ARTICLE IN PRESS

Cities xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Cities

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

# City dynamics through Twitter: Relationships between land use and spatiotemporal demographics

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

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

ARTICLE INFO

*Keywords:*
Social media
Twitter
Spatiotemporal demographics
Land use
Urban geography

ABSTRACT

Social network data offer interesting opportunities in urban studies. In this study, we used Twitter data to analyse city dynamics over the course of the day. Users of this social network were grouped according to city zone and time slot in order to analyse the daily dynamics of the city and the relationship between this and land use. First, daytime activity in each zone was compared with activity at night in order to determine which zones showed increased activity in each of the time slots. Then, typical Twitter activity profiles were obtained based on the predominant land use in each zone, indicating how land uses linked to activities were activated during the day, but at different rates depending on the type of land use. Lastly, a multiple regression analysis was performed to determine the influence of the different land uses on each of the major time slots (morning, afternoon, evening and night) through their changing coefficients. Activity tended to decrease throughout the day for most land uses (e.g. offices, education, health and transport), but remained constant in parks and increased in retail and residential zones. Our results show that social network data can be used to improve our understanding of the link between land use and urban dynamics.

## 1. Introduction

Internet users are no longer mere passive recipients of information, but have become producers of vast amounts of data, particularly through social networks. Many social media apps (Twitter, Foursquare and Facebook are three common examples) have geo-location features that (optionally) includes the ability to attach locational information in the form of coordinates provided by the units GPS or place names provided by the user, thereby enabling individual users to leave a digital 'geographic footprint' of their movement when posting a message (Blanford, Huang, Savelyev, & MacEachren, 2015). These digital data shadows are intimately intermingled with offline, material geographies of everyday life (Jin et al., 2017; Shelton, Poorthuis, & Zook, 2015). High spatial and temporal social media data reveals activity patterns information moving beyond the night-time residential geographies of conventional geodemographic data sources (Longley, Adnan, & Lansley, 2015). Social media data have been used to delineate the boundary of urban agglomeration based on people's activity (Zhen, Cao, Qin, & Wang, 2017), and can help to portray urban structure and related socioeconomic performance (Martí, Serrano-Estrada, & Nolasco-Cirugeda, 2017; Shen & Karimi, 2016; Zhang, Zhou, & Zhang, 2017).

Twitter has become an important sensor of the interactions between individuals and their environment (Frias-Martinez & Frias-Martinez, 2014). Twitter data contain precise spatial and temporal information, which can be used to perform spatiotemporal demographic analyses. It has been observed that residential zones on the periphery of cities generate more tweets in the evening, when people have returned to their homes, whereas areas of activity in the city centre are especially active during the day, when people visit them to undertake activities such as work or shopping (Ciuccarelli, Lupi, & Simeone, 2014). The spatiotemporal pattern of tweets posted from different zones in the city indicates the existence of the inherent linkage between the human urban activity pattern and the underlying land use structure (Zhan, Ukkusuri, & Zhu, 2014).

The aim of this study was to use information from the social network Twitter in order to analyse city dynamics throughout the day and the relationship between this and land use. We estimated the number of unique active users in each zone of the city of Madrid throughout the day, as a proxy for the changing location of the population. Then, we analysed Twitter activity (active users) throughout the day according to land use, yielding typical activity profiles for each land use. Lastly, we calculated multiple regression models (OLS) for each time slot to

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

http://dx.doi.org/10.1016/j.cities.2017.09.007
Received 13 June 2017; Received in revised form 5 September 2017; Accepted 16 September 2017
0264-2751/ 

explain Twitter activity based on square metres of each land use in each zone of the city. An analysis of the coefficients of the independent variables should indicate the changing influence of the different land uses according to time slots. These coefficients were used to predict activity patterns in a future urban development in Madrid according to a range of scenarios of land use mix and density. The underlying hypothesis was that if each land use has a specific influence on Twitter activity, then the built-up area of each land use should satisfactorily explain the distribution of tweeters in the city in each time slot. This approach provides useful information for urban planning, not only because it sheds light on urban dynamics in relation to land use, but also because it can be used as a tool in the planning process when it comes to evaluating future urban developments.

This paper contributes to the existing literature in several ways. Firstly, works related to Twitter use and land use attempt to infer land use on the basis of temporal profiles in Twitter use. This, in reality, is of limited use given that there are techniques (remote sensing) which are more reliable and easier to apply in order to obtain land use maps. In our case, the objective is not to infer land use, but to learn about the daily dynamics of the city with regard to land use. Secondly, we used explanatory OLS models which linked land use and Twitter activity. To the best of our knowledge, previous works have been limited to performing descriptive analyses. These explanatory models are of great interest for urban planning. On the basis of predicted land use in new urban developments it is possible to obtain a proxy of the activity (demand) at different times of day in those future new developments. Thirdly, in contrast to previous studies, the land use data in this paper do not come from land use maps but from the Cadastre. Cadastral data provide precise information on the mix of uses of built-up land in each spatial unit and not simply the predominant land use. This is of great importance for adjusting the explanatory models. In contrast with other previous works, in this paper we have worked with a detailed classification of land use, which is useful for urban planning. Fourthly, another distinguishing feature of our study is that the unit of analysis is the Twitter user rather than the geotagged tweet. We focused not on the spatiotemporal distribution of tweets but on that of Twitter users, as a proxy for the changing location of the population throughout the day. This approach mitigates the bias arising from varying intensities of Twitter use, which gives more weight to users who tweet more often, and above all, it provides useful information for the provision of public sector services and private sector business activities.

The remainder of this paper is structured as follows. Section 2 presents a literature review on the use of Twitter data in urban studies, especially in relation to land uses. Section 3 reports the methodology and data. Section 4 contains an analysis of the results and Section 5 presents the main conclusions.

## 2. Human activities and land uses: spatiotemporal patterns of digital footprints in the city

Social network users generate a vast amount of data. Worldwide, 500 million Twitter messages are posted every day, there are 7000 million check-ins on Foursquare, and > 80 million photos are uploaded on Instagram. Most studies based on social network data have used Twitter (Murthy, 2013), due in large part to the fact that the data (tweets) can be freely downloaded by connecting to the Twitter Streaming API. This real-time, public, and cost-free data stream covers only around 1–2% of the whole stream of tweets (Andrienko et al., 2013). Some of the tweets are geotagged. These facilitate profiling of usage across space as well as time and have been found to be a useful tool for urban research (Lansley & Longley, 2016).

Within a city from the centre to the periphery, the densities decrease in general, so tweet densities would be a good surrogate of population densities (Jiang, Ma, Yin, & Sandberg, 2016). Maps of the spatial distribution of tweets can be enriched with user IDs, tweet content and language. Thus, Longley et al. (2015) inferred the gender, age and ethnicity of tweeters from the user ID (username field) and analysed the spatial distribution of the tweets posted by each of the groups identified. Lansley and Longley (2016) and Andrienko et al. (2013) focused on tweet content in order to examine frequently tweeted words and their spatiotemporal patterns. Meanwhile, Mocanu et al. (2013) used the language of tweets to identify linguistically specific urban communities. It is also possible to analyse the degree of social mixing in the use of space, tracking the movement of demographic groups within cities (Netto, Pinheiro, Meirelles, & Leite, 2015; Shelton et al., 2015). In contrast to the information provided by official sources, which offer data on place of residence, these studies on multiculturalism and social mixing analyse the use of space throughout the day.

Through spatiotemporal monitoring of the tweets posted by each user, it is possible to deduce population mobility patterns (Blanford et al., 2015; Longley & Adnan, 2016; Wu, Zhi, Sui, & Liu, 2014), identify demographic groups based on user names (Luo, Cao, Mulligan, & Li, 2016) and obtain origin-destination matrices (Gao et al., 2014). The reliability of Twitter data in mobility studies has been validated by Lenormand et al. (2014), who compared Twitter data with mobile phone records and official data (census) and concluded that the three sources offer comparable results.

The spatial distribution of tweets in the urban core changes over the course of the day, reflecting changes in the location of the population. Areas of the city with similar time profiles can be grouped using clustering algorithms in order to identify clusters of common tweeting activity. Thus, Frias-Martinez, Soto, Hohwald, and Frias-Martinez (2012) and Frias-Martinez and Frias-Martinez (2014) identified four clusters with specific tweeting activity signatures that basically corresponded to the following types of land use: business, leisure/weekend, nightlife and residential. Using a similar methodology, Zhan et al. (2014) inferred four broad categories of land use based on a spatiotemporal analysis of Twitter data: residential, retail, open space/recreation and transportation/utility. A similar approach has been adopted in other studies (e.g., Pei et al., 2014; Ríos & Muñoz, 2017), in which clustering algorithms have been used to group city zones with similar profiles according to mobile phone activity, or Chen et al. (2017) in China using the social media "*Tencent*". These methodologies based on clustering the tweeting activity profiles of different areas in the city could be used as an alternative to satellite imagery, to infer urban land uses based on social network data. However, the land use categories obtained are very generic, and their usefulness for urban planning is therefore limited.

## 3. Data and methodology

### *3.1. Twitter data pre-processing*

One of the most widespread social networks is undoubtedly Twitter, a platform for posting messages with a maximum of 140 characters, known as tweets. The service has > 270 million active users around the world. Roughly 80% of active Twitter users access the service via a mobile telephone (Lansley & Longley, 2016). Since 2010, Twitter provided users with the ability to include their location either by attaching coordinates or a place name while tweeting, therefore making it possible to locate tweets geographically both in space and over time (Blanford et al., 2015). Tweets are automatically geotagged by the GPS of mobile devices, provided that the user has enabled this function. Geotagged Tweets account for approximately just 1% of all the messages that are sent using the Twitter service (Blanford et al., 2015).

Twitter data were downloaded and pre-processed as follows:

a) Data download

The data used for this study was downloaded via the Twitter Streaming API over two consecutive years (from January 2012 to December 2013). Only geotagged tweets were downloaded, selecting those that covered the municipality of Madrid. Besides the coordinates,