



# Exploring Smartphone Application Usage Logs with Declared Sociological Information

Vassili Rivron, Mohammad Irfan Khan, Simon Charneau, Isabelle Chrisment

## ► To cite this version:

Vassili Rivron, Mohammad Irfan Khan, Simon Charneau, Isabelle Chrisment. Exploring Smartphone Application Usage Logs with Declared Sociological Information. SocialCom 2016 - 9th IEEE International Conference on Social Computing and Networking, IEEE, Oct 2016, Atlanta, United States. pp.266-273, 10.1109/BDCLOUD-SocialCom-SustainCom.2016.49 . hal-01378795

**HAL Id: hal-01378795**

**<https://inria.hal.science/hal-01378795>**

Submitted on 1 Dec 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Exploring Smartphone Application Usage Logs with Declared Sociological Information

Vassili Rivron  
INRIA, France.  
vassili.rivron@inria.fr

Mohammad Irfan Khan  
INRIA, France.  
mohammad-irfan.khan@inria.fr

Simon Charneau  
INRIA, France.  
simon.charneau@inria.fr

Isabelle Chrisment  
Université de Lorraine, LORIA  
isabelle.chrisment@loria.fr

**Abstract**—In this paper we present an exploratory smartphone usage study with logs collected from users in the wild, combined with the sociodemographic, technological and cultural information provided by them. We observe a high diversity among users' most used applications, but by classifying applications into services we find significant correlations between service usage and socio-demographic profile. We discuss that sociological information has rich potential in characterizing smartphone usage and can be applied to interesting incentive strategies and use cases based on users' sociological context.

## I. INTRODUCTION

Since the launching of the touchscreen mobile phones in 2007, the number of smartphone users and the applications and services offered on mobile platforms have grown exponentially, transforming the communication, cultural, media and entertainment practices and, at the same time, the possibilities for scientists to observe them. This growth has mainly led to a rising interest by ISPs, content providers, governments and audience measurement companies to analyze the population's smartphone usage. However, the data collected by those entities are seldom open or made available to the scientific community.

Numerous studies have been conducted within the IT community over the last five years to study people's smartphone usage. These studies mainly look at the smartphone usage pattern, the usage context, the user privacy and security with implications such as enhancing quality of service, reducing power consumption, improving content caching, management and delivery.

However, one particular aspect that has been lacking among these scientific studies is that not enough consideration is given to socio-demographic and lifestyle aspects of users. Even a few interdisciplinary studies focus on usage patterns as behavioral or interaction indicators, but, due to the nature of their datasets and theoretical positions, they rarely explore further social and cultural implications of usage profiling, or reduce it to economic or generational factors. Our approach is not incompatible with uses and gratifications theory [1], but, inspired from sociology of cultural practices [2], it broadens a common "context" conception from actual situation [3], to social and cultural structures that are involved in the construction of the needs (which are not the same for all) and of the conditions of their satisfaction.

In this exploratory paper we analyze the application popularity among a small set of users and show the enormous diversity that exists among them. Due to this diversity it is not feasible to relate the popularity of specific applications with particular socio-demographic profiles. Nevertheless, when we group the applications according to the kind of service or content they provide, we find interesting relations among user profiles and popularity of particular types of application. Moreover, we

discuss how the combination of sociological declared data of users could be used to renew the statistics methodology, to improve application recommendation on application stores or to optimize the device's performance.

For our study we invited the general public to participate and contribute with anonymous smartphone usage logs, by respecting privacy and in the most natural settings (their own device and tariff plan). Besides, we used a questionnaire to accompany the usage logs, to gather anonymous socio-demographic data of the users along with some questions regarding lifestyle and smartphone usage. This experiment was conducted in the context of building a french country-wide open Internet observation platform called *MetroScope*<sup>1</sup>, for the scientific study of fixed and mobile Internet from a neutral perspective. Our main contributions in this paper are:

- 1) We show that most of the usage time is due to a set of few applications, which for most users remains relatively constant over time, illustrating different appropriation of the device in daily routine.
- 2) We show that there is a significant diversity in terms of application popularity, even among demographically similar users, making it infeasible to associate application usage with user profile.
- 3) We group applications into types of services and explore several methodologies to correlate service usage and users' socio-demographic, technological and cultural profile.

## II. RELATED WORK

Similarly to our approach, most past studies about smartphone usage involve either logs collection by deploying passive sensors on users' smartphones [4], [5], [3], [6], or from the ISP side [7] or a combination of both [8]. Some others [5] even gather traces from users' phones by implementing sensors in other applications downloaded from application stores, perhaps without users' informed consent. Only two studies [9], [10] collect usage logs in the wild as we do, but they do not collect profile information about users.

There have been several researches on application usage such as [3], [7]. They consider some common metrics of smartphone usage, such as usage duration, frequency, temporal and diurnal patterns. Some of these studies relate smartphone usage with contextual factors such as usage location and social surrounding [3], geography [7] etc. In a more qualitative approach, a previous exploitation of our own survey that combines usage logs, multiple choices questionnaires and interviews, was done to understand how the smartphone usage both reveals and modifies the social rhythms of student lifestyle [11].

Bohmer et al. [5] and Trinh et al. [12] study smartphone usage patterns by considering applications which are used together.

<sup>1</sup><http://www.metroscope.org>

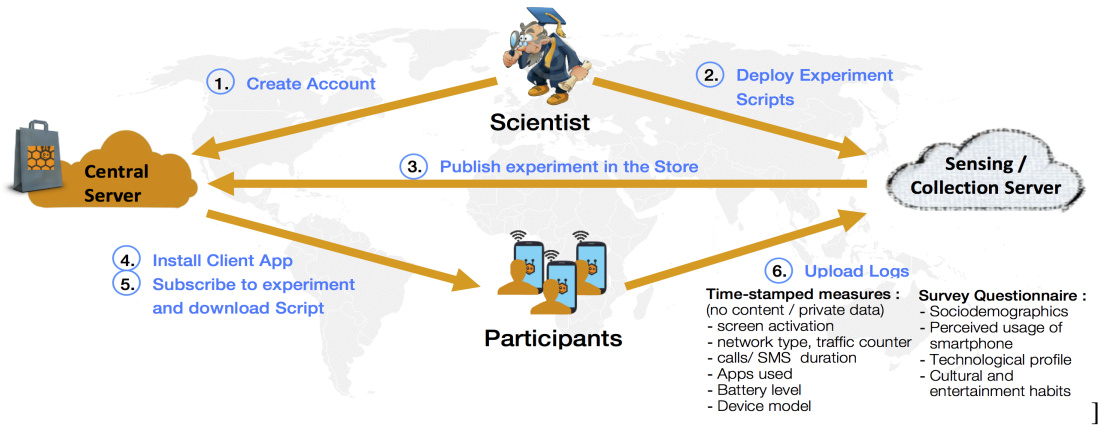


Fig. 1: Data Collection platform

Joon et al. [13] show some basic data on sessions, usage and traffic. They work on smartphone battery optimization as a use case. Stober et al. [14] fingerprint individual smartphones based on background traffic of popular applications. Rahmati et al. [15] study usage behavior but only specific to first time smartphone users. Falaki et al. [4] analyze various aspects of smartphone usage and present usage diversity among users in terms of usage duration, frequency, traffic volume, with some statistical distribution of diverse usage patterns. However none of these studies sufficiently face the huge diversity of application popularity among users.

Machine learning techniques have also been applied to learn individual application usage patterns and to exploit them for user experience improvement, such as faster app launching [16], [17], better app organization [18], app recommender system [19].

Although these studies produce interesting results about various aspects of smartphone usage, none of these thoroughly consider users' socio-demographic profile or find relations between demographic data and smartphone usage. We in our paper address this point and use social science methods to find correlation between users' socio-demographic data and mobile phone usage patterns.

There have been some sociological studies which show the importance of socio-demographic factors in understanding the diverse uses of digital technologies [2]. However, apart from some audience measurement companies (such as Nielsen or Mediamétrie in France), there are very few studies by sociologists which analyze usage traces in combination with demographic data. Some papers combine Internet usage logs on personal computers with demographic data for usage analysis [20], [21]. And, closer to our interests, some other studies [22] have experimented the translation of usage logs (such as phone calls) into sociological indicators such as social capital or employability.

### III. DATA COLLECTION AND DATASET

Our data collection platform called Apisense [23], which as shown in Figure 1, is composed of a client application running on the background (available on Android OS only, at the time of the experiment), a central server and a logs collection or sensing server.

In order to carry out an experiment, a scientist registers on the central server (1) and deploys an experiment's scripts on the sensing server (2) and the experiment appears on the Store of the central server (3). A participant installs the sensing

application (4) and after registration to the experiment, the sensing application downloads the scripts from the sensing server (5). The logs are stored on the phone and periodically uploaded to the collection server via Wi-Fi network (6). A Multiple Choice Questionnaire (MCQ) was also installed on the client application containing questions about socio-demographic, technological and cultural profile of the user, and could be uploaded similarly.

Convincing the crowd to contribute smartphone usage logs is quite difficult due to privacy and performance concerns (although our setup handles these issues). Therefore we organized a contest called PRACTIC [24] and invited the public to participate and win prizes based on their level of contribution and recruitment of other participants.

The campaign was organized between 10 March and 20 April 2014, for six weeks and attracted 260 participants from several cities in France. All 260 participants filled the multiple choices questionnaire while only 97 of them also installed the client application, because of system requirements (Android Only), privacy or incentive reasons. Out of those 97 users, only 42 users produced logs continuously for at least two weeks with complete questionnaire, which make our dataset for this paper. This filtration criteria of logs continuity for a minimum duration of two weeks was important because we study usage patterns in terms of lifestyle routines and cultural habits. Therefore we had to sacrifice on the quantity of users to get homogeneous quality logs, a selection that is common with other crowdsourced studies [25].

The participation duration of these 42 users varied from 17 to 139 days (some users continued even beyond the contest), with an average of 53 days and median 58 days. It represents 3621 hours of smartphone usage logs, from the usage of 929 different applications by the participants. Table I shows the

TABLE I: DataSet Statistics

<b>Gender</b>	66% male, 34% female
<b>Profession</b>	60% students, 40% professionals
<b>Field</b>	68% IT/engineering, 7% humanities/economics, 25% others
<b>Age</b>	59% 17-25 years, 29% 26-35 years, 12% over 36
<b>Android</b>	3% 2.3.X, 6% 4.0.X 60% 4.1-4.3.X, 31% 4.4.X
<b>Brand</b>	37% Samsung , 23% LG, 14% Sony 9% Wiko, 6% Motorola, 11% Others

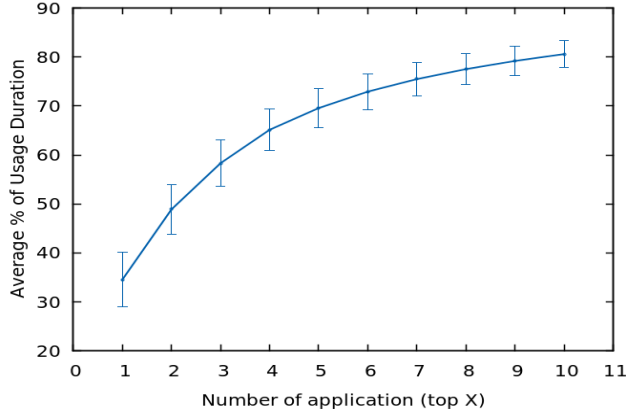


Fig. 2: Application usage distribution per user

statistics about the dataset and reveals the predominant IT scholar composition of our sample (due to the campaign’s dissemination launching channels). This composition leads to a representativity bias that we will discuss in section VI.

#### IV. APPLICATION USAGE DIVERSITY

In this section we analyze how the smartphone is used by users i.e. for what applications and services. We first present the distribution of application popularity among users at an aggregate time scale. This is followed by an analysis at a finer time scale i.e. whether the set of popular applications remains constant or how those vary on a daily or other cyclic basis. Subsequently we compare between users their sets of most popular apps and show the enormous diversity which exists among users.

##### A. Application Usage Distribution per User

We measure smartphone usage with two basic metrics: usage duration and usage frequency. Usage duration is calculated for only the period during which the user is interacting with the phone i.e. application is running on the foreground. Similarly usage frequency is the number of times the application is launched or comes to foreground. For most applications, both metrics are pertinent, while for some applications such as GPS, usage duration is more relevant. Whenever the two metrics tend to provide similar results, we mostly present our findings for duration only. For applications that can also be considered as directly used even when running in background, such as phonecalls or music, we only considered frequency.

In order to analyze application popularity for a user, we calculate how many applications consume what percentage of usage time and usage frequency. Figure 2 shows the distribution of the most popular applications in terms of usage duration. In the graph the x-axis represents the Top applications (1 to 10) resulting the most usage duration and the y-axis shows the percentage of usage time due to those applications, averaged for 42 users with 95% Confidence Interval. The graph shows that the most used application causes about 35% of usage time and only 3 top applications account for 60% of usage time. The next 7 applications among the Top 10 only add 20% of usage time, giving a ratio of 10 applications being used 80% of the time. The application popularity in terms of frequency shows very similar results. Even though at an average, each phone in our dataset contains around 84 applications, either pre-installed or installed by the user, but only less than 10 applications are actually used most of the time, a fact that is also demonstrated by other studies [4]. Thus in this paper, we

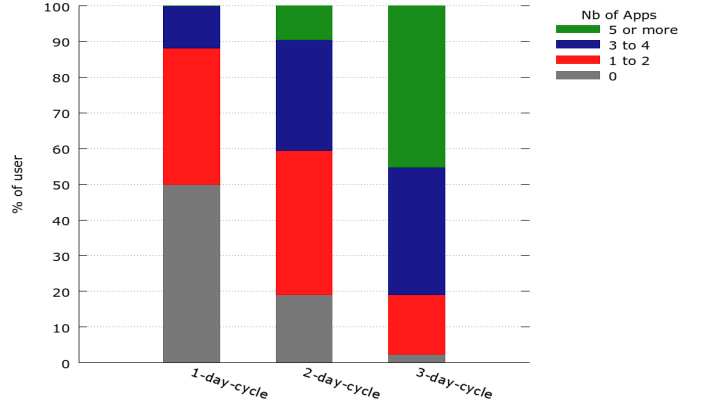


Fig. 3: Persistence of Popular Applications

will focus on the Top 3 and Top 10 most popular apps for each user. Table 2 shows the most common applications among the Top10 lists of the users and the number of users who have the application in his Top10 list.

TABLE II: Most common apps among Top10 list of users

AppName	Nb of User	AppName	Nb of User
Chrome	24	Youtube	10
Facebook	22	2048	9
SMS/MMS	21	Google Dialer	8
Hangout	16	WhatsApp	6
Telephone	15	Candy Crush	5
Gmail	12	Firefox	5
Contacts	11	Twitter	5
Maps	11	Browser	5
Clock	10	Camera	5
Gallery	10	Snapchat	5

##### B. Temporal Regularity of Most Used Apps

The ratio presented above (3 apps causing 60% and 10 apps causing 80% of usage time) is calculated at an aggregate time scale for the whole participation duration of each user. Here we analyze how do the Top10 popular apps vary or remain constant over time for each user.

Figure 3 shows how the popular apps for the users are regularly used or persistent, considering periods of 1 (daily), 2 and 3 days cycles.

**Everyday:** As shown by the leftmost bar, 12% users (in blue on top) have 3 or 4 apps which appear daily on his list of Top10 apps. These users make high usage of 3 to 4 specific applications every single day. However 50% of users (in gray) have no such application which ranks on the Top10 list every consecutive day.

**3 day cycle:** The number of regularly used applications improves drastically when a cycle of 3 days is considered (rightmost bar). We see that more than 40% of users (green portion) have at least 5 such apps which ranks on his Top10 list during every 3 day cycle.

In terms of granularity, a cycle of 3 days is relevant, indicating that a user may not use an application everyday, nevertheless that application is among his most used apps over a 3 day period. This is acceptable and still describes social routines,

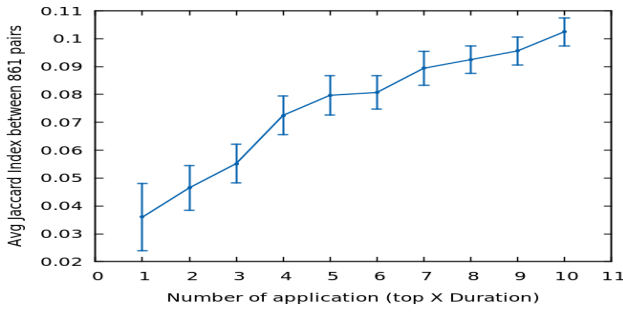


Fig. 4: Average Jaccard Index for comparison of most used applications of each user with others

because app usage patterns vary between weekdays/weekends and also due to temporal lifestyle change as we discussed in another paper [24].

Similarly we also found that the 3 most used apps which rank regularly among the Top10 of every 3 day cycle are also present on the user's list of 10 most used apps for the whole participation duration. Thus it can be concluded that the ratio of 3 apps causing of 60% usage time, holds true not only at an aggregate time scale but also over short time scale.

### C. Diversity of Application Popularity among Users

As each individual in our dataset can be associated to only a handful of applications he uses most of the time, we would like to know whether these applications are similar or diversified between one user and another. Here we compare the set of popular applications of each individual user with 41 other users and analyze the general trend of popular application commonality among the 42 users. We use the Jaccard Index to pairwise compare the set of the most used applications of each user with all other users. The index is calculated by dividing the intersection by the union of the two sets. Its value ranges from 0 (when intersection=0 i.e. maximum diversity), to 1 (when intersection=union i.e. minimum diversity).

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}; 0 \leq J(A, B) \leq 1 \quad (1)$$

In our set of 42 users we have:  $N(N-1)/2$  or 861 pairs and each pair is commutative as  $A \cap B = B \cap A$ . For each set size (Top 1 to Top 10 apps per set) the average of 861 Jaccard Indices is calculated, such as:

$$AvgJaccardIndex = \frac{\sum_{i=1}^{861} J_i(A, B)}{861} \quad (2)$$

Figure 4 shows the average Jaccard Index versus the set size. The Jaccard Indices result from the comparison of the top apps, in terms of duration, among users. When only top 1 application is compared among users, the level of diversity or the Jaccard Index is highest. The average Jaccard Index increases with increased set size as because the larger the set, the higher is the probability of obtaining common items between them. Nevertheless, for set size of 3 (60% usage time) and set size of 10 (80% usage time) the average Jaccard Index is quite close to zero, which is the level of maximum diversity.

Table III details the comparison results when the set size is 3 and 10 (Top3 and Top10 apps respectively). Among their Top 3 app set (middle column), 643 pairs do not have any application

in common. Some pairs have one or two applications in common and not a single pair has 3 applications in common. A similar pattern is visible for the comparison of Top 10 application set and only one pair of users has a maximum of 6 applications in common.

TABLE III: Number of user pairs having commons apps among their Top3 and Top10 most used apps

Nb. Of Common Apps	Nb. Of Pairs in Comparison of Top3	Nb. Of Pairs in Comparison of Top10
Zero	643	118
One	205	266
Two	13	251
Three	0	154
Four	0	59
Five	0	12
Six	0	1
Total:	861	861

Thus it indicates that there is a large degree of diversity among users in terms of usage duration for their most used applications (the sessions frequency show very similar results). Moreover it has to be noted that our population set is not highly varied demographically i.e. 60% are students, 59% aged between 17-25 years and 68% in the same domain of activity, as shown in Table I. However each user mostly uses applications which are different from other users.

As a result of this high diversity of application usage, it is not feasible to relate the social profile of each user with the exact applications he uses.

## V. USAGE ANALYSIS IN TERMS OF SERVICES

The last section demonstrated the large diversity in the application usage, which prevented the comparison of smartphone users in terms of application usage. In this section we group the applications into categories of services in order to show how the usage of certain services can be related to socio-demographic profiles. In addition to demographic data, we also discuss how smartphone usage can be related to other factors of lifestyle, technological and cultural habits.

### A. Application Categorization by Service

During their participation, the 42 users of our dataset used a total of 929 different applications and 149 of those applications were on the list of Top 10 apps of all users. In order to categorize the apps as services we could not fully rely on the Google Play Store as its categorization is not highly methodical. Rather, each application developer assigns a category based on strategies such as commercial motives, category competition, recommendation of ranking services (ex: www.applyzer.com).

Our application categorization is done in three steps: 1) assign our own category based on the app's description. 2) compare it to the category of Google Play Store. 3) in case of a mismatch or absent on Google Play Store, decide based on further information from the web. Nevertheless, for non-ambiguous types such as Weather, Transport, Games our own categorization matched mostly with Google Play Store. Thus we classified the 149 different applications in the Top10 list of all users into 26 different service types. Figure 5 shows the 17 most used types of services we defined through this method.



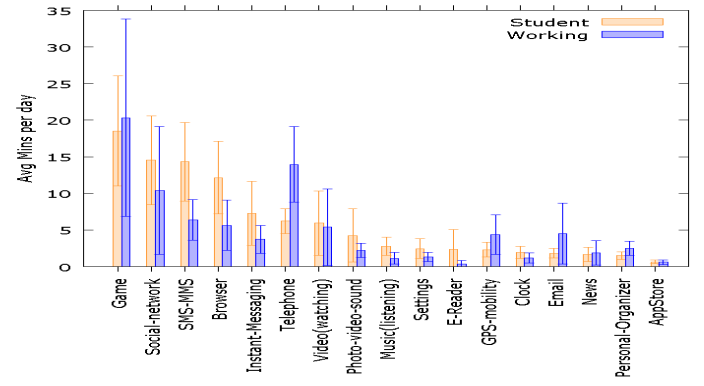
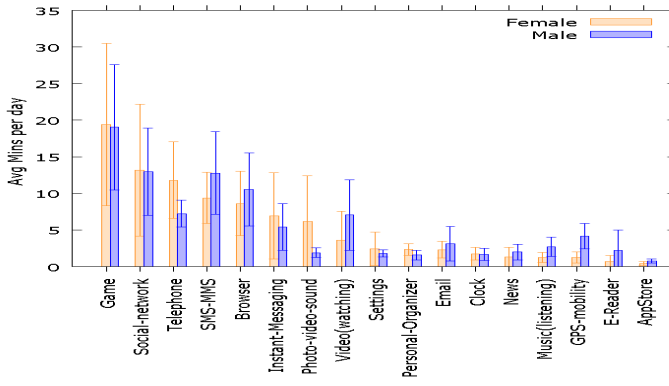


Fig. 5: Comparison of avg daily usage (metric 3) of different services between demographic groups

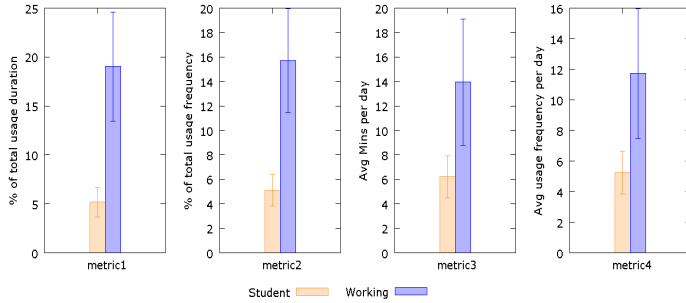


Fig. 6: Difference in Telephone usage between user groups of students and working people

### B. Methodology of Usage Logs Analysis and MCQ Declared Factors

The 10 MCQ declared parameters we analyze are shown in Table IV, which are just some of the many questions answered by the participants. The parameter values divide our user population into various demographic groups. For each sociological parameter, we compare the average usage of different services between the groups such as Male vs Female, Student vs Working, as shown in Figure 5.

TABLE IV: Sociodemographic Parameters Analyzed

Sociological Parameters	Parameter Values		
	Male	Female	
Gender	Male	Female	
Age	17 to 22	23 to 28	29 or +
Monthly revenue	<1.5k €	1.5k-3k €	3k € +
Marital Status	Single	Couple/Married	
Nb of Children	0	1 or 2	More
Profession	Student	Active	
Domain of Activity	Science/IT	Humanities/Commerce	Others
Daily Time Spent on PC	<1hr	1-3hr	3hr+
Preferred for Net Access	PC	Smartphone	
Smartphone Quality Range	Low/Med End	High-End	

As the smartphone usage is deeply embedded within people's lifestyle, profession, social behavior, amusement habits, we expected that some classes of the population would use some types of services more than other classes. To ensure whether the usage difference of a particular service between two groups is really due to the demographic parameter, we conducted the two-sample t-test (assuming equal variances, significance level: 0.05) and calculated the two-tailed p-value. If the p-value was 0.04 or less, we accepted the relation. In order to elaborate indicators that take into account the duration, the frequency and the most used apps per user, we have considered four metrics to evaluate and compare the usage of a service type by a demographic group. Metric 1 is the percentage of total usage duration resulted by a particular service type and Metric 2 is the percentage of total usage frequency due to that service type. Metric 3 is the average minutes spent on that service per day and Metric 4 is the average number of times the service is used per day.

Moreover, Metrics 1 and 2, have been multiplied by a coefficient and corrected to take into account the importance of the service for a particular user. To calculate the coefficient, the services used by a user are ranked (from highest to lowest) according to the percentage of contribution of each service towards the total usage. The actual percentage is lowered for lower ranked services as:

$$\text{Corrected Percentage} = \text{Actual Percentage} \times \frac{\text{Rank of the service}}{\text{Total Nb of services}}$$

**Example:** User 1 and 2 both use 10 services in total. For user 1, voice call is ranked 10 (the highest rank) and accounts for 20% of his total usage duration. For user 2, voice call is also responsible for 20% of his usage duration, but is ranked 7th. So for user 1, the corrected % of voice call is:  $(20 \times 10) \div 10 = 20\%$  and for user 2 it is:  $(20 \times 7) \div 10 = 14\%$ .

### C. Service Usage and Socio-demographic Factors

Here we analyze the usage difference between demographic groups for particular types of services and present some hypotheses.

**Professional Factors:** In our dataset, the most distinctive difference is in terms of telephone usage, i.e. people who are married, older, professionally active or with higher income use telephone significantly more than people in their opposite group. These 4 parameters are themselves co-related, as people who are students, are younger, earn lesser and are mostly

unmarried. Figure 6 shows the difference in telephone usage between students versus people with a job/income for all the 4 metrics. Our metrics show that telephone accounts for  $19.0 \pm 6.8\%$  of usage time for professionals whereas just  $5.2 \pm 1.8\%$  for students,  $p\text{-val} < 0.001$ . A possible factor for this difference is that people who are professionally active use telephone for their work, causing this higher usage.

Similarly, Email is more used by people who are older and professionally active compared to their opposite groups, possibly for professional reasons. Moreover, this difference of Email usage is apparent for Metrics 1 and 2 but not for Metrics 3 and 4. This indicates that although there is no significant difference between professionals/older people and students/younger people in terms of their daily Email usage, for the former group, Email is a higher ranked than for the latter group, which creates the differences of Metrics 1 and 2 compared to Metrics 3 and 4.

**Economic Factors:** Similarly there is also an economic factor i.e. unlimited telephone/voice plans are more expensive, which younger people with limited budget cannot afford. However, the least expensive phone plan in France, offer unlimited SMS. So student/younger people prefer SMS over phone call, and the number of SMS sent by younger people ( $33 \pm 18.4$  SMS/day), is superior to the ones of older people ( $7.2 \pm 5.6$  SMS/day).

**Age vs Technological Preference:** Another explanation of younger people using more SMS is related to the generational effect of technological preference. Text/SMS is the preferred medium of social conversation for teens compared to voice call [26]. There can be several reasons for this, such as SMS provides the required privacy desired by teens during communication, for example in a classroom or when surrounded by adults.

In our dataset, the average duration of a SMS session of people aged 17-22 years is 40.2 seconds with standard deviation 23 sec compared to average of 31.1 seconds, (SD 19 sec) for people aged 29+. Although the standard deviation is high, we also observed that more frequently for younger people, one SMS followed a chain of other SMSs in a short span of time, suggesting a session of conversation. Whereas older people's SMS sessions were shorter in duration and tended to be sporadic, suggesting the use of SMS for sending factual information or any short or rapid communication.

**GPS vs Gender & Smartphone Quality:** We characterize the phone models of our participants based on the classification specified by the phone manufacturer into high-end and low/medium-end. In our dataset, applications which require geo-localization such as travel/tourism are more used on higher-end phones,  $4.3 \pm 2.3\%$  of total usage duration vs  $0.3 \pm 0.3\%$  on lower-end models, as shown in Figure 7 (metric 1). Moreover, 85% of males in our dataset use a high-end smartphone while a lesser number of women (63%) use a high-end model. This is coherent with a survey of Nielsen<sup>2</sup>, which shows that males are more attentive towards technical specifications of smartphones, whereas women are more concerned about price, contract options and aesthetics.

However, we cannot hypothesize any economic relation in this case because there are more females who are professionally active in our dataset than males. Nevertheless, the males who are professionally active are mainly researchers and professors who have generally a wide-scale geographic mobility for

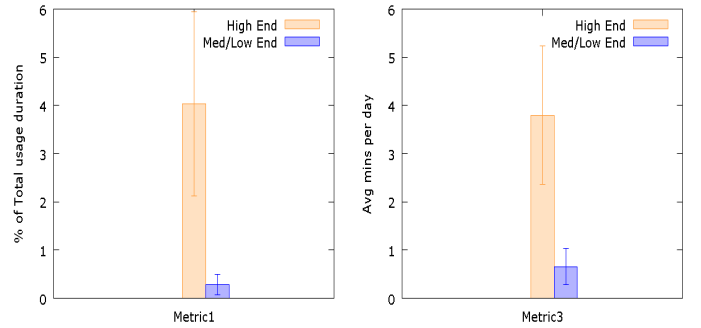


Fig. 7: Use of GPS wrt Smartphone Quality

reasons such as lectures, workshops, conferences etc. Thus they use applications which require GPS and applications related to transport and mobility.

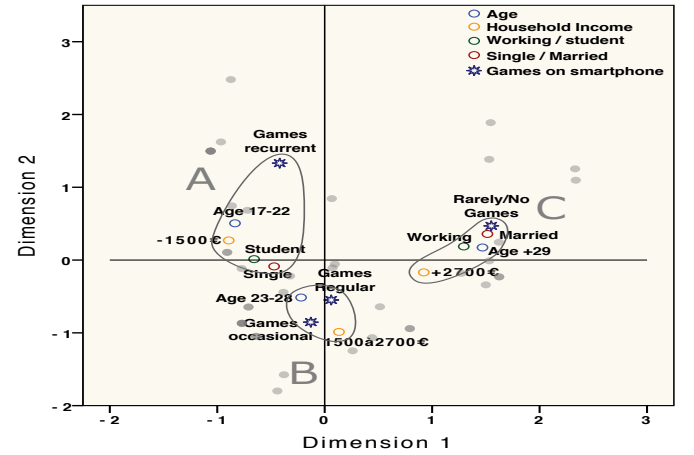


Fig. 8: Multiple Correspondence Analysis for Games

## VI. MULTIPLE CORRESPONDENCE ANALYSIS FOR TECHNOLOGICAL AND CULTURAL HABITS

### A. Sociodemographic Factors: Multiple Correspondence Analysis

In the last section, we showed the difference in service usage considering one sociodemographic factor at a time. However more advanced methods can be used such as **Multiple Correspondence Analysis (MCA)** [27] to find the correlation between measured service usage frequency and multiple declared factors. MCA is particularly adapted to find correlations in small and non-representative corpora. For example, in the case of Games usage, no clear distinction could be found among population groups by simply comparing the average usage among population groups. Therefore we use MCA on Games and find correlation trends simultaneously among multiple socio-demographic factors, as shown in Figure 8.

In our dataset, gaming applications are used frequently (at least once every 4 days) by everybody except 5 users. We elaborated a simple aggregated game usage frequency indicator with 4 degrees of routine usage as: No/Rarely (used only between 0-25% of the days), Occasional: used 25-50% of the days, Recurrent: used 50-75% of the days, Regular: used 75-100% of the days. In figure 8, we have highlighted 3 groups of factors

<sup>2</sup><http://www.nielsen.com/us/en/insights/news/2014/the-female-male-digital-divide.html>

(A, B & C) that are strongly correlated with the users around each zone (marked by grey spots).

Dimension 1 represents three demographic parameters: age, marital status and professional activity with discrimination measure = 0.69. Dimension 2 explains the regularity of playing games on smartphones with discrimination measure = 0.85. The reliability of the results is measured by Cronbach's alpha, which is 0.844 and 0.317 for dimensions 1 and 2 respectively. These reliability values are within the acceptable range of 0.7 (subject to discussion), considering our relatively small dataset.

Zone A shows users aged between 17-22 years, who are students and single, appear to play games on phone with recurrent frequency. On the opposite side, zone C shows users who do not play games and are more aged, married and professionally active. In the middle, zone B represents users aged 23-28 years, either student or professional, with an intermediate level of income, who play games either occasionally or regularly.

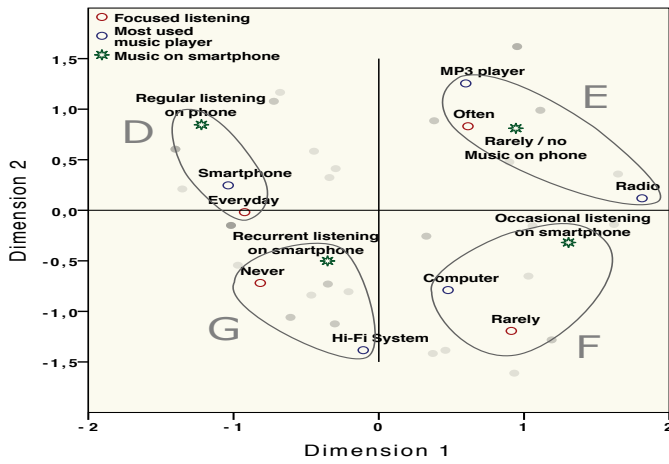


Fig. 9: Multiple Correspondence Analysis for Music

### B. Smartphone Usage Correlation with Cultural Habits

Although MCA is not a highly precise method i.e. does not divide the users into well-defined clusters, it can sense complex multifactorial correlations. Thus using MCA we can represent how smartphone usage is intertwined with complex sociological factors and parameters of lifestyle, technological preferences or cultural profiles.

For example, MCA allows us to identify two different long-term effects of the introduction of the smartphone as a device for entertainment: the different degrees of substitution of entertainment devices and the intensity of entertainment practices. As shown in Figure 9, we associated the degree of regularity of listening to music on smartphones with two factors: i) frequency of focused music listening in general, ii) most preferred device for music listening. These two factors are declared by users on MCQ and music listening regularity on smartphone is obtained from usage logs. The 4 levels of regularity are the same as used earlier for games. By focused music listening, we mean listening attentively to music, for its own sake.

In Figure 9, the discrimination measures do not associate each variable distinctly with any dimension unlike the previous MCA. Nevertheless Cronbach's alpha is satisfactory for both the dimensions i.e. 0.765 and 0.569 respectively. Zone D gives us two information, firstly users who declared to use smartphone as the preferred device for music listening, indeed

declared it correctly as determined from their usage logs. Secondly, these users are also music lovers who regularly spend time just listening to music without doing anything else. On the contrary, zones G and F show users who prefer listening to music on Hi-Fi systems or computers, tend to use music as a background ambiance and use smartphone irregularly for music listening.

## VII. DISCUSSION AND FURTHER APPLICATIONS

Thus we have shown how smartphone usage is related to not only users' demographic factors but also lifestyle factors such as cultural habits or technological preferences. Considering sociological factors for the analysis of smartphone usage, not only helps to better identify and interpret usage patterns, but on the other hand helps social science to understand the impact of smartphones on people's habits and lifestyle. Despite the limitations of our dataset, our method is an interesting path in order to analyze both the reproduction of cultural practices (as developed in French sociological tradition, [28]) from one technological generation of devices for cultural consumption to another, and also to analyze the modification of cultural behaviours due to portability of devices and accessibility of contents

In further studies, we aim to collect larger and more diversified samples of smartphone operating systems and users, which will enable us to examine more consistently the effects of user class or cultural group. A track to achieve this goal is to design incentive strategies, such as the sensing application would be associated to services that allow users to benefit from gathered data. There is an ongoing study based on similar methods but about network Quality of Experience<sup>3</sup>. The goal of this project is to articulate social factors of quality perception, usage patterns and network analysis in order to provide diagnosis and technical support.

However, this kind of perspective does not resolve by itself two major methodological challenges. Even if it is not a totally new topic [29], and beside the crowdsourcing/sensing disintermediation enthusiasm (skipping the expensive pollsters and collecting extensive datasets per user), these emergent survey techniques create methodological challenges for both social and data scientists. The insurmountable web diffusion bias and technological filters (availability of smartphones in the population and coverage of very diverse operating systems) impose us to rethink the paradigm of representativeness as it is conceived today in surveys and official statistics. Another major challenge for the consolidation of this field of research would be the elaboration of common methodologies and user classifications in order to make cumulative, comparative and periodical studies [30].

Further, another use case of implementing smartphone usage pattern would be to recommend apps based on usage pattern and sociological factors. It is important as we have seen that in our dataset, a smartphone contains 84 apps in average, but only 10 apps are used 80% of the time, suggesting that a lot of apps are downloaded but not used regularly. As the number of apps available on app stores continuously increases, the number of app downloads as surveyed by Statista, reached over 200bn downloads in 2015<sup>4</sup>. Therefore, the necessity for recommender systems to bring the right app to the right user will increase. Several propositions on recommender systems have been published, which are based on features such as usage pattern [31], usage location [32], social surrounding [33]. However, these studies have not considered sociological context :

<sup>3</sup><http://project.inria.fr/bottlenet/>

<sup>4</sup><http://www.statista.com/statistics/271644>



the closest parameters to social context are only related with social networks or social surroundings. If an app recommender system knows socio-demography, lifestyle, on-line and off-line cultural tastes, amusement habits as part of users' profile, then it can suggest a user about the popular apps currently trending among people sharing those commonalities such as common interests, hobbies, cultural preferences etc. Moreover, sociological information can help improve the precision of application categorization and app developers can provide sociological meta-data about their applications, concerning the potential audience to be interested by the application.

## VIII. CONCLUSION

In this study, we have explored several methods to exploit a new generation of tools and data based on the crowdsensing of smartphone usage and combining it to declared information about socio-demographics, lifestyle, technological preferences and cultural habits of their users. We have shown that application usage among users is highly diverse. However when the applications are grouped as services, interesting relations appear between user profiles and types of services used. We have discussed that socio-demographic and lifestyle information can be used for user profiling and characterizing smartphone usage pattern. We have proposed several possible use cases of how sociological information can be used to renew the official statistics, to recommend suitable applications to potential users. As part of our ongoing and future work we are building a crowdsourcing model that will increase the incentive for participants and give us a larger dataset. One use case we will consider is to combine sociological information, network access and usage patterns to improve QoS and network troubleshooting for mobile data users.

## REFERENCES

- [1] T. Ruggiero, "Uses and gratifications theory in the 21st century," *Mass Communication & Society*, p. 3–37, 2000.
- [2] O. Donnat, "Cultural practice and internet usage," *Culture études*, pp. 1–12, 2007.
- [3] T. M. Do, J. Blom, and D. Gatica-Perez, "Smartphone usage in the wild: A large-scale analysis of applications and context," in *Proc. of the International Conference on Multimodal Interfaces*, 2011.
- [4] H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R. Govindan, and D. Estrin, "Diversity in smartphone usage," in *Proc. of the International Conference on Mobile Systems, Applications, and Services*, 2010.
- [5] M. Bohmer, B. Hecht, J. Schoning, A. Kruger, and G. Bauer, "Falling asleep with angry birds, facebook and kindle: A large scale study on mobil application usage," in *Proc. of the International Conference on Human Computer Interaction with Mobile Devices and Services*, 2011.
- [6] H. Falaki, D. Lymberopoulos, R. Mahajan, S. Kandula, and D. Estrin, "A first look at traffic on smartphones," in *Proc. of the 2010 ACM Conference on Internet Measurement*.
- [7] Q. Xu, J. Erman, A. Gerber, Z. Mao, J. Pang, and S. Venkataraman, "Identifying diverse usage behaviors of smartphone apps," in *Proc. of the 2011 ACM Conference on Internet Measurement Conference*.
- [8] A. Gember, A. Akella, J. Pang, A. Varshavsky, and R. Caceres, "Obtaining in-context measurements of cellular network performance," in *Proc. of the 2012 ACM Conference on Internet Measurement Conference*.
- [9] D. T. Wagner, A. Rice, and A. R. Beresford, "Device analyzer: Understanding smartphone usage," in *Proc. of Mobile and Ubiquitous Systems: Computing, Networking, and Services*, 2014.
- [10] C. C. Tossell, P. Kortum, C. W. Shepard, A. Rahmati, and L. Zhong, "Getting real: a naturalistic methodology for using smartphones to collect mediated communications," *Advances in Human-Computer Interaction*, vol. 2012, p. 10, 2012.
- [11] A. Ouakrat, "Du rythme d'usage du smartphone aux rythmes de vie: les normes temporelles informelles des pratiques d'une population étudiante," *Questions de communication*, no. 1, pp. 301–321, 2015.
- [12] T.-M.-T. Do and D. Gatica-Perez, "By their apps you shall understand them: Mining large-scale patterns of mobile phone usage," in *Proc. of the International Conference on Mobile and Ubiquitous Multimedia*, 2010.
- [13] J. Kang, S. Seo, and J. W. Hong, "Usage pattern analysis of smartphones," in *Proc. of Network Operations and Management Symposium (APNOMS)*, 2011.
- [14] T. Stober, M. Frank, J. Schmitt, and I. Martinovic, "Who do you sync you are?: Smartphone fingerprinting via application behaviour," in *Proc. of the ACM Conference on Security and Privacy in Wireless and Mobile Networks*, 2013.
- [15] A. Rahmati and L. Zhong, "Studying smartphone usage: Lessons from a four-month field study," *IEEE Transactions on Mobile Computing*, vol. 12, no. 7, 2013.
- [16] T. Yan, D. Chu, D. Ganesan, A. Kansal, and J. Liu, "Fast app launching for mobile devices using predictive user context," in *Proc. of the International Conference on Mobile Systems, Applications, and Services*, 2012.
- [17] V. Srinivasan, S. Moghaddam, A. Mukherji, K. K. Rachuri, C. Xu, and E. M. Tapia, "Mobileminer: Mining your frequent patterns on your phone," in *Proc. of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2014.
- [18] C. Shin, J.-H. Hong, and A. K. Dey, "Understanding and prediction of mobile application usage for smart phones," in *Proc. of the ACM Conference on Ubiquitous Computing*, 2012.
- [19] M. Bohmer, L. Ganev, and A. Kruger, "Appfunnel: A framework for usage-centric evaluation of recommender systems that suggest mobile applications," in *Proc. of the International Conference on Intelligent User Interfaces*, 2013.
- [20] T. Beauvisage and H. Assadi, "From user-centric web traffic data to usage data," in *Special Interest Tracks and Posters of the International Conference on World Wide Web*, 2005.
- [21] T. Beauvisage, "Computer usage in daily life," in *Proc. of the SIGCHI conference on Human Factors in Computing Systems*, 2009.
- [22] C. Licoppe, D. Diminescu, Z. Smoreda, and C. Ziemlicki, "Using mobile phone geolocalisation for 'socio-geographical' analysis of co-ordination, urban mobilities, and social integration patterns," *Tijdschrift voor economische en sociale geografie*, vol. 99, no. 5, pp. 584–601, 2008.
- [23] N. Haderer, R. Rouvoy, and L. Seinturier, "Dynamic deployment of sensing experiments in the wild using smartphones," in *Distributed Applications and Interoperable Systems*, 2013.
- [24] V. Rivron, M. Khan, S. Charneau, and I. Chrisment, "Refining smartphone usage analysis by combining crowdsensing and survey," in *PERCOM Workshops*, 2015.
- [25] F. Hamka, H. Bouwman, M. De Reuver, and M. Kroesen, "Mobile customer segmentation based on smartphone measurement," *Telematics and Informatics*, vol. 31, no. 2, pp. 220–227, 2014.
- [26] A. Lenhart, "Teens, smartphones & texting," *Pew Internet & American Life Project*, 2012.
- [27] D. L. Hoffman and J. De Leeuw, "Interpreting multiple correspondence analysis as a multidimensional scaling method," *Marketing Letters*, 1992.
- [28] L. Fleury, *Sociology of Culture and Cultural Practices: The Transformative Power of Institutions*. Lexington Books, 2014.
- [29] J. Bethlehem, "Can we make official statistics with self-selection web surveys?" in *Proceedings of Statistics Canada Symposium*, 2008.
- [30] P. B. Brandtzæg, "Towards a unified media-user typology (mut): A meta-analysis and review of the research literature on media-user typologies," *Computers in Human Behavior*, vol. 26, no. 5, pp. 940–956, 2010.
- [31] B. Yan and G. Chen, "Appjoy: Personalized mobile application discovery," in *Proc. of the International Conference on Mobile Systems, Applications, and Services*, 2011.
- [32] C. Davidsson and S. Moritz, "Utilizing implicit feedback and context to recommend mobile applications from first use," in *Proc. of the Workshop on Context-awareness in Retrieval and Recommendation*, 2011.
- [33] A. Girardello and F. Michahelles, "Appaware: Which mobile applications are hot?" in *Proc. of the International Conference on Human Computer Interaction with Mobile Devices and Services*, 2010.