# Identifying Common Periodicities in Mobile Service Demands with Spectral Analysis

Cristina Marquez*, Marco Gramaglia*, Marco Fiore†, Albert Banchs*†, and Zbigniew Smoreda‡

*University Carlos III of Madrid, Madrid, Spain, Email: mcmarque@pa.uc3m.es, mgramagl@it.uc3m.es
†IMDEA Networks Institute, Madrid, Spain, Email: {marco.fiore, albert.banchs}@imdea.org
‡SENSe Department, Orange Labs, Châtillon, France, Email: zbigniew.smoreda@orange.com

*Abstract*—In this paper, we investigate the existence and prevalence of comparable dynamics in the temporal fluctuations for the traffic demands generated by mobile applications. To this end, we hinge upon a spectral analysis framework, by computing Discrete Fourier Transforms of the typical demands for tens of popular mobile services observed in an operational metropolitan-scale network. We filter, cluster, and analyse hundreds of frequency components, and identify a substantial set of regular patterns that are common across most service demands. We also unveil how several mobile services defy classification, and have instead highly distinguishing temporal dynamics.

## I. INTRODUCTION

Mobile services permeate today our lives, by providing seamless access to all kinds of information (*e.g.*, local and global news, messages of friends, or updates on acquaintances) and continuously assisting our decisions and actions (*e.g.*, in taking specific transportation means or trip routes or planning a daily schedule). In order to greatly facilitate many such daily tasks, mobile services generate a substantial amount of data traffic that flows from and to mobile devices. These traffic demands undergo complex fluctuations in both space and time, which are driven by the mobility and diverse endeavours of a vast user population.

Understanding the dynamics of the demands for mobile services is an important objective that helps to unveil the interactions between human activities and modern communication tools. Models explaining and possibly anticipating mobile service consumption in space-time have significant applications in both social sciences and technology. In social sciences, they can help drawing causal links between land use and the way mobile apps are used [1], highlighting cultural factors in apps adoption [2], or even detecting psychiatric disorder states at scale [3]. From an engineering and technology viewpoint, the knowledge of large-scale traffic volumes generated by each mobile service can enable a more efficient dimensioning and management of the communication infrastructure [4], the optimized caching of apps data at mobile devices [5], or the improved planning of urban transport systems based on app user flows [6].

However, disentangling mobile service demand patterns is a challenging exercise. The many and varied apps running on mobile devices entail strongly heterogeneous dynamics over a space that is high-dimensional along both its temporal (where measurement data can encompass long periods of months with a fine granularity of minutes) and geographical (with traffic information concurrently recorded at hundreds of locations within, *e.g.*, a single metropolitan area) facets. In addition, measurements are often noisy, due to inherent randomness in user access to apps [7], oscillations in device associations to the radio access infrastructure caused by signal strength fluctuations, or load balancing policies [8], or positioning accuracy limitations of the mobile network technology [9].

Previous studies aiming at characterizing consumption motifs of mobile services have typically adopted an *aggregate* perspective, by considering the cumulative mobile data traffic of all apps [10]–[12], or have investigated whole *service categories*, by considering the total network activity of, *e.g.*, video streaming, social network, or dating apps [13]–[17]. These works have revealed important properties of combined app demands, such as their locality or predictability; however, they could not unravel features that are specific to individual services. When focusing on single service-level demand analysis, the literature is much thinner. The peculiarities of a few popular services have been studied in isolation [18], [19] of for multiple apps within the same class [20], [21]. These works highlighted, for instance, the severe burst that underpin WeChat data traffic, or the uplink load that dominated mobile cloud storage services. To the best of our knowledge, only a couple of works analyzed a large set of heterogeneous individual mobile services to date. They focus on quantifying the statistical diversity of apps in terms of network-level performance indicators [22], and forecasting future demands for each app separately [23], respectively.

Unlike the studies above, we are interested in comparing the demand patterns of different mobile services, exploring eventual common behaviors. Previous research on this topic unveiled that mobile apps tend to yield fairly comparable geographical pattern of consumption at national or regional scales [24], [25]. However, the same does not hold for temporal dynamics: apps have very diverse time series and even apps belonging to the same class feature unique combinations of activity peaks. This makes any attempt at clustering mobile services along a time dimension vain [24].

In this paper, we show that a spectral approach can in fact reveal significant common properties in the time series of data traffic generated by many diverse mobile services. Such shared traits emerge in the form of periodicity in the demand fluctuations: we find that most apps have activity peaks occurring at similar frequencies. Therefore, our results indicate the existence of temporal regularities in the consumption of mobile services, which are most likely driven by the frequency of the same underlying human routines. As such, they are a first step towards a comprehensive classification of mobile service based on how they are used in time.

The paper is organized as follows. We introduce our measurement data in Section II. The spectral analysis of time series inferred from such data is in Section III. Section IV presents the results of a dedicated clustering of the harmonics returned by the spectral analysis. Finally, conclusions and open research directions are outlined in Section V.

## II. MEASUREMENT DATA

We perform our study with a metropolitan-scale dataset of mobile service demands collected by a major network operator in Europe. Below, we present the data in Section II-A, and detail preliminary denoising steps in Section II-B.

### A. Dataset

Our analysis relies on measurement data collected in the mobile network of Orange, one of the leading operators in Europe. Specifically, the data was recorded during three consecutive months in Fall 2016 by probes monitoring the Gn interfaces of Gateway GPRS Support Nodes (GGSN) as well as the S5 and S8 interfaces of Packet Data Network Gateways (P-GW), so as to capture all traffic flowing through the 3G and 4G infrastructures. The probes perform Deep Packet Inspection (DPI) of the IP traffic on the GPRS Tunneling Protocol user plane (GTP-U), and run proprietary fingerprinting algorithms that allow associating around of 90% of the IP sessions to specific mobile services.

The dataset employed in our study describes mobile service traffic generated in the metropolitan area of Paris, France. This is one of the largest conurbations in Europe, covering an area of over 100 km$^2$ and more than 2 million inhabitants. Orange had a 2016 market penetration of around 34% in the region, hence our data are representative of a large fraction of the local population. In order to bound traffic originated in Paris, IP sessions are geo-references at the antenna sector level, using information available on the GPRS Tunneling Protocol control plane (GTP-C), within the User Location Information (ULI) field of the 3G Packet Data Protocol (PDP) Contexts and 4G Evolved Packet System (EPS) Bearers. The localization accuracy of ULI is known to be limited due to irregular updates to the field; however, the typical precision, in the order of km [9], is sufficient to correctly locate the traffic produced within a large city such as Paris. The data have a temporal granularity of 5 minutes.

We focus our study on 37 mobile services. The rationale for this choice is that traffic volumes generated by mobile apps follow a well-studied power law [24], [26], hence only a very limited set of services yield considerable demands that are worth investigating. Our choice of mobile services includes heterogeneous applications that rank among the top 50 in terms of traffic load, and encompass video and audio streaming (*e.g.*, YouTube and iTunes), social media (*e.g.*, Facebook and Instagram), messaging (*e.g.*, Snapchat and WhatsApp), stores (*e.g.*, Apple Store and Google Play), gaming (*e.g.*, Pokemon Go and Supercell), as well as traffic generated by generic digital activities (*e.g.*, web browsing and electronic mail).

We remark that our study abides by high ethical standards. All data collection, processing, and storage procedures at Orange were carried out in compliance with applicable regulations, including the European Commission General Data Protection Regulation (GDPR). These activities were supervised by the Orange Data Privacy Officer as well as by the Commission Nationale de l'Informatique et des Libertés (CNIL), the French national body ensuring privacy in personal data use. Researchers outside the Orange premises only had access to traffic volumes aggregated at the antenna level over hundreds of users, which do not qualify as personal data and do not entail privacy risks.

### B. Preprocessing

Our analysis focuses on understanding *typical* patterns of mobile service demands. The raw time series of app traffic recorded over three months definitely capture such patterns, yet they also feature fast-varying noise (due to the inherent randomness of user access), and long-timescale trends (due to, *e.g.*, holiday periods or diverse daylight intervals). In order to filter out such phenomena, and work with cleaner time series, we preprocess the data, by computing a *median week* traffic demand for each mobile service [27].

Let the demand for a given mobile service $s$ be described by a three-month time series $d^s(t)$, where $t > 0$ has a resolution matching some time unit, *e.g.*, minutes. The median week is an ordered vector $\boldsymbol{w}^s = (w_n^s)$, $n \in [0, N-1 = \lfloor W/T \rfloor) \subset \mathbb{Z}$, where $W$ is the duration of one week in the considered time unit, *e.g.*, 10,080 in minutes, and $T$ is the duration of one time slot in the same unit. In our analysis we consider five-minute time slots, hence $T = 5$. Each element of $\boldsymbol{w}^s$ is expressed as:

$$w_n^s = \mu_{1/2} \left\{ \sum_{t=\tau}^{\tau+T-1} d^s(t) \; \middle| \; \frac{\tau \bmod W}{T} = n \right\}, \quad (1)$$

where the operator $\mu_{1/2}$ denotes the median of the value set in the argument. Equation (1) divides each week into slots spanning $T$, and computes the overall traffic observed for service $s$ within each such slot; then, slot $n$ of the median week $\boldsymbol{w}^s$ is assigned a single value $w_n^s$ equal to the median of (sum) values associated with the $n$-th slot of each week.

The preprocessing above has the effect of eliminating noise and seasonal effects and returning a more accurate representation of the ordinary demand generated by each mobile service. Note that we use the median instead of other statistical measures since it is more robust to strong outliers that risk for instance to bias averages. Figure 1 shows an example of a three-month time series compressed into a median week representation.

## III. SPECTRAL ANALYSIS

We employ Fourier decomposition on the time series presented before. We remark that spectral methods have already been applied to mobile data traffic time series [1]; however, they were only used for data denoising and not for interpretation, and considering aggregate traffic instead of service-level demands. Our approach allows deriving the frequency spectra of mobile service demands, as outlined in Section III-A, and discuss how they decompose into harmonics of different importance, as set forth in Section III-B.

### A. Service demand spectra

Fourier decomposition allows approximating complex time series as sums of simple trigonometric functions. Specifically, since our median week representations describe a discrete-time process over a finite interval, we apply a Discrete Fourier Transform (DFT). Given the signal $\boldsymbol{w}^s$ for service $s$, its DFT is a complex-valued function of discrete frequency, $\boldsymbol{X}^s = (X_k^s)$, $k \in [0, N) \subset \mathbb{Z}$. The granularity of the DFT in the frequency space is the reciprocal of the duration of the time signal, hence $1/(NT)$ in our case. Then, the $k$-th component of the DFT describes the value of the function at frequency $k/(NT)$, as:

$$X_k^s = \sum_{n=0}^{N-1} w_n^s \exp\left(-i\frac{2\pi}{NT}kn\right). \quad (2)$$
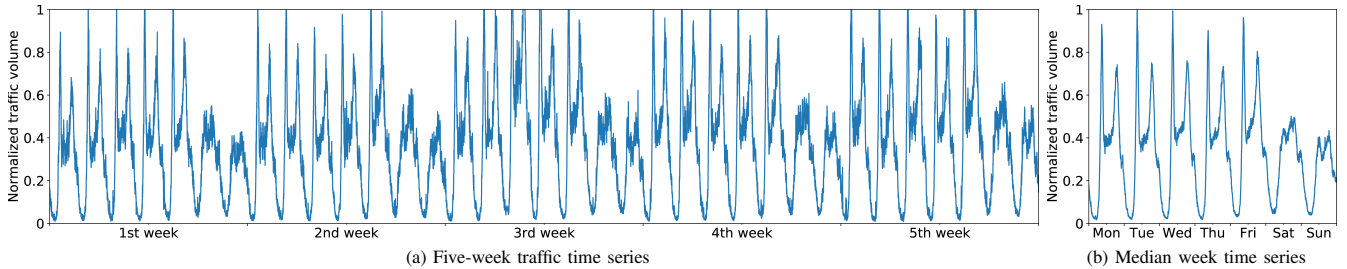
Fig. 1. Traffic time series generated by YouTube during five consecutive weeks in the target metropolitan area (left). Corresponding median week (right).

The DFT is a linear and invertible operation, and the inverse DFT allows reconstructing the time series from its DFT. Formally, the value of the median week demand for service $s$ at each time slot $n$ can be derived from $\boldsymbol{X}^s$ as:

$$w_n^s = \frac{1}{N} \sum_{k=0}^{N-1} X_k^s \exp\left(i\frac{2\pi}{NT}kn\right). \qquad (3)$$

Intuitively, upon inversion, each DFT component translates to a sinusoidal function of time with frequency $k/NT$, and amplitude and phase described by the phasor $X_k^s$. The time series of $s$ is expressed in Equation (3) as the sum of such sinusoidal functions.

We compute the DFT of all mobile service demands, so as to reveal their underlying frequency components. The DFT returns $N$ components for any given service $s$, where $N$ is typically large; however, these components have very different $X_k^s$ values. As mentioned above, these values embed the amplitude of the inverted sinusoidal function for component $k$, and only components with sufficiently high amplitude contribute in a significant manner to the original signal. In order to understand the importance of each component for mobile service demand, we resort to the DFT power spectrum, which is computed as $|\boldsymbol{X}^s|^2$ and describes the distribution of power across frequency components.

Figure 2 shows a representative subset of the power spectra for a selection of services. The spectra are centered at zero frequency, and the central value $|X_0^s|^2$ is the power[1] associated with the constant mean of the time series $\boldsymbol{w}^s$. As one moves away from the central value, the spectra portray the power of increasingly higher frequencies, up to $(N-1)/NT \sim 1/T$ for large $N$. All plots in Figure 2 have very similar shapes that highlight how the power spectra of median week traffic demands are dominated by low-frequency components for all services: as highlighted by the logarithmic ordinate, central frequencies have much higher $|X_k^s|^2$ values. This implies that the traffic demand for any service is strongly characterized by regular patterns with periods in the order of hours and days.

Faster dynamics with periods in the order of tens of minutes or less have a much reduced power. However, we also remark that the spectra flatten around the central peak, indicating that the original signals are also fairly noisy, and all high-frequency components are needed to perfectly reconstruct the time series.

### B. Component filtering

In order to provide better insights on the most important components, Figure 3 offers a detailed view of the central

[1]Note that power values in all figures are normalized with respect to the total signal power, so as to make them more easily understood.

| Services | # Components | Retained Power |
|---|---|---|
| WhatsApp | 3 | 99.34 % |
| MMS | 4 | 99.51 % |
| iCloud | 5 | 98.87 % |
| Youtube | 5 | 99.37 % |
| Generic messaging | 5 | 99.42 % |
| Instagram | 5 | 99.49 % |
| Instagram video | 5 | 99.57 % |
| News | 6 | 98.79 % |
| Generic video | 6 | 99.32 % |
| Facebook | 6 | 99.42 % |
| Google Services | 6 | 99.42 % |
| Ads | 6 | 99.43 % |
| DailyMotion video | 7 | 98.50 % |
| E-commerce | 7 | 98.74 % |
| iTunes | 7 | 99.26 % |
| Facebook video | 7 | 99.38 % |
| Generic web | 7 | 99.51 % |
| NetFlix | 8 | 98.97 % |
| Encrypted web | 8 | 99.38 % |
| Twitter | 8 | 99.40 % |
| Apple Store | 8 | 99.43 % |
| VoIP | 9 | 95.25 % |
| Google Drive | 9 | 98.98 % |
| Generic cloud | 9 | 99.29 % |
| Google Play | 9 | 99.39 % |
| Snapchat | 9 | 99.44 % |
| Supercell | 9 | 99.47 % |
| Generic gaming | 9 | 96.63 % |
| Gameloft | 10 | 85.15% |
| Mail | 10 | 99.37% |
| Adult | 11 | 98.67% |
| P2P | 15 | 96.31% |
| Gaming platforms | 17 | 88.83 % |
| Audio streaming | 17 | 98.72 % |
| King | 17 | 98.85 % |
| Updates | 18 | 96.72 % |
| Pokemon Go | 19 | 99.09 % |
| **Total number of components** | 326 | |
| **Average retained power** | | 98.24 % |

TABLE I
MINIMUM NUMBER OF COMPONENTS RETAINING AT LEAST 99% OF THE TOTAL SIGNAL POWER, FOR ALL MOBILE SERVICES. THE ACTUAL POWER IS INDICATED IN THE RIGHTMOST COLUMN, WHILE TOTAL FIGURES ARE REPORTED AT THE BOTTOM OF THE TABLE.

frequencies of the spectra in Figure 2, for $k \in [0, 50]$. This is equivalent to looking at periodicities $NT/k$ of roughly 4 hours or longer, as marked on the abscissa. These plots make it clear that diversity exists even among central components: for each service, specific frequencies have peaking $|X_k^s|^2$ values, i.e., are especially critical to the temporal dynamics of the demand.
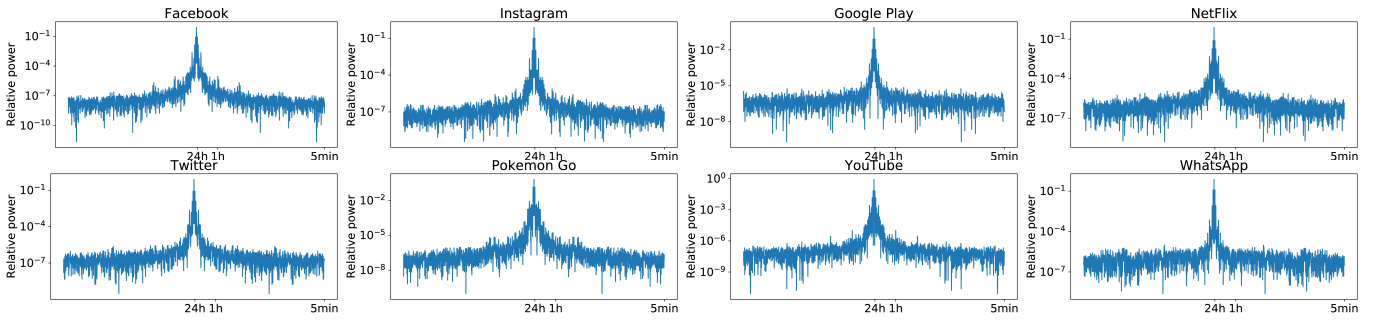
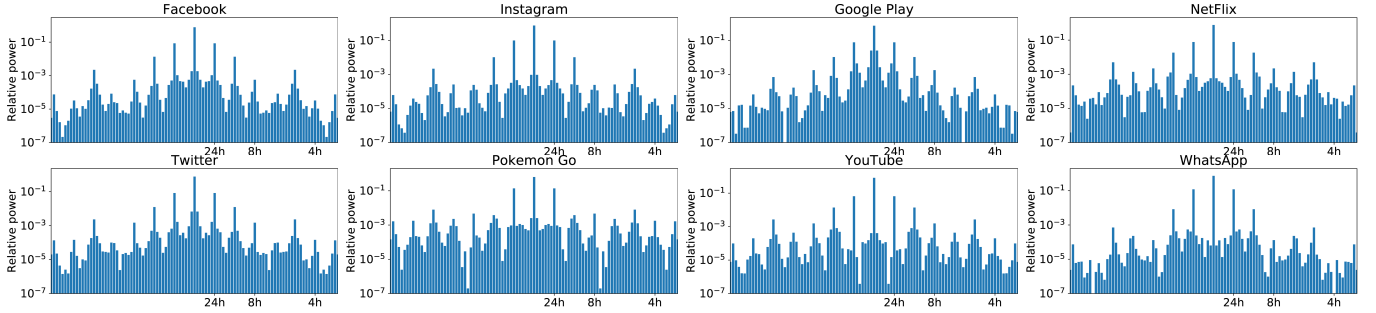Fig. 2. Power spectra of the DFT of selected mobile services.



Fig. 3. Low-frequency components ($k \in [0, 50]$) of the power spectra of the DFT of the selected mobile services in Figure 2.
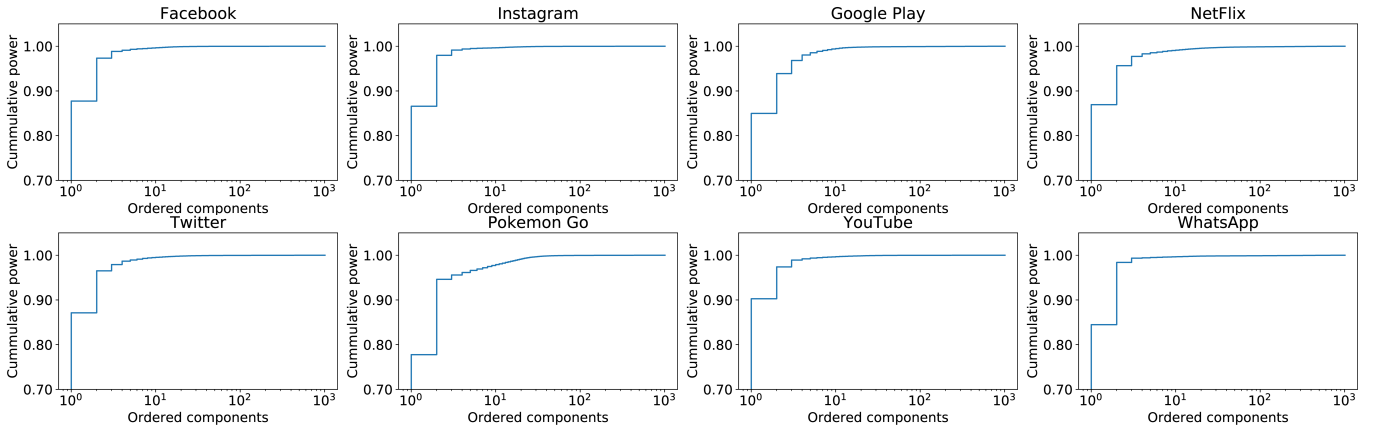


Fig. 4. The cumulative sum of the power associated to ranked components, for the selected mobile services in Figure 2.

| Component number | Power | Phase (degrees) | Period |
|:---:|:---:|:---:|:---:|
| 1 | 7.13% | 126.81 | 24h |
| 2 | 1.51% | -132.97 | 12h |
| 3 | 0.29% | 43.95 | 4.8h |
| 4 | 0.17% | -85.98 | 8h |

TABLE II
EXAMPLE OF RETAINED COMPONENTS FOR THE YOUTUBE SERVICE.

Given that components have heterogeneous power, and the vast majority only marginally contributes to the original signal, it makes sense to limit our analysis to a subset of relevant components for each service. To this end, we retain the minimum number of components whose summed $|X_k^s|^2$ values preserve at least 99% of the total signal power, excluding all components whose contribution is below 0.1%. This is equivalent to ranking components for each service based on their associated power, in descending order, and then computing the cumulative sum of such power following the ranking. Figure 4 illustrates the result for the services in Figure 2. The component ranking first is invariably that associated with the mean of the time series and accounts for the vast majority of the signal power. The most important

observation for our study is, however, that the number of additional components needed to attain the 99% threshold is small in the vast majority of cases.

An in-depth view of the number of components needed to reach the target cumulative power is provided in Table I, for all considered mobile services. The figures confirm that less than 10 components are sufficient to attain the 99% threshold in almost all cases, and the traffic demands of many popular applications such as YouTube, Instagram, or Facebook are mostly influenced by recurring patterns at five or six different frequencies. In light of these results, in the remainder of our analysis, we focus on the high-power frequencies listed in the table only and discard all other noisy components. Also, we do not consider the zero-frequency components in subsequent discussions, since the mean is a service-specific constant value that only captures the volume of traffic associated to each service, and is irrelevant to the temporal dynamics we are interested in. Ultimately, we retain 326 components across all service traffic demands for further analysis.

A detailed view of the retained components is provided in Table II for the specific case of the YouTube median
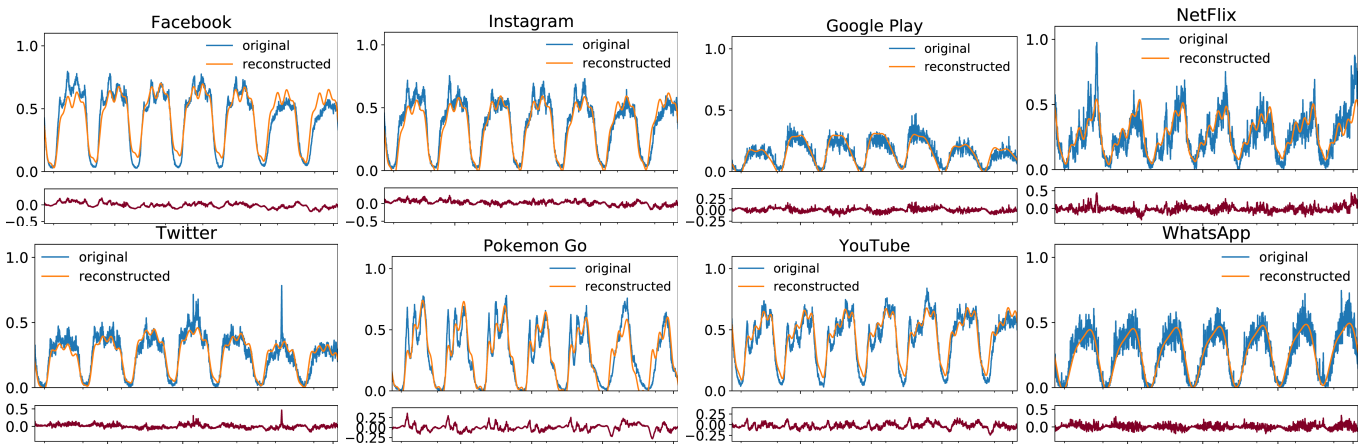
Fig. 5. Reconstructed traffic demands via the inverse DFT on retained components in Table I, compared with the original time series (top), and residual traffic associated to discarded components (bottom), for the selected mobile services in Figure 2.

week demand. The consumption of YouTube follows four main periodicities in time, namely every day, half-day, 8 and approximately 5 hours. The daily pattern, which is in fact determined by the circadian rhythm of human activities, has a clearly higher impact than the other dynamics in this case.

Figure 5 offers an intuitive illustration of the quality of the component filtering process above. Each plot refers to one representative service, and reports: (*i*) the original median week demand and the inverse DFT computed using only the components in Table I; and, (*ii*) the residual traffic that is not captured by the inverse DFT. In all cases, the inverse DFT allows reproducing the main temporal fluctuations of the original demand, and residuals are limited to low-volume noise. Remarkably, such a good approximation of the traffic time series is obtained with just a few components per service, as detailed in Table I.

## IV. COMPONENT ANALYSIS

By looking at the power spectra in Figure 3, it is apparent that many peaks are common to different services. This hints at the fact that the time dynamics of the demands for diverse services may yield patterns that recur at similar periodicity. Next, we design a systematic approach to explore this phenomenon, by developing a simple but efficient clustering algorithm for DFT components in Section IV-A, and applying it to our reference service demands in Section IV-B.

### A. Clustering components

As explained in Section III-A, each component is characterized by a frequency, an amplitude, and a phase. We implicitly used the amplitude to filter relevant components above, since the power $|X_k^s|^2$ associated to each component is proportional to the square of its amplitude. However, the amplitude is not relevant to the clustering problem; indeed, the amplitude is a measure of the magnitude of a given repetitive pattern, while our objective is to identify similar temporal periodicities across service demands, independently of their magnitude. In other words, if two services feature regular traffic peaks at, *e.g.*, noon every day, we would like to cluster together the components responsible for the peaks, no matter whether they have dissimilar amplitudes because the two services generate different traffic volumes.

Therefore, our clustering algorithm considers only the frequency and phase of each component. The two attributes

are in fact processed in two separate steps: in a first step, we group components that have identical frequencies; in the second step, components in the same group are further clustered based on their phase. The rationale for this design is that even slightly different frequencies lead to an increasing misalignment of the components in time, no matter what their phases are: during one week, misaligned components can determine peaks at very different times, which should not be assimilated in our analysis. As a result, we do not want to cluster components with non-identical frequency, even if they have phases that perfectly match. The constraint on equal frequency makes a clustering based on the joint frequency and phase inappropriate, and let us favor a simpler two-step approach instead.

In the first step, we cluster components on their frequency. For the reasons explained above, we require that only identical frequencies are grouped together. Therefore, the clustering operation is straightforward, and we simply gather components based on frequency identity.

The second step focuses on components within each frequency category. In this case, phases that are close but not perfectly matching may be clustered together, since recurring patterns with the same periodicity and small constant shifts in time capture semantically equivalent activity peaks during the whole observation time. Also, the distance measure for phases should be maximum in opposition of phase (*i.e.*, when the value difference is $K\pi$, $K \in \mathbb{Z}$), and null for phases that are $K2\pi$ apart, $K \in \mathbb{Z}$. To fulfill these specifications, we first map phases to a Cartesian plan; let us denote by $\phi_k^s$ the phase of the $k$-th DFT component for the demand of service $s$, then the Cartesian coordinates are

$$
\begin{aligned}
x &= \cos \phi_k^s \\
y &= \sin \phi_k^s.
\end{aligned}
\tag{4}
$$

The transformation above places components along a circle of unity radius, at an angle that it is proportional to their phase.

We then run DBSCAN, a well-known density-based clustering algorithm [28], on the bi-dimensional points that represent the components. We parametrize the algorithm so that at least 3 points shall be grouped to form a cluster, and the maximum distance between the two closest points in the same cluster is $\epsilon = 0.1$, which maps to a phase difference of roughly 5°.

| Cluster period | 24h | 12h | 1w | 4.8h | 1w | 28h | 84h | 8h | 21h | 21h | 8h | 84h | 8 | 84h | 33.6h | 42h | none |
| Cluster phase | 115° | 50° | -137° | 62° | -162° | -67° | -87° | 17° | 8° | 21° | -31° | -71° | -55° | -115° | -107° | -150° | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Ads** | 84.27 | 9.18 | 3.40 | 2.29 | 0 | 0.86 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Adult** | 19.86 | 0 | 12.02 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2.51 | 0 | 0 | 0 | 0 | 65.61 |
| **Apple Store** | 61.40 | 11.62 | 0 | 1.01 | 22 | 1.76 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2.21 |
| **Audio streaming** | 54.60 | 0 | 0 | 4.87 | 2.84 | 1.48 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.32 | 1.07 | 0 | 33.81 |
| **DailyMotion video** | 62.20 | 1.34 | 22.71 | | 4.01 | 0 | 0 | 4.06 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.34 |
| **E-commerce** | 65.32 | 25.49 | 2.36 | 2.95 | 0 | 0 | 1.07 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2.82 |
| **Encrypted web** | 75.97 | 9.55 | 8.29 | 2.35 | 0 | 1.91 | 0 | 0 | 1.06 | 0 | 0 | 0 | 0.88 | 0 | 0 | 0 | 0 |
| **Facebook** | 82.10 | 12.99 | 1.76 | 2.13 | 0 | 1.01 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Facebook video** | 79.40 | 13.26 | 0 | 2.68 | 2.46 | 0 | 1.20 | 0 | 0 | 0 | 0.99 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Gameloft** | 88.69 | 2.02 | 0 | 1.11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 8.18 |
| **Gaming platforms** | 80.70 | 4.58 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14.72 |
| **Generic cloud** | 76.37 | 9.65 | 7.47 | 0 | 0 | 1.50 | 1.74 | 0 | 0 | 1 | 0 | 0 | 1.34 | 0 | 0.94 | 0 | 0 |
| **Generic gaming** | 86.76 | 3.51 | 0 | 1.51 | 3.56 | 1.60 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.90 | 0 | 0.86 | 1.31 |
| **Generic messaging** | 96.14 | 2.27 | 0 | 0 | 0.97 | 0.61 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Generic video** | 66.62 | 22.25 | 4.34 | 3.19 | 0 | 0 | 0 | 3.60 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Generic web** | 82.97 | 8.47 | 4.78 | 1.92 | 0 | 1.03 | 0 | 0 | 0 | 0 | 0.84 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Google Drive** | 80.31 | 10.13 | 0 | 0 | 2.61 | 1.81 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.08 | 1.24 | 0 | 2.84 |
| **Google Play** | 61.78 | 8.45 | 0 | 0 | 20.38 | 3.58 | 0 | 0 | 1.23 | 0 | 0 | 0 | 1.43 | 0 | 0 | 0 | 3.14 |
| **Google+** | 93.27 | 4.06 | 0 | 0.83 | 0 | 0.81 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.04 |
| **Instagram** | 88.14 | 9.10 | 0.84 | 1.93 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Instagram video** | 82.16 | 14.71 | 0 | 1.56 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.57 |
| **King** | 57.82 | 0 | 9.12 | 3.54 | 0 | 0 | 8.20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 21.31 |
| **MMS** | 95.07 | 3.30 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.63 |
| **Mail** | 66.10 | 1.33 | 15.60 | 0 | 0 | 5.06 | 3.81 | 0 | 3.30 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4.79 |
| **NetFlix** | 0 | 17.30 | 0 | 4.71 | 0 | 0 | 0 | 2.06 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 75.92 |
| **News** | 73.67 | 13.62 | 6.59 | 4.76 | 0 | 1.36 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **P2P** | 36.34 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.28 | 0 | 0 | 0 | 0 | 62.38 |
| **Pokemon Go** | 78.98 | 0 | 1.46 | 4.54 | 0 | 0.51 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14.51 |
| **Snapchat** | 88.67 | 3.58 | 0 | 0 | 0 | 1.81 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.94 | 5 |
| **Supercell** | 81.84 | 3.57 | 6.26 | 0 | 0 | 2.15 | 2.34 | 0 | 0 | 1.53 | 0 | 0 | 0 | 0 | 0 | 0 | 2.30 |
| **Twitter** | 76.53 | 11.42 | 6.19 | 2.12 | 0 | 1.52 | 0 | 0 | 0 | 1.37 | 0.84 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Updates** | 50.78 | 10.59 | 16.85 | 0 | 0 | 2.84 | 0 | 0 | 0 | 2.89 | 0 | 0 | 0 | 0 | 0 | 0.66 | 15.40 |
| **VoIP** | 78.67 | 9.52 | 5.92 | 0 | 0 | 1.40 | 1.51 | 0 | 0.69 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2.28 |
| **WhatsApp** | 93.58 | 6.42 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **YouTube** | 78.32 | 16.62 | 0 | 3.20 | 0 | 0 | 0 | 0 | 0 | 0 | 1.86 | 0 | 0 | 0 | 0 | 0 | 0 |
| **iCloud** | 88.69 | 7.57 | 0 | 0 | 2.05 | 1.68 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **iTunes** | 0 | 31.71 | 3.12 | 3.52 | 0 | 0 | 0 | 7.10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 54.55 |
| **Total components** | 35 | 32 | 18 | 22 | 9 | 21 | 7 | 4 | 4 | 3 | 4 | 3 | 3 | 3 | 3 | 3 | 115 |
| **Percent power** | 70.69 | 9.23 | 3.14 | 1.68 | 1.64 | 0.98 | 0.53 | 0.45 | 0.17 | 0.14 | 0.13 | 0.12 | 0.09 | 0.08 | 0.08 | 0.06 | 10.79 |

TABLE III
OVERVIEW OF THE 16 CLUSTERS GROUPING THE 326 RETAINED SERVICE DEMAND COMPONENTS.
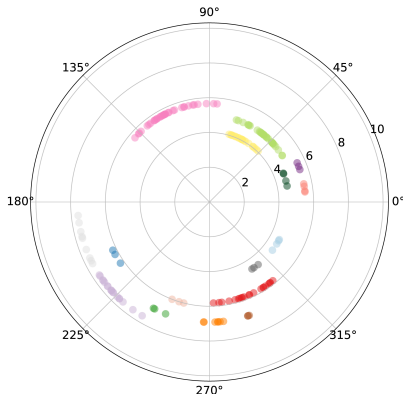


Fig. 6. Clusters of the 326 components from the demands for the 37 service considered in our study. Figure best viewed in colors.

### B. Commonalities and outliers in mobile service demands

Figure 6 illustrates the 16 clusters obtained with the two-step algorithm above. In order to show the quality of the result jointly for the frequency and phase attributes, we map the frequency to the distance from the origin, and the phase to the angle as per (4). Colors denote points, *i.e.*, components, labelled in the same cluster. This representation outlines clear groups of points, which are well identified by the algorithm, which thus assimilates components with the same frequency (*i.e.*, along the same circle) and close phases as desired.

Figure 7 provides a complementary view of the same clustering result. The four plots represent four different clusters, whose period (inverse of frequency) and phase (in °) are indicated below each image. In each plot, every component belonging to the cluster is represented as a sinusoidal function of time (in gold). We can observe that the sinusoids in the same cluster are very similar, hence they correspond to equivalent temporal patterns of activity peaks. The components differ in terms of amplitude, but, as previously mentioned, this is due to the heterogeneous popularity and traffic volume associated with each service. What is relevant to our analysis is the agreement in frequency and phase, which is confirmed

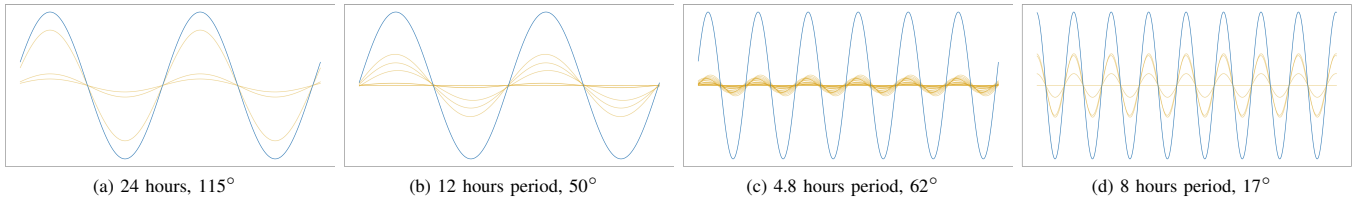| (a) 24 hours, 115° | (b) 12 hours period, 50° | (c) 4.8 hours period, 62° | (d) 8 hours period, 17° |

Fig. 7. Components in four different clusters, portrayed as sinusoidal functions of time. Figure best viewed in colors.

by the regular pattern of the sum of components (in blue).

A comprehensive summary of the 16 clusters is provided in Table III. Columns in the table correspond to clusters, identified by their frequency and phase indicated in the first row. Subsequent rows refer to each of the 37 services we consider in our analysis. The number in each element $(x, y)$ indicates the fraction of power associated with the component of service $x$ that is sorted in cluster $y$. The fractional power is calculated with respect to all service components retained for analysis, according to Table I; a value zero in element $(x, y)$ thus indicates that the demand for service $x$ does not have any relevant component with the frequency and phase associated to cluster $y$. The last column refers to outlier components that could not be associated to any cluster. The last two rows list the number of components in each cluster and the percent power of the whole cluster with respect to all clusters. Table III lets us provide the following insights.

*I.* Almost all (33 out 37) of services have a largely dominant component with a 24-hour periodicity. It is easy to map such a component to the circadian rhythm of human activities, which alternates low traffic overnight and high demand during the day.

*II.* Most (32 out of 37) services also show the same significant dynamic at a 12-hour periodicity. Many (22 out of 37) also share components that highlight regular patterns at every one week, 4.8 hours. An investigation of the causes for these sub-daily patterns is out of the scope here, and an object for future research; yet, we speculate that commuting affects the demands for many services and may be behind these dynamics.

*III.* Common regular behaviors are present also at periods longer than one day for many (18-21 out of 37) services. One week and 28 hours are the most relevant periods, and we consider that those are linked with different dynamics occurring during weekends.

*IV.* Several services tend to defy classification, and ($i$) have no or much less relevant components in the 24-hour cluster, ($ii$) have an unusually high weight associated with specific clusters, and/or ($iii$) have a high incidence of components that are not included in any cluster (*i.e.*, are outliers). Services in this category include games (King, Pokemon Go, generic Gaming platforms), audio streaming services (iTunes, generic Audio streaming), NetFlix, peer-to-peer, and adult web traffic. These are fairly specific categories of mobile applications, with different but reasonable reasons for their diversity. For instance, NetFlix is a fairly unique service providing long-lived video streams to niche mobile users. Audio streaming applications are the sole that do not need visual attention by the user. Or, adult web traffic is characterized by unique patterns due to its socially inconvenient nature.

## V. CONCLUSIONS

This paper provides an original perspective on the temporal analysis of mobile service traffic. By leveraging spectral methods, we could identify common periodic behaviors in the real-world traffic generated by a large set of applications, which were not detected by previous studies. The preliminary results we presented pave the road for further investigations, aimed at explaining the root causes for these temporal similarities, at assessing their generality at different spatial and geographical scales, and at exploiting them for applications in network planning and resource management.

## REFERENCES

[1] B. Cici, M. Gjoka, A. Markopoulou, and C. T. Butts, "On the Decomposition of Cell Phone Activity Patterns and Their Connection with Urban Ecology," in *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, ser. MobiHoc '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 317–326.

[2] E. Peltonen, E. Lagerspetz, J. Hamberg, A. Mehrotra, M. Musolesi, P. Nurmi, and S. Tarkoma, "The Hidden Image of Mobile Apps: Geographic, Demographic, and Cultural Factors in Mobile Usage," in *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, ser. MobileHCI '18. New York, NY, USA: Association for Computing Machinery, 2018.

[3] J. Alvarez-Lozano, V. Osmani, O. Mayora, M. Frost, J. Bardram, M. Faurholt-Jepsen, and L. V. Kessing, "Tell Me Your Apps and I Will Tell You Your Mood: Correlation of Apps Usage with Bipolar Disorder State," in *Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments*, ser. PETRA '14. New York, NY, USA: Association for Computing Machinery, 2014.

[4] C. Marquez, M. Gramaglia, M. Fiore, A. Banchs, and X. Costa-Perez, "How Should I Slice My Network? A Multi-Service Empirical Evaluation of Resource Sharing Efficiency," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 191–206.

[5] L. Ravindranath, S. Agarwal, J. Padhye, and C. Riederer, "Procrastinator: Pacing Mobile Apps' Usage of the Network," in *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 232–244.

[6] D. Elias, F. Nadler, J. Stehno, J. Krösche, and M. Lindorfer, "Somobil – improving public transport planning through mobile phone data analysis," *Transportation Research Procedia*, vol. 14, pp. 4478 – 4485, 2016, transport Research Arena TRA2016.

[7] D. Kim, J. Gluck, M. Hall, and Y. Agarwal, "Real World Longitudinal iOS App Usage Study at Scale," 2019.

[8] L. Qi, Y. Qiao, F. B. Abdesslem, Z. Ma, and J. Yang, "Oscillation Resolution for Massive Cell Phone Traffic Data," in *Proceedings of the First Workshop on Mobile Data*, ser. MobiData '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 25–30.

[9] Q. Xu, A. Gerber, Z. M. Mao, and J. Pang, "AccuLoc: Practical Localization of Performance Measurements in 3G Networks," in *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '11.   New York, NY, USA: Association for Computing Machinery, 2011, p. 183–196.

[10] C. Williamson, E. Halepovic, Hongxia Sun, and Yujing Wu, "Characterization of CDMA2000 cellular data network traffic," in *The IEEE Conference on Local Computer Networks 30th Anniversary (LCN'05)l*, Nov 2005, pp. Z000–719.

[11] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das, "Understanding traffic dynamics in cellular data networks," in *2011 Proceedings IEEE INFOCOM*, April 2011, pp. 882–890.

[12] F. Xu, Y. Li, H. Wang, P. Zhang, and D. Jin, "Understanding Mobile Traffic Patterns of Large Scale Cellular Towers in Urban Environment," *IEEE/ACM Transactions on Networking*, vol. 25, no. 2, April 2017.

[13] I. Trestian, S. Ranjan, A. Kuzmanovic, and A. Nucci, "Measuring Serendipity: Connecting People, Locations and Interests in a Mobile 3G Network," in *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC '09.  New York, NY, USA: Association for Computing Machinery, 2009, p. 267–279.

[14] Q. Xu, J. Erman, A. Gerber, Z. Mao, J. Pang, and S. Venkataraman, "Identifying Diverse Usage Behaviors of Smartphone Apps," in *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, ser. IMC '11.   New York, NY, USA: Association for Computing Machinery, 2011, p. 329–344.

[15] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, and J. Wang, "Characterizing geospatial dynamics of application usage in a 3G cellular data network," in *2012 Proceedings IEEE INFOCOM*, March 2012, pp. 1341–1349.

[16] R. Keralapura, A. Nucci, Z.-L. Zhang, and L. Gao, "Profiling Users in a 3G Network Using Hourglass Co-Clustering," in *Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '10.   New York, NY, USA: Association for Computing Machinery, 2010, p. 341–352.

[17] H. Li, X. Lu, X. Liu, T. Xie, K. Bian, F. X. Lin, Q. Mei, and F. Feng, "Characterizing Smartphone Usage Patterns from Millions of Android Users," in *Proceedings of the 2015 Internet Measurement Conference*, ser. IMC '15.   New York, NY, USA: Association for Computing Machinery, 2015, p. 459–472.

[18] P. Fiadino, M. Schiavone, and P. Casas, "Vivisecting Whatsapp through Large-Scale Measurements in Mobile Networks," in *Proceedings of the 2014 ACM Conference on SIGCOMM*, ser. SIGCOMM '14.   New York, NY, USA: Association for Computing Machinery, 2014, p. 133–134.

[19] Q. Deng, Z. Li, Q. Wu, C. Xu, and G. Xie, "An empirical study of the WeChat mobile instant messaging service," in *2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, May 2017, pp. 390–395.

[20] J. Erman, A. Gerber, K. K. Ramadrishnan, S. Sen, and O. Spatscheck, "Over the Top Video: The Gorilla in Cellular Networks," in *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, ser. IMC '11.   New York, NY, USA: Association for Computing Machinery, 2011, p. 127–136.

[21] Z. Li, X. Wang, N. Huang, M. A. Kaafar, Z. Li, J. Zhou, G. Xie, and P. Steenkiste, "An Empirical Analysis of a Large-Scale Mobile Cloud Storage Service," in *Proceedings of the 2016 Internet Measurement Conference*, ser. IMC '16.   New York, NY, USA: Association for Computing Machinery, 2016, p. 287–301.

[22] Y. Zhang and A. undefinedrvidsson, "Understanding the Characteristics of Cellular Data Traffic," in *Proceedings of the 2012 ACM SIGCOMM Workshop on Cellular Networks: Operations, Challenges, and Future Design*, ser. CellNet '12.   New York, NY, USA: Association for Computing Machinery, 2012, p. 13–18.

[23] H. Wang, Y. Li, S. Zeng, G. Wang, P. Zhang, P. Hui, and D. Jin, "Modeling Spatio-Temporal App Usage for a Large User Population," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 3, no. 1, Mar. 2019.

[24] C. Marquez, M. Gramaglia, M. Fiore, A. Banchs, C. Ziemlicki, and Z. Smoreda, "Not All Apps Are Created Equal: Analysis of Spatiotemporal Heterogeneity in Nationwide Mobile Service Usage," in *Proceedings of the 13th International Conference on Emerging Networking EXperiments and Technologies*, ser. CoNEXT '17.   New York, NY, USA: Association for Computing Machinery, 2017, p. 180–186.

[25] R. Singh, M. Fiore, M. Marina, A. Tarable, and A. Nordio, "Urban Vibes and Rural Charms: Analysis of Geographic Diversity in Mobile Service Usage at National Scale," in *The World Wide Web Conference*, ser. WWW '19.   New York, NY, USA: Association for Computing Machinery, 2019, p. 1724–1734.

[26] M. Z. Shafiq, L. Ji, A. X. Liu, and J. Wang, "Characterizing and Modeling Internet Traffic Dynamics of Cellular Devices," in *Proceedings of the ACM SIGMETRICS Joint International Conference on Measurement and Modeling of Computer Systems*, ser. SIGMETRICS '11.   New York, NY, USA: Association for Computing Machinery, 2011, p. 305–316.

[27] A. Furno, M. Fiore, R. Stanica, C. Ziemlicki, and Z. Smoreda, "A Tale of Ten Cities: Characterizing Signatures of Mobile Traffic in Urban Areas," *IEEE Transactions on Mobile Computing*, vol. 16, no. 10, Oct 2017.

[28] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, ser. KDD'96.   AAAI Press, 1996.