Second-level Digital Divide: A Longitudinal Study of Mobile Traffic Consumption Imbalance in France

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

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

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

ABSTRACT

We study the interaction between the consumption of digital services via mobile devices and urbanization levels, using measurement data collected in an operational network serving the whole territory of France. We unveil that such an interaction follows a power law, or, in other words, there exists an emergent behavior that prompts subscribers living in increasingly extended and populated urban areas to exhibit a surging individual consumption of mobile traffic. The result holds for the global traffic, but is also consistently observed across a range of mobile services, although with varying intensity. An unprecedented longitudinal analysis of the phenomenon unveils how the imbalance in the per-capita mobile data traffic usage across cities of different size has grown steadily and substantially in the 2014–2019 time frame in France. Our study raises questions on the presence of second-level digital divides in developed countries, and paves the road to further investigations.

CCS CONCEPTS

• **Applied computing** → *Sociology*; • **Social and professional topics** → *Geographic characteristics*; • **Networks** → *Network measurement*.

KEYWORDS

Mobile network traffic, digital divide, longitudinal analysis

ACM Reference Format:

Sachit Mishra, Zbigniew Smoreda, and Marco Fiore. 2022. Second-level Digital Divide: A Longitudinal Study of Mobile Traffic Consumption Imbalance in France. In *Proceedings of the ACM Web Conference 2022 (WWW '22), April 25–29, 2022, Virtual Event, Lyon, France.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3485447.3512125

1 INTRODUCTION

In countries worldwide, significant efforts in public policing and investments are made to reduce inequalities in the accessibility of digital services. Wireless communication technologies often play an important role in bringing broadband connectivity to smaller and less urbanized communities, as they can provide cost-effective coverage to wide and less populous regions. The ubiquitous availability of high-speed wireless Internet access granted by 4G and 5G mobile networks is therefore seen as a critical enabler to help closing the *digital divide* between citizens who can benefit from global ICT resources and those who cannot [4, 24].

In developed countries, 4G connectivity is nowadays available to a vast majority of the population [21]. However, a more homogeneous accessibility does not necessarily imply a more equal use. Even in presence of pervasive broadband connectivity, imbalances in the adoption of digital services across different population groups may emerge due to factors that transcend technology and root into living environment, education, or digital skills. This phenomenon is referred to as *second-level* digital divide, and is typically related to a disparity in the quantity and quality of consumption of mobile services, rather than in their sheer availability [10, 28].

In this paper, we investigate the presence of a second-level digital divide at a national scale in a prominent European country, *i.e.*, France. Specifically, we explore correlations between the number of inhabitants and the volume of consumed mobile data traffic in thousands of individual cities and towns in the country. Our analysis is based on substantial measurement data collected in the production network of a major operator between 2014 and 2019, and leads to the following key observations.

(*i*) The relationship between mobile traffic usage and number of inhabitants of an urban settlement is well described ($R > 0.80$) by a power law with exponent higher than 1. In other words, there exists an *emergent behavior* according to which the larger is the city a mobile subscriber lives in, the higher the mean volume of mobile data he or she consumes. The sizeable imbalance in the per-capita mobile service usage, which favors users in areas with a higher urbanization level, is a symptom of that a second-level digital divide may affect the French territory. This is rather unexpected in a country that vaunts the seventh economy worldwide, and has embraced governmental strategies to close traditional ICT accessibility gaps in recent years [2].

(*ii*) The behavior not only affects the global volume of mobile traffic, but it consistently holds across a wide range of individual mobile services, such as YouTube, Netflix, Twitter, or Spotify. However, the *intensity* of the observed imbalance varies depending on the specific application.

(*iii*) The aforementioned phenomenon has *amplified* over the 2014–2019 time frame, as inhabitants of larger cities have increased their consumption of mobile data traffic at a faster rate than inhabitants of smaller towns. In other words, the presumed second-level digital divide is not being reduced, but it has instead been growing in recent years.

Conference'17, July 2017, Washington, DC, USA

(*iv*) The growth of inequality in mobile service consumption over time is not explained by the geographical coverage of the network infrastructure, age or income of the local populations, or commuting of workers to larger cities. The robustness of the behavior to such potential confounding factors reinforces the hypothesis that the living environment (*i.e.*, a large conurbation opposed to a small village, with the sociological implications that the difference involves) is the root cause of the observed divide.

2 DATASETS

Our study hinges on open-source geospatial and demographics data, and on extensive measurements by a major mobile network operator in France, which we present next.

2.1 City shape data

Our interest is in examining how the mobile traffic consumption relates to the population size of cities and towns with diverse urbanization levels. As there is no unanimously recognized notion of city [7], we run all of our tests considering two different definitions that are employed for statistical analyses. Specifically, we study mobile traffic usage aggregated at the level of *communes*, *i.e.*, French local administrative units analogous to civil township in the United States [16]. In parallel, we look at how the same mobile traffic is utilized across *urban units*, *i.e.*, continuous built-up areas [17]. Shapefiles of the geographical surface of each of more than 34, 900 communes and 2, 400 urban units are obtained from public repositories [14, 15]. We note that using the two diverse definitions above ensures that behaviors observed in both cases are not artifacts of the specific way cities are designated.

2.2 Auxiliary demographic data

Information on the resident population of each commune and urban unit of France is obtained from national census databases maintained by the French National Institute of Statistics and Economic Studies (INSEE) [19]. Since censuses are conducted on a quinquennial basis in France, we estimated the number of inhabitants in the 2014–2019 time frame by linearly interpolating the population data available for 2012 and 2017, which is a reasonable approximation considering the very slow timescales of changes in resident populations.

In order to study confounding factors, we also employ demographic statistics compiled on a yearly basis by INSEE on the distributions of age [19] and income [20] of the population in each commune. From those, we derive for every commune (*i*) the fraction of population which is deemed more prone to consume mobile services according to sociological reviews [3], and (*ii*) the median income per dwelling unit. We also leverage surveys carried out by INSEE on the inter-commune home-work mobility of residents in France [18] to estimate the amount of people actually present in each commune during the working hours.

2.3 Mobile network traffic and coverage

We employ real-world mobile data traffic information collected in a nationwide production network servicing the whole territory of metropolitan France. The data was gathered by the network operator using passive measurement probes tapping at the Gi/SGi/Gn interfaces that connect the Gateway GPRS Support Node (GGSN) and the Packet Data Network Gateway (PGW) to external Public Data Networks (PDN). Proprietary and commercial classifiers were used to identify the individual mobile services associated to each traffic flow. This setup allows monitoring traffic volumes generated by 2G, 3G and 4G devices of the whole user base of the network operator. As a result, the data captures the total and per-service traffic consumption of around 30% of the mobile subscribers in France; we remark that the market share of the operator is fairly steady over the whole country, providing a solid statistical basis for a per-city study.

The data collection was performed in four years, *i.e.*, 2014, 2016, 2017, and 2019, during multiple weeks in the same period of the year. The yearly period was carefully selected not to include holidays, so as to avoid that user behavior changes during the vacation days may bias the analysis.

Traffic demands are geo-referenced by extracting the User Location Information (ULI) from the Packet Data Protocol (PDP) Contexts and Evolved Packet System (EPS) Bearers that transit over the GPRS Tunneling Protocol control plane (GTP-C). This allows associating flows to the Base Stations (BS) serving them[1]. Based on this information, the network operator built aggregates of the total (uplink plus downlink) traffic recorded at every BS in France, on a daily basis.

In addition, we employ open data on the evolution of the LTE and LTE-A coverage offered by the same operator over the national territory during the years above, which is provided by the French Authority for the Regulation of Electronics Communications and Postal Services (ARCEP) [1].

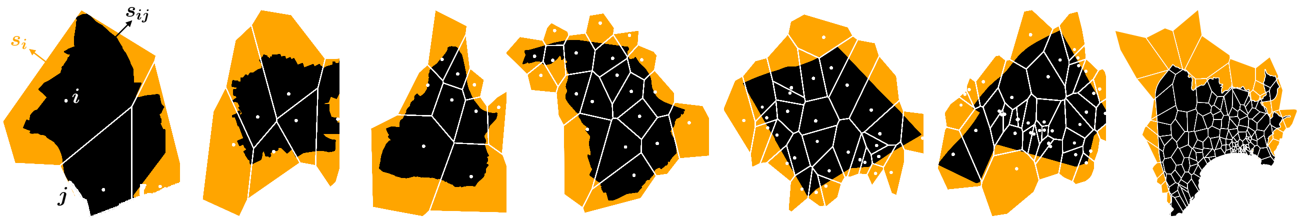
2.4 Ethics considerations

Ultimately, the mobile network dataset we use in this study consists of the daily volume of data traffic served by individual BS. The level of spatiotemporal aggregation ensures that no data subject can be re-identified, and that the data does not configure as personal data in the acceptation of the General Data Protection Regulation (GDPR) [26]. Therefore, the dataset and research do not involve risks for the mobile subscribers, while they provide new knowledge about the potential existence of a second-level digital divide in France, which may benefit a more informed social policing.

The raw traffic measurements used to derive the dataset above were stored and aggregated in a secure platform at the operator premises, in full compliance to article 89 of the GDPR, under the supervision of the Data Protection Officer (DPO) of the operator, and upon authorization by the French National Commission on Informatics and Liberty (CNIL). The raw measurements were deleted upon data aggregation.

No ethical issues are instead associated with the datasets on city shapes, demographics or LTE mobile coverage.

[1] It is worth noting that irregular ULI updates can limit the spatial accuracy of the data [30]. However, in Section 3.2 we map BS traffic at the level of communes and urban units, each of which covers tens of km^2 and encompassing multiple Routing Areas (RA) and Tracking Areas (TA). As changing RA/TA enforces a ULI refresh, the precision of the traffic localization is appropriate for a study at the level of communes or urban units.



(a) Portiragnes (b) Denain (c) Chamiers (d) Rezé (e) Levallois-Perret (f) Neuilly (g) Nice

Figure 1: Examples of weighted spatial interpolation in seven random French cities of growing size from (a) to (f). Voronoi cells (outlined by white contours) computed from the locations of each BS i (white dots) cover surfaces s_i (orange areas), and yield intersections s_{ij} with the territory of commune j (black areas). The total mobile network traffic associated to commune j is $\sum_i (s_{ij}/s_i)$. The cities are not in the same scale. Figure best viewed in color.

3 DATA PROCESSING

We process the datasets above so as to align them, while also removing artifacts of the data collection, as follows.

3.1 Filtering

First, we restrict the set of French communes that are the focus of our study. Indeed, metropolitan France is divided in over 34,900 communes, the vast majority of which map to rural areas with sparse settlements [7]. As we are interested in municipalities with non-negligible human presence, we retain communes with a minimum population count of 2,000 inhabitants during the target years. Indeed, that is the population threshold employed to define urban units [8], which also allows aligning the two different notions of city we work with. Ultimately, we consider 4,979 communes and 2,354 urban units, which cover all cities and towns in France, and encompass 70% and 75% of the total population in the country, respectively. Note that the resulting number of communes is around twice that of urban units: the reason is that urban units can bring together multiple communes, in case the latter form a continuous built-up area [17].

3.2 Alignment

Our research requires mapping the average daily demand recorded at each BS in France into the volume of mobile traffic consumed in every commune and urban unit.

We employ the geographical locations of the BS to draw a Voronoi tessellation of the French territory, and map the mobile traffic recorded in Voronoi cells into that generated within each target commune and urban unit via a weighted spatial interpolation. This is done by assuming that the average daily traffic measured at each base station is uniformly distributed within its Voronoi cell, which is a common approximation in absence of finer per-BS coverage information [23]. Then, each BS contributes to a given commune or urban unit a fraction of its traffic proportional to the intersection between the corresponding Voronoi cell and the shape of the commune or urban unit. We note that the Voronoi tessellation is computed by delimiting it along water surfaces or foreign territories, so that no traffic is incorrectly assigned to areas with no user presence. Illustrative examples of the interpolation process are provided in Figure 1.

3.3 Artifact removal

The assumptions on the Voronoi shape of BS coverage and on the uniform distribution of traffic over that shape may create artifacts in the traffic data at the level of communes and urban units. In order to quantify and address problems in the data alignment, we employ a Random sample consensus (RANSAC) technique [12]. RANSAC is an iterative method that can be used to automatically detect outliers in data whose distribution is assumed to be explained by a specific model [25]. In our case, we run RANSAC on data describing, for each commune or urban unit, the average traffic and number of inhabitants: this allows identifying situations where the relationship between the typical city-level mobile traffic usage and the local population is anomalous.

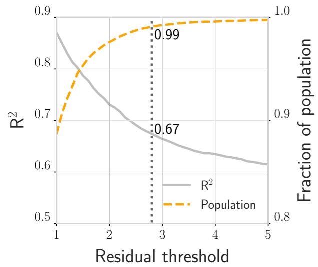
The sensitivity of RANSAC to outliers depends on a *residual threshold* parameter that controls the trade-off between the quality of the model and the amount of discarded data. We perform an exhaustive search for the best residual threshold for each year separately, as illustrated in Figure 2.

According to RANSAC, a total of 142 communes and 58 urban units are outliers in at least one of the years. A close inspection reveals that they are all very small towns covered by a reduced number of Voronoi cells with peculiar BS arrangements, as exemplified in Figure 3. The issue is due to the inherent limits of the Voronoi representation for the spatial mapping of mobile traffic, which are especially evident in minor towns with minute geographical surfaces. Instead, the impact of the Voronoi approximation is mitigated in the vast majority of the communes and urban units, which feature a larger size or more consistent BS deployment, as previously observed in Figure 1, and are considered inliers by RANSAC.

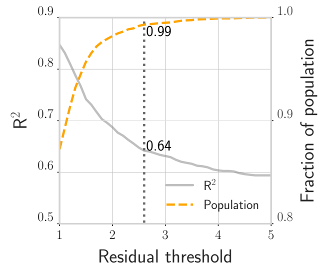
In order to avoid that anomalies induced by the spatial representation of the data bias our analyses, we discard the concerned communes for the rest of the study. By doing so, we retain 4,837 communes and 2,296 urban units with over 2,000 inhabitants, which cover 69% and 74% of the overall French population, respectively.

3.4 Binning

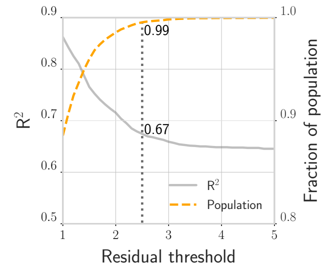
The data at both the commune and urban unit levels is affected by very high heteroskedasticity: as illustrated in Figure 4 and like in many other countries, there are hundreds of smaller towns in France, and a reduced number of major conurbations. For the purpose of our study, this creates substantial skewness in the number



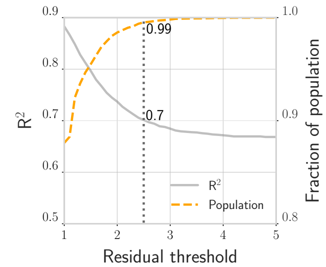
(a) Communes, 2014



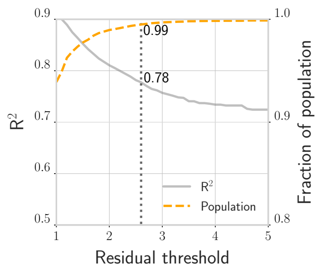
(b) Communes, 2016



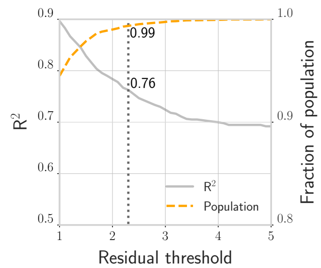
(c) Communes, 2017



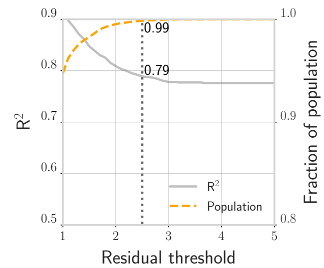
(d) Communes, 2019



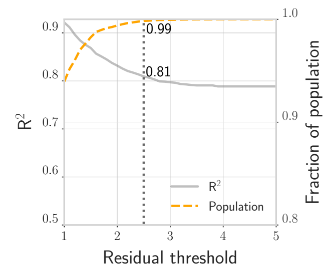
(e) Urban units, 2014



(f) Urban units, 2016

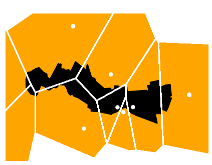


(g) Urban units, 2017

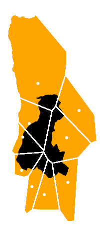


(h) Urban units, 2019

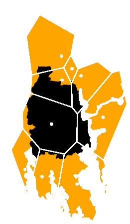
Figure 2: Calibration of RANSAC. For each year, we vary the residual threshold parameter to highlight the trade-off between model accuracy (measured as the coefficient of determination, R^2) and data loss (measured by the fraction of total population in the retained urban units). In all plots, removing outlying urban units that account for a minimal portion of the population improves the R^2. Further gains in the model quality come at the cost of substantial population loss, indicating that we are artificially improving the model quality by removing inlier urban units. The accepted values of the threshold (dotted vertical lines) preserve 99% of the population, with 0.64–0.81 R^2. Results are reported for (a)-(d) communes, and (e)-(h) urban units.



(a) Montsoult

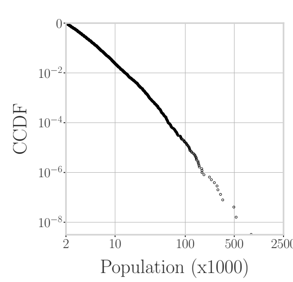


(b) Wimille

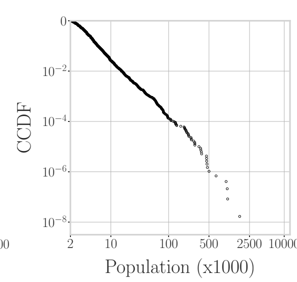


(c) Crac'h

Figure 3: Examples of outlying communes identified by RANSAC. The reduced territory of the municipalities and the BS locations cause especially small traffic fractions to be assigned to the communes. The cities are not in the same scale. Figure best viewed in color.



(a) Communes



(b) Urban units

Figure 4: Heteroscedasticity of the data, demonstrated by the Complementary Cumulative Distribution Function (CCDF) of the population size of all studied (a) communes and (b) urban units: 95% of samples are minor towns below 10,000 inhabitants.

of samples for settlements of different size, which becomes an issue in terms of model fitting. Indeed, minimizing the error on all points under such heteroskedastic data pushes a model to capture well the behavior of data-dense low-population towns, whereas the fitting quality on the few larger cities is neglected.

To explore the severity of the problem, we assess the quality of a global model fitted on the full data, *i.e.*, the mobile traffic samples recorded in all cities and towns during each observed day. The procedure we employ is illustrated in Figure 5. For a specific window τ, a local model is fitted on the data in the window only: such a model then represents the best model possible for the windowed data. We compute R^2_{max} as the coefficient of determination of the local model on the windowed data: this is a reference for the best quality that the model can attain on that subset of cities. Then, we compute the coefficient of determination R^2 of the global model (fitted on the whole data samples, whose accuracy we want to test) on the windowed data only. Finally, we associate to the central population of the window τ the ratio R^2/R^2_{max}, which measures the distance between the global and ideal models, for that specific window: the closer is the ratio to one, the better the global model represents the cities in the window. In the example in Figure 5, the ratio associated to the window τ is 0.75/0.80=0.94. By sliding the

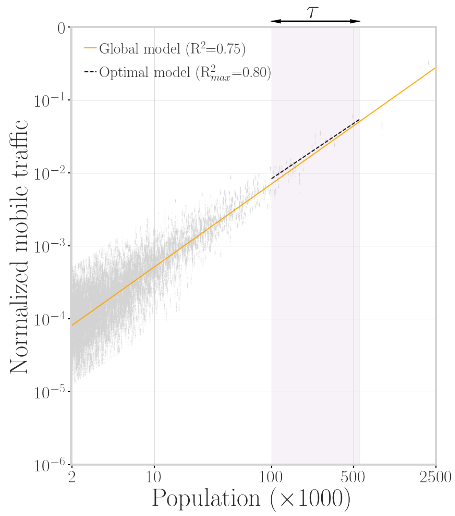
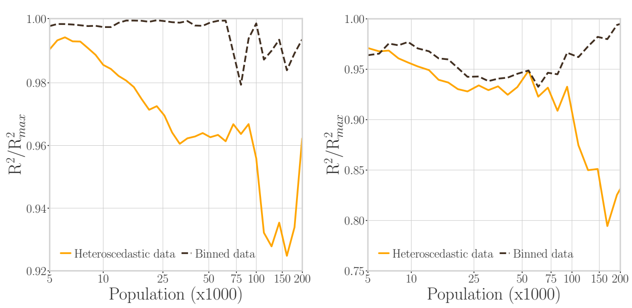


Figure 5: Illustration of the sliding window procedure adopted to assess the quality of model fitting versus the city population. The plot shows the daily traffic samples in all cities (communes in 2019, in this case, gray dots), the global model fitted on the whole data (solid orange line), and the local model fitted on data included in the window only (dashed black line). Figure best viewed in color.



(a) Communes, 2019 **(b) Urban units, 2014**

Figure 6: Fitting quality versus size of the settlement, obtained by a model fit on the full heteroscedastic data (solid orange), and on binned data (dashed black). Representative cases are shown: communes in 2019, urban units in 2014.

window τ across all city populations (*i.e.*, along the full abscissa), we obtain the orange curves in Figure 6. The heterogeneous quality of the global model is evident: it renders much better the traffic in smaller towns, whereas mobile usage in communes or urban units above 100,000 inhabitants is not captured well. Thus, conclusions drawn under such a model risk to be less reliable for larger cities.

To avoid data heteroscedasticity, we bin communes and urban units according to their population size; we employ 60 bins on a logarithmic scale to account for the substantial variety of settlement sizes that spans several orders of magnitude. For each bin, we compute the average daily mobile data traffic across all encompassed cities, for all observed days. The resulting single point per bin ensures a much more balanced distribution of samples: a global model fitted on the binned data yields improved quality across the full spectrum of cities, as shown by the black curves in Figure 6.

4 RESULTS

Building on the processed data above, we carry out an analysis of the traffic demand generated by mobile network subscribers in the cities and towns of France. Next, we discuss our findings on consumption imbalance, its evolution in time, and its relationship with potential confounding factors.

4.1 Imbalance in traffic consumption

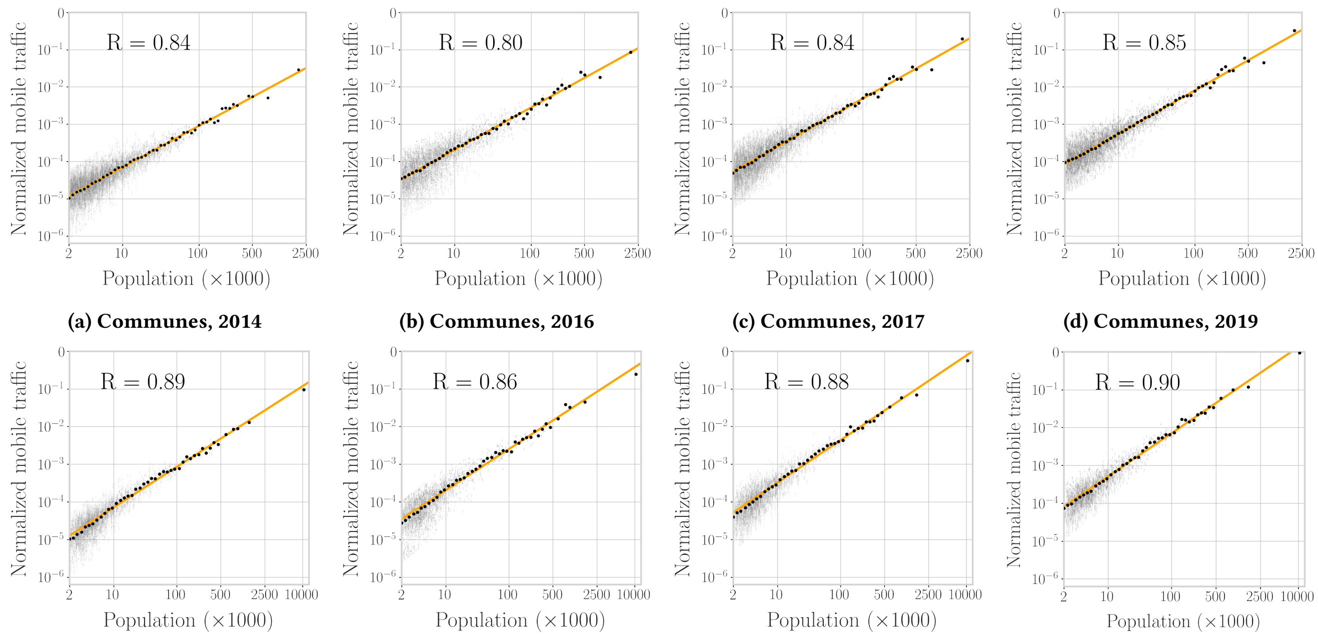
We are primarily interested in understanding how the per-capita traffic volume varies as a function of the city size. This is equivalent to identifying the scaling law of total traffic with respect to the resident population. Figure 7 shows that power laws fit well the relationship of traffic and population in all considered years. The Pearson correlation coefficient (R) of the model with respect to the per-city and per-day mobile traffic (gray dots) is consistently good, and ranges from 0.80 to 0.90 depending on the considered city definition and target year.

Formally, the result indicates that $t = k \cdot p^{\alpha}$, where t is the daily mobile network traffic demand measured in a commune (*e.g.*, in bytes, although all results are normalized by the maximum daily load observed in the data, so as not to disclose the actual volume of traffic of the operator), and p is the local population (in inhabitants). The multiplier k and the exponent α are the fitted model parameters.

Interestingly, $\alpha > 1$ in all years, which implies an emergent behavior, or, equivalently, positive returns to scale: the larger the city, the higher the amount of network traffic consumed individually by its inhabitants. The result highlights the existence of an imbalance in mobile service usage across the French territory. Citizens in smaller towns tend to generate less traffic per capita than inhabitants of larger cities, and such a disparity is magnified when considering people living in large metropolitan areas. For instance, by looking at the 2014 data, the average subscriber in a 10,000-inhabitant commune consumed 77% of the traffic of a user living in a commune of 100,000 inhabitants, and 60% of the traffic of subscriber in a metropolis with 1,000,000 residents. *The result unveils a form of second-level digital divide [10, 28], as individuals in more populous cities are remarkably keener to access mobile digital services than smaller town residents.*

We also explore how the phenomenon affects different mobile services in Table 1. We observe how the power law behavior characterizes not only the total traffic, but also the usage of individual mobile applications: models fitted on the traffic of each service yield R in the 0.69 to 0.90 range. The imbalance in service usage across urbanization levels is confirmed, with a power law exponent above one in all cases. Yet, the value of α is far from uniform. Some services, such as Google Play Store, have an α closer to one, *i.e.*, are consumed more homogeneously across cities. Others, like Google Meet, show an even higher inequality than the total traffic, with α values above 1.3.

As a final consideration, we remark the strong alignment of results in Figure 7 and Table 1 for communes and urban units, which shows how the hypothesis of a second-level digital divide is robust to different definitions of city.



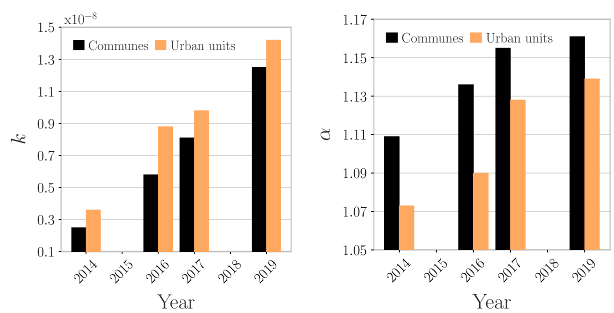
(a) Communes, 2014 (b) Communes, 2016 (c) Communes, 2017 (d) Communes, 2019

(e) Urban units, 2014 (f) Urban units, 2016 (g) Urban units, 2017 (h) Urban units, 2019

Figure 7: Scatterplots of normalized daily mobile traffic and resident population for 4, 837 communes (top) and 2, 296 urban units (bottom) of France, in four different years (columns). Both abscissa and ordinate are in logarithmic scale, resulting in a linear appearance of the fitted power law models (orange solid lines). Gray dots are the original per-city and per-day samples, and black dots are the bin averages used to remove data heteroscedasticity.

		Communes		Urban units	
Service	Total traffic	α	R	α	R
Instagram	17.99	1.258	0.83	1.253	0.88
Facebook	11.22	1.130	0.83	1.154	0.90
Facebook Live	8.20	1.105	0.83	1.14	0.90
Netflix	7.91	1.174	0.79	1.194	0.86
Youtube	7.85	1.157	0.82	1.177	0.88
SnapChat	5.84	1.141	0.82	1.192	0.89
Apple Video	5.17	1.189	0.76	1.215	0.82
Google Play Store	5.14	1.080	0.81	1.140	0.88
Apple iCloud	3.89	1.236	0.80	1.232	0.86
Deezer	3.18	1.200	0.79	1.218	0.85
Orange TV	2.26	1.097	0.69	1.163	0.77
Twitch	1.83	1.210	0.69	1.288	0.76
Twitter	1.75	1.267	0.80	1.284	0.86
DailyMotion	1.43	1.192	0.79	1.22	0.86
LinkedIn	1.30	1.404	0.76	1.410	0.82
Microsoft Store	1.10	1.195	0.72	1.239	0.79
Apple Music	0.92	1.313	0.74	1.304	0.79
WhatsApp	0.71	1.288	0.78	1.283	0.84
Pinterest	0.61	1.164	0.79	1.209	0.85
Google Drive	0.60	1.242	0.82	1.239	0.88
Microsoft Mail	0.48	1.251	0.78	1.264	0.85
Spotify	0.44	1.247	0.75	1.242	0.82
Google Maps	0.38	1.214	0.79	1.207	0.86
Icecast	0.34	1.049	0.77	1.072	0.84
Apple iTunes	0.30	1.260	0.78	1.273	0.84
Google Meet	0.26	1.337	0.79	1.309	0.84

Table 1: Exponent (α) and fitting quality (R) of power law models of the per-capita traffic, for individual mobile services across communes and urban units of France. Services are ranked based on decreasing order of their normalized traffic volumes. The intensity of the color of cells reporting the α values is proportional to the value itself, for better readability. All figures reported in the table refer to data for year 2019.



(a) Multiplier (b) Exponent

Figure 8: Evolution of the parameters k and α of the power law model of the total traffic and population across years, for communes and urban units.

4.2 Longitudinal analysis

We now explore how the observed divide has evolved in the 2014–2017 time frame. It is worth noting that this is a first-of-its-kind longitudinal study of the relationship between mobile traffic consumption and urbanization levels. The results are summarized in Figure 8, which reports the evolution of the power law parameters across years.

Unsurprisingly, the multiplier k in Figure 8a grows over time, indicating that the individual traffic demand has increased in *all* communes of France through the years. This is a well-known global trend, reported by a variety of industry reviews [6, 21] and linked to the more and more central role that mobile devices and services play in our lives. Importantly, the exact figures are fairly consistent for communes and urban units, which proves once again the reliability of the results across city definitions.

While k captures a nationwide trend, our model allows breaking down the increasing per-capita traffic behavior on a per-city basis, and a less expected tendency emerges there. As shown in Figure 8b, the exponent α of the power law has monotonically expanded between 2014 and 2019, from 1.11 to 1.16 for communes, and from 1.07 to 1.14 for urban units[2]. As α represents the unevenness among municipalities, its increase reveals that the imbalance in per-capita mobile traffic consumption has been in fact growing over the years in France. A similar example as that presented before exposes that in 2019 the average subscriber in a 10,000-inhabitant commune consumed just 69% and 48% of the traffic generated by typical users living in communes with 100,000 and 1,000,000 residents, respectively. When comparing these figures with 2014, in Section 4.1, it turns out that, in the considered five-year period, the per-capita traffic has been rising 13% and 26% faster in communes with 100,000 and 1,000,000 inhabitants than in communes of 10,000 people.

A spatial representation provides even neater evidence of the increasing divide in mobile traffic consumption over the French territory. Figure 9 illustrates the Compound Annual Growth Rate (CAGR) of the mobile data traffic per inhabitant, as measured between 2014 and 2019 across all cities we study. The two maps refer to the cases where communes or urban units are employed as the city definition. The darkest regions, characterized by a higher CAGR, clearly demarcate the major urbanized areas in the country. While mobile data traffic usage per capita has grown at 30–50% CAGR in smaller towns, major conurbations have witnessed their traffic per inhabitant increase by up to 75% on a yearly basis.

These insights are to some degree unexpected, as one could reasonably believe that in a developed country like France the adoption of mobile services in less populated regions is catching up with the standards set by metropolitan areas, and all forms of digital divide are being closed over the years. Our analysis proves that this is not the case, and an opposite tendency is in fact dominating mobile traffic usage.

4.3 Confounding factor analysis

A legitimate question is whether the disparity in per-capita mobile traffic consumption is biased by confounding factors. We investigate this possibility by considering several potential sources of disturbance, and exploring their heterogeneity across cities as well as their evolution in time. We carry out this part of the study at the level of communes, as the needed data is only available for these administrative units.

Broadband coverage. A first obvious aspect that may affect the results is the availability of broadband 4G coverage in the set of studied communes. Using the network coverage information presented in Section 2.3, available for years from 2017 to 2019, we compute for every commune its 4G coverage ratio, *i.e.*, the portion of its territory covered by 4G connectivity, separately in each year. We

[2]Note that the absolute value of α is lower in urban units than communes because the two definitions match in smaller towns, whereas larger urban units tend to aggregate many adjacent communes that form a single conurbation. This means that larger urban units ultimately capture a mixture of dense urban areas and suburbs, slightly curbing the emergent behavior observed in the case of communes that tell apart such areas.

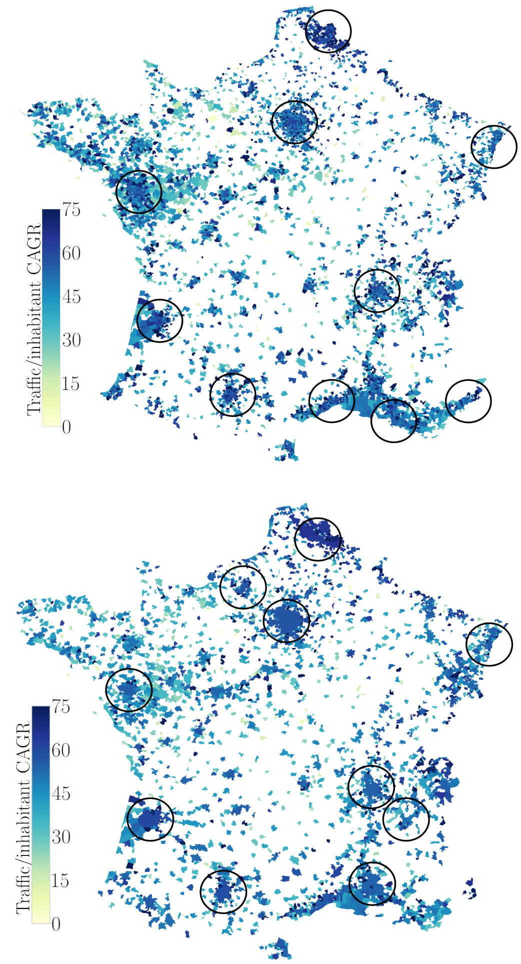


Figure 9: CAGR of mobile demand per inhabitant in 2014–2019, in 4,837 communes of France (top) and 3,242 urban units of France (bottom). The ten major cities (according to the corresponding definition) are outlined with dark circles: they are Paris, Marseille, Lyon, Toulouse, Nantes, Strasbourg, Bordeaux, Lille, Nice (communes only), Montpellier (communes only), Rouen (urban units only) and Grenoble (urban units only).

then group communes on five classes, based on their population: 3,000–5,000, 5,000–15,000, 15,000–30,000, 50,000–100,000 and over 100,000 inhabitants, according to convention. For each class, we compute the mean and standard deviation of the 4G coverage ratio of the associated communes.

The top-left plot in Figure 10 shows the result: as expected, 4G coverage has improved significantly between 2017 and 2019, reaching values close to 100% in almost all considered communes, with minimal differences among the five classes. As no 5G service was yet deployed in 2019 in France [5], the result highlights a *reduction* of the digital divide in terms of mobile broadband accessibility. This is in stark contrast with the inequality in usage, and further supports the hypothesis that a form of second-level digital divide may underpin the imbalance we unveil.

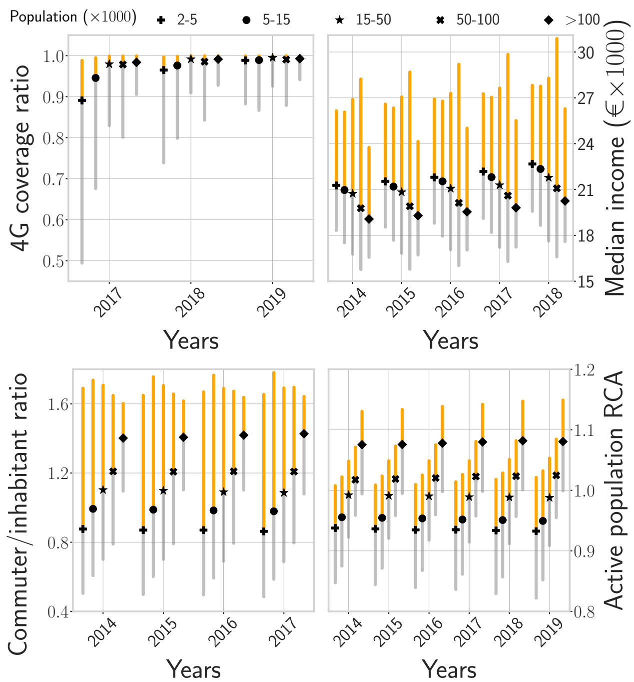


Figure 10: Mean (marker) and standard deviation (errorbars) of four potential confounding factors, for five population-based classes of communes (denoted by different markers) and in different years (in the abscissa). Years are not consistent across factors, due to the heterogeneous time span of the original data sources in Section 2.

Income. Previous studies on a digital usage divide have also pointed at income as a source of inequality in mobile service adoption [10]. By using income data presented in Section 2.2 and adopting a similar approach of clustering communes as done for 4G coverage, we compute the median value of the income across communes grouped by population, and over different years. The top-right plot in Figure 10 shows no clear relationship between economic wealth and traffic usage at a commune level. Moreover, median income distributions have only changed slightly in the 2014–2019 time frame, which cannot explain the observed growing disparity in mobile service consumption.

Commuting. Larger cities tend to attract commuters during the working hours, which inflates their actual mobile subscriber population with respect to the inhabitants recorded on census. The bottom-left plot in Figure 10 is computed using the work mobility survey data presented in Section 2.2, and highlights how the ratio of commuters per inhabitant actually grows with the number of residents in the commune. However, the effect is constant across years, hence cannot explain the increasing imbalance over years that we observe instead in per-capita mobile traffic consumption.

Active population. Mobile services are utilized in diverse ways and quantity by people of different age. In France, individuals in the 14–59 age range are the most active mobile users [3]. We thus investigate whether an uneven presence of inhabitants of that age may represent a confounding factor for our study. Specifically, we employ the population age data introduced in Section 2.2 to compute the Revealed Comparative Advantage (RCA) of residents who are 14 to 59 years old in each commune; the RCA measures the higher or lower incidence of people within that age range with respect to the average over all communes. The bottom-right plot in Figure 10 shows that a difference exists across cities, and larger municipalities indeed present a higher fraction of more active mobile network users. Yet again, this diversity does not vary across years, hence cannot be the cause of the growing divide we observe in traffic consumption.

5 RELATED WORK

Digital divides beyond accessibility have been studied in sociology, where qualitative analysis show that individuals with different socioeconomic status have diverse patterns of ICT use for economic, community, and political engagement. For instance, high-education people visit digital resources that increase their social capabilities [13], and low-skill individuals are more likely to use the communication network for entertainment purposes [27]. Our work takes a novel quantitative approach to the analysis of second-level digital divides, and reveals that an imbalance exists in the sheer volume of consumed traffic across cities of a developed country.

Also related to our study are previous researches showing that the interaction between the population residing in a specific geographical area and the volume of mobile traffic generated therein is governed by power laws. This property was proven to hold under several measures of mobile demands, including the number of voice calls or text messages [11], bytes of data traffic [29], or user activity [22]. Also, the property is robust across spatial scales, from neighborhoods in a city [11] to large regions in a country [9]. However, no previous work has tested the power law hypothesis at the level of thousands of municipalities in a whole country as we do. Our study not only confirms the validity of the law at such granularity and scale, but also introduces a first-of-its-kind longitudinal analysis of the phenomenon.

6 CONCLUSIONS

We empirically demonstrated a non-negligible and growing imbalance in the per-capita mobile traffic consumption across municipalities in France. The finding raises interesting questions on the potential presence and trends of covert forms of second-level digital divides in developed countries. Mobile broadband availability or socio-economic indicators do not explain the divide, which lets us speculate that it may reflect a stronger natural inclination (which has been reinforced over recent years) of inhabitants of larger cities to rely on and benefit from mobile services. However, we believe that more studies on the phenomenon are needed, and hope that our work spurs future research, *e.g.*, on the exploration of divides at even finer spatial granularity, or on the ultimate demonstration of their root causes.

ACKNOWLEDGEMENTS

This research was supported by Atracción de Talento Investigador grant no. 2019-T1/TIC-16037 "NetSense" funded by Comunidad de Madrid, and by the Agence Nationale de la Recherche (ANR) through the project ANR-18-CE25-0011 "CANCAN".

REFERENCES

[1] ARCEP. 2021. *Mon réseau mobile.* Retrieved October 18, 2021 from https://www.data.gouv.fr/fr/datasets/mon-reseau-mobile/

[2] ARCEP. 2021. *Observatoire des déploiements mobiles en zones peu denses.* Retrieved October 18, 2021 from https://www.arcep.fr/index.php?id=13111

[3] J. Baillet, P. Croutte, and V. Prieur. 2019. Baromètre du numérique 2019. *CRÉDOC* Sou2019-4761 (2019). https://www.credoc.fr/publications/barometre-du-numerique-2019

[4] R. Bamford, G. Hutchinson, and B. Macon-Cooney. 2021. The Progressive Case for Universal Internet Access: How to Close the Digital Divide by 2030. *Tony Blair Institute for Global Change* (2021). https://institute.global/policy/progressive-case-universal-internet-access-how-close-digital-divide-2030

[5] Elsa Bembaron. 2021. *La 5G représente "moins de 1% du trafic", affirme Bouygues Telecom.* Retrieved October 18, 2021 from https://www.lefigaro.fr/secteur/high-tech/la-5g-represente-moins-de-1-du-trafic-affirme-bouygues-telecom-20210520

[6] Cisco. 2020. Annual Internet Report (2018–2023) White Paper. (2020). Retrieved October 18, 2021 from https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[7] Clémentine Cottineau, Erez Hatna, Elsa Arcaute, and Michael Batty. 2017. Diverse cities or the systematic paradox of Urban Scaling Laws. *Computers, Environment and Urban Systems* 63 (2017), 80–94. https://doi.org/10.1016/j.compenvurbsys.2016.04.006 Spatial analysis with census data: emerging issues and innovative approaches.

[8] Marie-Pierre de Bellefon. 2018. A scenario for the new official French urban areas zoning. In *Proceedings of the 16th Conference of the International Association of Official Statisticians.* https://www.oecd.org/iaos2018/programme/IAOS-OECD2018_DeBellefon.pdf

[9] Pierre Deville, Catherine Linard, Samuel Martin, Marius Gilbert, Forrest R Stevens, Andrea E Gaughan, Vincent D Blondel, and Andrew J Tatem. 2014. Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences* 111, 45 (2014), 15888–15893.

[10] P. DiMaggio, E. Hargittai, C. Celeste, and S. Shafer. 2001. From unequal access to differentiated use: A literature review and agenda for research on digital inequality. In *Social inequality*, K. Neckerman (Ed.). Russell Sage Foundation, New York, NY, USA, 355–400. https://www.russellsage.org/research/reports/dimaggio

[11] 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.

[12] Martin A Fischler and Robert C Bolles. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* 24, 6 (1981), 381–395.

[13] E. Hargittai and A. Hinnant. 2008. Digital Inequality: Differences in Young Adults' Use of the Internet. *Communication Research* 35, 5 (2008), 602–621. https://doi.org/10.1177/0093650208321782 arXiv:https://doi.org/10.1177/0093650208321782

[14] INSEE. 2020. *Découpage administratif communal français issu d'OpenStreetMap.* Retrieved October 18, 2021 from https://www.data.gouv.fr/en/datasets/decoupage-administratif-communal-francais-issu-d-openstreetmap/

[15] INSEE. 2021. *Base des unités urbaines.* Retrieved October 18, 2021 from //www.insee.fr/fr/information/4802589

[16] INSEE. 2021. *Definition: Commune.* Retrieved October 18, 2021 from https://www.insee.fr/fr/metadonnees/definition/c1468

[17] INSEE. 2021. *Definition: Urban unit.* Retrieved October 18, 2021 from https://www.insee.fr/en/metadonnees/definition/c1501

[18] INSEE. 2021. *Mobilités professionnelles des individus : déplacements commune de résidence / commune de travail en 2017.* Retrieved October 18, 2021 from https://www.insee.fr/fr/statistiques/4507890

[19] INSEE. 2021. *Population selon le sexe et l'âge quinquennal de 1968 à 2017.* Retrieved October 18, 2021 from https://www.insee.fr/fr/statistiques/1893204

[20] INSEE. 2021. *Structure et distribution des revenus, inégalité des niveaux de vie en 2018.* Retrieved October 18, 2021 from https://www.insee.fr/fr/statistiques/5009218

[21] P. Jonsson, S. Carson, G. Blennerud, J. Kyohun Shim, B. Arendse, A. Husseini, P. Lindberg, and K. Öhman. 2019. *Ericsson Mobility Report – November 2019.* Retrieved October 18, 2021 from https://www.ericsson.com/en/mobility-report/reports

[22] G. Khodabandelou, V. Gauthier, M. Fiore, and M. A. El-Yacoubi. 2019. Estimation of Static and Dynamic Urban Populations with Mobile Network Metadata. *IEEE Transactions on Mobile Computing* 18, 9 (2019), 2034–2047. https://doi.org/10.1109/TMC.2018.2871156

[23] Utpal Paul, Anand Prabhu Subramanian, Milind Madhav Buddhikot, and Samir R. Das. 2011. Understanding traffic dynamics in cellular data networks. In *2011 Proceedings IEEE INFOCOM.* 882–890. https://doi.org/10.1109/INFCOM.2011.5935313

[24] L. Philip, C. Cottrill, J. Farrington, F. Williams, and F. Ashmore. 2017. The digital divide: Patterns, policy and scenarios for connecting the 'final few' in rural communities across Great Britain. *Journal of Rural Studies* 54 (2017), 386–398. https://doi.org/10.1016/j.jrurstud.2016.12.002

[25] Tilo Strutz. 2016. Data fitting and uncertainty. *A practical introduction to weighted least squares and beyond* (2016).

[26] 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/

[27] Alexander JAM van Deursen and Jan AGM van Dijk. 2014. The digital divide shifts to differences in usage. *New Media & Society* 16, 3 (2014), 507–526. https://doi.org/10.1177/1461444813487959 arXiv:https://doi.org/10.1177/1461444813487959

[28] J. Van Dijk (Ed.). 2005. *The deepening divide: Inequality in the information society.* SAGE Publications, Inc, Thousand Oaks, CA, USA.

[29] Fengli Xu, Pengyu Zhang, and Yong Li. 2016. Context-aware real-time population estimation for metropolis. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing.* 1064–1075.

[30] Qiang Xu, Alexandre Gerber, Zhuoqing Morley Mao, and Jeffrey Pang. 2011. AcuLoc: Practical Localization of Performance Measurements in 3G Networks. In *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services* (Bethesda, Maryland, USA) *(MobiSys '11)*. Association for Computing Machinery, New York, NY, USA, 183–196. https://doi.org/10.1145/1999995.2000013