



Reading urban land use through spatio-temporal and content analysis of geotagged Twitter data

Aminreza Iranmanesh · Nevter Zafer Cömert · Şebnem Önal Hoşkara

Accepted: 7 February 2021 / Published online: 18 February 2021
© The Author(s), under exclusive licence to Springer Nature B.V. part of Springer Nature 2021

Abstract This study explores the possibilities of reading urban land use through geotagged social media data using temporal and content analysis. The advent of social media into the everyday life of cities has transformed the natural complexity of urban space. People's interaction with space and with the social context happens in a new hybrid space that is becoming a part of the reality of city life. The publicly shared content that people produce as a side product of their digital routine can be utilized for developing new analytical studies. Social media data is not merely a new method of analysis, but a window into the emerging urban processes. Hence, understanding the potential of social media data in urban studies could provide new tools for future urban planning. The current study investigates the legibility of urban land-use patterns through classifications of geotagged Twitter data, with the aim of exploring the degree of empirical viability of using social media data for urban design processes. With this aim in mind, the study

proposes a framework for utilizing geotagged Twitter metadata. The framework is tested in a university campus in the city of Famagusta in Cyprus. First, the study establishes a data collection and filtering method. Second, data synthesis and classification of the data using GIS and Kernel Density Estimation is explained. Third, the paper explores possibilities for combining the content analysis and temporal analysis and aims to find the best fit for reading urban land use. The outcome shows promising results in reading urban land use through geotagged data.

Keywords Spatio-temporal · Twitter · Geotagged · Land use · Urban analytic · Content analysis

Introduction

The physical form of cities has been analyzed thoroughly in the field of urban morphology (Kostof et al. 1999). Although urban morphology often focuses on the physical form of cities, the social forms are often seen by analyzing their effects on the other tangible aspects of urban structures like centrality or land use. Urban settings can be seen as an ongoing dialect between social and spatial forms, where one cannot be conceptualized without the other (Hillier and Hanson 1984). This mutual relationship between the social and spatial form of the city has been central

A. Iranmanesh (✉)
Faculty of Architecture, Near East University, Nicosia/
TRNC, Mersin 10, Turkey
e-mail: aminreza.iranmanesh@neu.edu.tr

N. Z. Cömert · Ş. Ö. Hoşkara
Department of Architecture, Eastern Mediterranean
University, Famagusta/TRNC, Mersin 10, Turkey
e-mail: nevter.zafer@emu.edu.tr

Ş. Ö. Hoşkara
e-mail: sebnem.hoskara@emu.edu.tr

in urban studies. The dynamic transformation in the placement of urban spaces in relation to each other constitutes a slow and almost intangible force that reshapes land use (W. G. Hansen 1959; Penn and Turner 2004; Hillier 2012). Nevertheless, the intensity of the more public land uses is subject to social constructs that are manifested in social networks, meaning that people are more likely to use a place that others are using or recommending (Hasan et al. 2013).

Today, emerging means of communication, and ease of access to well-connected mobile devices have reshaped the way people interact with the urban setting. Similar to how the industrial revolution changed the physical form of the city, the mobile digital revolution is transforming the urban socio-spatial landscape. Citizens of the contemporary city produce an unprecedented amount of data through their mobile devices. This new data which is often referred to as “Big Data” could be considered a representation of the everyday life of the city (Kitchin 2014).

Big data as a rising topic in academic circles provides a new layer of reading upon the existing data in urban areas which renders a more refined picture for analytical research approaches (Hao et al. 2015; Feick and Robertson 2015). In this light, it seems important to investigate the potential of the new data types in unveiling emerging aspects of urban analysis that have gone unnoticed due to the lack of proper data. The potential use of the new data types is not merely a new approach to urban analytics, but a window into how these communication mediums are changing the forms of urban life. Thus, the development of methodological frameworks seems crucial as the majority of people are going to live in urban areas and at the same time are using social media to interact with each other and with space (Crooks et al. 2016).

This study aims to build a framework for reading urban land use at the micro-level via temporal and content analysis of geotagged Twitter data by searching for an answer to its research question: how accurately could the new data illustrate urban land-use features? The study contributes to the literature by developing a methodological framework for reading the functional features of land use through temporal and content properties of geotagged social media on a small scale. This process tries to showcase a way to read and predict urban land use using a small section of

social-media data. Here, the temporal stamps and the content of the data is analyzed concerning land use.

The proposed framework is tested in the city of Famagusta in an area where the rate of Twitter users is relatively high due to a major university campus that brings together active social media users (Iranmanesh and Atun 2018). The framework illustrates the possibilities for the automated generation of active land use which could be beneficial in areas without official land-use data. Moreover, in order to minimize the biases, the study targets the students of a university campus and explores the relationship between a university campus and the active land use of its host city through the framework.

Urban socio-spatiality in the age of social media

Place, as a socially constructed entity, is associated with human meaning and may be thought of as a process, a “process of becoming” (Pred 1984). People are active participants in the historically contingent process of the making of place: within the context of their times, they construct places by associating them with human meaning. Constructed places are not confined to the here and now but include places of past experiences –memory- (Lowenthal et al. 1976), or even those which exist in simulations and iconographies (Cosgrove et al. 1988; Duncan and Ley 1993). Therefore, it can be argued that place is composed of diverse but overlapping images and interpretations.

Another characterization of place-identity comes from Lynch (1981) who defines it as ‘the extent to which a person can recognize or recall a place as being distinct from other places’. The legibility and digital communication network of the city determine the ease with which a city’s parts can be recognized and can be organized into a coherent pattern. A legible city evokes a strong image in any given observer (Lynch 1960). All the discussions above indicate how people cognate their environment by understanding the important key elements of urban structure. Today social media takes part in this process via platforms such as Facebook, Instagram, or Twitter in which people share their interactions.

In recent literature, understanding human behavior on a larger scale is often linked with user-generated volunteer data which is being constantly produced through mobile communication networks (Goodchild

2007; Sagl et al. 2012). Through these global networking mediums, people record and share the experiences of their everyday lives (Ferrari et al. 2011). The advent of social media and the integration of location-aware devices into the dynamic of human interaction has created a necessity for developing new methodological and theoretical frameworks aiming to understand urban socio-spatial structures (Luo et al. 2016; Chen et al. 2016).

An emerging breakthrough in urban science is happening regarding the integration of geographically tagged metadata with networking websites such as Twitter, Instagram, and Facebook. People showcase segments of their lives publicly through photos, videos, or microblogs, linking it with a location in the urban setting, and this has opened up new horizons in understanding the spatial implications of collective movements and activities (Wakamiya et al. 2012). This data provides a new layer of interaction with cities that are becoming more vibrant as a new generation of people is emerging with mobile devices as an inseparable extension of their daily routine (Shelton and Poorthuis 2019). This could be critical in planning for urban amenity points since the success of such places is directly related to how the public would utilize them (Rout and Galpern 2018; Yang and Durarte 2021).

In this light, the willing contribution of active observers is regarded as ‘People as Sensors’; a phenomenon that creates new possibilities for remote public participation in design and decision making (Goodchild 2007). A significant benefit of model development for utilizing geotagged data lies in the possibility of remote classification of urban areas by identifying recurring patterns (Wakamiya et al. 2012). The data can be instantly updated, or in some cases created when no official data can be easily found or collected due to unprecedented circumstances (Iranmanesh and Alpar Atun 2020a; Haffner 2019). Geotagged social media interactions can illustrate an accurate image of urban points of interest and centers (Sun et al. 2016).

Spatio-temporal reading of urban spaces through geotagged data

The dynamic vibrancy of urban movement has a strong temporal dimension and similarly, location-

based social media shows strong temporal patterns in predicting human behavior (Santa et al. 2019). Ríos and Muñoz (2017) show the potential of reading urban land-use themes through the density of cellphone data, but social-media data, with its rich intrinsic metadata including timestamps, can describe these temporal shifts in urban spaces (García-Palomares et al. 2018).

The spatio-temporal profile of urban areas is a critical design tool that creates an effective framework for increasing urban vibrancy throughout the day-night cycle by providing a window into how people occupy these spaces in various time cycles (C. Wu et al. 2018). Overall, the understanding of urban systems can be significantly improved by utilizing the spatio-temporal analysis of geotagged social media data (Goodchild 2007). Tracing the digital footprint of movements through urban space in real-time allows us to map inherent complexities that might otherwise be overlooked. The spatio-temporal image of the city through Twitter data can go beyond the general day-night cycle and display a detailed unseen spectrum of activities (França et al. 2016).

On a macro scale, the spatio-temporal analysis of social media data has been used in exploring urban location patterns and hotspots (Ferrari et al. 2011), patterns of urban vibrancy (C. Wu et al. 2018), land use and urban dynamics (García-Palomares et al. 2018), land-use prediction (Soliman et al. 2017; Frias-Martinez and Frias-Martinez 2014), urban motilities (Luo et al. 2016), and monitoring for crowd behaviour (Wakamiya et al. 2012), but the smaller scale detailed analysis remains understudied. This study looks at a small section of big-data to explore the micro-interaction and spatio-temporal patterns of urban land-use related activities.

The social media data (Twitter in this case) is subject to many criticisms regarding the biases that are a natural side-effect of such datasets. In using these datasets, these biases need to be addressed, the noise that they might impose on the outcome, and the degree of accuracy must be accounted for in the analytical studies (Shaw et al. 2016). It is argued that the user demographic profile of Twitter is not a random homogenous representation of the society as a whole, so the possibility of having a specific profile affecting the outcome must be addressed (Steiger et al. 2015). The data might be biased on the basis of education level (García-Palomares et al. 2018), age, gender, or ethnicity (Mislove et al. 2011; Hecht and Stephens

2014). The biases might even be extended to the geotagged data that might not be a random sample of the Twitter user profile.

Moreover, the platform that the data is collected from can be subject to socio-spatial heterogeneity. Martí et al. (2019) show that the retrieved data from Instagram might render a different image from the check-in data of Foursquare. Nevertheless, some studies point to the contrary, such as Taubenböck et al. (2018) who show that in their temporal analysis of an informal settlement, although the number of tweets was very limited, the temporal patterns were similar to that of the wealthier formal settlements. It is argued that these biases are slowly fading with the increasing penetration of mobile devices and internet access (Mislove et al. 2011).

As the social media network weaves itself into the fabric of the everyday life of the city, the data becomes increasingly more accurate. The evolving body of literature indicates that Twitter data has the potential of reading urban socio-spatial patterns (Steiger et al. 2015; Shelton et al. 2015; García-Palomares et al. 2018; Arribas-Bel 2014), and contributing to bottom-up urban management approaches (Yao and Wang 2020). The current study addresses many of these biases by narrowing the data into a specific target group, and by the close examination of the outcome regarding its context. The proposed method collects public tweets from a group of Twitter handles that are regularly using the platform from within a city, this approach minimizes some of the concerns that were set forth by the previous studies.

Case study

The study was conducted in a major university located in the city of Famagusta, Northern Cyprus (Fig. 1). The case was selected due to its highly interactive Twitter users that have shown to be a reliable indicator of urban socio-spatial activities (Iranmanesh and Alpar Atun 2020b). The university was initially built on the periphery of the city, but the new developments in the city have since shifted toward the campus, making it an integrated part of the urban fabric (Önal et al. 1999).

The university campus was initially established with no clear planning position in relation to its host city of Famagusta, a left-over town in a post-conflict

area. Over time however, it has acted as a magnet pulling the development of the city towards the university. In part, this was due to the manmade and natural borders to the south, east, and west, but the resulting settlements could be seen as a transformation into “the campus as a city” with the increasing number of students reaching 18,000 in 2018. This number accounts for 45% of the city population during the active academic season.

Since a large portion of the city’s population consists of students, the land use around the campus has been adapted to address their needs. Although the university has borders and is connected to the city through a limited number of gates, it is working as a part of the urban grid especially at the pedestrian level. A recent study on the same case shows that the majority of students in the case study walk to the campus (80%), and do not own private vehicles (Dehghanmongabadi and Hoşkara 2018). Accordingly, the boundary of the study was set to include the campus and a walking distance buffer around it.

Data collection

The social media data was collected via Twitter public records from the students of the university using the NodeXL platform (Hansen et al. 2010). The study monitored the publicly available tweets addressing the university campus and the city and these were collated between September 2017 and September 2019 (within the limited access that the Twitter API provides). It must be noted that when creating a tweet, the user has the option of providing general locations like cities, neighborhoods, or venues, which generates a ‘bounding box’ and shows the general area where the tweet originated. Or the user can choose to provide an exact location which includes latitude and longitude tags. Only a very small fraction of tweets include precise geotagged information (Haffner et al. 2018). This ratio is reported 0.85% by Sloan et al. (2013), and 1% by Kumar et al. (2014). Nevertheless, contextual preferences toward the platform might increase the ratio reported as 3.1% by Morstatter et al. (2013). The study uses the precise location collected using the following step-by-step procedure.

- All tweets including the name of the university, and major public locations around the campus were

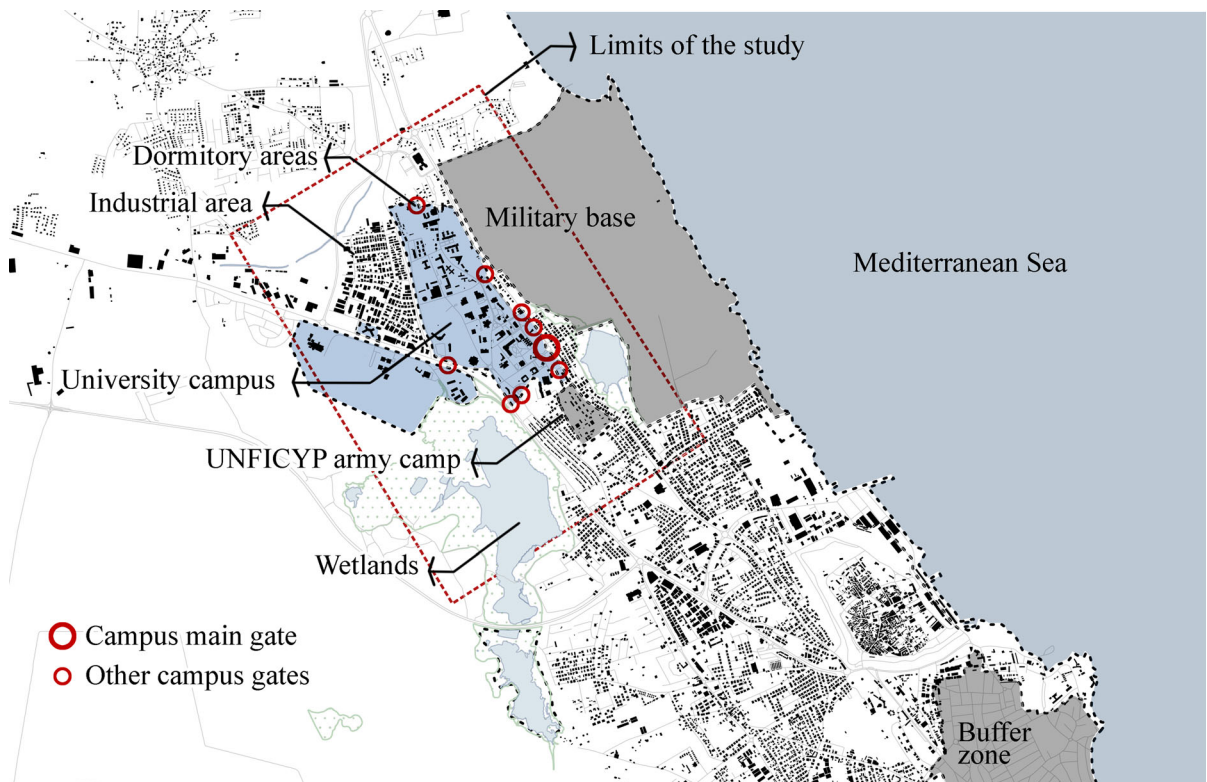


Fig. 1 The study area and its close surrounding borders

collected (within the limited access that the Twitter API provides).

- From the collected data, all handles that have regularly included precise geotagged information (latitude and longitude) in their public tweets were separated (990 handles, 1.29% of all).
- The selected handles were controlled for interaction with the university space. The handles that included tweets related to attending classes, taking an exam, or using the student dormitories were used as the final target group for the study ($n = 371$).
- Accordingly, the most recent 500 public Tweets (if applicable) from each account were collected between September 2017–2019. A total of 125,530 Tweets were collected from the selected handles including 37,100 tweets with intact geotagged metadata. It must be noted that these tweets are not distributed uniformly across Twitter handles. The average number of tweets, in this case, is 388 tweets per-handle for the two-year time-frame,

nevertheless, some of the users tweet more often on a daily basis, whereas some use it sparingly.

- Since the majority of the student body is composed of international students, not all collected tweets were from within the city, but 18,244 were identified within the boundary of the city. For this study, all tweets from inside the campus and its immediate surroundings were selected as the final dataset for the analysis ($n = 10,822$).

The final data set was classified on the basis of temporal and content classifications. The data was divided into 24 temporal subsets with 1 h intervals, and it was further classified into weekdays and weekends because the literature suggests different temporal patterns might occur within those time slots (Luo et al. 2016).

Analytical methods

The methodology of this study was built upon the spatial aggregation of geotagged Twitter data

regarding their temporal and contextual metadata. The paper tries to test the method in order to read the dynamic relationship between the urban form and its surrounding urban fabric. The land use is selected as a tangible variable that can be easily incorporated into the analytical method as a representation of people's daily needs and interactions. Accordingly, the study tries to predict land use through temporal and content classifications of geotagged Twitter feeds. Since the nature of data carries many variables, Kernel Density Estimation (KDE) was used for overlaying different layers to identify potential correlations.

In this study, KDE illustrates a function of the spatial density of tweets in different spatio-temporal categories for the study area and its close surroundings. The KDE function creates a continuous smooth estimate of the observed data points (Silverman 1986). Here, the methodology uses KDE to create a set of spatial representations according to the temporal and contextual aspects of the collected tweets. KDE has been used for analyzing various spatio-temporal aspects of urban space including road accidents (Prasannakumar et al. 2011), spatio-temporal criminology studies (Cusimano et al. 2010), criminology (Nakaya and Yano 2010), and analysis of urban centers (Thurstain-Goodwin and Unwin 2000). Furthermore, KDE has been widely used in analyzing spatio-temporal urban hotspot activities based on check-in, and social media data (C. Wu et al. 2018; L. Wu et al. 2014; Zhen et al. 2017; Zhang et al. 2019).

In using KDE the selection of a proper bandwidth diameter is critical for the accuracy of the analysis (Nakaya and Yano 2010). The bandwidth determines the smoothness of the density estimate by indicating the proximal effectiveness of each point on its neighboring points. A large bandwidth would often make the data dull by fading spatial details into each other. For this study, a standard walking distance was used for this purpose. Initially, 400 m was considered as the bandwidth for KDE as it is widely used as a general rule in planning for local walkable distance (Mulley et al. 2018; Park et al. 2015). The 400 m bandwidth could provide a meaningful representation of the distribution of activities at the local level (King et al. 2015). Borruso and Porceddu (2009) use a similar bandwidth for analyzing land use in two mid-size cities. Nevertheless, since the relationship between data points does not happen through straight-line connections, the reality of the physical

grid dictates a network distance which is often shorter than the direct linear distance between origin–destination points. To estimate an approximate radius representing the actual travel distance, the pedestrian route directness (PRD) introduced by Hess (1997) was used indicating a ratio of $\simeq 1.2$; the measurement has been since refined by Randall and Baetz (2001) between 1.40 and 1.48 for grid type street pattern which was used in this study. By applying the suggested ratio, $R = 300$ m was used as the bandwidth for KDE to make a better estimation of people's local interaction with space. Applying the method in a case with an organic street pattern needs further calibration. The properties of KDE remained uniform for all data types to ensure a smooth reliable probability detection. Figure 2 shows the general framework for data collection, synthesis, and analysis.

Figure 3 shows the distribution of all collected data throughout the day (1 h intervals), and days of the week. Comparing the day and night activities, the KDE illustrates an apparent movement between the educational zones and accommodation zones of the campus during the day-night shift. Most of the night activities shift outward toward the center of the city because the campus does not offer a very dynamic nightlife. This can be observed in the movement toward cafes and restaurants in front of the main entrance of the campus at night. There is a spike in activities as the campus wakes up at 5 o'clock, the number of tweets goes down slightly between 8 and 9 am as the morning lectures start, afterward, it increases slowly and picks at 7 pm. The graph shows an interesting spatial dynamic between tweets from inside and outside the campus indicating diametrically opposed picks in the number of tweets.

Figure 4 shows the university as a hub of activities in the morning. City and campus are equally vibrant around noontime and the majority of afternoon and night activities shift toward the city (García-Palomares et al. 2018). The only exception to this is the dormitory section of the campus that remains highly active throughout the night. The edge of the campus which is the main point of interaction between the two spatial structures is also a vibrant social hub that becomes active around noon and shows an increasing number of tweets up to midnight. The gap between the campus active hub and the city is due to the United Nation camp (UNFICYP army camp) which is closed to the public, this area creates an underwhelming lack

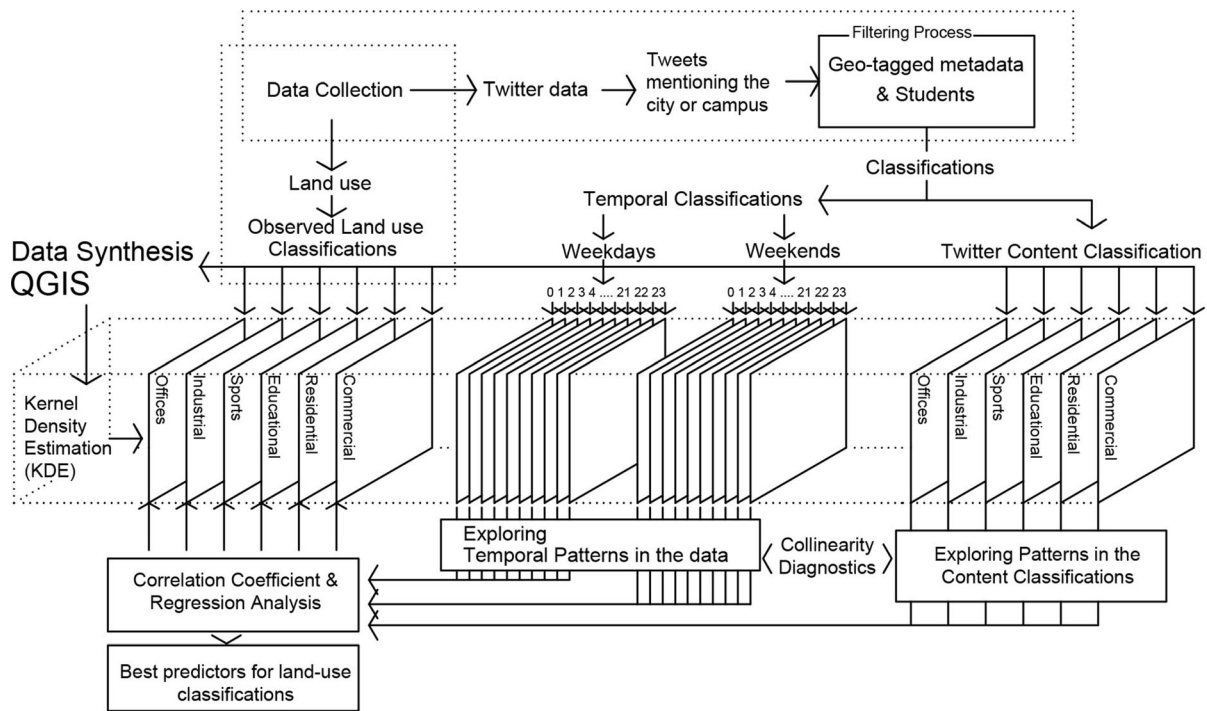


Fig. 2 The framework for data collection, synthesis, and analysis

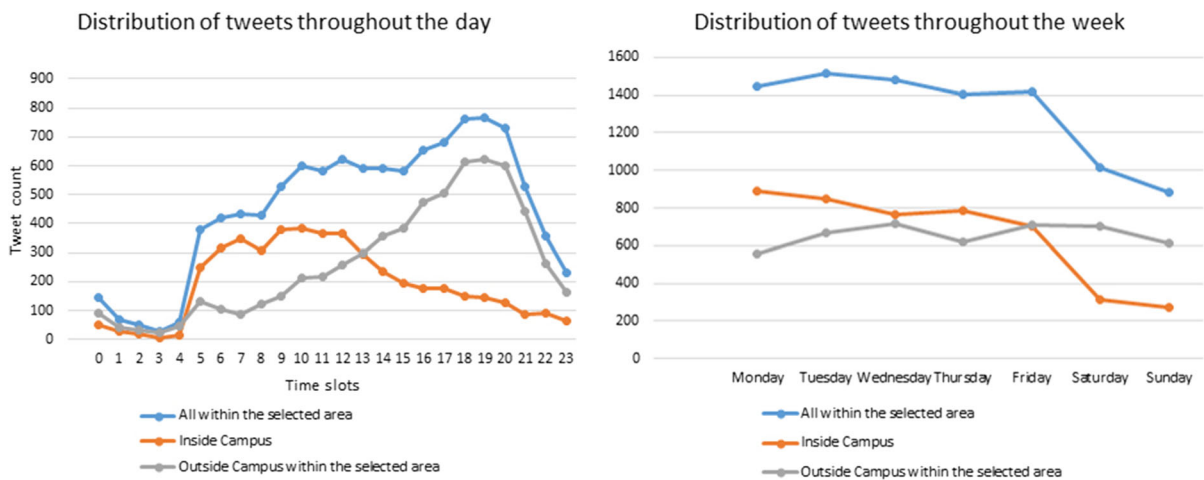


Fig. 3 The distribution of geotagged tweets across hourly time slots (left) and daily time slots (right)

of active forefront that divides the urban activities into two parts. The weekday classification shows a higher number of tweets compared to the weekend, directly caused by education-related tweets which drop down during the weekends while the out-campus tweets remain relatively high.

Land-use and spatio-temporal analysis

Similar to the Twitter feed data, a series of KDEs were produced for observed land use including commercial (cafes, restaurants, and shops), residential (dormitory and private housing units), sports activities, industrial functions, and offices (see Fig. 8: top row). For land use, the KDE was weighted using the area of the

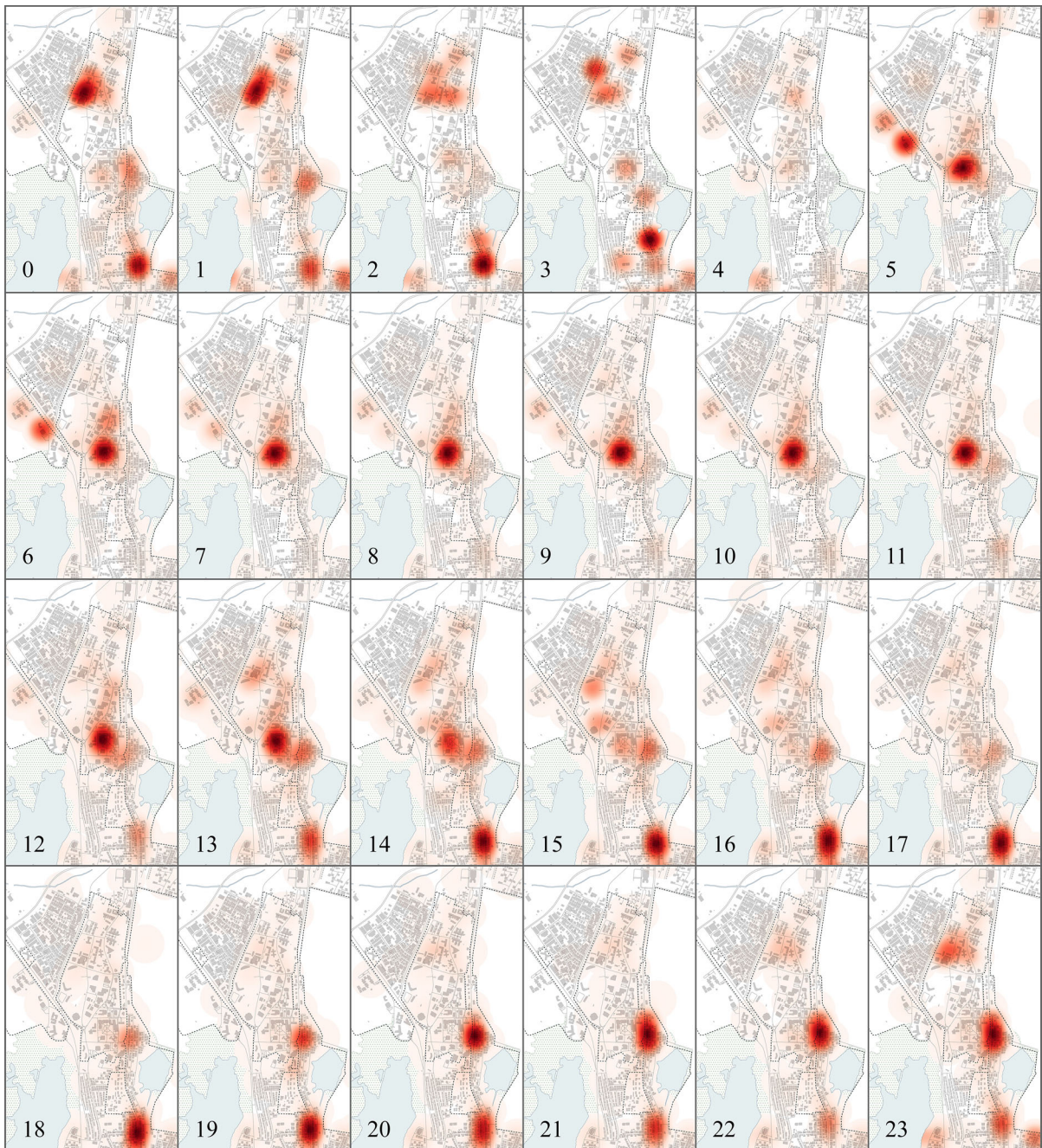


Fig. 4 KDE of all collected tweets for each time-slot

corresponding functions. Furthermore, the KDE was reproduced for all data (24-time slots) into weekdays and weekends as the initial analysis suggests that they might follow two different patterns (Luo et al. 2016; Frias-Martinez and Frias-Martinez 2014).

A linear regression model was used to explore the predictability of land use via spatio-temporal Twitter feed data in 24 time-slots for weekdays and weekends. The statistical model suggests a very strong correlation between the density of commercial land use and tweet density in the afternoon between 4 and 9 pm, (C.

Wu et al. 2018), reaching the maximum $R = 0.80$ at 8 pm. The correlation with commercial activities decreases at late night as the residential becomes a stronger predictor. In this case, the number of tweets in the early morning (2 am to 4 am) is relatively low and does not yield significant results. The activities become more random during these hours and individual tweets become effective in the regression model (Fig. 5).

The residential model shows a different pattern and correlates positively with weekend morning tweets and vice versa with weekday tweets. Both weekdays and weekends yield a similar result for the afternoon tweet rate reaching the maximum $R = 0.48$ at 8 pm which is similar to the highest time-slot correlation for commercial, and indicates the presence of mixed-use areas both inside and outside the campus because the ground floor is often occupied by commercial functions and the upper levels are often rental flats for students.

The educational land use shows strong predictability (max $R = 0.73$) during morning weekdays, and the tweet density declines significantly after 3 pm. The active hours of educational land use can be observed from the correlation table (7 am–3 pm). It must be mentioned that the university is not devoid of educational activities during weekends (especially Saturday), as the library is open, some classes are active (particularly in the department of health sciences), and some exams are being held on Saturday mornings, hence the high correlation with morning weekend tweets. The educational density shows a negative correlation with the evening tweet rate, although some tweets can be detected especially from the Department of Architecture with its 24 h open studios.

The sport-related functions show a significant correlation only during afternoon weekdays (2–4 pm) reaching a maximum of $R = 0.49$ at 3 pm. The sports fields are enveloped in their separated zone within the university campus, and the ones in the city

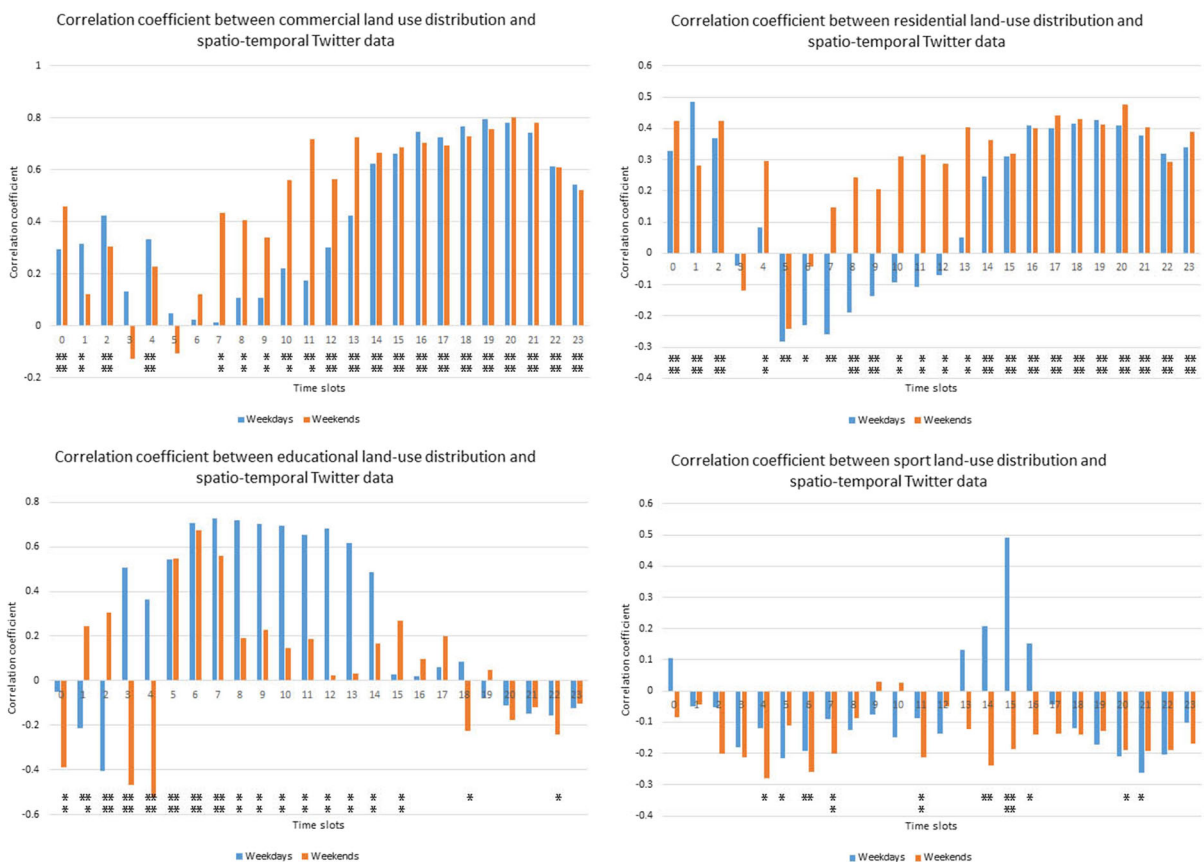


Fig. 5 The correlation coefficients between land use and time-slots

(gyms) are mixed with commercial. The mere size of the university sports facilities minimized the representation of smaller urban facilities via KDE. The industrial functions and the office functions do not show any statistically significant correlation with the temporal classifications of geotagged Twitter feeds due to the very limited number of tweets coming from those areas.

Content analysis

Although the temporal analysis shows promising results in reading active commercial and educational land use, the accidental nature of Twitter data (Arribas-Bel 2014) requires more investigation into the content of the data. We followed the example set by Lansley and Longley (2016); their research elaborates on methodological possibilities of reading land-use activities from Twitter topics. The process of analyzing the content starts with filtering the geotagged content. Therefore, the study expands the process of reading urban land use through Twitter feeds beyond mere temporal density by investigating the content of all collected tweets from within the selected area. The data classification and clean up are done using the following steps.

- The data was separated by urban blocks.
- The tweets from each block were analyzed using a word-counter, for this purpose all words longer than two letters were taken into account.
- All prepositions, pronouns, articles, and verbs were eliminated,
- Words related to the large-scale tags such as the university, city, region, country, which were the most frequent words throughout the tweet collection were eliminated (e.g. EMU, Famagusta, Gazimağusa, Ammochostos, Cyprus).
- The word count was conducted for the overall dataset and for each individual time-slot.
- The keywords were assigned to the geometric center of tweets containing each word (Fig. 6)

The keyword analysis denotes the location and importance of functions within the campus with a high level of accuracy and detail. Cafes, restaurants, faculties, public spaces, and dormitories can be identified. The content analysis has the potential to be used as a remote sensing technique for land use, socioeconomic, and point of interest analysis. Here,

the content analysis map illustrates a vibrant fabric of urban activities. Individual places can be easily identified. The data can be also utilized to explore the changing dynamics of the land use, Fig. 6 also includes locations that have been changed recently (marked with *). Some of the activities that might be overlooked in traditional land-use maps are visible here such as fan clubs (being held in one of the cafes), graduation ceremonies, and concerts (see Haffner 2018).

The spatial dimension of functions is not always related to the number of returned tweets; for instance, the “king of chicken burgers” is a very small and popular food vendor kiosk near the main gate. “Sandras” café that has the most number of tweets inside the campus is not a relatively large space, but its location is central. Although the proximity to the main gate seems to be a defining factor, the enclosed military areas are the most influential. The military base in front of the campus creates a rigid impermeable façade that eliminates any active functionality from one side of the street (see Fig. 2). The same phenomenon can be seen inside the city where the UN base occupies a large portion of the land between the university and the rest of the city, so the area in front of this base shows far fewer activities when comparing to other parts. It seems that a one-sided portion of the main street contains far less than half of the activities than the fully functional parts.

The prominent functions that were not detected properly from the temporal analysis are visible in the content analysis. These functions include sports facilities, the university health center, the general hospital, and industrial areas. The industrial areas have been mentioned in 30 tweets mostly related to car repair shops, these tweets although very limited in number, give an accurate representation of the location and the extent of the industrial area. The study could not find meaningful data related to administrative offices, and this is most likely due to the narrow focus of the target group including only students.

To utilize the vibrant potential of content analysis, a new set of KDEs were produced. The tweets from particular places were separated, with 9034 tweets having a clear relation with land use, and 1788 tweets were personal thoughts/interactions. The former group includes the check-in tweets, and those connected to land-use related activities (e.g. ‘juries at the department of architecture’, ‘exam’, ‘eating’, ‘residence’,

◀ **Fig. 6** The distribution of land-use related Tweets based on the content analysis

‘apartment’, ‘food’). All of the tweets not relating to land use were scanned for the same set of keywords. There are some keywords that must be scanned carefully, for instance, the word ‘home’ in both English and Turkish might come with ‘barbeque’ or food-related functions. Five KDE maps similar to the temporal classification were used to read the predictability of land use via geotagged data.

In general, the major patterns of interaction with urban space functions can be observed from the content analysis. Figure 7 (top) shows the most land-use related functions extracted from the content, and the bottom shows a small sample of words associated with those activities. The shift between education and commercial related tweets can be observed here. The cross-examination of the data shows that the morning commercial tweets (5–8 am) are connected to in-campus cafes, the late afternoon educational tweets, on the other hand, are mainly associated with labs, the library, and postgraduate courses. It should be noted that although in this case cafes and bars are assigned to the same class of activities, they show different temporal signatures between day and night. Furthermore, a set of correlation coefficients were conducted to detect the best predictors for different classes of land use (Fig. 7).

Overlaying temporal layers with the content analysis

The content-based KDE showed a strong improvement in the predictability of land use for less dominant activities ($R = 0.76$ for sports and $R = 0.47$ for industrial). These functions cannot be represented through temporal analysis because they are mixed with a variety of land-use related activities. The content KDE improves the predictability of land use very slightly for the dominant categories of commercial and educational. Here, the inter-correlation between the best temporal fit and the content analysis is extremely high (commercial: $R = 0.951$, educational: $R = 0.988$), indicating the land use could be estimated almost equally with either temporal or content classification (Fig. 8). Furthermore, the very high Variance Inflation Factor ($VIF = 1/1-R^2$) rejects the possibility of forming a multiple regression for better results, in fact, the VIF exceeds the suggested threshold (< 10) in both cases (Chatterjee and Hadi 2015).

The residential land use shows improvement in multiple regression estimates, although the outcome suggests that only 28% of all residential density can be predicted by temporal and content analysis (Fig. 8: fourth row). This could be caused by three phenomena:

- First, the sampling method of this survey, which is targeting only students. Although a very large

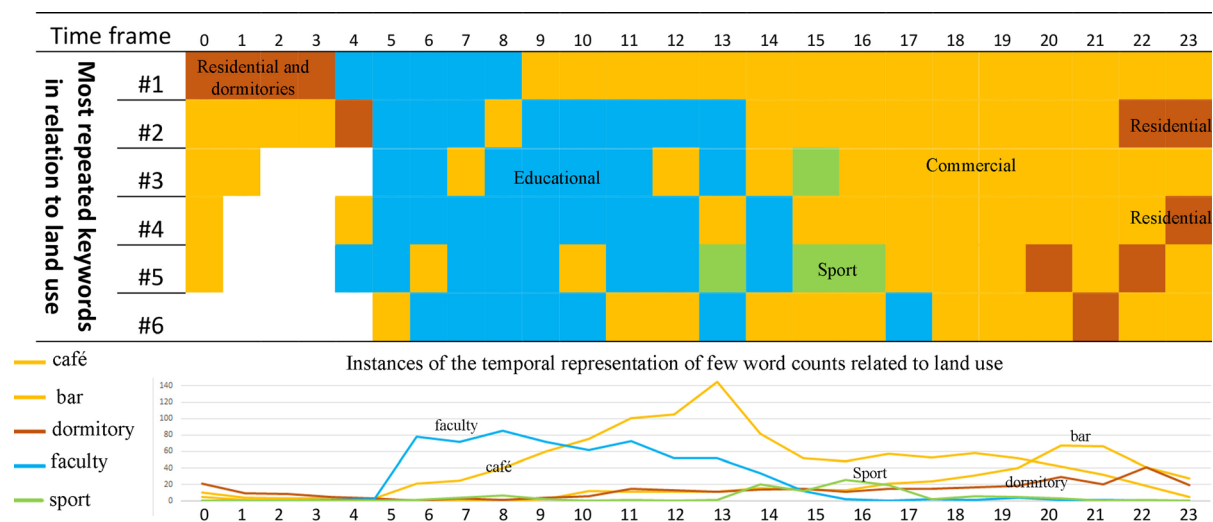


Fig. 7 Distribution of content related tweets across time-slots

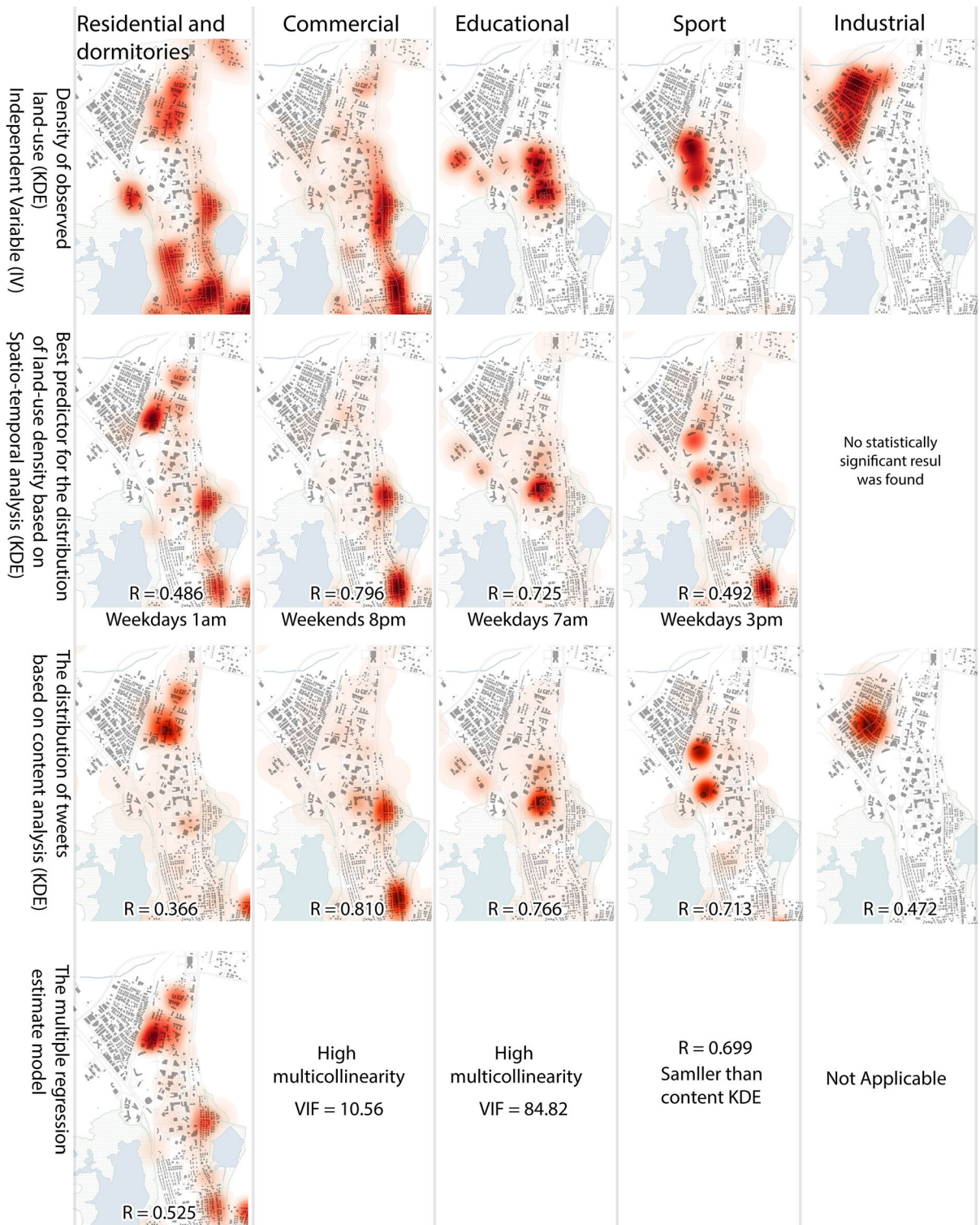


Fig. 8 Exploring the predictability of land use via the temporal/content analysis

portion of the city's population consists of students, many locals and non-students are living in the area. Hence, the lack of representation of these user types in the dataset could be influential.

- Second, the rate of tweets coming from residential areas is not enough for making a strong density estimation because they are the most common type of urban land use with a limited number of associated tweets. Moreover, people do not check-in to their places of residence compared to commercial places (Hasan et al. 2013).
- Third, the density of residential tweets might be affected by the typology of the built units. Li et al. (2016) shows that the typology of residential clusters has a significant impact on their representation through data, while gated communities disperse adjacent activities, the open community-based areas are better represented. Here, the dormitories are dominating the residential tweets undermining the small representation of other areas in the data.

Discussion and Conclusion

Mobile digital communication and social media have given a new face to urban space. People share their thoughts, showcase their activities, and simultaneously connect them to a precise time and location. This so-called “Big Data” opens up many new possibilities for research on urban structures. Looking through the glass of social media renders an emerging meaning of urban spaces in which the same digital mediums are in the process of re-defining what constitutes an urban space.

The current study used geotagged Twitter data as a source for reading urban land-use classifications. A systematic method of data collection and analysis was proposed using a small targeted section of social media data. The data was classified on the basis of temporal timestamps and content, and it was synthesized by Kernel Density Estimation to superimpose various layers. Although this study used geotagged Twitter data for the analysis, it can be extended to other social media that can provide sufficient public metadata. The fast-evolving history of social media suggests that a particular platform might not remain popular in an area or in general. The proposed

framework however can be adjusted for exploring urban settlements using emerging data types produced by the people. The study indicates three main findings:

First, the socio-temporal movement between spaces: different patterns were observed for weekdays and weekends showing the stronger commercial activities during weekends, and the educational activities during the weekdays. A general movement between inside and outside the campus was observed. Furthermore, the combination of hourly classification of tweets reveals a strong temporal movement between the commercial, educational, and residential. The best temporal predictor of land use was identified to be 1 am weekday for residential, 8 pm weekends for commercial, 7 am weekdays for educational, and 3 pm weekdays for sport-related urban land use. No statistically significant temporal predictor could be found for industrial areas or offices.

Second, the content analysis of tweets against actual land use: The geotagged content analysis shows a detailed map of urban land use. This map goes beyond the observed land use by showing an underlying network of pastime activities. The content analysis shows slight improvements in reading commercial, and educational land use. Furthermore, it shows a significant improvement in reading sport land uses, and even industrial areas that were not detectable in temporal analysis can be fairly predicted using the content analysis.

Third, temporal/content analysis combination: The combination of temporal and content analysis shows an improvement in reading residential areas. Nevertheless, the study concludes that the lower predictability of residential areas might be caused by the limitation of the target group, the naturally low rate of tweets from those areas, or typo morphological properties of the residential areas. The combination of temporal and content analysis also shows a subtle difference within the same class of land use (bars vs. cafes).

Furthermore, The analysis provides a new perspective that might be useful for planning and development in both problem diagnostics and problem-solving phases. Potential land use and zoning development could be evaluated based on social media analysis (see García-Palomares et al. 2018; Yao and Wang 2020). Furthermore, not only the data itself but the lack of data must be taken into account (Shelton et al. 2014). In this case study for example, vacant non-functioning

buildings located close to the main urban commercial centers have been identified. Finding an optimum location for new functions with a suitable temporal signature can be approached with this framework in three ways: First, to locate a new function close to its similar counterparts. This can improve the vitality of urban spaces hosting these functions, and it is a choice that many retailers do by instinct. Second, to complement areas with a variety of temporal land-use related activities may improve the overall safety and livelihood of the city (or campus in this case). Further studies are needed, however, to explore the changing landscape of urban tweets within longer time frames (Hu et al. 2015). Thirdly, it can be used for urban diagnostics, and provide a base for forthcoming studies oriented toward solving socio-spatial problems of cities. Further studies could elaborate more on the subject by adding more layers of data such as pedestrian movement, space quality, spatial network, and socio-economic census data. Separating the returned tweets during festivals, exams, concerts, and sports events might improve our understanding of the subject.

Finally, as this study has tried to showcase, the methodology must take into account the characteristics of the context. Although the methodology could be applied to any case that has a significant Twitter profile user, the socio-spatial forces that affect people's interaction with urban space must be examined thoroughly. In this case, the social nature of the university campus, and the inaccessible areas inside and around it have proven to be significant actors in how people use urban spaces.

Author contributions All authors have participated equally in developing this paper.

Funding No external funding or assistance was used for this research.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest regarding the submitted manuscript.

Human and animal rights The research does not include human or animal subjects.

References

- Arribas-Bel, D. (2014). Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography*, 49, 45–53. <https://doi.org/10.1016/j.apgeog.2013.09.012>.
- Borruso, G., & Porceddu, A. (2009). A tale of two cities: Density analysis of CBD on two midsize urban areas in Northeast Italy. In B. Murgante, G. Borruso, & A. Lapucci (Eds.), *Geocomputation and urban planning* (pp. 37–56). Berlin, Heidelberg: Springer.
- Chatterjee, S., & Hadi, A. S. (2015). *Regression analysis by example*. New York: John Wiley & Sons.
- Chen, X., Elmes, G., Ye, X., & Chang, J. (2016). Implementing a real-time Twitter-based system for resource dispatch in disaster management. *GeoJournal*, 81(6), 863–873. <https://doi.org/10.1007/s10708-016-9745-8>.
- Cosgrove, D., Daniels, S., & Baker, A. R. (1988). *The iconography of landscape: Essays on the symbolic representation, design and use of past environments* (Vol. 9). Cambridge: Cambridge University Press.
- Crooks, A. T., Croitoru, A., Jenkins, A., Mahabir, R., Agouris, P., & Stefanidis, A. (2016). User-Generated big data and urban morphology. *Built Environment*, 42(3), 396–414. <https://doi.org/10.2148/benv.42.3.396>.
- Cusimano, M., Marshall, S., Rinner, C., Jiang, D., & Chipman, M. (2010). Patterns of urban violent injury: A spatio-temporal analysis. *PLoS ONE*, 5(1), 1–9. <https://doi.org/10.1371/journal.pone.0008669>.
- Dehghanmongabadi, A., & Hoşkara, Ş. (2018). Challenges of promoting sustainable mobility on university campuses: The case of Eastern mediterranean university. *Sustainability*, 10(12), 4842. <https://doi.org/10.3390/su10124842>.
- Duncan, J., & Ley, D. (1993). *Representing the place of culture. Introduction. Place/Culture/Representation* (pp. 1–21). London: Routledge.
- Feick, R., & Robertson, C. (2015). A multi-scale approach to exploring urban places in geotagged photographs. *Computers, Environment and Urban Systems*, 53, 96–109. <https://doi.org/10.1016/j.compenvurbsys.2013.11.006>.
- Ferrari, L., Rosi, A., Mamei, M., & Zambonelli, F. (2011). *Extracting urban patterns from location-based social networks*. Paper presented at the Proceedings of the 3rd ACM SIGSPATIAL international workshop on location-based social networks, Chicago, Illinois.
- França, U., Sayama, H., Mcswiggen, C., Daneshvar, R., & Bar-Yam, Y. (2016). Visualizing the “heartbeat” of a city with tweets. *Complexity*, 21(6), 280–287. <https://doi.org/10.1002/cplx.21687>.
- Frias-Martinez, V., & Frias-Martinez, E. (2014). Spectral clustering for sensing urban land use using Twitter activity. *Engineering Applications of Artificial Intelligence*, 35, 237–245. <https://doi.org/10.1016/j.engappai.2014.06.019>.
- García-Palomares, J. C., Salas-Olmedo, M. H., Moya-Gómez, B., Condeço-Melhorado, A., & Gutiérrez, J. (2018). City dynamics through Twitter: Relationships between land use and spatiotemporal demographics. *Cities*, 72, 310–319. <https://doi.org/10.1016/j.cities.2017.09.007>.

- Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69(4), 211–221. <https://doi.org/10.1007/s10708-007-9111-y>.
- Haffner, M. (2018). A spatial analysis of non-English Twitter activity in Houston, TX. *Transactions in GIS*, 22(4), 913–929. <https://doi.org/10.1111/tgis.12335>.
- Haffner, M. (2019). A place-based analysis of #BlackLivesMatter and counter-protest content on Twitter. *GeoJournal*, 84(5), 1257–1280. <https://doi.org/10.1007/s10708-018-9919-7>.
- Haffner, M., Mathews, A. J., Fekete, E., & Finchum, G. A. (2018). Location-based social media behavior and perception: Views of university students. *Geographical Review*, 108(2), 203–224. <https://doi.org/10.1111/gere.12250>.
- Hansen, D., Shneiderman, B., & Smith, M. A. (2010). *Analyzing social media networks with NodeXL: Insights from a connected world*. Burlington: Morgan Kaufmann.
- Hansen, W. G. (1959). How accessibility shapes land use. *Journal of the American Institute of Planners*, 25(2), 73–76. <https://doi.org/10.1080/01944365908978307>.
- Hao, J., Zhu, J., & Zhong, R. (2015). The rise of big data on urban studies and planning practices in China: Review and open research issues. *Journal of Urban Management*, 4(2), 92–124. <https://doi.org/10.1016/j.jum.2015.11.002>.
- Hasan, S., Zhan, X., & Ukusuri, S. V. (2013). *Understanding urban human activity and mobility patterns using large-scale location-based data from online social media*. Paper presented at the Proceedings of the 2nd ACM SIGKDD international workshop on urban computing, Chicago, Illinois.
- Hecht, B., & Stephens, M. (2014). A tale of cities: Urban biases in volunteered geographic information. In *eighth international AAAI conference on weblogs and social media, Ann Arbor, Michigan*, (pp. 197–205): AAAI press.
- Hess, P. M. (1997). Measures of connectivity [Streets: Old paradigm, new investment]. *Places*, 11(2), 58–65.
- Hillier, B. (2012). Studying Cities to Learn about minds: Some possible implications of space syntax for spatial cognition. *Environment and Planning B: Planning and Design*, 39(1), 12–32. <https://doi.org/10.1068/b34047t>.
- Hillier, B., & Hanson, J. (1984). *The social logic of space*. Cambridge: Cambridge University Press.
- Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W., & Prasad, S. (2015). Extracting and understanding urban areas of interest using geotagged photos. *Computers, Environment and Urban Systems*, 54, 240–254. <https://doi.org/10.1016/j.compenvurbsys.2015.09.001>.
- Iranmanesh, A., & Alpar Atun, R. (2020a). Reading the changing dynamic of urban social distances during the COVID-19 pandemic via Twitter. *European Societies*, advance online publication. <https://doi.org/10.1080/14616696.2020.1846066>.
- Iranmanesh, A., & Alpar Atun, R. (2020b). Reading the urban socio-spatial network through space syntax and geo-tagged Twitter data. *Journal of Urban Design*, 25(06), 738–757. <https://doi.org/10.1080/13574809.2020.1814132>.
- Iranmanesh, A., & Atun, R. A. (2018). Exploring the spatial distribution of geo-tagged Twitter feeds via street-centrality measures. *Urban Design International*, 23(4), 293–306. <https://doi.org/10.1057/s41289-018-0073-0>.
- King, T. L., Thornton, L. E., Bentley, R. J., & Kavanagh, A. M. (2015). The use of kernel density estimation to examine associations between neighborhood destination intensity and walking and physical activity. *PLoS ONE*, 10(9), e0137402. <https://doi.org/10.1371/journal.pone.0137402>.
- Kitchin, R. (2014). The real-time city? *Big Data and Smart Urbanism. GeoJournal*, 79(1), 1–14. <https://doi.org/10.1007/s10708-013-9516-8>.
- Kostof, S., Castillo, G., & Tobias, R. (1999). *The city assembled: The elements of urban form through history*. London: Thames and Hudson.
- Kumar, S., Morstatter, F., & Liu, H. (2014). *Twitter data analytics (SpringerBriefs in Computer Science)*. New York: Springer-Verlag.
- 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>.
- Li, C., Wang, M., Wang, J., & Wu, W. (2016). The geography of city liveliness and land use configurations: Evidence from location-based big data in Beijing. *SERC Discussion Papers 0201*: Spatial Economics Research Centre, LSE.
- Lowenthal, D., Bowden, M. J., & Lamberty, M. A. (1976). *Geographies of the mind: Essays in historical geography in honor of John Kirtland Wright*. Oxford: Oxford University Press.
- Luo, F., Cao, G., Mulligan, K., & Li, X. (2016). Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago. *Applied Geography*, 70, 11–25. <https://doi.org/10.1016/j.apgeog.2016.03.001>.
- Lynch, K. (1960). *The image of the city* (Vol. 11). Cambridge, Massachusetts: MIT press.
- Lynch, K. (1981). *Good city form*. Cambridge, Massachusetts: MIT Press.
- Martí, P., Serrano-Estrada, L., & Nolasco-Cirugeda, A. (2019). Social Media data: Challenges, opportunities and limitations in urban studies. *Computers, Environment and Urban Systems*, 74, 161–174. <https://doi.org/10.1016/j.compenvurbsys.2018.11.001>.
- Mislove, A., Lehmann, S., Ahn, Y.Y., Onnela, J.P., & Rosenquist, J. N. (2011). Understanding the demographics of twitter users. In *Fifth international AAAI conference on weblogs and social media, Barcelona*, (pp. 554–557): AAAI Publications.
- 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. In *ICWSW-13: Proceedings of the 7th international AAAI conference on weblogs and social media Menlo Park, CA*, (pp. 400–408): AAAI.
- Mulley, C., Ho, C., Ho, L., Hensher, D., & Rose, J. (2018). Will bus travellers walk further for a more frequent service? An international study using a stated preference approach. *Transport Policy*, 69, 88–97. <https://doi.org/10.1016/j.tranpol.2018.06.002>.
- Nakaya, T., & Yano, K. (2010). Visualising crime clusters in a space-time cube: An exploratory data-analysis approach using space-time kernel density estimation and scan statistics. *Transactions in GIS*, 14(3), 223–239. <https://doi.org/10.1111/j.1467-9671.2010.01194.x>.

- Önal, Ş., Dağlı, U., & Doratlı, N. (1999). The urban problems of Gazimagusa (Famagusta) and proposals for the future. *Cities*, 16(5), 333–351.
- Park, S., Deakin, E., & Jang, K. (2015). Can good walkability expand the size of transit-oriented developments? *Transportation Research Record*, 2519(1), 157–164. <https://doi.org/10.3141/2519-17>.
- Penn, A., & Turner, A. (2004). Movement-generated land-use agglomeration: Simulation experiments on the drivers of fine-scale land-use patterning. *Urban Design International*, 9(2), 81–96.
- Prasannakumar, V., Vijith, H., Charutha, R., & Geetha, N. (2011). Spatio-temporal clustering of road accidents: GIS based analysis and assessment. *Procedia—Social and Behavioral Sciences*, 21, 317–325. <https://doi.org/10.1016/j.sbspro.2011.07.020>.
- Pred, A. (1984). Place as historically contingent process: Structuration and the time-geography of becoming places. *Annals of the Association of American Geographers*, 74(2), 279–297. <https://doi.org/10.1111/j.1467-8306.1984.tb01453.x>.
- Randall, T. A., & Baetz, B. W. (2001). Evaluating pedestrian connectivity for suburban sustainability. *Journal of Urban Planning and Development*, 127(1), 1–15. [https://doi.org/10.1061/\(ASCE\)0733-9488\(2001\)127:1\(1\)](https://doi.org/10.1061/(ASCE)0733-9488(2001)127:1(1)).
- Ríos, S. A., & Muñoz, R. (2017). Land use detection with cell phone data using topic models: Case Santiago, Chile. *Computers, Environment and Urban Systems*, 61, 39–48. <https://doi.org/10.1016/j.compenvurbsys.2016.08.007>.
- Rout, A., & Galpern, P. (2018). Using personal smartphone location histories in public engagement: Locating a new campus amenity. *Applied Geography*, 100, 68–77. <https://doi.org/10.1016/j.apgeog.2018.09.003>.
- Sagl, G., Resch, B., Hawelka, B., & Beinat, E. (2012). From social sensor data to collective human behaviour patterns: Analysing and visualising spatio-temporal dynamics in urban environments. In *GI_Forum 2012, Salzburg, Austria*, (pp. 54–63): Herbert Wichmann Verlag Berlin.
- Santa, F., Henriques, R., Torres-Sospedra, J., & Pebesma, E. (2019). A statistical approach for studying the spatio-temporal distribution of geolocated tweets in urban environments. *Sustainability*, 11(3), 595. <https://doi.org/10.3390/su11030595>.
- Shaw, S.-L., Tsou, M.-H., & Ye, X. (2016). Editorial: Human dynamics in the mobile and big data era. *International Journal of Geographical Information Science*, 30(9), 1687–1693. <https://doi.org/10.1080/13658816.2016.1164317>.
- Shelton, T., & Poorthuis, A. (2019). The nature of neighborhoods: Using big data to rethink the geographies of Atlanta's neighborhood planning unit system. *Annals of the American Association of Geographers*, 109(5), 1341–1361. <https://doi.org/10.1080/24694452.2019.1571895>.
- Shelton, T., Poorthuis, A., Graham, M., & Zook, M. (2014). Mapping the data shadows of Hurricane Sandy: Uncovering the sociospatial dimensions of 'big data.' *Geoforum*, 52, 167–179. <https://doi.org/10.1016/j.geoforum.2014.01.006>.
- Shelton, T., Poorthuis, 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. <https://doi.org/10.1016/j.landurbplan.2015.02.020>.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: CRC Press.
- Sloan, L., Morgan, J., Housley, W., Williams, M., Edwards, A., Burnap, P., et al. (2013). Knowing the tweeters: Deriving sociologically relevant demographics from twitter. *Sociological Research Online*, 18(3), 74–84. <https://doi.org/10.5153/sro.3001>.
- Soliman, A., Soltani, K., Yin, J., Padmanabhan, A., & Wang, S. (2017). Social sensing of urban land use based on analysis of Twitter users' mobility patterns. *PLoS ONE*, 12(7), e0181657. <https://doi.org/10.1371/journal.pone.0181657>.
- Steiger, E., Westerholt, R., Resch, B., & Zipf, A. (2015). Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data. *Computers, Environment and Urban Systems*, 54, 255–265. <https://doi.org/10.1016/j.compenvurbsys.2015.09.007>.
- Sun, Y., Fan, H., Li, M., & Zipf, A. (2016). Identifying the city center using human travel flows generated from location-based social networking data. *Environment and Planning B: Planning and Design*, 43(3), 480–498. <https://doi.org/10.1177/0265813515617642>.
- Taubenböck, H., Staab, J., Zhu, X., Geiß, C., Dech, S., & Wurm, M. (2018). Are the poor digitally left behind? Indications of urban divides based on remote sensing and twitter data. *ISPRS International Journal of Geo-Information*, 7(8), 304. <https://doi.org/10.3390/ijgi7080304>.
- Thurstain-Goodwin, M., & Unwin, D. (2000). Defining and delineating the central areas of towns for statistical monitoring using continuous surface representations. *Transactions in GIS*, 4(4), 305–317. <https://doi.org/10.1111/1467-9671.00058>.
- Wakamiya, S., Lee, R., & Sumiya, K. (2012). *Crowd-sourced urban life monitoring: urban area characterization based crowd behavioral patterns from Twitter*. Paper presented at the Proceedings of the 6th international conference on ubiquitous information management and communication, Kuala Lumpur, Malaysia.
- Wu, C., Ye, X., Ren, F., & Du, Q. (2018). Check-in behaviour and spatio-temporal vibrancy: An exploratory analysis in Shenzhen, China. *Cities*, 77, 104–116. <https://doi.org/10.1016/j.cities.2018.01.017>.
- Wu, L., Zhi, Y., Sui, Z., & Liu, Y. (2014). Intra-urban human mobility and activity transition: Evidence from social media check-in data. *PLoS ONE*, 9(5), e97010. <https://doi.org/10.1371/journal.pone.0097010>.
- Yang, L., & Durarte, C. M. (2021). Identifying tourist-functional relations of urban places through Foursquare from Barcelona. *GeoJournal*, 86(1), 1–18. <https://doi.org/10.1007/s10708-019-10055-9>.
- Yao, F., & Wang, Y. (2020). Tracking urban geo-topics based on dynamic topic model. *Computers, Environment and Urban Systems*, 79, 101419. <https://doi.org/10.1016/j.compenvurbsys.2019.101419>.
- Zhang, Y., Liu, L., & Wang, H. (2019). A new perspective on the temporal pattern of human activities in cities: The case of Shanghai. *Cities*, 87, 196–204. <https://doi.org/10.1016/j.cities.2018.10.002>.

Zhen, F., Cao, Y., Qin, X., & Wang, B. (2017). Delineation of an urban agglomeration boundary based on Sina Weibo microblog 'check-in' data: A case study of the Yangtze River Delta. *Cities*, *60*, 180–191. <https://doi.org/10.1016/j.cities.2016.08.014>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.