Open Geospatial Data Contribution Towards Sentiment Analysis Within the Human Dimension of Smart Cities

Tiago H. Moreira de Oliveira and Marco Painho

Abstract In recent years, there is a widespread growth of smart cities. These cities aim to increase the quality of life for its citizens, making living in an urban space more attractive, livelier, and greener. In order to accomplish these goals, physical sensors are deployed throughout the city to oversee numerous features such as environmental parameters, traffic, and the resource consumption. However, this concept lacks the human dimension within an urban context, not reflecting how humans perceive their environment and the city's services. In this context there is a need to consider sentiment analysis within a smart city as a key element toward coherent decision making, since it is important not only to assess what people are doing, but also, why they are behaving in a certain way. In this sense, this work aims to assemble tools and methods that can collect, analyze and share information, based on User Generated spatial Content and Open Source Geospatial Science. The emotional states of citizens were sensed through social media data sources (Twitter), by extracting features (location, user profile information and tweet content by using the Twitter Streaming API) and applying machine learning techniques, such as natural language processing (Tweepy 3.0, Python library), text analysis and computational linguistics (Textblob, Python library). With this approach we are capable to map abstract concepts like sentiment while linking both quantitative and qualitative analysis in human geography. This work would lead to understand and evaluate the "immaterial" and emotional dimension of the city and its spatial expression, where location-based social networks, can be established as pivotal geospatial data sources revealing the pulse of the city.

Keywords Smart cities · Open geospatial data · User generated spatial content (UGsC) · Sentiment analysis · Ambient geographic information (AGI)

T. H. M. de Oliveira (✉) · M. Painho
NOVA IMS, Universidade NOVA de Lisboa, Lisbon, Portugal
e-mail: tiago.oliveira@novaims.unl.pt

M. Painho
e-mail: painho@novaims.unl.pt

A. Mobasheri (ed.), *Open Source Geospatial Science for Urban Studies*, Lecture Notes in Intelligent Transportation and Infrastructure,
https://doi.org/10.1007/978-3-030-58232-6_5

1 Introduction

Cities, as government units, are becoming increasingly larger, more complex and more important as the population ranks of urban areas swell with ever increasing speed. According to the United Nations Population Fund, 2008 marked the year when more than 50% of all people, 3.3 billion, 23 lived in urban areas. By 2030 this number is expected to increase to 5 billion [1].

In this sense, a smart city's main purpose is to increase the quality of life for its citizens and to make the city more attractive, livelier, and greener [2]. To do so, physical sensors are deployed throughout the city to monitor various features such as environmental parameters (weather, pollution, etc.), traffic, and the consumption of resources [3]. This live state, however, includes only measurable quantities and disregards how the citizens feel. It is likely that correlations exist between the emotional states of the citizens and relevant statistics like well-being of the city inhabitants or quality of living [4]. When urban planners use the collected data to optimize parts of the city, the emotional state of the inhabitants can thus serve as valuable implicit feedback.

This gives rise to the vision of an emotion-aware city with the ability to understand and utilize the emotional states of its citizens to enable improved planning and decision making, in which urban citizens can be called upon to act as active sensors, sharing their spatiality. Therefore, the use of User Generated spatial Content (UGsC), including social media information, could truly lead to better urban planning [5–7].

This chapter aims to deliver a literature review regarding the definition of a smart city and the importance of its human/emotional component, while presenting a methodological approach towards sentiment analysis and mapping, to assess people's emotional response to their environment. This methodology, based in an Ambient Geographic Information (AGI) approach, gathers open source geospatial data through User Generated spatial Content (UGsC) retrieved from Social Media (Twitter) related with people's perception and feelings, and therefore characterize its emotional dimension.

This work is structured in the following sections: the current section will provide some insights regarding the definition of smart cities and concepts such as crowdsourcing, UGsC and AGI; Sect. 2 will address issues regarding emotion and sentiment analysis and its relevance smart cities; on Sect. 3, the methodological approach to harvest tweets with sentiment content will be presented; on Sect. 4, some preliminary results will be presented; finally, the last section will present some of the possible outcomes of the research along with proposed future work.

1.1 In the Pursuit of Smart Cities

Nowadays, since there is a rapid increase of urban population worldwide, cities face a variety of risks, concerns, and problems, namely: physical risks such as deteriorating conditions in air and transportation, and economic risks such as unemployment [1]. This unprecedented rate of urban growth creates an urgency to find new approaches and methods to tackle these challenges. Smart cities could be one of the reactions for these issues since they aim to improve "urban functions" and services provided to the population not only through innovation, but also through the combination of networks (especially wireless), sensors/actuators and the active commitment of citizens [8].

But how exactly can we define them? Table 1 presents some working definitions of what is classified as a smart city. While some authors [9] and focus on the use of smart computing technologies, others [10] highlight the performance of a smart city in economy, people, governance, mobility, environment, and living. In a Project of the Natural Resources Defense Council [11] the emphasis is given to the positive outcomes made by being "smarter". However, most of the definitions stress technologies [12–14], in which the city is dependent by the sourcing of real-time real-world

Table 1 Smart city definitions (Author)

Author	Definition	Smart city dimension focus
Washburn, D. (2009) [9]	*The use of Smart Computing technologies to make the critical infrastructure components and services of a city—which include city administration, education, healthcare, public safety, real estate, transportation, and utilities—more intelligent, interconnected, and efficient*	Use of smart computing technologies
Giffinger, R. (2007) [10]	*A city well performing in a forward-looking way in economy, people, governance, mobility, environment and activities of self-decisive, independent and aware citizens*	Smart city impacts
Appleyard, B. (2007) [11]	*A city striving to make itself "smarter" (more efficient, sustainable, equitable, and livable)*	Smart city impacts
Hall, R. (2000) [12]	*A city that monitors and integrates conditions of all of its critical infrastructures, including roads, bridges, tunnels, rails, subways, airports, seaports, communications, water, power, even major buildings, can better optimize its resources, plan its preventive maintenance activities, and monitor security aspects while maximizing services to its citizens*	Technology

(continued)

Table 1 (continued)

Author	Definition	Smart city dimension focus
Harrison, C. (2010) [13]	*An instrumented, interconnected, and intelligent city. Instrumentation enables the capture and integration of live real-world data through the sensors, kiosk, meters, personal devices, appliances, cameras, smart phones, implanted medical devices, the web, and other similar data-acquisition systems, including social networks as networks of human sensors. Interconnected means the integration of those data into an enterprise computing platform and the communication of such information among the various city services. Intelligent refers to the inclusion of complex analytics, modeling, optimization, and visualization in the operational business processes to make better operational decision*	Technology
Rios, P. (2012) [14]	*A city that gives inspiration, shares culture, knowledge, and life, a city that motivates its inhabitants to create and flourish in their own lives*	Technology
Partridge, H. (2004) [15]	*A city where the ICT strengthen the freedom of speech and the accessibility to public information and services*	Human dimension

data from both physical and virtual sensors. Such data may be interconnected across multiple processes, systems, organizations, industries, or value chains. The combination of instrumented and interconnected systems effectively connects the physical world to the virtual world. Few authors [15] shed light on the human dimension of a smart city.

There are four dimensions in which a smart city primarily operates, namely: intelligent city (its social infrastructure), the digital city (informational infrastructure), the open city (open governance) and the live city (a continuously adaptive urban living fabric) [8]).

Figure 1 identifies and clarifies the main conceptual variants and core factors of a smart city according to Nam and Pardo [1]: Technology (infrastructures of hardware and software); People (creativity, diversity, and education); and Institution (governance and policy).

The above mentioned human component of the smart city relates to the idea that a smart city is also a living urban fabric that is continuously being reshaped as is

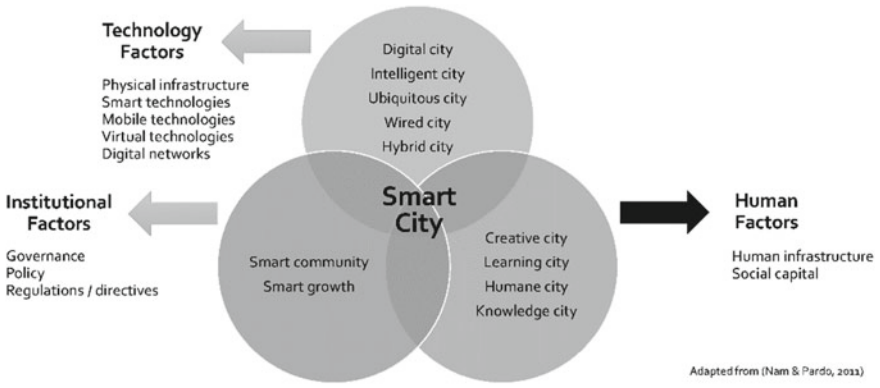

Fig. 1 Fundamental components of a smart city

adaptive to change [8]. Essentially, urban functions should not be disconnected from urban planning, and its immaterial dimensions, such as security perception, sense of belonging and joy, in which could affect positively or negatively health and urban population well-being [16].

In this context, there's a need to consider affect and emotion within a smart city, as a key element towards rational decision making [17]. Since emotion is a central component of human behavior and, for a city to be truly "smart", it is important not only to assess what people are doing, but also, why they are behaving in a certain way [18], as considering emotional states is essential for achieving real-time judgment and perceived life satisfaction [19]. With this information, city planners can make use of the gathered affective data to detect positive or negative trends developing in the city, managing to take early countermeasures. Answering subjective questions such as "which part of the city is the best?", requires affect and emotion [19].

But more important for a smart city is its capability to capture the sense of places. A city is not a machine, but rather made by people local actions and feelings. This could not be captured and represented without active citizens sensors (Volunteered Geographic Information [20], Ambient Geographic Information [21], crowdsourcing [22] connected to location based-social networks [23]. Due to the relative pre-eminence of place to understand urban environment, and following Professor Michael Goodchild insights [24], it is time to move from a space-based to a place-based geospatial infrastructure.

Smart urban solutions should be built on the vision of citizens as active sensors on one hand, and on the other hand on spatial enablement of citizens via social network. These kinds of solutions have also to be built on the potentials offered both by embedded sensors to crowdsource the process of collecting geo-referenced information about places in the city. These constructs gave rise to the vision of an emotion-aware city, in which the "immaterial" (and human) dimension is the main component, and the main focus is given on the need that smart cities have to assess their citizen's feelings, perception and well-being [25]. An emotion-aware city can be defined by its capacity on interpreting and harnessing the affective states of its citizens [26].

1.2 Crowdsourcing, UGsC and AGI

As explained by Roche [23], an active and engaged citizen is indeed the main driving force of a "smart city". Nowadays, there is a growing amount of location-based contents generated by connected "*produsers*", mainly equipped with smartphones. The exponential growth of ambient geographic information through social networks became the basic feature of a spatially enabled society, in which it behaves as a vessel where millions of people share their current thoughts, observations and opinions, showing to provide more reliable and trustworthy information than traditional methods like questionnaires and other sources [27]. A spatially enabled citizen is explained through his ability to express, formalize, equip (technologically and cognitively), and (un)consciously activate an efficiently use of his spatial skills [23].

Social media generated from many individuals plays a greater role in our daily lives while providing a unique opportunity to gain valuable insight on information flow and social networking within a society [28]. Through data collection and analysis of its content, it supports mapping and understanding of the evolving human landscape [21]. In this context, mobile technologies are definitely a valuable tool for collecting affective data in the context of an emotion-aware city [29], since they can simultaneously collect both spatial location and the user posts, which should contain emotion or mood content.

Harvesting this ambient geospatial information provides a unique opportunity to gain valuable insight on information flow and social networking within a society, support a greater mapping, understand the human landscape and its evolution over time [21]. This emergence of AGI represents a second step in the evolution of geospatial data availability, following on the heels of Volunteered Geographic Information (VGI) [30], in which harvesting and analyzing such ambient information represents a substantial challenge, needing new skillsets as it resides at the intersection of disciplines like geography, computational social sciences, linguistics and computer science [21].

Ambient Geographic Information (AGI) can be defined as geographic information that can be harvested from social media feeds. Since VGI is primarily about crowdsourcing with specific tasks outsourced to the public at large, AGI is different, since it focuses on "crowd-harvesting", with the general public broadcasting information that can be harvested in a meaningful manner [30]. Given that, the emergence of social media participation and information sharing is bringing forward a different model of geospatial information contribution, meaning that nowadays users' intent is not to directly contribute geospatial data (e.g. a map), but rather to contribute information (e.g. a geotagged picture from a family vacation, or text describing a planned event) that happens to have an associated geospatial component and a corresponding footprint.

Assembling and analyzing AGI provide us with unparalleled insight on a broad variety of cultural, societal, and human factors, particularly as they relate to human and social dynamics, for example: (1) mapping the manner in which ideas and information propagate in a society, information that can be used to identify appropriate strategies for information dissemination during a crisis situation. (2) Mapping people's opinions and reaction on specific topics and current events, thus improving our ability to collect precise cultural, political, economic and health data, and to do so at near real-time rates. (3) Identifying emerging socio-cultural hotspots.

Distinct from VGI where people are acting *only* as sensors, in AGI they are also the observations from which we can get a better understanding of various parameters of the human landscape. On the other hand, AGI focuses upon passively contributed data, unlike VGI.

Unfortunately, the geolocation features offered by social networks are not very commonly used, which generates a great variation in the availability of geolocation information. As an example, the most popular social networks, such as Instagram, Twitter and Facebook, do not have much posts with location attached to them, with respectively 30%, 15% and 4% [31].

Since social media has rose in popularity and scale, there is a growing need to extract useful information from huge amounts of data. Social networks like Twitter, with an estimated community of 332 million worldwide and more than 400 million posts every day [32], can potentially serve as a valuable information resource for various applications.

There are some case studies in which social media contributions lead to perceptions and images of real-world phenomena. Since twitter contains a large corpus of public real-time data that, it has been used successfully in studies regarding disasters and emergencies [33], public health monitoring [34], events exploration [35], among others.

The rise of social media and the ability for analysis raises several concerns with respect to the suitability of traditional mapping and GIS solutions to handle this type of information [22, 36]. It is now possible to map abstract concepts like the flow of information in a society, contextual information to place and linking both quantitative and qualitative analysis in human geography [21].

In a sense one could consider AGI to be addressing the fact that the human social system is a constantly evolving complex organism where people's roles and activities are adapting to changing conditions and affect events in space and time. By moving beyond simple mashups of social media feeds to actual analysis of their content we gain valuable insight into this complex system [21].

2 Sentiment Analysis

Mapping emotion builds on a tradition of studies in cognitive mapping, evaluative mapping, environmental preference and environmental affect [37], adding an approach in which people experience, evaluate, and describe their environment in situ through social media.

As stated before, one of the most easily accessible public data sources that can be used to detect emotion are social networks. People use them to share their opinions and/or express their feelings. It has been shown that users of social media tend to post authentic and reliable information about themselves [27]. This type of approach brings an opportunity to detect sentiment.

Approaches based on text analysis are suited for extracting moods and emotions from social networks, since linguistic characteristics analysis on the written posts made by individuals on their pages can be used to infer negative and positive moods [38]. By applying this technique to Twitter/Facebook/Flickr/Instagram update messages allows to reveal information about public moods and emotions [39], additionally Twitter "hashtags" can be harnessed to extract individual mood states [40].

As explained by Dramstad [41], most people, if questioned, will have an opinion as to whether a landscape is aesthetically pleasing or not, and how everyday landscapes reflect in the well-being of people is receiving increased focus in research. On the other hand, Weinreb [37] advocates that personal associations are a primary example of intangible and subjective feelings, related much more to memory than to anything immediately visual. Positive personal associations stemmed from memories about a range of personal experiences with friends and family and attachment to the place or locality. Negative personal associations can be articulated as disappointment regarding the ruined or unrealized potential of a space, often tied to a sentiment that municipal leaders could failed to follow through on promises to complete development, livability, or beautification projects.

A mental map refers to one person's point of view perception of their own world, and is influenced by that person's culture, background, mood and emotional state, instantaneous goals and objectives [31]. For instance, if we move along the streets of any city in a rush, trying to find a certain type of shop or building, our experience will be different than the one we would have had if we were searching for something else. There are examples [42] in which a concept of a pedestrian routing system was developed, who enabled users to consider factors and quantify their influences on the extraction of walking routes, factors such as street length, greenness, sociability, and quietness.

There is a multitude of reasons why a pedestrian may choose to avoid areas with negative affect: emotions like fear and anger indicate danger and could be avoided by travelers feeling afraid. Stress may be felt in areas of high traffic or crowdedness which are undesirable for pedestrians. The emotion disgust indicates places that are unsuitable for relaxation. Likewise, there are reasons to seek areas with positive affect: a detour through a relaxing area can be acceptable when someone is feeling stressed. The emotion relaxed may also be correlated with higher safety to walk in an area. Furthermore, someone may want to seek locations with increased surprise when they are curious and, in the mood, to explore. Reasons to seek or avoid areas with a certain affective state are as manifold and personal as affect itself [43].

Likewise, a person's socio-economic status, cultural ties, and past experiences influence how people perceive environmental quality. In the case of tourism, people using these areas can differ in many ways, including their personal characteristics and perception about the recreation environment [44].

Since urban planning is a process of regulating land use to optimize aspects like resource consumption, transportation and safety in the face of rapid urban growth, negative trends in the city should be recognized as early as possible, by using citizens as sensors [4, 45]. If problems are identified quickly, early countermeasures can be taken. For instance, there are problems that are obvious to inhabitants, but are hard to measure with statistics and facts. For instance, even without available crime records, people have a good intuition of the safety of a neighborhood. Similarly, the amount of visual degradation and littering is hard to measure but can be judged subjectively and trigger an emotion. These two examples state that affective states can be harnessed to understand how citizens truly perceive their environment. These subjective feelings may differ greatly from measurable statistics.

3 Materials and Methods

In this chapter we introduce the practical component of this study in which the main purpose is to present an AGI-based methodology, designed towards sentiment analysis, which aims to create tools and methods that can collect, analyze and share information, based on Open Source Geospatial technology and User Generated spatial Content (UGsC) linked with social networks and media.

3.1 Twitter as an UGsC Data Source

For this study, the data is harvested through Twitter, a social media service where both mobile and web users post and read short messages of up to 280 characters, called tweets, that can contain links and media content, such as pictures and videos [46]. Twitter is a popular microblogging service, with more than 332 million monthly active users, 400 million tweets sent per day and 80% of users using mobile platforms [32]. The introduction of the hashtag "#" syntax facilitates discussion on a specific topic and offers an important filter system for a specific subject [47]. Tweets can have both primary and secondary status, the first when the user posts a message and the second when the tweet is a reply to another user, or a retweet broadcasting of a prior message written by another user [48].

Fig. 2 Example of a geolocated tweet with precise coordinates and textual description

Usually, geolocation information in tweets can be provided directly by the contributing bloggers, if they decide to make this information available, or it can be deduced from IP addresses using any of the IP geolocation solutions [49].

For our approach we focus on geolocation information either contributed directly by the user or provided through the client application. This geolocation information is available through precise coordinates (as shown on the upper section of Fig. 2), or in a descriptive manner (similar as the example on the lower section of Fig. 2).

On the first example, at the top, we see the tweet as it appears on a follower's stream, and this is the Twitter message that is displayed in the browser. At the bottom we see highlights from the information retrieved through the search API for this particular tweet, including its geolocation information (marked by the box) in the form of precise coordinates. The coordinates correspond to a location in the Fairview borough of New Jersey. The second example contains the descriptive geolocation information recovered for another tweet similar to the one presented previously.

The study site for this work is continental/mainland Portugal, located on the southwestern side of Europe, from which geolocated tweets were harvested.

3.2 *Open Source Geospatial Methodology*

To harvest information from social media feeds is essentially a web-mining process. Mainly, it entails in general three operations [50]: extracting data from the data providers (various social media servers) via application programming interfaces

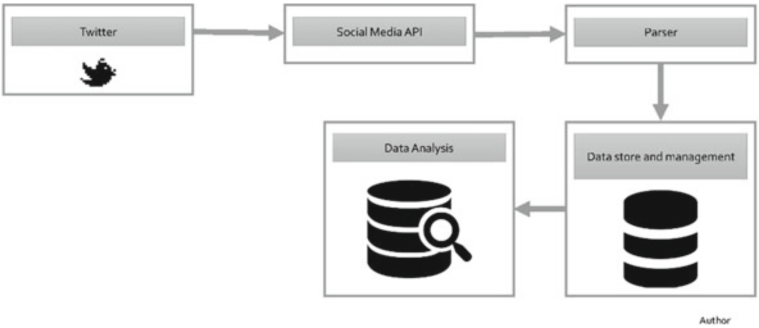

Fig. 3 General system architecture to harvest information from social media

(APIs); parsing, integrating, and storing these data in a resident database; and then analyzing these data to extract information of interest.

Data parsed from diverse sources are integrated by the parser, which organizes data from diverse sources (Fig. 3) into common categories like time of submission, username, originating location, keywords, as well as service-specific information (e.g. content, links to actual files). It allows us to establish an integrated multi-source local database that can be used to perform analysis of the harvested data that is not supported by the provider database interface (e.g. statistics on user activities) for various use cases.

The used methodological approach is as represented below in Fig. 4 and will be explained through the present section. The emotional states of citizens are sensed using Twitter, by extracting features (location, user information, posts, pictures and videos) and applying machine learning techniques, such as natural language processing, text analysis and computational linguistics.

The used method to extract tweets was through the Tweepy Python Library 3.0, which uses the Twitter Streaming API. This Open Source Python Library permits the extraction of several information related with the tweet, such as location (though precise coordinates and/or placenames/description), user profile information and the tweet content itself. The developed twitter stream service is enabled through a custom Python script made by the authors in which all the geolocated tweets from the study area are dynamically stored on a .JSON file format within a dedicated technological infrastructure with virtually zero downtime.

On our approach firstly a set of un-structured text documents is collected. Then, the pre-processing for the documents is performed to remove noise and commonly used words, stop words, stemming. This process produces a structured representation of the documents known as Term- document matrix, in which, every column represents a document and every row represents a term occurrence throughout the document. The final step is applying data mining techniques such as clustering, classification, association rules to discover term associations and patterns in the text and then, finally, visualizing these patterns using Geographic Information System tools.

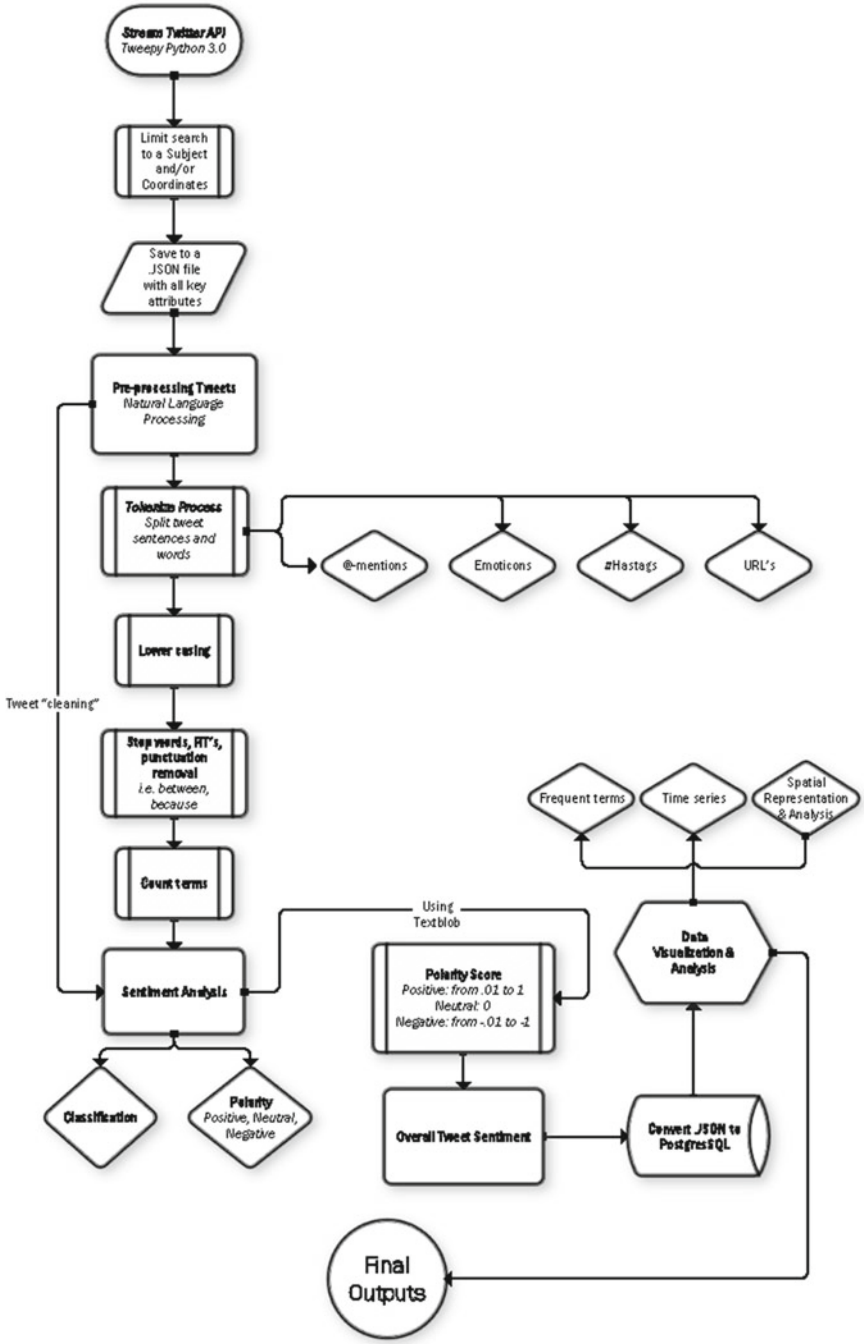

Fig. 4 Tweet harvesting and analysis methodological approach

According to Younis [32], text Mining is defined as the automated process of detecting and revealing new, uncovered knowledge and inter-relationships and patterns in unstructured textual data resources.

In this sense, and before the tweets can be used as training data, they need to be pre-processed. Since tweets are short text messages that are no longer than 280 characters and are often written on mobile devices, they typically contain many abbreviations, colloquial expressions and non-standard words [19]. Additionally, the language employed in Social Media sites is different from the one found in mainstream media and the form of the words employed is sometimes not the one we may find in a dictionary. Further on, users of Social Media platforms employ a special “slang” (i.e. informal language, with special expressions, such as “lol”, “omg”), emoticons, and often emphasize words by repeating some of their letters [51]. Mainly, during the tweet pre-processing the idea is to remove the variety from the messages which constitutes noise in the training data.

The tweet pre-processing stage is comprised by the steps depicted in Fig. 5, proceeding as follows:

1. Repeated punctuation sign normalization: In the first step of the pre-processing, repetitions of punctuation signs (“.”, “!” and “?”) are detected. Multiple consecutive punctuation signs are replaced with the labels “multistop”, for the fullstops, “multiexclamation” in the case of exclamation sign and “multiquestion” for the question mark and spaces before and after.
2. Emoticon replacement: The emoticons found are replaced with their polarity (“positive”, “negative” and “neutral”).
3. Tokenization and lower casing process: Lower case and split the tweet content into tokens, based on spaces and punctuation signs.

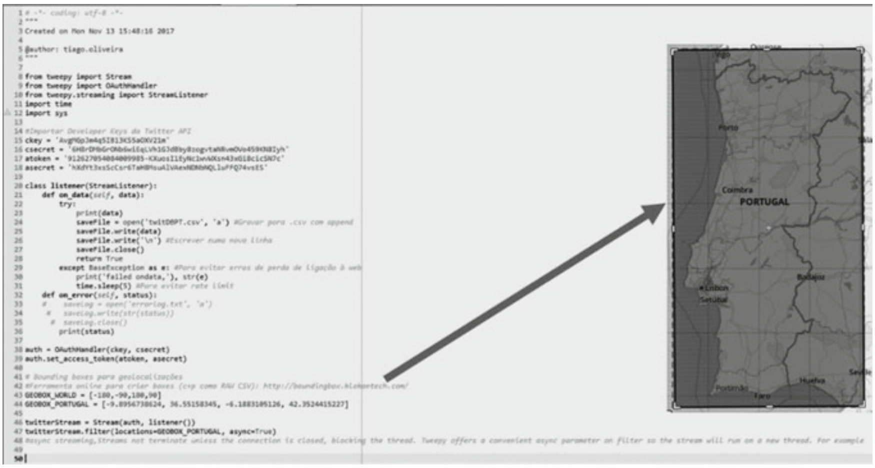

Fig. 5 Python script and study area bounding box

Balahur [51] framed sentiment analysis as the Natural Language Processing (NLP) task dealing with the detection and classification of sentiments in texts. In the literature, there is no standard method for mining and analyzing so-cial media business data [32]. On our approach, Sentiment analysis is executed with Textblob, an Open Source Python Library that processes textual data. It provides a simple API for diving into common NLP tasks such as: noun phrase extraction, part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more. In order to perform sentiment analysis with Textblob, there is the need to use also other Python libraries, namely tweepy and NLTK.

When this process is finished, the .JSON file which contains the tweets is converted to PostgresSQL database and therefore the extracted information can be used on any GIS application, both desktop and server oriented.

4 Preliminary Results and Discussion

In this chapter we present and discuss some preliminary results obtained by applying the methodology described in the previous section.

The implemented twitter streaming service so far extracted approximately 8 million georeferenced tweets, by using the below custom Python script (Fig. 5) whom harvested tweets from continental/mainland Portugal.

The harvested information was stored on a PostgresSQL database which contains three separated but interconnected tables, namely:

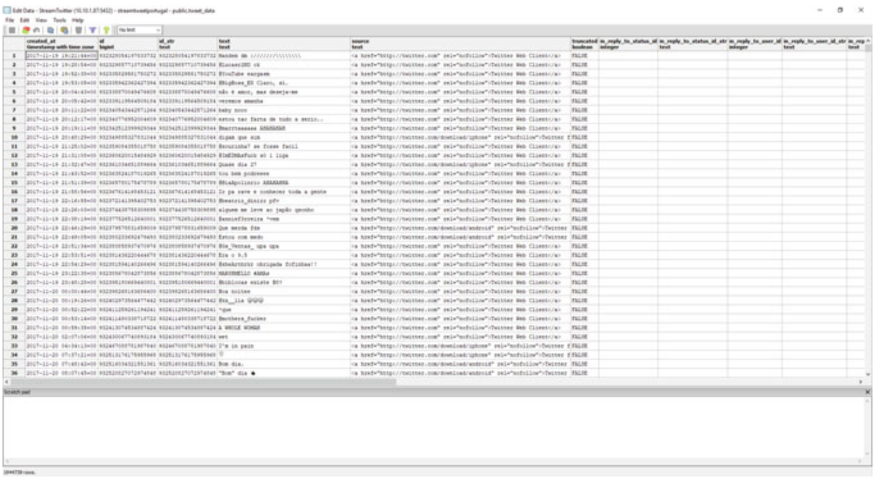

Fig. 6 Tweet data table example

1. Tweet_data (Fig. 6): which is the “parent” tweet object for the other tables, since it includes fundamental attributes such as id, created_at and text (the tweet content).
2. User_data (Fig. 7): it contains the public Twitter account metadata and describes the user which posted the tweet.
3. Place_data (Fig. 7): includes the tweet geo-tag. Tweet location can be an exact point location, or a Twitter place with a specific bounding box whose describes a larger area ranging from a venue to an entire region (Fig. 8).

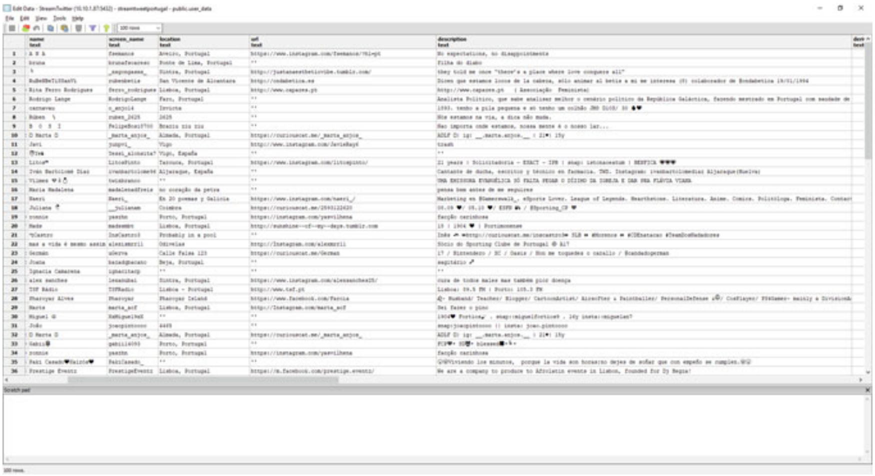

Fig. 7 User data table example

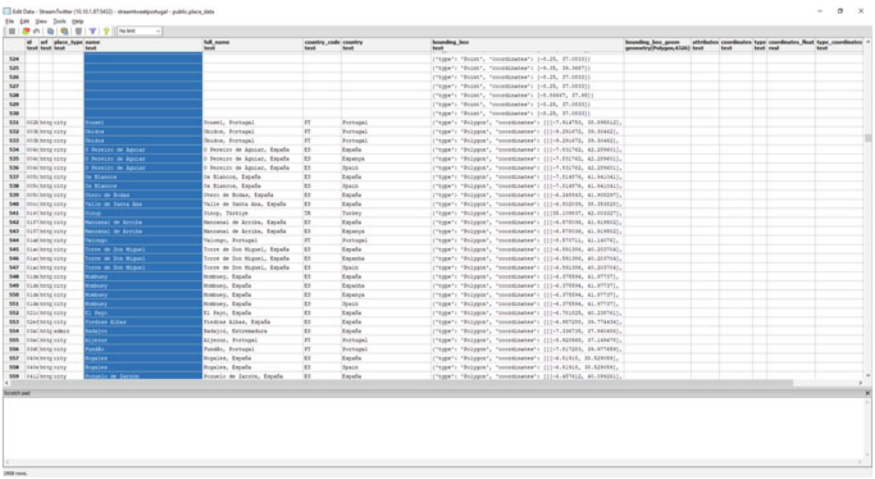

Fig. 8 Place data table example

Sentiment Analysis	
Text	If that is not cool enough for you than that is a you problem.
Polarity	-0.0875
Subjectivity	0.575
Classification	neg
P_Pos	0.344455873
P_Neg	0.655544127

Fig. 9 Sentiment analysis score example

As stated previously, with the use of NLP tools, such as Textblob (Python) on our methodology there was possible to perform sentiment analysis to each harvested tweet in which subjective information was identified and extracted. By analyzing the language used in the text, our custom Python script classified its polarity, measuring the negativity, neutralness, or the positivity of the tweet text/content, while giving it a score (from -0.01 to 1) representing if the tweet/post is negative, neutral or positive, as demonstrated on Fig. 9.

However, in Twitter, sometimes people use hashtags to notify others of the emotions associated with the message they are tweeting. However, tweets as the one showed on the fourth example (Table 2), prove that sometimes reading just the message before the hashtag does not convey the emotions of the tweeter. Here, the hashtag provides information not present (implicitly or explicitly) in the rest of the message.

On the other hand, there are also tweets, such as those shown in examples 5 and 7 that do not seem to express the emotions stated in the hashtags. This may occur for many reasons including the use of sarcasm or irony, which Textblob is not able to detect. Additional context is required to understand the full emotional impact of many tweets. Tweets tend to be very short, and often have spelling mistakes, short forms, and various other properties that make such text difficult to process by

Table 2 Smart city definitions (Author)

Tweet examples
1. RIP Ziggy Bowie Stardust…#sadness
2. My amazing memory saves the day again! #joy
3. Just saw the most awful dress ever! #Disgust
4. John used my photo on facebook. #anger
5. School is very boring today:/ #joy
6. Hope I get to school in time… #fear
7. Great morning stuck in traffic with a flat tire #love
8. I love when I pick up my car and it’s out of gas #surprise

natural language systems. Further, it is as well probable, that only a small portion of emotional tweets are hashtagged with emotion words [52].

5 Conclusions and Future Research Directions

Sentiment Analysis allows researchers to visually discern areas of strong feelings, either good or bad. This multidisciplinary approach can exhibit aggregations of positive ratings, negative ratings, or in some cases, a mixture of strong positive and negative ratings in the same place. Some other possible outputs are related with the identification of emotional patterns—those spaces where, at a specific or recurring time, a certain emotion is expressed powerfully and abundantly. It can lead us to some questions: do emotional landmarks change over time? Do they change according to the observer? To language? To the time of day, week, month or year? Additionally, several profiles of users can be established, based upon social-demographic characteristics, such as: gender; age; education level; motive of trip (leisure, business, e.g.); "level of acquaintance" of the place (old-timer, new-comer, tourist); origin (country and/or city).

This *emotional sensing* can be directed towards any topic or subject, such as: emergency scenarios such as revolts, riots or natural disasters; art and tourism; city planning and safety; entertainment and consumption, proving its relevance and replicability towards any other issue or thematic.

Contributing to cognitive mapping, emotional maps enable researchers to share a participant's position and views of the landscape as he or she articulates emotions and memories related to those views. Replicable in any setting, this technique could be used to create and maintain spaces that are attractive, inviting, and emotionally pleasing to a variety of users. These geospatial practices could highlight how emotions, subjectivities and spaces are mutually constitutive in particular places and at particular times, suggesting that people's shared feelings about specific places are influenced by the particular physical properties and characteristics of a given place, since this technique could be used to create and maintain spaces that are attractive, inviting and emotionally pleasing to a variety of users.

In this context the power of harvesting Open Source Geospatial data and UGsC stems from gaining a deeper understanding of groups rather than looking at specific individuals. As the popularity of social media is growing exponentially we are presented with unique opportunities to identify and understand information dissemination mechanisms and patterns of activity in both the geographical and social dimensions, allowing us to optimize responses to specific events, while the identification of hotspot emergence helps us allocate resources to meet forthcoming needs.

By engaging an emotion-aware city, new forms of communication can be generated. Traditionally, the choice of partners for online group communication is either based on pre-existing relationships or on similar interests or location [52]. In an emotion-aware city, communication groups can be formed spontaneously, based not only on a topic, but also on location and matching emotional state. These types of interactions can start interesting discussions about controversial projects or places within the city, since personal and sensitive issues are best shared with those who fell the same about a specific area of the city, creating participatory movements.

However, uncertainty over the data quality of UGsC (and VGI in general), is expected to be faced in this work. While using user generated information is common to struggle with positional accuracy, thematic accuracy, completeness, temporal quality and logical consistency [53, 54]. On this work it is expected to face some of those.

Besides data quality assessment, this work aims to be improved with the following developments:

1. Implementation of visual outputs of the gathered information and analysis, through maps, by producing alternative representations of space based on individuals' georeferenced experiences, thoughts and emotion;
2. Add other social media sources—as Facebook, Flickr and Instagram—and evaluate their potential as *emotional data sources*;
3. Analyze the possibility of replicate the same proposed methodology in other languages, besides English UGsCs;
4. How to understand and interpret the use of sarcasm or irony in tweets, as exemplified in Table 2;
5. Compare information retrieved from UGsC (subjective observations) with experimental data (objective measurements, such as socio-demographic statistics about a specific city), evaluating which can truly characterize and share the emotional dimension of the city;
6. Finally, to assess if there is a strong correlation between touristic sites and the *emotional landmarks* within the city, which can be defined as *emotional hot spots*.

Lastly, and revisiting the live city dimension of the smart city, it can greatly benefit from Open Source Geospatial Science, particularly from UGsC. The physical and *senseable* structure of a smart city can be analyzed through UGsC, since it provides innovative, creative, deliberative, uncertain, multi-actor, multi-scale, and multi-thematic methods and tools [25, 55].

An emotion and sentiment mapping methodology could lead to understand, assess and evaluate the *immaterial* and emotional dimensions of the city and its spatial expression. Open Source Geospatial Science can truly support the development of the intelligent city [56], due to crowdsourcing, Volunteered Geographic Information (VGI), Ambient Geographic Information (AGI), including location-based social networks which stand out as key geospatial data sources indicative of the pulse of the city.

References

1. Nam, T., Pardo, T.A.: Conceptualizing smart city with dimensions of technology, people, and institutions. The Proceedings of the 12th Annual International Conference on Digital Government Research.Challenging Times—dg.o '11, p 282 (2011)
2. Kehoe, M., Nesbitt, P.: Smarter cities series : a foundation for understanding IBM smarter cities. IBM J. Res. Dev. (2010)
3. Caragliu, A., Del Bo, C., Nijkamp, P.: Smart Cities in Europe. J. Urban Technol. **18**(2), 65–82 (2011)
4. Lathia, N., Quercia, D., Crowcroft, J.: The hidden image of the city: sensing community well-being from urban mobility. Pervasive Comput., 1–8 (2012)
5. Resch, B., Summa, A., Sagl, G., Zeile, P., Exner, J.: Urban emotions—geo-semantic emotion extraction from technical sensors, human sensors and crowdsourced data. In: Gartner, G., Haosheng, H. (eds.) Progress in Location-Based Services, pp. 199–212. Springer International Publishing, Switzerland (2014)
6. Exner, J.: Smart planning & smart cities. In: Corp 2014, vol. 8, pp. 603–610
7. Exner, J., Zeile, P., Streich, B.: Monitoring laboratory spatial planning: new benefits and potentials for urban planning through the use of urban sensing, geo- and mobile web. In: 12th International Conference on Computers in Urban Planning and Urban Management (CUPUM), pp. 1–18 (2011)
8. Roche, S.: Geographic information science I:Why does a smart city need to be spatially enabled? Prog. Hum. Geogr. **38**(5), 0309132513517365-, Feb 2014
9. Washburn, D., Sindhu, U.: Helping CIOs Understand 'Smart City' Initiatives. Massachusetts , Cambridge (2009)
10. Giffinger, R.: Smart cities ranking of European medium-sized cities, Vienna, Austria (2007)
11. Appleyard, B., et al.: Smart Cities: Solutions for China's Rapid Urbanization, New York (2007)
12. Hall, R.E., Bowerman, B., Braverman, J., Taylor, J., Todosow, H.: The vision of a smart city. In: 2nd International Life Extension Technology Workshop, p. 7 (2000)
13. Harrison, C., et al.: Foundations for smarter cities. IBM J. Res. Dev. **54**(4), 1–16 (2010)
14. Rios, P.: Creating 'The Smart City'. University of Detroit Mercy (2012)
15. Partridge, H.: Developing a human perspective to the digital divide in the smart city. In: ALIA 2004 Biennial Conference, Challenging Ideas, p. 7 (2004)
16. Schipperijn, J., Stigsdotter, U.K., Randrup, T.B., Troelsen, J.: Influences on the use of urban green space—a case study in dense, Denmark. Urban For. Urban Green. **9**(1), 25–32 (2010)
17. Naqvi, N., Shiv, B., Bechara, A.: The role of emotion in decision making a cognitive neuroscience perspective. Curr. Dir. Psychol. Sci. **15**(5) (2006)
18. Dolan, R.J.: Emotion, cognition, and behavior. Science **298**(5596), 1191–1194 (2002)
19. Guthier, B., Alharthi, R., Abaalkhail, R., El Saddik, A.: Detection and visualization of emotions in an affect-aware city. In: Proceedings of the 1st International Workshop on Emerging Multimedia Applications and Services for Smart Cities—EMASC '14, pp. 23–28 (2014)
20. Goodchild, M.F.: Citizens as sensors: The world of volunteered geography. GeoJournal **69**(4), 211–221 (2007)
21. Stefanidis, A., Crooks, A., Radzikowski, J.: Harvesting ambient geospatial information from social media feeds. GeoJournal **78**(2), 319–338 (2013)
22. Goodchild, M.F.: Twenty years of progress: GIScience in 2010. J. Spat. Inf. Sci. **1**(1), 3–20 (2010)
23. Roche, S.: Sensing places' life to make city smarter. In: ACM SIGKDD International Workshop on Urban Computing (UrbComp 2012), 12 Aug 2012, Beijing, China (2012)
24. Goodchild MF (2012) Reflections and visions, final keynote presentation. In: Global Geospatial Conference, pp. 1–22 (2012)
25. De Oliveira, T.H.M., Painho, M.: Using ambient geographic information (AGI) in order to understand emotion & stress within smart cities. In: Proceedings of AGILE 2015—18th AGILE International Conference on Geographic Information Science (2015)

26. De Oliveira, T.H.M., Painho, M.: Emotion & stress mapping—assembling an ambient geographic information-based methodology in order to understand smart cities. In: Atas da 10ª Conferência Ibérica de Sistemas e Tecnologias de Informação, vol. II, pp. 2–5 (2015)
27. Marwick AE, Boyd D (2010) I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. New Media Soc 13(1), 114–133 (2010)
28. Kar, B., Ghose, R.: Is my information private? Geo-privacy in the world of social media. ceur-ws.org, no. Webster, pp. 2009–2012 (2010)
29. Liu, Y., Piyawongwisal, P., Handa, S., Yu, L., Xu, Y., Samuel, A.: Going beyond citizen data collection with mapster: a mobile+cloud real-time citizen science experiment. In: 2011 IEEE Seventh International Conference on e-Science Work, pp 1–6 (2011)
30. Crooks, A., Croitoru, A., Stefanidis, A., Radzikowski, J.: #Earthquake: Twitter as a Distributed Sensor System. Trans. GIS **17**(1), 124–147 (2013)
31. Iaconesi, S., Persico, O.: An emotional compass harvesting geo-located emotional states from user generated content on social networks and using them to create a novel experience of cities. In: Proceedings of First International Workshop on Emotion and Sentiment in Social and Expressive Media approaches Perspectives From AI (ESSEM 2013) A Work. International Conference of the Italian Association for Artificial Intelligence, vol. 1096 (2013)
32. Younis, E.M.G.: Sentiment analysis and text mining for social media microblogs using open source tools : an empirical study. Int. J. Comput. Appl. **112**(5), 44–48 (2015)
33. Hossmann, T., Legendre, F.: Twitter in disaster mode: opportunistic communication and distribution of sensor data in emergencies. In: Proceedings of the 3rdExtreme Conference on Communication, No. 1, pp. 1–6 (2011)
34. Aramaki, E., Maskawa, S., Morita, M.: Twitter catches the flu : detecting influenza epidemics using Twitter. In: EMNLP ’11 Proceedings of Conference on Empirical Methods in Natural Language Process, pp. 1568–1576 (2011)
35. Marcus, A., Bernstein, M., Badar, O.: Twitinfo: aggregating and visualizing microblogs for event exploration. In: Proceedings of SIGCHI Conference Human Factors in Computing Systems (2011)
36. Sui, D., Goodchild, M.: The convergence of GIS and social media: challenges for GIScience. Int. J. Geogr. Inf. Sci. **25**(11), 1737–1748 (2011)
37. Weinreb, A.R.: Mapping feeling: an approach to the study of emotional response to the built environment and landscape. J. Archit. Plann. Res. 1–19 (2013)
38. Leshed, G., Kaye, J.: Understanding how bloggers feel: recognizing affect in blog posts. In: CHI’06 Extended Abstracts on Human Factors in Computing Systems, pp. 2–7 (2006)
39. Bollen, J., Mao, H., Pepe, A.: Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. ICWSM, pp. 450–453 (2011)
40. De Choudhury, M., Counts, S., Gamon, M.: Not all moods are created equal! exploring human emotional states in social media. In: The Sixth International AAAI Conference on Weblogs and Social Media, no. 1 (2012)
41. Dramstad, W.E., Tveit, M.S., Fjellstad, W.J., Fry, G.L.A.: Relationships between visual landscape preferences and map-based indicators of landscape structure. Landsc. Urban Plan. **78**(4), pp. 465–474 (2006)
42. Novack, T., Wang, Z., Zipf, A.: A system for generating customized pleasant pedestrian routes based on openstreetmap data. Sensors (Switzerland), 18(11) (2018)
43. Siemer, M., Mauss, I., Gross, J.J.: Same situation–different emotions: how appraisals shape our emotions. Emotion **7**(3), 592–600 (2007)
44. Petrosillo, I., Zurlini, G., Corlianò, M.E., Zaccarelli, N., Dadamo, M.: Tourist perception of recreational environment and management in a marine protected area. Landsc. Urban Plan. **79**(1), 29–37 (2007)
45. Zheng, Y., Liu, Y., Yuan, J., Xie, X.: Urban computing with taxicabs. In: Proceedings of 13th International Conference on Ubiquitous Computing—UbiComp ’11, p. 89 (2011)
46. Reips, U.D., Garaizar, P.: Mining twitter: a source for psychological wisdom of the crowds. Behav. Res. Methods **43**(3), 635–642 (2011)

47. De Longueville, B., Smith, R., Luraschi, G.: OMG, from here, I can see the flames!: a use case of mining location based social networks to acquire spatio-temporal data on forest fires. In: Conference: Proceedings of the 2009 International Workshop on Location Based Social Networks, no. c, pp. 73–80 (2009)
48. Schneider, J., Passant, A., Groza, T., Breslin, J.G.: Argumentation 3.0: how semantic web technologies can improve argumentation. In: Computational Models of Argument Proceedings, COMMA 2010, vol. 1380 (2010)
49. Eriksson, B., Barford, P., Sommers, J., Nowak, R.: A learning-based approach for IP geolocation. Passiv. Act. Meas., 171–180 (2010)
50. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes Twitter users: real-time event detection by social sensors. In: WWW'10 Proceedings of 19th International Conference on World Wide Web, p. 851 (2010)
51. Balahur, A.: Sentiment Analysis in Social Media Texts. Wassa **2013**, 120–128 (2013)
52. Ellison, N.B., Steinfield, C., Lampe, C.: Connection strategies: social capital implications of Facebook-enabled communication practices. New Media Soc. **13**(6), 873–892 (2011)
53. Foody, G., et al.: Mapping and the Citizen Sensor, First. Ubiquity Press Ltd., London (2017)
54. Mocnik, F.B., Mobasheri, A., Griesbaum, L., Eckle, M., Jacobs, C., Klonner, C.: A grounding-based ontology of data quality measures. J. Spat. Inf. Sci. **16**(16), 1–25 (2018)
55. Estima, J., Painho, M.: Investigating the potential of OpenStreetMap for land use/land cover production: a case study for continental Portugal. In: Arsanjani, J., Zipf, A., Mooney, P., Helbich, M. (eds.) OpenStreetMap in GIScience. Springer International Publishing (2015)
56. Jonietz, D., Antonio, V., See, L., Zipf, A.: Highlighting Current Trends in Volunteered Geographic Information, pp. 1–8 (2017)