



Modeling and understanding the impact of COVID-19 containment policies on mobile service consumption in French cities

André Felipe Zanella^{1,2*} , Stefania Rubrichi³, Zbigniew Smoreda³ and Marco Fiore¹

*Correspondence:

andre.zanella@imdea.org

¹IMDEA Networks Institute, Madrid, Spain

²Universidad Carlos III de Madrid, Madrid, Spain

Full list of author information is available at the end of the article

Abstract

Amid recent studies that have been exploring the wide impact that COVID-19 containment policies have had across sectors and industries, we investigate how mobility restrictions enacted in French cities during the later stages of the pandemic have affected the usage of smartphones and mobile applications. Leveraging a large-scale dataset of over 14 billion records, we unveil and quantify the substantial incidence of the different measures enforced in urban France to combat the COVID-19 epidemic on mobile service consumption. We present a simple but effective spatial linear model that can relate changes occurring at fine-grained spatial zoning in both global and per-service traffic to a limited set of socioeconomic indicators. Our model unveils some of the mechanisms that drove the significant evolution of mobile data traffic demands during the pandemic. It allows observing how the demand for mobile services has been affected by COVID-19 in very different ways across urban areas characterized by diverse population density, income levels and leisure area presence. It also discloses that usages of individual smartphone applications have been impacted in highly heterogeneous ways by the pandemic, even more so when considering the composite impacts of different transitions between periods characterized by diverse restrictions. Our results can aid governments in understanding how their measures were received across the space and different portions of population, and network operators to comprehend changes in usage due to extraordinary events, which can be used to optimize service provisioning.

Keywords: COVID-19; Mobile network traffic; Mobile services; Remote sensing

1 Introduction

The COVID-19 pandemic impacted people's lives in a manner never seen before in the digital era. Governments worldwide had to adopt measures to restrict the spread of the virus, resulting in major changes in the routine of populations worldwide. In addition to reshaping mobility [1–6], social relationships [7–9], work relationships [10–12] as well as dietary and exercise routines [13, 14], those measures also induced unprecedented alterations in the utilization of the many digital services that nowadays permeate our lives. As representative examples, substantial effects of COVID-19 containment policies have

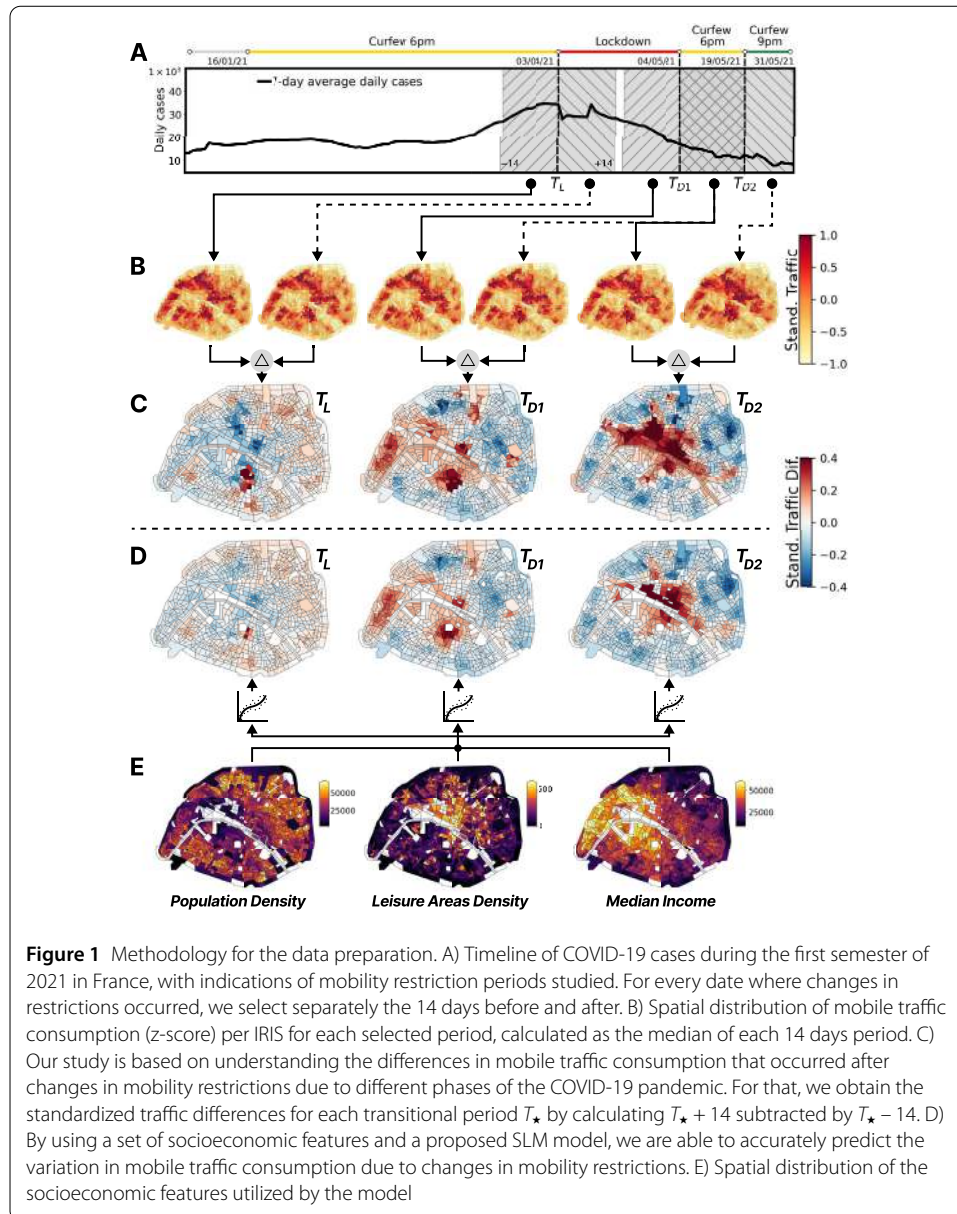
© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

been demonstrated on energy consumption [15–17], demands for both fixed [18, 19] and mobile [20, 21] Internet, or the operations of small to worldwide communication networks [22–25]. Social media consumption also helped researchers understand political narratives involving COVID-19 [26], in understanding how populations react through the different phases of the pandemic through the analysis of hashtags popularity [27], or even using spread out awareness and gather participants for classic surveys [28].

Prior studies on digital usages focus on the earlier stages of the pandemic, *i.e.*, the so-called first wave and its consequent lockdowns, and investigate its effects at coarse spatial granularity (*e.g.*, aggregating data at the level of countries or regions) [21] or without separating individual digital applications (*e.g.*, video streaming versus social media) [18, 20, 25]. The existing literature leaves knowledge gaps in terms of how the consumption of digital services has been impacted by the pandemic at a higher intra-city resolution. This is a relevant lacuna, since the utilization of mobile applications is known to be especially rich and heterogeneous within individual cities, where individual services present unique temporal usage patterns and spatial distributions dependent on the nature of urban neighborhoods [29, 30]. Understanding how a large-scale epidemics like COVID-19 has affected such strong spatiotemporal dynamics is an important piece of a comprehensive characterization of the effects of the pandemics on digital consumption.

In this work, we contribute to closing the gaps above and explore the impact of policies enacted during later stages of the COVID-19 pandemic (*i.e.*, waves following the first one), across individual neighborhoods in the largest cities of a major European country (*i.e.*, France) and at the level of single mobile services. Our approach yields a high-resolution analysis of the consequences of governmental policies for COVID-19 containment on digital service consumption from mobile devices, and carry substantial explanatory potential via the socioeconomic indicators typically associated to such geographical zoning. This study can be useful for many entities: Firstly, mobile operators and internet providers are able to have better visibility on how large-scale disruptions can affect their per-application demand, which can be useful in resource allocation problems [31]. These disruptions impacted applications in different levels of the chain, from the way users interacted with messages on WhatsApp [32] to the traffic demands observed at the infrastructure of Facebook [33], which connects to the access that mobile operators provide from end-user perspective all the way to the servers of the app, affecting directly the loads the operator experiences. The insights provided by mobile phone data can be also useful for governments and political entities, which can leverage passively collected data in order to understand the efficiency of their containment measurements in such large-scale events [34, 35]. Overall, the usage of non-pharmaceutical interventions [36] and digital technologies [37] throughout the COVID-19 pandemic provided an interesting new perspective in the study of epidemics, disrupting events and societies, as well as introducing new challenges in the field of data privacy [38]. Our end goal is to showcase the potentials of the passively collected, per application mobile phone traffic consumption for such activities.

This paper is organized as follows: Sect. 2 gives an in-depth look of our methodology, with details about the pipelines for data collection, processing and analysis. In Sect. 3 we present our results, separating the section in results related to overall trends in mobile traffic consumption and later regarding individual mobile services. Finally, Sect. 4 we present some implications about our observed results on the impacts of restriction policies in mo-



mobile services consumption and on Sect. 5 we conclude and present some perspectives on how this study can be of interest in key societal areas.

2 Methods

2.1 Timeline of COVID-19 policing in France

During the first two years of the COVID-19 pandemic, France experienced 3 nationwide lockdowns, intertwined with curfew periods. Day-by-day information is available in [39], and we provide here and in Fig. 1A a summary of the timeline of reported daily COVID-19 cases and associated containment measures during the studied period. The first year had an initial nationwide lockdown (17/03/20 - 10/05/20) with severe restriction measures that closed the majority of leisure locations and forbidding social gatherings; following a summer with looser restrictions, cases rapidly rose leading to a second nationwide lock-

down (31/10/20 - 14/12/20) to contain the virus during fall. As the second lockdown came to an end, the French government put in place a progressive lifting of restriction measures with a strict curfew from 8pm to 6am, which was later tightened (16/01/21) from 6pm to 6am. The majority of current works focus on periods until those dates.

Our study aims to fill the gap in understanding the later stages of the pandemic, specifically during the first half of 2021. The new year started with the aforementioned curfew, only allowing commuting to/from work or movements for family and medical reasons between 6pm and 6am; short displacements inside a radius of 1 km from the place of residence were also permitted, whereas non-essential trips outside of curfew hours were restricted to a radius of 10 km from the place of residence. This curfew encompassed the first 3 months of 2021, when cases started to quickly rise again pushing French authorities to enact a plan for a third nationwide lockdown (03/04/21 - 04/05/21), during which schools were closed and all non essential travel prohibited, but overall with lighter restrictions than the previous two lockdowns. After the third lockdown, France experienced a progressive lift of restrictions, with a first phase (starting on 05/05/21) initially characterized by a curfew between 6pm and 6am with restrictions similar to the one before the lockdown; throughout this phase, primary schools first reopened (from 25/04/21), 10-km restrictions for trips outside curfew hours were cancelled, and finally secondary schools also started operating again. A second phase (from 19/05/21) initiated with the reopening of most non essential shops, cinemas, theatres, museums and places to eat, and curfew start pushed to 9pm.

2.2 Data collection

To achieve our objective of understanding the consequences of COVID-19 containment policies in digital services consumption, we collect mobile data traffic information in the production network of Orange servicing the metropolitan France territory.

The mobile service generating each IP session was identified via a combination of Deep Packet Inspection and proprietary traffic classifiers deployed by the operator. Each IP session was geo-referenced at the level of Base Transceiver Station (BTS), leveraging the User Location Information (ULI) contained in Packet Data Protocol (PDP) Contexts and Evolved Packet System (EPS) Bearers over the GPRS Tunneling Protocol control plane (GTP-C). Based on this, the operator computed the hourly traffic demand for each mobile service at every BTS, by aggregating the uplink and downlink traffic of all users attached to the same BTS during one-hour intervals.

The traffic was generated by mobile subscribers using 2G, 3G and 4G connectivity. This basically represents the full mobile demand in the studied period, as 5G generated less than 1% of the total mobile data traffic as of May 2021 [40]. Our data was collected at 1 h intervals for each BTS and service.

2.3 Conversion of mobile traffic spatial granularity

The data collected for this study encompasses 370,189 BTSs servicing Metropolitan France, for the period between October 30th 2020 and May 31st 2021. The spatial granularity at BTS level can be represented by Voronoi geometries that roughly represent the coverage area of each BTS, with traffic uniformly distributed over each geometry [41]. This is not ideal to understand intra-city dynamics, which are better represented by sub-municipal divisions called Islets Regrouped for Statistical Information (IRIS), defined by France's National Institute of Statistics and Economic Studies (INSEE) to better

reflect changes in urbanization, geography and demographics in sub-municipal statistics in France [42], acting similarly to US census tracts. Metropolitan France is represented by over 49.000 IRIS units.

We are able to convert the mobile traffic associated to the Voronoi geometries to IRIS geometries by calculating the intersection ratio of the Voronoi that represents the space covered by each BTS with the set of IRIS from each city, as follows. Considering that T_n represents the traffic Voronoi n that intercept IRIS x . The traffic T_x of the respective IRIS will be composed by the sum of the ratio of traffic of every Voronoi $n \in N$ that intersects it, which is directly related to the area of intersection between the two geometries, and defined as:

$$T_x = \sum_n^N r_n T_n, \quad (1)$$

where N is the set of Voronois that intersect IRIS x , r_n is the area ratio of the intersection of Voronoi n with IRIS x and T_n is the total traffic of the respective Voronoi. We run this conversion for our entire BTS set covering metropolitan France.

2.4 Measuring changes in mobile service consumption across containment strategies

To explore the effects on mobile traffic consumption due to different restriction measures applied by the French government, our study explores three key transitions T_\star of mobility restrictions at the later stages of the COVID-19 pandemic. Specifically, $\star \in \{L, D1, D2\}$, denoting the change of measures when (i) entering the third lockdown, (ii) leaving the third lockdown and transitioning into the following curfew, and (iii) further lifting of restrictions into final light curfew, respectively, as seen in Fig. 1B. For every transition T_\star we obtain traffic snapshots that represent the mobile demands before and after the transitions. To this end, we aggregate mobile service demands in the period of 14 days immediately preceding the transition, as well as in the 14 days after the transitional moment has happened, as exemplified in Fig. 1B.

We obtain the median traffic of each IRIS over the periods before $T_\star - 14$ and after $T_\star + 14$, followed by dividing the median traffic of each IRIS by its area in km^2 to remove the bias in traffic volume at large areas. We standardized the before/after periods of each city by calculating the z-score of the traffic density, subtracting the value of each IRIS by the mean and dividing by the standard deviation of the selected set of IRIS of each city and period, so as to ensure that comparisons among different time periods are fair and not biased by confounding factors (e.g., changing traffic volumes). This results in values that highlight the importance of each IRIS for the traffic generated at the selected area and period. In addition, standardization allows anonymizing the raw traffic values, which we cannot share due to contractual obligations with the network operator that provided the data for this research. Finally, we calculate a traffic difference map that represents T_\star by subtracting $T_\star + 14$ and $T_\star - 14$, highlighting the effects caused by restrictions imposed on cities at the IRIS-level; as an example, Fig. 1C shows the resulting traffic difference maps for Paris (equivalent plots for all cities are in Figure S1 of SI).

The analysis of transitions focuses on the evenings hours of workdays (Monday through Friday, from 7pm to 1am). The rationale is that such periods best capture differences induced by COVID-19 restrictions, for several reasons: (i) these hours represent a moment

of the day when people conclude their workday and commute, are at their home or visit leisure locations (if allowed by the restrictions); (ii) the time interval matches the peak time of mobile traffic consumption in France during the pandemic [21]; and, (iii) these hours help relate changes of traffic with geo-referenced socio-economic information that is related to the place of residency as well as with data about leisure areas in each city. Our study considers the 10 cities with highest population in France: Paris, Marseille, Lyon, Toulouse, Nice, Nantes, Montpellier, Strasbourg, Bordeaux and Lille; in each city, mobile service demands are aggregated at the level of IRIS zones.

2.5 Socioeconomic data and features selection

To correlate changes of traffic at IRIS level, we utilize a set of socioeconomic features collected at the same granularity. We collected our features from three sources provided by INSEE. The first base comes from the 2017 Census in France [43], which is the most up to date Census in France that includes collection at IRIS level. From the complete set, we chose to use the area population and calculate its density by dividing population numbers by the area in km^2 of each IRIS; this results in a better geographical distribution of values, and also removing the bias that large areas may induce on the features. The second base is related to revenue; as this was not covered on the 2017 Census we utilize a 2018 survey on revenue, poverty and quality of live [44] to obtain the median income in Euros per IRIS. Finally, the third base utilized is SIRENE base of enterprises and establishments in France [45], which indicates the date of opening, if the place was still active at the time of collection (or if not, the date of closure), as well as its latitude and longitude. From this base, we extract the number of restaurants, bars and non essential stores (i.e. shopping centers, tech stores), and calculate the total number per IRIS of those establishments and later divide by the IRIS area to obtain the density of leisure locations.

The spatial distribution of those features for Paris can be seen on Fig. 1E, while the distribution for the remaining cities of this study can be seen on Figure S2 of the Additional file 1. To help guide the analysis, we also provide the classification of INSEE for IRIS usage based on their population on Figure S12 of the Additional file 1.

2.6 Calculating the spatial lag of traffic difference

By observing the spatial distribution of traffic differences for each IRIS, we note a presence of spatial noise in our collected data. This leads to spatial cluster patterns becoming less evident and an observed drop in fitting performance of our model. To overcome this problem, we will apply a spatial lag on the data [46], by calculating the neighborhood graph between IRIS using Queen's distance, and the resulted smoothed traffic difference will be the average between the value of the selected IRIS and its set of neighbors. This process will help highlight spatial patterns, as well as perform a filter on outliers (e.g., in a neighborhood all IRIS see an increase in traffic and there's a single one with a sharp decrease and no apparent reason for this behavior).

It is worth noting that the choice of using a spatial lag with Queen's weights is the result of extensive analyses. First, when testing the model to predict traffic differences caused by changes in government restrictions (detailed next on Sect. 2.7), the simple exclusion of spatial weights significantly reduces the correlation between real and predicted values, making the model unsuitable for downstream coefficient analysis (as observed in Figure S6A of the Additional file 1). There are then several options to calculate the spatial lag feature, based on different spatial weights techniques. One approach is to utilize

distance-based spatial weights, such as Kernel functions and k-nearest neighbors (KNN); While both techniques have good results when compared to not using spatial weights (as observed on Figures S6C and S6D of the Additional file 1), their performance is worse than contiguity-based techniques. For the latter, we test both Queen's and Rook's weights, which result in similar model performance and coefficient values (as observed over Figure S6B and S7). We believe results in these cases are similar due to the large number of IRIS geometries across cities and their complex geometries, which resulted in comparable spatial weight matrices. We choose to proceed with Queen's weights, but highlight that results would be consistent with Rook's weights.

2.7 Modeling changes in mobile traffic consumption in relation to socioeconomic features

We utilize a model for each transitional period T_* in order to use our selected set of socioeconomic features to predict the mobile traffic differences due to mobility restriction for all cities (which is performed first for total traffic and later for each mobile service's traffic). We initially evaluated using a simple linear regression model, but quickly noticed that R2 and Pearson correlation values were significantly low, which could be linked to spatial autocorrelation error on the residuals originally mentioned in Sect. 2.6. We concluded the necessity of a linear model that takes into account the spatial characteristics of the data and selected a spatial lag model (SLM) [47], which is a linear model and can be solved with traditional least squares methods and uses the spatial lagged version of the dependent variable as a regressor. It is described as:

$$y = \lambda W y + \beta_1 X + \epsilon \quad (2)$$

where y is our dependent variable, X is the matrix of features, $W y$ will be the spatial lagged feature of the dependent variable y and $\epsilon \sim N(0, \sigma^2)$ will be a Gaussian noise term. Our goal is to estimate $[\lambda, \beta]$. In addition, we have to deal with two additional problems in our data: heteroscedasticity and outliers. We can tackle both at the same time by treating our SLM as a robust regressor and solving it with an iteratively reweighted least squares (IRLS) [48, 49]. Those changes significantly reduced the spatial dependency of our model. Also, since our analysis is focused on coefficients values and not in the predictability power of the model, there's no necessity to split the set in train and validation.

2.8 Selection of mobile services

The collected data contains over 250 services with different degrees of traffic consumed. Specially when analyzing at IRIS level, noise can be significantly present in services that are less popular even with the measures applied in Sect. 2.6, so not all services can be studied on such a small level of geographical granularity. To overcome this and clean our data, we ranked all services by the total traffic generated by each. By looking at their spatial distributions of traffic consumption, we removed additional applications that had noisy patterns at IRIS level. We individually perform for each services the same data processing methodology described for total traffic in Sect. 2.4.

The selected set of 38 applications represents 85% of the traffic generated by the full set of services. Their traffic consumption rankings, as well as the full distribution of traffic from all applications, can be seen on Figures S8 and S9 of the Additional file 1. The considered set of applications involves mainly very popular applications (*e.g.*, Youtube, Netflix,

or WhatsApp, just to cite a few representative examples), as well as a few applications with lower traffic but with a typical usage that could be expected to be directly impacted by restriction measures put in place (e.g., Tripadvisor and Foursquare, which are mainly used to discover restaurants and locations within cities, as well as their ratings by other users).

A few notable removed applications were Instagram, Facebook and other related applications from Meta. Even though these applications had a significant traffic share, we detected problems during collection which resulted in visible changes in their usage patterns which were not related with COVID-19 restriction measurements and made impossible a long-term analysis. Other exclusions were mainly related to applications that did not have their usages clearly related to potential effects of the restriction measures, such as the case of application stores (e.g., Apple Store, Google Play and Samsung Galaxy Store), whose traffic is primarily determined by automated application updates that run in background and are unrelated to users' behavior.

2.9 Hierarchical clustering of coefficients

After fitting the model described in Sect. 2.7 to the traffic obtained for each selected service of Sect. 2.8, we are left with a significant sum of coefficients to analyze in regards to how socioeconomic features related to changes to each mobile service. This number scales quite rapidly: each of the 38 analyzed services will generate 3 coefficients per model of each transitioning period T_* , resulting in 114 unique coefficients per T_* and 342 coefficients in total. Therefore, to help guide this analysis, we choose to cluster coefficient values in order to highlight patterns across all 38 services.

While many techniques exist for unsupervised clustering where the ground-truth labels are not known, we employ the agglomerative clustering technique [50] due to its simplicity and high interpretability. This technique follows a *bottom-up* methodology, by initially considering that every entry (a set of 3 coefficients obtained from the model of a service during T_*) is a separate cluster, and then utilizing a distance metric to compare the similarity between clusters and iteratively group pairs of clusters until a single cluster is left. Our implementation of the agglomerative clustering employs the Euclidean distance to compare coefficients across services in the sample period T_* , as well as Ward's method [51] to group samples into clusters that minimize the total variance within each cluster.

When selecting the ideal number of clusters across all T_* , we first tried quantitative techniques such as the e.g., Silhouette score and Dunn index to identify the best grouping; yet these rules returned very inconsistent results across metrics and periods, making clusters unreliable and the identification of patterns through the different T_* intervals impossible. We then opted for a manual selection of all possible clusterings returned by the agglomerative algorithm at each iteration, based on a visual inspection of the dendrogram distances and on the heatmaps of distance matrices. This allowed us to identify more clear clusters in a same number of with homogeneous patterns across the periods. Ultimately, the agglomerative clustering returned 5 clusters of service coefficients in each period T_* . We note good separation between the clusters on their distance matrix, resulting in a balance between services that had significant similarity of coefficients, without resulting in far too many clusters (which goes against using this technique to find and aggregate patterns). These 5 clusters showcase clear major trends between the services belonging to each, as well as differentiate services across clusters. These patterns, the similarities between certain sets of applications, as well as how they relate to each transitional period T_* will be explored on Sect. 3.3.

3 Results

3.1 Changes in total traffic at city level during the pandemic

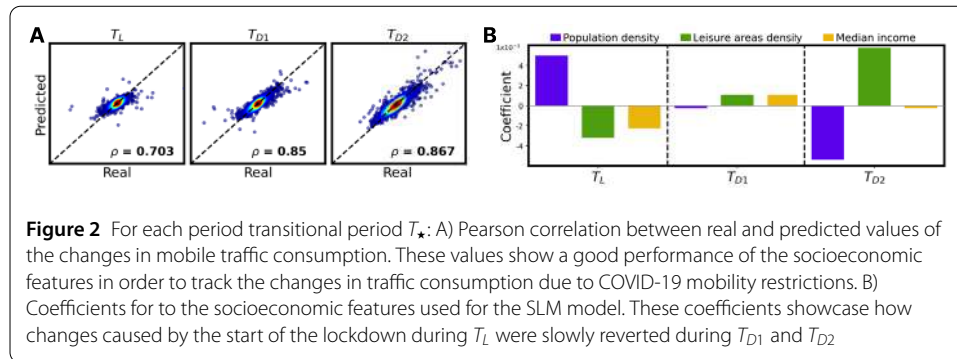
Figure 1C shows maps of the changes in the overall mobile service demand in Paris during the transitional periods T_L , T_{D1} , T_{D2} (plots for all other cities are in Figure S1 of SI), in terms of the increase or decrease of mobile traffic consumption generated by the transition with respect to the previous period. We note an evident heterogeneity of behaviors across space (*i.e.*, IRIS zones in a same city) and time (*i.e.*, transitions T_L , T_{D1} and T_{D2}), which suggest diverse responses of the urban tissue to the COVID-19 restrictions in terms of digital usages. Such a response is in fact so different that, at each transition, both growth and reduction of the relative mobile traffic demands are concurrently observed in a same city depending on the neighborhood considered. The phenomenon also exhibits clear spatial dependencies, as areas experiencing similar increase or decrease of mobile digital consumption tend to be geographically clustered for all studied transitions.

For instance, in Paris neighborhoods in the center and to the West of the urban region experience substantial reduction of mobile traffic activity induced by the lockdown in T_L , whereas areas to the East and South are those characterized by the highest growth of consumption during that same transition. West Paris recovers the lost digital service usage at the end of the lockdown in T_{D1} , whereas the city center does so only with when additional restrictions are lifted in T_{D2} . During both transitions to curfews T_{D1} and T_{D2} , the East and South regions display instead a progressive reduction of the demand for mobile services. As in the example of Paris above, all cities examined show cyclic patterns where neighborhoods are characterized by either (*i*) a reduction of demand during the lockdown, followed by an increased consumption in the subsequent curfews, or (*ii*) the opposite sequence of growth in lockdown and decrease with curfews.

3.2 Socioeconomic explanation to mobile traffic usage changes

The clustered patterns of IRIS zones that display different responses to the same governmental restrictions suggest a connection to the similarly clumped socioeconomic characteristics of urban areas. To explore this relation, we propose the spatial lag model described in Sect. 2.7 that uses socioeconomic indicators as regressors to predict mobile traffic variations in all of the 10 large cities considered in our study. All features input to the model are standardized by subtracting their mean and dividing by their standard deviation. The considered features are the population density, median income and leisure space presence (*i.e.*, the spatial density of restaurants and non essential stores), whose geographical distribution is depicted in Fig. 1E for Paris (equivalent maps for the remaining cities are shown in Figure S2 of SI). These socioeconomic features are loosely correlated among them (Pearson's correlation coefficients of 0.21–0.49, as detailed in Figure S3a of SI) and have low individual correlation with the mobile demand differences across transitions T_L , T_{D1} and T_{D2} (absolute value of Pearson's correlation coefficients below 0.2, as seen on Figure S3b of SI).

An SLM fed by the three socioeconomic features above can explain the changes in mobile service demands induced by COVID-19 regulations on each IRIS of the 10 cities with good accuracy. As shown in Fig. 2A, Pearson's correlation coefficients range between 0.70 and 0.86, depending on the transitional period considered. Correlation values stay satisfactory when computed per city, with Pearson's coefficient that are above 0.6 in the vast majority of cases (see Figure S4 of SI), which supports the generality of the model and of



the predictive power of the selected features. An illustrative example of the quality of the model is provided for Paris in Fig. 1D (estimated maps for all remaining cities are in Figure S5 of SI): the major spatial trends in mobile traffic changes are well reproduced across T_L , T_{D1} and T_{D2} .

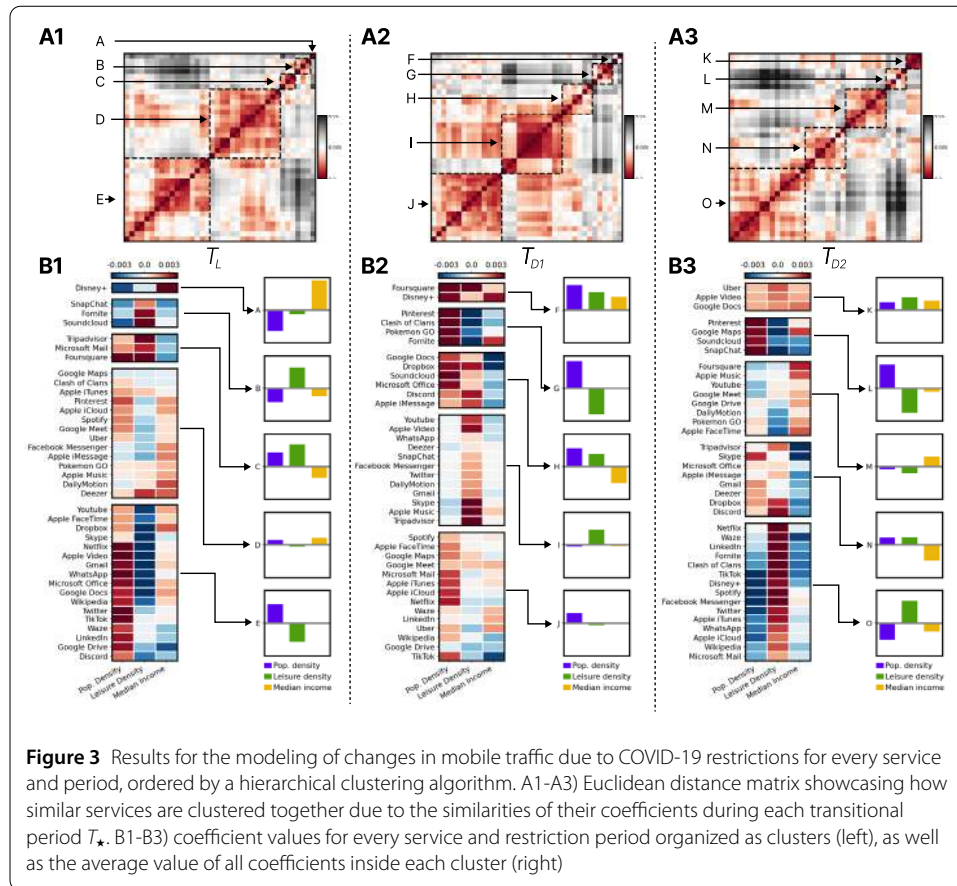
The SLM coefficients are shown in Fig. 2B and display notable swaps of sign across periods that align with the cyclic effects previously seen on the traffic difference maps. Specifically, densely populated areas of residential destination (*i.e.*, with low presence of leisure points of interest) and with lower income experienced traffic surges when the nationwide lockdown was put in place in T_L . In such neighborhoods, there is a tendency not to contract expensive fixed-line Internet access, and to use cellular connectivity as a cheaper replacement: being forced at home from the lockdown, local inhabitants consumed a substantially increased amount of mobile data traffic. This pattern is also consistent with multiple effects that include: higher-income households dropping cellular access in favor of high-speed fixed Internet at their home premises; richer families moving to second houses in the countryside during lockdowns, or leisure-dense areas suffering reduced visits under restrictions to non-essential mobility.

The transition to T_{D1} gives initial signs of reversed trends. With the end of the harsher lockdown limitations, the wealthier share of the population moved back to their urban residences, as indicated by the positive impact of median income on the growth of mobile data traffic usage. Similarly, partial resumption of leisure activities yielded an increase of mobile service demands in the associated neighborhoods.

Interestingly, the usage of cellular traffic in densely populated areas of the French cities remained unaffected by the transition from the lockdown to the first curfew period, *i.e.*, during T_{D1} . The recovery in those areas only occurred with the removal of additional restrictions, as shown by the highly negative SLM coefficient of this feature during T_{D2} , which denotes a substantial drop in mobile data service consumption, *i.e.*, a normalization of demands. This last transition is also linked to a full return to leisure activities in French urban areas, indicated by the positive SLM coefficient and the resulting increased demands for cellular access.

3.3 Impact of containment policies on the demand for individual services

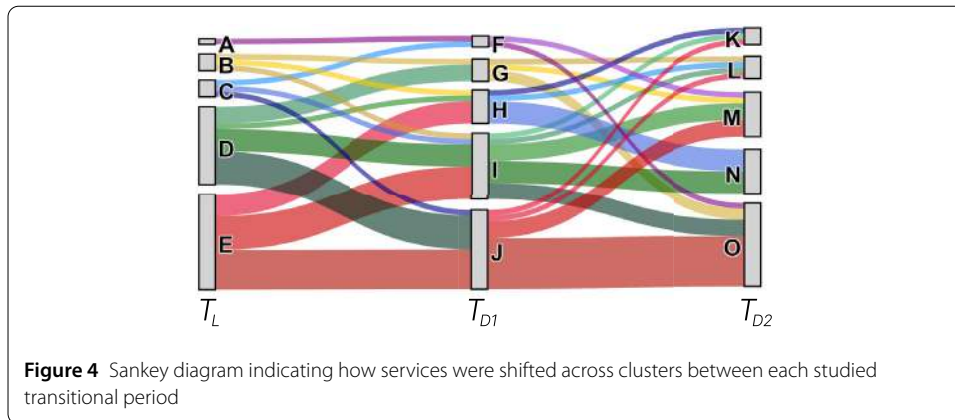
Mobile services are known to display a significant heterogeneity of spatiotemporal usages [52], and we verify that this diversity is also reflected on the impact of COVID-19 containment measures: the Pearson's correlation of the changes in the demands for total cellular capacity and for each service is as low as 0.08 on average across space and



transitions T_L , T_{D1} and T_{D2} (see Figure S10 of SI). As a consequence, the general effects of COVID-19 containment policies on the overall mobile traffic consumption that were observed above may not directly apply to the usage of individual mobile applications.

To investigate the existence and specificity of service-level effects, we repeat the same analysis described above on the consumption of the individual mobile applications; specifically, we focus on the 38 services that are responsible for the generation of the highest traffic demands in France (the traffic ranking of applications is presented on Figure S8 of SI). Interestingly, the SLM approach based on the same three features of population density, median income and leisure space presence retains a high accuracy when applied to the demands for individual services (Pearson correlation coefficients typically above 0.8 across all transitional periods and applications, see Figure S11 of SI).

We then identify groups of services whose consumption is similarly impacted by the different COVID-19 containment policies in each of the periods T_L , T_{D1} and T_{D2} . We do so by clustering applications based on their SLM coefficients. The agglomerative hierarchical clustering based on Euclidean distances among the SLM coefficients, explained in Sect. 2.9, returns 5 clusters in each period, which are highlighted in the correlation matrices in Fig. 3A1-A3. The result unveils how individual mobile applications display a relatively limited set of prototypical reactions to the imposed restrictions in major urban areas in France. Yet, when looking at the average SLM coefficients that characterize each of the identified clusters, we find that services in different clusters can exhibit highly diverse responses to COVID-19 policies, as depicted in Fig. 3B1-B3.



Finally, such heterogeneity in the response of individual applications is rendered even more complex by the combined behavior of a same service across transitions T_L , T_{D1} and T_{D2} . The Sankey diagram in Fig. 4 illustrates how applications move across the clusters of each transition: apart from some continuity in the services associated to the largest clusters, each mobile service reacts in a fairly unique way to the composition of the lockdown and curfew restrictions.

4 Discussion

Our study unveils several facets of how digital service usage changed during the COVID-19 pandemic in French urban areas, as an effect of the containment policies enacted by the local government. We find that the spatial impact on urbanized areas was not uniform, with significant differences in usage across areas focused on residential or leisure activities, as well as a relation to the income level. Moreover, we identified complex and highly diverse dynamics that the containment strategies have induced on the usage of each mobile application.

By selecting a simple linear model, we were able to effectively model changes in mobile traffic consumption related to mobility restrictions, and leverage its coefficients directly for the explainability of how changes consumption patterns related to smartphones within large urban centers in France. By utilizing mobile traffic consumption as a proxy of the spatial distribution of populations in urban areas, we are able to observe how socioeconomic status play a significant role in how mobility was affected due to COVID-19 imposed restrictions. We observe directly how denser areas with lesser income saw an increase of traffic during lockdowns, while richer and less dense areas saw a decrease of consumption. As this pattern started reversing when restrictions were lifted (specially during T_{D2}), we could relate this to how populations of lower income were confined within their neighborhoods of residence (leading to an increase of mobile traffic consumption). Previous works [53, 54] observed an increase in local mobility during the pandemic in more deprived areas, which could explain why our analysis showcases increases in traffic consumption in areas with a lower income status. Meanwhile, populations with a higher economic status either had a reduction in local mobility due to their confinement in their residences, utilizing less their mobile data plans and more their available fixed internet plans, or able to leave these urban centers for secondary residences (and therefore not moving within large urban centers). These patterns correspond to a previously observed increase in mobile traffic consump-

tion in known vacation destinations in France [21], where secondary home areas around the coast and mountains of France are located.

The polarity between the traffic difference maps of periods T_L and T_{D2} in plot C of Fig. 1 also highlights how the pattern changes introduced by restrictions of the later restrictions of the COVID-19 pandemic of urban centers in France were not long lasting, with a full reversal of their patterns around a month after the end of the third nationwide lockdown in France. This could be also validated by the reversal between the coefficients representing the Population and Leisure areas density between both periods. While the lockdown was in place and leisure areas were closed, traffic consumption (and population) presence were indeed confined more within residential areas; instead, a month after the end of restrictions and with the opening of restaurants and stores, mobile traffic began peaking again into leisure areas and decreased in residential, representing a shift back to previous patterns.

We also note that every transitional period, as seen in plots B1–B3 of Fig. 3, has one main cluster (E in T_L , I in T_{D1} , and O in T_{D2}) with SLM coefficients similar to those observed for the total traffic; in some cases, a secondary large cluster (D in T_L , J in T_{D1} , and N in T_{D2}) also exists that also show slight variations with respect to the behavior of the total traffic. These clusters include very popular applications that generate high volumes of traffic, including Twitter, WhatsApp, TikTok, or Netflix. They drive the overall mobile traffic demand by creating increased demands during the lockdown in densely populated areas of residential destination and with lower income.

The results also reveal different and specific patterns, such as that of applications strongly liked to Apple devices (*e.g.*, iMessage, iCloud, Apple Music) displaying an increased usage in higher income urban regions during the lockdown (cluster D), or a service of recent introduction at that time such as Disney+ being first adopted in richer and less populous areas of the French cities (cluster A). Similarly, it is worth highlighting how information about food providers (*e.g.*, Foursquare and Tripadvisor, in cluster C) was especially sought-after in both residential and leisure dense areas when the lockdown started, most likely due to local inhabitants looking for places that were open or accepting orders after the restrictions were enforced.

Also noteworthy are some of the effects induced by the end of the lockdown, such as younger shares of the population being allowed to stay outside but still barred from accessing leisure-dense places, so that they consumed mobile games (*e.g.*, Fornite, Clash of Clans, Pokemon Go) closer to home than usual, as shown in cluster G .

5 Conclusions and perspectives

Ultimately, the insights gathered by our analysis can help companies, including mobile network operators and mobile service providers, comprehend how their products might be affected by shifts in population mobility and allow a better policies to guarantee the quality of service requirements. Indeed, mobile application developers may employ spatiotemporal data about the impact of COVID-19 restrictions on the usage of their services to identify hidden adoption trends and discrepancies with respect to their competitors, leading a better offering on products inside their applications. Network operators can instead understand how to improve the scaling of their system for large-scale and persistent anomalous events such as epidemics.

Similarly, our results can help governments better understand the effectiveness of their actions across different indirect data sources, since the differences in smartphone usage

can indicate flows of populations across spaces of the city. They can take advantage of mobile network data, together with models like the one proposed on this study, to comprehend how mobility restrictions and extraordinary events can impact different income levels inside urban centers.

Abbreviations

BTS, Base Transceiver Station; CNIL, French National Commission on Informatics and Liberty; DPO, Data Protection Officer; EPS, Contexts and Evolved Packet System; GDPR, General Data Protection Regulation; GTP-C, GPRS Tunneling Protocol control plane; INSEE, National Institute of Statistics and Economic Studies; IRIS, Islets regrouped for statistical information; IRLS, iteratively reweighted least squares; PDP, Packet Data Protocol; SLM, Spatial lag model; ULI, User Location Information.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1140/epjds/s13688-024-00507-9>.

Additional file 1. (PDF 18.4 MB)

Acknowledgements

The work of A.F.Z. was supported by BANYAN project, which received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 860239. The work of M.F. was supported by NetSense, grant no. 2019-T1/TIC-16037 and 2023-5A/TIC-28944, funded by Comunidad de Madrid, by project PCI2022-133013 (ECOMOME), funded by MICIU/ AEI/10.13039/501100011033 and the European Union "NextGenerationEU"/PRTR, and by the research project CoCo5G (Traffic Collection, Contextual Analysis, Data-driven Optimizayion for 5G), grant no. ANR-22-CE25-0016, funded by the French National Research Agency (ANR).

Author contributions

The main idea of the paper was proposed by AFZ and MF, who also designed and performed research, analyzed data and wrote the paper. SR and ZS analyzed data and wrote the paper. All authors read and approved the final manuscript.

Funding

The work of A.F.Z. was supported by BANYAN project, which received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 860239. The work of M.F. was supported by NetSense, grant no. 2019-T1/TIC-16037 and 2023-5A/TIC-28944, funded by Comunidad de Madrid, and by the research project CoCo5G (Traffic Collection, Contextual Analysis, Data-driven Optimization for 5G), grant no. ANR-22-CE25-0016, funded by the French National Research Agency (ANR).

Data availability

The network traffic data utilized on this work is proprietary and cannot be made publicly available. Data from French Census at IRIS level is provided by INSEE and publicly available on <https://www.insee.fr/fr/information/4467366>. Data about establishments in France is provided by INSEE on the SIRENE database and is publicly available on <https://www.data.gouv.fr/fr/datasets/base-sirene-des-entreprises-et-de-leurs-etablissements-siren-siret/>.

Declarations

Ethics approval and consent to participate

Our work builds on mobile network traffic generated by users of a nationwide cellular infrastructure. The traffic measurements used to derive the data set were collected by the operator for network management and research purposes, and temporarily stored within a secure platform at their own premises. The aggregation at the level of the radio access antennas was also 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) of the European Commission. The data collection and processing was approved by the Data Protection Officer (DPO) of the operator, and authorized by the French National Commission on Informatics and Liberty (CNIL), within the context of a collaborative research project.

We remark that the original network measurements contained personal identifiers (e.g., the International Mobile Subscriber Identifier, or IMSI) and sensitive data (e.g., locations of visited antennas, or mobile services consumed) about individual users, and were deleted upon aggregation. Instead, the aggregated data consist of time series of total traffic at the antenna level with a temporal granularity of one hour, and do not contain personal identifiers or sensitive information, such as the device type, preference in terms of application consumption, or trajectories. In addition, the level of spatiotemporal aggregation ensures that no data subject can be re-identified, and that the statistics do not configure as personal data in the GDPR acceptance.

The researchers involved in the work presented in this paper only had access to such aggregated and privacy-preserving statistics for the purpose of carrying out the study. Ultimately, our dataset and research do not involve risks for the mobile subscribers.

Competing interests

The authors declared that there is no conflict of interest.

Author details

¹IMDEA Networks Institute, Madrid, Spain. ²Universidad Carlos III de Madrid, Madrid, Spain. ³Orange Innovation, Châtillon, France.

Received: 6 February 2024 Accepted: 29 October 2024 Published online: 07 November 2024

References

1. Pullano G, Valdano E, Scarpa N, Rubrichi S, Colizza V (2020) Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the covid-19 epidemic in France under lockdown: a population-based study. *Lancet Digit Health* 2(12):638–649. [https://doi.org/10.1016/S2589-7500\(20\)30243-0](https://doi.org/10.1016/S2589-7500(20)30243-0)
2. Valdano E, Lee J, Bansal S, Rubrichi S, Colizza V (2021) Highlighting socio-economic constraints on mobility reductions during COVID-19 restrictions in France can inform effective and equitable pandemic response. *J Travel Med* 28(4). <https://doi.org/10.1093/jtm/taab045>
3. Gauvin L, Bajardi P, Pepe E, Lake B, Privitera F, Tizzoni M (2021) Socio-economic determinants of mobility responses during the first wave of COVID-19 in Italy: from provinces to neighbourhoods. *J R Soc Interface* 18(181):20210092
4. Kim J, Kwan M-P (2021) The impact of the covid-19 pandemic on people's mobility: a longitudinal study of the U.S. from march to september of 2020. *J Transp Geogr* 93:103039. <https://doi.org/10.1016/j.jtrangeo.2021.103039>
5. Glodeanu A, Gullón P, Bilal U (2021) Social inequalities in mobility during and following the covid-19 associated lockdown of the Madrid metropolitan area in Spain. *Health Place* 70:102580. <https://doi.org/10.1016/j.healthplace.2021.102580>
6. Luca M, Lepri B, Frias-Martinez E, Lutu A (2022) Lutu, Andra: Modeling international mobility using roaming cell phone traces during covid-19 pandemic. *EPJ Data Sci* 11(1):22. <https://doi.org/10.1140/epjds/s13688-022-00335-9>
7. Sommerlad A, Marston L, Huntley J, Livingston G, Lewis G, Steptoe A, Fancourt D (2022) Social relationships and depression during the covid-19 lockdown: longitudinal analysis of the covid-19 social study. *Psychol Med* 52(15):3381–3390. <https://doi.org/10.1017/S0033291721000039>
8. Long E, Patterson S, Maxwell K, Blake C, Pérez RB, Lewis R, McCann M, Riddell J, Skivington K, Wilson-Lowe R, Mitchell KR (2022) Covid-19 pandemic and its impact on social relationships and health. *J Epidemiol Community Health* 76(2):128–132. <https://doi.org/10.1136/jech-2021-216690>
9. Philpot LM, Ramar P, Roellinger DL, Barry BA, Sharma P, Ebbert JO (2021) Changes in social relationships during an initial "stay-at-home" phase of the covid-19 pandemic: a longitudinal survey study in the U.S. *Soc Sci Med* 274:113779. <https://doi.org/10.1016/j.socscimed.2021.113779>
10. Spurk D, Straub C (2020) Flexible employment relationships and careers in times of the COVID-19 pandemic. Elsevier, Amsterdam
11. Juchnowicz M, Kinowska H (2021) Employee well-being and digital work during the covid-19 pandemic. *Information* 12(8):293
12. Bulińska-Stangrecka H, Bagieńska A (2021) The role of employee relations in shaping job satisfaction as an element promoting positive mental health at work in the era of covid-19. *Int J Environ Res Public Health* 18(4):1903
13. He M, Xian Y, Lv X, He J, Ren Y (2021) Changes in body weight, physical activity, and lifestyle during the semi-lockdown period after the outbreak of covid-19 in China: an online survey. *Disaster Med Public Health Prep* 15:23–28
14. Sidor A, Rzymiski P (2020) Dietary choices and habits during covid-19 lockdown: experience from Poland. *Nutrients* 12(6):1657
15. Raman G, Peng JC-H (2021) Electricity consumption of singaporean households reveals proactive community response to covid-19 progression. *Proc Natl Acad Sci* 118(34):2026596118. <https://doi.org/10.1073/pnas.2026596118>
16. Mustapa SI, Rasiah R, Jaaffar AH, Abu Bakar A, Kaman ZK (2021) Implications of covid-19 pandemic for energy-use and energy saving household electrical appliances consumption behaviour in Malaysia. *Energy Strategy Rev* 38:100765. <https://doi.org/10.1016/j.esr.2021.100765>
17. Surahman U, Hartono D, Setyowati E, Jurizat A (2022) Investigation on household energy consumption of urban residential buildings in major cities of Indonesia during covid-19 pandemic. *Energy Build* 261:111956. <https://doi.org/10.1016/j.enbuild.2022.111956>
18. Lutu A, Perino D, Bagnulo M, Frias-Martinez E, Khangosstar J (2020) A characterization of the covid-19 pandemic impact on a mobile network operator traffic. In: *Proceedings of the ACM Internet measurement conference. IMC '20*. Association for Computing Machinery, New York, pp 19–33. <https://doi.org/10.1145/3419394.3423655>
19. Liu S, Schmitt P, Bronzino F, Feamster N (2021) Characterizing service provider response to the covid-19 pandemic in the United States. In: Hohlfeld O, Lutu A, Levin D (eds) *Passive and active measurement*. Springer, Cham, pp 20–38
20. Feldmann A, Gasser O, Lichtblau F, Pujol E, Poese I, Dietzel C, Wagner D, Wichtlhuber M, Tapiador J, Vallina-Rodriguez N, Hohlfeld O, Smaragdakis G (2021) A year in lockdown: how the waves of covid-19 impact Internet traffic. *Commun ACM* 64(7):101–108. <https://doi.org/10.1145/3465212>
21. Zanella AF, Martínez-Durive OE, Mishra S, Smoreda Z, Fiore M (2022) Impact of later-stages covid-19 response measures on spatiotemporal mobile service usage. In: *IEEE INFOCOM 2022 - IEEE conference on computer communications*, pp 970–979. <https://doi.org/10.1109/INFOCOM48880.2022.9796888>
22. Böttger T, Ibrahim G, Vallis B (2020) How the Internet reacted to covid-19: a perspective from Facebook's edge network. In: *Proceedings of the ACM Internet measurement conference. IMC '20*. Association for Computing Machinery, New York, pp 34–41. <https://doi.org/10.1145/3419394.3423621>
23. Candela M, Luconi V, Vecchio A (2020) Impact of the covid-19 pandemic on the Internet latency: a large-scale study. *Comput Netw* 182:107495. <https://doi.org/10.1016/j.comnet.2020.107495>
24. Ukani A, Mirian A, Snoeren AC (2021) Locked-in during lock-down: undergraduate life on the Internet in a pandemic. In: *Proceedings of the 21st ACM Internet measurement conference. IMC '21*. Association for Computing Machinery, New York, pp 480–486. <https://doi.org/10.1145/3487552.3487828>
25. Favale T, Soro F, Trevisan M, Drago I, Mellia M (2020) Campus traffic and e-learning during covid-19 pandemic. *Comput Netw* 176:107290. <https://doi.org/10.1016/j.comnet.2020.107290>
26. Jing E, Ahn Y-Y (2021) Characterizing partisan political narrative frameworks about covid-19 on Twitter. *EPJ Data Sci* 10(1):53

27. Cui H, Kertész J (2021) Attention dynamics on the Chinese social media sina Weibo during the covid-19 pandemic. *EPJ Data Sci* 10(1):8
28. Perrotta D, Grow A, Rampazzo F, Cimentada J, Del Fava E, Gil-Clavel S, Zagheni E (2021) Behaviours and attitudes in response to the covid-19 pandemic: insights from a cross-national Facebook survey. *EPJ Data Sci* 10(1):17
29. Ucar I, Gramaglia M, Fiore M, Smoreda Z, Moro E (2021) News or social media? Socio-economic divide of mobile service consumption. *J R Soc Interface* 18(185):20210350
30. Tu Z, Cao H, Lagerspetz E, Fan Y, Flores H, Tarkoma S, Nurmi P, Li Y (2021) Demographics of mobile app usage: long-term analysis of mobile app usage. *CCF Trans Pervasive Comp Interact* 3:235–252
31. Banchs A, De Veciana G, Sciancalepore V, Costa-Perez X (2020) Resource allocation for network slicing in mobile networks. *IEEE Access* 8:214696–214706
32. Seufert A, Poignée F, Hoßfeld T, Seufert M (2022) Pandemic in the digital age: Analyzing whatsapp communication behavior before, during, and after the covid-19 lockdown. *Humanit Soc Sci Commun* 9(1)
33. Böttger T, Ibrahim G, Vallis B (2020) How the Internet reacted to covid-19: a perspective from Facebook's edge network. In: *Proceedings of the ACM Internet measurement conference*, pp 34–41
34. Oliver N, Lepri B, Sterly H, Lambiotte R, Deletaille S, De Nadai M, Letouzé E, Salah AA, Benjamins R, Cattuto C, et al (2020) Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *American Association for the Advancement of Science*
35. Szocska M, Pollner P, Schiszler I, Joo T, Palicz T, McKee M, Asztalos A, Bencze L, Kapronczay M, Petrecz P, et al (2021) Countrywide population movement monitoring using mobile devices generated (big) data during the covid-19 crisis. *Sci Rep* 11(1):5943
36. Perra N (2021) Non-pharmaceutical interventions during the covid-19 pandemic: a review. *Phys Rep* 913:1–52
37. Budd J, Miller BS, Manning EM, Lamos V, Zhuang M, Edelstein M, Rees G, Emery VC, Stevens MM, Keegan N, et al (2020) Digital technologies in the public-health response to covid-19. *Nat Med* 26(8):1183–1192
38. Fahey RA, Hino A (2020) Covid-19, digital privacy, and the social limits on data-focused public health responses. *Int J Inf Manag* 55:102181
39. Gouvernement of France (2023) Les actions du gouvernement. <https://www.gouvernement.fr/info-coronavirus/les-actions-du-gouvernement>
40. Bembaron E (2021) La 5G représente “moins de 1% du trafic”, affirme Bouygues Telecom. <https://www.lefigaro.fr/secteur/high-tech/la-5g-represente-moins-de-1-du-traffic-affirme-bouygues-telecom-20210520>
41. Paul U, Subramanian AP, Buddhikot MM, Das SR (2011) Understanding traffic dynamics in cellular data networks. In: *2011 proceedings IEEE INFOCOM*. IEEE Press, New York, pp 882–890
42. INSEE (2016) Definition - IRIS | Insee — insee.fr. <https://www.insee.fr/en/metadonnees/definition/c1523>. [Accessed 17-Nov-2022]
43. Insee (2020) Recensement 2017: résultats sur un territoire, bases de données et fichiers détail. <https://www.insee.fr/fr/information/4467366>
44. Insee (2021) Revenus, pauvreté et niveau de vie en 2018 (Iris). <https://www.insee.fr/fr/statistiques/5055909>. Accessed: 2022-10-2022-05
45. Base Sirene des entreprises et de leurs établissements (SIREN, SIRET). <https://www.data.gouv.fr/fr/datasets/base-sirene-des-entreprises-et-de-leurs-etablissements-siren-siret/>. Accessed: 2022-10-05
46. Anselin L, Smirnov O (1996) Efficient algorithms for constructing proper higher order spatial lag operators. *J Reg Sci* 36(1):67–89
47. Anselin L, Bera AK (1998) Spatial dependence in linear regression models with an introduction to spatial econometrics. *Stat Textb Monogr* 155:237–290
48. Huber PJ (2004) *Robust statistics*, vol 523. Wiley, New York
49. Holland PW, Welsch RE (1977) Robust regression using iteratively reweighted least-squares. *Commun Stat, Theory Methods* 6(9):813–827
50. Murtagh F, Contreras P (2012) Algorithms for hierarchical clustering: an overview. *Wiley Interdiscip Rev Data Min Knowl Discov* 2(1):86–97
51. Ward JH Jr (1963) Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 58(301):236–244
52. Marquez C, Gramaglia M, Fiore M, Banchs A, Ziemlicki C, Smoreda Z (2017) Not all apps are created equal: analysis of spatiotemporal heterogeneity in nationwide mobile service usage. In: *CoNEXT*, vol '17. Association for Computing Machinery, New York, pp 180–186. <https://doi.org/10.1145/3143361.3143369>
53. Do Lee W, Qian M, Schwanen T (2021) The association between socioeconomic status and mobility reductions in the early stage of England's covid-19 epidemic. *Health Place* 69:102563
54. Glodeanu A, Gullón P, Bilal U (2021) Social inequalities in mobility during and following the covid-19 associated lockdown of the Madrid metropolitan area in Spain. *Health Place* 70:102580

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.