Towards Connecting People, Locations and Real-World Events in a Cellular Network

R. Trestian, P. Shah, H. Nguyen, Q-T. Vien, O. Gemikonakli, B. Barn
School of Science and Technology, Middlesex University, London, UK
{r.trestian, p.shah, h.nguyen, q.vien, o.gemikonakli, b.barn}@mdx.ac.uk

ABSTRACT
The success of personal mobile communication technologies has led an emerging expansion of the telecommunication infrastructure but also to an explosion to mobile broadband data traffic as more and more people completely rely on their mobile devices, either for work or entertainment. The continuously interaction of their mobile devices with the mobile network infrastructure creates digital traces that can be easily logged by the network operators. These digital traces can be further used, apart from billing and resource management, for large-scale population monitoring using mobile traffic analysis. They could be integrated into intelligent systems that could help at detecting exceptional events such as riots, protests or even at disaster preventions with minimal costs and improve people safety and security, or even save lives. In this paper we study the use of fully anonymized and highly aggregate cellular network data, like Call Detail Records (CDRs) to analyze the telecommunication traffic and connect people, locations and events. The results show that by analyzing the CDR data exceptional spatio-temporal patterns of mobile data can be correlated to real-world events. For example, high user network activity was mapped to religious festivals, such as Ramadan, Le Grand Magal de Touba and the Tivaouane Maouloud festival. During the Ramadan period it was noticed that the communication pattern doubled during the night with a slow start during the morning and along the day. Furthermore, a peak increase in the number of voice calls and voice calls duration in the area of Kafoutine was mapped to the Casamance Conflict in the area which resulted in four deaths. Thus, these observations could be further used to develop an intelligent system that detects exceptional events in real-time from CDRs data monitoring. Such system could be used in intelligent transportation management, urban planning, emergency situations, network resource allocation and performance optimization, etc.

Keywords
Mobile Traffic Analysis, Cellular Networks, Human Mobility, Call Detail Records

1. INTRODUCTION
The outstanding progress of the telecommunication industry and the smart mobile computing devices, led to a significant growth in the number of advanced mobile users and their service demands. Users now are expecting uninterrupted, continuous and seamless services that satisfy the Quality of Service (QoS) demands of their applications when connecting to the Internet from any device type and at anytime, either while on the move or stationary. According to Cisco (2015), 66% of the IP traffic is generated from mobile and wireless devices and the Internet traffic will reach 18GB per capita by 2019. To be able to deal with this explosion of mobile broadband data traffic and satisfy their customers' traffic demands, the network operators started exploring the next generation of wireless infrastructure, which includes a high deployment of base stations and overlapping of different radio access technologies (RAT), such as Wireless Local Area Networks (WLAN), Long Term Evolution (LTE), Worldwide Interoperability for Microwave Access (WiMAX), etc. This heterogeneous wireless environment will provide the mobile users with high performance and wide coverage motivating the continuing uptake of the mobility around the world.

By continuously interacting with the mobile network infrastructure a digital signature can be easily recorded from each mobile computing device by the network operators. Thus, every time people interact with the mobile networks or any type of social media platform, they leave behind digital traces that network operators could use for different purposes, such as: billing or network resource management. This data collected from the mobile systems is referred to as Call Details Records (CDRs), which contain information details about every call carried within the cellular network, including information about the location, call duration, call time, and both parties involved in the conversation. The CDR traces have become a powerful tool to analyze human behavior patterns and an increased interest towards making use of CDRs to analyze the human mobility cheaply, frequently and especially at a very large scale has been recorder lately. Various studies have shown that several areas could benefit from understanding human mobility patterns, such as: network resource optimization, mobile computing, transportation systems, urban environment planning, events management, epidemiology, etc.

Within this context, our research questions are: can the CDR data be used to detect exceptional spatio-temporal patterns of the collective human mobile data usage? Can we correlate these exceptional usage patterns to real-world events?

In this work we provide a comprehensive literature review and we explore the use of anonymized CDRs containing data from voice-calls and SMS activities, collected over one year period from more than nine million mobile customers within a cellular network in Senegal. The data is analyzed and the results show that exceptional spatio-temporal patterns of the collective human activity could be identified from fully anonymized and highly aggregate cellular network data, like CDRs, and correlated with real-world events, such as religious festivals (e.g., Ramadan, Le Grand Magal de Touba and the Tivaouane Maouloud festival) and even conflicts (e.g., Casamance Conflict that resulted in four deaths). Apart from analyzing the correlation of the exceptional spatio-temporal patterns with real-world events, the study also analyses the telecommunication traffic flows. The observations from this study could be further used to develop an *intelligent system that detects exceptional events in real-time* from CDRs monitoring. The benefits of such systems could be threefold: (1) the *network operators* could benefit by detecting congested cells and optimize their network resources in advance of an exceptional event, e.g., make use of the Wi-Fi offloading solutions, enabling adaptive bandwidth allocation to their radio cells, etc.; (2) *the society* could benefit from intelligent transportation and urban planning and management; (3) *the individual* could benefit from traffic information and prediction, emergency management. For example, a real-time event detection system could be used in case of emergency situations, such as conflicts or riots protests which could be more efficiently handled if detected and handled on time.

2. RELATED WORKS

The use of user-generated traffic from mobile communications networks as a powerful tool to analyze human behavior, as well as mobile traffic analysis have become an extensive academic research area which is rapidly emerging over a wide range of disciplines. Moreover, considering the fact that billions of people around the world own at least one mobile device, the digital traces collected from the mobile communications networks could help study different aspects of human mobility and their interactions on a large scale. A comprehensive survey on the use of large-scale mobile traffic analysis in multidisciplinary activities is provided by Naboulsi et al. (2016). Saramäki et al. (2015) look into the use of mobile phone traffic datasets for extracting social graphs. Another extensive review on analysis of mobile phone datasets is provided by Blondel et al. (2015). This section provides a survey on user-generated traffic from mobile communications networks used to understand different aspects of human movement and their interaction. The existing solutions are divided into three main categories: (1) universal law for human mobility; (2) urban planning and traffic forecast; and (3) localization and mobility patterns.

2.1 Universal Law for Human Mobility

Several works tried to define some basic laws governing the human motion from CDR data in order to understand the human mobility patterns. In this regard, Gonzalez et al. (2008) tracked the position of 100,000 anonymized mobile phone users over a six month period and showed that individuals follow simple reproducible patterns despite the diversity of their travel history. As the users always return to several of their highly frequented locations such as home or work, significant regularity can be identified in their trajectories. Empirical data on human mobility was also used by Song et al. (2010a) to show that the predictions provided by the continuous time random walk (CTRW) models conflict with the empirical results. Two datasets were used for this study: (1) the CDRs of 3 million anonymized mobile phone users over a one year period; and (2) the hourly location record of 1,000 anonymized users who signed up for a location-based service, over a two week period. The authors look into the limitations of the traditional random walk models and show that humans tend to return to the highly visited locations, like home or workplace, which is not considered by the random walking models. To this extent, the authors propose a new individual mobility model that takes into consideration the exploration and the preferential return of the individual. The exploration is defined as the probability of the individual moving to a new location and the preferential return refers to the probability of the individual returning to one of the previously visited locations. Another study on the limitations of predictability in human mobility was conducted by Song et al.(2010b) using the CDRs of 50,000 anonymized mobile phone users over a three months period. The study shows that most of the individuals can be localized within a specific neighborhood with only few users traveling widely. Moreover, there is a 93% chance that the human location could be predicted regardless of how far the person travels within the preferred locations. According to Palchykov et al. (2014) human mobility can be predicted by using a simple model based on the frequency of the mobile phone calls between two locations and their geographical distance. The study was conducted on the data provided by Orange for Ivory Coast, consisting of CDRs from 50,000 anonymized mobile phone users collected over a 150 days period. Three different models were tested: the gravity model, the communication model based on the number of calls between two locations, and a modified version of the radiation model. The results showed that out of the three models the communication model is the most accurate in the given context. Table 1 presents a summary of the solutions focused on defining the universal law for human mobility.

Ref.	**Application**	**Number of Users**	**Period Covered**	**Location**	**Data Type**
(Gonzalez et al., 2008)	defining individual human mobility	-100,000 anonymized mobile phone users -206 anonymized mobile phone users	- 6 moths for the 100,000 users and one week for the 206 users	- not mentioned	- CDRs – identity of the closest communication tower when the user initiates or receives a call or a text message
(Song et al., 2010a)	defining individual human mobility	-3 million anonymized mobile phone users	- one year for the 3 million users	- not mentioned	- CDRs – identity of the closest communication tower when the user initiates or receives a call or a text

		-1,000 users who signed up for a location based service	- two week for the 1,000 users		message
(Song et al., 2010b)	limitations of predictability in human mobility	-50,000 anonymized mobile phone users	- 3months	- not mentioned	- CDRs – identity of the closest communication tower when the user initiates or receives a call or a text message
(Palchykov et al., 2014)	human mobility	-50,000 anonymized mobile phone users	-150 days from December 1, 2011 to April 28, 2012	- Ivory Coast	- CDRs – identity of the closest communication tower when the user initiates or receives a call or a text message

Table 1. Universal Law for Human Mobility- Solutions Summary

2.2 Urban Planning and Traffic Forecast

In terms of the use of cellular network data for urban planning and real-time traffic forecast Isaacman et al. (2010) looked at the mobility patterns for two cities: Los Angeles and New York. It was observed that the people living in Los Angeles tend to travel on a regular basis, two times farther than people in New York and that New Yorkers tend to take two to six times longer trips than Angelenos. These observations could help at investigating the environmental impact of daily commutes.

By using a combination of both human mobility data among different regions and the points of interests (POI), Yuan et al. (2012) propose a framework referred to as DRoF, Discovers Regions of different Functions (e.g., educational areas, entertainment areas, historic oriented areas, etc.) in a city. Large-scale and real-world datasets consisting of two POI datasets of Beijing collected in 2010 and 2011, and two 3 month GPS data used to represent human mobility, generated by 12,000 taxi cabs in Beijing in 2010 and 2011 were used to evaluate the proposed framework. Compared to other two baseline methods that make use of either POIs or mobility data only, the proposed framework offers a better solution.

A real-time urban monitoring platform is introduced by Calabrese et al. (2011) to provide a visualization map of the vehicular traffic status and the pedestrians' movement. The platform makes use of a broad range of datasets and was tested for the city of Rome in Italy. The visualization tool gives a qualitative understanding of how the mobile phone data and vehicle real-time location data could be used to provide valuable services in the context of urban planning and tourist management.

Di Lorenzo et al. (2011) propose a method for evaluating the human spatio-temporal activity patterns by combining the use of people trajectories and geographical preferences. Two datasets were collected over 4 months from one million unique devices and consist of the individual human trajectories extracted from anonymous mobile phone traces and the geographical features of places in the area, such as land use of the state of Massachusetts. The authors have identified four distinct patterns that could be mapped to a specific number of kilometers traveled in the day and that could be further integrated into activity-based transportation models. A summary of the above solutions is presented in Table 2.

Ref.	**Application**	**Number of Users**	**Period Covered**	**Location**	**Data Type**
(Isaacman et al., 2010)	-environmental impact of daily commutes	-hundreds of thousands of anonymized identifiers	-62 days from March 15, 2009 to May 15, 2009	Los Angeles and New York	-CDRs - incoming voice calls, outgoing voice calls, and data traffic exchange
(Yuan et al., 2012)	- urban planning	- 12,000 taxi cabs for GPS data	-year 2010 and 2011 for the POI data sets -two 3 month GPS data in 2010 and 2011	Beijing, China	-POI - coordinates and category like restaurants and shopping malls -GPS trajectory datasets representing human mobility
(Calabrese et al., 2011)	-real-time urban monitoring	- 30,000 calls -7268 buses for GPS data -43 taxies for GPS data	-weekday	Rome, Italy	-mobile phone position information, call in progress, SMS sending, handover, etc. -GPS data -real time traffic noise from sensor networks
(Di Lorenzo et al., 2011)	-human spatio-temporal activity patterns	-one million unique devices	- 4 months	Massachusetts	-individual human trajectories extracted from anonymous mobile phone traces -geographical features of places

Table 2. Urban Planning and Traffic Forecast – Solutions Summary

2.3 Localization and Mobility Patterns

Trestian et al. (2009) conducted one of the first studies to show evidence of geographic correlation between users' interests within a cellular network. The user interests were categorized into six groups based on the type of service they were using: mail, social networking, trading, music, news, and dating. The authors looked into correlating the users' interests with their location, e.g., home or work. The authors showed that the location affects the services the users are accessing and that they tend to spend a significant fraction of their time in their top three locations only. Isaacmanc et al. (2011) conducted another study trying to identify important locations in humans' live from mobile data traces. Several algorithms based on logistic regression were proposed to identify important places from CDRs and then used to apply semantic meaning to these important locations, namely Home and Work. The results show that the proposed algorithms could be used to identify the key locations with median errors under one mile.

A dynamic profile-based paging/location management technique to increase the efficiency of the location management process within a cellular network is proposed by Zang et al. (2007). The data used in the study was collected from hundreds of thousands of users over a one month period in three locations: Manhattan, Philadelphia, and Brisbane. The study shows that the proposed solution increases the average paging success rate across voice/data/SMS calls above 90% in Brisbane and 85% in Manhattan. Additionally the signaling overhead could be reduced by up to 90% at a cost of a small increase in paging delay.

The impact of temporal factors on the randomness and the size of mobility and the spatial distribution was investigated by Motahari et al. (2012). CDR data was collected from several thousands of users in the San Francisco area. The authors studied how temporal factors (e.g., the day of the week and the time of the day) impact four characteristics of human mobility: location entropy, radius of gyration, step size, and spatial probability distribution of user locations. The authors found that the spatial distribution is most concentrated during work hours and most scattered on the weekend and a different pattern is observed during the non-working hours of weekdays where the spatial distribution is concentrated around home, work and the commute path. The study shows that by considering the temporal factors, the location predictions mechanisms can improve their accuracy by 15% compared to the case without the temporal factors. Table 3 presents a summary of the above solutions.

Ref.	Application	Number of Users	Period Covered	Location	Data Type
(Trestian et al. 2009)	-correlation between people's application interests and mobility properties	-281,394 users	-seven days	large metropolitan area of 1,900 square miles (approx. 5,000 square kilometers)	-packet data session details containing: local timestamp, anonymized user identifier, anonymized IP-Address, correlation identifier, base station identifier, the URL accessed
(Isaacman et al., 2011)	-identifying important places in people's lives	-hundreds of thousands of phones	-78 consecutive days from November 15, 2009 to January 31, 2010	Los Angeles New York	-CDR containing information about voice calls and text messages
(Zang et al., 2007)	-location management	-1,061,000 users in Manhattan -543,000 users in Philadelphia -404,000 users in Brisbane	- one month, from February 2 to February 28, 2006	Manhattan, Philadelphia, Brisbane	-per call measurement data, containing: call starting time, call duration, initial cell, final cell, service type, call direction, number of pages, etc.
(Motahari et al., 2012)	-mobility characteristics and location prediction	- several thousands of users	- 4 days, 16 days, and 3 months	San Francisco	-CDRs including voice, SMS, and data sessions

Table 3. Localization and Mobility Patterns – Solutions Summary

2.4 Discussions

The massive increase in the amount of data generated by smart mobile computing devices, led to the appearance of new research domains across computing and social science which examine the issues in behavioral and social science from the Big Data perspective by making use of the mobile phone data sets collected by the network operators.

This section provides a comprehensive survey of the current research on this topic and classifies the existing solutions in three main groups: (1) universal law for human mobility; (2) urban planning and traffic forecast and (3) localization and mobility patterns.

The use of large-scale user-generated data traffic from mobile communications networks has great potential in several research directions, from the statistical modeling of human mobility and the definition of universal laws (Gonzalez et al., 2008) to real-time traffic monitoring conditions that would help in transportation planning and management (Isaacman et al., 2010). In this regard, Gonzalez et al. (2008) found that humans follow simple reproducible patterns, despite the diversity of their travel history. Moreover, Song et al. (2010a) showed that humans tend to return to the highly visited locations (e.g., home, workplace) in contrast to the well-known random walking models. In a different study (Song et al., 2010b), the authors showed that the human location could be predicted with a 93% probability, regardless of how far the person travels within the preferred locations. However all these human mobility prediction models are highly depended on the frequency of the mobile phone calls the mobile user is generated.

In terms or urban planning, Isaacman et al. (2010) investigated the environmental impact of daily commutes in two different cities, and found that there are significant differences in mobility patterns between different human populations. By making use of the mobile equipment location-based monitoring together with other type of real-time information, like taxis and buses positions could help developing a real-time urban traffic monitoring control system (Calabrese et al., 2011).

The use of CDRs could also help in terms of localization and mobility patterns. Motahari et al. (2012) showed that by considering the impact of temporal factors on mobility patterns, the accuracy of the prediction mechanisms could go up by 15%. Whereas, Isaacman et al. (2011) showed that by using temporally sparse and spatially coarse location information from CDRs, the users' key locations can be identified with median errors under one mile.

Apart from the benefits brought to the three categories identified above, the CDRs and big data analysis has also been used in the area of health management. For example, Wesolowski et al. (2012) and Tatem et al. (2014) made use of CDR data to map the malaria outbreaks in Kenya and Namibia, respectively. Whereas, Frias-Martinez et al. (2011) used the CDR data to monitor the public response to government health warnings during Mexico's swine-flu epidemic in 2009 and showed that by using an agent-based system to model the virus spreading, the peak number of individuals infected by the virus could be reduced by 10%, if government mandates would impose restricted mobility.

The network operators could also benefit from the use of big data analysis by understanding their subscribers churn. Han et al. (2013) made use of the CDRs and the tariff plans and showed that the churning probability increases by 3% if there is a friend who churns.

All these studies have shown that understanding the humans' mobility patterns could be a crucial component in several areas, such as: network optimization opportunities for cellular network operators in handling the explosive growth in traffic observed from CDRs; transportation planning and management, modeling commuting flows, content delivery services and context-aware applications, health management, etc. In our previous work (Trestian et al., 2016) we used the CDR data collected from 50,000 randomly selected customers from a cellular network in Ivory Coast over a 150 days period to identify the exceptional spatio-temporal patterns of the collective human activity and correlate these 'anomalies' with real-world events (e.g., parades, public concerts, soccer match, New Year's Eve, etc.). However these studies are limited by several factors: the dataset size, the location of the data collection, or the interaction frequency of a particular individual with the network.

In this work we make use of CDR data collected over one year from more than nine million Orange customers in Senegal with up to 300,000 randomly sampled customers per dataset in the case of high resolution individual movement data. Apart from looking into correlating the exceptional spatio-temporal patterns with real-world events, the study also analyses the telecommunication traffic flows.

3. DATA COLLECTION METHODOLOGY AND CHARACTERISTICS

3.1 Data Collection and Preprocessing

In this paper we use the anonymous CDR data provided by the Orange Group within the Orange *Data for Development* (D4D) challenge. The CDRs are anonymized phone calls and SMS exchanges between more than nine million Orange customers in Senegal. The anonymized CDRs were collected from a random set of cellular phones over one year, between January 1, 2013 and December 31, 2013. The original dataset was preprocessed and only the customers that meet the following criteria were retained: (1) customers that have more than 75% days with interactions for a given period; and (2) users having less than 1000 interactions on average per week. In order to maintain the privacy and avoid the potential risk of identifying the customers, the dataset was further divided in three datasets: (1) Antenna-to-Antenna Communication that covers a longer period of time and provides site-to-site traffic aggregated across all the users meeting the two criteria; (2)

Individual Trajectories –High Spatial Resolution Data that provides precise spatially and temporally information but is limited in the time span; and (3) Individual Trajectories – Long Term Data that provides geographically aggregated data. Thus the three datasets cannot be interlinked based on the customers' identifiers.
The territorial expanse of the dataset on Senegal is illustrated in Figure 1[1]. Senegal is located in West Africa having an area of 197,000 square kilometers and an estimated population of 13.5 million inhabitants. The capital is Dakar, however the country is subdivided in 14 administrative regions, each region having a regional capital. The country telecommunications sector is dominated by mobile telephony with Orange, owned by Sonatel, being one of the leaders in the market, recording two thirds of the cellular market. For the purpose of this study, there are three sets of data provided by Orange Group and described in the following sections.

Figure 1. Territorial expanse of the dataset – Senegal.[1]

3.2 Dataset 1: Antenna-to-Antenna Communication

The first dataset contains the aggregated number of calls as well as the calls durations within one hour, between any antennas pair from all the customers in the original dataset that satisfy the two criteria. The dataset was stored in 12 files each corresponding to a one month interval. All the datasets are provided in Comma Separated Values (CSV) file format. For the Antenna-to-Antenna dataset each line stores information about the date, time, originating antenna, terminating antenna, number of voice calls, and the duration of the voice calls in minutes for a given hour. A second set of 12 files is provided which contain the monthly aggregated number of SMS exchanged between any antenna pair within one hour.

3.3 Dataset 2: Individual Trajectories – High Spatial Resolution Data

The second dataset is stored in 24 CSV files and is split into consecutive two-week periods providing high resolution individual movement trajectories of 300,000 randomly sampled customers for each period. Each line in the file contains information about the customer identification number, the connection date and time and the antenna identification number they are connected to. However, to protect customers' privacy, for every two-week time period a new sample of 300,000 customers are selected and new random customer identifiers are generated. Thus, a single customer cannot be tracked across the two-week period datasets.

3.4 Dataset 3: Individual Trajectories – Long Term Data

The third dataset contains the long term, low spatial resolution trajectories of the 146,352 randomly selected customers that meet the two criteria mentioned before on a yearly basis. The low spatial resolution is obtained by replacing the antennas identifiers with the arrondissement identifier of the antenna the customer is connected to. Thus, the dataset was stored in 12 CSV files each corresponding to a one month interval which contain the user id, timestamp and the arrondissement identifier. Senegal has a total number of 123 arrondissement administrative regions. Figure 2 illustrates the arrondissements and their identifiers along with the Orange antennas locations as provided in the datasets. There are a total number of 1666 antennas. Figure 2 also highlights specific regions like Dakar, Pikine, Kaolack, Saint-Louis etc., where some identifiers are highlighted later on in the paper with respect to connecting people and real world events.

---

[1] https://maps.google.com/

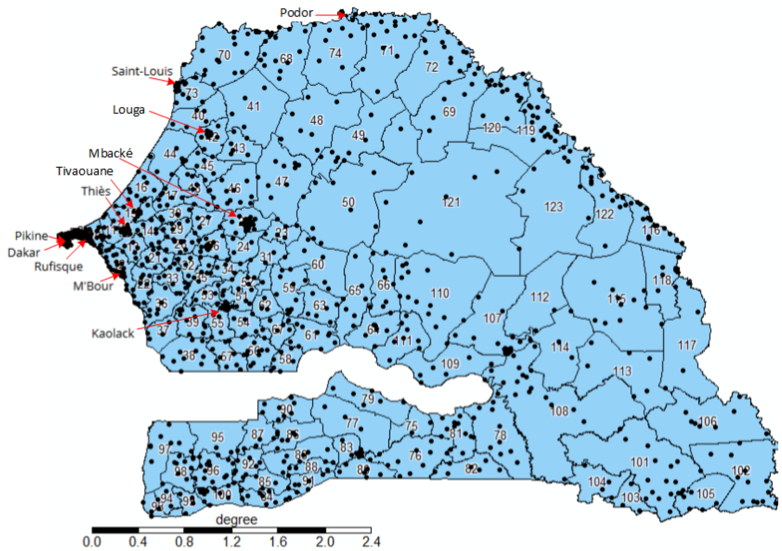

Figure 2. Senegal Arrondissement and Orange antennas location.

Each file in this dataset contains information on customer identification number, the connection date and time and the arrondissement identifier that contains the antenna the user is connected to.

3.5 Limitations of the Datasets

Even though the Call Detail Records represent a good source of location information they have several significant limitations:

- they are generated only when the mobile device is engaged in a voice call or exchanges text messages, thus no information about application usage type (voice/text/data) is available.
- the location granularity is at cell tower level or arrondissement, no information about the exact user location is provided.
- no information about the individual call duration is provided.

4. TELECOMMUNICATION FLOWS

This section aims at analyzing the characteristics of the telecommunication traffic flows from Dataset 1, in terms of number of voice calls, voice calls duration and SMS exchange.

4.1 Antennas Density vs. Population Density

Apart from the 14 regions, the country is further subdivided by 45 departments and 123 Arrondissements. We made use of the information provided by GeoHive[2] about the population census estimates for 2013. Based on this information, we projected the map of Senegal so that the 45 regions of the country are represented by a color proportional to its population as illustrated in Figure 3. Table 4 lists the main cities with the higher population sizes.

The placement of the antennas within a telecommunication network represents an important decision for any mobile service provider. This will determine how many people will be able to access the network and the quality of the calls. A crucial factor is given by the population density. Wherever there are more people, the density of the antennas should be higher. However the rural population is very important as well, as the access to the mobile communication network could impact their development, e.g., access to real-time agricultural services, access to education, etc.

City	Population (2013-11-19) Census
Pikine	1,101,859
Dakar	1,081,222
Mbacké	879,506
M'Bour	641,068
Thies	636,088
Kaolack	466,421

[2] http://www.geohive.com/cntry/senegal.aspx

Rufisque	462,741
Tivaouane	431,956
Podor	356,408
Louga	354,989

Table 4. Senegal Main Cities and Population

As illustrated in Figure 2, we mapped the 1666 antennas coordinates on a standard latitude-longitude projection, represented by the black dots on the map. It can be noticed that the antennas are spatially very unevenly distributed with a dense distribution around Dakar region, where the capital is located. However towards the center of the country the number of antennas is reduced. It can be seen that the antennas distribution correlates with the population density within the country. Thus in the most populous regions, there is an increased number of base station.

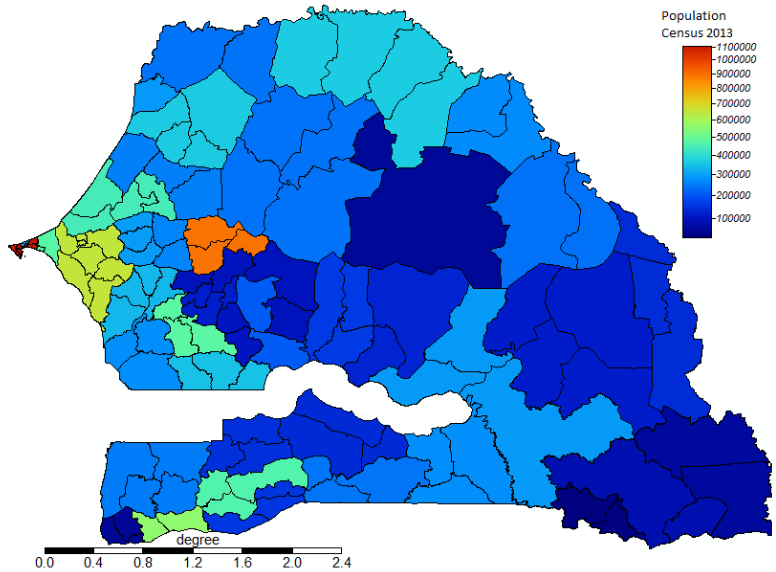

Figure 3. Senegal Population Census as per 2013-11-19.

4.2 Population Density vs. Telecommunication Traffic

The first dataset provided by Orange contains information about the traffic exchanged between every antenna pair. It provides information about the number of voice calls, voice calls duration, and number of SMS exchanged within one hour. In order to study if there is any correlation between the population size and the amount of traffic exchanged within the communication network, we aggregated the antenna level communication at a department level, and differentiated between incoming and outgoing traffic. Figure 4, 5 and 6 illustrate the population per department vs. the telecommunications traffic exchanged, in and out for each department along with the corresponding smoothed fit curve with confidence region. The information is aggregated over the full year of data.

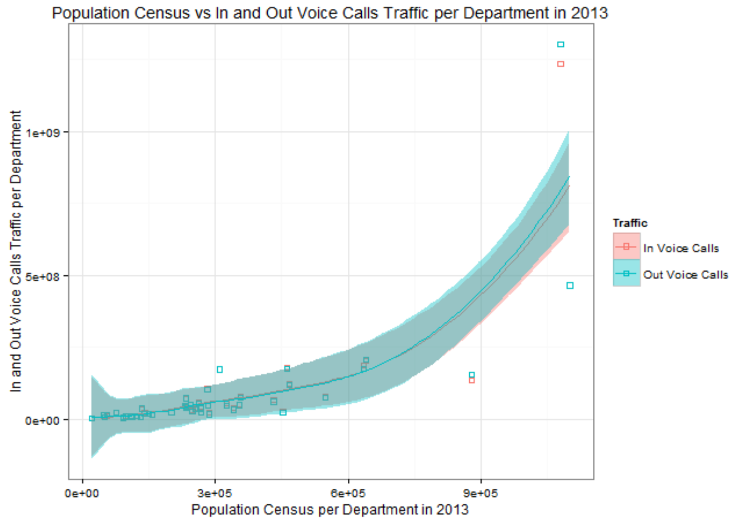

Figure 4. Population vs. In and Out Voice Calls Traffic per Dept.

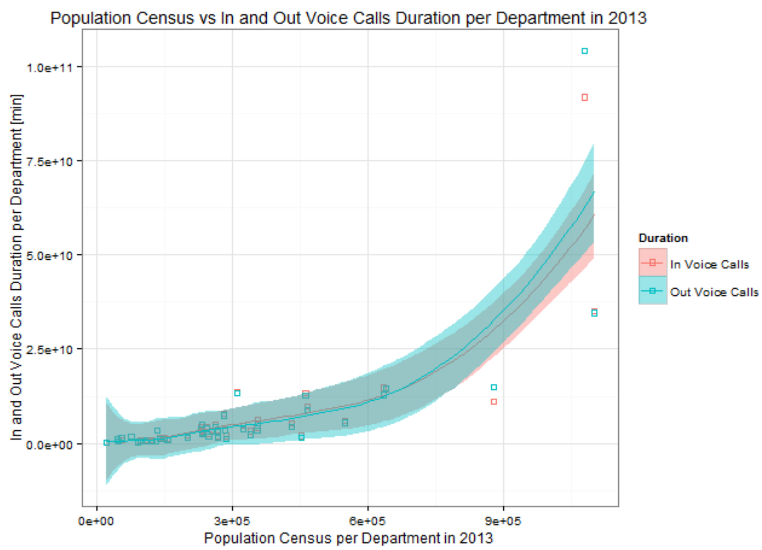

Figure 5. Population vs. In and Out Voice Calls Duration Traffic per Dept.

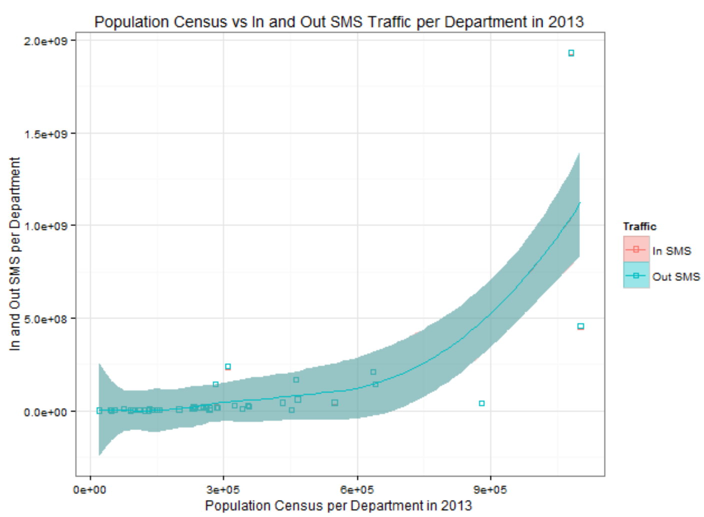

Figure 6 Population vs. In and Out SMS Traffic per Dept.

It can be noticed that along with the increase in the population size, the intensity of the traffic in that location is also increasing.

4.3 Inter-City Telecommunication Flows

After analyzing how the telecommunication traffic flows into and out of the 45 departments from Senegal, scale with the population size, we are interested in looking into the inter-antennas/inter- arrondissement telecommunication flows. Figure 7 illustrates the total traffic intensity for each antenna, in terms of the total number of in and out voice calls, with their size and color representation reported to the load intensity. Thus, heavily loaded antennas are represented by wider points and higher intensity color. It can be seen that the traffic load is correlated with the population size represented in Figure 3 and is mainly located in the main city areas as identified in Figure 2. In this context, an adaptive power allocation for the antennas could be employed so that, in the conditions where some antennas present high traffic load and dense user activity, their transmission power could be increased to improve the resource allocation capacity and the calls Quality of Service (QoS). Whereas for the antennas where the demand is low and there is light traffic load the transmission power could be reduced. In this way the network operator could also save on energy consumption.

When analyzing the inter-arrondissement telecommunication flows we compute the traffic matrix by associating the antennas to their corresponding arrondissement and we compute the total number of calls, total duration of calls, and total SMS exchanged between any two pairs of arrondissements. The logarithmic representations of the matrices are listed in Figure 8, 9 10. It can be noticed that for all three matrices the highest values are mostly concentrated along the diagonal and the most intensity in the origin. If we look at the previous analysis about the population and the main cities presented within the Arrondissements, as illustrated in Figure 2, we can notice that the Arrondissement IDs near the origin are the ones within the region of the capital and where most of the antennas are deployed.

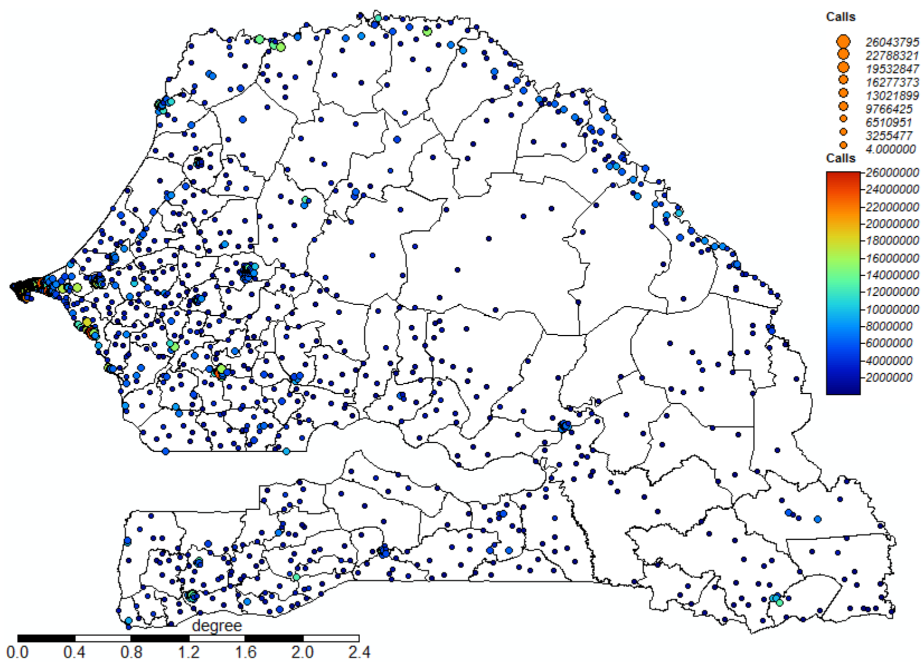

Figure 7 Antennas Total Voice Traffic over the Year.

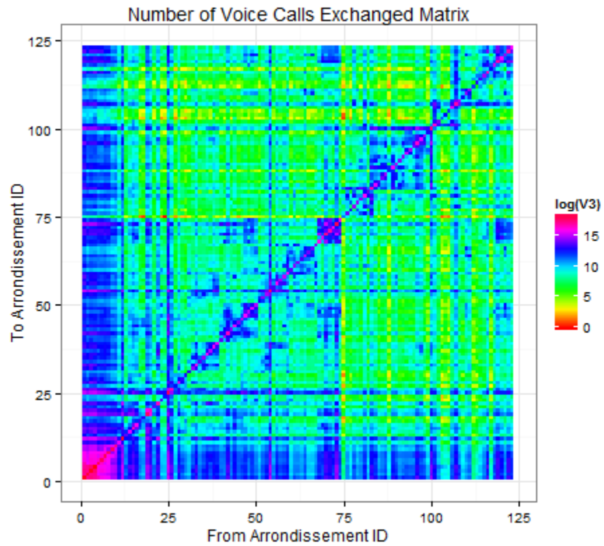

Figure 8 Inter-Arrondissement Voice Calls.

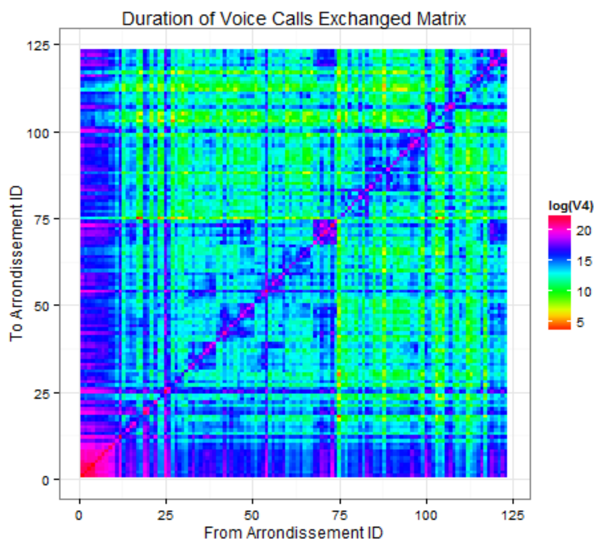

Figure 9 Inter-Arrondissement Voice Calls Duration.

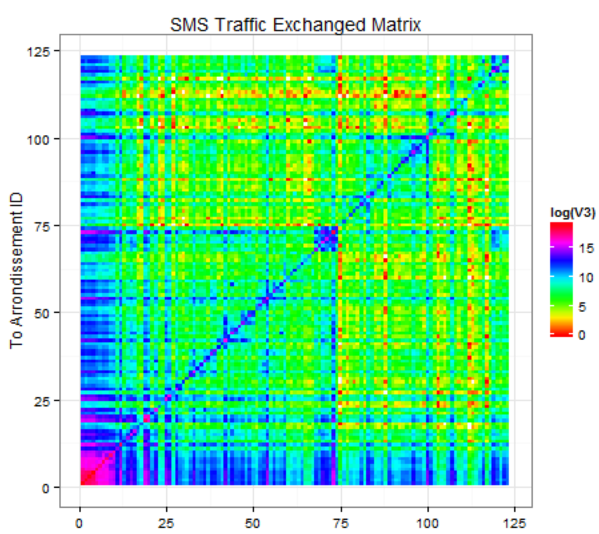

Figure 10 Inter-Arrondissement SMS Exchange.

Thus, from the Figures above we can notice that most of the communication is done with the Arrondissements near the capital, and within the same Arrondissements. Communication between arrondissements that are both far from the capital is highly uncommon. We can also notice that people are more likely to call than send an SMS, if we compare the voice matrices with the SMS traffic exchanged illustrated in Figure 10.

Looking into the inter-antenna traffic exchange for a more fine-grained view, Figure 11, 12, and 13 present the voice calls exchanged, the voice calls duration and SMS exchanged matrices. It is noticeable also at a fine-grained level that the communication is mainly happening within the same antenna, as given by the high values along the diagonal. It is also noticeable that the communication near the origin is more intense. This is because the antennas near the origin are within or very close to the Dakar region. It can also be noticed in terms of voice calls duration, the communication is more intense when calling customers located near the capital. In terms of SMS, it is visible that SMS exchange between antennas far away from the capital region is highly uncommon. This shows that calls are more common than SMS, as the calls matrices are visible denser than the SMS matrix.

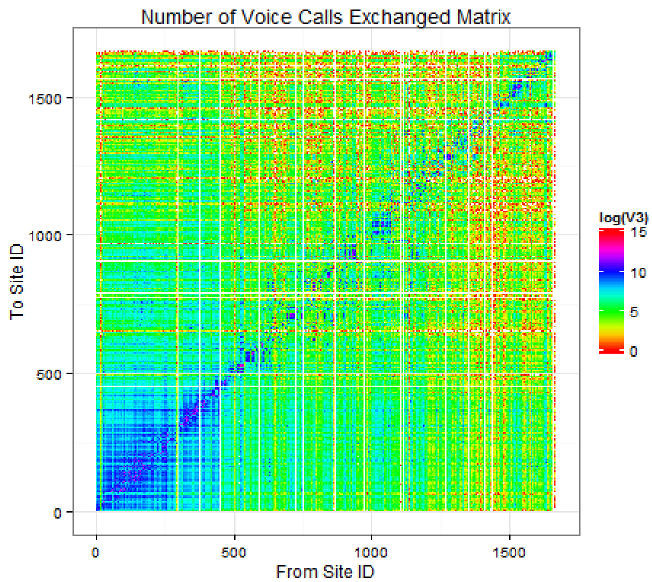

Figure 11 Inter-Antennas Voice Calls.

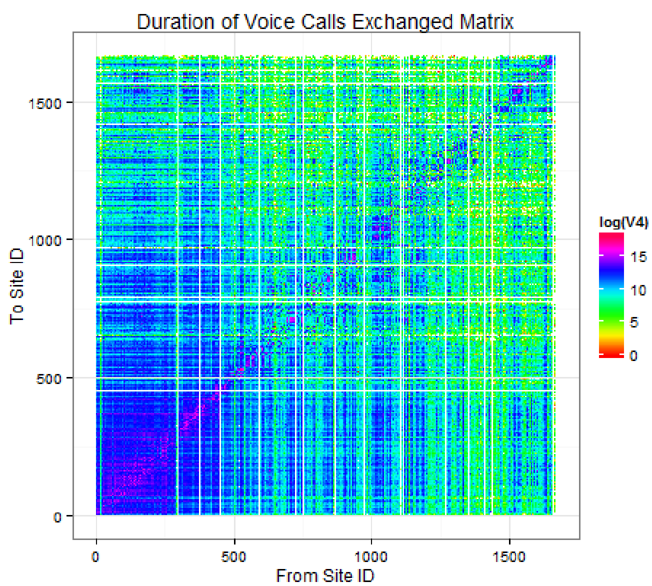

Figure 12 Inter-Antennas Voice Calls Duration.

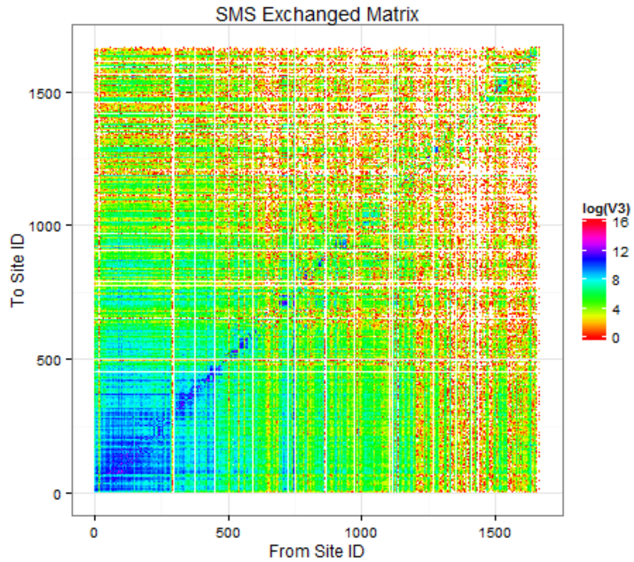

Figure 13 Inter-Antennas SMS Exchanged.

As null values are indicated using the white color, the white lines illustrated on the above matrices indicate that some antennas are not generating any kind of traffic over the full year. These findings are of importance for network operators when considering the deployment or development of their network for cost saving and efficient use of resources.

4.4 Distance vs. Telecommunication Traffic

Several studies (Krings et al., 2009; Kung et al. 2014) have stated that another parameter that influences the communication intensity between cities is the distance. In this section we look at the impact of the distance between antenna pairs and the telecommunication traffic exchange between antennas. To this extent, we have aggregated the antenna to antenna communication over the full year. Using the information provided about the antennas location, we make use of the latitude and longitude coordinates of each antenna and we calculate the distance in kilometers, between them using the haversine formula. Approximatively, a number of 2.4 million antenna-to-antenna interactions were found and the results are plotted in Figure 14, 15, and 16 indicating the distance vs. total number of voice calls, distance vs. total duration of voice calls, and distance vs. SMS exchanged, respectively. On each graph, the corresponding smoothed fit curve is plotted. It can be seen that the telecommunication exchanged traffic is mainly happening between antenna pairs close located, and as the distance increases the communication intensity is decreasing.

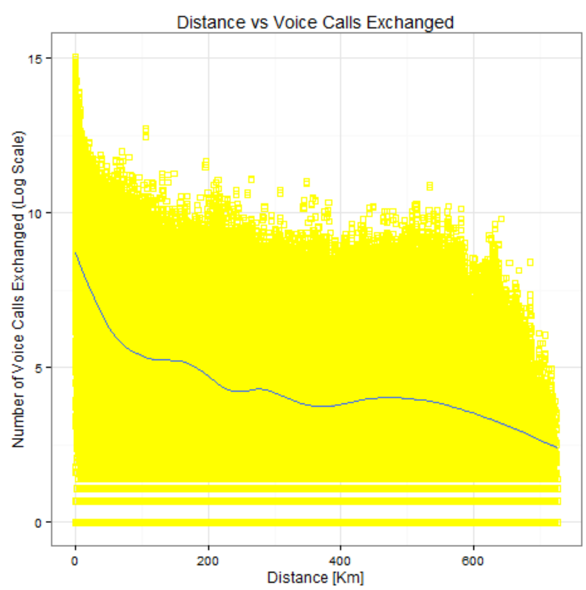

Figure 14 Distance vs. Voice Calls Exchanged.

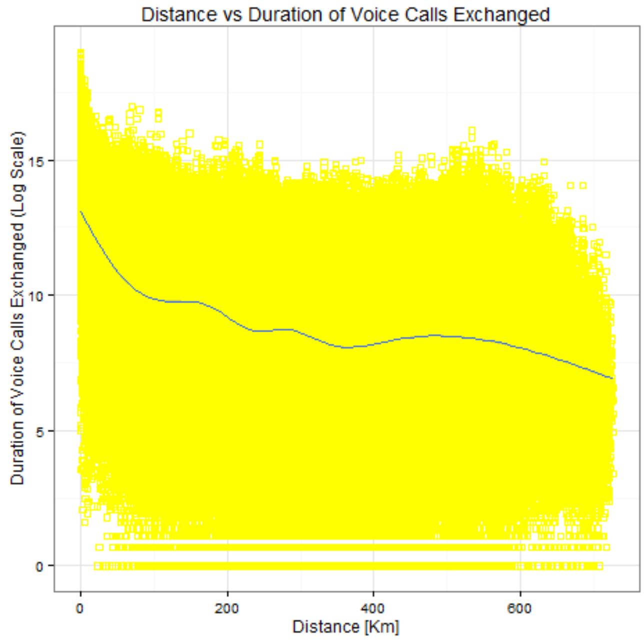

Figure 15 Distance vs. Voice Calls Duration Exchanged

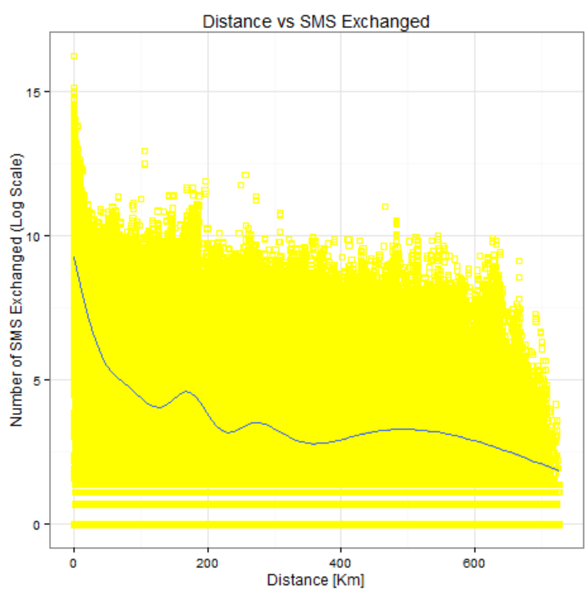

Figure 16 Distance vs. SMS Exchanged

4.5 Symmetry of the Telecommunication Traffic

In this section we analyze the symmetry of the telecommunication traffic at the arrondissement and antenna level. We look into how the communication flows into and out of the arrondissement/antenna. In Section 4.2 we have analyzed how the in and out traffic intensity scales with the population size. We have seen that the traffic is more intense in areas where the population is higher. Here we aggregated the number of voice calls, the total voice calls duration and the SMS exchanged in and out for each arrondissement and each antenna. The results are plotted in Figure 17, 18, and 19 for the arrondissements and Figure 20, 21, and 22 for the antennas. It can be visible that the incoming and outgoing communications are highly symmetric for both situations the arrondissements and the antennas. Thus the calls or SMS in one direction always find a match in the opposite direction.

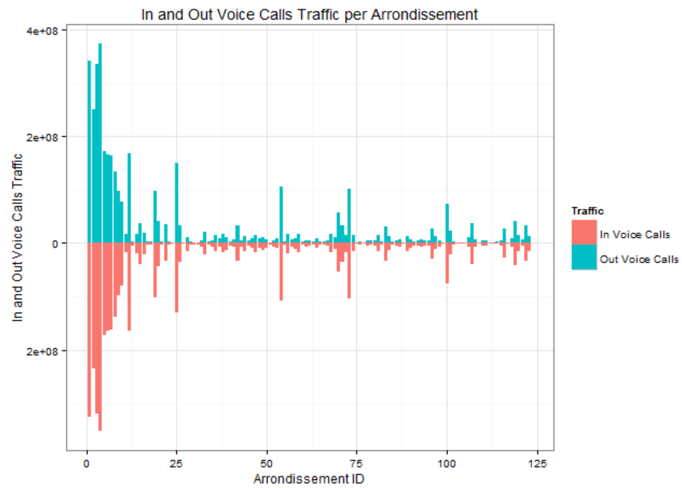

Figure 17 Voice Calls Traffic Symmetry per Arrondissement

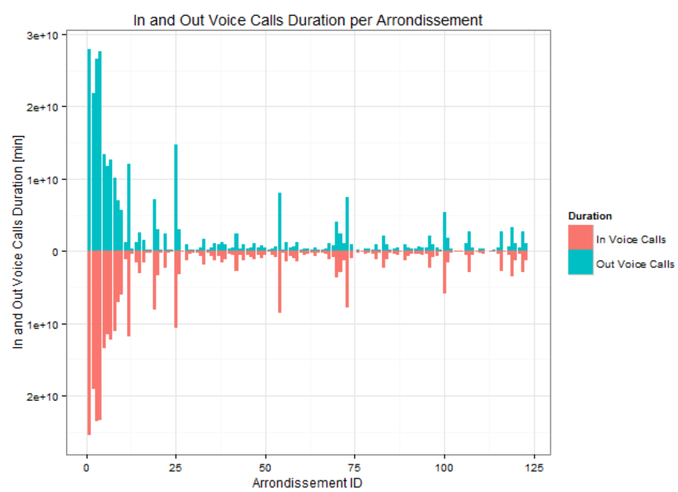

Figure 18 Voice Calls Duration Traffic Symmetry per Arrondissement

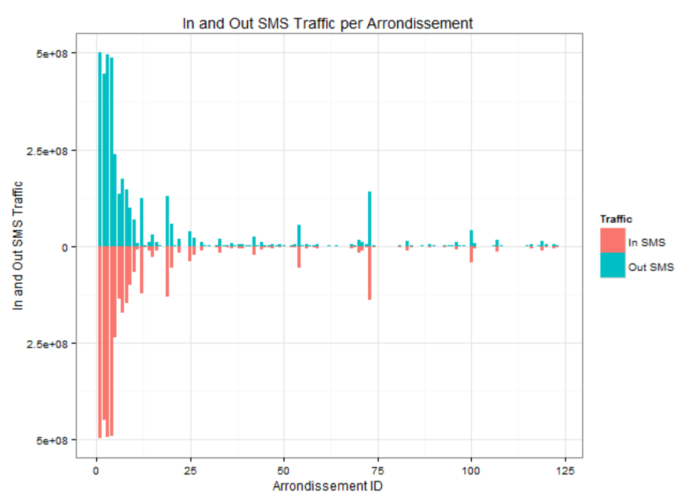

Figure 19 SMS Traffic Symmetry per Arrondissement

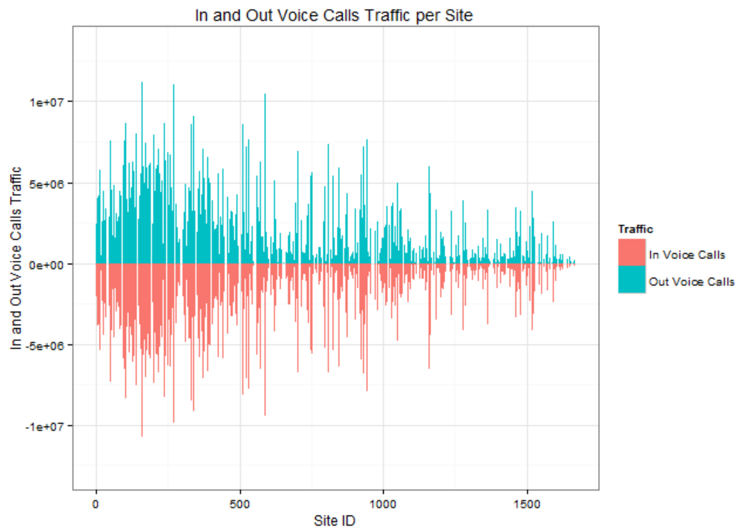

Figure 20 Voice Calls Traffic Symmetry per Antenna

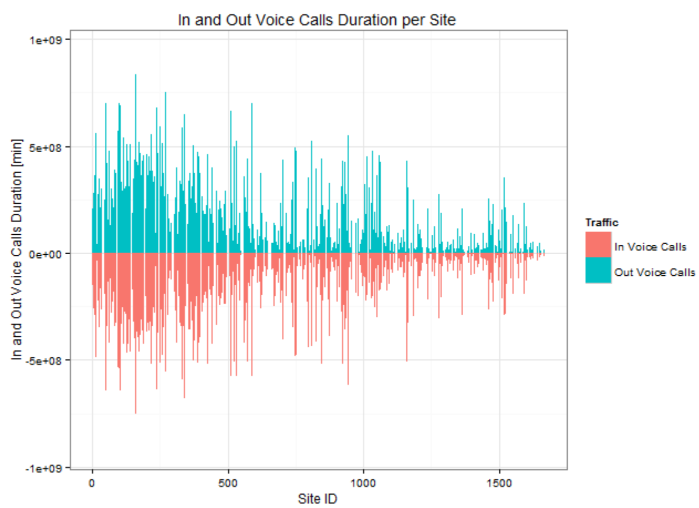

Figure 21 Voice Calls Duration Traffic Symmetry per Antenna

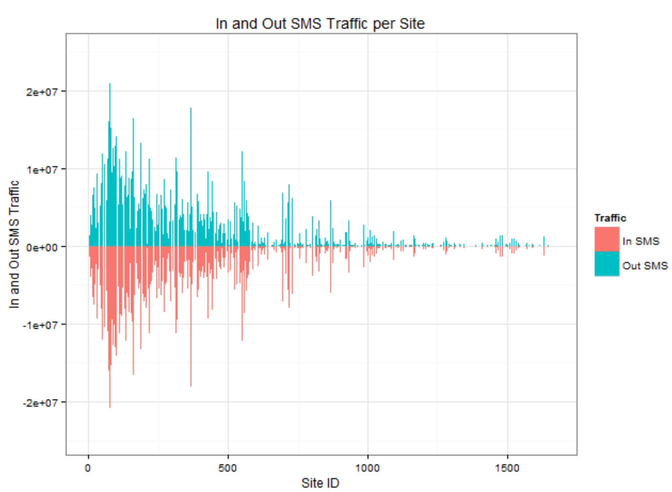

Figure 22 SMS Traffic Symmetry per Antenna

Figure 23 represents the voice in and out traffic flows between the 14 regions of Senegal. The source and destination of the voice traffic flows are represented by the circle's segments, where nearby regions are positioned close to each other. The width of the link also indicates the size of the traffic flow. It can be noticed again that most of the traffic is happening between the same region, and Dakar region occupies almost half of the traffic. Moreover the traffic is higher between the regions that are close to each other, as previously observed.

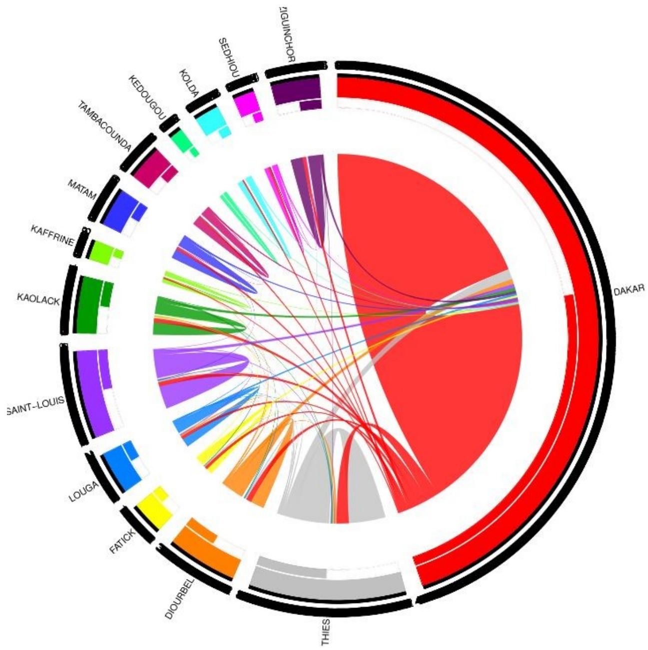

Figure 23 Traffic Symmetry between Regions

5. IDENTIFYING EVENTS IN DATA TRAFFIC

By analyzing the traffic exchange from dataset 1, we are able to identify several anomalies in the traffic.

5.1 Dead Antennas

The dataset should contain information from 1666 antennas spread around the country. However when analyzing the aggregated traffic exchanged over the full year in order to create the traffic matrices introduced in Section 4.3, we have noticed that several antennas do not generate any kind of traffic, indicated by the white lines in Figures 11-13. The reason could be that the information about them is missing from the datasets or they are dead antennas, meaning that they are not working. The dead antennas identified from dataset 1 are illustrated in Figure 24. An interesting fact is that 52 antennas (red points on Figure 24) out of 1666 are not recorded in the dataset to generate any kind of in or out voice nor SMS traffic, whereas 2 antennas (blue points on Figure 24) generate a very small amount of voice traffic, 4 and 64voice calls, but no SMS traffic.

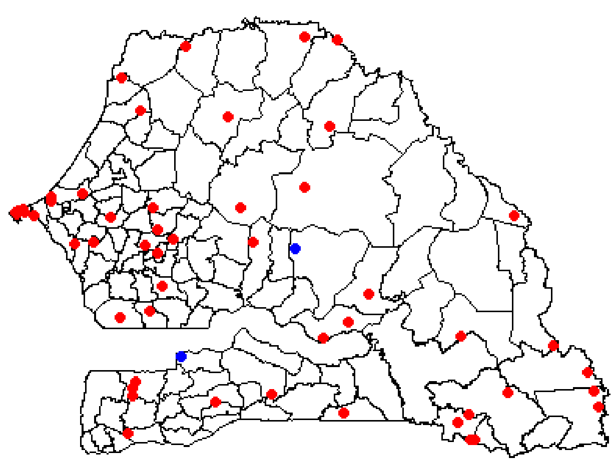

Figure 24 Dead Antennas

5.2 Anomalies in the Telecommunication Traffic

In order to identify anomalies in the telecommunication traffic, we first plot the aggregated total number of voice calls by month along with the median, the quartiles, and the maximum and minimum values, as illustrated in Figure 25. It can be seen that there is a very low value in the voice calls in March, several drops in traffic are registered in June, July and September, whereas peaks in traffic are registered in August and October.

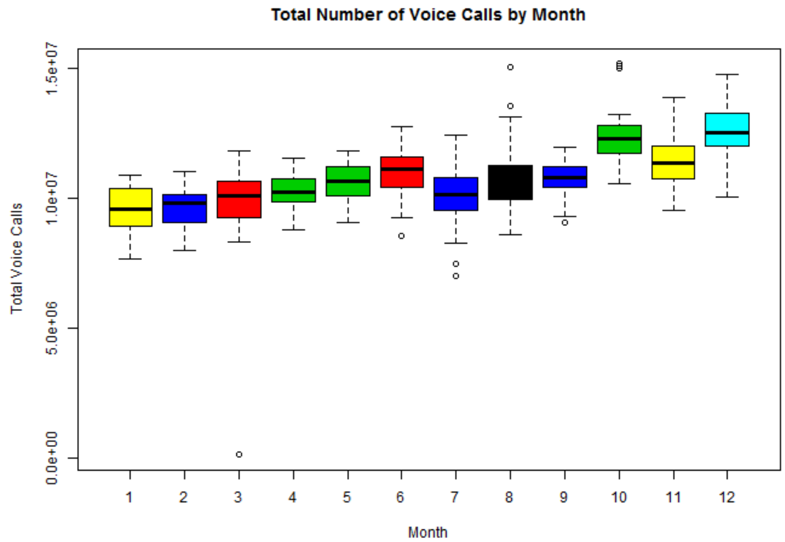

Figure 25 Aggregated Voice Calls per Month

To have a more fine grained level we plot the aggregated voice calls, duration and SMS exchanged per day of the year along with the corresponding smoothed fit curve with confidence region as illustrated in Figures 26, 27 and 28, respectively. It can be seen that for all three types of traffic there is a minimum traffic exchange in 29th of March. Several peaks are recorded on 15-16th of October and 9th of August for the voice calls exchanged, and 9th of August, 16th of October, and 7th of August for the voice calls duration, whereas for SMS traffic exchange the peaks are 16, 15, 17th of August.

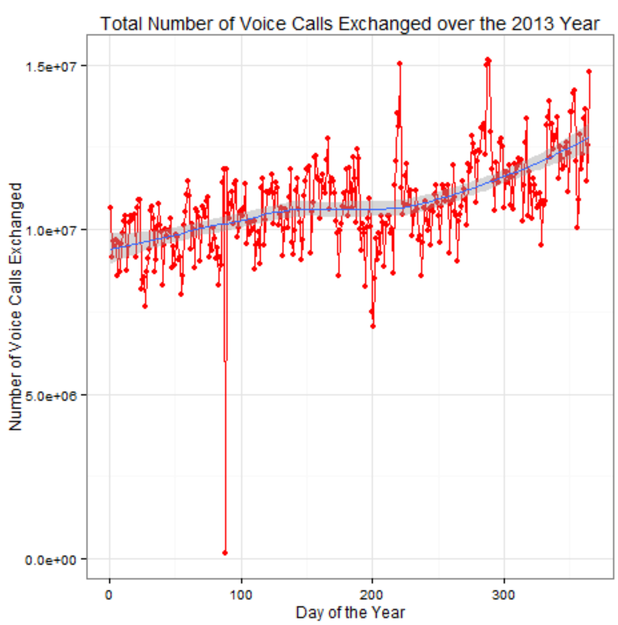

Figure 26 Aggregated Voice Calls per Day of the Year

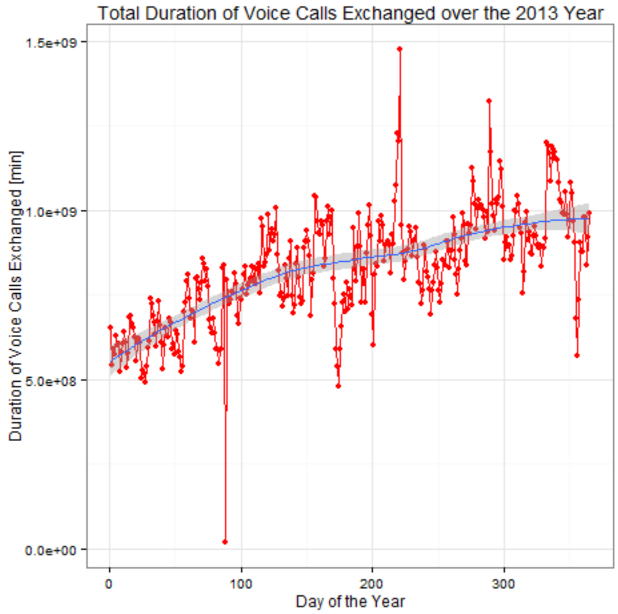

Figure 27 Aggregated Voice Calls Duration per Day of the Year

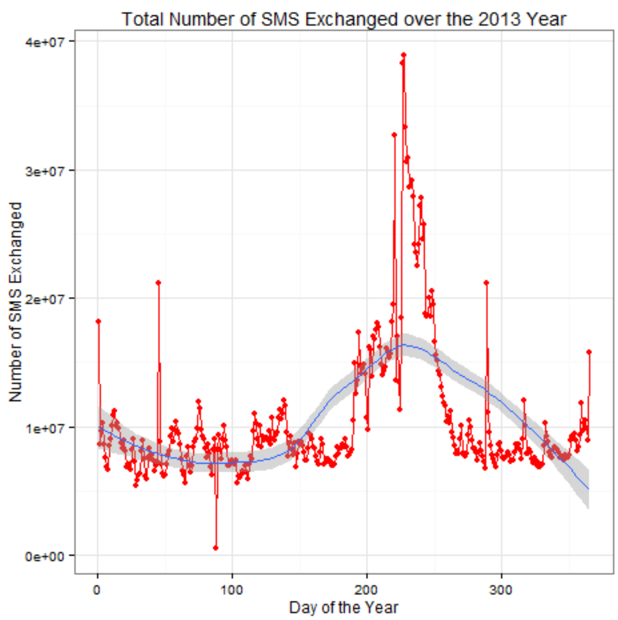

Figure 28 Aggregated SMS Traffic per Day of the Year

5.3 Low Activity on 29th of March

As seen previously there is a very low activity recorded on 29th of March which might be due to a gap in the Dataset 1 or it can be associated to an electric failure. Also it happens that the date is associated to Good Friday.

The traffic matrix for this particular day is listed in Figure 29. It can be noticed that the traffic is mainly local. The activity recorded is only between 12 to 1am, with a total number of voice calls of 153,119 and a total voice calls duration of 1,626,8224 minutes.

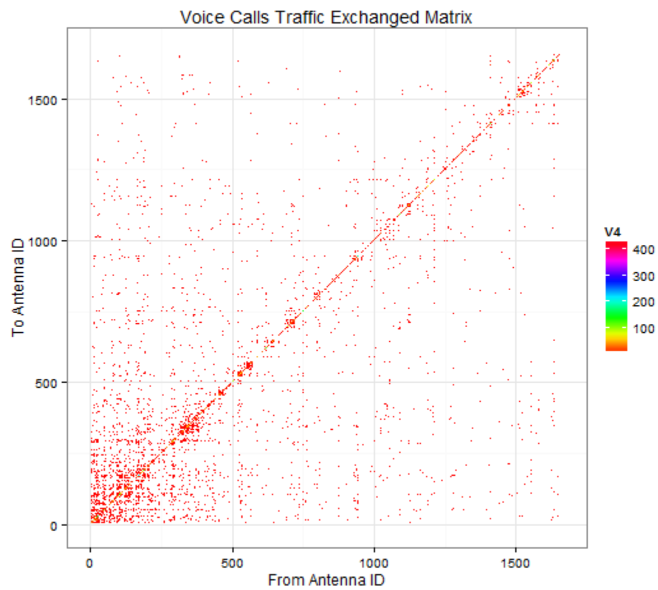

Figure 29 Voice Traffic Matrix on 29th of March

5.4 Ramadan Period

In Senegal, Islam is the predominant religion and is practiced by approximately 94% of the country's population, whereas Christian community is represented by almost 5% of the population. Ramadan represents the month the Quran was revealed and is celebrated by Muslims through fasting during the daylight hours from dawn to sunset. In 2013, the Ramadan started on 9th of July and continued for 30 days until the 7th of August. The marking of the end of the Ramadan period is celebrated through the Feast of Breaking the Fast, referred to as Eid al-Fitr. The celebration will start on the last day of Ramadan, on 7th of August and continues until the next day evening, 8th of August.

As during Ramadan, people fast during the daylight hours we want to analyze if this will impact their calling habits within the network operator. To this extent, we aggregated the number of voice calls and voice calls duration per month and we plot the aggregated values per hours of the day, including the median, the quartiles, and the maximum and minimum values, as illustrated in Figure 30 and 31, respectively. We can notice that there are several peak changes during the night period. The number of voice calls doubles starting from 11pm to 5am during the months of July and August. Thus, this means that during this month people are highly active at night and have a slow start during the morning and the day. This is reflected in the voice calls duration as well. People tend to speak more during July and August starting from 10pm until 6am, when the voice call duration doubles.

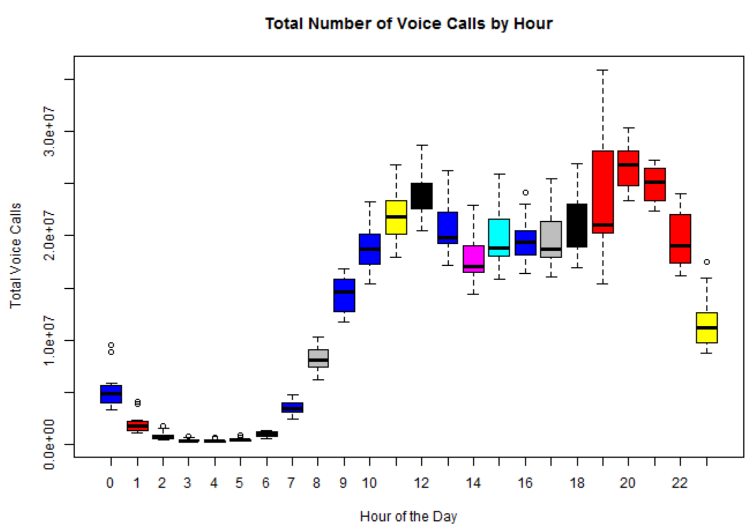

Figure 30 Voice Traffic by Hour of the Day

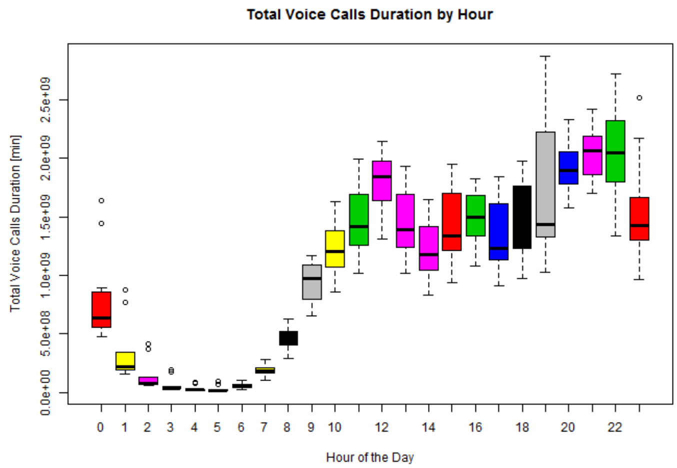

Figure 31 Voice Calls Duration by Hour of the Day

As identified in Section 6.2 and Figures 26-27, the peak days with the maximum voice calls exchanged over the full year are 15-16th of October and 9th of August and for the maximum voice calls duration over the full year are 9th of August, 16th of October, and 7th of August. These dates correspond to the end of Ramadan and the celebration of Eid al-Fitr 7-9th of August. Whereas on 15th of October the Feast of Sacrifice is celebrated referred to as Eid al-Adha. Eid al-Adha and Eid al-Fitr are the two most important religious Eid holidays celebrated by Muslims worldwide each year. Eid al-Adha lasts for three day, and it was celebrated in 2013 starting with 15th of October. This explains the high peak in the number of voice calls and voice calls duration over 15th to 16th of October.

6. CONNECTING PEOPLE, LOCATIONS AND REAL-WORLD EVENTS

From the spatio-temporal patterns of the collective customers' activity within the mobile network traffic datasets introduced previously, the correlation between people, locations and events is analyzed. Specifically the interest is on studying the correlation between exceptional patterns detected in the mobile usage within a cellular network and real-world events such as public concerts, parades, religious festivals, riots protests, etc.

Understanding the exceptional data usage patterns could significantly improve the spatial and temporal awareness when taking decisions. An example would be in the case of event management, when organizing parades/carnivals/concerts, etc.

For this analysis the data provided in Dataset 1 and Dataset 2 were used.

6.1 Connecting People and Locations

Considering the data from Dataset 2, we computed the overall antennas activity in terms of how many users are connected to it, over the full monitoring period. Taking the location coordinates of these antennas, it was possible to identify their position within a certain city. The results are illustrated in Figure 32. The dots represent the antennas locations whereas their size and color representation is reported to the load intensity over the 365 days period. Thus, heavily loaded antennas are represented by larger points and higher intensity color. The results show that the highly loaded antennas are spread across the main cities of Senegal as identified in Figure 2. Comparing the results with the indicative population map of Senegal as illustrated in Figure 3, it can be noticed that data usage activity is mostly registered in densely populated areas, as expected.

These findings have significant impact and they can be correlated to the important cities of the country. These observations led to the correlation between antennas activity within a cellular network and their geographical location. Thus by analyzing the user activity and their mobility patterns within a cellular network only, it is possible to identify the major cities/locations within a country/city.

Understanding the people-location interaction could represent a potential for location-based services. For example time-independent interactions refer to overlapping trajectories between distinct people irrespective of the actual time of overlap. This information is very useful in social recommender systems which are based on location-based tagging services.

The total number of active users over each period of Dataset 2 is represented in Figure 33. It is visible that there are some particular dates when the users are highly mobile. Several such events are identified in the following sections.

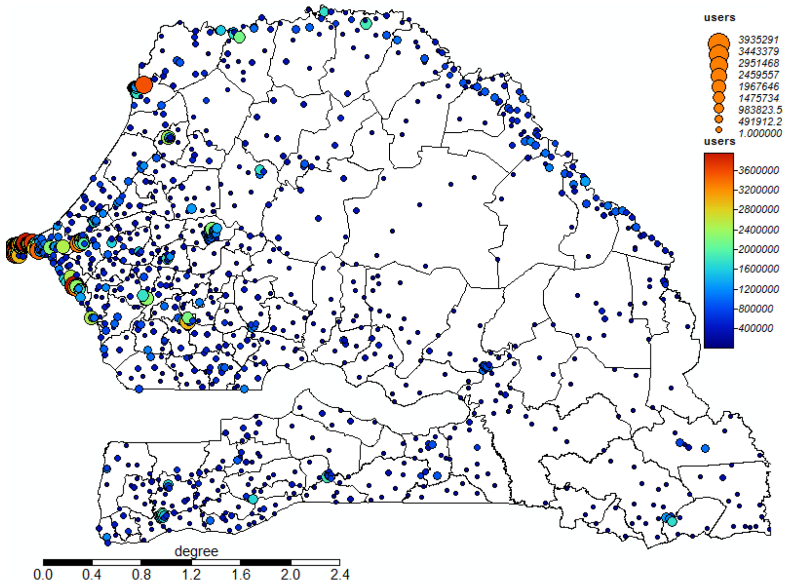

Figure 32 Total Users Activity per Antenna Over the Full Year

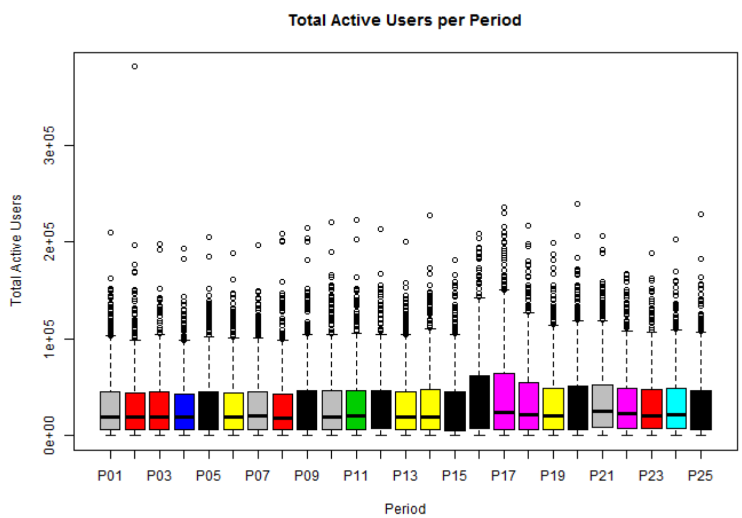

Figure 33 Total Active Users per Period

6.2 Le Grand Magal de Touba

By analyzing the antennas activity and user diversity (e.g., the number of distinct users connected to an antenna over the monitoring period), it was possible to identify particular religious festivals. The data showed that there was intensive users activity and users diversity in the area of Touba, Mbacké region, towards the end of December. Figure 34, illustrates the total number of active users over the December period for three antennas located in Touba area, specifically Antenna IDs: 1019, 1024 and 1025. It can be noticed that the user activity increases more than six times by 22nd of December. Taking these observations and looking at the real-time events happening in that specific location during exactly that period, we come to know about the Magal Festival[3] taking place on 22nd of December. Consequently, these pattern exceptions in the antenna usage are perfectly correlated with the real-world event, such as: Magal Festival.

[3]Senegal Magal Festival http://www.itnsource.com/en/shotlist/RT V/2013/12/23/RTV231213008/

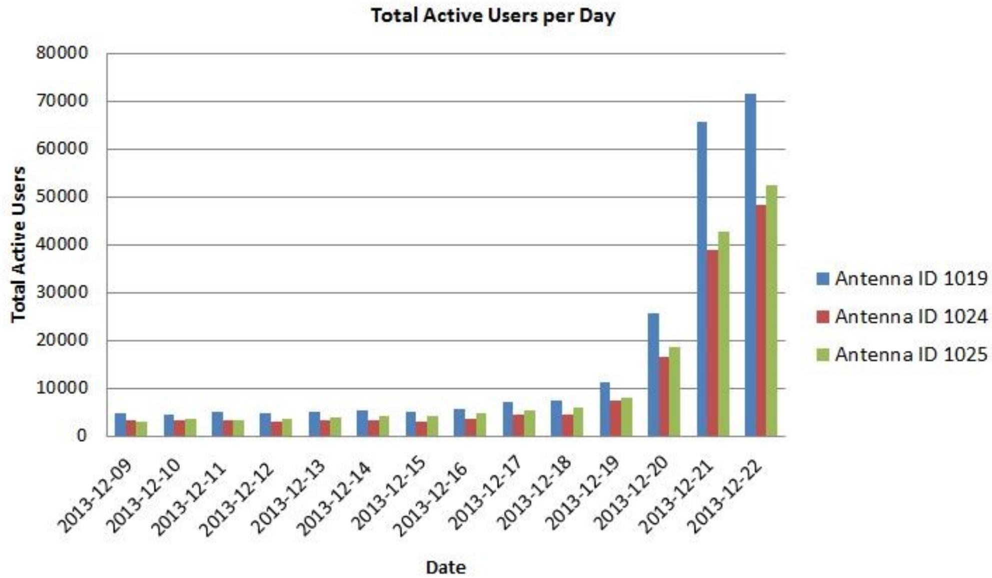

Figure 34 Total Users Activity per Day for Antenna IDs 1019, 1024 and 1025.

During the Magal Festival, more than a million pilgrims, members of the Mouride brotherhood flocked to 'Africa's Mecca' from all over the world in the holy town of Touba. On 22nd of December 2013 the 119th edition of the Magal festival was celebrated.

6.3 Tivaouane Maouloud Festival

Another exceptional event was registered in January near Tivaouane, Thies region. Figure 35, illustrates the total number of active users over the January period for five antennas located in the Thies area, specifically Antenna IDs: 599, 604, 606, 608, and 609. It can be noticed that the user activity increases more than ten times by 23rd of January. Taking these observations and looking at the real-time events happening in that specific location during exactly that period, we come to know about the Maouloud Festival[4]. Each year millions of visitors are celebrating the birth of the prophet Muhammad through the Maouloud festival, also known as Gamou.

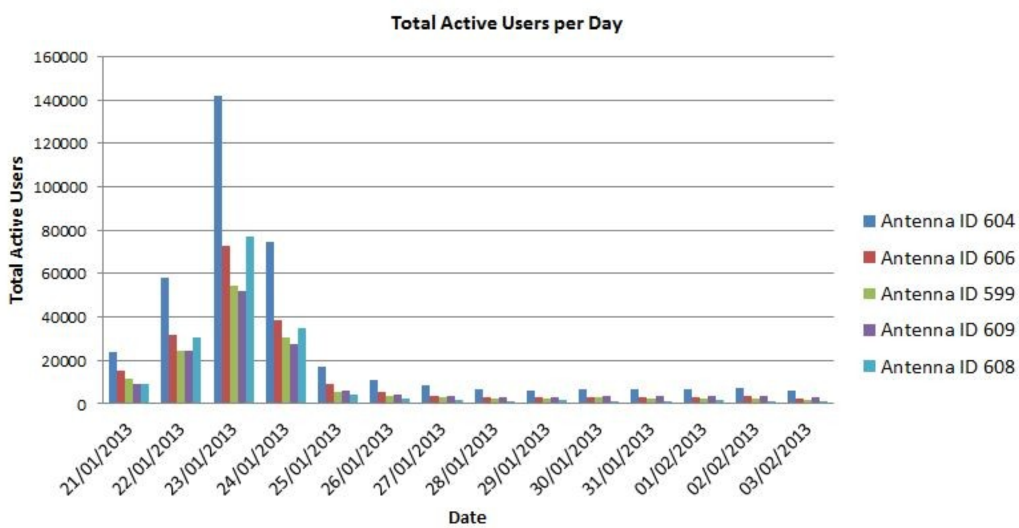

Figure 35 Total Users Activity per Day for Antenna IDs 604, 606, 599, 609, 608.

High user activity during the Maouloud festival was registered in Kaoulack as well, as illustrated in Figure 36 for Antenna ID 944.

[4] Senegal Maouloud Festival http://en.wikipedia.org/wiki/ Tivaouane

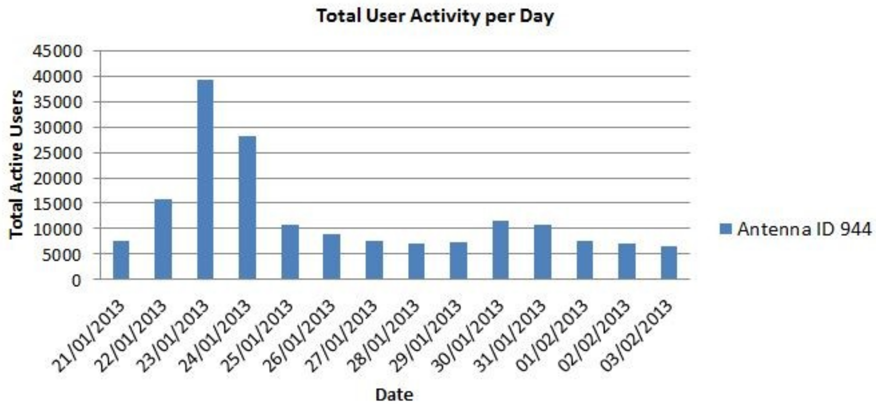

Figure 36 Total Users Activity per Day for Antenna ID 944.

6.4 Casamance Conflict

Apart from the religious festivals that can be detected from the CDR Datasets, we were able to detect conflict events as well. An exceptional event in the CDR datasets 1 and 2 was detected in the area of Kafoutine during February. The total number of voice calls along with the active users during the February period for Antenna ID 622 is listed in Figure 37. It can be seen that there is an anomaly in the traffic that starts increasing from 1st of February with a peak in the communication on the 2nd of February.

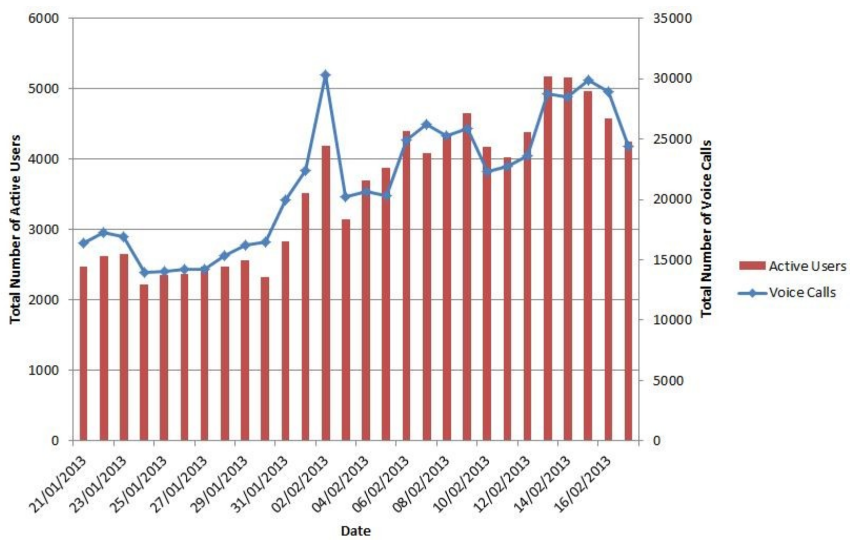

Figure 37 Total Number of Users and Voice Calls for Antenna ID 622.

The antenna is located near the Kafoutine city in the Casamance region. On 1st of February 2013 there was reported an attack of the rebels from the Casamance Movement for Democratic Forces over the Credit Mutuel bank in Kafountine[5]. After the clashes between the rebels and the government soldiers, four dead including a Frenchman were reported. The effect of this attack can be noticed in the CDRs. After the news about the attack came out on 2nd of February there was a peak in the number of voice calls made and a significant increase in the voice calls duration.

7. CONCLUSIONS

In this work we explore the use of anonymized Call Detail Records (CDRs) containing both voice-calls and SMS activities, from a cellular network to study the telecommunication traffic exchanged and connect people, locations and real-world events. The paper provides a comprehensive literature review where three wide categories are identified, such as: (1)

---

[5] Casamance Conflict http://www.bbc.com/news/world-africa-21306645

universal law for human mobility, (2) urban planning and traffic forecast, and (3) localization and mobility patterns. Moreover, the major findings from these categories are summarized and discussed.

The study presented in this paper, makes use of CDR data collected over one year from more than nine million Orange customers in Senegal to detect exceptional spatio-temporal patterns of the collective human mobile data usage and to correlate these 'anomalies' in the usage patterns to real-world events (e.g., religious festivals, riots, conflicts, etc.).

Despite the limitations of the datasets as identified in Section 3.5, which are imposed to maintain the anonymity of the datasets and to protect customers' privacy, there are still some important observations that can be drawn from the analysis provided.

The results show that as expected the antennas distribution within Senegal correlates with the population density, meaning that in the most populous regions, there is an increased number of base stations. Moreover, the traffic intensity in that location is also increased and is mainly located in the main city areas. Understanding the correlation between the population size, antenna density and the traffic load could help the network operators at developing an adaptive power allocation scheme for the base stations so that the transmission power where the antennas present high traffic load and dense user activity could be increased and the transmission power where the antennas have low demand and light traffic load could be decreased. In this way the network operator could improve the resource allocation capacity and the calls' QoS and also save energy consumption.

The results also show that the CDR data can be used to detect exceptional spatio-temporal pattern of the collective human mobile data usage that could then be mapped to real-world events. For example, high user network activity was identified during periods collocated with various religious festivals, such as Ramadan, Le Grand Magal de Touba and the Tivaouane Maouloud festival. In the case of the Ramadan period during July and August, we could notice that the communication pattern changed and there is more activity during the night period, with the number of voice calls doubling between 11pm and 5am compared to normal periods. During the Ramadan period it is noticeable that people are highly active at night and have a slow start during the morning and along the day. Understanding these exceptional data usage patterns could significantly improve the spatial and temporal awareness when taking decisions and this knowledge could be further used to develop an intelligent system that detects exceptional events in real-time from CDRs monitoring. For example, the results also showed that there was a peak increase in the number of voice calls made as well as in the voice calls duration in the area of Kafoutine during February which corresponds to the Casamance Conflict in the area. In this situation, a real-time event detection system could be of crucial importance to ensure people's safety in case of emergency situations, such as conflicts or riots protests which could be more efficiently handled if detected on time.

Moreover, the study presented in this paper validates the observations from our previous work (Trestian et al., 2016) on a smaller scale CDR dataset collected over 150 days in Ivory Coast. In that study we have seen that exceptional spatio-temporal patterns of mobile data usage could be associated to real-world events, such as parades, public concerts, football games and New Year's Eve. For example, in case of the football game, like the Africa Cup of Nations it was noticed that there was a decrease in users' activity during the day with the game between Ivory Coast and Zambia when compared to the users' activity during other days without these types of events. This marks in fact the importance of the event.

8. ACKNOWLEDGMENTS

Studies and Researches were performed using mobile communication data made available by SONATEL and Orange within the D4D Challenge.

9. REFERENCES

Blondel, V. D., Decuyper, A., & Krings, G. (2015). A survey of results on mobile phone datasets analysis. arXiv:1502.03406 [physics]. Retrieved from http://arxiv.org/abs/1502.03406

Calabrese, F., Colonna, M., Lovisolo, P., Parata, D., & Ratti, C. (2011). Real-Time Urban Monitoring Using Cell Phones: A Case Study in Rome. IEEE Transactions on Intelligent Transportation Systems, 12(1), 141–151. doi:10.1109/TITS.2010.2074196

Cisco Visual Networking Index: Forecast and Methodology, 2014-2019 White Paper - Cisco. (2015). Retrieved 17 March 2016, from http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.html

Di Lorenzo, G., & Calabrese, F. (2011). Identifying human spatio-temporal activity patterns from mobile-phone traces. In 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC) (pp. 1069–1074). doi:10.1109/ITSC.2011.6082974

Frias-Martinez, E., Williamson, G., & Frias-Martinez, V. (2011). An Agent-Based Model of Epidemic Spread Using Human Mobility and Social Network Information. *In 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom)* (pp. 57–64). doi:10.1109/PASSAT/SocialCom.2011.142

González, M. C., Hidalgo, C. A., & Barabási, A.-L. (2008). Understanding individual human mobility patterns. Nature, 453(7196), 779–782. doi:10.1038/nature06958

Han, Q., & Ferreira, P. (2013). Determinants of Subscriber Churn in Wireless Networks: The Role of Peer Influence. *26th European Conference on Operational Research, Telecommunication, Networks, and Social Networks Stream, Roma, Italy*. doi: 10.1145/2639968.2640057

Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Rowland, J., & Varshavsky, A. (2010). A Tale of Two Cities. In Proceedings of the Eleventh Workshop on Mobile Computing Systems & Applications (pp. 19–24). New York, NY, USA: ACM. doi:10.1145/1734583.1734589

Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Martonosi, M., Rowland, J., & Varshavsky, A. (2011). Identifying Important Places in People's Lives from Cellular Network Data. In Proceedings of the 9th International Conference on Pervasive Computing (pp. 133–151). Berlin, Heidelberg: Springer-Verlag. Retrieved from http://dl.acm.org/citation.cfm?id=2021975.2021988

Krings, G., Calabrese, F., Ratti, C., & Blondel, V. D. (2009). Urban Gravity: a Model for Intercity Telecommunication Flows. arXiv:0905.0692 [Physics]. Retrieved from http://arxiv.org/abs/0905.0692

Kung, K. S., Greco, K., Sobolevsky, S., & Ratti, C. (2014). Exploring Universal Patterns in Human Home-Work Commuting from Mobile Phone Data. PLOS ONE, 9(6), e96180. http://doi.org/10.1371/journal.pone.0096180

Motahari, S., Zang, H., & Reuther, P. (2012). The Impact of Temporal Factors on Mobility Patterns. In 2012 45th Hawaii International Conference on System Science (HICSS) (pp. 5659–5668). doi:10.1109/HICSS.2012.572

Naboulsi, D., Fiore, M., Ribot, S., & Stanica, R. (2016). Large-Scale Mobile Traffic Analysis: A Survey. IEEE Communications Surveys & Tutorials, 18(1), 124–161. http://doi.org/10.1109/COMST.2015.2491361

Palchykov, V., Mitrović, M., Jo, H.-H., Saramäki, J., & Pan, R. K. (2014). Inferring human mobility using communication patterns. Scientific Reports, 4. doi:10.1038/srep06174

Saramäki, J., & Moro, E. (2015). From seconds to months: an overview of multi-scale dynamics of mobile telephone calls. The European Physical Journal B, 88(6). http://doi.org/10.1140/epjb/e2015-60106-6

Song, C., Koren, T., Wang, P., & Barabási, A.-L. (2010a). Modelling the scaling properties of human mobility. Nature Physics, 6(10), 818–823. doi:10.1038/nphys1760

Song, C., Qu, Z., Blumm, N., & Barabási, A.-L. (2010b). Limits of Predictability in Human Mobility. Science, 327(5968), 1018–1021. doi:10.1126/science.1177170

Tatem, A. J., Huang, Z., Narib, C., Kumar, U., Kandula, D., Pindolia, D. K., … Lourenco, C. (2014). Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning. *Malaria Journal*, 13, 52. doi:10.1186/1475-2875-13-52

Trestian, I., Ranjan, S., Kuzmanovic, A., & Nucci, A. (2009). Measuring Serendipity: Connecting People, Locations and Interests in a Mobile 3G Network. In Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference (pp. 267–279). New York, NY, USA: ACM. doi:10.1145/1644893.1644926

Trestian, R., Zaman, F., & Muntean, G. (2016). Spotted: Connecting People, Locations, and Real-World Events in a Cellular Network. In J. Martins, & A. Molnar (Eds.) *Handbook of Research on Innovations in Information Retrieval, Analysis, and Management* (pp. 1-40). Hershey, PA: Information Science Reference. doi:10.4018/978-1-4666-8833-9.ch001

Wesolowski, A., Eagle, N., Tatem, A. J., Smith, D. L., Noor, A. M., Snow, R. W., & Buckee, C. O. (2012). *Quantifying the Impact of Human Mobility on Malaria*. Science, 338(6104), 267–270. doi:10.1126/science.1223467

Yuan, J., Zheng, Y., & Xie, X. (2012). Discovering Regions of Different Functions in a City Using Human Mobility and POIs. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 186–194). New York, NY, USA: ACM. doi:10.1145/2339530.2339561

Zang, H., & Bolot, J. C. (2007). Mining Call and Mobility Data to Improve Paging Efficiency in Cellular Networks. In Proceedings of the 13th Annual ACM International Conference on Mobile Computing and Networking (pp. 123–134). New York, NY, USA: ACM. doi:10.1145/1287853.1287868