University of Luxembourg

SCILABTEC 2015, Friday 22 May.

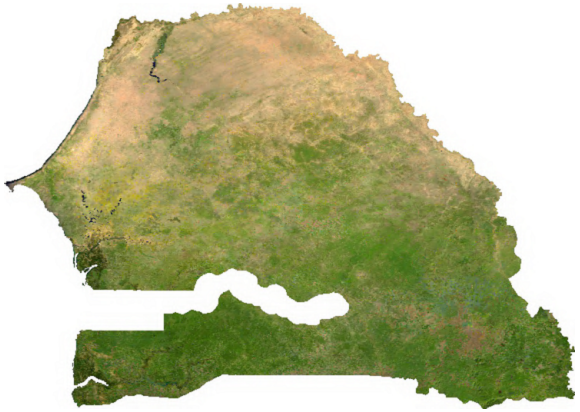
ANALYSIS Of Call Detail Records Based on Scilab

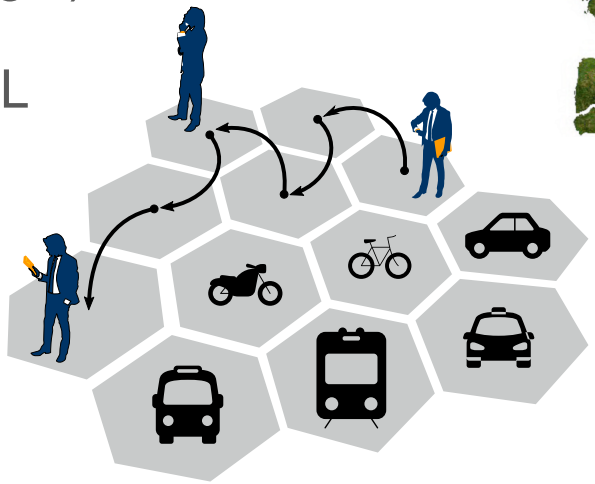
Foued Melakessou

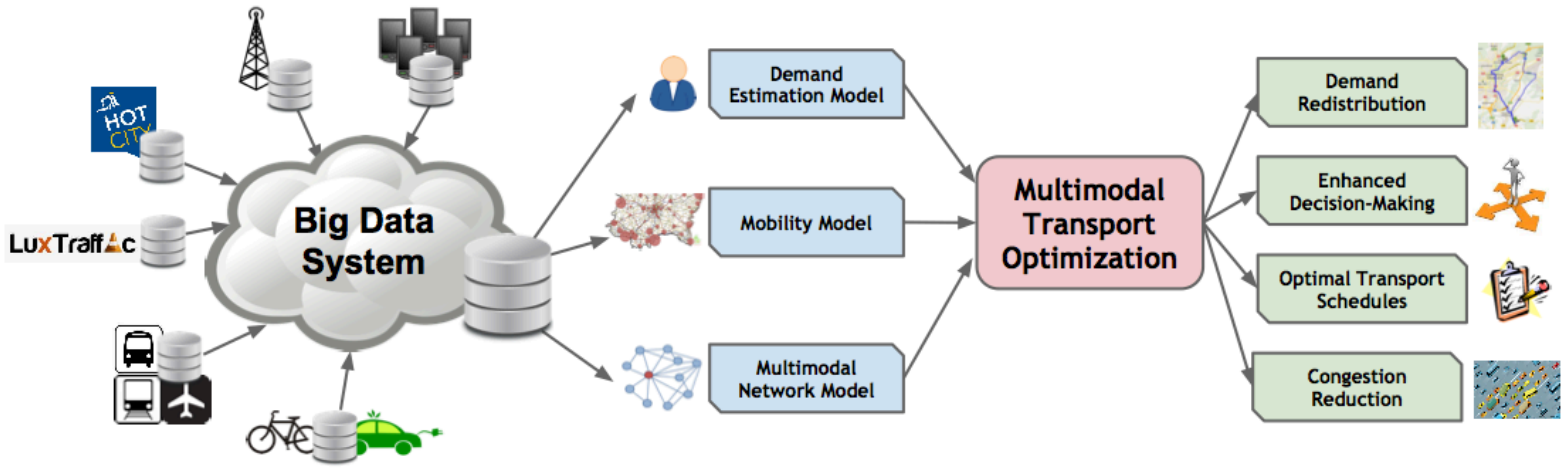
uni.lu
UNIVERSITÉ DU
LUXEMBOURG

Outline

- Project Description
- Problematic
  - Big Data Analysis
  - Population Mobility Modeling
- D4D Challenge (2015 Senegal)
- Scilab Contribution: NARVAL
- Conclusion & Perspective







- FNR Core Project (April 2014-March 2017)
- The **MAMBA** (**M**ultimod**A**l **M**o**B**ility **A**ssistance) project intends to propose and validate a multimodal mobility platform that relies on new Internet technologies to interconnect different mobile services with the aim of providing relevant travel advice based on users' contexts, so as to optimize overall system performance
  - Real time traffic conditions
  - Status of existing public transport services (e.g., buses, trains)
  - User preferences
- **Analysis of large mobility datasets (e.g., mobile phone traces)**
- Influence the itinerary of the users by **suggesting new multimodal routes based on their preferences**

Problematic

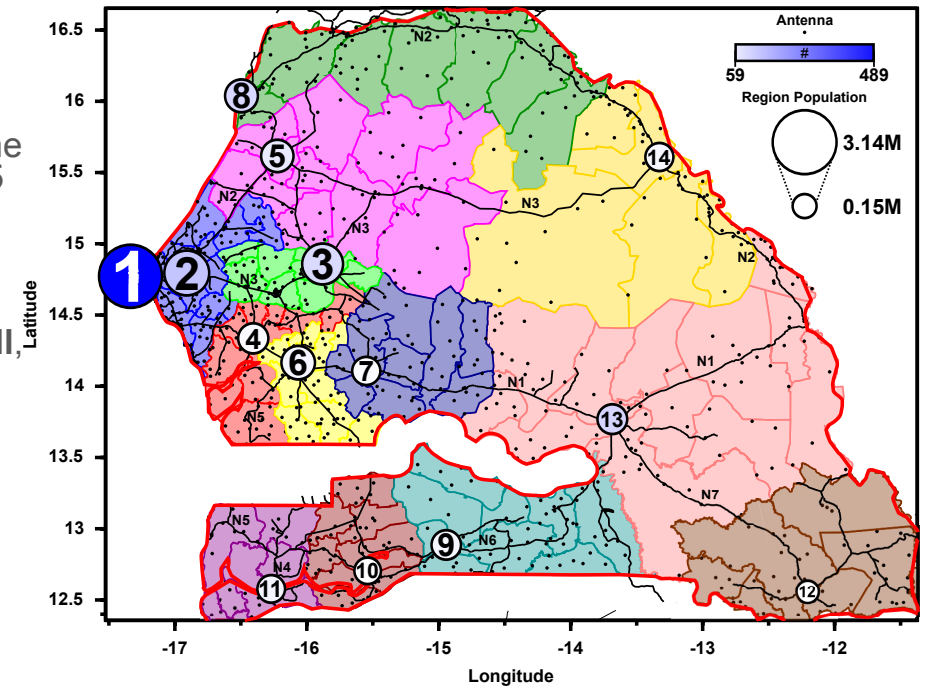
- Transportation Research: human mobility
  - Urban planning
  - Traffic forecasting
  - Resource management
- The evaluation of mobility models helps to better design and develop future infrastructure in order to better support the actual demand
- Lack of tools to monitor the time-resolved location of individuals
  - Observation studies
  - Forms: population census
- Call Detail Records (CDRs), generated by mobile phone operators can be used to retrieve mobility patterns of the population under study.
  - Extraction of realistic mobility models adapted to the Senegal use case

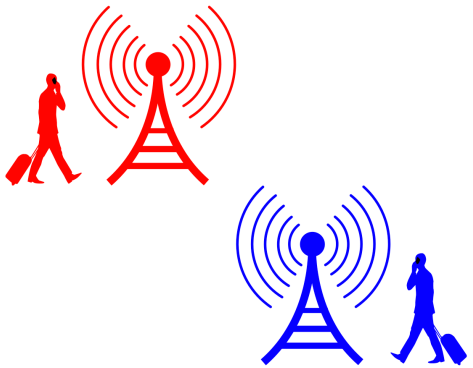
Data 4 Development Challenge: Senegal

NetMob
The main conference on the scientific analysis of mobile phone datasets

- D4D Senegal is an innovation challenge open on ICT Big Data for the purposes of societal development
  - NETMOB'15, Media Lab, MIT, April 8-10, 2015
- Sonatel and the Orange Group are making **anonymous data**, extracted from the mobile network in Senegal, available to international research laboratories
- 5 priority subject matters: health, agriculture, transport/urban planning, energy and national statistics
  - Advancing research in the field of Big Data (anonymisation, datamining, visualization and cross-referencing)
  - Involving local stakeholders and guaranteeing benefits in education and development of the ecosystem of local start-ups
  - Advancing in anonymisation techniques to allow sharing of data that is relevant for society while respecting privacy

- Sonatel/Orange (Société nationale des télécommunications) is the main mobile telecommunication operator
- Anonymous call patterns of Orange's mobile phone users in Senegal have been released for the 2015 D4D challenge: phone calls and text exchanges between **9M** of Orange customers in **2013**
  - **1666** Base Stations: **Antenna ID**, **Lat**, **Lon**
  - **SET1** (**53.8GB**): Voice and Text traffic (**#Call**, **TCall** and **#Text**) for each month (antenna-to-antenna traffic on an hourly basis)
    - 2013-01-01 00,1,1,1,54
    - 2013-01-01 00,1,2,1,39
  - **SET2** (**38.4GB**): Fine-grained mobility data on a rolling 2-week basis (trajectories of 300000 randomly sampled users)
    - 1,2013-01-07 13:10:00,461
    - 1,2013-01-07 17:20:00,454
  - **SET3** (**16.5GB**): Coarse-grained mobility data, month by month, at the district level (trajectories of 150000 randomly sampled users)
    - 37509, 2013-01-29 15:00:00,3
    - 84009, 2013-01-14 07:00:00,3

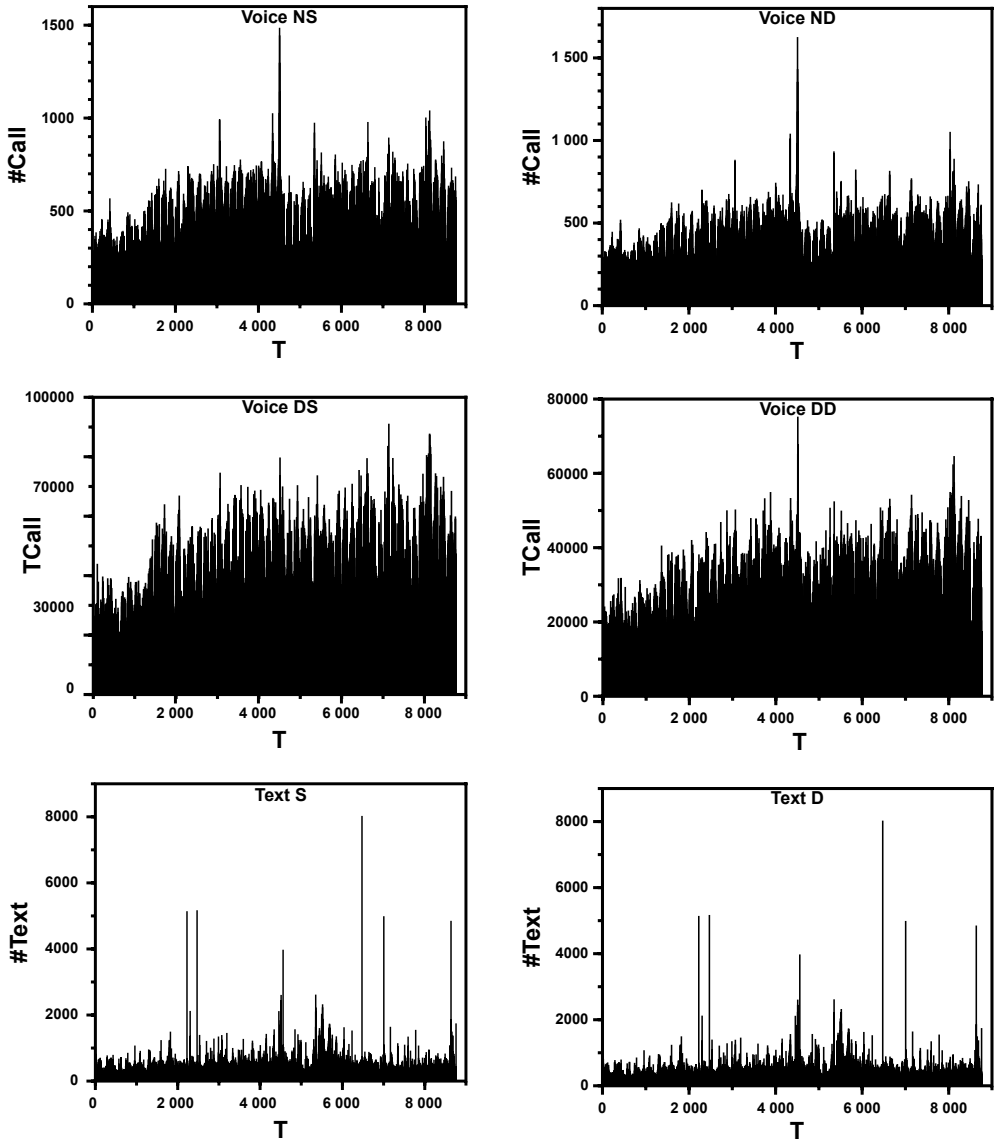




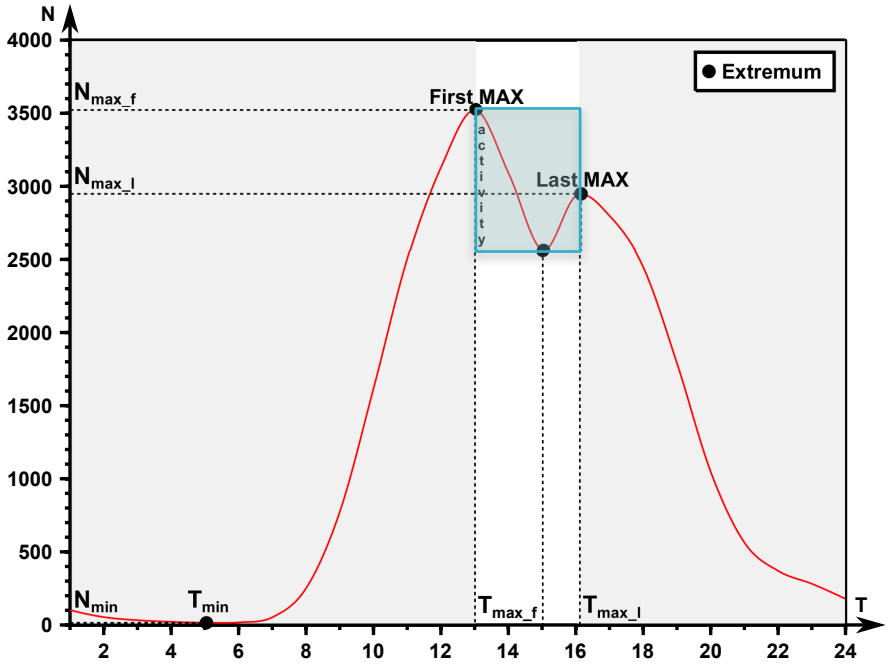
1	Dakar	8	Saint-Louis
2	Thies	9	Kolda
3	Djourbel	10	Sedhiou
4	Fatick	11	Ziguinchor
5	Louga	12	Kedougou
6	Kaolack	13	Tambacounda
7	Kaffrine	14	Matam

Daily Profile Model

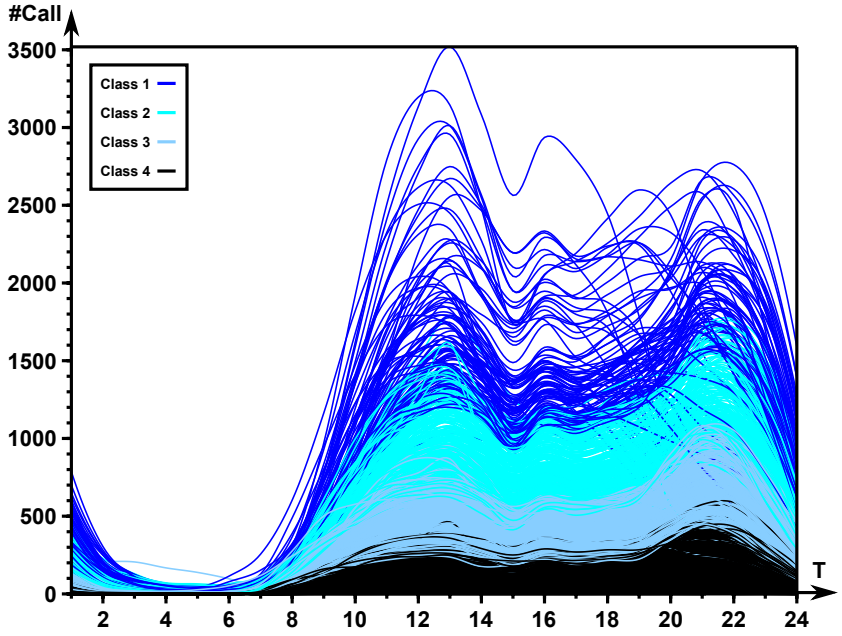
- Human activities heavily impact the behavior of calls and messages emitted and received during each period of the day
  - Model the daily traffic profile supported in each base station (level of aggregation=1h)
    - Traffic variable: number of calls #Calls, duration of calls Tcall and number of messages #Text
    - Outliers removal process at 2σ
  - Normal traffic behavior
  - Detection of traffic anomalies (cultural or sport events)

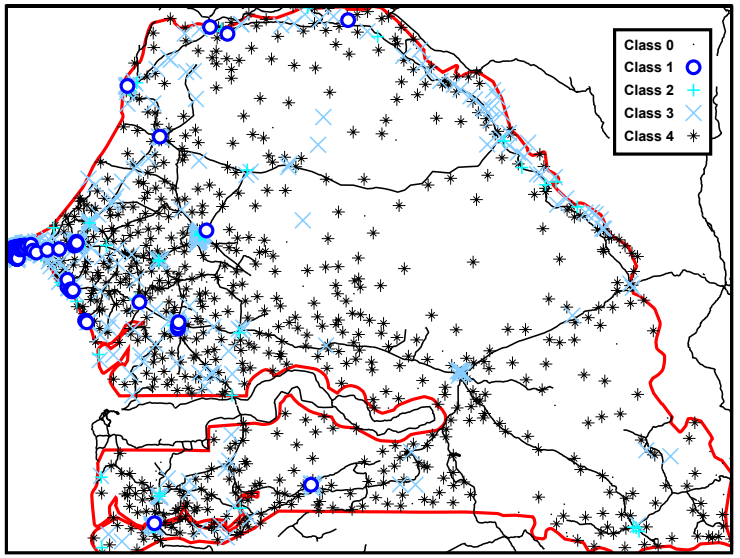


- $P_X(a)=\{X_1X_2X_3\ldots X_{24}\}$
  - X_i is the expected value of the traffic variable X collected in the antenna a during the i^{th} one hour time slot
- $E(a)=\{E_1E_2\ldots E_e\}$
  - List of extrema computed with a gradient algorithm
  - Succession of local minimums and maximums
- **NaN**, a statistics and machine learning toolbox for Scilab focusing on data with and without missing values encoded as NaN's
  - T_{max_f}: time when the first local maximum N_{max_f} occurs (beginning of diurnal activities)
  - T_{max_l}: time when the last local maximum N_{max_l} occurs (end of diurnal activities)
  - **M** (respectively **S**) is the average traffic (respectively the standard deviation) of the diurnal activities
  - N_e: number of extrema
  - T_{min}: time when the global minimum N_{min} occurs
- Classification based on k-Means algorithm to find a set of maximally disjoint clusters (**k=4**)
  - Urban (Class 1)
  - Suburban (Class 2 and 3)
  - Rural (Class 4)

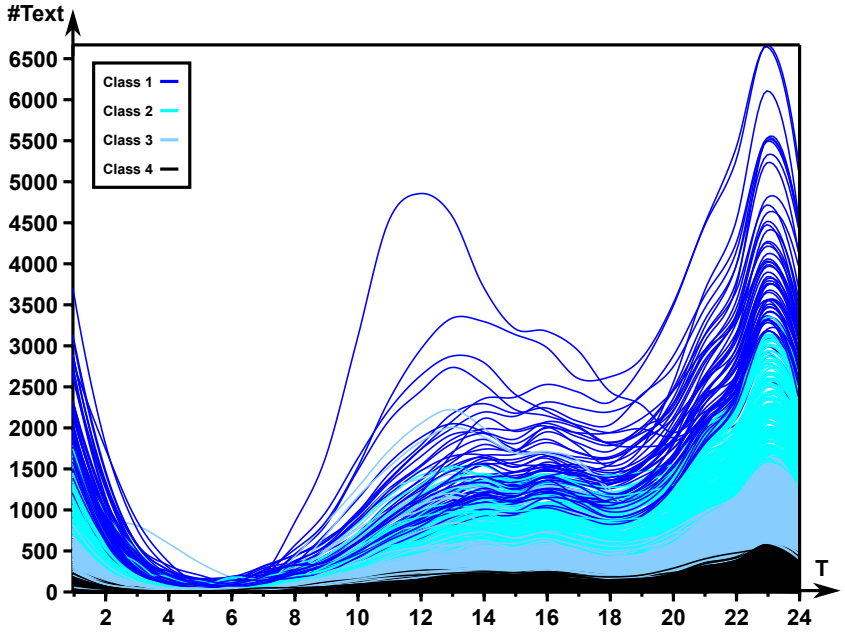


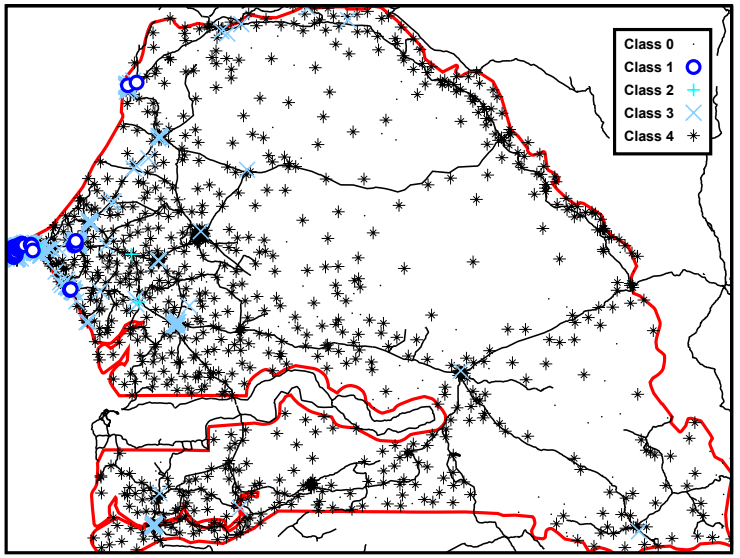
- First and largest peak centered at 12AM
- Diurnal activities start at 08AM and finish at 11PM
- Rural population represents **55%** of the total population (**14M**)
  - Highest population density in the area of Dakar, the capital of Senegal





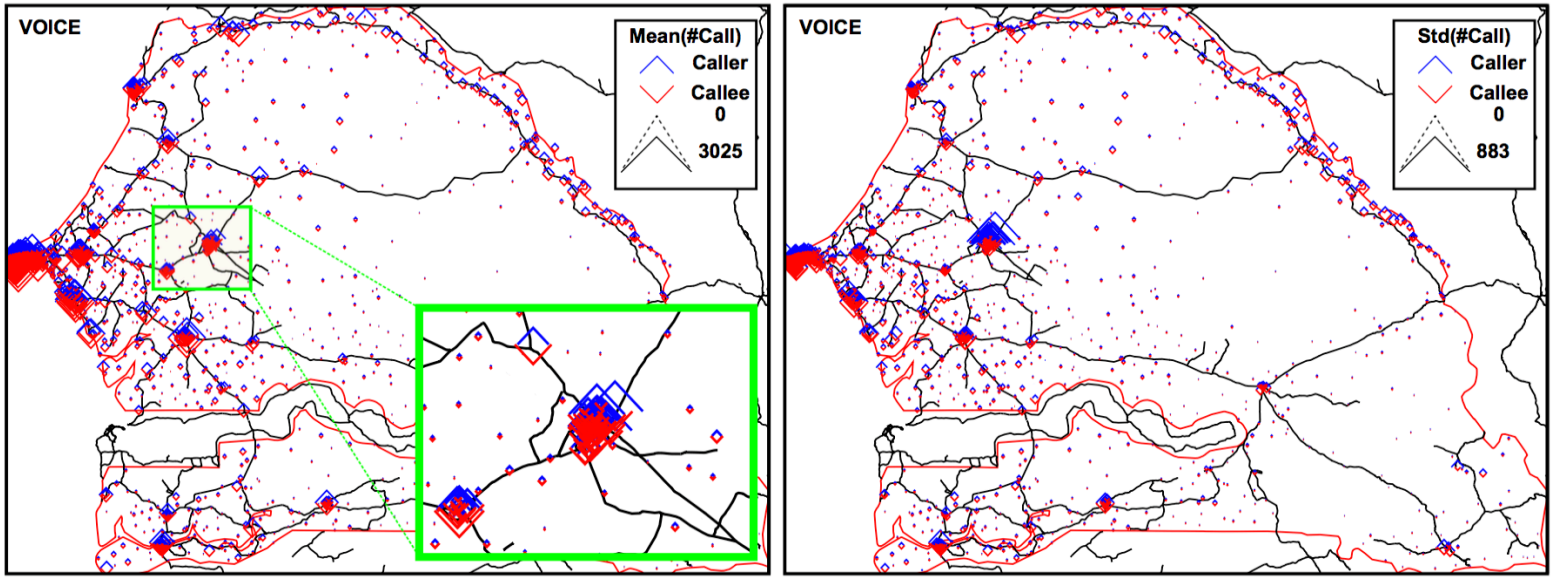
- High activity during the night (08PM to 02AM)
- Young population
  - The median age is **18.4** years
  - 60% of the population are less than 24 years old



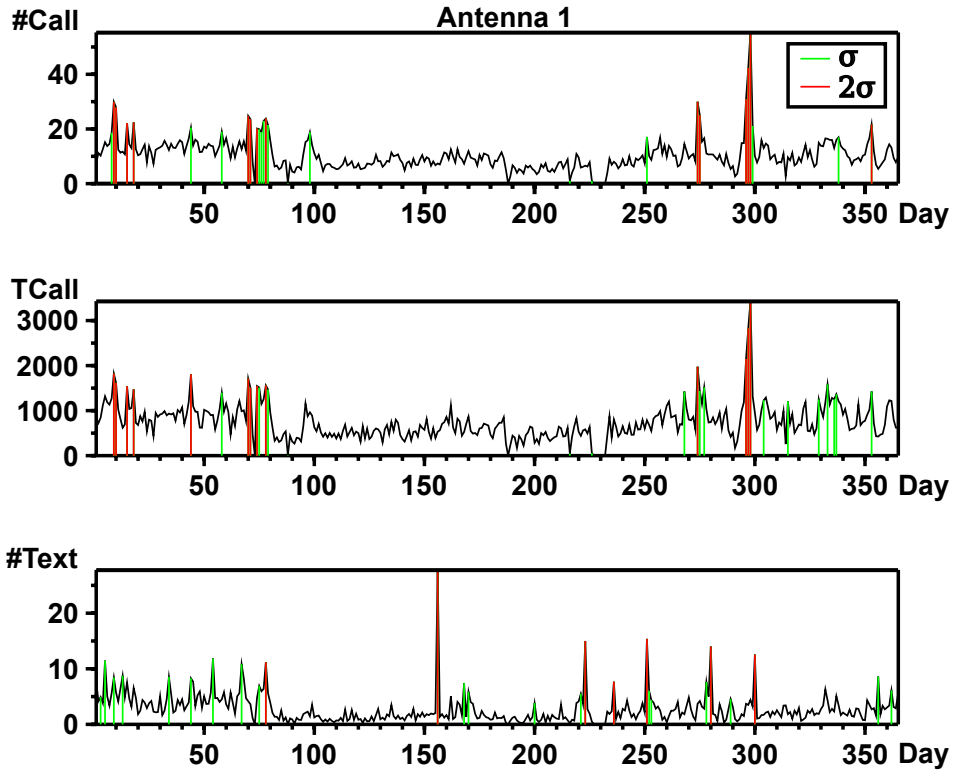


UNIVERSITÉ DU LUXEMBOURG

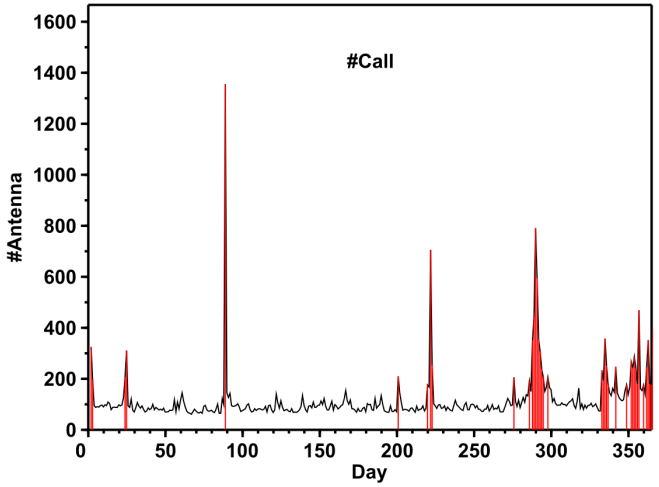
- Mobile phone activity between the first and last peaks
  - Mean
  - Standard deviation
- Local vs Distant calls

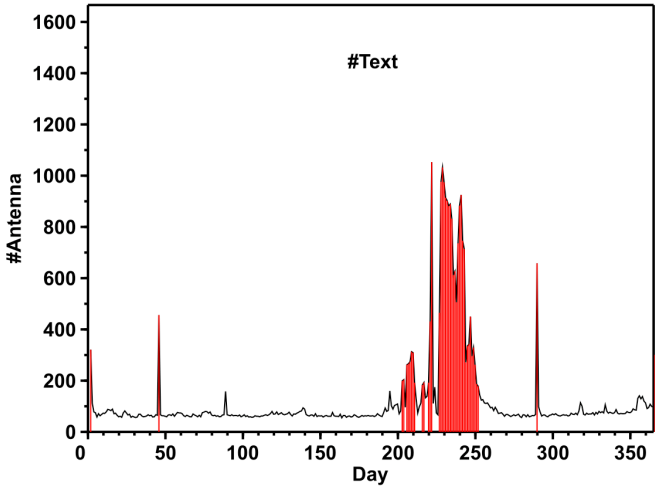


- Level of aggregation: 1 Day
- Detection of local anomalies
  - Correlation #Call and Tcall



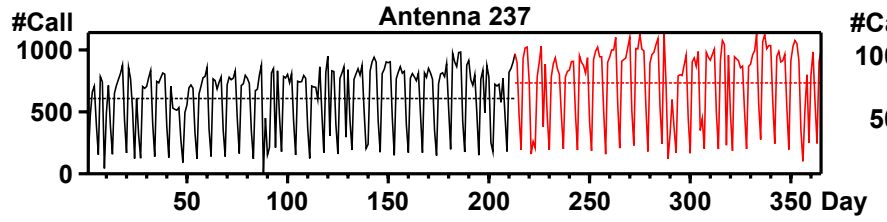
- Level of aggregation: 1 Day
  - Number of antennas that present an anomaly the same day

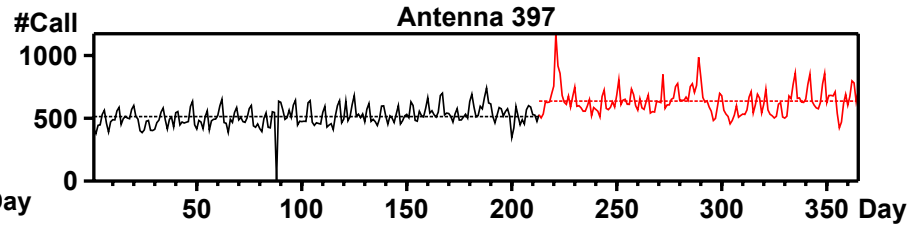


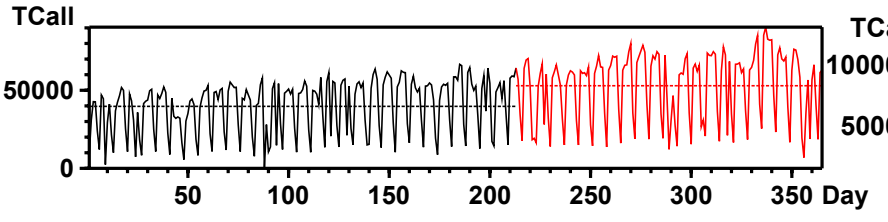


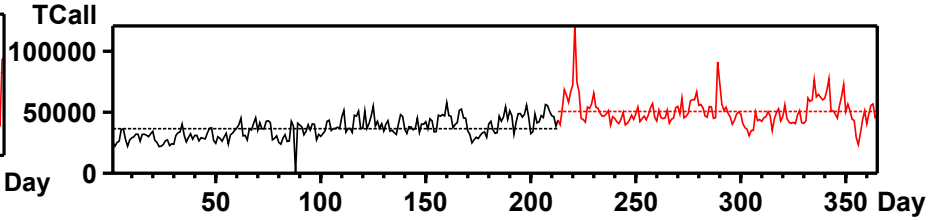
Date (2013)	Event
1 January	New Year
24 January	Mawlid (Prophet birth)
31 March - 1 April	Easter
4 April	Independence day
9 July - 7 August	Ramadan
4 August	Laylat Al Qadr
8 August	Aid El Fitr
15 August	Assomption
15 October	Aid El Kebir
13-14 November	Tamkharit (Achoura)
15 December to beginning of January	Magal of Touba
25 December	Christmas

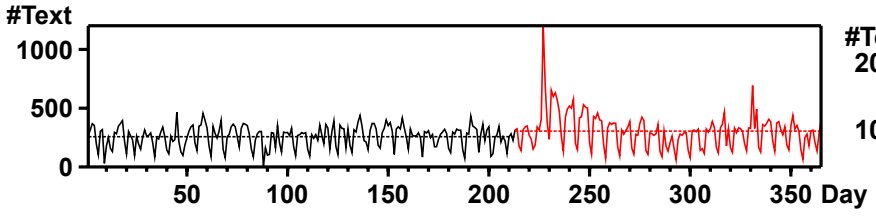
- On the 1st August 2013, the new highway between Dakar and Diamniadio has been inaugurated

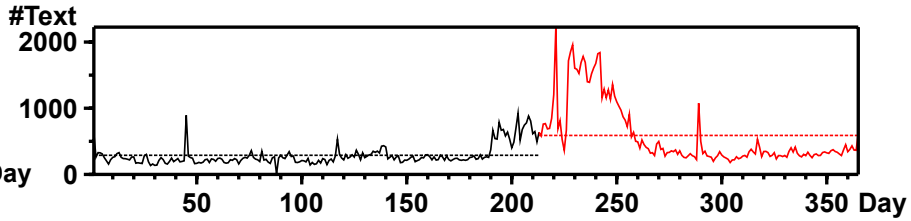






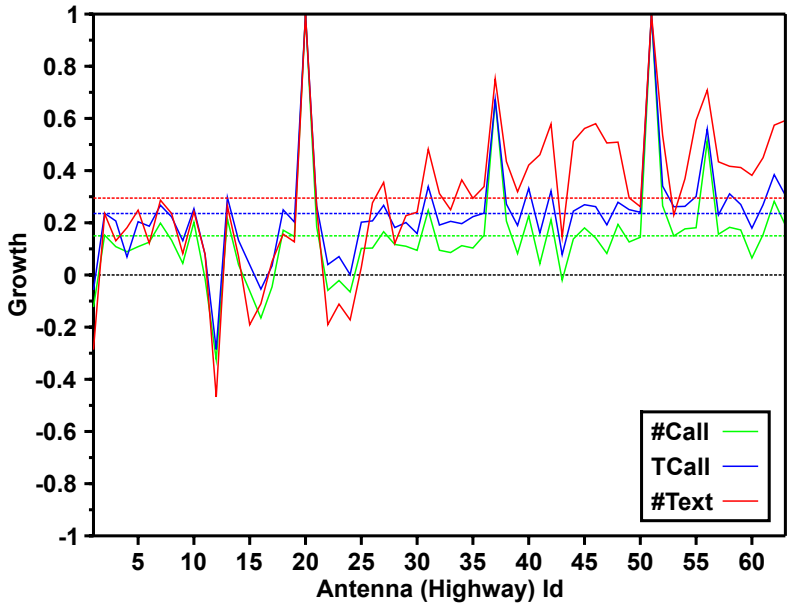






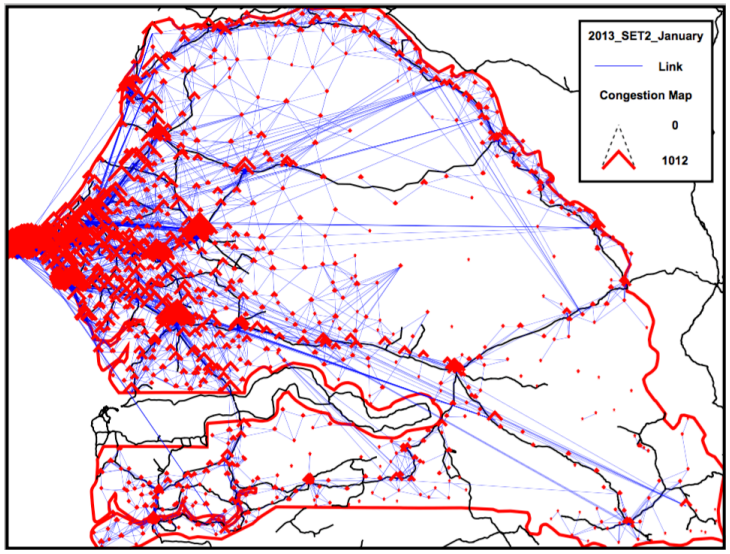
UNIVERSITÉ DU LUXEMBOURG

- 67 antennas located near to the new highway
- Computation of the traffic growth after the inauguration
  - #Call: +15%
  - TCall: +23%
  - #Text: +29%



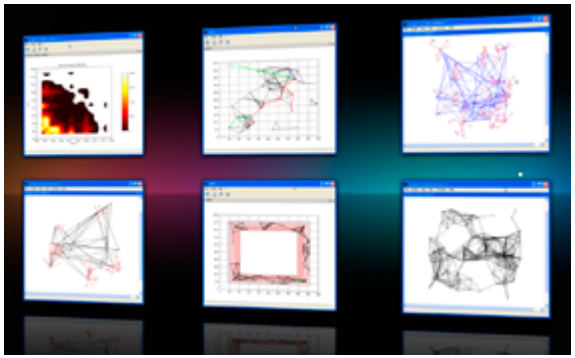
UNIVERSITÉ DU LUXEMBOURG

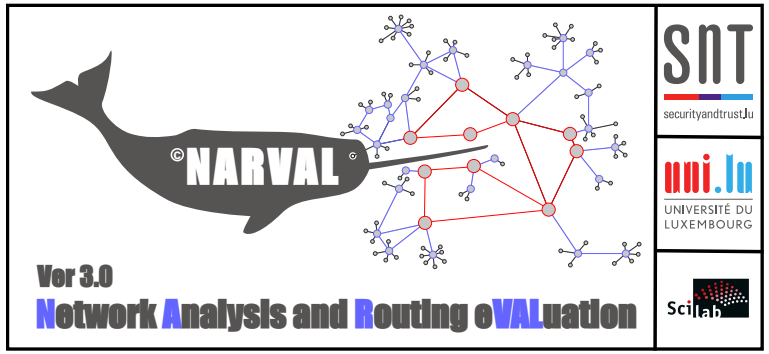
- Maps of primary flows
  - Mobility graphs are composed by all interconnections between antennas (respectively districts) where users moved during the duration of the traffic collection
  - Congestion map provides for each antenna **A**, the number of users that have crossed **A** during the duration of the traffic measurement
- Each congestion map presents high values near to urban areas
- Mobility graphs are highly correlated with the road network
- Long links: there is a lack of information about intermediate locations where the users traveled, if the inter-arrival time between two mobile activities is too large



- Other data sets used in this project
  - Type of data: OpenStreetMap
  - Perl Scripts (**automatic extraction of national boundaries, highway/primary/secondary/ trunk roads**)

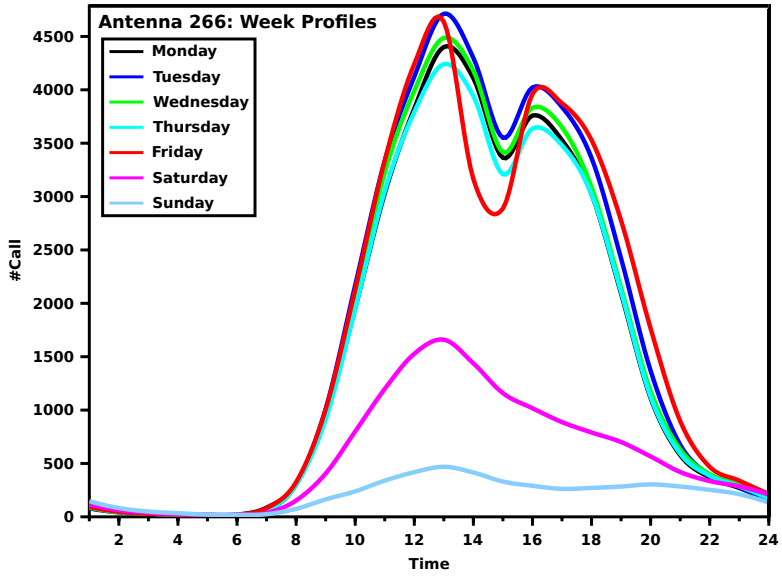
- **New Scilab module (CDRs analysis) added to the NARVAL toolbox**
  - Complete software environment enabling the understanding of available communication algorithms, but also the design of new schemes
  - Graph Optimization, Topology, Internet Traffic, Routing, Transmission Protocol, Route Diversity, Mobility, Security, Anonymity, Path Planning, Wireless Sensor Network, etc.
  - Target audience: academics, students, engineers and scientists
  - **http://atoms.scilab.org/toolboxes/NARVAL**





Conclusion

- Daily Profile Model
  - Characterization of each base station traffic (amount of calls, duration of calls and amount of text messages)
  - Classification: urban, suburban and rural modes
  - Correlation between each base station traffic and population of its covering area
- Traffic Anomalies Detection
  - Analysis of national anomalies
  - Analysis of local anomalies, e.g. inauguration of the new highway between Dakar and Diamniado
- Mobility graphs
  - Computation of mobility flows
  - Performance of congestion maps
- Future Work
  - Separation of daily profiles into 7 daily profiles
    - Working day and weekend
  - Anomaly detection with an aggregation level of **1h**



UNIVERSITÉ DU LUXEMBOURG

UNITED NATIONS GLOBAL PULSE
Harnessing big data for development and humanitarian action

- A global open innovation initiative for climate resilience
- Now is the time to use big data to fight climate change
  - World leader will decide on a universal climate agreement at COP21, potentially the most decisive climate conference in 25 years
  - COP21 (Paris, 2015, Nov 30th to Dec 11th) is critical as the impacts of climate change are accelerating, and time to develop solutions is limited
    - Achieve a new international agreement on the climate, applicable to all countries, with the aim of keeping global warming below 2°C
  - Analyzing anonymised, aggregated data from digital sources such as mobile phones and bank transactions can provide valuable insights on human behavior and climate risks.

COP21 • CMP11
PARIS 2015
UN CLIMATE CHANGE CONFERENCE

UNIVERSITÉ DU LUXEMBOURG

- Thank You!
- Acknowledgement
  - This Data was made available by ORANGE/SONATEL within the framework of the D4D Challenge. The author would like to thank Orange and SONATEL for the availability of these CDR's datasets.
  - The author would like to thank the National Research Fund of Luxembourg (FNR) for providing financial support through the CORE 2013 MAMBA project (C13/IS/5825301)
- Contact
  - foued.melakessou@uni.lu