Article

Public Transport Tweets in London, Madrid and Prague in the COVID-19 Period—Temporal and Spatial Differences in Activity Topics

Martin Zajac [1], Jiří Horák [1,*], Joaquín Osorio-Arjona [2], Pavel Kukuliač [1] and James Haworth [3]

1   Department of Geoinformatics, Faculty of Mining and Geology, VSB-Technical University of Ostrava, 70800 Ostrava, Czech Republic
2   Department of Population, Centro de Ciencias Humanas y Sociales CSIC, 28037 Madrid, Spain
3   Department of Civil, Environmental and Geomatic Engineering, University College London, London WC1E 6BT, UK
*   Correspondence: jiri.horak@vsb.cz

Abstract: Public transport requires constant feedback to improve and satisfy daily users. Twitter offers monitoring of user messages, discussion and emoticons addressed to official transport provider accounts. This information can be particularly useful in delicate situations such as management of transit operations during the COVID-19 pandemic. The behaviour of Twitter users in Madrid, London and Prague is analysed with the goal of recognising similar patterns and detecting differences in traffic related topics and temporal cycles. Topics in transit tweets were identified using the bag of words approach and pre-processing in R. COVID-19 is a dominant topic for both London and Madrid but a minor one for Prague, where Twitter serves mainly to deliver messages from politicians and stakeholders. COVID-19 interferes with the meaning of other topics, such as overcrowding or staff. Additionally, specific topics were discovered, such as air quality in Victoria Station, London, or racism in Madrid. For all cities, transit-related tweeting activity declines over weekends. However, London shows much less decline than Prague or Madrid. Weekday daily rhythms show major tweeting activity during the morning in all cities but with different start times. The spatial distribution of tweets for the busiest stations shows that the best-balanced tweeting activity is found in Madrid metro stations.

Keywords: Twitter; public transport; COVID-19; time analysis; text mining

1. Introduction

Public transport as an important public service requires continuous monitoring, timing corrections and improvements to satisfy the multitude of travellers. Operational data must be quickly updated to tackle problems as they arise. With conventional methods of data acquisition, getting these near constant updates would be both financially and technically difficult. Additionally, traditional information channels, such as sensor networks, are limited, especially in the case of transport-related issue detection due to sparse physical sensor coverage and the labour intensiveness of reporting incidents to the emergency response system [1]. It is also difficult to maintain them regularly [2]. Traditional sociological tools, such as questionnaires, are expensive, time consuming and often suffer from vague answers, incomplete information from respondents and are biased by predefined answers [3]. Answers might also be skewed because of the sample of respondents [3]. Researchers often complement traditional surveys with crowdsourced data from social networks. Social media data are dynamic, user-generated, and include a combination of spatial, temporal, and textual information [4]. They grows by not only number of users and number of posts or messages, but are also increased by social interactions, providing faster and more open reactions.

Penetration of social media into the population is not equal, however. Currently, one of the major problems in this method of data collection is the separation of social groups by different types of active social media usage. Social exclusion of the elderly is traditionally documented [5–7], but differences in the number of users within each ageing adult population group decreases. While the older population is more active on older types of media, such as Facebook, generation Y and Z prefer Instagram, Tik-Tok and YouTube [8]. Despite such a common separation tendency, there are social media which seem to more equally penetrate society and they are widely used for information spreading, such as Twitter. Twitter is a promising resource of a big volume of data for studying social interactions, behaviours, and attitudes, as well as for monitoring large distributed systems to detect problems early. Twitter data are comprised of a wider range of respondents than would be covered in a sample from a traditional research method [9]. The popularity of Twitter as a social medium differs from country to country. In the UK, penetration reached 28.34% of the population in October 2021, and Twitter became the most used platform for coronavirus news [10]. In Spain, Facebook is the most popular platform; however, Twitter is also used by 8.8% of the population [11]. Czechia is on the tail end with Twitter usage by only 5.31% of the population [12,13] (representing the sixth order). While the number of users and tweets is low, tweets are still widely used to disseminate comments by politicians, information for citizens delivered from central or local governments, police and other institutions as well as information for customers from transit providers.

Public transport in London represents the main transport mode for commuters, carrying 1.35 billion passengers annually. The London tube is the oldest metro system in the world and many of the ongoing public transport issues in the city relate back to this fact; older station design and heavy passenger traffic are hardly conducive to meeting modern requirements for comfort, safety and punctuality. The biggest public transport issues during the pre-pandemic period were overcrowding on the tube and delays in bus connections due to heavy traffic. The public transport of Madrid in the pre-pandemic period faced issues caused by a reduction of trains and services, making it difficult to provide services to an increasing number of users. The main problems were breakdowns on the lines, delays in the timetables, overcrowded platforms and trains, lack of ventilation and air conditioning, and lack of attention to the peripheral stations due to unavailability of staff [14]. Similar issues can be found in Prague. Periodically, partial service breakdowns occur due to reconstruction works. Often, they are coupled with delays and the overcrowding of operating vehicles. A shortage of transit drivers restricts potential service extension.

The traffic situation during the COVID-19 pandemic has been analysed in several studies. Bansal et al. 2022 analysed the traffic situation in London, which was affected by the pandemic in many ways. The initial lockdown resulted in a drop in public transport usage. After the introduction of mandatory face masks on public transport premises, the ridership increased. On the other hand, the travel time decreased, perhaps due to the discomfort of wearing face masks [15]. The behaviour of travellers also changed. For example, people were more sensitive to crowding levels, when COVID-19 cases were high. In Madrid, ridership decreased by 95% at the peak of the pandemic. The only passengers that kept using public transport during the pandemic were commuters who needed to continue travelling to and from their job site. Ridership patterns also changed, showing lower afternoon peaks and a morning peak hour before its usual time [16].

Analyses of Twitter are widely used for various purposes, e.g., assessment of happiness or evaluation of opinions about a given topic or detecting problems. An interesting utilisation of Twitter analysis is the detecting overall distribution of topics in a city. For example, 20 topics of discussion were identified in London on Twitter [17]. Popular topics such as sport and games are common, (4.79%), humour and informal conversations (5.5%), politics and current affairs (5.18%) or slang and profanities (5.06%). People also talk about transport and travel, which accounts for 5.52% of all discussed topics. When expressing perceptions on social sites, informal language is often used. Social sites create a feeling of anonymity, and profanities are used more often [18]. On the other hand, it can be seen that

the usage of profanities portrays the highest level of user dissatisfaction and they utilise them accordingly.

Practical usage of Twitter data analyses is still low. Transit companies utilise Twitter mainly as a channel of promotion and communication of news and information such as special timetables. Their Twitter accounts also reply to users who ask them questions or report issues such as lost items, and they occasionally tweet out data on transport usage.

The goal of this paper was to compare selected transit related topics discussed on Twitter in London, Prague and Madrid. Our interest is in discovering how people tweet in these cities. Focus was placed on the understanding of tweeting course and pattern during the pandemic period, comparison of tweeting activity on a weekly cycle and daily cycle, and differences in discussed topics. This knowledge is essential to improving the processing of data from social networks, discovering useful transit related comments, attitudes and proposals and gaining a deeper understanding of passenger behaviour and ability to influence their opinions.

The paper is organised as follows: in the second section, the relevant literature is reviewed, underlining data collection methods and approaches to semantic analysis. The third section explains the data processing used in this study, namely identification of relevant topics. The fourth section presents the main results of the study consisting of tweet evolution analysis, comparison of daily and weekly cycles of tweeting activity, as well as different rhythms of topic in tweets and spatial distribution of activity among main transit stations. The fifth section provides a discussion of the results and the conclusion summarises the main findings.

2. State-of-The-Art

One of the main public ground transport providers, Transport for London (TfL) widely uses Twitter data to analyse interactions with customers [19]. Another frequent usage is to provide traffic updates. Public transport operators use social media mainly for [4,20]: (a) real-time updates and information, (b) information for customers regarding services, fares, and service disruptions, (c) engaging citizens by handling complaints and inquiries, (d) employee recognition and recruitment of staff and (e) video entertainment and contests. These interactions between users and operators on social media can provide a cost-effective, reliable, and timely mechanism for sharing information with passengers and other travellers [21]. Public perception regarding transit systems can be captured as an essential part of creating an appropriate and equitable service. Gathering such perceptions also holds the potential for identifying ways of increasing ridership and for exploring sources of transport-related social exclusion [22,23].

2.1. Semantic Analysis

Semantic analysis enables an understanding of the topics discussed in tweets. Semantic analyses are focused on selected thematic categories which enable concentration of manifold and vibrant information expressed in natural language into specific classes which are more suitable for useful interpretation and decision making. The class names may directly indicate how to deal with customers' satisfaction or dissatisfaction in relevant classes. To discover relevant topics, two approaches were applied—an expert interpretation of the content (bag of words where relevant keywords are defined in advance, followed by a supervised interpretation) and artificial intelligence-based unsupervised analysis (i.e., Latent Dirichlet Allocation LDA), where the data is first processed automatically, followed by an expert interpretation of the emerged topics [24].

In general, transport topics cannot be narrowly specified, and one topic usually covers a wide range of problems. For example, when speaking about COVID-19, there may be problems with not respecting precautions, not wearing masks or overcrowded trains where people fear getting infected. Effective semantic analysis requires the selection of an exhaustive set of topics well-aimed at the specified interest. More specific topics should be validated by surveys or customer/traffic observations. For example, Alshehri et al. [9]

who studied user satisfaction with "London's Oyster" applied the following categories: system (system faults, problems with usage), Twitter communication (feedback, support, helpdesk), organisation (technical support, competence of employees), and market coverage (ubiquity). Such a large framing of categories invokes additional exploration even on the level of individual tweets. A much more detailed exploration enables a good understanding of customer complaints.

Liu et al. [7] identified the five following topics in the Chinese Weibo discussion: "description of unfair events" (19.8%), "Provision of transportation events" (33.4%), "Status quo of elderly people" (33.9%), "Pandemic prevention" (7.3%), and "Government response" (5.6%). The proposed topics are specific and narrowly targeted. They also explored the temporal pattern of topics' occurrences.

The approach to semantic analysis might be supervised, e.g., by using predefined topics with a predefined set of words typical to these topics or unsupervised using topic modelling techniques, e.g., LDA [25]. For determining the relevance of tweets to predefined topics, machine learning techniques are often used, for example, Almohammad and Georgakis [2] used the Random Forest model [26] or TFIDF (Term-Frequency-Inverse-Document-Frequency) algorithm implemented by [7].

2.2. Methods of Data Collection

Twitter has developed an API for a distant approach to its data [2,3,22,25,27,28]. The Twitter API can be accessed through various tools such as Python tweepy library [29], twitter4j Java library [30], twitteR or other relevant packages for R [26]. They evaluate tweets in real time or download archives of the past seven days. This seven-day time window is one of many restrictions and limitations of previous Twitter APIs [31]. The majority of these restrictions were removed in Twitter API v.2 which was launched in November 2021 [32].

This time window selection is typical for such social network APIs. Liu et al. [7] worked similarly with Weibo and its API. It allows for the downloading of data from several past timeframes, usually from 7 to 9 days [7]. Due to this restriction, they decided to use a crawler made in Python to download a seven month data archive for given keywords.

Tweets can be downloaded for either text selection (key words, users' profile, hashtags) or for geographic selection with a spatial filter (buffer around a point) [2] or a bounding box [33].

Similar to other social networks, the feed of messages on Twitter contains both individual user-generated content and tweets from authorities and robots (bots). The tweets from local authorities are more structured than tweets from regular users, therefore a transfer learning technology with a pretrained CRF model was used to process these data. To enhance the performance of the model, spatial rules based on spatial prefixes were used [3].

Usually, data are aggregated into several categories according to different research priorities. Temporal categories (month, day of the week, hours, peak hours, etc.) can be specified according to transport schedules and operation conditions into, e.g., normal operation days, days with disruption, and days with information surges [34]. Spatial categories may include boxes or other relevant shapes, e.g., Almohammad and Georgakis [2] divided the city of Manchester into 55,448 bounding boxes of approximately 111×133 m. A specific approach was used to handle user mentions, URL addresses and emoticons, while a standard approach would be to delete them as redundant parts of text.

Twitter data are frequently complemented with data from transport surveys, customer satisfaction surveys [34], data from official traffic information systems [9], data from transport service providers (e.g., shared bike services [30]), traffic restrictions and disruptions reported by the public/passengers [2], accident-reporting information systems [2], and statistical data.

2.3. Data Processing

Data were downloaded from the official transit company accounts from London, Madrid and Prague, including all tweets mentioning and addressed to these accounts, for a period from 26 March 2020 to 31 January 2021 (hereafter the 'pandemic period'). Detailed steps for data downloading using Twitter API v.1 are explained in Figure 1. Selection of tweets addressed to official transit providers enables the focus to be set on transit-related tweets. For London (@TfL), 545,295 tweets were received and for Madrid (@metro_madrid), 102,822 tweets were received, while for Prague (@DPPoficialni), only 4949 tweets were obtained. London represents a city with a very high usage of Twitter. When analysing English language tweets, no translation or adjusting processing tools needed to be used. Contrarily, Prague is a city with a very low usage of Twitter and Czech language tweets require more demanding processing. Madrid is situated between those cities where the wider usage of Twitter is combined with a non-English environment. All three capitals utilise an extended urban public transport system centred around the Metro system.
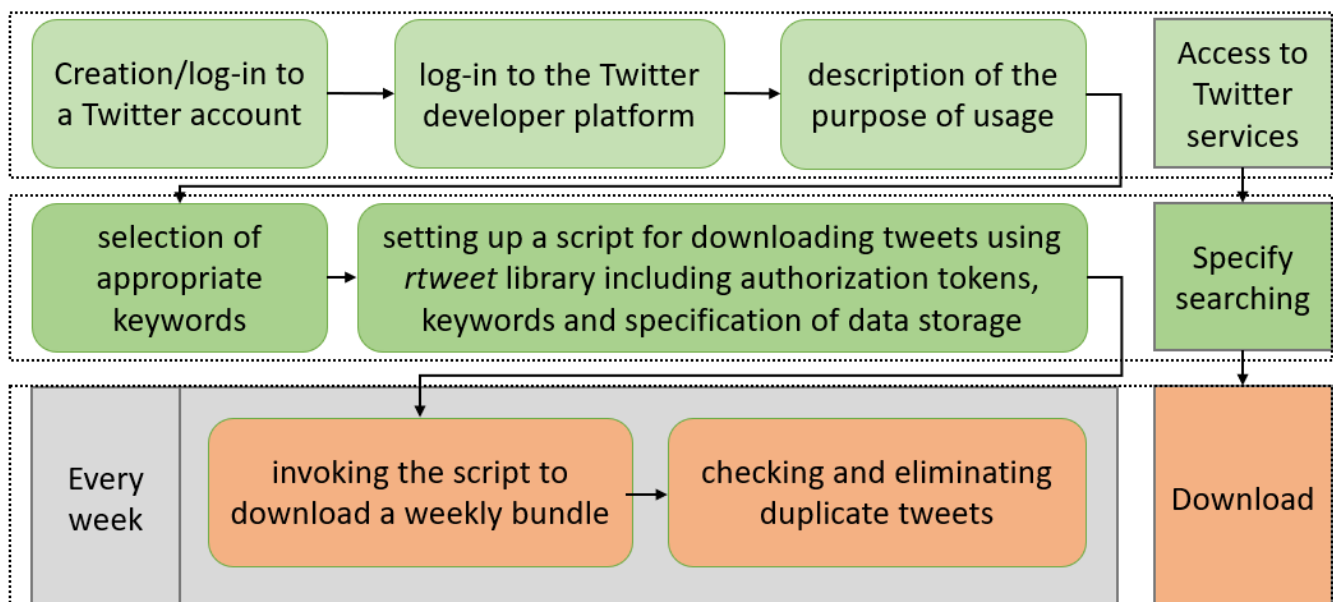

Figure 1. Twitter data downloading process with Twitter API version 1.

For comparison with the situation before the COVID-19 pandemic, an additional download using Twitter API v.2 was conducted for the same interval one year before (26 March 2019 to 31 January 2020, hereafter the 'pre-pandemic period'). The new download is much easier than when using the previous version of Twitter API.

For further analysis, tweets had to be pre-processed. The pre-processing function utilises basic commands from tidytext and dplyr packages of R applied to each tweet. It converts all letters to lower case, transfers utf-8 characters to characters in the basic ASCII table, and transforms every user mention, hashtag, punctuation mark, emoticon and stopword. URL addresses were changed to string "URLs", everything starting with an "at" sign was changed to "USER_MENTION" and emoticons were changed according to their emotion to "EMO_POS" and "EMO_NEG" [33]. It is in such a way that tokens in each tweet were identified (tokenisation process [35]).

An exploratory data analysis of tweets provided an overview of the most commonly used words. This analysis enabled the identification of the following important topics: issues in public transport, comfort at stations and transportation hubs, problems with payment services for public transport and COVID-19-related issues.

COVID-19 represents the most discussed theme in almost every area during the selected temporal period due to the pandemic situation. The delay and congestion topics arose from the fact that delayed connections are the most common issue related to public

transport. Quite often, users mention official accounts and accounts of activists in tweets, especially when they want to announce an issue and a certain response is anticipated. The topic of staff was chosen due to the direct relationship to transport provider's services and interesting temporal and contextual aspects. The curse word topic arose from the fact that people use profanities as the highest display of dissatisfaction. Therefore, it was decided to track these words during the day and by location to explore their distribution and aspects forming peaks of customer frustration.

For each topic, the list of tokens was prepared. Personal auditing of individual tweets containing these words was required to confirm relationships between words and topics and exclude ambiguous tokens. Additionally, the frequency of words used is important for the decision to include the word into the list of tokens. After completion of the list of tokens for a given topic, each token was looked up in the full list of tokens created from the tokenisation process. In this phase, all word shapes for the given token were searched for and added to the token list to minimise any omission error.

For the sake of simplicity, each token was assigned to only one topic despite the classification ambiguity of some tokens. Some of the topics were respecified according to results of the detailed analysis. The cross-frequency table was created to compare the occurrence of topics across time of day and city. Not every topic shows sufficient activity in all three cities, therefore, two sets of topics were created. The first set of topics (Table 1) includes general topics that can be documented in each city and covers typical themes currently discussed worldwide such as COVID-19, transport delays, or messages of official representatives. The first set enables comparison between cities. The second set (Table 2) includes topics that are specific to London, Madrid or Prague.

Table 1. General topics and associated tokens in cities.

Topic	Tokens in London	Tokens in Prague	Tokens in Madrid
COVID-19	COVID, Masks, Wearing	Rouska, Koronavir, Pandemie	COVID, Mascarilla, Pandemia
Delay	Delay, Delayed, Congestion	Zpozdeni, Odklon, Zpozdene	Retraso, Tarde, esperando
Transport Card	Oyster, Oystercard	Litacka, Litackapraha	Tarjeta, Abono, billete
Officials	Sadiqkhan, mayoroflondon, willnorman	Adamvojtechano, zdenekhrib, scheinherr	Ayuso, Alme, ComunidadMadrid
Activists	Airqualitynews, Cleanairlondon, Lucyfacer	Pavelnovotnak, Tramvajak, Janerdubes	SufridoresMetro
Staff	Staff, Workers	Personal, Obsluha, Revizor	Trabajadores, personal

Note: Translation of selected tokens: Rouska—mask, Koronavir—Coronavirus, Pandemie—pandemic, Zpozdeni—delay, Odklon—detour, Zpozdene—delayed, Personal—staff, Obsluha—steward, Revizor—conductor.

Table 2. Specific topics identified in cities.

Topic	Words	City
Air quality	Airpollution, cleanair, mumsforcleaneir	London
Dirt	Disgust, filth, malodor	Prague
Curse	Shitty, sucks, morons	London
Racism	Racism, racist	Madrid
Maintenance [1]	Malfunction, broken, maintenance	Prague
Maintenance [1]	Breakdowns, works, stairs, elevators	Madrid

[1] The presence of the maintenance topic was revealed both in Prague and Madrid but with different relevant tokens.

The COVID-19 topic occurs in messages containing complaints about people (not) wearing masks, not maintaining appropriate distance and not sanitising hands when

entering public places. The Delay topic is often associated with congestion, which is why this token was included. The topic is also complemented with words expressing time intervals (minutes, mins). The Transport Card topic is linked with the name of these cards in the given city system (i.e., Oyster card in London and "Lítačka" in Prague). Usually, tweets are related to problems and usage of the public transport card such as payment issues, terminals not working and basic functional questions. The Officials topic contains all usernames of official accounts, while Activists remembers the usernames of activists in public transport discussions. The last general topic, Staff, is associated with transport provider employees. These topics were tracked among all three cities.

The majority of the results utilised full data sets. To analyse spatial distribution, a better spatial resolution was required. According to [36], an extraction of transit stops' names ought to be applied. For their study, names of the most frequent stations were searched in the collected tweet sets for each city. Only the top seven busiest stations in the explored cities had a sufficient amount of tweets for further analysis. The names of stations were searched in different forms and variations to suppress the omission error. Bias can also be caused by the name of the station being a commonly used word. For example, the busiest station in the Madrid metro system is called Sol, which means sun.

3. Results

3.1. Tweeting Course

Tweeting activity in the explored pandemic period is relatively stable without apparent trends in all three cities (Figures 2–4). A certain decrease in activity is visible in December for Madrid. The variability in Madrid and Prague is higher than in London. Coefficients of variation are 117%, 117% and 64%, respectively. The curves are strongly modulated by weekly cycles, with a higher amplitude in London. The Prague curve is more randomly variable due to the lower volume of tweets.
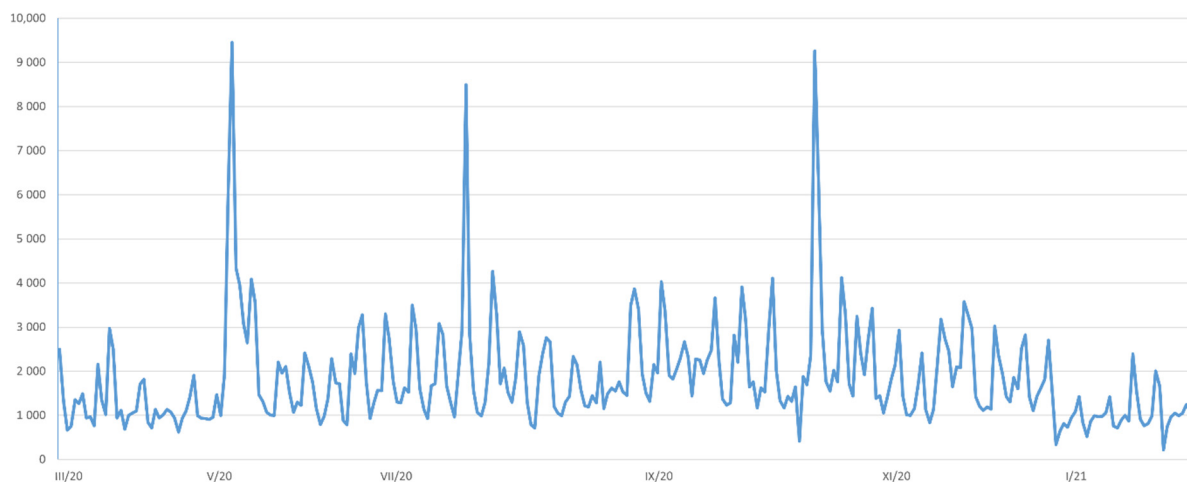

Figure 2. Course of tweeting activity in London (26 March 2020–31 January 2021).

The usual run is disturbed by extraordinary events when the tweeting activity booms to spikes. In London, there are three events that can be detected visually. On 15 July 2020, people discussed wiping Banksy's COVID-19-related graffiti off tube trains. On 21 October 2020, a fire alarm in the Victoria station caused a closure of the station and evacuation of passengers [37]. In Madrid, the March peak corresponds with Holy week holidays. Peaks in the middle of October and the middle of January can be credited to popular football matches in the city. Peaks in December correspond with holidays; many people travel to Madrid on the sixth and eighth holidays for shopping and the period of Christmas, and New Year is typically accompanied by a lot of activity and mobility in Madrid. In Prague, three events were detected. One of them relates to the pandemic situation when the announcement of automatic door opening on all transport vehicles due

to hygiene reasons was discussed on 9 September 2020 [38]. Another large discussion was on 16 September 2020 when Prague introduced a new colour scheme for public transport vehicles [39].
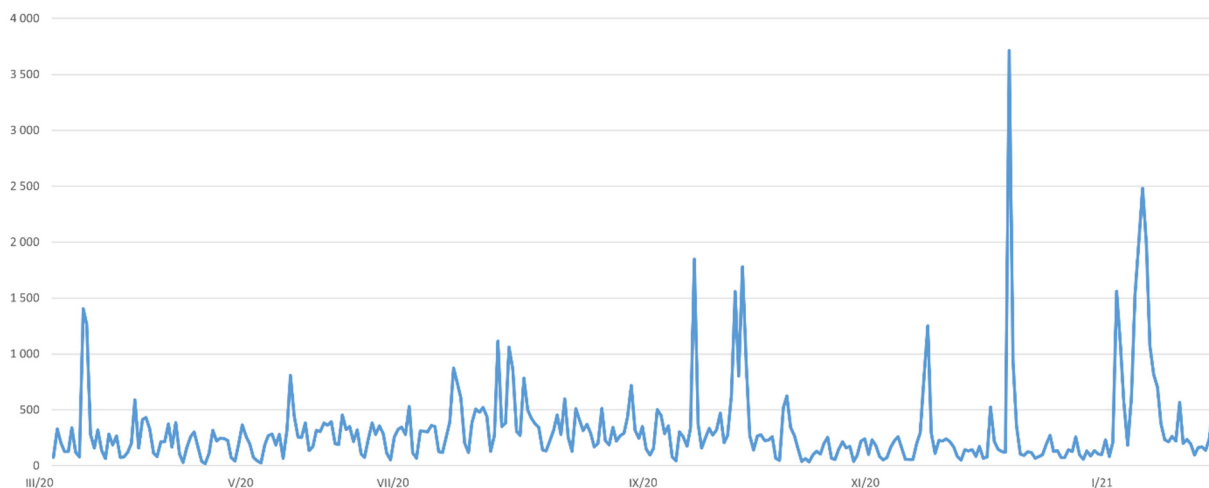

Figure 3. Course of tweeting activity in Madrid (26 March 2020–31 January 2021).
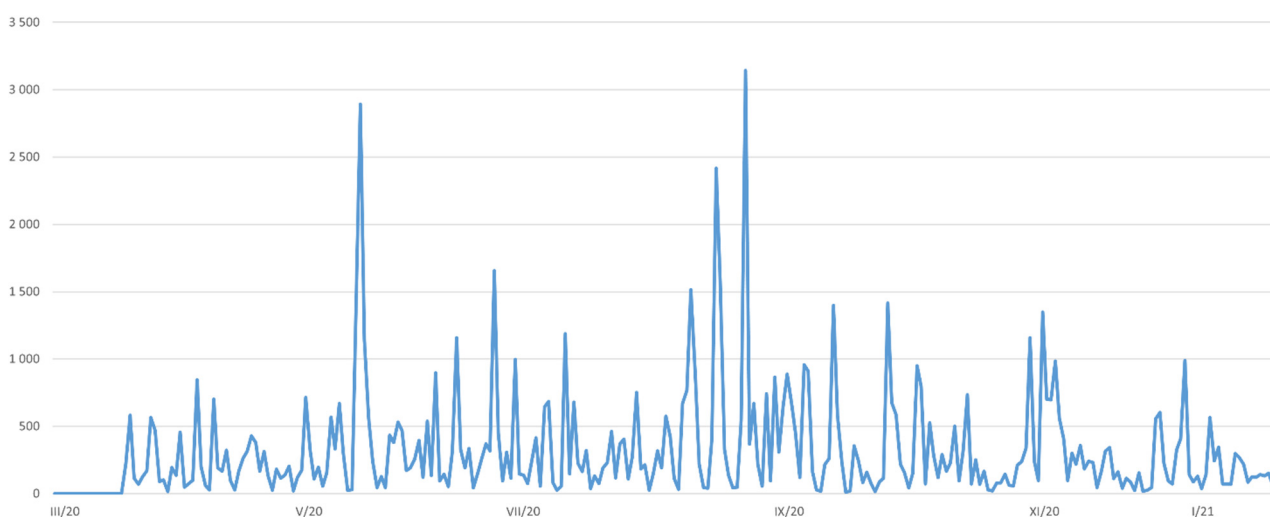

Figure 4. Course of tweeting activity in Prague (8 April 2020–31 January 2021).

3.2. Temporal Distribution of Tweeting Activity—Differencies in Week and Day Cycles

Tweeting activity in the pandemic period during workdays was almost constant with the exception of some local peaks for Wednesday in London and Prague and Friday in Madrid (Figure 5).

Analysis of outliers in London discovers two extremely frequent tweeting Wednesdays influenced by two extraordinary events discussed previously. Elimination of these two events causes a shift of the peak to Thursdays. These examples demonstrate the important role of specific events which may bias the distribution of tweets.

Similarly, in Prague, Wednesdays show a higher frequency than other workdays but there is no known reason for this increased activity.

In Madrid, there are two peaks of tweeting activity on Monday and Friday. This increased activity is caused by the fact that, according to official data [40], most users travel to work on these days. Friday's increased activity could be also caused by young people using the metro to go to parties to start the weekend.
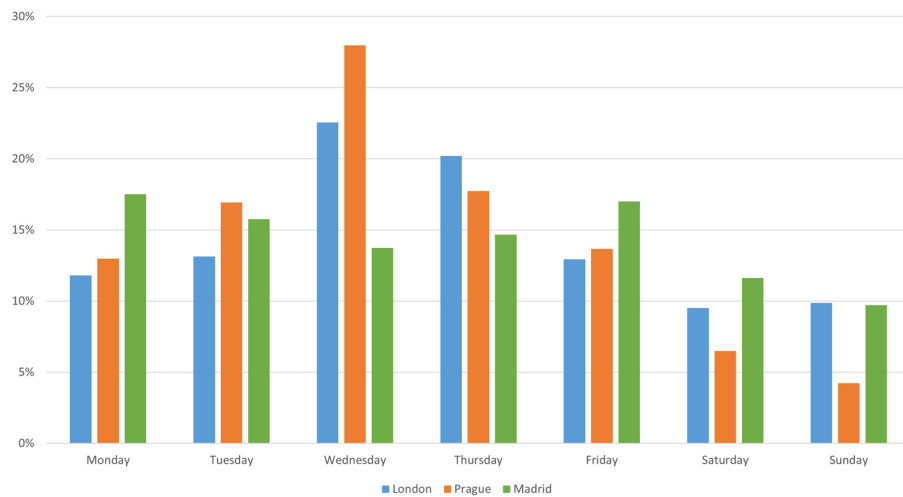
Figure 5. Share of traffic tweets in London, Prague and Madrid per day in the pandemic period.

Tweeting frequency declines on the weekend to 63% (London), 31% (Madrid), and approximately 28% (Prague) of average workday activity. This indicates less travel activity in Madrid and Prague over weekends, compared to London. People complain about traffic when it directly affects their day, e.g., when it makes them late for work or university, and these problems are less likely to occur during weekend.

The distribution of tweeting activity in the pre-pandemic period is more evenly distributed between working days and weekends with a strong decline towards weekends (Figure 6). The patterns for London and Madrid in the pre-pandemic period are similar to those during the pandemic—maximal activity is in the middle of working days in London, while at the end of working days for Madrid.
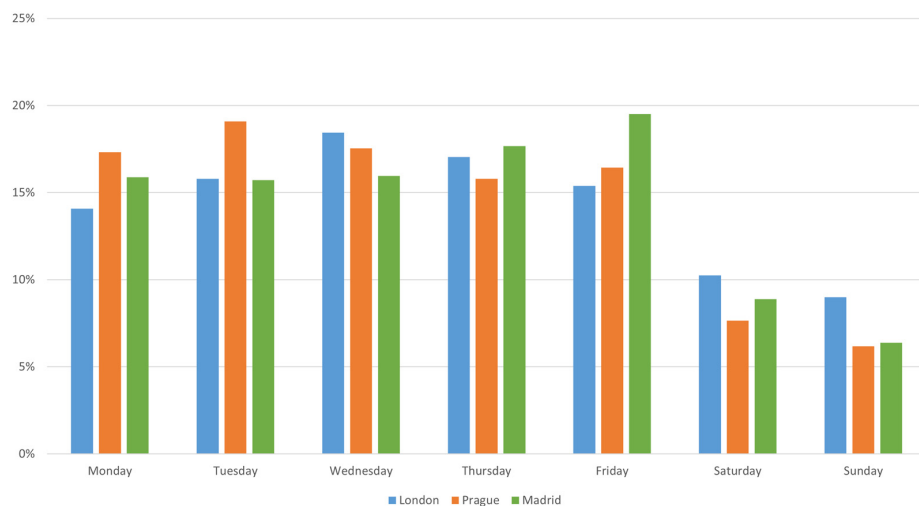

Figure 6. Share of traffic tweets in London, Prague and Madrid per day in the pre-pandemic period.

Due to the periodic character of the data, circular properties of the data set were analysed, which confirm Thursday as a mean tweeting activity day for London, and Wednesday as a mean day for Madrid and Prague.

During the day in the pandemic period, tweeting activity continuously rises from 5:30 in the morning as people start to commute to work (Figure 7). The activity is fast-growing until 8:00 in every city. Later, the activity differs between cities. In Madrid and Prague, the activity slows down and rises again, reaching the daily peak at 10:00 and 11:00, respectively. In London, tweeting continuously grows until 11:00. In the afternoon, the activity slowly decreases until the evening. The afternoon peak in London is much earlier than in Madrid (17:00 and 20:00, respectively). The London curve is smoother than those for Prague or

Madrid due to higher tweeting activity. The time shift for Madrid could be explained by a shift of the whole daily cycle in Spain.
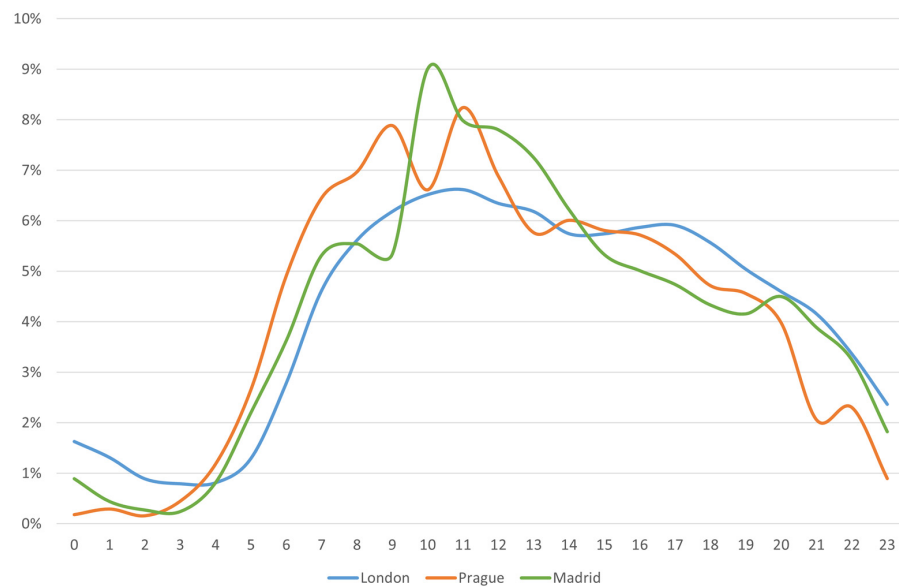

Figure 7. Tweeting daily rhythm for transport companies in London, Prague and Madrid in the pandemic period.

In the pre-pandemic period, a different daily rhythm is portrayed (Figure 8). In London, clear morning and afternoon peaks are formed at 8:00 and 17:00, respectively. In the pandemic period, both peaks are mild and the morning peak is shifted to later (at 11:00). Even stronger changes in the daily rhythm are seen in Madrid; in the pre-pandemic period, strong morning and afternoon peaks are situated at 6:00 and 13:00, respectively. During the pandemic period, the peaks merged. In Prague, one larger morning peak is reached between 8:00 and 11:00 and several small afternoon peaks occur between 15:00 and 19:00.
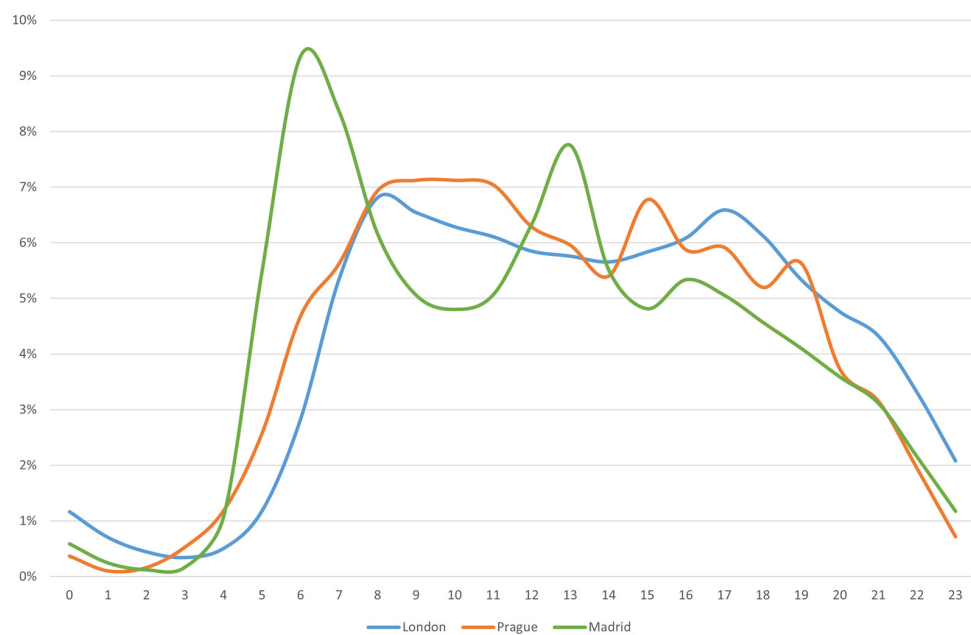

Figure 8. Tweeting daily rhythm for transport companies in London, Prague and Madrid in the pre-pandemic period.

3.3. Daily Rhythm of General Topics

The day-cycle distribution of topics was evaluated according to the frequency of relevant tokens. Frequencies were recalculated to the relative share of the daily frequency for the given topic. Their daily rhythms in the pandemic period are presented in Figures 9–11. Significant differences between cities are recognised mainly in various shares of **COVID-19**, **Delay** and **Officials** topics. The most important presence of the **COVID-19** topic is in London (22–38%), followed by Madrid (about 22% except in night extremes) and Prague (around 4%). **Delay** possesses the stable share during the day in London (approx. 10%) and also in Prague (5%). According to the highly variable share of the Delay topic in Madrid's tweets, it is possible to assume big daily differences in delays in the Madrid public transport system. The largest share of Officials mentions is in Prague, followed by London and then Madrid. Contrarily, the role of activists seems to be highest in Madrid. These differences may indicate different usages of Twitter between cities. The high variability of shares in morning hours in Madrid (Figure 10) is caused by very low tweeting activity during these hours.
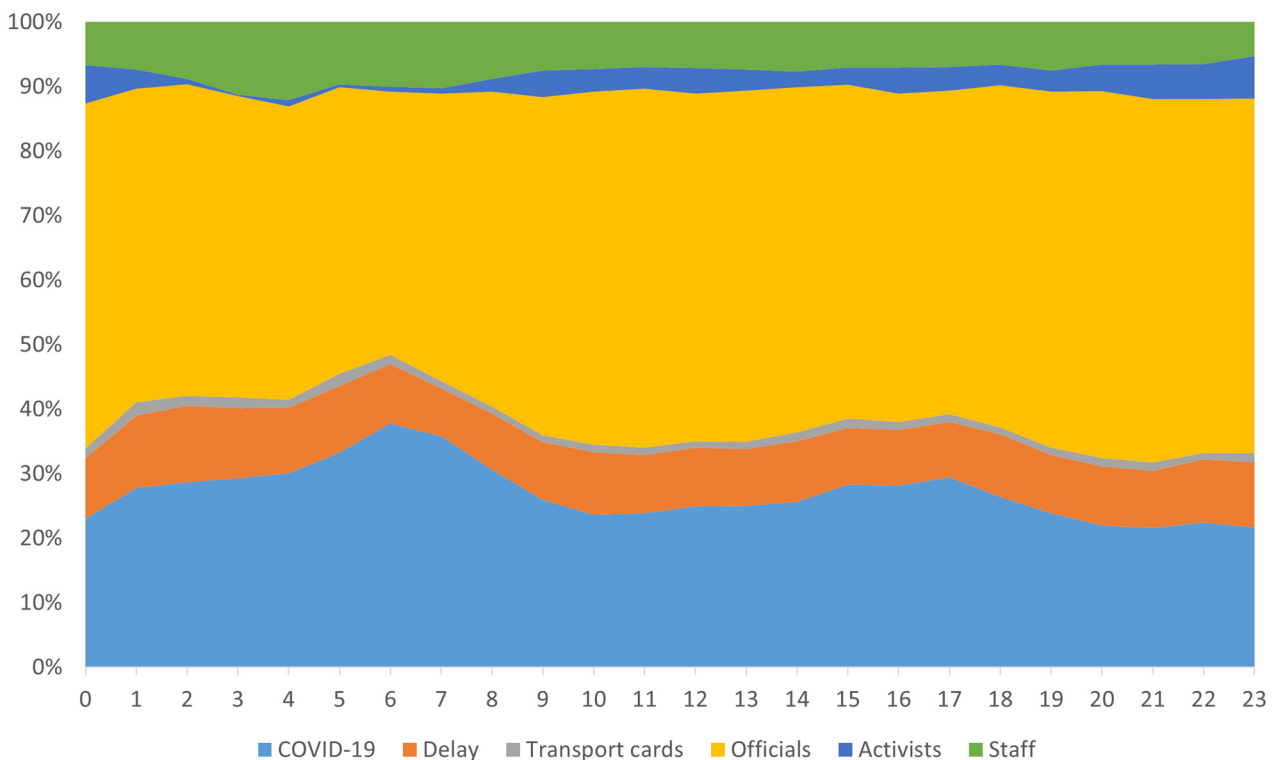

Figure 9. Daily rhythm of selected topics in public transport tweets in London.

In the pre-pandemic period, obviously COVID-19 was seldom discussed. In this period, discussion about delays and transport cards were much stronger both in London and Madrid. Likely, during the pandemic period, customers encountered less stressful transit with fewer delays due to the drop in the number of passengers. Interestingly, a much larger share of tweets from activists in Madrid were found in the pre-pandemic period. The daily rhythm in Prague in the pre-pandemic period for selected topics is practically the same.

Differences in daily topic rhythms is better recognisable from the share of the day (Figures 12–17). With the exception of **COVID-19** and **Delay**, each topic appears around 5:00 in the morning, when the demand for public transport is highest as people commute to work. Most transport services decline at night, so it is believed that these tweets are mostly from people coming back from nights out and include complaints about lack of connections in the night and maladaptive people not obeying the pandemic restrictions.
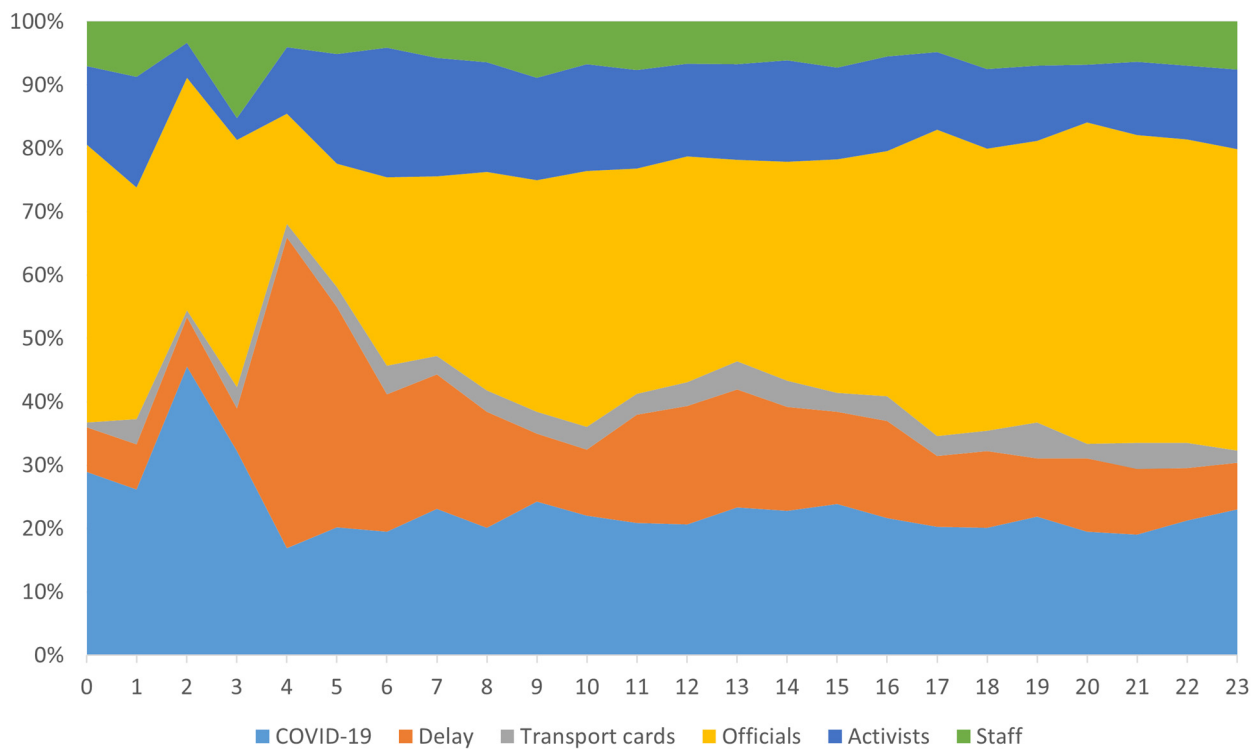
Figure 10. Daily rhythm of selected topics in public transport tweets in Madrid.
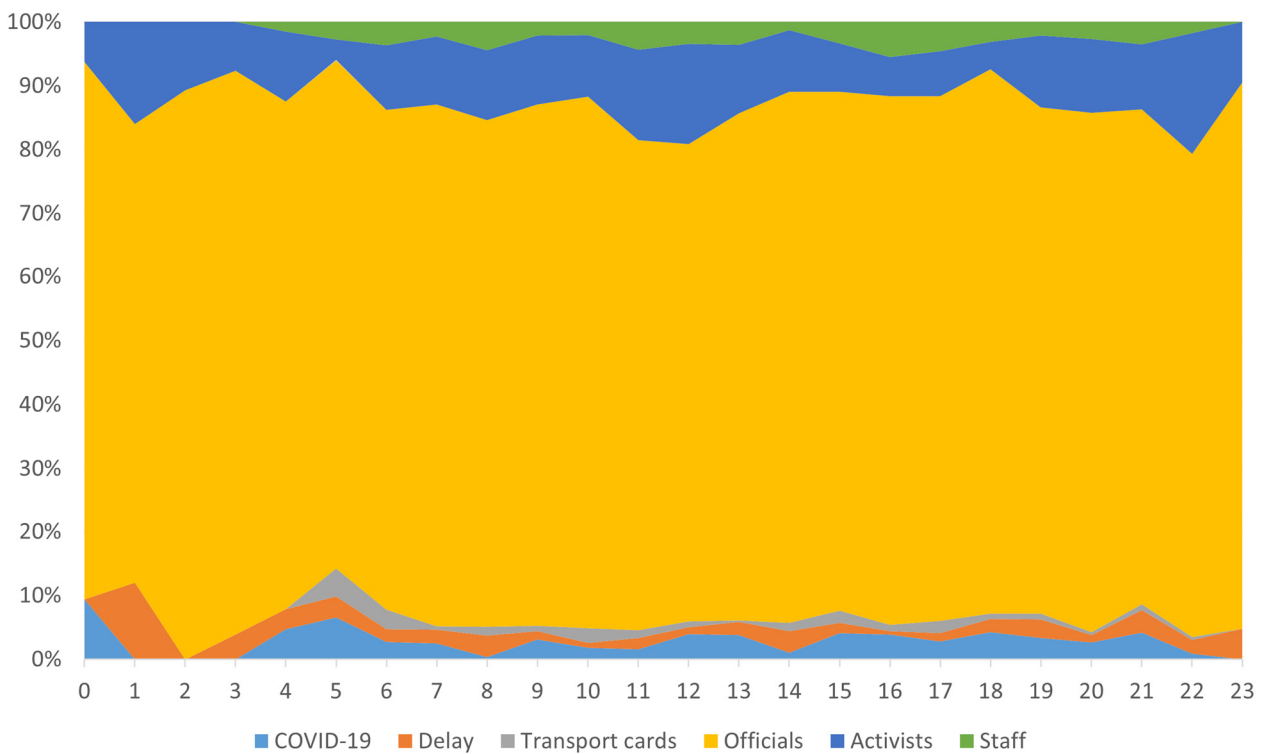

Figure 11. Daily rhythm of selected topics in public transport tweets in Prague.

Figure 12. Daily distribution of activity concerning COVID-19 (**left**) and transport cards (**right**) in the pandemic period.
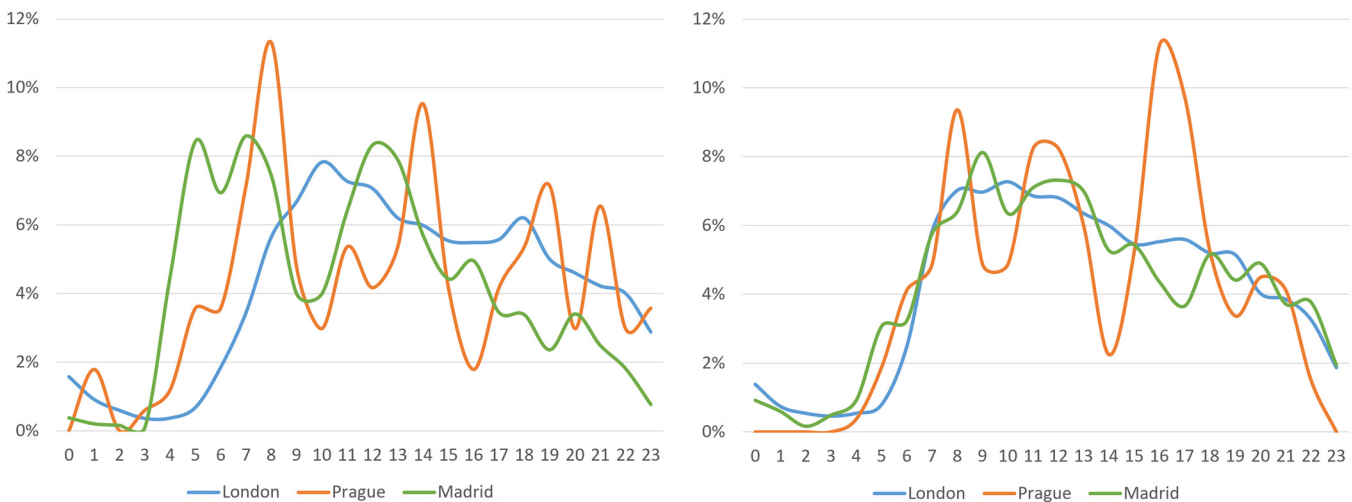

Figure 13. Daily distribution of tweeting activity concerning delays (**left**) and staff (**right**) in the pandemic period.

The **COVID-19** topic was the overall most common problem discussed in the year 2020. Tweeting about this topic started early in the morning at 4:00 in Madrid and Prague as people start to commute to work (Figure 12 left). In London, top tweeting activity occurs from 7:30 to 12:00 (probably related to the overcrowded underground); later, the activity about this topic decreased. The following peak is around 17:00. In London, the topic is highly present in rush hours. In Madrid, tweets commenting on COVID-19 issues are more frequent in morning, from 7:00 (when users tend to leave their homes to go work) until 13:00 (lunch time hour). It can be seen that the activity hours are coincident with the working and break hours, while, in the afternoon, the percentage of tweets decreases due to leisure, sports, or other free time activities.

Discussion concerning **Transport Cards** in the pandemic period reaches a higher peak in Madrid in the early morning (Figure 12 right). In London, the activity course is similar to the discussion about delays—slowly starting to reach a peak at 10:00. The second peak in London, occurring around 18:00, contains tweets mostly regarding refunds and information about cards or proposals for improvement of the application. Madrid presents higher activity in the morning, with a peak from 12:00 to 13:00. There is another minor peak after 19:00, the time when commuters travel back from dinners or work.
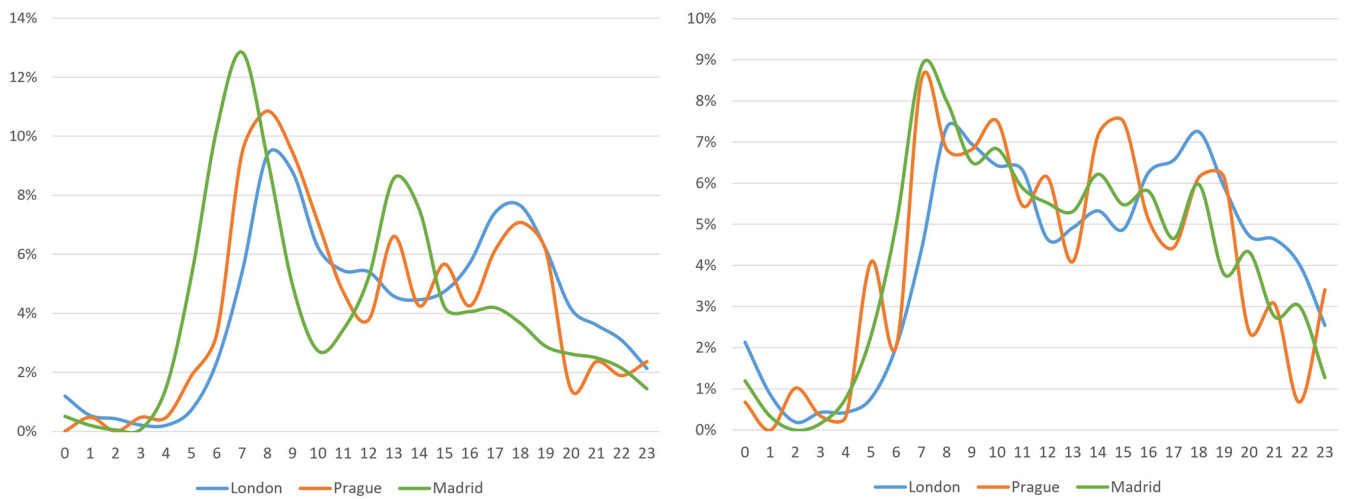
Figure 14. Daily distribution of tweeting activity concerning delays (**left**) and staff (**right**) in the pre-pandemic period.
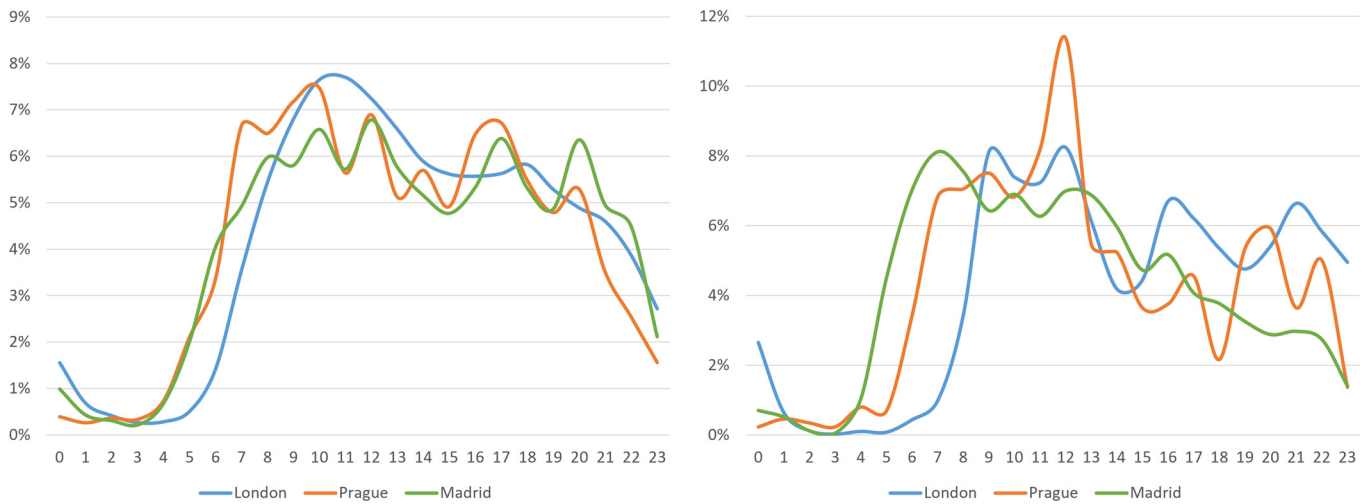

Figure 15. Daily distribution of mentions of officials (**left**) and activists (**right**).
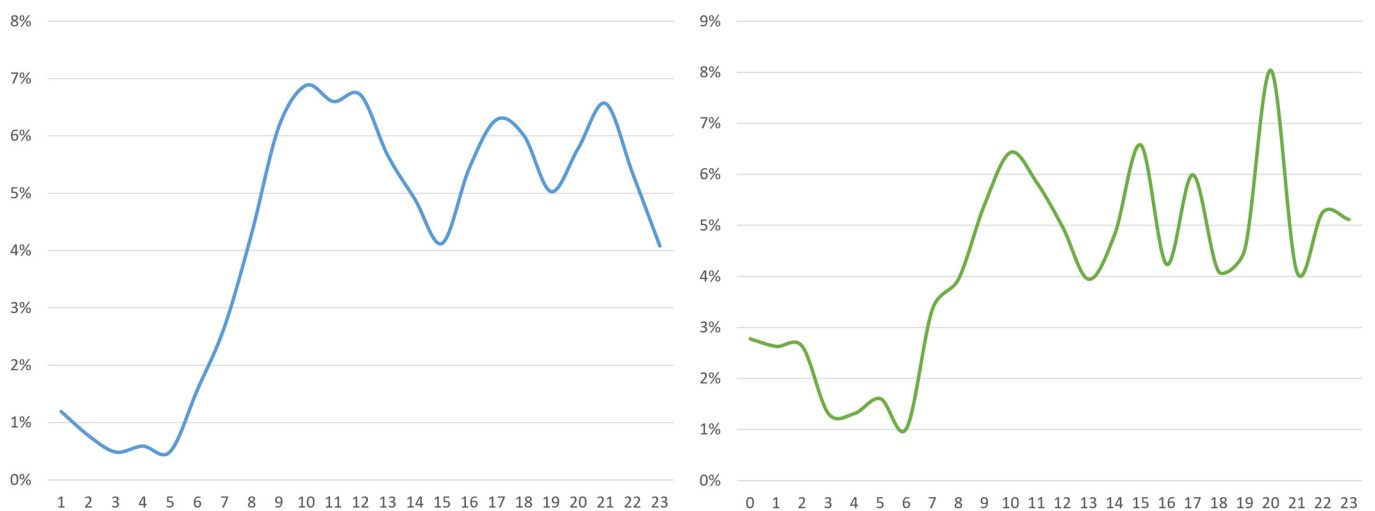

Figure 16. Profanity in London transit tweets (**left**) and racism in Madrid transit tweets (**right**).
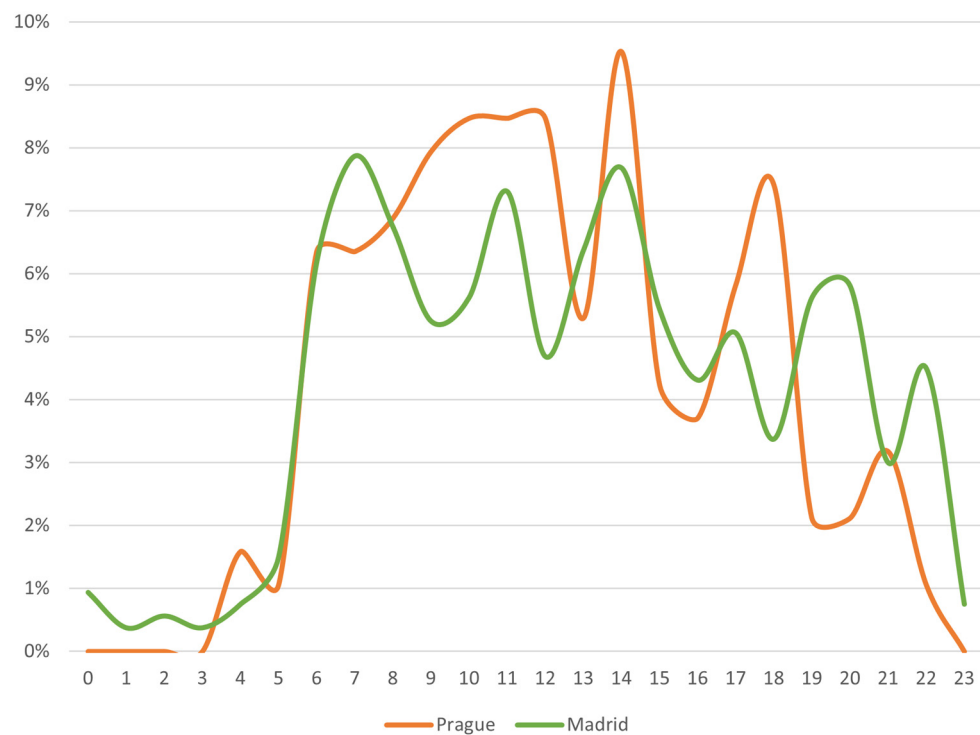
Figure 17. Daily distribution of activity concerning maintenance in Prague and Madrid.

The overall activity for this topic is halved compared with the pre-pandemic period; thus, this topics discussed during the pandemic can be seen as much less important. The daily pattern in London in the pre-pandemic period shows a steeper rise in the morning (passengers during COVID-19 commute later and more distributed) and the afternoon peak is much later (17:00) than in the pandemic period. In Madrid, the morning peak in the pre-pandemic period was highly concentrated around 7:00.

The temporal distribution of the **Delay** topic (Figure 13 left) in the pandemic period shows that delays in Prague are most frequently commented on during the morning rush hour at about 8:00 and in the afternoon between 14:00 and 15:00. Delays are expected in these hours due to insufficient transit capacity for peaks of travellers. The distribution in London is slightly different; activity starts slowly increasing from 6:00 with significant increase during rush hours, the first peak occurring later at around 10:00, with a second smaller peak in the late afternoon at around 18:00. The reason for this could be that during the pandemic most office workers in London worked from home. When people were allowed to go back to work, they did not go back to their traditional 9–17 h jobs. When looking at cycle hire data we have found that, while the afternoon peak is still strong, the morning peak has become much smaller. This might be what is seen here—people are more flexible in their travel and can start work later in the day. In Madrid, the percentage of users complaining about delays is similar to that seen in Prague (peaks during morning rush hour and in the midday when users tweet during lunch breaks). The percentage of complaints falls continuously over the afternoon and may be due to a smaller coincident concentration of passengers and less time stress than in the morning.

Tweeting about this topic is almost double that in the pre-pandemic period in London, while almost the same level of tweeting is seen in Madrid and a slight decrease is seen in Prague. In all cities, stronger peaks are found in the pre-pandemic period (Figure 14). In London, morning and afternoon peaks are nicely distinguished by the noon saddle with almost half the complaints about delays. As well as this, the morning peak was reached earlier than in the pandemic period. All these findings confirm that travelling and complaints about delays were more clearly distributed in time in the pre-pandemic period. Madrid shows stronger peaks, but the similar basic pattern as in the pre-pandemic era

remains; there is no significant time shift of peaks. In Prague, the basic daily pattern is the same as in pre-pandemic period.

Discussions about **Staff** in the pandemic period start shortly after 4:00 in each city (Figure 13 right). In Prague, there is a peak around 8:00, followed by Madrid with a peak around 10:00. In London, the most frequent discussion about staff occurs between 8:00 and 11:00 and then slowly decreases.

Tweeting about this topic has risen London six-fold (!) during the pandemic period, mostly due to a relationship with the COVID-19 topic. Contrarily, in Prague and Madrid, tweeting about staff declined (91% and 77%, respectively). In London, there is a strong afternoon peak of tweeting about staff which has fully dissolved during the pandemic period. In Madrid, the morning peak is situated much earlier during the pandemic period. Prague's patterns are different in the afternoon hours—in the pandemic period one big peak arises at about 17:00, while previously two peaks were formed at about 15:00 and 19:00.

Tracked mentions of **officials** and **activists** in the pandemic period show different behaviour (Figure 15). In London, both topics have a similar course. There is a peak at 11:00, followed by a second peak at 18:00, after which the activity decreases. The morning curve copies the behaviour of the **Delay** curve, and supports the assumption that these two topics can be bound; hence, tweets about delays in the morning are usually addressed to officials, while in the afternoon this binding is weaker. On the other hand, in Prague, there is no significant peak in mentioning official representatives. The activity starts after 6:00 and increases rapidly until 7:00. Activity remains relatively consistent until 19:00, when a decrease begins. Contrarily, mentions of activists show a significant peak at 12:00, which is probably connected with lunch time. In Madrid, the percentage of tweets regarding official representatives is stable the whole day, while addressing activists' accounts is more popular in the morning, corresponding to the basic rhythm of Twitter users.

Compared with the pre-pandemic period, tweeting about officials significantly increased in London and Madrid (323% and 170%, respectively), while in Prague a slight decrease was found (89%). Different evolution is discovered for activists in cities; the number of these mentions has risen almost tenfold in London, while in Madrid it drops down to less than half. This indicates different roles and behaviour of activists in these cities.

3.4. Daily Rhythm of Specific Topics in the Pandemic Period

Air pollution is a frequent topic in London, in particular around London Victoria station and Victoria Line. Discussion of this topic is intensive all day with peaks during rush hours. This problem is almost missing in other stations (in London as well as in other explored cities).

People on social media use informal language which is often punctuated with **profanities**. Likely, it expresses the highest level of dissatisfaction and complaints, meaning the most negative sentiment. The most common English profanities were taken and their usage tracked across the day (Figure 16 left). Frequency of profanities in London tweets has increased by more than three times in the pandemic period. This reflects the growing frustration of the population. Peaks in the pandemic period are situated in morning rush hours, at noon and in the evening. The evening peak at about 20:00 is specific and only racism in Madrid shows the same evening peak. The frequency of curse words is partly related to the number of passengers in the station. People swear the most about "Victoria station", followed by "Stratford" (the number goes down to half) and then "London bridge". Usage of profanities in Prague compared to London is negligible (about 1% in Prague, about 5% in London). Czech people do tweet about traffic in a negative way; however, they do not swear as much. It may be related to different penetrations of social media in the population. In Madrid, Twitter users do not tend to use profanities either; they prefer to use both exaggeration and irony in their text to strongly signal their discontent, especially to do with punctuality and overcrowding issues. An example of such a tweet follows: *"Arriving half an hour late for work after waiting 10 min for the subway. Bravo Metro Madrid! Thank you for this great service and this great punctuality that you offer!"*

The volume of tweeting about racism in Madrid boomed to 10 times higher in the pandemic period compared with the pre-pandemic period. The tweets containing **racism** (Figure 16 right) peak at 10:00 (when users arrive at work), 15:00 (when they leave work), 17:00 (the start of the afternoon shift) and the highest peak at 20:00. These tweets mainly report racist activities they see in the metro network at that point of travel). Racism issues are more common in the late afternoon than in the morning due to the increased activity of teenagers and young adults, who are travelling to parties or have already started consuming alcohol. They tend to be prone to racism against people from Latin America, who are also more active on transit in the afternoon. The morning peaks are weaker due to the profile of metro users being mainly adult workers.

The final documented specific topic is **maintenance,** which is discussed mainly in Prague and Madrid but significantly less than in the pre-pandemic period (16% and 87%, respectively). The maintenance shows a different pattern from other topics (Figure 17). The curve steeply rises at 5:00, followed by two peaks at 15:00 and 18:00 in Prague. We hypothesised that people complain about maintenance issues during all operation hours, but the daily pattern differs from that of total daily tweeting activity. Madrid is similar to Prague, however, Madrid peaks around 7:00 with variable activity during the day and overall activity is higher in the evening compared to Prague. This may be due to maintenance discussion activity in Madrid starting and ending later, similar to other activity in the city.

3.5. Spatial Distribution of Tweeting in the Pandemic Period

This section utilises data for tweets selected for the names of stations. Obviously, the number of relevant tweets drops, resulting in only 3455 tweets for Madrid and 6507 tweets for London (about 1% in both cases).

London is served by a wide spectrum of public transport means, including the underground (the Tube), Docklands Light Railway (DLR), buses, riverboats, local trains, trams, etc. Current transport strategy (established 2017) is seeking to serve a projected population growth to 10.5 million people and 32 million daily trips by 2024 with 80% of people being delivered by public transport (there were 26.7 million daily trips with 36% by public transport in 2015) [41]. The busiest part of the London public transport system is the underground, where the top 14 major stations account for more than 60% of all transfers. The underground often operates close to maximal capacity in peak periods and directions, and overcrowding is a big problem in many parts of the system [42]. Similar to London, Madrid public transport means include metro, light rail, buses and regional trains. Madrid registers 4 million daily trips, 69% of which are by public transport [43]. Madrid is linked to its surrounding municipalities by six corridors covering all public transport services, which experience regular congestion problems on a daily basis [44]. Prague public transport consists mainly of a metro system (made up of only three lines), trams and buses. In the following figures, metro networks of London (Figure 18) and Madrid (Figure 19) with the numbers of served passengers are presented. Focus was set on the seven busiest stations marked in the figures. Unsurprisingly, they are main transport hubs.

Relationships between the tweet frequency and the number of passengers are weak (Figure 20) and not significant (R^2 are 0.21 and 0.24 for London and Madrid, resp.) caused mainly by the small number of points. The influence of outliers is high. Madrid's R^2 increases from 0.24 to 0.65, excluding the station Nueva Numanica. This station is a metro stop from Line 1 in Vallecas, a peripheral neighbourhood which is one of the most populated zones of Madrid and where the average income of the population tends to be one of the lowest in the city. It is assumed that the number of tweets is quite low, probably due to socio-economic reasons causing residents to be unaccustomed to making complaints on Twitter. The outlier in London is Victoria station where the high number of tweets is likely caused by the local problem of air pollution (discussed in this paper) and also two incidents related to COVID-19.
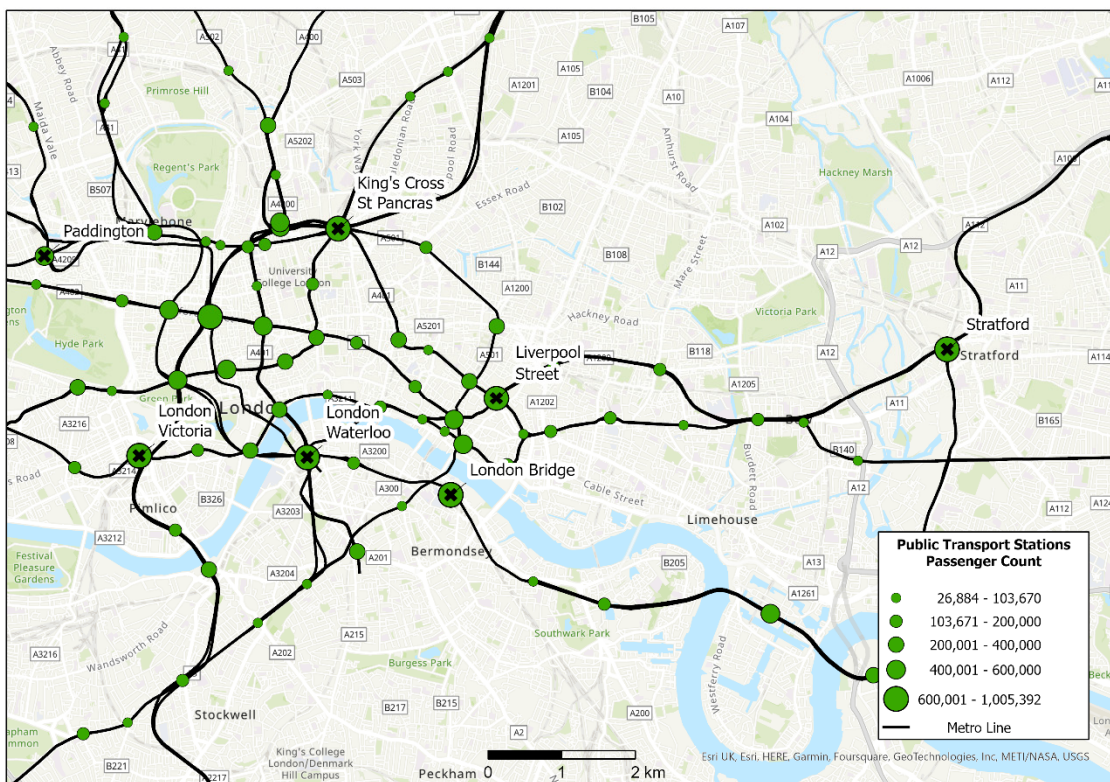
Figure 18. Number of passengers at main public transport stations in London (year 2021) (analysed stations are marked and labelled).
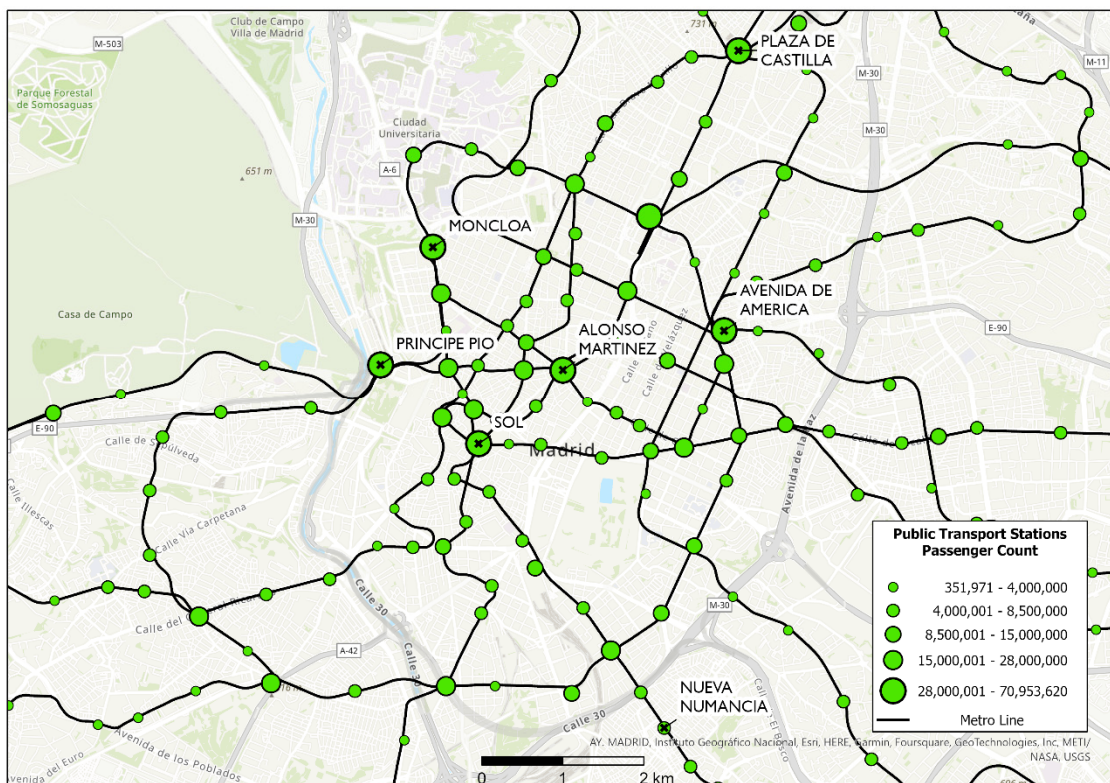

Figure 19. Number of passengers at main public transport stations in Madrid (analysed stations are marked and labelled).
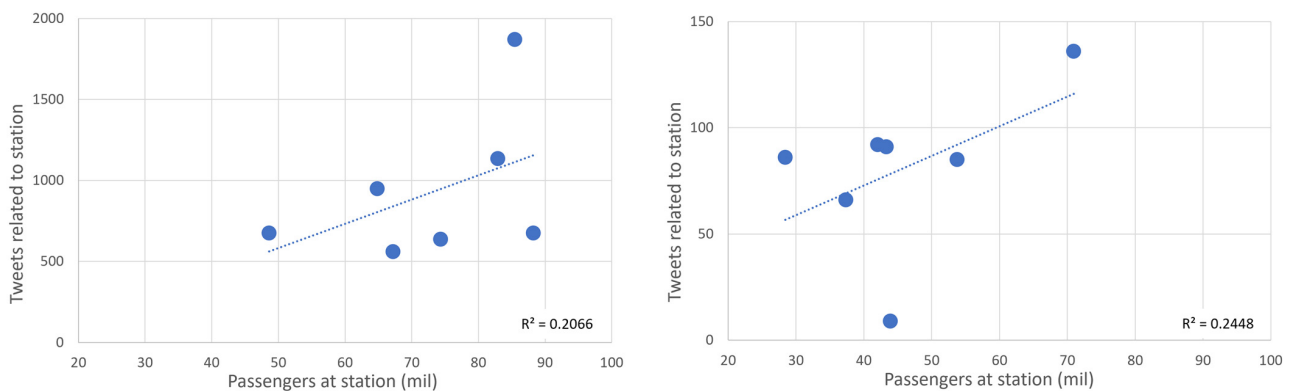
Figure 20. Relationship between number of tweets and number of passengers at stations for London (**left**) and Madrid (**right**).

Lorenz curves document the heterogeneity of a distribution [45]. Lorenz curves for the busiest stations (Figure 21) show that the distribution of tweets among major stations varies between almost equal to highly uneven. Only the top seven stations were evaluated to keep the conditions equal for each city due to a limited number of tweets from Prague. The closest to equality is Madrid where 60% of tweets refer to 55% of stations, while in London they refer to 42% and in Prague only to 28% of tweets. The uneven distribution in Prague documents large differences between the top stations.
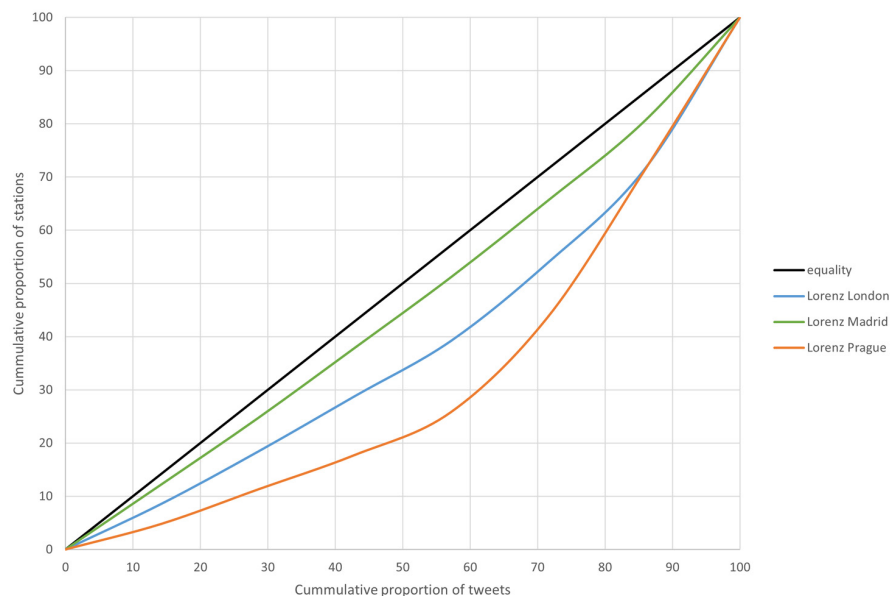
Figure 21. Lorenz curves for stations in London, Madrid and Prague.

4. Discussion

The course of transit-related tweeting in the explored time period can be characterised as stable, modulated by weekly and daily cycles and sporadic peaks. The influence of extraordinary events (lasting from several hours to one day) is confirmed. The detection of such events is an important goal of Twitter analysis [46,47]. Such events are often identified as public service problems or urban events related to traffic and security [29]. In the studied cities, the following types of peaks were distinguished: holidays with increased mobility (Madrid), sport events (Madrid), transport problems (London), behaviour of staff (London) and new measures and rules (Prague). It is interesting that each city showed different behaviour and a massive tweet volume is triggered by different types of events. However, the time period should be lengthened to validate such findings.

The daily course of tweeting activity has been discussed by many authors. Agarwal and Toshniwal [1] discovered in tier-one cities in India (Mumbai, Delhi, Hyderabad, Chennai, Kolkata, and Bengaluru) that the morning peak is between 8:00 and 10:30 and the evening peak is between 16:30 and 18:00. The tweeting behaviour there is explained by [33] who distinguished three time-blocks: (a) people wake up in the morning and activities increase throughout the day, (b) activities decrease after getting off work followed by a steady decline into the night, and (c) activity reaches its lowest point during the midnight hours when most people are sleeping. They documented a steep rise in tweeting in the early morning (7:00 to 8:00), and large decreases during evening commuting hours (17:00 to 18:00), which can be interpreted to mean: (a) users tend to post multiple tweets early in the morning and (b) the evening commute is the least favoured time for posting multiple tweets. It was observed that users often tweet when they are stuck in traffic congestion, however, some tweet during the working hours as well [1].

A steep morning rise in tweeting can be documented in our study as well. In daily courses, the biggest peak in Prague is at noon and might be caused by the fact that Czech people use their mobile phones more during their lunch break and have time to complain about morning events on Twitter. Similar activity around noon is also seen in Madrid, which was confirmed in the pre-pandemic period as well. Nothing similar can be found in our results for the London daily course; London tweeting activity in the pandemic period usually peaked in the late morning rush hours, and in the morning and afternoon rush hours in the pre-pandemic period.

Large decreases in tweeting activity were found right after the evening commuting hours. The case might be that the focus was on transport tweets and people generally tweet about transport while commuting home in the evening. Haghighi et al. [48] pointed out the late-night peak for tweeting about quality of transport service in the Salt Lake City region when people might have more free time to talk about their experience during the day. In the study of London, Madrid and Prague, the existence of a late-night peak is not confirmed. In London, peaks are mostly in the morning until noon as people get to the work and discuss experiences or issues while commuting. The afternoon peak occurs in London at about 17:00 and in Madrid at about 20:00 in the pandemic period.

Haghighi et al. [48] also found that peak periods for tweet frequency do not coincide with transit service peak periods. The daily rhythm of public transport usually shows a typical course between 4:00 and 23:00, with almost equivalent morning and afternoon peaks no matter whether sensor data is used [49] or whether public transport timetables are analysed [50]. The course of transit related tweeting is different.

In the weekly cycle, the most activity in London and Prague was found to be in the middle of workdays, while in Madrid it was found to be only in workdays close to weekends. This can indicate more discussion in Madrid about travelling to leisure activities close to the weekends, partly emphasised by home office time usually also taken close to weekends. Contrarily, Twitter users in London or Prague do not tweet in a similar way in such volume.

Transit-related tweeting activity in all explored capitals is substantially decreased during the weekend. Even though people are less busy during the weekend and have more time to spend scrolling and discussing on social networks, the frequency of transit-related tweets is low. This finding is different from the study of Agarwal and Toshniwal [1] who reveal that tweet frequency is relatively higher during weekends than the weekdays. It could be concluded that the activity is not lower in general because people tend to discuss more on weekends, however, the tweeting is shifted towards other topics than public transport.

The topics identified in the transit tweets can be organised into two groups based on international comparison. General topics such as COVID-19, Delay, Transport Card, Officials, Activists and Staff play important roles across all cities. These topics partly overlap with those from a previous study of Madrid metro tweeting [36]. Punctuality of transport in that study is almost equal the topic of delay used in this study. The Comfort

topic was analysed more deeply to distinguish various specific reasons for discomfort in public transport such as air quality, cleanliness, and racism. The Breakdowns topic is related to the Maintenance topic in our study, classified as a specific topic, due to the ability to cover various issues with transit service. Our hypothesis was that maintenance would be one of the most discussed topics in London as the traffic and public transport services are several times bigger than in Prague and such large transportation systems require a lot of maintenance and exclusions. In fact, the number of tweets about maintenance in London was low, indicating a good level of maintenance in London where people complain relatively less about it. However, the problem may be partly overshadowed by COVID-19. The last topic from Arjona et al.'s previous study, Overcrowding, could have also been shifted to complaints about COVID-19 as people more often mentioned masks and safe distance rather than not enough space and overcrowded premises.

Some topics have different meanings in different cities, e.g., "Staff". "Staff" in London is linked mostly with COVID-19 where people complain about staff not obeying the pandemic precautions and rules. On the other hand, in Prague, this topic is clearly only related to staff sensu stricto (working, inspecting validity of tickets) with outlying minor discussions of COVID-19. In Madrid, "staff" is most related with services (usually complaining about the lack of staff) and maintenance works in the stations.

The share of COVID-19 discussion in cities is different. The highest share is in London. Reasons can be found in transit loads and overcrowding in London public transport due to the substantially higher number of inhabitants (14 million) in comparison to Madrid (6.4 million) and Prague (1.3 million). Overcrowded vehicles, as well as stations, result in a lower possibility to keep safe distances and a higher chance of meeting maladaptive people without masks. Another reason might be that Madrid ridership dropped to 5% of its regular operative numbers [16], thus less people were discussing COVID-19 related to public transport. The same situation occurred in London where, at the end of March, ridership in the London underground dropped to 5% and stayed under 10% until the middle of June [51]. In Prague, ridership dropped at the beginning of pandemic to 17% [52].

When the activity profiles in general are summarised, the main tweeting activity hours range from 8:00 to 10:30, where almost all topics are frequently discussed. The biggest time shift between activity peaks in cities in the pandemic period is documented for delay complaints. High tweeting activity about delays occurs in Madrid and Prague from 7:00 to 9:00, contrary to London, where it is at its highest around 10:00. The early start of discussion may be linked with ridership pattern changes due to the pandemic [16] as well as intervals of connections not being adapted, such as early in the morning. In the evening hours, the activity in Madrid and London tends to decrease whereas in Prague the activity spikes, probably caused by complaints from young people trying to get to the centre. Discussion about staff in each city is highest around noon. The secondary peak in Prague around 16:00 may be caused by office workers complaining about the transport system staff on their commute home. The tweeting course for officials and activists is rather interesting in Madrid and Prague. Discussions about officials persist from morning to late evening (22:00) at about the same pace, while, contrarily, discussions targeting activists, which start at the same time, rapidly decrease after 14:00.

A kind of "dark hour" can be detected at 20:00, when maximal profanities in London and racism in Madrid occur. This may be linked with a different demographic of transit passengers during this time. Frequency of these topics increased several times in the pandemic period compared with the pre-pandemic period.

In Prague, Twitter is used mainly to deliver messages to politicians and main stakeholders, and is much less common in reporting issues to staff or discussing topics within a community. The share of announced issues in Prague is quite low, which can be interpreted in either of two ways: (a) the given problem is smaller than in London or Madrid, which is probably true for overcrowding or delays but hardly for COVID-19; (b) the Twitter users do not report or discuss these issues because they do not expect any feedback and complaints

seem to be useless. We believe this is the case for discussion of COVID-19 in Prague's public transport system.

5. Conclusions

Public transport systems in large cities require various sources of information to provide punctual, reliable, safe and pleasant transport for citizens. Assembling and analysing customer feedback is essential for the improvement of the service and attracting more travellers, which in turn helps to mitigate negative impacts of individual transport and contributes to the sustainability of cities.

Twitter enables the collection of a large amount of transit-related user messages with its freely available API and continuous monitoring of the traffic, various situations and events reflected in transport issues. Understanding of users' behaviour and early warnings are two of many reported benefits of this system.

This study aimed to compare temporal and spatial differences in activity and topics regarding public transport during the COVID-19 period in London, Madrid and Prague. The analysed Twitter data includes direct replies to the public transport service accounts in London, Madrid and Prague, collected in the first part of the pandemic period between 26 March 2020 and 31 January 2021.

Exploring the overall course of tweeting activity, several events were revealed to increase transit tweeting activity. Surprisingly, activity in each city was affected by different types of events. In Madrid, common events such as holidays and sport matches with indirect impact to public transport due to higher spatial mobility were distinguished. London's peaks, however, were linked directly to transit issues and, in Prague, new transport rules elicited large discussions on Twitter.

In contrast to some previous results, the transit tweeting activity in our cities declined over weekends. However, London shows much less of a decline than Prague or Madrid, where the activity drops to almost a quarter of average weekly activity.

Speaking about the daily rhythm in the pandemic period, tweeting activity starts, in general, at 5:30 as people begin to commute to work or school. Maximal activity is reached between 8:00 and 13:00. The usual afternoon decline is shaped differently in each city. In Madrid the activity decreases with insignificant peaks in the evening at around 20:00. In London, the afternoon peaks occur earlier (around 17:00) and are even smaller than in Madrid. In Prague there are no significant afternoon peaks. The patterns differ from those in the pre-pandemic period mainly in Madrid, where strong morning and afternoon peaks are situated at 6:00 and 13:00, respectively.

The topics in tweets were identified using the 'bag of words' approach and *tidytext* and *dplyr* packages of R. Two sets of topics were established. General topics are well-saturated in tweets for each city. They include COVID-19, Delay, Transport Card, Officials, Activists and Staff. Specific topics are limited to one or two cities or only to some stations.

COVID-19 is a dominant topic for both London and Madrid, while in Prague it is less discussed, probably due to Twitter users not believing they would receive feedback. COVID-19 also indirectly influences the meaning of other topics. Contrary to previous research [36], we were unable to follow overcrowding as its own topic due to links to fear of infection, mask wearing, etc. Another significant difference between cities was recognised in the Delay topic. Complaints about delays start in Madrid and Prague much earlier than in London.

Results indicate that discussions about topics with high impact on a user's day, such as delays (being late to work or school), are frequent mainly in rush hours when the events are more likely to occur. Contrarily, topics with long-term effects, such as transport payment cards, are discussed throughout the day with almost the same intensity.

Analysing the spatial distribution of tweeting for busiest stations, we found that distribution of tweets among major stations varies between almost equal to highly uneven. Madrid stations are best balanced, where 60% of tweets refer to 55% of stations, opposed to Prague with only to 28% of tweets.

Limitations of the study can be found in several factors. A high dissemination of Twitter usage among young people (i.e., age 20–39) biases results towards the behaviour of the younger population. The assembled sample depends on the technique of data collection, selection of accounts, filtering and pre-processing steps. The collection of tweets containing mentions of official account of the main transit provider does not cover all transit-related tweets. The sample for Prague is quite small. Another problem is the need for better language (slang) processing, namely for non-English environments. The identification of names of stations to provide more effective spatial analysis is not error-prone and the collected samples are only enough for analysis of major stations which limits some related findings.

The main challenges for future research seem to be a better profiling of Twitter users, automatic detection of events (similar to e.g., [53]) and improved geocoding of places and events to support better targeting of taken measures.

The findings could contribute to improving public transport companies' understanding of their system in a holistic way. We believe discussion in social networks represents one of the best sources for customer feedback for various measures and issues in the transport system. Such analyses could be substantially beneficial in unpredictable situations such as the COVID-19 pandemic, which strongly affected people's behaviour.

Author Contributions: Conceptualization, M.Z. and J.H. (Jiří Horák); data curation, M.Z. and J.O.-A.; formal analysis, P.K.; methodology, M.Z. and J.H. (Jiří Horák); software, P.K.; visualization, M.Z.; writing—original draft preparation, M.Z. and J.O.-A.; writing—review & editing, J.H. (James Haworth). All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Data available upon request due to privacy restrictions.

Conflicts of Interest: The authors declare no conflict of interest. The sponsors had no role in the design, execution, interpretation, or writing of the study.

References

1. Agarwal, A.; Toshniwal, D. Face off: Travel Habits, Road Conditions and Traffic City Characteristics Bared Using Twitter. *IEEE Access* **2019**, *7*, 66536–66552. [CrossRef]
2. Almohammad, A.; Georgakis, P. Public Twitter Data and Transport Network Status. In Proceedings of the 2020 10th International Conference on Information Science and Technology (ICIST), Bath, London, Plymouth, UK, 9–15 September 2020; pp. 169–174.
3. Das, R.D. Understanding Users' Satisfaction towards Public Transit System in India: A Case-Study of Mumbai. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 155. [CrossRef]
4. Georgiadis, G.; Nikolaidou, A.; Politis, I.; Papaioannou, P. How Public Transport Could Benefit from Social Media? Evidence from European Agencies. In *Advances in Mobility-as-a-Service Systems*; Nathanail, E.G., Adamos, G., Karakikes, I., Eds.; Advances in Intelligent Systems and Computing; Springer International Publishing: Cham, Switzerland, 2021; Volume 1278, pp. 645–653. ISBN 978-3-030-61074-6.
5. Monachesi, P.; de Leeuw, T. Analyzing Elderly Behavior in Social Media Through Language Use. In *HCI International 2018— Posters' Extended Abstracts*; Stephanidis, C., Ed.; Communications in Computer and Information Science; Springer International Publishing: Cham, Switzerland, 2018; Volume 851, pp. 188–195. ISBN 978-3-319-92278-2.
6. Coto, M.; Lizano, F.; Mora, S.; Fuentes, J. Social Media and Elderly People: Research Trends. In *Social Computing and Social Media. Applications and Analytics*; Meiselwitz, G., Ed.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2017; Volume 10283, pp. 65–81. ISBN 978-3-319-58561-1.
7. Liu, X.; Ye, Q.; Li, Y.; Fan, J.; Tao, Y. Examining Public Concerns and Attitudes toward Unfair Events Involving Elderly Travelers during the COVID-19 Pandemic Using Weibo Data. *Int. J. Environ. Res. Public Health* **2021**, *18*, 1756. [CrossRef] [PubMed]
8. Carvalho, J.M.S.; Faria, S. Social Media Choice of Generations Y and Z in the Portuguese Market. In *Marketing and Smart Technologies*; Reis, J.L., Peter, M.K., Cayolla, R., Bogdanović, Z., Eds.; Smart Innovation, Systems and Technologies; Springer Singapore: Singapore, 2022; Volume 280, pp. 377–389. ISBN 9789811692710.
9. Alshehri, A.; O'Keefe, R. Analyzing Social Media to Assess User Satisfaction with Transport for London's Oyster. *Int. J. Hum.-Comput. Interact.* **2019**, *35*, 1378–1387. [CrossRef]
10. Top 25 Surprising Twitter Statistics UK Edition 2022. Available online: https://cybercrew.uk/blog/twitter-statistics-uk/ (accessed on 20 November 2022).

11. Twitter En España—Datos Estadísticos | Statista. Available online: https://es.statista.com/temas/3595/twitter-en-espana/#dossierKeyfigures (accessed on 20 November 2022).

12. Digital in Czechia: All the Statistics You Need in 2021. Available online: https://datareportal.com/reports/digital-2021-czechia (accessed on 20 November 2022).

13. Hladík, R.; Štětka, V. The Powers That Tweet: Social Media as News Sources in the Czech Republic. *J. Stud.* **2017**, *18*, 154–174. [CrossRef]

14. Más Usuarios y Menos Trenes Hacen Que el Metro de Madrid ya no Vuele. Available online: https://www.abc.es/espana/madrid/abci-mas-usuarios-y-menos-trenes-hacen-metro-madrid-no-vuele-201810121750_noticia.html (accessed on 11 December 2022).

15. Bansal, P.; Kessels, R.; Krueger, R.; Graham, D.J. Preferences for Using the London Underground during the COVID-19 Pandemic. *Transp. Res. Part Policy Pract.* **2022**, *160*, 45–60. [CrossRef]

16. Fernández Pozo, R.; Wilby, M.R.; Vinagre Díaz, J.J.; Rodríguez González, A.B. Data-Driven Analysis of the Impact of COVID-19 on Madrid's Public Transport during Each Phase of the Pandemic. *Cities* **2022**, *127*, 103723. [CrossRef]

17. Lansley, G.; Longley, P.A. The Geography of Twitter Topics in London. *Comput. Environ. Urban Syst.* **2016**, *58*, 85–96. [CrossRef]

18. Wong, S.C.; Teh, P.L.; Cheng, C.-B. How Different Genders Use Profanity on Twitter? In Proceedings of the 2020 the 4th International Conference on Compute and Data Analysis, Silicon Valley, CA, USA, 9 March 2020; pp. 1–9.

19. Perriam, J. A Tweet is not just a Tweet: Public Sector Understandings and Analysis of Social Media Customer Service Data. In Proceedings of the 10th International Conference on Social Media and Society, Toronto, ON, Canada, 19 July 2019; pp. 33–40.

20. Howard, J.M. Trains, Twitter and the Social Licence to Operate: An Analysis of Twitter Use by Train Operating Companies in the United Kingdom. *Case Stud. Transp. Policy* **2020**, *8*, 812–821. [CrossRef]

21. Cottrill, C.; Gault, P.; Yeboah, G.; Nelson, J.D.; Anable, J.; Budd, T. Tweeting Transit: An Examination of Social Media Strategies for Transport Information Management during a Large Event. *Transp. Res. Part C Emerg. Technol.* **2017**, *77*, 421–432. [CrossRef]

22. Casas, I.; Delmelle, E.C. Tweeting about Public Transit—Gleaning Public Perceptions from a Social Media Microblog. *Case Stud. Transp. Policy* **2017**, *5*, 634–642. [CrossRef]

23. Polat, I.; Kocak, B.B. Determination of Twitter Users Sentiment Polarity toward Airline Market. *Pressacademia* **2016**, *2*, 684. [CrossRef]

24. Politis, I.; Georgiadis, G.; Kopsacheilis, A.; Nikolaidou, A.; Papaioannou, P. Capturing Twitter Negativity Pre- vs. Mid-COVID-19 Pandemic: An LDA Application on London Public Transport System. *Sustainability* **2021**, *13*, 13356. [CrossRef]

25. Brzustewicz, P.; Singh, A. Sustainable Consumption in Consumer Behavior in the Time of COVID-19: Topic Modeling on Twitter Data Using LDA. *Energies* **2021**, *14*, 5787. [CrossRef]

26. TwitteR Package—RDocumentation. Available online: https://www.rdocumentation.org/packages/twitteR/versions/1.1.9 (accessed on 20 November 2022).

27. Gong, Y.; Deng, F.; Sinnott, R.O. Identification of (near) Real-Time Traffic Congestion in the Cities of Australia through Twitter. In Proceedings of the ACM First International Workshop on Understanding the City with Urban Informatics, Melbourne, Australia, 22 October 2015; pp. 7–12.

28. Peplow, A.; Thomas, J.; AlShehhi, A. Noise Annoyance in the UAE: A Twitter Case Study via a Data-Mining Approach. *Int. J. Environ. Res. Public Health* **2021**, *18*, 2198. [CrossRef]

29. Gonzalez, M.; Viana-Barrero, J.; Acosta-Vargas, P. Text Mining in Smart Cities to Identify Urban Events and Public Service Problems. In *Advances in Artificial Intelligence, Software and Systems Engineering*; Ahram, T., Ed.; Advances in Intelligent Systems and Computing; Springer International Publishing: Cham, Switzerland, 2021; Volume 1213, pp. 84–89. ISBN 978-3-030-51327-6.

30. Serna, A.; Ruiz, T.; Gerrikagoitia, J.; Arroyo, R. Identification of Enablers and Barriers for Public Bike Share System Adoption Using Social Media and Statistical Models. *Sustainability* **2019**, *11*, 6259. [CrossRef]

31. Congosto, M.; Basanta-Val, P.; Sanchez-Fernandez, L. T-Hoarder: A Framework to Process Twitter Data Streams. *J. Netw. Comput. Appl.* **2017**, *83*, 28–39. [CrossRef]

32. Twitter API for Academic Research | Products | Twitter Developer Platform. Available online: https://developer.twitter.com/en/products/twitter-api/academic-research (accessed on 20 November 2022).

33. Yao, Z.; Yang, J.; Liu, J.; Keith, M.; Guan, C. Comparing Tweet Sentiments in Megacities Using Machine Learning Techniques: In the Midst of COVID-19. *Cities* **2021**, *116*, 103273. [CrossRef]

34. Shalaby, A.; Hosseini, M. Linking Social, Semantic and Sentiment Analyses to Support Modeling Transit Customers' Satisfaction: Towards Formal Study of Opinion Dynamics. *Sustain. Cities Soc.* **2019**, *49*, 101578. [CrossRef]

35. Shinde, T.; Thatte, P.; Sachdev, S.; Pujari, V. Monitoring of Epidemic Outbreaks Using Social Media Data. In Proceedings of the 2021 2nd International Conference for Emerging Technology (INCET), Belagavi, India, 21 May 2021; pp. 1–6.

36. Osorio-Arjona, J.; Horak, J.; Svoboda, R.; García-Ruíz, Y. Social Media Semantic Perceptions on Madrid Metro System: Using Twitter Data to Link Complaints to Space. *Sustain. Cities Soc.* **2021**, *64*, 102530. [CrossRef]

37. Victoria Station Evacuated: Fire Alert Closed Busy London Underground Station—MyLondon. Available online: https://www.mylondon.news/lifestyle/travel/london-victoria-evacuation-live-reports-19139294 (accessed on 20 November 2022).

38. Dopraváček DPP Zavedl u Tramvají Automatické Otevírání Dveří ve Všech Zastávkách, od Pátku Uzavře pro Cestující Přední Dveře. Available online: https://dopravacek.eu/2020/09/09/dpp-zavedl-u-tramvaji-automaticke-otevirani-dveri-ve-vsech-zastavkach-od-patku-uzavre-pro-cestujici-predni-dvere/ (accessed on 20 November 2022).

39. Praha Ukázala První Autobus v Nových Barvách PID, Vyrazí Na Linku 176—Deník.Cz. Available online: https://www.denik.cz/ekonomika/pid-autobus-mhd-logo-barvy-metro-tramvaj.html (accessed on 20 November 2022).

40. Madrid, C.R.D.T.D. Consorcio Regional de Transportes de Madrid—EDM 2018. Available online: https://www.crtm.es/conocenos/planificacion-estudios-y-proyectos/encuesta-domiciliaria/edm2018.aspx (accessed on 21 November 2022).

41. Song, Z.; Cao, M.; Han, T.; Hickman, R. Public Transport Accessibility and Housing Value Uplift: Evidence from the Docklands Light Railway in London. *Case Stud. Transp. Policy* **2019**, *7*, 607–616. [CrossRef]

42. Guo, Z.; Wilson, N.H.M. Assessing the Cost of Transfer Inconvenience in Public Transport Systems: A Case Study of the London Underground. *Transp. Res. Part Policy Pract.* **2011**, *45*, 91–104. [CrossRef]

43. Garcia-Martinez, A.; Cascajo, R.; Jara-Diaz, S.R.; Chowdhury, S.; Monzon, A. Transfer Penalties in Multimodal Public Transport Networks. *Transp. Res. Part Policy Pract.* **2018**, *114*, 52–66. [CrossRef]

44. Romero, F.; Gomez, J.; Paez, A.; Vassallo, J.M. Toll Roads vs. Public Transportation: A Study on the Acceptance of Congestion-Calming Measures in Madrid. *Transp. Res. Part Policy Pract.* **2020**, *142*, 319–342. [CrossRef]

45. Sarabia, J.M.; Castillo, E.; Pascual, M.; Sarabia, M. Mixture Lorenz Curves. *Econ. Lett.* **2005**, *89*, 89–94. [CrossRef]

46. Suma, S.; Mehmood, R.; Albeshri, A. Automatic Detection and Validation of Smart City Events Using Hpc and Apache Spark Platforms. In *EAI/Springer Innovations in Communication and Computing*; Springer International Publishing: Cham, Switzerland, 2020; pp. 55–78. [CrossRef]

47. Pászto, V.; Darena, F.; Marek, L.; Fuskova, C. Spatial Analyses of Twitter Data—Case Studies. In Proceedings of the 14th International Multidisciplinary Scientific Geo Conference (SGEM), Albena, Bulgaria, 17–26 June 2014; pp. 785–792.

48. Haghighi, N.N.; Liu, X.C.; Wei, R.; Li, W.; Shao, H. Using Twitter Data for Transit Performance Assessment: A Framework for Evaluating Transit Riders' Opinions about Quality of Service. *Public Transp.* **2018**, *10*, 363–377. [CrossRef]

49. Kraft, S.; Blažek, V.; Marada, M. Exploring the Daily Mobility Rhythms in an Urban Environment: Using the Data from Intelligent Transport Systems. *Geografie* **2022**, *127*, 127–144. [CrossRef]

50. Osman, R.; Ira, V.; Trojan, J. A Tale of Two Cities: The Comparative Chrono-Urbanism of Brno and Bratislava Public Transport Systems. *Morav. Geogr. Rep.* **2020**, *28*, 269–282. [CrossRef]

51. Vickerman, R. Will COVID-19 Put the Public Back in Public Transport? A UK Perspective. *Transp. Policy* **2021**, *103*, 95–102. [CrossRef]

52. Pražské MHD Ubylo Za Pandemie 40 Procent Cestujících. Nejméně Jezdili Metrem—Aktuálně.Cz. Available online: https://zpravy.aktualne.cz/ekonomika/doprava/prazskou-mhd-loni-vyuzilo-kvuli-covidu-mezirocne-o-asi-40-pr/r~{}3c34872ee9fd11eba1070cc47ab5f122/ (accessed on 20 November 2022).

53. Cheng, T.; Wicks, T. Event Detection Using Twitter: A Spatio-Temporal Approach. *PLoS ONE* **2014**, *9*, e97807. [CrossRef]