Characterizing, Modeling and Exploiting the Mobile Demand Footprint of Large Public Protests

André Felipe Zanella
IMDEA Networks Institute
Universidad Carlos III de Madrid
Madrid, Spain
andre.zanella@imdea.org

Diego Madariaga
IMDEA Networks Institute
Madrid, Spain
diego.madariaga@imdea.org

Sachit Mishra
IMDEA Networks Institute
Universidad Carlos III de Madrid
Madrid, Spain
sachit.mishra@imdea.org

Orlando E. Martínez-Durive
IMDEA Networks Institute
Universidad Carlos III de Madrid
Madrid, Spain
orlando.martinez@imdea.org

Zbigniew Smoreda
Orange Innovation
Châtillon, France
zbigniew.smoreda@orange.com

Marco Fiore
IMDEA Networks Institute
Madrid, Spain
marco.fiore@imdea.org

ABSTRACT

Smartphones and mobile applications are staple tools in the operation of current-age public demonstrations, where they support organizers and participants in, *e.g.*, scaling the management of the events or communicating live about their objectives and traction. The widespread use of mobile services during protests also presents interesting opportunities to observe the dynamics of these manifestations from a digital perspective. Previous studies in that direction have focused on the analysis of content posted in selected social media so as to forecast, survey or ascertain the success of public protests. In this paper, we take a different viewpoint and present a holistic characterization of the consumption of the whole spectrum of mobile applications during social protests. Hinging upon pervasive measurements in the production network of the incumbent network operator and focusing on the 2023 French pension reform strikes, we unveil how large masses of protesters generate a clearly recognizable footprint on mobile service demands in the examined events. In fact, the footprint is so strong that it lets us develop models informed by the usage of selected mobile applications that are capable of (*i*) tracking the spatiotemporal evolution of the target demonstrations and (*ii*) estimate the time-varying number of attendees from aggregate network operator data only. We demonstrate the utility of such privacy-preserving models to perform a-posteriori analyses of the public protests that reveal, *e.g.*, the precise progression of the marches, alternate minor routes taken by participants or their dispersal at the end of the events.

CCS CONCEPTS

• **Computing methodologies → Modeling methodologies**; • **Networks → Network measurement**.

KEYWORDS

Mobile network measurement; Traffic modeling; Public protests

IMC '24, November 4–6, 2024, Madrid, Spain

ACM Reference Format:
André Felipe Zanella, Diego Madariaga, Sachit Mishra, Orlando E. Martínez-Durive, Zbigniew Smoreda, and Marco Fiore. 2024. Characterizing, Modeling and Exploiting the Mobile Demand Footprint of Large Public Protests. In *Proceedings of the 2024 ACM Internet Measurement Conference (IMC '24), November 4–6, 2024, Madrid, Spain.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3646547.3688421

1 INTRODUCTION

Among many other usages, smartphones and other mobile devices have become paramount instruments for instant communication, advertising and reporting of significant social events. Organizers of gatherings of political, cultural or sports nature take advantage of a variety of mobile applications to announce, coordinate and disseminate the outcome of the happening, and journalists or online commentators use social media platforms as a mean to report or opine about the ongoing events. Mobile services thus create a bridge connecting the physical world to the online ecosystem in presence of large social events.

In the case of public demonstrations, recent years have witnessed a sparking interest in utilizing data from mobile phones and social networks to understand not only how protesters organize and communicate among them via these tools, but also how the digital sphere affects and is affected by large manifestations of social unrest. Existing works have explored in particular the relationship between protests and the X (formerly Twitter) social media platform, by either analyzing geo-tagged tweets in the area of the rallies [2] or performing sentiment analysis of posted content related to the subjects of the demonstrations [33].

In this paper, we propose a different approach that embraces a more comprehensive view of the relationship between large protests and their impression on the digital world. We first investigate *whether such manifestations social unrest induces patterns in the demands for mobile services*, by studying the relationship between the occurrence of public0 marches and the fluctuations in the overall demands for a variety of mobile services in the base stations surrounding the events. To this end, we look at widespread social unrest episodes that happened in France in 2023 as a consequence of pension reforms proposed by the local government, and analyze traffic measurements performed in the affected regions and periods by Orange, the major network operator of the country.

Our investigation unveils that large protests leave a very recognizable footprint on the overall mobile network traffic, especially when considering the consumption of individual services: indeed, a set of specific applications –mainly associated with social media and map navigation– experiences substantially increased demands, whereas others –generally related to entertainment– have significantly reduced relative usage. Hinging upon such strong effects, we further provide a detailed *characterization of the spatiotemporal evolution*, prompted by the different utilization patterns in mobile services. We develop a *post-hoc* proof-of-concept model capable of reconstructing the dynamics of the target marches, based solely on variations of traffic aggregates measured across a relevant subset of mobile services. We also exemplify that a positive indicator for public participation in a large march is the total volume of traffic recorded by the network operator. Ultimately, our findings let us generate a-posteriori estimates of the number of participants over time through each march of the 2023 French pension reform strikes.

Our study reveals the potential of passive mobile network measurements for better understanding large-scale events of social unrest. We show that, if used judiciously, it can offer an opportunity to inspect public marches after they have occurred and in a privacy-preserving manner –as de-personalized aggregates do not offer opportunities for individual surveillance or re-identification of protesters (see Appendix A). These measurements can unveil the precise progression of marches, the attendance across time, alternate routes and the dispersion of participants after the event. This can be useful to support local administrators for planing and management of similar future events. Additionally, these results can inform mobile operators and help them plan for additional capacity in areas of interest in advance.

2 RELATED WORK

Digital media consumption and dissemination is known to have today a critical impact in how protests demonstrations occur online and break barriers offline [30]. Social media act as an alternative medium to traditional sources of news, and provide virtual stages where civil and political organizations can perform online recruitment of different segments of the population, and later move such masses of individuals to physical-world manifestations [8].

The role of smartphones in how a distributed population of protesters organizes, coordinates, and reports in real time about the ongoing demonstration has thus become a subject of multidisciplinary research. Previous studies have proven that social media platforms in particular act as facilitators for information exchange *prior* to protests by shaping online relationships that then become a driving force for individuals to embrace the causes of dissent [18]. The effect is strong to the point that activity surges in geo-located tweets can be used to predict protests hours in advance [26] or to anticipate the rate of success or failure of the future demonstration [15]. Instant messaging applications, on the other hand, play a critical role *during* the social gatherings, with applications such as WhatsApp creating new ways for activists to meet, discuss, and organize in real time [19]. Telegram is sometimes favored in that role as it mitigates the privacy concerns of the organizers [1, 29].

Mobile applications have also opened the door to unprecedented live coverage of public protests, as they offer a very immediate way for any participant or local observer to report facts as they happen to global audiences, factually serving as a form of public journalism [22]. Protesters have been found to often share their visual expressions to help shaping the narrative of the manifestation [20] and developing a community around their cause [14].

Real-time digital communication is also a double-edged sword when it comes to control enforcement of marches. On the one hand, it can be leveraged by government agencies for monitoring purposes [11] or as a surveillance tool to counter-act against protests [21]. On the other hand, social media coverage is a means to announce violent repression of social protests by the authorities [6, 12] or even to coordinate counter-actions against police forces [27].

It is worth noting that not only protests but also many other categories of large social events have been studied through the lenses of the usage of mobile devices and applications. Across all different studies of the digital facets of public events, the most popular source of data is X (or Twitter at the time when most works were carried out): the textual analysis of geo-located posts allows for thorough analysis [25, 33] and accurate forecasting [2, 9, 17, 24, 31] of the target event. Images contained in posts can also be used for the same purposes [32].

Unlike prior work, we do not concentrate on a single mobile application nor we examine the content of material shared therein. Instead, we shift attention to the entirety of data traffic that encompasses a wide range of mobile services beyond single social media and messaging applications. While earlier studies have adopted a similar perspective to investigate different phenomena, none has ever applied it to public protests. For instance, network traffic measurements have been used to characterize cultural or sports events [4], understand the impact of holidays during elections [5] or monitor road traffic congestion [13].

Closer to our approach, a few studies have employed signaling data from mobile networks to study the attendance to mass protests [23] and to link it with the home location of participants [28]. Yet, such works depend on personal network localization information to investigate the dynamics of demonstrations in real time, and do not look into mobile traffic demands or application consumption that are the focus of our study. Also, our research is not targeted at live tracking of protesters that entails significant privacy risks and raises ethical questions.

Ultimately, ours is the first investigation of the effects that large public demonstrations induce on the whole mobile data traffic and on the demands for a varied range of applications.

3 MEASUREMENT DATA

Our study leverages network traffic measurements from the production network of Orange, a major global operator with a leading market position in France. The traffic data collection was performed in a continuous manner from January 31 to May 31, 2023 over 29, 171 carriers covering five main urban areas of France, *i.e.*, Paris, Lyon, Toulouse, Nantes, and Bordeaux. In these regions, we monitored all available radio access network (RAN) technologies, *i.e.*, 2G, 3G, 4G, and 5G: while 4G and 5G serve today the vast majority of the demand, it is not uncommon that 2G and 3G serve non-negligible portions of the data traffic in heavily loaded scenarios such as those produced by large manifestations.

3.1 Mobile network traffic measurements

At the time of the collection, Orange had a non-standalone (NSA) 5G deployment, where 5G gNodeBs providing faster wireless communication to 5G-capable devices coexist with 4G eNodeBs in the RAN. In such a configuration, mobility control operations for both 4G and 5G access technologies happen via the 4G Mobility Management Entity (MME). Thanks to this, monitoring the S1-MME interface is sufficient to localize both 4G and 5G equipment via their signalling traffic, by associating them to a carrier (hence a base station geographical site) at every time they interact with the network. For 2G and 3G traffic, geo-referencing is performed using the User Location Information (ULI) element contained in the Packet Data Protocol (PDP) Contexts that indicates the carrier where the device is currently registered. This is achieved by inspecting GPRS Tunneling Protocol control plane (GTP-C) signalization over the Gn interface of the Gateway GPRS Support Node (GGSN).

The traffic generated by the mobile devices is analyzed by Orange proprietary classifiers that hinge upon Deep Packet Inspection (DPI) and tap at the Gn interface of the GGSN for 2G and 3G equipment, or at SGi interface of the 4G Packet Gateway (PGW) that is serving both 4G and 5G users in NSA 5G deployments. The classifiers associate each traffic flow to one mobile application. They are systematically used for network monitoring and optimization by the operator and are reported to yield high accuracy, although their precise operation stays confidential.

By crossing the localization information with the service-level classification data, and aggregating the result over all mobile devices attached to every RAN carrier, the operator computed the demands for each application served by every carrier. For the purpose of this study, such information also accumulated over time into 5-minute intervals. The result are time series of the volume of traffic (in bytes) generated by over 400 mobile services at each of the target 29, 171 carriers at every 5 minutes. As discussed in detail in Appendix A, the data measurements and processing abide by all applicable regulations, and the resulting aggregates are privacy-preserving since they do not allow re-identifying individuals or retrieving personal information.

3.2 Baseline carrier-level service demands

During the observation period, a number of protests linked with the French pension reform took place in the considered cities. The full list can be found in Table 2 in Appendix B.

In order to assess the impact of protests on the mobile network traffic, we need to establish a baseline period where the consumption of mobile services is deemed *normal*, *i.e.*, not affected by similar demonstrations. As the majority of pension reform strikes in France occurred during work days, we consider a set of 10 baseline work days that are not holidays, are not marked by manifestations according to the available records, and are not immediately adjacent to days characterized by protests: February 8, 9, 14, 15, 28, March 1, 2, and April 25, 26, 27, 2023.

For each carrier and mobile service, we compute a reference daily demand profile by using the median volume of traffic observed at every 5 minutes in the 10 baseline days. This captures the typical expected traffic in normal conditions.

4 THE MOBILE TRAFFIC FOOTPRINT OF THE 2023 FRENCH PROTESTS

Our first objective is characterizing the impact of large social unrest on local mobile network traffic demands. To investigate, we use data from France, where protesters marched against the 2023 pension reform issues by the national government. To this end, we analyze the changes in the dynamics of the overall traffic consumption during the dates when the target protests occurred, with respect to the *normal* baseline days. More precisely, we aim at gaining insights from two perspectives, as follows.

(1) We explore how the protests impact the *overall consumption of mobile data traffic* at each carrier of the network near the protest route; this allows pinpointing the approximate areas and times where network traffic is substantially influenced by the events.
(2) We investigate how protests affect the *demands for individual mobile applications*; it is expected that services are consumed in different –positive or negative– ways during the manifestations, which may offer a more precise view into the footprint.

In both parts of our analysis, we only consider aggregates of the overall and service-level traffic at each carrier and devise original metrics to quantify the impact of the target public protests on such demands.

4.1 Pinpointing macroscopic effects of protests on overall mobile traffic

As an initial step in our analysis, we aim at identifying deviations in the patterns of the total traffic served by individual carriers caused by the presence of protesters. The general objective is to single out clear *true positives*, *i.e.*, carriers that are evidently impacted by a protest and that can be used to establish a *ground truth* of affected areas and times to be later utilized for a finer service-level analysis.

Our methodology is illustrated in Figure 1. For a given protest day, the first step involves an initial triage of carriers based on their proximity to the protest route authorized by the local police prefecture. This classification gives reasonable certainty that the observed patterns are happening near the expected protest location (and they are not uncontrolled outliers away from it). We set $d = 1$ km as the proximity threshold. Considering only this set of nearby carriers, the second step involves identifying carriers that see their traffic pattern clearly affected by protesters' presence. To achieve this, we compute the deviation of the traffic consumption during protest days with respect to the baseline days defined in Section 3.2. It is important to note that at this stage, our methodology is set to be conservative, *i.e.*, it seeks to select only carriers greatly affected by the protest. As this is an intermediary step of the analysis, the objective is to find carriers to be used as true positives to characterize the impacts of protests on mobile traffic consumption.

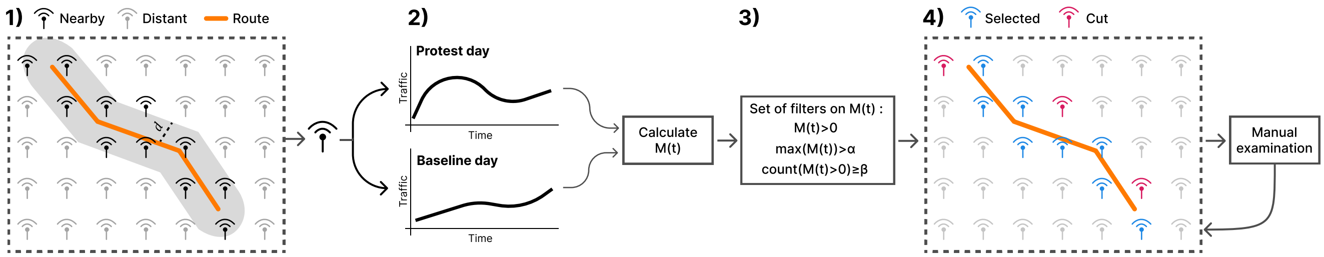


Figure 1: Methodology adopted to identify true positive carriers that are affected in a macroscopic way by a protest. Carriers in proximity of a protest path (1) are assessed based on the variation of their served mobile traffic during the day of the protest (2) using a dedicated metric and filters (3). Finally, true positive (in blue) carriers are told apart from cut (in red) carriers that do not show sufficiently clear change in their traffic pattern (4).

In order to compute the traffic pattern deviation, we first normalize the traffic on carriers during both protest and baseline days. Here, a volume-oriented (*e.g.*, min-max) standardization would not be ideal as it puts emphasis solely on volume variations, which may lead to incorrect interpretations. For instance, one carrier may yield a higher demand during the protest day than in the baseline and yet have aligned peaks of usage, meaning that the observed traffic consumption behavior is semantically the same and most likely not caused by the protest (which is limited to a specific time interval). Instead, we opt for a z-score standardization, which avoids volume bias and allows for a direct comparison of the traffic dynamics between different days.

For each 5-minute interval $t \in T$, let us denote $P(t)$ the z-scored traffic observed during the protest day. The z-scored traffic at a given baseline day $n \in N$ will be defined as $B_n(t)$, with the set $\mathbf{B}(t) = \{B_1(t), ..., B_n(t)\}$ containing the time series of all baseline days, with the same range of values T as $P(t)$. The mean of $\mathbf{B}(t)$ will be $\mu_b(t) = \frac{1}{N}\sum_n^N B_n(t)$ at time t; similarly, the instantaneous standard deviation will be $\sigma_b(t) = \sqrt{\frac{1}{N}\sum_n^N (B_n(t) - \mu_b(t))^2}$. Our metric for overall traffic pattern deviation is then defined as

$$M(t) = P(t) - (\mu_b(t) + 3\sigma_b(t)), \tag{1}$$

and time t is considered to be affected by a significant change in mobile activity due to protesters roaming within coverage of the carrier if $M(t) > 0$. In other words, we identify as significant deviations those exceeding $3\sigma_b(t)$, which, in case the distribution of baseline traffic over a given instant t on the set $\mathbf{B}(t)$ follows a normal distribution, corresponds to values outside the 99.7% percentage.

We then run two additional filters on $M(t)$ to remove potential outliers. The filters are defined as follows.

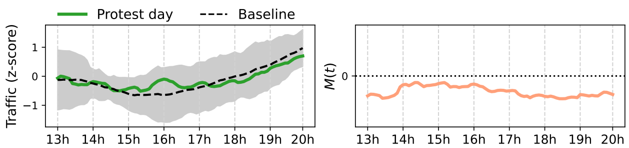
- It is important to filter out cases where $M(t) > 0$ just because the overall traffic at the carrier is very low or noisy and the condition above is met only as a result of random demand fluctuations. To this end, we introduce a minimum threshold α on the maximum value of $M(t)$ for a given carrier, and only flag as true positive carriers for which $max(M(t)) > \alpha$. After examining results, we determine $\alpha = 0.5$, which indicates strong effects in the traffic dynamics during the protest day that cannot be solely caused by random fluctuations.
- We consider that whenever a protest affects a carrier in a significant way, a $M(t) > 0$ condition shall be observed through the set of consecutive time instants t during which the protesters are under coverage of the carrier. We thus define as $\tilde{T} \in T$ the set of time instants during which $M(t) > 0$, and define a threshold β over its cardinality, *i.e.*, we force $count(M(t) > 0) \geq \beta$. Based on extensive trials, we set $\beta = 10$ during a protest day.

We also calculate a set of clear *true negative* carriers, which are denoted as the carriers (*i*) far from the protest route based on the initial triage, and (*ii*) presumably not affected by the protest, *i.e.*, $P(t)$ is significantly similar to all values within $\mathbf{B}(t)$. After calculating $M(t)$ for all far away carriers, we define the true negatives as any carrier where its 97^{th} percentile of $M(t)$ is smaller than -0.5. This means that at least 97% of the 5-min time instants for this carrier of $P(t)$ are within the interval estimate of $\mathbf{B}(t)$.

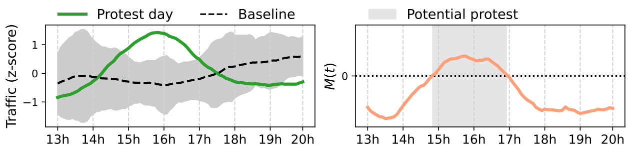
All the true positive and true negative carriers are then manually inspected, in order to obtain the final ground truth set. For the sake of clarity, we present two examples in Figure 2. Figure 2a presents a *true negative* carrier, where both $P(t)$ (in green) and $\mathbf{B}(t)$ (represented by $\mu_b(t)$ as the dashed line and the $\pm 3\sigma_b(t)$ interval around it) have a similar pattern for the z-scored traffic (left plot). This results in a traffic pattern deviation $M(t)$ (right plot, in orange) that is never greater than zero, implying that the temporal consumption pattern of this carrier was not affected by the presence of the protesters. On the opposite hand, Figure 2b shows a *true positive* carrier, with clear changes in its traffic consumption patterns. It can be specifically noted that this carrier experiences an uncommon surge of traffic demand $P(t)$ just after 15:00, which is only normalized after 17:00. Indeed, this surge of demand is completely outside the 99.7% interval estimate of $3\sigma_b(t)$, as observed over $M(t)$ on the right plot. The set of time intervals $\tilde{T}$ which were flagged as the potential hours that the protesters were within the coverage area of this carrier are marked in light gray.

4.2 Ground truth validation via temporal and spatial analyses

Having established a ground truth of carriers clearly affected by the presence of the protesters, we proceed to its visual validation. Specifically, we explore how the set of true positive carriers obtained



(a) Carrier not affected by the protest



(b) Carrier affected by the protest

Figure 2: Examples of (a) true negative and (b) true positive carriers, in relation to how their traffic patterns deviated during the protest day against the baseline days. The proposed $M(t)$ analyzes the difference of protest (in green) and baseline (in black) traffic, highlighting the potential hours when the protest affected the selected carriers whenever $M(t) > 0$, as indicated by the gray range, together with additional filters.

for each protest day was affected over space and time, in relation to the intervals $\tilde{T}$ that are marked as potential moments when the protest passes by the coverage area of the selected carriers. The temporal evolution of $M(t)$ across the true positive carriers of each protest day can be observed in Figure 3, where red values indicate $M(t) > 0$ and black lines denote the interval $\tilde{T}$ where each carrier is affected. As expected, an evolution of $M(t)$ happened over time: some carriers are flagged only at the initial points hours of the protest (14:00), some in the middle and others near the end (19:00). While this is an expected behavior, it is an initial validation for the ground truth set, as it shows that the set of carriers affected changes over time in a consistent way with respect to the path followed by the protesters. Indeed, no single carrier displays the traffic pattern disturbance throughout the full 14:00 to 19:00 time interval of the protest, which confirms the natural handover of mobile network users as they progress through the route.

The natural movements of protesters along the designed route, following the true positive carriers, can be observed over Figure 4, where lighter/darker colors indicate carriers that are affected earlier/later. As expected, carriers that are impacted earlier are nearby the official starting points of each protest, while the ones affected later are nearby the official end points, with intermediary carriers located along the announced route of the march.

These results over both temporal and spatial dimensions offer initial insights on the dependability of our methodology, since the way in which carriers are influenced by protests according to our metric and filters is very consistent with expectations. This suggests that the obtained ground truth can be trusted for the next steps of the analysis.

4.3 Analyzing the impact of protests on individual application demands

Leveraging the identified true positives, we delve into a more detailed study of the impact of the target protests on the demands for specific mobile services. When it comes to per-application traffic, simply comparing traffic dynamics as done for the overall traffic in Section 4.1 is not a reliable approach: due to the natural diversity of traffic volumes, services like video streaming will generate much higher network traffic loads than lower-traffic services, such as instant messaging. Moreover, carriers in general may expect different loads, according to the network optimization routines defined by the operator, *i.e.*, a higher capacity macro cell will expect more users attached than a specific beaming micro cell, which will as well lead to different orders of magnitude volume between both. Consequently, those differences in magnitudes may just hide away anomalies and create volume biases for models. Therefore, an alternative method for quantifying the impact of manifestations of social unrest on mobile application usage is required.

We propose a new metric for the characterization of changes in the volume of mobile applications' traffic due to massive events. Let $i \in I$ be a specific application from the set I of all mobile applications. For each carrier, let $T_{p,i}(t)$ be the traffic on the protest day at time t, and $T_{b,i}(t)$ be baseline traffic at time t. Then the proposed metric is defined as

$$F_{i,ns}(t) = \frac{T_{p,i}(t)/\sum_i^I T_{p,i}(t)}{T_{b,i}(t)/\sum_i^I T_{b,i}(t)} \tag{2}$$

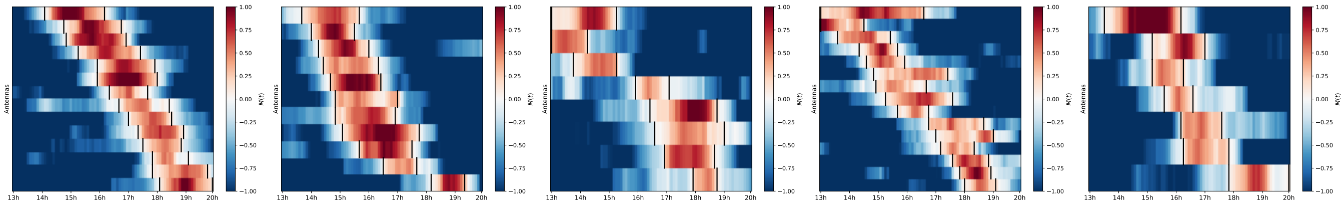
$$F_i(t) = \frac{F_{i,ns}(t) - 1}{F_{i,ns}(t) + 1} \tag{3}$$

where $F_{i,ns}(t)$ is the non-symmetric metric and $F_i(t)$ is its symmetric version that takes values in $[-1, 1]$.

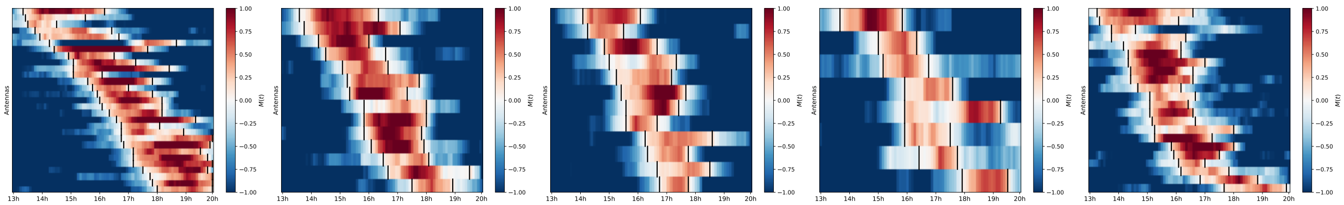
$F_i(t)$ represents how the percentage of traffic each mobile application generates in relation to the remaining set has changed on the day of the protest, with respect to the baseline. Values of $F_i(t) > 0$ mean that the application i at instant t experiences a rise in importance against other applications in regards to the baseline, while $F_i(t) < 0$ indicates a loss of importance against other applications during the protest day in relation to the normal period.

In order to check the validity of our proposed $F_i(t)$ metric, we compare its values for the set of true positive carriers and for the set of true negative carriers separately. As previously mentioned, the ground truth set is established from the overall traffic analysis only, where no distinction between applications is considered. Therefore, this metric can help find applications whose consumption patterns change because of the presence of protesters. Figure 5 represents the distribution of $F_i(t)$ for true positive carriers (in green) versus true negative carriers (in black). The following important remarks are in order.

(1) A first set of applications have a clear separation in their $F_i(t)$ values between true positive and true negative carriers, where there is a growth of the share during the protest day for the positives and a decrease for the negatives. This set includes instant messaging, localization, and news applications, such as WhatsApp, Twitter, Google Maps, and News

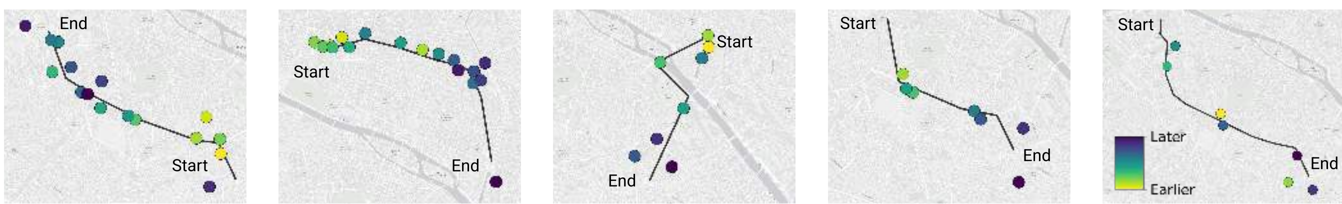


(a) Jan 31 (b) Feb 7 (c) Feb 16 (d) Mar 7 (e) Mar 15

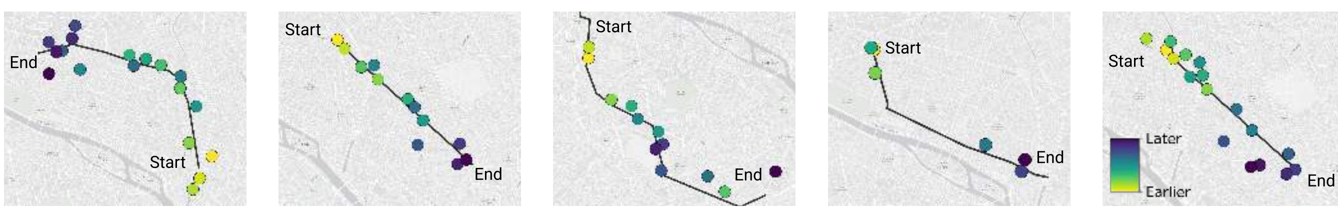


(f) Mar 23 (g) Mar 28 (h) Apr 6 (i) Apr 13 (j) May 1

Figure 3: Evolution of the protests, represented by the standardized traffic of the flagged carriers, ranked by the average time when $M(t) > 0$. Results refer to the 10 protests that occurred in Paris (see Table 2).



(a) Jan 31 (b) Feb 7 (c) Feb 16 (d) Mar 7 (e) Mar 15



(f) Mar 23 (g) Mar 28 (h) Apr 6 (i) Apr 13 (j) May 1

Figure 4: Evolution of the protests represented by the spatial distribution of the flagged carriers, color coded according to the time ranking. Results refer to the 10 protests that occurred in Paris (see Table 2).

websites. This suggests that those applications have their consumption greatly increased by the protests.

(2) A second set presents also a clear separation between true positives and true negatives, although not an increase of share. This set includes Telegram and Transport. We believe that these applications are moderately affected by the protest since they have significant lesser decline in their share against the baseline for the true positives.

(3) A third set of applications had a growth of share in the protest day against the baseline both for the true positives and negative sets. This includes popular social media applications, such as Instagram, Facebook, and Snapchat. This can be related to not only people generating content within the march, but also users searching for such content throughout the city.

(4) A final set can be characterized by applications with significant decrease of usage in the carriers affected by the protest, which includes specially applications related to audio (Spotify, Deezer) and video streaming (Netflix), as well as work-related (LinkedIn).

Overall, the insights presented above prove that large public protests that occurred in 2023 in France effected a very noticeable footprint on the mobile network traffic. This conclusion holds at the level of both macroscopic dynamics of the overall traffic and more fine-grained service-level demand shares. The results imply

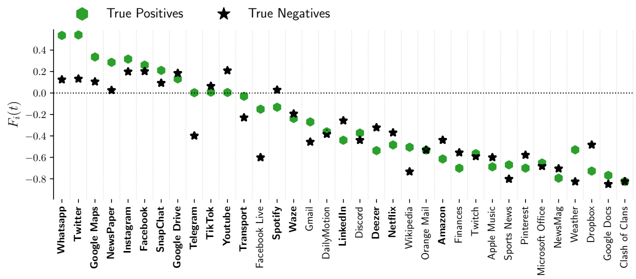


Figure 5: Distribution of $F_i(t)$ across all days in Paris. Applications that represent a significant change in usage are the ones where a clear separation between the distribution of protest and baseline happens, such as WhatsApp and Twitter. Applications selected as features are highlighted in bold.

clear variations in the consumption of applications at the carriers affected by the presence of protesters, with services related to instant messaging, transportation, and news seeing a surge of usage, which matches what previous works reported in Section 2.

5 LEVERAGING MOBILE TRAFFIC DEMANDS TO CHARACTERIZE THE 2023 FRENCH PROTESTS

The effect of the considered social unrest on mobile traffic is so strong that it paves the way for a data-driven modeling of the phenomenon. Specifically, we explore to what extent the footprint left by protesters in the mobile network traffic can be leveraged to characterize the evolution of their march over space and time.

Following the procedure described in Section 4.1, we obtain a ground truth set composed of more than $100,000$ 3-tuples with type (*carrier_id*, *time*, *label*), where *carrier_id* corresponds to a specific geographical location, *time* is a 5-min interval, and *label* classifies each data entry as a true positive or true negative sample. The label distribution is inherently imbalanced, with a ratio of 1 to 30 between positive (protest) and negative (non-protest) classes.

The ground truth set, mostly based on our traffic pattern deviation metric, can then be used to derive a model to classify carriers affected by public protests at a 5-minute time resolution. The classifier relies on the relative usage values $Fi(t)$ for all applications $i \in I$ that we find to be affected in a significant manner by the protests: specifically, the set of services whose $Fi(t)$ values are employed as features in our model are highlighted in bold in Figure 5.

5.1 Model description and evaluation

Our classification model aims at predicting, at a 5-minute resolution, which carriers are affected by large public protests, solely based on the changes in the application traffic demands with respect to a baseline day. In our experiments, we employ an XGBoost classifier (XGBC), an iterative algorithm that relies on an ensemble of decision trees capable of handling class imbalance through a weighing class technique. After hyperparameter tuning, the resulting XGBC consists of $1,000$ decision tree estimators, reinforcing the complexity of the model. Also, we select logistic regression as a loss function for binary classification, with a learning rate of 0.05. Moreover, our XGBC deals with class imbalance by weighing

		Predicted	
		Protest	Non-protest
Actual	Protest	970	56
	Non-protest	10	31,116

Table 1: Confusion matrix of trained XGBC model when applied to test dataset.

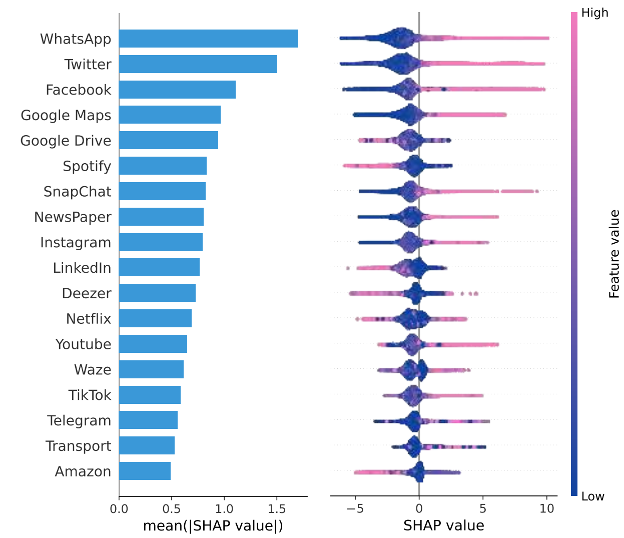
the true positive examples relative to true negative examples and favoring better performance in the positive protest class.

To analyze the classifier model performance, we randomly sample our ground truth dataset into training and testing sets using a 70:30 splitting ratio. Then, we train the model on the training dataset using a subset of 18 of the application consumption features as outlined in Figure 5. The classification results using the XGBC are summarized in Table 1, where its high accuracy in distinguishing the carriers affected by public protests is evident. Accordingly, we also find that the provided set of features lets the XGBC model determine with a 0.97 F1-score whether a large public protest has impacted a carrier during a specific 5-minute interval. The high F1-score value also implies the good performance of the model in terms of precision (0.99) and recall (0.95).

In order to provide insight into the explainability of our classifier model, we calculate SHAP values for the trained model to analyze the contribution of individual features to the classification outcomes. As shown in Figure 6a, we employ absolute SHAP values to determine feature importance. From the figure, it can be seen that changes in the consumption of WhatsApp and Twitter contribute the most to form the model prediction. Additionally, Figure 6b provides insights into the associations learned by the classifier model, *i.e.*,, whether protest-affected carriers are related to increased or decreased mobile service consumption. The figure suggests that affected carriers relate to increased consumption of services such as WhatsApp, Twitter, Facebook, and Google Maps, as well as a decreased consumption of Google Drive, Spotify, LinkedIn, and Netflix, among others. It is important to note that these trends are well-aligned with the insights into service consumption changes presented in Section 4.3. Indeed, Figure 5 shows that, in general, the selected features (highlighted in bold) present high differences when comparing true positives and true negatives data points.

5.2 Generalization analysis

To further validate our proposed method, we perform different experiments running our classification model over an entire target day. As a result, we obtain a list of potentially affected carriers by a protest during specific time intervals. Besides, given that public demonstrations are assumed to be continuous in space and time, we incorporate a secondary step to reduce the number of false positives and identify any potential protest as a complex but consistent event. More precisely, we apply a density-based spatiotemporal clustering algorithm (ST-DBSCAN) [3] over the pairs of carriers and time intervals previously classified as protest by the XGBC. Therefore, a potential protest is associated to a spatiotemporal cluster of carriers classified as affected by the model. In all our experiments, we



(a) Feature importance **(b) Explanation summary**

Figure 6: Contribution of application features to the XGBC model prediction.

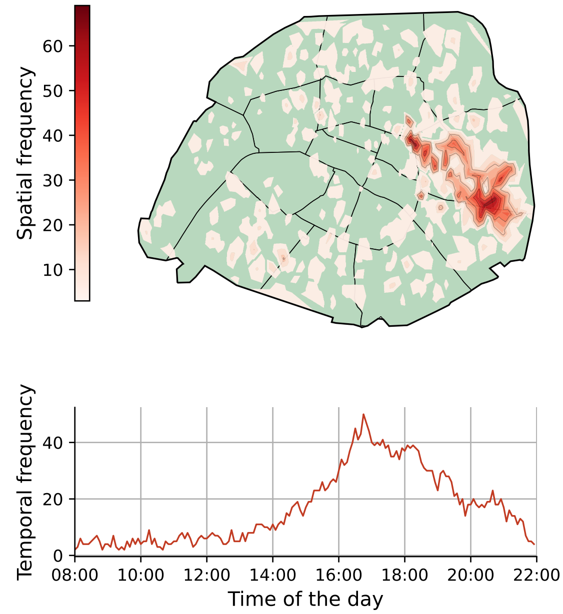


Figure 7: Carriers labeled as affected by XGBC on May 1 in Paris. The plots illustrate the frequency of labeled carriers in space (top) and time (bottom).

employ a specific parametrization for the ST-DBSCAN algorithm to ensure consistency in detected protests. Accordingly, the neighborhood around a point is defined by a spatial radius of 1.2 km and a temporal radius of 45 minutes. In addition, the minimum number of points to form a dense region is set to 30.

5.2.1 Intra-city testing. Firstly, we carry out an intra-city test, where we train our XGBC using a subset of the ground truth dataset to later classify all the carriers of Paris at a 5-minute resolution. We experiment using our 2-step methodology (combining XGBC and ST-DBSCAN) to detect each of the ten large demonstrations in Paris, previously mentioned in Section 4.1. Given that our ground truth contains data from all ten protests under consideration, we first remove the target day from the training set. Thus, only the remaining nine protest days are used to train the XGBC in each case. Then, we apply our 2-step protest detection approach over the entire target day to corroborate the identification of the protest along the officially authorized route.

In order to illustrate the importance of employing a density-based clustering algorithm, Figure 7 depicts the spatiotemporal temporal patterns of the carriers classified as affected during a protest day (May 1). Those results are built on the output of the XGBC, which classified more than 1, 500 carriers in Paris for each 5-minute interval between 8:00 and 22:00. According to the figure, less than 3% of all the carriers are labeled as affected at each time interval. Notably, a faction of the detected carriers emerges from the rest, exhibiting high density in both spatial and temporal scales. Therefore, the incorporation of ST-DBSCAN is essential, as it can help separate false positive data points from the actual group of carriers affected by the protest. Indeed, in all ten experiments, ST-DBSCAN distinguishes a single dense cluster and labels the rest of the points returned by XGBC as noise.

Figure 8 shows the output of our protest detection approach, where the carriers labeled as affected are represented by their corresponding Voronoi cell, and the coloring depicts the average time of the day when each carrier is classified as affected. These results show that our methodology consistently identifies the ten public protests in Paris, densely covering the official routes and exhibiting a clear time consistency referable to the protesters moving from the start to the end of the march. In fact, these results greatly enhance the set of impacted carriers shown in Figure 4: not only they fill all gaps along the routes but also exhibit some interesting phenomena, such as the dispersion of people at the end of the demonstration (*e.g.*, Figures 8a, 8b, and 8i) and the existence of alternative less-pronounced routes (*e.g.*, Figures 8d, 8h, and 8i). Those alternative itineraries were not part of the authorized routes, yet we found that some were non-officially advertised during the marches.

In addition, we focus on inspecting our proposed detection method in terms of its precision in identifying public protests. Accordingly, we train our XGBC model over the full ground truth set and apply the 2-step approach to four normal non-protest days in Paris: May 16, 17, 23, 24, and 25. During those days, our model correctly does no find any potential protest.

5.2.2 Inter-city testing. Secondly, we perform an inter-city test to assess the feasibility of using an XGBC trained in one city to detect public demonstrations in different cities. For this analysis, we train the XGBC using the ground truth dataset containing information about the ten protests in Paris, and apply our proposed methodology to identify all other protests occurring in the other four French cities of Lyon, Toulouse, Nantes, and Bordeaux. The test set then consists of 24 events: Table 2 summarizes the selected protest days and cities, showing a great diversity among the magnitude of the social unrest events studied.

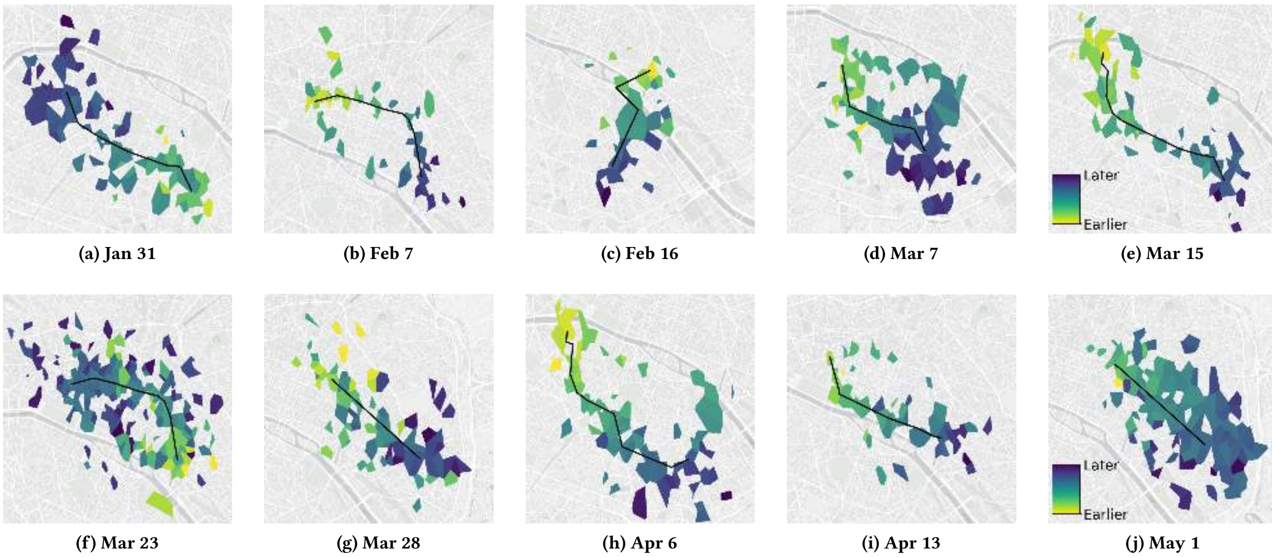


Figure 8: Intra-city testing: Protest events identified in Paris based on our proposed detection methodology. Early affected carriers colored yellow, while later affected carriers colored purple.

Figure 9 presents the obtained results according to the 24 test cases, where carriers are represented by Voronoi cells and the colors relate to the average time of the day each carrier was affected. For each target event, our methodology is able to identify a spatiotemporal dense set of affected carriers, separating the single discovered cluster from the noise data. Closer inspection of the figures exhibits consistency between the coverage area of affected carriers and the officially authorized route, which further supports the soundness of our proposed method.

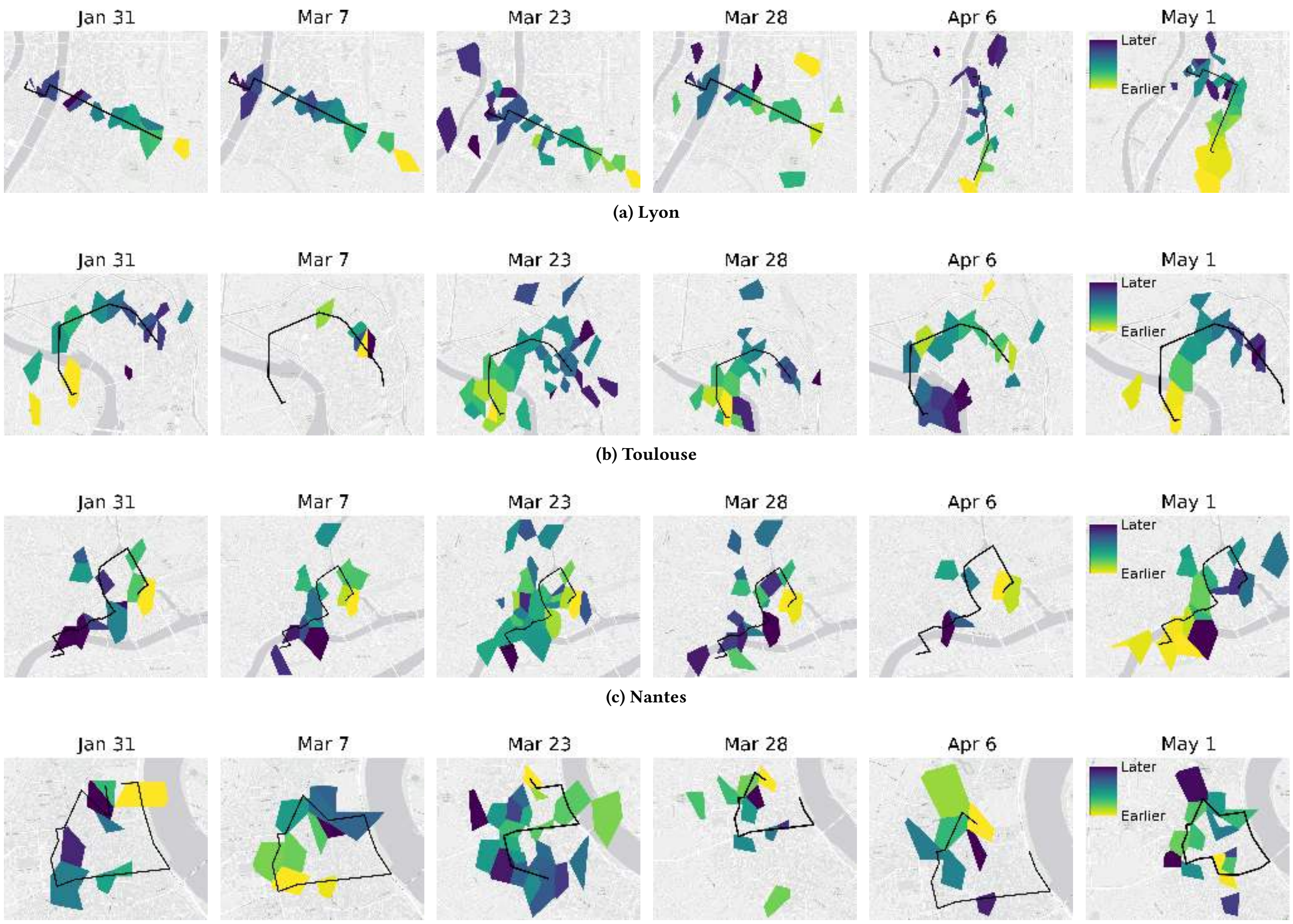
Overall, these results indicate that the combined XGBC and ST-DBSCAN model can identify public marches in various locations based on traffic patterns observed in an initial city. Moreover, given that the successfully detected events were also diverse in terms of magnitude, our detection approach is an important step for developing a unified model working at a nationwide scale.

6 ESTIMATION OF SPATIOTEMPORAL PROTEST ATTENDANCE

Public demonstrations are mainly quantified in terms of the number of participants, which provides qualitative evidence of the magnitude of the event. Thus, organizers and local authorities employ different methods for people counting to announce their attendance estimations, typically during the course of the protest. Based on the outcomes of our proposed framework, we explore the possibility of extending our model so as to estimate of the number of protest attendees over space and time from aggregated network traffic.

We analyze the ten identified protests in Paris, previously shown in Figure 8, and compute a protest-related traffic time series from the total traffic consumed across affected carriers every 5-minute interval. Then, we associate the maximum value of the traffic volume time series with the organizers' and police's attendance estimations. It is well established from a variety of studies the existence of a power relationship between mobile network activity and population [7, 16]. Thus, we model the protest attendance as a power function of the peak traffic volume related to affected carriers. As shown in Figure 10, we find a consistent relationship for both sources of estimations, having $R^2 \geq 0.69$, and thus, supporting the existence of such functional power relationships. However, a closer inspection of the data points in the Figure suggests that the relationship between attendance and peak total traffic could also be explained using a simpler linear regression model. Indeed, when adjusting a linear function to those points, we obtain R^2 values similar to the ones obtained by the power regression: 0.73 and 0.79 for the organizers' and police's estimations, respectively. However, the optimized intercept values of both linear regressions are overly high, associating zero network traffic with almost 230,000 (organizers) and 17,000 (police) attendees. Therefore, modeling the relationship between attendance and the peak of total traffic as a linear function, leaves no room for analyzing medium-sized events, such as most of the demonstrations outside Paris illustrated in Figure 9 and listed in Table 2. In contrast, we show next that a power regression adjusted to massive protests in Paris can be employed to estimate attendance in demonstrations of lower magnitude.

Indeed, we next experiment with estimating the number of participants in the protests outside Paris already considered in Section 5.2.2. We compute the peak traffic volume from the affected carriers and apply the power regression models for organizers' and police's estimations previously adjusted to protests in Paris. Then, we analyze how well the regression values approximate the actual estimations of both entities. On the one hand, the regression model for the organizers' estimations predicted the announced protest attendance with a MAE of 57,045. On the other hand, the regression model for the police's estimations obtained a lower MAE of 7,821. These results suggest that the police's methodology for estimating



(a) Lyon

(b) Toulouse

(c) Nantes

(d) Bordeaux

Figure 9: Inter-city testing: Protest events identified across different French cities based on our proposed detection methodology. Early affected carriers colored yellow, while later affected carriers colored purple.

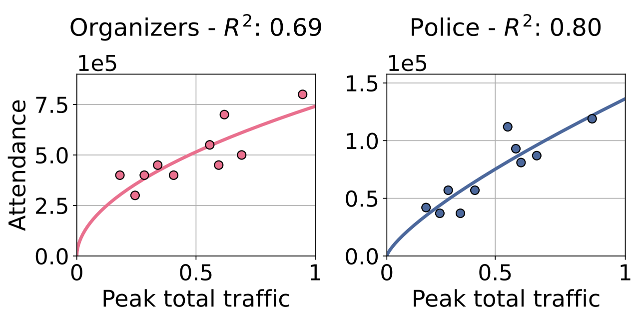


Figure 10: Power regressions of the attendance estimations against the peak total traffic related to a protest event. Real traffic volumes are anonymized.

participation during protests is more consistent across cities. Consequently, the following experiments devise a dynamic estimation of attendees, considering the police's power regression only.

We take advantage of the previously designed models to extend our a-posteriori analysis of public protests. Thus, we extrapolate the power regression models to the entire protest-related traffic time series, obtaining a dynamic estimation of protest attendance along the duration of the identified event. Then, we select two target days of nationwide demonstrations to study the results of the proposed dynamic attendance estimation, March 23 and April 6, which cover a wide range of protest sizes. Indeed, for each city, the demonstration on March 23 is consistently listed among the protests with higher attendance, whereas the demonstration on April 6 is usually one of the less crowded events (see Table 2). Figure 11 shows the time-variant number of participants in both protest days across the five cities analyzed in this study. Intuitively, the dynamic number of participants is generally characterized by a bell-shaped curve, showing the increasing number of participants from the start of

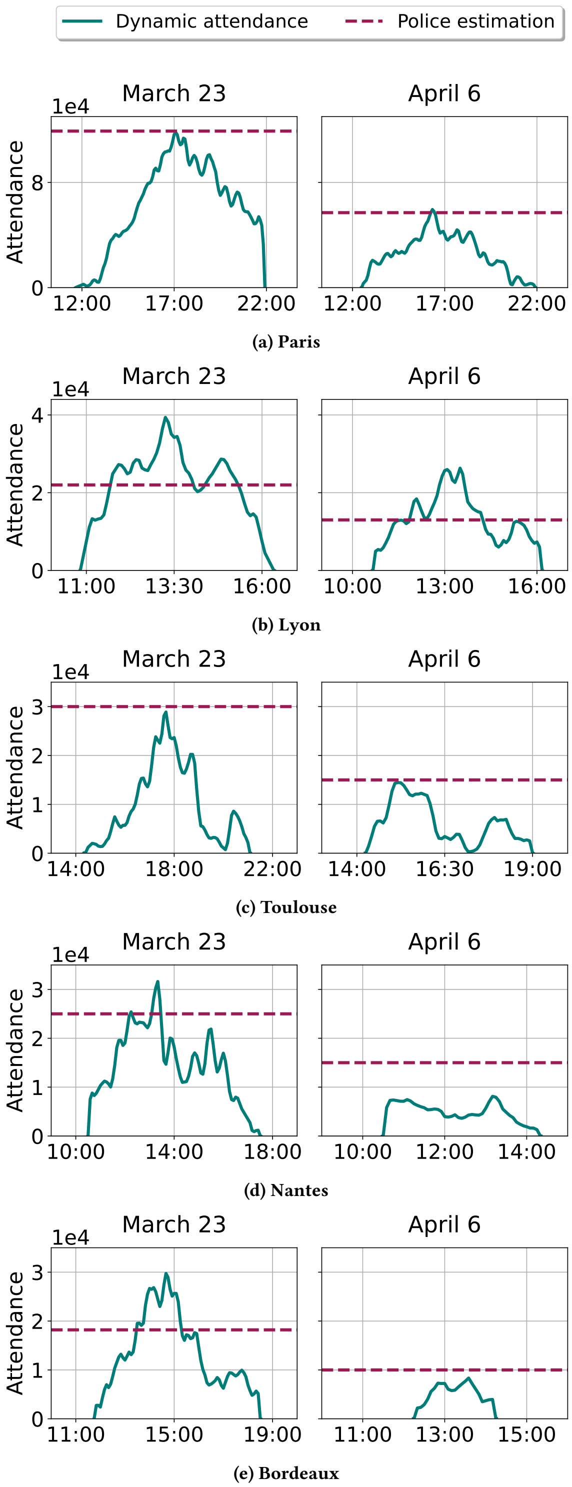


Figure 11: Dynamic attendance estimation in five French cities during the course of two nationwide protest events.

the demonstration achieving a maximum value around the median time of the event's duration. However, we also observe a surge in participants close to the end of some demonstrations (*e.g.*, Figures 11b and 11d), which may be related to external factors such as the geographical area where the protests occurred. Additionally, as these attendance time series are based on the regression of police estimations, we include in the Figures the official numbers (dashed line), which can be compared against the maximum value of the time-variant attendance. Thus, these Figures provide further support for the suitability of a power regression to estimate the number of participants from aggregated volumes of mobile traffic.

Finally, we highlight that the attendance estimation capabilities above can be combined with the spatiotemporal modeling of protests developed in Section 5 so as to perform a high-resolution post-hoc analysis of the progression of a public march. For example, Figure 12 provides a visualization of the evolution of the public protest on March 15 in Paris that associates a time-varying number of participants with a well-defined area and time. These results provide an important basis for the development of a privacy-preserving framework to investigate similar events a-posteriori.

7 CONCLUSIONS AND DISCUSSION

We present a first analysis of the effects of large public protests on the demands for mobile services, and reveal the recognizable footprint that such events leave on the data traffic –in particular at the level of individual services. By leveraging these findings, we show that it is possible to use measurements passively collected by cellular network operators to perform an a-posteriori characterization of the spatiotemporal dynamics of the demonstrations that includes the identification of the paths taken by protesters as well as the estimation of their time-varying number. On the one end, this provides insights that can be useful for operators to respond to these special circumstances, or to local administrations to better organize safety measures in similar future events. On the other end, these results are preliminary to the development of a full-fledged framework to better understand protests or similar public events while preserving the privacy of the participants.

In this work, we could only explore one specific class of public protest linked with the 2023 pension reform in France. While we proved that our modeling approach generalizes well across different events and cities, we do not have evidence supporting its applicability to other types of large manifestations of social unrest occurring in the same or other countries. Yet, we believe that the methodological framework we devised (*i.e.*, the labeling metrics, traffic-based features and models for classification and clustering) may be employed with minimum changes to identify the footprint of a wide variety of large social events (besides protests) in different countries. Instead, the results presented in this work must be considered specific and limited to the targeted case: the mobile service demands affected by these events may not generalize for other forms of events and regions, and may require future exploration from interested parties.

On that note, it is important to stress that the proposed study targets a-posteriori analysis of large events where tens of thousands of people gather, and (*i*) it does not target live monitoring or prediction and (*ii*) it does not apply to smaller populations that

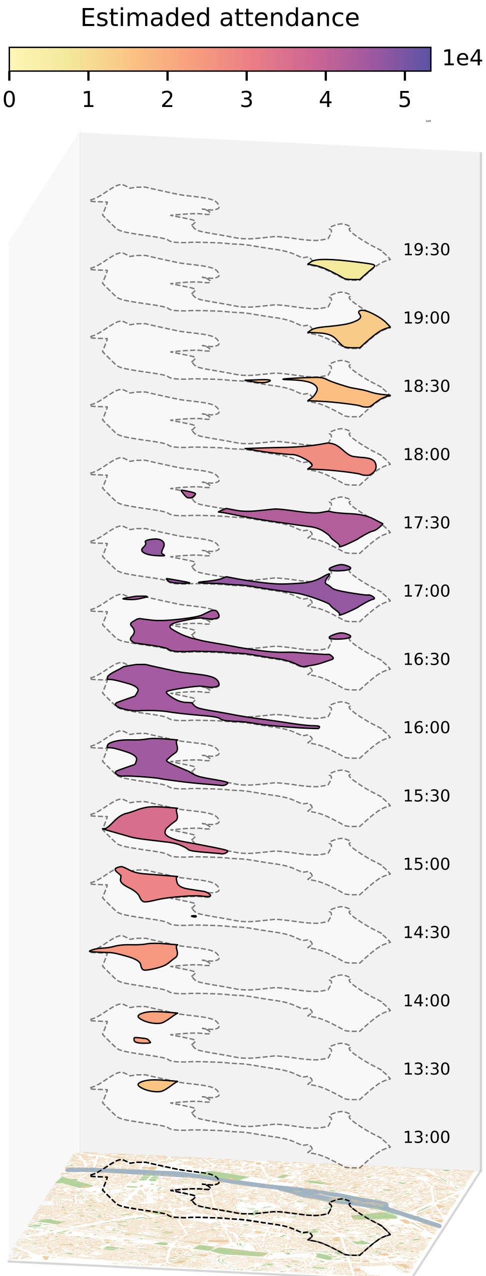


Figure 12: Comprehensive characterization of the public demonstration on March 15 in Paris, showing the spatiotemporal reconstruction of the event along with the dynamic estimation of participants.

are not sufficient to generate detectable changes in the mobile network traffic demand aggregates. Indeed, the footprint on the overall mobile network traffic we observe hinges upon a process of mutual reinforcement in the utilization of specific mobile applications from a large mass of protest participants (*e.g.*, a volume of Twitter posts proportional to the success of the event in terms of number of attendees), which is likely absent in smaller gatherings. While these are limitations to the capabilities and resolution achievable by our methodology, they also provide inherent protection towards secondary uses of our study aimed at surveillance: building on top of our study to develop an on-line or predictive tool is not straightforward (*e.g.*, the clustering stage cannot be performed live), and in all cases, such a tool would not be usable to identify small groups (*e.g.*, tens) of people and even less so individuals, whose impact on the global traffic demand is not detectable from the carrier-level aggregates we have.

Finally, the insights produced by this study, such as the precise progression of the marches, alternate minor routes taken by participants, or their dispersal at the end of the events, do not present risks for the individual organizers or participants as they cannot reveal any personal data and do not allow for individual tracking. Extending the methodology proposed in the paper to track protesters at the user- or device-level would require re-thinking the framework. One would need to look at individual IP flows transiting in the network instead of the coarse aggregates we consider, which is a completely different problem than the one we tackle and which indeed hinges upon personal traffic measurements. We refer the reader to Appendix A for further considerations on the privacy-preserving nature of our study and all related ethics considerations.

ACKNOWLEDGMENTS

This work was supported by NetSense, grants no. 2019-T1/TIC-16037 and 2023-5A/TIC-28944, funded by Comunidad de Madrid; by CoCo5G, grant no. ANR-22-CE25-0016, funded by the French National Research Agency; by BANYAN, grant no. 860239, funded by European Union's Horizon 2020 program; and by AEON-ZERO, project no.TSI-063000-2021-52, funded by Spanish Ministry of Economic Affairs and Digital Transformation and the European Union-NextGenerationEU. We'd also like to thank the reviewers and shepherd for their comments and support.

REFERENCES

[1] Martin R Albrecht, Jorge Blasco, Rikke Bjerg Jensen, and Lenka Mareková. 2021. Collective Information Security in {Large-Scale} Urban Protests: the Case of Hong Kong. In *30th USENIX security symposium (USENIX Security 21)*. 3363–3380.
[2] Nasser Alsaedi, Pete Burnap, and Omer Rana. 2017. Can we predict a riot? Disruptive event detection using Twitter. *ACM Transactions on Internet Technology (TOIT)* 17, 2 (2017), 1–26.
[3] Derya Birant and Alp Kut. 2007. ST-DBSCAN: An algorithm for clustering spatial–temporal data. *Data & knowledge engineering* 60, 1 (2007), 208–221.
[4] Francesco Calabrese, Francisco C Pereira, Giusy Di Lorenzo, Liang Liu, and Carlo Ratti. 2010. The geography of taste: analyzing cell-phone mobility and social events. In *Pervasive Computing: 8th International Conference, Pervasive 2010, Helsinki, Finland, May 17-20, 2010. Proceedings 8*. Springer, 22–37.
[5] M Keith Chen and Ryne Rohla. 2018. The effect of partisanship and political advertising on close family ties. *Science* 360, 6392 (2018), 1020–1024.
[6] Darin Christensen and Francisco Garfias. 2018. Can you hear me now? How communication technology affects protest and repression. *Quarterly journal of political science* 13, 1 (2018), 89.
[7] Rex W Douglass, David A Meyer, Megha Ram, David Rideout, and Dongjin Song. 2015. High resolution population estimates from telecommunications data. *EPJ Data Science* 4 (2015), 1–13.
[8] Bernard Enjolras, Kari Steen-Johnsen, and Dag Wollebaek. 2013. Social media and mobilization to offline demonstrations: Transcending participatory divides? *New media & society* 15, 6 (2013), 890–908.
[9] Ali Mert Ertugrul, Yu-Ru Lin, Wen-Ting Chung, Muheng Yan, and Ang Li. 2019. Activism via attention: interpretable spatiotemporal learning to forecast protest activities. *EPJ Data Science* 8, 1 (2019), 5.
[10] European Union. 2016. *EU General Data Protection Regulation (GDPR): Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on*

the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Retrieved October 18, 2021 from https://gdpr-info.eu/

[11] Gabriel Grill. 2021. Future protest made risky: Examining social media based civil unrest prediction research and products. *Computer Supported Cooperative Work (CSCW)* 30, 5 (2021), 811–839.

[12] Alberto Hermida and Víctor Hernández-Santaolalla. 2018. Twitter and video activism as tools for counter-surveillance: the case of social protests in Spain. *Information, Communication & Society* 21, 3 (2018), 416–433.

[13] Andreas Janecek, Danilo Valerio, Karin Anna Hummel, Fabio Ricciato, and Helmut Hlavacs. 2015. The cellular network as a sensor: From mobile phone data to real-time road traffic monitoring. *IEEE transactions on intelligent transportation systems* 16, 5 (2015), 2551–2572.

[14] Olu Jenzen, Itir Erhart, Hande Eslen-Ziya, Umut Korkut, and Aidan McGarry. 2021. The symbol of social media in contemporary protest: Twitter and the Gezi Park movement. *Convergence* 27, 2 (2021), 414–437.

[15] John T Jost, Pablo Barberá, Richard Bonneau, Melanie Langer, Megan Metzger, Jonathan Nagler, Joanna Sterling, and Joshua A Tucker. 2018. How social media facilitates political protest: Information, motivation, and social networks. *Political psychology* 39 (2018), 85–118.

[16] Ghazaleh Khodabandelou, Vincent Gauthier, Marco Fiore, and Mounim A El-Yacoubi. 2018. Estimation of static and dynamic urban populations with mobile network metadata. *IEEE Transactions on Mobile Computing* 18, 9 (2018), 2034–2047.

[17] Rostyslav Korolov, Di Lu, Jingjing Wang, Guangyu Zhou, Claire Bonial, Clare Voss, Lance Kaplan, William Wallace, Jiawei Han, and Heng Ji. 2016. On predicting social unrest using social media. In *2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)*. IEEE, 89–95.

[18] Jun Liu. 2015. Communicating beyond information? Mobile phones and mobilization to offline protests in China. *Television & New Media* 16, 6 (2015), 503–520.

[19] Stefania Milan and Sérgio Barbosa. 2020. Enter the WhatsApper: Reinventing digital activism at the time of chat apps. *First Monday* (2020).

[20] Christina Neumayer and Luca Rossi. 2018. Images of protest in social media: Struggle over visibility and visual narratives. *New Media & Society* 20, 11 (2018), 4293–4310.

[21] Christina Neumayer and Gitte Stald. 2014. The mobile phone in street protest: Texting, tweeting, tracking, and tracing. *Mobile Media & Communication* 2, 2 (2014), 117–133.

[22] Susan C Pearce and Jaylen Rodgers. 2020. Social media as public journalism? Protest reporting in the digital era. *Sociology Compass* 14, 12 (2020), 1–14.

[23] Assaf Rotman and Michael Shalev. 2022. Using location data from mobile phones to study participation in mass protests. *Sociological Methods & Research* 51, 3 (2022), 1357–1412.

[24] Polina Rozenshtein, Aris Anagnostopoulos, Aristides Gionis, and Nikolaj Tatti. 2014. Event detection in activity networks. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. 1176–1185.

[25] Angelica Salas, Panagiotis Georgakis, and Yannis Petalas. 2017. Incident detection using data from social media. In *2017 IEEE 20th International conference on intelligent transportation systems (ITSC)*. IEEE, 751–755.

[26] Zachary C Steinert-Threlkeld, Delia Mocanu, Alessandro Vespignani, and James Fowler. 2015. Online social networks and offline protest. *EPJ Data Science* 4, 1 (2015), 1–9.

[27] Tin-yuet Ting. 2020. From ‘be water’to ‘be fire’: nascent smart mob and networked protests in Hong Kong. *Social Movement Studies* 19, 3 (2020), 362–368.

[28] Vincent A Traag, Rick Quax, and Peter MA Sloot. 2017. Modelling the distance impedance of protest attendance. *Physica A: Statistical Mechanics and its Applications* 468 (2017), 171–182.

[29] Aleksandra Urman, Justin Chun-ting Ho, and Stefan Katz. 2021. Analyzing protest mobilization on Telegram: The case of 2019 anti-extradition bill movement in Hong Kong. *Plos one* 16, 10 (2021), e0256675.

[30] Jeroen Van Laer. 2010. Activists online and offline: The internet as an information channel for protest demonstrations. *Mobilization: An International Quarterly* 15, 3 (2010), 347–366.

[31] Steven Lloyd Wilson. 2017. Detecting mass protest through social media. *The Journal of Social Media in Society* 6, 2 (2017), 5–25.

[32] Donghyeon Won, Zachary C Steinert-Threlkeld, and Jungseock Joo. 2017. Protest activity detection and perceived violence estimation from social media images. In *Proceedings of the 25th ACM international conference on Multimedia*. 786–794.

[33] Congyu Wu and Matthew S Gerber. 2017. Forecasting civil unrest using social media and protest participation theory. *IEEE Transactions on Computational Social Systems* 5, 1 (2017), 82–94.

A ETHICS

Our work builds on mobile network traffic generated by users of a nationwide cellular infrastructure. Specifically, we leverage traffic demands aggregated over 5-minute intervals at the level of individual cells, which are generated from network measurements carried out in the target infrastructure.

The traffic measurements used throughout this work were collected by the operator for network management and research purposes, and temporarily stored within a secure platform at their own premises. The temporal aggregation was carried out in the same platform by personnel of the network operator, in full compliance with Article 89 of the General Data Protection Regulation (GDPR) [10] of the European Commission and other applicable national and international regulations. The data collection and processing were approved by the Data Protection Officer (DPO) of the operator in the context of a collaborative research project.

While the original network measurements contained personal identifiers (*e.g.*, the International Mobile Subscriber Identifier, or IMSI) and sensitive data (*e.g.*, locations of visited cells, or mobile services consumed) about individual users, the aggregated traffic demands do not contain personal identifiers or sensitive information. The spatiotemporal aggregation (*i.e.*, carrier-level over 5-minute intervals) merges information from tens to hundreds of users and ensures that no single data subject can be re-identified, and that the resulting data does not configures as personal data in the GDPR acceptation.

The researchers involved in the work presented in this paper only had access to such aggregated and privacy-preserving traffic data for the purpose of carrying out the study. Our dataset and research do not involve risks for the mobile subscribers, while generating important insights about its potential as an alternate type of data capable of remotely sensing large-scale manifestations.

B PROTEST INFORMATION

Table 2 presents the estimations of attendance of the protests considered in our study across multiple cities, as provided by the protest organizers as well as by local police forces.

Protest		Attendance estimation	
City	Date	Organizers	Police
Paris	January 31	500,000	87,000
	February 7	400,000	57,000
	February 16	300,000	37,000
	March 7	700,000	81,000
	March 15	450,000	37,000
	March 23	800,000	119,000
	March 28	450,000	93,000
	April 6	400,000	57,000
	April 13	400,000	42,000
	May 1	550,000	112,000
Lyon	January 31	45,000	25,000
	March 7	50,000	25,000
	March 23	55,000	22,000
	March 28	30,000	12,500
	April 6	32,000	13,000
	May 1	45,000	17,000
Toulouse	January 31	80,000	34,000
	March 7	120,000	27,000
	March 23	150,000	30,000
	March 28	150,000	23,000
	April 6	90,000	15,000
	May 1	100,000	13,500
Nantes	January 31	60,000	28,000
	March 7	75,000	30,000
	March 23	80,000	25,000
	March 28	60,000	18,000
	April 6	50,000	15,000
	May 1	80,000	17,500
Bordeaux	January 31	75,000	16,500
	March 7	100,000	16,500
	March 23	110,000	18,200
	March 28	80,000	11,000
	April 6	60,000	10,000
	May 1	130,000	12,000

Table 2: City, date, and estimated participation of the protest events investigated in our study.