Not All Apps Are Created Equal: Analysis of Spatiotemporal Heterogeneity in Nationwide Mobile Service Usage

Cristina Marquez
Universidad Carlos III Madrid
mcmarque@pa.uc3m.es

Marco Gramaglia
Universidad Carlos III Madrid
mgramagl@it.uc3m.es

Marco Fiore
CNR-IEIIT
marco.fiore@ieiit.cnr.it

Albert Banchs
Universidad Carlos III Madrid
banchs@it.uc3m.es

Cezary Ziemlicki
Orange Labs
cezary.ziemlicki@orange.com

Zbigniew Smoreda
Orange Labs
zbigniew.smoreda@orange.com

ABSTRACT

We investigate how individual mobile services are consumed at a national scale, by studying data collected in a 3G/4G mobile network deployed over a major European country. Through correlation and clustering analyses, our study unveils a strong heterogeneity in the demand for different mobile services, both in time and space. In particular, we show that: *(i)* somehow surprisingly, almost all considered services exhibit quite different temporal usage patterns; *(ii)* in contrast to such temporal behavior, spatial patterns are fairly uniform across all services; *(iii)* when looking at usage patterns at different locations, the average traffic volume per user is dependent on the urbanization level, yet its temporal dynamics are not. Our findings do not only have sociological implications, but are also relevant to the orchestration of network resources.

CCS CONCEPTS

•Networks → Mobile networks; Network measurement;

1 INTRODUCTION

As the mobile Internet traffic grows along with the quantity and diversity of offered services, it becomes increasingly important to understand the demands generated by such services. Indeed, characterizing the traffic dynamics associated to different mobile services is of paramount importance in order to properly dimension and orchestrate the mobile network, and also offers an opportunity to unravel broader societal behaviors in general.

Our comprehension of current mobile service usages is however still superficial. As later elaborated in our review of related works, most previous works in the area are concerned with coarse-grained service categories or rely on relatively small-sized datasets. As a result, they provide limited knowledge about the specificities of individual services and their differences.

In this paper, we contribute to fill the gap above, by analyzing the usage of a selection of mobile services at a national scale. Our study investigates the traffic behavior of specific services across time (*i.e.*, temporal usage patterns) and space (*i.e.*, at different locations). In doing so, we offer a global overview of the traffic dynamics for specific applications in a large-scale operational cellular network.

The results and insights gained from our analysis may find applications in different areas. In future-generation mobile networks, the understanding of when, where and how different mobile services are consumed will be essential to dynamically tailor resources to the actual fluctuations of the subscribers' activity. Indeed, many novel architectural paradigms aim at enabling the dynamic management of system resources, across multiple network functions at the network edge or core [1–3]. For instance, an effective orchestration of network slices builds on the spatial complementarity of the demands for the different services [4].

The characterization of mobile service consumption carried out in this paper is also relevant to disciplines beyond networking. In fact, it allows observing social phenomena at unprecedented scales, unveiling interplays between the digital and physical worlds that are relevant to, *e.g.*, urban development [5] or planning [6, 7].

Related work. The vast majority of analyses of cellular traffic builds on measurements and accounting data of voice calling and texting activities, such as Call Detail Records (CDR); a thorough review is in [8]. Also, a number of works have investigated the properties of mobile data traffic from a high-level perspective, aggregating the load of all services [9–11]. The approaches above allow inferring important information on the communication patterns of users and on the overall data traffic they generate, but clearly do not explore subscribers' behavior on a per-service basis.

The literature becomes fairly thin when considering the usage of specific mobile services. Previous works have almost exclusively addressed the traffic dynamics of broad service categories (*e.g.*, *video* or *chat*) that encompass all mobile services of a kind. Service categories were proven to display interesting properties, including strong locality [12–14], high predictability [15], and adoption by well-outlined user groups [16, 17]. However, such broad categories hide the peculiarities of each service that, as we show in this paper, are not negligible and deserve a dedicated investigation.

Considering individual mobile services (*e.g.*, *YouTube* or *WhatsApp*), they have been studied in isolation [18, 19], or within the scope of single categories such as video streaming [20] and mobile cloud [21]. As far as we know, the only work to consider a huge number of heterogeneous mobile services at once is that in [22].

CoNEXT '17, Incheon, Republic of Korea

DOI: 10.1145/3143361.3143369

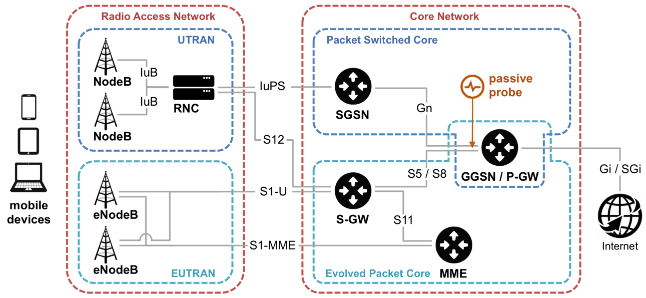

Figure 1: Simplified 3G/4G mobile network.

There, the aim is comparing cellular and wireline traffic statistics for a relatively small (20-50,000) user population: both the purpose and the scale of the analysis are sensibly different from

Key insights. In contrast to previous works, in this paper we analyze fine-grained service consumption from a nationwide dataset. Our analysis reveals that:

(i) Mobile services tend to have different temporal dynamics, characterized by unique patterns of activity peaks. Interestingly, such a diversity is also observed across services of a same kind, unveiling a heterogeneity of behaviors that was not observed before.

(ii) The nationwide geographical distribution of usage is instead very similar for our choice of mobile services. In all cases, the demand variations across space are strongly driven by the urbanization level and the layout of long-distance transport networks.

(iii) Although the level of urbanization has an impact on how much subscribers consume mobile services, it has less influence on when they do so: the temporal dynamics of a given service are indeed similar at all geographical locations.

2 MEASUREMENTS AND DATASET

The dataset we employ in our study was collected in the core network of Orange, a major European mobile operator. It describes the mobile traffic generated by the whole Orange subscriber base in France, *i.e.*, a user population of approximately 30 million individuals distributed over more than 550,000 km^2. The data cover one week, starting on September 24, 2016. The time frame allows capturing the vast majority of mobile traffic dynamics, which are known to occur over weekly timescales [15, 16], while avoiding that the dataset size becomes unmanageable. To maximise the generality of our results, the measurement week was carefully selected so as to avoid major nationwide events like holidays or strikes.

A simplified representation of the Orange 3G/4G mobile network architecture is portrayed in Fig. 1. The figure is limited to the 3G UTRAN and packet switched core, and to the 4G EUTRAN and evolved packet core, as our focus is on data traffic. The data were recorded by passive probes at the Gn and S5/S8 interfaces of the Gateway GPRS Support Node (GGSN) and of the Packet Data Network Gateway (P-GW). The 3G and 4G gateways are conveniently co-located, which eases the probe deployment, management and synchronization. The probes inspect IP traffic on the GPRS Tunneling Protocol user plane (GTP-U), and extract information on the transport- and application-layer protocols of each user session. The specific mobile service associated to each IP session is detected by the mobile network operator via Deep Packet Inspection (DPI)

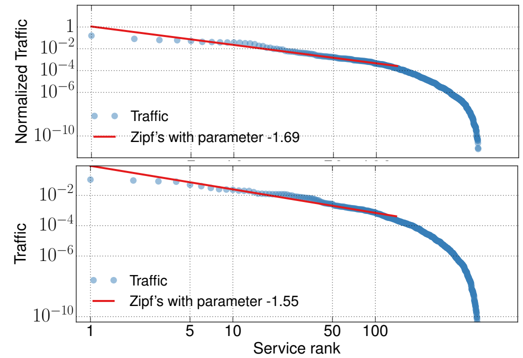

Figure 2: Ranking of mobile services on downlink (top) and uplink (bottom) traffic volume.

and multiple fingerprinting techniques, each tailored to a specific traffic type. These operations can classify 88% of the mobile traffic; however, their implementation is proprietary to the mobile network operator, which prevents us from providing further details[1].

Geo-referencing of the IP sessions, and of the corresponding mobile service usages, is performed by examining the User Location Information (ULI) contained in the 3G Packet Data Protocol (PDP) Contexts and 4G Evolved Packet System (EPS) Bearers. These data structures are transferred over the GPRS Tunneling Protocol control plane (GTP-C), which also transits through Gn and S5/S8 interfaces, making their inspection straightforward. The localization granted by this approach is fairly coarse, since the ULI is updated upon possibly infrequent events, *i.e.*, the establishment of a new IP session, and handovers across access technologies (2.5G, 3G, 4G, Wi-Fi) or Routing/Tracking Areas (RA/TA).

This limits the spatial accuracy of the dataset. As prior analyses showed that the median error of ULI is around 3 km [23], in our study we consider an appropriate tessellation of space at the level of *communes*. These are over 36,000 administrative regions whose union covers the whole France, and whose average surface is around 16 km^2. We thus associate each base station to the commune where it is deployed, and aggregate at the commune level all traffic mapped by ULI to base stations in that commune.

Privacy, ethics and legal issues. Our study does not breach user privacy, or raises ethical or legal issues. On the one hand, we do not process individual or personal data; indeed, the dataset is implicitly and strongly anonymized by the geographical aggregation at the commune level, which ensures that mobile service demands are merged over several thousands of subscribers. On the other hand, all data used in the analysis classify as secondary use data; *i.e.*, the data were not collected by the mobile network operator specifically for our study, but prior to it under the control of Orange Data Privacy Officer and in compliance with applicable regulations.

3 MOBILE SERVICES: OVERVIEW

The dataset contains information about over 500 mobile services that generate some traffic during the measurement period. Fig. 2 shows their ranking on the normalized traffic volume, in downlink and uplink. In both cases, rankings for the top half of services

[1]Due to similar confidentiality restrictions, we do not disclose the absolute values of traffic volumes in the paper, and limit our analysis to percentages of the total demand.

fit Zipf's distributions with similar parameters, at 1.69 and 1.55 respectively, before a cut-off intervenes that separates the bottom half of services.

When comparing this plot to the equivalent one in [15], referring to 3G traffic measured in a US state in 2010, we remark that: *(i)* as in [15], the per-service traffic volumes span around 10 orders of magnitude, denoting a strong imbalance among service loads; *(ii)* however, the distribution of traffic volume differs significantly from that in [15]: as mentioned before, only the top half of the services follows a Zipf distribution, and furthermore this Zipf distribution exhibits much lower parameters than the 4.74 value recorded in [15]. This latter observation highlights how per-service mobile traffic has evolved over the past six years: top-ranked applications now share more evenly the global demand, but a large number of very low-traffic services has also emerged.

In this paper, we focus on the head of the distribution, which is composed of 20 representative services summarized in Fig. 3. This subset of mobile services covers a large fraction (over 60%) of the overall network traffic and spans across a variety of service categories with diverse requirements in terms of network performance, such as *video streaming*, *gaming*, and *social networks*. The plots list the services (names on bars), categorized (colors of bars, as per the legend) and ranked on the relative traffic volume they generate.

In downlink, video streaming services tend to dominate mobile downloads, with an aggregate figure at over 46% of the total traffic. This is a non-negligible increase over the 36% performance recorded six years ago in downstream cellular traffic [20]. It is also quite a different value from the 60% reported by Cisco[2] in their yearly forecast [24], for both cellular and Wi-Fi traffic: this lets us speculate that subscribers may drastically reduce access to video streaming services when Wi-Fi is not available. YouTube emerges as the dominant provider, followed at a distance by iTunes.

Things change significantly in the uplink direction. Here, social networks and messaging services occupy the top three positions. This is not surprising, since services such as SnapChat and Facebook are oriented at content-sharing within limited circles (*e.g.*, of friends or contacts). Such a small potential audience of the content reduces the number of visualizations for the published material, ultimately resulting in high upstream-to-downstream traffic ratios.

4 NATIONWIDE TIME DYNAMICS

Our first goal is investigating the temporal dynamics of different mobile services. We focus on traffic at the national scale, aggregating the weekly demand for each service in space, over all communes.

Examples of the resulting time series are shown in Fig. 4, for four sample mobile services in downlink. In all cases, classic patterns can be observed, *i.e.*, higher diurnal activity versus much reduced overnight traffic, and a distinctive dichotomy between weekends and working days. In addition, the time series of each service is characterized by a variety of fluctuations. For instance, the first plot in Fig. 4, which refers to Facebook, displays a major traffic peak at midday of working days, plus several other minor peaks. However, other services in Fig. 4 show other traffic peak arrangements.

[2]Cisco statistics refer to the aggregate downlink and uplink traffic. However, the latter accounts for less than one twentieth of the total network load in our case, hence it does not affect our conclusions.

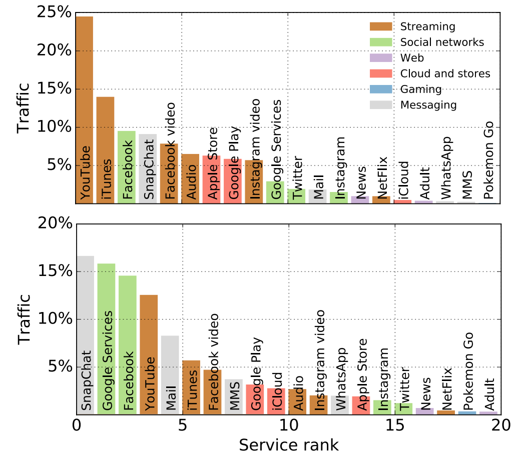

Figure 3: Selected mobile services, ranked on downlink (top) and uplink (bottom) traffic volume.

Motivated by this last observation, we study how similar are the temporal patterns of mobile service usages in the whole France. Our aim is understanding whether the diversity of behaviors holds in general or is specific to the services considered in Fig. 4.

We first make an attempt at grouping our 20 selected services based on similarity of their time series. Even if the original number of services is limited, a simple visual inspection of their time series can lead to subjective observations, challenging the validity and reproducibility of conclusions. We thus favor a sounder approach, and relay on a suitable clustering algorithm to carry out the classification task. Our choice is *k-shape*, which is the current state-of-the-art unsupervised technique for time series clustering, as proven by extensive tests [25]. We then carry out an exhaustive search for mobile service clusters, by testing k-shape on all possible values of k, in combination with multiple indices of clustering quality. The latter are the (modified) Davies-Bouldin, Dunn, and Silhouette indices, which constitute a representative selection of popular indices used in the literature to rank different cluster sets generated from the same original elements [26]. Unfortunately, the outcome is inconclusive. As shown in Fig. 5, none of the indices pinpoints a value of k as a clear winner. Instead, all indices indicate steadily decreasing clustering quality as k grows. Also, a thorough manual examination of the internal structure of the clusters generated by k-shape for different k does not reveal any consistent grouping of mobile services. We interpret these results as an indication that the temporal dynamics of our considered mobile services may be very distinctive, which makes them not easily equated.

The observations from Fig. 4 suggest that the cause for the above behaviour may lie in the different patterns of activity peaks that characterize each service. In order to verify this possibility, we detect the activity peaks in the per-service time series using the *smoothed z-score algorithm*[3]. It compares the original signal versus a smoothed version of its z-score, and tags values higher than a threshold as peaks. It takes three parameters, controlling the threshold value (*threshold*) and the z-score smoothing, via the interval of

[3]Implementation available at https://gist.github.com/ximeg/.

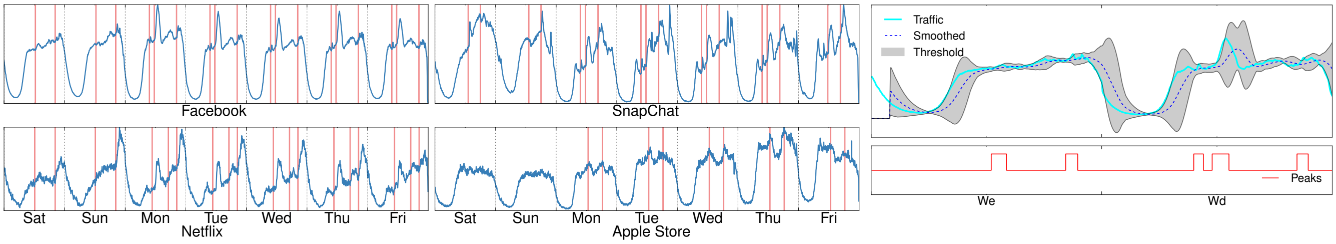

Figure 4: Sample time series of mobile services (left): vertical lines highlight activity peaks detected in the time series by the smoothed z-score algorithm. Example of smoothed z-score algorithm operation (right): Facebook.

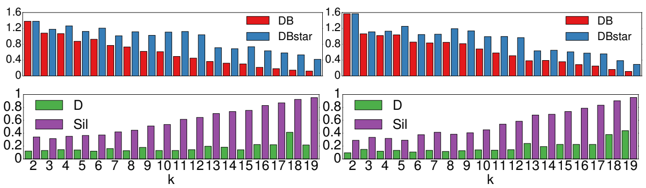

Figure 5: Clustering quality indices versus the cluster number, in downlink (left) and uplink (right). Plots for Davies-Bouldin, modified Davies-Bouldin (top, minimum is best) and Dunn, Silhouette (bottom, maximum is best) indices.

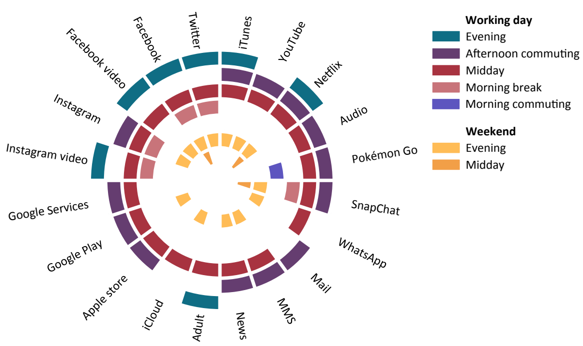

Figure 6: Activity peak times of mobile services.

past samples (*lag*) and their weight with respect to the current sample (*influence*). We set these parameters to 3 z-scores, 2 hours and 0.4, respectively, upon an extensive tuning process. An illustration of smoothed z-score peak detection is provided in the right plots of Fig. 4, for the Facebook case. The top plot portrays the original signal, its smoothed version, and the range around it determined by the z-score threshold: when the original signal exceeds the range boundaries, a peak is detected in the bottom plot.

Examples of peaks inferred by the smoothed z-score are in the left plots of Fig. 4, where vertical red lines denote the rising front of peaks. Interestingly, by applying this methodology to all mobile services, we find that peaks only appear at seven specific moments during the week: at midday (around 1pm) and evenings (9pm) during weekends, and during the morning commuting time (8am), morning break (10am), midday (1pm), afternoon commuting time (6pm) and evenings (9pm) during working days.

This lets us summarize the peak patterns observed for all services as done in Fig. 6. In this figure, each sector refers to one mobile service, and each ring to a different topical time, as per the legend. We remark that: *(i)* individual services tend to have very diverse patterns even when looking only at when they show peaks of activity; *(ii)* this heterogeneity also separates services that belong to a same category, *e.g.*, video streaming behaves quite differently in YouTube, Facebook, Instagram, Netflix and iTunes platforms.

Other specific behaviors have interesting implications. For instance, almost all services show increased usage on midday of working days. Similarly, large (different) sets of services have activity peaks during the afternoon commuting time and during weekend evenings. In all these cases, the increased usage affects services of various nature, indicating that subscribers with different interests all tend to consume mobile services at those times. On the contrary, we speculate that morning break activity peaks may highlight services that are popular among students, who access them during the pause between classes: coherently, these include SnapChat, Instagram, Facebook, and Twitter.

Fig. 7 offers an in-depth view of the activity peaks: it displays, for each topical time, the actual activity peak intensity of every service. This is computed as the ratio between the maximum and minimum traffic volumes recorded during the peak intervals as detected by the smoothed z-score algorithm. The key observation here is that services with demand peaks at a same time in fact undergo very diverse variations of activity. Overall, the diversity of activity peaks, both in timing and intensity, corroborates the intuition that temporal fluctuations in the usage of individual mobile services are very heterogeneous. These results explain the poor outcome of a clustering based on time series, and ultimately demonstrate how each mobile service has unique dynamics, dictated by the classes of subscribers consuming it.

5 SERVICE USAGE GEOGRAPHY

Mobile services show significant peaks of activity not only in time, but also in space. An example is provided in Fig. 8 for the specific case of Twitter. The left plot shows the cumulative weekly downlink and uplink traffic recorded in the ranked communes. We observe that the top 1% and 10% of the communes generate over 50% and 90% of the Twitter traffic, respectively.

The strong imbalance in the demand recorded across communes is expected, considering the high heterogeneity of population density that characterizes France. What is less obvious is that such variability remains strong also when considering the average traffic generated by a subscriber, obtained as the ratio of the traffic volume to the average number of users in each commune. The right plot in Fig. 8 shows the Cumulative Distribution Function (CDF) of the *per-subscriber* Twitter usage over all communes. The distribution is highly skewed: subscribers in half of the communes consume

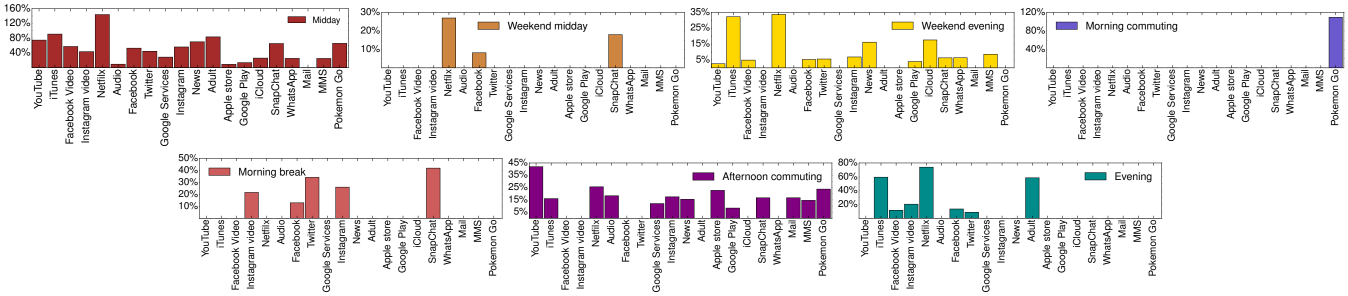

Figure 7: Peak-to-average ratios measured for each mobile service at different topical times.

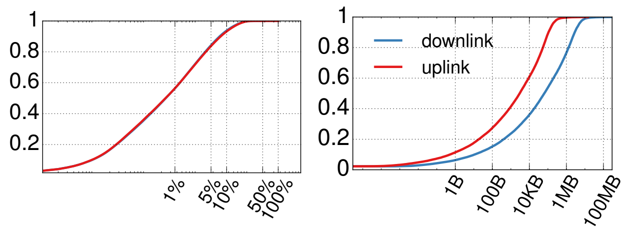

Figure 8: Twitter. Cumulative traffic on ranked communes (left). CDF of per-subscriber traffic on all communes (right).

a negligible weekly Twitter load of a few 1 Kbytes, whereas users in other areas download tens of Mbytes of Twitter contents per week. Basically, individual mobile users who live in distinct areas of France tend to use Twitter services in very different quantity.

The left plot in Fig. 9 reports a map of the weekly per-subscriber Twitter traffic in downlink, and helps to visualize this phenomenon. The map evidences how users living in large cities (*e.g.*, Paris, Lyon and Marseille) and traveling along major transportation arteries (*e.g.*, the high-speed TGV train lines connecting the three cities above) tend to generate significant demands. Instead, subscribers in rural areas located far from major cities and transportation infrastructures are prone to make little use of mobile services.

The considerations above refer to Twitter, but they are valid for any mobile service. This is shown in Fig. 10. The left plot is a CDF of the geographical correlation of mobile service usage: we represent each mobile service as a vector of the weekly per-subscriber traffic recorded in each commune, and compute the coefficient of determination between the vectors of each pair of services. The majority of values in the CDF are strongly positive, with an average of 0.60 and 0.53 for downlink and uplink respectively. The middle and right plots detail the correlation among specific service pairs: low correlations are only experienced with Netflix (almost completely absent in rural areas) and iCloud (pushing uplink data from all iPhones, and thus more uniformly distributed over the country). These outlier cases apart, our results let us conclude that mobile services tend to be consumed similarly over the French territory. Such mobile data consumption features geographical distributions of the per-user traffic that are highly skewed, as in Fig. 8, with subscribers within cities and on inter-city routes that tend to be more active than those in other areas, as per the left plot in Fig. 9.

The middle plot in Fig. 9 provides instead additional detail on the Netflix outlier, showing the corresponding map of weekly per-subscriber traffic. Densely inhabited city centers and major transportation lines stand out much more clearly than in the typical case, in the left plot of the same figure. This occurs also because Netflix usage is dramatically low, or even absent, in large regions of rural France. We partly ascribe such a strong duality of adoption to the high-end nature of Netflix as a mobile service, with users in cities more prone to embrace novel applications. In addition, the Netflix case also gives us the opportunity to discuss how the mobile network technology is an important factor that can further explain the success of specific services. Streaming high-quality long-duration videos requires substantial capacity and quality of service, and indeed the 3G and 4G coverage in France, in the right plot of Fig. 9, seems to drive Netflix usage over the country. In the case of other services, such as Twitter, the spatial distribution is more uniform; when looking at the 3G and 4G coverage, this suggests that (pervasive) 3G already provides sufficient performance, and makes demands less dependent on the cellular technology.

In order to further explore the impact of urbanization on the way mobile services are consumed, we group communes in France into *urban*, *semi-urban* and *rural*, according to classifications of the French National Institute of Statistics[4]; in addition, we consider rural communes that are crossed by a high-speed train line into a separate *TGV* category. We then aggregate all traffic recorded in the communes of a same group, and investigate their relationships.

The top plot in Fig. 11 summarizes, for each service, *how much* traffic is generated by the average individual subscriber in semi-urban, rural and TGV regions with respect to a typical user located in a urban area of France. More precisely, each bar represents the slope of the linear least square regression of per-subscriber time series in urban and (from darker to lighter) semi-urban, rural and TGV regions. The plot highlights that: *(i)* semi-urban and urban areas present similar individual service usage levels, *i.e.*, the coefficient is close to 1; *(ii)* subscribers in rural areas consume around a half of the mobile service data than their counterparts in urban zones do; *(iii)* users on high-speed trains generate on average twice or more the volume of traffic of urban users.

Interestingly, all these results are fairly consistent across services. They unveil how users in cities, from medium-sized towns to large metropolis, show equivalent mobile services consumptions. On the other hand, urban mobile service consumptions are twice as large as those of people living in the countryside, which may be covered by 3G connectivity only. That trend is exacerbated for passengers on high speed trains, who are much more prone to use mobile services during their travels: rural communes belonging to

[4] https://www.insee.fr/fr/information/2115011

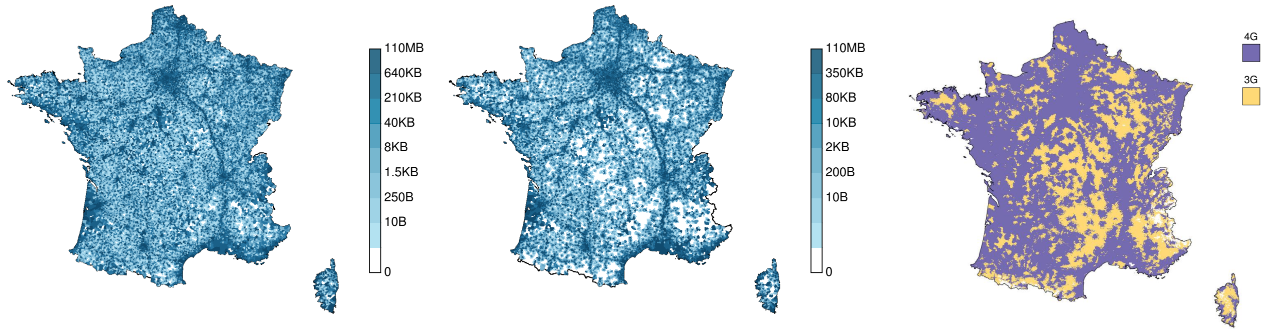

Figure 9: Maps of the average per-subscriber activity for downlink Twitter (left), Netflix (middle). Coverage of Orange 3G and 4G cellular technologies in France (right).

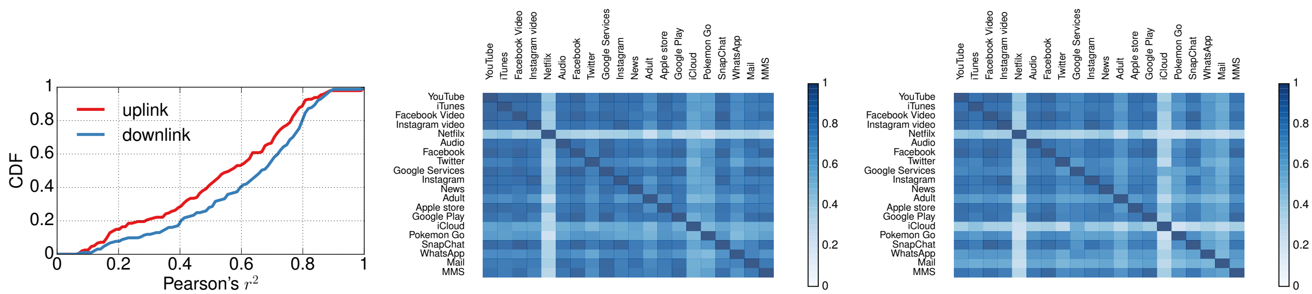

Figure 10: Per-user traffic spatial correlation between services. Pearsons' r^2 CDF computed between the per-user traffic maps of all service pairs (left). Pairwise coefficients of Pearsons' r^2 for downlink (middle) and uplink (right).

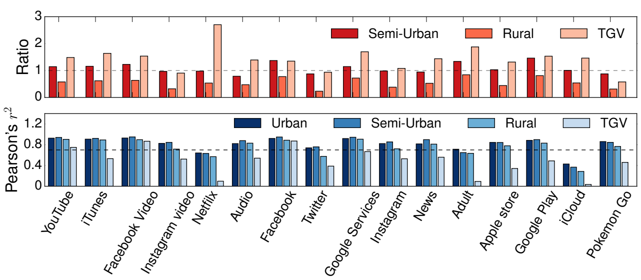

Figure 11: All services. Per-user traffic volume ratios among urbanization levels (top). Correlation of per-user traffic time series among urbanization levels (bottom).

the TGV category show completely opposite relative consumption trends when compared to non-TGV ones.

The bottom plot in Fig. 11 assesses instead if the urbanization level plays a role in *when* the typical subscriber access mobile services. The bars represent the mean coefficient of determination between the time series of a same service recorded in one type of region and those of the other types. Across the vast majority of services, correlations are high for combinations involving urban, semi-urban and rural areas; thus, we conclude that the level of urbanization has very little impact on the temporal dynamics of service usage. Instead, subscribers on TGVs have quite different temporal patterns than users in the rest of France. We argue that the cause is a combination of train schedules (constraining usages in time), and the intrinsic nature of some services (*e.g.*, TGV seats are probably not the best environment for adult websites browsing).

6 CONCLUSIONS

We investigated the temporal and spatial dynamics of mobile services at a national scale which allowed us to unveil interesting macroscopic properties of traffic that had not been previously observed. A first finding is that no two services exhibit similar time patterns in their nationwide aggregate traffic: although expected for different service categories, this is less obvious for akin services, e.g., diverse applications that all provide video streaming. A second key insight is that mobile services have very comparable geographical distributions of both total and the per-user traffic demands. That is, different services have different temporal patterns (*i.e.*, they are consumed at different times), but their geographical patterns (*i.e.*, locations where they are consumed) are very similar. Along these lines, our third takeaway message is that spatial distributions of per-subscriber service usage are in fact driven by land use, *i.e.*, the urbanization level plays a major role in influencing how much mobile services users consume. Nonetheless, it has a much lower impact on when they do so, as the average subscribers in urban, semi-urban and rural regions all follow similar service access patterns; a notable exception is represented by users on high-speed trains, who show instead unique time dynamics.

ACKNOWLEDGEMENTS

This research work has been performed in the framework of the H2020-ICT-2014-2 project 5G NORMA (Grant Agreement No. 671584).

REFERENCES

[1] China Mobile Research Institute, "C-RAN: The Road Towards Green RAN," White Paper, 2011

[2] T.X. Tran, A. Hajisami, P. Pandey, D. Pompili, "Collaborative Mobile Edge Computing in 5G Networks: New Paradigms, Scenarios, and Challenges," IEEE Communications Magazine, 55(4):54-61, 2017.

[3] D. Tuncer, M. Charalambides, S. Clayman, G. Pavlou, "Adaptive Resource Management and Control in Software Defined Networks," IEEE Transactions on Network and Service Management, 12(1):18-33, 2015.

[4] P. Rost, C. Mannweiler, D. S. Michalopoulos, C. Sartori, V. Sciancalepore, N. Sastry, O. Holland, S. Tayade, B. Han, D. Bega, D. Aziz, H. Bakker, "Network Slicing to Enable Scalability and Flexibility in 5G Mobile Networks," IEEE Communications Magazine, 55(5):72-79, 2017.

[5] M. De Nadai, J. Staiano, R. Larcher, N. Sebe, D. Quercia, B. Lepri, "The Death and Life of Great Italian Cities: A Mobile Phone Data Perspective," ACM WWW, Montréal, QC, Canada, 2016.

[6] B. Cici, M. Gjoka, A. Markopoulou, C.T. Butts, "On the Decomposition of Cell Phone Activity Patterns and their Connection with Urban Ecology," ACM MobiHoc, Hangzhou, China, 2015.

[7] A. Furno, M. Fiore, R. Stanica, C. Ziemlicki, Z. Smoreda, "A Tale of Ten Cities: Characterizing Signatures of Mobile Traffic in Urban Areas," IEEE Transactions on Mobile Computing, 16(10):2682-2696, 2017.

[8] D. Naboulsi, M. Fiore, S. Ribot, R. Stanica, "Large-scale Mobile Traffic Analysis: A Survey," IEEE Communications Surveys and Tutorials, 18(1):124-161, 2016.

[9] C. Williamson, E. Halepovic, H. Sun, Y. Wu, "Characterization of CDMA2000 Cellular Data Network Traffic," IEEE LCN, Sydney, Australia, 2005.

[10] U. Paul, A.P. Subramanian, M.M. Buddhikot, S.R. Das, "Understanding Traffic Dynamics in Cellular Data Networks," IEEE INFOCOM, Shanghai, PRC, 2011.

[11] F. Xu, Y. Li, H. Wang, P. Zhang, D. Jin, "Understanding Mobile Traffic Patterns of Large Scale Cellular Towers in Urban Environment," IEEE/ACM Transactions on Networking, to appear.

[12] I. Trestian, S. Ranjan, A. Kuzmanovic, A. Nucci, "Measuring Serendipity: Connecting People, Locations and Interests in a Mobile 3G Network," ACM IMC, Chicago, IL, USA, 2009.

[13] Q. Xu, J. Erman, A. Gerber, Z. Mao, J. Pang, S. Venkataraman, "Identifying diverse usage behaviors of smartphone apps," ACM IMC, Berlin, Germany, 2011.

[14] M.Z. Shafiq, L. Ji, A.X. Liu, J. Pang, J. Wang, "Characterizing Geospatial Dynamics of Application Usage in a 3G Cellular Data Network," IEEE INFOCOM, Orlando, FL, USA, 2012.

[15] M.Z. Shafiq, L. Ji, A.X. Liu, J. Wang, "Characterizing and Modeling Internet Traffic Dynamics of Cellular Devices," ACM SIGMETRICS, San Jose, CA, USA, 2011.

[16] R. Keralapura, A. Nucci, Z.-L. Zhang, L. Gao, "Profiling Users in a 3G Network Using Hourglass Co-Clustering," ACM MobiCom, Chicago, IL, USA, 2010.

[17] H. Li, X. Lu, X. Liu, T. Xie, K. Bian, F.X. Lin, Q. Mei, F. Feng, "Characterizing Smartphone Usage Patterns from Millions of Android Users," ACM IMC, Tokyo, Japan, 2015.

[18] P. Fiadino, M. Schiavone, P. Casas, "Vivisecting WhatsApp through Large-scale Measurements in Mobile Networks," ACM SIGCOMM Computer Communication Review, 44(4), 2014.

[19] Q. Deng, Z. Li, Q. Wu, C. Xu, G. Xie, "An Empirical Study of the WeChat Mobile Instant Messaging Service," Global Internet Symposium, Altanta, GA, USA, 2017.

[20] J. Erman, A. Gerber, K.K. Ramadrishnan, S. Sen, O. Spatscheck, "Over the top video: the gorilla in cellular networks," ACM IMC, Berlin, Germany, 2011.

[21] Z. Li, X. Wang, N. Huang, M.A. Kaafar, Z. Li, J. Zhou, G. Xie, P. Steenkiste, "An Empirical Analysis of a Large-scale Mobile Cloud Storage Service," ACM IMC, Santa Monica, CA, USA, 2016.

[22] Y. Zhang, A. Arvidsson, "Understanding the Characteristics of Cellular Data Traffic," ACM CellNet, Helsinki, Finland, 2012.

[23] Q. Xu, A. Gerber, Z. Morley, J. Pang, "AccuLoc: Practical Localization of Performance Measurements in 3G Networks," ACM MobiSys, Bethesda, MA, USA, 2011.

[24] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021," 2017.

[25] J. Paparrizos, L. Gravano, "k-Shape: Efficient and Accurate Clustering of Time Series," ACM SIGMOD, Melbourne, Victoria, Australia, May 2015.

[26] G.W. Milligan, M.C. Cooper, "An Examination of Procedures for Determining the Number of Clusters in a Data Set," Psychometrika, 50(2):159–179, 1985.