Chapter 28
Towards Developing Smart Cities: Evidence from GIS Analysis on Tourists' Behavior Using Social Network Data in the City of Athens

Athanasios Koutras, Ioannis A. Nikas and Alkiviadis Panagopoulos

Abstract The evolution of web along with the technological explosion in web technologies and social networks has made internet a "place" where someone can post his/her tourist digitalized experiences easily, while these are made available and accessed worldwide almost instantly. In addition, such information is usually associated with features related to users' location and temporal position. This rapid development gives the motive for the creation of smart cities and corresponding smart tourism ecosystems. In this work, a density-based spatial clustering method is applied on posts from photo-sharing social media, to analyze geo-tagged data emanating from an urban area. The aim of this research is mainly to reveal the most important spots in well-known landmarks inside an urban area, and explore the visiting tendencies of tourists of these spots. The proposed method was applied on the city of Athens, and our results are presented using data from the Flickr photo-sharing web service.

A. Koutras (✉)
Department of Informatics and Mass Media, Technical Educational Institute of Western Greece, Patra, Greece
e-mail: koutras@teiwest.gr

I. A. Nikas · A. Panagopoulos
Department of Tourism Management, Technical Educational Institute of Western Greece, Patra, Greece

1   Introduction

The importance of Information and Communications Technologies (ICTs) in the distribution of tourism product is indisputable and seems to be an inevitable step towards the future of tourism development. This can be vividly experienced in urban areas. Cities as tourism destinations should be highly competitive and beyond that should co-create their tourism products along with their potential visitors (Smith 2015; Buhalis 2000). Furthermore, the available urban touristic resources, as well local infrastructures should be efficiently managed and routed under the visitors and native citizens demands and needs (Bădiţă 2013; Smith 2015).

The importance of ICTs in the functionality of a city as a tourism destination has resulted the notion of smart tourism and consequently the necessity for developing smart cities. The exact definitions of "smart tourism" and its derivatives are currently active issues for discussion and research (e.g. Gretzel et al. 2015a; Li et al. 2017). However, its continuously development motivates researchers to establish some new ideas, where "digital ecosystems and smart business networks" are combined to develop notions like Smart Tourism Ecosystems (Gretzel et al. 2015b). These kinds of environments allow tourism stakeholders to elevate and enrich their services using the existing city's digital environment (Gretzel et al. 2015b; Brandt et al. 2017).

The sustainability of these environments is crucially depended on the information regarding spatial-, as well as, temporal-oriented data. These data are usually related on visitors' habitudes, available infrastructures, urban landmarks etc. The official statistics provide accurate and reliable data, but present the drawback that they are published with a significance delay, so there is an inherent difficulty in extracting information about tourist behavior, especially in urban area environments (García-Palomares et al. 2015). In a smart city environment behavioral information is required and, ideally, should be accessed immediately. Furthermore, researchers are allowed to "cut" their dependence to official statistics (Shelton et al. 2015).

On the other hand, an enticing alternative is the available web information which is provided indirectly from social-network users. These data, in most cases are carrying attributes concerning temporal and spatial characteristics of user's activities all over the world. Social media-derived data constitute a source of value to a variety of tourism stakeholders, including tourism suppliers, destination marketing organizations (DMOs), and government agencies (Brandt et al. 2017).

In this direction, Brandt et al. (2017) modeled a Smart Tourism Environment which is based on social media analytics. Although, data originated from social media have been criticized on their reliability, representativeness, and the lack of theoretical background, they still offer a significant statistical alternative since big data datasets provide "a way to ask different kinds of questions than is possible with census data" (Shelton et al. 2015).

The analysis of social media data is not a new idea in tourism, but is an evolving field which needs further development and investigation. In this work, a set of web-originated data are collected and visualized based on their geolocation. Specifically, a set of published pictures in Flickr social network, concerning the city

of Athens in Greece, are collected and based on their geolocation attributes they were visualized. A spatial analysis on these data revealed the main landmarks that Flickr users visit. Based on this discovery, a temporal analysis was performed that revealed tourists' visiting habits, as well their statistical trends. It is worth mentioning that for the city of Athens the resulted landmarks coincide with the city's main landmarks.

2 Literature Review

Web 2.0 popularization and development change partially its content as it become user driven (User-Generated Content) (Bruns 2007; Straumann et al. 2014). At the same time, all the users' devices (smartphones, GPS devices etc.) generate geolocated information, which, aptly, is named big (geo)data (García-Palomares et al. 2015). Nowadays, the most common source of producing geolocated data are social networks (Batty 2013), which can be employed in various ways (Kitchin 2013).

The overall tourism experience includes trip preparation to the end of the trip back at home, which this usually involves internet and social networks (Fotis et al. 2012; Leung et al. 2013; Zeng and Gerritsen 2014; Munar and Jacobsen 2014). So, geotagged data beyond the extraction of tourists' destination preferences and points of interest in an urban area (Paldino et al. 2015; Junker et al. 2017) can also be used as consulting material for future visitors (Buhalis and Law 2008; Xiang and Gretzel 2010).

Several studies have developed to analyze social media data in connection with tourism activity, using geo-tagged data sourced on web services like Flickr and Panoramio: Wood et al. (2013) used photographs to estimate number of visitors in tourism sites; Girardin et al. (2008) studied the tourists' behavior in Rome analyzing their flows between various points of interest; Popescu et al. (2009) identified places people were visited as well as the duration of their stay; Gavric et al. (2011) extract Berlin's preferred locations and tourist dynamics; Kisilevich et al. (2013) identify popular city landmarks and events; Kurashima et al. (2013) used geotagged photos as a sequence of visited locations and then they recommend travel routes between landmarks; De Choudhury et al. (2010) introduced the creation of automated travel itineraries aiming in creating meaningful travel itineraries for individuals and professionals; Mamei et al. (2010) also recommend personalized routes using tourist experiences, behavior and tastes; Lu et al. (2010) suggest tourist trips based on phot-sharing geodata; Li (2013) used geotagged photographs to approximate the optimal solution for tourists' multi-day and multi-stay travel planning using the Iterated Local Search heuristic algorithm; Tammet et al. (2013) they develop sightmap exploiting photos density; Koerbitz et al. (2013) approximate the overnight stays in Austria and compare their results with the official statistics; Straumann et al. (2014) distinguish and study foreign and domestic

visitors in Zurich; Sun and Fan (2014) identify social events; García-Palomares et al. (2015) identify tourist hot spots in European metropolis using spatial statistical techniques to analyze location patterns.

The spatial-temporal analysis of these data constitutes, also, the main interest of this work.

3 Methodology

3.1 Experimental Setup

To collect our data for the analysis, we have used the Flickr (www.flickr.com) photo-sharing platform. Flickr offers publicly geo-tagged photo data to registered users via Flickr's Application Programming Interface (API) (www.flickr.com/services/api/) for non-commercial use. Among the photos that users upload to the platform, a great number is taken by tourists while they travel and visit places using GPS enabled photo cameras or smartphones. Picture information such as user name, geo-location of the photo, date and time that the photo was taken and uploaded, type of camera that was used, photo metadata, photo EXIF data, etc., can be queried and retrieved using requests in REST, XML-RPC or SOAP format. The platform's API response formats include REST, XML-RPC, SOAP, JSON, or PHP depending on the developer's request. In this work, we used REST requests to the API, while the responses were returned in JSON format.

To collect pictures for our experiments, we have used the flickr.photos.search method that provides a great number of optional arguments that can be used to narrow our search spatially as well as temporally. For the first case, a user has the choice of querying for pictures taken inside a geographic area in the form of a rectangle bounding box with user-defined bottom-left and top-right corner of the box. Additionally, the API offers the choice of including temporal information in the query as the user can define the minimum and maximum date that pictures were taken.

To extract photos using the Flickr API, we have written a script in PHP language using the following parameters in our query:

1. **Area of interest**: since we want to study visitors of the city of Athens, we define a bounding box with long/lat 23.5923,37.8058 (corner A), and 23.9178,38.1479 (corner B) that includes the city's administrative boundaries.
2. **Temporal period of interest:** we queried photos taken in the above area of interest in a period from 1/1/2009 to 15/10/2017.

Successful API calls return query results in JSON format. The results contain a rich amount of information about the photos, including photo ID, owner ID, photo title, photo geo-location (latitude and longitude), date that the picture was taken, date that the picture was uploaded, textual tags, as well as other type of information, such as the ID of the host server. Among the metadata, photo title, textual tags,

and location are optionally provided by users, while the other fields (e.g., photo ID and server ID) are automatically filled by Flickr when the photos are uploaded. Location information is available for each metadata record, since we have only retrieved geotagged photos. For our analysis, only the owner ID, the date that the photo was taken, the time that the photo was taken, the latitude/longitude, and the place_ID of the photo were retained.

The total number of photos we collected was 201.100. All data were pre-processed prior to the analysis to initially remove multiple photos taken from the same user in the same location in a very short period of time (multiple pictures taken within a minute). Additionally, to exclude locals from tourists that are our main subject of interest, we searched for users who appear in the database for a period longer than one month (García-Palomares et al. 2015), we categorized them as locals and excluded them. After the preprocessing step, the final number of photos in the database was 193.554.

3.2   Experiments

The purpose of this work is to explore big data publicly available from the internet to extract information about tourist concentration in urban regions. Using exploratory data analysis techniques on photograph geolocation data inside specific areas, we want to find places of interest (POIs) within a city that show big concentration of visitors. To achieve this, we have applied clustering techniques on the geolocation data of our database.

Due to the large number of clustering methods that exist in the literature (K-means, fuzzy C-means, Neural Network based etc.), we have set the following four specifications that must be met by an algorithm in order to be chosen for the purpose of our analysis:

1. A clustering technique should be used that doesn't require a pre-determined number of clusters, rather it works in an unsupervised manner as the exact number of clusters (POIs) cannot be not known beforehand.
2. The clustering method should result in clusters with a user defined minimum distance to each other. This is important as it can help us find POIs nearer to or further away from each other inside the city boundaries.
3. The user can define the minimum number of members in every cluster and thus define the minimum desired concentration of tourists in every POI
4. The clustering method must be robust to noise and classify only the significant examples, rejecting any noisy data.

Taking the above specifications into consideration, the DBSCAN (Ester et al. 1996) is a density-based spatial clustering method and it is used in our experiments. It is configured by two parameters Eps, the search radius, and MinPts, the minimum number of points within the search radius. These two parameters together define a

minimum density threshold, and clusters are identified at locations where the density of points is larger than the threshold.

For the selection of the parameters, different values of Eps and MinPts were tested. Parameter Eps is attributed to the scale of the regions that we want to cluster. A larger value results in broader POIs, while a smaller value creates smaller areas in the city. The second parameter MinPts defines the minimum number of points (cluster members) that is required to form a new cluster. A larger value of MinPts ensures higher significance for the detected clusters but may exclude some interesting different areas as it tends to unify them in a larger one. A smaller value of this parameter extracts more clusters but may also include noisy results.

In this paper, we have selected the value of Eps = 100 and MinPts = 2000 after experimentation that showed consistency in all our experiments.

4   Results

Application of the DBSCAN algorithm on the database, resulted in seven clusters that are shown in Fig. 1. Figure 1 has been created using the QGIS software (QGIS 2009). QGIS is a free and open source geographic visualization system that has been used by a large community to create, visualize, analyze and publish geospatial information in various applications that include but are not limited to tourist data, health data, etc. By inspection of this figure we can see that DBSCAN finds seven clusters covering the following areas inside the city of Athens: (a) Acropolis of Athens (b) Panepistimiou Street (c) Lycabettus Monument (d) Panathenaic Stadium

Fig. 1 DBSCAN algorithm applied on photo-sharing data from the city of Athens

Table 1 Results of the DBSCAN algorithm on Flickr geotagged photos in the city of Athens

Cluster	POI	Members
1	Acropolis of Athens	15 464
2	Panepistimiou Street	108511
3	Lycabettus Monument	1133
4	Panathenaic Stadium	21 416
5	Syntagma Square	12 658
6	Temple of Olympian Zeus	3743
7	National Archaeological Museum	2384
Noise		28 245

(e) Syntagma Square (f) Temple of Olympian Zeus and finally (g) National Archaeological Museum of Athens. The clusters with their members are shown in Table 1.

It is evident that the DBSCAN algorithm works very well and manages to cluster the geolocation data of the pictures into seven of the most important and well visited places of interest in the region of Athens. The above landmarks actually describe the city of Athens in its entity involving ancient monuments, as well as downtown important sites and buildings (www.visitgreece.gr).

To further analyze the POIs, we explored the temporal dynamics of each cluster. Using each photo's date and time from the database, we explored the temporal tourist concentration in every POI using four different time scales: (a) hourly within a day (b) daily within a week (c) monthly within a year and finally (d) yearly within the time data were gathered (2009–2017).

The results of this analysis are shown in Figs. 2, 3, 4 and 5.

By closer inspection of these Figures, the following remarks can be drawn:

(a) **Acropolis of Athens:** *"Acropolis" is the site of some of the most important masterpieces of worldwide architecture and art, the most renowned of which is*
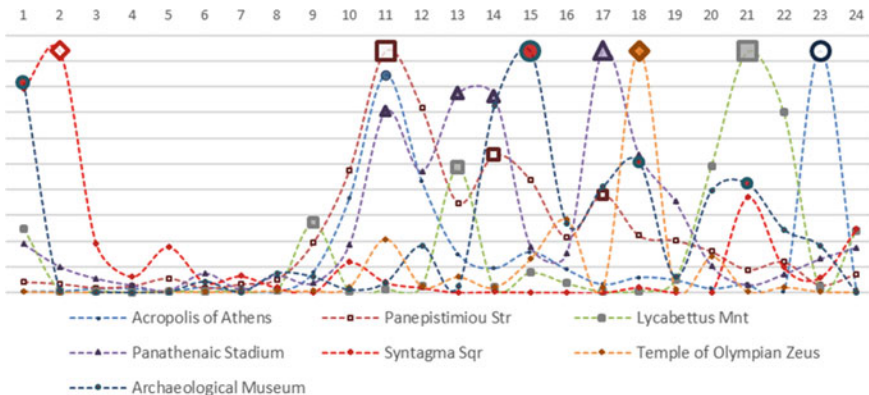

Fig. 2 Temporal tourist concentration in every POI using hourly time within a day scale
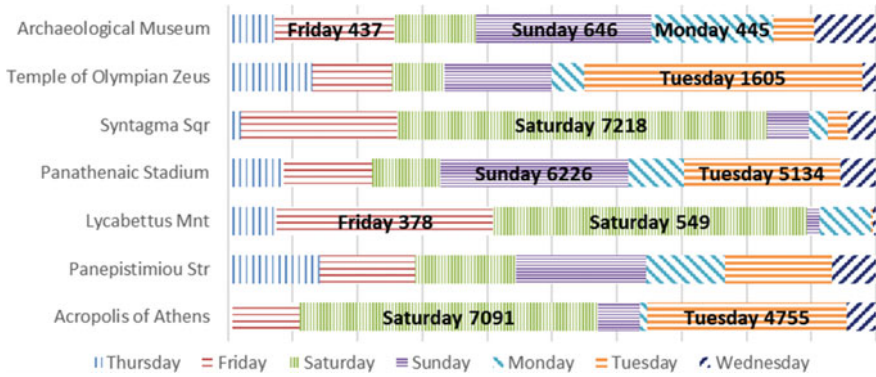
Fig. 3 Temporal tourist concentration in every POI using a daily within a week scale
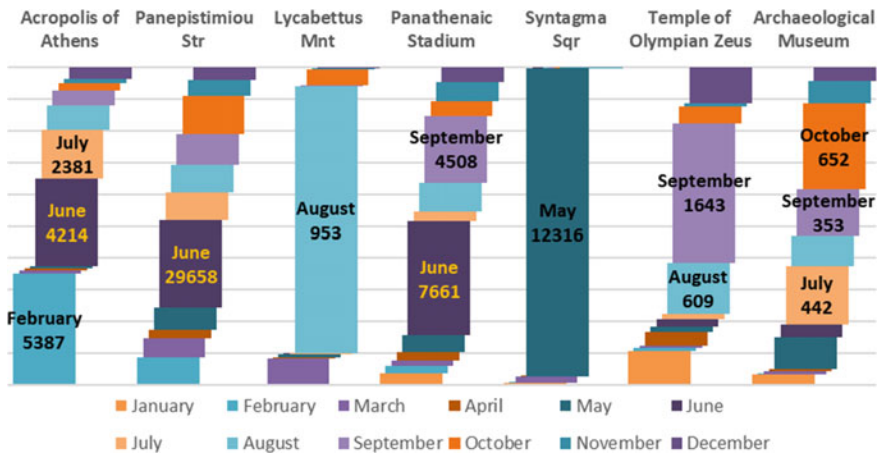

Fig. 4 Temporal tourist concentration in every POI using monthly within a year scale
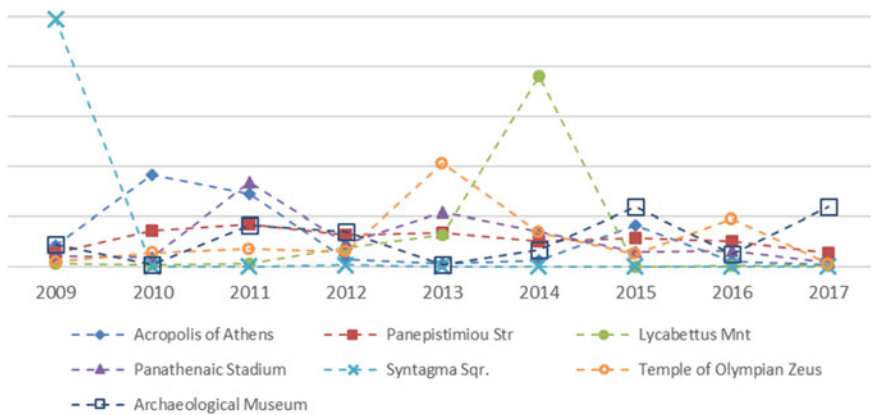

Fig. 5 Temporal tourist concentration in every POI using yearly within time data scale

the Parthenon temple. The Acropolis of Athens is mostly visited by tourists early in the morning 10–12 am and late at night, while a smaller group of visitors get there at around 3 pm. On a weekly basis, the days with the biggest crowd are Saturdays and Tuesdays, while on Mondays, Wednesdays and Thursdays, visits are very limited. Additionally, February, June and July are the months with the biggest attendance, while in January and April attendance is almost negligible.

(b) **Panepistimiou Street:** *Panepistimiou Street is a major street in Athens where some buildings of particular importance are there (some of them are Bank of Greece, the University of Athens, the Academy of Athens, the National Library, the Numismatic Museum, Titania Hotel, a part of the Grande Bretagne Hotel, and the Catholic Cathedral of Athens).* This area of the center of Athens shows a rise in concentration from early in the morning around 9 am until late in the evening 8 pm, with a peak value at 11–12 am. On a weekly basis, visits seem to be stable regardless the day, with an exception on Wednesdays. On a monthly basis, tourists visit this POI in all months with an exception of January.

(c) **Lycabettus Hill:** *Lycabettus Hill is the highest point of Athens which is known for nice view of the Acropolis, the Temple of Olympian Zeus, Panathenaic Stadium and the Ancient Agora.* Lycabettus Hill presents biggest tourist concentration in August during summer on the monthly scale, while most tourists choose to visit the hill Fridays and Saturdays, early in the morning (9–10 am), at lunch time (12–2 pm), or as most of this POI's visitors do, late at night 8–11 pm.

(d) **Panathenaic Stadium:** *The Panathenaikon (Kallimarmaro) Stadium is the old Olympic Stadium of Athens. It is the only stadium in the world built entirely of marble. The first Olympic Games in modern history were held there (1896).* Panathenaic Stadium similar to other open archaeological areas presents the biggest tourist concentration from early morning to late evening (9 am–8 pm) with a big decrease at 3 pm (coincides with the hottest time of a summer day). On a weekly basis, Sundays and Tuesdays are the two days with the most visitors, but in the other weekdays tourists also visit this premise as well. On a monthly scale, June and September are the months that present the most visitors.

(e) **Syntagma Square:** *Syntagma is the central square of Athens. It is located in front of the Old Royal Palace which is housing the Greek Parliament.* Syntagma square is mostly a place where people meet, with hotels nearby and the shopping streets around. Data analysis on the Syntagma cluster showed that most people visit the place from evening till late at night 8 pm–2 am almost exclusively on weekends (Friday night and Saturday night).

(f) **Temple of Olympian Zeus:** *The temple of Olympian Zeus was one of the largest temples in antiquity and close to Hadrian's Arch, which forms the symbolic entrance to the city.* This archaeological site presents the biggest concentration early in the morning 9–10 am or in the evening 5–8 pm. Tourist visits seem to be spread during all days of the week, with Tuesday being the

busiest day. On the monthly basis, December–January, and August–September show the biggest proportion of the tourist attraction.

(g) **National Archaeological Museum:** *The National Archaeological Museum in Athens houses some of the most important artifacts from a variety of archaeological locations around Greece from prehistory to late antiquity.* The area inside and round the old Archaeological Museum of Athens shows a big concentration after 1 pm until the late night hours, especially on Saturday and Sunday during all year with a small exception in February and April.

5 Discussion

In this work a spatial-temporal analysis on geolocation data was presented. The used data came from Flick photo-sharing web service and concerns the city of Athens downtown, capital of Greece. The analysis was performed using a density-based spatial clustering algorithm (DBSCAN). The results of the algorithm were seven clusters corresponding to seven lively places in the city of Athens, as well as some patterns regarding foreigner visitors' temporal preferences.

This research is the primary stage of studying visitors' tendencies, since data from more social media should be considered and the correlation with the official statistics should be studied.

However, the city authorities should consider these studies (even in this primary stage) to provide up-to-date tourism information and improve the tourist flows in well-known POIs in the peak of day, week, month, or year, respectively. On the other hand, visiting voids in major landmarks should be a starting point for discussion and new strategic plans concerning all partners involved in tourism production aiming to increase, respectively, tourist flows in these POIs.

Finally, from the perspective of tourism product consumers, especially in urban environments, the availability of this kind of information will take the place of a "tourism flow regulator", giving them the opportunity to develop personalized visiting plans, thus improving their tourism experience.

References

Bădiță, A. (2013). Approaches to the analysis and evaluation of urban tourism system within urban destinations. *Revista de turism-studii si cercetari in turism, 16,* 58–66.

Batty, M. (2013). Big data, smart cities and city planning. *Dialogues in Human Geography, 3*(3), 274–279.

Brandt, T., Bendler, J., & Neumann, D. (2017). Social media analytics and value creation in urban smart tourism ecosystems. *Information & Management, 54*(6), 703–713.

Bruns, A. (2007, June). Produsage. In *Proceedings of the 6th ACM SIGCHI Conference on Creativity & Cognition* (pp. 99–106). ACM.

Buhalis, D. (2000). Marketing the competitive destination of the future. *Tourism Management, 21*(1), 97–116.

Buhalis, D., & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research. *Tourism Management, 29*(4), 609–623.

De Choudhury, M., Feldman, M., Amer-Yahia, S., Golbandi, N., Lempel, R., & Yu, C. (2010, June). Automatic construction of travel itineraries using social breadcrumbs. In *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia* (pp. 35–44). ACM.

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996, August). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (Vol. 96, No. 34, pp. 226–231).

Fotis, J., Buhalis, D., & Rossides, N. (2012). *Social media use and impact during the holiday travel planning process* (pp. 13–24). Springer.

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,* 408–421.

Gavric, K. D., Culibrk, D. R., Lugonja, P. I., Mirkovic, M. R., & Crnojevic, V. S. (2011, October). Detecting attractive locations and tourists' dynamics using geo-referenced images. In *2011 10th International Conference on Telecommunication in Modern Satellite Cable and Broadcasting Services (TELSIKS)* (Vol. 1, pp. 208–211). IEEE.

Girardin, F., Calabrese, F., Dal Fiore, F., Ratti, C., & Blat, J. (2008). Digital footprinting: Uncovering tourists with user-generated content. *IEEE Pervasive Computing, 7*(4).

Gretzel, U., Sigala, M., Xiang, Z., & Koo, C. (2015a). Smart tourism: foundations and developments. *Electronic Markets, 25*(3), 179–188.

Gretzel, U., Werthner, H., Koo, C., & Lamsfus, C. (2015b). Conceptual foundations for understanding smart tourism ecosystems. *Computers in Human Behavior, 50,* 558–563.

Junker, C., Akbar, Z., & Cuquet, M. (2017, September). The network structure of visited locations according to geotagged social media photos. In *Working Conference on Virtual Enterprises* (pp. 276–283). Cham: Springer.

Kisilevich, S., Keim, D., Andrienko, N., & Andrienko, G. (2013). Towards acquisition of semantics of places and events by multi-perspective analysis of geotagged photo collections. In *Geospatial Visualisation* (pp. 211–233). Berlin, Heidelberg: Springer.

Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography, 3*(3), 262–267.

Koerbitz, W., Önder, I., & Hubmann-Haidvogel, A. C. (2013). Identifying tourist dispersion in Austria by digital footprints. In *Information and Communication Technologies in Tourism 2013* (pp. 495–506). Berlin, Heidelberg: Springer.

Kurashima, T., Iwata, T., Irie, G., & Fujimura, K. (2013). Travel route recommendation using geotagged photos. *Knowledge and Information Systems, 37*(1), 37–60.

Leung, D., Law, R., Van Hoof, H., & Buhalis, D. (2013). Social media in tourism and hospitality: A literature review. *Journal of Travel & Tourism Marketing, 30*(1–2), 3–22.

Li, X. (2013, November). Multi-day and multi-stay travel planning using geo-tagged photos. In *Proceedings of the Second ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information* (pp. 1–8). ACM.

Li, Y., Hu, C., Huang, C., & Duan, L. (2017). The concept of smart tourism in the context of tourism information services. *Tourism Management, 58,* 293–300.

Lu, X., Wang, C., Yang, J. M., Pang, Y., & Zhang, L. (2010, October). Photo2trip: Generating travel routes from geo-tagged photos for trip planning. In *Proceedings of the 18th ACM International Conference on Multimedia* (pp. 143–152). ACM.

Mamei, M., Rosi, A., & Zambonelli, F. (2010, July). Automatic analysis of geotagged photos for intelligent tourist services. In *2010 Sixth International Conference on Intelligent Environments (IE)* (pp. 146–151). IEEE.

Munar, A. M., & Jacobsen, J. K. S. (2014). Motivations for sharing tourism experiences through social media. *Tourism Management, 43,* 46–54.

Paldino, S., Bojic, I., Sobolevsky, S., Ratti, C., & González, M. C. (2015). Urban magnetism through the lens of geo-tagged photography. *EPJ Data Science, 4*(1), 5.

Popescu, A., Grefenstette, G., & Moëllic, P. A. (2009, November). Mining tourist information from user-supplied collections. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management* (pp. 1713–1716). ACM.

QGIS Development Team. (2009). *QGIS Geographic information system*. Open Source Geospatial Foundation. http://qgis.osgeo.org.

Shelton, T., Poorthuis, A., & Zook, M. (2015). Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landscape and Urban Planning, 142,* 198–211.

Smith, R. (2015). *SMART tourism tools: Linking technology with the touristic resources of city destinations* (Doctoral dissertation, NHTV Breda University of Applied Sciences).

Straumann, R. K., Cöltekin, A., & Andrienko, G. (2014). Towards (re) constructing narratives from georeferenced photographs through visual analytics. *The Cartographic Journal, 51*(2), 152–165.

Sun, Y., & Fan, H. (2014). Event identification from georeferenced images. In *Connecting a digital Europe through location and place* (pp. 73–88). Cham: Springer.

Tammet, T., Luberg, A., & Järv, P. (2013). Sightsmap: Crowd-sourced popularity of the world places. In *Information and communication technologies in tourism 2013* (pp. 314–325). Berlin, Heidelberg: Springer.

Wood, S. A., Guerry, A. D., Silver, J. M., & Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. *Scientific reports, 3,* 2976.

Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management, 31*(2), 179–188.

Zeng, B., & Gerritsen, R. (2014). What do we know about social media in tourism? A review. *Tourism Management Perspectives, 10,* 27–36.