Sustainable Cities and Society 64 (2021) 102530

Contents lists available at ScienceDirect

Sustainable Cities and Society

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

Social media semantic perceptions on Madrid Metro system: Using Twitter data to link complaints to space

Joaquín Osorio-Arjona [a,*], Jiri Horak [b,1], Radek Svoboda [c], Yolanda García-Ruíz [d]

[a] Department of Geography. Universidad Complutense de Madrid, C/Profesor Aranguren, s/n. 28040, Madrid, Spain
[b] Department of Geoinformatics. VŠB - Technická Univerzita Ostrava, 17.listopadu 15, 708 00, Ostrava-Poruba, Czechia
[c] Department of Computer Science. VŠB - Technická Univerzita Ostrava, 17.listopadu 2172/15, 708 00, Ostrava-Poruba, Czechia
[d] Department of Computer Systems and Computing, Universidad Complutense de Madrid, C/Profesor José García Santesmases, 9, 28040, Madrid, Spain

ARTICLE INFO

Keywords:
Twitter
Public transport
Text mining
Sentiment analysis
Geographically Weighted Regression

ABSTRACT

Social networks are platforms widely used by travelers who express their opinions about many services like public transport. This paper investigates the value of texts from social networks as a data source for detecting the spatial distribution of problems within a public transit network by geolocating citizens' feelings, and analyzes the effects some factors such as population or income have over that spatial spread, with the goal of developing a more intelligent and sustainable public transit service. For that purpose, Twitter data from the Madrid Metro account is collected over a two-month period. Topics and sentiments are identified from text mining and machine learning algorithms, and mapped to explore spatial and temporal patterns. Lastly, a Geographically Weighted Regression model is used to explore the causality of the spatial distribution of complaining users, by using official data sources as exploratory variables. Results show Twitter users tend to be mid-income workers who reside in peripheral areas and mainly tweet when traveling to workplaces. The main detected problems were punctuality and breakdowns in transfer stations or in central areas, mainly in the early morning of weekdays, and affected by density of points of interest in destination areas.

1. Introduction

Mobility is one of the biggest challenges for metropolitan areas. Urban growth entails a rise in mobility demand, meaning an increase in the number of trips, greater diversity of travel motives, motorized transport intensification, and longer, more time-consuming routes (Banister, 2011). The promotion of public transport is vital for cities, since they seek to decongest the traffic and reduce the level of pollution (Hosseini, El-Diraby, & Shalaby, 2018). Hence, public transport represents the main sustainable mode of urban mobility (Chen et al., 2018). The constant increase in mobility demand has led to a rise in congestion in public transport systems. In this paradigm, public transport agencies need to have updated information about the functionality of their services to detect disruptions (Ji, Fu, Self, Lu, & Ramakrishnan, 2018). Citizen's opinions are fundamental for understanding the necessities, motivations and sensibilities of public transport usage, providing useful insight into planning models that seek to respond to these necessities (El-Diraby, Shalaby, & Hosseini, 2019). However, data from traditional sources seems insufficient due to its high cost, low updating frequency, and low spatial and temporal resolution (Gutiérrez-Puebla & García-Palomares, 2016; Miralles-Guasch & Martínez, 2013; Wang, Phillips, Small, & Sampson, 2018).

The large variety, velocity, and volume of new data sources based on Information and Communication Technologies (ICT) are valuable for mobility and land use studies, allowing analysis to be performed on space-time patterns that cannot be studied by traditional means (Gutiérrez-Puebla & García-Palomares, 2016). Public transport agencies have adopted approximations to communicate with ICT users, providing them with information about their services (Manetti, Bellucci, & Bagnoli, 2017). The volunteered information created through social applications establishes a range of opportunities for these agencies, allowing them a better understanding of the necessities and opinions of public transport users (Casas & Delmelle, 2017).

Messages and opinions shared on social networks can be used not only to detect current transport events, but also to create large volumes of useful data for discovering users' opinions, long-term trends, etc.

* Corresponding author.
E-mail addresses: joaquoso@ucm.es (J. Osorio-Arjona), jiri.horak@vsb.cz (J. Horak), radek.svoboda@vsb.cz (R. Svoboda), ygarciar@fdi.ucm.es (Y. García-Ruíz).
[1] https://www.scopus.com/authid/detail.uri?authorId=18036987500.

https://doi.org/10.1016/j.scs.2020.102530
Received 20 March 2020; Received in revised form 18 August 2020; Accepted 28 September 2020
Available online 2 October 2020

Tweets are rich data resources for extracting opinions and feelings (Kocich, 2017). There is a wide range of research in many fields that uses the text included in tweets to obtain valuable results in application fields like disaster management, health management, or traffic management (Steiger, de Albuquerque, & Zipf, 2015). The content of a tweet is difficult to interpret and analyze due its unstructured nature, but recently there have been developments in text mining techniques applied to tweets to interpret them quantitatively (Lansley & Longley, 2016). Text mining needs to work with very short, unstructured messages, containing noises and errors such as casual or non-usual language, abbreviations, symbols, spelling errors or acronyms (Haghighi, Liu, Wei, Li, & Shao, 2018; Hiltz et al., 2014). Besides being low in cost, data produced continually and compiled almost in real time is beneficial for measuring sentiments using social data like tweets over traditional mobility surveys, including being able to observe users' specific necessities regarding a topic, and useful insight about a particular sentiment (Collins, Hasan, & Ukkusuri, 2013).

Although the use of Twitter data for opinion mining is popular in many fields, its use in transport administration, management and planning sector remains limited (Luong & Houston, 2015). This investigation aims to explore the perceptions of Twitter users when they travel in public transport systems, and showcase the utility of Twitter to spatially locate problems within a public transit network, using the Madrid Metro network as a practical case. Twitter data is free, and its high temporal resolution allows the constant updating of results over time. This investigation also seeks to understand the effects some variables (like population, income, density of points of interest, or connections with other public transit services) load over the spatial distribution of problems in the network with the goal of generating useful knowledge for the development of a more intelligent and sustainable public transit network design and planning.

However, geotagged tweets tend to be only about 1 % of the total of a sample (Graham, Hale, & Gaffney, 2014). When studying a determined topic, these samples can be small-sized and mixed with noise from non-relevant tweets for the investigation. Therefore, we propose using the texts of non-geotagged tweets to extract spatial information by geocoding keywords, and then extract the most commented topics and the sentiments from the Madrid Metro system users. As explained later, this approach allows us to obtain less noise and a larger sample of tweets.

Many studies explore factors influencing satisfaction with public transit (e.g. Shen, Xiao, & Wang, 2016; Stathopoulos & Marcucci, 2014; Wong, Szeto, & Yang, 2017) but we know very little about factors influencing comments and emotions in social media. To overcome that, this paper proposes the elaboration of a Geographically Weighted Regression (GWR) model to explore the causality of variables that spatially affect the number of users with negative sentiments, using official data from official sources such as population, income level, or density of points of interest as exploratory variables.

This paper is divided into six sections. Following the introduction, Section 2 summarizes existing literature on investigations using a Twitter semantic and sentimental analysis in the public transport field. Section 3 describes the study area and the data used, while Section 4 defines the methodology. Section 5 explains the results obtained, and Section 6 provides a series of conclusions.

2. Literature review

Text mining and semantic analysis from tweets has been a prolific area of investigation over the last decade thanks to the easy collection of data samples with a wealth of opinions and feelings within a short time. Text mining has been used to map and compare the frequency of feelings in cities during the day (Kocich, 2017; Lansley & Longley, 2016; Steiger, Resch, & Zipf, 2016; Wachowicz & Liu, 2016), to detect possible natural phenomena like hurricanes or earthquakes (Hiltz et al., 2014; Sakaki, Okazaki, & Matsuo, 2010), or to extract the spatial patterns of feelings in different events, such as the 2016 United States Elections (Chin, Zappone, & Zhao, 2016), 2015 baseball games in Boston (USA) (Steiger, Ellersiek, Resch, & Zipf, 2015), 2014 Sochi Winter Olympic Games (Kirilenko & Stepchenkova, 2017), or 2012 London Summer Olympic Games (Kovacs-Gyori, Ristea, Havas, Resch, & Cabrera-Barona, 2018). Another field of growing importance has been studying the perceptions and feelings about green spaces during different times of the day such as urban green spaces and parks in Melbourne (Lim et al., 2018) or in London (Kovacs-Györi et al., 2018). Few papers have used Twitter as a data source for semantic and sentiment analysis in transport management and planning, for example Collins et al. (2013) for Chicago, Schweitzer (2014) for the Philadelphia region, or Luong and Houston (2015) for Los Angeles. Common topics and their spatial patterns were analyzed in Waterloo by Zhang and Feick (2016) and similar semantic and sentiment analyses were conducted for Salt Lake City by Haghighi et al. (2018). Hosseini et al., 2018 and El-Diraby et al. (2019) performed a more detailed semantic and sentiment analyses for Vancouver TransLink, Toronto Transit Commission, and Toronto GO transit.

All these studies challenge similar issues. The first issue is how to assemble an appropriate data set for analysis, usually using temporal, spatial, and topic or keywords filtering. Setting an appropriate temporal range is useful for analyzing specific events like the Olympic games (Kirilenko & Stepchenkova, 2017). For non-event studies, two main approaches can be found: using long-time intervals around 1 year (e.g. El-Diraby et al., 2019; Steiger, Ellersiek, et al., 2015; Lansley & Longley, 2016; Zhang & Feick, 2016) or short intervals of 1–2 weeks (e.g. Casas & Delmelle, 2017; Collins et al., 2013; Haghighi et al., 2018). Long intervals are required to assemble large volumes of data, especially when using Twitter Streaming API which provides only about 1 % of tweets (Haghighi et al., 2018). Longer periods are advantageous to study long-term trends and distributions at the cost of sacrificing homogeneity which is often an issue for transport studies. For example, Lansley and Longley (2016) removed Mondays, Fridays and weekends to obtain a more homogeneous dataset of tweets. Occurrences of transport failures and non-periodic specific conditions will substantially change the frequency and pattern of tweets. Hence, El-Diraby et al. (2019) clustered the days into three categories: normal days; days with disruptions and days with an information surge. Using long periods is not well suited for quick and continuous monitoring and updating of customer sentiments and topics. Thus, we suggest using a 2-months interval instead.

Regarding the spatial filtering of tweets, the majority of studies are focused on selected localities. For transport and urban purposes, results can be better utilized by the public transit providers and urban and transport planners. To filter relevant tweets, the name of the location is typically used frequently with an appropriate buffet to eliminate boundary issues (e.g. Casas & Delmelle, 2017; Haghighi et al., 2018) A boundary box for coordinates is also usually applied (e.g. Kovacs-Gyori et al., 2018; Steiger et al., 2016; Zhang & Feick, 2016). Another strategy is to use more focused filters coupled with the city name and the name of line or rail (Collins et al., 2013; Luong & Houston, 2015).

The majority of investigations (Hiltz et al., 2014; Kovacs-Gyori et al., 2018; Steiger et al., 2016; Sakaki et al., 2010; Steiger, de Albuquerque, et al., 2015; Steiger, Ellersiek, et al., 2015; Zhang & Feick, 2016) harvest only geotagged tweets. As stated earlier, geotagged tweets represent only about 1 % of all tweets posted in the given period. This approach may also be ineffective when studying specific topics such as public transit due to large noise and the inclusion of non-relevant tweets. Zhang and Feick (2016) found over 99 % of tweets from their sample were off-topic. However, analyzing the spatial distribution of topics and sentiments needs a good accuracy of point locations and for that it is necessary to use coordinates (for example, spatiotemporal analysis for park visits by Kovacs-Gyori et al., 2018 or advanced spatiotemporal methods such as Geo-SOM by Steiger et al., 2016). Some scholars also utilize user metadata to recognize user home locations (Kirilenko and Stepchenkova (2017) estimated a place of residence in roughly 35 % of tweets). For public transit analysis, a good approach is detecting the location based on the tweet texts using geocoding (Fojtík, Horák, Orlíková, Kocich, & Inspektor, 2016).

The third form of data filtering is to select appropriate Twitter accounts for the study. Usually, individuals with poor tweeting activity (e.g. 4 tweets in parks according to Kovacs-Györi et al., 2018) or highly active accounts (e.g. users with more than 100 tweets per year in El-Diraby et al., 2019) are removed. For many purposes it is convenient to distinguish residents and "visitors" (e.g. Kovacs-Gyori et al., 2018). Separation of the groups is usually based on temporal or spatiotemporal patterns where residents have some continuous frequency of tweeting (Kovacs-Gyori et al., 2018). More advanced methods combine tweeting activity and distances to potential residences (Kovacs-Györi et al., 2018), analysis of social networks based on graph theory (Hosseini et al., 2018) or analysis of users' accounts (Luong & Houston, 2015). However, sociological profiling of users is still limited and unreliable.

Social media analysis usually contains semantic and sentiment analysis. Semantic analysis enables to capture the meaning and content of the text, which is challenging especially for short unstructured text messages such as tweets. Semantic analyses are usually based on vocabularies (Casas & Delmelle, 2017; Hosseini et al., 2018; Zhang & Feick, 2016) or on unsupervised machine learning approaches (Haghighi et al., 2018; Kovacs-Gyori et al., 2018; Lansley & Longley, 2016; Zhang & Feick, 2016). Some studies use a basic schema with two identified topics such as actual user experience of the transit system and transit agency decisions (Haghighi et al., 2018) where only the first one is suitable for analyzing users' feedback on the quality of public transit service. Other studies applied a more detailed scheme with e.g. eight topics for customer satisfaction issues. El-Diraby et al. (2019) found most concerns were related to travel time and service delivery (on both weekdays and weekends), safety and security, followed by information availability and spatial availability (important only on weekends). Casas and Delmelle (2017) also designed eight topics (routes, stations, buses, infrastructure, accidents, safety, exclusion, technology) and found buses and safety to be the most referenced categories (the main concerns were bus riders' behavior, inappropriate use of dedicated seats, crimes and fights). Semantic Twitter data has also been used for the creation of predictive and event-detection models (Ji et al., 2018; Zhañay, Cordero, Cordero, & Urigüen, 2019). Spatial distribution of topics was investigated only in a few non-transport focused studies. For example, Lansley and Longley (2016) describe the dominancy of leisure topics in tweets written at home, while business and information topics prevail in tweet from not-domestic buildings.

Sentiment analysis enables the identification of appreciations and feelings in texts. Sentiment analysis is based on two main techniques: the use of vocabulary where each word is assigned a value, and machine learning techniques, which use counting methods to determine the sentiment of a body of text (Collins et al., 2013). Sentiment analysis is performed in only one language, mainly in English. Some exceptions are bilingual sentiment evaluation by Kirilenko and Stepchenkova (2017) and multilingual sentiment mapping using Google Translate API by Kocich (2017). Usually, tweets get a positive, negative or neutral score, but some studies classified only negative and positive sentiment for tweets with high scores strongly deviating from 0 (e.g. Kovacs-Gyori et al., 2018). Hiltz et al. (2014) provided a two-step evaluation where, firstly, they classified tweets as polar versus neutral using the *SentiStrength* algorithm, and then they classified polar tweets as positive or negative using machine-learning classifiers like *Naïve Bayes* and *Support Vector Machine* (SVM). Negative sentiments are dominant in practically all investigations for the performance of transit systems (Haghighi et al., 2018), especially due to delays (Collins et al., 2013) and incidents or common disruptions (Hosseini et al., 2018).

Almost all the mentioned investigations have in common a similar methodology (text-mining techniques to extract topics from words, and sentiment analysis to add positive or negative value to tweets). Frequently, temporal patterns have been analyzed. For example, El-Diraby et al. (2019) discovered negative sentiments are predominant on weekdays at peak hours (7–9 AM and 4–6 PM), while there were more satisfied users on weekends. However, few papers have dealt properly with the spatial dimension, usually being limited by the detection of events. There are almost no works investigating the geographical factors that could explain the spatial distribution of user complaints or topics. In addition, few papers support the results with official data (mainly surveys are used). This paper plans to continue this line of investigation, seeking to go deeper, establishing this main research question: What factors influence the spatial distribution of negative tweets and associated topics inside a city?

To explore the geographical distribution of factors spatial regression models are recommended (Elhorst, 2010). Two different approaches in spatial regression analysis are usually applied: spatial autoregressive models and GWR. Spatial autoregressive models evaluate a spatial component (typically as an independent variable) and include it into the regression model (Anselin, 2002; LeSage, 1998). The most applied variants of spatial autoregressive models are spatial lag and spatial error models (Smith, Goodchild, & Longley, 2018). However, the spatial influence in these models is assumed to be the same in the whole study area. Meanwhile, GWR coefficients are allowed to vary spatially (Brunsdon, Fotheringham, & Charlton, 1996), allowing the study of spatial differences in relationships and to address a possible spatial heterogeneity. Thus, we suggest mapping the main problems (based on negative sentiments) in the study area, adding a GWR analysis to understand the factors contributing locally to the spatial distribution of negative evaluations in Madrid Metro, using official data sources as exploratory variables.

3. Study area and data

3.1. Study area

Madrid Metropolitan Area has an estimated population of 6.2 million inhabitants (2019), of which 3.2 live in the city of Madrid. Madrid Metro is the main public transport system of the metropolitan area. It is a 294-kilometer network, composed of 12 conventional rail lines, and 242 stations. The Madrid Transport Consortium owns the network system. It has an estimated use of 2.3 million travelers per day.[2] Over 43 % of Madrid citizens use Madrid Metro to travel, while 27 % of people travel by bus, and only 13 % of travelers use the commuter train (Cercanías). The main traveling reasons for metro users are to go to work (55.03 %), or study (15.50 %)[3] . Madrid Metro network currently connects all 21 districts in the Madrid municipality and reaches 12 municipalities (Fig. 1).

3.2. Data

Tweets were collected by the streaming Twitter API using a *Python* code, selecting replies to the Madrid Metro account, and excluding retweets. These tweets are replies to the Madrid Metro system public account *(@metro_madrid*). The main reason we selected this approach lies in the assumption, proven by previous papers, that public transport users tend to reply directly to the Twitter account of transport agencies when they complain about or praise the service offered, or when they detect a failure in a specific location, so the frequency is much higher than usual tweeting. In addition, texts from direct replies tend to give insight on very specific complaints or information related with the service, while usual tweets tend to talk about common issues (Haghighi et al., 2018). While being in a middle urban scale (slightly bigger than the bus service, and smaller than the metropolitan train), Madrid Metro

[2] https://www.metromadrid.es/es/quienes-somos/metro-de-madrid-en-cifras.

[3] https://www.metromadrid.es/sites/default/files/documentos/Portal%20de%20transparencia/Memorias/INFORME%20CORPORATIVO%202018.pdf

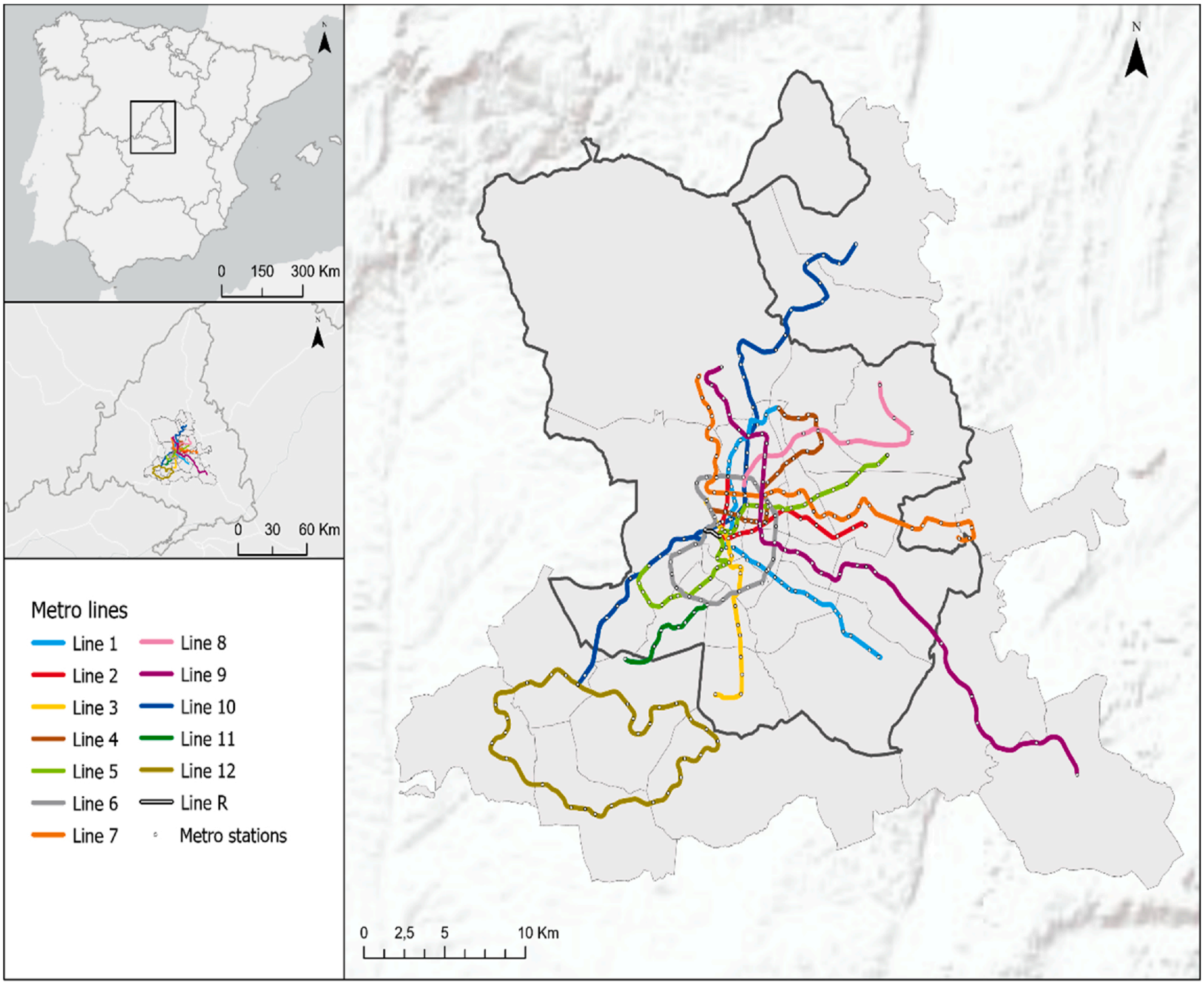

Fig. 1. Madrid Metro network in the Madrid Metropolitan Area.
Source: Own elaboration.

is the public transport service account that receives the most replies on Twitter (Table 1).

The initial tweet database contains 27,603 tweets from 12,361 users, compiled over a two-month period (from September 16 to November 17, 2019). A two-month interval is used to find a compromise between dynamic changes to the transport situation (like seasonal changes or events) and satisfactory data volume. Each tweet has information related to the user identification number (ID), date, time, and language. However, these tweets are not geotagged. We have chosen to use non-geotagged data because the sample is bigger for a smaller period of time, and the noise of not-related tweets is smaller. In this case, geocoding is applied using the text of the tweet, under the assumption that public transport users tend to use a very specific location vocabulary, frequently mentioning the metro station where they are (Haghighi et al., 2018).

Table 1
Tweet frequencies for different transit service accounts in a one-week period (September 16th–22th).

Twitter account	Service	Scale	Number of tweets
@metro_madrid	Metro	Urban	4004
@CercaniasMadrid	Train	Metropolitan	2686
@EMTMadrid	Bus	Urban	769

Source: Own elaboration.

Metro line and station shapefiles are available on the Madrid Transport Consortium webpage. For GWR analysis, data on the resident population was downloaded from the 2019 census by the National Institute of Statistics (INE), selecting the working-age population (19–55 years of age). Income data was provided by the 2015 Madrid Hall Urban Audit for the Madrid city districts, and the 2016 Madrid Community Statistics Institute for the municipalities. The number of points of interest (POI) was calculated using the 2019 *OpenStreetMap* shapefile.

4. Methodology

4.1. Data preprocessing

In order to obtain information, text from tweets needs to be preprocessed, prepared and cleaned using the *Python Pandas* library. For that, a set of transformations are carried out on the text of the tweets in order to reduce the complexity of our analysis algorithms and increase the accuracy. The cleaning process contains the following steps:

1 Conversion of all letters to lowercase.
2 Transformation of letters with diacritical marks into ASCII characters.
3 Removal of hyperlinks (words starting with “http” or “https”).

4 Removal of mentioned accounts (words starting with the character @).
5 Removal of special characters (like #, /, or – characters).
6 Removal of punctuation signs.
7 Removal of emoticons.

Other typical transformations performed in text analysis, such as stopwords removal (meaningless words like articles or prepositions) or text tokenization, were not carried out in this phase. Many station names contain stopwords which are needed for geocoding. In the same way, lemmatization and tokenization are not recommended because the built-in topic classifier is based on pattern recognition (terms as a set of words).

4.2. Geocoding

The next step consisted in giving spatial information to the tweets by geocoding. A *Python* dictionary of keywords with the name of all metro stations was elaborated for this purpose. Then, potential abbreviations were sought and replaced by each station's full name (for example, tweets including the term "ppio" were geocoded in the Principe Pio metro station). That way, 3454 tweets from 2418 users were geocoded (12.5 % of the initial sample). For geocoded tweets located in a station with several metro lines (752 tweets), a second dictionary of keywords with the name of each metro line was employed to extract individual lines from the text of the tweets.

4.3. Semantic and sentiment analysis

Users share opinions on specific topics through the text contained in their tweets. To infer topics in texts, a semantic perspective based on pattern recognition and limited to a set of predefined topics was defined. Before the pattern recognition can take place, a second phase of pre-processing is required. Firstly, words with less than 3 characters were removed from the texts (Lansley & Longley, 2016). Removal of stopwords and text tokenization steps were also carried out. Then, a series of four topics was formulated, based on the main complaints the metro system received on social networks.[4] The topics reflect the time and origin of the tweets. They help to discover personal behaviors and regular patterns within a transit system, but they also enable the detection of extraordinary events occurring within the transport network such as train delays or failure of services (Cheng, Wicks, & Bejon, 2014). A dictionary containing a collection of specific terms (unigrams and bigrams) specific to each topic was elaborated and used to identify topics in every tweet (Table 2). This method was able to identify 1769 tweets that contained an applicable topic (51 % of the sample).

To classify a larger number of tweets, an unsupervised classification model was used. The unclassified tweets were clustered into groups using a probabilistic modeling approach called *Latent Dirichlet Allocation* (LDA) (Blei, Ng, & Edu, 2003). This model finds topics based on the word frequency from a set of documents, where a topic is represented as a weighted list of words. These words are divided into topics, and these are named using frequently used words combined into a sentence that makes sense. In this project, an implementation of a *Gibbs* sampling algorithm was used (Saura & Bennett, 2019). This algorithm is available in the free library *Gensim.*

The number of topics for the LDA model included in this work was set to five. Four clusters were related with the described topics with an accuracy of 69 %. The fifth cluster was composed of miscellaneous tweets with topics different from the previously defined four (298 tweets).

Sentiment analysis classifies the feeling expressed in a text into one of several predefined categories. This analysis was used to classify each tweet in the text corpus into two feeling categories: positive and negative. For this purpose, a BERT (*Bidirectional Encoder Representations from Transformers*) deep learning model was employed (Devlin, Chang, Lee, & Toutanova, 2018). BERT is a modern model based on deep learning that has been recently used to perform sentiment analysis with high accuracy. Unlike other methods traditionally used, BERT takes into account the context of each of the words present in a tweet, which implies a better understanding of the full context of each tweet. In technical terms, the multilingual base model consists of 12 transformer blocks previously trained on Wikipedia corpuses of 104 languages, including Spanish. The model was then specific all tuned for sentiment analysis of tweets in Spanish. Two datasets were used for fine tuning. The first one contained 20,474 tweets mainly about sports bets (with 9474 negative sentiment, and 11,000 positive sentiment tweets) (Malafosse, 2019). The second one contained 46,915 tweets (with 20,111 negative and 26,804 positive) divided into three categories (general, economics, and *InterTASS* corpus) (Martínez-Cámara, García-Cumbreras, Villena-Román, & García-Morera, 2016; Sobrino Sande, 2018).

Two different experiments were performed on the datasets. In the first experiment, the first dataset was employed as the training data for BERT model fine tuning, and the second dataset was used for testing. In the second experiment, both datasets were merged, and then split into a training dataset (80 % of the tweets) and a testing dataset (the remaining 20 %). The results of the experiments can be seen in Table 3. In the metrics calculated in the first experime, the F1-Score and accuracy are both above 0.80 values (although the model was fine-tuned on tweets about sports betting and the test set contained tweets about different topics). Despite the lack of available labeled tweets about public transport in Spanish, the metrics show the model is able to generalize very well. The second experiment shows the model trained on larger text corpora training had even higher evaluation scores with both the F1-score and accuracy metrics being over 0.90. According to these results, the second experiment model was used to label the tweets from the database with sentiment categories (1 for positive tweets, 0 for negative tweets). The generalization abilities of this approach enable us to overcome the obstacle of lack of labeled tweets about public transit.

Table 2
List and description of formulated topics.

Topic	Description	Terms	Number of used terms
Punctuality	Tweets referring to frequency or slowness issues.	late, slow, delay, frequency, waiting time, etc.	82
Comfort	Tweets regarding user wellness from issues like temperature, cleaning, or security.	ventilation, hot, dirty, smell, suffocated, etc.	91
Breakdowns	Tweets reporting breakdowns or failures of the system.	breakdown, interrupted, failure, suspended, works, etc.	137
Overcrowding	Tweets complaining about oversaturation problems in stations or trains.	full, saturated, overcrowded, overflowed, agglomeration, etc.	125

Source: Own elaboration.

Table 3
Results of model fine tuning on a merged test dataset.

Evaluation metrics	Test dataset	
	Experiment #1	Experiment #2
Precision	0.89	0.95
Recall	0.75	0.87
Accuracy	0.81	0.90
F1-Score	0.81	0.91

Source: Own elaboration.

[4] http://t-hoarder.com/metro_madrid/.

After tuning the BERT model, the set of tweets is geocoded and every tweet has a time stamp, a topic and a sentiment category. Using summary statistics by user ID field is possible to perform exploratory da analysis to extract the tweets with negative sentiments and visualize the distribution of users and topics in space and time. We use only negative tweets because we can identify them as problem reports in the metro network. Working with users instead of tweets gives less biased results since users can repeat the same tweet or elaborate on the same complaint many times, generating noise. Another reason is to satisfy the aim to analyze influencing factors which are mainly linked to individuals and not to tweeting activity, such as population, density, income or POI density.

4.4. Ordinary least square (OLS) and geographic weighted regression (GWR)

An Exploratory Data Analysis (EDA) was done, including distribution analyses of every variable. To improve the behavior of variables and to avoid modifiable areal unit problems, a set of secondary variables such as shares (for example, share of complaining Twitter users of the total number of travelers) or densities (users/objects per km^2) was calculated. For all variables, Z-standardized variants were prepared and lognormal transformation of positively skewed variables was applied. Three types of regression models were developed: based on original values of variables, on densities and on Z or Z-LN transformations. OLS and GWR models were tuned for all of them and results were compared. Finally, the model based on densities was selected taking into account its statistical quality and exploratory effect.

OLS analyses were conducted for assessment of global relationships and behavior of variables. To avoid global multicollinearity and information redundancy, a Variance Inflation Factor (VIF) of each explanatory variable as well as an average VIF in OLS models was checked to eliminate those models with values above 5 (Akinwande, Dikko, & Samson, 2015). In addition, multicollinearity tests, heteroscedascity tests (Breusch-Pagan, Koenker-Bassett), residual normality (Jarque--Bera) and model significance tests (F-test for ANOVA, *t*-test for each variable) were performed to build statistically relevant models. The OLS linear regression model enabled us to select an appropriate baseline model and appropriate variables for GWR modelling.

GWR analysis (Brunsdon et al., 1996) was developed in response to heterogeneous conditions with spatially varying relationships when global spatial regression models such as Spatial Autoregressive Models (SAR) fail (LeSage, 1998). While methods like the OLS model don't interpret how variables affect to the spatial distribution of users or topics, the GWR model allows us to analyze variations over space within the sample, obtaining local coefficients that reflect the influence of these variables in the number of users and topics, and recognize where independent variables have greater or lesser explanatory power (Cardozo, García-Palomares, & Gutiérrez, 2012).

In GWR, β_n coefficients of the predictors x_n are allowed to vary spatially. Coefficients are evaluated at each target point (polygon centroid) using a spatially weighted least squares regression (Smith et al., 2018) for a set of points within radius r. Usually, instead of fixed values for r, a distance-decay function $f(d)$, or more commonly, a certain spatial weighting scheme is used to express different influences of the neighborhood on local estimation. The GWR model takes the following form (Fotheringham, Brunsdon, & Charlton, 2002; Rybarczyk, 2018):

$$y_i = \beta_0(u_iv_i) + \beta_1(u_iv_i)\,x_1 + \beta_2(u_iv_i)\,x_2 + \ldots + \beta_n(u_iv_i)\,x_n + \varepsilon_i$$

where y denotes the dependent variable, (u_iv_i) location coordinates i, β_n coefficients of the intercept and predictors x_n, and ε_i is the random error term.

The GWR analysis was performed for a district/municipality spatial resolution (the 21 districts of Madrid City plus 11 adjacent municipalities with metro services, 32 spatial units in total). Different spatial weighting schemes were evaluated. Due to large differences in size, shape, and distance between polygons, non-distance or contiguity-based measures are recommended. Several options with different configurations of neighbors were tested. Finally, the k-Nearest Neighbors method with k =12 was applied. GWR modelling was based on optimization of selected statistical indicators (R^2, AICc, residual squares, sigma), testing of normality and spatial autocorrelation of errors, checking standardized residuals within a threshold 2.5 standard deviation, distribution of local R^2 values, and an explanatory contribution of independent variables.

Table 4 summarizes the employed variables. OLS and GWR analyses were conducted in *GeoDa v.1.14*, *SPSS (IBM Inc.) v.25*, and *ArcGIS Pro*.

5. Results

5.1. Exploratory analysis of Twitter users with a negative sentiment score

2097 users wrote negative sentiment tweets, while only 475 users published positive tweets. The percentage of negative sentiment users (86.47 %) is higher than the percentages obtained in similar works (60.31 % in El-Diraby et al., 2019), and follows the assumption that suggests users usually tweet complaints when they interact with public transport service official accounts (Schweitzer, 2014). Metro lines 6, 9 and 12 present higher percentages of users who posted negative tweets while lines 4, 8 and 11 are the ones with lesser negative score percentages. Lines 6 (circular line that surrounds the Madrid central area and includes all the stations serving as bus transit nodes), 1 (line that connects the center of the city, both train stations, the northern business area, and the southern residential district of Vallecas), and 10 (line that crosses the metropolitan area in longitudinal direction) are the most commented lines (Fig. 2).

The most commented metro stations in the sample correspond with the most used stations according to official data: stations located in the Madrid city center or stations from line 6 where different metro lines connect with the central stations, and other transport services like buses. Other major stations are located in peripheral key zones (stations connecting with Madrid airport and the two city train stations), or are stations that serve as transit points with the metro network in peripheral municipalities (Fig. 3).

By comparing the results obtained with the number of metro travelers in 2019 according to official data sources, coefficient R^2 score 0.81, showcasing Twitter distribution is close to a real-life situation. This

Table 4
Variables used in OLS and GWR modelling.

Variable	Type	Source
Twitter users with negative sentiment scores	Dependent	Twitter
Shares of Twitter users complaining about punctuality (%)	Dependent	Twitter
Shares of Twitter users complaining about comfort (%)	Dependent	Twitter
Shares of Twitter users complaining about overcrowding (%)	Dependent	Twitter
Shares of Twitter users complaining about breakdowns (%)	Dependent	Twitter
Density of residents of working age (per km^2)	Exploratory	2019 INE Census
Average income of residential workers aged 18–55 (€/year)	Exploratory	2015 Madrid Hall Urban Audit, 2016 Madrid Community Statistics Institute
Density of POI (POI/km^2)	Exploratory	2019 *OpenStreetsMap*
Shares of intermodal transit stations (%)	Exploratory	Madrid Transport Consortium

Source: Own elaboration.

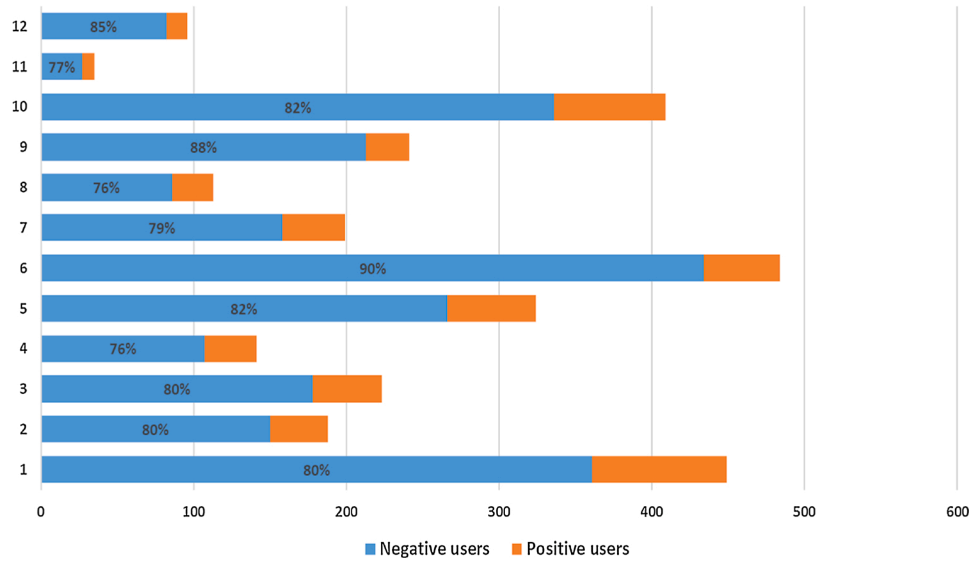

Fig. 2. Number of Twitter users per metro lines by sentiment score.
Source: Own elaboration.

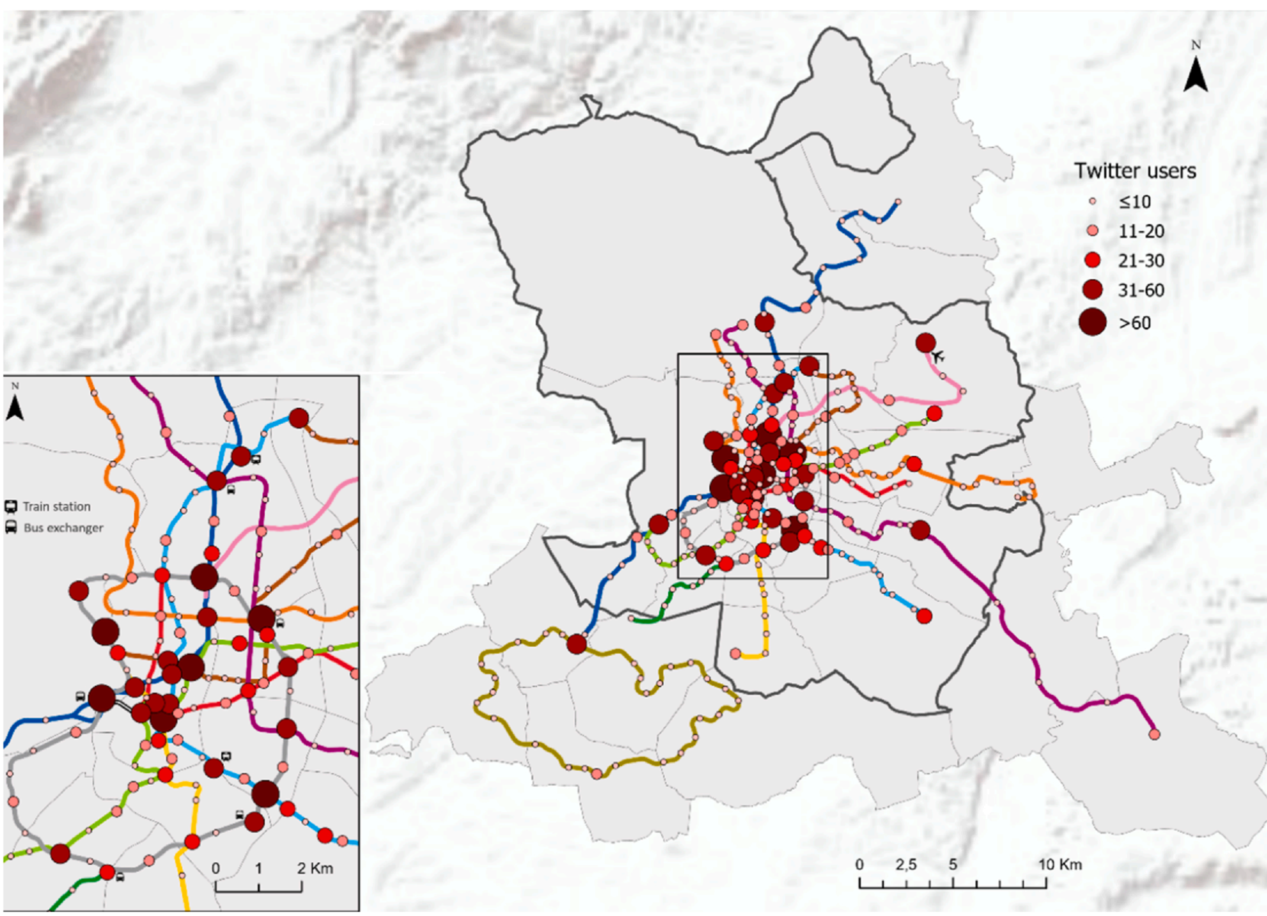

Fig. 3. Distribution of Twitter users with a negative sentiment score in the Madrid metro network.
Source: Own elaboration.

coefficient rises to 0.92 if we add the Twitter users by metro lines, indicating a better accuracy. There are some stations showing over-representation of Twitter complaining travelers, especially stations from line 10 or located in the central area. Two stations (Pacífico in lines 1 and 6, and Bilbao in lines 1 and 4) stand out as the most overrepresented stations in the sample. Meanwhile, stations belonging to metro line 4 and from peripheral southern districts (line 3) and other municipalities (mainly from line 12) are under-represented (Fig. 4).

There is a constant percentage of users during workdays, except on Fridays, the day with the most users in the week. This situation contrasts

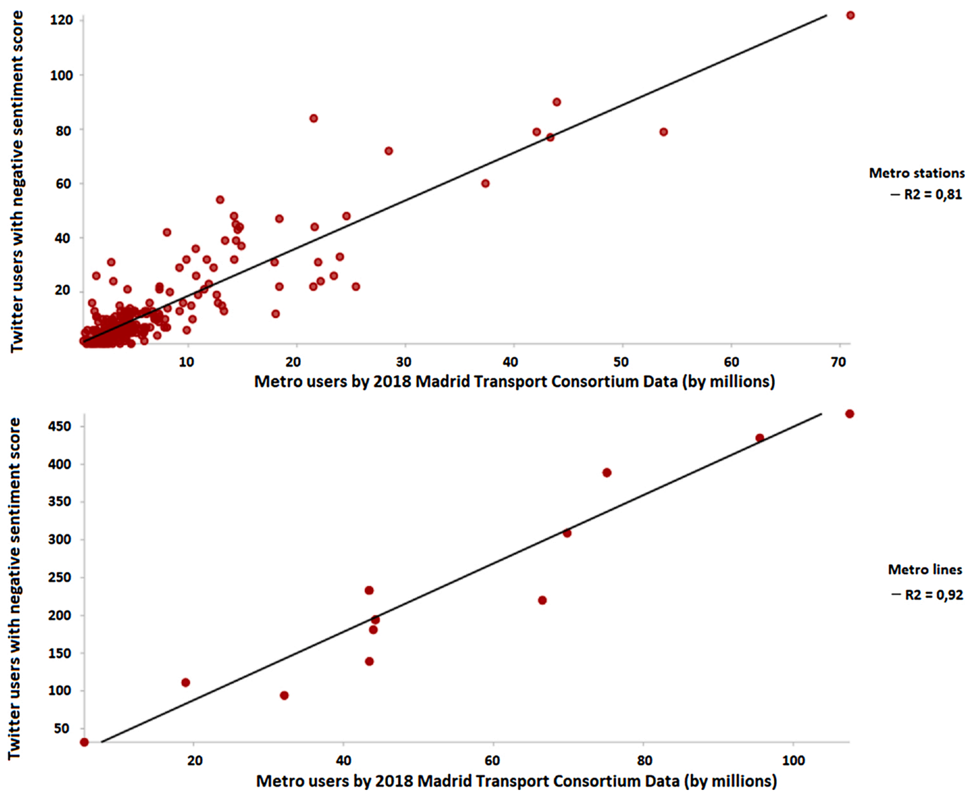

Fig. 4. Relationship between the number of Twitter users with a negative sentiment score and the number of travelers according to Madrid Transport Consortium official data from 2018.
Source: Own elaboration.

with weekends, where the number of Twitter users is low compared with workdays. The main reason may lie in travel motive, since according to official data, most users travel to work. Previous investigations cited in literature review (El-Diraby et al., 2019; Zhang & Feick, 2016) have also recollected weekdays as the most prominent days with Twitter activity.

From analyzing the percentage of users by hour, two peak moments can be observed: a major peak in the early morning (8 AM) where people travel to work or study, and a minor peak in the early afternoon (3 PM), where people travel back home. This observation is in line with previous studies (e.g. El-Diraby et al., 2019). In contrast, midday registers a low use of the metro system, since people are already working or studying, so mobility in the metropolitan area decreases. There is also a continuous, decrease in activity of metro users in the afternoon, which extends into the night, parallel to the decreasing service (Fig. 5). This temporal profile contrasts with the temporal distribution of Twitter users in Madrid not related with transport, which can be downloaded without any transport account filtering. Twitter users generally present little activity in the morning and tend to tweet at night, contrary to

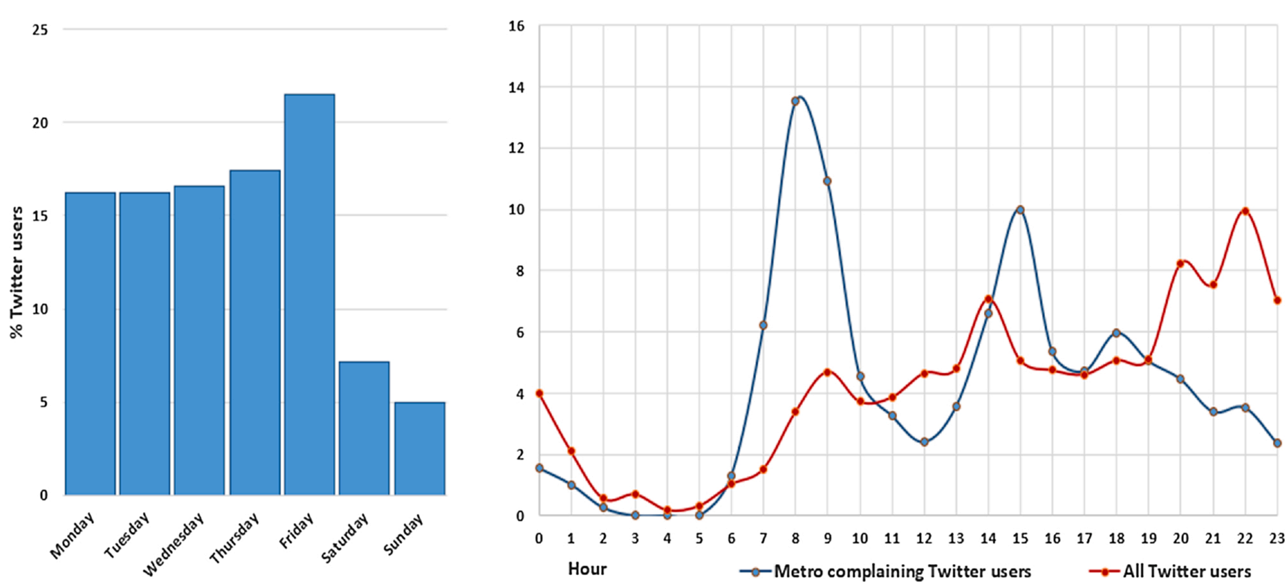

Fig. 5. Percentage of Twitter users with a negative sentiment score in Madrid metro by day and hour.
Source: Own elaboration.

morning-active metro Twitter users. This pattern was also observed in Salt Lake City (Haghighi et al., 2018).

5.2. Distribution of topics in space and time

Punctuality is the issue with the biggest number of total Twitter users. This finding is consistent with results from previous papers (Casas & Delmelle, 2017; Hosseini et al., 2018) that showed Twitter users tend to complain about punctuality and facility problems. While overcrowding has the smallest number of users, it presents the biggest percentage of users with a negative score over the total. On the contrary, comfort is the most positive topic commented on by Twitter users (Table 5).

Fig. 6 shows the distribution of the main topics with negative scores in each metro station. Some patterns can be recognized. Breakdowns are the most reported topic in every station of the lower half of line 1 or in the eastern part of the line 12 circle, while stations on line 9 are affected by punctuality problems. Punctuality and breakdowns are the most distributed topics in the network, with more visibility of punctuality in central area stations, and more breakdown issues in the periphery. Comfort issues also stand out in the central area.

The most used metro lines (lines 1, 6, and 10) are the ones with the most negative reports. However, problems can be different for each line. Punctuality is a major issue in almost all metro lines (and the most reported problem in the most used metro lines), but it is most visible in lines 9 and 12. Comfort is an issue in long lines with many stations (lines 1, 5, 6), but it is mainly apparent in line 8 (shortest line of the network, specifically designed to connect with the airport). Lines that reach the city periphery (lines 2, 4, 7, 11) have breakdowns as one of their main worries. Overcrowding stands out in lines with most stations that connect with other metro lines (lines 6, 10), but also in line 8 (Fig. 7).

During the day, all four issues have maximum negative scores during the two peak mobility times (early morning and early afternoon). Punctuality is once again the main issue during almost the whole day, being clearly visible in these two peak points. A larger amount of negative sentiment users can be seen in the early morning peak, presenting a larger relative difference with regard to other issues in that moment. The temporal profile of overcrowding is similar to the punctuality profile, but shows a similar number of users in both peak times (as a consequence, the relative difference with regard to punctuality is smaller in the early afternoon). While being the main issue in evening hours, users complaining about breakdown issues are also more active in the morning, similar to comfort, despite being also prominent at midday and in the evening (Fig. 8).

5.3. Geographic weighted regression

Analyzing the statistics of the GWR model applied on the density of users with negative score tweets; R^2 shows high scores in the general model. AICc and sigma values are low, showing the validity of the model (Table 6). High absolute values of residuals are located mainly in the central Madrid districts. Local R^2 indicates a very good accuracy of the model (Fig. 9).

The population density variable positively influences the number of complaining Twitter users in the northern units of the city of Madrid (work areas with low population density), while it has little effect on the southern units (residential areas with a high population density). At the same time, there is little impact of this variable in central areas since there the main users of the metro are tourists. Income has no negative coefficients, meaning that it only has a positive influence on the number of complaining users. The biggest power in this variable lies in the southern units, residential areas inhabited by mid-level income workers. These results can also demonstrate that high-level income citizens tend to travel to work by car.

Table 5
Number of Twitter users per topic.

Topic	Number of users	Number of users with negative scores	% users with negative scores
punctuality	1016	859	84.54 %
comfort	729	555	76.13 %
breakdowns	809	666	82.32 %
overcrowding	372	345	92.74 %

Source: Own elaboration.

POI density is the variable that influences the model the most. This result matches previous variables (points of interest are related with points located in the travel destination areas, and the results in the population density variable show that Twitter users tend to send messages while traveling to work areas). It has high explanatory power in almost the entire study area (especially in the central units, the areas with the highest number of infrastructures and services in the study area). However, this variable has negative values in eastern units (peripheral residential areas with less population and metro service than southern areas). Transfer to non-metro station coefficients is negative in southern zones while having a high positive influence in northern zones. This can be interpreted as Twitter users tending to travel by bus or train to northern work zones, and then transfering from the central or northern stations to metro transport to complete their trips. Meanwhile, metro users don’t usually travel in southern areas (Fig. 10).

The GWR analysis by topics shows slightly lower R^2 values than the first model. Distribution of Twitter users complaining about punctuality is mainly related with income in the southern units (where the model presents the highest local R^2 values), and transfer to non-metro stations in the northern areas, showcasing that Twitter users living in the periphery tend to have punctuality problems when transferring to metro services. For overcrowding complaints, transfer to non-metro stations also has high explanatory power in the northern areas (with high local R^2 scores), indicating that punctuality problems related with transport transfers also cause overcrowding. The other standing variable is population density in the southeastern zones. That variable also presents high explanatory value in the models that analyze distribution of breakdowns with complaining users in the same residential municipalities hinting that breakdowns in these areas trigger overcrowding problems for residents who travel to work. Breakdowns have high local R^2 values in the central and western areas of the city, where POI density is the variable with the highest coefficients. That can mean breakdowns are reported quickly not only by workers but also other groups like tourists attracted by frequent POIs. For users elaborating on comfort problems (with highest R^2 values in the central areas), the main variable that affects the model is population density, which extends to the entire central and southern part of the study area. Income also strongly affects the southern units. As a result, it can be interpreted that Twitter users suffer comfort issues during long trips from the peripheral zone to the central or northern work areas. Fig. 11 shows two examples of coefficient distribution by variable and topic.

6. Conclusions

In recent years, rapid urban growth has increased the use of public transport, leading to an unsustainable situation where traditional source data is not enough to understand citizens necessities. In this paradigm, the rise in Big Data sources allows us to obtain large volumes of updated data that can be used to model sustainable public transport scenarios. Twitter stands out for its capacity to provide meaningful information about public transport users perceptions, opinions and sentiments over several potential issues.

In this article, we worked with non-geotagged Twitter data that includes direct replies to the Madrid metro system account. While previous articles focused on sentiment analysis and didn’t work properly on the spatial dimension, this paper has mapped the distribution of metro users and problems detected in the study area. Also, this investigation

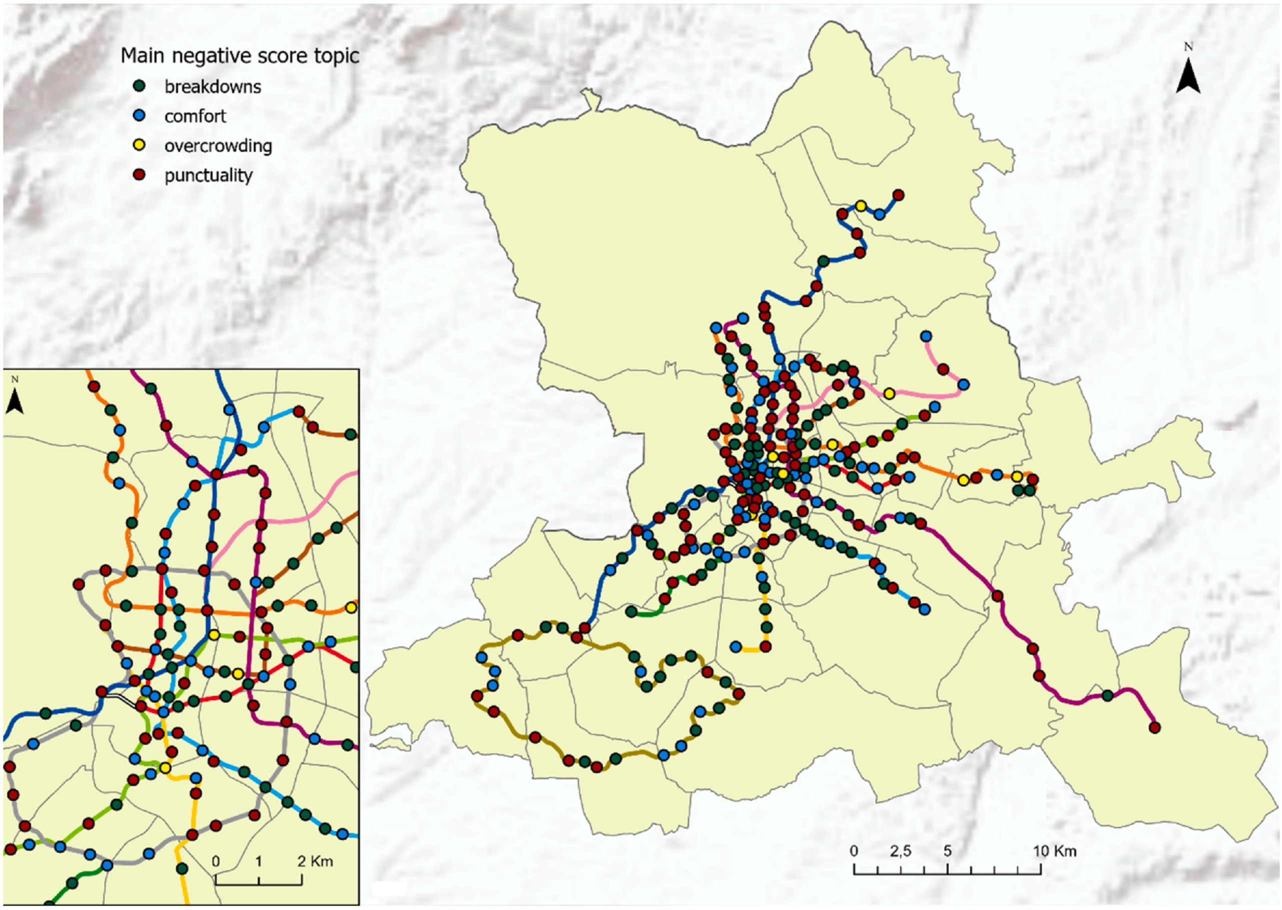

Fig. 6. Main topic with a negative score in Madrid metro stations.
Source: Own elaboration.

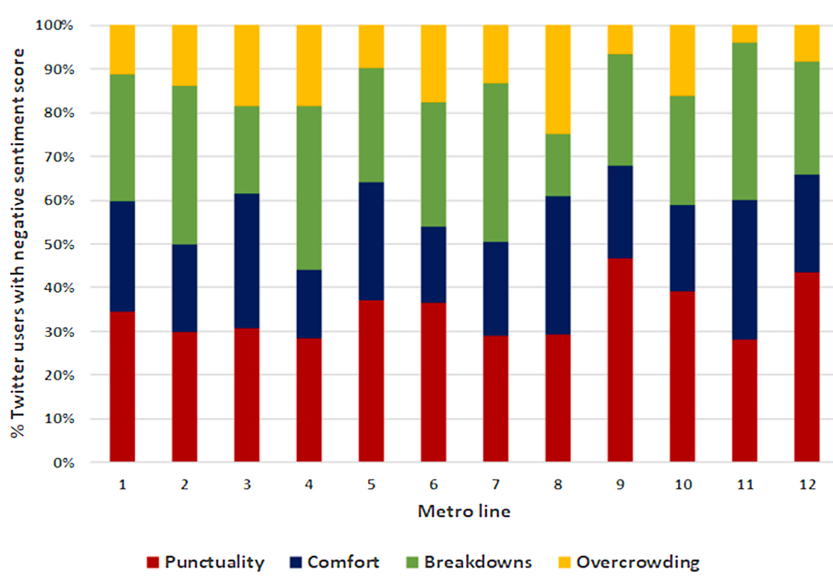

Fig. 7. Percentage of Twitter users with a negative sentiment score by topic and line.
Source: Own elaboration.

has searched for the factors behind Twitter complaints in the metro network. For that, a GWR model was employed to analyze the causality of the spatial distribution of Twitter users with negative sentiments.

Using non-geotagged tweets, it was possible to obtain a larger number of tweets with less noise than using solely geotagged data (since geotagged tweets represent 1% of the sample, it is difficult to obtain a large number of tweets in a short time, and these tweets tend be mixed with non-related messages that create noise). However, geolocating the messages in the study area is essential to be able to perform a spatial analysis, so we geocoded the data by finding the name of stations on a

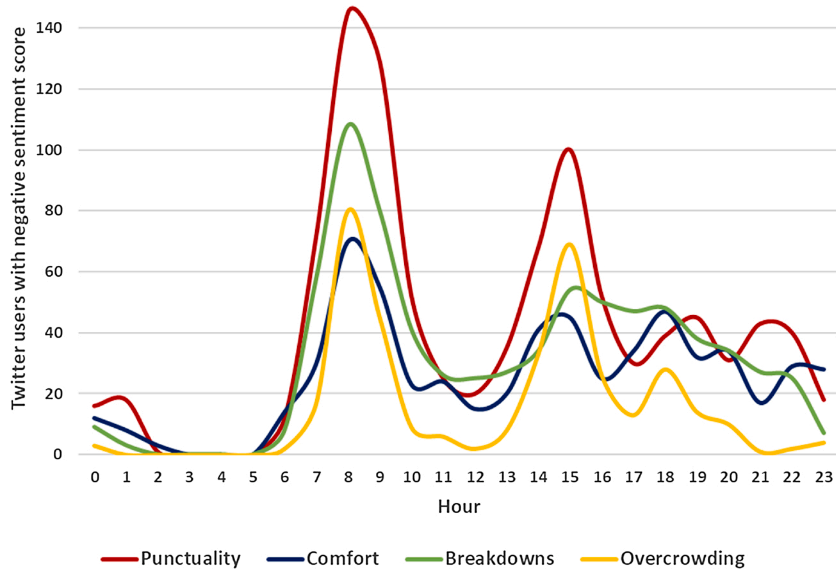

Fig. 8. Number of Twitter users with a negative sentiment score by topic and hour.
Source: Own elaboration.

Table 6
GWR optimized model statistics.

Spatial weights	Variables	R^2	Adjusted R^2	AICc	Sigma	Residual sum of squares
12 neighbors	4	0.976	0.909	184.613	3.930	77.872

Source: Own elaboration.

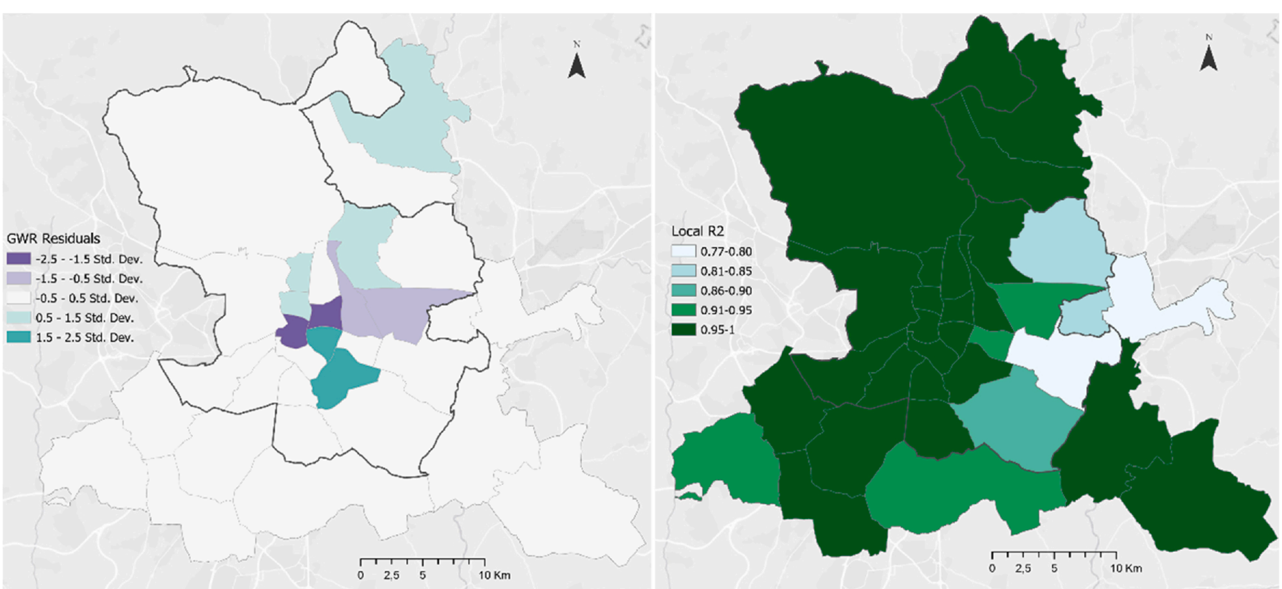

Fig. 9. GWR Residuals and local R^2 values in Madrid Metropolitan Area.
Source: Own elaboration.

metro line in the tweet text. Only 12.5 % of the sample data was able to be geocoded, but it still provided us with a significantly larger number of useful tweets with less noise than if we had used GPS geotagged tweets. To extract opinions and sentiments from Twitter users, texts needed to be pre-processed and cleaned. After that, this paper partially followed the methodology employed by (Saura & Bennett, 2019) to cluster tweets in topics and score sentiments. It was more efficient than selecting appropriate topics from literature such as public transport quality guidelines (Casas & Delmelle, 2017), local media (Zhang & Feick, 2016) or word clouds (Collins et al., 2013).

Some exploratory results are similar to previous investigations: an important amount of negative score tweets, written mainly on weekdays. In this study case, we found the metro stations with the most users who reported issues in a negative sentiment are located in the center of the city or in the circular line 6 that surrounds the central area. Punctuality is the main issue for Madrid Metro users, while comfort problems

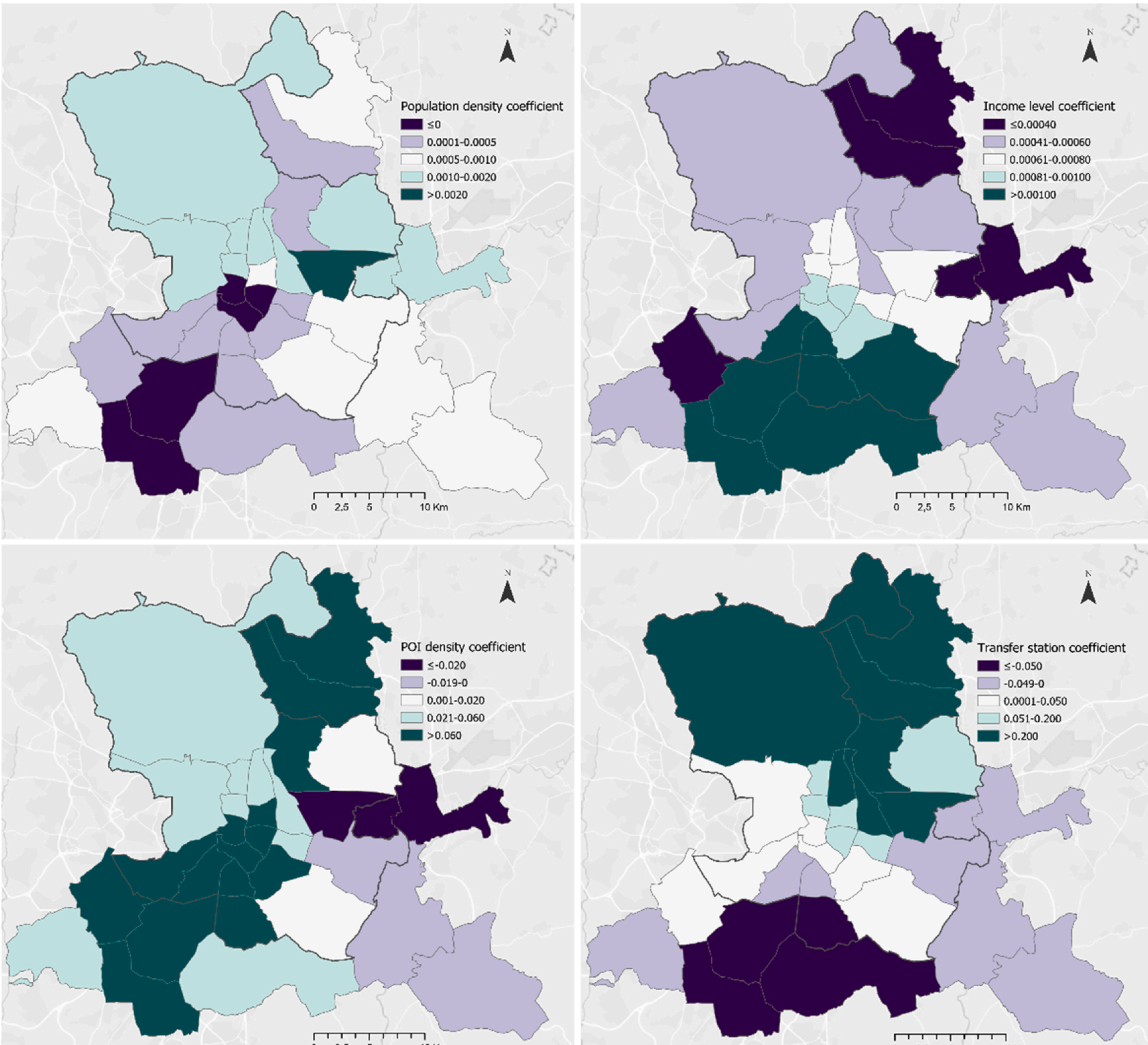

Fig. 10. GWR variables coefficient distribution in Madrid Metropolitan Area. Source: Own elaboration.

are mainly reported in the central area and breakdowns happen more frequently in the periphery. Two peak moments during the day were observed, which coincide with the time when citizens travel from home to work and vice versa which is consistent with the results from previous studies (e.g. El-Diraby et al., 2019). While punctuality is the main issue for users in peak hours, overcrowding mainly during the short afternoon peak, and percentage of users complaining about breakdown of services except during the morning peak is relatively high during the evening especially in the 15–20 PM range. Such findings may be directly used by transport agencies to improve the service in the given time slot.

The main research contribution of this investigation has been the spatial mapping of exploratory variables to explain the reasons behind the distribution of problems in a Metro network. Instead of analyzing tweeting frequency, we focus on complaining users which should decrease a bias of repeated activity as well as better correspond to investigated environmental factors. The most influential factor is POI density, associated with facilities and services located in travel destinations (except in eastern zones, where other factors prevail). Population density shows a stronger influence in northern work areas. Punctuality, the most frequent topic, is related to income level in southern parts, and the density of intermodal transit stations in the northern part. These results also allow us to visualize the profile of traveling Twitter users: mid-income workers, who live mainly in the south of the study area, and tend to travel to work areas mainly located in the north of Madrid (these users travel directly by metro or they use other transport systems and then they transfer to the metro in central or north units).

These findings bring a new view of public transport problems and enable fast and targeted responses. The updated regression modelling of Twitter users may help transport agencies to better understand the emotions, expectations and requirements of Metro travelers. The relationships between factors and frequencies of Twitter topics enable better predictions of changes in volume and emotions of the social media feed due to future changes in urbanization and transport networks. This is essential for appropriate assessment of temporal trends in Twitter activity and customer evaluation. An example may be the ability to decide if the observed local increasing density of negative sentiment (increased volume negative of tweets without a corresponding increase in the volume of passengers) is really caused by higher dissatisfaction of travelers or driven by changes in urban factors such as an increase in

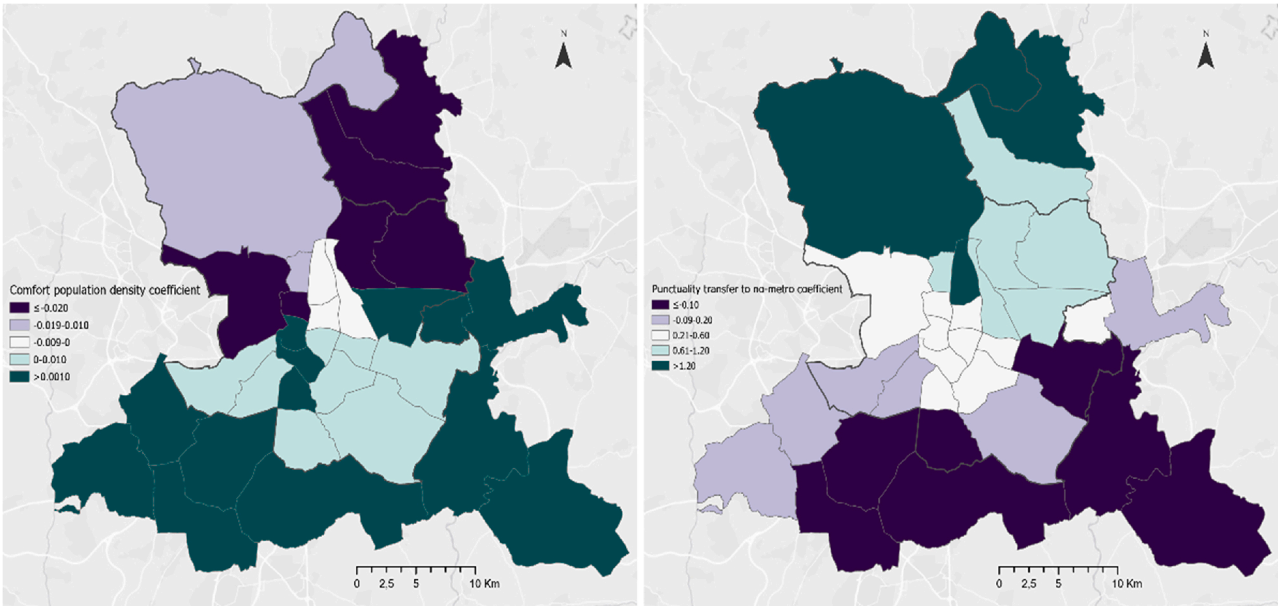

Fig. 11. GWR coefficient distribution by population density in comfort Twitter users (left) and by transfer to non-metro stations in punctuality Twitter users (right). Source: Own elaboration.

population density, changes in personal income, etc.

Nevertheless, some problems were detected during the investigation. Firstly, while the employed data sample is larger and less noisy than geotagged data, accuracy is lower, and there is a loss of data with no information for geocoding, while all geotagged tweets are useful for spatial visualization. Another problem lies in the accuracy of text-mining techniques to extract topics and sentiments from texts. While techniques like geocoding, the LDA model or sentiment extraction algorithms are useful to extract data, they are difficult to implement for the personal decontextualized and specific nature of tweet texts (short messages that usually use abbreviations and emoticons), so complementary methods like abbreviation dictionaries are needed for further improvement. Current sentiment analyses also do not utilize a flow of tweets (re-tweeting). Processing of especially non-English texts is still less satisfactory namely when statements contain irony or sarcasm. Other limitations are data bias (predominant Twitter use by 20–39-year-old users), unknown social profiles of users, and lack of sample representativeness. Working with the text of the tweets also entails having problems with privacy data so, to minimize the problem, tweets have been added by metro station or line, or municipality.

Some of the detected problems can be improved by increasing the size of the samples, for example increasing the temporal period of the sample. Future lines of investigation include comparative analysis of the public metro network with other public transport services like buses or metropolitan trains, the combined used of non-geotagged data and GPS geotagged data to improve accuracy, or the use of a longer timeframe to explore annual patterns and problems detected in exceptional situations or events, how these events are differently reflected in different urban areas, variable spatio-temporal tweeting patterns for visitors and residents or advanced social typology of users accounts (Casas & Delmelle, 2017).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowlegment

This work was supported by Community of Madrid: [Grant Number SOCIALBIGDATA-CM, S2015/HUM-3427]; Spanish Ministry of Education, Science and Universities and the European Regional Development Fund: [Grant Number DynMobility, RTI2018-098402-B-I00]; Spanish Ministry of Edu- cation, Culture and Sport: [Grant Number FPUAP2015-0147].

References

Akinwande, M. O., Dikko, H. G., & Samson, A. (2015). Variance inflation factor: As a condition for the inclusion of suppressor variable(s) in regression analysis. *Open Journal of Statistics, 05*(07), 754–767. https://doi.org/10.4236/ojs.2015.57075

Anselin, L. (2002). Under the hood: Issues in the specification and interpretation of spatial regression models. *Agricultural Economics, 27*(3), 247–267.

Banister, D. (2011). Cities, mobility and climate change. *Journal of Transport Geography, 19*(6), 1538–1546. https://doi.org/10.1016/j.jtrangeo.2011.03.009

Blei, D. M., Ng, A. Y., & Edu, J. B. (2003). Latent dirichlet allocation Michael I. Jordan. *Journal of Machine Learning Research, 3*.

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis, 28* (4), 281–298. https://doi.org/10.1111/j.1538-4632.1996.tb00936.x

Cardozo, O. D., García-Palomares, J. C., & Gutiérrez, J. (2012). Application of geographically weighted regression to the direct forecasting of transit ridership at station-level. *Applied Geography*. https://doi.org/10.1016/j.apgeog.2012.01.005

Casas, I., & Delmelle, E. C. (2017). Tweeting about public transit — Gleaning public perceptions from a social media microblog. *Case Studies on Transport Policy, 5*(4), 634–642. https://doi.org/10.1016/j.cstp.2017.08.004

Chen, Y., Bouferguene, A., Li, H. X., Liu, H., Shen, Y., & Al-Hussein, M. (2018). Spatial gaps in urban public transport supply and demand from the perspective of sustainability. *Journal of Cleaner Production, 195*, 1237–1248.

Cheng, T., Wicks, T., & Bejon, P. (2014). Detection using Twitter: A spatio-temporal approach. *PloS One, 9*(6). https://dx.plos.org/10.1371/journal.pone.0097807.

Chin, D., Zappone, A., & Zhao, J. (2016). Analyzing twitter sentiment of the 2016 presidential candidates. *Applied Informatics and Technology Innovation Conference (AITIC 2016)*.

Collins, C., Hasan, S., & Ukkusuri, S. V. (2013). A novel transit rider satisfaction metric a novel transit rider satisfaction metric: Rider sentiments measured from online social media data. *Journal of Public Transportation, 16*(2), 21–45.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of deep bidirectional transformers for language understanding*. Retrieved from http://arxiv.org/abs/1810.04805.

El-Diraby, T., Shalaby, A., & Hosseini, M. (2019). Linking social, semantic and sentiment analyses to support modeling transit customers' satisfaction: Towards formal study of opinion dynamics. *Sustainable Cities and Society, 49*. https://doi.org/10.1016/j.scs.2019.101578

Elhorst, J. P. (2010). Spatial panel data models. *Handbook of applied spatial analysis: Software tools, methods and applications* (pp. 377–407). Berlin Heidelberg, Berlin: Springer. https://doi.org/10.1007/978-3-642-03647-7_19
Fojtík, D., Horák, J., Orlíková, L., Kocich, D., & Inspektor, T. (2016). Smart geocoding of objects. *Proceedings of the 2016 17th International Carpathian Control Conference, ICCC 2016*, 200–205. https://doi.org/10.1109/CarpathianCC.2016.7501093
Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationships*. Wiley.
Graham, M., Hale, S. A., & Gaffney, D. (2014). Where in the world are you? Geolocation and language identification in twitter. *The Professional Geographer, 66*(4), 568–578. https://doi.org/10.1080/00330124.2014.907699
Gutiérrez-Puebla, J., & García-Palomares, J. C. (2016). Big (Geo) Data en Ciencias Sociales: Retos y Oportunidades. *Revista de Estudios Andaluces, 33*(331), 1–23. https://doi.org/10.12795/rea.2016.i33.0
Haghighi, N. N., Liu, X. C., Wei, R., Li, W., & Shao, H. (2018). Using Twitter data for transit performance assessment: A framework for evaluating transit riders' opinions about quality of service. *Public Transport, 10*(2), 363–377. https://doi.org/10.1007/s12469-018-0184-4
Hiltz, S., Pfaff, M., Plotnick, L., Robinson, A., Caragea, C., Squicciarini, A., … Tapia, A. (2014). Mapping moods: Geo-mapped sentiment analysis during hurricane Sandy. *Proceedings of the 11th International ISCRAM Conference – University Park, Pennsylvania, USA*. Retrieved from http://reliefweb.int/sites/reliefweb.int/files/resources/report_36.pdf.
Hosseini, M., El-Diraby, T., & Shalaby, A. (2018). Supporting sustainable system adoption: Socio-semantic analysis of transit rider debates on social media. *Sustainable Cities and Society, 38*, 123–136. https://doi.org/10.1016/j.scs.2017.12.025
Ji, T., Fu, K., Self, N., Lu, C.-T., & Ramakrishnan, N. (2018). Multi-task learning for transit servide disruption detection. *ASONAM 2018: Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*.
Kirilenko, A. P., & Stepchenkova, S. O. (2017). Sochi 2014 Olympics on Twitter: Perspectives of hosts and guests. *Tourism Management, 63*, 54–65. https://doi.org/10.1016/j.tourman.2017.06.007
Kocich, D. (2017). Multilingual sentiment mapping using twitter, Open source tools, and dictionary-based machine translation approach. *GIS Ostrava 2017*. https://doi.org/10.1007/978-3-319-61297-3_16
Kovacs-Gyori, A., Ristea, A., Havas, C., Resch, B., & Cabrera-Barona, P. (2018). #London2012: Towards citizen-contributed urban planning through sentiment analysis of twitter data. *Urban Planning, 3*(1), 75. https://doi.org/10.17645/up.v3i1.1287
Kovacs-Györi, A., Ristea, A., Kolcsar, R., Resch, B., Crivellari, A., & Blaschke, T. (2018). Beyond spatial proximity—Classifying Parks and their visitors in London based on spatiotemporal and sentiment analysis of twitter data. *ISPRS International Journal of Geo-information, 7*(9), 378. https://doi.org/10.3390/ijgi7090378
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
LeSage, J. P. (1998). *Spatial econometrics*. Retrieved from http://www.econ.utoledo.edu.
Lim, K. H., Lee, K. E., Kendal, D., Rashidi, L., Naghizade, E., Winter, S., & Vasardani, M. (2018). The grass is greener on the other Side. *Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW' 18* (pp. 275–282). New York, New York, USA: ACM Press. https://doi.org/10.1145/3184558.3186337
Luong, T. T. B., & Houston, D. (2015). Public opinions of light rail service in Los Angeles, an analysis using twitter data. *IConference 2015 Proceedings*, 2–5.
Malafosse, C. (2019). *FastText sentiment analysis for tweets: A straightforward guide*. Retrieved January 23, 2020, from https://towardsdatascience.com/fasttext-sentiment-analysis-for-tweets-a-straightforward-guide-9a8c070449a2.
Manetti, G., Bellucci, M., & Bagnoli, L. (2017). Stakeholder engagement and public information through social media: A study of Canadian and American public transportation agencies. *The American Review of Public Administration, 47*(8), 991–1009. https://doi.org/10.1177/0275074016649260
Martínez-Cámara, E., García-Cumbreras, M. A., Villena-Román, J., & García-Morera, J. (2016). *The evolution of the Spanish opinion mining systems*. Retrieved January 23, 2020, from http://rua.ua.es/dspace/handle/10045/53559.
Miralles-Guasch, C., & Martínez, M. (2013). Las fuentes de información sobre movilidad: La visión de los profesionales. Ejemplo de aplicación de metodología DELPHI. *Revista Transporte y Territorio*, (8), 99–116. Retrieved from http://www.rtt.filo.uba.ar/RTT00806100.pdf.
Rybarczyk, G. (2018). Toward a spatial understanding of active transportation potential among a university population. *International Journal of Sustainable Transportation, 12*(9), 625–636. https://doi.org/10.1080/15568318.2017.1422301
Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes twitter users: Real-time event detection by social sensors. *WWW2010*. Retrieved from http://mecab.sourceforge.net/.
Saura, J. R., & Bennett, D. R. (2019). A three-stage method for data text mining: Using UGC in business intelligence analysis. *Symmetry, 11*(4). https://doi.org/10.3390/sym11040519
Schweitzer, L. (2014). Planning and social media: A case study of public transit and stigma on twitter. *Journal of the American Planning Association, 80*(3), 218–238. https://doi.org/10.1080/01944363.2014.980439
Shen, W., Xiao, W., & Wang, X. (2016). Passenger satisfaction evaluation model for Urban rail transit: A structural equation modeling based on partial least squares. *Transport Policy, 46*, 20–31.
Smith, M. J., Goodchild, M. F., & Longley, P. A. (2018). *Geospatial analysis: A comprehensive guide to principles, techniques and software tools*. Retrieved from http://www.spatialanalysisonline.com/.
Sobrino Sande, J. C. (2018). *Análisis de sentimientos en Twitter*. Retrieved January 23, 2020, from http://hdl.handle.net/10609/81435.
Stathopoulos, A., & Marcucci, E. (2014). De Gustibus Disputandum Est: Non-linearity in public transportation service quality evaluation. *International Journal of Sustainable Transportation, 8*(1), 47–68.
Steiger, E., Resch, B., & Zipf, A. (2016). Exploration of spatiotemporal and semantic clusters of Twitter data using unsupervised neural networks. *International Journal of Geographical Information Science, 30*(9), 1694–1716. https://doi.org/10.1080/13658816.2015.1099658
Steiger, E., de Albuquerque, J. P., & Zipf, A. (2015). An advanced systematic literature review on spatiotemporal analyses of twitter data. *Transactions in GIS, 19*(6), 809–834. https://doi.org/10.1111/tgis.12132
Steiger, E., Ellersiek, T., Resch, B., & Zipf, A. (2015). Uncovering latent mobility patterns from Twitter during mass events. *International Journal of Geographical Information Science: IJGIS, 1*, 525–534. https://doi.org/10.1553/giscience2015s525
Wachowicz, M., & Liu, T. (2016). Finding spatial outliers in collective mobility patterns coupled with social ties. *International Journal of Geographical Information Science*. https://doi.org/10.1080/13658816.2016.1144887
Wang, Q., Phillips, N. E., Small, M. L., & Sampson, R. J. (2018). Urban mobility and neighborhood isolation in America's 50 largest cities. *Proceedings of the National Academy of Sciences of the United States of America, 115*(30), 7735–7740. https://doi.org/10.1073/pnas.1802537115
Wong, R. C. P., Szeto, W. Y., Yang, L., et al. (2017). Elderly users' level of satisfaction with public transport services in a high-density and transit-oriented city. *Journal of Transport & Health, 7*, 209–217.
Zhañay, B. A., Cordero, G. O., Cordero, M. O., & Urigüen, M.-I. A. (2019). *A text mining approach to discover real-time transit events from twitter* (pp. 155–169). Cham: Springer. https://doi.org/10.1007/978-3-030-02828-2_12
Zhang, S., & Feick, R. (2016). Understanding public opinions from geosocial media. *ISPRS International Journal of Geo-information, 5*(6). https://doi.org/10.3390/ijgi5060074