Urban Activity Detection Using Geo-located Twitter Data

GI_Forum 2020, Issue 1
Page: 15 - 31
Full Paper
Corresponding Author:
anna.kozlowska@ait.ac.at
DOI: 10.1553/giscience2020_01_s15

Anna Kozlowska[1] and Klaus Steinnocher[1]
[1]Austrian Institute of Technology GmbH, Austria

Abstract

More and more studies are based on freely available social media data. Using microblogs, a midpoint between instant messaging and content production, analyses of urban activities are possible. This paper focuses not only on mapping human activities but also on defining urban function in the city. Using geotagged Twitter data, the research carried out separate spatial and temporal analyses, in conjunction with combined spatio-temporal analyses. Tweets were categorised into six activity groups: *Working*, *Eating*, *Shopping*, *Leisure*, *Home* and *Education*, based on selected keywords. The results show stronger performance for the detection of *Leisure*, *Eating*, *Shopping* and *Education* activities and less successful performance for *Working* and *Home* activities. The first four cluster near the centre of the city, while the rest are scattered all over the city. Moreover, each activity shows its own temporal pattern. This study finds characteristic patterns for everyday activities and shows the possibility of using social media data to define urban function for places where land-use information is not available.

Keywords:

geotagged Twitter data, spatial analysis, temporal analysis, urban data, social media

1 Introduction

Leveraging location-based data offers new perspectives on, and better understanding of, events taking place in the world. Studying human behaviour and activity using social media was not possible a decade ago. That changed when social networks grew and the use of the Internet increased. The availability of data has made Twitter one of the most popular data sources for scientific research. With 320 million accounts creating over 500 million messages a day (*The Number of tweets per day in 2019*, 2019), Twitter is one of the largest social networks. It is also one of the preferred platforms for large-scale studies of human behaviour, thanks to its openness, global range, and the large number and variety of its users (Steinert-Threlkeld, 2018). Microblogs such as tweets help to validate socio-economic theories, predict social phenomena, or find spatial, temporal or thematic patterns in society. Moreover, according to Juhász & Hochmair (2019), among social-media microblogs, tweets relate the best to locations of daily activities.

While most of the literature focuses on the text of the tweets (Miller, 2011), few studies use the geographic information attached to the tweets (Hawelka et al., 2014; Leetaru et al., 2013). Geotagged text strings from Twitter are used mostly in research on social relationships and human dynamics. For example, they can be a support for analysing spatio-temporal patterns of happiness and public sentiment (Cao et al., 2018; Dodds et al., 2011; Nguyen et al., 2016), mobility (Hawelka et al., 2014; Kurkcu et al., 2016), or crime counts (Vomfell et al., 2018). Several studies use geo-located social-media data to track activities. Sakaki et al. (2010) used tweets to detect certain big events, like earthquakes or accidents, by searching for keywords related to the events. A different approach was presented by Martín et al. (2019), who used top tweeted words to obtain a clear idea about activity on a specific day. Zhang et al. (2018) used geo-tagged photos collected from social media to learn about principal tourist destinations.

Contrary to previous studies, the work presented here focuses more on detecting everyday activities than looking at extraordinary events. The objective is to learn about urban function in different parts of the city in order to determine urban land use. Land-use classification using social media data has already been carried out by Jiang et al. (2015), but their work was based on POI data rather than textual information.

The work closest to our approach was done by Andrienko et al. (2013). Although they used spatial and temporal clustering to analyse different activities, the choice of activities was based on the most frequently used keywords in their dataset. Also, unlike our study, it did not show hourly or daily spatial distributions. In pursuing the goal of this study, spatial, temporal and spatio-temporal descriptive analyses of geo-located data from the City of Manila, Philippines were carried out. The spatial analyses are based on Kernel density estimation. As shown in previous studies, this method is useful when analysing changes in density distribution of chosen events (Ma et al., 2009; Polonczyk & Lesniak, 2018; Zhang et al., 2009).

The paper is organised as follows: Section 2 outlines the data collection, pre-processing and activity classification; in Section 3, we present the analyses and results; Section 4 provides discussion and concludes the paper.

2 Methodology

2.1 Data collection

Twitter, as described by its owners, is 'what's happening in the world and what people are talking about right now' (*Twitter About Page*, 2019). It is an online social networking service where anyone can post short text messages ('tweets', max. 280 characters) and interact. Communication takes place in real time, by posting a message, commenting on a message, or redistributing another user's message (retweet). The message may also include a picture, a video or a link. Certain information can be marked with a hashtag '#', which facilitates the search for tweets within a chosen topic.

The study is based on geo-located Twitter data, which means that only tweets with an assigned location are used. On Twitter, the location can be set automatically by activating the precise

location option from the user account or the mobile device; alternatively, it can be set manually, each time a post is uploaded, by selecting a location from a predefined list. While the first option gives precise information on longitude and latitude, the degree of precision for the second ranges from the name of a city or neighbourhood to that of a specific public place (e.g. the name of a restaurant or other point of interest recognised by the Twitter service). Only tweets with original content can be georeferenced. Retweets, which are not classified by Twitter as original content, cannot be geotagged.

The Twitter data was acquired using the streaming API (Application Programming Interface) and the R Studio environment, with twitteR and streamR packages. The connection to the Twitter Search API was created through a Twitter account and Access-Token. Twitter provides data encoded in JavaScript Object Notation (JSON), which is based on key value pairs with named attributes and related values. All core attributes that accompany the tweet are encapsulated in that format. Each record stores the text of the tweet, the exact time of its publication and, in this case, information on geolocation. Whenever a tweet is georeferenced, a combination of the JSON keys 'geo', 'coordinates' and 'place' is filled with values. Specifically, each geo-tweet contains exact coordinates (longitude/latitude) in WGS84 as a single point.

The subject of this study was the City of Manila, Philippines. Data was collected for 9 months (20.06.2016 – 03.04.2017), with a total number of 608,667 tweets. The datasets were stored in CSV (Comma Separated Values) files.

Manila is a 'perfect' use case for any twitter data analysis. It is the world's most densely populated city, with an area of 42.88 km² and 1.78 million inhabitants (*Manila Population*, 2019). At the time of this study, the Republic of the Philippines was one of the most Twitter-active spots in the world (Figure 1), with approximately 200,000 tweets posted per day (*The one million tweet map*, 2017).

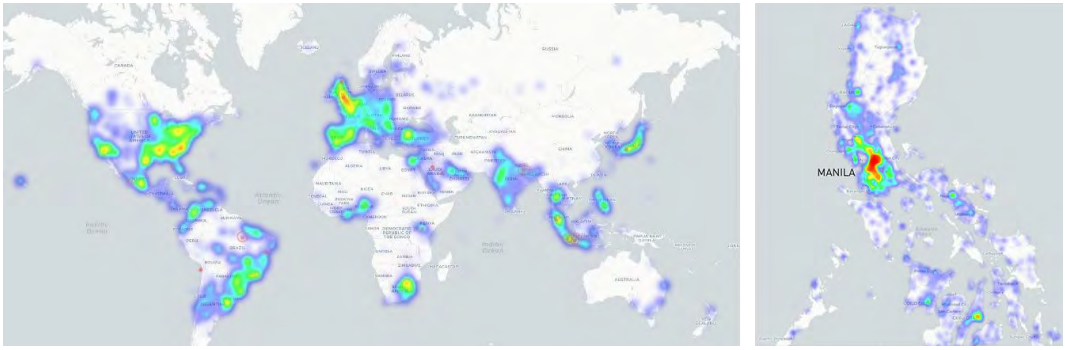

Figure 1: Twitter activity around the world (The one million tweet map, 2017).

2.2 Data processing

As argued by Symeonidis et al. (2018), pre-processing is a necessary and very important step in any analysis of text strings. Before addressing the actual content of the tweets, the geo-

location information was verified. All tweets lacking spatial reference and all location outliers (tweets lying outside the area of interest) were removed from the dataset.

In general, social media data are characterized by a large amount of noise. This noise includes all the special characters and punctuation embedded in a text string. In order to perform a successful classification and limit erroneous results, a clean text string is essential. Therefore, it is necessary to carry out several steps of text cleansing (Hangya & Farkas, 2013; Symeonidis et al., 2017). As most studies use tokenization – dividing the text string into separate words (Balazs & Velásquez, 2016) – this approach was also applied here. Furthermore, it was necessary to detect and delete duplicates and retweets. Later, the following techniques were used:

- Removing Unicode characters like comma (u002c) and unnecessary characters <, >, “, $
- Removing URLs which are part of most tweets but are not useful for the analysis and might also release sensitive information
- Unifying user tags (user account preceded by the ‘@’ symbol)
- Removing whitespaces
- Removing ‘#’ (commonly used on Twitter to categorize tweets).

After normalization, the next step was to identify and remove spam messages. A large group of tweets deemed not to be useful for this study were those generated automatically. Two types were detected and removed: tweets created by a bot (web robot), e.g. job offers and weather forecasts; tweets created automatically though external web sources like apps for music, running or games.

The final preparatory step was sorting date and time information. In this dataset, the time zone had to be corrected, from Greenwich Mean Time, by adding 8 hours (GMT+8). The time formats were normalized and additional information about the month (January–December), number of the week (1–52) and day of the week (Monday–Sunday) was assigned.

In the Philippines, there are around 200 unique languages and dialects, and two official languages: Filipino (Tagalog) and English. Therefore, before the text analysis the predominant language for all tweets was identified. If the tweet was too complex or there was no leading language in the text, the dominant language was not defined. The language detection showed that at least 2/3 of tweets in the dataset were written in English (Figure 2). The other 1/3 of the tweets may still contain English words. Hence, it was decided to proceed with text analysis for the entire dataset in English.

Figure 2: Results of language detection

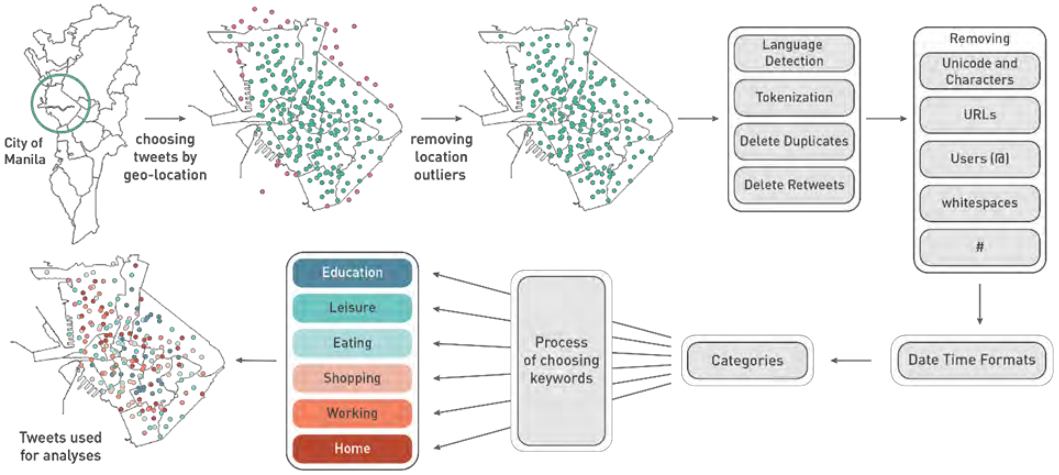

Figure 3: Cleansing data and activity classification process

2.3 Activity classification

Activity mapping and pattern analyses were based on tweets allocated to one of the six chosen activities: *Education*, *Eating*, *Leisure*, *Working*, *Home* and *Shopping*. *Education* covers studying at the locations of universities and schools; *Leisure* includes indoor and outdoor free-time activities, e.g. music events, cinema; *Eating* refers to eating and drinking, and places like restaurants, bars and cafés; *Shopping* refers to buying products in certain locations, e.g. malls; *Working* and *Home* refer to posts from work or home respectively, and do not match to any of the other groups. The allocation was done by choosing the relevant keywords for each activity. In this step, manual classification was chosen over automated text detection techniques. According to Hahmann (2014), classification of tweets by humans, although a subjective process, is more accurate.

The first simplified attempt to categorize tweets gave misleading results. Some of the tweets were wrongly assigned due to an ambiguity of individual word combinations. Expressions like 'after' or 'before' in tweets like 'Lunch before going to work' might suggest an action not necessarily happening at the place the tweet is posted. Phrases like 'going to', 'on my way', 'off to' express movement rather than an activity. Posts like 'Shopping for office supplies' are incorrectly classified to more than one group (*Shopping* and *Working*). Expressions such as 'feel like home' or 'second home' ought not to be classified as *Home*, just as 'working future' or 'work angels' do not concern actual *Working*. Moreover, activities related to *Eating*, *Shopping*, *Education* or *Leisure* are not considered if a location fitting another activity is included, e.g. 'Having lunch at University'. To correct these errors, a further group of keywords and phrases to be excluded from activity groups was created. Figure 4 shows a sample of keywords included and excluded from groups. As a result, only 14% of tweets (86,007 tweets) were successfully categorized (Figure 5).

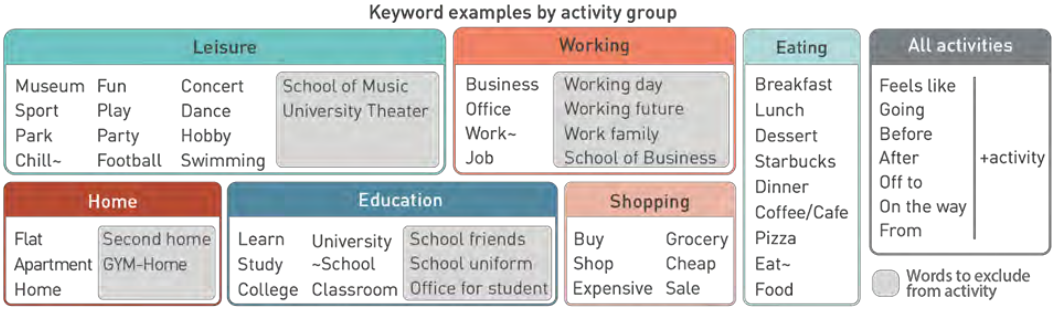

Figure 4: Selection of keywords for each activity group

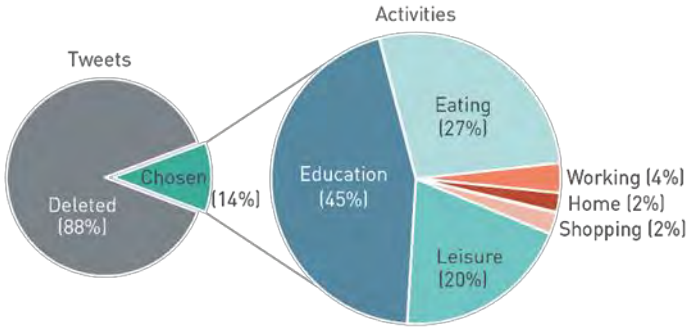

Figure 5: Final dataset used for analysis

3 Spatial and temporal analyses

Once the tweets have been correctly sorted into the six activity groups, the data is ready to be explored and analysed within the spatio-temporal context. The dataset stores information about date and time, as well as longitude and latitude. Thus, both temporal and spatial analyses can be performed. The results from the analyses presented below are therefore divided into three types: temporal analysis, spatial analysis and spatio-temporal analysis. By carrying out these analyses separately, we can gain an insight into when the activities are more present in the city, where people spend time depending on the activity, and how the patterns change throughout the day or week.

3.1 Temporal analysis

Looking at the total number of tweets and their temporal distribution, it is clear from the results that the numbers of posts vary throughout the period analysed (Figure 6). It can also be noted that for some dates data are missing (gaps in Figure 6). In general, most of the peaks are observed on Saturdays, when Leisure appears to be more present. The fluctuation in data can also be explained by the Education cycle of school holidays and the major exam period. Some changes might be related to national holidays or celebrations. It is interesting to note that Eating appears to have the least fluctuation throughout the period investigated.

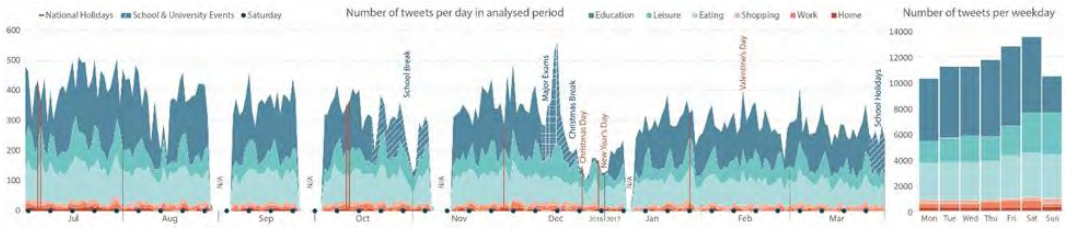

Figure 6: Number of tweets and major events taking place in Manila during the period analysed

The temporal patterns of each activity group were analysed according to daily, weekly and monthly distributions. It was expected that each activity group would have its own characteristic temporal pattern. Using the example of *Home* and *Working*, these activities were expected to have opposite patterns, because most people live and work in different places. Another expectation was that tweets related to *Education* or *Working* would be more intense during traditional working hours (Monday to Friday, 8am-6pm), while all activities analysed were expected to reduce to zero during night-time hours.

The temporal analyses of the general weekly trend (Figure 7a) show an expected pattern. Starting in the morning, the number of tweets rises throughout the day, reaching a peak in the evening hours, especially on Fridays and Saturdays, and finally declining during the night. The results for each activity group (Figure 7b) show more differentiated patterns.

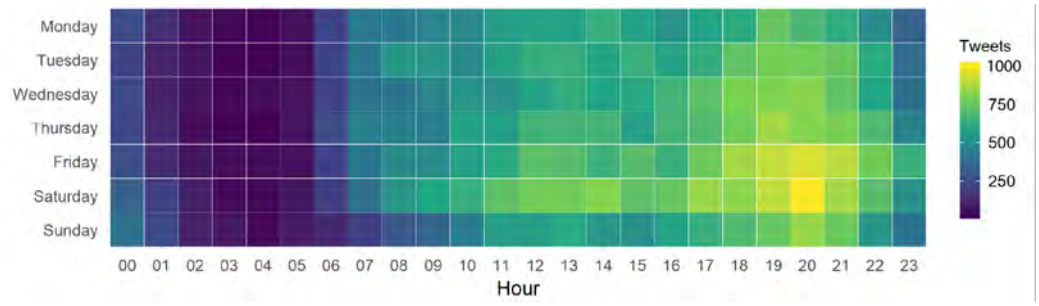

Figure 7a: Temporal heatmaps for each day of the week for all tweets

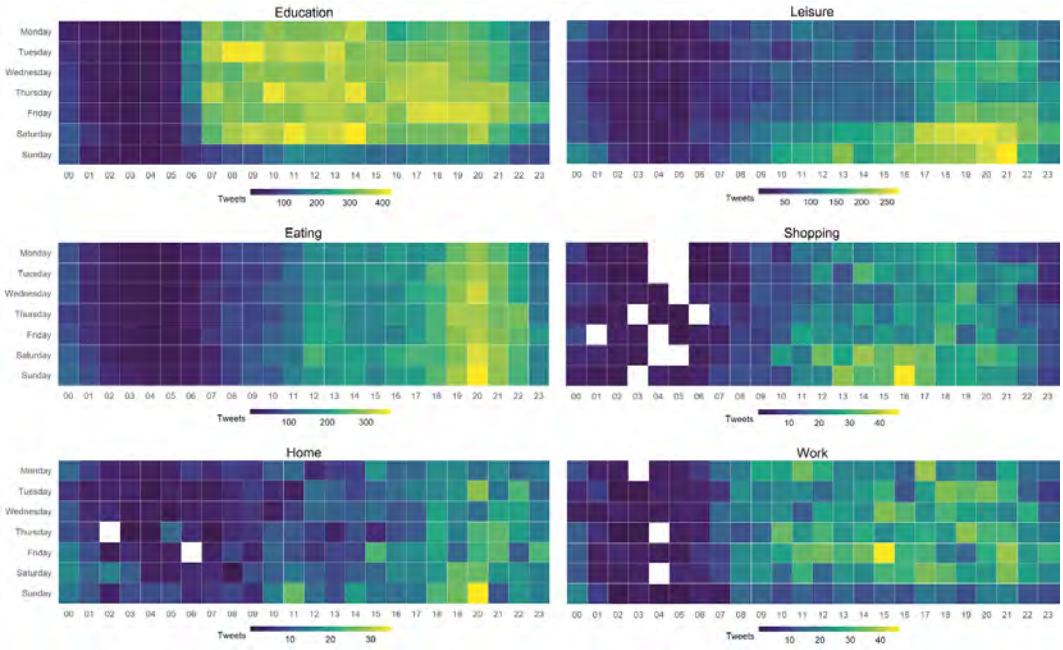

Figure 7b: Temporal heatmaps for each day of the week for each activity group

Tweeting activity related to *Education* is characterized by a regular pattern for the entire week excluding Sunday. Tweeting starts at 7am, with the number staying at a high level until the end of the day. There are no significant peaks or troughs. *Leisure* and *Eating activities* are also characterized by definable and stable, but contrasting, weekly-cycle patterns. For *Eating*, the first concentrated activity takes place between 12am and 2pm, and the main peak is between 7pm and 9pm regardless of the day of the week. By contrast, *Leisure* is characterized by significant peaks, mostly on weekend afternoons and in the evenings. *Shopping*, *Home* and *Working* show irregular temporal distribution. While *Working* differentiates between inactive Sundays and the rest of the week, the other two activities do not present a significant pattern.

3.2 Spatial analysis

To identify where activities take place across the city, a statistical analysis of spatial point patterns was carried out. The study focuses on visual analysis using 2D Gaussian Kernel density estimation, where a spatial relationship of tweets is visualized as a density surface using a graduated colour scheme. The result is a collection of density maps – heatmaps – with a spectrum of 'high' and 'low' point densities. The Kernel bandwidth in this case was based on a number of educated trial runs. The aim was to show the targeted distributions in a more interpretable way and to avoid over- or underfitting. First, the analysis was run for the entire dataset, then for each of the six activity groups separately.

It was expected that each activity would have its hotspots (clustering occurrences) in multiple locations throughout the entire city. This could reveal both overlaps and clear distinctions between activities. Social activities like *Leisure*, *Eating* and *Shopping* were predicted to take place

in close proximity to each other or even at the same location, and most likely in the centre of the city. *Education* activity was expected to show up around the main university campuses and schools, while it was anticipated that *Working* activity would be spread through the entire city, with a focus on the main business areas. *Home* activity was foreseen to be the most widespread activity, as citizens live in different parts of the city. To allow an informed analysis of the spatial distribution of tweets and their underlying localities, a land-use map of the City of Manila was used to compare the results of the analysis with the actual distribution of land use (Figure).

Figure 8: Land use map of City of Manila. Information source: City Planning and Development Office Manila (2017)

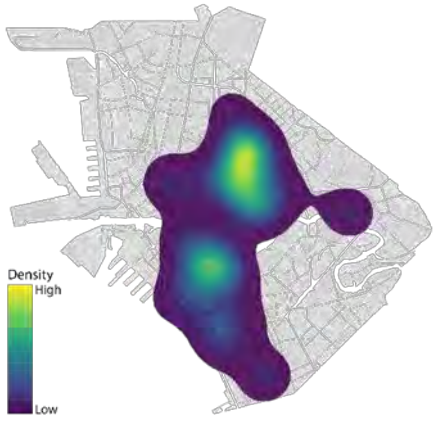

Figure 9: Heatmap for all activities combined

The density analysis shows that tweeting is not spread evenly across the city, with most tweets being located in the city centre (Figure). Analysis shows clear variations between categories (Figure). The most widespread activities are *Home* and *Working*. Tweeting from work occurs in most locations in the city. *Home* activity omits industrial and recreational areas along the riverbank and the northern industrial area. Moreover, *Home* hotspots do not overlap with *Working* activities. The third most widespread activity is *Eating*, covering almost half of the city, mostly where commercial, retail, recreational and institutional areas are located. Much more

concentrated and showing more details are *Shopping*, *Education* and *Leisure*. The hotspots for these activities were compared with corresponding places in the land-use map (Figure). *Education* tweets occur around the main universities and schools. *Shopping* coincides with areas where large shopping centres are located, but does not cover most of the commercial areas and local markets. *Leisure* is a very complex group as it comprises several kinds of activity. The analysis reveals that the main concentration is on the south side of the city, with a smaller spot in the north. These areas are where most of Manila's museums and parks are located. There is no spatial link detected for other entertainment and nightlife locations. The spatial analysis shows that there is a higher chance of finding a more precise location corresponding to *Education*, *Shopping*, *Leisure* or *Eating*, than for *Working* or *Home*.

Figure 10: Heatmaps for each activity: Shopping, Eating, Leisure, Working, Education, Home

Figure 11: Heatmaps for selected activity groups checked against identified locations in the city of Manila. Left to right: Education, Shopping, Leisure

The contours representing the extent of the spatial distribution of six activities were overlaid, as shown in Figure, a simplified urban function map which helps to narrow down the areas where everyday activities take place in the city.

Figure 12: Urban function map based on Twitter analysis

3.3 Spatio-temporal analysis

Finally, the combination of both temporal and spatial aspects was analysed. Using time stamps for each point of data allows the mapping of tweets for selected time periods. As for the previous analyses, this analysis focused on different time scales, for various activity groups. This time, it was expected that the spatial distribution of the tweets would differ depending on the time frame. It seemed more likely that more significant results would be obtained from daily or weekly distributions than monthly ones.

Figure shows monthly and weekly distributions of all tweets, for which there are no significant variations. They vary only slightly from one month to another, and between days of the week, showing slight differences in intensity for some areas. As the interest lies in differences between activities, the next step was to explore hourly differences depending on the results from the temporal analysis. Due to the irregularity in the size of the activity groups, the analysis was done only for *Eating*, *Education* and *Leisure*.

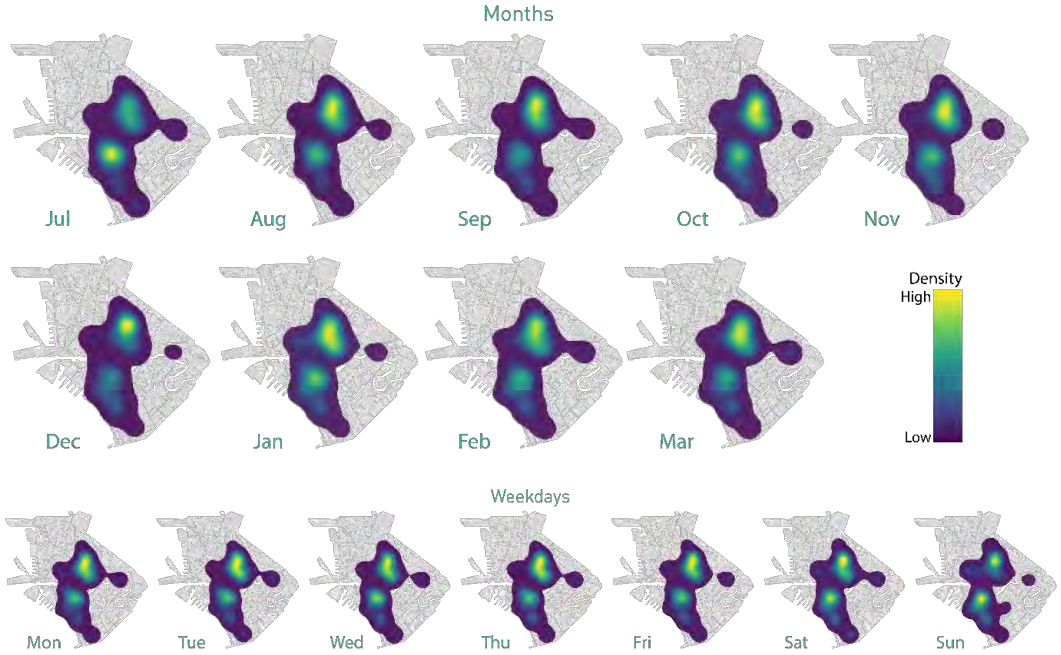

Figure 13: Heatmaps for all tweets, months (top) and weekdays (bottom)

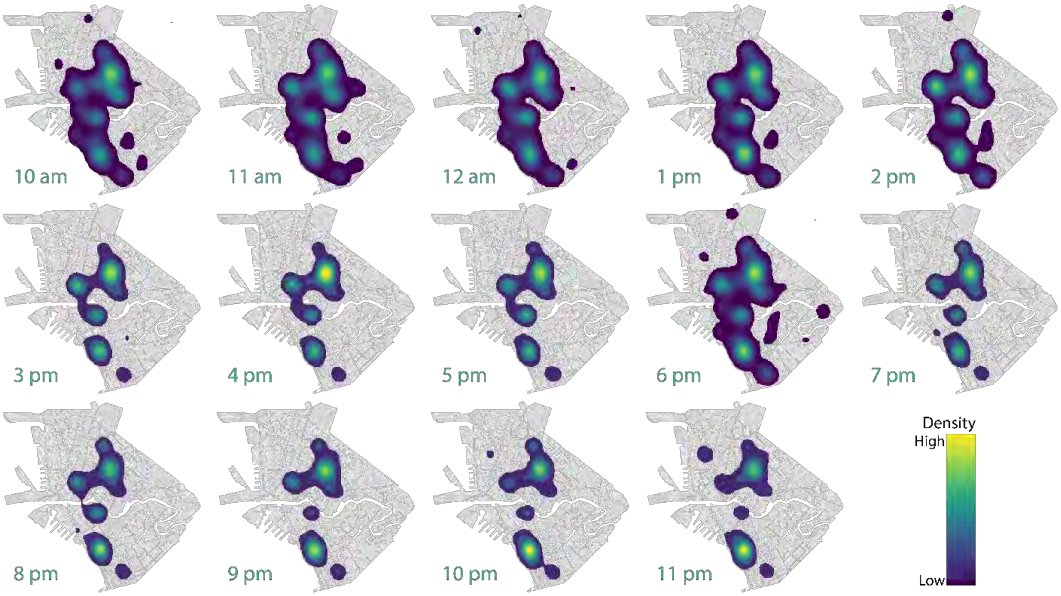

Figure 14: Heatmaps for Eating (10am to 11pm)

Eating and *Leisure* were examined for hourly patterns for a reference day derived as an average from the whole period under investigation (Figure 14, Figure 15). Results show that tweets

related to *Eating* are posted from a wider range of locations between 10am and 2pm and at 6pm. Their locations can be seen to intersect with the *Working* spatial distribution. Between 3pm and 5pm, and from 9pm to 11pm, the action is more focused on certain spots. Between 3pm and 5pm, *Eating* overlaps with *Education* and *Shopping* locations, with the strongest focus in the north where one of the universities is located. The second timeframe (9pm to 11pm) shows a pattern similar to *Leisure* and *Shopping*, with activity in areas where bars and restaurants are concentrated.

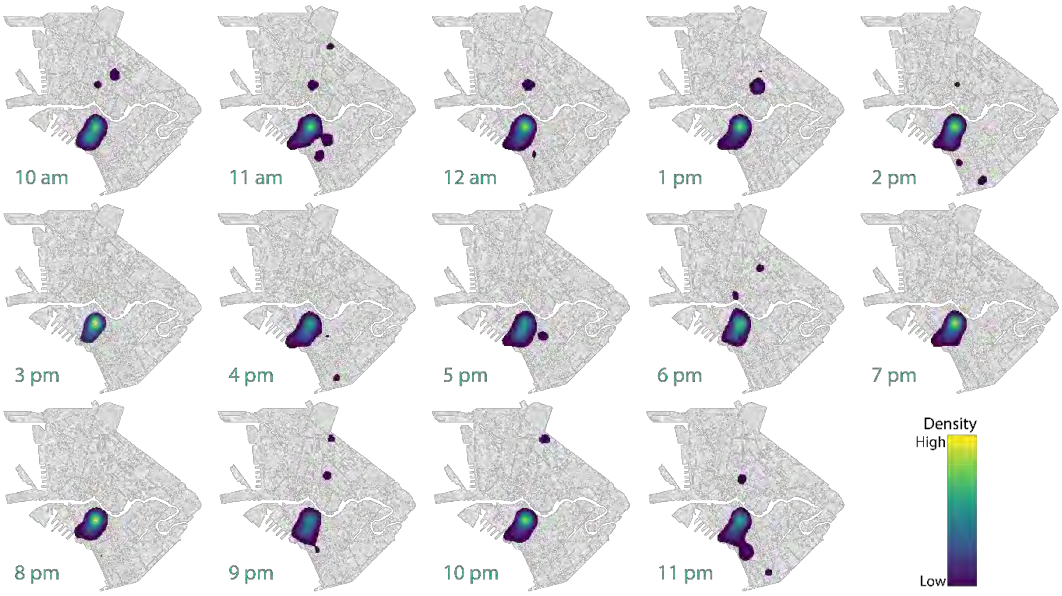

Figure 15: Heatmaps for Leisure (10am to 11pm)

The spatial daily distribution of *Leisure* is presented for a typical Saturday derived as an average over the whole period (Figure 15). The plot shows no significant hourly changes. The main hotspot identified earlier is still visible, with additional spots occurring with no apparent pattern. *Education* was divided into two temporal periods: 1) more activity from Monday to Saturday; 2) less activity on Sunday. *Leisure* shows more hotspots spread across the city, while *Education* has its focus mostly in the University area (Figure 16).

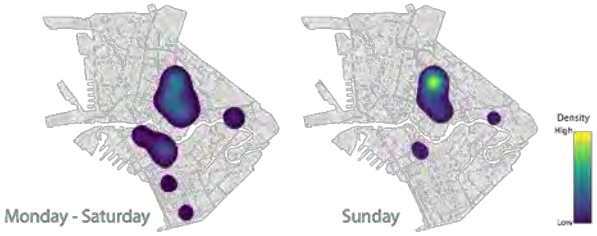

Figure 16: Heatmaps for Education

The results show the change in spatial distribution throughout the day or week but, in most cases, this distribution does not relate to the temporal patterns examined earlier. The intersecting spatial distribution of two or three activities suggests the hourly or daily changes of functionality of different parts of the city. *Working* or *Education* can be replaced by *Eating* during lunch breaks or by *Leisure* in the evening.

4 Discussion and Conclusion

The analyses presented here show the advantages of using data from social media for defining temporal and spatial patterns in the city. Results derived from analysing Twitter data can be used to determine urban function and list major activities taking place in certain parts of the city. The spatial aspect represents gatherings of people and main points of activity. The temporal aspect makes it possible to estimate an average time pattern for individual activities, which can be used to improve various aspects of the urban context, such as public transport or safety. The combination of time and location helps us to understand how the spatial distribution changes during the day. The study shows the effectiveness of using social media for detecting activities which are connected to social interactions or public spaces. It also shows that mapping *Education*, *Leisure* and *Eating* is more precise than mapping *Working* and *Home*; *Working* and *Home* had significantly fewer tweets than the rest of the activities. There are several reasons for this. Firstly, not everyone is willing to share their real location on social media, preferring instead to tag one of the locations from Twitter's predefined list of POIs. As Ludford et al. (2007) show, most people are likely to share about activities in public places, but they are not willing to share the exact location of their home or workplace. Moreover, most people are active on social media when they want to share exciting news, new locations or interesting events in which they are participating, and these are most likely to happen outside their work and home. Furthermore, tweeting about spending time at home or work is not particularly common among users of Twitter. The small sample size for *Working* and *Home* might have been a reason for the fluctuations observed in the temporal analyses.

The temporal and spatial patterns do not show a 1:1 relationship. It was expected that a temporal peak would be reflected in an even stronger peak in the spatial distribution. However, while there is an increase in tweets reflected in higher activity levels at individual hotspots, there is no significant increase in the number of activity hotspots.

Leisure differs from other groups because it combines more than one activity and can take place in different locations. It can be associated with morning or evening activities, or both. The activities can also be referred to as outdoor and indoor events. As a result, this category is very complex, which explains the lack of conclusive results in the *Leisure* analyses. This situation could be improved by predefining more activity groups related to leisure.

Although multiple studies show the strength of Twitter data for understanding urban processes, there are several drawbacks and limitations to using social media data in analyses. They depend on people being willing to share their opinions and feelings with the public. Some text messages comprise incomplete sentences or words reserved for certain social groups, which might be misinterpreted during the word classification. Moreover, the reliability of Twitter data analysis for cities like Manila is compromised by the high number of languages

used, some of which are not easily translated or widely understood. As this study was carried out entirely in English, the results might not reflect all activities typical for the region. According to Longley et al. (2015), analyses using microblogs do not represent the whole of society but only certain demographic groups. However, this should not have a severe impact on this study, as land-use classification does not necessarily depend on demographic or social groups, and thus does not require a complete representation of society. Another source of error are wrong locations assigned to tweets. Hecht et al. (2011) stated that for 34% of the tweets they analysed, the location was wrongly assigned, mainly due to the deliberate indication of a false location. Moreover, only 1% of tweets can be freely downloaded, while the full dataset is very expensive (Morstatter et al., 2013). Choosing only geo-located tweets reduces the sample size even more.

However, despite these limitations, the study brings a new perspective to using social-media data. Using Twitter data only, it is possible to learn about everyday activities taking place in a chosen area. Twitter data can be leveraged to provide information about hourly, daily or weekly patterns for common activities, especially those taking place in public spaces. More importantly, this study shows that by using Twitter data, it is possible to define urban function for places where land-use information is not available.

Acknowledgements

This study is a part of the project INTERSENSE funded by FFG, Vienna, in the frame of the Austrian Space Applications Programme, contract number 865977.

References

Andrienko, G., Andrienko, N., Bosch, H., Ertl, T., Fuchs, G., Jankowski, P., & Thom, D. (2013). Thematic Patterns in Georeferenced Tweets through Space-Time Visual Analytics. *Computing in Science Engineering*, *15*(3), 72–82. https://doi.org/10.1109/MCSE.2013.70

Balazs, J. A., & Velásquez, J. D. (2016). Opinion Mining and Information Fusion: A survey. *Information Fusion*, *27*, 95–110. https://doi.org/10.1016/j.inffus.2015.06.002

Cao, X., MacNaughton, P., Deng, Z., Yin, J., Zhang, X., & Allen, J. G. (2018). Using Twitter to Better Understand the Spatiotemporal Patterns of Public Sentiment: A Case Study in Massachusetts, USA. *International Journal of Environmental Research and Public Health*, *15*(2). https://doi.org/10.3390/ijerph15020250

City Planning and Development Office Manila. (2017). *Existing Land Use Map 2017*. https://upload.wikimedia.org/wikipedia/en/8/82/Existing_Land_Use_Map_of_Manila_2017.jpg

Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., & Danforth, C. M. (2011). Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PLoS ONE*, *6*(12), e26752. https://doi.org/10.1371/journal.pone.0026752

Hahmann, S., Purves, R., & Burghardt, D. (2014). Twitter location (sometimes) matters: Exploring the relationship between georeferenced tweet content and nearby feature classes. *Journal of Spatial Information Science*, *9*, 1–36. https://doi.org/10.5311/JOSIS.2014.9.185

Hangya, V., & Farkas, R. (2013). Target-oriented opinion mining from tweets. *2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom)*, 251–254.

https://doi.org/10.1109/CogInfoCom.2013.6719251
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. https://doi.org/10.1080/15230406.2014.890072
Hecht, B., Hong, L., Suh, B., & Chi, E. H. (2011). Tweets from Justin Bieber's heart: The dynamics of the location field in user profiles. *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI '11*, 237.
https://doi.org/10.1145/1978942.1978976
Jiang, S., Alves, A., Rodrigues, F., Ferreira, J., & Pereira, F. C. (2015). Mining point-of-interest data from social networks for urban land use classification and disaggregation. *Computers, Environment and Urban Systems*, *53*, 36–46.
https://doi.org/10.1016/j.compenvurbsys.2014.12.001
Juhász, L., & Hochmair, H. (2019). Comparing the Spatial and Temporal Activity Patterns between Snapchat, Twitter and Flickr in Florida. *GI_Forum*, *1*, 134–147.
https://doi.org/10.1553/giscience2019_01_s134
Kurkcu, A., Ozbay, K., & Morgul, E. F. (2016). Evaluating the Usability of Geo-located Twitter as a Tool for Human Activity and Mobility Patterns: A Case Study for New York City.
Leetaru, K., Wang, S., Cao, G., Padmanabhan, A., & Shook, E. (2013). Mapping the global Twitter heartbeat: The geography of Twitter. *First Monday*, *18*(5).
https://doi.org/10.5210/fm.v18i5.4366
Longley, P. A., Adnan, M., & Lansley, G. (2015). The Geotemporal Demographics of Twitter Usage. *Environment and Planning A: Economy and Space*, *47*(2), 465–484.
https://doi.org/10.1068/a130122p
Ludford, P. J., Priedhorsky, R., Reily, K., & Terveen, L. (2007). Capturing, sharing, and using local place information. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1235–1244. https://doi.org/10.1145/1240624.1240811
Ma, J., Yang, S., You, J., & Zhang, M. (2009). Spatial Pattern Detection and BP Neural Network Analysis of Bank Mesh Point in Urban Area. *2009 Fifth International Conference on Natural Computation*, *3*, 639–643. https://doi.org/10.1109/ICNC.2009.473
Manila Population. (2019). http://worldpopulationreview.com/world-cities/manila-population/
Martín, A., Julián, A. B. A., & Cos-Gayón, F. (2019). Analysis of Twitter messages using big data tools to evaluate and locate the activity in the city of Valencia (Spain). *Cities*, *86*, 37–50.
https://doi.org/10.1016/j.cities.2018.12.014
Miller, G. (2011). Social Scientists Wade Into the Tweet Stream. *Science*, *333*(6051), 1814–1815. https://doi.org/10.1126/science.333.6051.1814
Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. M. (2013). Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose. 9.
Nguyen, Q. C., Kath, S., Meng, H.-W., Li, D., Smith, K. R., VanDerslice, J. A., Wen, M., & Li, F. (2016). Leveraging geotagged Twitter data to examine neighborhood happiness, diet, and physical activity. *Applied Geography*, *73*, 77–88.
https://doi.org/10.1016/j.apgeog.2016.06.003
Polonczyk, A., & Lesniak, A. (2018). The Impact of Generalised Spatial Data on the Incidence Density of Selected Offences in Krakow. *2018 Baltic Geodetic Congress (BGC Geomatics)*, 328–334. https://doi.org/10.1109/BGC-Geomatics.2018.00068
Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: Real-time event detection by social sensors. *Proceedings of the 19th International Conference on World Wide Web*, 851–860. https://doi.org/10.1145/1772690.1772777
Steinert-Threlkeld, Z. C. (2018). *Twitter as Data* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781108529327

Symeonidis, S., Effrosynidis, D., & Arampatzis, A. (2018). A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis. *Expert Systems with Applications*, *110*, 298–310. https://doi.org/10.1016/j.eswa.2018.06.022

Symeonidis, S., Effrosynidis, D., Kordonis, J., & Arampatzis, A. (2017). DUTH at SemEval-2017 Task 4: A Voting Classification Approach for Twitter Sentiment Analysis. *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, 704–708. https://doi.org/10.18653/v1/S17-2117

The Number of tweets per day in 2019. (2019). David Sayce. https://www.dsayce.com/social-media/tweets-day/

The one million tweet map. (2017). http://onemilliontweetmap.com

Twitter About Page. (2019). https://about.twitter.com/en_gb.html

Vomfell, L., Härdle, W. K., & Lessmann, S. (2018). Improving crime count forecasts using Twitter and taxi data. *Decision Support Systems*, *113*, 73–85. https://doi.org/10.1016/j.dss.2018.07.00

Zhang, W., Tan, G., Lei, M., Guo, X., & Sun, C. (2018). Detecting tourist attractions using geo-tagged photo clustering. *Chinese Sociological Dialogue*, *3*(1), 3–16. https://doi.org/10.1177/2397200917752649

Zhang, Z., Li, J., & Liu, Y. (2009). GIS-Based Spatial Distributions and Evolvement Analysis of Urban Affordable Housing: A Case Study. *2009 International Conference on Environmental Science and Information Application Technology*, *2*, 419–422. https://doi.org/10.1109/ESIAT.2009.130