Human Mobility Flows in the City of Abidjan

Diala Naboulsi, Marco Fiore, Razvan Stanica

Human Mobility Flows in the City of Abidjan

Diala Naboulsi, Marco Fiore, Razvan Stanica
INSA Lyon / Inria, Lyon, France
firstname.lastname@insa-lyon.fr

***Abstract*—The growing ubiquity of mobile communications has offered researchers new possibilities to understand human mobility over the last few years. In this work, we analyze Call Detail Records (CDR) made available within the context of the Orange D4D Challenge, focusing on calls of individuals in the city of Abidjan over a period of five months. Our results illustrate how aggregated CDR can be used to tell apart typical and special mobility behaviors, and demonstrate how macroscopic mobility flows extracted from these cellular network data reflect the daily dynamics of a highly populated city. We discuss how these macroscopic mobility flows can help solve problems in developing urban areas.**

I. Introduction

Understanding human movements is critical for different scientific domains. In order to deploy efficient networking solutions, a clear view of human mobility patterns is required. The same applies for urban planning, where the global mobility flows can determine the optimal deployment of infrastructure. Human mobility also plays a major role when analyzing the ways diseases can spread in a population.

Significant research efforts have been conducted in this direction, aiming at understanding how people move as a first step, and proposing models of such mobility as a second step. Recently, as people are more and more connected, network traces have received particular attention as a source of information about human mobility at large scales.

However, previous studies have focused on developed countries, and whether the observed patterns and models are applicable to developing countries remains an open question, due to differences in the lifestyle, country's infrastructure and modes of transportation. Indeed, a clear understanding of human movements would be crucial for the progress of such countries, especially in highly populated urban regions where new transportation infrastructures are being deployed.

In this paper, we explore Call Detail Records (CDR) of Orange customers in Abidjan, the economic capital of Ivory Coast. The dataset, made available within the context of the D4D Challenge, provides the position of each caller – approximated as the base station's location – at every time he/she initiates a call or sends an SMS. We start by analyzing the temporal, spatial and geographical characteristics of the calls, which allows us to capture differences between distinct times of the day and different days of the week over multiple geographical regions of the city. We propose a method to distinguish between typical and outlying behaviors in the CDR dataset, enabling the detection of special events such as the New Year's Eve and football games played during the Africa Cup of Nations. Our approach also allows us to infer which moments can be aggregated in order to characterize macroscopic mobility flows that provide a view of the global and local mobility flows in Abidjan, as well as of their daily evolution.

The rest of the paper is organised as follows. Sec. II discusses previous studies focused on human mobility traces obtained from wireless network data. In Sec. III, we analyze the CDR dataset. We introduce our methodology to detect typical and special behaviors in Sec. IV. We then aggregate typical behaviors and extract the global mobility flows in Sec. V. Finally, we draw conclusions in Sec. VI.

II. Related work

Human mobility has been drawing significant research efforts over the last few years. Previous work mostly explores real-world movement traces in a variety of contexts and for diverse goals. Kim et al. [1], use wireless network traces from WiFi access points at Dartmouth College to extract a human mobility model. Their study stays valid within the limited context of a university campus, thus cannot be generalized to the level of a city where more complex human movements emerge. More recent works consider social networking platforms. Data from Foursquare is analyzed in [2] to propose a mobility model reflecting movements within a city. Girardin et al. [3] analyze geographically tagged photos from Flickr to uncover the movements of tourists in Rome. Similarly, in [4] the authors separate the behavior of tourists and that of residents in the city of New York. These studies are capable of capturing users' most preferred locations [2] and the main paths followed by tourists [3] [4], however they do not account for the temporal properties of human movements, while in our work we generate O/D matrices capable of capturing both temporal and spatial characteristics of human movements.

Analyses based on CDR have offered studies on human mobility a much wider perspective, uncovering important characteristics at large scales. González et al. [5] analyze the CDR obtained from a European mobile operator providing data on 100,000 mobile phone users over a 6 month period. Their aim is to understand individual users' movements, proving that these mobility patterns present high temporal and spatial regularities. In [6], the authors check the level of movement predictability that can be achieved knowing the history of individual movements. Both of these papers focus on the individual human mobility, while in our study we are interested in working on a more macroscopic level.

Isaacman et al. [7] consider the analysis of CDR by considering a population as a whole. They use the CDR of more than 100,000 individuals randomly chosen in Los Angeles and New York and track their movements over a period of 4 months. The authors compare the moving patterns between population in both cities. The same dataset is also applied in [8] to build a mobility model. Although in these studies the authors consider the metrics at the level of populations, their

main goal in [7] is to check the difference between population behaviors in both cities, and generate synthetic CDR in [8], while we are interested in detecting mobility flows. Closer to our study, the work by Pulselli et al. [9] shows how general trends of movements through the city of Milan can be detected based on call volume variations. However, our methodology, leveraging individual call records, allows to obtain a more accurate description of the global mobility flows.

In conclusion, all previous works consider urban areas of developed countries: their results do not necessarily reflect the mobility observed in cities of developing countries, which is our goal. As a corollary contribution, we focus on separating typical CDR behaviors from outlying ones – a subject that has drawn minor attention to date.

III. Dataset characterization

In our analysis we study datasets provided by Orange within the context of the D4D Challenge, based on the Call Detail Records of 5 million anonymized Orange customers in Ivory Coast. The information obtained span over 5 months, from December 5th, 2011 until April 22nd, 2012. We focus on two of the four datasets provided, detailed below.

Dataset D1: Antenna-to-antenna traffic. This dataset includes the call traffic volume exchanged between any two base stations in Ivory Coast on an hourly basis. It provides both the number of calls and their total duration for the whole observation period.

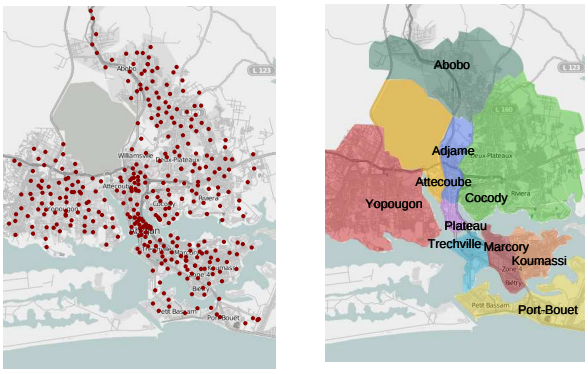
Dataset D2: Individual trajectories. For each two weeks of the observation period, this dataset provides the CDR of 50,000 individuals randomly chosen over the whole Ivory Coast Orange customer population.

Our study focuses on the city of Abidjan, the economic capital of Ivory Coast. Thus, we filter both D1 and D2 by keeping only the information involving the antennas in Abidjan. This leaves us with information about 364 antennas out of the 1231 antennas covering the whole country. In Fig. 1(a) we show the position of these base stations together with the city's street layout obtained from OpenStreetMap [10]. Fig. 1(b) presents the different *communes* of the city, providing a reference for discussions in the rest of the paper. We remark that, in the following, we will use the term *snapshot* to refer to data aggregated over each one-hour interval, and the term *call* to denote a call or an SMS indifferently.

A. Temporal and spatial inter-call properties

We start our analysis by considering the temporal and spatial properties of the second dataset. Specifically, our goal is to determine whether the D2 dataset can be reliably leveraged for our objective of characterizing human movements patterns. To that end, we generate the distributions of two relevant metrics: the inter-call duration that we define as the time elapsed between two consecutive calls made by the same person, and the inter-call distance defined as the distance separating the position of the person between two consecutive calls.

Fig. 2(a) displays the Probability Distribution Function (PDF) of the inter-call duration. We can clearly notice the concentrantion of high probability values for small inter-call durations, and the exponentially decreasing trend with peaks on a daily basis. The plot outlines the heterogeneity in the inter-call duration, as we can observe that, despite the low probability, very high values of inter-call duration can appear. Moreover, the probability peaks indicate the periodicity in calling patterns as successive calls are mainly separated by days. More importantly, 90% of the calls occur at one hour distance or less: the high time granularity makes the D2 dataset fit to our analysis from a temporal viewpoint.



(a) (b)

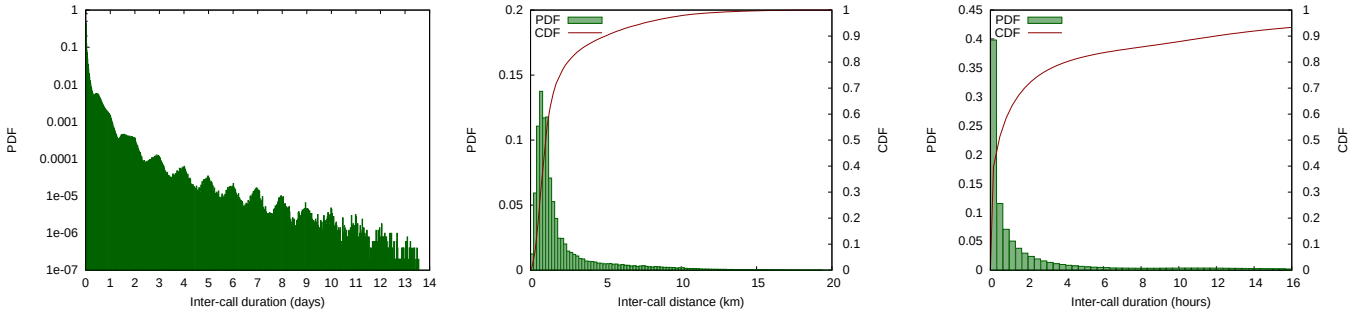
Fig. 1. (a): Distribution of the base stations in Abidjan over the street layout. (b): The 10 communes of Abidjan.

After understanding the global temporal trend of calls, we check the distributions of the inter-call distance in Fig. 2(b). For a clearer representation, we remove from this figure the consecutive calls that occur from the same base station, i.e., data that cannot be exploited from a mobility point of view. We remark that this concerns around 70% of the overall records, and we thus focus on the remaining 30% of the data, still statistically sufficient for our study. We can observe in this figure that more than 70% of the inter-call distances are concentrated within a 2km range, which means that when the position of an individual changes between two succesive calls, he/she tends to move within a limited area. Again, short traveled distances between calls positively affect the reliability of human movements inferred from the dataset.

Finally, in Fig. 2(c) we focus on inter-call durations corresponding to successive calls occuring from different locations only. This figure indicates that movements can still be captured with a high temporal precision, as 60% of the calls yielding a physical user movement occur within intervals smaller than one hour. Overall, our analysis shows that the D2 dataset can effectively be leveraged towards the characterization of user mobility, since a significant portion of the records yield low inter-call distances and durations.

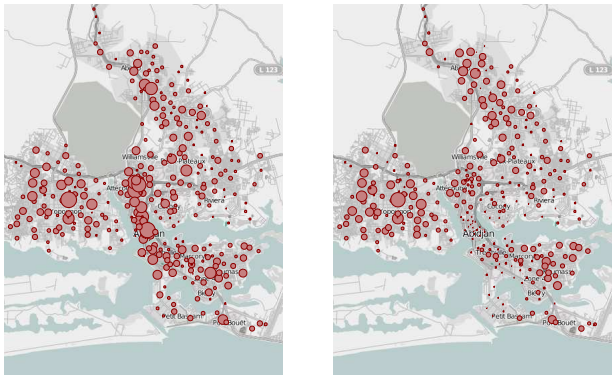
B. Geographical call volume diversity

Although the inter-call duration and the inter-call distance are capable of capturing the global temporal and spatial characteristics of the dataset, they cannot reveal the behavior in different regions of the city. Thus, we examine the distribution of calls over the city obtained from the first dataset D1, through a set of geographical plots. These geographical plots are capable of capturing what happens in different regions for different times of the day at an hourly basis. In Fig. 3, we present an extract of these plots, in which we represent the number of calls at each base station with a disk, whose radius



(a) (b) (c)

Fig. 2. Distributions of inter-call duration and inter-call distance. (a): PDF of inter-call duration for the whole observation period. (b): Distributions of the inter-call distance excluding consecutive calls from the same base station. (c): Distributions of the inter-call duration excluding consecutive calls from the same base station.



(a) Thursday, April 12th at 10:00 (b) Thursday, April 12th at 20:00

(c) Sunday, April 22nd at 10:00 (d) Sunday, January 1st at 0:00

Fig. 3. Geographical distributions of calls.

is proportional to the number of calls detected there. Fig. 3(a) and Fig. 3(b) present the geographical distribution of calls on Thursday, April 12th, 2012 respectively at times 10:00 and 20:00. We pick these times as typically time 10:00 reflects the concentration of individuals at the working and studying areas while at time 20:00 individuals are mostly present at their home locations.

At 10:00, we can clearly see the explosion of cellular traffic at the city center, the region in which most of the working acitivities take place. In some other places, such as the region of Cocody, we can still observe a certain number of base stations with a high density; these are mostly the regions where different universities of Abidjan are located. As for the commune of Yopougon, we can see it presents a heterogeneity at the level of base stations in terms of cellular traffic: base stations with high and low traffic coexist. That is due to the fact that this region is a mix of both residential and industrial areas.

On the other hand, if we consider the same day at time 20:00, we can observe that, in the city center, the traffic strongly decreases; the same applies for the university campuses. As for some residential areas in Abobo, they encounter a traffic increase, since people will be present at their homes, while Yopougon still presents a heterogeneity in the base stations in terms of traffic.

Fig. 3(c) illustrates the distribution of calls on Sunday, April 22nd, 2012 at time 10:00. Compared with the previous 2 figures, we can observe a behavior diverging from the typical behavior of a normal week day at the same time, and resembling more to what can be seen at time 20:00 of a normal weekday. This comes confirming the fact that on Sunday morning people mostly stay at their homes.

As for the New Year's Eve portrayed in Fig. 3(d), first we remark that some of the base stations do not appear due to technical problems that we will discuss in more detail in the following section. Second, for the existing antennas, we can see that the volume of all the base stations explodes except for the working areas.

Based on these observations, we conclude that different hours of the same day show different patterns for the traffic on the cellular network, and these patterns can be linked with the human geographic distribution and mobility. More interestingly, different days at the same time present different behaviors. Finally, some special days can present strongly diverging behaviors. In the next section, we shed some light on the detection of typical days as well as special days and anomalies that we call outliers.

IV. Typical behavior and outliers

In this section, we first discuss the reliability of individual snapshots, and then explain how we tell apart typical and special behaviors in the CDR.

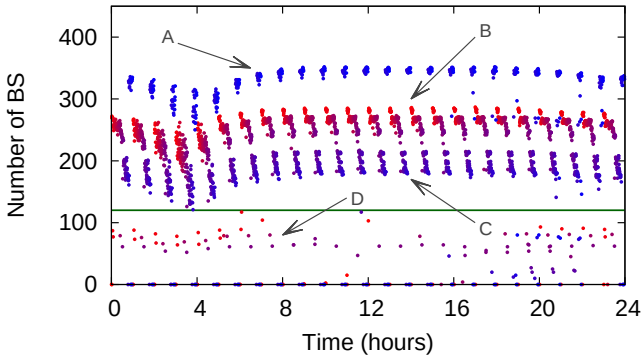


Fig. 4. Number of antennas over the whole observation period for the different hours of the day.

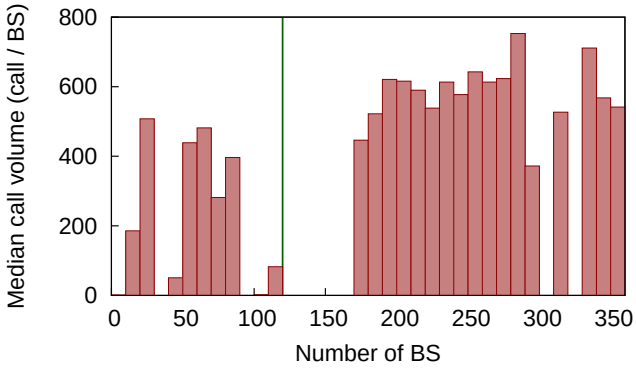


Fig. 5. Median volume per base station as a function of the number of existing base stations, aggregated over the entire observation period for high traffic hours: between 10:00 and 20:00.

A. Determining the snapshot reliability

As we mentioned in the previous section, we notice that some snapshots do not include all the antennas. That is due to technical problems encountered by Orange as well as electricity failures that can occasionnally occur in Ivory Coast. Thus, we now focus on the evolution of the number of antennas throughout the 5-month observation period.

In Fig. 4, we plot the number of antennas detected in D1 with respect to the time of the day. Each point on this figure represents the number of antennas included in the dataset at a specific hour of one day. We distinguish between the different days with a colour degradation from red to blue, such that the red colour maps to the first day in the dataset and the blue colour maps to the last one.

Three different behaviors are detected: the first one (labeled as A), includes the highest number of base stations – around 350 – and goes from March 28th, 2012 until April 22nd, 2012; the second one (B), with almost 250 antennas, goes from February 22nd, 2012 until March 27th, 2012; and the third behavior (C), featuring the smallest number of base stations – around 170 – and goes from December 7th, 2011 until February 21st, 2012.

We also point at the appearance of fluctuations in each of these behaviors, where local minima appear in the night hours between times 3:00 and 6:00, while the number of antennas

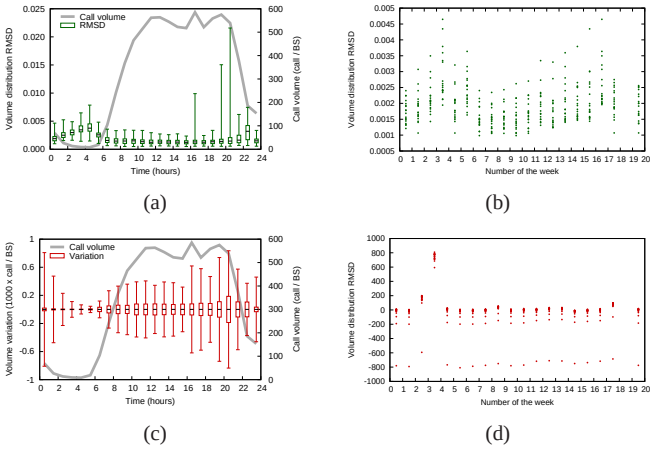


Fig. 6. (a): The volume distribution RMSD for all Sundays at different times of the day. (b): The volume distribution RMSD for all Sundays at time 0:00. (c): The volume variation for all Sundays at different times of the day. (d): The volume variation for all Sundays at time 0:00.

stays almost stable for the rest of the day. The presence of these minima can be explained by the fact that during night hours a very small number of individuals use the cellular network, which means that there is less chance to detect calls from base stations all over the city.

However, we observe that a limited number of points (tagged as D) with a very small number of antennas also arise. These are the result of missing information or malfunctioning of the majority of base stations. We verify that such a significantly reduced number of base stations has an impact on the reliability of the call traffic information in Fig. 5. There, we draw the evolution of the median per-base station call traffic volume with respect to the number of base stations, for all the snapshots with notable call traffic, i.e. between times 10:00 and 20:00, over all the 5-month observation period. The three behaviors A, B and C map to a consistently high median volume per base station. The two intervals centered around the values of 305 and 325 antennas present null median volume values as no snapshot exists with a number of base stations lying in one of these intervals. Conversely, for the snapshots with less than 120 antennas, falling in the D case above, the behavior differs: we observe highly variable median values with a lower average. This leads us to exclude from our analysis snapshots that fall in the D category, i.e., for which less than 120 antennas are recorded, since they yield irregular traffic volume information, due to technical issues in the network and that would risk to bias our analysis.

B. Identifying outlying behaviors

After having pruned the D1 and D2 datasets from unreliable snapshots, we focus on separating typical and special snapshots. To that end, we compare different snapshots through two metrics, defined next.

Let us use $C_i(t)$ to refer to a cell i appearing in the 1-hour interval from time t until time $t+1$. $\mathbb{C}(t) = \{C_i(t)\}$ represents the set of cells detected between time t and $t+1$. We denote the volume of each cell $C_i(t)$ as $v_i(t)$. We use $\mathcal{V}(t)$ to refer to the total volume of calls obtained between time t and $t+1$. We define the intersection of the appearing

antennas between the two 1-hour intervals at times t and $t+k$ as: $I(t,t+k) = \mathbb{C}(t) \cap \mathbb{C}(t+k)$ We define the root mean square deviation of the volume distribution distance, or *volume distribution RMSD*, between two different 1-hour intervals at times t and $t+k$ as:

$$\mathcal{D}(t,t+k) = \frac{1}{\sqrt{|I(t,t+k)|}} \sqrt{\sum_{C_i \in I(t,t+k)} (\frac{v_i(t)}{\mathcal{V}(t)} - \frac{v_i(t+k)}{\mathcal{V}(t+k)})^2}.$$

The $\mathcal{D}$ metric captures the distance between two snapshots, in terms of distribution of volumes among the base stations present in both snapshots.

We also account for the overall volume variations, by calculating the *volume variation* per base station as:

$$\mathcal{DV}(t,t+k) = \frac{\sum_{C_i \in I(t,t+k)} (v_i(t) - v_i(t+k))}{|I(t,t+k)|}$$

The $\mathcal{DV}$ metric captures large positive or negative variations in the calling volume.

We leverage the two metrics above to compare the same hour of the same days of the week at different dates: for example we compare every Saturday at 7:00 with all the other Saturdays at 7:00.

In Fig. 6(a) and Fig. 6(c), we show the evolution of $\mathcal{D}$ and, respectively, $\mathcal{DV}$ for all the Sundays for different times of the day. The candlesticks show the median, minimum and maximum values, together with the first and third quartile of each of the two metrics. We also show on each of these figures the median volume detected per base station. It is clear from Fig. 6(a) that $\mathcal{D}$ stays mostly small, but still presents some relatively high outlying values. We can also notice the candlesticks positioned slightly higher for the night hours. If we map that with the variation of the median volume per base station, we notice that this shift is caused by the variation of traffic: in the case of small traffic, it is more probable to detect variations in the volume distribution RMSD. It is noteworthy that these variations are not necessarily linked to the overall cellular traffic volume: different volumes distributed in a similar geographic manner result in small $\mathcal{D}$, while a similar total volume distributed differently can produce high $\mathcal{D}$ values.

As for the volume variation in Fig. 6(c), we can see the larger fluctuations are detected in the day hours, when high volumes are reached; this can be the result of large increase in the calling volumes on special events or large decrease in the calling volume due to the technical problems.

In the next step, we detect outliers based on the boxplot technique [11] such that: if Q1 and Q3 represent the first and third quartile, and IQR = Q3 - Q1 the interquartile range, then every value that is smaller than Q1-1.5*IQR or greater than Q3+1.5*IQR is detected as an outlier. We eliminate every snapshot causing outliers detected based on $\mathcal{D}$ since it reflects untypical cellular traffic distribution. As for the outliers detected based on $\mathcal{DV}$, we distinguish between negative and positive variations. The negative variations represent moments of relatively very low traffic, which can be due to technical problems in producing the datasets. As these problems are not uniformly distributed over all the base stations, such negative variations usually produce outliers on the $\mathcal{D}$ set; therefore these snapshots are removed from the aggregation process described in the next section. However, a different trend can be observed for positive variations of the total volume, usually produced by special events when people use their phones simultaneously. Our analysis shows that these snapshots do not necessarily result in outliers on the volume distribution RMSD, meaning that the use of the cellular network changes (as more people make a call), but the geographical distribution of the users does not modify.

For example, this technique allows us to detect special days such as the New Year's Eve, which happened to be on a Sunday and it is therefore shown in Fig. 6. This outlying value is detected by a high positive volume variation and a high value of $\mathcal{D}$, which means that it does not only imply a volume variation, but also a different volume distribution RMSD. This can be caused by the fact that possibly on such events, people gather in special places to celebrate. This outlying value can be better noticed in Fig. 6(b) and Fig. 6(d). In these figures, we show the obtained $\mathcal{D}$ and, respectively, $\mathcal{DV}$ values at time 0:00 for the different Sundays of the dataset. We refer to every Sunday based on the number of the week to which it belongs such that the first Sunday of the observation period is designated by 0. In these plots, every point represents the value of $\mathcal{D}$ and $\mathcal{DV}$ obtained by considering the snapshot at time 0:00 on the Sunday specified on the x axis, with another Sunday snapshot at the same time. We remark that The New Year's Eve is the fourth column on these figures, and the high relative traffic volume and RMSD distance can be distinguished.

However, the volume variation $\mathcal{DV}$ also allows us to detect 5 out of 6 football games of the Africa Cup of Nations 2012, hosted by Gabon and Equatorial Guinea, in which the Ivory Coast participated, with the last one left undetected because it presents less than 120 antennas and thus is excluded of this part. As an example, we detect the final game that took place on Sunday, February 12th, 2012, causing the positive outlying values at 23:00 in Fig. 6(c). Nevertheless, these events do not present an abnormal distribution of the cellular traffic, despite the increased overall volume, therefore they are not detected as outliers on the $\mathcal{D}$ set. This is an interesting result, as it shows that, despite the common practice in other studies which recommend excluding high traffic moments when studying human mobility based on CDRs, these snapshots can still be useful from a mobility point of view when they present classical geographical distribution patterns. Moreover, their use can even bring benefits, as the increased traffic on the cellular network allows us to detect more individual trajectories during these moments.

Finally, our results allow us to conclude that it is possible to aggregate the same times of the typical days presenting acceptable values of the volume distribution RMSD and the volume variation.

V. Mobility vectors

In this section, we construct a series of origin-destination (O/D) matrices that accurately represent the human mobility flows over the Abidjan urban area and discuss the major mobility flows detected using this procedure.

A. Building O/D matrices

While the analysis proposed in the previous section is based on the first dataset, D1 can not be used as the main source for an O/D matrix. The reason is that, although it provides us with a view of the entire traffic over the cellular network, the dataset D1 is focused on values aggregated per base station, without any information on the individual trajectories. Therefore, in order to construct the mobility flows in the Abidjan area, we make use of the dataset D2 which contains individual information for a subset of the users.

However, using only a subset of the individuals, sampled by the operator following an unknown distribution, can reduce the accuracy of the mobility matrices. To alleviate this problem, we decided to aggregate multiple days, which allows us to follow a larger set of individuals; for example we aggregate multiple Tuesdays at time 10:00, therefore creating a *classical 10am Tuesday*. In order to establish which days can be aggregated, we use input from the detailed analysis of the dataset D1, as follows:

i) Because in dataset D2 the position of every individual is associated to the geographical location of the serving base station, to accurately represent the mobility flows in the city of Abidjan it is preferable to use the snapshots containing the highest number of antennas. We therefore chose the consecutive days from April 2nd, 2012 until April 15th, 2012 as they include the complete set of antennas, capable of offering a clear picture of the movements all around the city.

ii) Even during this period, during some time intervals the cellular traffic might be affected by special events or situations. As our goal is to detect the mobility patterns in a *classical* day, we use the outliers detection technique described in the previous section to eliminate these special moments from the aggregation process. Overall, this leaves us with more than 32k individual trajectories, representing almost 1% of the entire Abidjan population.

In the following, we denote as $\mathcal{OD}(t)$ the O/D matrix representing the mobility flows between time t and $t+1$. $\mathcal{OD}(t)$ is a square matrix of order n_a, where $n_a = \max_t |\mathbb{C}(t)|$ represents the highest number of base stations simultaneously present in the dataset. Every line and column i of the matrix represents the outgoing, respectively the incoming, human mobility flow in the geographical area covered by base station C_i during the time interval between t and $t+1$. We denote as Δt the *time step* of the O/D matrix, practically the time duration between t and $t+1$. As D1 provides data on per-hour basis, we decided to keep this time granularity in our mobility analysis, therefore $\Delta t = 1$ hour. This means that every element o_{ij} from $\mathcal{OD}(t)$ represents the number of trips with a starting point at base station C_i and arriving in the area covered by C_j during a Δt time interval.

However, building the O/D matrix from dataset D2 is not a trivial task. Indeed, when using CDR to infer human trajectories, the accuracy of an individual movement is given by the frequency of the CDR data. As shown in Fig. 2(a), the inter-call duration presents a very heterogeneous distribution, which makes it difficult to map the detected displacements to a mobility matrix using a fixed time-step. For example, an important question is how to include in $\mathcal{OD}(t)$ a movement that spans over multiple hours. In this sense, we distinguish among three types of movements:

i) A movement with a duration smaller than Δt where both the starting and end time are situated between t and $t+1$. An example from this category is a person who calls two times between 9am and 10am, the first time using base station i and the second time while being associated to base station j. We consider such a movement adds a weight of 1 to the element o_{ij} from $\mathcal{OD}(9)$.

ii) A movement with a duration smaller than Δt spanning over two snapshots. This is the case of a person making a call from base station i at 9:45am and another one at 10:20am using base station j. The problem in this case is that such a displacement needs to be shared by the two matrices $\mathcal{OD}(9)$ and $\mathcal{OD}(10)$. While different strategies could be applied in such a scenario, in our study we decided to add a weight of 1/2 to the element o_{ij} in both matrices.

iii) A movement with a duration higher than Δt. Such an event clearly covers multiple snapshots, therefore we denote as n_s the number of O/D matrices that need to take into account the movement. For example, a person detected near base station i at 9:10 and who makes the next call at 15:30 using base station j has obviously moved between these two moments and needs to be considered in at least one of the $n_s = 7$ Δt intervals covering this time period. Although it is highly probable that the person in question did not continuously move for the entire interval, and the movement could actually be restricted to a smaller number of snapshots, the limited resolution of CDR data does not allow this operation. Therefore, we decided to add a weight of $1/n_s$ to o_{ij} in each of the n_s matrices covering the detected time interval. However, we need to point out that, according to Fig. 2(c), the probability of high inter-call durations is relatively low, meaning that the impact of these *long movements* on the mobility flow is not very important.

B. Identifying major mobility flows

The O/D matrices obtained using the process described above are an essential input for studies in different scientific areas, such as intelligent transportation systems, urban infrastructure planning, or cellular network deployment. However, presenting numerical values in a row/column format is not the best choice for visualization purposes. The solution we adopted to solve this issue consists in associating a vector $\vec{V}_{ij}(t)$ to every value o_{ij} in $\mathcal{OD}(t)$. The origin of $\vec{V}_{ij}(t)$ is the location of base station C_i, while its direction is given by the segment connecting base stations C_i and C_j. Finally, the length of the vector $|\vec{V}_{ij}(t)| = o_{ij}$.

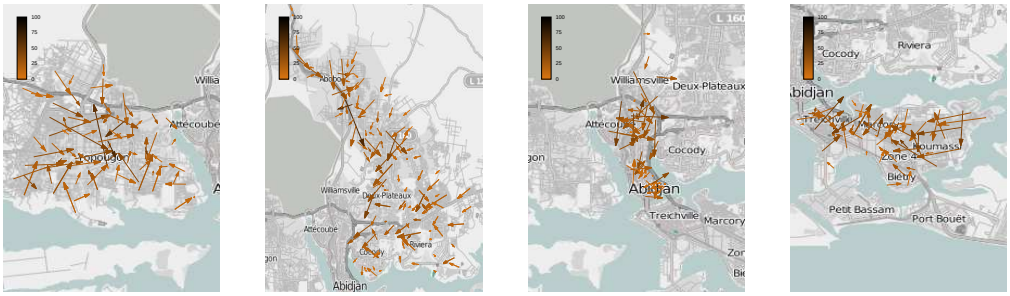
As each pair of antennas results in a vector, this implies that every base station C_i has a number of $n_a - 1$ associated vectors created this way. In the next step, we compute the resultant vector $\vec{V}_i(t)$ for every C_i, as follows:

$$\vec{V}_i(t) = \sum_{j \in \mathbb{C}(t), j \neq i} \vec{V}_{ij}(t)$$

While $\vec{V}_i(t)$ practically represents an aggregation of the human flows, we believe it can represent very well the general trends of the urban mobility. Moreover, as it can be noticed in Fig. 8, even after this aggregation the number of vectors on a figure

(a) 6:00 (b) 8:00 (c) 13:00 (d) 19:00 (e) 22:00

Fig. 7. Mobility vectors of each group of base stations for different times of the day for Thursdays.



(a) Yopougon (b) Cocody and Abobo (c) Plateau and Adjame (d) South of Abidjan

Fig. 8. Mobility vectors of each base station for different regions of Abidjan for Thursdays at 8:00.

is high, which makes the mobility flows difficult to follow. Therefore, we use these per base station vectors only when we want to zoom on a geographical area to uncover local flows. For city-scale mobility, we add another aggregation step, where we divide the entire map in squares with an area of 1.2 km^2, and we group together the base stations in every square, by summing all the associated vectors.

In Fig. 7 we represent the global mobility flows in Abidjan, for different daytimes of a classical Thursday, while in Fig. 8 we represent the mobility flows per base station at time 8:00 for separated regions of the city. These vectors are represented with arrows whose sizes map to their lengths, while the colors map to the total volume of outgoing movements. It is noteworthy that these two metrics are not necessarily correlated, as a geographical area with a high number of outgoing trips uniformly distributed in the space would result in a high total volume, but a small length of the resultant vector.

In Fig. 7(a), at time 6:00 we can observe a general trend of vectors oriented towards the city center, however their sizes and colors reflect the presence of only minor movements.

At time 8:00, represented in Fig. 7(b), we notice that the vectors present an increased length, meaning that a higher level of mobility is detected, mainly due to the movement of people between their homes and their working or studying areas. Most of the vectors are oriented towards the Plateau, as it constitutes a working and studying area itself, and a bridging region between the north and south of Abidjan.

At time 13:00, we observe a different behavior in the multiple regions of the city, shown in Fig. 7(c). Important flows emerge from the center; this pattern is probably related to the part time jobs that a part of the population has, which means that people in this category go back to residential regions at this time of the day, such as in Cocody. More interestingly, at time 19:00 presented in Fig. 7(d), the behavior detected is almost the opposite of what can be seen for the case of 8:00 as this is the time when most of the people go back to their homes using the same paths taken in the morning, but in the opposite direction. Finally, at time 22:00, Fig. 7(e) clearly shows that the city calms down: less movements can be detected, it is the time when most people are already back at their homes.

While these general trends are very clear, it is interesting to observe local mobility flows, using a per base station representation. If we focus on each commune of Abidjan separately at time 8:00, we can observe the following:

Yopougon: In this commune, represented in Fig. 8(a), we observe important flows directed towards the center of the commune. This can be explained by the fact that the bus station is located in that area, with public transportation offered to the residents to reach the other regions of Abidjan. In addition to that, we notice the presence of arrows directed towards the center of the city. This behavior is expected as people will mostly go to this region for their daily activities or transit to the communes of Cocody or Marcory for the same reason, as already observed on the general flows in Fig.7(b). We can also detect some flows oriented towards the north of the commune, going in the direction of the highway that links Yopougon to the other regions of the city. Vectors pointing at the south east of Yopougon are also detected, and they could be explained by the presence of the boat station, where another public transportation service is offered.

Abobo and Cocody: The arrows in these regions, drawn in Fig. 8(b), are mostly oriented towards the south following the direction of the main streets and the highway leading towards the Plateau.

Plateau and Adjame: In the Plateau we can observe from Fig. 7(b) a small dark arrow, which means that in this region an important volume of movement is detected but spread in different directions. This is clearly confirmed by the individual base station's vectors in Fig. 8(c), showing that in this area the destinations of the individual trips are much more heterogeneous. The same effect holds as well for the commune of Adjame represented on the same figures.

Marcory, Treichville, Koumassi: In these communes we can see from Fig. 7(b) that the largest arrows are mostly oriented towards the center of the southern part they form. In this case, it can be due to the working nature of Marcory or the fact that in this central area people can reach the highway to go to the north of Abidjan. We can also observe a large flow oriented towards Cocody, which is mostly representative of the students' flow. The vectors per base station in Fig. 8(d) come to confirm this behavior.

Our observations reflect the high commuting activity between Northern Abidjan and Southern Abidjan, notably at peak traffic hours such as 8:00 and 19:00. While the two parts of the city are joined by two bridges, they both seem to suffer from an insufficient capacity regarding the high traffic congestions during rush hour times. Thus, our results confirm the need for a third bridge linking both parts to reduce the level of traffic congestions.

VI. Conclusion and future works

In this paper, we analyze Call Detail Records of mobile phone users in the city of Abidjan. We introduce a method that allows to distinguish between typical and special calling behaviors of the population. We extract the global mobility flows across the whole city, which we prove to reflect the dynamics of the lifestyle in Abidjan. We believe that our results can help solve important problems in the city. They can be exploited to improve the public transportation services, adapting the paths taken by buses and taxis as well as their number to the mobility flows. They can also be used when considering traffic problems and the road infrastructure, showing where new roads are needed: our results clearly confirm the need for the construction of the third bridge linking the northern and southern parts of the city. Mobile phone services can be ameliorated as well based on our results, by adapting them to the macroscopic movements detected taking into account the geographical locations where people mostly cluster at different times of the day. As for the accuracy of the obtained flows, we remark that it can be improved by considering larger, more complete datasets. Finally, our results are limited to one city in Ivory Coast, a similar study can be achieved for the other areas of the country, also it can be interesting to check the mobility flows between different regions of Ivory Coast when considering the improvement of services in the country as a whole.

References

[1] M. Kim, D. Kotz, and S. Kim. Extracting a mobility model from real user traces. In *IEEE INFOCOM*, 2006.

[2] A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, and C. Mascolo, A tale of many cities: universal patterns in human urban mobility, In *PLoS One 7*, 2012.

[3] F. Girardin, F. Calabrese, F. Dal Fiorre, A. Biderman, C. Ratti, and J. Blat. Uncovering the presence and movements of tourists from user-generated content. In *Intnl Forum on Tourism Statistics*, 2008.

[4] F. Girardin, A. Vaccari, A. Gerber, A. Biderman, and C. Ratti. Towards estimating the presence of visitors from the aggregate mobile phone network activity they generate. In *Intl. Conference on Computers in Urban Planning and Urban Management*, 2009.

[5] M. C. González, C. A. Hidalgo, and A.-L. Barabási. Understanding individual human mobility patterns. *Nature*, 453, June 2008.

[6] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási. Limits of predictability in human mobility. *Science*, 327, February 2010.

[7] S. Isaacman, R. Becker, R. Cceres, S. Kobourov, M. Martonosi, J. Rowland, and A. Varshavsky. Ranges of human mobility in los angeles and new york. In *Eighth IEEE Workshop on Managing Ubiquitous Communications and Services*, 2011.

[8] S. Isaacman, R. Becker, R. Cceres, M. Martonosi, J. Rowland, A. Varshavsky and W. Willinger, Human mobility modeling at metropolitan scales. In the *10th ACM Conference on Mobile Systems, Applications, and Services*, 2012.

[9] R. Pulselli, P. Ramono, C. Ratti, and E. Tiezzi. Computing urban mobile landscapes through monitoring population density based on cellphone chatting. *Int. J. of Design and Nature and Ecodynamics*, 3, 2008.

[10] OpenStreetMap, http://www.openstreetmap.org.

[11] J.W. Tukey, Exploratory Data Analysis (limited preliminary ed.), Reading, MA: Addison-Wesley.