, otherwise the frequency of the group that visited more than four nodes is much lower than for those who visited less than four nodes. The results show that the behavioral ways of short-stay and long-stay visitors are not as significantly different as one might expect. Both types of visitors tend to visit the same number of popular places but the long-stay visitors just tend to do so more time extensively (spending longer studying the exhibits).

We examine in more detail the visitors whose path length is less than four nodes. Within them, the most frequently appearing path for each category (ie, visited 1, 2, or 3 nodes) coincides well between the groups of short-stay and long-stay visitors. Figures 7(a), (b), (c) present the probability of visiting 1 node, 2 nodes, or 3 nodes, respectively. We can observe that in each case only one path has a strong influence on the probability as a whole, especially in figure 7(a), where 89.4% of those visitors took the path E-S-E).

Similarly, visitors who followed the E-S-B-E path, which is the most frequently appearing path for those visiting two nodes (37.52%), added node B at the end of their visit, while for E-D-S-B-E, the most frequently appearing path for those visiting three nodes (38.94%), nodes D and B were added at the beginning and end of their visit, respectively. There is no clear difference between long-stay and short-stay visitors to 1, 2, or 3 nodes; rather, their behavior seems very similar, other than the substantial difference in the length of the visit to the museum.

## 5 Discussion

The previous sections revealed that many features of the behavior of the long-stay and short-stay visitors, including the path sequence length and the unique nodes visited, do not appear to be strikingly different between visits of different duration, and are sometimes even independent or nearly independent of duration. In this section we show that visitors' paths and their variations are quite selective, with visitors mostly choosing the same paths in terms of the path sequence length and sequential order although many other options exist. This creates an uneven distribution of visitors among spaces, and is possibly one of the main causes of high congestion and vacant spaces in the museum.



**Figure 7.** [In color online.] The probability that visitors take particular path lengths visiting (a) 1 node, (b) 2 nodes, (c) 3 nodes, versus the length of their visit to the museum.

---

### 5.1 Uneven spatial distribution of visitors

The interplay between sensor locations and the spatial layout of the museum determines the specific and possible route(s) used by visitors. All sensors were placed logistically in the determinant positions for visitors' route choice in the museum. Therefore, the transition between two places makes it possible to estimate the determinant route that visitors take. Thus, we can clarify the uneven spatial use of the museum accesses for visitors' entry and exit behaviors by analyzing the first two and last two locations, respectively, in their sequence: 71.6% of visitors took E–D, E–B, or E–S, meaning that they entered through the Denon access, which indicates that only 28.3% used the Sully or Richelieu access (ie, E–V, E–C, E–G). 57.3% of visitors exited from Denon, 14.3% fewer than entered at Denon. This technique enables us to determine the rooms visited without observing their exact trajectories. Also, this indicates that we could speculate on the volume of visitors and their concentration along specific paths without knowing the exact load per room.

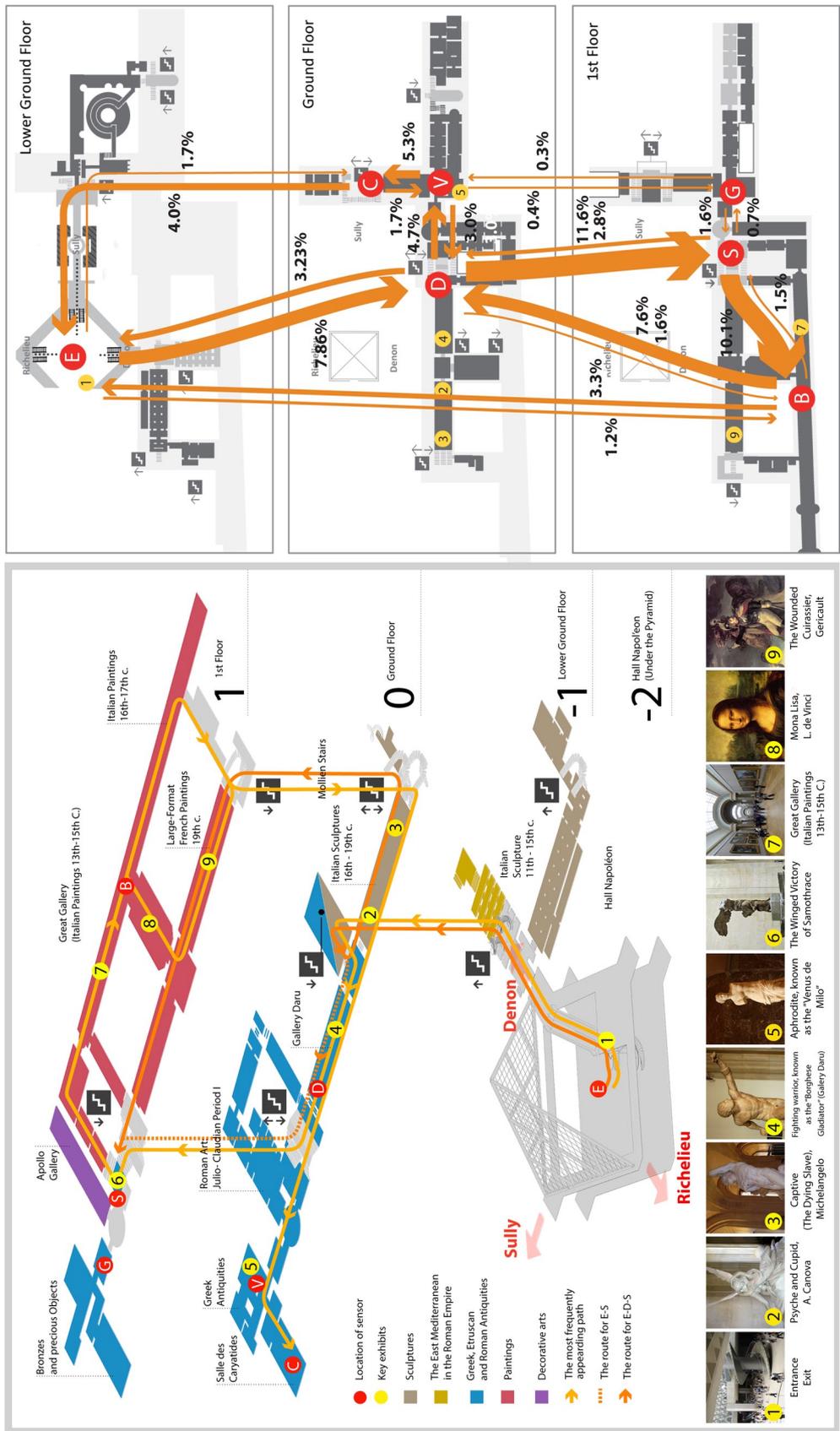
In the previous section it was revealed that 13.5% of all visitors to the museum only visited the Victory of Samothrace (node S—one of the most iconic exhibits in the museum), not visiting any of the other five nodes. Considering the museum's spatial layout, these visitors used the Mollien stairs, which connect the 16th-century to 19th-century Italian sculpture rooms on the ground floor to the 19th-century French painting room on the first floor, to visit node S instead of using the Victory of Samothrace staircase where node D is located (see the orange line in figure 8).

From the spatial point of view this is an intriguing result because the shortest path from the entrance (node E) to the Victory of Samothrace (node S) is the one which passes node D, meaning that they turned to the left at the intersection between exhibit No. 2 and No. 4 (see the orange dotted line in figure 8). To use the route through the Mollien stairs signifies a detour, both spatially and temporally, to reach node S.

Table 6 reveals visitors' route choice in more detail; from node E almost 40% of visitors turned to the left to reach node D (ie, E–D), while around 20% of visitors turned to the right (ie, E–S). Again, there is no significant difference between the behaviors of the long-stay and short-stay visitors, meaning that both start their museum experience in a similar way. In addition, since nodes D, V, B, and G are installed in some key points after exiting the Denon Wing, all those visitors whose path was E–S–E tended to stay in a very confined area of the Denon Wing during their visit. They just explore and stay in the small area during their visit, and this tendency is stronger for the long-stay visitors than the short-stay ones (see table 5).

On the other hand, the most frequently appearing path for both groups who visited at least four nodes is E–D–S–B–D–V–C–E, where the visitor visited the Gallery Daru, the Victory of Samothrace, the Great Gallery, and the Venus de Milo (see the yellow line in figure 8). This path starts from a trail of E–D and finishes with C–E, indicating that the visitor entered the museum at the Denon access, and exited from the Richelieu or Sully access. This suggests that these visitors tend to explore the museum extensively through covering most of the iconic exhibits rather than staying in only one part of the museum. In addition, the frequency of this path for the short-stay visitors is much higher than that of the long-stay visitors. This could indicate that the short-stay visitors might tend to select the most spatially optimized paths to visit all the possible iconic exhibits within their limited available time in the museum.

We believe that short-stay visitors explore fewer of the popular places due to the limited time that they have to spend in the museum. This is intuitive since a visitor's movement and their activities would be limited when the length of their visit to the museum is short. Consequently, the trajectories of the long-stay visitors would be expected to be more complex than those of the short-stay visitors, and vice versa. However, the results show that the behavioral patterns of short-stay and long-stay visitors are not as significantly different as



**Figure 8.** [In color online.] (a) The map of the spatial layout of The Louvre Museum and the routes visitors used. (b) The transition percentages between locations, showing only major links between each pair of nodes.

**Table 6.** Three types of visitor transition rates from node E to the subsequent node expressed as a percentage.

Subsequent node from node E	All visitors	Short-stay visitors	Long-stay visitors
D	43.32	42.41	40.34
V	11.25	12.80	11.38
C	9.59	10.02	11.32
B	6.80	9.20	7.51
S	21.53	21.02	20.82
G	7.51	4.54	8.63

one might expect. Both types of visitor tend to visit a similar number of the popular rooms, but the long-stay visitors tend to do so more time extensively.

The results imply that visitors' trajectories seem to be quite limited in terms of the path sequence length and its order, although there exist a number of possible routes including repeating nodes. More generally, we might say that—and this partially agrees with Choi's (1999) statement—the more spaces available, the more the visitor's path tends to be selective. That is, when the number of the rooms with exhibits increases, visitors seem not to visit them all, but visit a few of them selectively. But our findings tell us more; these limited paths and their use are almost independent of the length of the visit to the museum, meaning that most visitors, irrespective of whether their visit is short or long, tend to use the same trajectories.

We speculate that this similarity/dissimilarity of the patterns makes the distribution of the quantity of visitors in the museum space uneven; for instance, the route E–D–S–B–D is frequently observed, independent of the length of the visit, suggesting that there can be a high concentration of visitors in those enclosed areas. In contrast, some spaces can be found to be quite vacant; the sequential pattern between node S and node G is rarely found, especially, in the short-stay visits. This indicates that the topological proximity and the attractiveness of a node can be changed depending on the visitor's length of stay (see figure 5). It could be that node G, which tends to be visited when people have more time to explore the museum, is not seen as a necessity or 'priority' during the museum visit. Thus, the distribution of visitors is uneven and the number of visits that each room receives varies.

## 6 Conclusion

In this study we examined visitors' mobility styles and their respective spatial impacts by analyzing large-scale datasets obtained through Bluetooth proximity detection in a bottom-up methodology. This analysis and the results obtained give a great scientific advancement to improving visiting conditions, which strongly affect the quality of a visitor's experience in the museum.

The results indicate that the behavior of short-stay and long-stay visitors is not as different as one might expect. The path lengths grow at a much slower rate compared with increasing duration of stay. Even more surprisingly, the number of unique nodes visited remains almost constant, independent on the length of the visit. The correlation coefficient between these two variables quantitatively indicates that the unique number of nodes visited is independent of the duration of the visit to the museum, and vice versa. Both short-stay and long-stay groups visit mostly the same number of sensor locations, while the long-stay visitors just tend to do so more time extensively. Moreover, the probability of the appearance of visitors whose path sequence length is small (<4 nodes), is constant across all time divisions, meaning that there always exists a certain category of visitors who do not try to explore museum space extensively no matter how much time they have to do so. Also, we discovered that the

---

frequency of node visits per hour is almost constant and independent of the length of time spent in the museum.

Conversely, we can point out key differences in visitors' behavior within each of two groups—those who visited more than four nodes and those who visited fewer than four. The average number of locations visited, for each of the groups, does not depend on the time people have to spend in the museum (ie, it is independent of a visitor being classified as a short-stay and long-stay visitor). For both short-stay and long-stay visitors the most frequently occurring path in the group that visited at least four nodes is E–D–S–B–D–V–C–E. We might suggest that this path could be one of the most optimized paths, enabling visitors to explore all the interesting places as quickly as possible. Alternatively, the group that visited just a few nodes (less than four), which appears to be of relatively the same size among both short-stay and long-stay visitors, might be interested in just a few of the iconic art works, or just not motivated or informed enough to explore the bigger space.

All of this suggests that some routes used to explore the museum appear frequently for both short-stay and long-stay visitors even though the latter might be expected to be much more diverse in their choices given the longer time available. This implies that visitors' sequential movement in The Louvre Museum is quite limited in terms of path sequence length and order, though there are a number of possible routes including repeating the same nodes. We speculate that these similarities/dissimilarities could cause uneven distribution of the number of visitors, resulting in congestion or sparsity in some museum spaces.

These findings present a significant advancement in describing patterns in visitors' activity and behavior in a museum, and might enable us to foresee visitor movement. This also indicates the possibility of dynamically managing visitor flow and museum congestion, taking into account time-related factors, and the possible advantages of the design of the spatial arrangement. In addition, the transition rate and the probability of movement between nodes makes it possible to foresee the specific quantity and flow of visitors at a certain time and space, helping the development of more flexible and dynamic policies for space control. For instance, the similarities/dissimilarities of both types of visitor, which were unknown prior to this study, might make the practitioner reconsider the target of some management techniques that should be applied carefully on the proper and segmented group types (Krebs et al, 2007; Maddison and Foster, 2003). Also, a dynamic visitor-control system might be developed, based on our findings, by using the audioguides to change suggested visitor routes dynamically depending on the congestion level as calculated by the data gathered from sensors installed throughout the museum.

Finally, these results might enable improvement in the quality of information that can be provided to visitors at an adequate place and time in order to maximize their fulfillment of the social and cultural experience, thereby optimizing the museum infrastructure.

**Acknowledgements.** We would like to thank the Studies and Research Department of The Louvre Museum for their support, Kyuha Shim, Namju Lee, and Miriam Roue for the graphics materials, and Alexander Amini and David Lee for text proofreading. We are indebted to many people for providing an extremely stimulating research discussion. In particular, we would like to thank Jaume Barceló, Catherine Guillou, Anne Krebs, Thomas Besancon, Marie-Anne Corniou, Michael Szell, Riccardo Campari, Sebastian Grauwin, Markus Schlapfer, Chaogui Kaug, Tao Pei, Kael Greco, Dietmar Offenhuber, Rodrigo de Oliveira, Kenichi Nakamura. We further thank the MIT SMART Program, the Center for Complex Engineering System (CCES) at KACST and MIT CCES program, Volkswagen Electronics Research Lab, BBVA, The Coca Cola Company, Ericsson, Expo 2015, Ferrovial, The Regional Municipality of Wood Buffalo, and all the members of the MIT Senseable City Lab Consortium. RS is supported by the James S. McDonnell Foundation. Of course, any shortcomings are our sole responsibility.

---

## References

- Asakura Y, Iryob T, 2007, "Analysis of tourist behaviour based on the tracking data collected using a mobile communication instrument" *Transportation Research Part A: Policy and Practice* **41** 684–690
- Barceló J, Montero L, Marqués L, Carmona C, 2010, "Travel time forecasting and dynamic origin–destination estimation for freeways based on Bluetooth traffic monitoring" *Transportation Research Record: Journal of the Transportation Research Board* number 2175, 19–27
- Bitgood S, 2006, "An analysis of visitor circulation: movement patterns and the general value principle" *Curator* **49** 463–475
- Choi Y K, 1999, "The morphology of exploration and encounter in museum layouts" *Environment and Planning B: Planning and Design* **26** 241–250
- Delafontaine M, Versichele M, Neutens T, Van de Weghe N, 2012, "Analysing spatiotemporal sequences in Bluetooth tracking data" *Applied Geography* **34** 659–668
- Eagle N, Pentland A, 2005, "Reality mining: sensing complex social systems" *Personal and Ubiquitous Computing* **10** 255–268
- Falk J H, Dierking L D, 1992 *The Museum Experience* (Walesback Books, Washington, DC)
- González M C, Hidalgo C A, Barabási A L, 2008, "Understanding individual human mobility patterns" *Nature* **453** 779–782
- Grauwin S, Sobolevsky S, Moritz S, Gódor I, Ratti C, 2014, "Towards a comparative science of cities: using mobile traffic records in New York, London and Hong Kong", arXiv preprint arXiv:1406.4400
- Hamnett C, Shoval N, 2003, "Museums as "flagships" of urban development" in *Cities and Visitors: Regulating People, Markets, and City Space* Eds L M Hoffman, D Judd, S S Fainstein (Blackwell, Oxford) pp 217–236
- Hawelka B, Sitko I, Beinat E, Sobolevsky S, Kazakopoulos P, Ratti C, 2014, "Geo-located Twitter as proxy for global mobility patterns" *Cartography and Geographic Information Science* **41** 260–271
- Hein G, 1998 *Learning in the Museum* (Routledge, London)
- Hillier B, 1996 *Space is the Machine: A Configurational Theory of Architecture* (Cambridge University Press, Cambridge)
- Hillier B, Hanson J, 1984 *The Social Logic of Space* (Cambridge University Press, Cambridge)
- Hillier B, Tzortzi K, 2006, "Space syntax: the language of museum space", in *A Companion to Museum Studies* Ed. S MacDonald (Blackwell, Oxford) pp 282–301
- Hood M G, 1983, "Staying away: Why people choose not to visit museums" *Museum News* **61**(4) 50–57
- Hoteit S, Secci S, Sobolevsky S, Ratti C, Pujolle G, 2014, "Estimating human trajectories and hotspots through mobile phone data" *Computer Networks* **64** 296–307
- Hui S K, Bradlow E T, Fader P S, 2009, "Testing behavioral hypotheses using an integrated model of grocery store shopping path and purchase behaviour" *Journal of Consumer Research* **36** 478–493
- Kanda T, Shiomi M, Perrin L, Nomura T, Ishiguro H, Hagita N, 2007, "Analysis of people trajectories with ubiquitous sensors in a science museum", in *Proceedings 2007 IEEE International Conference on Robotics and Automation (ICRA'07)* pp 4846–4853
- Kirchberg V, Tröndle M, 2012, "Experiencing exhibitions: a review of studies on visitor experiences in museums" *Curator: The Museum Journal* **55** 435–452
- Klein H, 1993, "Tracking visitor circulation in museum settings" *Environment and Behavior* **25** 782–800
- Kostakos V, O'Neill E, Penn A, Roussos G, Papadongonas D, 2010, "Brief encounters: sensing, modelling and visualizing urban mobility and copresence networks" *ACM Transactions on Computer–Human Interaction* **17**(1) 1–38

- 
- Krebs A, Petr C, Surlbled C, 2007, “La gestion de l’hyper fréquentation du patrimoine: d’une problématique grandissante à ses réponses indifférenciées et segmentées” [Managing the hyper-congestion of cultural heritage sites: from a growing problematic to its unspecialized and segmented responses], in Communications of the 9th International Conference on Arts and Culture Management University of Valencia, <http://www.adeit.uv.es/aimac2007/index.php>
- Kung K S, Greco K, Sobolevsky S, Ratti C, 2014, “Exploring universal patterns in human home/work commuting from mobile phone data” *PLoS ONE* **9**(6): e96180
- Laetsch W, Diamond J, Gottfried J L, Rosenfeld S, 1980, “Children and family groups in science centers” *Science and Children* **17**(6) 14–17
- McKercher B, Shoval N, Ng E, Birenboim A, 2012, “First and repeat visitor behaviour: GPS tracking and GIS analysis in Hong Kong” *Tourism Geographies* **14** 147–161
- Maddison D, Foster T, 2003, “Valuing congestion costs in the British Museum” *Oxford Economic Papers* **55** 173–190
- Mayer-Schönberger V, Cukier K, 2013 *Big Data: A Revolution That Will Transform How We Live, Work and Think* (John Murray, London)
- Melton A W, 1935 *Problems of Installation in Museums of Art. American Association of Museums Monograph New Series No. 14* (American Association of Museums, Washington, DC)
- Miller H J, 2005, “A measurement theory for time geography” *Geographical Analysis* **37** 17–45
- Parsons M, Loomis R, 1973 *Visitor Traffic Patterns: Then and Now* (Office of Museum Programs, Smithsonian Institution, Washington, DC)
- Paulos E, Goodman E, 2004, “The familiar stranger: anxiety, comfort, and play in public places” *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* **6** 223–230
- Pei T, Sobolevsky S, Ratti R, Shaw S L, Li T, Zhou C, 2014, “A new insight into land use classification based on aggregated mobile phone data” *International Journal of Geographical Information Science* (ahead-of-print), arXiv preprint arXiv: 1310.6129
- Pekarik A J, Doering Z D, Karns D A, 1999, “Exploring satisfying experiences in museums” *Curator: The Museum Journal* **42** 152–173
- Ratti C, Pulselli R, Williams S, Frenchman D, 2006, “Mobile landscapes: using location data from cell phones for urban analysis” *Environment and Planning B: Planning and Design* **33** 727–748
- Ratti C, Sobolevsky S, Calabrese F, Andris C, Reades J, Martino M, Claxton R, Strogatz S, 2010, “Redrawing the map of Great Britain from a network of human interactions” *PLoS ONE* **5**(12): e14248. doi:10.1371/journal.pone.0014248
- Robinson E S, 1928 *The Behavior of the Museum Visitor. American Association of Museums Monograph New Series No. 5* (American Association of Museums, Washington, DC)
- Sanfeliu A, Llàcer M R, Gramunt M D, Punsola A, Yoshimura Y, 2010, “Influence of the privacy issue in the deployment and design of networking robots in European urban areas” *Advanced Robotics* **24** 1873–1899
- Schuster J M, 1995, “The public interest in the art museum’s public”, in *Art in Museums* Ed. S Pearce (Athlone, London) pp 109–142
- Serrell B, 1998 *Paying Attention: Visitors and Museum Exhibitions* (American Association of Museums, Washington, DC)
- Shoval N, McKercher B, Birenboim A, Ng E, 2013, “The application of a sequence alignment method to the creation of typologies of tourist activity in time and space” *Environment and Planning B: Planning and Design* forthcoming doi:10.1068/b38065
- Sobolevsky S, Szell M, Campari R, Couronné T, Smoreda Z, Ratti R, 2013, “Delineating geographical regions with networks of human interactions in an extensive set of countries” *PLoS ONE* **8**(12) e81707
- Sobolevsky S, Sitko I, Grauwin S, Tachet des Combes R, Hawelka B, Arias J M, Ratti C, 2014, “Mining urban performance: scale-independent classification of cities based on individual economic transactions”, arXiv preprint arXiv: 1405.4301
- Sparacino F, 2002, “The Museum Wearable: real-time sensor-driven understanding of visitors’s interests for personalized visually-augmented museum experiences” in *Proceedings of Museums and the Web 2002* Eds D Bearman, J Trant (Archives and Museum Informatics, Tronto)

- 
- Stallings W, 2011 *Cryptography and Network Security: Principles and Practice* 5th edition (Prentice-Hall, Boston, MA)
- Tröndle M, Greenwood S, Kirchberg V, Tschacher W, 2014 “An integrative and comprehensive methodology for studying aesthetic experience in the field: merging movement tracking, physiology, and psychological data” *Environment and Behavior* **46** 102–135
- 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** 208–220
- Weiss R, Boutourline S, 1963, “The communication value of exhibits” *Museum News* November, pp 23–27
- Yoshimura Y, Girardin F, Carrascal J P, Ratti C, Blat J, 2012, “New tools for studying visitor behaviours in museums: A case study at the Louvre”, in *Information and Communication Technologies in Tourism 2012. Proceedings of the International Conference Helsingborg (ENTER 2012)* Eds M Fuchs, F Ricci, L Cantoni (Springer, New York) pp 391–402