# Urban Vibes and Rural Charms: Analysis of Geographic Diversity in Mobile Service Usage at National Scale

Rajkarn Singh, Marco Fiore†, Mahesh K. Marina, Alessandro Nordio† and Alberto Tarable†
The University of Edinburgh UK, †CNR-IEIIT Italy
Email: r.singh@ed.ac.uk, marco.fiore@ieiit.cnr.it, mahesh@ed.ac.uk, alessandro.nordio@ieiit.cnr.it, alberto.tarable@ieiit.cnr.it

## I. Problem and Results

As mobile data traffic keeps surging worldwide, knowledge of where, when, how and why mobile services are consumed by network subscribers becomes increasingly relevant across research and technology domains. Still, our comprehension of mobile service adoption is currently limited and many questions remain unanswered. We focus on one such open question, namely: *"how similar (or different) are demands for mobile services across a whole country?"* We answer by analyzing a real-world dataset of mobile network traffic collected by a major operator that describes the demands for individual services in 10,000 *communes* (*i.e.*, administrative areas) in France. Our study yields the following key insights.

- The demand for most popular mobile services is fairly uniform across the whole country, and only a reduced set of peculiar services (mainly operating system updates and long-lived video streaming) yields geographic diversity.
- Just 9 (respectively, 50) service consumption patterns are sufficient to retain 23% (respectively, 35%) of the overall usage diversity, implying that a small number of distinct behaviors is sufficient to characterize the many thousands of areas in the whole of France.
- The spatial distribution of these behaviors correlates well with the urbanization level, ultimately suggesting that the adoption of geographically-diverse mobile applications is linked to a dichotomy of cities and rural areas.

## II. Methodology

In order to derive our results, we model the amount of traffic generated by mobile services in different communes as jointly distributed random variables, and adopt an approach that hinges on information theory. We first assess the global geographic diversity of service usages in France, by computing the *mutual information* $I(C;S)$ of mobile service demands ($\mathcal{S}$) and geographical locations ($\mathcal{C}$). The mutual information captures how much can be inferred about the consumed services by knowing the commune, and we find it to be close to zero in our nationwide scenario, implying that mobile services demand distributions are homogeneous across the whole of France.

However, usage distributions are highly skewed, with a few services generating the vast majority of traffic. Thus, the mutual information primarily captures the (absence of) diversity in the consumption of popular apps. We investigate if, among less popular services, there exist some that are *informative*, *i.e.*, are characterized by a non-negligible diversity of usage across geographical areas. In information theoretical terms, this can be formulated as the combinatorial problem of identifying a subset $\mathcal{S}' \subset \mathcal{S}$ of all mobile services whose mutual information with the spatial areas is maximized, *i.e.*,

$$\mathcal{S}'_{\text{opt}} = \arg\max_{\mathcal{S}' \subseteq \mathcal{S}} I(C|_{\mathcal{S}'}; S|_{\mathcal{S}'}).$$

An efficient approximation of the solution is provided by the Blahut-Arimoto algorithm [1], [2], which maximizes $I(C;S)$ by weighting all services according to their informativeness. By applying the approach to our scenario, 13 informative services are singled out. They are listed in Figure 1.

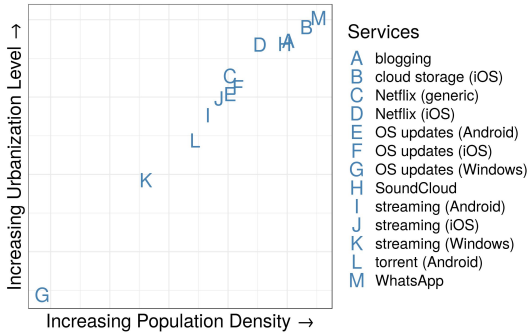


Fig. 1. Relative positioning of informative mobile services in the space of population density and urbanization levels. People living in rural regions of France have a preference to use Windows Mobile devices. Instead, inhabitants of metropolitan areas prefer Apple iPhones, and display a remarkable tendency to significantly use WhatsApp and long-lived streaming services like Netflix.

In order to understand *where* the informative services differ in usage, we cluster communes in France based on how their inhabitants consume such services. To this end, we measure the distance between two communes $i$ and $i'$ as the loss of mutual information $d(i,i') = I(K;S) - I(K_{i,i'};S)$ incurred when they are merged [3]. Here $K_{i,i'}$ is the set of (clusters of) communes in $K$ such that $i \in K$ and $i' \in K$ are merged. We feed such distances to a scalable two-phase greedy hierarchical clustering algorithm, and find that a significant portion of the system information is captured by a small number of clusters, *i.e.*, mobile service consumption patterns. Interestingly, the geographical distribution of these usage patterns over the French territory is strongly correlated with population density and urbanization levels, as in Figure 1.

## References

[1] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," IEEE Trans. on Inf. Theory 18(1), 1972.
[2] R. Blahut, "Computation of channel capacity and rate-distortion functions," IEEE Transactions on Information Theory 18(4), 1972.
[3] P.S. Chodrow, "Structure and information in spatial segregation," Proceedings of the National Academy of Sciences 114(44), 2017.