The use of public spaces in a medium-sized city: from Twitter data to mobility patterns

María Henar Salas-Olmedo a,b and Carolina Rojas Quezada b

aDepartamento de Geografía Humana, Facultad de Geografía e Historia, Universidad Complutense Madrid, Madrid, Spain; bDepartamento de Geografía, Facultad de Arquitectura, Urbanismo y Geografía, Universidad de Concepción, Centro de Desarrollo Urbano Sustentable, CEDEUS, Concepción Región del Biobío, Chile

ABSTRACT
This research evidences the usefulness of open big data to map mobility patterns in a medium-sized city. Motivated by the novel analysis that big data allow worldwide and in large metropolitan areas, we developed a methodology aiming to complement origin-destination surveys with a la carte spatial boundaries and updated data at a minimum cost. This paper validates the use of Twitter data to map the impact of public spaces on the different parts of the metropolitan area of Concepción (MAC), Chile. Results have been validated by local experts and evidence the main mobility patterns towards spaces of social interaction like malls, leisure areas, parks and so on. The Main Map represents the mobility patterns from census districts to different categories of public spaces with schematic lines at the metropolitan scale and it is centred in the city of Concepción (Chile) and its surroundings (∼10 kilometres).

1. Introduction

The study of social networks has long been a field of interest from different perspectives. Transport-oriented studies are a major research line as they aim to uncover mobility patterns in order to improve transport and urban planning. There is however another raising field of interest related to the use of public spaces, with the focus on the shared spaces. The aim is to help reducing social exclusion by promoting the shared use of public spaces as opposed to the existence of segregated places.

In this research, we explore the utility of big open data through the Twitter platform for small–to medium-sized cities, which have been paid much less attention and still host almost 50% of the urban population (United Nations, 2014). We use the city of Concepción (South of Chile) as a study area to analyse mobility patterns to public spaces. Concepción is a medium-sized city with over 200,000 inhabitants and strong spatial relationship with Talcahuano. Both cities are the articulators of Concepción metropolitan area with nearly 1 million inhabitants (MAC).

Twitter data are publicly available, even at no cost if one uses the Streaming API. For this reason, Twitter has become a popular data source and is present in millions of research documents. However, most of them do not make use of its geographical dimension (Leetaru, Wang, Padmanabhan, & Shook, 2013), or do it at a worldwide or regional scale (Hawelka et al., 2014; Li, Goodchild, & Xu, 2013; Liu, Zhao, Khan, Cameron, & Jurdak, 2015; Sobolevsky et al., 2015). The utility of Twitter data to obtain regular mobility data in large urban areas was demonstrated by Huang and Wong (2015) in their case study in Washington DC. They were able to obtain space–time paths from Twitter geolocated data similar to traditional travel diaries. We aim to extract similar material for our case study area in order to obtain information about the use of public spaces as opportunities for spatio-temporal interaction.

Our aim is to provide complementary and easily updatable information of mobility patterns to current origin-destination surveys that take place every 10 years, especially considering the last one in Concepción was done in 1999, and it has not even been actualized in 2009. While being aware of the potential over-valorization of this type of big datasets, we agree with Shelton, Poorthuis, and Zook (2015) that their high spatial and temporal granularity provide useful information additional to that extracted from official datasets.

Another advantage is the free delimitation of the study areas, which is essential for studying specific places. This is the case of the public space, which is relevant because many social activities outside homes occur in public spaces (Rojas, Carrasco, Pérez, Araneda, & Lima, 2015). Even park and plazas play an important role for people (Villagra-Islas & Alves, 2016). The challenge is to extract useful information from a smaller...
In order to fulfill the objective of improving spatio-temporal resolution and reducing data collection costs, we made use of geolocalized Twitter data. This is freely available on a real-time basis; thus an effort was made to download all of the geolocalized tweets trough the streaming API in order to build a proper database. We used Python language to screen and save the tweets within a certain bounding box, and then we transformed it into a point layer for use in a Geographic Information System (GIS).

This research was made with the geolocalized tweets published in the study area (Concepción metropolitan area) between 1st January and 31 March 2016, which partly covers two different seasonal behaviours, that is, summer vacation patterns (January–February) and normal business and students’ behaviour (March).

Data treatment included the removal of identical tweets referring to the same emergency phenomena and at the same location, since this information is not relevant for the purpose of this research, that is, identifying the use of public space. In addition, user accounts with more than 250 tweets over the whole period (>2.7 tweets per day on average) were checked in order to remove those not corresponding to individuals (i.e. Twitter accounts devoted to disseminate news or emergency issues). This led to the removal of 17 of the 26 top active users in the entire metropolitan area. Table 1 already allows the identification of a potential bias in using Twitter as a proxy for individual mobility, although only the latter are present in the final map.

The generation of mobility maps implies the need to set Twitter users’ place of residence, which was done assuming they live in the district where they tweeted the most between 22:00 and 07:59 on regular weekdays (Monday to Thursday). Our methodology includes an exploratory analysis of the spatial autocorrelation between official resident-based data and our estimated place of residence prior to the flow maps representing mobility patterns. This way we aim to confirm the null hypothesis that the number of users of public spaces from a particular place is not directly proportional to that place population volume and distance, and that this relationship varies across the space.

In particular, we used global and local bivariate Moran’s spatial autocorrelation Index, which indicates whether there is a strong or weak relationship between high (and low) population volume and Twitter users at the district level. This index was developed by Anselin, Syabri, and Kho (2010) and was computed using their GeoDa software.

The mobility analysis was made in base to those users that tweeted at night on working days at any part of the metropolitan area, and during the day and weekends within the most popular public spaces according to local researchers. Data treatment and map production was done with the commercial software ArcGIS 10.3.1 and Simantel’s Flow Map Generator (Simantel, 2012) toolbox.

We followed Shelton (2016) and proceed to normalize our data to a tweet usage baseline. The mobility analysis was therefore performed with ‘raw’ data (i.e. number of Twitter users moving from one district to a particular public space) and with normalized data (i.e. the proportion of Twitter users from one district to a particular public space). The comparison between the former (number of people moving) and the latter (the impact of each public space on each district) allows the identification of a potential bias in using Twitter as a proxy for individual mobility, although only the latter are present in the final map.

### Table 1. Main figures of downloaded geolocated tweets in Concepción, Chile.

| Area                  | Total tweets | Valid tweets | Users | Users that moved* |
|-----------------------|--------------|--------------|-------|-------------------|
|                       | No.          | No. | % | No. | No. | % |
| Regional area         | 52,536       | 37,838 | 72.02 | 4113 | 2258 | 54.90 |
| Metropolitan area     | 52,345       | 37,708 | 72.04 | 4101 | 2255 | 54.99 |
| Central area          | 21,568       | 17,422 | 80.78 | 2494 | 1695 | 67.96 |

*Users that tweeted more than once and from different locations.
Source: Own elaboration from data obtained from Twitter Public API between 1 January 2016 and 31 March 2016, local time (GMT-3).
services, retail and most employments are located; while the other categories refer to more specialized areas. Malls includes the biggest shopping centres of MAC, which are: Mall del Centro, Mall Mirador Costanera and Mall El Trébol; Leisure are places to practice sports as a Club Hípico, and Soccer Stadium ‘Ester Roa’, or gastronomic landmarks as Diagonal-Perú and Plaza España. Campus locations represent the main Universities of MAC, which are UdeC (Universidad de Concepción), UBB (Universidad del Bío Bío), USS (Universidad San Sebastian), UCSC (Universidad Católica Santísima Concepción), UFSM (Universidad Federico Santa María) and DUOC (Technical education of Universidad Católica de Chile), while Transport collects national passenger infrastructures: two bus stations (Collao Terminal and Camilo Henríquez Terminal), and Carriel Sur national airport. Finally, public places under the category of Parks include some large green open areas: Cerro Caracol, Laguna Tres Pascualas, Parque Ecuador and Parque Jurásico.

The exploratory analysis of spatial autocorrelation evidences a weak relationship between resident population (according to district delimitations of the latest Census, 2002) and the estimated place of residence of Twitter users that visited the selection of public spaces (Figure 2, left). This indicates that, globally, the resulting flows between place of residence and public spaces are not influenced by the population volume in each census district. Figure 2’s right panel shows the spatial pattern of this relationship, thus adding useful insights. Districts in red are those with high volume of both variables, while light blue indicates low values of population surrounded by high volume of Twitter users that visited the public spaces. These two categories are statistically significant in the northern and central part of Concepción metropolitan area. On the contrary, the southern part, which is a rural area, registers statistically significant low values of both variables, with the exception of Santa Juana district (high population, low Twitter users).

We then compared the results of non-normalized vs. normalized data in order to understand the mobility patterns to public spaces (Figure 3, top vs. bottom). These maps show the number of Twitter users in each district that visited each public space (top, non-normalized data) vs. the impact of each public space on residential areas by district (bottom, normalized data). Line thickness grows proportionally with the number of users/size of the impact. Lines start at the furthest district in a given direction and accumulate the flow from nearby districts which lines merge together (like a hydrological basin). The scale growth is constant in all maps, thus making them easily comparable. We produced generalized lines that are topological and quantitative accurate, and that maximize map simplicity and readability.
CBD’s impact is much larger than any other public space, thus deserving a prominent place. The rest of the public spaces are grouped in maps by categories. Colours representing individual public spaces in each map have been carefully chosen from ColorBrewer qualitative ramps (Brewer, 2002) in order to ensure equitable visibility to each one.

In general terms, the resulting maps, especially the ones showing the impact of each public space over the districts, reproduce current mobility patterns towards public places, based on the expertise of the authors’ local knowledge. As expected, the CBD is the main attraction place with users from all districts and a greater impact on the northwest and south part.

Figure 2. Twitter users that visited public spaces by district of residence and resident population scatterplot (left) and local Moran’s I (right). Source: own elaboration from data obtained from Twitter Public API between 1 January 2016 and 31 March 2016, local time (GMT-3) and 2002 Census (Instituto Nacional de Estadística de Chile INE).

Figure 3. Twitter users’ flows from estimated resident district to public spaces. Source: own elaboration from data obtained from Twitter Public API between 1 January 2016 and 31 March 2016, local time (GMT-3). Colour symbols: ColorBrewer.org.
of the metropolitan area (flows from Talcahuano, Penco, San Pedro de la Paz and Coronel). This mainly occurs due to the high concentration of main services of MAC within this area. The CBD is an important employment centre and it is also the meeting place for people from different income neighbourhoods (Rojas et al., 2015).

Normalized flows (also plotted in the Main Map) indicate that Trebol mall (the largest in size, with more than 300 shops), the furthest away from the city centre and the easiest to reach by car, has the largest impact on most districts, including the city centre. Usually, Trebol mall has been indicated as a new centrality within MAC favoured by the low of land price in the decade of nineties and high accessibility by car both from Talcahuano and Concepción (Pérez Bustamante & Salinas Varela, 2007; Rojas Quezada, Muñiz Olivera, & García-López, 2009). Indeed, this centrality is now being enhanced by an increasing number of in dwellings, roads, hotels, colleges and business buildings. On the contrary, Centro mall shows a sharp decrease in its impact outside the inner city boundaries (i.e. walkable and with higher transit density), and the Mirador mall extends most of its impact along the northwest-southeast axis, which is coincident with the railway and bus network.

University and college campus also concentrate a high proportion of Twitter visitors with differences across the metropolitan area. Universidad San Sebastián (USS), on the northern part of the city centre, gets the highest impact from its surroundings and, to a lesser extent, from the western area. In the second position, University of Concepción (UdeC) has a large impact that spreads in all directions. This might be related to the proximity of these campuses to natural areas (Laguna Tres Pascualas for USS and Parque Ecuador for UdeC, which are the largest and most accessible green areas for students), given the size and the recreation role of these campuses. In addition, USS and UdeC are the largest campuses in terms of number of students, both of them having more than 20,000 students, while DUOC and UBB’s, with around 10,000 students, see their impact decrease sharper with distance. UdeC’s campus is particular because it is an open access campus, and is usually used for recreational and cultural activities for citizens during the weekends.

Leisure areas not related to retail activities but to entertainment and gastronomy receive fewer visitors. In some cases, the impact is clearly concentrated on some specific parts of the metropolitan area. For example, Diagonal-Perú, an outdoor bar and restaurant area in the city centre, has a higher impact on the northern districts, whereas Casino, a private entertainment area in the northern outskirts, impacts especially in the districts located on the high-income residential area north of the city centre and on the southwest part of the metropolitan area. Both areas also receive a larger impact from the airport than other districts, which is coincident with the higher income of these new neighbourhoods, and entertainment facilities associated (Casino), that can better afford air travel. Other places, like Lenga, Plaza España or Laguna Grande have an impact that decreases with distance in a consistent way. In the case of parks, it is clear that Parque Ecuador, the one with a clear urban character, is the only one with a hinterland covering the whole metropolitan area. Similarly, Terminal Collao bus station has an extended impact which is more intense towards the north and west.

4. Conclusions

In this research, we used a 3-month inter-season sample data to investigate the potential of geolocated tweets to map mobility patterns in a medium-sized city. We were interested in mapping the mobility patterns associated with different types of public spaces in order to unveil potential areas of social exclusion: public spaces which are less visited by people living relatively near, which prefer further spaces.

Our methodology is based on GIS and it includes an exploratory bivariate spatial autocorrelation analysis to help the interpretation of the results as well as flow maps to compare normalized vs. non-normalized data.

Our results suggest that Twitter users serve as a proxy for population in urban areas, which served to evaluate the reliability of the dataset. In addition to that, exploratory analysis indicates that the spatial distribution of public space users differs from population density, with a significant gap between the northern and central metropolitan area, and the southern part in terms of Twitter usage.

Flow maps between the estimated district of residence and a selection of public spaces were then computed and represented in the form of schematic lines representing accumulative flows (raw data) or impact (normalized data) towards each public space. According to local experts, maps showing normalized data (i.e. the impact of each public space over each district of residence) represent close-to-reality mobility patterns (Main Map). One clear example is the case of the malls, which impact’s spatial pattern is similar to the transport network of the mode that provides higher accessibility in each case. These results are of great value to complement current mobility surveys and other data sources.

Further research includes the validation of a 6-month sample with the results of a recent origin-destination survey and prospections to produce finer time and spatial resolution.
Software
We used several technologies in each step. First, data collection was based on the Tweepy Python library (Roesslein, 2009), which was modified to fit our needs. Data treatment to convert geolocated tweets to a GIS point layer was done with Stata 12 and ESRI’s ArcGIS 10.3.1.

Map design
These maps represent the strength of the relationship between each district and the most popular public spaces in the metropolitan area of Concepción. The novelty is in the data source that has been used: we have extracted the origin-destination matrices solely from big open data (i.e. Twitter geolocated messages). Each line represents the cumulative proportion of users that resides in an origin district and has visited a particular public space. We did not intend to reproduce the exact route. Instead we produced generalized lines that are topological and quantitative accurate, and that maximize map simplicity and readability.

Acknowledgements
Special thanks must be given to Gaspar Cascallana, Guillermo Dorado and Borja Moya-Gómez for their valuable help to access the data, and to the reviewers that helped to improve the original map and manuscript.

Disclosure statement
No potential conflict of interest was reported by the authors.

Funding
The authors would like to thank Santander Universidades España for financing a three-month mobility grant to Universidad de Concepción (Chile) and the Ministerio de Economía, Industria y Competitividad of Spain and Universidad Complutense Madrid for the funding provided for a post-doctoral fellowship (FPDI-2013-17001). This research is also supported by Centro de Desarrollo Urbano Sustentable CEDEUS (FONDAP CONICYT 15110020), research project TRA2015-65283-R (Ministerio de Economía y Competitividad of Spain), and programme S2015/HUM-3427 (Comunidad de Madrid and European Structural Funds).

ORCID
Maria Henar Salas-Olmedo http://orcid.org/0000-0003-3234-1478
Carolina Rojas Quezada http://orcid.org/0000-0001-9505-4252

References
Anselin, L., Syabri, I., & Kho, Y. (2010). GeoDa: An introduction to spatial data analysis. In M. M. Fischer & A. Getis (Eds.), Handbook of applied spatial analysis: Software tools, methods and applications (pp. 73–89). Berlin: Springer Berlin Heidelberg. 10.1007/978-3-642-03647-7_5
Brewer, C. (2002). ColorBrewer. Retrieved July 29, 2016, from http://www.colorbrewer.org
Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., & Ratti, C. (2014). Geo-located Twitter as proxy for global mobility patterns. Cartography and Geographic Information Science, 41(3), 260–271. 10.1080/15230406.2014.890072
Huang, Q., & Wong, D. W. S. (2015). Modeling and visualizing regular human mobility patterns with uncertainty: An example using Twitter data. Annals of the Association of American Geographers, 105(6), 1179–1197. doi:10.1080/00045608.2015.1081120
Leetaru, K., Wang, S., Padmanabhan, A., & Shook, E. (2013). Mapping the global Twitter heartbeat: The geography of Twitter. First Monday, 18(5). doi:10.5210/fm.v18i5.4366
Li, L., Goodchild, M. F., & Xu, B. (2013). Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. Cartography and Geographic Information Science, 40(2), 61–77. doi:10.1080/15230406.2013.777139
Liu, J., Zhao, K., Khan, S., Cameron, M., & Jurda, R. (2015). Multi-scale population and mobility estimation with geotagged Tweets. 2015 31st IEEE International Conference on Data Engineering Workshops (ICDEW) (pp. 83–86). doi:10.1109/ICDEW.2015.7129551
Pérez Bustamante, L., & Salinas Varela, E. (2007). Crecimiento urbano y globalización: transformaciones del Área Metropolitana de Concepción, Chile, 1992–2002. Scripta Nova. XI(251), 1–33.
Roesslein, J. (2009). Tweepy. An easy-to-use Python library for accessing the Twitter API. Retrieved from http://www.tweepy.org/
Rojas, C. A., Carrasco, J. A., Pérez, F., Araneda, G., & Lima, C. (2015). Análisis de Redes Sociales. Paper presented at XV Conferencia Iberoamericana de Sistemas de Información Geográfica. Valparaíso.
Rojas Quezada, C., Muñiz Olivera, I., & García-López, M. Á. (2009). Estructura urbana y policentrismo en el Área Metropolitana de Concepción. Eure (Santiago), 35(105), 47–70. doi:10.4067/S0250-71612009000200003
Shelton, T. (2016). Spatialities of data: Mapping social media ‘beyond the geotag’. GeoJournal, 1–14. doi:10.1007/s10708-016-9713-3
Shelton, T., Poorhuis, A., & Zook, M. (2015). Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. Landscape and Urban Planning, 142, 198–211. doi:10.1016/j.landurbplan.2015.02.020
Simantel, B. (2012). Flow Map Generator. ArcGIS toolbox. Retrieved from http://www.arcgis.com/home/item.html?id=3bb50838ee0497df18f83b116cb8c1
Sobolevsky, S., Bojic, I., Belya, A., Sitko, I., Hawelka, B., Arias, J. M., & Ratti, C. (2015). Scaling of city attractiveness for foreign visitors through big data of human economical and social media activity. 2015 IEEE International Congress on Big Data (pp. 600–607). doi:10.1109/BigDataCongress.2015.92
United Nations. (2014). World Urbanization Prospects: The 2014 Revision, Highlights (ST/ESA/SER.A/352). New York, NY: Author. doi:10.4053/ DemRes.2005.12.9
Villagra-Islas, P., & Alves, S. (2016). Open space and their attributes, uses and restorative qualities in an earthquake emergency scenario: The case of Concepción, Chile. Urban Forestry & Urban Greening, 19, 56–67. doi:10.1016/j.ufug.2016.06.017