Computers, Environment and Urban Systems 74 (2019) 161–174

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

Social Media data: Challenges, opportunities and limitations in urban studies

Pablo Martí*, Leticia Serrano-Estrada, Almudena Nolasco-Cirugeda

University of Alicante, Building Sciences and Urbanism Department, Carretera San Vicente del Raspeig s/n. 03690, San Vicente del Raspeig. Alicante, Spain

ARTICLE INFO

Keywords:
Urban planning
Data analysis
Location based social networks
Urban analysis

ABSTRACT

Analysing the city through data retrieved from Location Based Social Networks (LBSNs) has received considerable attention as a promising method for applied research. However, the use of these data is not without its challenges and has given rise to a stream of polemical arguments over the validity of this source of information. This paper addresses the challenges and opportunities as well as some of the limitations and biases associated with the collection and use of LBSN data from Foursquare, Twitter, Google Places, Instagram and Airbnb in the context of urban phenomena research. The most recent research that uses LBSN data to understand city dynamics is presented. A method is proposed for LBSN data retrieval, selection, classification and analysis. In addition, key thematic research lines are identified given the data variables offered by these LBSNs. A comprehensive and descriptive framework for the study of urban phenomena through LBSN data is the main contribution of this study.

1. Introduction

In an era of ever increasing user-generated content —via new data sources using fixed or mobile sensors such as GPS, credit cards, smartphones, etc.— patterns of human activity are revealed. Thus, obtaining meaningful information from these sources represents both a challenge and an opportunity. Specifically, digital data sources provide researchers with a new approach to the study of urban phenomena.

Social media is accepted by scholars as a valuable resource to advance research on specific urban aspects (Anselin & Williams, 2015; Arribas-Bel, Kourtit, Nijkamp, & Steenbruggen, 2015; Roick & Heuser, 2013; Shelton, Poorthuis, & Zook, 2015). Practitioners argue that social media offers different visions on diverse aspects of social, economic and political urban life reflected by the user's interests and activities (Bawa-Cavia, 2011; Cerrone, 2015; Graham, Hale, & Gaffney, 2014; Huang & Wong, 2015). In fact, the representation and interpretation of data retrieved from Location Based Social Networks —LBSNs hereafter— provide a means by which to assess different urban dynamics, such as, mobility (Cheng, Caverlee, Lee, & Sui, 2011; Luo, Cao, Mulligan, & Li, 2016; Noulas, Scellato, Lambiotte, Pontil, & Mascolo, 2012; Quercia, Aiello, Schifanella, & Davies, 2015a, 2015b); land uses and urban activity (García-Palomares, Salas-Olmedo, Moya-Gómez, Condeço-Melhorado, & Gutiérrez, 2017; Hamstead et al., 2018; Quercia & Saez, 2014; Van Canneyt, Schockaert, Van Laere, & Dhoedt, 2012a, 2012b); human behaviour (Hochman & Manovich, 2013; Lee, Wakamiya, & Sumiya, 2013; Peña-López, Congosto, & Aragón, 2014; Quercia, Aiello, Schifanella, & Davies, 2015a, 2015b); event detection (Béjar et al., 2016; Chen & Roy, 2009); and, the issue of adding value to urban planning, decision making processes and city design (Dunkel, 2015; Tasse & Hong, 2014).

This research presents an innovative, comprehensive and descriptive method which has been developed for retrieving, processing and interpreting LBSN geolocated data for the study of cities. Furthermore, inferences that have been drawn from previous research are provided to illustrate how this method can be applied to different urban contexts. The novelty of this research lies in the proposed strategy for addressing the opportunities, limitations and difficulties associated with the process of retrieving, validating, classifying and filtering LBSN datasets for the study of specific urban phenomena. Five LBSNs —Twitter, Foursquare, Google Places, Instagram and Airbnb— are considered for their unique characteristics and varied metadata to exemplify these processes.

The paper is structured as follows: i) a recap of the existing literature on the main opportunities and limitations found in the use of LBSN data for urban analysis; ii) an explanation of the proposed comprehensive data retrieval, selection, filtering and usage method that overcomes many of the most important drawbacks; and, iii) a discussion of the opportunities and, also, the limitations and difficulties of

* Corresponding author.
E-mail addresses: pablo.marti@ua.es (P. Martí), leticia.serrano@ua.es (L. Serrano-Estrada), almudena.nolasco@ua.es (A. Nolasco-Cirugeda).

https://doi.org/10.1016/j.compenvurbsys.2018.11.001
Received 29 May 2018; Received in revised form 23 September 2018; Accepted 2 November 2018
Available online 15 November 2018

using LBSN data for the analysis of cities.

1.1. LBSN's in urban analysis: the opportunity

Some inherent features of LBSN data render them a valuable resource for developing urban studies.

Firstly, LSBN data are generated by millions of people from different countries throughout the world (Hu et al., 2015) and, as the number of social network users grows, so does the amount, quality and usability of data. In 2018, the number of active social media users worldwide reached 3.196 billion and the number of active mobile social media users was 2.958 billion (Kemp, 2018).

Secondly, automatic retrieval of social media user-generated content represents a technological advance for urban analysis. This is mainly due to the ease with which data collection can be done, removing many of the constraints associated with traditional methods such as, collection time, accurate geolocation marks, etc. Traditionally, large surveys and long periods of observation were required to collect an adequate amount of data for research. Relevant work developed thus far demonstrates the usefulness of these data for urban analysis. For instance, some of the most significant research that originally used traditional methods in urban studies: *The Image of the city of Boston* (Lynch, 1960) and the *Death and Life of Great American Cities* (Jacobs, 1961), have been revisited using LBSN data (Al-Ghamdi & Al-Harigi, 2015; Lee et al., 2013; Liu, Zhou, Zhao, & Ryan, 2016; Quercia, Aiello, Schifanella, & Davies, 2015a, 2015b). These studies concur that although closer scrutiny of the data is necessary for more effective data filtering and mining, crowdsourcing technologies —including social networks— provide great opportunities for researchers and designers involved in the analysis of urban environments (Granell & Ostermann, 2016).

Thirdly, the information contained in LBSN data enables the exploration of intangible aspects of urban life that are linked to places (McLain et al., 2013). Some social exchanges and events happening in the city remain concealed (Soja, 1989) in morphological or physical studies (Cerrone, 2015). However, they leave a virtual trail linked to a specific location, which provides a more thorough analysis of users experiences and perceptions of the city (Saker & Evans, 2016; Silva, Vaz de Melo, Almeida, Salles, & Loureiro, 2014).

Fourthly, LBSN data can be recognised as volunteered geographic information —VGI— (Campagna, 2016; Jiang, Alves, Rodrigues, Ferreira, & Pereira, 2015; Kitchin, 2013) since the expressed perceptions, interests, needs and behaviours are published online voluntarily by the users and refer to unique and specific places in cities. Data are generally collected "unobtrusively" (Quercia, Aiello, Schifanella, & Davies, 2015a, 2015b) and users are generally not constrained when generating information. This is an advantage because according to the Hawthorne effect, subjects may alter their behaviour in a study on realization that they are being observed (McCarney, Warner, Iliffe, van Haselen, R Griffin, & Fisher, 2007).

Lastly, the diversity of LBSNs, and the content retrieved from them, offer a multi-perspective approach to the study of cities. There is considerable research using data from Facebook, Twitter and Instagram, some of the most globally-renowned LBSNs, that covers different topics in relation to diverse fields of knowledge. However, other LBSNs, such as Foursquare and Google Places have demonstrated their relevance as supplementary georeferenced data sources (Jiang et al., 2015; Milne, Thomas, & Paris, 2012; Serrano-Estrada, Marti, & Nolasco-Cirugeda, 2016; Van Canneyt, Schockaert, et al., 2012a). Moreover, different LBSNs, with the same functionality as the renowned global ones, are more commonly used in specific geographical areas. For instance, Weibo —China— or Mastodon —India— are an alternative to Twitter. In both South Korea and Japan, Never, is an alternative to Google. Thus, methods developed for collecting and analysing LBSN data for research purposes can be transferrable to other LBSNs.

As evidenced by the previously cited works, social media user-generated data are a valuable by-product for the study of the city (Arribas-Bel, 2014), and when the information is georeferenced, that provides added value for urban research since specific phenomena can be analysed in a determined urban area. That is the reason why this study adopts exclusively geolocated data from Social Networks.

1.2. Challenges and limitations of using LBSN data for urban research

Some of the most commonly cited limitations associated with the use of LBSNs refer to the lack of consistency in the provision of an acceptable amount of valid geocoded data for each sample (Boyd & Crawford, 2012; Cerrone, 2015; Leetaru, Wang, Cao, Padmanabhan, & Shook, 2013; Sloan & Quan-Haase, 2017). For instance, a study conducted in the metropolitan area of Pittsburgh indicated a greatly reduced amount of LBSN data generated from urban areas with lower median income compared to the rest of the city, probably due to lower smartphone ownership (Tasse & Hong, 2014). Therefore, the amount of data are largely conditioned by ownership of a smartphone and access to an internet connection (Arribas-Bel, 2014). Also, there is a difference in terms of the quantity of information retrieved from LBSNs between rural and urban areas (Hecht & Stephens, 2014); and, in the specific case of Twitter, the fact that only a small portion of its users activate the geocoded function when publishing tweets is also an important consideration (Sloan, 2017). Furthermore, the reasons for using the geocoding function in Twitter messages are certainly biased by factors such as social-economic status, political context or education (Graham et al., 2014). Certain social networks are more popular in some places than in others, impacting the quantity of information available from a specific social network (Sloan & Quan-Haase, 2017). That is why research is usually applied to case studies involving large metropolitan cities with a high population density given that there is a considerably greater amount of LBSN data available for study.

Even if the dataset is acceptable in terms of quantity, lack of transferability and representativeness in the information provided has been acknowledged as a problem in the two following circumstances.

Firstly, LBSN data retrieved about specific locations reveal important details about the everyday urban life in those places (Lee et al., 2013; Sui & Goodchild, 2011). Thus, research on single case studies is limited to a specific place and it is difficult to know with certainty if the conclusions obtained from the selected sample are transferrable to other locations (Goodchild, 2013).

Secondly, there are contrasting opinions about whether LBSNs represent the entire population. Some studies argue that LBSN data provide a representative sample of citizen preferences, opinions and activities (Agryzkov et al., 2015; Barbera & Rivero, 2015; Martí, Serrano-Estrada, & Nolasco-Cirugeda, 2017; Morstatter, Pfeffer, Liu, & Carley, 2013; Tufekci, 2014), given the increasing diversity of user profiles (Pew Research Center, 2017). Others claim that LBSN users are not necessarily a representative sample (Quercia, Aiello, Schifanella, & Davies, 2015a, 2015b) based on the assumption that social media users comprise only part of the population whose use of a particular social network tends to be aligned to a specific interest. However, since no personal details are retrieved when collecting user data, the sample cannot be rigorously characterised in terms of user profiles as is possible in a controlled environment —interviews, focus groups, etc.— (Chorley, Whitaker, & Allen, 2015). Evidently, some users of social networks are not private individuals but represent organisations, institutions, businesses, public figures, and influencers whose tailored comments reach and, potentially influence huge audiences. This case implies that the data generated is driven by public relations and external communications executives whose role is to comply with the organization's communications strategy, for example (Cerrone, 2015; Marwick & Boyd, 2011).

The aforementioned concerns are acknowledged in studies that use LBSNs for addressing city dynamics. However, the challenges and limitations associated with the process of data retrieval, verification, selection and filtering have received scant coverage in the literature. The accuracy of these methods is crucial for obtaining valid datasets and dealing with different research problems concerned with the field of urban studies and this paper seeks to bridge this gap.

2. Method for retrieving and using LBSN data in the study of cities

This section presents a comprehensive method for obtaining, verifying, filtering and classifying data from five LBSNs: Foursquare, Google Places, Twitter, Instagram and Airbnb. Additionally, some inferences are included from previous analysis of urban phenomena using these data.

2.1. Data retrieval process and tools

There are various methods for retrieving LBSNs data: via Application Programming Interface —API— (Jagadeesan & Venkatesan, 2015; Leetaru et al., 2013; Tsou et al., 2013; Wang, 2013; Wilken, 2014; S. Williams, 2012); via crawled from the website (Mahto & Singh, 2016); and via purchased by official resellers; among others (Mayr & Weller, 2017). Specifically, this study takes the case of a web-based application that retrieves data from Foursquare, Google Places, Twitter and Instagram: SMUA —Social Media Urban Analyser—. As for Airbnb data, it is obtained through AirDNA, a third party company that "gathers information publicly available on the Airbnb website" (AirDNA, 2017).

SMUA's functionality and interface —Fig. 1— has been specifically designed to collect geolocated social network data —Table 1—. Launched in 2013, the first version retrieved data from the social networks Foursquare and Panoramio. However, the latter has been removed after its closure date on November 4, 2016. Currently, SMUA retrieves data from Foursquare, Google Places, Twitter and Instagram.

Some aspects involved in the experience with SMUA's retrieval procedure that are worth highlighting are as follows: first, the limitations and requirements imposed by each social network regarding the shape and size of the search area; and, second, the maximum number of records provided by the API for each data request. These conditions, met by SMUA data requests, are commonly found among other LBSNs, especially those whose data are harvested through an API. Furthermore, although LBSNs frequently change the requirements in terms of the number of results per request, the retrieval process remains the same and, thus could be transferrable to collect data from other similar social networks. Therefore, the overarching principles of the LBSNs data retrieval process through API could be narrowed down to:

1. Request type
2. Search polygon shape
3. Search polygon size
4. Number of requests and/or results allowed per request
5. Timeframe up to data retrieval
6. Retrieved data

These principles, listed in Table 1, have been adopted by SMUA for each LBSN and will be further explained. It can be observed that Twitter data can be obtained through two complementary methods: Streaming and Rest.

The overall procedure for requesting and retrieving data from the APIs —Foursquare, Google Places, Twitter and Instagram— is as follows: firstly, a search polygon area is delineated —of regular or irregular shape— in the Open Street Map cartography (Liftn & Parad, 2018) within SMUA's interface; secondly, SMUA delineates a Superimposed Regular Shape —SRS—, rectangular or circular, onto the search polygon area according to the shape and size restrictions imposed by the social network's API; and thirdly, the data request is processed. The data retrieval time will vary according to the size of the SRS. All the information in the API is retrieved; however, as explained in Section 2.2, a selection, validation and filtering of data is performed before conducting analysis and drawing any conclusion.

2.1.1. Foursquare and Google Places

The requirements to define the search area are similar for both Foursquare and Google Places. Foursquare web service requires the search area to be a rectangular polygon whose sides cannot exceed the length of 100 km —Fig. 2— and the maximum number of results provided per request is 50 records —*venues*—. Google Places requires the search area to be a circle, whose radius cannot exceed 5 km in length —Fig. 3— and the maximum results provided per request is 60 records —*places*—.

In commercially active areas, for example, the original search polygon is very likely to contain more than 50 *venues* or 60 *places*. Therefore, a search algorithm has been incorporated into SMUA to

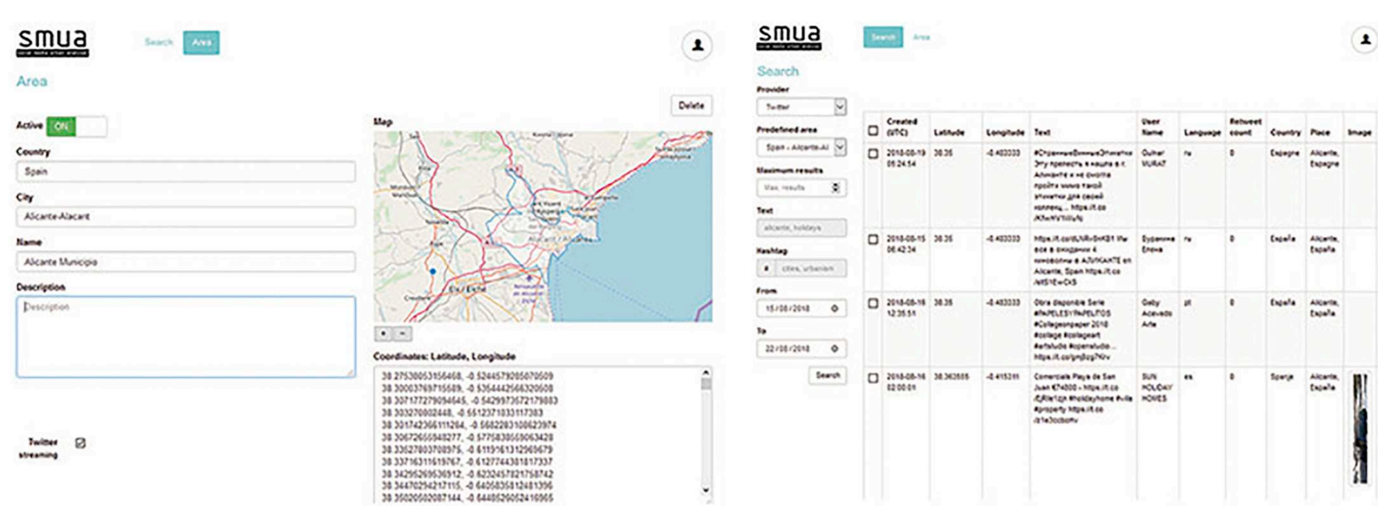

Fig. 1. SMUA's user interface. The left image shows the set up and definition of a search area and the right image shows the search results display page.

Table 1
Summary of the social networks' API requirements incorporated into SMUA for the data request process.

	Foursquare	Twitter		Google Places	Instagram
1. Request type	Rest	Streaming	Rest	Rest	Rest
2. Search polygon shape	Rectangular	Rectangular	Circular	Circular	Circular
3. Search polygon size	The sides cannot exceed 100 km	No limitation	No limitation	Radius cannot exceed 5 km	5 km
4. Number of requests and/or results allowed per request	50 results	450 requests per each 15-min window. No limitation on the number of results.	Max. 1% of all world-wide generated tweets.	60 results	5000 calls per hour. No limitation on the number of results.
5. Timeframe up to data retrieval	*Venues*' cumulative and updated data	Real time tweets	Recently shared tweets —approx. Up to seven days prior to the retrieval date—	*Places*' updated data	Pictures' updated data
6. Retrieved data	Spreadsheet with all *venues* registered within the search area	Spreadsheet with a representative sample of tweets collected within the geographic filter while Streaming is activated	Spreadsheet with a listing of tweets is obtained with no guarantee that all tweets within the geographic filter will be retrieved.	Spreadsheet with all the *places* registered within the search area	photographs tagged within the area are retrieved in individual jpg files.

guarantee that all data available from the source is retrieved. The algorithm known as the quadtree decomposition method (Samet, 1984), which is similar to divide-and-conquer methods (Aho, Hopcroft, & Ullman, 1974, p. 60), recursively divides the SRS into four quadrants and, if necessary, the partial quadrants are again subdivided into four sub-quadrants until the following two conditions are satisfied: the shape sides or circle's radius do not exceed the size limitation set by Foursquare and Google Places and, concurrently, the number of registers obtained is less than 50 *venues* for Foursquare or 60 *places* for Google Places —see Fig. 2 and Fig. 3 respectively.

The resulting dataset includes the cumulative list of registers in Foursquare *venues* and a list of registered establishments on Google Places up to the retrieval date.

Table 2 provides details of the number of datapoints retrieved from Foursquare and Google Places using SMUA from four cities of the Mediterranean Spanish Arc. The number of datapoints is compared to the measured area of the continuous urban fabric.

2.1.2. Twitter

Geolocated and non-geolocated tweets can be collected. Twitter spatiotemporal analyses are conditioned by the amount of data available taking into account that only part of the tweet traffic is geocoded (Sloan & Morgan, 2015). This is because a tweet geocode can only be generated from GPS-enabled devices (Han, Cook, & Baldwin, 2014), and, even though users have full control of whether their tweets are geolocated or not, the geolocation option in the Twitter app is off by default.

There are two ways to include the tweet location: enabling the precise location of the browser or devise from which the twitter is broadcast, or select a location label suggested by default by the Twitter app. In some locations, the latter option includes Twitter places labels of specific landmarks, businesses or points of interest that are sourced from Foursquare (Twitter, 2018a). The presented methodology focuses on collecting and analysing geolocated tweets retrieved in both ways: with exact coordinates and those whose location is defined through Twitter places.

As for retrieving data, there are three different ways to access Twitter API (González-Bailón, Wang, Rivero, Borge-Holthoefer, & Moreno, 2014): Twitter's Streaming API; Twitter's Search API —Rest API—; and Twitter's Firehose. SMUA accesses free and open Twitter data using the first two.

The Streaming API is based on real-time data collection. Previous research has demonstrated that the data search method via Streaming HTTP protocol using a geographic boundary box as a filter returns a very representative sample of tweets (Morstatter et al., 2013).

Twitter's Streaming API requires a rectangular area of any size defined by two pairs of latitude and longitude coordinates. SMUA's algorithm superimposes a SRS and "listens" to the tweets shared within the defined area. This search method provides a sample of user-geolocated tweets that are occurring real-time within the boundary box. The

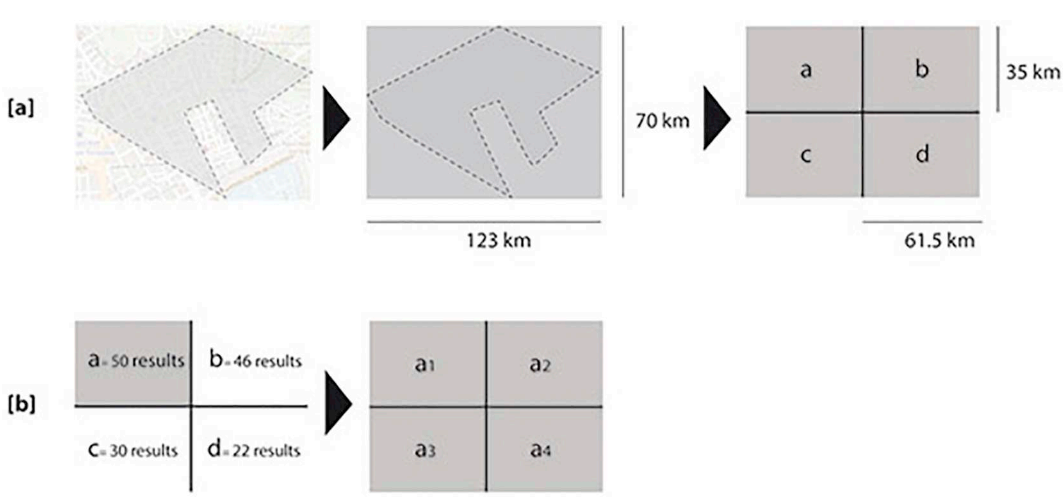

Fig. 2. Foursquare. Sub-quadrants derived from the search polygon in compliance with the Foursquare API requirements on size and maximum number of *venues* retrieved.

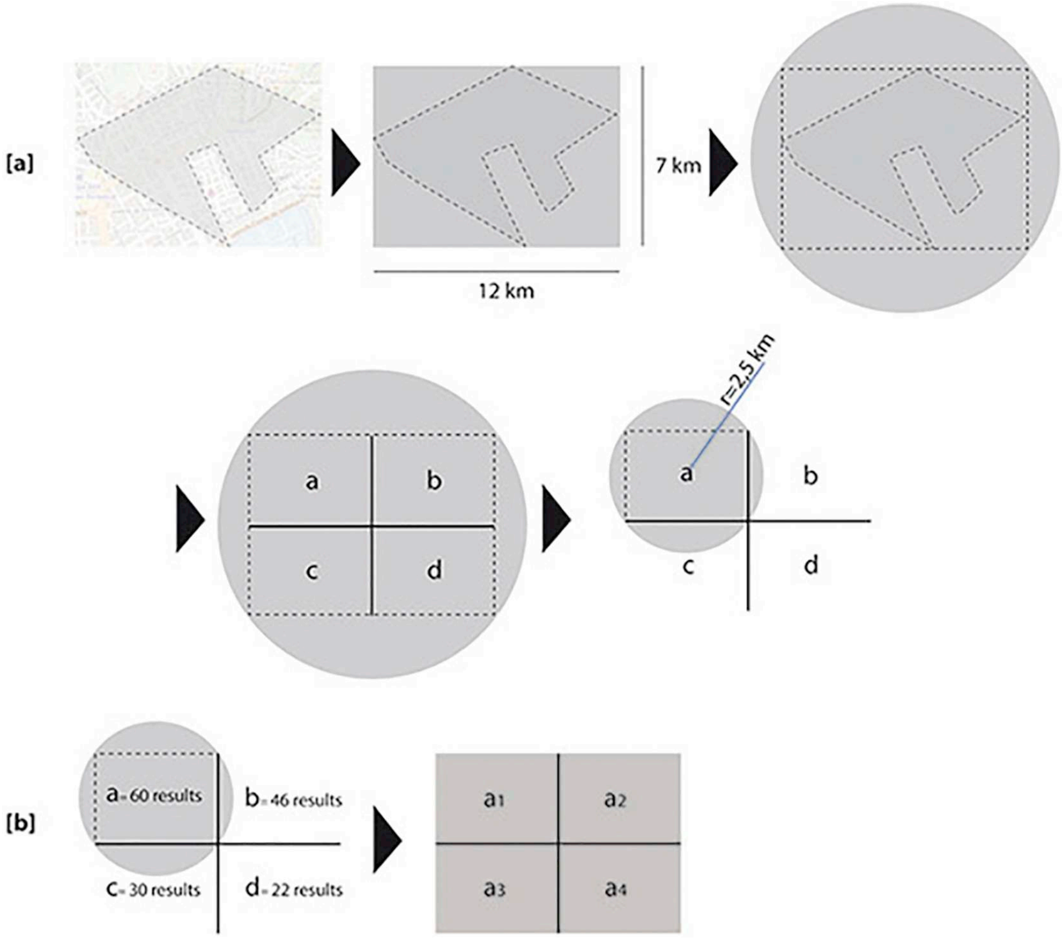

Fig. 3. Google Places. Sub-quadrants derived from the search polygon in compliance with the Google Places API requirements on circle size and maximum number of *places* retrieved.

Table 2
Number of datapoints retrieved from Google Places and Foursquare in relation to the measured area of the continuous urban fabric for four Spanish Mediterranean Arc cities.

	Population within the continuous urban fabric area (INE, 2011)	Area (km^2) (Instituto Geográfico Nacional, 2018)	Google Places Datapoints (data retrieved: 16 Feb 2018)	Foursquare Datapoints (data retrieved: 16 Feb 2018)
VALENCIA	782,657	46,68	70,214	15,262
ALICANTE	309,651	35,1	30,758	6417
ELCHE	188,951	19,95	14,880	2633
CASTELLON	153,295	11,94	14,179	2492

sample includes those tweets that were shared by the user with a precise location and those that were tagged with a specific Twitter place label. The data collection rate of the Streaming API is limited to 1% of all world-wide generated tweets (Boyd & Crawford, 2012). Therefore, it is possible to retrieve all the tweets within a specific area as long as the total quantity of tweets requested by the filter —geographic boundary box— does not exceed this limitation.

The second method of retrieval, the Rest API, works on requests and requires the delineation of a circular area with neither a size nor a limit set for the quantity of results. For the case of SMUA, the limit on the number of requests is 450 requests per each 15 min window (Twitter, Inc., 2018b). Despite this data collection method often being used by researchers (Roberts, 2017; Villatoro, Serna, Rodríguez, & Torrent-Moreno, 2013), Twitter does not guarantee that the Rest API method will list all the tweets shared within the search area. In fact, the final dataset per Rest search in Twitter will include a list of tweets that have been shared in the last seven days approximately.

In both methods, retweets generated by the retweet command on the Twitter app are not considered original content, and therefore, are not geolocated. However, copy-pasted tweets generated as new tweets are considered original content and thus, the user can geolocate them (Sloan & Morgan, 2015).

A visual comparison of tweets collected using both methods in the case of Central Park area, New York, —Fig. 4— shows that when requiring the maximum amount of results, the Streaming API method is preferable, but in terms of obtaining a tweet location pattern over a short period of time —one week, for example—, the Rest method provides a random but representative sample. Furthermore, the Rest method is rather useful in cases where the Streaming API search is not available due to technical reasons, such as when the internet connection is interrupted. That said, the combination of both methods would allow a completer and more accurate dataset.

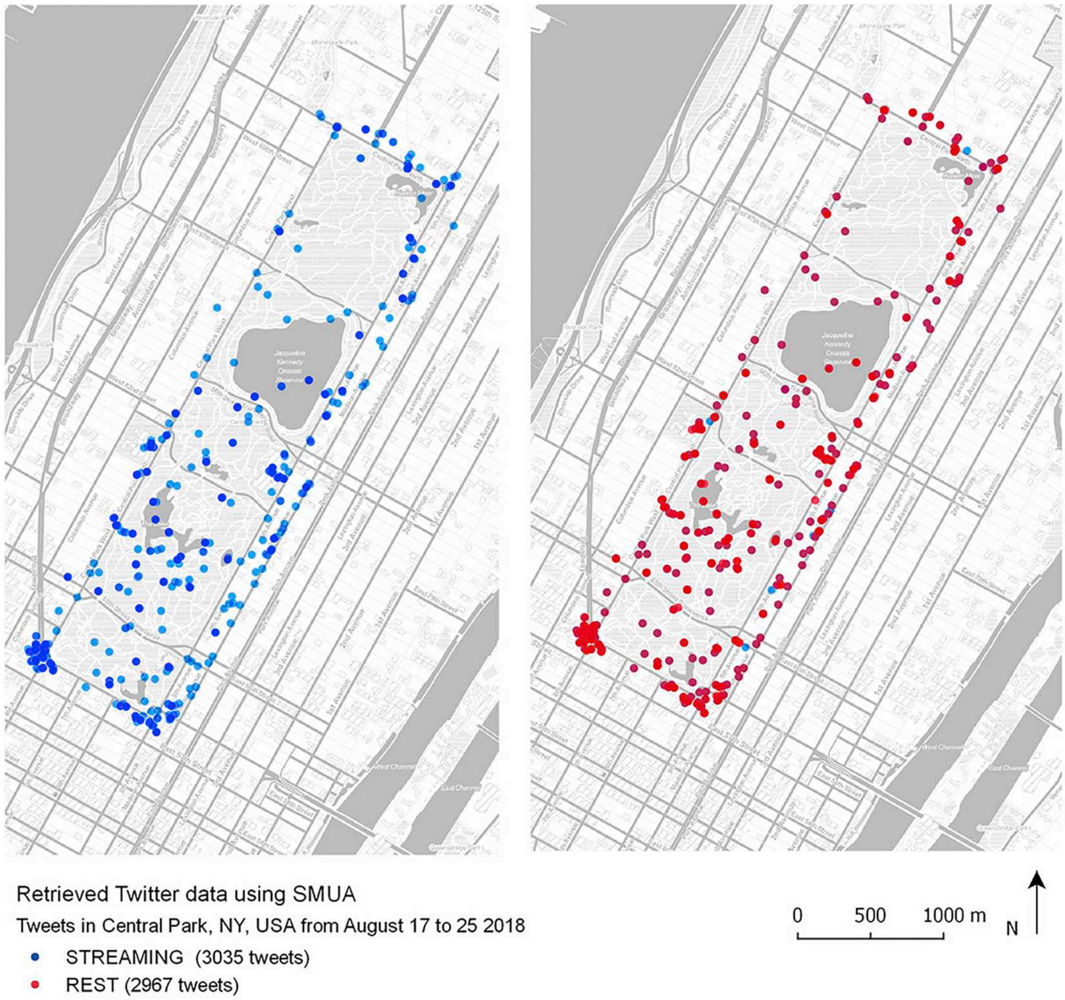

Fig. 4. Comparison between Twitter datasets obtained via Streaming and Rest API methods.

2.1.3. *Instagram*

The Instagram data retrieval is conducted by manual and automated means. The manual method implies downloading photos directly from Instagram webpage using third-party download plugins, and the automated download is performed through SMUA. According to previous research (Boy & Uitermark, 2016; López Baeza, Serrano Estrada, & Nolasco-Cirugeda, 2016), manual collection is known to have advantages in terms of the closer analysis of data, especially in qualitative research where granularity of detail is crucial since data can provide valuable insights that would not be obtained otherwise from large datasets (Laestadius, 2017). This is because each post is searched and extracted from Instagram's web service, and more sense can be made of the pictures in the context of the user's profile page than by using automated data extraction.

SMUA's automated data retrieval process for Instagram's API search method consists of a circular shape with a required maximum 5 km radius. The search area is then covered first by a rectangular SRS and then a circular shape and, if the radius exceeds the allowed distance, the SRS is subdivided into four sub-quadrants until the circular size complies with the Instagram API requirements. There is no limitation in terms of the quantity of registers delivered by the Instagram API. However, there is a limit of calls per hour which used to be 5000 but has recently changed to 200 —as of April 2018—.

There are two important differences between Instagram and the previously explained three social networks Foursquare, Google Places and Twitter. Firstly, data retrieved from Instagram —pictures and their metadata— are not georeferenced to the exact location from where they were posted. Instead, Instagram has delimited areas with a geolocated centre point to which all data within the area will be associated —Fig. 5— For example, all the pictures shared on Instagram in a certain urban area —namely, downtown area— may be geolocated to the city's cathedral. Secondly, as of June 2016, Instagram has placed important restrictions on its API access, one of which is limiting the quantity of data accessed. Any app or program that intends to retrieve data from Instagram's API requires system approval first. Otherwise, only a "sandbox" version of the data is available which provides only a very limited amount of data for retrieval.

2.2. *Data variables and usage*

Two considerations have an important impact on the analysis of data retrieved from the social network APIs. Firstly, the LBSNs user-generated information differs significantly from one social network to the other since they have been designed for different purposes. For instance, users can broadcast their presence by checking-in on Foursquare *venues*; register and rate businesses in Google Places, share a tweet in Twitter, upload a photograph to Instagram and/or comment on users' images, or register a short-term rental property on Airbnb. Thus, each social network provides unique data variables —metadata—. Depending on each research topic, one or several variables from different sources can be considered allowing a more comprehensive assessment process —Fig. 6—. Although data from different LBSNs are not comparable, the resulting analysis from each can be complementary for research purposes.

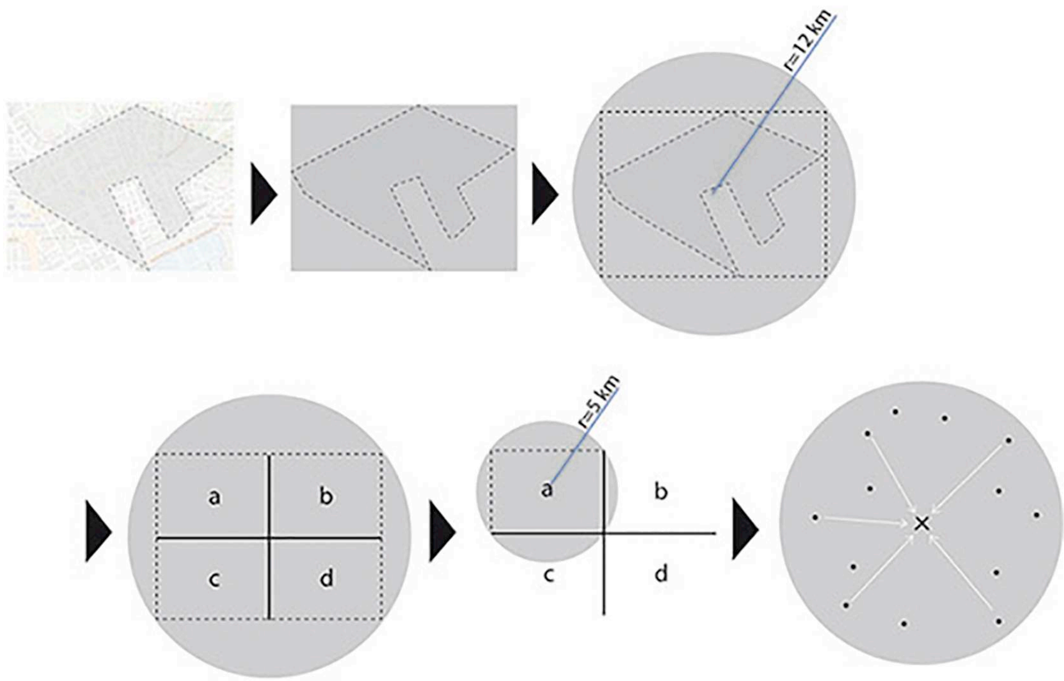

Fig. 5. Instagram API search method.

Fig. 6. Comprehensive assessment process for the interpretation of data from different LBSNs.

Secondly, and as a consequence of the previous consideration, reflection on the research topic is required prior to selecting suitable data variables for analysis because not all metadata information offered by the social network API may be useful for the study of a specific urban phenomena. Specifically, SMUA has been programmed to retrieve only specific data variables relevant to the urban phenomena being addressed. These variables can be grouped into 5 categories, as shown in Table 3: location [LOC], temporal information [TEMP], user generated data [UGDAT], data categorization [CAT] and data ID [ID].

The collected data variables have different formats depending on each LBSN: geographic coordinates [coor]; text [txt] —tweets, tips, comments, hashtags, reviews, photo name or description—; rating values —check-ins, visitors, rating value— [rat]; photographs [pho]; *place*, *venue* or accommodation listing ID [id]; data categories [cat], and temporal information [temp].

These specific data variables can be grouped and combined to address different research topics in the field of urban studies. Some of these potential topics are presented in Section 2.2.2.

2.2.1. Data verification and validation

Data harmonization prior to visualization or analysis strengthens the validity of the data, avoiding errors, over-presence or duplicated information.

Some users can be extremely active on LBSNs and, thus, skew the interpretation of the pattern, especially if the size of the sample is reduced (Mayr & Weller, 2017). For example, a single user can actively generate a large number of tweets from a fixed location (Lloyd & Cheshire, 2017) or, in the case of Foursquare, special promotions exclusive to users checking-in a specific business establishment may skew the results for identifying user presence and preference.

Moreover, duplicate *venues* and *places* were found in Foursquare and Google Places, but not in Airbnb.

In the case of Foursquare, duplicate *venues* can be detected easily when they have the same name; these kinds of duplicates have been found to account for less than 10% of dataset listings —i.e. 3.14% in Prague, Czech Republic; 6.5% in Tallinn, Estonia; and 8.78% in Valencia, Spain. All datasets retrieved by 11 April 2018—. Other duplicate *venues* that need to be carefully sorted are those that might be a typo or a different name to the same *venue* and thus be listed twice. For instance, in the case of the dataset of Alicante, Spain, the *venue* of the municipal cemetery is listed twice as: "Cementerio De Alicante", with 15 check-ins and 12 users; and "Cementerio de Alicante", with 15 check-ins, 12 users, in dataset retrieved on 16 February 2018. In this case, both *venues* are considered as only one and the final number of check-ins and users corresponds to the addition of the two previous *venues*.

Table 3
LBSNs' general data variables.

General variables	Data format/ type	FOURSQUARE	TWITTER	GOOGLE PLACES	INSTAGRAM	AIRBNB
1. Location [LOC]		Longitude	Longitude	Longitude	Longitude	Longitude
	coor	Latitude	Latitude	Latitude	Latitude	Latitude
	txt	Address, city, country	City, country	Street, number, neighbourhood, district, city, country	Geolocated pin	Neighbourhood, city, country
2.Temporal information [TEMP]	temp	Cumulative data on *venues*	Time the tweet was posted	Updated data on registered *places*	–	Listing creation date
3.User generated data [UGDAT]	txt	*Venue* name	Tweets text	*Place* name	Photo description	Listing title/ description
	num	Check-ins	–	–	–	
	num	Users	–	–	–	Number of Bedrooms
	rat	Rating	–	Rating	–	Rating
	txt	Tips, reviews		-	-	Average rate
	pho	Photographs	Photographs	–	Photographs	Photographs
4. Data categorization [CAT]	cat	Hierarchy of categories and sub-categories	Tweet language Hashtags	Categories, sub-categories, sub-sub-categories	Hashtags	Listing type, Property type —host selects from drop down menu—
5. Data ID [ID]	id	*Venue* ID and URL	User ID, Tweet ID	*Place* ID	Image ID and URL	Property ID

In Google Places some *places* are registered twice with a different name. For example, the same restaurant could be referred to as a "bar" or "cafeteria". Previous experience has shown that a Google Places dataset could include up to 2% duplicate listings. One exemplary case is the raw dataset of the cities of Alicante and Valencia in Spain, with 32,995 and 72,621 *place* listings, respectively —datasets retrieved on 16 Feb 2018—. After the deletion of duplicate listings, the unique datapoints amounted to 32,392 and 72,019, respectively.

With regard to Twitter and Instagram duplicated data, the duplicate verification of these datapoints is rather simple since every single tweet and post on Instagram has its own unique ID. However, two relevant considerations should be taken into account while validating data from Twitter, especially related to the tweets' locative features. Firstly, copy-pasted tweets that are generated as new tweets have the same content with a different user ID. In the case where the research requires unique text, one tweet would need to be deleted. Secondly, when a user or business is highly active on Twitter, it might skew the analysis of the spatial tweet pattern distribution as there may be a disproportionate number of tweets generated from a single location by the same user ID.

Once duplicated data have been removed, perusal of the data before analysis guides decision making on the appropriate filtering for the research purpose (Chiera & Korolkiewicz, 2017). The consistency and organization of datasets largely depend on data categorization, hierarchy, and structure, which are determined by the LBSN and the users' criteria for registering and classifying information. Two distinguishable cases emerge on how data are organized into categories: by tags and/or user-generated keywords, as in the case of Instagram and Twitter; and by predetermined categories as in the case of Foursquare, Google Places and Airbnb.

In the case of LBSNs using keywords to classify the information, data from Twitter —texts— and Instagram —images— are grouped according to the hashtags included in the user's post. The '#'—hashtag— and '@'—at— symbols before a keyword or a user allow all posts related to the same topic or user to be grouped together.

For those LBSNs that use distinctive predetermined categories, such as Foursquare, Google Places and Airbnb, the validation, refinement and re-assignment of categories to data is necessary depending on the research topic and the database's consistency.

Foursquare has 10 general categories (Foursquare Inc., 2017). Each category is divided into a wide range of sub-categories that provide more information about the *venue's* description. Foursquare users registering a *venue* on the platform can assign a category and a sub-category; however, the logic behind why some sub-categories are assigned to a category is not always clear (M. J. Williams & Chorley, 2017). Although there are some strategies to promote consistency across *venue* data (M. J. Williams & Chorley, 2017) —such as a "style guide" and voluntary reviewers called "Superusers"—, a careful revision of categories and subcategories is needed.

As for Google Places, when users register a *place* on the platform, they assign one or more *place types* —Google Places categories—. There are over 120 predefined *place types* (Google Developers, 2018), thus, user-assigned categorization of *places* is even less accurate than in the case of Foursquare. Therefore, Google Places datasets need to be revised, refined and many *places* require recategorization prior to analysis for five reasons:

(i). As previously mentioned, sometimes *places* are registered twice with a different name which needs to be considered as one *place*.
(ii). Some Google Places categories are too general, and/or some *places* may not have assigned a specific sub-category, thus it is not clear what type of place they represent. Specifically, the categories "establishment", "premise" and "point of interest" could include all kinds of *place* types, for instance, restaurants, hotels, offices, lawyer offices, banks, etc. These *places* may account for over 32% of the unique datapoints in a dataset. Taking the previous case examples, 10,242 listings in Alicante and 22,408 listings in Valencia, out of the 32,392 and 72,019 unique datapoints, respectively, belong to those three non-specific categories.[1] Since there is a large quantity of these datapoints in the datasets, reassigning a category and subcategory is important prior to any analysis.
(iii). Some data listings do not represent an economic activity or a *place* but refer to a larger geographic area or region. That is the case of *places* categorised as "street_address"; "postal_town"; and, "sublocality_level_4" categories. The number of places that fall within these categories may represent up to 40% of all unique data listings. Alicante city dataset has 12,557 non-economic activity *places* while *Valencia* has 27,708; out of the 32,392 and 72,019 unique datapoints, respectively.
(iv). While recategorizing a *place*, existing Google Places categories may not be applicable to businesses and places within a specific location, thus new categories need to be created. For instance, in

[1] Since February 16, 2017 some non-specific general categories such as "establishment" and "point of interest" have been deprecated (Google Developers, 2018), although places registered prior to that date remain in their originally assigned categories.

Table 4
Example research topics that can be addressed by combining LBSNs data variables.

	FOURSQUARE	TWITTER	GOOGLE PLACES	INSTAGRAM	AIRBNB
Research topic	Identification of the most visited/checked-in *venues*. [1]	Spatiotemporal patterns of people presence, activities and languages. [3]	Quantity and diversity of economic activities in an area. [2]	Identify relevant spatial features/ character related to the user experience and perception. [4]	location and clusters of accommodation typology —single family house, multifamily/ apartment building— [5]
Variables selected	UGDAT, ID, TEMP	LOC, TEMP	LOC, CAT, ID	LOC, UGDAT (pho)	LOC, CAT

the case of the Alicante and Valencia datasets, new categories had to be created such as “lottery”, with 188 and 478 establishments.

(v). There are cases where the listing location descriptors and addresses are not homogeneous. For example, in the case of Alicante and Valencia datasets, the address field of 4123 and 719 listings, respectively, have “Avenida”; 112 and 184 listings respectively have “Av.”; and, 1172 and 7025 listings respectively have “Avinguda”. In these cases, harmonization of the terms should be considered prior to analysis.

Airbnb's temporary accommodation listings are classified into two main groups: *property type* and *listing type*. These categories have sub-categories that provide further details about the accommodation characteristics. For example, a *property type* could be an apartment, bed & breakfast, boutique hotel, Bungalow, Camper, Dorm, Loft, etc.; and a *listing type* refers to whether the accommodation listing includes the entire apartment, a private room or a shared room. Even though Airbnb has predefined sub-categories, anybody listing a property can create a new *property type* in, for example, a different language. For instance, datasets of many Spanish cities have “casa particular” as a *property type*. Thus, as in the case of Google Places, where the listing categories are reconsidered, the Airbnb's *property type* categories need to be revised and possibly grouped into fewer categories —”apartment” and “serviced apartment” listings could fall within the same *property type* category, for example—.

2.2.2. Data selection, reclassification and interpretation

Thorough data variables selection and data reclassification, followed by detailed examination, is important to ensure that the sample results are valid for the specific research purpose and thus can be then interpreted to obtain representative conclusions (Lansley & Longley, 2016). The variable data selection and classification is necessary as the data have not been generated for urban research purposes. Also, as previously mentioned, an appropriate selection of data variables is required, which is conditioned by the research topic to be addressed. The following Table 4 presents five example research topics, relevant for the field of urban studies, that will be used to explain filtering methods for the interpretation of the LBSN data selected: [1] people's perception and preference over *venues* can be assessed using Foursquare by ranking the number of visitors and check-ins and by analysing the *venue*'s user-shared images and opinions; [2] the diversity and quantity of economic activities in a specific urban area can be analysed using Google Places' listing of businesses; [3] spatiotemporal patterns of people presence, activities and languages can be assessed using Twitter's geolocated tweets; [4] the perception and character of the urban environment can be depicted from the analysis of Instagram images and hashtags; and, [5] location patterns and building typologies of unregistered temporary accommodation can be identified by using data from Airbnb.

Considering the different groupings and combinations of data variables, there are several key points about the filtering methods used for studying the five research topics —Table 4— that will subsequently be dealt with.

2.3. Foursquare: People's perception and preferences

Foursquare datasets include information that is valuable for identifying people's perception (Quercia, 2015, 2016) and preferences in cities (Agryzkov, Martí, Tortosa, & Vicent, 2016; Tasse & Hong, 2014; Van Canneyt, Schockaert, Van Laere, & Dhoedt, 2012b). It is possible to ascertain the cumulative total amount of visitors and check-ins registered for each Foursquare *venue*. Filtering *venues* by number of visitors and check-ins allows identification of the most visited and thus, the most preferred *venues*. However, deciding whether to use the number of check-ins rather than the number of visitors depends on the research question itself considering that a single visitor can check-in multiple times in a venue. Many authors consider the check-ins number for analysing *venue* preferences or identifying key points of interest (Ferreira, Silva, & Loureiro, 2016; Jiang et al., 2015); while other scholars take the cumulative number of visitors to identify how many people have checked-in a venue at least once (Bentley, Cramer, & Müller, 2015; Martí et al., 2017; Noulas, Scellato, Mascolo, & Pontil, 2010). For example, to identify which public plaza is the most socially relevant in Foursquare, the dataset is filtered so that *venues* are ranked according to the number of unique visitors registered under the sub-category “plaza”, within the general category “outdoors & recreation” (Martí et al., 2017). This process could be applied to a different kind of *venue*, for instance, to identify the most preferred restaurants or stores.

Furthermore, the pictures and opinions —tips— shared by Foursquare users on each *venue* provide an indication of how the space is perceived and used (Aliandu, 2015; Y. Chen, Yang, Hu, & Zhuang, 2016). The photographed activities of users —for instance, kids playing in a plaza— and the urban/architectural features in the background —fountains, sculptural elements, etc.— could be useful perceptual indicators of a *venue's* safety for children, or of whether the *venue* is a youth-oriented space. However, it is often found that in some cities, the sharing of pictures and tips is scant because the social network is mostly intended to showcase presence with check-ins or because the social network's penetration is low. For example, the case of the most visited *venues* categorised as plazas in Foursquare: Plaza Catalunya in Barcelona and Plaza Luceros in Alicante. Barcelona has a population of 1.6 million people, and Alicante, a population of just over 300,000. As of 24 August 2018, these plazas had, respectively, 8199 photographs and 656 tips; and, 419 photographs and 41 tips.

2.4. Google Places: The diversity and quantity of economic activities

Information on Google Places listings —classified by category and sub-category— reveals clusters of economic activities as well as quantity, diversity, and complexity in the spatial distribution of these activities and places of interest. The regrouping of categories into much fewer and more general categories is helpful not only for making easier reading and interpretation of cartographies, avoiding the colour coding

Table 5
Example categorization of Twitter data by daily time periods, tweet language and frequency of hashtags.

	Axis Paseo de la Castellana, Madrid	Axis Diagonal Avenue, Barcelona
Total tweets collected from 21-09-2016 to 17-02-2017	61,716	14,849
Tweets aggregated by daily time periods		
Early morning [0:00 to 6:59 h.]	7.02%	7.85%
Morning [7:00 to 12:59 h.]	27.88%	30.54%
Afternoon [13:00 to 18:59 h.]	36.44%	37.03%
Evening-night [19:00 to 23:59 h.]	28.66%	24.58%
Tweet languages		
Spanish	68.55%	35.84%
English	17.34%	25.91%
Italian	0.60%	1.61%
Portuguese	3.89%	1.54%
Undefined	4.43%	27.50% Mostly Catalan
Others	5.19%	7.60%
Word cloud with most repeated *hasthags.*		

of 120 or so *place* types in Google Places, but also for studying specialization of economic activities. For instance, previous experiences have proven that the recategorization of *places* into the Land Based Classification Standards —LBCS—, specifically into the "functional dimension" hierarchical categories (American Planning Association, 2018), enables the identification of location patterns and spatial distribution of economic activities at different scales and granularity. This classification provides a fine-grain land use class taxonomy based on three levels: 7 main categories, 47 sub-categories and 159 sub-sub-categories (Deng & Newsam, 2017). As an example, in the case of Alicante city dataset —retrieved on 16 Feb 2018—, the allocation of Google Places *place* types into the first level APA categories resulted as follows:

1000- Residence or accommodation functions- 3.6%.
2000- General sales or services- 46.9%.
3000- Manufacturing and wholesale trade- 4.5%.
4000- Transportation, communication, information and utilities- 9%.
5000- Arts, entertainment and recreation- 14.4%.
6000- Education, public admin, health care and other institutions- 17.12%.
7000- Construction-related businesses- 4.6%.

2.5. Twitter: Spatiotemporal patterns of people presence

In general, selecting variables related to either geo-location or the tweet content to analyse temporal patterns of activities and people presence can provide two different filtering approaches: spatiotemporal and/or message content.

Firstly, tweet representations in a cartography by using tweet timestamps and geolocation is an easy and straightforward way to observe the concentration patterns of tweets shared in a certain area (Adnan, Longley, & Khan, 2014; Fujita, 2013; Steiger, Westerholt, Resch, & Zipf, 2015). The time-based aggregation of tweets could be useful to understand regular activities happening at a certain hour on a certain day of the week. For example, in the case of the urban axes that run and extend along both sides of Paseo de la Castellana in Madrid —6.3 km long— and the Diagonal avenue in Barcelona —10.2 km long—, respectively, 61,716 and 14,849 Twitter datapoints were retrieved between 21 September 2016 and 17 February 2017. These datasets were aggregated into four daily time periods as shown in Table 5, showing that daily tweeting patterns in Madrid and Barcelona are quite similar. This type of filtering process is also applicable to recognise when one-off events or demonstrations happen. In the latter case, the number of tweets increases substantially in a certain urban area and fades away once the event is finished (Bolognesi & Galli, 2017; Panteras et al., 2015).

Secondly, categorization of data by the tweet or the user language has also proved to be a useful way to identify, for example, the geographical location of the different cultures and nationalities in a city. It is possible to find out what kind and how many foreign languages are spoken in a certain area (Fisher, 2011; Lange & Waal, 2013). For instance, in the case of Madrid and Barcelona, Spanish in both cities is the most spoken language among tweeters; however, Barcelona presents a greater amount of English and "undefined" language tweets, most of which are in Catalan —Table 5—.

Lastly, the recognition of certain activities, opinions, ideas and trending topics that are predominant in a given place and at a given time can be detected by using the information related to the tweet content —text, hashtags— and sentiment analysis (Cheng et al., 2011; Yang, Sun, Zhang, & Mei, 2012). Word count techniques can be applied to a Twitter dataset and represented, for example, in a word cloud using scaled text size where the higher the frequency of words in a dataset the larger the font size (Sang & Van Den Bosch, 2013) —Table 5—.

2.6. Airbnb: Location patterns and building typologies of unregistered temporary accommodation

Airbnb geolocated data provide useful information about the spatial distribution and concentration patterns of temporary accommodation by *property type* or *listing type* in a given area (Moreno Izquierdo, Ramón Rodríguez, & Such Devesa, 2016; Temes Cordóvez, Simancas Cruz,

Table 6
Airbnb's property type categories grouped into building's typology classification.

Multifamily	Single family	Private room	Others
Apartment	Bungalow	Bed & breakfast	Camper
Boutique hotel	Cabin	Casa particular	Boat
Condominum	Chalet	Dorm	Igloo
Entire floor	House	Guest suite	
Loft	Villa	Guest house	
Timeshare	Townhouse	Hostel	
Others	Nature lodge	Timeshare	
Serviced apartment	Earth House	In law	
Vacation home			

Peñarrubia Zaragoza, Moya Fuero, & García Amaya, 2016). The definition of the different accommodation categories is vague —for example, it is difficult to know the difference between the *property types* service apartment vs. apartment— and the user-generated information is not homogeneously classified. Therefore, it becomes necessary to regroup *property type* categories. For instance, in a study conducted of 9 Spanish cities —Alicante, Benidorm, Calpe, Castellón, Gandía, Peñíscola, Teulada, Torrevieja and Valencia—, a new categorization of listings by *property type* was proposed to specify the listing's building typology: i) multifamily housing apartment; ii) single family housing, iii) private room, and iv) others —Table 6.

2.7. Instagram: The perception and character of the urban environment

Instagram data offer relevant insights about what is interesting in the urban environment for people. Pictures shared through Instagram “promote visual rather than textual communication” (Laestadius, 2017), thus the analysis of the character and identity of the urban environment can be depicted from the filtering and studying of a much smaller dataset than in other types of data. However, unlike the previously explained social networks whose information is retrieved in the form of a spreadsheet, filtering large sets of Instagram images can be rather challenging and still remains largely inaccessible for researchers (Laestadius, 2017). There are open tools available that can classify pictures automatically according to their hue and luminosity (Hochman & Manovich, 2013; Manovich, 2016). These techniques are useful to identify, for example, which pictures are taken indoors or outdoors and to gauge the extent to which users are interested in outdoor and/or indoor activities.

The manual filtering and geocoding of pictures retrieved via screen-captured posts (Laestadius, 2017) and/or Instagram webpage downloads (López Baeza, Serrano Estrada, Nolasco-Cirugeda, Serrano-Estrada, & Nolasco-Cirugeda, 2016) is often done by using place hashtags. These correspond to a geolocated point that represents a place or a region —i.e. #centralpark; #newyork—, thus filtering by these hashtags is a straightforward way to obtain a sample with images that are shared in a specific urban location. Another type of picture aggregation and filtering is done by categorizing the content of the picture; for instance: a selfie; a person posing nearby a specific urban element —tree, monument—; landscape; scenery; and, architecture. Moreover, ascertaining people's activities in photos could provide an indication of the perception of the surrounding space.

2.8. Other research topics

A compilation of other potential research topics using the aforementioned social networks and their respective data variables are listed in Table 7.

Table 7
Potential research topics in the field of urban studies that can be approached by using LBSN data variables.

	FOURSQUARE	TWITTER	GOOGLE PLACES	INSTAGRAM	AIRBNB
Researchtopic	Offer of economic activities and *venues* of public interest	People presence in the urban public or private space.	Public opinion/evaluation of a business or service.	Keywords/ hashtags related to the user experience/opinion of a place	location and clusters of the different residential rental types—single room, entire property—
Variables selected	LOC, UGDAT (txt, rat, pho), ID	LOC	UGDAT (rat), ID	LOC, UGDAT (txt)	LOC, CAT
Research topic	Cumulative people presence in a *venue* up to the retrieval date.	Text and/or hashtags to depict user location —district, neighbourhood, city—	Particularities of the case study derived from the business offer. Economic activities and services that characterize an urban area.	Identify the social/public activities developed in a space.	Geographical distribution of rental homes with their respective rating values.
Variables selected	LOC, UGDAT (rat), ID	LOC, UGDAT (txt)	LOC, CAT, ID	LOC, UGDAT (pho)	LOC, UGDAT (rat)
Research topic	Tips, reviews or comments —public opinion—	Depict cultural features, traditions, routines, habits of residents through the text they share.	Economic activities on the main floor that contribute to the livability of urban spaces.	User Profile of frequenters—gender, approximate age—	Average rating value of rental homes for selected urban areas.
Variables selected	UGDAT (txt)	LOC, UGDAT (txt)	LOC, CAT, ID	UGDAT (pho)	LOC, UGDAT (rat)
Research topic	Type of activity that takes place in the space.	Frequency of people tweeting in an urban area.	Predominant economic activities and specialization of a neighbourhood.	Description of a space through hashtags as keywords	Physical qualities of the best rated rental types.
Variables selected	LOC, CAT, ID, UGDAT (txt)	LOC, TEMP, ID	LOC, CAT, ID	UGDAT (txt)	UGDAT (pho, rat)
Research topic	Physical characteristics of the *venues* relevant to user experiences	Opinion, emotions about relevant events, social and political matters.			
Variables selected	UGDAT (pho)	LOC, UGDAT (txt)			
Research Topic	Local habits and social behaviour in the space.	Opinion, perception about urban spaces.			
Variables selected	LOC, GDAT(pho)	LOC, UGDAT (pho)			

3. Discussion and conclusions

The findings of this research back the many previously cited urban scholars who support the use of LBSN data for the study of cities. This trend is set to continue given that content generated by an exponentially growing community of LBSN users cannot be neglected in urban research of a qualitative nature. These data can potentially trigger more discussion about current trends in urban reality than traditional sources, which cannot compete in terms of immediacy, availability and quantity of data.

This study underscores the importance of addressing the challenges, limitations as well as the opportunities provided by LBSN data for the field of urban studies. A new framework is presented in this study for overcoming several challenges associated with the retrieval, validation, selection, filtering and interpretation of geolocated user-generated data —from Twitter, Foursquare, Google Places, Instagram and Airbnb—. The findings evidence that a close review and manual verification are required to avoid losing the implicit nuances of each dataset and thereby, of each case study.

Furthermore, two issues may compromise the rigorous procedure and the reproducibility of this type of research. First, reliance on data accessibility makes the retrieval process vulnerable to the changes in access conditions; and, second, the excessive amount of data makes manual verification of large datasets impractical and implies certain automatization processes —a script, for example—. Accordingly, the increasing availability of free and open data means a more representative sampling, but, as argued by Boyd & Crawford (2012) "bigger data are not always better data".

LBSN-oriented methods present limitations for the study of cities in terms of the representability and applicability of the data, according to some previously cited scholars. This research recognizes the constraints associated with using LBSN data for the analysis of urban phenomena, with specific reference to: [1] the complexity involved in requesting and retrieving data according to each LBSN; [2] the amount of data retrieved, whether the sample is too small to be representative or too large to manage; [3] the validation, selection, filtering and interpretation of data, as a process that is conditioned by the complexity of the research topic and the distinctive variables obtained from each social network.

In relation to the complexity involved in requesting and retrieving data [1], this research underscores the importance of dealing properly with the API requirements in terms of shape and size of the search polygon and the number of results per request. Precisely,

one of the main methodological contributions is the recognition of key aspects involved in the data retrieval, making the method transferrable to other LBSNs. For example, even though most common social media APIs use the Rest method (Brown, Soto-Corominas, Suárez, & de la Rosa, 2017), an approach to Twitter Streaming API request method is rather similar not in terms of quantity but in terms of data representability, as explained in Section 2.1.2 Twitter, Fig. 4.

As for the number of datapoints retrieved [2], the information from a specific location can be far richer if the resulting analyses of two or more sources are considered when approaching a single research case study. There are some aspects that can be related among social networks such as the relation between the number of datapoints and the measured area of cities —Table 2—, or apparently common types and formats of data variables —such as *venues* and *places*—. However, the raw data from two different sources should not be compared until the data have been independently analysed—Fig. 6—. These analysed data can be complementary to address a research topic. For example, Foursquare and Google Places both provide a listing of points of interest. However, the size of the dataset, the data variables and the purpose for which users share data is rather different.

That is why the verification and selection [3] processes are important as they may show that there are *places* registered in Google Places that are not present in Foursquare and vice versa. In this case, a business or urban area may not be considered relevant by Foursquare users —not checked-in—, but the establishment may be listed in Google Places. Similarly, a recently opened *venue* may not yet be listed in Google Places, but it may have check-ins on Foursquare. Thus, the combination of the resulting analyses of filtered and selected data from different LBSNs can supplement the information on a sample to produce a more complete and accurate research approach.

Comprehensive research on a specific urban topic may require the consideration of validated information from different LBSN datasets and, therefore, the selection of different variables. Notably, analysing variables related to images —Instagram, Twitter and Foursquare— remains challenging in terms of the slowness of the procedure. Although advances in image recognition software can facilitate this task, each image still needs to be viewed manually to appreciate local nuances (Boy & Uitermark, 2016) related to social activity, for example.

Finally, the main contribution of this work is a comprehensive framework for the study of cities that effectively deals with the challenges and opportunities provided by readily accessible user-generated LBSN data. The approach presented could benefit urban design and planning intervention criteria.

Acknowledgements

This work was supported by the Council of Education, Research, Culture and Sports – Generalitat Valenciana (Spain). Project: Valencian Community cities analysed through Location-Based Social Networks and Web Services Data. Ref. no. AICO/2017/018.

References

Adnan, M., Longley, P. A., & Khan, S. M. (2014). Social dynamics of Twitter usage in London, Paris, and New York City Citation Format. *First Monday, 19*(5).
Agryzkov, T., Nolasco-Cirugeda, A., Oliver, J. L., Serrano-Estrada, L., Tortosa, L., & Vicent, J. F. (2015). Using data from Foursquare Web Service to represent the commercial activity of a city. *International Journal of Computer, Control, Quantum and Information Engineering. World Academy of Science, Engineering and Technology, 9*(1), 69–76.
Agryzkov, T., Martí, P., Tortosa, L., & Vicent, J. F. (2016). Measuring urban activities using Foursquare data and network analysis: A case study of Murcia (Spain). *International Journal of Geographical Information Science,* 1–22.
Aho, A. V., Hopcroft, J. E., & Ullman, J. D. (1974). *The Design and Analysis of Computer Algorithms.* Reading: Addison-Wesley Publishing Company.
AirDNA (2017). Short-Term Rental Data Methodology - The AI and science behind AirDNA. Retrieved 28 August 2018, from https://www.airdna.co/methodology.
Al-Ghamdi, S. A., & Al-Harigi, F. (2015). Rethinking image of the City in the Information Age. *Procedia Computer Science, 65*, 734–743. https://doi.org/10.1016/j.procs.2015.09.018.
Aliandu, P. (2015). Sentiment Analysis to Determine Accommodation, Shopping and Culinary Location on Foursquare in Kupang City. *Procedia Computer Science, 72*, 300–305. https://doi.org/10.1016/j.procs.2015.12.144.
American Planning Association (2018). LBCS Function Dimension with Descriptions. Retrieved 18 January 2018, from https://www.planning.org/lbcs/standards/function.htm.
Anselin, L., & Williams, S. (2015). *Digital Neighborhoods*.
Arribas-Bel, D. (2014). Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography, 49*, 45–53. https://doi.org/10.1016/j.apgeog.2013.09.012.
Arribas-Bel, D., Kourtit, K., Nijkamp, P., & Steenbruggen, J. (2015). Cyber Cities: Social Media as a Tool for Understanding Cities. *Applied Spatial Analysis and Policy, 8*(3), 231–247. https://doi.org/10.1007/s12061-015-9154-2.
Barbera, P., & Rivero, G. (2015). Understanding the Political Representativeness of Twitter users. *Social Science Computer Review, 33*(6), 712–729. https://doi.org/10.1177/0894439314558836.
Bawa-Cavia, A. (2011). Sensing the urban: using location-based social network data in urban analysis. *Pervasive PURBA Workshop* (pp. 1–7). .
Béjar, J., Álvarez, S., García, D., Gómez, I., Oliva, L., & Tejeda, A. (2016). Discovery of spatio-temporal patterns from location-based social networks. *Journal of Experimental & Theoretical Artificial Intelligence, 28*(1–2), 313–329. https://doi.org/10.1080/0952813X.2015.1024492.
Bentley, F., Cramer, H., & Müller, J. (2015). Beyond the bar: The places where location-based services are used in the city. *Personal and Ubiquitous Computing, 19*(1), 217–223. https://doi.org/10.1007/s00779-014-0772-5.
Bolognesi, C., & Galli, A. (2017). Mapping Socials a Voluntary Map of a Great Event in Monza Park. *Proceedings. Vol. 1. Proceedings* (pp. 917–). . https://doi.org/10.3390/proceedings1090917.
Boy, J. D., & Uitermark, J. (2016). How to Study the City on Instagram. *PLoS One, 11*(6),

e0158161. https://doi.org/10.1371/journal.pone.0158161.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information Communication and Society, 15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878.

Brown, D. M., Soto-Corominas, A., Suárez, J. L., & de la Rosa, J. (2017). Overview- the social media data processing pipeline. In A. Q.-H. Luke Sloan (Ed.). *The SAGE Handbook of Social Media Research Methods* (pp. 125–145). London: SAGE Publications Ltd.

Campagna, M. (2016). Social Media Geographic Information: Why social is special when it goes spatial? *European Handbook of Crowdsourced Geographic Information* (pp. 45–54). .

Cerrone, D. (2015). *A Sense of Place.* Turku.

Chen, L., & Roy, A. (2009). Event detection from flickr data through wavelet-based spatial analysis. *Proceedings of the 18th ACM Conference on Information and Knowledge Management* (pp. 523–532). . https://doi.org/10.1145/1645953.1646021.

Chen, Y., Yang, Y., Hu, J., & Zhuang, C. (2016). Measurement and analysis of tips in foursquare. *2016 IEEE International Conference on Pervasive Computing and Communication Workshops, PerCom Workshops 2016* (pp. 4–7). .

Cheng, Z., Caverlee, J., Lee, K., & Sui, D. Z. (2011). Exploring millions of Footprints in Location sharing Services. *Icwsm, 2010*, 81–88.

Chiera, B. A., & Korolkiewicz, M. W. (2017). Visualizing big Data: Everything Old is New again. In F. P. García Márquez, & B. Lev (Eds.). *Big Data Management*Springer International Publishinghttps://doi.org/10.1007/978-3-319-45498-6_1.

Chorley, M. J., Whitaker, R. M., & Allen, S. M. (2015). Personality and location-based social networks. *Computers in Human Behavior, 46*, 45–56. https://doi.org/10.1016/j.chb.2014.12.038.

Deng, X., & Newsam, S. (2017). Quantitative Comparison of Open-Source Data for Fine-Grain Mapping of Land Use. *Proceedings of the 3rd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics - UrbanGIS. Vol. 17. Proceedings of the 3rd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics - UrbanGIS* (pp. 1–8). . https://doi.org/10.1145/3152178.3152182.

Dunkel, A. (2015). Visualizing the perceived environment using crowdsourced photo geodata. *Landscape and Urban Planning, 142*, 173–186. https://doi.org/10.1016/j.landurbplan.2015.02.022.

Ferreira, A. P. G., Silva, T. H., & Loureiro, A. A. F. (2016). Beyond Sights: Large Scale Study of Tourists' Behavior using Foursquare Data. *Proceedings - 15th IEEE International Conference on Data Mining Workshop, ICDMW 2015* (pp. 1117–1124). . https://doi.org/10.1109/ICDMW.2015.234.

Fisher, E. (2011). Language communities of Twitter. Retrieved 20 July 2001, from https://flic.kr/p/ayDr8X.

Foursquare Inc (2017). Foursquare Category Hierarchy. Retrieved 1 January 2018, from https://developer.foursquare.com/docs/resources/categories.

Fujita, H. (2013). Geo-tagged Twitter collection and visualization system. *Cartography and Geographic Information Science, 40*(3), 18. https://doi.org/10.1080/15230406.2013.800272.

García-Palomares, J. C., Salas-Olmedo, M. H., Moya-Gómez, B., Condeço-Melhorado, A., & Gutiérrez, J. (2017). *City dynamics through Twitter: Relationships between land use and spatiotemporal demographics Cities.* https://doi.org/10.1016/J.CITIES.2017.09.007.

González-Bailón, S., Wang, N., Rivero, A., Borge-Holthoefer, J., & Moreno, Y. (2014). Assessing the bias in samples of large online networks. *Social Networks, 38*(1), 16–27. https://doi.org/10.1016/j.socnet.2014.01.004.

Goodchild, M. F. (2013). The quality of big (geo)data. *Dialogues in Human Geography, 3*(3), 280–284. https://doi.org/10.1177/2043820613513392.

Google Developers (2018). Place Types. Retrieved 30 April 2018, from https://developers.google.com/places/supported_types.

Graham, M., Hale, S. A., & Gaffney, D. (2014). Where in the world are you? Geolocation and Language Identification in Twitter. *The Professional Geographer,* 1–11.

Granell, C., & Ostermann, F. O. (2016). Beyond data collection: Objectives and methods of research using VGI and geo-social media for disaster management. *Computers, Environment and Urban Systems, 59*, 231–243. https://doi.org/10.1016/j.compenvurbsys.2016.01.006.

Hamstead, Z. A., Fisher, D., Ilieva, R. T., Wood, S. A., McPhearson, T., & Kremer, P. (2018). Geolocated social media as a rapid indicator of park visitation and equitable park access. *Computers, Environment and Urban Systems*. https://doi.org/10.1016/j.compenvurbsys.2018.01.007.

Han, B., Cook, P., & Baldwin, T. (2014). Text-based twitter user geolocation prediction. *Journal of Artificial Intelligence Research, 49*, 451–500. https://doi.org/10.1613/jair.4200.

Hecht, B., & Stephens, M. (2014). *A Tale of Cities: Urban Biases in Volunteered Geographic Information. Icwsm.* 197–205. http://doi.org/papers3://publication/uuid/B13C63A5-B3B8-4619-9558-86BCAFE5E2CA.

Hochman, N., & Manovich, L. (2013). *Zooming into an Instagram City: Reading the local through social media.*

Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W., & Prasad, S. (2015). Extracting and understanding urban areas of interest using geotagged photos. *Computers, Environment and Urban Systems, 54*, 240–254. https://doi.org/10.1016/j.compenvurbsys.2015.09.001.

Huang, Q., & Wong, D. W. S. (2015). Modeling and Visualizing regular Human Mobility patterns with uncertainty : An example using Twitter Data Modeling and Visualizing regular Human Mobility patterns with uncertainty : An example using Twitter Data. *Annals of the Association of American Geographers, 105*(6), 1179–1197 November.

INE (2011). Instituto Nacional de Estadística. Retrieved 11 May 2018, from http://www.ine.es/censos2011_datos/cen11_datos_resultados.htm.

Instituto Geográfico Nacional (2018). Centro Nacional de Información Geográfica. Retrieved 4 April 2018, from http://www.ign.es/web/ign/portal/inicio.

Jacobs, J. (1961). *The death and life of great American cities.* New York: Vintage Books.

Jagadeesan, J., & Venkatesan, N. (2015). Study of API for web applications. *International Journal of Contemporary Research in Computer Science and Technology, 1*(7), 257–261. Retrieved from http://www.ijcrcst.com/papers/IJCRCST-OCTOBER15-07.pdf.

Jiang, S., Alves, A., Rodrigues, F., Ferreira, J., & Pereira, F. C. (2015). Mining point-of-interest data from social networks for urban land use classification and disaggregation. *Computers, Environment and Urban Systems, 53*, 36–46. https://doi.org/10.1016/j.compenvurbsys.2014.12.001.

Kemp, S. (2018). Digital in 2018: World's internet users pass the 4 billion mark. Retrieved from https://wearesocial.com/blog/2018/01/global-digital-report-2018.

Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography, 3*(3), 262–267. https://doi.org/10.1177/2043820613513388.

Laestadius, L. (2017). Instagram. In A. Q.-H. Luke Sloan (Ed.). *The SAGE Handbook of Social Media Research Methods* (pp. 573–592). London: SAGE Publications Ltd.

de Lange, M., & de Waal, M. (2013). *Owning the city: New media and citizen engagement in urban design.* (First Monday).

Lansley, G., & Longley, P. A. (2016). The geography of Twitter topics in London. *Computers, Environment and Urban Systems, 58*, 85–96. https://doi.org/10.1016/j.compenvurbsys.2016.04.002.

Lee, R., Wakamiya, S., & Sumiya, K. (2013). Urban area characterization based on crowd behavioral lifelogs over Twitter. *Personal and Ubiquitous Computing, 17*(4), 605–620. https://doi.org/10.1007/s00779-012-0510-9.

Leetaru, K., Wang, S., Cao, G., Padmanabhan, A., & Shook, E. (2013). Mapping the global Twitter heartbeat: The geography of Twitter. *First Monday, 18*.

Liftn, J., & Parad (2018). Dual reality: Merging the real and Virtual. OpenStreetMap Wiki. Retrieved from http://wiki.openstreetmap.org/w/index.php?title=Browsing&oldid=1550720.

Liu, L., Zhou, B., Zhao, J., & Ryan, B. D. (2016). C-IMAGE: City cognitive mapping through geo-tagged photos. *GeoJournal, 81*(6), 817–861. https://doi.org/10.1007/s10708-016-9739-6.

Lloyd, A., & Cheshire, J. (2017). Deriving retail Centre locations and catchments from geo-tagged Twitter data. *Computers, Environment and Urban Systems, 61*, 108–118. https://doi.org/10.1016/j.compenvurbsys.2016.09.006.

López Baeza, J., Serrano Estrada, L., & Nolasco-Cirugeda, A. (2016). Percepción y uso social de una transformación urbana a través del social media. *Las setas gigantes de la calle San Francisco. I2 Innovación e Investigación En Arquitectura y Territorio. Vol. 4. Las setas gigantes de la calle San Francisco. I2 Innovación e Investigación En Arquitectura y Territorio* (pp. 2–). . https://doi.org/10.14198/i2.2016.5.03.

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. https://doi.org/10.1016/j.apgeog.2016.03.001.

Lynch, K. (1960). *The image of the city.* MIT Press.

Mahto, D. K., & Singh, L. (2016). A dive into Web Scraper world. *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 689–693). .

Manovich, L. (2016). Notes on Instagrammism and mechanisms of contemporary cultural identity (and also photography, design, Kinfolk, k- pop, hashtags, mise-en-scène, and состояние). *Instagram and Contemporary image*.

Martí, P., Serrano-Estrada, L., & Nolasco-Cirugeda, A. (2017). Using locative social media and urban cartographies to identify and locate successful urban plazas. *Cities, 64*, 66–78. https://doi.org/10.1016/j.cities.2017.02.007.

Marwick, A. E., & Boyd, d. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society, 13*(1), 114–133. https://doi.org/10.1177/1461444810365313.

Mayr, P., & Weller, K. (2017). Think before you collect: Setting up a data collection approach for social media studies. In L. Sloan, & A. Quan-Haase (Eds.). *The SAGE Handbook of Social Media Research Methods* (pp. 108–124). London: SAGE Publications Ltd.

McCarney, R., Warner, J., Iliffe, S., van Haselen, R., Griffin, M., & Fisher, P. (2007). The Hawthorne effect: A randomised, controlled trial. *BMC Medical Research Methodology, 7*(30), https://doi.org/10.1186/1471-2288-7-30.

McLain, R., Poe, M., Biedenweg, K., Cerveny, L., Besser, D., & Blahna, D. (2013). Making sense of Human Ecology Mapping: An Overview of Approaches to Integrating Socio-Spatial Data into Environmental Planning. *Human Ecology, 41*(5), 651–665. https://doi.org/10.1007/s10745-013-9573-0.

Milne, D., Thomas, P., & Paris, C. (2012). Finding, Weighting and describing Venues : CSIRO at the 2012 TREC Contextual Suggestion Track. *The Twenty-first Text REtrieval Conference (TREC 2012) Proceedings*.

Moreno Izquierdo, L., Ramón Rodríguez, A., & Such Devesa, M. J. (2016). Turismo colaborativo stá Airbnb transformando el sector del alojamiento? *Economistas, 150*, 107–119.

Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. M. (2013). *Is the Sample good enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose.* 400–408. https://doi.org/10.1007/978-3-319-05579-4_10.

Noulas, A., Scellato, S., Mascolo, C., & Pontil, M. (2010). An Empirical Study of Geographic User activity patterns in Foursquare. *Fifth International AAAI Conference on Weblogs and Social Media* (pp. 570–573). .

Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., & Mascolo, C. (2012). A tale of many cities: Universal patterns in human urban mobility. *PLoS One, 7*(5), https://doi.org/10.1371/journal.pone.0037027.

Panteras, G., Wise, S., Lu, X., Croitoru, A., Crooks, A., & Stefanidis, A. (2015). Triangulating Social Multimedia Content for Event Localization using Flickr and Twitter. *Transactions in GIS, 19*(5), 694–715. https://doi.org/10.1111/tgis.12122.

Peña-López, I., Congosto, M., & Aragón, P. (2014). Spanish Indignados and the Evolution of the 15M Movement on Twitter: Towards Networked Para-institutions. *Journal of*

Spanish Cultural Studies, 1–28.
Pew Research Center (2017). Social media fact sheet. Retrieved 15 May 2016, from http://www.pewinternet.org/fact-sheet/social-media/.
Quercia, D. (2015). *Chatty, Happy, and Smelly Maps. Proceedings of the 24th International Conference on World Wide Web, 741.* https://doi.org/10.1145/2740908.2741717.
Quercia, D. (2016). *Playful Cities: Crowdsourcing Urban Happiness with Web Games. 42,* 3.
Quercia, D., & Saez, D. (2014). Mining urban deprivation from Foursquare: Implicit crowdsourcing of city land use. *IEEE Pervasive Computing, 13*(2), 30–36. https://doi.org/10.1109/MPRV.2014.31.
Quercia, D., Aiello, L. M., Mclean, K., & Schifanella, R. (2015a). *Smelly Maps: The Digital Life of Urban Smellscapes.* AAAI Publications327–336.
Quercia, D., Aiello, L. M., Schifanella, R., & Davies, A. (2015b). *The Digital Life of Walkable Streets.* 875–884. https://doi.org/10.1145/2736277.2741631.
Roberts, H. V. (2017). Using Twitter data in urban green space research: A case study and critical evaluation. *Applied Geography, 81*, 13–20. https://doi.org/10.1016/j.apgeog.2017.02.008.
Roick, O., & Heuser, S. (2013). Location based social networks–definition, current state of the art and research agenda. *Transactions in GIS, 17*(5), 763–784.
Saker, M., & Evans, L. (2016). Locative Media and Identity: Accumulative Technologies of the Self. *SAGE Open, 6*(3), https://doi.org/10.1177/2158244016662692.
Samet, H. (1984). The Quadtree and Related Hierarchical Data Structures. *ACM Computing Surveys, 16*(2), 187–260. https://doi.org/10.1145/356924.356930.
Sang, E. T. K., & Van Den Bosch, A. (2013). Dealing with big data: The case of Twitter. *Computational Linguistics in the Netherlands Journal, 3*, 121–134. https://doi.org/10.1126/science.345.6193.148-a.
Serrano-Estrada, L., Marti, P., & Nolasco-Cirugeda, A. (2016). Comparing two Residential Suburban areas in the Costa Blanca, Spain, Articulo. *Journal of Urban Research, 13*. https://doi.org/10.4000/articulo.2935.
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. https://doi.org/10.1016/j.landurbplan.2015.02.020.
Silva, T. H., Vaz De Melo, P. O. S., Almeida, J. M., Salles, J., & Loureiro, A. A. F. (2014). Revealing the City that we cannot see. *ACM Transactions on Internet Technology (TOIT), 14*(4), 26.
Sloan, L. (2017). Social Science 'Lite'? Deriving Demographic Proxies from Twitter. In L. Sloan, & A. Quan-Haase (Eds.). *The SAGE Handbook of Social Media Research Methods* (pp. 90–104). London: SAGE Publications Ltd.
Sloan, L., & Morgan, J. (2015). Who tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on twitter. *PLoS One, 10*(11), 1–15. https://doi.org/10.1371/journal.pone.0142209.
Sloan, L., & Quan-Haase, A. (2017). *The SAGE Handbook of Social Media Research Methods.* London: SAGE Publications Ltd. Retrieved from https://www.amazon.es/Handbook-Social-Media-Research-Methods/dp/1473916321.
Soja, E. (1989). *Postmodern geographies. The reassertion of space in critical social theory.* London, New York: Verso.
Steiger, E., Westerholt, R., Resch, B., & Zipf, A. (2015). Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data. *Computers, Environment and Urban Systems, Vol. 54*, 255–265. https://doi.org/10.1016/j.compenvurbsys.2015.09.007.
Sui, D., & Goodchild, M. (2011). The convergence of GIS and social media: Challenges for GIScience. *International Journal of Geographical Information Science, 25*(11), 1737–1748. https://doi.org/10.1080/13658816.2011.604636.
Tasse, D., & Hong, J. I. (2014). Using social media data to understand cities. *NSC workshops on big data and urban informatics, Chicago*.
Temes Cordóvez, R. R., Simancas Cruz, M. R., Peñarrubia Zaragoza, M. P., Moya Fuero, A., & García Amaya, A. M. (2016). Characterization and spatial identification of holiday tourist assessments in the city of Valencia. In J. Rivas Navarro, & B. Bravo Rodríguez (Eds.). *6th Sustainable Development Symposium - Book of Abstracts*. Granada: Godei.
Tsou, M. H., Yang, J. A., Lusher, D., Han, S., Spitzberg, B., Gawron, J. M., & An, L. (2013). Mapping social activities and concepts with social media (Twitter) and web search engines (Yahoo and Bing): A case study in 2012 US Presidential Election. *Cartography and Geographic Information Science, 40*(4), 337–348. https://doi.org/10.1080/15230406.2013.799738.
Tufekci, Z. (2014). Big questions for social media big data: Representativeness, validity and other methodological pitfalls. *ICWSM '14: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media* (pp. 505–514). .
Twitter, I. (2018a). Rate limiting. Retrieved from https://developer.twitter.com/en/docs/basics/rate-limiting.html.
Twitter, I. (2018b). Tweet location FAQs. Retrieved 14 May 2018, from https://help.twitter.com/en/safety-and-security/tweet-location-settings.
Van Canneyt, S., Schockaert, S., Van Laere, O., & Dhoedt, B. (2012a). Detecting places of interest using social media. *Proceedings - 2012 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2012* (pp. 447–451). . https://doi.org/10.1109/WI-IAT.2012.19.
Van Canneyt, S., Van Laere, O., Schockaert, S., & Dhoedt, B. (2012b). Using social media to find places of interest. *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information - GEOCROWD '12*https://doi.org/10.1145/2442952.2442954.
Villatoro, D., Serna, J., Rodríguez, V., & Torrent-Moreno, M. (2013). The TweetBeat of the City: Microblogging used for Discovering Behavioural patterns during the MWC2012 BT. In J. Nin, & D. Villatoro (Vol. Eds.), *Citizen in Sensor Networks. Lecture Notes in Computer Science*. *Vol. 7685*. *Citizen in Sensor Networks. Lecture Notes in Computer Science* (pp. 43–56). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-36074-9_5.
Wang, W. (2013). Using Location-based Social Media for Ranking Individual Familiarity wih Places: A Case Study with Foursquare Check-in Data. In G. Gartner, & H. Huang (Eds.). *Progress in Location- Based Services 2014* (pp. 171–183). Springer.
Wilken, R. (2014). Places nearby: Facebook as a location-based social media platform. *New Media & Society, 16*(7), 1087–1103. https://doi.org/10.1177/1461444814543997.
Williams, S. (2012). We are here now. *Social media and the psychological city*. Retrieved from http://weareherenow.org/about.html.
Williams, M. J., & Chorley, M. J. (2017). Foursquare. In L. Sloan, & A. Quan-Haase (Eds.). *The SAGE Handbook of Social Media Research Methods* (pp. 610–626). London: SAGE Publications Ltd.
Yang, L., Sun, T., Zhang, M., & Mei, Q. (2012). We know what@ you# tag: Does the dual role affect hashtag adoption? *WWW'12 Proceedings of the 21st International Conference on World Wide Web* (pp. 261–270). . https://doi.org/10.1145/2187836.2187872.