Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS

Juan Carlos García-Palomares*, Javier Gutiérrez, Carmen Minguez

Departamento de Geografía Humana, Universidad Complutense de Madrid, C/Profesor Aranguren, s/n, 28040 Madrid, Spain

ABSTRACT

New sources of geolocated information, associated with big data and social networks, show great promise for geographical research, especially in the field of tourism geography. Photo-sharing services comprise one of these sources. The aim of this article is to demonstrate the potential of photo-sharing services for identifying and analyzing the main tourist attractions in eight major European cities: Athens, Barcelona, Berlin, London, Madrid, Paris, Rome and Rotterdam. Geotagged photographs on Panoramio were differentiated according to whether they had been taken by tourists or local residents, and their spatial distribution patterns were analyzed using spatial statistical techniques in a GIS. The results indicated the concentration and dispersion of photographs in each city and their main hot spots, and revealed marked differences between tourists’ and residents’ photographs, since the former showed higher spatial concentrations. In addition, differences were observed between cities; Barcelona and Rome presented a strong spatial concentration compared with London or Paris, which showed much greater dispersion.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

The generation of geolocated information is rocketing, popularizing the concept of big data (one could say big (geo)data in the case of geolocated data). In today’s technological age, all human activity leaves a trace. Networks of fixed and mobile sensors such as smartphones, GPS, credit cards, travel cards, etc. monitor the public’s spatial behavior throughout the day. As Batty (2013, 274) has noted, much if not most of what we now call big data is produced automatically and routinely by various types of sensor. To all these sources can be added the information generated on social networks. The geolocation of tweets and the analysis of Facebook relationships are two of the best-known examples, and they offer vast possibilities for the analysis of space, social networks and their spatial traces.

These new sources of geolocated information present numerous opportunities (Kitchin, 2013). Perhaps the greatest of these is that they offer the possibility of working with high spatial and temporal resolution data and render it feasible to study cities dynamically using constantly updated information, thus paving the way for a new theoretical perspective associated with this opportunity for analyzing cities in the “short term”. The breadth of available data sources has allowed researchers to end their dependence upon official statistics on demographics, economic activity, traffic, and any number of other urban indicators (Shelton, Poorthuis, & Zook, 2015). The new data is certainly enriching our experiences of how cities function (Batty, 2013).

One of the fields which offer the most opportunities is tourism geography (Chareyron, Da-Rugna, & Raimbault, 2014). These days, enjoyment of the tourist experience commences in the stage of trip preparation prior to the activity itself, and this frequently involves use of the Internet and social networks (Fotis, Buhalis, & Rossides, 2012; Leung, Law, Van Hoof, & Buhalis, 2013; Zeng & Gerritsen, 2014). Furthermore, it concludes at home, when photographs and experiences are made available to networks of friends or other potential tourists (Munar & Jacobsen, 2014). Along the way, an enormous amount of information about tourist destinations is generated, which potential tourists can consult before their trip, to the point that they often feel they have already been there before they have actually gone (Buhalis & Law, 2008; Xiang & Gretzel, 2010). This phenomenon is also of great interest to urban
planners and researchers, who can access hitherto virtually unknown information. Consequently, studies have begun to appear using new sources of data to complement the information provided by the traditional sources and official statistics employed to date to analyze tourism. Various sources have been used, including email use, phone calls, and social networks such as Twitter and Facebook, but some of the most promising sources are social networks for sharing geotagged photographs (Leung et al. 2013). Communities as Instagram, Flickr and Panoramio enable users to share photographs of the places they have visited, with the exact location of the site where they were taken (Kachkaev & Wood, 2013).

The aim of this article is to demonstrate the possibilities offered by these photo-sharing services for identifying and analyzing the most popular visual attractions of a city. To this end, we used geo-location information contained in photographs downloaded from the Panoramio website, having previously processed the data in order to classify the photographs according to whether they had been taken by tourists or locals. The information obtained was then analyzed in a GIS using spatial statistical indicators. This provided insight into tourist movements in each city, and enabled us to conduct a comparison between cities. Here, we analyze spatial patterns in Europe’s major cities. Although some previous studies have mapped the intensity of photographs on one or another of these photo-sharing services (for example, Kisilevich, Krstajic, Keim, Andreienko, & Andreienko, 2010; Kadar, 2014), we did not find any papers in the literature concerning the kind of research reported here. In the present study, we mapped the intensity of the phenomenon, facilitating a comparison between cities (processing the data in order to differentiate between tourists’ and locals’ photographs), and this approach was combined with an analysis of the spatial distribution patterns presented by these photographs using spatial statistical techniques in a GIS. Our analysis revealed the intensity of tourism in each of the cities, the degree of concentration-dispersion and the location of spatial clusters.

The use of geo-located photographs fills a gap in tourism data sources, offering useful tourist information for planning and management, public and private, and for tourists themselves. The density of photographs reveals the distribution of tourists’ presence in the whole city, and consequently its carrying capacity, compared to traditional head counting at specific points and surveys. Data can be cross-checked by correlation regression models. With this information it is possible to identify areas of saturation and establish controls for not surpassing carrying capacity. In addition, with the data obtained in this paper it is possible: 1) To know the most visited places in the city and make decisions in choosing accommodation, depending on the accessibility of tourist attraction points, which represents a major business opportunity. 2) To provide optimal location of services (eg. information points) or commercial activities to tourists. 3) To know the most visited places, being essential for marketing destinations (BMD) (Hays, Page, & Buhllis, 2013), as it is closely related to the new images of cities. 4) To study densities of photographs which helps identify both touristically unexploited areas and allows planners to devise strategies to enhance their value. Sometimes these strategies are related to physical characteristics (public planners) and in others to the creation of tourist products and services (private planners). 5) To be used in visitor planning trips. 6) To report the most photographed streets and spaces, and therefore most popular, to manage tourist flows and to identify and propose new tourist routes. Also, from the perspective of tourists themselves, this information is very useful for trip planning and decision making during their visit.

The present article is divided into five sections. Following this introduction, in Section 2 below we discuss the potential of photo-sharing services and review previous studies which have used these data. In Section 3, we define the study areas and describe the data used and the indexes applied for their analysis. The results are given in Section 4. Lastly, the main conclusions are presented in Section 5.

2. Background

Big data presents considerable challenges and opportunities for tourism geography because of the conjunction of two elements: the difficulty involved in extracting information about tourist behavior from official statistics (especially in the inner cities), and the quantity of new Web 2.0 sources that have emerged connected with tourism activity. Big data unquestionably represents a new challenge for tourism (Chareyron et al., 2014).

New data sources make it possible to conduct new analyses, or complete those already in existence, of tourists’ use of space at different scales. Several studies have analyzed email and social network information to determine tourist flows on a global scale. Geographic location data on more than 100 million anonymous Yahoo email users have been used to study flows, differentiating between those of a migratory nature and tourism (Weber & Zagheni, 2013; Zagheni & Weber, 2012). Hawelka et al. (2014) has also studied flows at a global scale, in this case using geotagged tweets, and has attempted to validate this information by comparing the results obtained with official statistics on world tourism.

However, it is on an urban scale that these new sources show the greatest promise. Web user communities contain various sources of geolocated information, and this has enormous potential for leverage in tourism research. For example, Foursquare (https://es.foursquare.com/) is a web-based location service applied to social networks. The main idea is to mark (check-in) specific places where the user is located and earn points by “discovering” new places. Based on the information entered by users, the service has evolved into a recommender system that makes smart suggestions about places of interest. On Facebook Places (https://es-es.facebook.com/about/location), users can share the place where they are located, where they have been or where they are going, or record the location of their favorite photographs. Both sources can be used to conduct an analysis of service supply and use in any city. Zhai et al. (2015) use these sources to analyze patterns of urban restaurant locations, according to user assessments.

Crowdsourced data available from photo-sharing services have the potential for serving as a measure of space attractiveness (Kachkaev & Wood, 2013). Three communities in particular provide the option of geotagging photographs:

- Instagram (https://instagram.com/), a social network that allows users to apply effects to their photographs or videos and then share them on various social networks such as Facebook, Twitter, Tumblr and Flickr. By the end of 2014, Instagram had more than 300 million users.
- Flickr (http://www.flickr.com/map), owned by Yahoo, has a large community of users who share photographs and videos that they have created. Both provide the option of geotagging photographs.

1 Whereas the early Web was primarily one-directional, allowing a large number of users to view the contents of a comparatively small number of sites, the new Web 2.0 is a bi-directional collaboration in which users are able to interact with and provide information to central sites, and to see that information collated and made available to others (Goodchild, 2007). Diaz et al. (2012) have shown the possibilities Web 2.0 offers to manage and integrate crowd sources as new sources of information in applied geography.
photographs. However, with both Instagram and Flickr, the emphasis is on editing and retouching photographs.

- **Panoramio** ([http://www.panoramio.com](http://www.panoramio.com)) places more stress on geotagging photographs shared by users. The site enables users to view photographs of places or landscapes that other users have taken and uploaded once they have been geotagged. Its objective is to allow users to learn more about a given area of the world by viewing the photographs that other users have taken at that place. These images can be viewed on the Panoramio website or via Google Earth. Since its launch in October 2005, it has received more than 5 million photographs in less than two years. In July 2007, Panoramio was acquired by Google. Since then, the number of uploads has increased exponentially to exceed 100 million images in December 2013.

The volume of information in these three major photo-sharing services reduces the impact of incorrect geotagging, but there is also a correction process whereby any user who believes that another user's photograph is incorrectly located can suggest to the author that he or she corrects the location, using a map to indicate the position he or she considers correct and even sending a short message explaining the reasons for the suggestion. Several authors have also analyzed the positional quality of these sources (Kachkaev & Wood, 2013). Zielstra and Hochmair (2013) analyzed the positional accuracy of 1433 images for 45 areas in four selected world regions by comparing the geotagged position of photographs to the manually corrected camera position based on the image content. Their analysis revealed a better positional accuracy for Panoramio than for Flickr images.

Other authors have created maps using geodata from these photo-sharing services, or combining geotagged photographs with other sources. Based on the density of Panoramio photographs, Tammet, Luberg, and Järv (2013) developed Sightsmap, a map viewer showing the most popular sights. Eric Fischer used photo-locations of the public Flickr and Picasa search APIs to create three map albums: The viewer showing the most popular sights. Eric Fischer used photo-sharing services, or combining geotagged photographs with the positional accuracy of 1433 images for 45 areas in four selected world regions by comparing the geotagged position of photographs to the manually corrected camera position based on the image content. Their analysis revealed a better positional accuracy for Panoramio than for Flickr images.

In addition to these cartographic representations, other studies have begun to appear that use geodata from photo-sharing services for different types of applications. In line with previous research in which photo-sharing services were used to identify the main tourist attractions and the intensity of their use, Popescu, Grefenstette, and Moellic (2009) identified the places people visited, the duration of their stay and panoramic spots at these destinations from photographs uploaded on Flickr and covering 183 cities in different parts of the world. Kislevich, Keim, Andrienko, and Andrienko (2013) used Flickr and Panoramio data and spatio-temporal analysis to identify attractive places and points of interest. In another study, photo-sharing services were used to identify popular city landmarks and events (Kislevich et al., 2010). Gavric, Culibrk, Lugonja, Mirkovic, and Crnojevic (2011) used Flickr data to identify attractive locations in the city of Berlin and determine tourist dynamics. Meanwhile, Straumann, Cöltken, and Andrienko (2014) conducted a very exhaustive study in Zurich. They analyzed the number of photographers, how many pictures they took (and shared), when (studying various temporal patterns) and where people took (and shared) these pictures (and where not) as well as how they moved throughout the city. These authors distinguished foreign visitors from domestic visitors and examined differences and similarities of the coverage and the movements of these two groups in space. Sun and Fan (2014) used geotagged photographs and binary logistic regression to identify social events (e.g., festivals, parades, protests, sports, etc.). Koerbitz, Onder, and Hubmann-Haidvogel (2013) used Flickr data and polynomial regression to estimate actual tourist bed nights, and evaluate according to these estimates how representative Flickr data is in comparison to actual tourist numbers in Austria. To do this, they divided Flickr users into tourists and local residents, based on their activity time span. Their results show that Flickr data can be used as an estimation of actual tourist numbers in Austria. Zhou, Liu, Oliva, and Torralba (2014) attempted to characterize city identities using a dataset with 2 million geotagged images from 21 cities over 3 continents collected from Panoramio. First, they estimated the scene attributes of these images and used this representation to build a higher-level set of 7 city attributes, tailored to the form and function of the cities. Then, they conducted city identity recognition experiments on the geotagged images, identifying images with a salient city identity for each city attribute. Based on the misclassification rate of the city identity recognition, they analyzed the visual similarity among different cities. Finally, they discussed the potential application of computer vision to urban planning.

Various studies have used information from photo-sharing services to analyze tourist movements and propose or assess tourist routes. Girardin, Calabrese, Fiore, Ratti, and Blat (2008) examined uploaded photographs of Rome. They analyzed tourist movements and created a map that shows where tourists go and the density of tourists in these areas. Kurashima, Iwata, Irie, and Fujimura (2013) proposed a travel route recommendation method that makes use of the photographers’ histories as held by social photo-sharing sites. Assuming that each photographer’s collection of geotagged photos is a sequence of visited locations, photo-sharing sites are important sources for gathering tourists’ location histories. By following their location sequences, representative and diverse travel routes can be found that link key landmarks. De Choudhury et al. (2010) examined the creation of automated travel itineraries by using data from Flickr that show when the photograph was taken and its associated geographical location and semantic tags. The automatically created travel itineraries were compared with popular professional bus tours by crowdsourcing, and the results show that Flickr data is useful for creating meaningful travel itineraries. Also with the help of Flickr, Mamei, Rosi, and Zambonelli (2010) developed a system that learns and stores users’ tourist experiences, behavior and tastes to recommend personalized routes. Kurashima et al. (2013) and Lu, Wang, Yang, Pang, and Zhang (2010) have leveraged these sources to suggest tourist trips. Li (2013) used Panoramio data and the Iterated Local Search heuristic algorithm to find an approximate optimal solution for tourists’ multi-day and multi-stay travel planning (different places of accommodation and transportation).

However, despite this incipient interest in geotagged photograph data, none of these studies has used spatial statistical techniques to analyze location patterns and thus contribute to our knowledge about tourist behavior in cities.

---

1 http://www.panoramio.com/
2 http://www.sightsmap.com/.
3. Methodology and sources

3.1. Study areas

To analyze the potential of Panoramio, we selected eight cities in the European Union: Athens, Barcelona, Berlin, London, Madrid, Paris, Rome and Rotterdam. The study areas were defined in accordance with criteria that would permit their comparison, establishing a similar size in terms both of the data downloaded and the analyses performed. Thus, study areas were defined from a central point in each city to the edge of a buffer zone 12 km from the center (a total surface area of about 314 km²). Based on the location of photographs taken within these 12 km radius areas and downloaded from Panoramio, analyses were conducted of a smaller area with a radius of 10 km. Our goal in working with 10 km to the city center plus 2 km outside the final study area was to try to avoid an “edge effect” in the results of the analyses.

3.2. Data

The data were downloaded from the Panoramio website API. To ensure comparability, the parameters used for downloading were the same for each city. A similarly sized bounding box was established for all the cities, divided into a grid of 400 sub-regions in order to increase the number of photographs downloaded and streamline processes. The data obtained by downloading were samples of all the photographs stored, and contained information about the geographic coordinates, the ID of the owner of the photograph, a URL link to the photograph and the date on which it was uploaded. Downloaded locations corresponded to photographs uploaded to Panoramio between 2005 and 2014.

Once downloading had concluded, the location of each of the photographs was georeferenced and analyzed in a GIS, using ArcGIS 10.3. Downloading generated "csv" files, which contained the geographical coordinates of the location of each of the photographs. These coordinates were used to create a layer of points for each of the locations. The files contained a large volume of locations, ranging from 126,000 for Rotterdam to over 335,000 for London (Table 1).

3.3. Methods

Geotagged photographs on Panoramio show a clear and high concentration around tourist areas in cities. However, since these are the most visually attractive places in cities, it was possible that a considerable proportion of the photographs had been taken by local residents. The criterion used to determine whether the photographs were taken by visitors or local residents was the period during which each user had taken pictures: if this period exceeded one month, then the photographs were attributed to residents; if the period was less than one month, then they were attributed to tourists. This methodology is similar to that used by Fischer for his Geotaggers' World Atlas.

Density maps provided an initial visual overview of the density distribution of the photographs in the cities studied. Descriptive statistics were used to determine those cities with a higher density and spatial concentration of tourists’ photographs.

For specific location analyses, the standard distance of the photographs was calculated to measure the degree to which features were concentrated or dispersed around the geometric mean center.

Lastly, data aggregated in hexagonal grids were used to perform an analysis of location patterns using spatial statistical indicators. Two indicators were calculated to determine global location patterns: the Getis-Ord General G statistic and the Global Moran’s I statistic. The General G index measures the degree of clustering for either high or low values, while Moran's I index measures spatial autocorrelation based on feature locations and attribute values. The Anselin Local Moran’s I (LISA statistic) was used to identify local trends in the location of the intensity of tourism in each of the study areas. LISA analysis identifies concentrations of high values, concentrations of low values, and spatial outliers (Anselin, 1995). Contiguity by edges was the method employed to define neighborhood. Megler, Banis, and Chang (2014) have used a similar methodology to analyze graffiti location patterns in the city of Los Angeles.

Although the main goal of the analysis was to identify the spatial distribution patterns of tourists’ photographs, pictures taken by local residents were also analyzed in order to determine whether both sets of photographs presented similar patterns.

4. Results

4.1. Intensity

Fig. 2 shows the distribution of the number of photographs in each of the hexagons into which the cities were divided. Fig. 2a and c shows the distribution of photographs differentiating between those taken by tourists or locals. London presented the highest density of tourists’ photographs, with a mean of 22 photographs per hexagon, followed by Paris (Table 1). In London, the phenomenon extended over large areas of the city, as reflected by the low coefficient of variation (CV); Barcelona also presented a very high density (mean values were influenced by its coastal location) and showed the most highly concentrated distribution of tourists’ photographs, with the highest CV.

The most photographed sites were the cities' main monuments, but photographs were also taken of other tourist attractions, such as...
Table 1
Photograph density statistics.

<table>
<thead>
<tr>
<th>City</th>
<th>Tourists’ photographs</th>
<th>Locals’ photographs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>City Number of photographs</td>
<td>Mean density (Photographs/hexagon)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>London</td>
<td>99,679</td>
<td>22.1</td>
</tr>
<tr>
<td>Paris</td>
<td>85,154</td>
<td>19</td>
</tr>
<tr>
<td>Barcelona</td>
<td>77,552</td>
<td>17.3</td>
</tr>
<tr>
<td>Berlin</td>
<td>58,320</td>
<td>13</td>
</tr>
<tr>
<td>Madrid</td>
<td>43,296</td>
<td>9.8</td>
</tr>
<tr>
<td>Rome</td>
<td>56,817</td>
<td>12.6</td>
</tr>
<tr>
<td>Athens</td>
<td>38,998</td>
<td>8.7</td>
</tr>
<tr>
<td>Rotterdam</td>
<td>28,054</td>
<td>6.3</td>
</tr>
<tr>
<td></td>
<td>All photographs</td>
<td>74.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1. Downloaded photographs of the city of Madrid, differentiating between those taken by tourists or local residents.
Fig. 2. a: Density of photographs (total). b: Density of photographs (tourists vs locals). c: Density of photographs (tourists vs locals).
as football stadiums. Thus, for example, the highest number of photographs per hexagon in London corresponded to the British Museum, Camden Market and Tower Bridge, and to Nou Camp (Barcelona FC Stadium) and Gaudí’s Park Güell and Sagrada Familia church in Barcelona.

Although tourists’ photographs presented similar spatial patterns to those of residents’ photographs, their distribution was clearly more highly concentrated, with markedly higher CVs for tourists’ photographs in all the cities. Both tourists and locals tended to photograph the cities’ most representative sites, generally located in historic centers. However, locals also extended their radius of action to sites rarely frequented by tourists, such as parks or recreational areas on the periphery of the cities, resulting in more scattered distribution patterns.

4.2. Distance

The standard distance was used to describe the degree of dispersion of the photographs with respect to the center, and thus the extent to which tourism extended out from the historic center in each city. Table 2 shows the standard distances (in meters) calculated for each city according to specific distributions. Athens presented the highest value for deviation from the center, due to the presence of two peripheral but high density focal points, one in the southwest (Piraeus Port) and to a lesser extent one in the northeast (Olympics area), in addition to the central focus around the Parthenon. Berlin and London also showed a scattered distribution, as indicated earlier by the CV data on density distribution. Paris and Madrid presented a medium degree of dispersion, whereas the highest concentrations were observed in Barcelona, Rome and Rotterdam. There is no clear trend when comparing tourist and locals distances. In some cities distances of the former are somewhat larger than the latter, and in other cities the opposite is true.

4.3. Spatial autocorrelation: identifying spatial clusters

Indexes such as the Getis-Ord General G statistic and Moran’s Index can be used to analyze the global spatial distribution patterns presented by photograph locations. The G statistic revealed a very marked trend toward a concentration of high values (high clusters), with very high statistical significance in all cities (p-value 0.00000). Similarly, Moran’s Index also indicated a marked spatial autocorrelation, with a very strong tendency in all cases toward the formation of spatial clusters (p-value 0.00000).

Anselin Local Moran’s I statistic was calculated in order to map the presence of these clusters at local level\(^9\) (Fig. 3). As might be expected, High–High (HH) clusters abounded in all cities, but again showed different spatial patterns. While the location of HH clusters tended to be central in cities such as Rome, Rotterdam and to a lesser extent Barcelona and Athens, in Berlin, Paris and London they tended to occupy more hexagons and were more dispersed.

A few cases presented outliers, in the form of Low - High (LH) or High-Low (HL) clusters. These latter are interesting in that they indicate attractions outside tourist areas. A good example of this is the HL cluster located in northwest London, in a hexagon containing the famous Abbey Road zebra crossing used on the cover of the Beatles’ twelfth album (Fig. 4). This HL cluster appeared in the analysis of total photographs and of those taken by locals, but not in the case of those taken by tourists (although it still presented a high numerical concentration of photographs).

Tourist hot spots presented the greatest extension in London, whereas they were most highly concentrated in Madrid, followed by Athens and Berlin (Table 3). In all the cities analyzed, the extension of hot spots was clearly higher in the distribution of tourists’ photographs that in that of those taken by locals.

A comparison between the location of the spatial clusters and maps of the cities, using Open Street Map, indicated that these results were consistent, since the spatial clusters coincided with the cities’ main tourist attractions and museums. The advantage of the methodology employed in this study is that clusters were identified objectively based on spatial statistical techniques.

These techniques revealed a very dispersed distribution of hot spots in London: the main clusters in the center were located around the Tower of London, Westminster, Buckingham Palace, Piccadilly Circus and the British Museum, whereas in less central areas, they appeared in some green areas (Hyde Park), distinctive shopping areas (Portobello, Camden Town), railway stations (Saint Pancras, Paddington) and football stadiums (Emirates and Stamford Bridge). Other tourist attractions (for example, Wembley Stadium) are located outside the radius analyzed. Local residents’ photographs presented similar cluster patterns to those of the tourists, although clusters in the center showed a much larger extension, i.e. locals photographed many more spaces in the center than tourists.

Berlin also presented a very scattered distribution of tourism clusters. Those in the center corresponded to historical monuments (Museum Island, Brandenburg Gate), shopping areas (Kurfürstendamm) and renovated urban areas (Alexander Platz, Postdamer Platz). More peripheral locations presented clusters at Charlottenburg Palace and the East Side Gallery (an open air art gallery located on a stretch of the Berlin Wall). Peripheral clusters were also observed in local residents’ photographs, corresponding to green areas and Ostkreuz station.

Clusters in the center of Paris were primarily located along the Seine River from Notre Dame to the Eiffel Tower, but also occurred in other nearby areas such as the Champs Elysées, Montparnasse and Montmartre. On the periphery, they appeared mainly around historical monuments (Château de Vincennes), parks (Parc de Villette) and football stadiums (Stade de France). Residents’ photographs were more highly concentrated in the center, forming a very extensive cluster along the Seine.

Athens presented three widely separated areas with a high concentration of tourists’ photographs: a cluster in the center of Athens (Acropolis, Temple of Olympian Zeus, Roman Stadium, National Gardens, Parliament Building, Lykavittos Hill, Omonia Square), another at the port (to the southwest) and a third around the Olympic area (to the northeast). In residents’ photographs, the clusters in the center and at the port covered a greater area, and new areas with a high concentration of photographs also appeared in green areas on the periphery and along the coast.

In Barcelona, tourists’ photographs formed clusters around the historic center, the works of Gaudí (Sagrada Familia, Casa Batlló,
Casa Milà, Park Güell), the beach, the Forum of Cultures site, Nou Camp stadium (Barcelona FC) and green areas with scenic views (Montjuïc, Tibidabo). The clusters formed by residents’ photographs basically coincided with those formed by tourists’ photographs, but extended further in the center and in several city parks.

Fig 3. a: Anselin Local Moran’s I statistic (totals). b: Anselin Local Moran’s I statistic for locals or tourists. c: Anselin Local Moran’s I statistic for locals or tourists.
In the case of Madrid, the clusters formed by tourists’ photographs were very compact and corresponded to the historical center, Real Madrid FC stadium, the Las Ventas bullring and the new towers along the Castellana. Residents’ photographs also revealed hot spots in peripheral areas, particularly in parks and recreation areas (the zoo, amusement parks). The cluster maps indicated that there was a high concentration of residents’ photographs in the Madrid Riverside park, but that this had still not been ‘discovered’ by tourists, even though the park offers an attractive view of the west wing of the Royal Palace as well as access to the Perrault Bridge. The only hot spot for tourists in this area was the Atlético de Madrid stadium.

Rome presented a branched cluster in the historic center, covering emblematic areas such as the Coliseum, the Forum, the Church of Santa Maria Maggiore, Piazza Navona and the Trevi Fountain. The Vatican constituted another central cluster. Peripheral clusters corresponded to the Olympic Stadium and several historic churches. The spatial distribution of residents’ clusters was very similar, but occupied a greater number of hexagons in the historic center.

Lastly, clusters in Rotterdam included distinctive sights such as the Erasmus bridge, the Cube Houses, the Town Hall and the Euromast Tower. The zoo constituted a cluster close to the center. Further away, clusters were formed around the windmills in Schiedam, and Feijenoord Stadium.

5. Conclusions

Geotagged photographs taken by tourists make it possible to identify and analyze the main visual tourist attraction areas in a city and to perform comparisons between cities. Most studies based on geolocated information from social networks have been limited to creating visualization products such as maps of points and animations. The present study represents a step forward since geotagged photographs from Panoramio were used to create density maps and spatial statistics.

Photographs taken by tourists and locals were differentiated according to the duration of the period in which each user was taking pictures. The results show that the distribution patterns observed in tourists’ and residents’ photographs presented both similarities and differences. Tourists’ photographs showed a higher spatial concentration than those taken by locals, as evidenced by a higher coefficient of variation and more compact size of the spatial clusters. The comparison between spatial clusters and maps of the cities (Open Street Map) indicated that these results were consistent, since the spatial clusters coincided with areas in the cities that are known to attract a high concentration of tourists. Consequently, future studies using geotagged photographs from social networks as a proxy for the distribution of tourists in cities may focus on photographs taken by tourists using the segmentation procedure, being not necessary to compare spatial patterns of tourists and residents.

The advantage of geotagged photographs is that they can be used as a proxy for the spatial distribution of tourists, allowing to measure, map and make comparisons within cities and between cities anywhere in the world, providing useful information for tourist researchers and managers. In the present case, they made it possible to identify different patterns in the cities analyzed, reflecting a different spatial distribution of tourist resources. Hence, tourists’ photographs of Athens presented a medium spatial concentration (coefficient of variation) but only formed three widely separated clusters (the center, port and Olympic city). The two peripheral clusters explain why Athens presented the highest standard distance. Barcelona showed the highest spatial concentration of photographs (a large number of photographs in focal points in hexagons) and the lowest standard distance, but it had a higher number of spatial clusters than Athens. London, the city with the lowest spatial concentration of photographs, presented numerous clusters, indicative of a greater dispersion of tourist attractions.

Areas of interest can be defined by studying the density of tourists’ photographs and the operations done with GIS. Currently, many of them are not considered tourist attractions by the government nor by businesses (visitor planning trips editors, product and service creators, hotel professionals, etc.). This way, all of them can know which areas are the most attractive to visitors and what this area’s characteristics are (sports or monumental venues, places with modern architecture, etc.), fundamental aspects in order to market these destinations (Hays et al., 2013).

These city-specific patterns have important implications for the industry and the authorities. Hospitality professionals, store and restaurant owners can identify the areas of highest tourist concentration in order to place their businesses accordingly. Companies responsible for tourism products and services can know where best to install their facilities. They can also contribute to the enhancement and adaptation of certain places by designing, for example, routes which can also be a business opportunity for them. In doing this planners will know points of greatest concentration in order to identify the spaces of saturation. This will enable them to establish carrying capacity controls that take into consideration the

<table>
<thead>
<tr>
<th>City</th>
<th>Tourists</th>
<th>Locals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Athens</td>
<td>123</td>
<td>182</td>
</tr>
<tr>
<td>Barcelona</td>
<td>131</td>
<td>171</td>
</tr>
<tr>
<td>Berlin</td>
<td>124</td>
<td>185</td>
</tr>
<tr>
<td>London</td>
<td>177</td>
<td>236</td>
</tr>
<tr>
<td>Madrid</td>
<td>107</td>
<td>238</td>
</tr>
<tr>
<td>Paris</td>
<td>147</td>
<td>215</td>
</tr>
<tr>
<td>Rome</td>
<td>135</td>
<td>172</td>
</tr>
<tr>
<td>Rotterdam</td>
<td>152</td>
<td>188</td>
</tr>
</tbody>
</table>

Table 3

Surface area (number of hexagons) included in the hot spots (HH clusters).
conservation of sites, security and the happy coexistence of tourists and locals. In addition to this, they can also manage tourism flows by designing alternative routes to help visualize less-visited areas and reduce excessive tourist pressure on others.

These results are also important for tourists because information is immensely helpful for planning the trip and making decisions. Specifically, knowing the dimensions of the city and the degree of dispersion-concentration of the sights beforehand helps to determine the total of time allocated to the visit; select the accommodation according to their proximity to places of interest; or hire tourism services (guides, sightseeing routes or a sightseeing bus) and even purchase tickets for specific transport tourists (city pass, travelcard, etc.).

The data from geotagged photographs posted on social networks can be used to achieve progress in geographic research on tourism, but present some limitations. First, the information is biased in so far as not all tourists make use of these networks, and those who do, do so with varying degrees of intensity. Secondly, it is not allowed to take pictures inside of some buildings, especially museums, so that this data source is more reliable for open spaces than for indoors. Additionally the information refers exclusively to aesthetically attractive places visited by most tourists, and does not fully reflect the attraction of other, less “photogenic” sites. This particularly affects visually undistinguished places of business, study and shopping, which may be under-represented. For shopping tourism, for example, some shopping areas (street markets) do appear in these data sources. One possible means to correct this bias might be to compare geolocated data with credit card data.

Acknowledgments

This paper forms part of the INSIGHT project (Innovative Policy Modeling and Governance Tools for Sustainable Post-Crisis Urban Development), funded under the ICT Theme of the European Union’s Seventh Framework Program. The authors thank Luca Piovano from CEDINT (Polytechnic University of Madrid) for help downloading the data.

References


Anselin, L. (1995). Local indicators of spatial association


Li, X. (2013). Multi-day and multi-stay travel planning using geo-tagged photos. In Proceedings of the second ACM SICeSPATIAL international workshop on crowd-sourced and volunteered geographic information (pp. 1—8). ACM.


Zagheni, E., & Weber, I. (2012). You are where you e-mail: using e-mail data to estimate international migration rates. In Proceedings of web science (pp. 348—358).


