Tourism Management 66 (2018) 13–25

Contents lists available at ScienceDirect

Tourism Management

journal homepage: www.elsevier.com/locate/tourman

Tourists' digital footprint in cities: Comparing Big Data sources

María Henar Salas-Olmedo, Borja Moya-Gómez, Juan Carlos García-Palomares*, Javier Gutiérrez

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

HIGHLIGHTS

- The paper analyses the digital footprint of urban tourists through Big Data.
- Panoramio, Foursquare and Twitter reflect different tourism activities.
- The methods used are density maps, OLS, spatial self-correlation and cluster analysis.
- The results show different tourist spaces: multifunction (several activities) vs specialising.

ARTICLE INFO

Article history:
Received 29 July 2017
Received in revised form
30 October 2017
Accepted 2 November 2017

Keywords:
Urban tourism
Big Data
Photo-sharing services
Social networks
Spatial analysis
GIS

ABSTRACT

There is little knowledge available on the spatial behaviour of urban tourists, and yet tourists generate an enormous quantity of data when they visit cities. These data sources can be used to track their presence through their activities. The aim of this paper is to analyse the digital footprint of urban tourists through Big Data. Unlike other papers that use a single data source, this article examines three sources of data to reflect different tourism activities in cities: Panoramio (sightseeing), Foursquare (consumption), and Twitter (being connected-accommodation). The results show that the data from the three activities are partly spatially redundant and partly complementary, and allow the characterisation of multifunction tourist spaces and spaces specialising in one or various activities. The main conclusion is that it is not sufficient to use one data source to analyse the presence of tourists in cities; several must be used in a complementary manner.

1. Introduction

The tourism product of large cities has enormous capacity and is highly diversified (Jansen-Verbeke, 1986). Tourists make very selective use of the city. They reduce uncertainty in their exploration of an area by visiting sites perceived to give the greatest reward for effort (Cooper, 1981). In fact, it is impossible for tourists to consume the entire urban tourism product on an average 2–3-day visit to such a city (Mazanec, 1997), so they must choose which of the attractions they wish to visit, and which to skip. The result is the creation of typical tourism-product consumption patterns based on the preferences and limitations of different tourist types (Shoval & Raveh, 2004). Studies analysing spatial patterns of tourist mobility in cities show that they tend to be concentrated in specific areas of city centers (Hayllar & Griffin, 2009; Shoval & Raveh, 2004), where they find the main tourist attractions (historical buildings and parks, museums, theatres, concert halls, etc.), along with leisure and shopping facilities and accommodation services for tourists (Pearce, 1987). Not surprisingly, most tourists seek hotels or apartments that are within walking distance of major attractions in the city (Arbel & Pizam, 1977) and spend a large share of their time budget in the immediate vicinity of the hotel (Shoval, McKercher, Ng, & Birenboim, 2011). As a result, city centers are profoundly transformed by the pressure of tourism. They become more and more oriented towards meeting the needs of tourists (hotels and apartments, souvenir shops, restaurants, etc.), and less toward residents' needs, who tend to abandon central locations. These processes are known as tourismification (Jansen-Verbeke, 1998) and tourism gentrification (Gotham, 2005).

Traditionally the spatial behaviour of tourists in cities has been studied through surveys (for example, Cooper, 1981). However,

* Corresponding author.
E-mail addresses: msalas01@ucm.es (M.H. Salas-Olmedo), bmoyagomez@ucm.es (B. Moya-Gómez), jcgarcia@ucm.es (J.C. García-Palomares), javiergutierrez@ghis.ucm.es (J. Gutiérrez).

https://doi.org/10.1016/j.tourman.2017.11.001
0261-5177/

surveys do not provide great spatial and temporal data on the spatial behaviour of tourists. Over the past nine years, the rapid advancement and availability of small, inexpensive and reliable tracking devices that draw on GPS technology is assisting researchers in developing new methods of spatial research (Edwards & Griffin, 2013). GPS allows for the precise and continuous tracking of individuals and provides spatially rich data, making it possible to accurately track the paths tourists take and to provide greater understanding of their socio-spatial behaviour (Asakura & Iryo, 2007). Not surprisingly, several studies have used GPS tracks in recent years to analyse the spatial behaviour of tourists (for example, Shoval et al., 2011). However, most of these studies use small samples, since the collaboration of tourists is necessary to obtain their tracks.

Big Data offers new opportunities in tourism research by providing high spatial and temporal data that make it possible to analyse the spatiotemporal patterns of a large numbers of tourists. Big Data is a new concept that has become widely popularised in recent years to describe the production of massive quantities of data. Big Data covers a range of very different areas: Internet searches, bank card transactions, records of mobile phone activity, social networks, data on water and electricity consumption, meteorological data, images recorded with video cameras and many more. The main characteristics of these new data sources include particularly the following three Vs: *volume*, in terabytes or petabytes of data; *velocity*, created in or at near real time; and *variety*, taken from a wide variety of sources, either structured (data that can be stored in the form of tables), semistructured (htlm files) or unstructured (texts, photographs, videos) (McAfee, Brynjolfsson, Davenport, Patil, & Barton, 2012; Kitchin, 2013; Sagiroglu & Sinanc, 2013).

Big Data supplies a large quantity of information to complement the traditional sources. Tourists leave a digital "footprint" in most of their activities, and these new data sources now make it possible to analyse tourists' behaviour in the cities they visit. They take vast numbers of photographs and upload them to photo-sharing services, they make payments with bank cards, they talk and send messages via their mobile phones, they are active on social networks, and so on. All this activity produces an enormous quantity of digital data (Big Data) which can be analysed to study behaviour patterns (Shoval & Isaacson, 2007; Asakura & Iryo, 2007; Girardin, Fiore, Ratti, & Blat, 2008b, 2008a;). Much of these data is geolocated, so tourists' activity can be analysed spatially. However, there are very few papers that apply Big Data to examine the spatial distribution of tourists in cities, probably due to the novelty of these information sources and the fact that some are difficult to access.

Photo-sharing services provide very useful information for identifying the presence of tourists when they go sightseeing in cities. Although there are several photo-sharing communities that allow the geolocation of photos (such as Flickr or Instagram), Panoramio is probably the most useful service for measuring tourist hotspots, as this website shows photographs taken of places or landscapes when sightseeing, which are then posted online once they have been georeferenced. The records of geolocated photographs can be used not only to identify sightseeing spots (García-Palomares, Gutiérrez, & Mínguez, 2015), but also to analyse the spatial and temporal patterns of tourist flows in cities (Girardin et al., 2008b).

However, tourists not only visit the most photographed spaces. They also go shopping, go to restaurants and stay in hotels, and they leave their digital footprint in all these establishments when they pay with a bank card or check-in their location on social networks. These digital footprints of tourists offer information which is largely complementary to the data from photo-sharing social networks. The most photographed areas often have very little offer of accommodation and shopping. In the case of business tourists, their hotel and the spaces they frequent tend to be close to business sectors, and not necessarily in the most photographed areas. During their stay in the city, tourists also log onto the Internet to confirm details of their visit, check their e-mail, engage in the social networks, and so on. This activity also leaves a digital footprint in many of the places they visit. Tourists often use the facilities in hotels, hostels, restaurants and certain open spaces to connect to Internet through free WiFi networks, so their activity on social networks may particularly reflect this type of spaces.

The main aim of this paper is to compare three geolocated data sources to identify the presence of tourists in cities in terms of their different activities: geolocated photographs from the Panoramio platform area used as a proxy for sightseeing, Foursquare check-ins as a proxy for consumption, and interaction on the social network Twitter as a proxy for being connected-accommodation. The study area is the city of Madrid, one of the European cities with the highest volume of tourists.

This paper contributes to the literature on Big Data and tourism activity from a threefold perspective: 1) Three different data sources are compared to obtain the most comprehensive analysis of locations of tourists. 2) The data for tourist activity (photos, check-ins, tweets) are not analysed directly as in previous works, but the tourists themselves are the unit of analysis. The data are processed to allow the number of tourists to be counted in each place in the city according to each data source, thus avoiding problems of multiple counting, and making the results comparable. 3) The information from the different data sources is integrated through cluster analysis and spatial autocorrelation analysis to characterise the areas of tourist concentration according to the type of activity.

The remainder of this paper is structured as follows. Section 2 summarises the existing literature on the use of photo-sharing services, Foursquare check-ins and Twitter in urban studies, with a particular focus on tourism. Sections 3 and 4 describe the data and the methodology respectively. Section 5 describes and discusses the results, while Section 6 presents the main conclusions and suggests further directions for research.

2. Related literature

2.1. Photo-sharing services

Sightseeing is one of the main tourist activities in cities, and leaves its digital footprint on social networks for photo-sharing such as Instagram, Flickr and Panoramio. All three offer the possibility of geolocating photographs, but Panoramio (http://www.panoramio.com) particularly allowed the georeferencing of the photos, as it focused on images of places or landscapes shared by its users, which can be seen on the Panoramio website (until November 4, 2016) or through Google Earth and Google Maps. In fact, Panoramio was a Google service, and had over 120 million geolocated photographs (2015 data according to Panoramio data API).

Photo-sharing services have been used for several purposes in the field of tourism, including identifying social events such as festivals, demonstrations, sporting events and so on (Sun & Fan, 2014), estimating tourist numbers (Koerbitz, Önder, & Hubmann-Haidvogel, 2013), identifying the presence of tourists (Girardin, Calabrese, Fiore, Ratti, & Blat, 2008a; Gutiérrez, García-Palomares, Romanillos, & Salas-Olmedo, 2017; Kisilevich, Keim, Andrienko, &