

Article

Beyond Spatial Proximity—Classifying Parks and Their Visitors in London Based on Spatiotemporal and Sentiment Analysis of Twitter Data

Anna Kovacs-Györi ^{1,*} , Alina Ristea ¹ , Ronald Kolcsar ², Bernd Resch ^{1,3} ,
Alessandro Crivellari ¹ and Thomas Blaschke ¹ 

¹ Department of Geoinformatics—Z_GIS, University of Salzburg, 5020 Salzburg, Austria; mihaela-alina.ristea@sbg.ac.at (A.R.); bernd.resch@sbg.ac.at or bresch@fas.harvard.edu (B.R.); alessandro.crivellari@sbg.ac.at (A.C.); thomas.blaschke@sbg.ac.at (T.B.)

² Department of Physical Geography and Geoinformatics, University of Szeged, 6722 Szeged, Hungary; kolcsar@geo.u-szeged.hu

³ Center for Geographic Analysis, Harvard University, Cambridge, MA 02138, USA

* Correspondence: anna.gyori@sbg.ac.at; Tel.: +43-662-8044-7555

Received: 17 August 2018; Accepted: 11 September 2018; Published: 14 September 2018



Abstract: Parks are essential public places and play a central role in urban livability. However, traditional methods of investigating their attractiveness, such as questionnaires and in situ observations, are usually time- and resource-consuming, while providing less transferable and only site-specific results. This paper presents an improved methodology of using social media (Twitter) data to extract spatial and temporal patterns of park visits for urban planning purposes, along with the sentiment of the tweets, focusing on frequent Twitter users. We analyzed the spatiotemporal park visiting behavior of more than 4000 users for almost 1700 parks, examining 78,000 tweets in London, UK. The novelty of the research is in the combination of spatial and temporal aspects of Twitter data analysis, applying sentiment and emotion extraction for park visits throughout the whole city. This transferable methodology thereby overcomes many of the limitations of traditional research methods. This study concluded that people tweeted mostly in parks 3–4 km away from their center of activity and they were more positive than elsewhere while doing so. In our analysis, we identified four types of parks based on their visitors' spatial behavioral characteristics, the sentiment of the tweets, and the temporal distribution of the users, serving as input for further urban planning-related investigations.

Keywords: urban parks; urban green areas; spatial analysis; GIS; sentiment analysis; temporal analysis; livability; social media analysis; accessibility analysis; urban planning

1. The Importance of Urban Green Areas and Ways to Analyze Their Role or Characteristics in the Urban System

While every city is unique in its characteristics, the universal aspect of all cities is their complexity [1]. One element of this complexity is the constant movement of hundreds of thousands or even millions of people, who also spend time in public places, such as in parks. Parks are essential public places and play a central role in a city's livability, primarily because of their role in offering social contact, exercise and restorative recreation. Furthermore, urban green areas have various effects on humans [2], partially as ecosystem services [3]. It is proven that access to green spaces is directly related to well-being through the influence of these areas on physical and mental health [4–11]. This influence is discernible mostly on changes in air quality [6,7,12,13], land surface temperature [6,14], physical activity [6,8,10,12], social cohesion [7,12], community identity [15,16], and stress reduction [7,8,12,17].

Therefore, analyzing the various effects of parks and how they are perceived is gaining increasing interest among researchers from different fields [18,19]. For instance, a growing body of literature deals with the analysis of factors determining urban green space use among the residents. The most relevant factors for parks are the functionality and facilities [7,20–24], safety and access [11,20,23,24] or even size or perceived greenness [25,26]; whereas, from the park visitors' side there are many personal characteristics ranging from age or ethnicity to health conditions that are determinant when selecting a park to visit [12,20,21,24,26].

Good access to urban green spaces is of increasing relevance in the design of livable, healthy and sustainable cities [7,27]. Having a park within 10–15 min walking distance from the residents' homes is also often considered as a factor of livable cities. However, in the literature, there are contradictory observations regarding the distance to a park from home and its relevance in people's decisions regarding which park to go to. There are studies that completely neglect spatial aspects (mostly when using Twitter data and performing sentiment analysis) [7,28–30], or, on the contrary, studies that only consider spatial aspects of park visits but not the functionality or other attracting factors [31–33]. In some other studies, either only the closest green area is considered, or the results show that having the park within less than a kilometer is more important than other factors [20,34], also for improving health [35]. However, there are also results showing that people visited parks that are further away, even if they had green areas nearer to their home, partially due to the differences between perceived and real distances [29,36,37]. Also, if the purpose of the park visit is performing physical activity, distance might be less likely to be a predictor of choice [38]. Only a limited number of studies focused on the issue of accessibility in a holistic way [39,40] even analyzing its direct effect on physical activity or health [41,42].

Most of the decision makers and urban planners intend to make public places livable [43–45]. However, livability strongly depends on the people's values and, therefore, their expectations, which means that planners should try to explore these expectations on an individual scale [46]. Asking people directly about their trips' characteristics or, for example, their expectations when visiting a park—as a traditional method in the form of questionnaires, which may even be combined with in situ observations—might be time- and resource-consuming while providing less transferable and only site-specific results. Also, the information produced as a result of such investigations still only represents a subset of temporal and spatial characteristics. At the same time, Twitter data analysis is mostly limited in data accessibility, thereby, once the required data is available, the analysis can be performed on scales ranging from intra-urban to even global for any period ranging from a few hours to several years. Recent developments in Geographic Information Systems (GIS)-based social media analysis offer the possibility to explore spatial, temporal and even affective aspects of users' behavior, even for public spaces and park visits [28–30,47,48]. However, some of these analyses still have limitations due to the manual interpretation of only a relatively low number of social media posts.

Several analysis efforts have used social media data for urban planning purposes over the last years, and the field of application is diverse and growing, ranging from more straightforward tasks to rather complex analysis, e.g., the detection of urban form and function [49]. In general, Twitter and other social media platforms are often used to analyze human activity and mobility on scales ranging from intra-urban to global [50–58], because these two phenomena are almost impossible to trace on finer spatial and temporal scales by using traditional methods such as questionnaires or quantitative observations (e.g., population counts). Furthermore, social media data can be used for socio-spatial analysis [59], for instance, by extracting the content of the tweets [60–63] or by investigating emotions and how they vary over space and time [64–68] also considering health factors such as diet or physical activity [65,69]. Campagna [70] proposed the concept of "Social Media Geographic Information" (SMGI) as a way of investigating "people[s] perceptions and interest in space and time" and thereby supporting spatial planning and geodesign, also by means of Spatial-Temporal Textual Analysis (STTx). Combined with other sources of data, such as mobile phone data, spatiotemporal characteristics of the urban environment can be described even more accurately [71]. Due to their fine spatial and temporal

scale, another great potential of social media data is the detection [72] and analysis of events [73–77], or disasters [78,79], and their effect on daily urban planning routines [80].

The goal of our analysis—similarly to SMGI—was to illustrate the possibilities of using social media (Twitter) data to extract spatial and temporal patterns of park visits for urban planning purposes, along with the sentiment of the tweets to represent how positive or negative a given post was, focusing on frequent Twitter users. Thereby, we intended to answer the following research questions:

1. *Spatial aspects*: What are the spatial characteristics of the selected users' tweeting behavior and how do these characteristics relate to their park visits? In terms of parks, how far do the visitors travel on average to visit a given park from their center of activity?
2. *Content aspects*: Are tweets in parks more positive than in other urban areas? What feelings do the visitors have when spending time in a park? How does this vary between parks?
3. *Temporal aspects*: How do the spatial and sentiment characteristics vary over time? Are there any significant differences during the day, week or year?
4. *Profiles*: What types of parks and park visitors can we classify based on the identified spatial, temporal, and sentiment characteristics? What do we learn about them?

Indubitably, every park and park visitor can be unique, and, in a large city, it is hard to answer these questions for every individual. Compared to traditional questionnaires where most of the focus is on only one or a few locations, big data or social media data allows every park and thousands of visitors to be considered within the city—not only as individual entities in isolation but also as a set of comparable characteristics. To overcome some of these limitations, a combined approach has emerged in planning, which can use the advantages of both quantitative and qualitative data analysis to some degree. Geo-questionnaires and public participatory GIS (PPGIS) has been developed over the past decade and has advanced our understanding of public preferences or even legitimizing decisions [81–84]. At the same time, we must recognize that, depending on the purpose of the study, social media analysis may not reach accuracies comparable to individual on-site studies [25], but can still produce valuable input or added value as an overview of the general patterns. In that sense, social media analysis should be considered a complement to, not a replacement of, on-site field studies [85].

In this paper, we analyzed the spatiotemporal park visiting behavior of more than 4000 Twitter users for almost 1700 parks along with users' feelings extracted from over 78,000 tweets posted in London, UK. The novelty of our research is the combination of spatial and temporal aspects of Twitter data analysis for park visits in a transferable way while applying sentiment and emotion extraction to also explore the content of the tweets, to overcome the limitation of traditional methods. In summary, the findings are aggregated to identify different types of parks and their visitors, serving as an input for further investigations.

2. The Core Data Sets of the Analyses

2.1. Input Data Sets

Our analysis is based on 11,372,967 tweets from Greater London for the year 2012. All tweets are geolocated, i.e., they have latitude and longitude coordinates to identify the location at which they were posted. The data are accessed through the Twitter Streaming Application Programming Interface (API) [86], using a bounding box around the Greater London (Table 1). In addition to the coordinates, the data set contains the user ID, the text, and the timestamp of each tweet as attributes.

Table 1. General overview of the Twitter data set.

	Value
Bounding box (WGS84)	51.225808, -0.560455; 51.734863, 0.319181
Total tweets	11,372,967
Total unique users	374,700
Temporal extent	1 January 2012–31 December 2012

The polygons representing areas of interest (parks, urban green spaces) are defined using OpenStreetMap, which is globally available—an important criterion for the possible transferability of the presented methods. Unfortunately, there is no single tag or keyword to extract the required polygons, so a combination of tags is used containing the words “park”, “green”, “garden” or even “forest” for the fields “natural”, “amenity”, “landuse” and “leisure”. The query resulted in a total of 5007 polygons for the same spatial extent as our tweets. For the sake of simplification and clarification, we will refer to any type of urban green space or area as “park” throughout the rest of the paper.

2.2. Preprocessing of the Data

Figure 1 provides an overview of our data preprocessing workflow. The first step was to define our study area by performing a spatial query, selecting a subset of elements from both input data sets (tweets, polygons) located in Inner London (surrounded by a 5 km buffer to reduce the edge effect of the administrative boundary). We then joined the two data sets spatially to identify “park tweets”. These tweets, according to their coordinates, were posted from one of the green areas we identified. This resulted in 341,888 park tweets.



Figure 1. Overview of the data preprocessing workflow.

Figure 2 shows the long-tailed distribution of tweet frequency per user. Around 70% of the users posted less than ten tweets in 12 months, which would make the analysis of their behavioral patterns less reliable. In fact, 50% of all tweets were posted by only 2.2% of all users.

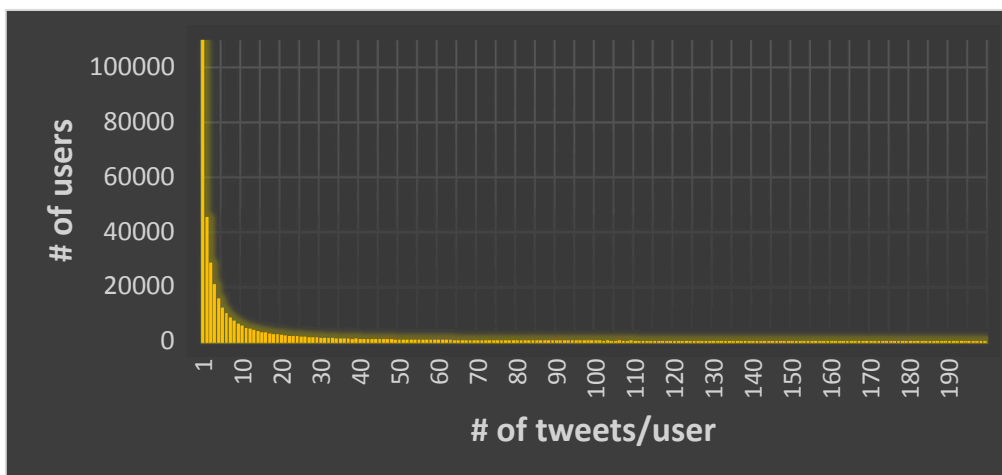


Figure 2. Frequency of tweet count per user.

Furthermore, there is a general difference in the park visiting activity of residents compared to tourists. We take this into account to restrict our analysis to only presumable residents. Therefore, we apply further filters based on the number of tweets per user and the temporal distribution of the tweets throughout the year to identify frequently tweeting users. Some of these users might not be residents in an administrative sense but, using our filters, we can select those who tweet in a larger temporal range. By combining the temporal filter with a higher number of tweets per user, we have more information to characterize more representatively the (possible) residents' park visiting behavior. Tourists tend to have a different park visiting pattern and motivation than residents, as they usually have just a few tweets in a short period throughout the whole year, mostly from the popular parks that are considered to be tourist attractions. This different nature of park visits between residents and tourists is a relevant aspect in urban planning, and therefore, we intended to focus only on presumable residents in our analysis. We only consider users with at least 12 tweets (1 tweet/month on average) within at least two non-consecutive quarters of the year (Table 2). Every quarter is three months long (e.g., 1 April–30 June), so a user is selected for further analysis if they have a tweet, for example, from May and another one from October (Table 2—Option 3), also to represent various seasons.

Table 2. Minimum requirement for the temporal distribution of the user's tweets (✓ = At least one tweet in that period. Tweeting activity in the other two quarters of the year are optional).

	Option 1	Option 2	Option 3
Quarter 1 (January–March)	✓	✓	
Quarter 2 (April–June)			✓
Quarter 3 (July–September)	✓		
Quarter 4 (October–December)		✓	✓

This method has some limitations, as it will not identify less-active Twitter users who still could be residents. However, using a data set where only one city is represented instead of each city where the user tweeted in the given period, it would require more complex methods to extract residents with high accuracy, which is beyond the scope of our paper. At the same time, using only (geolocated) tweets to represent a person's spatial behavioral pattern adequately requires larger number of tweets. Thereby, we can rely on the results of the method by excluding less-active residents. Recurring tourists with an interval of three months at least, cannot be excluded either but their contribution to the overall data set might be low. The selection of presumable residents resulted in 41,967 users with 157,760 park tweets out of 4,502,364 total tweets by these users.

In 2012 the Olympic Games were held in London. The main venue of the event (Queen Elizabeth Olympic Park) is also part of our study area, and during the Olympic Games (and the Paralympics) extraordinary Twitter activity was observable, which could distort our results. To avoid the bias, we excluded tweets from the Olympic Park from 24 July–13 August (Olympics) and 29 August–9 September (Paralympics). After this filtering, we checked the above-mentioned criteria for residents, to exclude users who no longer fulfilled the defined requirements. In the end, we obtained 141,542 park tweets.

Finally, we also set up a threshold for the proportion of park tweets per user, to be more representative of users' behavior. Due to the high variance of users' tweet frequency, we set a minimum of four park tweets (1/3 of the overall minimum tweet count for a user) or if someone has more than 80 tweets in total, then at least 5% should be park tweets to represent park visiting behavior on an individual level. Consequently, our pre-processed data set, ready for spatiotemporal and content analysis, as well as for defining user and park profiles, consisted of 78,597 park tweets out of 636,917 tweets, from 1754 parks and posted by 4337 unique users (Figure 3).



Figure 3. Map of tweets and parks (input data sets of the analysis).

3. Methodology

3.1. Overview

As an overview, Figure 4 shows the main components of our analysis. After the preprocessing of the data, the first group of analyses is performed to study the spatial characteristics of park visitors’ behavior. In a second step, the content of the tweets is analyzed to provide a general interpretation of the mood of the people while tweeting in a park. This comprises sentiment and emotion extraction, and then an aggregation of the gathered information on park level for both steps (sentiments, emotions). Regarding the temporal variability of both park visits and the sentiment of the tweets, we also analyze how the results of the previous analyses vary over time. The analysis of daily, weekly and seasonal trends are essential characteristics for the study of park visits. These temporal patterns are not only relevant for studying the number of visitors per hour, day or season, but also to trace the changes in the sentiment of the tweets or the emotions of the users accordingly. In summary, the results of the spatial, temporal and content analyses are combined to identify different user and park profiles.

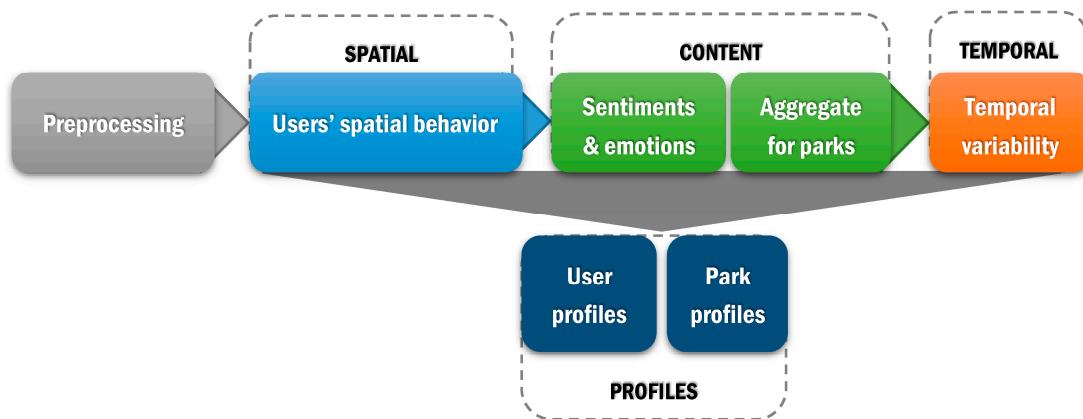


Figure 4. Methodology overview.

3.2. Spatial Analysis

In this part of the analysis, we had two goals:

1. To describe the main characteristics of the relevant users' spatial behavior: Where is their main center of activity based on their tweets? What is the average distance between this center and each tweet from the same user?
2. To measure the average distance between a park visitor's main activity center and a given park: What is the median and mean distance from the activity center of the users who tweeted from the given park?

As the home location or any other reliable information is not available for the users, we use the centroid of their tweets as the main attribute to investigate a user's spatial behavior [57]. This centroid or "center of mass" (COM) is the coordinate of the users' main center of activity representing the average of each unique tweet by a user (Figure 5a). As a result, of the preprocessing, we can distinguish park tweets from non-park tweets, which is an important aspect for the investigation of users' spatial behavior. Thereby, the COM is also calculated only for the park tweets of a user. Figure 5b illustrates how the average and median distance from the users' COM (for all tweets) to park tweets was calculated.

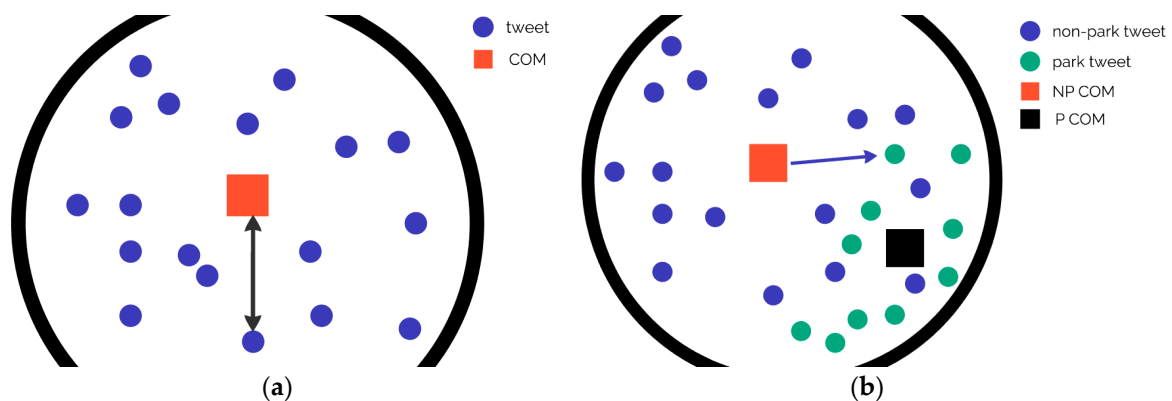


Figure 5. The illustration of how COM is interpreted (a) and how it is used to measure the average distance between COM and a park tweet (b).

For both centroids (park, non-park), the average and median distance to each tweet are calculated, which shows some general insights as to whether the users are more mobile in terms of their tweeting behavior. The median is used to offset the negative effect of outlier tweets with very high distances. The shift between the two types of COM coordinates also provides valuable information: if the shift and the average distance for park tweets are relatively low, we can conclude that the user mostly visits parks closer to their main center of activity. Table 3 summarizes all the derived variables for the user-specific spatial behavior. The values of these variables are then visualized on histograms to represent trends in their distributions (see Section 4.1). All the distances calculated in this section are Euclidian distances.

Table 3. Derived spatial variables for each user.

Variable	Description
COM_all	Center of the coordinates of all tweets (per user)
COM_park	Center of the coordinates of all park tweets (per user)
COM shift	Distance between the two different COMs
COM to all distance	Distance between COM and all the tweets (average and median)
COM to park distance	Distance between COM and park tweets (average and median)

3.3. Semantic Content Analysis

In the first step, we create the tweets' corpus from the text which we clean in a few preprocessing steps, including tokenization, removal of stop words, anything other than Latin characters, URLs, numbers and punctuation symbols, a procedure also suggested by Steiger et al. [87]. After the preprocessing, we apply two sentiment analysis methods to extract polarity followed by emotion extraction.

3.3.1. Sentiment Scores

To define sentiment values for each tweet's text, we use the lexicon by Hu & Liu [88], which contains positive and negative words. The polarity value shows the difference between the data set's negative and positive attributions. If the difference value is higher than zero, the tweet is assumed to have an overall "positive sentiment", while below zero, it is considered "negative", and when it equals zero, then the text message is "neutral". To avoid possibly misclassified tweets (difference score is close to zero), we define "positive sentiment" by a score equal or higher than two and "negative sentiment" by a score equal or lower than minus two, thus excluding weak and potentially unreliable sentiment scores of $[-1,1]$ which is in line with previous research [89].

3.3.2. Emotion Detection

To determine emotions included in written text, we use the National Research Council Canada Emotion Lexicon (NRC Emolex) [90,91] through the Syuzhet package in R [92]. This lexicon includes a list of 14,182 unigrams and their associations with eight emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and polarities (negative and positive). The words were manually annotated through Amazon's Mechanical Turk that is a crowdsourced marketplace, where users sign up for simple tasks that gives them small rewards. At least three annotators annotated every word. This procedure helps to define a scale of association between emotions or sentiments and the tweet text (not associated, weakly, moderately, or strongly associated), all of which are used in this study except for "not associated".

3.4. Temporal Analysis

Using the timestamps of each tweet, we divided them into different temporal categories to trace changes in the number of tweeting visitors and how the sentiments and emotions of their tweets vary over time. Daily patterns show the hourly distribution of the tweets during the day. We aggregated them on the park level, to determine whether the park is more popular and "positive" in the morning, throughout the day, or the evening. On the other hand, the main advantage of the weekly pattern is to distinguish which parks are more favorable at the weekends than during the week or on what days they have more tweets that are positive. Seasonal patterns can reflect the effect of climatic factors along with the functionality of a park. Especially the winter trends are interesting because a park functions well if it can also attract people during the colder periods of the year. In our analysis, we did not define seasons by precise dates (e.g., equinox to solstice) but rather used simpler groupings of months: spring was defined as March to May; summer as June to August; fall as September to November; and winter as December to February.

3.5. Profiles

The Partitioning Around Medoids (PAM) clustering algorithm was used in R (RStudio) to combine and interpret the results of all the analysis steps discussed above. Compared to the traditional K-means clustering algorithm, PAM is more robust to noise and outliers [93]. The number of clusters was defined by using the *fviz_nbclust* (https://www.rdocumentation.org/packages/factoextra/versions/1.0.5/topics/fviz_nbclust) R package as a starting point, and complemented by manual interpretation if the R tool provided no clear suggestion. The spatial, temporal, and content analysis each represents

one important attribute of parks or user behavior. Our analysis is mainly done on tweet or user level, which is then aggregated on park level as well. For example, the COM is calculated for each user considering all their tweets. However, to derive the average distance from a park to its visitors' COM we need to consider individual tweets to select which users were tweeting from the park to measure the distance to their COM. After we have all the required distances to a park (for each user tweeting from there), we can calculate the average value for the whole park.

Regarding the users, we consider the spatial tweeting behavior to be the most relevant aspect in accordance with the aim of this paper, especially park tweets and their distance from the users' activity center. We extracted clusters of users based on the factors described in Table 3. In terms of sentiments and emotions, we compared non-park and park tweets for every user, to see if a user is more positive while in a park.

For parks, all three types of analysis were considered to be of equal importance. Therefore, parks were clustered according to their visitors' spatial characteristics (how "mobile" they are for parks and in general), how positive or negative the tweets were in that given park along with which emotions are present or more prevailing, and how the frequency of visits and the content of the tweets varies over time.

4. Results

To synthesize the results for all our investigated aspects (spatial, temporal, content) we cluster similar features into groups. Identifying different types of parks and visitors is not just useful to represent vast amounts of information, but it also aids planning, as similar types of parks might face similar problems and thereby require similar actions. Park types can also help to define a hierarchy among different urban green areas based on the distances users travel to them or the time of the day/week when a peak in the number of visitors occurs.

4.1. Spatial Profiles

As shown in Table 3, we calculated the average and median distances from COM to each tweet both for only park tweets and for all tweets. Every user had one value for each calculation; Figure 6 shows the frequency of these distances.

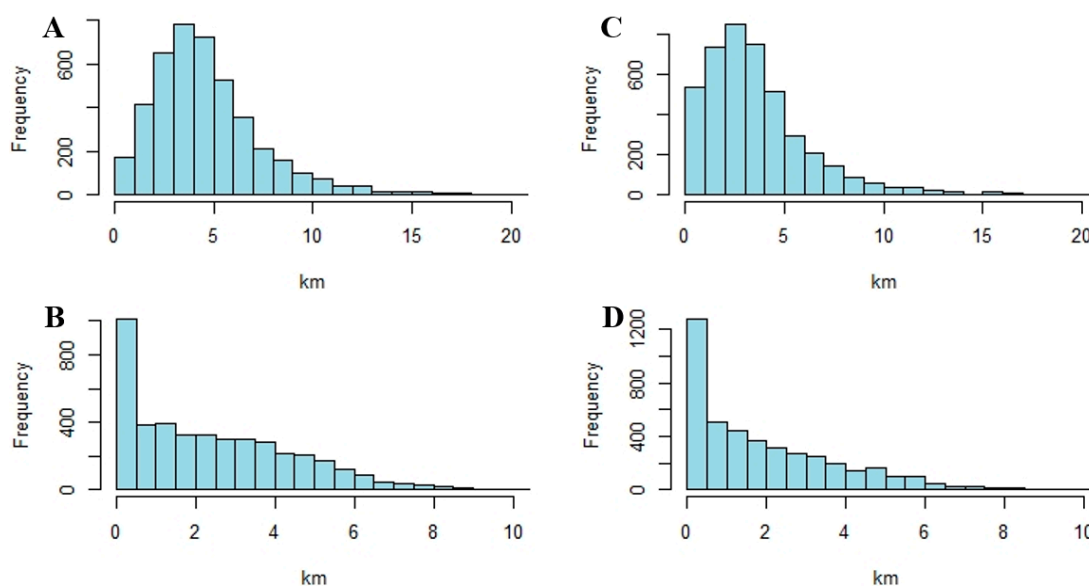


Figure 6. (A) Average distance to each tweet from COM—all tweets; (B) Average distance to each park tweet from park COM; (C) Median distance to each tweet from COM—all tweets; (D) Median distance to each park tweet from park COM.

The graphs show that most of the people tweeted around 3–4 km away from their main activity center (Figure 6A,C). However, if we only consider park tweets, most users have rather small distances between the park COM and each park tweet, which means they mostly tweeted from parks close to each other (Figure 6B,D) or even from the same park. Furthermore, it is very interesting to see the average distances between park tweets and the COM of all tweets (Figure 7). The results show that people were not tweeting very close to their COM. For those Twitter users, who live closer to the city center or at least, not too close to the edges of the study area, the COM can serve as an approximation for their home location [94]. Thereby, this distance between park tweets and the COM of all tweets can reflect the average home-park distance for the given user. Most users tweeted in a park around 3–4 km away from their COM (of all tweets) on average.

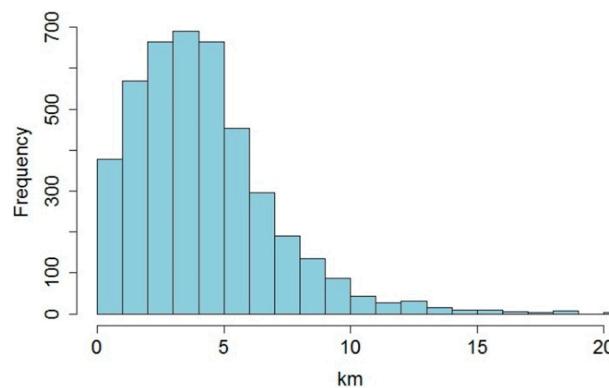


Figure 7. Frequency of average distances from COM (of all tweets) to park tweets.

This is where one of the limitations of social media analysis becomes obvious, as we are not able to identify or verify the causal relationship for this trend. The observed higher distance could have various explanations, but without a further in-depth investigation, we can only hypothesize them. However, some of the hypotheses can also be rejected by the combination and cross-checking of results. The distribution of the parks in the study area regarding accessibility can be considered good, as from almost any point of the city there is a park within 500 m of walk. This means that it is not necessary for someone to visit a park 3–4 km away from their home/COM because they have parks closer to them. However, the case might simply be that the closest park is not adequate for their needs. Another explanation is that due to self-imposed privacy constraints or some other reasons they would not tweet near their home. Also, the motivation behind posting a tweet is often to report something extraordinary, so when a user regularly visits a nearby park, it is not something that they would tweet about, but when visiting an unfamiliar location further away they might do so.

In Section 1 we pointed out the ambiguous role of spatial factors in park visits based on existing research. Although social media data analysis by itself cannot reveal direct motivation behind visiting a given park over other parks, the observed trends in distances between COM and park tweets along with accessibility imply that short distance (especially visiting the closest park to home/COM) has no primary importance. Perceived distances and accessibility might still be relevant in the overall decision but other factors (such as functionality, perception of safety/beauty, etc.) can have a more significant influence in selecting a park for visit, which can serve as an input for further research.

If we aggregate the user-specific measurements on the park level, we can see the average distances traveled (as network distances) to a given park by the users who tweeted from there (Figure 8). Only parks with tweets from at least ten different users are visualized. There is a clear concentric trend observable; the further away from the city center a park is, the bigger the average traveled distances. This means that parks in the outer parts of the city are visited not just by people living nearby but really by visitors from all over the city.

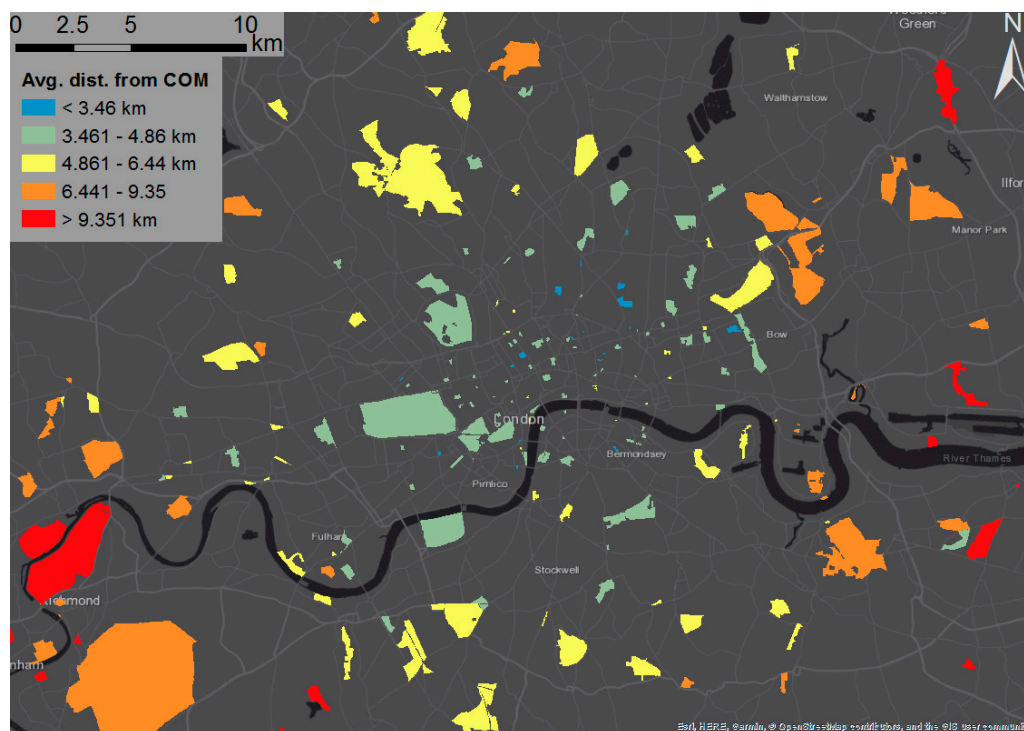


Figure 8. Average distance from users' COM to the park, aggregated on park level based on the tweets posted from a given park.

As stated above, considering the analyzed spatial aspects, we clustered users and parks according to their characteristics. Among users, we identified four groups (Figure 9). The users in the first group have low values for each distance category, which means that based on their tweeting activity they are not so mobile compared to other groups. Their movements in general, not just for park visits, is constrained to a relatively small area within the city. The second group is quite similar, except that they have the highest value among all groups for the average and median distance between the park COM and park tweets. Interestingly, their median value is even higher than the average. This high value means that these users usually move around in a small area, except for when visiting parks. The third group is exactly the opposite of the previous one—they have relatively high values for each variable except for the park COM to park tweets. This means that these users travel large distances in the city; however, when it comes to visiting parks, they opt for parks that are close to each other (or even only one park) and usually not close to the visitors' COMs. Finally, the fourth group is similar to the first one, they have very similar values for each variable, but they are higher than in the first group. These users are mobile, they visit and tweet from various places around the city both in parks and for other activities.

Figure 10 shows park types based on the proportion of different user types visiting a given park. There were four types of park according to which user cluster their visitors mostly belong to. Parks in the first category are mostly visited by users from Cluster 1, indicating visitors who generally traveled short distances (blue), while the second category with users from Cluster 4 represents large distances (yellow). The third group of parks is exactly the opposite of the second one as those parks have visitors from all user clusters except for Cluster 4 (green). The last park category also has multiple user types, dominantly from Cluster 1 and 4 (purple). If we investigate the spatial distribution of different park types, we can see that the parks in the first group (blue) are mostly outside the city center, which indicates that most of the visitors live close to the park and residents of the inner city visit them less frequently. The second group (yellow) contains smaller parks, which means that they might provide some specialized functionality and therefore people might also visit them from larger distances. Parks in the third group (green) are on average larger in extent and visited by all types

of users. The reason for this is probably quite the opposite as it is for the second group, as a bigger park would provide various functionalities and thereby attract different people. The last group of parks (purple) also has visitors from a wider range according to their spatial behavior. These parks are (except for a few) closer to the city center so people living closer to the city center will visit them, but the parks can also attract visitors from larger distances. Parks located next to each other but belonging to different groups (especially blue or yellow) represent interesting scenarios, the cause of which can be investigated in further studies.

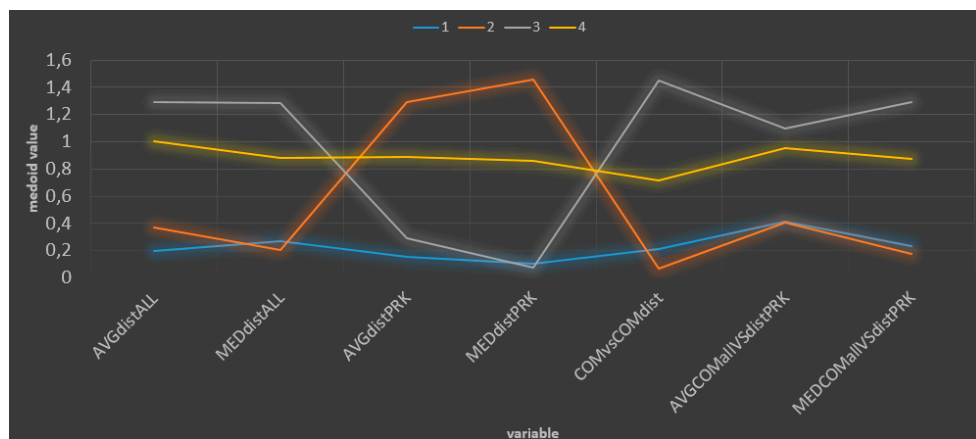


Figure 9. Medoid values of user clusters.

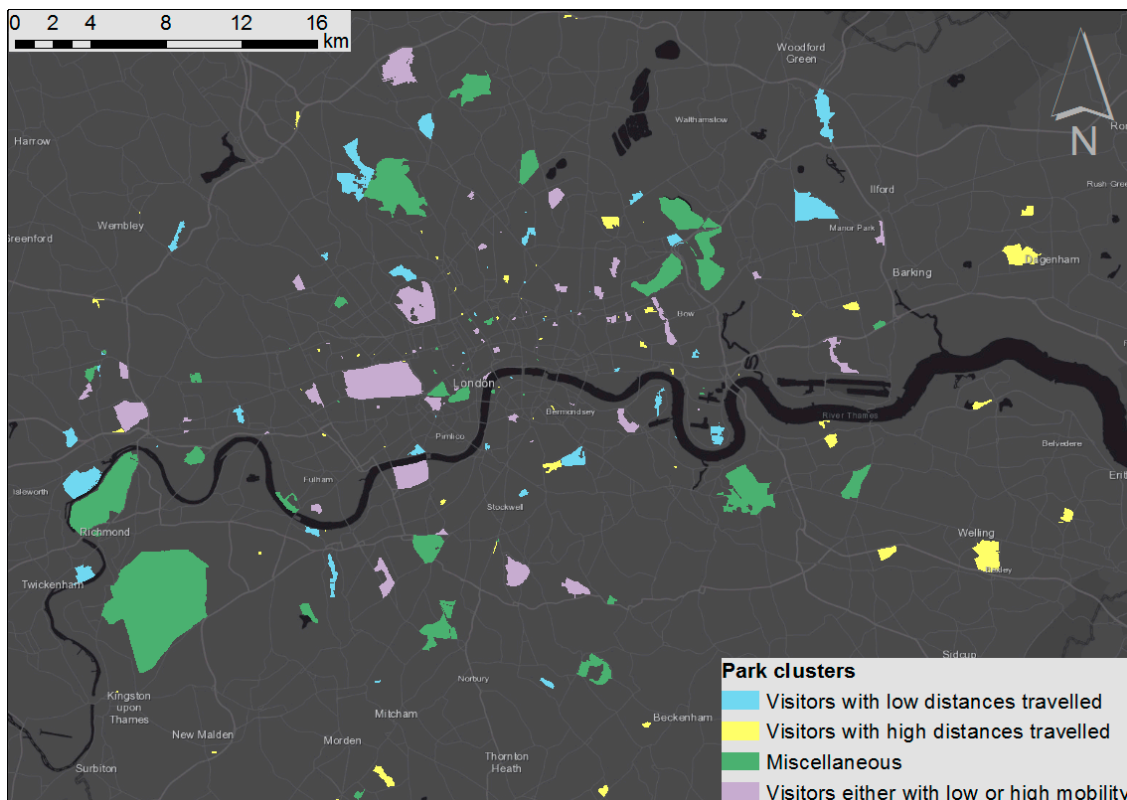


Figure 10. Park categories based on the spatial characteristics of their visitors' behavior.

4.2. Sentiments and Emotions

In this part of the analysis, positive and negative sentiments along with eight different emotions were extracted for each tweet. After the algorithm assigned a value to each of these new attributes,

we compared whether there is a difference between park tweets and non-park tweets. Based on previous studies (e.g., [47]) the hypothesis was that park tweets could be more positive. However, our results only partially confirm this. If we consider all the tweets (Figure 11A), not distinguishing them based on users, the proportion of positive non-park tweets (among all non-park tweets) is higher than the proportion of positive park tweets among park tweets. However, if we first calculate the user-level proportions of positive and negative sentiments for park and non-park tweets, we get exactly the opposite results (Figure 11B). The proportion of positive tweets in the parks is higher than the proportion of other positive tweets posted outside the parks. This means that, in general, Twitter users are more positive while being in a park.

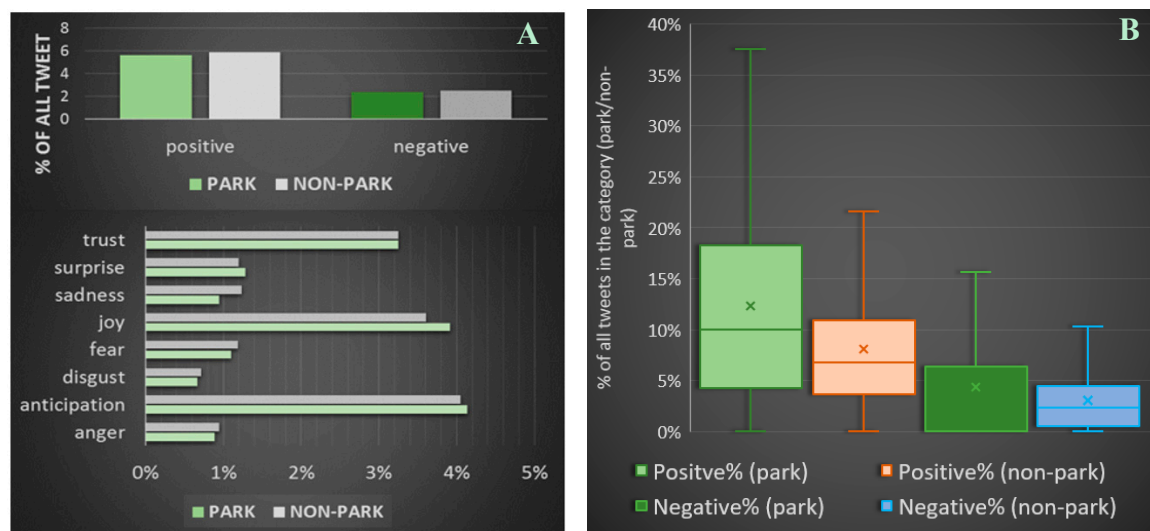


Figure 11. Percentage of sentiments and emotions for park tweets and non-park tweets ((A) all tweets considered in one step; (B) aggregated user-level values).

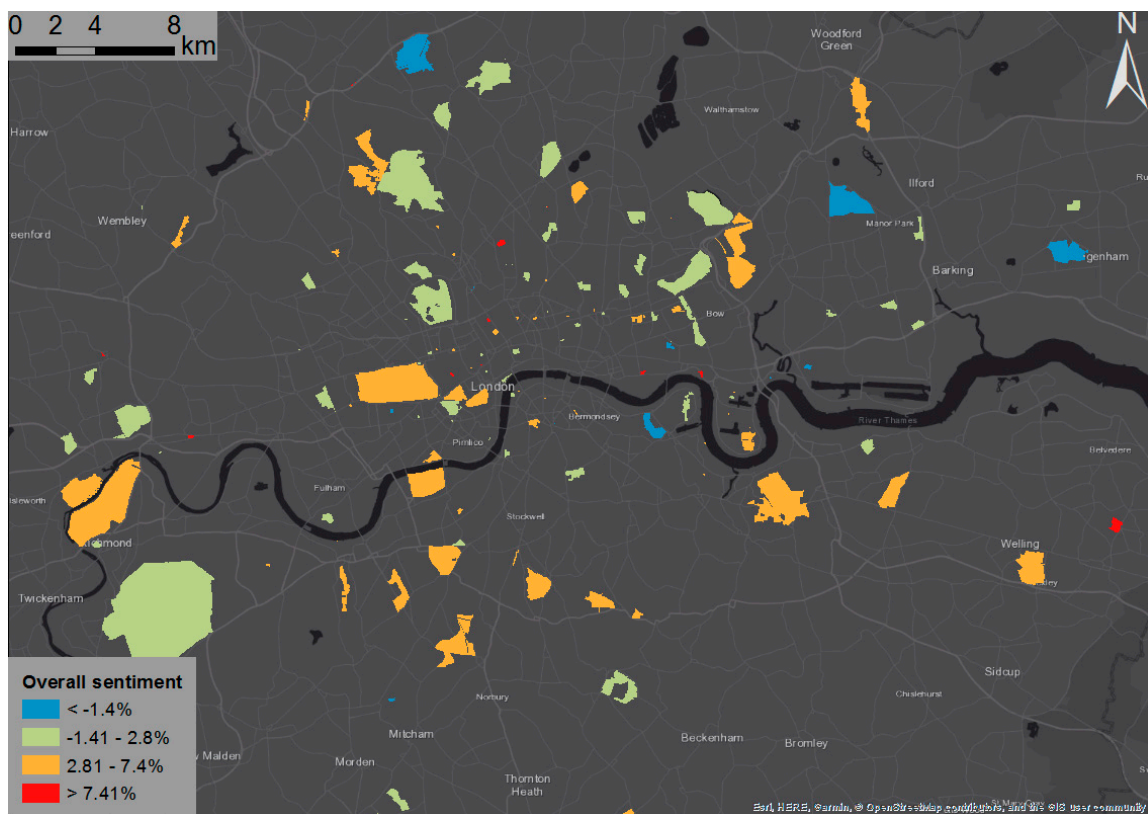
Regarding the emotion categories; surprise, joy, and anticipation are higher in proportion for park tweets than among non-park tweets (Figure 11). We used a one-factor ANOVA (Analysis Of Variance) test to compare park and non-park tweets to determine whether this difference is statistically significant (Table 4). Bold font denotes significant differences between the two data sets. Polarity means that all sentiment values were considered (negative sentiments got a negative sign). Except for the polarity, anticipation, and trust, all other differences are significant between the two data sets. However, the values representing differences are relatively low.

Figure 12 shows the polarity on park level. Polarity represents the overall sentiment, which means that the percentage of tweets with negative sentiments is subtracted from the percentage of positive tweets. For example, if 6% of the tweets in a park are positive and 2% are negative, the overall score will be 4%. In this way, we can see that there are parks with more negative than positive sentiments in the tweets posted from there (blue color in the map). Parks tend to have a higher overall sentiment (=considerably higher number of positive tweets than negative, orange and red polygons) south of the river Thames; however, the parks with the highest overall sentiment (red) are mostly located in the inner city, on the opposite side of the river. A detailed analysis including a more in-depth content analysis, which is beyond the scope of the present study, can investigate whether the high number and proportion of negative tweets reflect some serious issues regarding those parks or there are different reasons for it.

Table 4. Difference between park tweets and non-park tweets for sentiments and emotions.

Sentiment or Emotion	Significance (<i>p</i>)	Result	Difference (%)
Sentiment polarity	0.482183466	Not significant	
Anger	<0.001	less anger in parks	0.50%
Anticipation	0.170095853	Not significant	
Disgust	0.001601268	less disgust in parks	0.34%
Fear	<0.001	less fear in parks	1.19%
Joy	<0.001	more joy in parks	0.71%
Sadness	<0.001	less sadness in parks	1.49%
Surprise	<0.001	more surprise in parks	0.50%
Trust	0.875638565	Not significant	
Positive sentiment *	0.015684988	less positive in parks	0.22%
Negative sentiment *	0.00279889	less negative in parks	0.18%
Positive emotions	<0.001	less positive in parks	1.07%
Negative emotions	<0.001	less negative in parks	1.14%

* were considered as binary value (1 = sentiment identified, 0 = no sentiment).

**Figure 12.** Overall sentiment scores in parks with at least 100 tweets.

4.3. Temporal Variability of the Results

4.3.1. Number of Tweets

As a first step, the absolute number of tweets was grouped to yearly, seasonal, weekly, and daily periods (Figure 13). The yearly and seasonal distribution clearly reflects a higher proportion of tweets occurring in parks during spring and summer, in accordance with similar research (e.g., [28]). Interestingly, there are slightly fewer tweets during fall than winter, reflected by the high number of tweets in January and February. Considering the weather conditions in January and February, it is surprising that these two winter months have almost the same amount of park tweets as some periods during the spring. The weekly pattern is quite regular with almost the same number of park tweets

on every weekday, while at weekends the numbers are slightly higher than during the week but still almost identical on Saturday and Sunday. The daily trend follows an obvious, ordinary pattern with almost no tweets during the night, one peak in the afternoon at 2:00 p.m. and another relative peak in the evening around 9:00 p.m.

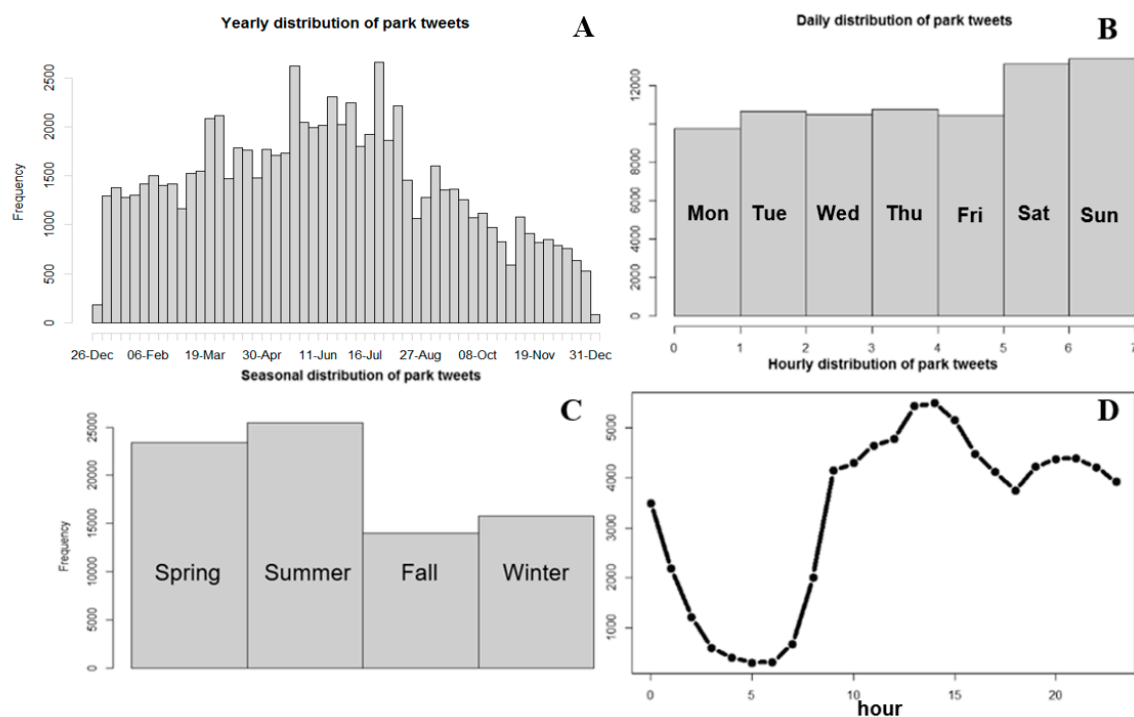


Figure 13. Temporal distribution of tweet frequency ((A) yearly; (B) weekly; (C) seasonal; and (D) hourly).

Parks were also clustered according to the temporal characteristics in terms of visits (Figure 18). The daily pattern was divided into four groups, one where the proportion of tweets is almost constant and the other three with a peak for each temporal unit (morning, afternoon, evening). The results were similar for the seasons as well—one group with a clear peak in spring and another in winter—whereas the other two groups have similar values in spring-summer or spring-fall. The weekly pattern is not shown; there were two groups, in both of which the proportion of weekday tweets was higher than weekend tweets, and in one of the two groups the difference between the weekend and weekday values was slightly bigger. As the groups were almost identical based on this weekly pattern, it has no significant effect on the final park categories, so we excluded them.

Figure 14 shows park clusters according to the daily patterns. The first group (red) contains the most parks compared to the other three groups, and their size and location vary to a large degree. These parks experience a visitor peak both in the afternoon and the evening. The second group (light blue) mostly encompasses smaller parks and has an evening peak, whereas the parks in the third group (yellow) are relatively small, located closer to the center and people visit them mostly in the morning. The last group (green) has larger parks with an afternoon peak.

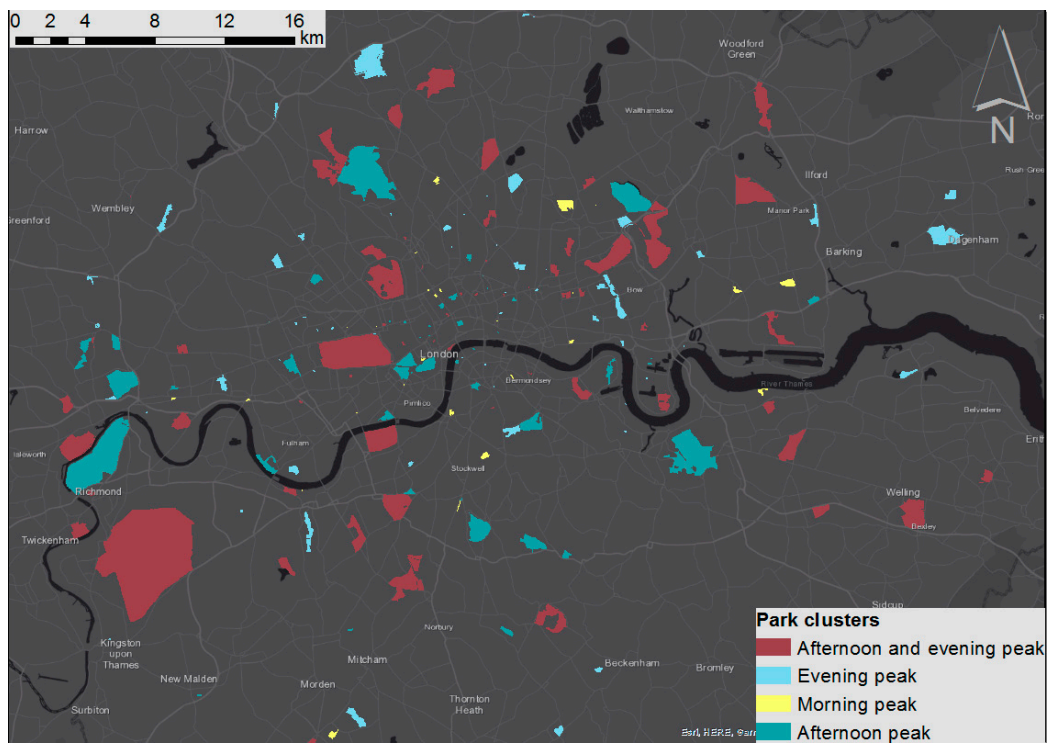


Figure 14. Park clusters according to visitors’ spatial behavior.

4.3.2. Temporal Patterns of Positive Tweets

Figure 15 shows how the proportion of positive tweets vary over the day on weekdays. For visualization purposes we selected parks with the highest number of positive tweets in the given temporal category. There is no clear pattern observable, there are parks both in the inner city and towards the edge of the study area with higher or lower values for example in the afternoon.

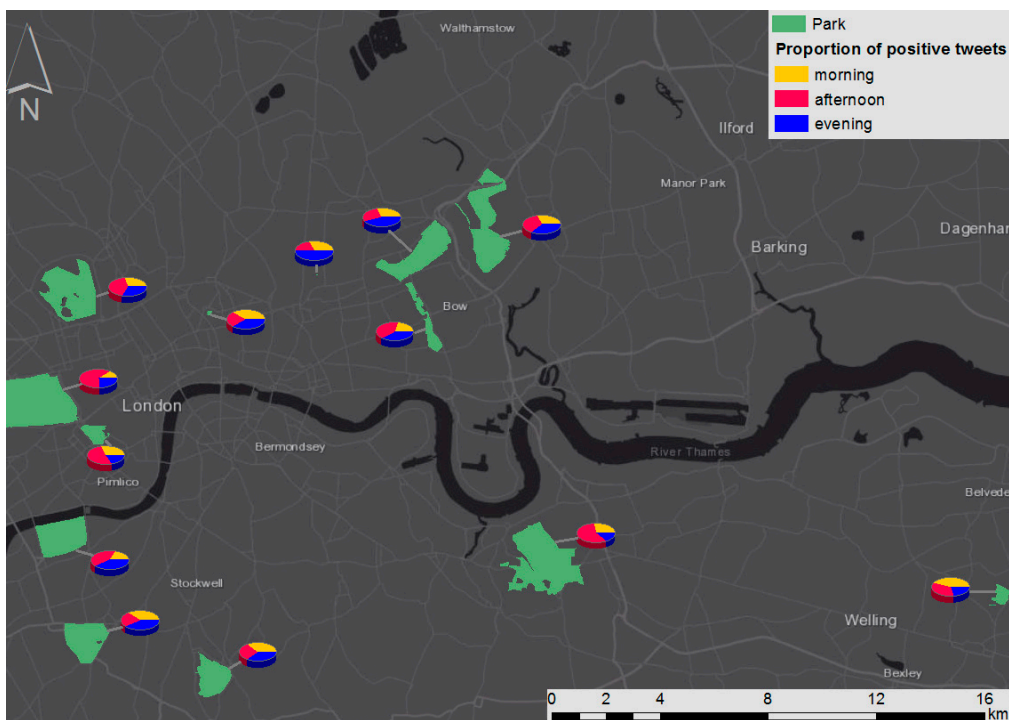


Figure 15. Proportion of positive tweets during the day on weekdays.

Figure 16 shows the proportion of positive tweets at weekends. There were only six parks selected for visualization, using the same criterion (at least 5 tweet/temporal category) as for Figure 15. The parks located closer to the city center (north of the river Thames) has relatively low number of positive tweets during the morning, while the visitors are rather positive in the evening and the afternoon. The numbers are absolute values, but these trends are following a different pattern than the number of visits. Both Figures 15 and 16 are illustrations for a few parks and the same graph can be generated for any park in the analysis, also representing seasonal differences and negative tweets, depending on the purpose of a more in-depth investigation.

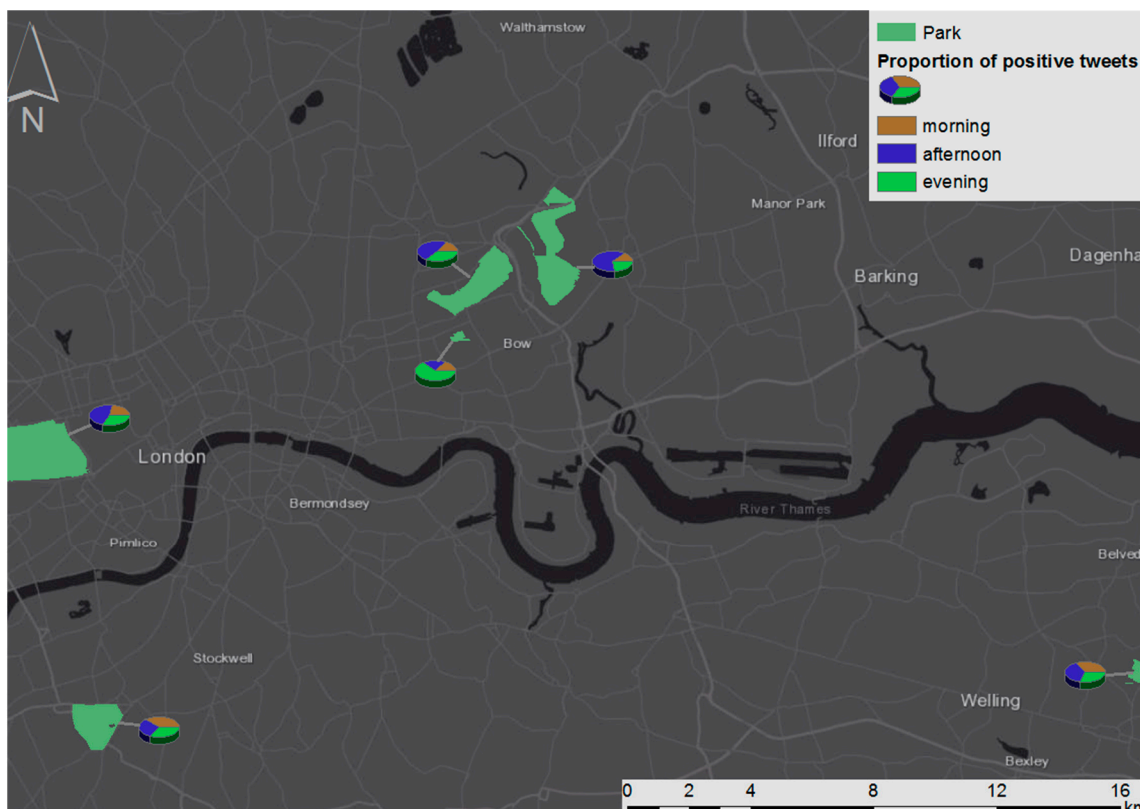


Figure 16. Proportion of positive tweets during the day at the weekends.

4.3.3. Temporal Patterns of Emotions

Figure 17 shows the temporal variation of tweets with “fear” as the identified emotion. The emotion of fear was selected for this illustration because this had the highest difference between non-park and park tweets (similarly to e.g., [30]), but it can be generated for any of the eight emotions. Regarding the variation of proportions during the day, we can see that many parks have a higher number of fear tweets in the evening, which might indicate safety problems. However, other parks where this is not the case, and most of the “fear tweets” were posted in the morning, might be also interesting for further analysis to identify the reasons. Parks shown in blue (from light to dark blue) have more tweets during the weekends, while the minority of the parks shown in pink and purple are visited more often on weekdays.

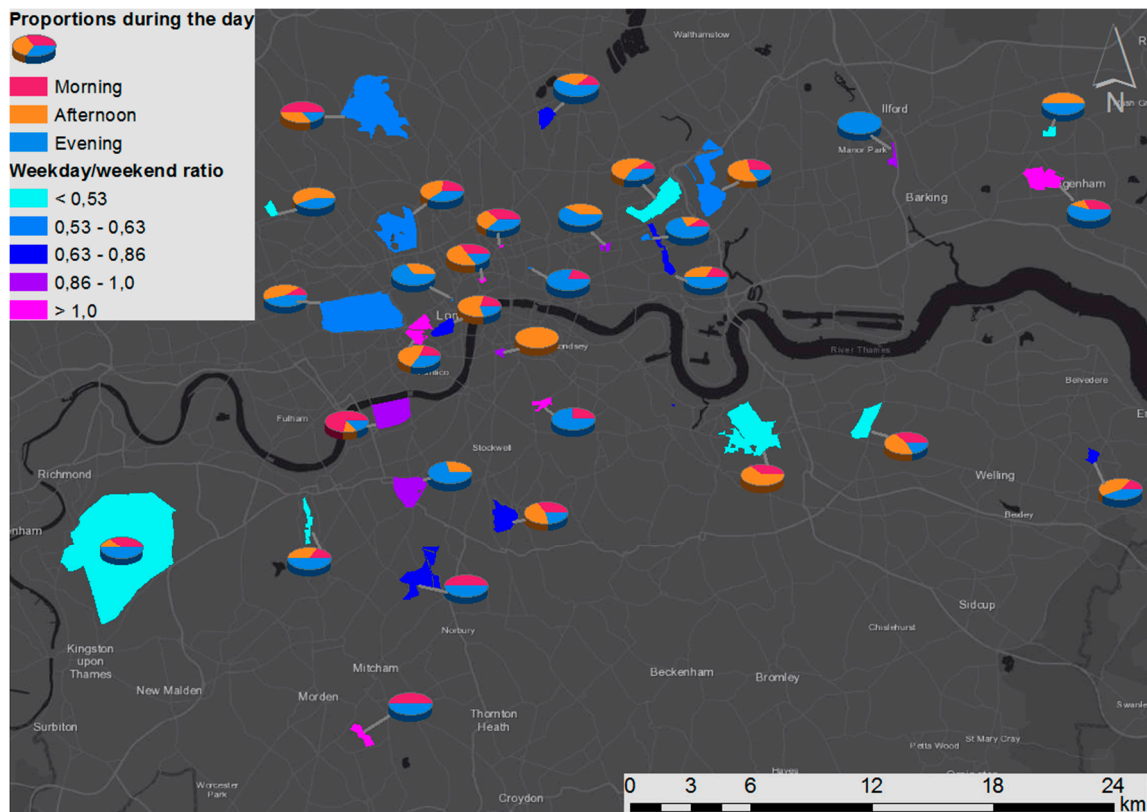


Figure 17. Proportion of fear tweets during the day, indicating weekday/weekend ratio as well.

4.4. Comprehensive Park Profiles

As the basis of the final clustering to define comprehensive park profiles, we used four sub-clusters representing user types, sentiments, and temporal (daily, seasonal) characteristics (Figure 18). The sub-clustering was useful for the interpretation of each final cluster (see Sections 4.1 and 4.3). With regard to the negative and positive emotions or sentiments, there were only two meaningful groups, one with a lower and one with a higher proportion of both negative and positive sentiments and emotions, so it is not shown in Section 4.2.

Table 5 shows the final clusters for parks. We clustered 197 parks based on the average number of tweets per park for each of the sub-clusters (at least 45 or 70 tweets). The categorization could be done for all parks. However, due to the low number of tweets, the reliability of the results is reduced. The first row of the table describes the first park category. The visitors of these parks tend to visit various parks around the city, also ones that are further apart from each other (user Cluster 2). In terms of sentiments, the visitors belong to Cluster 1, with lower values regarding mobility. Parks in the first category have most visitors in the evening, and they have more visitors tweeting during spring and summer. The second category contains parks whose visitors mostly visit the same park (user Cluster 3), while the proportion of the tweets of these users' sentiments and emotions are lower. Parks in the second category are mostly active during the day, and, similarly to the first category, spring and summer are dominant seasons with more tweets from visitors than in other seasons. The third category is identical to the first one except that the parks are visited more in the morning, and, instead of summer, they are more popular during the spring and fall. The last category comprises all the parks where most users belong to Cluster 4; there are fewer sentiments and emotions, they are mostly active during the day, and partly in the evening. Figure 19 shows the different types of parks on the map.

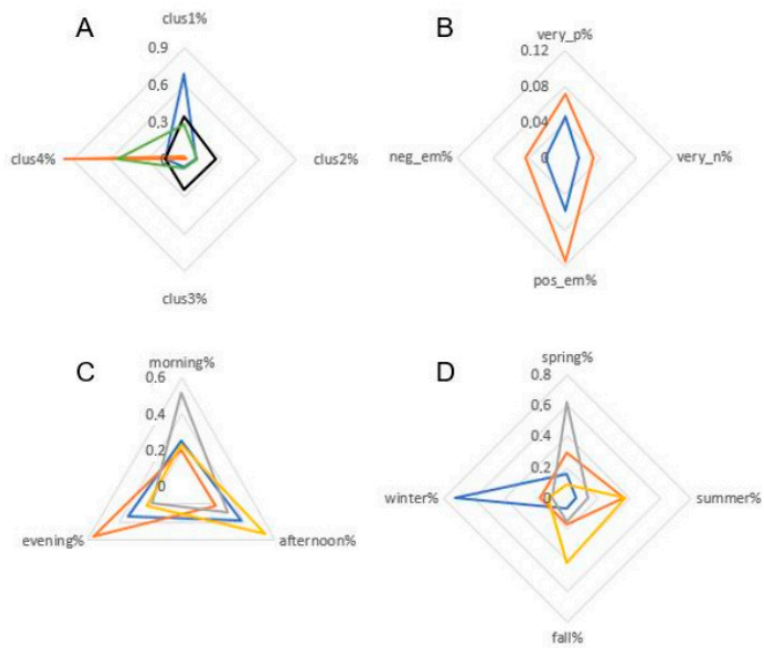


Figure 18. Sub-clusters for parks (A) user clusters; (B) sentiment and emotion proportions; (C) daily pattern; (D) seasonal pattern).



Figure 19. Final park categories.

Table 5. Park clusters.

Category	User Types	Sentiment and Emotions	Daily Pattern	Seasonal Pattern
1	Park COM to park distance is larger (Cl. 2)	Higher	Evening peak	Spring and summer higher
2	Every distance value is high except park to park COM (Cl. 3)	Lower	Afternoon peak	Spring and summer higher
3	Park COM to park distance is larger (Cl. 2)	Higher	Morning peak	Spring and fall higher
4	Every distance value is in the mid-range (Cl. 4)	Lower	Afternoon and evening peak	Spring and summer higher

5. Discussion

As this analysis demonstrates, Twitter data is one promising resource to assess the characteristics of urban parks, analyzing spatial, temporal and content-specific aspects. The advantage of using this type of social media for urban green space analysis is that we can derive qualitative, fine-scale information for the entire city as input for more specific, in-depth investigations.

Despite the potential of big data and social media data in this kind of analysis, the representativeness of the gathered information does have limitations due to uncertainties in the demographics of the users. However, we can infer at least that extremes in age (both very young and very old populations) and social situation (mostly poorer populations) translate into lower rates of social media use, and therefore under-representation in the data. In urban planning and in the analysis of urban green areas, such demographic data, and other data such as sex, ethnicity, etc., are important factors. It is a problem that social media data usually contains no direct information regarding these factors, unlike traditional census data, but there are methods to extract them indirectly [59]. While this was not part of the current study, it could add new insights to the analysis results.

The main limitation in our case, besides the general one of representativeness, was the low number of tweets per user, especially in the case of park tweets. This resulted in a less than ideal number of data points per user. Similarly, when we extracted sentiments and emotions, due to the limitations of the algorithm, most of the tweets were classified as neutral or had no identifiable emotion. This again strongly affected the number of tweets used for the analysis. Also, the sentiment analysis itself has uncertainty as it considers words individually and not in context, while some “strong” words can bias the overall sentiment score. Finally, our selection criteria did not specify whether the content of a given tweet from a park is about the park or not. Therefore, a higher number of negative tweets does not necessarily indicate bad park quality, but further investigation can specify the connection and reason for the observation.

Another issue emerged from the categorization of users as residents. The method has uncertainties, as we are not able to validate it with official data sources. Although the selected users might not be residents, based on the number of tweets and their temporal distribution, we can at least conclude that their behavior is appropriate for investigation, as it has the potential to represent the activity patterns of average users, including patterns over the course of the year. Through this averaging process, we can derive more information, which is a common practice to improve the credibility of the results, e.g., [50,52,95].

Regarding the tweeting behavior in general, another consideration is that while users will not necessarily tweet every time they are in a park, the tweet frequency is still able to show relative differences between parks. On the other hand, even if users tweet, they might not share their location. As with demographic information, location data can also be extracted indirectly through various methods, which might be relevant for future investigations.

Finally, the transferability of the methods is important and was considered in our analysis. As mentioned above, this is the reason we used OpenStreetMap, because, depending on the availability of Twitter data, urban green areas in any city can be analyzed following this methodology, making

even global comparisons possible. However, access to tweets, in general, is usually not free of charge, especially for longer periods of time, and this constraint can negatively influence the transferability of the methods.

6. Conclusions

This study has tested an exploratory methodology to investigate spatial, temporal, and affective patterns of park visits for urban planning purposes using Twitter data of frequent users, and thereby to define profiles of parks and their visitors. The performed analyses yielded new insights about the visitors and use patterns of urban parks in London. In particular, we found that most users tend to tweet from parks that are located 3–4 km away from their COM and the average distance between a park and its visitors' COM increases towards the outer areas of the city. Even though social media data is not appropriate to investigate motivation and determining factors of park visits directly, these higher average distance values suggest that absolute distance to a park has a lower priority in deciding which park to visit. Nevertheless, the larger absolute distance can still imply good accessibility.

In terms of sentiments, statistical analysis confirmed a significantly higher number of positive tweets in parks than in other urban areas, when considering tweets on an individual level. However, if we do not distinguish individual users in the analysis, this difference is already less obvious. Regarding emotions, joy and anticipation are significantly more frequent in parks than outside of them, but all the other emotions are more common in non-park areas, and these proportions can also vary from park to park. The temporal distribution of the tweets mostly corresponded to general expectations with more tweets in the afternoon, weekend, and summer, although surprisingly there were more tweets from parks during the winter than fall in our analysis period. Interestingly, on the park level, there was hardly any observable temporal trend for the number of visitors or the sentiment and emotion of the tweets. In summary, we identified four groups of parks based on their visitors' characteristics, emotions, and how the visiting of these parks and the emotions in the tweets posted there varies over time.

While the methodologies and technologies of spatiotemporal social media analyses are developing fast, it seems that GIScience needs to work towards the exploration of causal relationships and the realization of a GIS of place. This study may contribute to the incorporation of the traditional quantitative spatial analytical tools of GIS with "non-traditional" data towards the realization of GIS as a hypothesis-generator. What is needed is not so much the development of many more analysis models, but rather, new ways of integrating mixed-methods approaches that incorporate a sense of place.

Although social media data analysis has its limitations, it could be shown that an exhaustive spatial, temporal and content analysis can provide valuable information through grasping general trends, serving as input for more in-depth analysis and field research, providing more specific purposes for urban planners and decision makers.

Author Contributions: All authors have contributed to this paper. A.K.-G. Proposed the main idea. A.K.-G., A.R., R.K. and A.C. were involved in the design of methodology and performed the analyses. A.K.-G. and A.R. drafted the manuscript. B.R. and T.B. contributed to the final version of the paper.

Funding: This research was funded by the Austrian Science Fund (FWF) through the Doctoral College GIScience at the University of Salzburg (DK W 1237-N23). We would also like to express our gratitude to the FWF for supporting the project "Urban Emotions", reference number I-3022 and the project "The Scales and Structures of Intra-Urban Spaces" (reference number P 29135-N29).

Acknowledgments: We thank Michael Mehaffy for his assistance with the validation of the concept, and for his comments that greatly improved the manuscript.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

- Batty, M. *The New Science of Cities*; MIT Press: Cambridge, MA, USA, 2013; ISBN 9780262019521.
- Keniger, L.E.; Gaston, K.J.; Irvine, K.N.; Fuller, R.A. What are the benefits of interacting with nature? *Int. J. Environ. Res. Public Health* **2013**, *10*, 913–935. [[CrossRef](#)] [[PubMed](#)]
- Costanza, R.; D'Arge, R.; De Groot, R.; Farber, S.; Grasso, M.; Hannon, B.; Limburg, K.; Naeem, S.; O'Neill, R.V.; Paruelo, J.; et al. The value of the world's ecosystem services and natural capital. *Nature* **1997**, *387*, 253–260. [[CrossRef](#)]
- Sturm, R.; Cohen, D. Proximity to urban parks and mental health. *J. Ment. Health Policy Econ.* **2014**, *17*, 19–24. [[PubMed](#)]
- Van den Berg, M.; Wendel-Vos, W.; van Poppel, M.; Kemper, H.; van Mechelen, W.; Maas, J. Health benefits of green spaces in the living environment: A systematic review of epidemiological studies. *Urban For. Urban Green.* **2015**, *14*, 806–816. [[CrossRef](#)]
- Wolch, J.R.; Byrne, J.; Newell, J.P. Urban green space, public health, and environmental justice: The challenge of making cities “just green enough”. *Landsc. Urban Plan.* **2014**, *125*, 234–244. [[CrossRef](#)]
- Chiesura, A. The role of urban parks for the sustainable city. *Landsc. Urban Plan.* **2004**, *68*, 129–138. [[CrossRef](#)]
- Roe, J.J.; Ward Thompson, C.; Aspinall, P.A.; Brewer, M.J.; Duff, E.I.; Miller, D.; Mitchell, R.; Clow, A. Green space and stress: Evidence from cortisol measures in deprived urban communities. *Int. J. Environ. Res. Public Health* **2013**, *10*, 4086–4103. [[CrossRef](#)] [[PubMed](#)]
- Akpınar, A.; Barbosa-Leiker, C.; Brooks, K.R. Does green space matter? Exploring relationships between green space type and health indicators. *Urban For. Urban Green.* **2016**, *20*, 407–418. [[CrossRef](#)]
- Lee, A.C.K.; Maheswaran, R. The health benefits of urban green spaces: A review of the evidence. *J. Public Health* **2011**, *33*, 212–222. [[CrossRef](#)] [[PubMed](#)]
- Takano, T.; Nakamura, K.; Watanabe, M. Urban residential environments and senior citizens' longevity in megacity areas: The importance of walkable green spaces. *J. Epidemiol. Community Health* **2002**, *56*, 913–918. [[CrossRef](#)] [[PubMed](#)]
- Hartig, T.; Mitchell, R.; de Vries, S.; Frumkin, H. Nature and Health. *Annu. Rev. Public Health* **2014**, *35*, 207–228. [[CrossRef](#)] [[PubMed](#)]
- Liu, H.L.; Shen, Y.S. The impact of green space changes on air pollution and microclimates: A case study of the taipei metropolitan area. *Sustainability* **2014**, *6*, 8827–8855. [[CrossRef](#)]
- Maimaitiyiming, M.; Ghulam, A.; Tiyyip, T.; Pla, F.; Latorre-Carmona, P.; Halik, Ü.; Sawut, M.; Caetano, M. Effects of green space spatial pattern on land surface temperature: Implications for sustainable urban planning and climate change adaptation. *ISPRS J. Photogramm. Remote Sens.* **2014**, *89*, 59–66. [[CrossRef](#)]
- Budruk, M.; Thomas, H.; Tyrrell, T. Urban green spaces: A study of place attachment and environmental attitudes in India. *Soc. Nat. Resour.* **2009**, *22*, 824–839. [[CrossRef](#)]
- Kim, J.; Kaplan, R. Physical and psychological factors in sense of community: New urbanist Kentlands and nearby orchard village. *Environ. Behav.* **2004**, *36*, 313–340. [[CrossRef](#)]
- Tyrväinen, L.; Ojala, A.; Korpela, K.; Lanki, T.; Tsunetsugu, Y.; Kagawa, T. The influence of urban green environments on stress relief measures: A field experiment. *J. Environ. Psychol.* **2014**, *38*, 1–9. [[CrossRef](#)]
- Anguluri, R.; Narayanan, P. Role of green space in urban planning: Outlook towards smart cities. *Urban For. Urban Green.* **2017**, *25*, 58–65. [[CrossRef](#)]
- Hartig, T.; Kahn, P.H. Living in cities, naturally. *Science* **2016**, *352*, 938–940. [[CrossRef](#)] [[PubMed](#)]
- Schetke, S.; Qureshi, S.; Lautenbach, S.; Kabisch, N. What determines the use of urban green spaces in highly urbanized areas?—Examples from two fast growing Asian cities. *Urban For. Urban Green.* **2016**, *16*, 150–159. [[CrossRef](#)]
- Lindberg, M.; Schipperijn, J. Active use of urban park facilities—Expectations versus reality. *Urban For. Urban Green.* **2015**, *14*, 909–918. [[CrossRef](#)]
- Goličnik, B.; Ward Thompson, C. Emerging relationships between design and use of urban park spaces. *Landsc. Urban Plan.* **2010**, *94*, 38–53. [[CrossRef](#)]
- Ives, C.D.; Oke, C.; Hehir, A.; Gordon, A.; Wang, Y.; Bekessy, S.A. Capturing residents' values for urban green space: Mapping, analysis and guidance for practice. *Landsc. Urban Plan.* **2017**, *161*, 32–43. [[CrossRef](#)]
- Bedimo-Rung, A.L. The Significance of Parks to Physical Activity and Public Health: A Conceptual Model. *Am. J. Prev. Med.* **2005**, *28* (Suppl. S2), 159–168. [[CrossRef](#)]

25. Kothencz, G.; Blaschke, T. Urban parks: Visitors' perceptions versus spatial indicators. *Land Use Policy* **2017**, *64*, 233–244. [CrossRef]
26. Ode Sang, Å.; Knez, I.; Gunnarsson, B.; Hedblom, M. The effects of naturalness, gender, and age on how urban green space is perceived and used. *Urban For. Urban Green.* **2016**, *18*, 268–276. [CrossRef]
27. United Nations General Assembly Transforming Our World: The 2030 Agenda for Sustainable Development. 2015. Available online: <https://sustainabledevelopment.un.org/content/documents/7891Transforming%20Our%20World.pdf> (accessed on 26 June 2018).
28. Roberts, H.; Sadler, J.; Chapman, L. Using Twitter to investigate seasonal variation in physical activity in urban green space. *Geo Geogr. Environ.* **2017**, *4*, e00041. [CrossRef]
29. Roberts, H.V. Using Twitter data in urban green space research: A case study and critical evaluation. *Appl. Geogr.* **2017**, *81*, 13–20. [CrossRef]
30. Roberts, H.; Sadler, J.; Chapman, L. The value of Twitter data for determining the emotional responses of people to urban green spaces: A case study and critical evaluation. *Urban Stud.* **2018**. [CrossRef]
31. Lee, G.; Hong, I. Measuring spatial accessibility in the context of spatial disparity between demand and supply of urban park service. *Landsc. Urban Plan.* **2013**, *119*, 85–90. [CrossRef]
32. La Rosa, D. Accessibility to greenspaces: GIS based indicators for sustainable planning in a dense urban context. *Ecol. Indic.* **2014**, *42*, 122–134. [CrossRef]
33. Kolcsár, R.A.; Szilassi, P. Assessing accessibility of urban green spaces based on isochrone maps and street resolution population data through the example of Zalaegerszeg, Hungary. *Carpath. J. Earth Environ. Sci.* **2018**, *13*, 31–36. [CrossRef]
34. Cohen, D.A.; McKenzie, T.L.; Sehgal, A.; Williamson, S.; Golinelli, D.; Lurie, N. Contribution of public parks to physical activity. *Am. J. Public Health* **2007**, *97*, 509–514. [CrossRef] [PubMed]
35. Maas, J.; Verheij, R.A.; De Vries, S.; Spreeuwenberg, P.; Schellevis, F.G.; Groenewegen, P.P. Morbidity is related to a green living environment. *J. Epidemiol. Community Health* **2009**, *63*, 967–973. [CrossRef] [PubMed]
36. Scott, M.M.; Evenson, K.R.; Cohen, D.A.; Cox, C.E. Comparing perceived and objectively measured access to recreational facilities as predictors of physical activity in adolescent girls. *J. Urban Heal.* **2007**, *84*, 346–359. [CrossRef] [PubMed]
37. Wang, D.; Brown, G.; Liu, Y. The physical and non-physical factors that influence perceived access to urban parks. *Landsc. Urban Plan.* **2015**, *133*, 53–66. [CrossRef]
38. Kaczynski, A.T.; Potwarka, L.R.; Saelens, B.E. Association of park size, distance, and features with physical activity in neighborhood parks. *Am. J. Public Health* **2008**, *98*, 1451–1456. [CrossRef] [PubMed]
39. Dony, C.C.; Delmelle, E.M.; Delmelle, E.C. Re-conceptualizing accessibility to parks in multi-modal cities: A Variable-width Floating Catchment Area (VFCA) method. *Landsc. Urban Plan.* **2015**, *143*, 90–99. [CrossRef]
40. Gupta, K.; Roy, A.; Luthra, K.; Maithani, S. Mahavir GIS based analysis for assessing the accessibility at hierarchical levels of urban green spaces. *Urban For. Urban Green.* **2016**, *18*, 198–211. [CrossRef]
41. Ekkel, E.D.; de Vries, S. Nearby green space and human health: Evaluating accessibility metrics. *Landsc. Urban Plan.* **2017**, *157*, 214–220. [CrossRef]
42. Bauman, A.E.; Bull, F.C. *Environmental Correlates of Physical Activity and Walking in Adults and Children: A Review of Reviews*; National Institute of Health and Clinical Excellence: Loughborough, UK, 2007.
43. Kashef, M. Urban livability across disciplinary and professional boundaries. *Front. Archit. Res.* **2016**, *5*, 239–253. [CrossRef]
44. IMCL. The Value of Rankings and the Meaning of Livability. Available online: <http://www.livablecities.org/blog/value-rankings-and-meaning-livability> (accessed on 19 April 2017).
45. Salzano, E. Seven Aims for the Livable City. In *Making Cities Livable—Wege zur Menschlichen Stadt*; Lennard, S.H.C., von Ungern-Sternberg, S., Lennard, H.L., Eds.; Gondolier Press: Carmel, CA, USA, 1997; pp. 18–20, ISBN 0-937824-08-1.
46. Veenhoven, R. The Four Qualities of Life. *J. Happiness Stud.* **2000**, *1*, 1–39. [CrossRef]
47. Lim, K.H.; Lee, K.E.; Kendal, D.; Rashidi, L.; Naghizade, E.; Winter, S.; Vasardani, M. The grass is greener on the other side: Understanding the effects of green spaces on Twitter user sentiments. In Proceedings of the WWW '18 Companion, The Web Conference 2018, Lyon, France, 23–27 April 2018; pp. 275–282. [CrossRef]
48. Martí, P.; Serrano-Estrada, L.; Nolasco-Cirugeda, A. Using locative social media and urban cartographies to identify and locate successful urban plazas. *Cities* **2017**, *64*, 66–78. [CrossRef]

49. Crooks, A.; Pfoser, D.; Jenkins, A.; Croitoru, A.; Stefanidis, A.; Smith, D.; Karagiorgou, S.; Efentakis, A.; Lamprianidis, G. Crowdsourcing urban form and function. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 720–741. [[CrossRef](#)]
50. Hasan, S.; Ukkusuri, S.V. Urban activity pattern classification using topic models from online geo-location data. *Transp. Res. Part C Emerg. Technol.* **2014**, *44*, 363–381. [[CrossRef](#)]
51. Ciuccarelli, P.; Lupi, G.; Simeone, L. *Visualizing the Data City—Social Media as a Source of Knowledge for Urban Planning and Management*, 1st ed.; Springer International Publishing: Berlin, Germany, 2014; ISBN 978-3-319-02194-2.
52. Wu, L.; Zhi, Y.; Sui, Z.; Liu, Y. Intra-urban human mobility and activity transition: Evidence from social media check-in data. *PLoS ONE* **2014**, *9*, e97010. [[CrossRef](#)] [[PubMed](#)]
53. Aubrecht, C.; Ungar, J.; Freire, S. Exploring the potential of volunteered geographic information for modeling spatio-temporal characteristics of urban population: A case study for Lisbon Metro using foursquare check-in data. In Proceedings of the 7th International Conference Virtual Cities and Territories, Lisbon, Portugal, 11–13 October 2011.
54. Fujisaka, T.; Lee, R.; Sumiya, K. Exploring urban characteristics using movement history of mass mobile microbloggers. In Proceedings of the Eleventh Workshop on Mobile Computing Systems & Applications—HotMobile '10, Annapolis, MD, USA, 22–23 February 2010; pp. 13–18.
55. Fujisaka, T.; Lee, R.; Sumiya, K. Discovery of user behavior patterns from geo-tagged micro-blogs. In Proceedings of the 4th International Conference on Ubiquitous Information Management and Communication—ICUIMC '10, Suwon, Korea, 14–15 January 2010; p. 36.
56. Hawelka, B.; Sitko, I.; Beinart, E.; Sobolevsky, S.; Kazakopoulos, P.; Ratti, C. Geo-located Twitter as proxy for global mobility patterns. *Cartogr. Geogr. Inf. Sci.* **2014**, *41*, 260–271. [[CrossRef](#)] [[PubMed](#)]
57. Blanford, J.I.; Huang, Z.; Savelyev, A.; MacEachren, A.M. Geo-located tweets. Enhancing mobility maps and capturing cross-border movement. *PLoS ONE* **2015**, *10*, e0129202. [[CrossRef](#)] [[PubMed](#)]
58. Cranshaw, J.; Hong, J.I.; Sadeh, N. The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. In Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media, Dublin, Ireland, 4–8 June 2012.
59. Longley, P.A.; Adnan, M. Geo-temporal Twitter demographics. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 369–389. [[CrossRef](#)]
60. Shelton, T.; Poorthuis, A.; Zook, M. Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landsc. Urban Plan.* **2015**, *142*, 198–211. [[CrossRef](#)]
61. Cheng, Z.; Caverlee, J.; Lee, K. You Are Where You Tweet: A Content-Based Approach to Geo-locating Twitter Users. In Proceedings of the 19th ACM International Conference on Information and Knowledge Management, Toronto, ON, Canada, 26–30 October 2010; pp. 759–768. [[CrossRef](#)]
62. Kinsella, S.; Murdock, V.; Hare, N.O. “I’m Eating a Sandwich in Glasgow”: Modeling Locations with Tweets. In Proceedings of the 3rd International Workshop on Search and Mining User-Generated Contents, Glasgow, UK, 28 October 2011; pp. 61–68.
63. Birkin, M.; Harland, K.; Malleson, N. The classification of space-time behaviour patterns in a British city from crowd-sourced data. In *Computational Science and Its Applications—ICCSA 2013*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2013; pp. 179–192.
64. Frank, M.R.; Mitchell, L.; Dodds, P.S.; Danforth, C.M. Happiness and the Patterns of Life: A Study of Geolocated Tweets. *Sci. Rep.* **2013**, *3*, 2625. [[CrossRef](#)] [[PubMed](#)]
65. Mitchell, L.; Frank, M.R.; Harris, K.D.; Dodds, P.S.; Danforth, C.M. The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place. *PLoS ONE* **2013**, *8*, e64417. [[CrossRef](#)] [[PubMed](#)]
66. Quercia, D.; Seaghdha, D.O.; Crowcroft, J. Talk of the City: Our Tweets, Our Community Happiness. In Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media, Dublin, Ireland, 4–8 June 2012; pp. 555–558.
67. Quercia, D.; Ellis, J.; Capra, L.; Crowcroft, J. Tracking “gross community happiness” from tweets. In Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work—CSCW '12, Seattle, WA, USA, 11–15 February 2012; p. 965.
68. Resch, B.; Summa, A.; Zeile, P.; Strube, M. Citizen-Centric Urban Planning through Extracting Emotion Information from Twitter in an Interdisciplinary Space-Time-Linguistics Algorithm. *Urban Plan.* **2016**, *1*, 114–127. [[CrossRef](#)]

69. Nguyen, Q.C.; Kath, S.; Meng, H.W.; Li, D.; Smith, K.R.; VanDerslice, J.A.; Wen, M.; Li, F. Leveraging geotagged Twitter data to examine neighborhood happiness, diet, and physical activity. *Appl. Geogr.* **2016**, *73*, 77–88. [[CrossRef](#)] [[PubMed](#)]
70. Campagna, M. The Geographic Turn in Social Media: Opportunities for Spatial Planning and Geodesign. In *Computational Science and Its Applications—ICCSA 2014*; Murgante, B., Misra, S., Rocha, A.M.A.C., Torre, C., Rocha, J.G., Falcão, M.I., Taniar, D., Apduhan, B.O., Gervasi, O., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 598–610.
71. Sagl, G.; Resch, B.; Hawelka, B.; Beinat, E. From Social Sensor Data to Collective Human Behaviour Patterns—Analysing and Visualising Spatio-Temporal Dynamics in Urban Environments. In *GI-Forum 2012: Geovisualization, Society and Learning*; Jekel, T., Car, A., Strobl, J., Griesebner, G., Eds.; Wichmann Verlag: Salzburg, Austria, 2012; pp. 54–63, ISBN 978-3-87907-521-8.
72. Lee, R.; Sumiya, K. Measuring geographical regularities of crowd behaviors for Twitter-based geo-social event detection. In Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks—LBSN '10, San Jose, CA, USA, 2 November 2010.
73. Gupta, A.; Kumaraguru, P. Credibility ranking of tweets during high impact events. In Proceedings of the 1st Workshop on Privacy and Security in Online Social Media—PSOSM '12, Lyon, France, 17 April 2012.
74. Zhang, Z.; Ni, M.; He, Q.; Gao, J. *Mining Transportation Information from Social Media for Planned and Unplanned Events*; University at Buffalo, SUNY: Buffalo, NY, USA, 2016.
75. Li, R.; Lei, K.H.; Khadiwala, R.; Chang, K.C.C. TEDAS: A twitter-based event detection and analysis system. In Proceedings of the 2012 IEEE 28th International Conference on Data Engineering, Washington, DC, USA, 1–5 April 2012.
76. Weng, J.; Yao, Y.; Leonardi, E.; Lee, F. Event Detection in Twitter. In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, Barcelona, Spain, 17–21 July 2011.
77. Panteras, G.; Wise, S.; Lu, X.; Croitoru, A.; Crooks, A.; Stefanidis, A. Triangulating Social Multimedia Content for Event Localization using Flickr and Twitter. *Trans. GIS* **2015**, *19*, 694–715. [[CrossRef](#)]
78. Fraustino, J.D.; Liu, B.; Yan, J. *Social Media Use during Disasters: A Review of the Knowledge Base and Gaps*; National Consortium for the Study of Terrorism and Responses to Terrorism: College Park, MD, USA, 2012.
79. Resch, B.; Usländer, F.; Havas, C. Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment. *Cartogr. Geogr. Inf. Sci.* **2017**, *45*, 362–376. [[CrossRef](#)]
80. Kovács-Györi, A.; Ristea, A.; Havas, C.; Resch, B.; Cabrera-Barona, P. #London2012: Towards citizen-contributed urban planning through sentiment analysis of twitter data. *Urban Plan.* **2018**, *3*, 75–99. [[CrossRef](#)]
81. Jankowski, P.; Czepkiewicz, M.; Młodkowski, M.; Zwoliński, Z. Geo-questionnaire: A Method and Tool for Public Preference Elicitation in Land Use Planning. *Trans. GIS* **2016**, *20*, 903–924. [[CrossRef](#)]
82. Brown, G.; Raymond, C.M. Methods for identifying land use conflict potential using participatory mapping. *Landsc. Urban Plan.* **2014**, *122*, 196–208. [[CrossRef](#)]
83. Innes, J.E.; Booher, D.E. Reframing public participation: Strategies for the 21st century. *Plan. Theory Pract.* **2004**, *5*, 419–436. [[CrossRef](#)]
84. Pietrzyk-Kaszyńska, A.; Czepkiewicz, M.; Kronenberg, J. Eliciting non-monetary values of formal and informal urban green spaces using public participation GIS. *Landsc. Urban Plan.* **2017**, *160*, 85–95. [[CrossRef](#)]
85. Bluemke, M.; Resch, B.; Lechner, C.; Westerholt, R.; Kolb, J.-P.; Kolb, J.-P. Integrating Geographic Information into Survey Research: Current Applications, Challenges and Future Avenues. *Surv. Res. Methods* **2017**, *11*, 307–327. [[CrossRef](#)]
86. Twitter Developers. Available online: <https://developer.twitter.com/> (accessed on 21 February 2017).
87. Steiger, E.; Resch, B.; Zipf, A. Exploration of spatiotemporal and semantic clusters of Twitter data using unsupervised neural networks. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 1694–1716. [[CrossRef](#)]
88. Hu, M.; Liu, B. Mining and summarizing customer reviews. In Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining—KDD '04, Seattle, WA, USA, 22–25 August 2004.
89. Breen, J.O. Mining Twitter for Airline Consumer Sentiment. In *Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications*; Academic Press: Orlando, FL, USA, 2012; ISBN 9780123869791. pp. 133–149. [[CrossRef](#)]

90. Mohammad, S.M.; Turney, P.D. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In Proceedings of the CAAGET '10 NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, Los Angeles, CA, USA, 5 June 2010.
91. Mohammad, S.M.; Turney, P.D. Crowdsourcing a word-emotion association lexicon. *Comput. Intell.* **2013**, *29*, 436–465. [[CrossRef](#)]
92. Jockers, M. Syuzhet: Extracts Sentiment and Sentiment-Derived Plot Arcs from Text (Version 1.0.1). Available online: <https://github.com/mjockers/syuzhet> (accessed on 14 July 2018).
93. Van der Laan, M.J.; Pollard, K.S.; Bryan, J. A new partitioning around medoids algorithm. *J. Stat. Comput. Simul.* **2003**, *73*, 575–584. [[CrossRef](#)]
94. Lin, J.; Cromley, R.G. Inferring the home locations of Twitter users based on the spatiotemporal clustering of Twitter data. *Trans. GIS* **2018**, *22*, 82–97. [[CrossRef](#)]
95. Abbasi, A.; Rashidi, T.H.; Maghrebi, M.; Waller, S.T. Utilising Location Based Social Media in Travel Survey Methods: Bringing Twitter Data into the Play. In Proceedings of the 8th ACM SIGSPATIAL International Workshop on Location-Based Social Networks, Bellevue, WA, USA, 3–6 November 2015.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).