# Generating demand responsive bus routes from social network data analysis

Lidia Sala[1], *Steve Wright[2], Caitlin Cottrill[2], Emilio Flores-Sola[1]

[1] Mosaic Factor S.L., Barcelona, Spain

[2] Centre for Transport Research, School of Engineering, University of Aberdeen, UK

* Corresponding author: s.d.wright@abdn.ac.uk

Postal address:
Dr Steve Wright
Research Fellow
School of Engineering
University of Aberdeen,
Fraser Noble Building,
King's College,
Aberdeen, AB24 3UE,
United Kingdom.

**ABSTRACT:**

Many European cities are establishing mandatory obligations for large mobility demand generators such as business and retail parks, tourist sites and events to develop Mobility Management Plans (MMP). Developing MMPs for events with uncertain spatial demand is a particular challenge.

This paper investigates whether reliable demand data can be extracted from mining social network (Twitter) content and using the resulting information to inform the design of commercially viable bus routes from peri-urban areas of Barcelona to a large music event (Canet Rock). Using data from relevant Twitter users, a Twitter influence score was established for each of the 947 municipalities in the Barcelona Region, providing a spatially distributed picture of the demand to attend the event, prior to event ticket purchase. This was used as the basis for planning and delivering 11 new commercially viable event bus routes transporting over 450 additional passengers from peri-urban and more rural areas in the Barcelona Region.

This paper demonstrates that the innovation of information mining from Social Networks can provide better comprehension of the demand to support Mobility Management Planning for large events and can radically improve the ability of bus services to serve demand from peri-urban and rural areas.

Keywords: Mobility demand prediction; Social Media; User-generated data; Demand responsive bus; Mobility management planning

# 1. Introduction

Public transport (PT) services in large cities and their surrounding areas are typically operated on a radial route structure linking peripheries and the metropolitan centre (McLeod et al., 2017). These services are generally designed based on known typical daily demands, resulting in services which predominantly cater to 'peak hour' travel (e.g., morning and evening commute times). This approach means that services are often inflexible and infrequent, particularly at night. As a consequence, for those living in the outskirts of the city and neighbouring towns, the possibility of accessing large demand generators located outside the city centre, particularly if they take place outside of working hours, by public transport becomes challenging or impossible.

As part of the Sustainable Urban Mobility Plan (SUMP) process (May et al., 2017), cities are being tasked with developing Mobility Management Plans (MMP) which bring together a package of measures tailored to the needs of a location that generates significant travel demand. These packages generally include measures to promote and improve the attractiveness of using public transport, providing dedicated buses, cycling, walking, car-sharing, flexible working or a combination of these as alternatives to drive-alone journeys. Where these relate to employment sites or business parks, employee addresses disclose where the demand originates and MMPs can be designed accordingly. For retail parks it is less certain but historic records and surveys can be used to identify where the majority of the demand originates. For tourist sites and non-regular large-scale events it is more complex, as demand is continuously changing and the challenge in providing tailored public transport or other mobility services is that the location of the transport demand is not known in advance.

A potential method for addressing this lack of information on demand to attend non-regular large-scale events is, however, emerging. Widespread uptake of digital social media services (such as Facebook and Twitter) that allow for publicly accessible 'conversations' to take place have significantly increased the data resources available for gauging interest in and travel needs for infrequent events. Effectively utilising this generally unstructured data to plan more efficient travel options, however, requires careful consideration of analysis and interpretation. In this paper, we utilise a case study approach to demonstrate a potential method for integrating social media data with known transport, demographic and geographic characteristics of an area to contribute to the development of targeted transport options for an infrequent event. Focussing on a major music festival (Canet Rock) held overnight on the peri-urban fringe of Barcelona, in Canet de Mar (see Figure 1), this paper investigates whether reliable demand data can be extracted from mining social network content and using the resulting information to inform the design of commercially viable bus routes to the event from peri-urban areas of Barcelona and neighbouring towns/rural areas.

We first provide background on both the use of public transport for purposes of travel to infrequent events and on the use of social media data in transport contexts. We next present an overview of the case study, describing both the Canet Rock festival and the transport constraints that have previously been experienced. This description is followed by discussion and analysis of the process used to generate demand data using data sourced through the Twitter platform, combined with known characteristics of the Barcelona urban and peri-urban region. Finally, we discuss findings and implications for broader application.

# 2. Literature review

*Event travel*

Travel associated with large events has been linked to local increases in congestion, particularly at the start and end times of the event (Leilei, et al. 2012); (Fernando 2019). A combination of large numbers of persons travelling and constrained transport infrastructure providing access to the event, particularly if most persons travel by private vehicles requiring parking (Chang and Lu 2013), contribute to these effects. Public transport, generally utilising buses or park-and-ride systems, is often encouraged to mitigate these challenges (Collins and Potoglou 2019); (Chirieleison and Scrucca 2017). An additional benefit is that for events where alcohol is served, the use of public transport may reduce incidents of impaired driving and increase the safety of event-related travel (Scagnolari, Walker and Maggi 2015); (Evans 2012).

As noted above, however, public transport networks are generally designed to serve known regular demands, often largely reflecting employment patterns and travel between districts containing clusters of known transport generators (such as dense residential, retail or service areas) (Schöbel 2012); (Webster and Bly 1981). Travel services to infrequent or irregular events may be more difficult to plan given the varying characteristics of both the event and attendees (Latoski, et al. 2003). While events that occur on a regularly recurring basis (such as yearly or monthly events) may have associated historic origin-destination data available, these may not be indicative of actual demand as they will only reflect the available transport options (Kuppam, et al. 2011). Forecasting travel demand under special event conditions has seen a great deal of interest in the literature, with (Li, et al. 2017) exploring demand forecasting of subway travel during special events, (Pereira, Rodrigues and Ben-Akiva 2015) examining forecasting of special event demand using web-based data, and (Kwoczek, Di Martino and Nejdl 2014) trialling methods for forecasting special event related congestion. The interest in this area is evident; however, studies related to pre-event demand forecasting of bus travel are limited.

*Social network data in transport planning*

The use of online social network platforms has seen rapid growth in tandem with the increasing adoption rates of smartphones. According to StatCounter, as of July 2019 nearly 79% of Europeans used Facebook, 9% used Pinterest, and just over 5% used Twitter, with the majority of interaction taking place on mobile devices (StatCounter Global Stats 2019). While users of online social media platforms often exhibit demographic skew, with users tending to be younger and some platforms seeing decided gender splits (for instance, Instagram users tend to be more female dominated, while platforms such as LinkedIn and Twitter tend to be more heavily male dominated (The London School of Economics and Political Science N.D.)), the resulting data has become a rich source of information within a number of areas in the transport sector. For example, a number of researchers have explored the use of Twitter for incident detection (Unankard, Li, and Sharaf 2015; Walther and Kaisser 2013; Gu, Qian, and Chen 2016), while others have explored the perceptions of public transport riders based on sentiment analysis of related social media posts (Collins, Hasan, and Ukkusuri 2013; Schweitzer 2014, Candelieri and Archetti 2015). The use of Twitter to create real-time dialogues between transport service providers and travellers has also been explored (Cottrill et al. 2017), finding that Twitter provided a useful and low-cost method of sharing and relaying travel-related information during a large event.

While such research indicates the extent to which social media may be used within the transport sector, it is perhaps more interesting to look directly at how the information and messages collected from social media platforms may be used to fill in existing information gaps related to the demand for travel. (Abbasi et al. 2015), for example, explored the use of Twitter as a complementary data source to conventional household travel surveys, and noted that an added benefit of this data source is the potential to identify and analyse travel patterns of visitors in addition to area residents. Van Eggermond et al. (2017) also used Twitter data to gather insights into mobility patterns and activity locations, finding that the data generated would be unlikely to replace, but could be used to supplement, traditional survey data. Yao and Qian (2020) demonstrated that people's tweeting activities in the night before and early morning are statistically associated with congestion in morning peak hours and hence can be used to build a predictive framework which forecasts morning commute congestion. Hi and Jin (2017) develop a model using Location Based Social Networking (LBSN) service data to accurately determine travellers' trip arrivals and trip purposes. When aggregated, such data can provide a new secondary data source for the estimation of urban travel demand. Cui et al. (2018) explore the potential for predicting both current and next trip purposes using social media data. They developed a model incorporating Google Places and Twitter information that was found to greatly improve the overall accuracy of prediction for certain activities and identify a possible application in activity-based travel demand modelling.

Analysis of social media data has also been used to improve management and control of the New York City subway network during sporting events. Zhang et al (2016) conducted a study using Twitter hashtags to predict the subway passenger flow generated for major baseball games. This work showed that the prediction of the subway passenger flow is improved in both accuracy and precision and suggests social media data can provide a good indicator of public transport demand to events. Kaiseret al. (2017) highlighted the possibilities of predicting attendance at large scale events from analysis of social media data to help bus and shuttle companies to plan and advertise collective transport services to the event; however, no application of this was demonstrated. Gal-Tzur et al. (2014) describe the potential for Twitter data to be used in planning and policy making with regards to sporting events with an example using football matches in the UK. De Lira et at. (2019) propose a classification approach to infer event attendance from Twitter posts before the event takes place. Unlike previous work, their approach does not rely on geotagged posts but infers location by classifying the non-geotagged content of the users' posts, thereby including a much larger number of posts to predict user attendance to a given event. Analysing 90K pre-event Tweets relating to the Creamfield music festival resulted in identifying 35,239 Tweets positively suggesting attendance from 3856 users' whose hometown location could be extracted from within their Twitter profile. Analysis of Tweets during and after the event identified a total of 10,788 inferred participants where their hometown information could be obtained from their public Twitter profiles. This showed the predicted demand derived from the pre-event Tweets provided a good approximation of the final distribution of attendees by city. It is suggested that such information could be useful for planning efficient bus routes or ride-sharing transportation services to the event, but as with other studies this is suggested as possible future work.

Such research shows the promise of social media data in transport planning and operations; however, it does not present a perfect option. As noted above, not all persons use online social media platforms and the demographic disparities may introduce unintentional bias into planning and operational decisions (Boyd and Crawford 2012). In addition, generating meaningful information from large amounts of unstructured textual data is not a trivial task (Grant-Muller et al. 2015). Despite these reservations, however, online social media services promise a useful platform for engaging with the public and gathering useful data on travel desires and needs.

## 3. Research questions

The previous section identified a number of studies where analysis of social media data offers promise and potential for identifying demand for new or tailored bus services; however, none of these studies followed up with designing, delivering and evaluating practical solutions. This paper explores a case study which uses data from Twitter to enhance predictions of demand to attend a music festival and uses this enhanced demand estimation to design and deliver tailored bus services to the event.

The specific research questions the paper addresses are:

- Can information mining of Social Networks (Twitter) provide enhanced indication of spatial distribution of demand to attend large scale events?
- Can the social media derived demand data form the basis to establish commercially viable bus routes to large scale events from peri-urban and rural areas?

## 4. Introduction to the case study site

The focus of this work is to reduce transport accessibility barriers for residents in peri-urban areas of Barcelona and neighbouring towns wishing to attend the annual 'Canet Rock' festival of music and Catalan culture, held overnight on the peri-urban fringe of Barcelona, in Canet de Mar, 50km north-east of the city centre (see Figure 1). The Canet Rock festival attracts 25,000 attendees and is held in July every year. A large proportion of the attendees are young people with no access to their own mobility solutions and who are from locations where PT access is poor. Figure 2 shows there is a rail link R1 connecting Canet de Mar to the Barcelona city centre, but the bus network only connects Canet de Mar with Girona to the north and is disconnected from the main population to the south. Additionally, the festival ends on a Sunday, during the holiday period, when rail and bus service schedules are particularly limited. As a result, for many, travelling to the event by car is the only viable option despite the fact that it is more expensive, less sustainable, unsafe (the festival takes place through the night) or simply, no fun. Parking limitations and alcohol consumption during the festival generates further demand and need for PT alternatives. However, the limited PT options available prevent many young people without their own car and who are uncertain about getting a lift from friends or family from purchasing festival tickets.

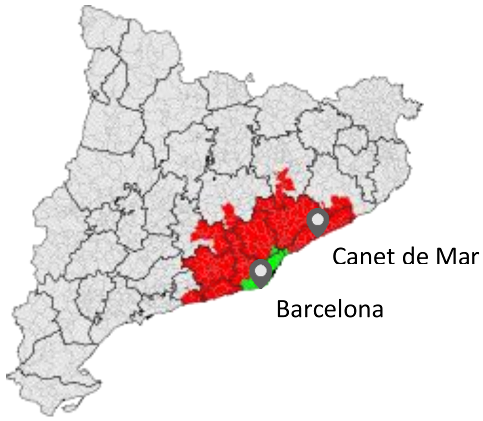


*Figure 1 - Map of festival location in relation to greater Barcelona Metropolitan Region, depicted in red. Green Area represents the first zone.*

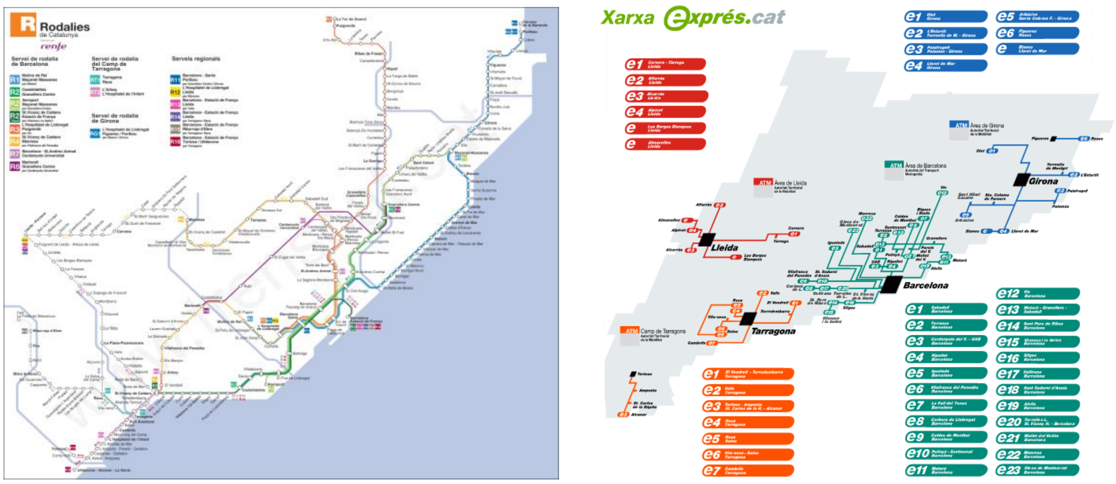


*Figure 2 - Existing public transport services. On the left the rail network and on the right the express bus services.*

## 5. Methodology

The goal of this work is to develop a model for demand prediction that will allow BusUp (a PT service provider) and the event organizer (Canet Rock) to form a better understanding of the latent mobility demand to the event (in terms of geographical location) prior to event ticket purchase. This information is then used to provide more informed design and delivery of tailored bus routes that better meet the determined demand and increase the overall transport accessibility.

The solution applies information mining from Twitter to identify the demand from potential users who want to attend the festival event. It should be noted that the festival attracts predominantly younger attendees from age groups that are also active Twitter users. This approach involves two different types of analysis: first, establishing the Twitter accounts that are relevant to the analysis, which involves identifying Twitter relationships of users connected with the “festival accounts core”; and second, analysing non-structured data of Tweets to infer users who display intention to attend the event. The social network data is then combined with data on demographic distribution (highlighting the key geographic areas in the region), transport connectivity (between each geographic area and Canet de Mar), and historic data of festival attendees from previous years. This analysis enables planners to aggregate the mobility demand (potential event attendees) from different geographic areas which are poorly served by existing PT service provision and use this to propose the most suitable demand responsive tailored bus routes and bus-stop locations for the unserved demand. These three distinct phases are illustrated in Figure 3.

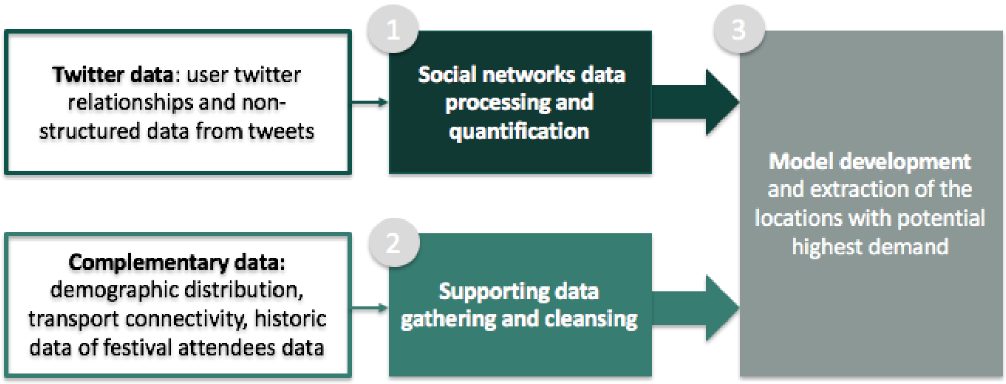


*Figure 3 – Three phases of analysis*

## *5.1 Social network data processing*

One of the key challenges of mining social networks to extract data is to interrogate and interpret the unstructured information content. Therefore, it is crucial to identify the types of information the social media content contains and to understand the knowledge that can be extracted. In this work, Twitter provides the social media content. The main goal is to link all the useful information to the municipalities, to understand the geographic picture of interest in attending the festival and to derive the resulting PT demand.

The first part of the Twitter data analysis consists of identifying the Twitter accounts per municipality which are most relevant to the event. This festival's network graph, Figure 4, has been categorised as the *festival organiser core* accounts and the *festival's influence network* accounts. The *festival organiser core* are all the official Twitter accounts that belong to the festival including the organiser, all the gold and silver sponsors, official media and the artists that performed at the festival the previous year. The *festival's influence network* considered here, takes into account all the Twitter users following the festival organiser core, as well as those users following those users respectively. The level of the network considered is limited to the second order relations.

The festival's network graph is driven by the festival organiser core and its followers. These followers' network relations are weighted depending on the type of users that are followed (Hangal et al. 2010). Firstly, as shown in panel b) of Figure 4, the users that directly follow the festival organiser core (including the event organisation, institutions, event sponsors, or the people that participate actively on it) are selected. Secondly, shown in panel c) in Figure 4, the follower's interrelations are analysed to select the followers of the festival organiser core accounts with the most influence. Followers are associated with a municipality if the location given by their account is accessible, and so each municipality is rated with a Twitter influence score depending on the composition of the follower influence network position. The Twitter influence score for a given municipality $k$ is defined as $T^k = \sum_{m=1}^{N} c_m n_m^k$, where $n$ is the number of followers of type $m$, and $c$ is the weight of each level defined under the restriction of $c_m > c_{m+1}$. The types, as aforementioned, are listed in order of importance from festival organizers, current-year artist repeating, current-year artist, previous-year artist and sponsors.

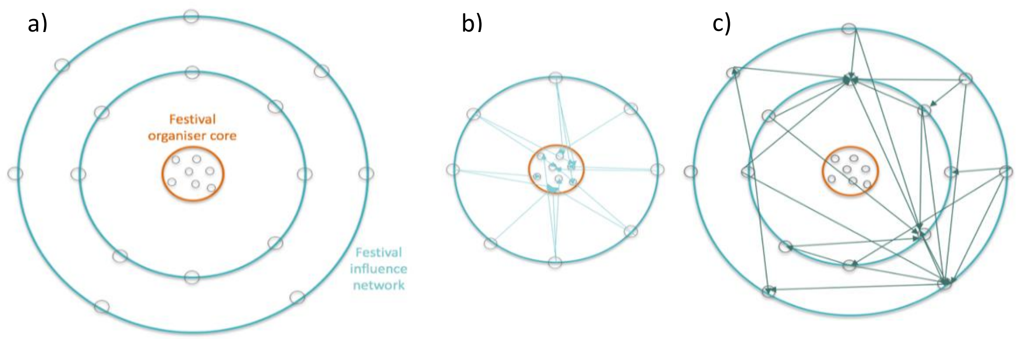


*Figure 4 – Festival's network graph levels and connections.*

The distribution of this Twitter score over the municipalities in Catalonia is shown in **Error! Reference source not found.**.

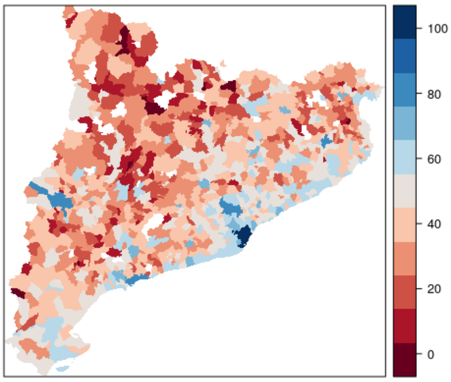


*Figure 5 – Twitter Score distribution, percentage respect to the maximum, for municipalities of Catalunya estimated from Canet Rock social network.*

The second part of the social network analysis is focused on the analysis of non-structured Twitter data obtained from the accounts in the festival network graph. For this exercise, the MORIARTY[1] software, developed by ITAINNOVA[2] (Montañés et al. 2018), was used. MORIARTY detects and stores big data information from Twitter enabling identification of Twitter activity related to general topics in different areas. For this project, specific dictionaries containing CanetRock-related keywords, artists, accounts and hashtags for the current year are set to properly and accurately identify Tweets, which are later collected and stored in a database. The quantity of Tweets, number of retweets and likes generated in different locations through time are analysed. These Tweets are related with the predefined keywords related to *Music* (e.g., artists attending to the festival), *Event* (e.g., sponsors, organisers) or *Transport* (e.g., transport services available to arrive to the festival). These indicators are generated based on a continuous tracking of the Twitter activity. For the planning of bus services to the festival, it is the home address of the Twitter user that is of relevance and not necessarily the geotag location where Tweets were made. Location is firstly derived from the registered location associated with the account. Where this is not available, but geolocation is enabled by the user, then the most frequent geotag location revealed from all Tweets by that user usually indicates their home location.

## *5.2 Model Development*

The model described in this section establishes the bus stop locations most likely to attract passengers for new tailored bus services to and from the event. The model developed is based on two different phases (see Figure 6); the first phase is *data extraction and pre-processing*, while the second is the iterative process of *data analytics and cluster development.* This approach enables planners to aggregate the mobility demand (potential event attendees) from different geographic areas which are poorly served by existing PT service provision and use this to propose the most suitable demand responsive tailored bus routes and bus-stop locations for the unserved demand.

[1] *http://www.ita.es/moriarty/?home.json*
[2] ITAINNOVA is the Brand name for the *Instituto Tecnológico de Aragón.*

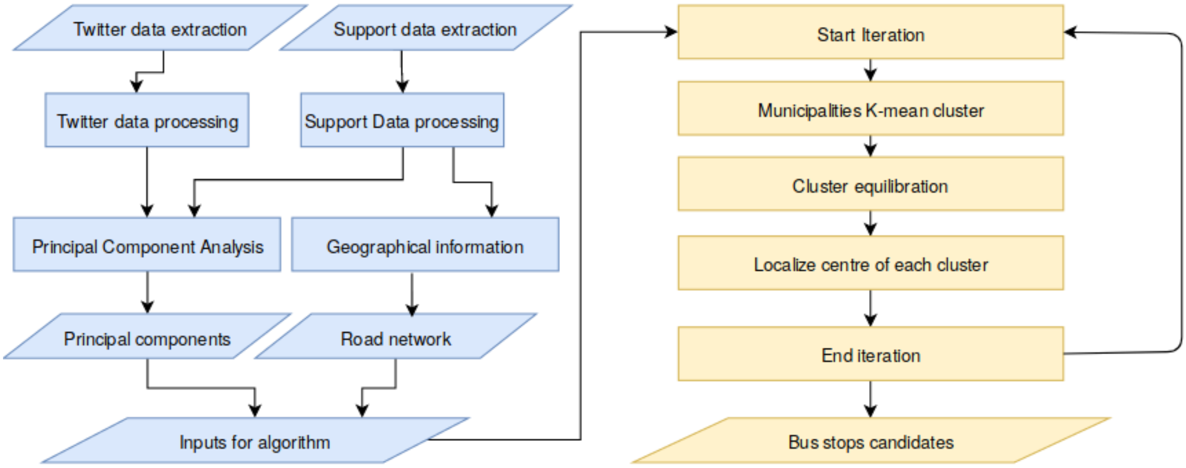


*Figure 6 - Algorithm for searching potential stops.*

In the model development phase, the Social Network data is combined with other data sources to increase the robustness of identifying areas with interest in using PT to attend the event, and then identifying specific locations where demand exists but there is a lack of public transport and other mobility services to reach the event. To gauge how better transport connectivity could affect the attendees' accessibility from each municipality to the location of the event, a comparison of public and private modes of transport has been made through the Google Maps API. To estimate the connectivity level to the event from each municipality, multiple starting and ending times (including night hours) for both modes of transport have been considered, with the shortest travel time by public and private routes selected for the comparison. The resultant connectivity by each mode of transport is plotted in Figure 7, where the values represent the average of all the considered schedules.

Another useful data set considered here is the number of event attendees registered by postal code from the previous year (2018), provided by the Canet Rock organizers, shown on the left-hand side of Figure 8. The data is prepared by associating each postal code with its respective municipality. Supporting data applied by the model includes the distribution of the population aged between 15 and 65 (see image on the right in Figure 8) and the political results of the last elections, due to the political character of the event, which celebrates Catalan culture.

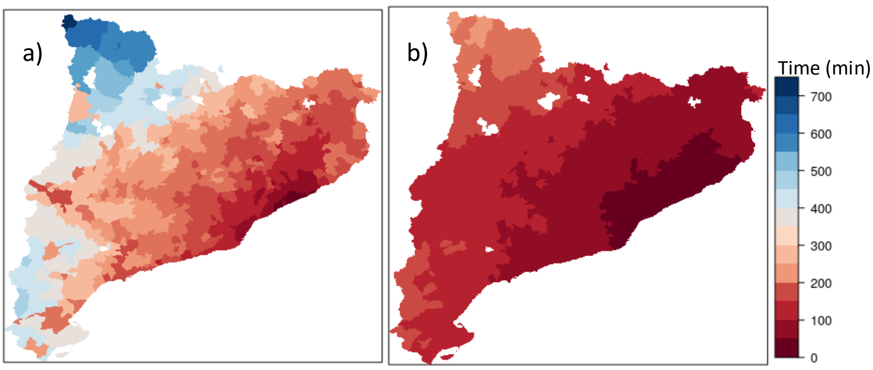


*Figure 7 – a) Travel time by public transport, in minutes, from each municipality to Canet Rock. b) driving time by private vehicle*

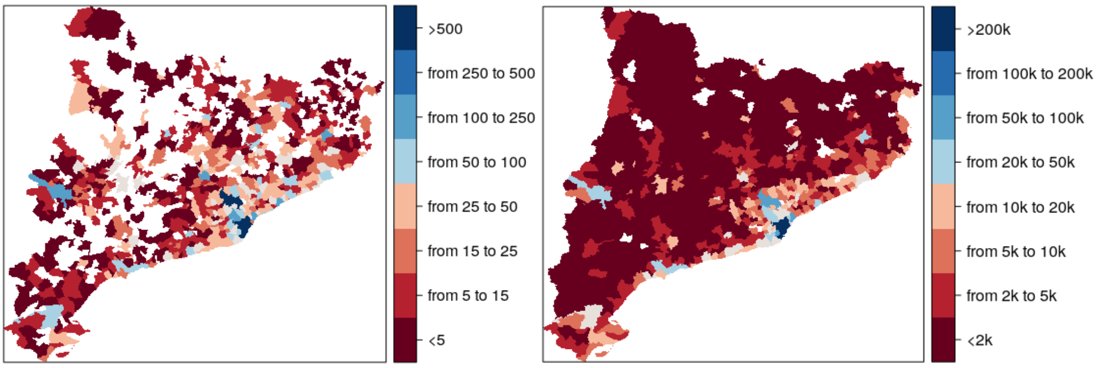


*Figure 8 – Canet Rock 2018 attendees on the left, and population from 15 to 65 years old on the right.*

After completing the data extraction and pre-processing phase, the geographic structure network is created by analysing the road network's connectivity among the municipalities. The nodes are defined by the urban centres, one per municipality, with the geographic location defined by the distribution of the population. The edges that connect the nearest urban centres are weighted by the driving time between two nodes. Each urban centre of all 947 Catalan municipalities will be considered as a possible candidate to locate a route stop.

The algorithm starts by performing a demographic constrained k-mean clustering that is a modified version of a standard k-mean (Forgy 1965; Lloyd 1982). This is based on forcing each cluster to be formed by a group of neighbouring municipalities under the constraint that each cluster requires a minimum number of n past attendees. At this point, the initial clusters are restructured, moving municipalities from one cluster to another, to make them similar in numbers of past attendees using a modified version of a union find algorithm (Newman and Ziff 2001). This restructuring process picks, at random, two adjoining municipalities in different clusters and attempts to reach equilibrium in the number of attendees between neighbouring clusters. A municipality is moved to a neighbouring cluster if the resulting clusters' configuration decreases the difference in number of attendees between them. The model ensures that no municipality becomes physically detached from its cluster and limits the "infinite" expansion of each cluster. This iterative process ends when the number of attendees in each cluster are in equilibrium and the system achieves convergence. Ultimately, 65 clusters were formed.

From there, the geographically weighted mean of each cluster is computed, i.e., the centre of the cluster, using the processed Twitter data and the most relevant information from all other data sources analysed (demographic distribution, road transport connectivity, event attendees from 2018, political results). In relation to this, the following list of metrics are considered for each municipality: population from 15 to 65 years old, population density per km$^2$, number of Canet Rock purchases in 2018, numbers of followers, number of Tweets, Twitter influence score, distance, driving time and transit time in public transport to Canet Rock, political votes related to concert ideology together with metrics derivate from those indicators in densities and absolute values.

*Table 1:List of features per municipality included in the model before applying PCA*

| Features | *Description* |
|---|---|
| *Demographic distribution* | Population from 15 to 65 years old |
| *Population density* | Population density per km$^2$ |
| *Attendees from 2018* | Canet Rock attendees from 2018 |
| *Road transport connection* | Network of travel distance among all the municipalities setting the urban centres as the nodes |
| *Distance to Canet Rock* | Straight distance from Canet to each municipality |
| *Driving time* | Driving time to Canet Rock from each municipality |
| *Transit time* | Public transport transit time (any) to Canet Rock from each municipality |
| *Political Results* | GenCat parliament elections result 2017 |
| *Number of followers* | Quantity of followers of CanetRock influence orbit up to second order |
| *Twitter influence score* | Metric extracted from twitter analysis based on the type of the follower |
| *Number of Tweets* | Number of Tweets related to CanetRock orbit |
| *Miscellaneous density indicators* | Absolute values have been divided by total population since PCA works on linear relations |

To avoid the inclusion of redundant data, the Principal Component Analysis (PCA) technique (Jolliffe 1986) is carried out to understand the model variability and behaviour by identifying the most uncorrelated variables of the data sources analysed. The centre of each cluster is then associated with the municipality that has the closest urban centre. This urban centre becomes a candidate for a bus stop and the municipality it belongs to is recorded. The iterative process of restructuring the clusters is repeated with the municipality at the centre of each cluster recorded in every iteration along with the sum of all estimated attendees (derived from Twitter data analysis) that belong to each cluster. Then, the attraction factor is defined for a given municipality $k$ by $A^k = 1/N \sum_n^N C_n^k$, where $n$ is the $n$th time that the municipality $k$ was selected as cluster centre during the iterative process with the sum of $C$ attendees of the cluster.

So, in summary, the model incorporates a fusion of demographic and historic demand data with Twitter scores to identify the municipalities that feature as cluster centres and hence have the highest potential demand to attend the Canet Rock Festival. In order to identify the bus stop locations with highest predicted demand, the municipalities are ranked according to the frequency with which they feature as a cluster centre and the sum of estimated attendees for their cluster catchment from Twitter data analysis. The top 25% of these municipalities are selected as potential candidates for a bus stop.

## 6. Model Results

This section presents results to answer the first research question on whether information mining of Social Networks (Twitter) provides an enhanced indication of the spatial distribution of demand to attend a large-scale event.

The model has identified 119 municipalities suitable for potential bus stops. These have been categorised by the model as having a high (32), mid (36) or low (51) level of potential demand. From those identified, 20 stops were already stops offering a BusUp service to Canet Rock in 2018, hence the model has identified 99 new municipalities in which to locate a bus stop. Of these, 16 are categorised as having high potential demand, 33 as having mid potential demand and 50 with low potential demand.

To assess the accuracy of the model for estimating the spatial distribution of demand, the predicted attendance from Twitter influence score for each municipality is measured against the actual attendees at Canet Rock 2019 from each municipality. These results are then compared with simply using population as the predictor for attendance.

Table 2 presents the log-log correlation, its coefficient of determination $R^2$ from the log-log fit $\log y = a + c \log x$, for the relation between attendance and population, and attendance and Twitter score respectively. The log-log correlation coefficient describes the strength of the relationship between the key parameter (i.e., Twitter score or population) and the actual attendees, with values closer to 1 reflecting a stronger (linear) relationship. The coefficient of determination ($R^2$) gives a measure of the 'goodness of fit' of the model and indicates the extent to which variation in demand to attend the festival can be explained by the key parameter (i.e., Twitter score or population). Again, the closer to 1, the better the key parameter at estimating demand to attend the festival. It can be seen from these results in Table 2 that Twitter score is superior to population variable when describing the attendance to Canet Rock 2019. This is observed, either considering all municipalities, including those not classified as candidates for a new bus stop, or only considering candidates for a potential new stop with an improvement of the coefficient $R^2$ of 7.4% and 67.8% respectively, as indicated in the last column of Table 2. What is most notable is that the Twitter score correlates much better with attendance than population does for the municipalities categorised as having low and mid potential demand. This suggests that use of Twitter data for estimating demand is especially useful in identifying areas with lower population but where there is relatively high demand to attend the event.

*Table 2 – Accuracy of using either municipality 'population' or municipality 'Twitter score' as basis for estimating demand from each municipality to attend the event (values of log-log correlation closer to 1 represent stronger linear relationship between population or Twitter score and attendance; value of $R^2$ closer to 1 reflect better 'goodness of fit' of population or Twitter score to estimate attendance)*

| | *population* | | *Twitter score* | | |
|---|---|---|---|---|---|
| *Level* | **Log-log correlation** | **$R^2$** | **Log-log correlation** | **$R^2$** | **$R^2$ improvement** |
| *All* | 0.770 | 0.592 | 0.798 | 0.636 | 7.4% |
| *Low* | 0.256 | 0.045 | 0.594 | 0.339 | 653.3% |
| *Mid* | 0.392 | 0.135 | 0.523 | 0.249 | 84.4% |
| *High* | 0.752 | 0.554 | 0.805 | 0.636 | 14.8% |
| *Candidates* | 0.622 | 0.382 | 0.803 | 0.641 | 67.8% |

The plots in Figure 9 contain a single dot for each municipality: blue diamonds denote municipalities containing bus stops from last year; green circles denote municipalities with high modelled potential demand (1) for new bus stops; orange triangles denote municipalities with medium potential demand (2) for new bus stops; red squares denote municipalities with low potential demand (3) for new bus stops; and black dots represent municipalities not selected as potential candidates for a bus stop. As the model already takes into consideration factors such as historic demand, the proximity of potential stops and the existing connectivity to the concert location, the selection of municipalities to receive a bus stop is not simply based on those with highest predicted attendees based on Twitter score. This is why some municipalities with a higher Twitter score, but which are already well connected, only have a medium modelled demand for a bus stop (orange triangles), while other municipalities with a lower Twitter score, but which are less well connected, have a high modelled demand for a bus stop (green circles).  In other words, the model accounts for a higher proportion of the predicted attendees from better connected municipalities using existing PT connections to access the festival. As a result, there is considerable overlap in the position of the plotted points for each bus stop demand category (high - green circles, medium - orange triangles, and low - red squares) in the right hand side of Figure 9, as there is not a clear distinction in the demand for new bus stops based on Twitter score.

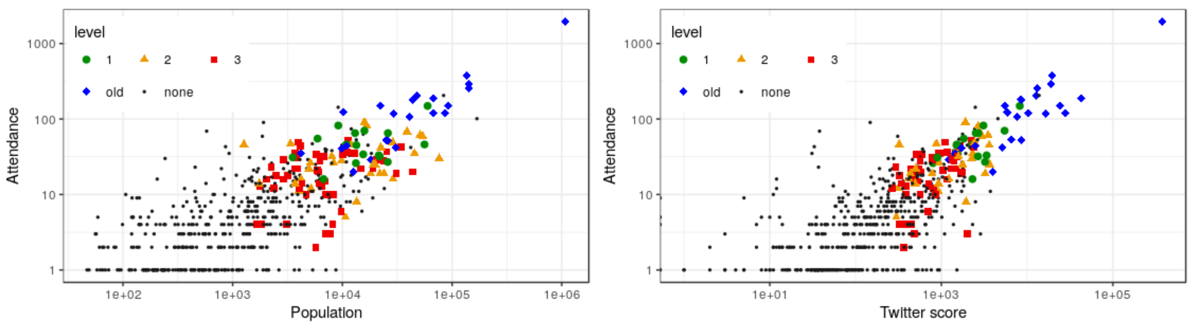


*Figure 9 – Attendance as a function of population, left, and Twitter score, right. Colours denote potential level of demand, green (1 - high), orange (2 - medium), red (3 - low).  Blue diamonds denote stops from last year, and black dots represent municipalities not selected as a potential candidate for a bus stop.*

Although Twitter scores proved to be better than population at estimating attendance ($R^2$ values closer to 1 in Table 1), some municipalities with high Twitter scores showed low attendance and so would not be suitable bus stop candidates. This weakness is overcome within the model by integrating the demographic and historic demand data (attendance by postcode for 2018) with the Twitter score in the clustering process as described in Section 5.2. The function describing this fusion of demographic and historic demand data with future demand derived from Twitter data analysis results in a parameter termed the 'attraction factor' and further enhances the estimation of demand providing a more consistent fit to attendance than using the Twitter score by itself.

Incorporating this fusion of demographic and historic 2018 attendance data with Twitter scores improves the coefficient of determination ($R^2$), giving a measure of the 'goodness of fit' of the model: from 0.592 based on population; and 0.636 based only on Twitter score; to 0.86 based on attraction factor (fusion of Twitter score with historic demand and demographic data).

In order to test the contribution of each of these parameters to the attendance, multiple linear regression is applied following a log-log model. This is given by the following function

$$\log y_{at} = a + c_{tw} \log x_{tw} + c_{af} \log x_{af} + c_p \log x_p$$

Where $y_{at}$ is the attendance, and $x_{tw}, x_{af}, x_p$ corresponds to the Twitter score, attraction factor and population, respectively. To reveal the most significant parameter to the attendance we compare the model results with all parameters ($x_{tw}$, $x_{af}$, $x_p$) included to the results obtained when removing each parameter, one at a time. Applying this methodology, the parameter which results in the most negative variation in the $R^2$ value indicates the parameter with most importance for the fit. Table 3 presents the results for all the municipalities, those selected as potential candidates (high, medium and low levels). It also considers results when only considering the bottom 50% of municipalities in terms of connectivity to the festival venue, as well as the bottom 50% of municipalities taking account of connectivity and population.

*Table 3 – Coefficients of determination for the log-log multi linear regression. There are several fits removing concrete variables.*

| | | *log-log Attendees fit* | *Removing Population* | | *Removing Twitter score* | | *Removing Attraction factor* | |
|---|---|---|---|---|---|---|---|---|
| *level* | **subset** | **$R^2$** | **$R^2$** | **$\Delta R^2$ %** | **$R^2$** | **$\Delta R^2$ %** | **$R^2$** | **$\Delta R^2$ %** |
| *All* | - | 0.860 | 0.845 | -1.7 | 0.840 | -2.6 | **0.819** | **-4.7** |
| *Potential* | - | 0.853 | 0.845 | -1.0 | 0.815 | -4.7 | **0.800** | **-6.2** |
| *All* | 50% less connected | 0.872 | 0.867 | -0.6 | 0.847 | -2.8 | **0.836** | **-4.1** |
| *Potential* | 50% less connected | 0.844 | 0.845 | 0.1 | **0.776** | **-8.1** | 0.816 | -3.3 |
| *All* | 50% less connected & populated | 0.722 | 0.713 | -1.2 | 0.694 | -3.9 | **0.688** | **-4.7** |
| *Potential:* | 50% less connected & populated | 0.842 | 0.847 | 0.6 | 0.790 | -6.1 | **0.766** | **-9.0** |

The results of the variation of the $R^2$ indicates that attraction factor is the most important parameter in almost all the subsets examined. This is to be expected as it draws on data related to demography, historic demand and Twitter score. Removing only Twitter score also obtained significant variation showing the importance this parameter in the estimation of demand, especially for less connected and less populated municipalities. This emphasises the value of Twitter data in identifying where there may still be sizeable demand from less densely populated peri-urban and rural areas which would otherwise be missed.

## 7. Delivery of new bus routes

This section focusses on the second research question, namely, whether the social media derived demand data can form the basis to establish commercially viable bus routes to large scale events from peri-urban and rural areas.

Sections 5 and 6 have shown how Twitter data analysis can be used to enhance the prediction of potential demand to the Canet Rock festival. The best approach for predicting potential demand involved the combination of demographic data, historic attendance data and Twitter data (attraction factor selection) to identify the 99 bus stop locations with the highest predicted demand for new public transport services to the event. The locations of these are illustrated in Figure 10, which also highlights the 16 bus stops with highest potential (green dots), 33 bus stops with medium potential (orange dots) and 50 bus stops with lower potential (red dots). Blue dots show the locations of bus stops which have been established and retained from previous Canet Rock events.

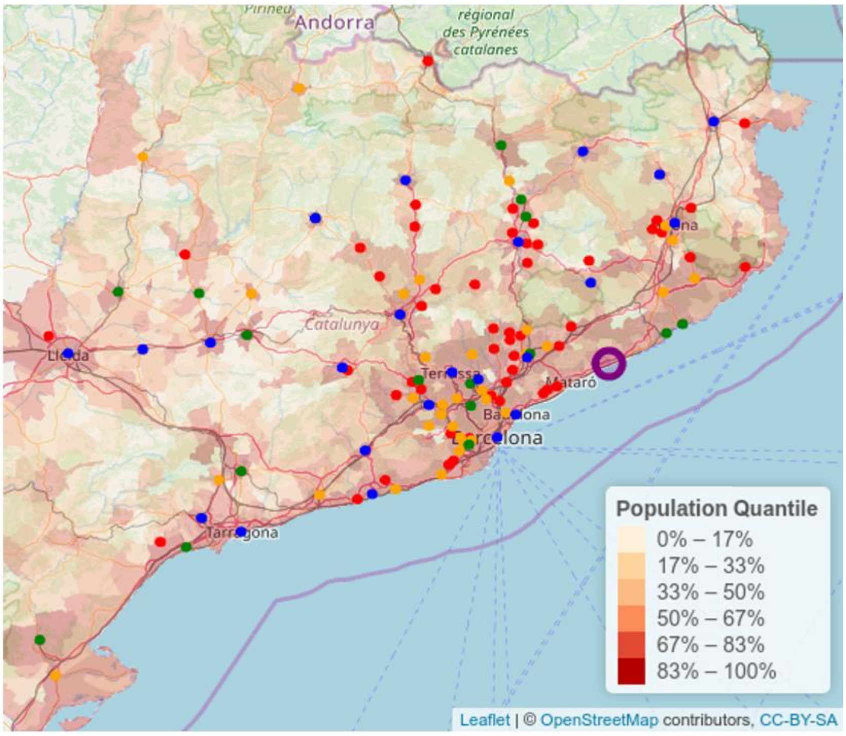


*Figure 10 - Road network and location of the stops for BusUp services to Canet Rock 2019 (location depicted by a purple circle).*

The process to establish the bus routes to adequately match the offer of bus services to the predicted demand to attend the festival (based on attraction factor selection) needs to take account of several factors. These include the location of stops, the driving time distances between them, the expected demand from each stop, as well as logistics and business considerations. For instance, knowledge of the budget available and cost of hiring different sized buses is required along with the number and availability of buses of differing size. This is undertaken by BusUp, which acts as a broker to hire bus service provision for the event from local bus and coach operators. A set of core routes are identified by BusUp, each of which have a minimum predicted demand of 50 passengers. Having established the core routes, these are then publicised to potential festival attendees via the festival website and Twitter account. Those wishing to book travel are then directed to the BusUp bus ride-sharing App/platform to complete the booking process. The ticket prices for passengers vary depending on the driving time and distance with the BusUp services on average costing passengers 0.17€/km, compared to an average cost by car of 0.35 €/km. So, unless sharing lifts, the cost of travelling by car is around double that of using the BusUp bus service. The cost of parking at the event can also add further cost to driving.

Of the 16 high potential stops output by the model, based on 'attraction factor' selection, 10 of these were published as new stops and made available to receive bookings. Of the 33 stops identified by the model as having medium potential, 11 were published as new stops and made available to receive bookings. Finally, out of the 50 high potential stops output by the model only 4 were published as new stops and made available to receive bookings. In addition to this, there were a total of 20 'old' stops (i.e., those established in previous years) which were published and made available to receive bookings. The reduction in new stops between those output from the model and those chosen to be published for booking is the result of filtering out stop locations where there is already some form of reasonable public transport access to the event location. Most of these fell in the more urban areas of Barcelona and large towns such as Girona (see Figure 2 for maps of existing PT network).

The majority of hired buses allocated to the core routes contain 52 seats. A route is deemed to be economically viable once a minimum vehicle occupancy of 70% is achieved. If all 52 seats become booked and expected demand can justify a second bus for the route, then this is hired. If the 70% occupancy is not achieved 2 weeks prior to the event (2-week cut-off), then the route is cancelled; however, it will be checked if a smaller vehicle can be hired which is economically viable with lower numbers of passengers.

Table 4 provides an overview of the high, medium and low potential stops which were a) ‘proposed’ by the model; b) ‘published’ to receive bookings; and c) ultimately ‘operated’ within commercially viable routes. We see that of the new stops which were published, 5 out of 10 high potential stops had received sufficient bookings 2 weeks prior to the event to be commercially viable (i.e., tickets sold > 70% of bus capacity). Only 3 out of 11 medium potential stops and none of the 4 low potential stops received sufficient bookings to be economically viable. Meanwhile, 19 of the 20 old stops which were published received sufficient bookings to be viable. The location of the ‘operated’ bus stops served by BusUp services to Canet Rock 2019 are shown in Figure 11.

*Table 4 – Proposed, published and confirmed (operated) stops for services to Canet Rock 2019*

| Demand levels stops | Proposed | Published | Published/ proposed % | Operated | Operated/ proposed % | Operated/ published % |
|---|---|---|---|---|---|---|
| New stops (added for 2019) | 99 | 25 | 26.00% | 8 | 8.08% | 32.00% |
| High – new | 16 | 10 | 62.50% | 5 | 35.29% | 50.00% |
| Mid – new | 33 | 11 | 33.33% | 3 | 9.09% | 27.27% |
| Low – new | 50 | 4 | 8.00% | 0 | 0.00% | 0.00% |
| Existing stops (repeated from 2018) | 20 | 20 | 100.00% | 19 | 95.0% | 95.00% |
| High – old | 16 | 16 | 100.00% | 16 | 100.00% | 100.00% |
| Mid – old | 3 | 3 | 100.00% | 2 | 66.67% | 66.67% |
| Low – old | 1 | 1 | 100.00% | 1 | 100.00% | 100.00% |
| **Total** | 119 | 45 | 37.82% | 27 | 22.69% | 60.00% |
| High | 32 | 26 | 81.25% | 21 | 65.63% | 80.77% |
| Mid | 36 | 14 | 38.89% | 5 | 13.89% | 35.71% |
| Low | 51 | 5 | 9.80% | 1 | 1.96% | 20.00% |

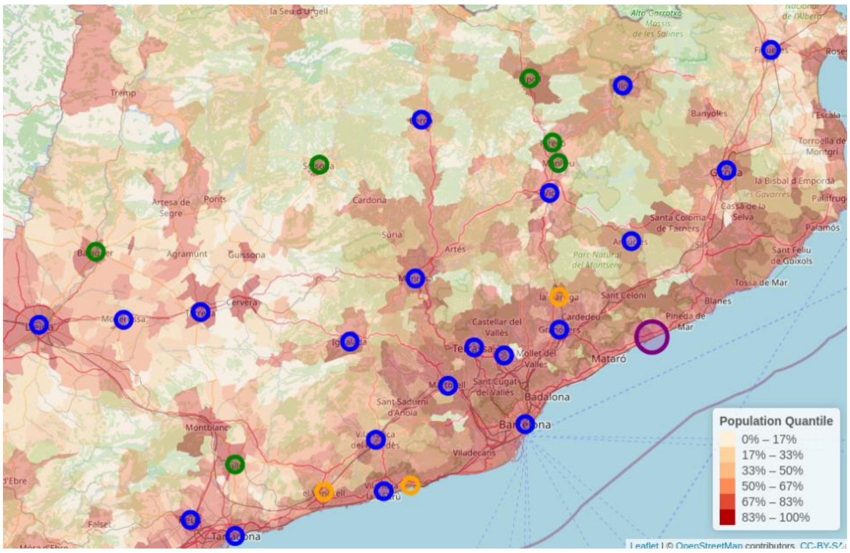


*Figure 11 – Confirmed bus stop locations where BusUp services operated overlayed on population density. Blue circles denote old stops, green and orange denote high and medium demand for new stops. Purple circle marks the location of the Canet Rock festival.*

Overall, there was a 42% increase from last year on the number of bus stops served (see Table 5), with the majority of these operating as direct routes between the new stop and the event. This suggests the model successfully identified, from the outset, stop locations with sufficient demand to fill buses without the need to divert or stop en-route. The number of tickets sold increased by 37%.

*Table 5 – Summary of the BusUp services operated to Canet Rock festival*

| | *2018* | *2019* | *% change* |
|---|---|---|---|
| *Number of Bus Stops* | 19 | 27 | +42% |
| *Number of Routes* | 24 | 35 | +46% |
| *Number of Direct Routes* | 4 | 14 | +250% |
| *Bus Tickets sold* | 1217 | 1666 | +37% |

The majority of the existing BusUp routes (developed through traditional demand estimation based on population and previous attendance) are serving more densely populated urban areas or are intermediate stops on a route serving a densely populated urban area. For the 19 municipalities served by the existing BusUp routes the average density of population is 2100 persons/km$^2$, and all stops on direct routes have density of 1200 persons/km$^2$ or above.

It is notable that the majority of new stops (illustrated by the green and yellow circles in Figure 11), suggested by incorporating the Twitter influence score into the stop selection, resulted in commercially viable direct routes from less densely populated areas. For the 8 municipalities served by the new BusUp routes the average density of population is 600 persons/km$^2$ and all have a density of 1000 persons/km$^2$ or below.

This confirms that deriving estimated demand from social media data can form the basis to establish commercially viable bus routes to large scale events from peri-urban and more rural areas.

## 8. Discussion and conclusion

The results presented in this paper provide an evaluation of the effectiveness of the demand estimation, prior to event ticket purchase, from social network data (Twitter) as a basis for identifying suitable bus stop locations in order to generate routes for new bus services to large scale events such as music festivals.

The model validation compared the predicted demand with the real demand for Canet Rock 2019. The predicted demand, according to location, was computed by analysing non-structured data of Tweets from accounts of users who connected with the festival core accounts to infer users who display intention to attend the event. From this, a Twitter influence score can be established for each of the 947 municipalities in the Region to get estimates of the spatially distributed demand to attend the event. The real demand was provided by the database of sold tickets for Canet Rock 2019.

The results showed that the Twitter influence score is superior to population as a predictor of attendance, particularly in identifying peri-urban and more rural areas with lower population but where there is relatively high demand to attend the event. The additional knowledge gained from Social Media (Twitter) data has led to the expansion of the BusUp service offering pre-bookable bus services to the Canet Rock event. BusUp provided services from 19 stops in 2018. The location of these stops, established through traditional demand estimation based on population and previous attendance, are serving more densely populated urban areas, or are intermediate stops on a route serving a densely populated urban area. The demand estimation from Twitter data resulted in publication of 25 potential new stops for 2019 which the demand analysis suggested could be viable, mainly from less densely populated areas.

Ultimately 8 of the stops modelled from Twitter data analysis attracted enough bookings to make the services commercially viable, providing improved access to around 450 new users (a 37% increase on the previous year). All eight new stops are located in municipalities with population density less than 1000 persons/km$^2$ (average 600 persons/km$^2$). This is lower than all but three of the 20 municipalities where existing BusUp stops are located. In all of these three cases the existing stop does not form a direct route but is on the route of another established service (average population density for established stops is 2100 persons/km$^2$). Hence these new stops are all in lower density municipalities which would not have been identified (and indeed were not identified previously) as having sufficient demand for commercially viable BusUp services without the additional information on interest to attend the festival from these locations indicated by the Twitter data. In other words, without the Twitter data, these locations would not have been recognised as having sufficient demand to make a BusUp service viable and so would not have had a service offered. Hence it is not unreasonable to claim that these new services and the additional passengers on these new services are a direct result of the inclusion of Twitter data in the analysis/model.

While this confirms that deriving estimated demand from social media data can form the basis to establish commercially viable bus routes to large scale events from peri-urban and more rural areas, it also highlights that there is scope to further improve the number of bookings from the stops on routes which were ultimately cancelled. More marketing of the published routes across a wider range of media is required and stronger partnership with festival organisers could tie in festival ticket sales with bus bookings. Other ideas being explored include early bird discounts offered for advance booking of bus tickets to encourage more passengers to book in advance and well ahead of the 2-week cut-off.

To better understand who was using the BusUp service, surveys were conducted with a sample of BusUp users (sample size 209) and a sample of non-BusUp users (sample size 60) at the entrance to the festival. Of those surveyed, three quarters were female and over three quarters were under 24 years of age.

- Of the BusUp users, 15% stated they would not have been able to attend without BusUp, while 36% would have come by car, the remainder stating they would have come by train.
- 40% of the BusUp users had attended the festival in previous years, and on those occasions 59% came by car, 37% by train and only 2.5% by bus.
- Satisfaction with the BusUp service by its users was very high, with 64% stating they were very satisfied and 33% satisfied with the service. This contrasts to the train users, where 50% were unsatisfied with the service received.
- For the routes where alternative public transport is possible, the BusUp services are estimated to give a 60% faster travel time.

In the example featured in this paper, the attendees to the annual Canet Rock music festival were predominantly under 24 of age and the majority were social media active. In addition, there was a reliance on lifts from parents due to a lack of public transport options from peri-urban and more rural areas. There characteristics were important factors in the appropriateness and success of the approach described.

This paper has shown that the innovation of information mining from Social Networks can provide a better comprehension of the demand and can radically improve the capacity of bus ride-sharing services, such as BusUp, to offer more tailored services and to better aggregate and serve the demand to large events (>10.000 attendees). Practitioners should be aware that the transferability of this approach to other scenarios needs to be considered carefully. Three characteristics of this example were important factors in contributing to its successful application:

1. Gaps in knowledge of demand: one-off or irregular large-scale travel demand generators such as events where attendees are changing/not known in advance
2. Social Media active target groups
3. Shortage of suitable existing transport options for the target groups

A more general limitation of the approach relates to the quantity and quality of the Twitter data sources. Twitter users often don't make their registered location publicly available for their accounts and seldom have geolocation access activated (in this case, only 15% of accounts). Other fields related to user location data can also be deprecated or fictional. Therefore, even when the target audience is likely to be active social media users, such as those attending music festivals, there may only be a relatively low percentage of their Twitter data that is useable and useful in estimating spatial distribution of demand. Hence, for best results, it becomes essential to complement the social media analysis with other data sources to increase robustness. This has been explored by combining demographic data, historic attendance data and Twitter influence score according to the methodology described in Section 5.2.

Nevertheless, in the future, increasing penetration of social media data, innovations in text mining and analysis methods, alongside more user centred transport solutions will continue to leverage social media as a supplemental transport planning data source.

**Acknowledgment**

The research reflected in this paper has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 770115.

## References

Abbasi, A., T. H. Rashidi, M. Maghrebi, and S. T. Waller. 2015. "Utilising location based social media in travel survey methods: bringing Twitter data into the play." *Proceedings of the 8th ACM SIGSPATIAL international workshop on location-based social networks.* ACM. 1.

Boyd, D., and K. Crawford. 2012. "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon." *Information, communication & society, 15(5), 662-679.* 15 (5): 662-679.

Chang, M. S., and P. R. Lu. 2013. "A multinomial logit model of mode and arrival time choices for planned special events. Journal of the Eastern Asia Society for Transportation Studies, 10, 710-727." *Journal of the Eastern Asia Society for Transportation Studies* 10: 710-727.

Candelieri, A., and Archetti, F. (2015) Detecting events and sentiment on twitter for improving urban mobility. CEUR Workshop Proceedings. 1351. 106-115.

Chirieleison, C., and L. Scrucca. 2017. "Event sustainability and transportation policy: A model-based cluster analysis for a cross-comparison of hallmark events." *Tourism Management Perspectives* 24: 72-85.

Collins, A., and D. Potoglou. 2019. "Factors influencing visitor travel to festivals: challenges in encouraging sustainable travel. Journal of Sustainable Tourism." *Journal of Sustainable Tourism* 27 (5): 668-688.

Collins, C., S. Hasan, and S. V. Ukkusuri. 2013. "A novel transit rider satisfaction metric: Rider sentiments measured from online social media data." *Journal of Public Transportation* 16 (2): 21-45.

Cottrill, C.D., P. Gault, G. Yeboah, J. D. Nelson, J. Anable, and T. Budd. 2017. "Tweeting Transit: An examination of social media strategies for transport information management during a large event." *Transportation Research Part C: Emerging Technologies* 77: 421-432.

Cui, Y., Meng, C., He, Q., Gao, J., 2018. Forecasting current and next trip purpose with social media data and google places. Transportation Research Part C: Emerging Technologies 97, 159-174.

de Lira, V.M., Macdonald, C., Ounis, I., Perego, R., Renso, C., and Times, V.C (2019) Event attendance classification in social media, Information Processing & Management, Volume 56, Issue 3, 2019, Pages 687-703, ISSN 0306-4573, https://doi.org/10.1016/j.ipm.2018.11.001.

Evans, G. 2012. "Hold back the night: Nuit Blanche and all-night events in capital cities." *Current Issues in Tourism* 15 (1-2): 35-49.

Forgy, E. 1965. "Cluster Analysis of Multivariate Data: Efficiency versus Interpretability of Classifications." *Biometrics* 21: 768–80.

Fernando, R. 2019. "The impact of Planned Special Events (PSEs) on urban traffic congestion." *EAI Endorsed Transactions on Scalable Information Systems* 6 (23): 1-11.

Gal-Tzur, A., S. M. Grant-Muller, T. Kuflik, E. Minkov, S. Nocera, and I. Shoor. 2014. "The potential of social media in delivering transport policy goals." *Transport Policy* 32: 115-123.

Grant-Muller, S. M., A. Gal-Tzur, E. Minkov, S. Nocera, T. Kuflik, and I. Shoor. 2014. "Enhancing transport data collection through social media sources: methods, challenges and opportunities for textual data." *IET Intelligent Transport Systems* 9 (4): 407-417.

Gu, Y., Z. S. Qian, and F. Chen. 2016. "From Twitter to detector: Real-time traffic incident detection using social media data." *Transportation research part C: emerging technologies* 67: 321-342.

Hangal, Sudheendra, Diana MacLean, Monica S Lam, and Jeffrey Heer (2010) “All Friends Are Not Equal: Using Weights in Social Graphs to Improve Search.” In *Workshop on Social Network Mining & Analysis, ACM KDD*. http://vis.stanford.edu/papers/weighted-social-graphs.

Hu, W., Jin, P.J., 2017. An adaptive hawkes process formulation for estimating time-of-day zonal trip arrivals with location-based social networking check-in data. Transportation Research Part C: Emerging Technologies 79, 136-155.

Jolliffe, I. T. (1986) *Principal Component Analysis*. Springer Verlag.

Kaiser, M.S., Lwin, K.T., Mahmud, M., Hajializadeh, D., Chaipimonplin, T., and Sarhan, A. (2018) "Advances in Crowd Analysis for Urban Applications Through Urban Event Detection," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 10, pp. 3092-3112, Oct. 2018, doi: 10.1109/TITS.2017.2771746.

Kuppam, A., R. Copperman, T. Rossi, V. Livshits, L. Vallabhaneni, T. Brown, and K. DeBoer. 2011. "Innovative methods for collecting data and for modeling travel related to special events." *Transportation research record* 2246 (1): 24-31.

Kwoczek, S., S. Di Martino, and W Nejdl. 2014. "Predicting and visualizing traffic congestion in the presence of planned special events." *Journal of Visual Languages & Computing* 25 (6): 973-980.

Latoski, S. P., W. M. Dunn, B. Wagenblast, J. Randall, and M. D. Walker. 2003. *Managing travel for planned special events.* Office of Transportation Management, United States: Federal Highway Administration.

Leilei, D., S. Zheng-liang, G. Jin-gang, and Q. Hong-tong. 2012. "Study on traffic organization and management strategies for large special events." *2012 International Conference on System Science and Engineering (ICSSE).* IEEE. 432-436.

Li, Y., X. Wang, S. Sun, X. Ma, and G. Lu. 2017. "Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks." *Transportation Research Part C: Emerging Technologies* 77: 306-328.

Lloyd, Stuart P. (1982) “Least Squares Quantization in PCM.” *IEEE Transactions on Information Theory*. https://doi.org/10.1109/TIT.1982.1056489.

May, A., Boehler-Baedeker, S., Delgado, L. et al. (2017) Appropriate national policy frameworks for sustainable urban mobility plans. Eur. Transp. Res. Rev. 9, 7. https://doi.org/10.1007/s12544-017-0224-1

McLeod, S., Scheurer, J., & Curtis, C. (2017). Urban public transport: planning principles and emerging practice. *Journal of Planning Literature*, 32(3), 223-239.

Montañés, R., Aznar, R., Nogueras, S., Segura, P., Langarita, R., Meléndez, E., Peña, P., and Del Hoyo, R. (2018) “Monitorización de Social Media.” *Procesamiento Del Lenguaje Natural* 61: 177–80. http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/5664.

Newman, M. E.J., and R. M. Ziff. 2001. “Fast Monte Carlo Algorithm for Site or Bond Percolation.” *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*. https://doi.org/10.1103/PhysRevE.64.016706.

Pereira, F. C., F. Rodrigues, and M. Ben-Akiva. 2015. "Using data from the web to predict public transport arrivals under special events scenarios." *Journal of Intelligent Transportation Systems* 19 (3): 273-288.

Scagnolari, S., J. Walker, and R. Maggi. 2015. "Young drivers’ night-time mobility preferences and attitude toward alcohol consumption: A hybrid choice model." *Accident Analysis & Prevention* 83: 74-89.

Schöbel, A. 2012. "Line planning in public transportation: models and methods." *OR spectrum* 34 (3): 491-510.

Schweitzer, L. 2014. "Planning and social media: a case study of public transit and stigma on Twitter." *Journal of the American Planning Association* 80 (3): 218-238.

StatCounter Global Stats. 2019. *Social Media Stats Europe.* July. Accessed August 2019. https://gs.statcounter.com/social-media-stats/all/europe.

The London School of Economics and Political Science. N.D. "Digital Communications: Social Media Platforms and Demographics." Accessed August 2019. https://info.lse.ac.uk/staff/divisions/communications-division/digital-com-munications-team/assets/documents/guides/A-Guide-To-Social-Media-Platforms-and-Demographics.pdf.

Unankard, S., X. Li, and M. A. Sharaf. 2015. "Emerging event detection in social networks with location sensitivity." *World Wide Web* 18 (5): 1393-1417.

van Eggermond, M. A., H. Chen, A. Erath, and M. Cebrian. 2017. "Investigating the potential of social network data for transport demand models." *arXiv preprint arXiv:1706.10035.*

Walther, M., and M. Kaisser. 2013. "Geo-spatial event detection in the twitter stream." *European conference on information retrieval.* Berlin, Heidelberg: Springer. 356-367.

Webster, F. V., and P. H. Bly. 1981. "The demand for public transport: Part I. the changing environment in which public transport operates." *Transport Reviews* 1 (4): 323-351.

Yao, W., Qian, S., 2020. From twitter to traffic predictor: Next-day morning traffic prediction using social media data. arXiv preprint arXiv:2009.13794.

Zhang, Z.; Ni, M.; He, Q.; Gao, J. *Mining Transportation Information from Social Media for Planned and Unplanned Events*; University at Buffalo: Buffalo, NY, USA, 2016; pp. 1–68.