



Identifying tourist-functional relations of urban places through Foursquare from Barcelona

Liya Yang · Carlos Marmolejo Durarte

Published online: 25 July 2019
© Springer Nature B.V. 2019

Abstract Previous research has mainly focused on spatial relations among tourist attractions in urban areas. However, few studies have examined the functional relations between tourist attractions and other urban places (i.e. the flows of tourists between them). Therefore, this study focuses on quantification of the tourist-functional relations among Places of Interest (POIs) using Foursquare data from Barcelona. This represents an effort to highlight the important functional closeness between different types of POIs whose significance is not usually obvious from their spatial relationships. In order to quantify these functional relationships, this paper classifies Foursquare POIs into 22 categories according to their different usages and constructs a matrix of usage-flows to depict the connections among these different usages. A model of interaction values is introduced to describe the strength of relations and identify the dominant tourist usages. The results confirm that the functional centroids differ from these spatial distributions which only focused around tourist attractions. In addition to tourist attractions, places in the categories of Restaurants, Transport, and Hotels play important roles in

functional relationship of tourism. The typical urban usages of tourists can be distinguished by the interaction values between these categories. Our model provides a practical method to quantify the interlinkage of usages of POIs based on tourist flows. In particular, LBSN data has potential as a method to observe the tourist-functional relations among places.

Keywords LBSN · Tourism · Functional relation · Foursquare · Urban usage · Big data · Tourist flows · Barcelona

Introduction

Tourism is treated as an important part of the development strategy in many cities. According to the Cities Report 2018 from the World Travel and Tourism Council, the Travel and Tourism sector directly contributed 3.2% of global GDP in 2017. Therefore, analysis of the dynamics of tourism is a popular issue for academia, business, and governments.

The emergence of Location-Based Social Networks (LBSNs) has provided abundant data for tourism studies, such as travelling reviews (Marine-Roig and Clavé 2015), footprints (Girardin et al. 2009), and photos of destinations (Nikjoo and Bakhshi 2019). When compared with the traditional field survey, these

L. Yang (✉) · C. M. Durarte
Department of Architectural Technology, Polytechnic University of Catalonia, Av. Diagonal, 649,
08028 Barcelona, Spain
e-mail: li.ya.yang@upc.edu

C. M. Durarte
e-mail: carlos.marmolejo@upc.edu

sources of data provide increased opportunities to explore large-scale human activities in detail. By utilizing LBSN data, many studies have investigated the spatial distribution of tourist attractions or the travel patterns of tourist activities (Hasnat and Hasan 2018; Vu et al. 2015; García-Palomares et al. 2015). However, the functional relations between tourist places (i.e. flows of tourists between these places) have not been investigated in depth. In fact, visiting tourist attraction (e.g. a landmark) is just one component of tourist activities, since tourist visits necessarily imply itineraries. Other places, such as restaurants, public transport and accommodations comprise the functional bases of tourism in a city. Such relations between tourist facilities are usually described based on practical experiences. Therefore, the quantification of the relation is beneficial for better understanding the dynamics of tourist activities.

This study aims to quantify the important tourist-functional relations between urban places, which are usually not noticeable on spatial proximity relations. Here, the functional relation between tourist places is defined as the flows of tourist users between Four-square POIs. Further, we use the functional interaction value to measure the functional proximity of different groups of POIs which are classified by different categories of usages. The case study proposed in this paper investigates functional relations between tourist places in Barcelona using 18 months of Foursquare (an LBSN) check-in activity. On Foursquare, users make “check-ins” at venues which are predefined or created by themselves. Because Foursquare data provides the attributes of POIs and their geographic location, it is a popular source of information for mining online-user behaviors (Preotiuc-Pietro and Cohn 2013), urban mobility (Hasan et al. 2013), improvement of categorization systems (Hu et al. 2015), among other applications.

The main contribution of the study lies in its practical method of quantifying the functional relations of places, rather than trying to depict the dynamics of tourist activities in a specific city. The study process mainly comprises four steps: identification of tourist users, classification of POIs based on usage, calculation of flows among different categories of POIs, and analysis of functional relations. A network graph is used to present the significant relations among different usages. Finally, we also

compare the spatial proximity of POIs with their functional proximity for finding out their differences.

The remainder of this paper is structured as follows. The literature review summarizes the main methods to identify tourist users and related studies of tourists’ activities using LBSN data. The part of methods introduces the methodology and implements the identification of tourists within the entire sample. The result analyzes the spatial characteristics of tourist movement and the functional relationships among tourist places in Barcelona. Finally, the conclusion discusses the limitations and future work.

Study of tourist activities based on LBSN data

The spatial distribution of tourist activities based on LBSN data has been investigated in-depth globally. It turns out that tourist movement have, on a macro level, some similar spatial characteristics in different cities. In general, they are more spatially concentrated than locals’ movements. Urban centers, airports, and tourist attractions, such as famous churches or museums, are typical clustering places for tourists. For instance, Girardin et al. (2009) identified the differences of spatial activities between tourists and locals in New York via cell phone data and Flickr data. They confirmed that the movement range of visitors is limited, especially for foreign visitors. Vu et al. (2015) combined GPS and Flickr data to cluster tourists’ main routes in Hong Kong and showed that tourists tend to travel to and between adjacent areas. Hasnat and Hasan (2018) described tourist movements as clustered around tourist attractions based on Twitter data in Florida. Béjar et al. (2014) extracted the spatial-temporal characteristics of Twitter and Instagram users in Barcelona; they found that the main tourist attractions were important connecting nodes in both datasets, though their study did not distinguish tourist users from locals. Kádár (2014) concluded that the majority of geo-tagged images generated from Flickr tourist users were gathered around tourist attractions or landmarks in Budapest.

However, such spatial-temporal characteristic fails to uncover how places interact with each other functionally, since it only shows the degree of spatial aggregation of human activities in separate places or the tendency of people’s movements. Although the quantification of flows between places have been

involved in studies of travel patterns and land-use detection, the focuses are still different from functional linkages. Travel patterns focus on the temporal patterns of human mobility (Sagl et al. 2012; Luo et al. 2016; Thuillier et al. 2017) Land-use detection aims to reveal the relation between spatial distribution of POIs and urban land-uses (Gao et al. 2017), rather than the connections between different urban usages.

To date, few studies have investigated the functional linkages among POIs; so, it still offers room for exploration. The most closely related study to the present work is from Ferreira et al. (2015). They collected Foursquare data in London, New York, Rio de Janeiro, and Tokyo. They classified POIs into nine categories and compared the check-in patterns of tourists and residents in these categories. They found that some categories show a significant temporal difference of check-in pattern between tourists and residents, such as locations classified as Arts and as Transport. However, they focused on the mobility patterns between specific places, rather than the functional closeness between different categories. Preotiuc-Pietro and Cohn (2013) discussed the probability of transition of users between different types of Foursquare POIs, but did not explain the functional relationships among POIs. Therefore, our study actually proposes a new perspective to analyze the functional relations between places.

With regard to the identification of tourists, LBSN data offers the potential to identify vast numbers of tourists at low cost. According to previous studies, the identification mainly relies on the geo-location (Vu et al. 2015; Da Rugna et al. 2012; Luo et al. 2016) or the time threshold (Kádár 2014; Girardin et al. 2008; García-Palomares et al. 2015). For example, Luo et al. (2016) distinguished residential Twitter users from visitors in Chicago by determining their locations during the night. Users are identified as locals if most of their check-ins during nights are in residential areas. However, this method is effective only if hospitality services are segregated from residential areas. It is also difficult to apply the approach in a compact city which has large amounts of mixed-use land, such as Barcelona. Manca et al. (2017) combined the time threshold and the geo-location of Twitter users to distinguish tourist users from locals. Users who posted tweets less than 20 days in Barcelona as considered as tourists. In general, a longer time threshold could be more reliable; however, it requires a dataset with more

long-term data. Moreover, the determination of the time threshold is usually derived from empirical experiences or the advice of tourism experts, and so lacks objectivity. Above all, this study sought to classify users as tourists or local residents by examining users' behaviors on Foursquare and applying a threshold based on statistical analysis of the dataset. A semi-supervised model, described in the next section, was adopted to identify tourist users. This method allowed further exploitation of the dataset, because both active and inactive users are included.

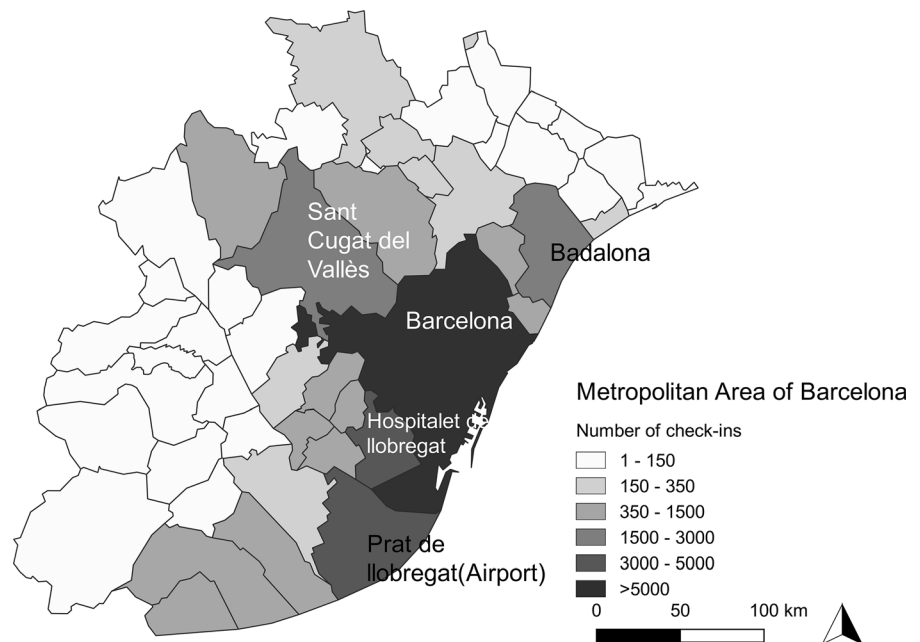
Methods

Study scope and description of datasets

Barcelona is an ideal city for an analysis of tourism; it has been known as a tourist destination since the early 1900s. The government established a Commission for the Attraction of Foreigners and Tourists in 1906 and aimed to build up the city as a tourist destination known as the “Pearl of Mediterranean”. Its importance as a tourist destination only increased after the Summer Olympic Games of 1992. According to the Annual Tourism Sector of Barcelona Report 2014, the total number of overnight tourists who stayed in hotel accommodation reached more than 7.5 million, ranking as the 20th most-visited city in the world. In 2017, this number rose to nearly 9 million. Currently, according to the local government of Barcelona, tourism accounts for about 12% of the city's GDP and generates approximately 9% of employment in Barcelona.

This paper extracted data from a global Foursquare check-in dataset (Yang et al. 2016). The studied area includes the inner area of the Barcelona Metropolitan Region (RMB), due to the monitoring range of Foursquare data (Fig. 1) which is slightly larger than the first zone of the RMB. The first zone is also called the Metropolitan Area of Barcelona (AMB), which comprises 36 municipalities. According to the official statistics, the population was 3,239,337 in 2014, about half of whom lived in the city of Barcelona. The land uses of Barcelona city is highly mixed. Most check-ins, 57,764 items, occurred in Barcelona city (Fig. 1). Other than Barcelona, only four cities in the region have more than 1000 check-ins: L'Hospitalet de

Fig. 1 Distribution of check-ins in Barcelona Metropolitan Region.
Source: Authors



Llobregat, El Prat de Llobregat, Badalona, Sant Cugat del Vallés and Conellà de Llobregat.

The dataset includes data collected between 2012-04-03 and 2013-09-16. A total of 80,936 check-ins were made by 4527 users in the area of Barcelona. To eliminate noise, users who only checked-in one time in the area were excluded from the dataset. Therefore, 3350 “valid” users with a combined total of 79,798 valid check-ins were included (Table 1).

The method for measuring the functional relations of POIs included eight steps: (1) identifying tourist users as a group distinct from local users and classifying all Foursquare POIs into 22 different categories for purpose of analysis; (2) constructing the chronological path of activity of each tourist user based on their check-ins; (3) calculating the number of connections between two categories of POIs using the paths; (4) building a heat map of the matrix of flows; (5) introducing a model of interaction values to

investigate the functional relation of these categories; (6) representing the matrix of interaction values in two dimensions through PROXCAL multidimensional scaling (MDS); (7) visualizing the prominent functional relationships via a network graph; (8) comparing the functional proximity with the spatial proximity based on different categories of POIs.

Identification of tourists

This study uses a semi-supervised method to distinguish tourist users from locals using Foursquare data. This method consists of K-means clustering with manual improvement. Departing from previous research, it is based on the assumptions that the number of tourists’ check-ins, total travel distance, and duration of stay are lower (on average) than those of residents. The method of calculation of duration is the following:

$$\text{Total Duration} = T_{last} - T_{first} \quad (1)$$

Table 1 Summary of valid users. Source: Authors

Total users		Mean	SD	Min	Max
3350	Check-ins (times)	23.8203	54.14576	2	1182
3350	Duration (days)	112.5728	179.0711	0	531

where T_{last} is the latest timestamp of a check-in of a user and T_{first} is the earliest timestamp.

The travel distance is calculated via ArcGIS software:

Total Travel Distance

$$= \sum_i^n \sqrt{(X_{P_{t+1}} - X_{P_t})^2 + (Y_{P_{t+1}} - Y_{P_t})^2} \quad (2)$$

where P_t is the position of a given check-in at timestamp i , with coordinates (X_{P_t}, Y_{P_t}) of the Universal Transverse Mercator (UTM) system. The distance between P_{t+1} and P_t is calculated by the straight-line distance between two points.

Before clustering, several users were chosen to form a sample to test the validity of the K-means clustering by checking whether they are classified correctly. One user was initially randomly selected from each group of different frequency of check-ins (Fig. 2a), combined with the different duration of stay (Fig. 2b). In other words, each selected user was from different groups of check-ins and duration.

Therefore, except null values of inquiries, 23 users were selected out and their identities were manually verified. According to the numbers of check-ins, the places where they checked-in and their duration of stay, we identified whether the selected user is a tourist or a local. For example, if a user checked-in 10 times and the duration of stay was 2 days, and all places where the user checked-in are places of amusement or tourist attractions; we classified this user as a tourist.

Next, we utilized K-means clustering to divide all users into the two groups. For discrete data, algorithms of grouping data are classification and clustering. Classification requires a training dataset which contains samples whose category is known. As the characteristics of tourist behavior were unknown in our case, clustering was the better approach to divide users. K-means clustering is widely applied due to its simplicity. Moreover, K-means algorithms perform very well with huge datasets (Abbas 2008). The Z-score was used to standardize the three indicators for clustering:

$$z = \frac{x - \mu}{\sigma} \quad (3)$$

where z is the standardized score of each of the indicators, x is the value of indicator, μ is the mean of x , and σ is the standard deviation.

Manual examination showed that two local users of the sample group were included in the tourist group because they had comparatively lower duration of stay. On the other hand, four tourist users of the sample group were included in the local group because their duration of stay were too long. Such difference is partly caused by our method of calculation of duration. It is possible that the duration of stay was incorrect for some visitors returning in the second year. Therefore, it was necessary to use a threshold of check-ins and duration to improve the classification. Four different combinations of threshold were tested (Table 2). Those users whose indicators were above the threshold were categorized as locals.

The results of classification were more stable when time span was more than 84 days. Therefore, the strictest threshold, 4, was adopted as the final standard for improvement. Users who stayed more than 113 days and made check-ins over 24 times are classified as locals. In total, 2770 tourist users were identified. These users generated 19,180 check-ins during the monitoring period (Table 3).

Classification of Foursquare POIs

With regard to the features of Foursquare POIs, it seems that the distribution of check-ins is mainly clustered into a few categories: Travel and Transport, Food, and Shopping. Abbasi and Alesheikh (2018) mentioned that shopping and eating places contributed 59% of check-ins in Manhattan. Li et al. (2013) investigated the popularity of Foursquare POIs in 14 urban regions all over the world. Their results show that the Travel and Transport category occupies the highest frequency of check-ins. Preotiuc-Pietro and Cohn (2013) collected Foursquare data of frequent users worldwide for 1 month. They used the basic category from Foursquare website and learned that Shopping and Services, Food, and Travel and Transport accounted for 53.8% all check-ins.

There are 13,887 unique Foursquare POIs in Barcelona, which are labeled by 385 sub-categories. Restaurants form a large portion of all types of POIs. According to the official website, the venue categories of Foursquare classifies them by nine major categories: Art and entertainment, Faculty and University, Event, Food, Night spots, Outdoor Recreation, Professional and others, Store and services, Travel and transportation. However, these categories need to be

Fig. 2 Foursquare users' check-ins and stay duration.
Source: Authors

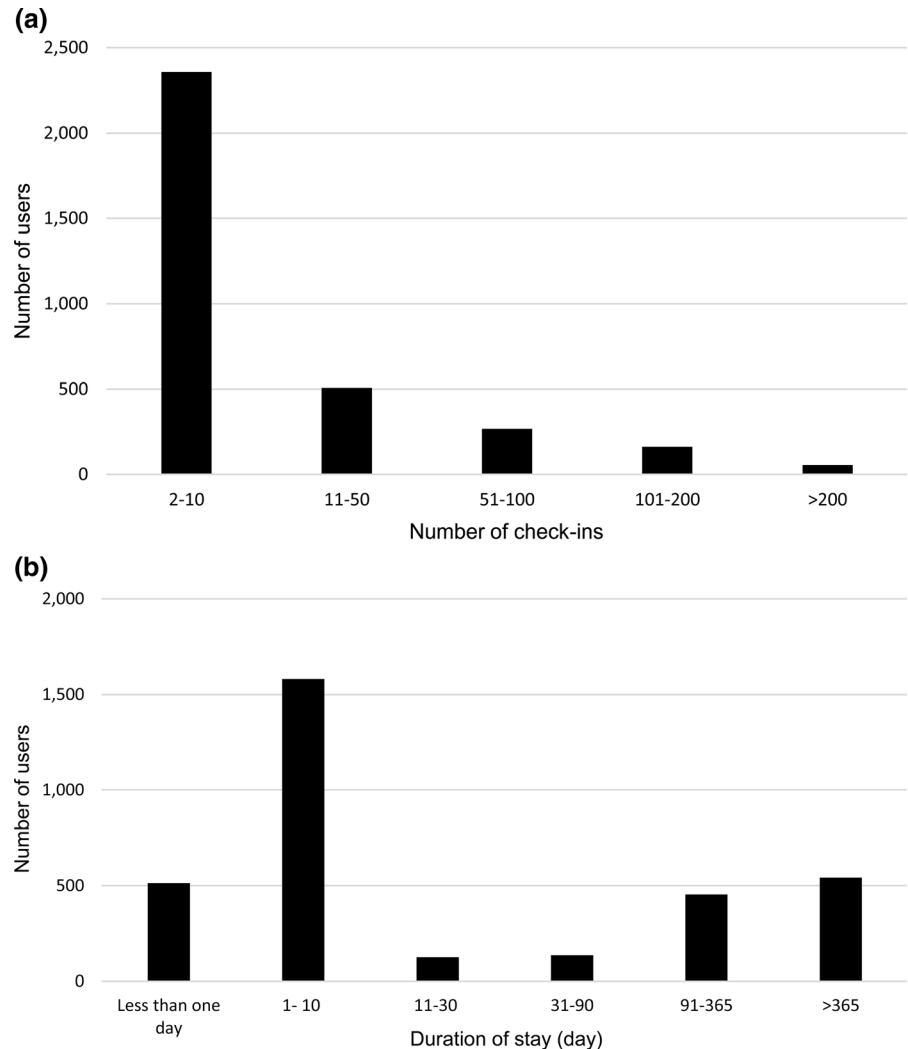


Table 2 Four thresholds of improvement. Source: Authors

Threshold	Total duration (days)	Total check-ins (times)	Description
1	84	18	Mean values of duration and check-ins based on all 4527 users
2	90	24	Empirical threshold
3	98	21	Mean values of duration and check-ins based on threshold 1 and 4
4	113	24	Mean values of duration and check-ins based on 3350 valid users

improved for conducting the tourist-functional analysis. For example, Hotel is under the category of Travel and transportation, but Hotel actually belongs to the category of accommodation. Moreover, some names of the category are too vague. For example, “Event”

actually contains temporary and permanent markets. It is clearer to use “market” rather than “event”.

Therefore, this paper assembles these categories into 22 main types considering their usages, for example, all kinds of restaurants are grouped as “Restaurant”. Table 4 lists the new classification with

Table 3 The summary of locals and tourists. *Source:* Authors

	Number	Check-ins	Average number of check-ins
All users	3350	79,798	23.82
Residents	580	60,618	104.51
Tourists	2770	19,180	6.92

Table 4 New category of POIs. *Source:* Authors

Types of POIs	N. of POIs	N. of check-ins	Avg. check-in per POI	Description
Restaurant	3318	2726	0.82	Mediterranean Restaurant, Japanese Restaurant, Food, Diner, etc.
Public services	1321	746	0.56	Medical, Finance, Post Office, Bakery, Salon, Barbershop, Spa, Tattoo, etc.
Workplace	1174	585	0.50	Building, Campaign Office, Co-working Space, Design Studio, etc.
Bar	1138	1031	0.91	Bar, Beer Garden, Cocktail Bar, Jazz Club, Nightclub, etc.
Outdoor resorts	1061	2017	1.90	Tourist attractions, Rest Area, Park, Scenic Lookout, etc.
Shop	940	789	0.84	Bike Shop, Dessert Shop, Frozen Yogurt, Gift Shop, etc.
Store	839	1398	1.67	Kids Store, Pet Store, Paper/Office Supplies Store, Video Store, etc.
Transport	668	2792	4.18	Train Station, Subway, Airport, Boat, Airport Terminal, Light Rail, etc.
Education places	634	428	0.68	University, College, Elementary School, Student Center, etc.
Café	540	330	0.61	Café, Tea Room, Cafeteria, etc.
Residential place	506	563	1.11	Neighborhood, Residential Building (Apartment/Condo), etc.
Hotel	478	1576	3.30	Motel, Hotel, etc.
Sports center	308	581	1.89	Athletic and Sport, Baseball Field, Basketball Court, Football Stadium, Golf Course, etc.
Museum, Art, Historical Place (MAH)	206	1156	5.61	Public Art, Performing Arts Venue, Museum, Historic Site, Castle, etc.
Opera, Concert, Cinema (OCC)	192	186	0.97	Indie Movie Theater, Concert Hall, Movie Theater, Opera House, etc.
Gym	148	154	1.04	Gym Pool, Gym, Yoga Studio, etc.
Infrastructure	131	443	3.38	Bridge, Harbor/Marina, River, etc.
Conference center	88	412	4.68	Conference Room, Meeting Room, Convention Center, etc.
Market	74	177	2.39	Fair, Farmers Market, Flea Market, Fish Market, etc.
Plaza	57	1062	18.63	Stables, Track, Planetarium, etc.
Others	51	24	0.47	Plaza
Tourist info center	5	4	0.80	Tourist Information Center

brief descriptions. The Transport, Restaurant, Hotel and Outdoor Resorts make up a combined 47.5% of check-ins. It is worth noting that we extracted Plaza as a separate category, because it is a compound urban place which mixes multiple functions, such as leisure,

transport hub, food, shopping services, etc. The volume of check-ins in the Plaza category also indicates that it is an important functional hinge for tourists in Barcelona.

Construction of paths of Foursquare users

In this study, a path is defined as the time-sequential check-ins of a user; each user has a unique path. Based on the study of Scholtes (2017), the construction of the paths is built on two assumptions:

- (1) Each user's path has a chronological order, hence, it is directed.
- (2) Paths are not transitive. Only direct connections count. For example, assuming there is a path: $a \rightarrow b \rightarrow c$, $[a,b]$ and $[b,c]$ are valid connections and $[a,c]$ is not counted.

This method avoids the duplicate calculation of connections among nodes. We constructed the functional paths of a user in terms of the category of usages.

The matrix of usage-flows

Based on the paths, we calculated the number of direct connections between each pair of POIs, including the connections within the same category. The "inflow" from the category i to j is the total number of connections from i to j . The reverse is the "outflow". The "flow" of paired usage is the sum of the "inflow" and the corresponding "outflow":

$$f_{ij} = C_{ij} + C_{ji} \quad (4)$$

where f_{ij} is the flow between i and j , and C_{ij} is the number of connections from i to j .

Visualization of spatial paths

To visualize the spatial distribution of tourist paths, all paths taken by tourist between two POIs were counted. There are several methods to aggregate paths, such as edge bundling (Graser et al. 2019) or aggregating paths through characteristic points (Andrienko and Andrienko 2011). However, the former method requires very high-capacity computation, because it needs to compare the similarity of each path and then implement aggregation. The latter method extracts the characteristic points of each path first, and then calculates the centroids of these points in term of the distance between points to reduce the cost of computation. The aggregated flows are then generated from these centroids. The shortfall of this method is that the

centroids are not the actual places, and thus it is hard to reflect the precise spatial relationships between places on small scales. Therefore, this study used the original traces to display the spatial relationships between POIs and tourist movement. To delimitate the major flows of tourists, only traces that repeated more than once are visualized.

The interaction values analysis

This study introduces the improved model of interaction value from Roca Cladera and Moix (2005) to depict the functional interactions among different usages. The first model of interaction value was created by Smart (1974), and was developed from the gravity model. The advantage of that model is that it explores the functional relation between two areas or objects without the inference of physical distance. It also eliminates distortion caused by differences in the "masses" of objects (i.e. the number of tourists visiting each of the POIs in our case), because the function divides the product of total flows of "sender" and "receiver" POI. Thus, this model can uncover the interaction relation between two objects effectively. It has been invoked in different studies of interactions, such as commuting flows between two areas (Roca Cladera et al. 2009), immigration flows (Dou et al. 2018) and air passenger flows (Burns et al. 2008). It takes the form

$$IV_{ij} = \frac{f_{ij}^2}{O_i \cdot I_j} + \frac{f_{ji}^2}{I_i \cdot O_j} \quad (5)$$

where IV_{ij} is the interaction value between the category i and j ; f_{ij} is the existing flow from category i to j ; O_i is the sum of outflows of category i , I_i is the sum of inflows of category i . Moreover, a statistic threshold was set to delimit the prominent relations of interaction:

Prominent interaction value

$$= \text{Mean value} + 1 \text{ standard derivation of IVs}$$

(6)

where mean value and standard derivation are the values of the whole matrix of interaction value.

To visualize the functional proximity between different usages, we used a PROXCAL multidimensional scaling (MDS) method to reduce the original matrix of interaction values to only two dimensions.

MDS is a fast way to visualize the level of similarity of objects. In our case, the similarity is understood as the strength of interaction values between usages. The closer is one usage to other, the stronger their functional relationship. The closer is one usage to the center of the graph, the stronger its relationship with all other usages.

Comparison between spatial proximity and the functional proximity

To further analyze the difference between functional proximity and the spatial proximity of POIs, we also plotted the average geo-distance and the “functional distance” of each types of POIs. The functional distance is represented as the distance to the zero point of the PROXICAL plot:

$$Functional_dist_i = \sqrt{PX_i^2 + PY_i^2} \quad (7)$$

where (PX_i, PY_i) are the coordinates of i usage on the PROXICAL two-dimension plot. The spatial distance is calculated using the Euclidean distance from the weighted median center of all POIs containing tourists' check-ins. Considering Foursquare check-ins are highly concentrated in Barcelona city (Fig. 1), it is reasonable to use the weighted median center as the central point to measure the spatial distribution of different types of POIs approximately. The advantage of median center is that it could indicate the spatially central tendency meanwhile is robust to outliers. Based on the algorithm of Arcgis, the weighted median center (Burt and Barber 1996; Kulin and Kuenne 1962) is given as:

$$D_p^t = \sqrt{(X_p - X^t)^2 + (Y_p - Y^t)^2 + (Z_p - Z^t)^2} \quad (8)$$

where (X_p, Y_p) is the geo-coordinates of a POI, Z_p is the weight of a POI which is the number of tourist check-ins in our case, (X^t, Y^t) is a candidate median center at the t step of the iterative process, D_p^t is the distance between the candidate center and other POIs at step t . The final median center minimizes the Euclidean distance to all other points in the dataset. Therefore, the average geographic distance of each type of POI is the corresponding number of POIs divide by the total distance of the corresponding POIs:

$$Spatial_dist_i = \frac{\sum_i^n \sqrt{(X_{P_i} - X_M)^2 + (Y_{P_i} - Y_M)^2}}{N_i} \quad (9)$$

where (X_{P_i}, Y_{P_i}) is the geo-coordinates of a POI of the i type of usage, (X_M, Y_M) is the median center, N_i is the total number of POIs of the i type.

Results

The spatial characteristics of tourist trajectories

Figure 3 depicts the spatial distribution of tourists' check-ins in Barcelona. The place with the densest population of tourists is the airport, which garnered about 8% of total check-ins. 77% of check-ins were concentrated in the municipality of Barcelona, while nearby municipalities split the remainder of the check-ins.

In regard to the specific locations of active tourist activities, 64% of POIs do not have records from tourists, suggesting that the range of tourists' activity is limited. Only 18 POIs had more than 100 check-ins (Fig. 4). Except for the airport, all are located near the city center. Figure 4 also overlays the aggregated flows of tourist trajectories and restaurant-POIs that contains more than two check-ins: purple lines are tourist trajectories, and the big white dots the most popular POIs, small white dots are restaurant-POIs.

Tourist attractions are clearly important nodes driving the spatial flows of visitors. La Sagrada Família, Casa Milà, Casa Batlló and Park Güell belong to famous historical heritage of Barcelona, while Camp Nou is a world-famous soccer center. These POIs were also on the official list of the top 15 most-visited places in Barcelona during 2012–2015 (Table 5). Hence, the hot spots of Foursquare POIs matched with the official data, meaning that tourist attractions organize the spatial movement of tourists.

Plaça de Catalunya takes the second rank of check-ins after La Sagrada Família. It is one of most important transport hubs and public spaces in Barcelona city, and it connects to the historical center of Barcelona. The Apple store, the main building of El Corte Inglés (the main department store in Spain), and the Hard Rock Café are located around Plaça de Catalunya. It is probable that each of these locations are meeting points for tourists since they have an outstanding position and/or buildings. Hence, the

Fig. 3 Distribution of tourist check-ins. *Source:* Authors

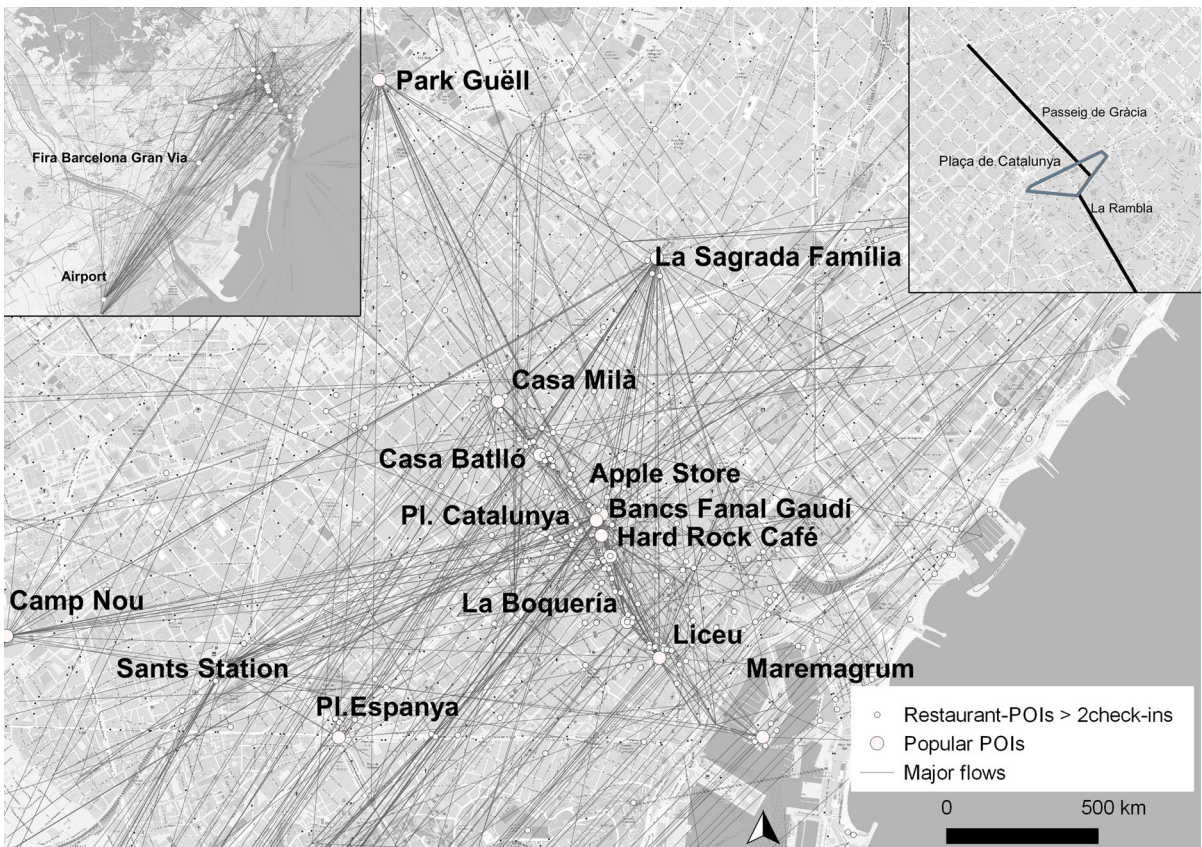
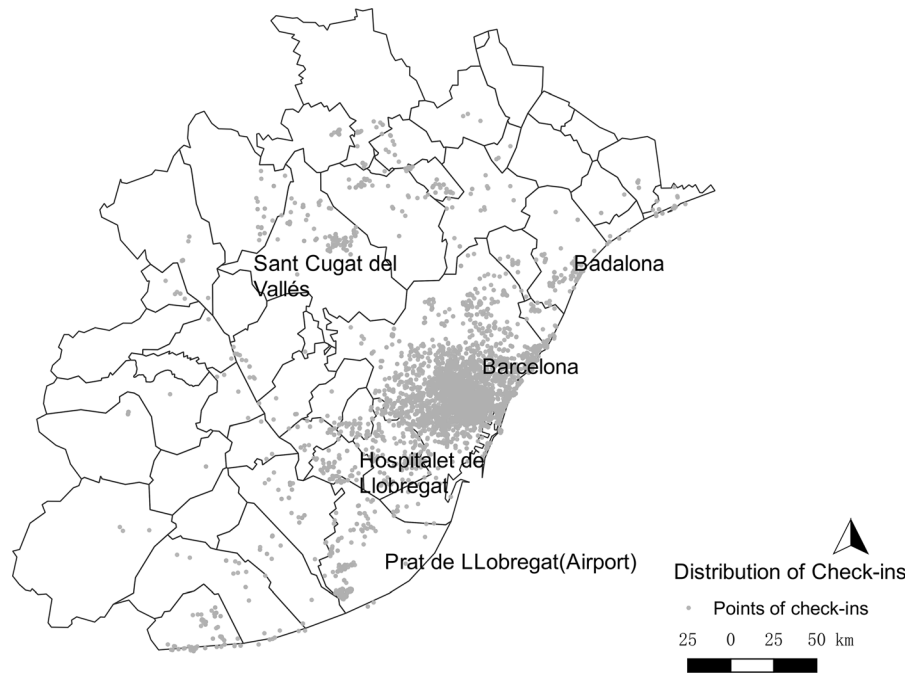


Fig. 4 Major tourist spatial flows and popular POIs. *Source:* Authors

Table 5 Number of visitors to major attractions in Barcelona. *Source:* <http://www.bcn.cat/estadistica/castella/dades/anuaris/anuari14/cap13/C1304010.htm>

Place	2012	2013	2014	2015
La Sagrada Família	3,233,526	3,176,970	3,260,880	3,722,540
Park Güell	–	–	2,598,732	2,761,436
Museum FC Barcelona (Camp Nou)	1,540,648	1,506,022	1,530,484	1,785,903
Aquarium of Barcelona	1,647,163	1,718,380	1,590,420	1,549,480
El Born Centre cultural	–	675,726	1,894,400	1,486,228
Poble Espanyol de Montjuïc	1,223,875	1,258,645	1,236,664	1,221,647
Museum of Picasso	948,869	915,226	919,814	1,008,125
Zoo Park of Barcelona	1,080,187	1,070,104	1,057,188	1,004,069
Casa Batlló	–	796,301	930,000	992,126
La Pedrera (Casa Milà)	861,583	944,509	932,356	990,112
History Museum of Barcelona	–	556,730	973,034	916,517
CaixaFòrum Museum of Barcelona	971,101	686,151	775,068	775,020
CosmoCaixa Barcelona	788,176	716,877	739,649	733,778
Castell de Montjuïc	1,159,042	1,072,000	577,639	670,526

Plaça de Catalunya can be considered another activity center for tourists. Two other commercial places received high numbers of check-ins. Maremagnum is a shopping mall near the Aquarium of Barcelona that provides multiple services for tourists. Fira Gran Via Barcelona is an international conference center which hosts many important conferences every year, such as the World Mobile Conference.

The Matrix of usage-flows

The following figure is the heat map of the matrix of usage-flows. It is a symmetric matrix which displays the bilateral flows of usages. Each cell represents the number of direct connections between two categories or within the same category. It is obvious that Restaurant–Transport, Restaurant–Hotel, and Restaurant–Outdoor Resorts have the highest flows between each other, with over 400 connections. Transport has its highest flows with itself.

This result fits with the spatial distribution of check-ins, as well as it uncovers some relationships among different categories. First, those higher flows of usages, such as Transport, Tourist attractions and Plaza, were also present in popular venues. Meanwhile, as Fig. 4 shows, those POIs of restaurants mainly concentrate along with Passeig de Gràcia street and La Rambla street which are pedestrian streets and gather many tourist attractions. Secondly, it shows the typical usage of tourist activities. Restaurant, Transport, Hotel and Outdoor Resorts have more intense

connections with each other than with non-tourist locations such as Gyms, Opera, Concert, Cinema (OCC) and Workplace (Fig. 5).

Interaction values

Similarly, the bilateral interaction values are visualized via a heat map (Fig. 6). The highest interaction value is between Hotel and Transport, followed by Outdoor Resorts-MAH, and Restaurant–Hotel. This result indicates that tourists tend to move directly between Transport–Hotel more than other categories. The number of movements between Museum, Arts and Historical place (MAH) and Outdoor resorts is also higher than between other categories. Compared with the matrix of flows, the matrix of interaction values clearly shows the functional proximity of connections between two categories of usages, without the distortion of the scale of flows. For example, Outdoor Resorts-MAH does not have a high volume of flows, but their interaction value is at the second-highest rank of interaction values (Table 6).

The PROXCAL-MDS reduce the original number of dimensions (i.e. 22²) to only two dimensions (Fig. 7). It is clear that Hotel, Transport and Restaurant have the closest functional interaction with each other, while Outdoor Resorts, MAH and Plaza are closer to each other. Both of these groups are located at the central part of the graph, indicating that they dominate the tourist flows. The usages located at the peripheral positions, such as Educational places, Gym,

Fig. 5 Heat map of the usage-flow matrix. *Source:* Authors

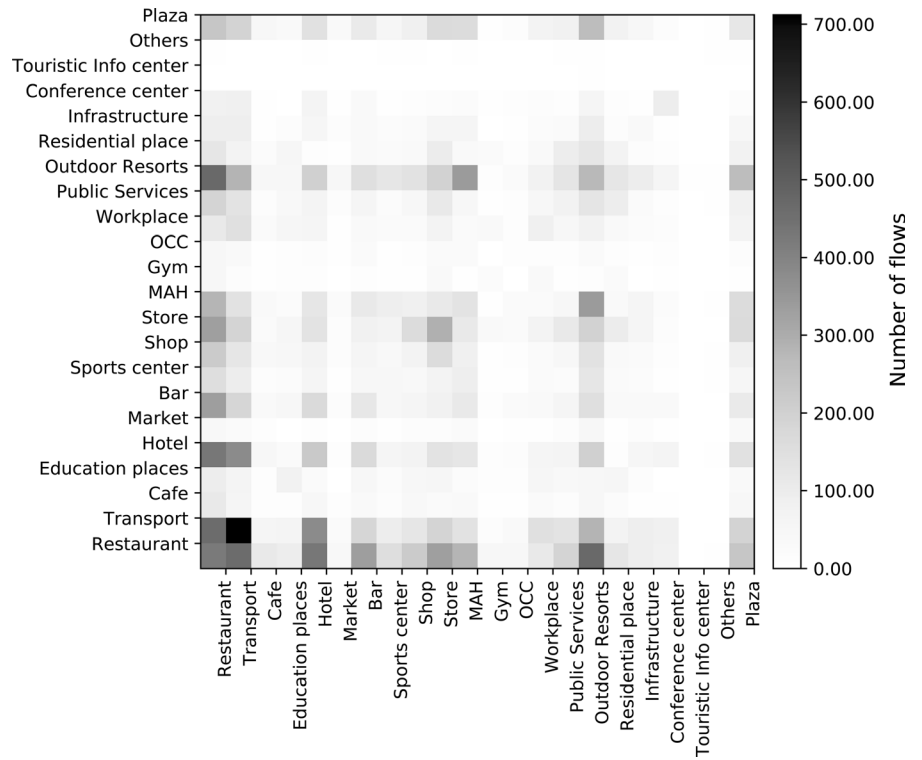


Fig. 6 Heat map of the interaction values matrix. *Source:* Authors

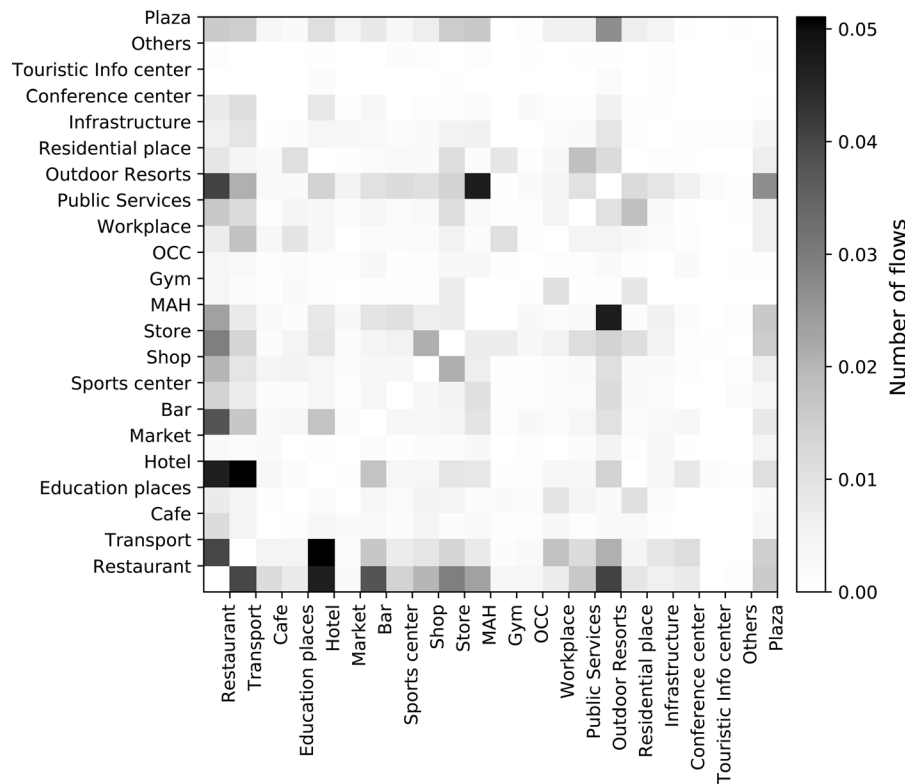


Table 6 Prominent interaction values. *Source:* Authors

Usage	Usage	Interaction value
Transport	Hotel	0.051061
MAH	Outdoor resorts	0.046973
Restaurant	Hotel	0.046562
Restaurant	Outdoor resorts	0.04087
Restaurant	Transport	0.039965
Restaurant	Bar	0.038003
Restaurant	Store	0.029328
Outdoor resorts	Plaza	0.026766
Restaurant	MAH	0.023513
Shop	Store	0.021034
Transport	Outdoor resorts	0.020784
Restaurant	Shop	0.019955
Public services	Residential place	0.017956
Transport	Workplace	0.01775
Hotel	Bar	0.017413
Transport	Bar	0.016533
Restaurant	Services	0.016266
MAH	Plaza	0.016008
Restaurant	Plaza	0.015661
Store	Plaza	0.015453
Transport	Plaza	0.014737
Restaurant	Sports center	0.01381
Hotel	Outdoor resorts	0.013807
Store	Outdoor resorts	0.0138
Transport	Store	0.013465

and Workplace, have minimal relations with other usages. This may imply that inside our tourist sample there are different kind of temporal visitors to the city: the first clearly attracted by heritage, cultural and leisure venues; the second, more linked to places intensively used by local population.

Next, we apply the chosen threshold (Formula 6) to show usages with prominent interaction values (Table 6). A total of 25 pairs of usages were qualified. Residential place and Workplace are also on the list; however, they are only connecting with Services and Transport separately. It is possible that Residential appears on the list due to trips to visit friends and relatives, or due to an error in the classification of tourists. According to the official survey, the motivation of travel for business accounted for 40% of all

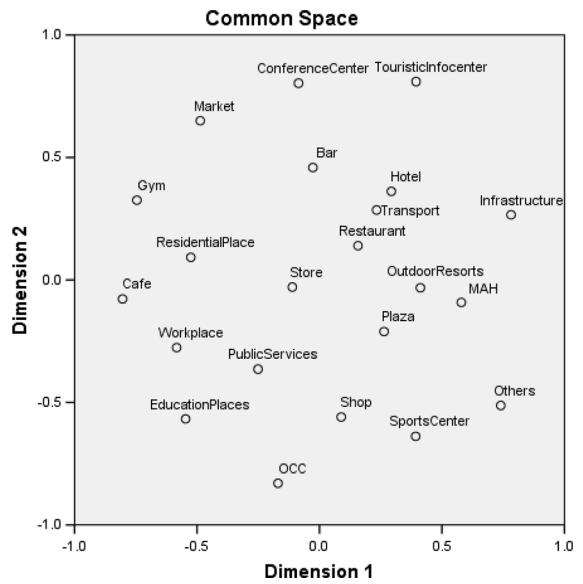


Fig. 7 Proximal plot of interaction value matrix. *Source:* Authors

visitors in 2014. Thus, it is reasonable that Transport-Workplace is shown on the list.

To depict the main relations in a simple way, Fig. 8 exhibits the prominent interaction values. The nodes are the category of usages and edges are the interaction values. The graph is visualized by Networkx program using Fruchterman-Reingold force-directed algorithm. Because the algorithm of visualization is aimed to reduce the crossing edges as few as possible, the position of nodes and the distance between them do not have specific meanings. However, the nodes with less edges tend to be placed at the periphery. The value is represented by different colors. The black lines represent values equal or above 0.0216 which is the mean value of all prominent interaction values, and the grey dashed line those below 0.0216. This graph embodies the basic tourist activities: eating, travelling, visiting, shopping and getting accommodation. Those categories that do not correspond to typical tourist places are at periphery. “Restaurant” has the largest number of connections with other categories, i.e. it is the central vertex of this network.

Comparison between spatial proximity and functional proximity

As examples, Fig. 9 depicts the location of the median center and the distribution of POIs of Restaurant,

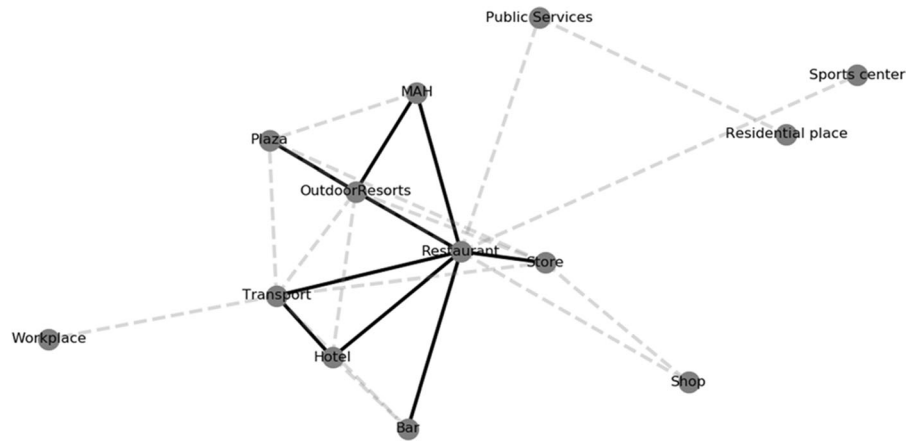


Fig. 8 Paired usages with prominent interaction values. *Source:* Authors

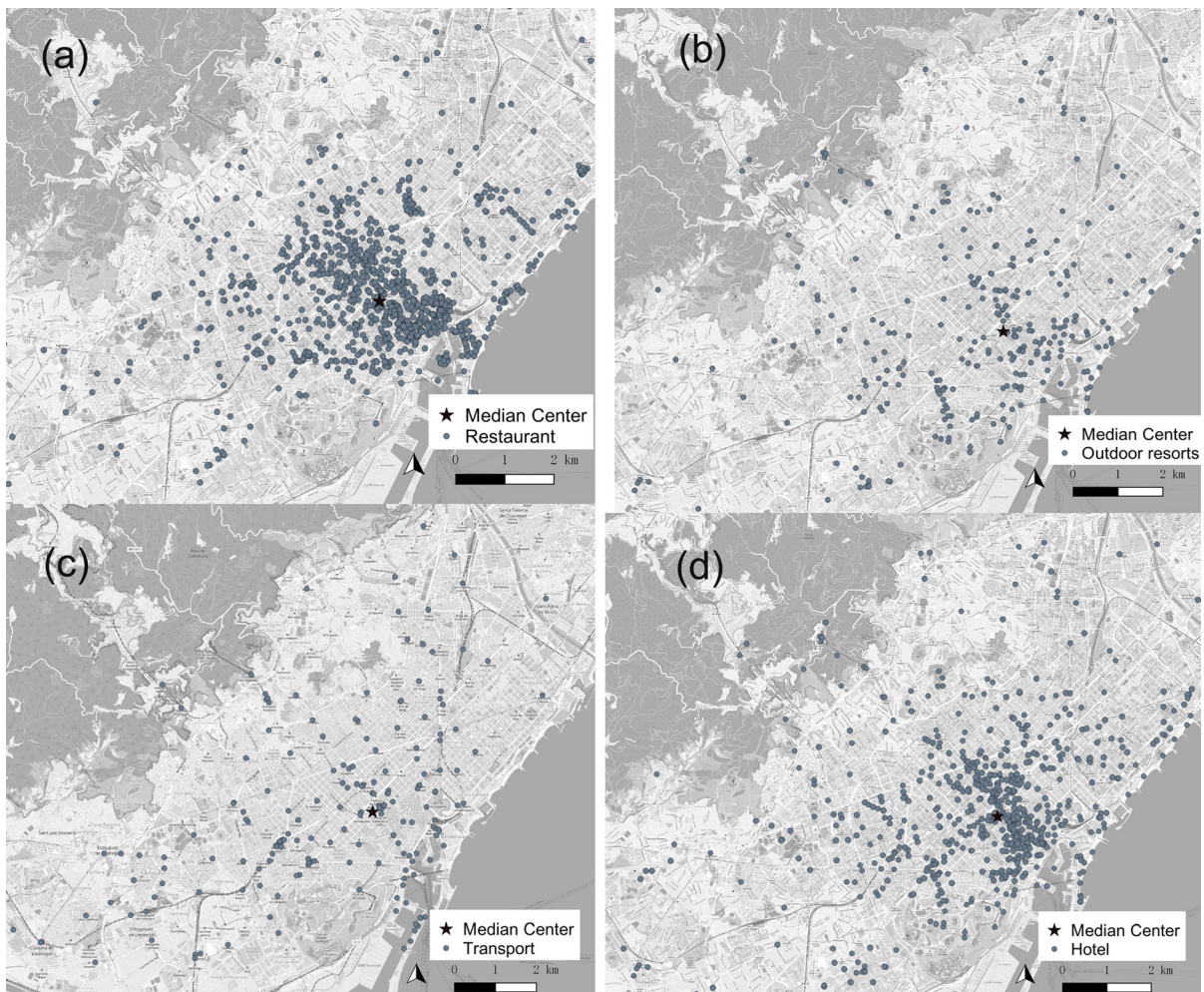


Fig. 9 The spatial location of median center and four types of POIs. *Source:* Authors

Transport (except the airport), Outdoor resorts and Hotel separately. Transport-POIs are the sparsest and far away from the center. It is evident that restaurants and hotels gather in the city center and are close to the median center. The distribution of POIs of Outdoor resorts is comparatively farther than them.

Figure 10 plots the functional distance (Y-axis) based on PROXCAL graph and the spatial distance of each type of POIs to the median center. For easier reading, the average spatial distance is rescaled by min–max normalization. The plot clearly shows that the Transport-POIs has the largest average spatial distance to the median center, but it has highly importance in tourist-functional relations. Hotel–Transport has the highest interaction value of all paired categories. Tourist information center is nearest to the median center, but its functional distance is distant due to few check-ins was generated in here. One possible explanation is that tourists would not like to check-in at places which only provide practical information.

Secondly, some important tourist attractions, such as La Sagrada Família, Park Güell and Camp Nou, are at “peripheral” area of tourist activities. However, as Fig. 4 shows, these tourist attractions play a pivot role among spatial flows of tourists, thus, their functional importance are highlighted. Although it seems that Public services is at the similar position with Outdoor Resorts, the functional distance of the two categories is at the opposite direction (Fig. 7). Similarly, Residential places is closer to the PROXCAL center, but is far away from POIs of Outdoor Resorts, Transport etc.

Discussion and conclusions

This research analyzes the tourist-functional relations of POIs in Barcelona and adopts a new method to identify tourist and non-tourist users of an LBSN. Essentially, in accordance to previous researches, the delimitation of tourists relies on the definition of tourists, i.e. the duration of stay is the key variable.

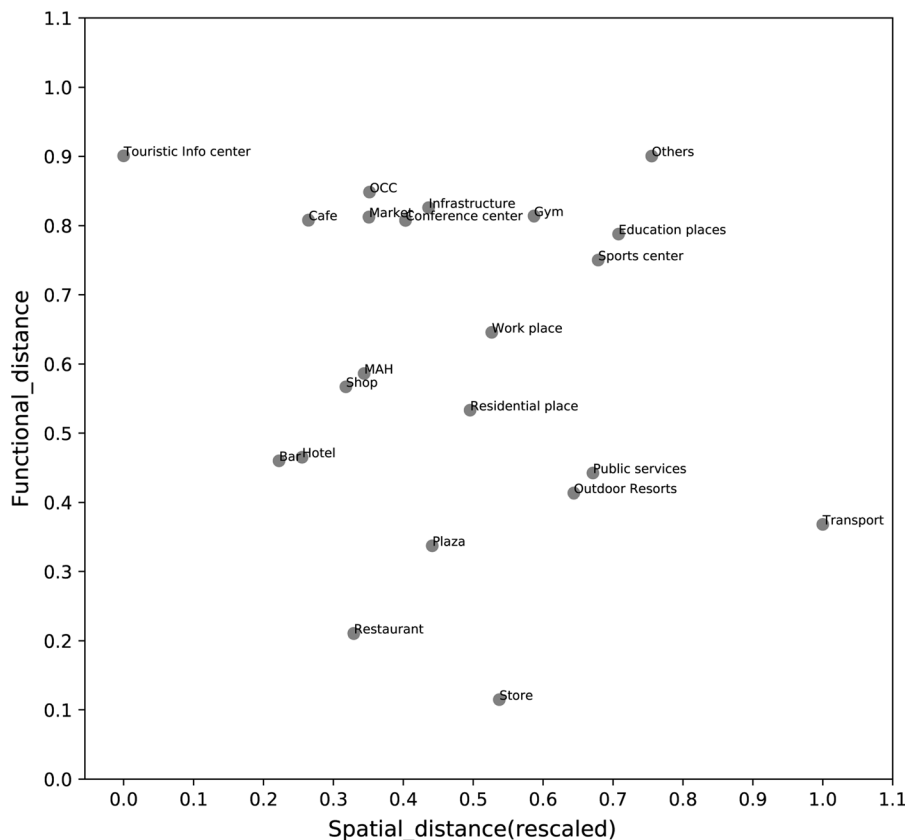


Fig. 10 Comparison between functional proximity and spatial proximity. *Source:* Authors

The longer period a person stays in one place; the more traces tend to be recorded. Hence, the unsupervised method (K-means clustering) can identify users who have extreme attributes. However, for those users without typical characteristics, manual intervention is unavoidable.

Secondly, this study confirms that the functional centroids of tourist activities differ from the centroids of the spatial distribution. The spatial distribution of tourist activities is concentrated in the airport and the central area of Barcelona city where most of the landmarks are located. The spatial relations of tourist activities, in essence, are decided by the locations of tourist attractions. Tourists by their nature tend to make check-ins around tourist attractions. However, their core functional usages are led by daily habits. Eating out is one of the most important activities, and thus POIs in the Restaurant category become the hinge of all functions.

Thirdly, the tourist-functional proximity of POIs highlights the major nodes of places with tourism functions. As Fig. 8 shows, the chain of functional places appeared when we raised the threshold for interaction values. The categories of Restaurant, Transport, Outdoor Resort, Hotel, etc. are the places of basic functions for tourists. Moreover, such functional proximities are distinct from the spatial proximities of POIs. The closer spatial proximity between two categories of POIs does not lead to closer functional proximity with each other.

Furthermore, the interaction values depict the main patterns of tourist flows. For example, Outdoor Resorts has a higher interaction value with Museums, Arts and Historical Places. This indicates that tourists tend to travel directly from one tourist attraction to another. Transport–Hotel has the highest interaction value because people usually need to drop their luggage at the hotel when they arrive and carry their luggage from the hotel to transport hubs as they leave.

Indeed, it is undeniable that Foursquare data has a bias. Because the function of Foursquare is to provide users with practical information about places, a high proportion check-ins are in the category of restaurants. However, the average number of check-ins at a restaurant is only 0.82, while the mean number of check-ins at Outdoor resorts and MAHs is much higher. Hence, such bias does not affect the general results in this study. Moreover, the result reflects the typical land uses of tourists in a city, such as hotel,

transport and tourist attractions; these categories have a higher intensity of flows. Essentially, Foursquare POIs are able to reflect tourist activities to some degree.

In summary, this paper reveals how tourist POIs functionally interact with each other. The method of evaluation of tourist-functional relations is possible to spread to other groups of people, e.g. a comparison between locals and tourists based on the same functional analysis. Moreover, the spatial distribution of our research also coincides with the pattern of tourist distribution in some other cities, as described in the literature review. This result is similar to the conclusion of Béjar et al. (2014) which utilizes Instagram data, a finding that indicates different datasets may be comparable, as long as the volume of the dataset is substantial.

It should be noted that the defect of data itself limits the scope of the present study. The meta-database that we received eliminated the personal profiles due to the potential privacy issues. Therefore, it lacks of any analysis of the background of users. For example, we cannot utilize the user's profile to identify their home country or to discuss whether the Foursquare behavior of foreign visitors differs from that of domestic visitors. Secondly, with the decline of popularity of Foursquare, the availability of its data has shrunk in Barcelona. The representative nature of more recent Foursquare data is questionable. The use of new data sources from currently popular LBSNs, such as Instagram and Twitter, will be necessary to examine and understand the latest dynamics of tourists. Last but not the least, it is worth to noting that the availability of Wi-Fi would impact the density of check-ins. Despite of most of the public spaces do have free Internet Access that provided by the City Council and Private Firms, there is a small risk of overrepresentation in best-served premises such as the Airport.

References

- Abbas, O. A. (2008). Comparisons between data clustering algorithms. *International Arab Journal of Information Technology (IAJIT)*, 5(3), 320–325.
- Abbasi, O. R., & Alesheikh, A. A. (2018). Exploring the potential of location-based social networks data as proxy variables in collective human mobility prediction models. *Arabian Journal of Geosciences*, 11(8), 173.

- Andrienko, N. V., & Andrienko, G. L. (2011). Spatial generalization and aggregation of massive movement data. *IEEE Transactions on Visualization and Computer Graphics*, *17*(2), 205–219.
- Béjar, J., Álvarez-Napagao, S., García-Gasulla, D., Gómez-Sebastià, I., Oliva, L., Tejada-Gómez, A., et al. (2014). Discovery of spatio-temporal patterns from Location Based Social Networks. In *CCIA* (pp. 126–135).
- Burns, M. C., Cladera, J. R., & Bergadà, M. M. (2008). The spatial implications of the functional proximity deriving from air passenger flows between European metropolitan urban regions. *GeoJournal*, *71*(1), 37–52.
- Burt, J. E., & Barber, G. M. (1996). *Elementary statistics for geographers*. New York: Guilford Press.
- Da Rugna, J., Chareyron, G., & Branchet, B. (2012). Tourist behavior analysis through geotagged photographs: A method to identify the country of origin. In *IEEE 13th international symposium on computational intelligence and informatics (CINTI)* (pp. 347–351). IEEE.
- Dou, X., Arellano Ramos, B., & Roca Cladera, J. (2018). China's inter-provincial population flow based on the interaction value analysis. *Geographical Research*, *37*(9), 1848–1861.
- Ferreira, A. P. G., Silva, T. H., & Loureiro, A. A. F. (2015). Beyond sights: large scale study of tourists' behavior using foursquare data. In *IEEE international conference on data mining workshop (ICDMW)* (pp. 1117–1124). IEEE.
- Gao, S., Janowicz, K., & Couclelis, H. (2017). Extracting urban functional regions from points of interest and human activities on location-based social networks. *Transactions in GIS*, *21*(3), 446–467.
- García-Palomares, J. C., Gutiérrez, J., & Mínguez, C. (2015). Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS. *Applied Geography*, *63*, 408–417.
- Girardin, F., Blat, J., Calabrese, F., Dal Fiore, F., & Ratti, C. (2008). Digital footprinting: Uncovering tourists with user-generated content. *IEEE Pervasive Computing*, *7*(4), 36–43.
- Girardin, F., Vaccari, A., Gerber, A., Biderman, A., & Ratti, C. (2009). Quantifying urban attractiveness from the distribution and density of digital footprints. <https://doi.org/10.2902/1725-0463.2009.04.art10>.
- Graser, A., Schmidt, J., Roth, F., & Brändle, N. (2019). Untangling origin-destination flows in geographic information systems. *Information Visualization*, *18*(1), 153–172.
- Hasan, S., Zhan, X., & Ukkusuri, S. V. (2013). Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. In *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing* (p. 6). ACM.
- Hasnat, M. M., & Hasan, S. (2018). Identifying tourists and analyzing spatial patterns of their destinations from location-based social media data. *Transportation Research Part C: Emerging Technologies*, *96*, 38–54.
- Hu, Y., McKenzie, G., Janowicz, K., & Gao, S. (2015). Mining human-place interaction patterns from location-based social networks to enrich place categorization systems. In *Proceedings of the workshop on cognitive engineering for spatial information processes at COSIT 2015*.
- Kádár, B. (2014). Measuring tourist activities in cities using geotagged photography. *Tourism Geographies*, *16*(1), 88–104.
- Kulín, H. W., & Kuenne, R. E. (1962). An efficient algorithm for the numerical solution of the generalized Weber problem in spatial economics. *Journal of Regional Science*, *4*(2), 21–33.
- Li, Y., Steiner, M., Wang, L., Zhang, Z. L., & Bao, J. (2013). Exploring venue popularity in foursquare. In *Proceedings IEEE on INFOCOM* (pp. 3357–3362). IEEE.
- Luo, F., Cao, G., Mulligan, K., & Li, X. (2016). Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago. *Applied Geography*, *70*, 11–25.
- Manca, M., Boratto, L., Roman, V. M., i Gallissà, O. M., & Kaltenbrunner, A. (2017). Using social media to characterize urban mobility patterns: State-of-the-art survey and case-study. *Online Social Networks and Media*, *1*, 56–69.
- Marine-Roig, E., & Clavé, S. A. (2015). Tourism analytics with massive user-generated content: A case study of Barcelona. *Journal of Destination Marketing & Management*, *4*(3), 162–172.
- Nikjoo, A., & Bakhshi, H. (2019). The presence of tourists and residents in shared travel photos. *Tourism Management*, *70*, 89–98.
- Preoțiuc-Pietro, D., & Cohn, T. (2013). Mining user behaviours: a study of check-in patterns in location based social networks. In *Proceedings of the 5th annual ACM web science conference* (pp. 306–315). ACM.
- Roca Cladera, J., Marmolejo Duarte, C. R., & Moix, M. (2009). Urban structure and polycentrism: Towards a redefinition of the sub-centre concept. *Urban Studies*, *46*(13), 2841–2868.
- Roca Cladera, J., & Moix, M. (2005). Critical surveys edited by Stephen Roper. *Regional Studies*, *39*(3), 357–373. <https://doi.org/10.1080/00343400500087372>.
- Sagl, G., Resch, B., Hawelka, B., & Beinat, E. (2012). From social sensor data to collective human behaviour patterns: Analysing and visualising spatio-temporal dynamics in urban environments. In *Proceedings of the GI-Forum* (pp. 54–63). Berlin: Herbert Wichmann.
- Scholtes, I. (2017). [arXiv:1702.05499](https://arxiv.org/abs/1702.05499).
- Smart, M. W. (1974). Labour market areas: uses and definition. *Progress in planning*, *2*, 239–353.
- Thuillier, E., Moalic, L., & Caminada, A. (2017). Dynamic purpose decomposition of mobility flows based on geographical data. In *24th International symposium on temporal representation and reasoning (TIME 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

- Vu, H. Q., Li, G., Law, R., & Ye, B. H. (2015). Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. *Tourism Management*, *46*, 222–232.
- Yang, D., Zhang, D., & Qu, B. (2016). Participatory cultural mapping based on collective behavior data in location-based social networks. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *7*(3), 30. Retrieved from

<https://sites.google.com/site/yangdingqi/home/foursquare-dataset?authuser=0>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.