Cross-city Analysis of Location-based Sentiment in User-generated Text

Christopher Stelzmüller
christopher.stelzmueller@gmail.com
Johannes Kepler University
Linz, Austria

Sebastian Tanzer
s.tanzer98@gmail.com
Johannes Kepler University
Linz, Austria

Markus Schedl
markus.schedl@jku.at
Johannes Kepler University and
Linz Institute of Technology
Linz, Austria

ABSTRACT

Geolocated user-generated content is a promising source of data reflecting how citizens live and feel. Information extracted from this source is being increasingly used for urban planning and policy evaluation purposes. While a lot of existing research focuses on the relationship between locations and sentiment in social media postings, we aim to uncover relations between location and sentiment that are consistent over cities around the world. In this paper, we therefore analyze the relationship between multiple categories of points of interest (POIs) in the *OpenStreetMap* dataset and the sentiment of English microblogging messages sent nearby using a three-stage processing pipeline: (1) extract sentiment scores from geolocated microblogs posted on Twitter, (2) spatial aggregation of sentiment in cities and POIs, (3) analyze relationships in aggregated sentiment. We identify differences in Twitter users' sentiments within cities based on POIs, and we investigate the temporal dynamics of these sentiments and compare our findings between major cities in multiple countries.

KEYWORDS

sentiment analysis, microblogs, spatial analysis

ACM Reference Format:
Christopher Stelzmüller, Sebastian Tanzer, and Markus Schedl. 2021. Cross-city Analysis of Location-based Sentiment in User-generated Text. In *Companion Proceedings of the Web Conference 2021 (WWW '21 Companion), April 19–23, 2021, Ljubljana, Slovenia.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3442442.3451889

1 INTRODUCTION

User-generated content on social media platforms is gradually replacing traditional sources to gather information in the context of urban and transportation planning [11, 14, 23]. By utilizing extracted information about citizens' sentiment, planners can more easily gain an understanding of the public's usage of city infrastructure and demand [23]. As opposed to traditional methods of data gathering (e.g. via surveys), analyzing social media data allows for faster feedback from the population, provides more fine-grained information and is more cost-efficient. Social media data furthermore allow combining data about local citizens' and tourists' use of infrastructure.

While most existing work quantifies sentiment in general [6] or in one city's area, e.g. [11, 14, 23], we compare sentiment over multiple cities. By combining methods from geospatial analysis with sentiment analysis we aim to first identify the relationship between spatial regions (e.g. a park, a football stadium, or a traffic hub) and general sentiment expressed in nearby social media postings. To this end, we analyze user-generated postings on the social media network Twitter.

We strive to answer the following research questions:

- RQ1: Are there differences in Twitter users' sentiments expressed in their tweets between cities?
- RQ2: Are there differences in Twitter users' sentiments expressed in their tweets within cities depending on the geographic location (only considering POI)?
- RQ3: Are there differences in Twitter user's sentiments expressed in tweets at special time intervals? Are these differences more distinct in certain regions or near POIs?

We base our analysis on a world-wide dataset of geolocated tweets that were posted between September 1st and 17th, 2012, acquired via the Twitter streaming API. The Twitter streaming API[1] provides access to a random sample of tweets that make up approximately one per cent of the total daily tweets. By providing the coordinates for a bounding box, the location of tweets to sample from can be restricted. We further use a pre-defined set of metropolitan areas, focusing on urban and suburban areas, since our analysis depends on the availability of sufficiently many geolocated tweets, and the number of geolocated tweets generally is higher in areas with high population density [1]. Since we only analyze the sentiment of tweets in the English language, we select only cities that provide sufficient coverage of English tweets. Furthermore, we do not include cities in which the use of Twitter is either officially blocked or only possible through the use of VPNs. Most notably, this constraint includes highly populated areas in Iran and mainland China [2]. For defining urban areas, we work with the *World Urban Areas* dataset by Kelso et al., which combines urban area shapes with population estimates from the LandScan population database.[2] We research both urban areas within one country in order to find relations between location, time, and expressed sentiment, as well as world-wide urban areas in order to verify if potentially found correlations are stable over distinct geographical

WWW '21 Companion, April 19–23, 2021, Ljubljana, Slovenia

ACM ISBN 978-1-4503-8313-4/21/04.
https://doi.org/10.1145/3442442.3451889

[1] https://developer.twitter.com/en/docs/twitter-api/v1/tweets/filter-realtime/api-reference, visited on 2021-02-27
[2] https://purl.stanford.edu/yk247bg4748, visited on 2021-02-27

areas. As opposed to existing research in this field, we especially focus on the comparison of findings across metropolitan regions in multiple countries and continents.

The remainder of this paper is organized as follows: In Section 2, we discuss related work. Section 3 details our methodology, which is followed by the presentation and discussion of results in Section 4. Limitations of our work are discussed in Section 5. Section 6 rounds off this work and provides pointers for future research.

2 BACKGROUND AND RELATED WORK

2.1 Sentiment Analysis in Microblogging

Sentiment analysis has been used in many different areas. One of them being politics where tweets of people are analyzed to get the public's opinion on forthcoming elections and using this as a base to predict their outcome [26]. Companies use sentiment analysis in marketing to help them determine the success of a product launch or a marketing campaign or what product version or services are popular [7]. Similarly, Makrehchi et al. use opinions towards public companies and certain products in order to offer investors additional insights on possible stock price changes [15].

Different approaches can be used in order to acquire opinions, feelings, or personal expressions from textual data shared via microblogging services. Most related research employs either unsupervised or supervised learning mechanisms for extracting sentiment from textual data. For example, Yaqub et al. [26] use subjective and polarity analysis in order to analyze tweet sentiment regarding election campaigns. Subjective analysis is used to calculate a subjective score for each candidate, by evaluating whether the shared content of Twitter users during the election was fact-based or opinion-based. This evaluation is done by looking for the presence of certain adjectives, adverbs, groups of verbs and nouns which are subsequently used as indicators of subjective opinion. Polarity analysis is used to obtain an average sentiment for both candidates, by determining the emotional attitude of a text's author with regard to the topic under discussion. This is done by giving each relevant word of a sentence a score that is defined in a score lexicon and afterwards computing the total score of the text, by adding up all individual scores. For both analyses predefined datasets from Python TextBlob,[3] a library for text processing, are used.

Makrehchi et al. [15] propose a method based on supervised learning. In general, supervised learning yields better results, but is often not feasible because of the lack of labeled data needed, especially regarding Twitter posts. The proposed method tries to eliminate this problem by automatically extracting labels based on significant stock market events. The labeled data then is used to build a sentiment classifier, which again offers the base to built a model for identifying factors that affect stock prices.

Modern approaches often adopt supervised learning using pretrained Transformer models based on BERT. Kim et al. [10] use this approach to define public sentiment toward solar energy by performing a classification task on a dataset of tweets specific to this topic, using Robustly optimized Bidirectional Encoder Representations from Transformers (RoBERTa).

[3]https://textblob.readthedocs.io/en/dev/, visited on 2021-02-27

Müller et al. [18] release COVID-Twitter-BERT, a BERT model trained on a specific target domain. The model is trained on a preprocessed and cleaned dataset of 22.5M tweets. The preprocessing and cleaning steps include replacement of user names and links by neutral tokens, the replacement of unicode emoticons with textual ASCII representations, as well as the removal of retweets, duplicates and close duplicates.

2.2 Sentiment Analysis of Points of Interest

With an increasing amount of user-generated and georeferenced content being created on social media and interactive web platforms, more and more research is put into combining and aggregating extracted information from content with geospatial information. Especially the combination of information sentiment and its location has been researched in multiple papers [4, 8].

Generally it is worth noting that the existence of a vast number of different social media and interactive web platforms requires numerous different analysis and sentiment extraction methods. For instance, Hauthal et al. [8] focus on analyzing text-based metadata of Flickr and Panoramio photos. They argue that photo metadata commonly does not contain emotional descriptions, but rather more objective information concerning the content of the photo. In order to measure metadata sentiment, they therefore employ an algorithm based on the *affective connotation* of individual words leading to a *valence* and *arousal* score for each metadata. Furthermore, they consider special grammatical constellations that influence the sentiment of a data point (e.g. negations). The authors focus on the city of Dresden and propose implementing an emotional travel guide that could, for instance, suggest travel destinations based on the user's current feelings. Arguably, this constitutes a novel approach for exploring and traveling through an unknown city.

Bertrand et al. [4] focus on analyzing geo-tagged tweets. Since the majority of tweets are not written in English, they create their own language-independent classification function instead of using a dictionary-based approach. To this end, they use a set of tweets including emoticons as training data, from which they infer the general sentiment of a tweet's remaining content. It is worth noting that in 2012, when the paper was written, emojis were not as widely used on the internet as they are at the time of writing in 2020. A comparison between the sentiment map and the city's POIs shows that certain POIs (e.g. parks) are mostly connected to strong positive sentiment, whereas jails and hospitals are usually connected to a negative sentiment. However, this connection between sentiment and POI type cannot be generalized, as demonstrated by the fact that areas surrounding cemeteries can be both connected to good or bad sentiment.

The location of a tweet can also be combined with its content in order to improve the accuracy of sentiment analysis methods [13]. In their paper, Lim et al. find that concatenating the tweet text with types of nearby POIs often helps increase prediction accuracy, but has significant implications on model training time. Possible reasons for this increase in accuracy include the spatial correlation with (especially positive) sentiments [9] and the fact that certain areas of countries often associate a topic with a specific sentiment [24]. While geolocation information can improve the accuracy of sentiment analysis and is essential for location-based analyses, only around 0.85% of all, and 3.1% of non-retweet posts include geotags [25].

2.3 Sentiment in Spatial Analysis

Since geolocated microblogs (and especially tweets) provide a unique mix of textual content and geospatial information, they are used increasingly often as a data source for tasks involving country- and city-wide spatial analysis considering topics like urban planning [22]. In New York City, Twitter data has been used to analyze the influence of parks and green spaces on visitors' sentiments [21]. For instance, Plunz et al.'s [21] analysis focuses only on geotagged English tweets that were sent from a predefined geographical location within New York's bounding box. Furthermore, they only consider tweets of users that have at least once tweeted from within a park or green space. They find that the average sentiment of tweets sent from within parks and outside is generally different; though depending on the type of district, the sentiment change can be both positive or negative. They also point out that the number of Twitter users only represents a fraction of total park visitors that is not necessarily representative. A tweet's sentiment furthermore is often influenced by events happening at the user's location or in their personal life and, especially in urban, densely built-up environments, smartphone GPS accuracy often is lacking. Despite these limitations, insights gained from sentiment analysis of tweets are potentially helpful in augmenting existing urban design and planning measures as well as substituting currently manually conducted surveys. While Plunz et al. [21] find a difference in sentiment between the general population and park visitors, our goal is to validate if that difference is consistent between major metropolitan areas in multiple countries.

Another interesting application of Twitter sentiment analysis is proposed by Li et al. [12]. They analyze tweet sentiments in order to give smart cities insights about their citizens' current status and feelings. Similar to Bertrand et al. [4], they leverage user-generated tweets including emoticons and emojis, split into positive and negative affections, as a training set for a sentiment classification model. For recognizing objective tweets, they use regular text passages from newspaper articles. Even though the newspaper articles constrain the model to the articles' language, the model can easily be trained on articles in other languages without manual labeling. Before they start the classification task, they first employ a set of pre-processing steps in order to standardize the Twitter data. In particular, this concerns removing *noise* (e.g. URLs, user mentions) from tweets, correcting spelling mistakes, tokenizing the texts, and specifically handling negations in the tweet content. They also propose a visualisation system that can be used by city governments to recognize spatial and temporal changes in their citizens sentiment.

User-generated geo-tagged data from the Chinese microblogging platform Sina Weibo has been used to classify waterfront areas in Wuhan according to users' sentiment [14]. In addition to sentiment data and location of posts, Ma et al. [14] also consider the post author's gender in the analysis and therefore are able to recognize significant differences in the sentiment connected with certain lakes between male and female users. This sentiment difference could help city planners recognize the need for infrastructural changes before they are even communicated by their citizens. Similarly, differences in mean sentiment and variance can also be found connected with other kinds of POIs like cultural locations and restaurants. The authors point out that social media and microblogging data can be helpful in gathering immediate feedback of public opinion and can be utilized for understanding, recognizing, and solving problems in the areas of urban planning, traffic analysis, tourism and public health. Unlike Ma et al. [14] who analyze data from Sina Weibo, we leverage Twitter data and focus on comparing similar locations in different cities and countries, whereas they analyze the difference in expressed sentiment between male and female social media users.

Mitchell et al. [17] analyze geotagged tweets gathered from 373 urban areas within the US to examine sentiment (happiness) across states and urban areas. To measure sentiment they use the Language Assessment provided by Mechanical Turk. They further map areas of high and low happiness and score individual states and cities to obtain average word happiness. While Mitchell et al. [17] focus on the analysis and comparison of all cities within the US, we analyze and compare tweets form selected cities across the whole world and further analyze the sentiment of certain types of POIs.

3 METHODS

3.1 Selecting Metropolitan Regions

Since the Twitter API only returns a small sample of tweets and generally only 0.85% of tweets also contain location information [25], we need to select regions with a sufficient number of tweets. Because the density of tweets scales linearly with population density [1], we primarily select densely populated metropolitan areas. Since we focus on analyzing sentiment of social media posts written in English, the local language of a metropolitan region also has to be considered. In order to find regions for our task, we start by extracting all tweets from the dataset that are located within one of the 28 most populous metropolitan areas, according to the *World Urban Areas* dataset.[4] In order to cover all continents, we also extract tweets sent from within Sydney's and Melbourne's metropolitan areas. Next, we utilize an n-gram-based text classification approach [5] in order to estimate the language of each tweet. We finally select the top 15 metropolitan areas in terms of their number of English tweets (see Table 1).

We can observe that the metropolitan areas with most English tweets are not exclusively located in countries whose official language is English, but distributed among various countries on all continents. Furthermore, a high percentage of English tweets does not necessarily mean that the city's native language is English, as can be seen by the high percentage of English tweets in Manila and Lagos. Cities that have diverse native languages, but utilize English as a lingua franca also tend to have a comparably high percentage of English tweets. This is especially noticeable in Johannesburg, where less than 20% of citizens are native English speakers,[5] but about 77% of tweets are written in English.

[4] https://purl.stanford.edu/yk247bg4748, visited on 2021-02-27

[5] According to the Census data from 2011: http://www.statssa.gov.za/?page_id=993&id=city-of-johannesburg-municipality, visited on 2021-02-27

Metropolitan Area	# EN Tweets	% EN Tweets	Population
New York	303.3 K	80.9%	9.4 M
Los Angeles	225.4 K	76.9%	5.0 M
London	198.2 K	78.2%	7.7 M
Manila	145.6 K	55.5%	3.1 M
Chicago	138.7 K	79.5%	3.7 M
Jakarta	87.7 K	12.3%	9.7 M
Tokyo	27.4 K	15.0%	13.8 M
Paris	27.3 K	11.8%	7.5 M
Mexico City	24.0 K	13.3%	10.8 M
Sydney	21.0 K	81.6%	2.7 M
Lagos	20.7 K	66.2%	7.1 M
Melbourne	19.6 K	80.8%	1.9 M
Istanbul	17.7 K	5.9%	9.9 M
Sao Paolo	16.5 K	7.5%	12.5 M
Johannesburg	12.0 K	77.4%	3.9 M

Table 1: Distribution of number of tweets and population in selected metropolitan regions, sorted by number of English tweets per week.

Table 1 shows the distribution of tweets in the investigated regions. It is worth noting that the selected regions are distributed among multiple continents: We focus on New York, Los Angeles, Chicago and Mexico City in North America, Sao Paolo in South America, London and Paris in Europe, Istanbul in both Europe and Asia, Manila, Jakarta and Tokyo in Asia, as well as Lagos and Johannesburg in Africa.

3.2 Pre-processing and Filtering Tweets

The investigated dataset consists of geolocated tweets posted between September 1st and 17th, 2012. Each tweet includes detailed information about its author, intended recipient, source, location, content, and creation timestamp. For our study, we extract information about the timestamp, location, and text. While the creation timestamp and text of a tweet are straightforward to process, the location is more complex, since each tweet contains up to three different pieces of spatial information. In every geotagged tweet, the *place* object contains information about a named location in the *Twitter Places*[6] database. The place object consists of a bounding box of the specified place in the GeoJson format[7] and additional information about the selected place (e.g. country, name, type). If a user decides to use an exact location, the tweet also contains a *geo* and *coordinates* object.[8] Both objects encode the exact location information, but while the *coordinates* object contains the location in GeoJson format, the *geo* object utilizes a similar format, but changes the coordinate order. In order to be able to connect tweet locations with POIs, we exclusively work with tweets that contain the *coordinates* object, and therefore contain exact location information.

[6]https://developer.twitter.com/en/docs/twitter-api/v1/geo/place-information/overview, visited on 2021-02-27
[7]https://tools.ietf.org/html/rfc7946, visited on 2021-02-27
[8]https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/geo, visited on 2021-02-27

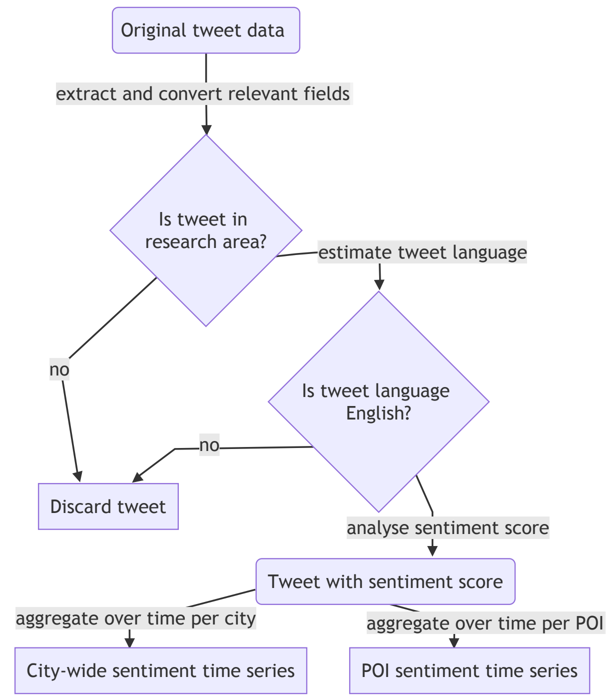

Figure 1: Pre-processing steps for analyzing tweet sentiment.

Figure 1 outlines the general steps needed in pre-processing of tweets. The result of the pre-processing process is not dependent on the order in which steps are performed: First determining if a tweet's location is within an investigated area and then validating its language leads to the same results as performing the steps in reverse order. However, since the duration of pre-processing steps is different, it is advisable to perform less time-consuming tasks first. In regard to the proposed pre-processing steps, this means that the optimal order of steps is to first check if a tweet's location is within one of the researched areas, then validate if the tweet's language is English and last calculate the content's sentiment score.

The first pre-processing step in our pipeline therefore is to check if a tweet's location is within one of our selected metropolitan regions. In addition to validating the tweet's location, we also add the information which city this tweet belongs to at this point, in order to speed up further pre-processing and analysis steps needed afterwards. To validate a tweet's language, we use *langdetect*,[9] a Python language detection library ported from Nakatani Shuyo's Java language detection library [19], which utilizes n-gram-based text classification in order to calculate the probabilities of the text being in a language from features of spelling of a text using naive Bayesian filter. After having filtered all non-English tweets, we perform several additional pre-processing steps as suggested in [3]: converting text to lower case, removing repeated characters, links, user and place names, and special characters such as € and #.

[9]https://pypi.org/project/langdetect, visited on 2021-02-27

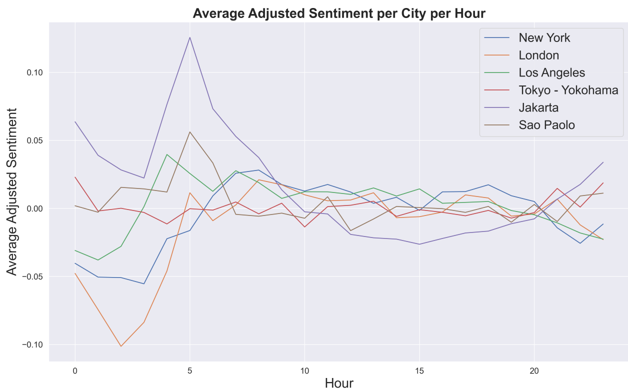

Figure 2: Average adjusted sentiment for selected cities over hours of day (adjusted by time zones).

3.3 Sentiment Analysis

For analyzing tweet sentiment, we adopt the approach proposed by Barbieri et al. [3] in the TweetEval benchmark, and use a pre-trained model for sentiment analysis in English tweets. This model is based on BERT and is available online.[10] It returns the sentiment $s(t_i)$ of a tweet t_i as three different values: the probability that the analyzed tweet is negative $p(t_i = neg)$, neutral $p(t_i = neu)$, or positive $p(t_i = pos)$. The individual probabilities $p(t_i = neg)$, $p(t_i = neu)$, and $p(t_i = pos)$ always sum up to 1 for each tweet t_i. For our further analysis we combine these three values into one sentiment score. The sentiment score of a tweet $s(t_i)$ is calculated according to Equation 1.

$$s(t_i) = \begin{cases} 0, & \text{if } p(t_i = neu) > p(t_i = pos) \wedge \\ & \quad p(t_i = neu) > p(t_i = neg) \\ p(t_i = pos) - p(t_i = neg), & \text{otherwise} \end{cases} \tag{1}$$

For the comparison of sentiment between different cities, two further problems need to be addressed. The first one is different writing styles, as well as general sentiment, based on the geographical location of each city. We will go more into the details of solving this problem in Section 3.4 by introducing an *adjusted sentiment score* that is defined in Equation 2. Second, the creation time of a tweet is always returned as universal coordinated time (UTC). Since we want to compare tweets at similar local times, we need to correct the creation date by the local time offset defined by each tweet's location.

In Figure 2 we can see a timeline of hourly mean sentiment after correcting for time zone. Notably, all shown cities have both a negative and a positive peak between midnight and 8 am in the morning.

3.4 Spatial Analysis

Comparison of sentiment between cities. To compare the sentiment between cities, we first calculate the average sentiment score

[10]https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment, visited on 2021-02-27

Figure 3: Average and adjusted average sentiment over time per city for selected English-speaking cities, respectively, in upper and lower plot.

per city per day. We then plot the resulting time series to get a first impression of the different sentiment scores (see upper plot in Figure 3). However, as mentioned in Section 3.3 this only gives us a general overview of each city's sentiment per day. To be able to better compare sentiment changes between cities, we adjust the sentiment score of each tweet t_i with the "background sentiment" of the city it was posted from, according to Equation 2, where T_{city} denotes the set of tweets posted in t_i's city, and $s_{adj}(t_i)$ is the adjusted sentiment score of t_i.

$$s_{adj}(t_i) = s(t_i) - \frac{1}{|T_{city}|} \sum_{t_j \in T_{city}} s(t_j) \tag{2}$$

Figure 3 shows a comparison between mean sentiment and adjusted mean sentiment in English-speaking metropolitan regions. In the upper plot, the difference in absolute sentiment between cities is visible, whereas the lower plot makes it easier to compare changes in sentiment of a given city over the displayed time frame.

Combination with POI data. In order to address our second research question, we use POI data from OpenStreetMap (OSM). We focus on analyzing sentiment in and nearby parks, cemeteries, and public transportation hubs. The first step in our POI analysis is to retrieve available POI data from OSM in the form of a planet file [20], which contains all current information mapped in the OSM project. Generally, all data within an OSM file is represented as

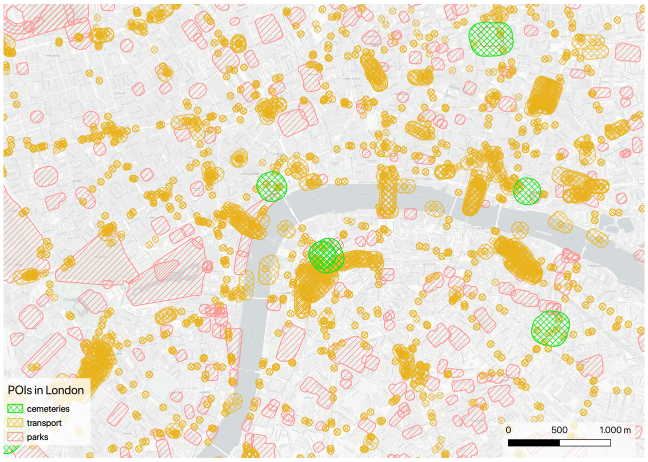

Figure 4: Distribution of POIs in the center of London.

nodes (points), ways (lines or polygons) and relations.[11] Each of the three elements can contain tags, such as *leisure=park*, *name=Hyde Park*, or *dog=no*.

The selected POIs are represented in different ways in OSM: while most public transport stops are, for instance, stored as points, bigger train stations are often represented by their building outline. In order to extract the data needed for our study, we utilize the *osmfilter* tool, which allows selecting only elements with specific tags, thus we are able to extract information about all POIs in our areas under investigation.[12]

In order to find out which tweets are within a POI's area of influence, we first convert all POIs to polygons by applying a buffer to the underlying geometry. Depending on the type of POI, we define different reaches of spatial influence that also take into account the POI's type in OSM (e.g., a point or a polygon). All parks that are stored as polygons and public transport stops stored as points receive an influence area of 50 meters around the POI itself, bigger public transport stops of 100 meters, and cemeteries of 200 meters. For further processing, we assign each tweet to the corresponding POI, considering these influence areas. Figure 4 illustrates the distribution of POIs and their influence areas within the inner city of London. As can be seen in the figure, the influence areas often overlap, hence one tweet can be within multiple POI influence areas at once.

4 RESULTS

Comparison of sentiment between metropolitan regions (RQ1). To answer RQ1, we calculate the kernel density distribution of tweet sentiment scores within a city, excluding tweets with a sentiment score of 0. We use the raw sentiment value in order to accurately depict areas of very high and very low sentiment scores. The results of this calculation are shown along with each city's mean sentiment score as a vertical line in Figure 5. Cities in English-speaking

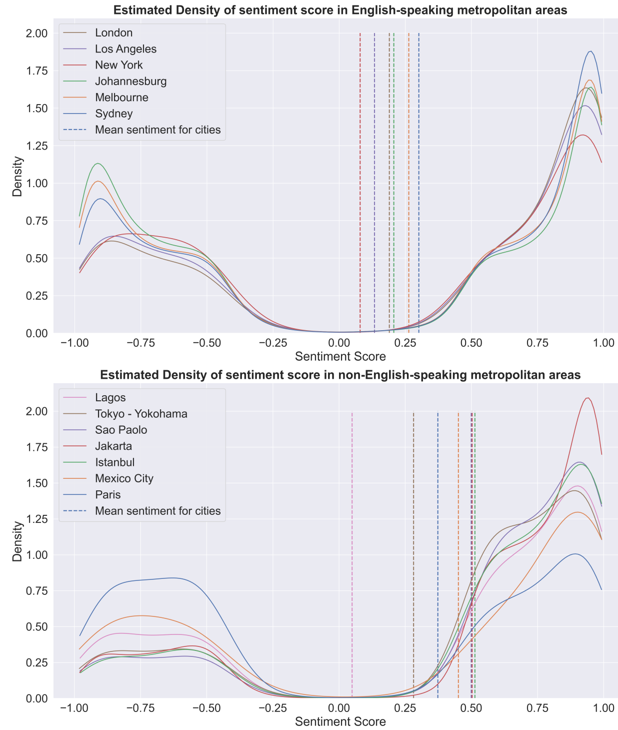

Figure 5: Comparison of sentiment distribution in English-speaking (upper plot) and non-English-speaking (lower plot) metropolitan areas.

countries (i.e., London, Los Angeles, New York, Melbourne, and Sydney) have a high density of tweets with very high (around 0.9) and very low (around -0.9) sentiment scores. Conversely, cities in non-English-speaking countries have a flatter peak of sentiment concentration especially in the sentiment range from -0.8 to -0.5. Some non-English-speaking metropolitan areas (e.g., Tokyo–Yokohama and Mexico City) also have a significantly higher sentiment density in the range of 0.5 to 0.6.

It is worth noting that restricting the investigation to English tweets can introduce further limitations. For instance, we assume that in the metropolitan area of Tokyo–Yokohama a significant number of English tweets is authored by tourists rather than locals, which influences the mean sentiment in comparison to English-speaking metropolitan regions.

Comparison of sentiment at different locations within metropolitan regions (RQ2). By aggregating sentiment nearby different kinds of POIs, we find that for most cities, there is a difference in sentiment nearby the investigated POIs and the city's mean. However, as Figure 6 shows, these differences are not consistent over different metropolitan regions. For instance, while tweets in New York tend to be more positive when created nearby parks and transportation hubs, this is not noticeable in any other city, except for London, where a similar trend is observable for parks. It is also worth noting that Figure 6 does not show the absolute sentiment, but displays the adjusted sentiment in order to allow inter-city comparison. While the displayed metropolitan regions of New York and London show

[11] https://wiki.openstreetmap.org/wiki/Elements, visited on 2021-02-27

[12] We use the tags *leisure=park* and *amenity=park* for selecting parks, *landuse=cemetery*, *amenity=grave_yard*, *cemetery=** and *amenity=cemetery* for selecting cemeteries, as well as *building=transportation*, *building=train_station*, *public_transport=station*, *aeroway=terminal*, *shelter_type=public_transport*, *public_transport=stop_area*, *public_transport=platform* and *amenity=bus_station* for selecting public transport POIs.

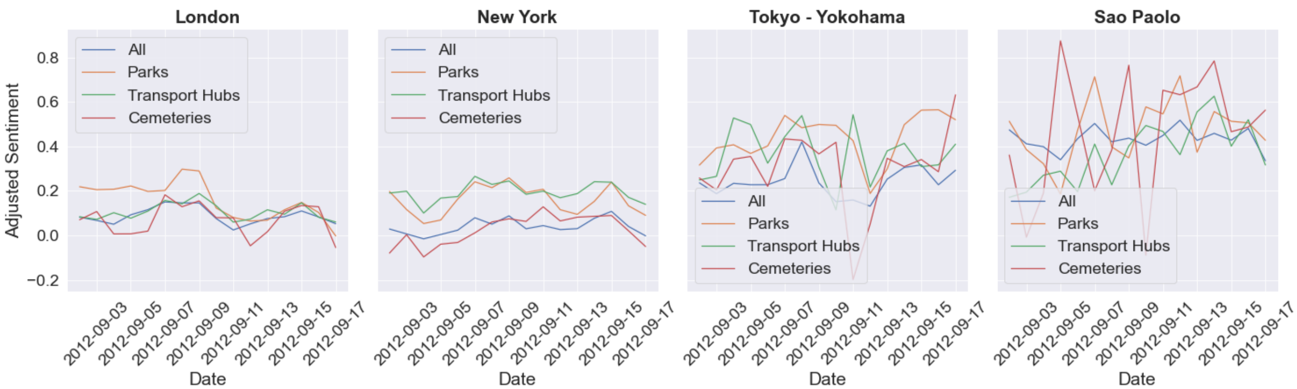

Figure 6: Comparison of sentiment nearby POIs and general sentiment for different metropolitan areas.

similar time series for general sentiment and POI-based sentiment, Sao Paolo and Tokyo–Yokohama contain big outliers in the POI-based time series. A potential cause of this is the generally lower number of tweets in these regions that decreases further when only tweets nearby POIs are taken into account.

Comparison of sentiment change in time series (RQ3). By comparing the adjusted and time-corrected sentiment series in different cities (see Figure 2), we notice that the sentiment generally has two trends. First and foremost, the average sentiment in all Western cities has both a negative followed by a positive peak in the morning (between 5 am and 10 am) and then decreases during the day. These peaks can be explained by lower number of neutral tweets during this time. Another interesting insight is that English-speaking cities generally have higher sentiment scores during weekends than during working days. For English-speaking cities, this relation is also temporally stable. However, this is not the case for non-English-speaking metropolitan areas, where a distinction between workdays and weekends is not clearly noticeable, as can be seen in Figure 7.

A possible cause of the difference in sentiment over weekdays between English- and non-English-speaking cities could be the disparity between sentiment of tourists and natives, or different cultural backgrounds between English-speaking (mostly Western) and non-English-speaking countries. Analyzing these difference in different areas within cities is a particularly interesting topic for further research.

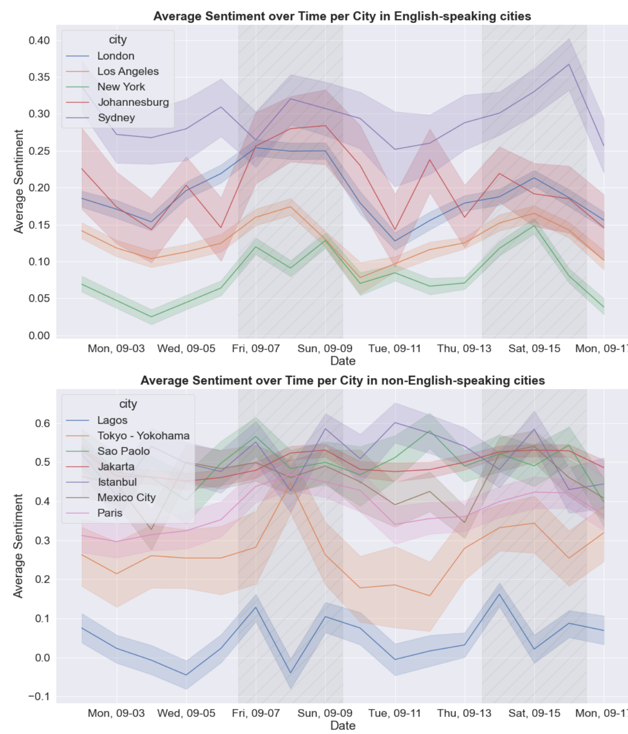

Figure 7: Comparison of average sentiment per day for selected English-speaking (upper plot) and non-English-speaking (lower plot) cities; the semi-transparent colored area around a city's sentiment line illustrate its 95% confidence interval, darker gray areas illustrate weekends.

5 LIMITATIONS

While our study yielded interesting insights, there are a few limitations that should be noted. The most prominent one is that microblogging users do not necessarily represent the general population of cities, as already outlined by Plunz et al. [21]. Social media data, therefore, cannot totally substitute regular surveys and measures for gathering public opinion, but can be a powerful complementary tool. Second, we are aware that the amount of data we were able to process is limited. We are currently analyzing a larger and up-to-date dataset, spanning the years 2012–2020.

Furthermore, we assume that in non-English-speaking cities many English tweets are authored by tourists or foreigners whose daily activities are often different from locals'. This is likely to introduce distortions to our results. In addition, the filtering with POI locations also leads to further decreasing the number of existing data points.

Another limiting factor for POI-based analysis is the existence and quality of OSM data about local infrastructure. While OSM data quality is generally acceptable in most urban areas, coverage in emerging economies tends to be lackluster. This has been especially conspicuous during our POI analysis of the metropolitan area of Lagos, where the total number of existing POIs in OSM was extremely small. The geospatial analysis of tweet location is furthermore constrained by the fact that GPS sensors in smartphones can be less accurate than expected. Merry et al. [16] find that GPS sensors can be up to 100 meters off in urban settings. To take this inaccuracy into account, we defined our POI influence areas appropriately big. However, this increases the risk of misclassifying tweets that should not belong to a POI. This limitation is particularly problematic in areas with a dense population and a good public transportation network, since our POI analysis method would lead to many tweets sent from home being classified as POI-related.

6 CONCLUSION AND FUTURE WORK

In this paper, we studied differences in sentiment reflected by Twitter users' postings, at different levels: between major global cities, within places of interests, and during different days of the week. We show that sentiment in different major cities around the world differs between English-speaking and non-English-speaking cities, and we show that sentiment changes of the former follow a similar pattern. Furthermore, we inspect sentiment change for specific types of POIs (e.g., parks and cemeteries). Although there is no clear general relationship between sentiment and POIs over all investigated cities, we were able to observe that for all English-speaking cities the sentiment measured nearby parks is generally more positive than nearby cemeteries. We also find that daily and hourly mean sentiment follow a similar pattern for all analyzed English-speaking metropolitan areas.

There are several avenues for further research. The inspection of English speaking and non-English-speaking cities revealed certain differences in sentiment scores that can be further looked into by using a language-independent classifier for the sentiment analysis. This offers the opportunity to analyze a greater number of tweets from non-English-speaking countries, as well as to look further into the subject of distinguishing between citizens and tourists to understand how tourists might influence sentiment scores for non-English-speaking cities. Another further research topic is the analysis of sentiment changes during the COVID-19 pandemic in different cities. Social media data could represent a potential new viewpoint into the change in mobility patterns during an ongoing pandemic. Furthermore, considering user metadata (e.g., age and gender) could yield interesting insights into possible different sentiment levels for certain POIs based on such user characteristics. Eventually, adopting a clustering method that encompasses both the spatial and sentiment component of a tweet could help identify areas of especially positive or negative sentiment within cities, irrespective of a particular POI category.

REFERENCES

[1] Rudy Arthur and Hywel T. P. Williams. 2019. Scaling Laws in Geo-Located Twitter Data. *PLoS ONE* 14, 7 (July 2019). https://doi.org/10.1371/journal.pone.0218454

[2] David Bamman, Brendan O'Connor, and Noah Smith. 2012. Censorship and Deletion Practices in Chinese Social Media. *First Monday* (March 2012). https://doi.org/10.5210/fm.v17i3.3943

[3] Francesco Barbieri, Jose Camacho-Collados, Leonardo Neves, and Luis Espinosa-Anke. 2020. TweetEval: Unified Benchmark and Comparative Evaluation for Tweet Classification. *arXiv:2010.12421 [cs]* (Oct. 2020). arXiv:2010.12421 [cs]

[4] Karla Z. Bertrand, Maya Bialik, Kawandeep Virdee, Andreas Gros, and Yaneer Bar-Yam. 2013. Sentiment in New York City: A High Resolution Spatial and Temporal View. *arXiv:1308.5010 [physics]* (Aug. 2013). arXiv:1308.5010 [physics]

[5] William Cavnar and John M. Trenkle. 1994. N-Gram-Based Text Categorization. In *In Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval*. 161–175.

[6] Peter Sheridan Dodds, Kameron Decker Harris, Isabel M. Kloumann, Catherine A. Bliss, and Christopher M. Danforth. 2011. Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter. *PLOS ONE* 6, 12 (12 2011), 1–1. https://doi.org/10.1371/journal.pone.0026752

[7] Xing Fang and Justin Zhan. 2015. Sentiment analysis using product review data. *Journal of Big Data* 2, 1 (2015), 5.

[8] Eva Hauthal and Dirk Burghardt. 2016. Mapping Space-Related Emotions out of User-Generated Photo Metadata Considering Grammatical Issues. *The Cartographic Journal* 53, 1 (Jan. 2016), 78–90. https://doi.org/10.1179/1743277414Y.0000000094

[9] T. Hu, B. She, L. Duan, H. Yue, and J. Clunis. 2020. A Systematic Spatial and Temporal Sentiment Analysis on Geo-Tweets. *IEEE Access* 8 (2020), 8658–8667. https://doi.org/10.1109/ACCESS.2019.2961100

[10] Serena Y Kim, Koushik Ganesan, Princess Dickens, and Soumya Panda. 2020. Public Sentiment Toward Solar Energy: Opinion Mining of Twitter Using a Transformer-Based Language Model. *arXiv preprint arXiv:2007.13306* (2020).

[11] Anna Kovacs-Györi, Alina Ristea, Clemens Rudolf Havas, Bernd Resch, Pablo Cabrera-Barona, and Mihaela-Alina Ristea. 2018. #London2012: Towards Citizen-Contributed Urban Planning Through Sentiment Analysis of Twitter Data. *Urban Planning* 3, 1 (2018), 75–99. https://doi.org/10.17645/up.v3i1.1287

[12] M. Li, E. Ch'ng, A. Chong, and S. See. 2016. The New Eye of Smart City: Novel Citizen Sentiment Analysis in Twitter. In *2016 International Conference on Audio, Language and Image Processing (ICALIP)*. 557–562. https://doi.org/10.1109/ICALIP.2016.7846617

[13] W. L. Lim, C. C. Ho, and C.-Y. Ting. 2020. Sentiment Analysis by Fusing Text and Location Features of Geo-Tagged Tweets. *IEEE Access* 8 (2020), 181014–181027. https://doi.org/10.1109/ACCESS.2020.3027845

[14] Yue Ma, Changlong Ling, and Jing Wu. 2020. Exploring the Spatial Distribution Characteristics of Emotions of Weibo Users in Wuhan Waterfront Based on Gender Differences Using Social Media Texts. *ISPRS International Journal of Geo-Information* 9, 8 (Aug. 2020), 465. https://doi.org/10.3390/ijgi9080465

[15] M. Makrehchi, S. Shah, and W. Liao. 2013. Stock Prediction Using Event-Based Sentiment Analysis. In *2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, Vol. 1. 337–342. https://doi.org/10.1109/WI-IAT.2013.48

[16] Krista Merry and Pete Bettinger. 2019. Smartphone GPS Accuracy Study in an Urban Environment. *PLoS ONE* 14, 7 (July 2019). https://doi.org/10.1371/journal.pone.0219890

[17] Lewis Mitchell, Morgan R. Frank, Kameron Decker Harris, Peter Sheridan Dodds, and Christopher M. Danforth. 2013. The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place. *PLoS ONE* 8, 5 (May 2013), e64417. https://doi.org/10.1371/journal.pone.0064417

[18] Martin Müller, Marcel Salathé, and Per E Kummervold. 2020. Covid-twitter-bert: A natural language processing model to analyse covid-19 content on twitter. *arXiv preprint arXiv:2005.07503* (2020).

[19] Shuyo Nakatani. 2010. Language Detection Library for Java. https://github.com/shuyo/language-detection

[20] OpenStreetMap contributors. 2020. Planet dump retrieved from https://planet.osm.org . https://www.openstreetmap.org.

[21] Richard A. Plunz, Yijia Zhou, Maria Isabel Carrasco Vintimilla, Kathleen Mckeown, Tao Yu, Laura Uguccioni, and Maria Paola Sutto. 2019. Twitter Sentiment in New York City Parks as Measure of Well-Being. *Landscape and Urban Planning* 189 (Sept. 2019), 235–246. https://doi.org/10.1016/j.landurbplan.2019.04.024

[22] Helen Victoria Roberts. 2017. Using Twitter Data in Urban Green Space Research: A Case Study and Critical Evaluation. *Applied Geography* 81 (April 2017), 13–20. https://doi.org/10.1016/j.apgeog.2017.02.008

[23] Jisoo Sim, Patrick Miller, and Samarth Swarup. 2020. Tweeting the High Line Life: A Social Media Lens on Urban Green Spaces. *Sustainability* 12, 21 (Oct. 2020), 8895. https://doi.org/10.3390/su12218895

[24] Prabhsimran Singh, Ravinder Singh Sawhney, and Karanjeet Singh Kahlon. 2018. Sentiment Analysis of Demonetization of 500 & 1000 Rupee Banknotes by Indian Government. *ICT Express* 4, 3 (Sept. 2018), 124–129. https://doi.org/10.1016/j.icte.2017.03.001

[25] Luke Sloan and Jeffrey Morgan. 2015. Who Tweets with Their Location? Understanding the Relationship between Demographic Characteristics and the Use of Geoservices and Geotagging on Twitter. *PLoS ONE* 10, 11 (Nov. 2015). https://doi.org/10.1371/journal.pone.0142209

[26] Ussama Yaqub, Nitesh Sharma, Rachit Pabreja, Soon Ae Chun, Vijayalakshmi Atluri, and Jaideep Vaidya. 2020. Location-based Sentiment Analyses and Visualization of Twitter Election Data. *Digital Government: Research and Practice* 1, 2 (2020), 1–19.