Data for Development Challenge Senegal

Book of Abstracts: Posters

At NetMob 2015

Organized by

Universidad Carlos III de Madrid

Sponsored by

RealImpact ANALYTICS

www.d4d.orange.com / Tweeter : @O4Dev
Contact: Nicolas De Cordes, Orange, VP Marketing Anticipation, nicolasdecordes@orange.com

Contents :

Session 1 : Agriculture

1. [A01] Mobility profiles and calendars for food security and livelihoods analysis 6
   Pedro Zufrina
2. [A02] Genesis of millet prices in Senegal: the role of production, markets and their failures 7
   Damien Jacques
3. [A03] Visual analysis on call data records for improving disaster resilience 8
   Brian Tomaszewski
4. [A04] Unraveling correlations between agricultural events and phone traffic 9
   Rosa Benito

Session 2: Energy

5. [E01] Using mobile phone data for rural electrification planning in developing countries 10
   Markus Schläpfer

Session 3: Health

6. [H01] Quantifying the effect of movement associated with holidays on malaria prevalence using cell phone data 11
   Sveta Milusheva
7. [H02] Uncovering the impact of human mobility on schistosomiasis via mobile phone data 12
   Marino Gatto
8. [H03] Impact and scale of mobility on the spatial epidemiology of tuberculosis 13
   Bruno Lepri
9. [H04] Human contact and the diffusion of Ebola in Senegal 14
   Rich Ling
10. [H07] Human mobility and the spreading of waterborne diseases 15
    Enrico Bertuzzo
11. [H08] Progmosis: evaluating risky individual behavior during epidemics 16
    Antonio Lima
12. [Ho9] Developing an agent based migration model for Senegal for malaria transmission 17
    Adrian Tompkins

13. [H10] Forecasting Influenza in Senegal with call detail records 18
Hao Wu
14. [H11] Mobile data as public health decision enabler: a case study of cardiac and neurological emergencies 19
Edward Mutafungwa
15. [H12] Modeling Ebola virus diffusion in Senegal using mobile phone datasets and agent-based simulation 20
Jonathan Leidig

Session 4: National Statistics

16. [N01] Towards connecting people, locations and real-world events in a cellular network 21
Ramona Trestian
17. [N02] Spatial structure and efficiency of commuting in Senegalese cities 22
Thomas Louail
18. [N03] Construction of socio-demographic indicators with digital breadcrumbs 23
Timo Schmid
19. [N04] Mapping and Measuring of social disparities in Senegal using mobile phone subscribers data 24
M. Saravanan
20. [N05] Detection of population mobility anomalies in Senegal from base station profiles 25
Foued Melakessou
21. [N07] Virtual networks and poverty analysis in Senegal 26
Neeti Pokhriyal, Wen Dong, Venu Govindaraju
22. [N09] Estimating population density and climate vulnerability using CDRs 27
Emmanuel Letouze

Session 5: Transports / Urbanism

23. [T01] Roads+: estimation of travel times between cities in Senegal from CDR's 28
Jari Sarämaki
24. [T02] Deviations from the norm: detecting regularities and anomalies in human mobility patterns 29
Gijs Joost Brouwer
25. [T03] National and regional road network optimization for Senegal using mobile phone data 30
Erik De Romph

26. [T04] Building workers' travel demand models based on mobile phone data 31
Liu Feng
27. [T05] Urban road construction and human mobility: evidence from Dakar 32
Amadou Sy
28. [T06] Travel demand analysis with differentially private releases 33
David Gundlegard
29. [T07] Cars and Calls: using CDR data to approximate official traffic counts 34
Vanessa Frias-Martinez
30. [T08] Towards mobile social crowd-sensing for transport information management 35
Georgios Bouloukakis
31. [T09] Operational planning of transport & infrastructure using mobile data analysis 36
Ankit Singla
32. [T10] The municipality and the territory: two scales of understanding the city 37
Jerome Chenal
33. [T11] Sanitization of call detail records via differentially-private summaries 38
Mahammed Tuhin
34. [T12] A cell dwell time model for synthetic population generation from call detail records 39
Foued Melakessou
35. [T13] Improving mobility predictability based on user clustering on D4D Dataset 40
Jaeseong Jeong
36. [T14] From digital footprints to the dynamic population distribution and road network efficiency 41
Ruiqi Li, Dong Lei
37. [T15] Using mobile phone data for spatial planning simulation and optimization technologies 42
Serigne Gueye
38. [T16] Where do we develop? Discovering regions for urban investment in Senegal 43
Derek Doran
39. [T17] Understanding traffic matrix for transportation planning with interregional connectivity 44
Seungho Seungho Kim
40. [T18] High resolution traffic maps generation using cellular Big Data 45
Ahmed El Mahdy
41. [T19] Structure and resilience of call networks in Senegal 46
Anil Vullikant

42. [T20] Exploring relationships between human mobility motifs and rainfall 47
Akihiro Fujihara
43. [T21] Text me - if you can: literacy, networks and mobility in Dakar 48
Martina Kirchberger, Christopher Small
44. [T22] Cellphone's users trajectories clustering from call detail records 49
Paulin Melatagia Yonta

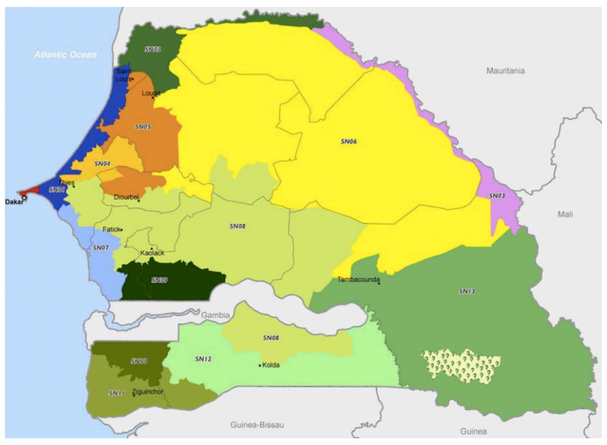
Session 6: Others

45. [O01] Generating anonymous sequence preserving datasets 50
Sigal Shaked
46. [O02] On the anonymizability of mobile traffic datasets 51
Marco Gramaglia
47. [O04] Human mobility during religious festivals in Senegal 52
Christelle Scharff
48. [O05] Data for Development reloaded: visual matrix techniques for the exploration and analysis of massive mobile phone data 53
Stef Van Den Elzen
49. [O06] Adaptive power load balancing in cellular networks 54
Petko Bogdanov
50. [O07] Senegal Time-Space 2013 55
Christopher Small
51. [O08] Call data network structure, revealed mobility & social networks 56
David Meyer

A01

Mobility profiles and calendars for food security and livelihoods analysis

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network

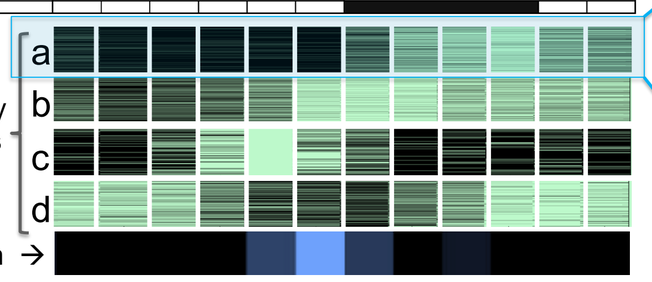


Livelihood zones

- SN01 - Niayes Horticulture and Fishing Zone
- SN02 - Senegal River Valley: Rice and Gardening Zone
- SN03 - Senegal River Valley: Out-migration and Remittence Zone
- SN04 - Agropastoral: Cassava Zone
- SN05 - Agropastoral: Cowpea Zone
- SN06 - Sylvo-pastoral
- SN07 - Petite-Côte: Fishing, Tourism and Arboriculture Zone
- SN08 - Agropastoral: Peanut Zone
- SN09 - Agriculture Zone
- SN10 - Food Crops and Forestry Zone
- SN11 - Agroforestry, Fishing and Tourism Zone
- SN12 - Agro-sylvo-pastoral: Peanuts and Cotton Zone
- SN13 - Agro-sylvo-pastoral: Food Crops Zone
- Not Zoned: Urban Area

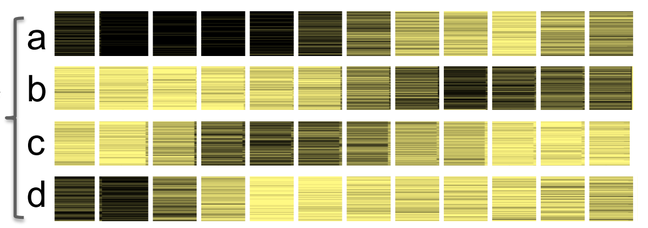
Zone 8: Agro-pastoral Peanut Zone

Source of income/activity	January	February	March	April	May	June	July	August	Septembe	October	November	December
millet, cowpea, watermelon												
Peanuts												
Maize												
Sorghum												
Cattle/poultry sales												
Baobab gathering												
Jujube gathering												
Salt harvesting												
Market gardening												
Transhumant livestock rearing												
Trade												
Crafts												
Seasonal migration												
International migration												
Local labor												

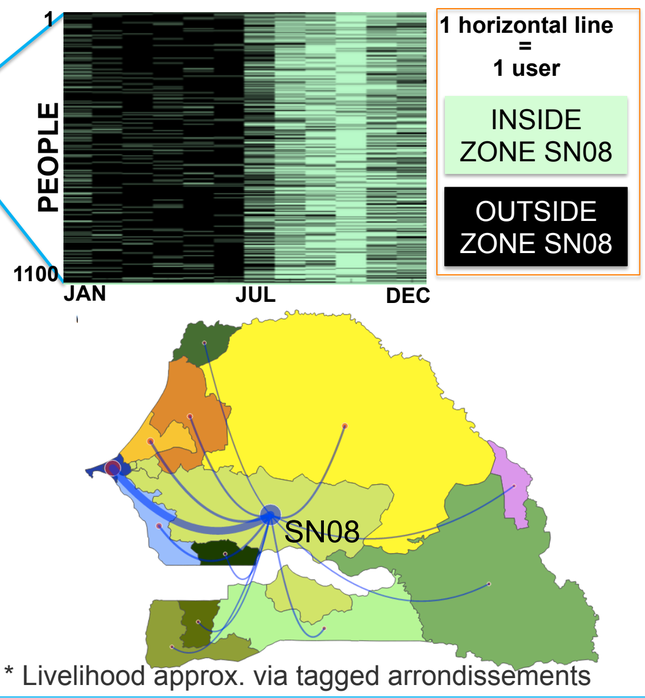


Zone 6: Sylvo-pastoral Zone

Source of income/activity	January	February	March	April	May	June	July	August	Septembe	October	November	December
Cattle/poultry sales												
Milk sales												
Transhumant livestock rearing												
Wild food gathering												
Petty trade												
Millet												



Population subgroup with similar mobility behavior during the year



* Livelihood approx. via tagged arrondissements

Pedro J. Zufiria, David Pastor-Escuredo, Luis Úbeda-Medina, Miguel A. Hernández-Medina, Iker Barriales-Valbuena, Alfredo J. Morales
Universidad Politécnica de Madrid

John Quinn, Paula Hidalgo-Sanchis, Miguel Luengo-Oroz
Pulse Lab Kampala, United Nations Global Pulse

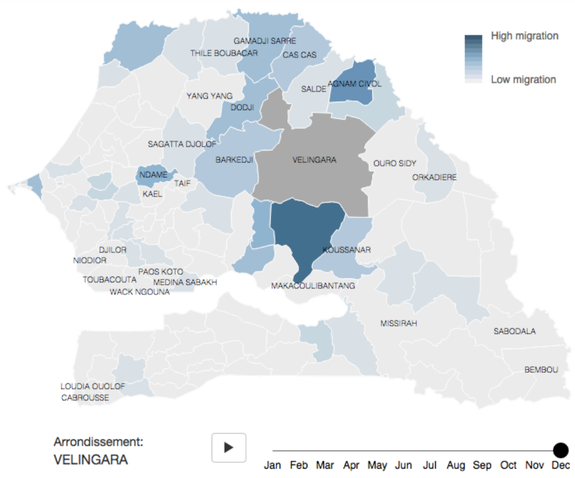
Wilfred Nkwambi, **World Food Programme Senegal**

Summary

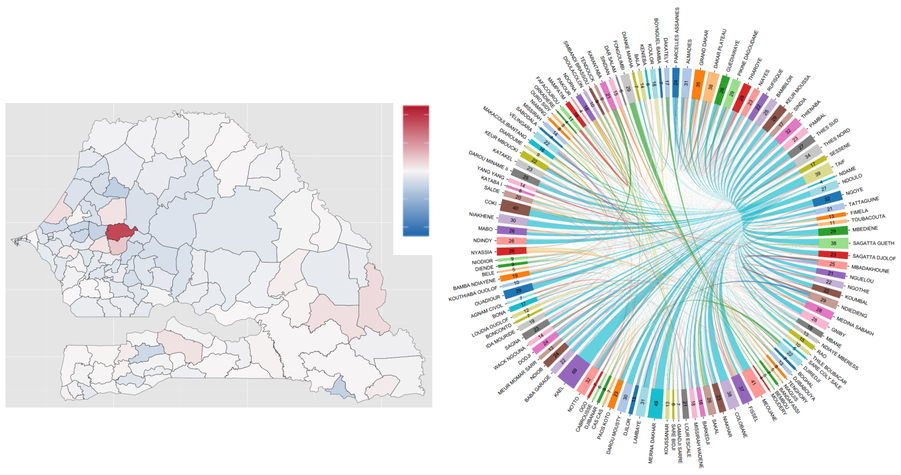
Senegal has a number of livelihood zones, where pastoralism, agriculture or fishing, for example, are the main activities. We have developed statistical measures for mobility profiling in the context of the livelihood zones and seasonal activity patterns of Senegal. The mobility information at the livelihood level has been packed into mobility calendars accessible for food security experts and aligned with agricultural and seasonal activity calendars.

Use for development

For vulnerable population groups, changes in mobility patterns can indicate a change in livelihoods or coping strategies as a result of shocks. Monitoring changes in mobility patterns can thus be a powerful early warning mechanism. Statistics about the movements of different population groups, and in particular deviations from normal seasonal patterns, can be a new quantitative dimension of analysis for the Seasonal Monitor reports used for decision making and crisis management.



Arrondissement-centered mobility profiling



National level population movements

Main results:

- We have developed a methodology to establish a baseline quantification of the mobility patterns related to livelihood zones for the year 2013. We show that different regions of the country exhibit mobility patterns that change by season, and illustrate the effects of particular known events on movements.

- We have implemented an interactive online dashboard to prototype the migration calendars at livelihood and arrondissement level that can be aligned and compared with seasonal activity calendars.

Methods:

- For each of the 13 “livelihood zones” and arrondissements, we have segmented the population according to their mobility profiles by clustering individual mobility trajectories along the year into mobility classes. Each user has been characterized by a 12xN - eg. “Each month “1” if the user is inside the region of interest and “0” outside” (based on the monthly estimation of home location).

- We have computed aggregated population daily movements between all arrondissements at national level and we have characterized/filtered populations according to displacements and/or bandicoot indicator properties.

- Average rainfalls at different geographical and temporal resolutions using remote sensing information have been calculated.

Full paper:
http://erdos.mat.upm.es/d4d-senegal/research-report.pdf

Interactive data visualizations:
http://pulselabkampala.ug/d4d-senegal/
http://goo.gl/QJKfl4
http://goo.gl/aZ6UJu
http://erdos.mat.upm.es/d4d-senegal/rain.mp4

Data sources:
+ D4D DATASET 3- Movement at low resolution
+ SENEGAL LIVELIHOOD ZONES MAP
http://fews.org/pages/remote-monitoring-country.aspx?gb=sn&l=en
+ COMPREHENSIVE FOOD SECURITY AND VULNERABILITY ANALYSIS (CFSVA), SENEGAL
http://www.fews.net/sites/default/files/documents/reports/sn_livelihoodzonedescriptions2011_en.pdf
+ RAIN DATA, TRMM-NASA
http://disc.sci.gsfc.nasa.gov/recipes/?q=datacollection/TRMM_3B42_daily.007/description

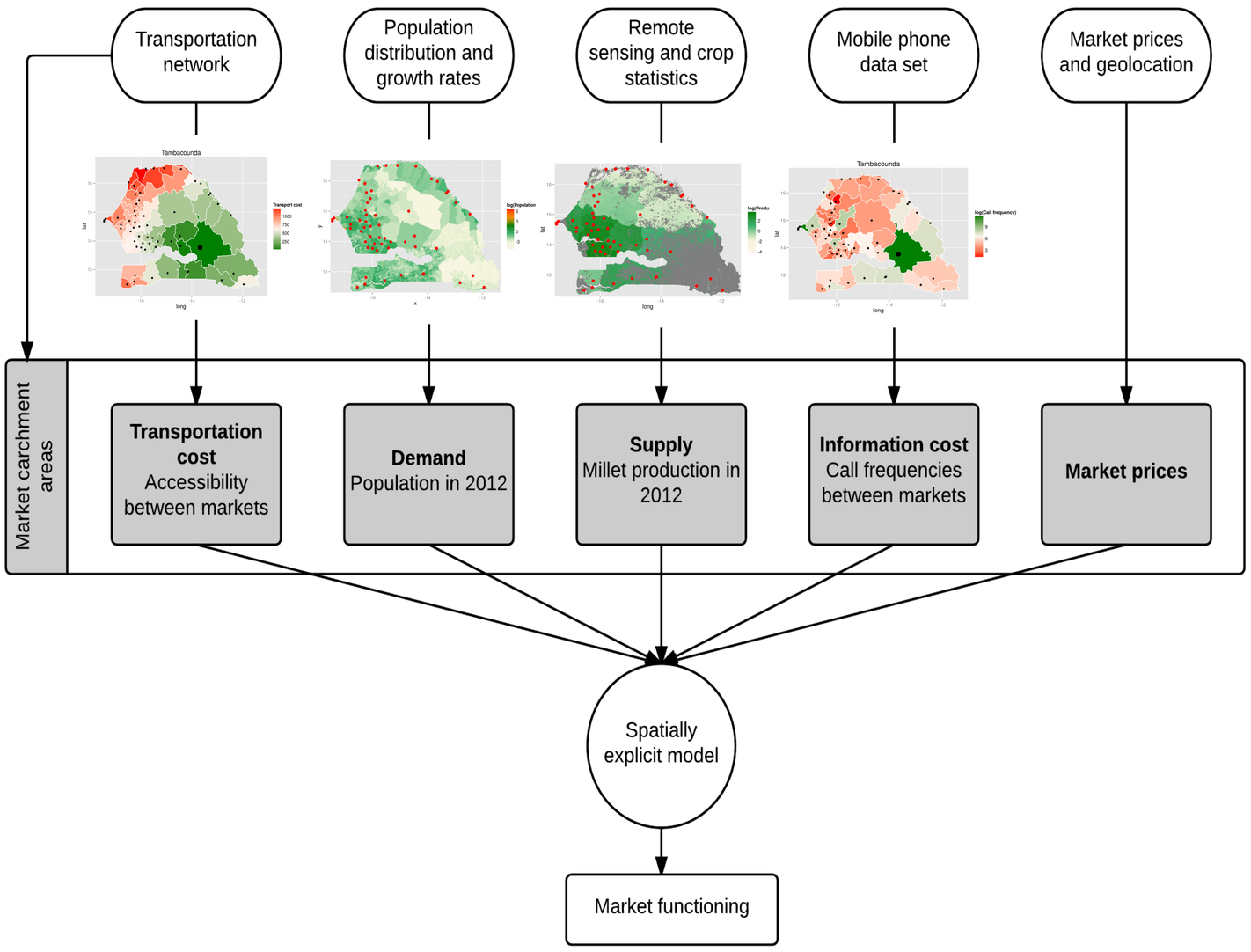
Main Tools used:
- Python
- Matlab
- R
- Algorithms: clustering (hierarchical), times series processing
- Visualization: d3.js

Open Code available upon request

A02

Genesis of millet prices in Senegal: the role of production, markets and their failures

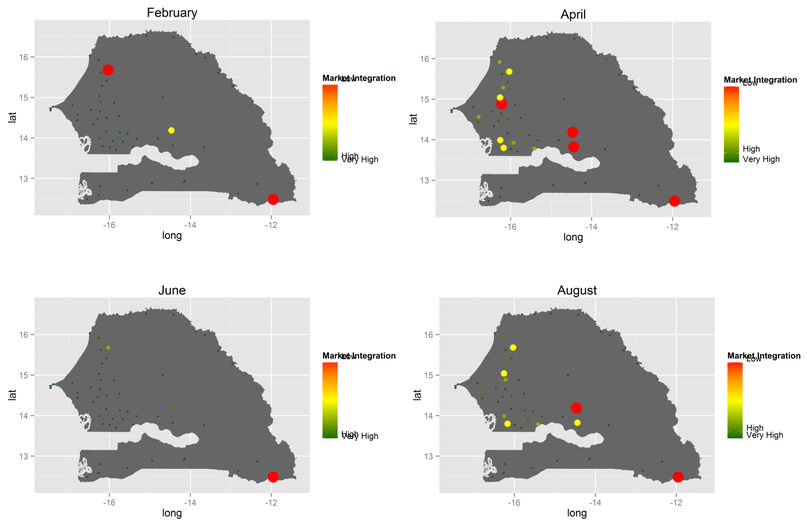
Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



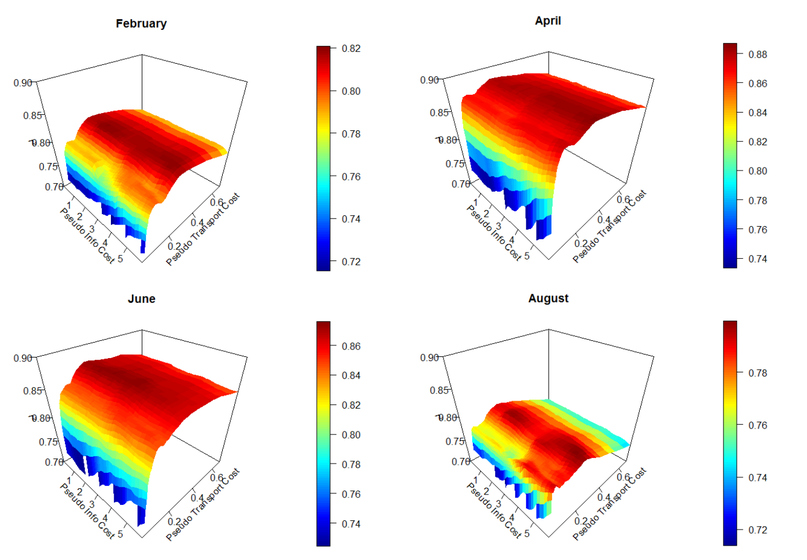
Jacques (Ir, UCL), d'Andrimont (Ir UCL), Radoux (PhD UCL), Waldner (Ir UCL) and Marinho (PhD)

Project Summary:

Staple prices, the main indicator of food access and a key determinant of the revenues of the poorest, result from the equilibrium between supply and demand at the local and regional levels. This work simulates the millet prices formation process in Senegal in a spatially explicit model that integrates the markets failures emerging from transportation costs and information asymmetries.
The findings suggest that, thanks to the recent development of the mobile phone network in the country, information asymmetries are responsible for price differentials in only the few areas where the mobile phone coverage has not yet reached its full potential, while transportation costs remain an important burden for food security.



Market integration derived from the mobile phone data



Correlation between pseudo-prices and actual millet price

Main results:

We model the millet prices formation process in Senegal in a spatially explicit model that accounts for both the per kilometer transportation cost and the information asymmetry resulting from low levels of mobile phone activity between markets. The model integrates a unique and diversified set of data in a framework coherent with economic theory. It explains more than 80% of the price differentials observed in the 40 markets. These results can be used in the assessment of the social welfare impacts of the further development of both road and mobile phone networks in the country. The model could be further developed as a valuable tool for the prediction of future staple prices in the country.

Methods:

The departure point is a local demand and supply model around each market having its catchment areas determined by the road network. We estimate the local supply of agricultural commodities from satellite imagery while the demand is assumed to be a function of the population living in the area. From this point on, profitable transactions between areas with low prices to areas with high prices are simulated for different levels of per kilometer transportation cost and information flows (derived from the mobile phone data). The simulated prices are then compared with the actual millet prices.

Full paper is here:
-

DataViz or video are here:
-

Login:
Pw:

Data sources used for this project:
- D4D data set 1, com between antenna ■
- D4D data set 2, movement routes high res ☐
- D4D data set 3, movement routes low res ☐
- D4D synthetic data set ☐

Other data sets used in this project:
- Type of data: Production Source: SPOT-veg, Ministry of economy
- Type of data: Demography Source: Afripop
- Type of data: Market prices Source: UN WFP
- Type of data: Road network Source: Global Insight

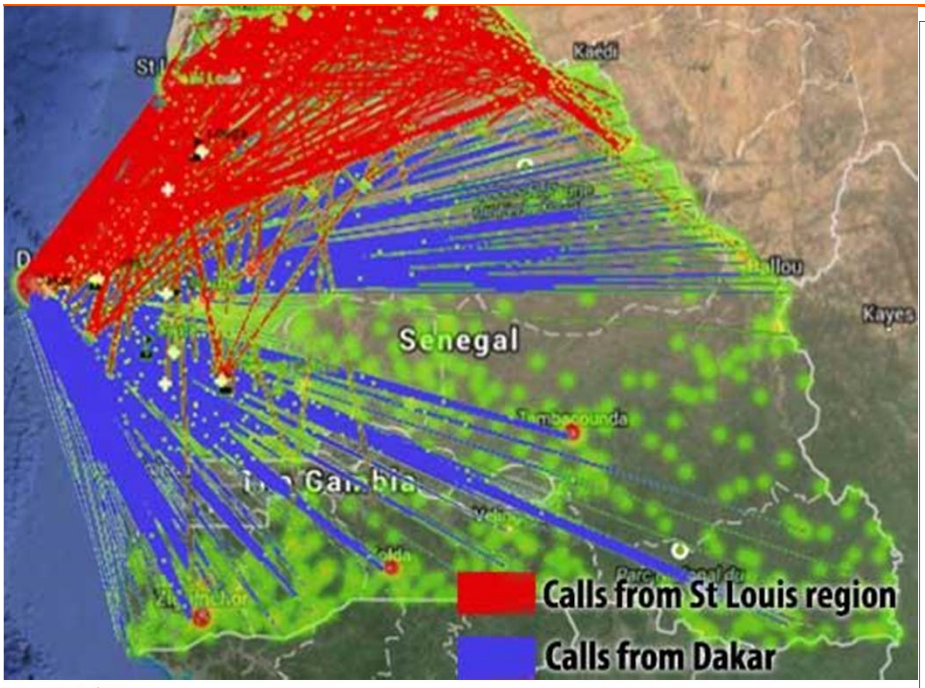
Main Tools used:
- Remote sensing
- Accessibility analysis
- Spatially explicit modelling

Open Code available:
- Yes ■
- No ☐

A03

Visual Analysis on Call Data Records for Improving Disaster Resilience

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



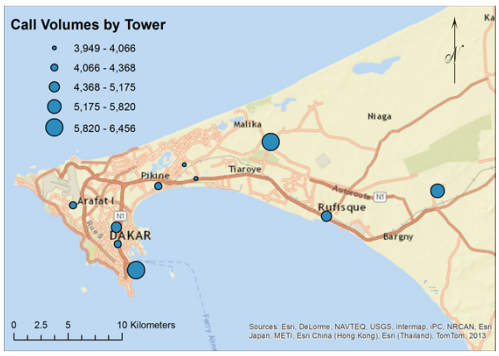
- Chitturi S, Davis J, Farooqui S, Gohel J, Gonsalves S, Gunda R, Raina A, Sawant A, Sawant T, Vijayasekharan A, Wasani J, and Tomaszewski, B.
- Department of Information Sciences and Technologies
- Rochester Institute of Technology, Rochester NY, USA.

Project Summary:

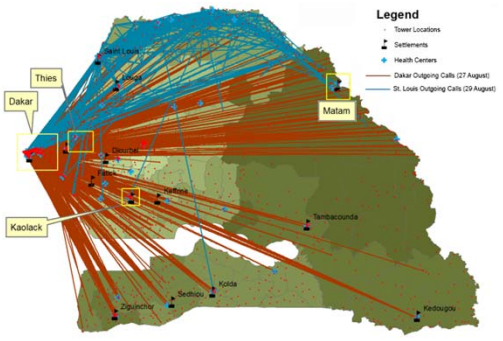
Natural disasters like floods have compounded effects on agriculture, which in turn, result in food insecurity and malnutrition. The Senegalese population is particularly vulnerable to floods due to existing poverty. We analyzed Data for Development (D4D) Call Detail Records (CDR) datasets to (1) find statistically significant spatial clusters of areas vulnerable to floods and (2) identify spatial interactions between call origins and destinations to understand calling behaviors during flooding events.

Possible use for development:

Identifying spatial interactions during floods using CDRs can help build resilience to natural disasters and related events such as food insecurity and other flood-related health issues.



Top call volumes during 2013 floods.



Dakar and Saint Louis flood call patterns.

Main results:

- A significant pattern was found in analyzing Call Data Records (CDR) dataset 1 (Antenna to Antenna) with respect to specific flood dates in Dakar and Saint Louis. The results for the pre- and post- flood dates for Dakar and Saint Louis yielded a visual pattern which showed a drastic outreach to regions such as Kaolack, Thies and Louga identified through hot-spot analysis.
- We wish to suggest that the reason for this behavior is that people affected by floods were likely contacting relatives, friends or other people more frequently than normal in order to provide social or financial support as this is a very common behavior during disasters.

Methods:

- A database schema design was implemented that mirrored the structure of each file provided by Orange. Each .csv file in the dataset had a corresponding table created in the database. These files were then loaded into their appropriate table using SQL Server Business Intelligence Development Studio. We leveraged views to abstract table complexities and provide a cleaner virtual table containing all the information needed for further in-depth analysis.
- A geo-database was created with different reference datasets such health centers, cellular tower locations, settlement areas and nutrition. Using Geographic Information Systems (GIS), these reference datasets were then overlaid with CDRs extracted from our central database for specific flood dates identified for Dakar and Saint Louis. We used the XY to Line Management tool of ArcGIS to identify call tower origin and destination pairs for the flood dates using tower latitude and longitude coordinates combined with relevant call data (duration, call date, etc.). The Getis-Ord-Gi* spatial statistic was then used to identify the statistically significant clusters or "hot spots" where the most calls were being made during flood dates using outputs from the XY to Line tool.

Full paper QR code

Full paper is here:
http://geoapps64.main.ad.rit.edu/d4d2014/RIT_D4D_Senegal_Scientific_Paper_31_December_2014.pdf

Data sources used for this project:

- D4D data set 1, com between antenna [x]
- D4D data set 2, movement routes high res [ ]
- D4D data set 3, movement routes low res [ ]
- D4D synthetic data set [ ]

Other data sets used in this project:

- Type of data: GIS Reference Datasets for Senegal
  Source: UN OCHA Common Operational Datasets (CODs)
  URL: http://www.humanitarianresponse.info/applications/data/datasets/locations/senegal

Main Tools used:

- MS SQL Server
- ArcGIS 10.2
- Custom Python Algorithm for data retrieval

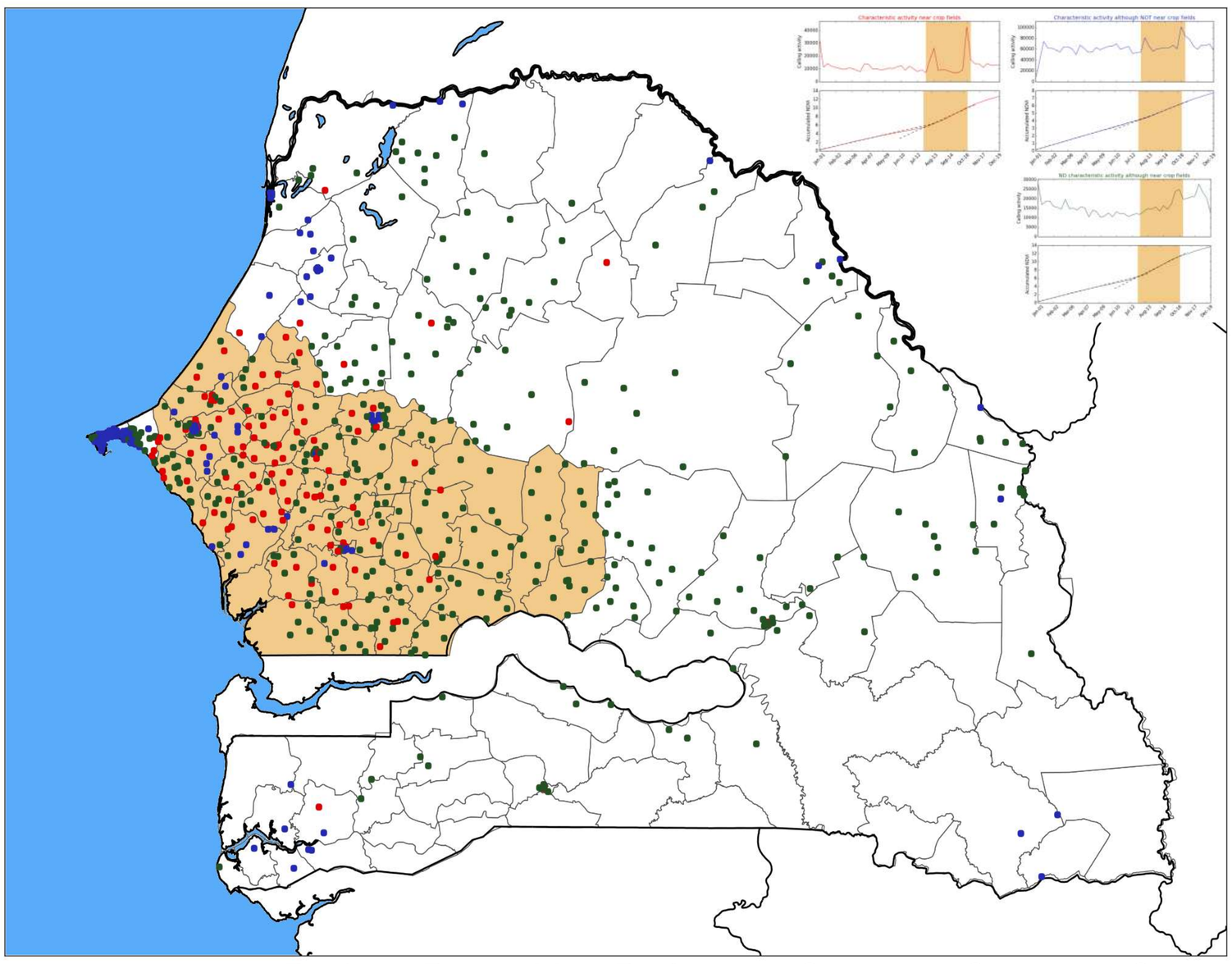
Open Code available:

- Yes [ ]
- No [x]

A04

Unraveling correlations between agricultural events and phone traffic

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



GRUPO DE SISTEMAS COMPLEJOS
UNIVERSIDAD POLITÉCNICA DE MADRID

POLITÉCNICA "Ingeniamos el futuro" CAMPUS DE EXCELENCIA INTERNACIONAL

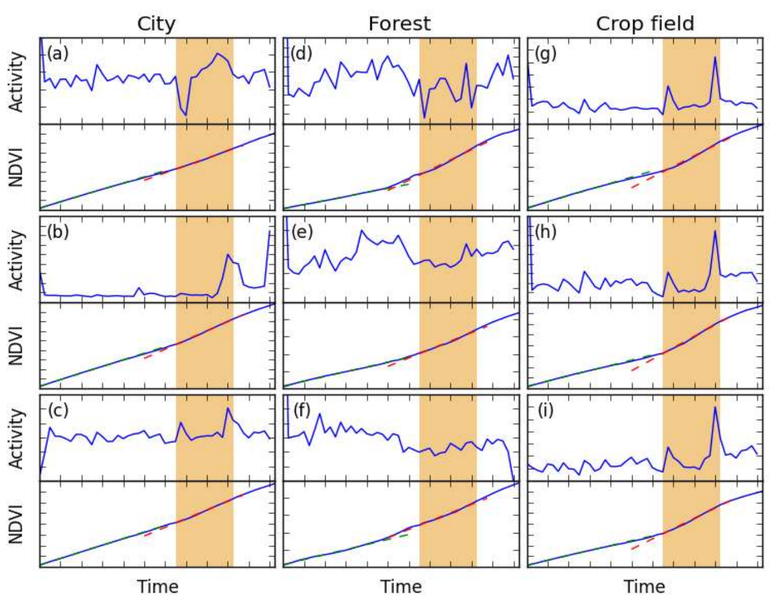
- Benito, Rosa M., Professor, Complex Systems Group, Technical University of Madrid (lead author)
- Martín, Samuel, BS, Borondo, Javier, Msc, Morales, Alfredo, PhD, Losada, Juan C., PhD, Tarquis, Ana M., Complex Systes Group, Technical University of Madrid

Project Summary:

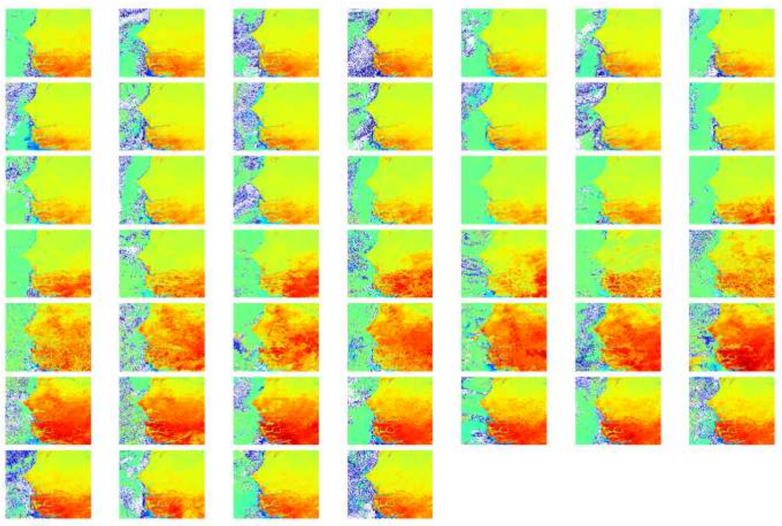
We have found differential patterns of phone calls activity during the growing season of groundnuts in antennas installed close to crop fields that allow us to distinguish them from those installed in cities and forests. Correlations with the underlying agricultural and economic context have been detected.

Possible use for development:

The identification of different patterns in the communication during the growing seasons allow us to detect the regions of high activity and mobility. It will help the government to improve planning transportation and employment policies contributing to the development of the country.



Time-series of calling activity and NDVI in different locations



NDVI evolution during 2013

Main results:

- We have found that, during the growing season, a characteristic calling pattern emerges in antennas installed near crop fields allowing us to distinguish them from those located in cities or forest areas.

Methods:

- Classification of the antennas considering the behavior of their activity and vegetation index (NDVI) time-series:
  - The vegetation index time-series have been generated taking a three-by-three pixels lattice centered at the antenna coordinates and computing the average value.
  - The antennas have been classified considering whether they show characteristic activity during the growing season and whether their vegetation index time-series can be associated to crop fields.
- Built of callings networks : each antenna is a node and two nodes are connected if a call is made from on to the other.
- The community structure of the networks has been extracted with a modularity optimization algorithm.

Full paper:
www.gsc.upm.es/gsc/paper_D4D
Video of phone activity:
www.gsc.upm.es/gsc/Video_D4D
Login:
Pw:

Data sources used for this project:
- D4D data set 1, com between antenna [x]
- D4D data set 2, movement routes high res [ ]
- D4D data set 3, movement routes low res [ ]
- D4D synthetic data set [ ]

Other data sets used in this project:
Type of data:Normalized difference vegetation index.
Source:NASA's Land Processes Distributed Active Archive Center.

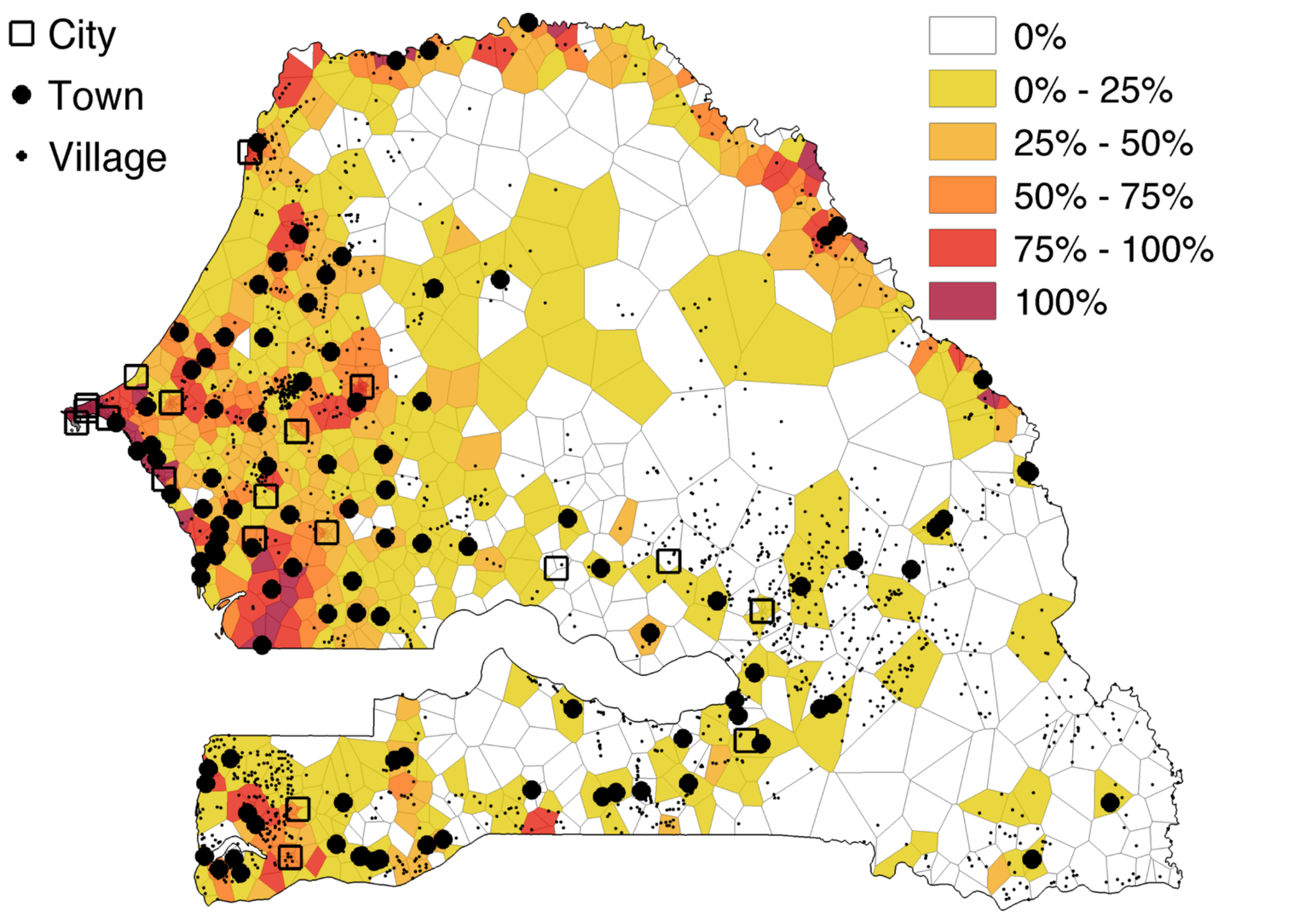
Main Tools used:
- NetworkX python module
- Peak detection algorithm (goo.gl/Edq7cW)
- Self developed code in Python

Open Code available:
- Yes [ ]
- No [x]

E01

Using Mobile Phone Data for Electrification Planning

Health	Transport Urban	National Statistics	Other
Agriculture	**Energy**	DataViz	Network



MANCHESTER 1824
The University of Manchester

SANTA FE INSTITUTE

- Dr. E. A. Martínez Ceseña, The University of Manchester
- Dr. P. Mancarella, The University of Manchester
- Dr. M. S. Schläpfer, Santa Fe Institute

Project Summary:

The dramatic increase of mobile phone use and the recent availability of the corresponding anonymized data offer unprecedented insights into human activity in Senegal. This new data can be extrapolated into valuable electricity needs, which are otherwise very scarce, particularly in rural areas.

An electrification framework based on mobile phone, electricity and geo-referenced data is developed to assess electrification throughout electricity grid extension, and installation of Diesel engine-based Microgrids and solar photovoltaic systems. The tool is used to meet all energy needs or just lighting and mobile phone charging.

The mobile phone data can support efficient and economically attractive electrification plans for Senegal.

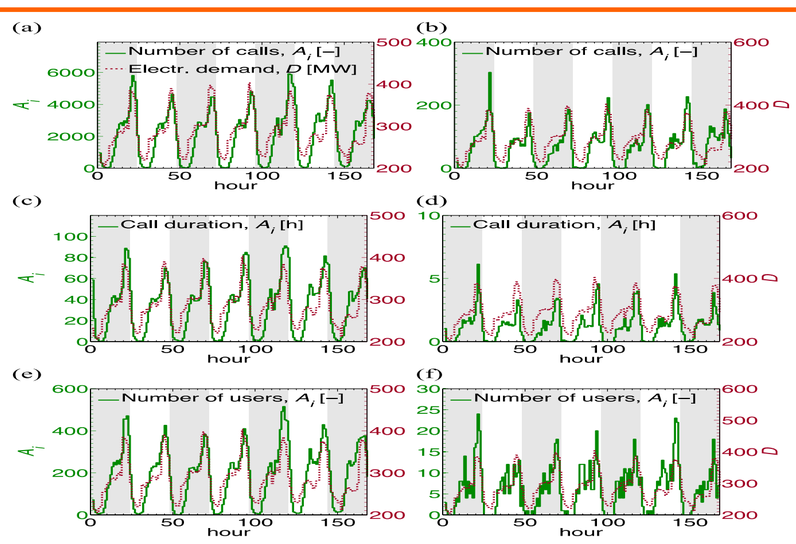


Fig. 1: Comparison of mobile activity and electricity consumption

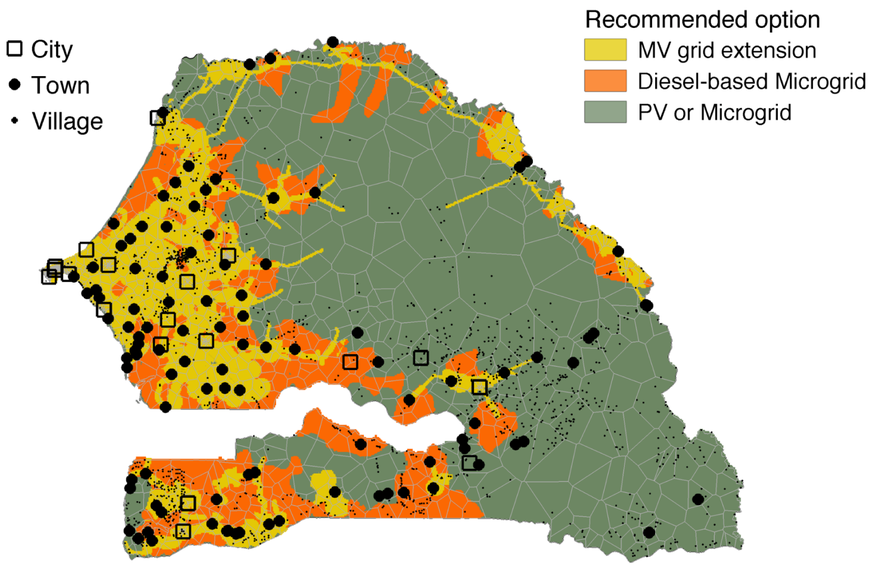


Fig. 2: Electrification recommendations

Main results:

- The mobile phone data has proven to be an accurate proxy of electricity needs (see Fig. 1 for a visual comparison), and facilitate population growth estimation in rural areas. This allows detailed assessments of electrification technologies and, thus, more economically efficient investment strategies (see fig. 2 for the recommended investments)
  - Electricity grid extensions can become more economically effective for zones with high potential population growth due to migration in the vicinity of the grid.
  - Diesel engine-based community level Microgrids become the preferred technology for most zones located far from the grid, particularly large villages or areas with significant potential population growth due to migration.
  - Individual household level PV systems are convenient for the smallest villages; particularly when only electrifying for lighting and mobile phone charging.

Methods:

- The proposed electrification planning methodology brings together mobile phone, socio-economic and geo-referenced data analysis, and state-of-the-art energy infrastructure planning engineering techniques to quantify the techno-economic feasibility of different electrification options in Senegal. Some of these studies include Pearson coefficient-based correlation assessment, discounted cash flows analysis, time series analysis, battery degradation studies, heuristic optimization and image processing.

Data sources used for this project:

■ D4D data set 1, com between antenna	X	■ D4D data set 3, movement routes low res	
■ D4D data set 2, movement routes high res	X	■ D4D synthetic data set	

Other data sets used in this project:

Description:	Source:	Description:	Source:
■ Geo-referenced population and access to electricity, health and education databases	http://modi.mech.columbia.edu	■ Hourly insolation and temperature profiles	http://www.soda-is.com/
■ Aggregated hourly electricity profiles	Senelec	■ OpenStreetMap	https://www.openstreetmap.org

Main Tools used:

- Pearson coefficient-based correlation analysis
- Discounted cash flows studies
- Time series analysis

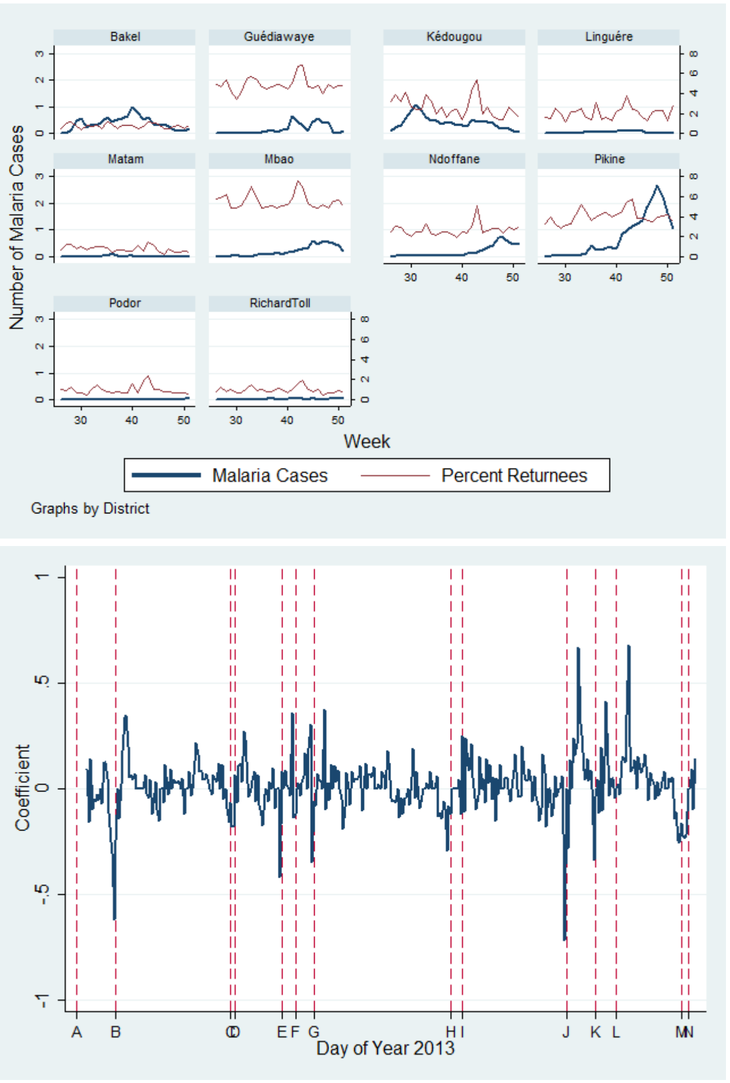
Open Code available:

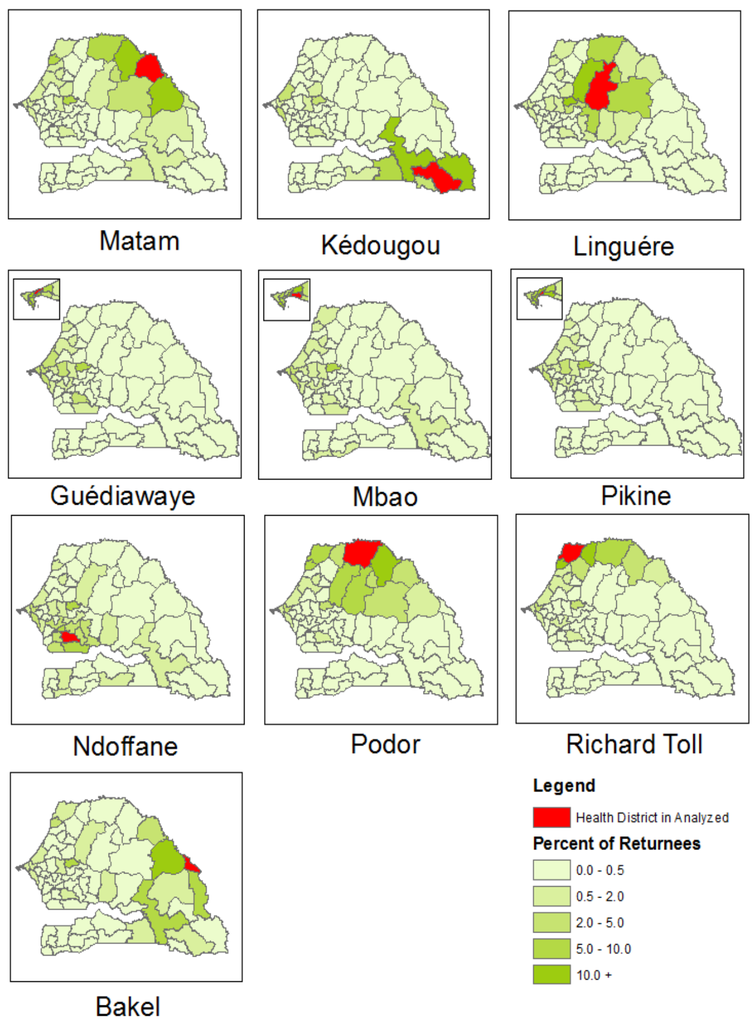
- Yes
- No X

H01

Quantifying Effect of Movement Due to Holidays on Malaria Prevalence

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network





Top Left: Malaria Cases in 10 Health Districts vs Returnees; **Bottom Left:** Short term vs Long term movement and holidays **Right:** Percent of people Returning from other health districts

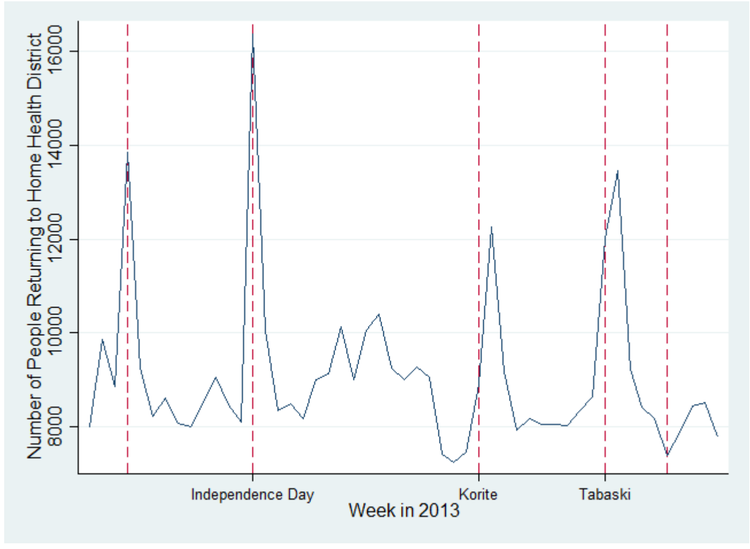
BROWN
Department of Economics
Sveta Milusheva, Brown University

Project Summary:

This paper expands previous literature by quantifying the effect of short term movement in Senegal on number of malaria cases. Using mobile phone data from the provider Orange Telecom, the study establishes short term patterns of movement around Senegal and demonstrates how they are associated with certain holidays. Using economic analysis techniques, it shows a significant correlation between these movements and malaria prevalence.

Possible use for development:

This research can be used to craft policies that target the spread of malaria during and after holidays because it shows which holidays are associated with increased travel and what locations are at highest risk. Such targeting would allow for more effective decreases at a lower cost



Number of People Returning to Home Districts after 3-14 Days

	(1)	(2) Include Low Malaria Return	(3) Include Lagged Malaria	(4) No Pikine	(5) Weeks 25-48
Return, High Malaria	3.128**	3.120**	1.136**	1.624**	3.708***
	(1.231)	(1.224)	(0.487)	(0.698)	(1.269)
Return, Low Malaria		-0.150			
		(0.145)			
Lagged Malaria Cases			0.865***		
			(0.0596)		
Constant	282.4***	284.8***	58.72	380.3***	371.4***
	(73.24)	(73.43)	(39.51)	(55.38)	(71.01)
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	230	230	230	207	190
R-squared	0.542	0.545	0.905	0.524	0.560
Number of panels	10	10	10	9	10

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Regression of Malaria Cases on People Returning Home from High Malaria Districts, 4 Lags

Main results:

- There is a large increase in short term movement around the following holidays: the Prophet's birthday, Independence Day, Korite, and Tabaski
- This movement is in the reverse direction from regular long term migration patterns, implying it represents people returning to their original homes for the holidays
- During the high malaria season of July to December, the jumps in movement seem to be associated with jumps in malaria
- Regression results confirm this association by showing a significant positive correlation between return from high malaria places and number of malaria cases in the district
- Results are robust to several checks: including returnees from low malaria places, including lagged number of malaria cases, and removing the health district that is an outlier in number of malaria cases

Methods:

- Assigned a home location to each individual in the annual low res data
- Tracked movement of 3 to 14 days for each person away from home
- Used weekly malaria data for 20 health posts around the country
- Ran regressions of total malaria cases on number of people returning to that district from a high malaria place
- Ran similar regressions with visitors and found no effects
- Conducted a robustness check by looking at people returning from low malaria places
- All regressions control for time and district location

Data sources used for this project:

- D4D data set 1, com between antenna ☐
- D4D data set 2, movement routes high res ☐
- D4D data set 3, movement routes low res ☒

Other data sets used in this project:

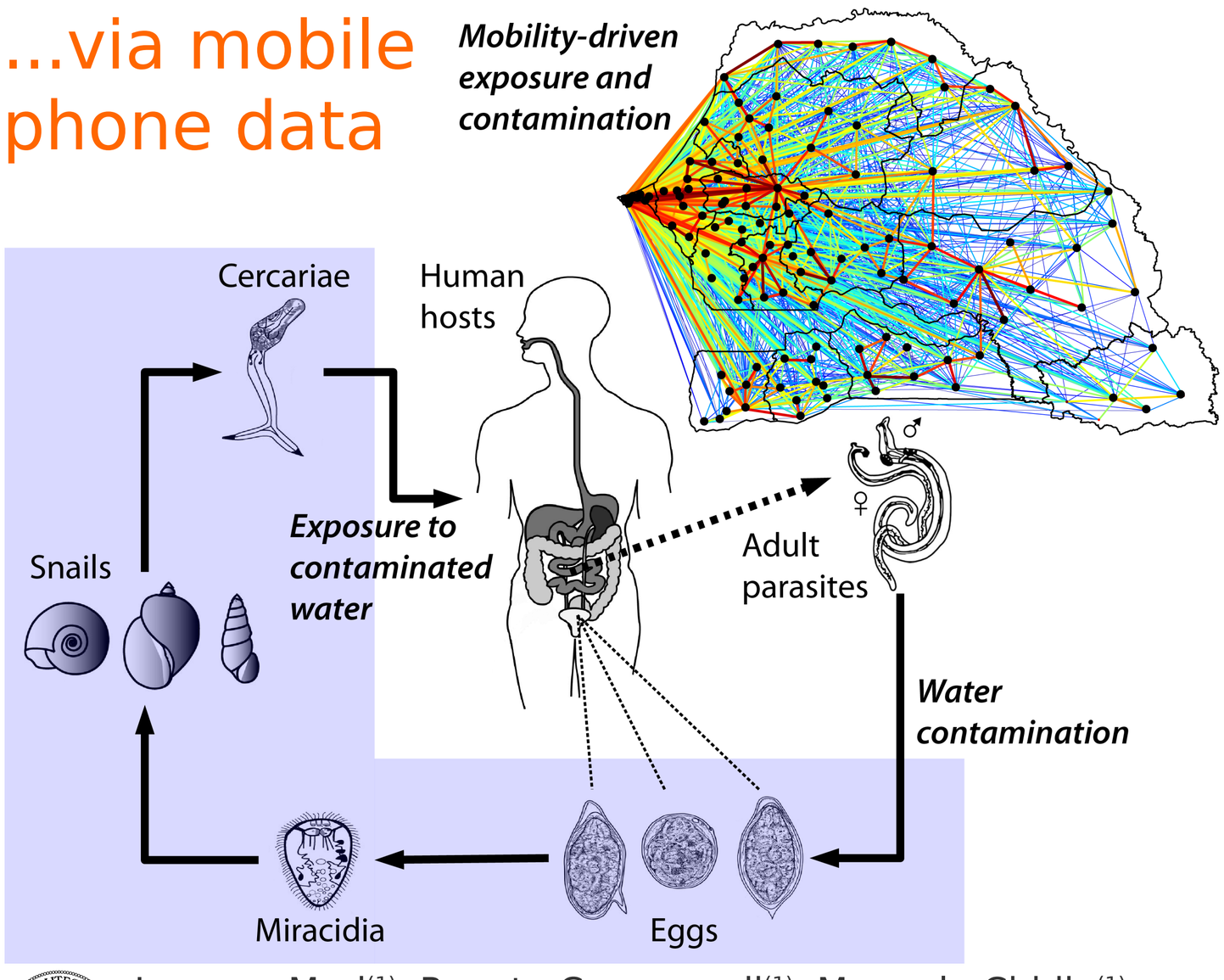
- Type of data: Weekly Malaria Data
  Source: http://www.pnlp.sn/Bulletin-Surveillance-Palu/
- Type of data: Malaria prevalence for every district
  Source: http://www.pnlp.sn/Rapports/

H02

Uncovering the impact of human mobility on schistosomiasis...

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network

...via mobile phone data



Lorenzo Mari[(1)], Renato Casagrandi[(1)], Manuela Ciddio[(1)], Susanne H. Sokolow[(2)], Giulio De Leo[(2)], Marino Gatto[(1,*)]

[(1)] Politecnico di Milano, Italy [(2)] Stanford University, USA

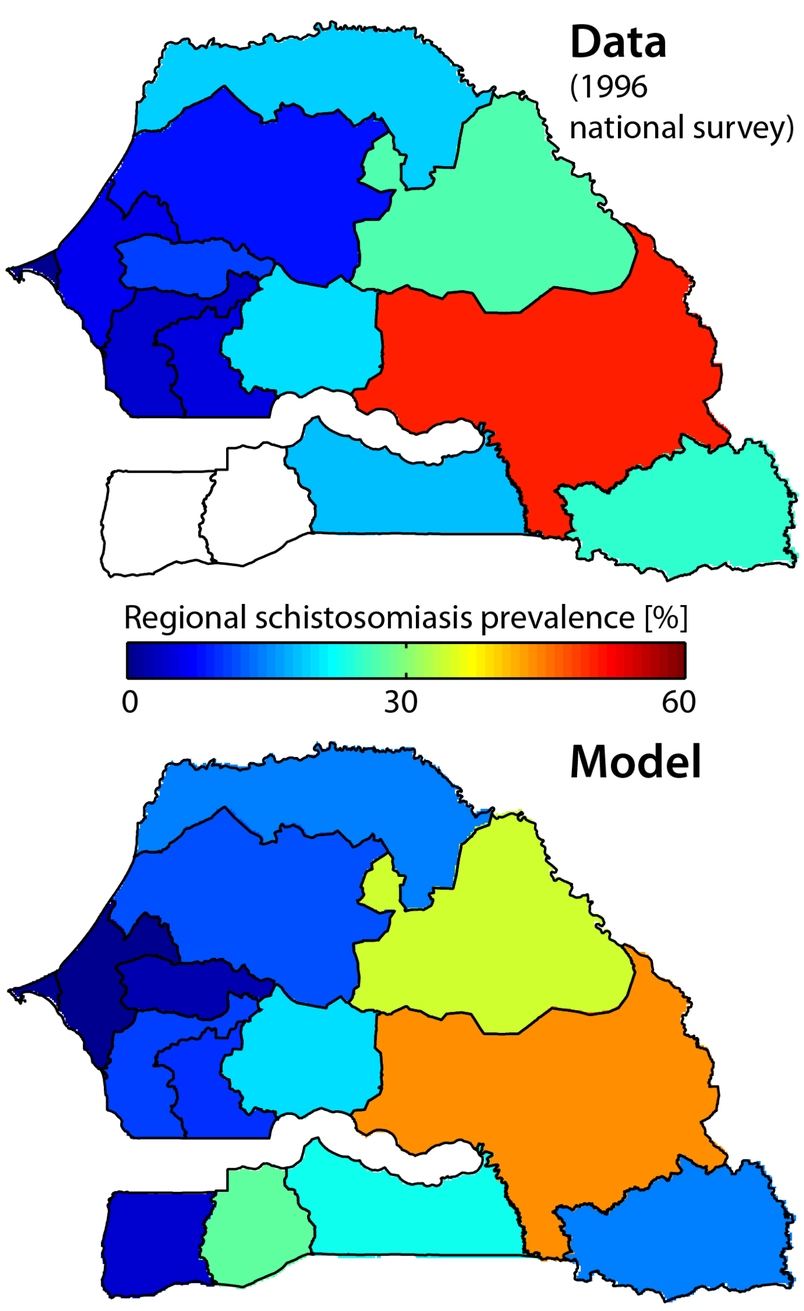
[(*)] marino.gatto@polimi.it

Project summary:

- We study schistosomiasis transmission in Senegal accounting for both local epidemiological processes and human mobility
- We show that a relatively simple model can reliably reproduce regional patterns of schistosomiasis prevalence across the country
- We use the model to study the role of human mobility on disease dynamics and to analyze intervention strategies aimed at reducing disease burden

Possible use for development:

- A thorough understanding of disease transmission dynamics is fundamental to designing effective plans for the fight against schistosomiasis, a parasitic infection with chronic debilitating symptoms that represents a major health problem in Senegal
- Our modeling framework represents a first step towards the implementation of a quantitative decision-support tool to help eradicate schistosomiasis from Senegal



Main results:

- The model is able to reproduce the regional patterns of schistosomiasis prevalence throughout Senegal quite accurately (Pearson's $r = 0.89$)
- Human movement plays an important (and nontrivial) role in schistosomiasis transmission : at small spatial scales mobility can either increase or reduce infection risk, with the latter effect being predominant at large spatial scales
- Structural interventions and educational campaign, respectively aimed at improving access to safe water/sanitation and promoting awareness, can contribute to schistosomiasis control and eradication

Methods:

- Schistosomiasis dynamics are described by means of a spatially-explicit model for macroparasite transmission accounting for both epidemiological processes and human mobility
- Mobility-driven exposure and contamination are evaluated from low-resolution movement routes of anonymous mobile phone owners : specifically, they are assumed to be proportional to the time spent in a given administrative unit, as estimated from call detail records
- Georeferenced data on demography, water supply/sanitation and schistosomiasis prevalence are used for model calibration
- The effects of human mobility and different intervention strategies are evaluated via sensitivity analysis

The full paper can be found here:
http://tinyurl.com/d4dsenegal

Username: **d4d**
Password: **ch@ll3ng3!**

Data sources used for this project:

- D4D data set 1, com between antenna [ ]
- D4D data set 2, movement routes high res [ ]
- D4D data set 3, movement routes low res [x]
- D4D synthetic data set [ ]

Other data sets used in this project:

- Population distribution (AfriPop/WorldPop)
- Water and sanitation (Global Atlas of Helminth Infections)
- Schistosomiasis prevalence (1996 national survey, GAHI)

Main tools used:

- Spatially-explicit model
- GIS manipulation
- Data mining
- Parameter calibration
- Sensitivity analysis

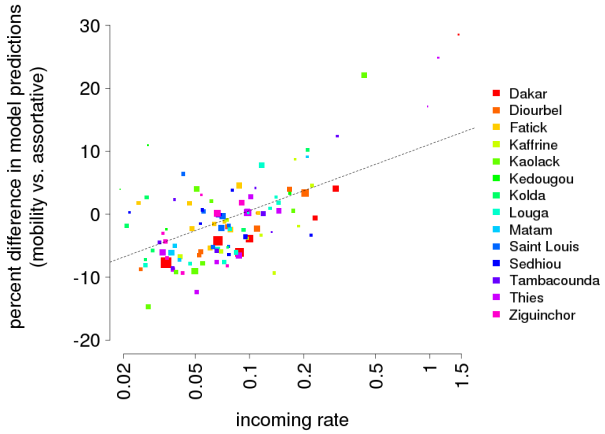
Open code available:

- Yes [ ]
- No [x]

H03

Impact and scale of mobility on the spatial epidemiology of tuberculosis

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



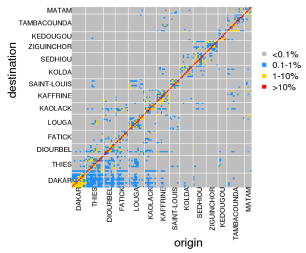
- Giorgio Guzzetta, Fondazione Bruno Kessler/TrentoRise
- Jacopo Stajano, Fondazione Bruno Kessler
- Marco Ajelli, Fondazione Bruno Kessler
- Denise Kirschner, University of Michigan Medical School
- Bruno Lepri, Fondazione Bruno Kessler
- Stefano Merler, Fondazione Bruno Kessler

Project Summary:

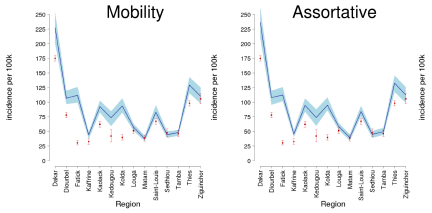
We investigated the use of Call Detail Records (CDRs) to identify human mobility patterns in Senegal and we evaluate their impact on the spatial epidemiology of tuberculosis (TB) through a mathematical model.

Possible use for development:

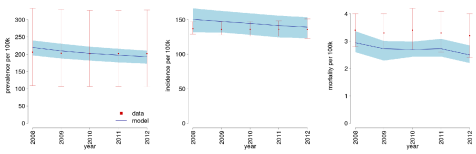
Understanding spatial transmission of tuberculosis may help optimize the localization of control efforts and increase accessibility to health services, eventually contributing to reduce the burden of TB through a reduction in diagnostic times and an increase in adherence to treatment.



1. Mobility matrix



2. Impact of mobility at regional scale



3. Calibration

Main results:

- Individuals in Senegal spend the majority of their time in the district (arrondissement) they live in, and most of the remaining time in districts within the same region (see Fig. 1, Mobility matrix) .
- Human mobility does not influence TB transmission at the regional scale, as a model with mobility is substantially equivalent with one having within-district transmission only (see Fig. 2, Impact of mobility at regional scale)
- Mobility has impact on district-level incidence (see main figure above):
  - strong positive correlation between the amount of incoming individuals and predicted incidence
  - strong negative correlation between the amount of outgoing individuals and predicted incidence

Methods:

- For the mobility matrix:
  - Each subject is assigned to the district of his last daily communication.
  - Movement between districts occurs at the midpoint between the time spent of two consecutive communications in different districts
  - We compute the mobility matrix as the average of the percentage of time spent by the subject in each district.
- The epidemiological model considers 9469 squared cells, each cell is assigned to a district and region based on geographical coordinates and is calibrated to TB prevalence, incidence and mortality in Senegal in 2008-2012 (see Fig. 3, Calibration)

Full paper is here:
put your link here

DataViz or video are here:
put your link here
Login:
Pw:

Data sources used for this project:
- D4D data set 1, com between antenna ☐
- D4D data set 2, movement routes high res ☐
- D4D data set 3, movement routes low res ☒
- D4D synthetic data set ☐

Other data sets used in this project:
- Population density (SEDAC-CIESIN, University of Columbia)
- Demographic data (Demographic and Health Surveys Program)
- Epidemiological data (WHO, UNAIDS, Senegal Ministry of Health)

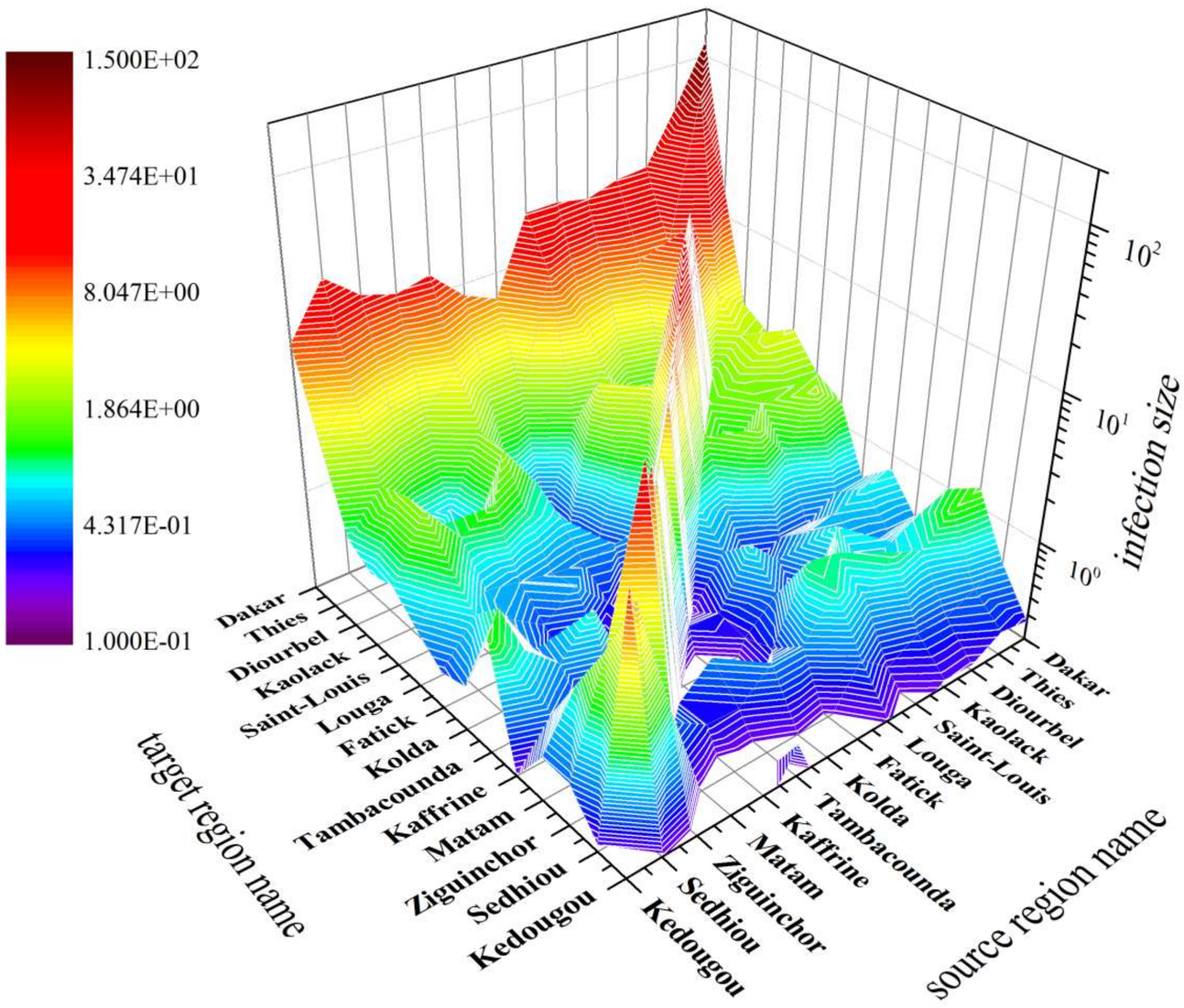
Main Tools used:
- Tool 1
- Tool 2
- Algorithm

Open Code available:
- Yes ☐
- No ☒

H04

Human contact and the diffusion of Ebola in Senegal

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



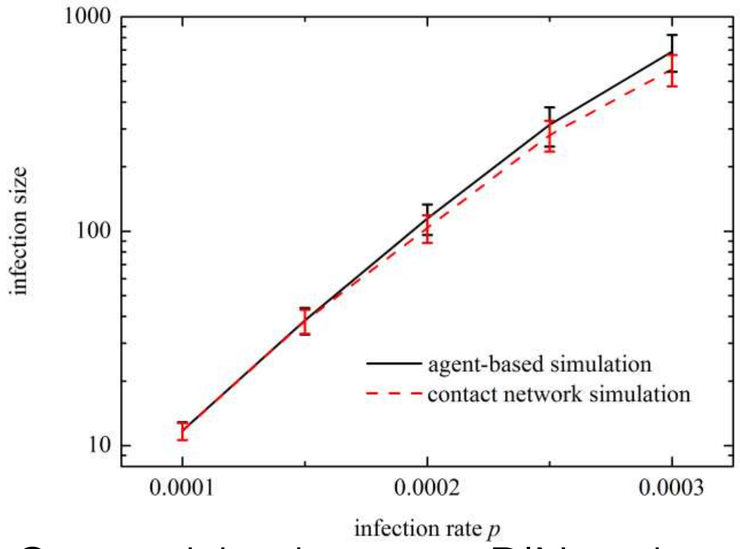
- Yi Yu, Mr., Ph.D. student
- Gaoxi Xiao, Dr., Associate professor
- Tai-quan Peng, Dr., Assistant professor
- Rich Ling, Dr., Professor

Project Summary:

This project compares the use of agent-based modeling to a data-driven individual-based network modeling (DIN) approach to examine the spread of contagious disease. We find that the DIN approach is efficient computationally and it avoids theoretical/statistical assumptions for human mobility/contact patterns. The DIN approach is verified by comparing with agent-based simulation and a good match is achieved. Several epidemic scenarios are evaluated using samples from D4D dataset.

Possible use for development:

The DIN approach can be a useful tool in managing the spread of contagious disease by giving authorities the ability to determine where and when a disease can spread within a country such as Senegal.



Comparision between DIN and agent-based simualtion.

Demonstration of a generic infection starting in Zinguinchor, Senegal.

Main results:

- Our work verifies the DIN approach by comparing agent-based simulation. We conduct epidemic simulations on samples to understand the dynamics of epidemics in Senegal. Our findings show that:
  - The DIN approach results in a good match with the results of agent-based simulations while being much more efficient computationally.
  - In Senegal, the capital Dakar is dominate in population and as a traffic hub. The outbreak of an infection soon propagates into the capital and from there propagates a large infection.

Methods:

- The development of the DIN approach relies on the empirically tested fact that people mostly stay within a small radius and that their travel patterns are repetitive.
  - People's contact network are built by tracing the total number of 10 minute time steps they are co-located with others (this can be adjusted based on the granularity of the data)
  - The epidemic process can be mapped on this contact network. We can calculate the potential for contamination between the two persons during a given period. The DIN approach can be conveniently extended to evaluate more complicated diffusion trajectories with more real-life parameters such as incubation period, etc.

Full paper is here:
To be developed

DataViz or video are here:
put your link here
Login:
Pw:

Data sources used for this project:
- D4D data set 1, com between antenna ☐
- D4D data set 2, movement routes high res ☐
- D4D data set 3, movement routes low res [X]
- D4D synthetic data set ☐

Other data sets used in this project:
- Type of data: Source:
- Type of data: Source:
- Type of data: Source:

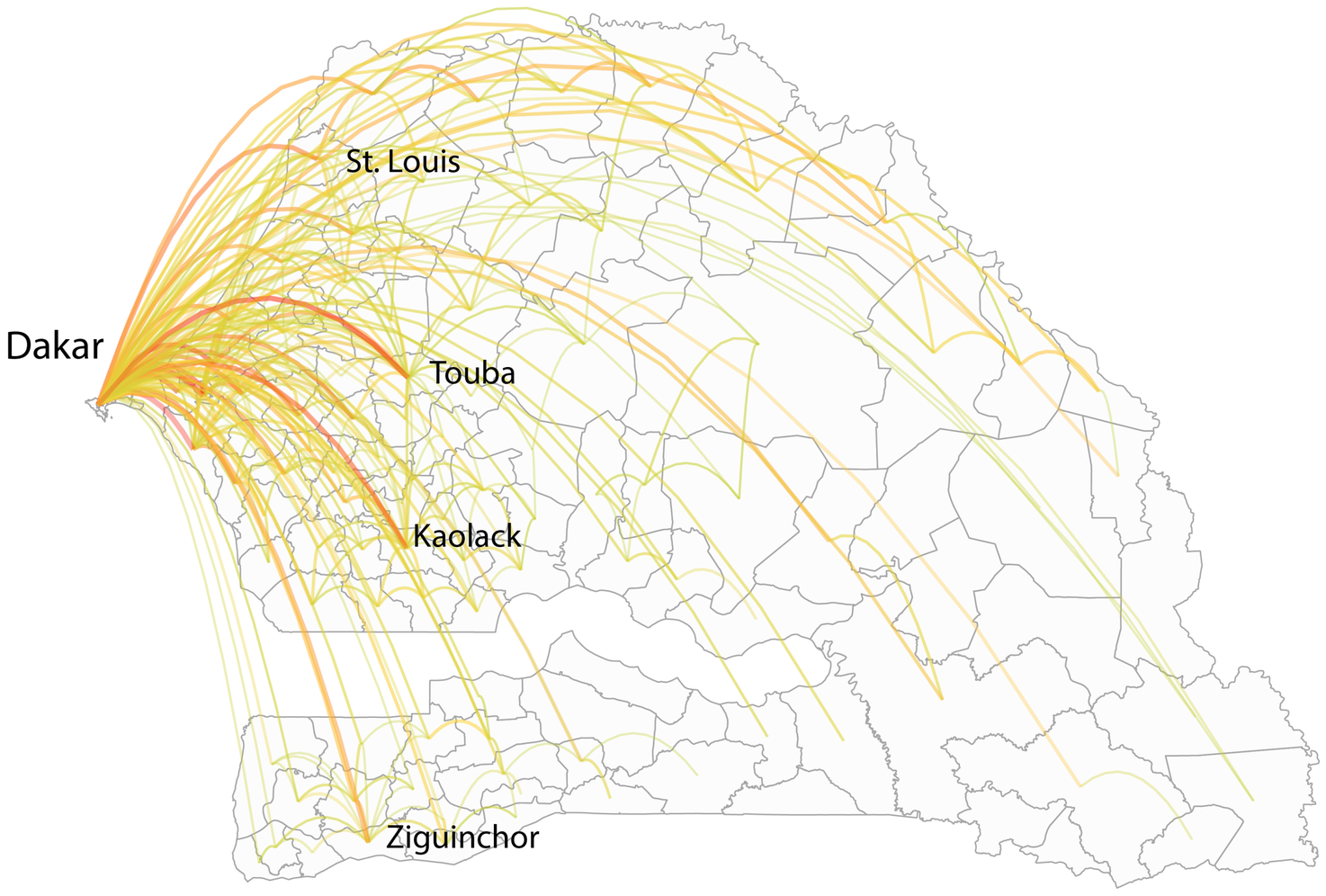
Main Tools used:
- Tool 1
- Tool 2
- Algorithm

Open Code available:
- Yes ☐
- No ☐

H07

Human mobility and the spreading of waterborne diseases

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DtaViz	Network



ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

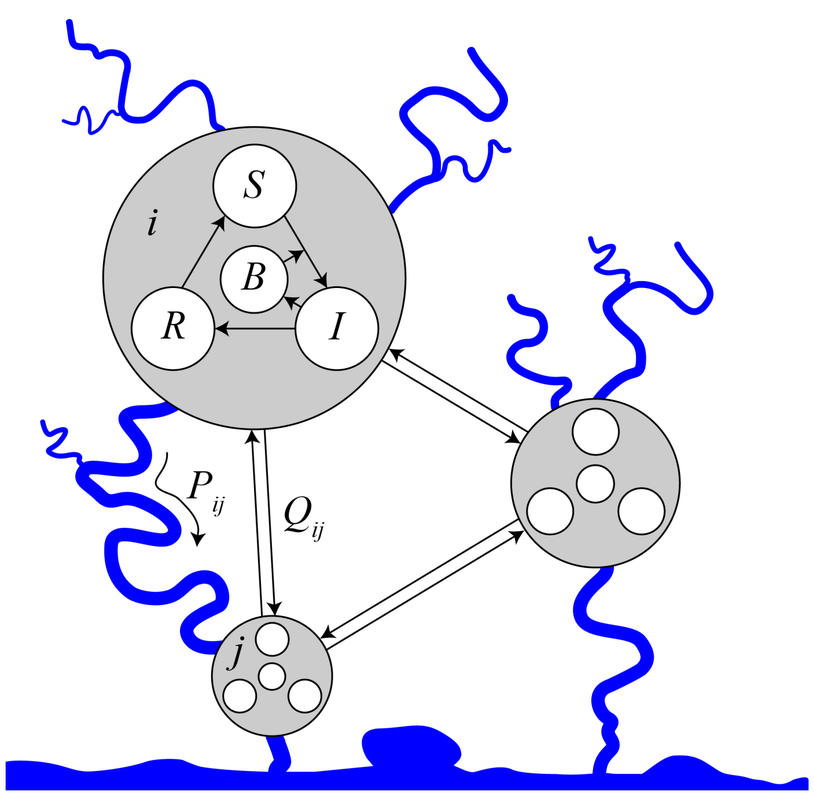
- Perez Saez, Javier, EPFL
- Finger, Falvio, EPFL
- Mari, Lorenzo, Ph.D., EPFL
- Rinaldo, Andrea, Ph.D., EPFL
- Bertuzzo, Enrico, Ph.D., EPFL

Project Summary:

Human mobility patterns are crucial determinants for the spreading of waterborne diseases. In this project we use mobile phone records to estimate the fraction of people moving outside their resident community and the distribution of trip destinations: two quantities vital to develop reliable spatially explicit epidemiological models.

Possible use for development:

Development of appropriate mechanistic epidemiological models to understand, control, and predict waterborne disease epidemics, as well as to devise appropriate intervention measures.



Schematic representation of a spatially explicit model of waterborne disease dynamics. Both hydrological and human mobility networks are accounted for.

Main results:

- The fraction of people moving outside their resident community and the distribution of trip destinations can be robustly estimated through the analysis of mobile phone records.
- A comparative analysis with alternative conceptual models of human mobility used in epidemiological contexts when mobility data are not available, namely gravity and radiation models, shows that mobile phone records are more informative and can be used to infer non trivial mobility patterns that cannot be captured by simple conceptual models.

Methods:

- Analysis of mobile phone records to determine the resident community of each user, the distribution of trip destinations and the time spent at destination.
- Analysis of road maps and digital terrain model to determine travel time for trips between all possible pairs of locations.
- Travel time is used as a proxy for distance in conceptual models of human mobility, namely gravity and radiation model.

Full paper is here:

http://tinyurl.com/pgew5td

Data sources used for this project:

- D4D data set 3, movement routes low res

Other data sets used in this project:

- Road Network — Source: DIVA Gis
- Digital Terrain Model — Source: NASA
- Population distribution — Source: Worldpop

Main Tools used:

- ArcGIS
- MatLab
- Dijkstra Algorithm

Open Code available:

- No

H08

Progmosis: Evaluating Risky Individual Behavior During Epidemics

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network

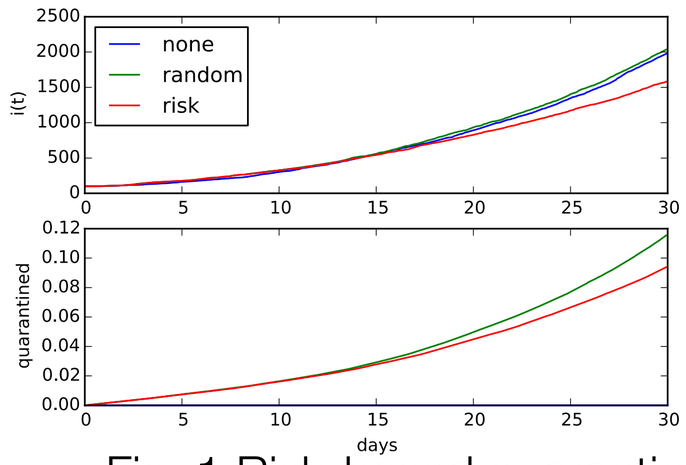
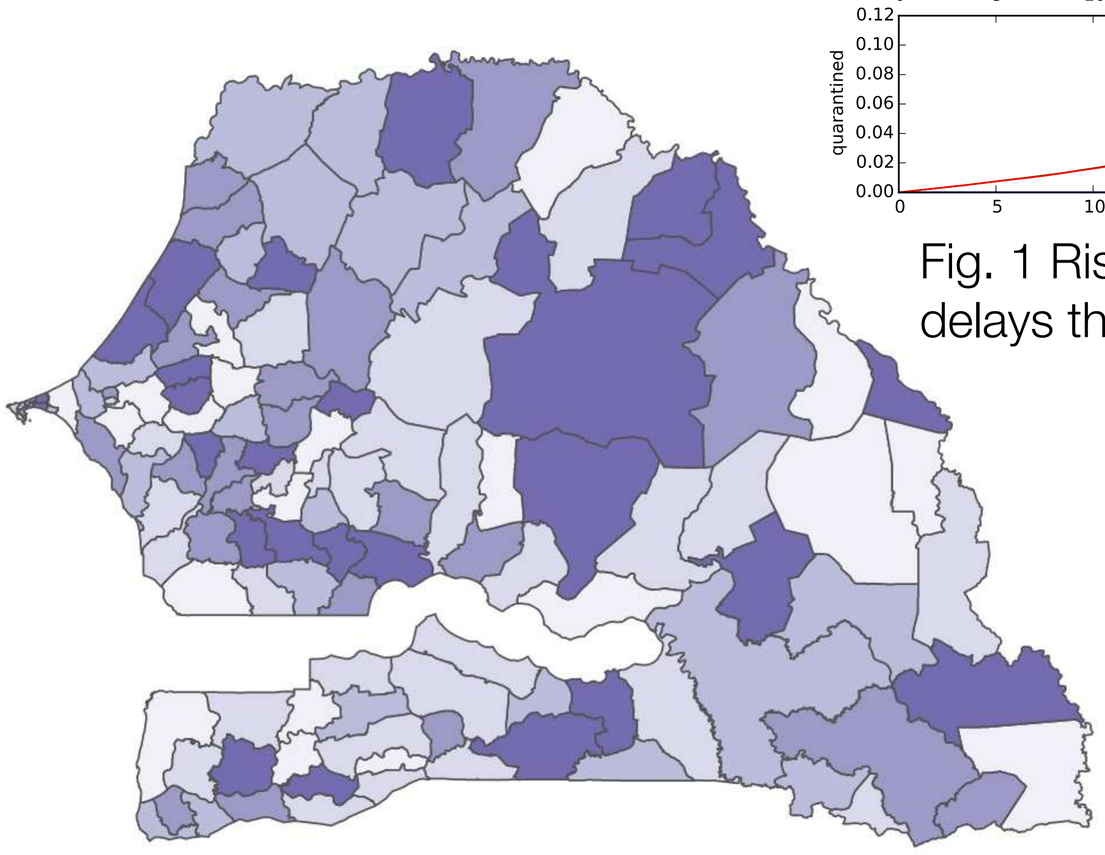


Fig. 1 Risk-based quarantine delays the disease spread.



Regional disease prevalence and travel behavior are heterogeneous.

UNIVERSITY OF BIRMINGHAM

Massachusetts Institute of Technology

- A. Lima, PhD Stud, Univ. of Birmingham - Vest Scholar, MIT
- V. Pejovic and L.Rossi, PostDocs, University of Birmingham
- M. Musolesi, Reader, University of Birmingham
- M. Gonzalez, Assistant Professor, MIT

Project Summary:

We propose a methodology that leverages data generated by mobile carrier networks to evaluate individual contagion risk. The risk represents the loss incurred by not isolating or treating a specific person, both in terms of how likely it is for this person to spread the disease as well as how many secondary infections it will cause.

Possible use for development:

A system can be built to evaluate this risk for a large number of users. The system can be accessed by healthcare officers, who can deliver appropriate and targeted actions (sending SMSes, asking people to get tested or keep them under observation). Individuals can also access the system, receive information about their own risk profile and receive tailored suggestions.

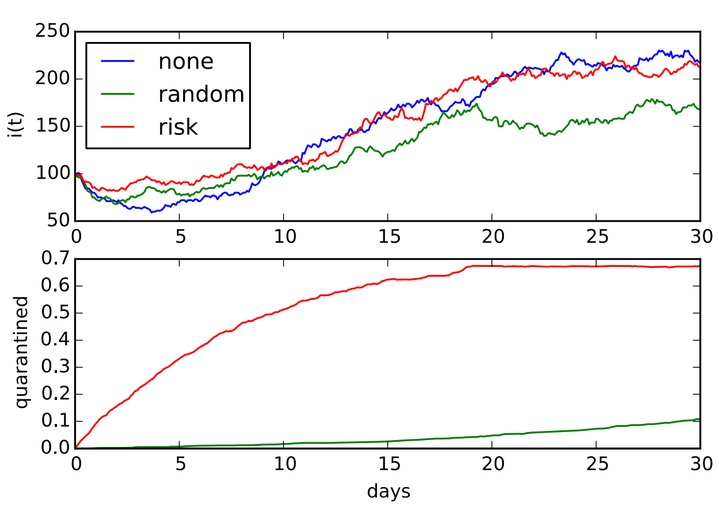


Fig. 2 Number of infected and quarantined in the source region.

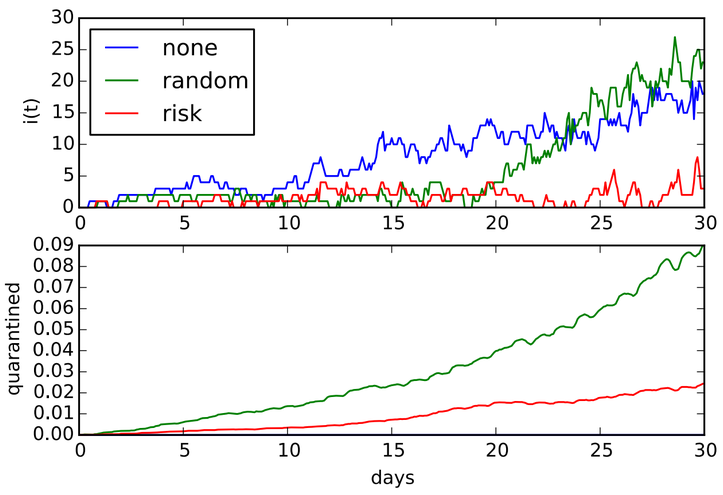


Fig. 3 Number of infected and quarantined in another region.

Main results:

- Targeted quarantine based on risk, instead, manages to delay the disease spread; at the end of the month there are 24% fewer infected individuals than in the baseline cases (Fig.1)
- Random quarantine with the same rate does not delay the spreading of the disease, despite involving a high number of people (Fig. 1).
- Targeted quarantine manages to delay the disease by keeping under observation individuals in high-risk region who travel to lower-risk regions (Fig. 2). This approach causes a decrease in the number of infection cases in low-risk regions (Fig. 3).

Methods:

- We initialize computer-based simulations supported with real-data by introducing 100 cases in a single region.
- We use the first 6 months of the year 2013 (Jan-Jun) to learn the habits of the users. We perform simulations using data from July 2013.
- Then we perform simulations under the three scenarios mentioned above: no countermeasures, people quarantined randomly and people quarantined according to their risk rank.
- We set an adaptive quarantine rate of $\xi=\beta i(t)$ to match the countermeasure efforts with the speed of growth of the outbreak.

Full paper:
http://bit.ly/progmosis

Risk evaluation tool:
http://github.com/themiurgo/progmosis-risk

Data sources used for this project:

- D4D data set 1, com between antenna [ ]
- D4D data set 2, movement routes high res [ ]
- D4D data set 3, movement routes low res [X]
- D4D synthetic data set [ ]

Other data sets used in this project:

- Disease epidemic parameters were estimated from data about 2014 Ebola outbreak (Althaus 2014).

Main Tools used:

- Python

Open Code available:

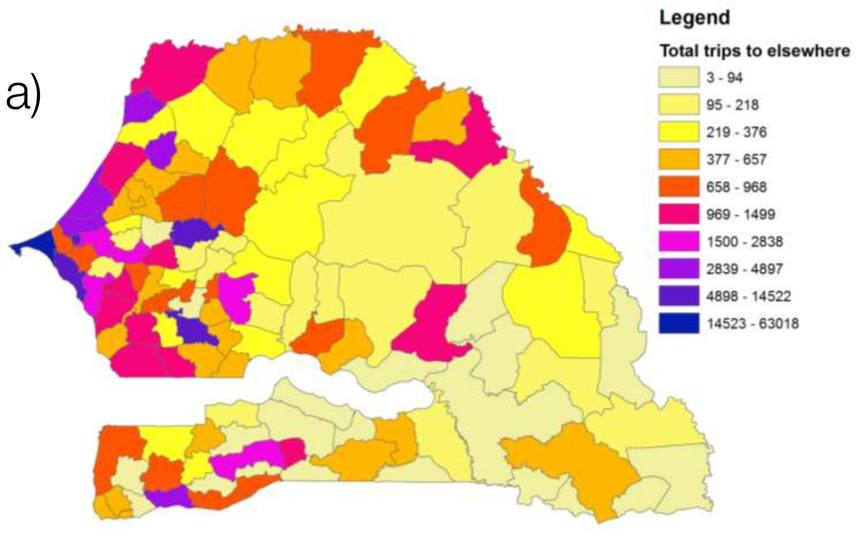
- Yes [X]
- No [ ]

H09

Developing an agent based migration model for Senegal for malaria transmission.

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network

a)



b)

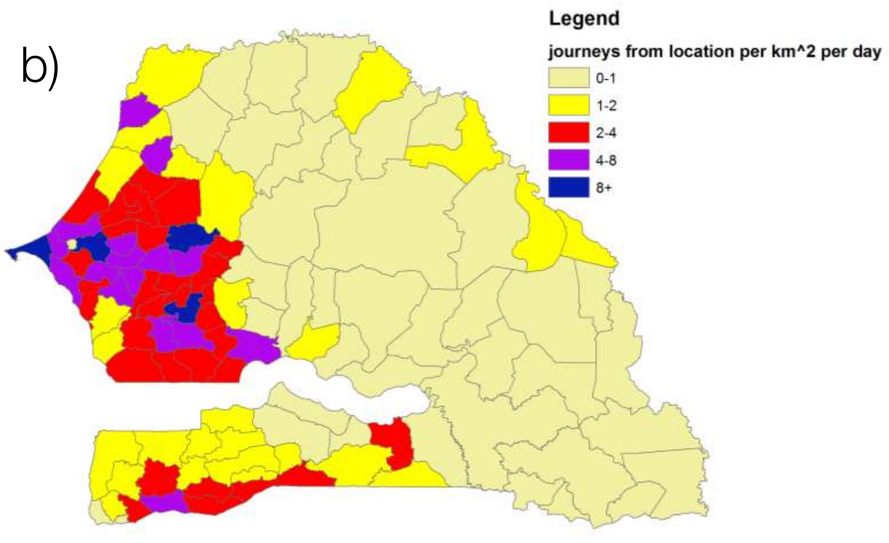


Figure 1. Comparison of empirical data and model output.

a) Total number of trips involving a night away made by people living in each arrondissement. Data from 18,384 people were analysed.

b) Preliminary results of the WISDOM model simulations. Units are individuals leaving per square km perday.

Tompkins, Adrian, Dr, and McCreesh* Nicky: Abdus Salam International Centre for Theoretical Physics

Email: Tompkins@ictp.it

* present affiliation: London School of Hygiene and Tropical Medicine

Project Summary:

•A preliminary analysis of the D4D dataset is made to determine the characteristics of jounreys that result in an overnight stay that are considered relevant for malaria transmission.

•We then introduce a new, highly memory and computationally efficient agent-based migration model called WISDOM that sets some of its parameters using the D4D analysis.

•First simulations of the agent based model are shown. These show the agent based model WISDOM reproduces the zero-order patterns of migration involving overnight stays but further improvements are required before the next step of coupling WISDOM to the spatially explicit malaria model VECTRI can be undertaken

Possible use for development:

While the key goal is to couple the WISDOM migration model to the VECTRI malaria model to investigate the impact of migration on malaria transmission, the highly efficient agent-based transmission code can be widely applied to other commicable diseases or other migration applications.

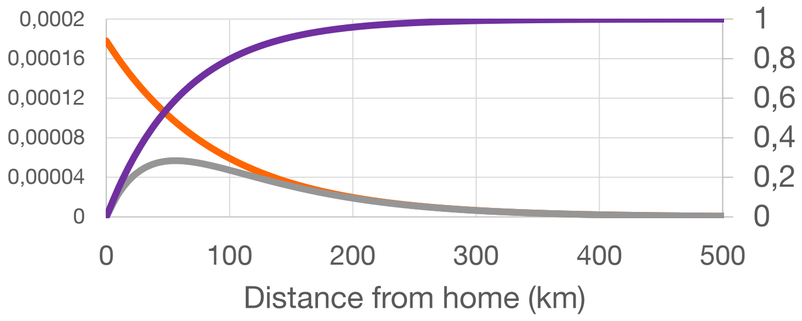


Figure 2. Best fit relationship to probability of making a trip with an overnight stay by distance.

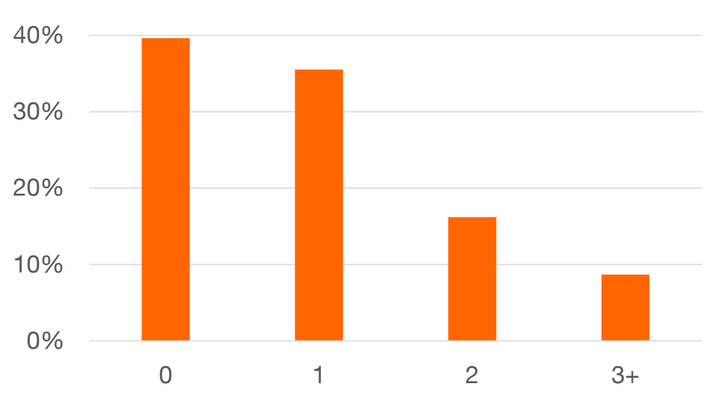


Figure 3. Proportion of people with 0, 1, 2, and 3+ regular non-home locations.

Main results:

- D4D phone data analysis
  - Probability of overnight stay increases with 62km e-folding length-scale.
  - Peak distance for an overnight stay is at a separation of 50km.
- WISDOM migration model
  - can produce the zero order spatial distribution of the journeys from the western provinces and in Casamance.
  - Performs poorly for some eastern provinces and the northern border region of Saint Louis.
  - The northern region is well connected to the capital by the N2 highway and is partly majority Wolof, thus high numbers of migrations from the capital are expected.
  - Ethnic background and transport network need to be added to WISDOM

Methods:

- WISDOM journey probability map based on D4D data and modified implementation of the Simini et al. (2012) radiation model of migration
- Model implements 3 million agents in a map covering Senegal that uses 5km by 5km grid-cells.
- Each agent has daily journeys chosen from a small number of “regular” destinations or a random location.
- When away from their assigned “home” location, they may also return home with a high probability

Full paper is here:
clima-dods.ictp.it/data/d10/tompkins/wisdom/d4d_wisdom.pdf

Data sources used for this project:

- D4D data set 1, com between antenna ☐
- D4D data set 2, movement routes high res ☐
- D4D data set 3, movement routes low res ☒
- D4D synthetic data set ☐

Other data sets used in this project:

- Type of data: Source:
- Type of data: Source:
- Type of data: Source:

Main Tools used:

- WISDOM migration model
- R and

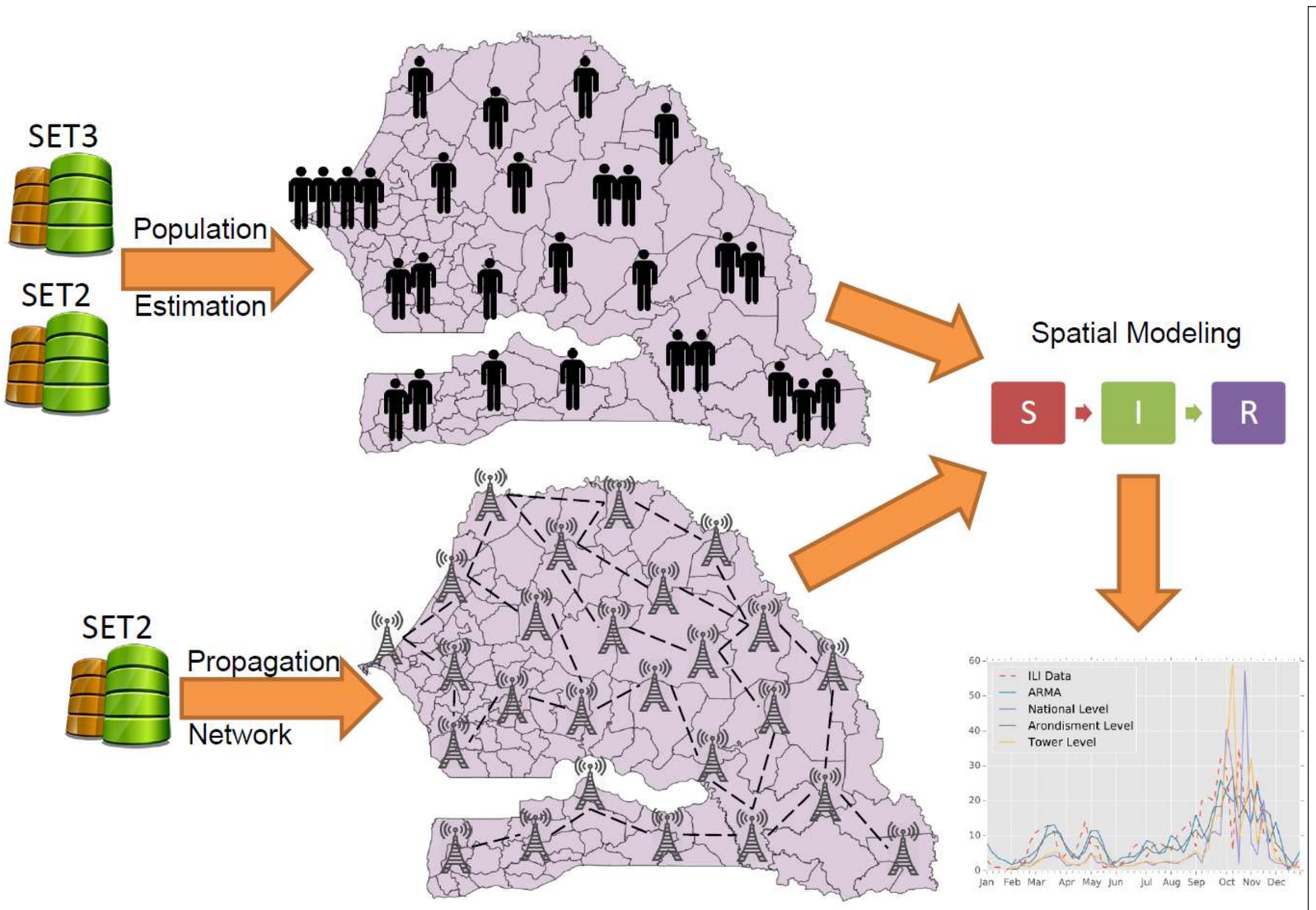
Open Code available:

- Yes ☐
- No ☐

H10

Forecasting Influenza in Senegal with Call Detail Records

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



- Hao Wu, Virginia Tech
- Prithwish Chakraborty, Virginia Tech
- Saurav Ghosh, Virginia Tech
- Naren Ramakrishnan, Virginia Tech (lead author)

Project Summary:

In this project, we use mobile call detail records from the D4D Senegal Challenge dataset to seed parameters of an epidemiological model prescribed over meta-populations, and study the spread of influenza-like illnesses using the model. The comparison between our simulation results and surveillance data demonstrates promise of our approach.

Possible use for development:

Epidemiologists and public policy makers are constantly seeking novel, surrogate, non-intrusive data sources to study disease spread and propagation. This is especially important in emerging countries that lack sophisticated public health infrastructure, such as Senegal. Our work aims to not just model people's activities but demonstrates a predictive approach to influenza-like illnesses (ILI) that can be used by public policy makers.

Table 1. Comparison of Accuracy

Method	Percentage Relative Accuracy
ARMA	43.25
Nation Level	67.50
Arrondissement Level	**80.25**
Tower Level	70.30

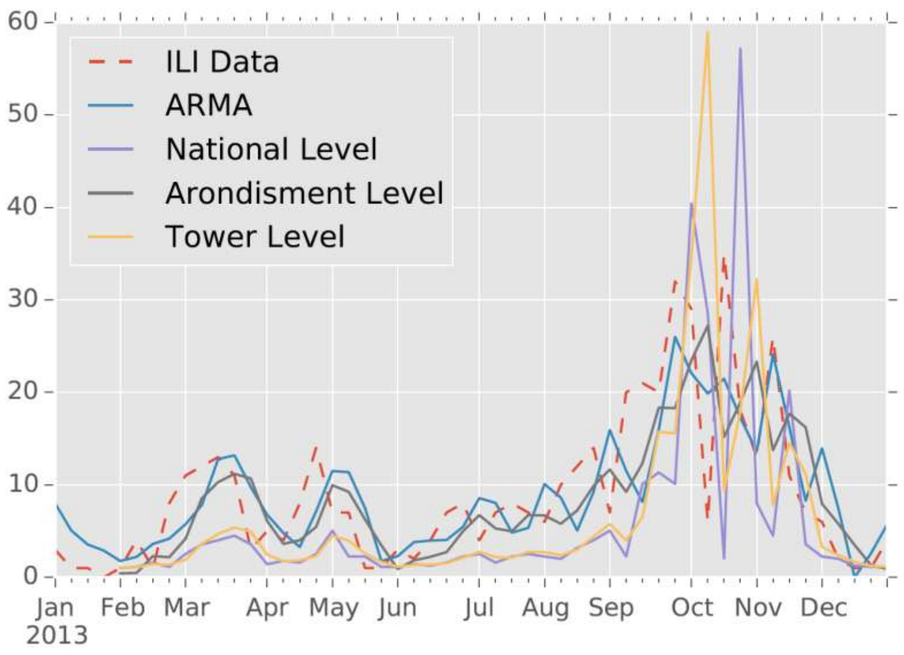


Fig.1 Average Curvers

Main results:

- Mobility Network derived from call records can be used as indicators networks for ILI propagation.
- Network-based Spread Simulations (Tower and Arrondissement level) give better accuracy than two baseline methods (ARMA and National Level).
- Simulations on Arrondissement level networks provide better accuracy than Tower level simulations.

Methods:

- SET2 and SET3 are used to estimate the population and the propagation network.
- Discrete Stochastic SIR Meta-population models used to model Influenza Spreads on Underlying Propagation Network.
  - Negative Binomial Process used to approximate the discretization.
  - Modified Gravity Model used to describe the spatial coupling.
  - Gamma Distribution on Spatial Coupling to describe the spatial force of infection.
- Meta-population Assumptions:
  - Arrondissement level : Users within same arrondissement governed by Single SIR dynamics. Interacts with users of other arrondissements through spatial force of infection.
  - Tower level : Users within the same Tower Cell are governed by Single SIR dynamics. Interacts with users of other Cell Towers through spatial force of infection.

Full paper is here:
put your link here

DataViz or video are here:
put your link here
Login:
Pw:

Data sources used for this project:

- D4D data set 1, com between antenna [ ]
- D4D data set 2, movement routes high res [x]
- D4D data set 3, movement routes low res [x]
- D4D synthetic data set [ ]

Other data sets used in this project:

- Type of data: Influenza data in Senegal Source: WHO FluNet
- Type of data: Senegal mobile phone population Source: World Bank, mobile

Main Tools used:

- Tool 1: Epidemiology model over meta-population
- Tool 2: Python Disco Map-Reduce framework

[x]

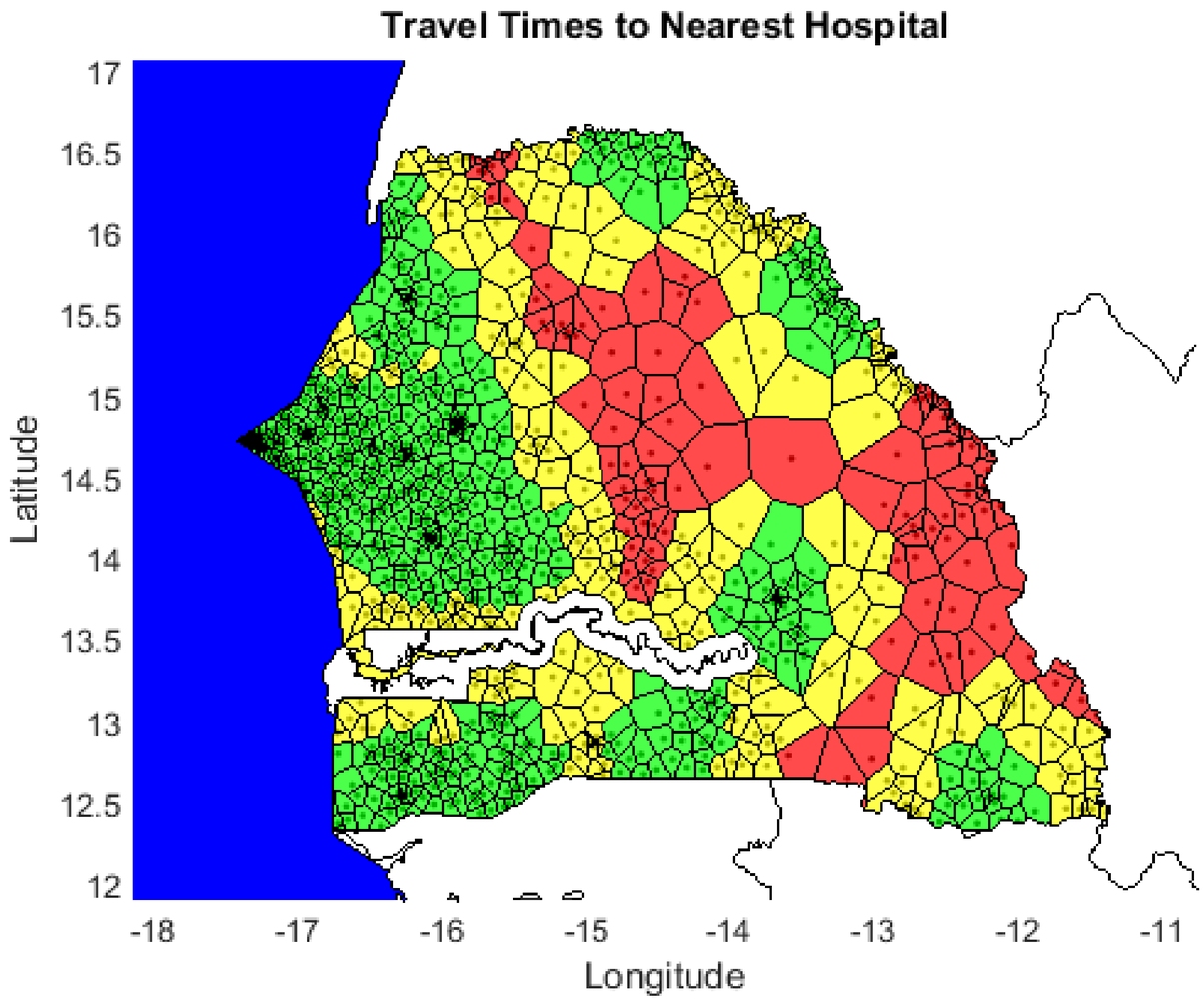
Open Code Available:

- Yes
- No

H11

Mobile Data as Public Health Decision Enabler: A Case Study

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



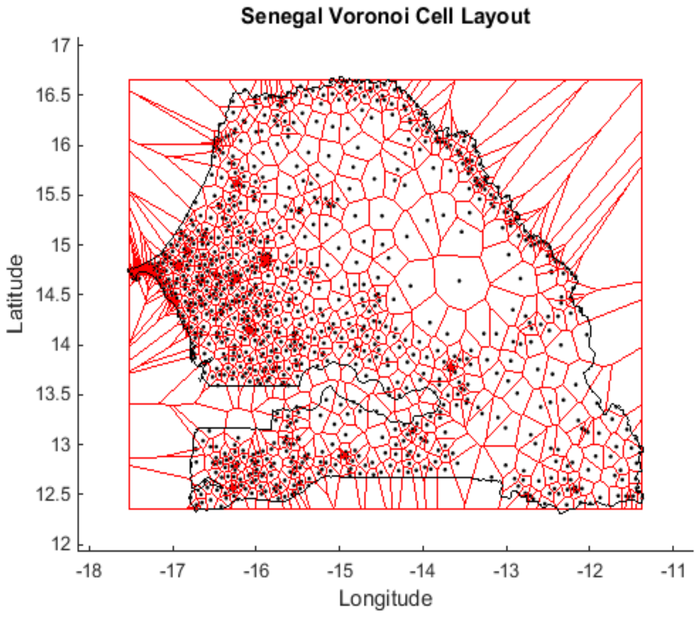
- Mutafungwa, E., Aalto University (lead author)
- Ben Yahia, S, University of Tunis
- Diallo, G. University of Bordeaux
- Diallo, M. P. University of Bordeaux
- Gore, R., Old Dominion University
- Hämäläinen, J. Aalto University
- Jouhet, V., University of Bordeaux
- Karray, C., University of Tunis
- Kheder, N., University of Tunis
- Mougin, F., University of Bordeaux
- Saddem, R. University of Tunis
- Thiessard, F., University of Bordeaux

Project Summary:

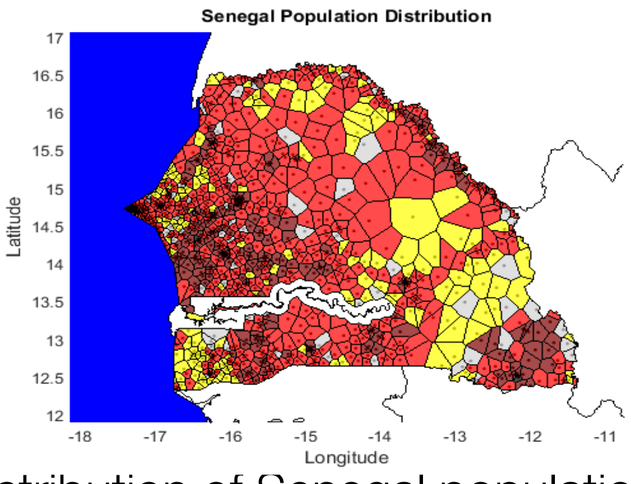
The establishment of hospitals in an area depends on many parameters taken into account by health authorities. We to investigate whether data from the use of mobile phones could feed this reflection. In order to do this, we chose two diseases that require rapid hospitalization for their care: myocardial infarction and stroke. The objective of the study is to show the areas in which the absence of a nearest hospital can result in death or serious sequelae.

Possible use for development:

The identification of areas at high risk in case of stroke of myocardial infarction could help Public Health decision makers to take the required actions on the earlier.



Voronoi cell layout for Senegal based on provided 1666 site locations



Distribution of Senegal population according to the mobile antennas

Main results:

- Thanks to the analysis of anonymized mobile dataset of Senegal customers, we have been able
  - To identify all areas at high risk in case of stroke or myocardial infarction
  - To estimate the number of people at risk at antenna coverage and therefore at regional level of Senegal
  - And provide a web prototype for demonstrating our proposal

Methods:

- Our approach is based on the following process
  - Estimating travel times to reach the nearest hospital
  - Computing the population density at antenna coverage area
  - Estimating Stroke and Myocardial Infarction rate at antenna coverage are
  - Highlighting white zones: areas where people are at high risk in case of Stroke or Myocardial Infarction

Full paper is here:
put your link here

DataViz or video are here:
put your link here

Data sources used for this project:

- D4D data set 1, com between antenna ☐
- D4D data set 2, movement routes high res ☒
- D4D data set 3, movement routes low res ☐
- D4D synthetic data set ☐

Other data sets used in this project:

- SenDoctor source: www.sendoctor.com
- Annuaire Médical Sen. http://www.annuairemedical-senegal.com/

Main Tools used:

- Languages: MatLab, R, Java
- Qgis, ArcGis
- Voronoï Tesselation

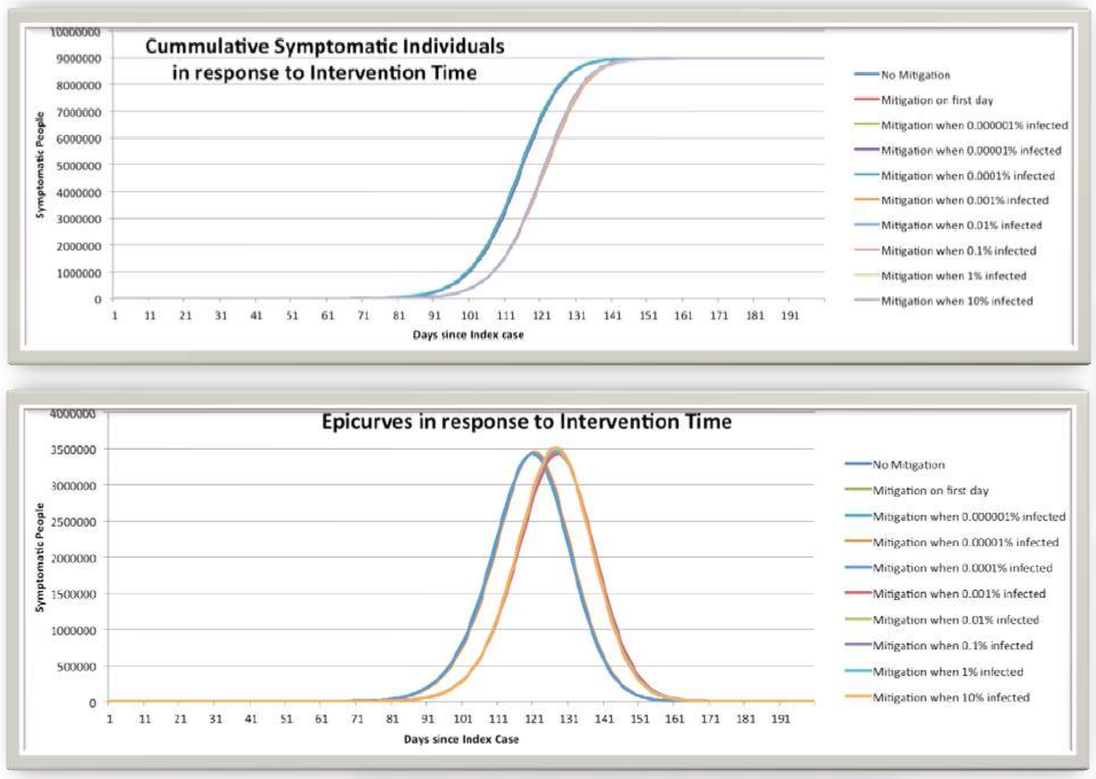
Open Code available:

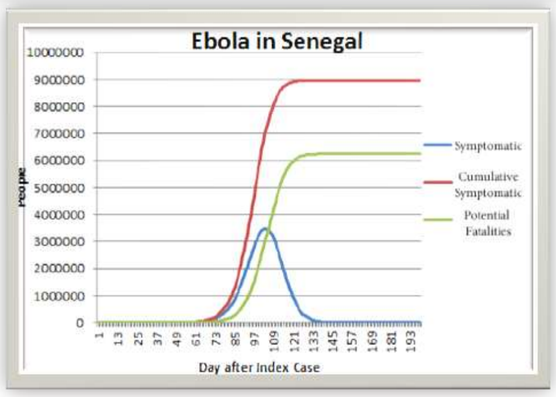
- Yes ☐
- No ☒

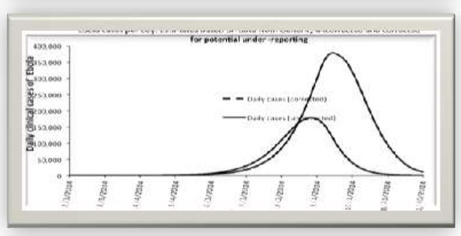
H12

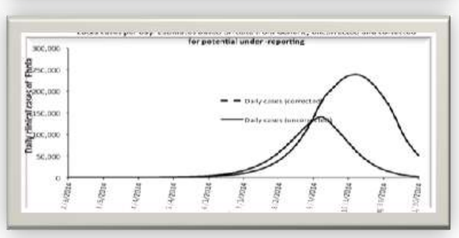
Modeling Ebola Diffusion in Senegal using Agent-based Simulation

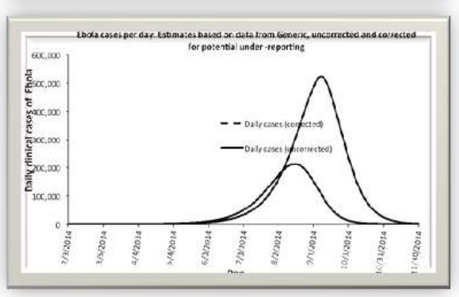
Health	Transport Urban	National Statistics	Emergency
Agriculture	Energy	DataViz	Network











Jonathan P. Leidig*, Christopher Theisen*, Nicholas Vogel*, Doug H. Graham†, Jerry Scripps*, Greg Wolffe*

*School of Computing and Information Systems

†Department of Biomedical Sciences

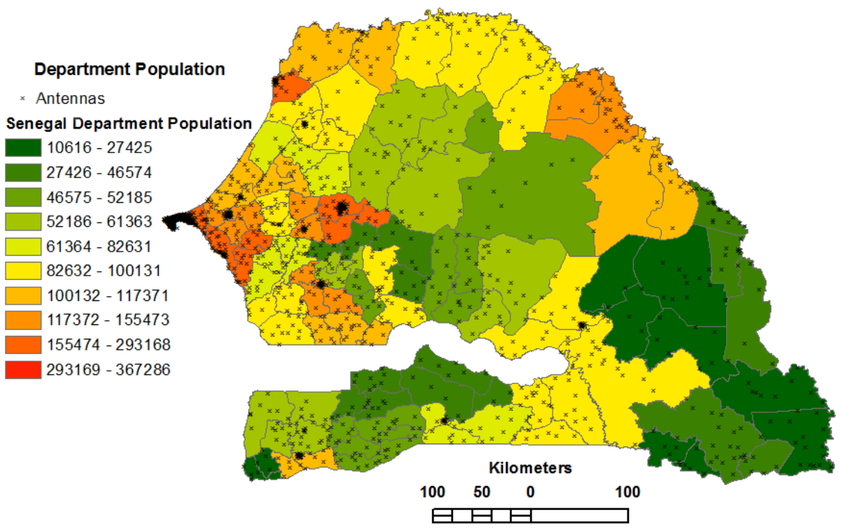
Contact: jonathan.leidig@gvsu.edu

Project Summary:

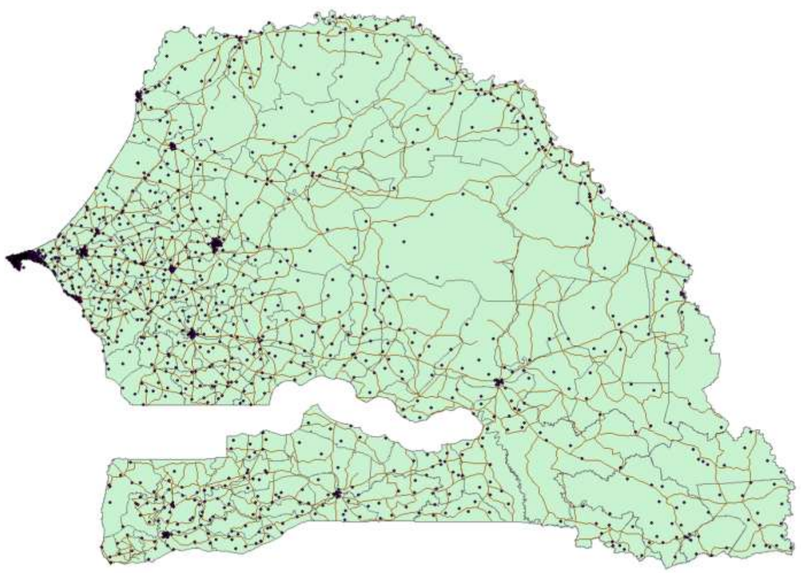
The D4D datasets were analyzed to develop latent population and mobility models of the underlying population of Senegal. Simulation software was developed to model Ebola virus diffusion and transmission in Senegal, based on existing computational epidemiology software. Experiments were performed to study the effect of disease outbreak mitigation strategies and governmental policies for optimization of resources and efforts, e.g., quarantine and border closures.

Possible use for development:

The software may be utilized (by conducting simulation studies) to inform Senegalese government policy and best practices on disaster prediction, preparation, public health response, and recovery.



Population model for each antenna region being simulated



Sampling locations (antennas) and border crossings

Main results:

- Models produced
  - Synthetic population model for Senegal
  - Mobility and activity models for Senegal
  - Ebola disease model
- Computational Epidemiology software produced
  - Vuvuzela: an Ebola virus modeling and simulation platform for the analysis of public health policies in Senegal
- Policy - provides quantitative support for decisions regarding:
  - Closing/opening borders, closing schools, limiting social and economic gathering, isolation and quarantine of infected individuals, travel restrictions, individual behavior recommendations, body washing, etc.
- Software analyzes effects of Ebola mitigation strategies

Methods:

- Datasets 2 and 3 were utilized to produce models of Senegal.
- Open-source FluTE simulation software was modified to handle Ebola.
- A simulation platform is now available to study the effects of governmental policy on a potential Ebola epidemic in Senegal.

Full paper is here:
TBD

Data sources used for this project:

- D4D data set 1, com between antenna [ ]
- D4D data set 2, movement routes high res [X]
- D4D data set 3, movement routes low res [X]
- D4D synthetic data set [ ]

Other data sets used in this project:

- Type of data:Ebola characteristics Source: multiple
- Type of data:Population models Source: multiple
- Type of data:Ebola predictions Source: CDC

Main Tools used:

- Epicurve visualizations
- Vuvuzela (modified FluTE version)
- CDC generic EbolaResponse

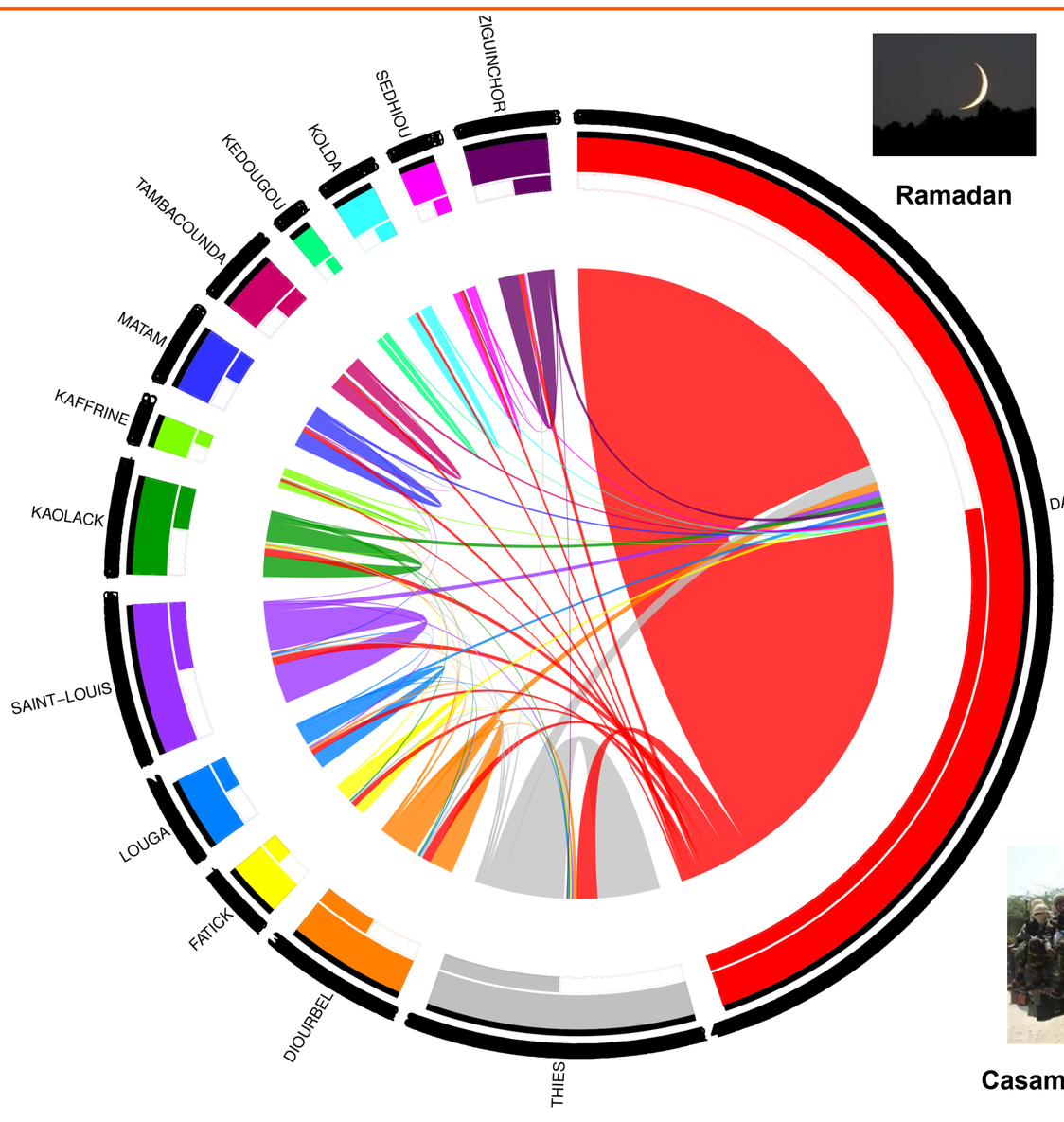
Open Code available:

- Yes [X]
- No [ ]

N01

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

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



Ramadan

Le Grand Magal de Touba

Tivaouane Maouloud

Eid al-Fitr and Eid al-Adha

Casamance Conflict – 4 dead

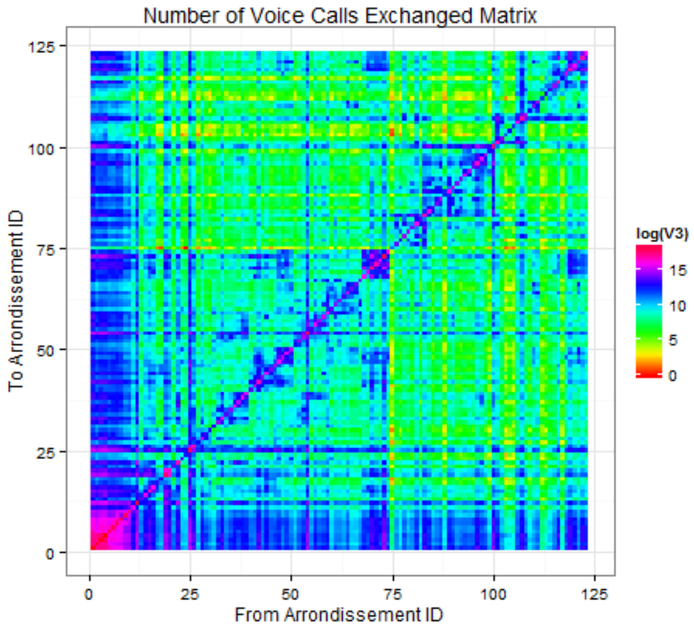
Project Summary:

Being able to react fast to exceptional events such as riots protests or disaster preventions is of paramount importance, especially when trying to ensure peoples' safety and security, or even save lives. We study the use of Call Detail Records (CDRs) to analyze the telecommunication traffic and connect people, locations and events. The goal of this study is to see if the CDR data can be used to detect exceptional spatio-temporal patterns of the collective human mobile data usage and correlate these 'anomalies' with real-world events (e.g., religious festivals, conflicts etc.).

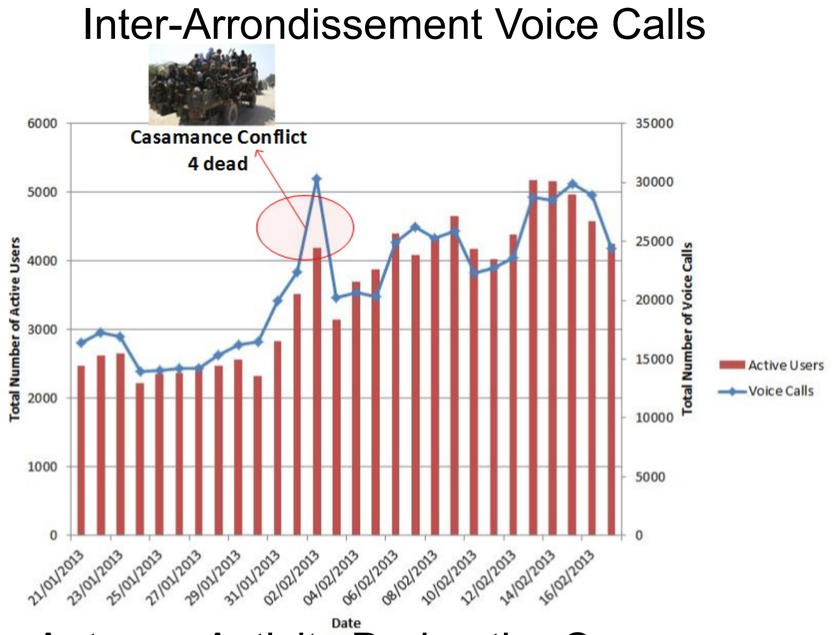
Possible use for development:

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, network resource allocation, performance optimization, etc. For example, a real-time event detection system could be used in case of emergency situations, such as riots, protests, conflicts, etc. which could be more efficiently handled if detected on time.

- Ramona , Trestian, Dr., Middlesex University (lead author)
- Purav, Shah, Dr., Middlesex University
- Huan, Nguyen, Dr., Middlesex University
- Quoc-Tuan, Vien, Dr., Middlesex University
- Orhan, Gemikonakli, Prof., Middlesex University
- Balbir, Barn, Prof., Middlesex University



Inter-Arrondissement Voice Calls



Antenna Activity During the Casamance Conflict – 4 dead reported

Main results:

- The results of the analysis on the telecommunication traffic pattern show that the communication is highly symmetrical. The SMS or voice calls in one direction always find a match in the opposite direction. Moreover, the communication is mainly within the same region or with the capital region.
- By analysing the anomalies in the CDR data we detected exceptional spatio-temporal patterns of the collective human mobile data usage and these 'anomalies' were then correlated to real-world events, such as:
  - **Ramadan** - people tend to speak more at night during Ramadan, starting from 10pm until 6am, when the number of voice calls and voice calls duration doubles. They also have a slower start during the day.
  - **Eid al-Fitr**, representing the end of Ramadan with peaks in the voice traffic on 7-9th of August.
  - **Eid al-Adha**, the Feast of Sacrifice with peaks in the voice traffic on 15th of October.
  - **Tivaouane Maouloud Festival**, peak in user mobility near Tivaouane, Thies region on 23rd of January.
  - **Le Grand Magal de Touba Festival**, peak in user mobility near the Mbacké region on 22nd of December.
  - **Casamance Conflict** near Kafoutine with peaks in traffic and user mobility during February, when an attack of the rebels from the Casamance Movement for Democratic Forces over the Credit Mutuel bank in Kafountine, was reported with four dead including a Frenchman.

Methods:

- The data provided in Datasets 1 and 2 was used for analysis in order to detect spatio-temporal patterns and show that these 'anomalies' in the usage patterns can be correlated to real-world events using external information, such as: population estimates and local news.

Full paper is here:

NA

DataViz or video are here:

NA

Login:

Pw:

Data sources used for this project:

- D4D data set 1, com between antenna [X]
- D4D data set 2, movement routes high res [X]
- D4D data set 3, movement routes low res [ ]
- D4D synthetic data set [ ]

Other data sets used in this project:

- Type of data: Population Census — Source: http://www.geohive.com/
- Type of data: Local News — Source: Internet
- Type of data: — Source:

Main Tools used:

- Tool 1: R+Hive+Hadoop+OpenStack, RStudio
- Tool 2: SAGA GIS
- Algorithm

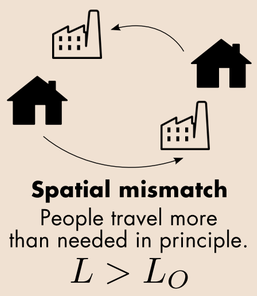
Open Code available:

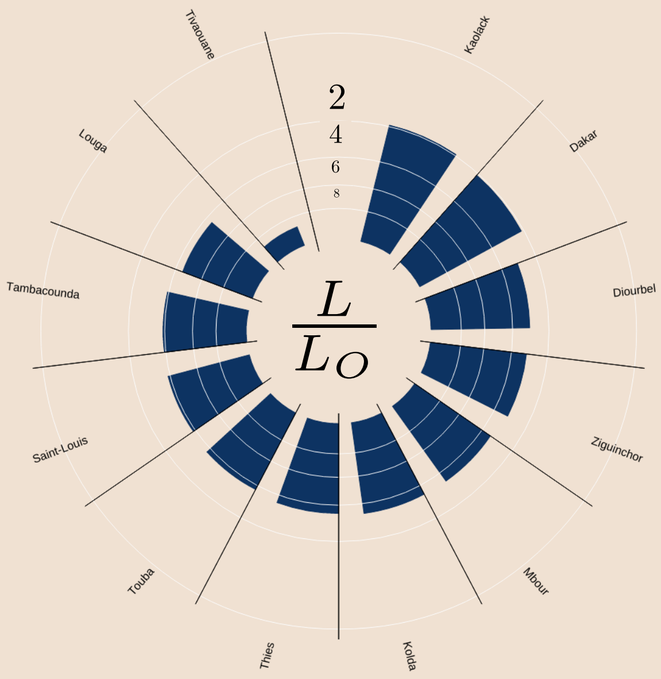
- Yes [ ]
- No [X]

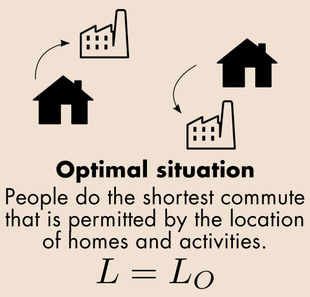
N02

Spatial *mis*match in Senegalese cities

Health	Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network







Spatial mismatch in Senegalese cities. We compare the total distance travelled by individuals L to the distance L_O that would be travelled if cities were perfectly organised. We find that, apart from Tivaouane, Senegalese cities are very close to the perfect match.

Rémi Louf, Giulia Carra, Riccardo Gallotti, Thomas Louail, Marc Barthelemy
Institut de Physique Théorique, CEA, CNRS-URA 2306, F-91191, Gif-sur-Yvette, France

Hadrien Commenges
Laboratoire Eau, Environnement et Systèmes Urbains, Ecole des Ponts, 77455, Marne la Vallée, France

Maxime Lenormand
Instituto de Física Interdisciplinar y Sistemas Complejos (CSIC-UIB), 07122, Palma de Mallorca, Spain

Jean-Marie Dembele
Université Gaston Berger, UFR SAT. Saint-Louis, BP 234, Sénégal

Project Summary:

In this project, we show that mobile phone data and open-source tools can be used to investigate the spatial structure of mobility in cities.

From the Call Detail Record detained by mobile phone companies, we extract the pendular flows of commuting to and from Senegalese cities.

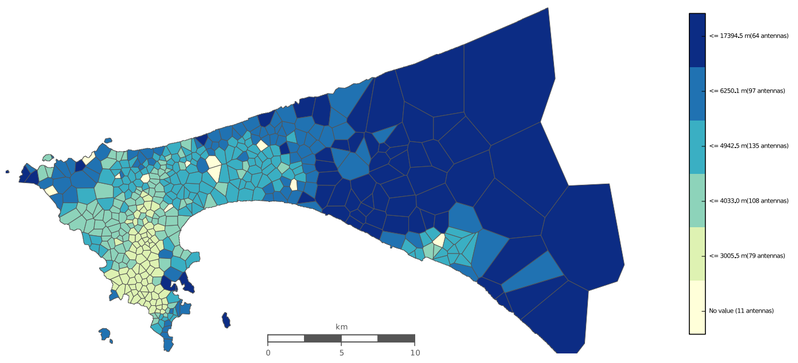
We then compare the existing structure of flows to (1) The `optimal' situation where everyone's daily activity would be at the closest location available. (2) The `anarchy' situation where everyone's daily activity would be located at random.

This comparison gives us a measure of the spatial mismatch in each city *given the existing residential and employment structure*.
Our method allows to reliably compare cities to one another.

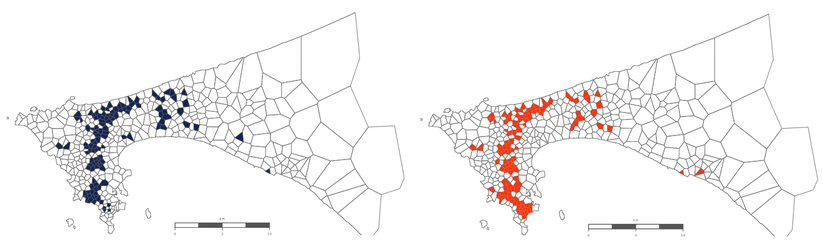
Possible use for development:

The identification of commuting flows within cities allows policy-makers to identify the areas that would benefit the most from investments in infrastructure.

Used along with socio-economic data, measures of spatial mismatch provide critical insights for urban policies aiming at reducing inequalities.



Average distance commuted by individuals living in the different zones of the metropolitan area of Dakar. The longer distances travelled by people in the outskirts suggest an important mismatch.



Home (in blue) and Daily activity (in red) hotspots in Dakar. They respectively contain 37% and 39% percent of the total population. 80% of them overlap, suggesting an important matching.

Main results:

Large Senegalese cities, with the exception of Tivaouane, have a highly integrated labour market.

We can identify the populations that do the longest commutes within cities.

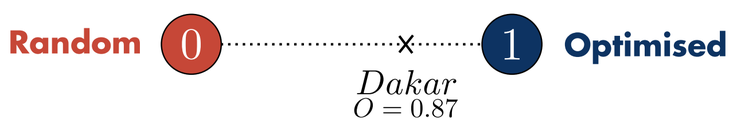
Senegalese cities exhibit surprisingly low levels of spatial mismatch (Tivaouane excluded).

Methods:

We extract origin-destination matrices using the most frequent locations of users during the day and during the night. The results are then compared to computation of the relative intensity of calls between nighttime and daytime to check for accuracy.

Cities are defined based on the density of antennas.

The Home and Daily activity hotspots are extracted using the LouBar method previously developed by members of the group.



Null model. A way to conclude is to compute the organisation index O, obtained by comparing the real situation with the following idealized ones
(i) people choose their location at random (O = 0)
(ii) people work at the nearest activity location (O = 1).

For each city, we compute 100 random origin-destination matrices that conserve the number of inhabitants and workers observed at each antenna in the data.
The optimal origin-destination matrix (with the same number of inhabitants and workers at each antenna as observed in the data) is computed using a simulated annealing method.

Full paper is here:
put your link here

Data sources used for this project:

D4D data set 1, com between antenna	[x]
D4D data set 2, mouvement routes high res	[x]
D4D data set 3, mouvement routes low res	[ ]
D4D synthetic data set	[ ]

Main Tools used:
- *QGIS*
- *Python*
- *R*
- *Shell scripts*

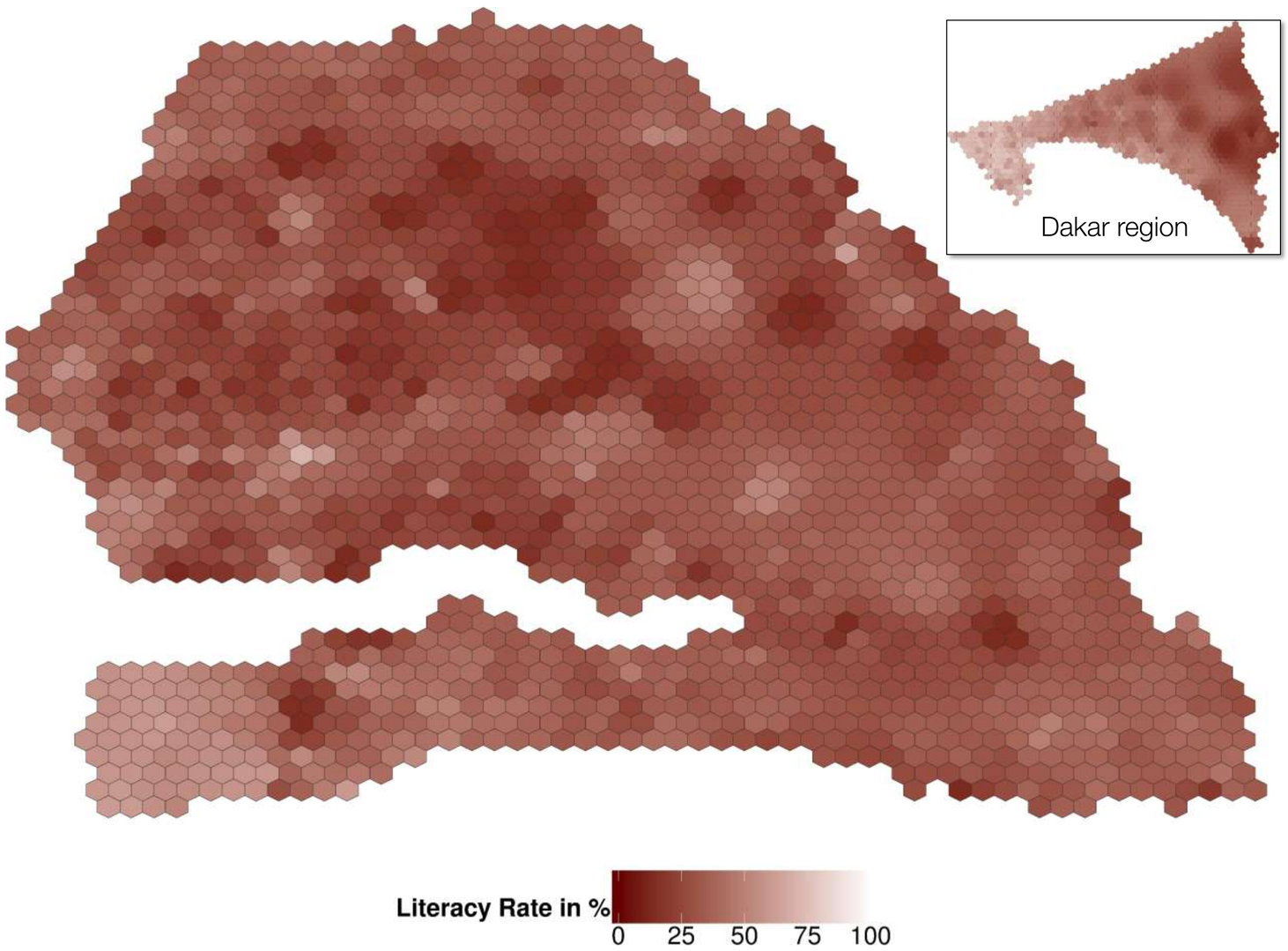
Open Code available:

Yes	[ ]
No	[x]

N03

Construction of socio-demographic indicators with digital breadcrumbs

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



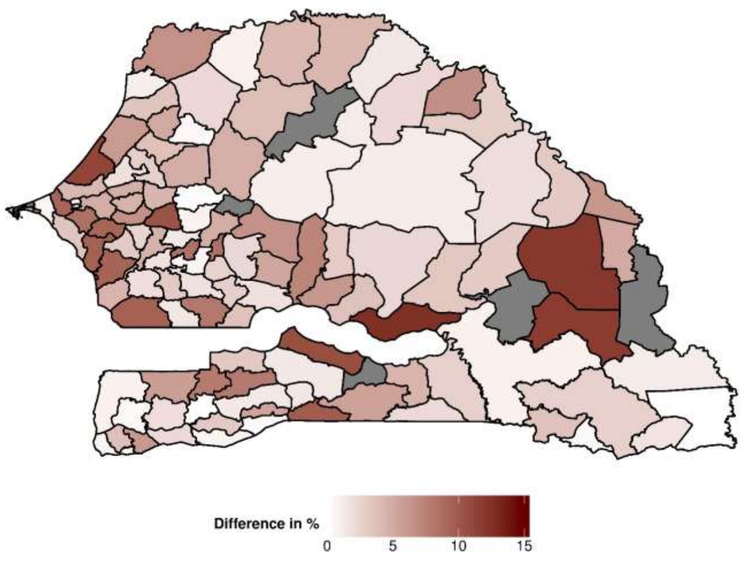
- Bruckschen, Fabian, HU Berlin
- Schmid, Timo, FU Berlin (lead author)
- Zbiranski, Till, HU Berlin

Project Summary:

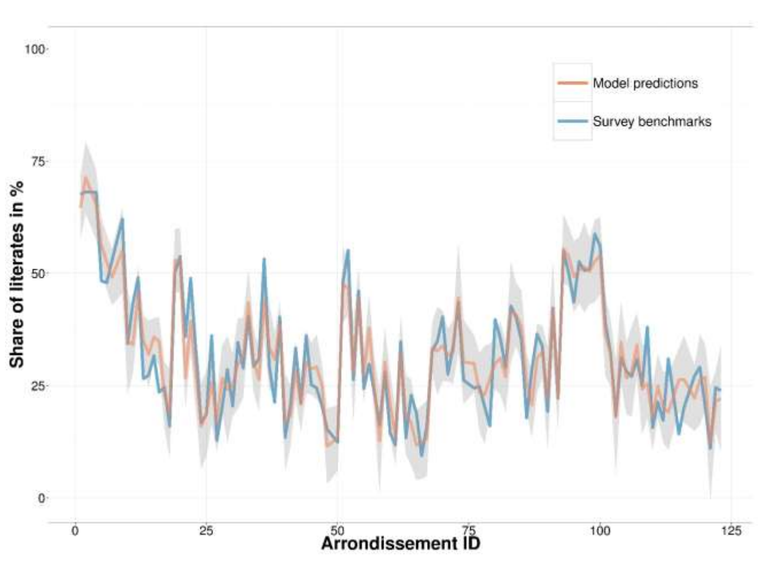
The aim of this research is to capture so far hidden local socio-demographic heterogeneity, such as pockets of illiteracy or poverty. We offer a uniform approach that can be easily extended to other variables. It is based on aggregated antenna traffic data only - data that is less prone to privacy concerns than e.g. mobility patterns, thus facilitating implementation. Models are fitted to geocoded survey data and used for prediction on the tower level. Results are tested for spatio-temporal robustness and visualized as heat maps.

Possible use for development:

In the short-term, uncovering local socio-demographic heterogeneity at little costs can facilitate timely & targeted relief.
In the medium-term, variables that can be modelled reliably could be collected less frequently, thereby reducing the scope and thus the costs of surveys.



Deviance of predictions to benchmarks



Precision of predictions

Main results:

- Successful rebuilding of socio-demographic indicators on administrative area level
- Robust tower level predictions

Methods:

- Multi-level linear regression models
  - Region-specific random intercepts and slopes
  - Shapiro normality test
- Performance measuring
  - Adjusted R^2
  - RMSE
- Selection procedures
  - Stepwise forward selection
  - Backward elimination
- Robustness techniques
  - Winsorizing
  - Inter-Spatial RMSE for validation
- Multiple Imputation for handling missing values
- Spatial Smoothing
  - Inverse distance weights
  - Heat map via hexagon grids

Full paper is here:
https://www.dropbox.com/s/mri4h1dbrf72sok/cookbook_socio_demographic_kpi_basket.pdf?dl=0

Data sources used for this project:

- D4D data set 1, com between antenna [X]
- D4D data set 2, movement routes high res [ ]
- D4D data set 3, movement routes low res [ ]
- D4D synthetic data set [ ]

Other data sets used in this project:

- Type of data: Survey Data — Source: DHS 2011
- Type of data: Census Data — Source: RGPHAE 2013
- Type of data: App Downloads — Source: Priori Data

Main Tools used:

- R (Open Source)

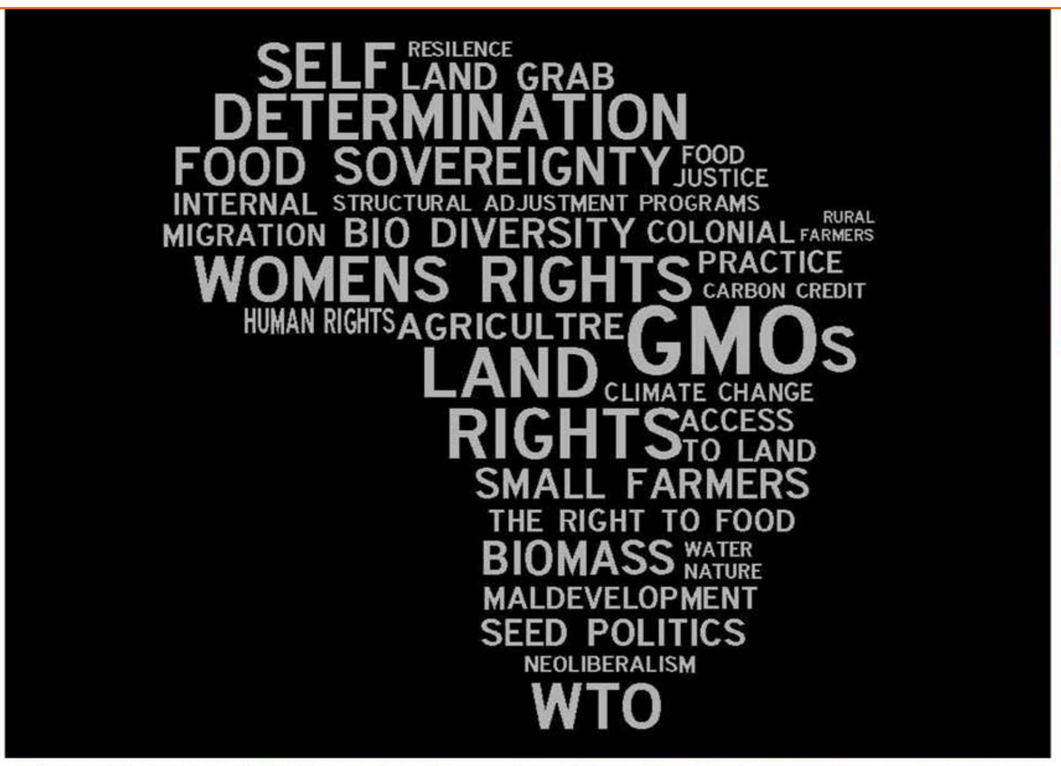
Open Code available:

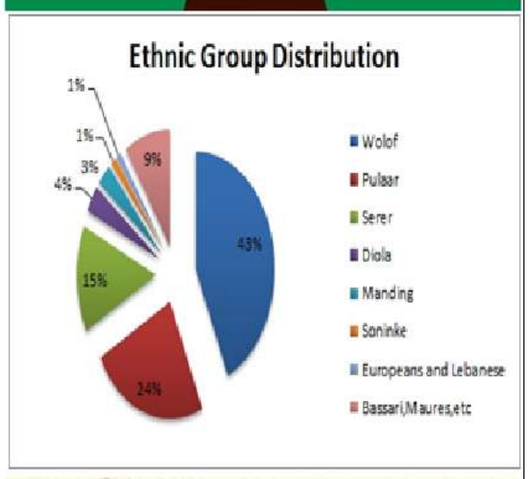
- Yes [X]
- No [ ]

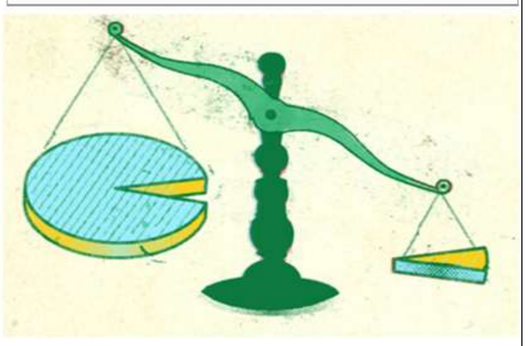
N04

Mapping and Measuring of social disparités in Senegal using mobile phone subscribers data

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network







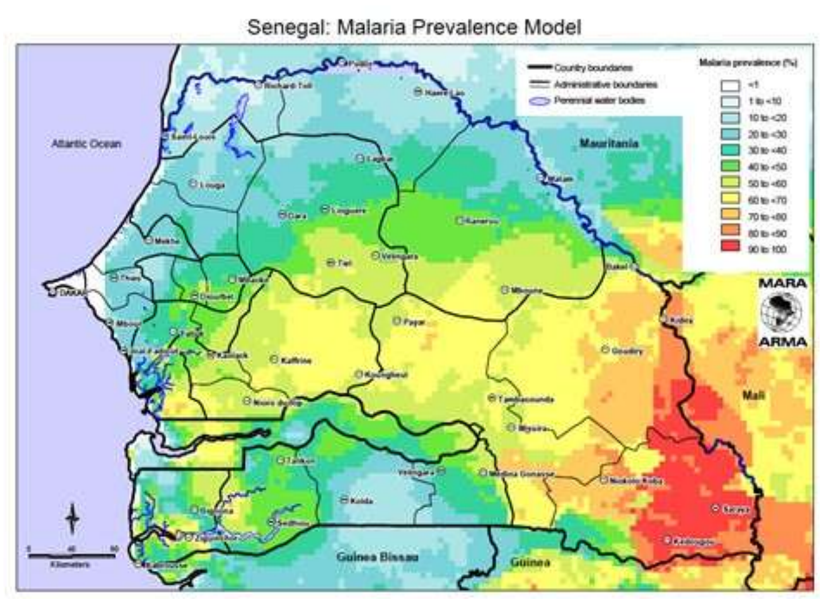
- M. Saravanan, Senior Researcher, Ericsson Research India
- Aarthi, Student, SSN College of Engg, India
- Nikil Bharatwaj, Student, SSN College of Engg, India

Project Summary:

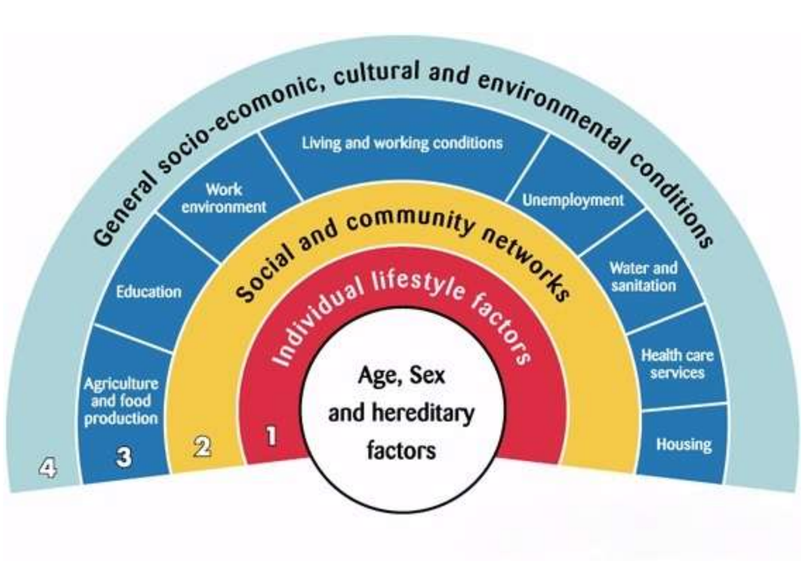
The project addresses various social disparities existing among the regions of Senegal and proposes a statistical methodology to analyze and understand the existence of patterns among the ethnic groups, with the help of mobile phone subsribers log data, thereby identifying the problems existing in the society and hence suggesting rectiifcation procedures for the same.

Possible use for development:

The identification of social disparities and the reasons behind them could provide more insight to the Senegal government on the issues to concentrate on, and help equip the deprived with necessary faciliites, thereby accelerating the development of the nation on various fronts.



Disease spread in Senegal



Factors for inequalities in Senegal

Main results:

- The analysis of the regionally categorized mobile phone data in Senegal, together with the associated statistics, identify a multitude of issues that impact the society.
  - Gender disparity prevails in employment due to the orthodox nature of women of certain ethnic groups
  - High disease spread and infant mortality rates are observed due to the lack of quality educational institutions and health care facilities, especially in rural regions.
  - Urbanization severely impacts the agricultural income and hence the economy of Senegal
  - Though there are governmental policies that address issues like sustainable development and environmental protection, their effects are not satisfactory.

Methods:

- The data available included both statistical and mobile phone logs. These were preprocessed and seggregated into region wise details.
  - The data derived from mobile phone dataset was used to identify the mobility patterns using location analysis, with the help of select attributes.
  - The statistical data was analysed based on various socio-economic parameters such as health, education, occupation, gender, etc.
  - The reults from both ends were combined to infer from and relevant conclusions were drwan upon, thus proving or disproving the hypotheses stated on the same.

Full paper is here:
put your link here

DataViz or video are here:
put your link here
Login:
Pw:

Data sources used for this project:

- D4D data set 1, com between antenna ☐
- D4D data set 2, movement routes high res ☐
- D4D data set 3, movement routes low res ☒
- D4D synthetic data set ☐

Other data sets used in this project:

- Type of data: Statistical Data from Senegal Government Websites Source:
- Type of data: Source:
- Type of data: Source:

Main Tools used:

- Hypothèses Testing
- Régression Line
- Inferential Statistics

Open Code available:

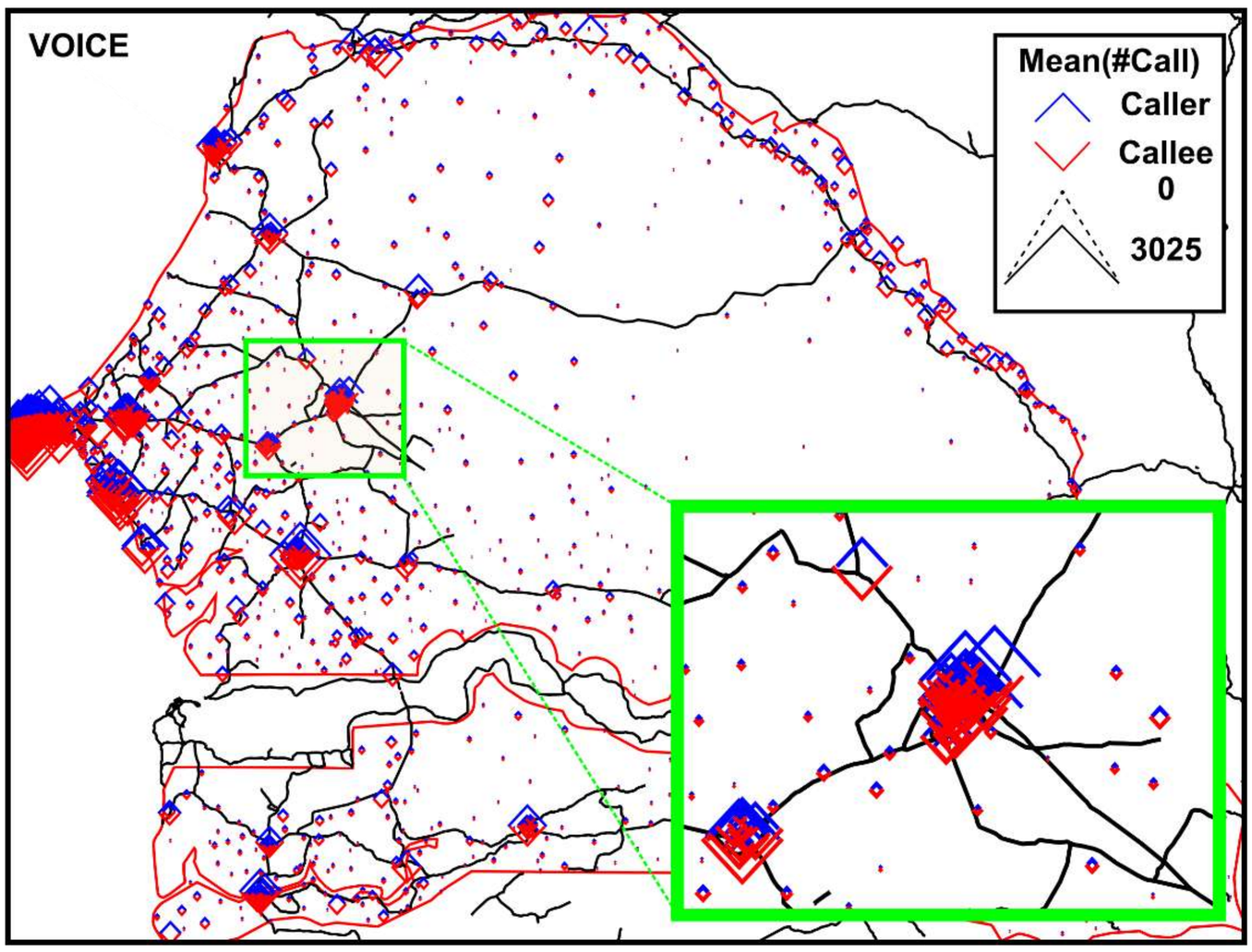
- Yes ☐
- No ☐

N05

Detection of Population Mobility Anomalies in Senegal from Base Station Profiles

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network

Spatial Distribution of Voice Activities



UNIVERSITÉ DU LUXEMBOURG

- Melakessou, Foued, Dr., University of Luxembourg
- Derrmann, Thierry, MCS., University of Luxembourg
- Frank, Raphaël, Dr., University of Luxembourg
- Castignani, German, Dr., University of Luxembourg
- Engel, Thomas, Prof. Dr., University of Luxembourg

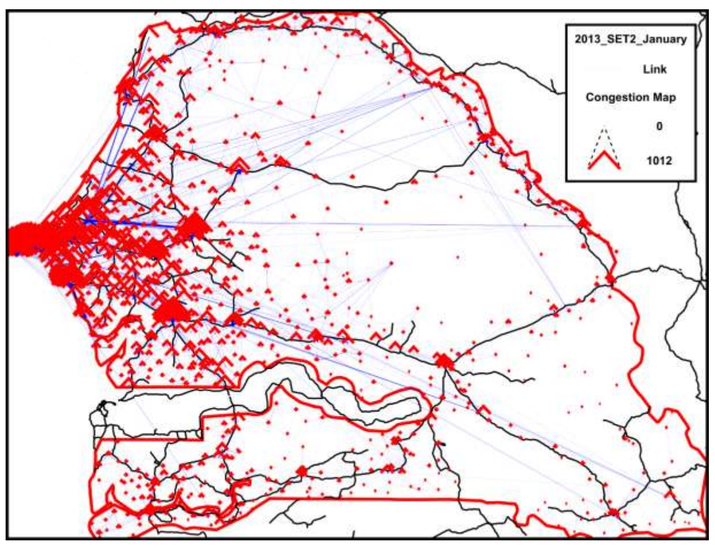
Project Summary:

The analysis of Call Detail Records has captured the attention of traffic and transportation researchers. Our project is focusing on the modeling of the daily traffic demand profile of each base station by considering voice and also short message services.

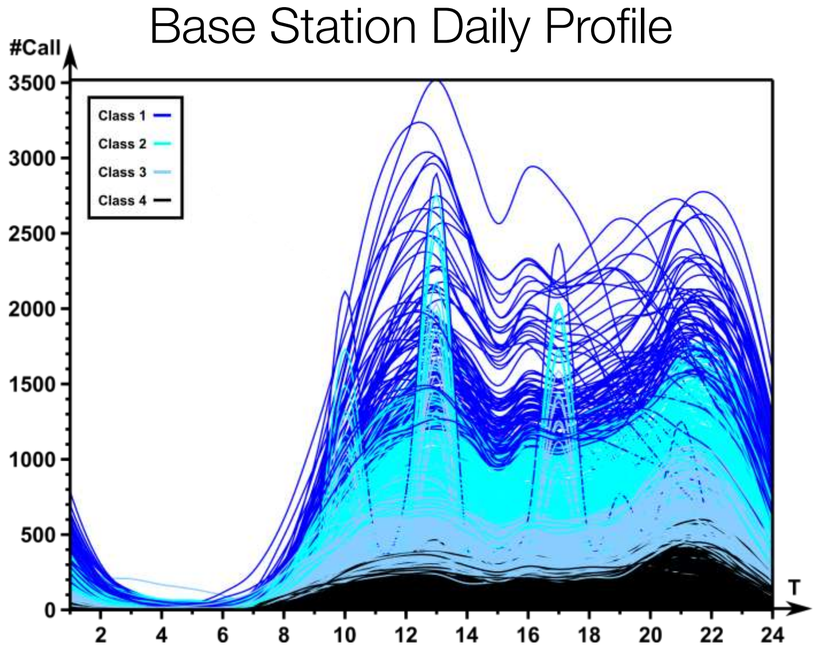
Possible use for development:

The evaluation of mobility models will help to better design and develop future infrastructures in order to support the actual demand. The modeling of each base station profile will be used to develop an anomalies detection system focusing on unusual situations of mass population movements.

Fine-Grained Mobility Graph



Base Station Daily Profile



Main results:

- 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. Magal of Touba and inauguration of the new highway between Dakar and Diamniado
- Mobility graphs
  - Computation of mobility flows
  - Performance of congestion maps

Methods:

- K-Means (classification)
- Anomaly detection algorithm (outliers rejections)

Full paper is here:
http://goo.gl/J0d7YG

Data sources used for this project:

D4D data set 1, com between antenna	X
D4D data set 2, movement routes high res	X
D4D data set 3, movement routes low res	X
D4D synthetic data set	

Other data sets used in this project:
- Type of data: OpenStreetMap
- Source: http://www.openstreetmap.org

Main Tools used:
- Tool 1: Scilab
- Tool 2: Perl scripts
- Algorithm: k-Means, Outliers Rejection

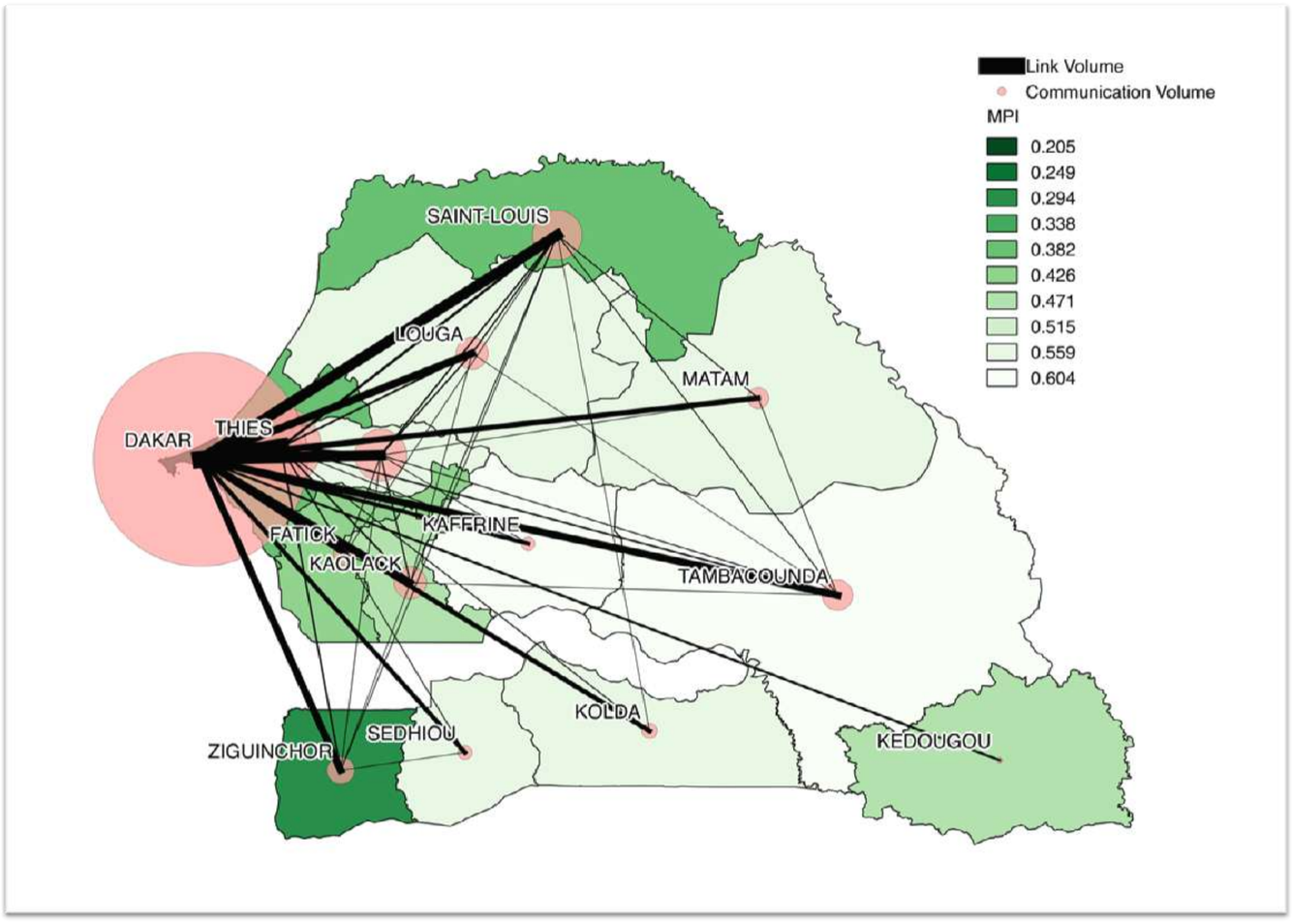
Open Code available:

Yes	X
No	

N07

Virtual Networks and Poverty Analysis in Senegal

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



Virtual network for Senegal at region level with MPI (Multi-dimensional Poverty Index) as an overlay. Thickness of links indicates the volume of calls and texts exchanged between a pair of regions. Size of the circle at each region indicates the total number of incoming and outgoing calls and texts for the region. Note that regions with plenty of strong links have lower poverty, while most poor regions appear isolated.

University at Buffalo
The State University of New York

- Pokhriyal, Neeti, PhD Student, SUNY Buffalo
- Dong, Wen, Assistant Professor, SUNY Buffalo
- Govindaraju, Venugopal, Professor, SUNY Buffalo

Project Summary:

Can the accessibility of mobile technology be used to identify, characterize and alleviate poverty? This project is an attempt to answer this question. We conduct two studies.

1.Using the cellular-communications data, we construct virtual connectivity maps for Senegal, which are then correlated with the poverty indicators to learn a model. Our model predicts poverty index at any spatial resolution, thus generating a poverty map for Senegal at an unprecedented resolution.

2.We study how user behavioral statistics, gathered from cellular-communications, correlate with the poverty indicators. Can this relationship be learnt as a model to generate poverty maps at a finer resolution? Using only this relationship can give us a poverty map.

Since poverty is a complex phenomenon, poverty maps showcasing multiple perspectives, such as ours, provide policymakers with better insights for effective responses for poverty eradication.

Use for development:

- Poverty maps at arrondissement/department levels, or at any spatial levels, will enable targeted policies for inclusive growth of all the regions in Senegal.
- Poverty maps built using the behavioral indicators can focus policies for certain demographics of the society that are specially vulnerable to poverty, such as women and specific ethnic groups.

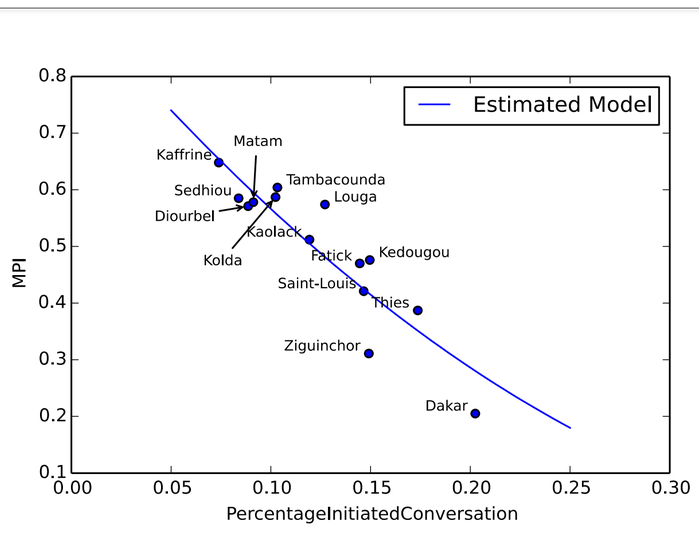


Fig 1a. Estimated model for predicting MPI using behavioral indicators.

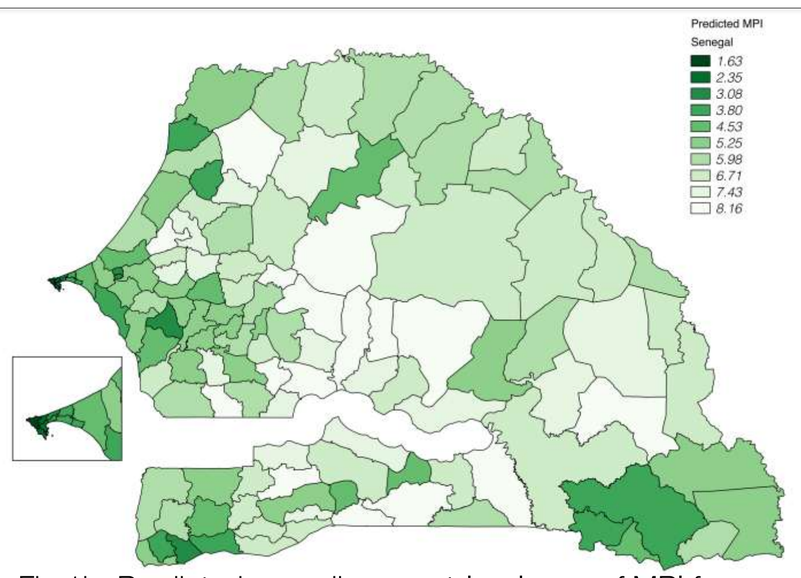


Fig 1b. Predicted arrondissement-level map of MPI for Senegal using behavioral indicators.

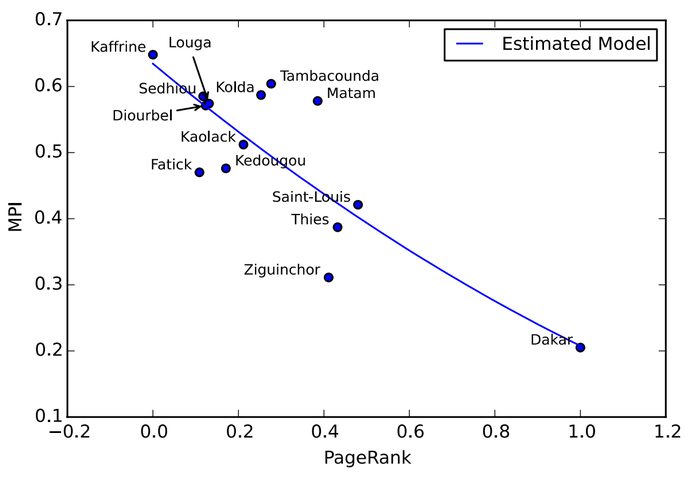


Fig 2a. Estimated model for predicting MPI using the virtual network.

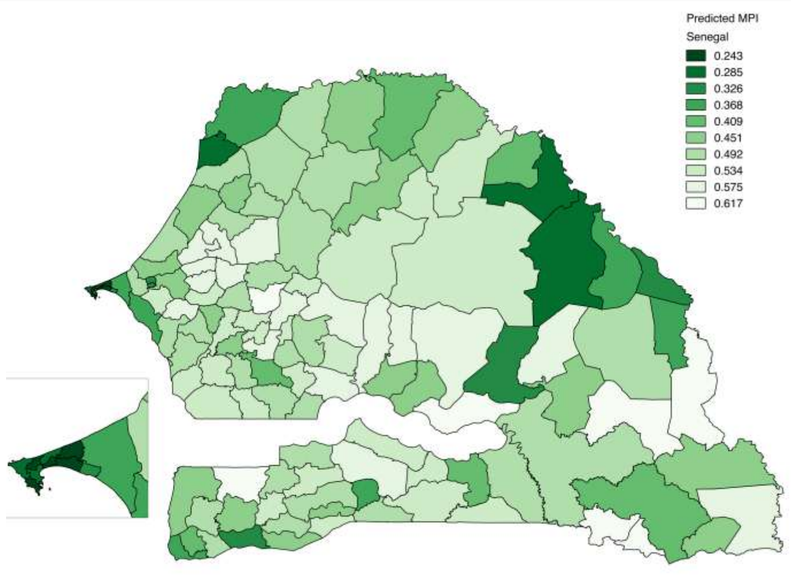


Fig 2b. Predicted arrondissement-level map of MPI for Senegal using the virtual network.

Main results:

- Model to predict MPI using user behavior indicators (Fig 1a)
- Poverty map at arrondissement level using behavioral indicators (Fig 1b)
- Model to predict MPI using centrality features extracted from virtual network (Fig 2a)
- Poverty map at arrondissement level using the virtual network (Fig 2b)

Methods:

Virtual network analysis

- We use Dataset 1 to construct the raw communication matrix at region level.
- The matrix is normalized to remove influence of population and geographical distances.
- Network centrality measures (page rank, Eigen vector centrality) are computed from the normalized matrix.
- Learnt a linear regression model to predict MPI using the centrality measures.
- Predicted MPI at arrondissement level using the learnt model and generated a poverty map.

User Behavior analysis

- We localize users to arrondissements by analyzing their yearly trajectories from Dataset 3. For analysis we focus on 33,323 users with reliable localization information.
- From the user sample, we compute region level aggregates for 33 behavior indicators.
- Region level indicators are correlated with MPI and a model is learnt using the best indicator and generated a poverty map.

Full paper is here:
http://www.cse.buffalo.edu/~neetipok/D4DSenegal/pokhriyal_fullpaper.pdf

Code is here:
https://github.com/neetip/D4D_Senegal

Data sources used for this project:

- D4D data set 1, com between antenna [X]
- D4D data set 2, movement routes high res [ ]
- D4D data set 3, movement routes low res [X]
- D4D synthetic data set [ ]

Other data sets used in this project:

- Multidimensional Poverty Index Data http://www.ophi.org.uk/multidimensional-poverty-index/
- Human Development Index (2014) Data http://hdr.undp.org/en/content/human-development-report-2014

Main Tools used:

- iPython Notebook, SciPy
- QGIS
- Networkx library

Open Code available:

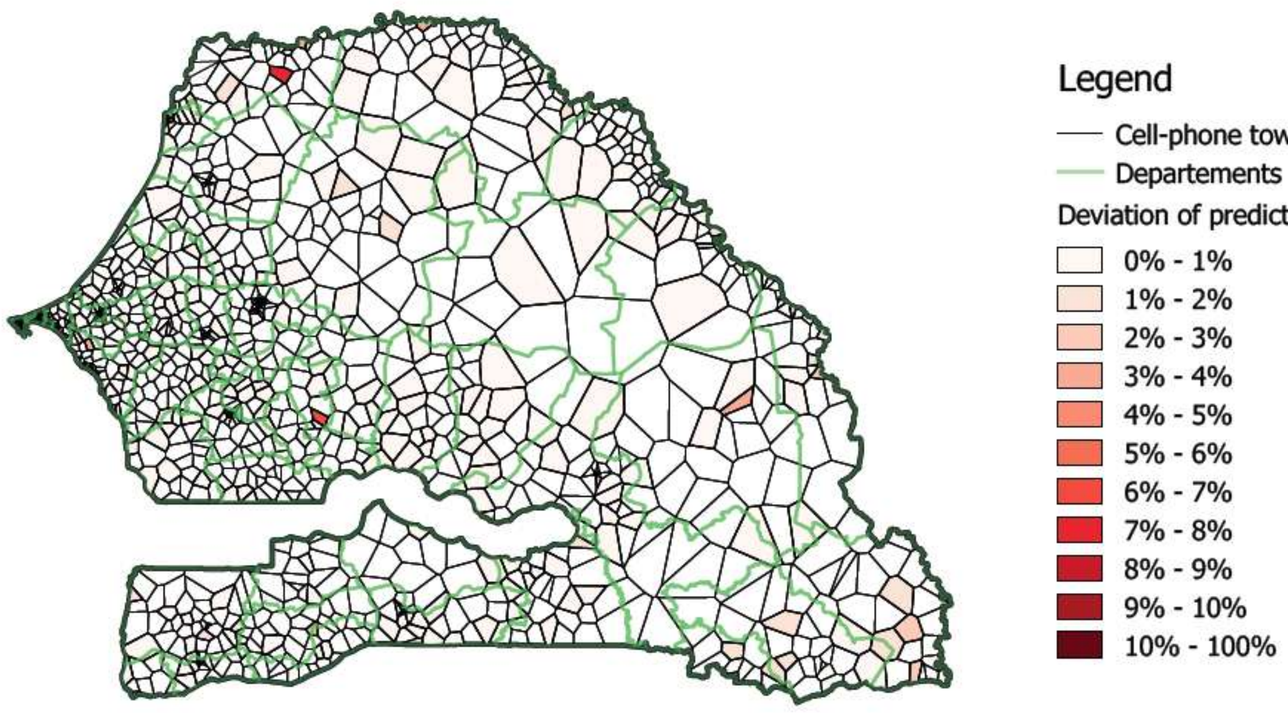
- Yes [X]
- No [ ]

N09

Estimating Population Density and Climate Vulnerability Using CDRs

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network

CDR-based vs. Official Population Densities in Senegal



- Letouzé, Emmanuel, Director, Data-Pop Alliance (lead author)
- Areias, Ana, Program Manager, Data-Pop Alliance
- Pestre, Gabriel, Research Assistant, Data-Pop Alliance
- Schwarz, Bessie, Project Coordinator, Yale University
- Tellman, Beth, PhD Student, Arizona State University
- Zagheni, Emilio, Assistant Professor, University of Washington

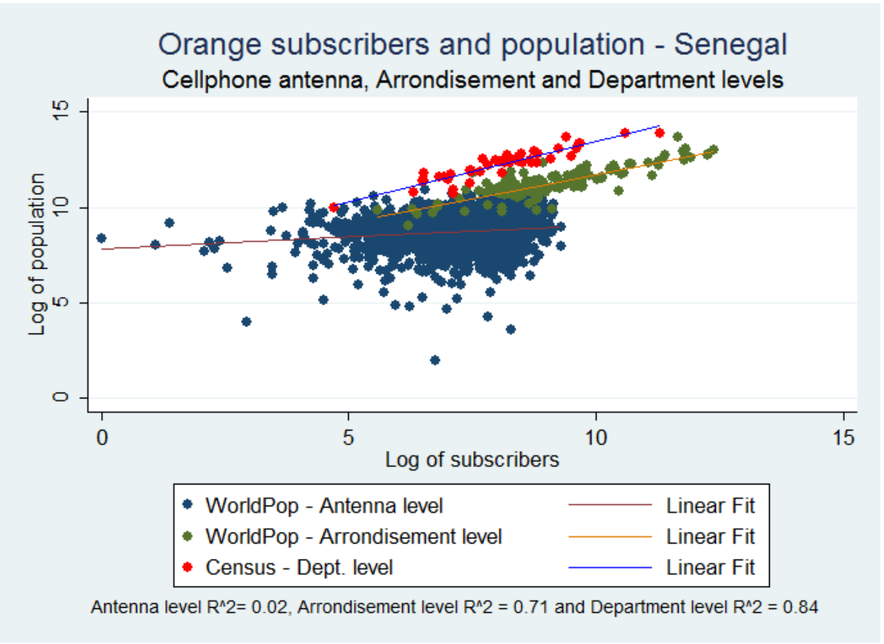
Project Summary:

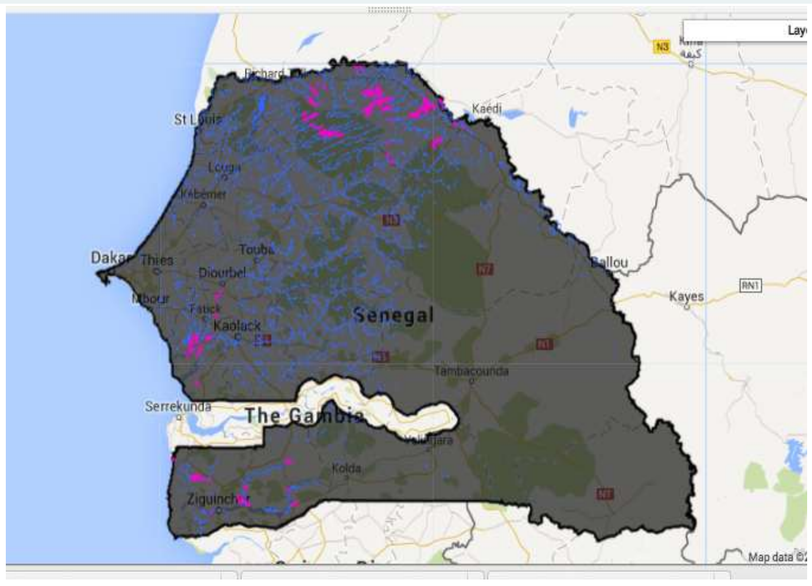
We investigate whether cell-phone activity could yield more accurate and timely population density estimates than official statistics, looking in particular at how this may affect climate vulnerability monitoring.

Controlling for cell-phone penetration, our initial results show that estimates from cell-phone activity are very close to census data at high aggregation levels, but suggest large deviations at finer geographical and temporal granularities that give notably different pictures of vulnerability to climate shocks.

Possible use for development:

Getting reliable real-time population density estimates would help better target and coordinate assistance during sudden-onset crises such as flooding.





Flooding vulnerability using cell-phone data

Main results:

- Our results can be summarized in three main points:
  - The correlation between CDR-based population estimates and official population figures is strong at the department level, but weakens as the observed geographic units decrease in size, as shown in the graph (left).
  - Using the CDR-based population proxy instead of official population figures yields different flood vulnerability maps, which has an impact on which communities should be targeted for climate crisis monitoring.
  - The empirical differences we observe at higher geographic and temporal granularities lead us to question which figures are more reliable for real-time applications: does the accuracy of 'official' population figures hold as spatial and temporal granularity increase, or would CDR-based measures of population density, adjusted for sample biases, provide a clearer picture of how vulnerable populations are distributed in space and time?

Methods:

- Using data from Senegal's census, Demographic and Health Surveys (DHS) program, and WoldPop, we gathered population data at the department, arrondissement, and cell-phone tower levels over several time periods.
- We then used a linear regression model to estimate the parameter k in the following equation: $log(P_i) = k\ log(U_i) + e_i$ where is the official population of geographic region i and U_i is the number of active cellphone users in the Orange sample in the geographic region.
- We then assumed that variations in mobile phone penetration rates across the geographic areas generate some systematic bias in the model: $log(P_i) = k\ log(U_i) + bias_i + \varepsilon_i$ where $bias_i = f(mobile\ phone\ penetration_i)$
- Using data from the DHS for the Senegal, conducted in 2013, we looked for the best way to adjust the CDR-based model to account for mobile phone penetration.
- We then ran the flood vulnerability model developed and owned by two co-authors of this paper, Bessie Schwarz and Beth Tellman, but replaced the initial population density data (from WorldPop) with our CDR-based estimate, maintaining the same area of analysis: the number of highly vulnerable people exposed to flooding increases to 22,000 people in our new prediction.

Full paper is here:

https://www.dropbox.com/l/L0Bf2uL4w39rP7jhKL1Aws

Data sources used for this project:

- D4D data set 1, com between antenna [X]
- D4D data set 2, movement routes high res [ ]
- D4D data set 3, movement routes low res [ ]
- D4D synthetic data set [ ]

Other data sets used in this project:

- Type of data: Senegal Preliminary Census data 2013
- Type of data: DHS data 2013
- Type of data: WoldPop data 2014

Main Tools used:

- Google Earth Engine
- QGIS / ArcGIS
- R / STATA

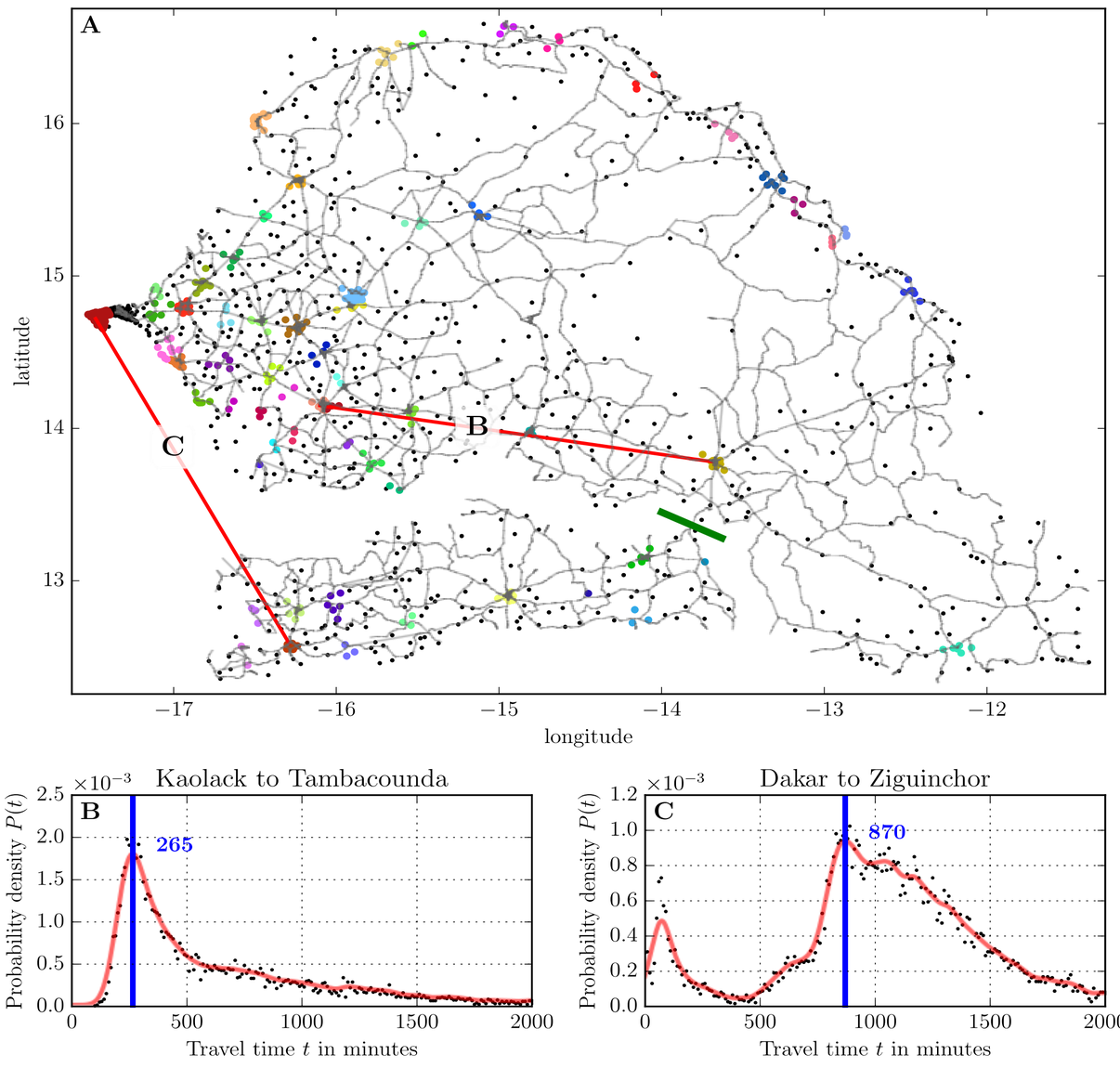
Open Code available:

- Yes [X]
- No [X]

T01

Roads+: Estimation of Travel Times Between Cities in Senegal From CDR's

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



- Saramäki, Jari, Associate Professor, Aalto University
- Kujala, Rainer, PhD student, Aalto University
- Aledavood, Talayeh, PhD student, Aalto University

Project Summary:

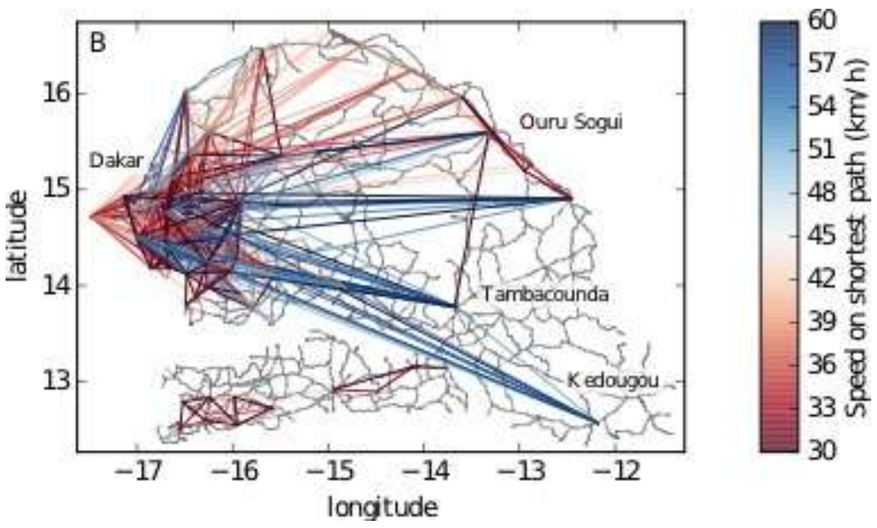
We have developed a method for estimating travel times between cities from CDR's, and applied it to data from Senegal. As the outcome, we present tables of typical travel times and speeds between cities in Senegal. This information fills an existing gap for Senegal (there are no accurate travel time estimates available), but the method is readily applicable anywhere where CDR information is available.

Possible use for development:

Estimation of travel times
Monitoring of road conditions
Information for road infrastructure planning

	Dakar	Thies	Kaol.	M'Bour	Sain.	Tamb.	Mbac.
Dakar		125	280	130	300	620	290
Thies	125		165	80	200	405	155
Kaolack	265	170		135	320	265	125
M'Bour	125	80	145		300	445	230
Saint-Louis	305	200	330	300		660	255
Tambacounda	580	440	270	395	700		415
Mbacke	270	145	115	205	270	370	

Travel times between selected cities (in minutes)



Comparison of travel speeds between different cities

Main results:

- We have used CDR's to estimate travel times between cities in Senegal
  - Results appear meaningful even with limited CDR time resolution (10 min) and limited spatial resolution of tower locations
  - Suggested use for method: near-real-time monitoring done by mobile operator, using high-resolution data
  - Method detects roads with lower-than-average travel speeds, points out problems

Methods:

- Travels between cities are extracted from mobility trace data
- Distributions of travel times are then analyzed for the typical travel time (peak detection)
  - Validation: travel time symmetry (from A to B vs from B to A), linear distance dependence, comparison with other sources
  - Results improve with larger sample; very good results expected for larger amounts of data with better temporal and spatial resolution
  - Allows visualization (see Fig on the left)

Full paper is here:
http://becs.aalto.fi/~rmkujala/d4d/RoadsPlus_paper.pdf

Data sources used for this project:

- [ ] D4D data set 1, com between antenna
- [x] D4D data set 2, movement routes high res
- [ ] D4D data set 3, movement routes low res
- [ ] D4D synthetic data set

Other data sets used in this project:

- Type of data: Source:
- Type of data: Source:
- Type of data: Source:

Main Tools used:

- Python
- QGIS

Open Code available:

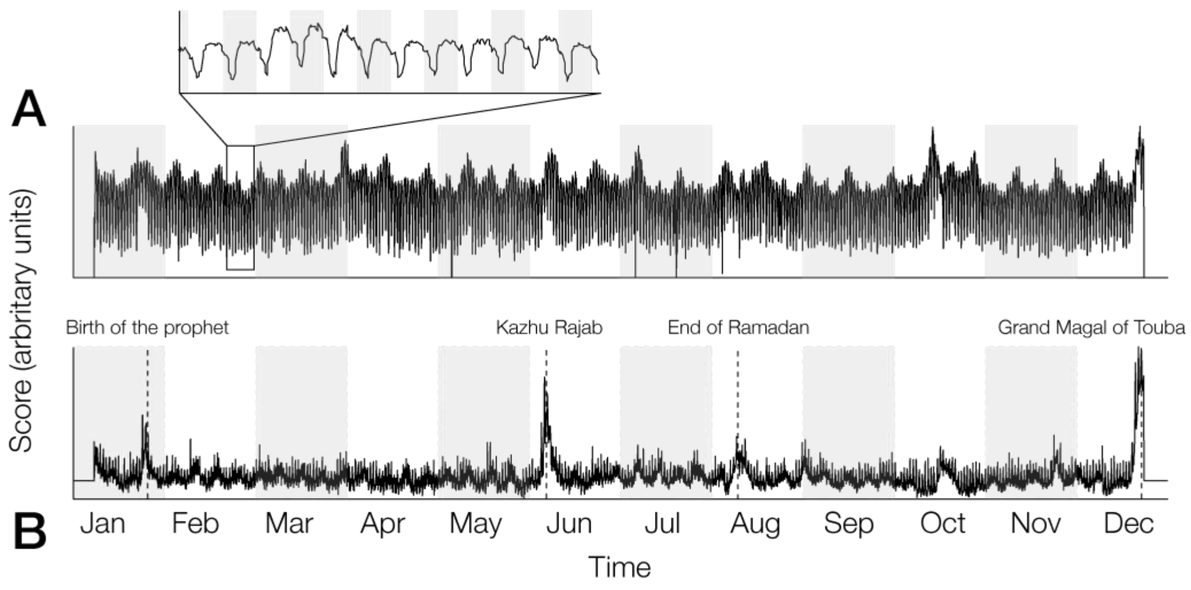
- Yes [x]
- No [ ]

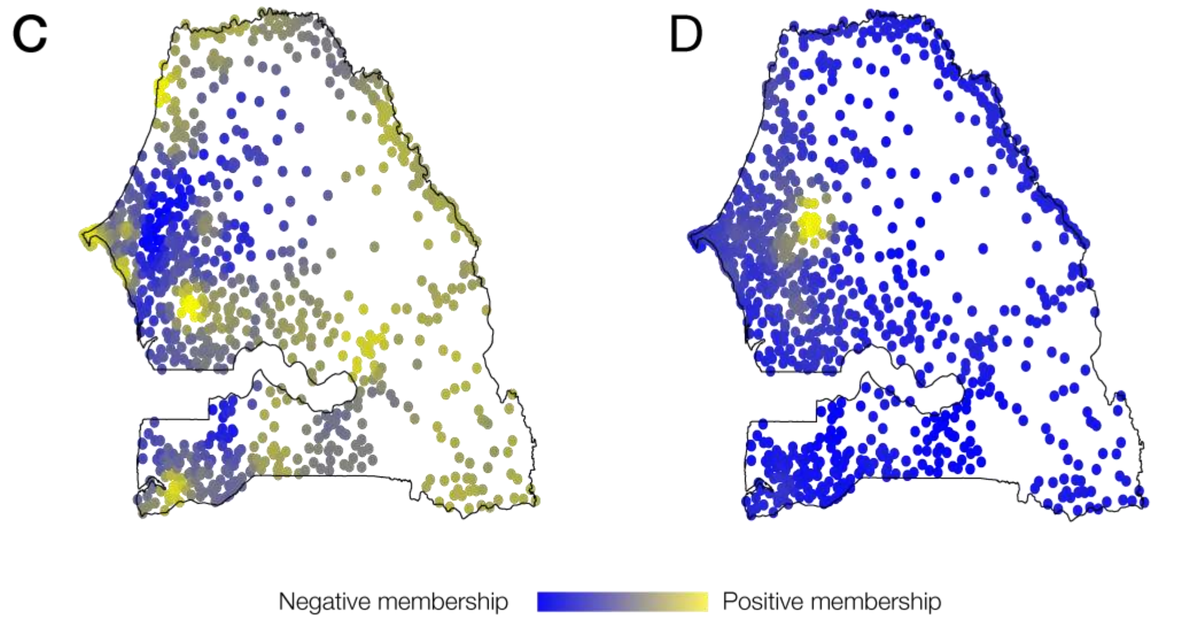
T02

Deviations from the norm:

Detecting regularities and anomalies in human mobility patterns

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network





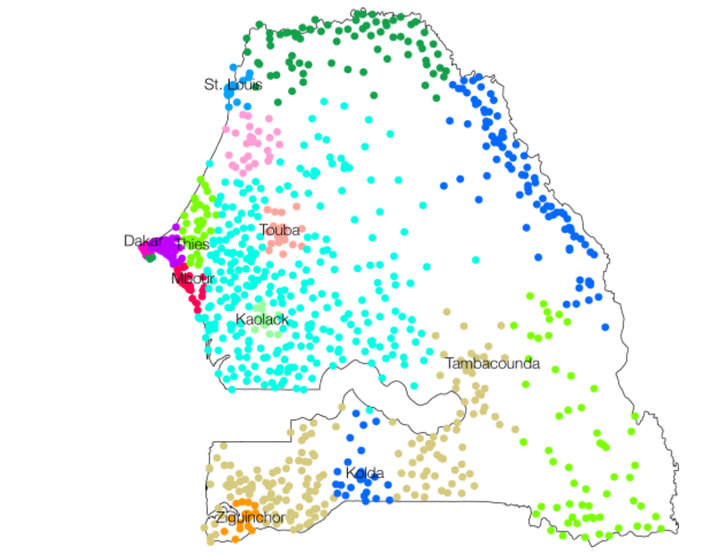
- Dr. Gijs Joost Brouwer, Data scientist, Integral Ad Science
- Prof. Foster Provost, New York University

Project Summary

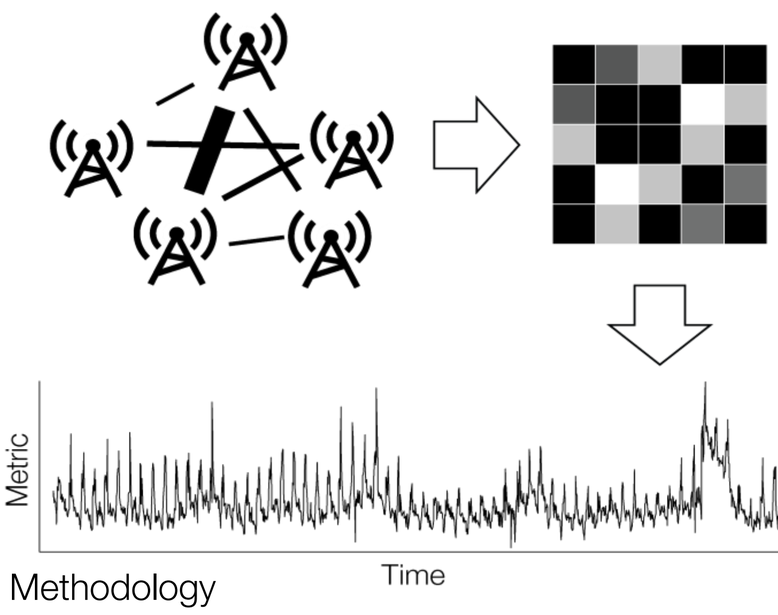
We defined a set of metrics that capture both regularities and anomalies in human mobility patterns. For each user, we extracted the most common cell phone tower, the surprise of that user visiting a particular cell phone tower, and the probability of seeing transitions between two specific cell phone towers. Tracking these metrics over time allowed us to observe both regular and anomalous patterns developing.

Possible use for development

By detecting regularities, our metrics can be applied to infrastructural planning, while by detecting anomalies, the metrics become an early warning system for events that change or even disrupt daily life.



Metric based communities



Methodology

Main results

▪We provided proof of principle that a set of simple metrics assigned to each cell phone tower allows for the detection of both regular and anomolous human mobility patterns. Furthermore, we showed that regions cluster together in terms of weights assigned to each detected mobility pattern.

Methods

▪For each user, we used the sequence of transitions between cell phone towers to derive 1) a transition matrix T, 2) the most frequently observed cell phone tower and 3) the frequency of visiting destination each tower, regardless of the origin.

▪Replaying the same sequence of events gave us the probability of observing a particular transition, the surprise of seeing a user visit a specific cell phone tower (1 – frequency) and the distance between the destination cell phone tower and the most common cell phone tower for that user.

▪We grouped transitions by their destination tower and the hour in which they occurred, averaging the metrics associated with these transitions.

▪Principal component analysis was used to extract mobility patterns that were shared by some or many cell phone towers.

Full paper is here:
put your link here

Data sources used for this project

- D4D data set 1, com between antenna
- D4D data set 2, movement routes high res
- D4D data set 3, movement routes low res
- D4D synthetic data set

Other data sets used in this project

- Type of data: news archives
- Type of data: meterological data

Main Tools used

- [ ] Apache Hadoop, Spark and Pig
- [x] Python, Matlab
- [ ] Principal component analysis, information theory, community detection.
- [ ] Other

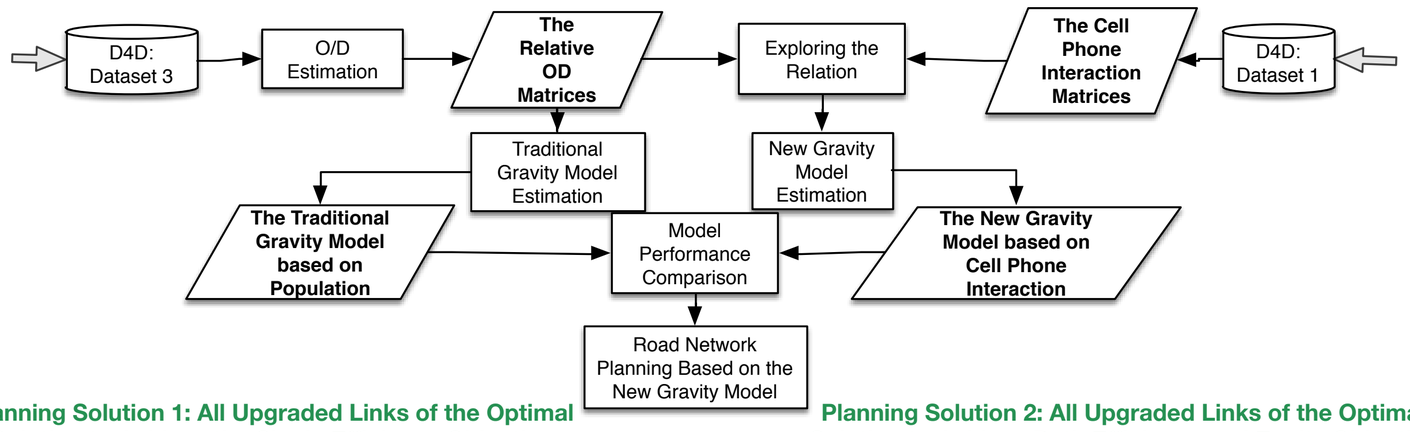
Open Code available

- [x] Yes
- [ ] No

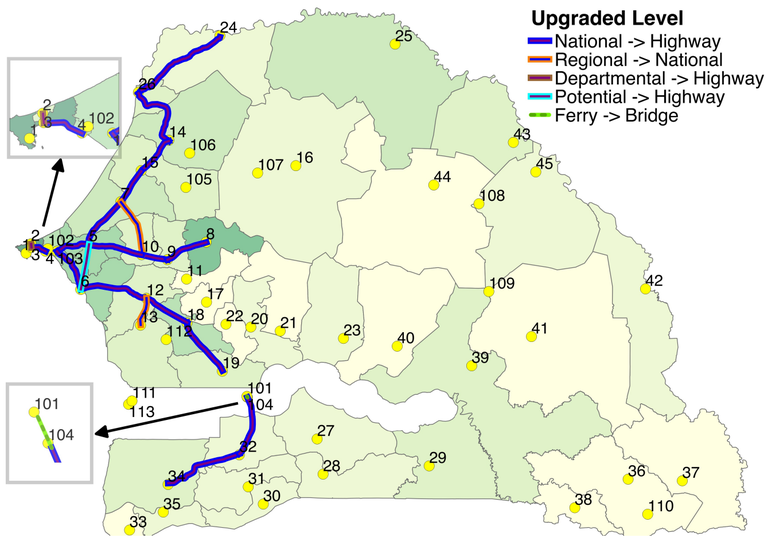
T03

National and Regional Road Network Optimisation for Senegal Using Mobile Phone Data

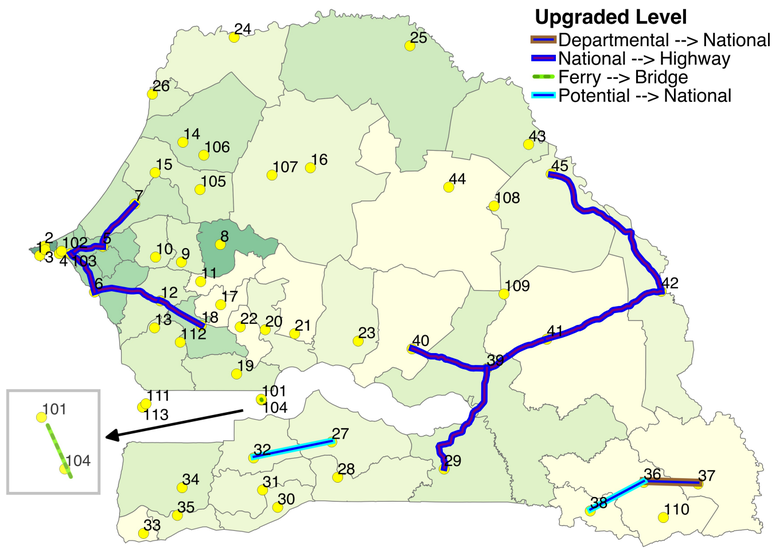
Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



Planning Solution 1: All Upgraded Links of the Optimal Solution to Achieve the Single Efficiency Objective



Planning Solution 2: All Upgraded Links of the Optimal Solution to Achieve Both Efficiency and Equity Objective



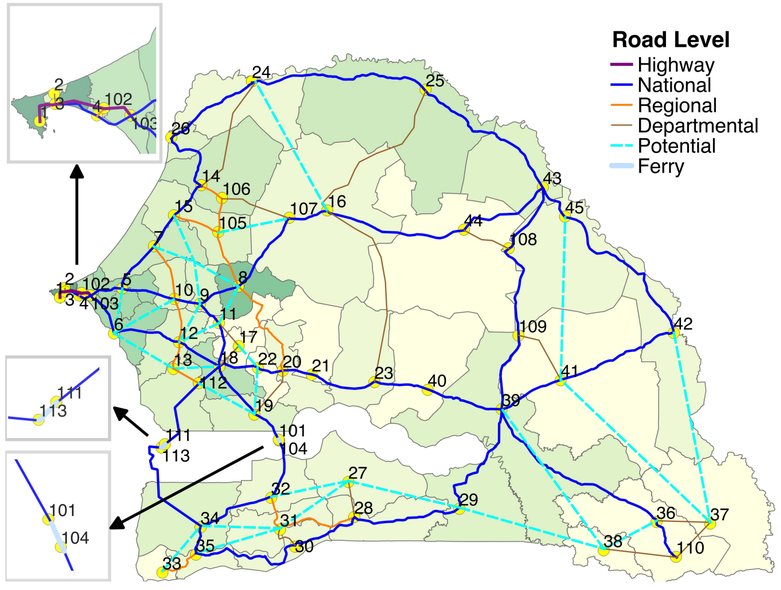
- Erik de Romph, Professor
- Gonçalo H. A. Correia, Assistant Professor
- Yihong Wang, Master Student
- Department of Transport and Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology

Project Summary:

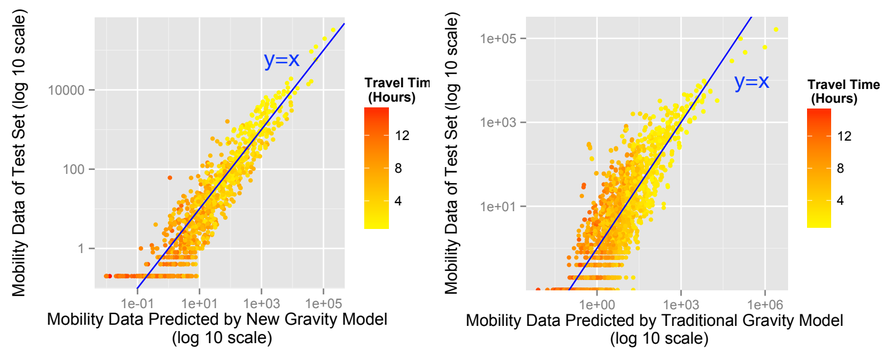
Based on the cell phone interaction data (Dataset 1) and the mobile phone traces (Dataset 3), we found that the mobility between departments was proportional to the aggregated number of cell phone interactions between departments and inversely proportional to the travel costs between departments in Senegal. To that extent, using the filtered mobile phone traces, we estimated a new gravity model based on the number of cell phone interactions, and compared it with the estimated traditional gravity model based on population regarding the model fitness and the predictive accuracy. Because of the better model performance, the estimated new gravity model based on the number of cell phone interactions was used to solve the road network design problem in Senegal.

Possible use for development:

1. The filtering algorithm introduced in this project can be used to filter the mobile phone traces (Dataset 3) and to improve the OD estimation.
2. The empirically found relation between tele-communication and travel, and the new gravity model based on cell phone interactions (Dataset 1), allow the government to better understand and predict mobility patterns in Senegal.
3. Based on the actual needs of planning, the policy-makers can determine the weights of different objectives and the actual available budget in the optimisation model by themselves, in order to obtain the best solution under a certain scenario.



Graph 1: Potential Links to Be Added and Existing Links to Be Upgraded



Graph 2: The Comparison between the New Gravity Model Based the Number of Cell Phone Interaction (RMSE: 5590) and the Traditional Gravity Model Based on Population (RMSE: 157229) Regarding Their Model Performance of Predicting the Test Set

Main results:

- The relative OD matrices, which can reflect the current mobility pattern in Senegal, are estimated.
- The new gravity model based on the number of cell phone interactions, which can predict the future mobility pattern in Senegal, performs better than the traditional gravity model.
- Under assumed budget constraints, our optimisation model gives some interesting results:

A. The Trans-Gambia ferry service is suggested to be replaced by a bridge for the objective either of efficiency or of equity. Actually, this has been planned for a long time, though the plan has not come to fruition.

B. The Dakar-Diamniadio highway is suggested to be extended to Thies and Mbour for the efficiency objective and for the objective of 50% efficiency and 50% equity even under a rather low budget constraint. This suggestion is exactly similar with what the government of Senegal is planning as the phase 2 of the Dakar Toll Road Project.

C. For the single efficiency objective, the focus of road development is suggested to be on the western part of Senegal, where the departments are densely populated, while for the single equity objective, the focus of road development is suggested to be on the southeastern part of Senegal, where the departments are less densely populated.

Methods:

- The OD estimation using mobile phone traces (Dataset 3): a filtering algorithm is applied to filter the mobile phone traces and estimate the relative OD matrices, incorporating the number of trips between departments made by sampled users in 2013.
- The cross-validation technique for gravity model selection: the estimated relative OD matrices are classified into two sets, the training set, which is used to estimate the gravity models, and the test set, which is used to test how accurately the estimated gravity models can predict the test set.
- The non-linear optimisation-based multi-objective road network design model: the objective functions of efficiency and equity, which are originally based on the traditional gravity model, are adapted to the ones based on the new gravity model. In addition, a local search algorithm including the steps of add, interchange and drop is applied for solving the heuristic problem efficiently.

Full paper is here:
put your link here

DataViz or video are here:
put your link here
Login:
Pw:

Data sources used for this project:

- D4D data set 1, com between antenna [x]
- D4D data set 2, movement routes high res [ ]
- D4D data set 3, movement routes low res [x]
- D4D synthetic data set [ ]

Other data sets used in this project:

- Type of data: Census Data Source: ANSD
- Type of data: GIS Data Source: ArcGIS, AATR and OpenStreetMap

Main Tools used:

- R: the language used for data processing
- QGIS and ArcGIS: the tools for network analysis and GIS visualisation
- Algorithm 1: a filtering algorithm for OD estimation
- Algorithm 2: a local search algorithm for solving network design problem

Open Code available:

- Yes [x] (all algorithms are available.)
- No [ ]

T04

Building workers' travel demand models based on mobile phone data

Transport
Urban

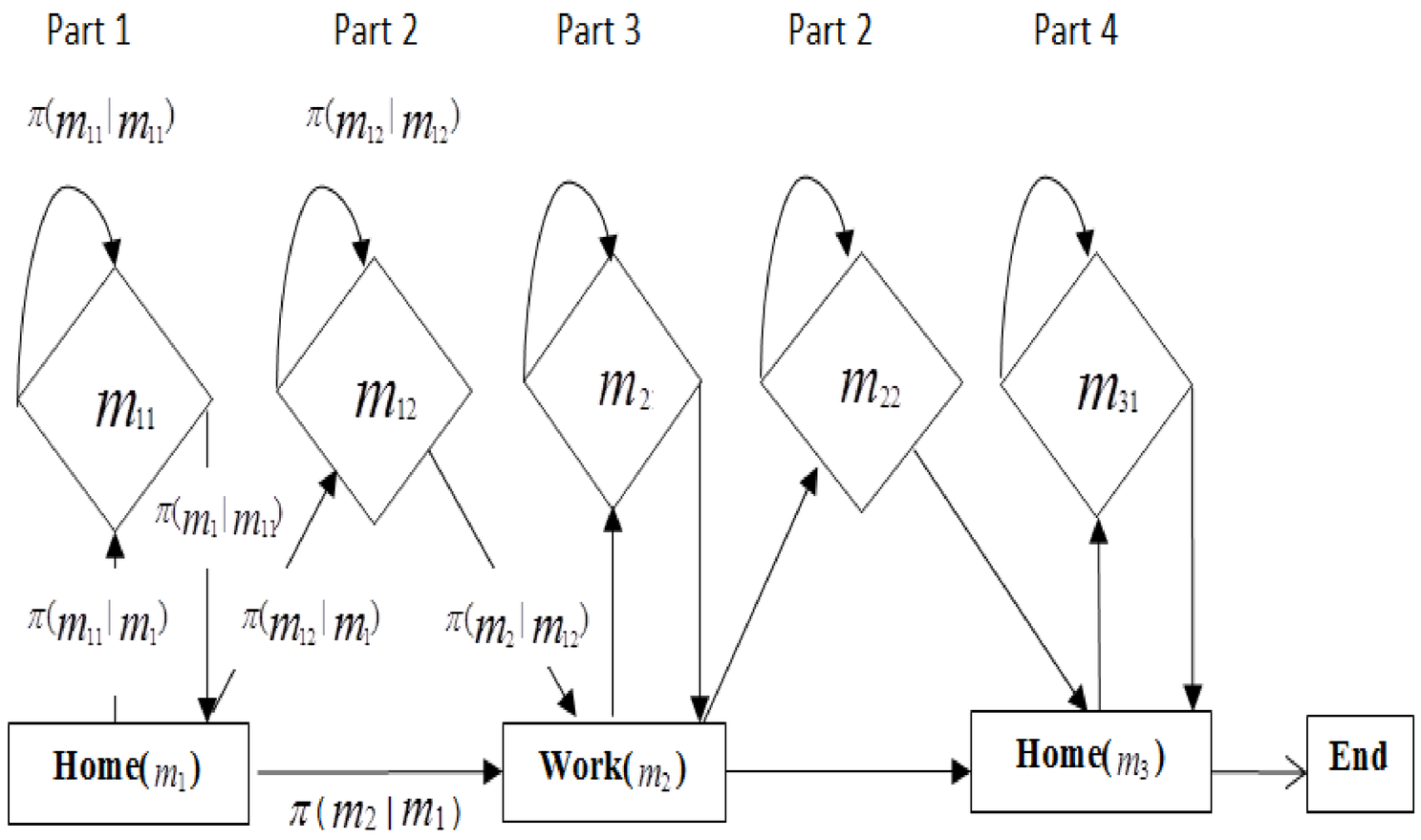


Fig. The HMM for a home-based-work-tour

- Liu, Feng , dr., Transportation Research Institute (IMOB), Hasselt University
- Janssens, Davy , dr., IMOB, Hasselt University
- Wets, Geert , dr., IMOB, Hasselt University

Project Summary:

We have developed a workers' travel demand model based on mobile phone data. This model captures the probabilistic distribution of activity locations and their sequential orders revealed by the call location trajectories.

Possible use for development:

Daily activity-travel sequences of each of the employed people in a region can be simulated by this model. The obtained sequences can serve as a key input for travel demand analysis and forecasting in the area.

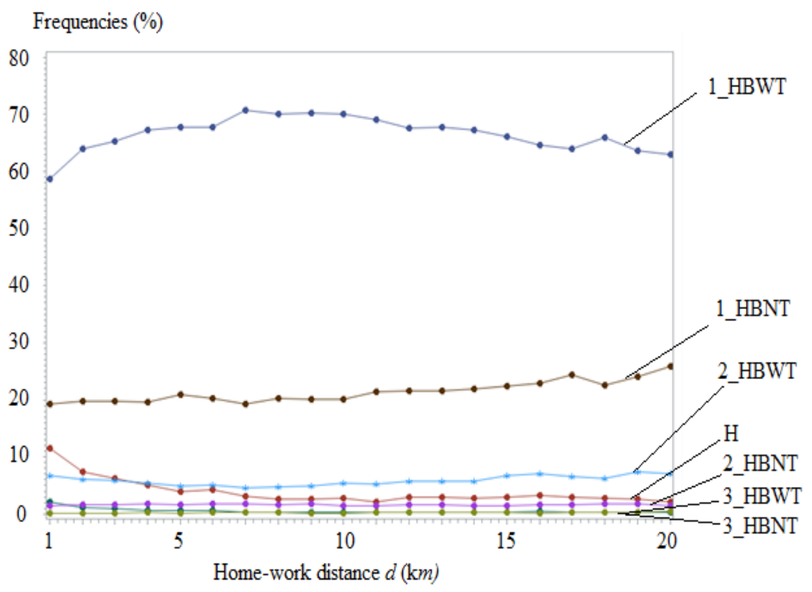


Fig. Distribution of sequence frequencies in each class over home-

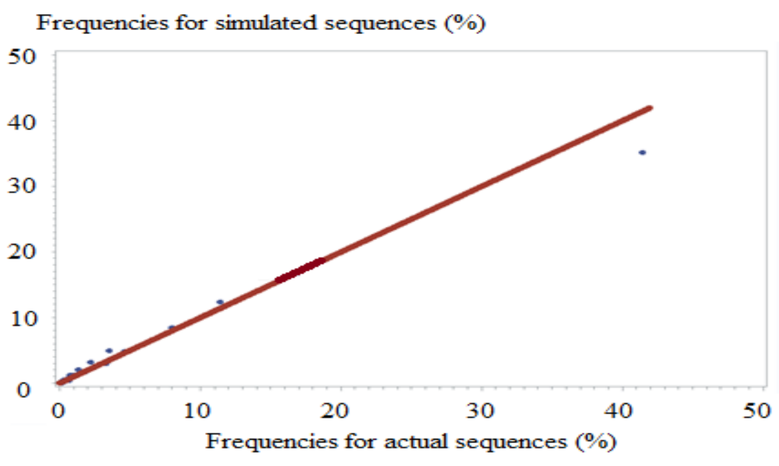


Fig. Correlation of sequence frequencies for each type between call location sequences and simulated ones

Main results:

- The average length of daily sequences drawn from the call-location-trajectories and the simulated results is 4.55 and 4.72, respectively.
- Among all the 677 types of the call-location-trajectories, 520 (e.g. 76.8%) are observed from the simulated sequences, and the correlation of sequence frequency distribution over all the types between these

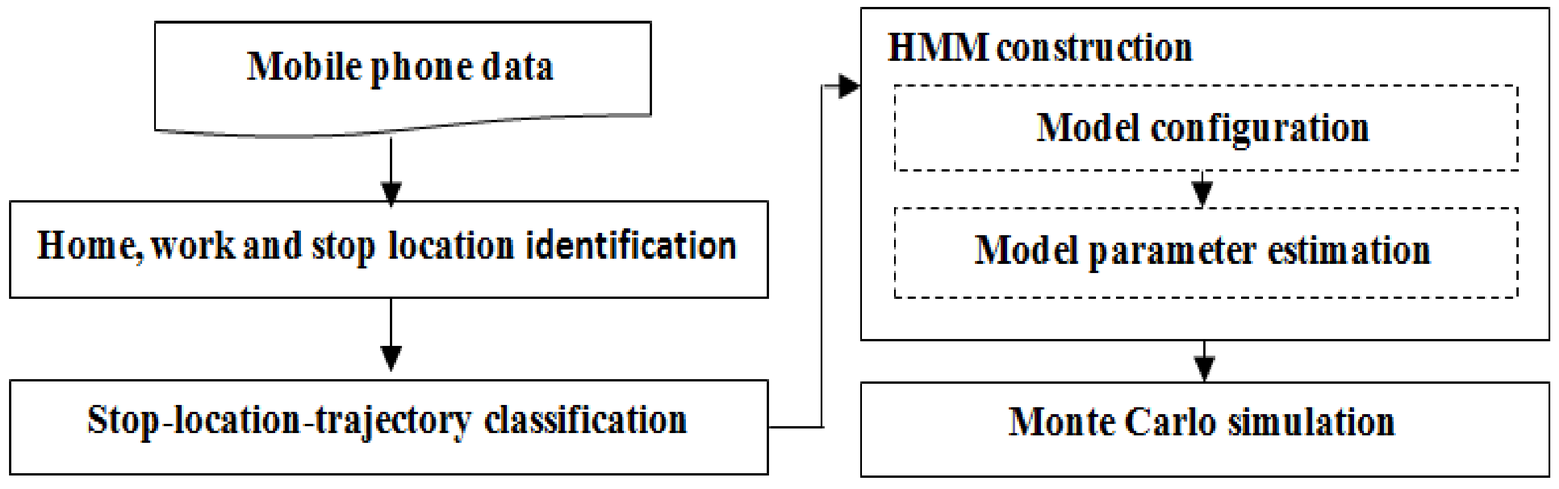


Fig. The overall structure of the methodology

Data sources used for this project:

- D4D data set 1, com between antenna
- D4D data set 2, movement routes high res

[ ]
[ ]
[x]
[ ]

Main Tools used:

- Tool 1: **Statistical Analysis System (SAS)**
- Algorithm: **Hidden Markov Model (HMM)**
- Algorithm: **Monte Carlo simulation**

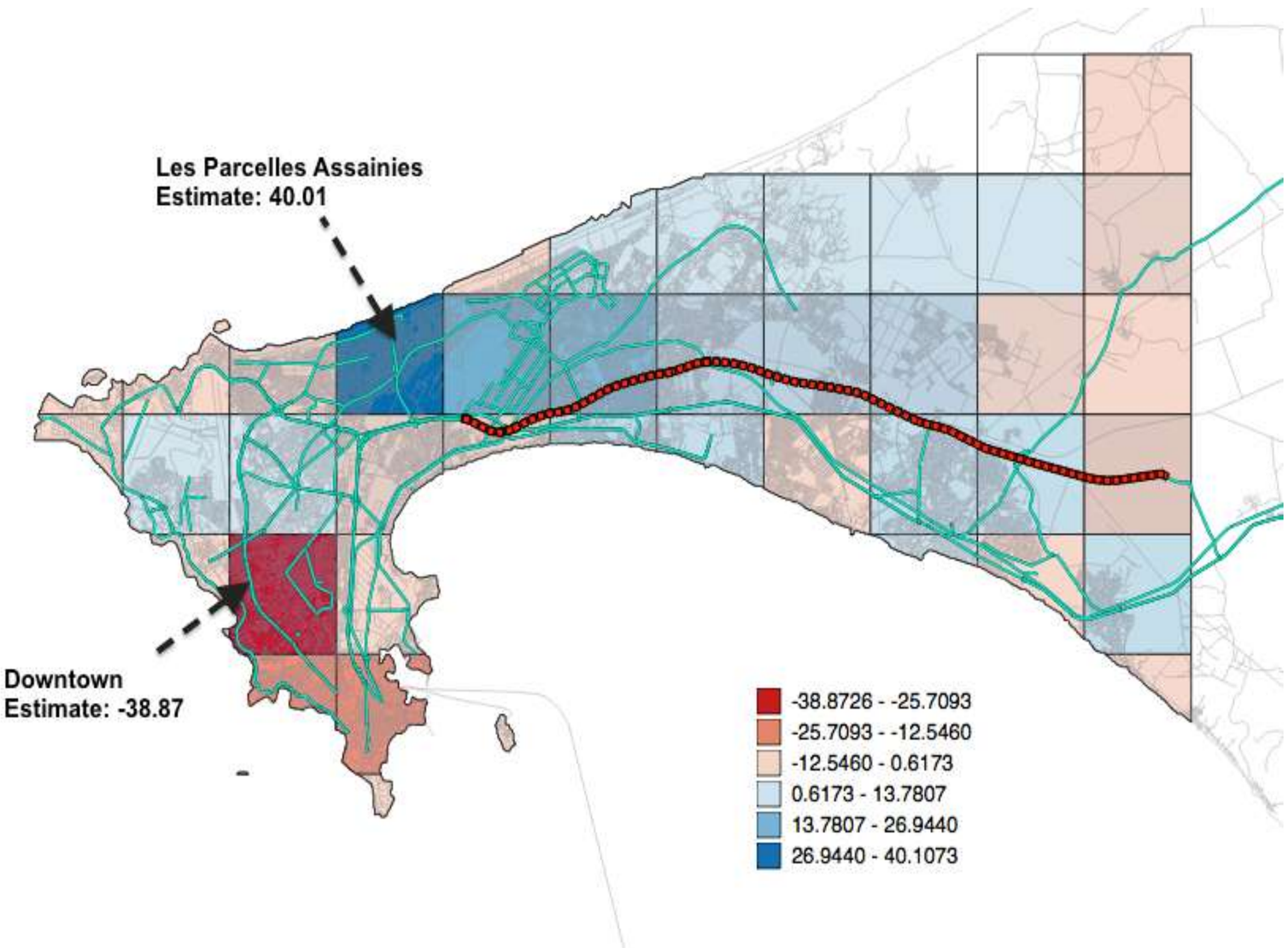
Open Code available:

- No

T05

Impact of Transport Infrastructure on Urban Mobility: Evidence from Dakar

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



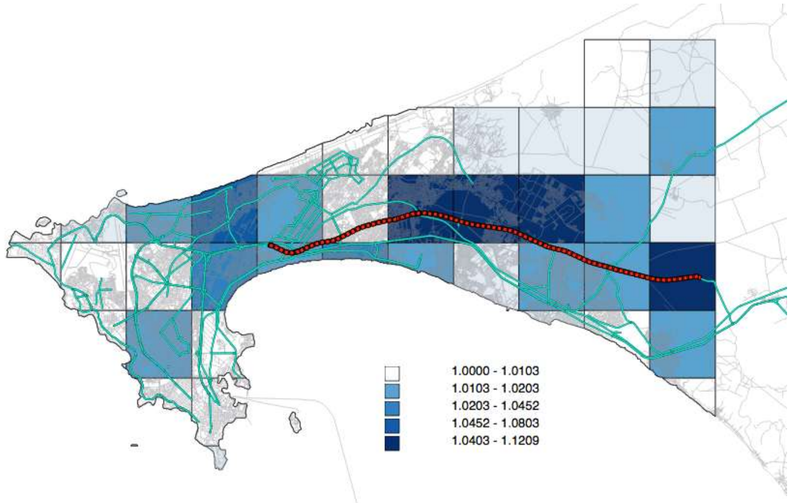
- Fetzer, Thiemo, London School of Economics (Lead Author)
- Sy, Amadou, Brookings Institution
- Arezki, Rabah and Chan-Lau, Jorge, IMF

Project Summary:

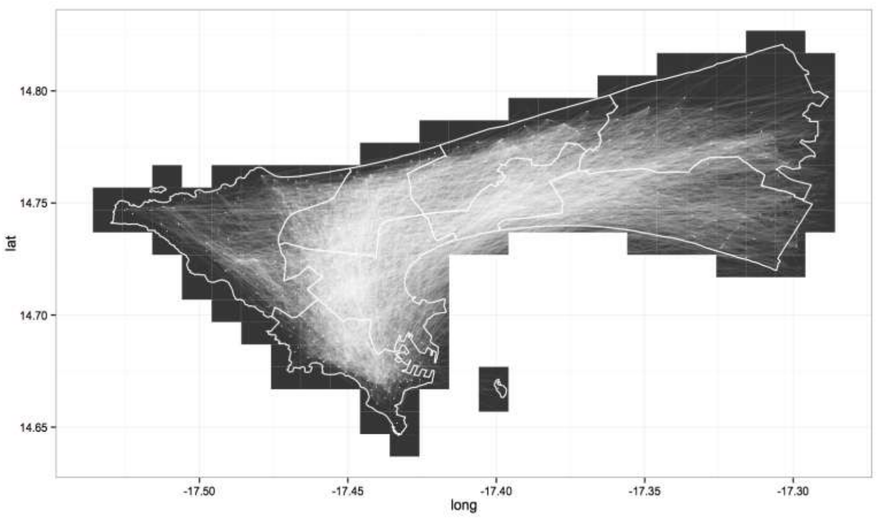
We use mobile phone data to assess how the new Dakar Diamniadio Toll Highway is changing human mobility in the Dakar metropolitan area. We find that although the new toll road increases human mobility on average, its impact differs by location. While some suburban areas (such as Les Parcelles Assainies) experience significant mobility from the new infrastructure, other areas (such as the downtown) experience decreased mobility.

Possible use for development:

Assessing how the impact of the new toll highway differs by area and how it changes over time can help policymakers benchmark the performance of their investment and better plan the development of urban areas.



Dakar Diamniadio Toll Highway (in red) and predicted travel time reductions in blue.



Mobility of individuals in Dakar, Senegal

Main results:

- Human mobility in the metropolitan Dakar area increased on average by 1.34% after the opening of the Dakar Diamniadio Toll Highway. However, this increase masks important disparities across the different sub-areas of the Dakar metropolitan areas.

- In particular, the Parcelles Assaines sub-area benefited the most for the toll road with an increase in mobility of 26 percent. The Centre Ville (downtown) area experienced a decrease in mobility of about 20 percent.

Methods:

- Using mobile phone usage data, we assume that the average cell tower location during daytime is a good proxy for the location of where people work, while the location in early mornings and evenings is a good indicator for where there live. We assess changes in human mobility by tracking the flow of people between their place of residence and their place of work before and after the opening of the new toll road.

- We simulate the reduction in the distance or time to travel between the place on non-economic activity and the place of economic activity, following the introduction of the toll road. We then run a flexible linear regression to estimate how the relationship between mobility and time savings changes over time. We also estimate how mobility varies across different geographic areas.

Full paper is here:
http://bit.ly/173WqN9
DataViz or video are here:

Login:
Pw:

Data sources used for this project:
- D4D data set 1, com between antenna ☐
- D4D data set 2, movement routes high res ☐
- D4D data set 3, movement routes low res ☐
- D4D synthetic data set ☐

Other data sets used in this project:
- Type of data: travel time Source: Google Maps

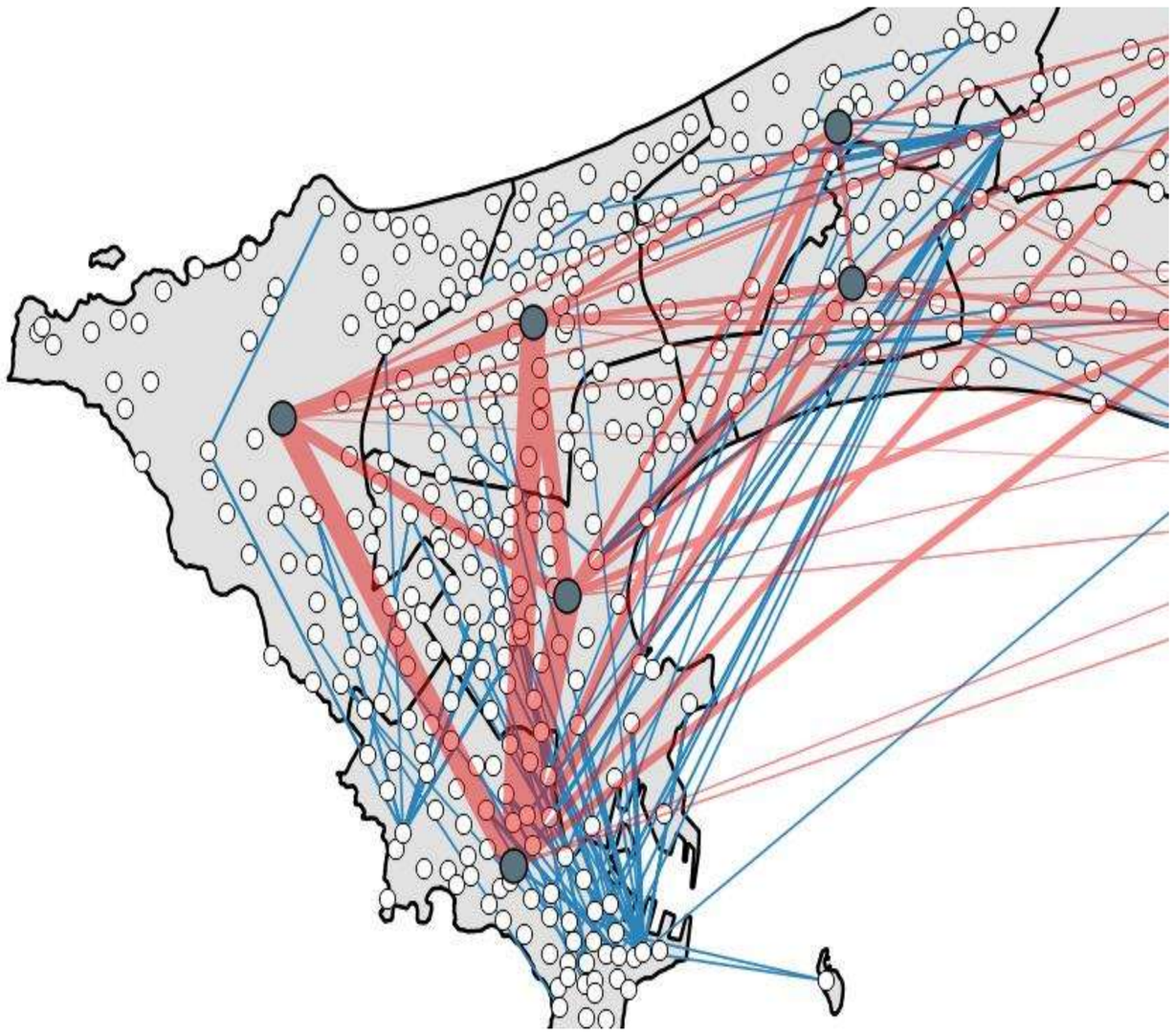
Main Tools used:
- R
- QGIS
- Python
- Open Street Maps

Open Code available:
- Yes ☐
- No ☐

T06

Travel demand analysis with differentially private releases

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



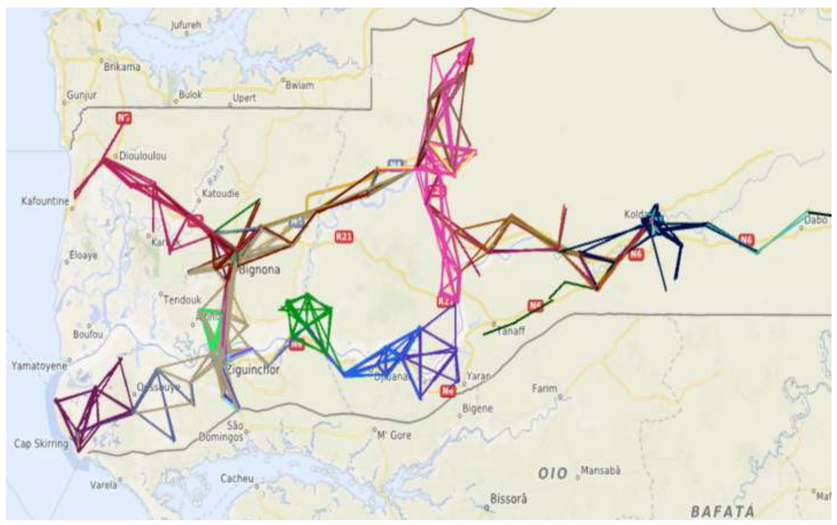
- David Gundlegård, Clas Rydergren, and Jaume Barcelo, Linköping university, Sweden
- Nima Dokoohaki, Olof Görnerup, and Andrea Hess, SICS Swedish ICT, Sweden

Project Summary:

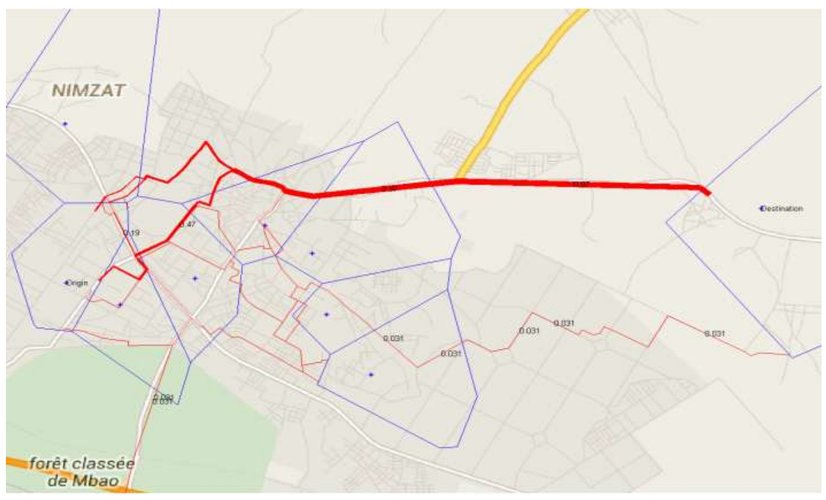
The use of mobile phone data for planning of transport infrastructure has been shown to have great potential in providing a means of analyzing the efficiency of a transportation system.

We describe how this type of data can be used in order to act as both enablers for traditional transportation analysis models, and provide new ways of estimating travel behavior. The raw data is aggregated with the goal to retain relevant information while discarding sensitive user specifics.

The approach presented for the estimation of travel demand and route choices, and the additional privacy analysis, makes a comprehensive framework usable in the processing of mobile phone data for transportation planning.



Extracted frequent travel sequences.



Travel demand in one origin destination relation assigned to the road network.

Main results:

- An approach for the estimation of travel demand and route choices, and privacy analysis of the aggregated data, which makes a comprehensive framework usable in the processing of mobile phone data for transportation planning.
  - Based on users activities, home and work place, the travel demand is described in time sliced origin destination matrices.
  - Frequent sequences for travelling are used for estimating route choices. The computed demand is assigned to the estimated routes, resulting in an estimation of the network load.
  - The information contained in the frequent travel sequences and the demand matrices are analyzed from a privacy perspective, with the aim of identifying privacy preserving data for use in transport planning.

Methods:

- The methods for processing and aggregation of the data into useful descriptions for transport planning is developed in the paper are:
  - Frequent sequence extraction
  - Site sequence clustering
  - Resolution adaptive demand matrix construction
  - Reviews and experiments with differentially private frameworks for estimating budgets allowed on data queries.

Full paper is here:
http://transportanalyticslab.se/d4d.html
DataViz or video are here (Initial viz of non-aggregated data):
http://n7.se/t/
Login: d4d
Pw: d4d

Data sources used for this project:

- D4D data set 1, com between antenna ☐
- D4D data set 2, movement routes high res ☒
- D4D data set 3, movement routes low res ☐
- D4D synthetic data set ☐

Main Tools used:

- PostgreSQL and OSM
- Seqwog and MG-FSM

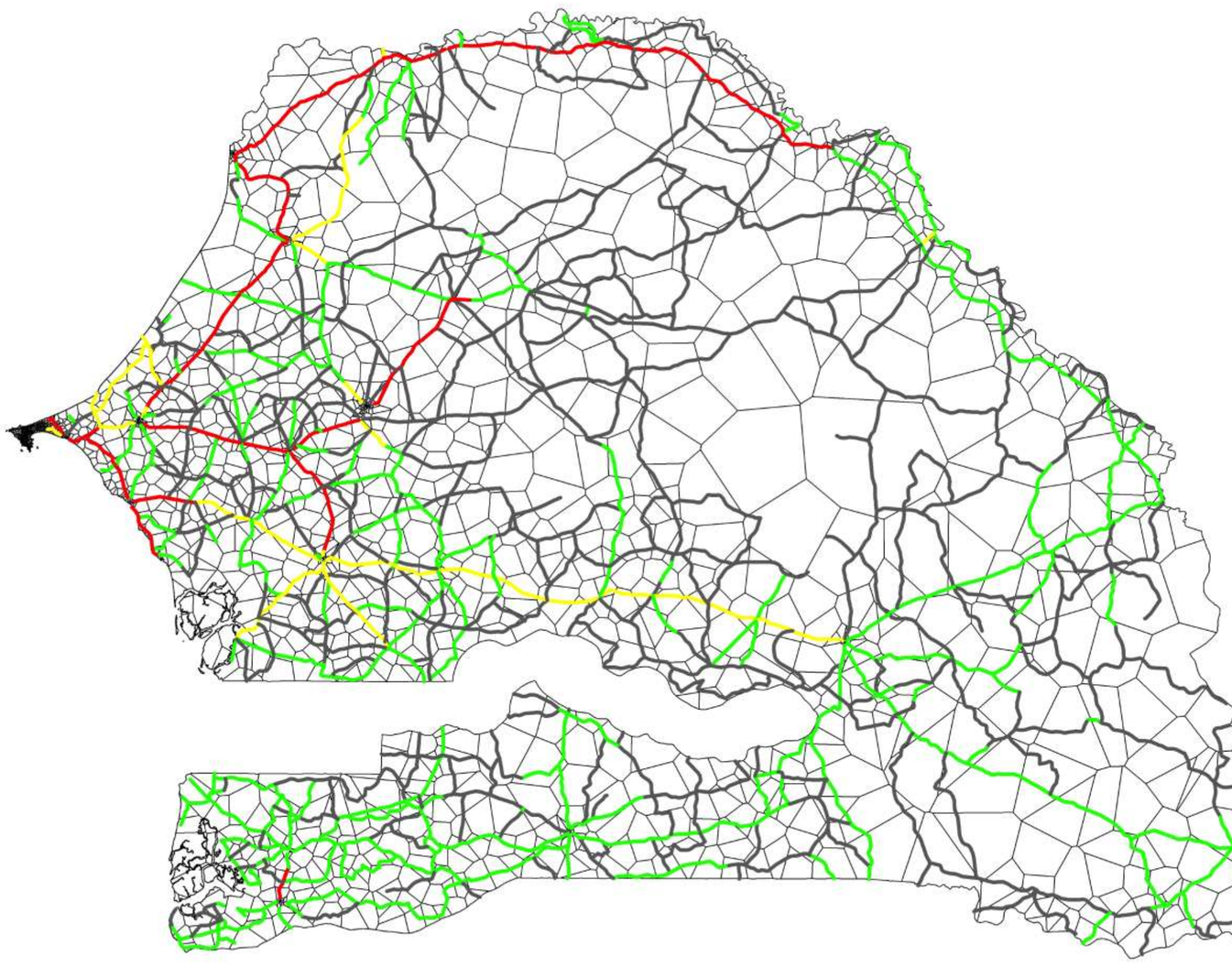
Open Code available (some):

- Yes ☒
- No ☐

T07

Cars and Calls: Using CDR Data to Approximate Official Traffic Counts

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



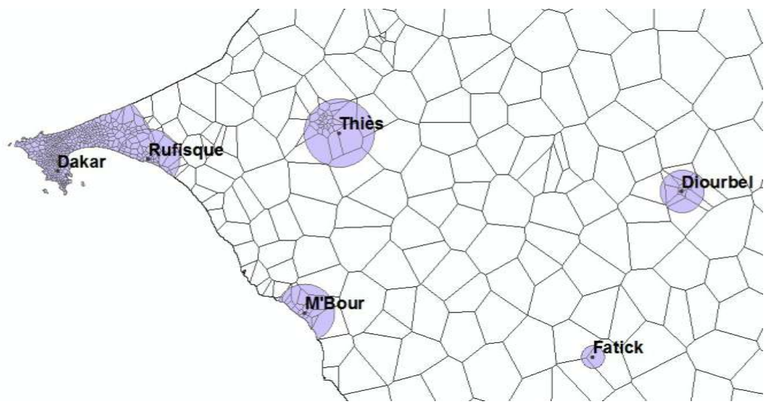
- Frias-Martinez, Vanessa, Assistant Professor, iSchool, University of Maryland
- Liang, Tony, Student, Computer Science Department, University of Maryland

Project Summary:

Official traffic counts approximate the amount of traffic observed in roads. These counts are computed by local authorities to understand transportation needs. However, automatic traffic collection techniques tend to be highly expensive. We propose the use of cell phone data as a proxy for traffic count estimation. Specifically, our approach automatically computes official traffic counts using mobility features extracted from Call Detail Records.

Possible use for development:

Countries with limited economic resources tend to compute traffic counts with manual and inaccurate techniques. Our approach provides a reliable technique to measure traffic counts at large-scale and in affordable manner.



I. Filter Regions Approach: Eliminating urban-type coverage areas

II. Regression Results: Filter Users Approach

Methods:

- We propose to use cellular activity as a proxy to predict the traffic counts in roads. However, the cellular activity might be due to individuals who are walking or driving by. We explore two filtering techniques to disentangle motorized from non-motorized traffic:
  - *Filter Regions*: consider only the cellular activity at towers that give coverage to geographical areas that are mostly inhabited (see Figure I)
  - *Filter Users*: consider only the cellular activity of users that we can tell for for sure are driving from their CDR traces' speeds.
- We analyze two predictive models:
  - *Linear Regressions: fit with Ordinary Least Squares.*
  - *Support Vector Regressions*: with a Radial Basis Function kernel, to explore non-linear regression approaches

Main Results:

- We combine the two predictive models with the filtering approaches and report prediction accuracy (as correlation btw. real and predicted):
  - SVR performs better than Linear Regressions for any type of filters: **a non-linear approach provides a better fit for the official traffic counts that we want to approximate (see Figure II).**
  - Filter Users shows better predictive results indicating that filtering users based on speed is better than filtering out urban-like areas.
  - Best results: r=0..698 for all roads and r=0.889 for only national roads

Full paper is here:

www.urbancomputinglab.org/d4dchallenge

Data sources used for this project:

- D4D data set 1, com between antenna
- D4D data set 2, movement routes high res
- D4D data set 3, movement routes low res
- D4D synthetic data set

Other data sets used in this project:

- Official Traffic Counts for Senegal Roads
- Source::http://www.infrastructureafrica.org/documents/type/arcgis-shape-files/senegal

Main Tools used:

- Scripting languages
- Google Eartn
- R

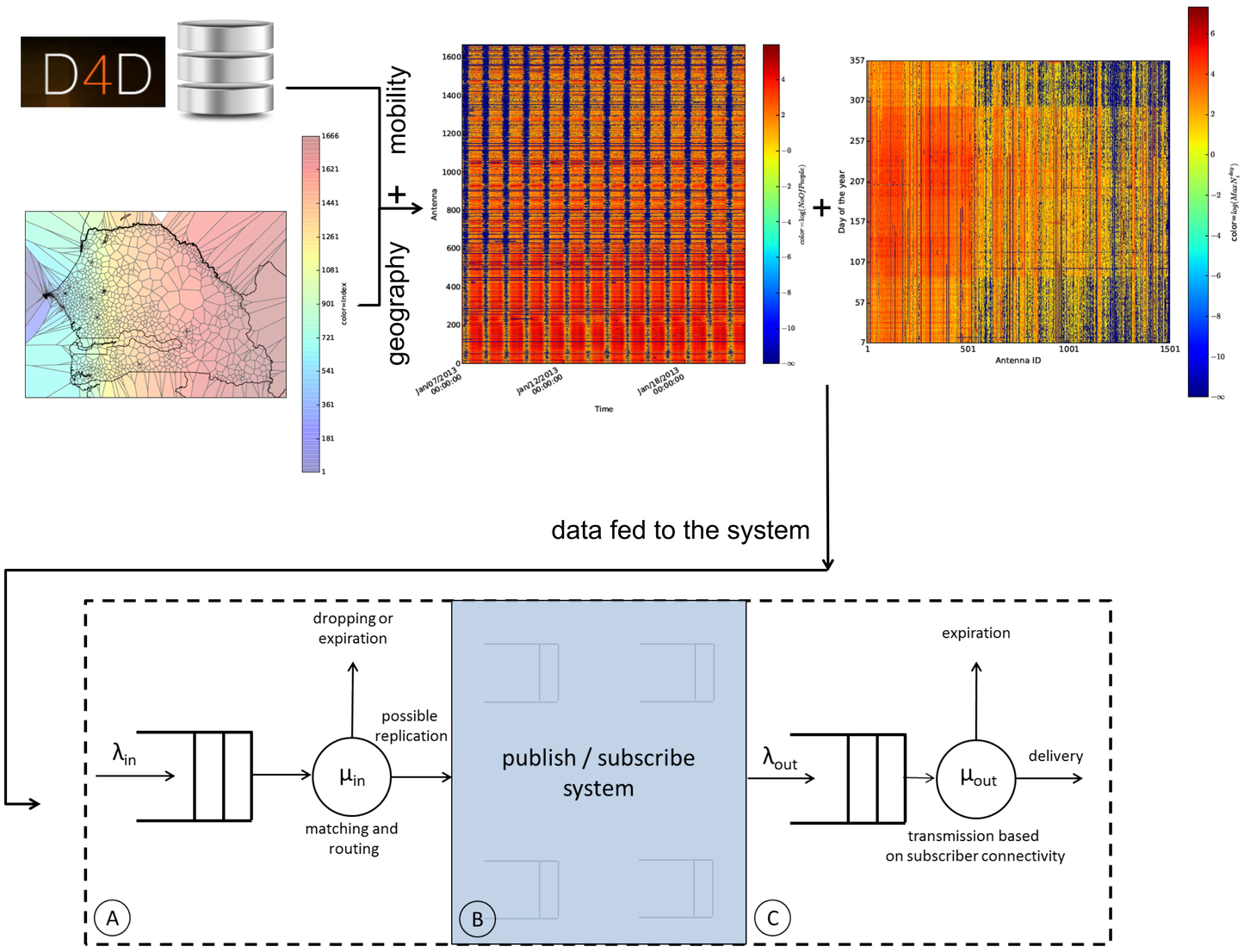
Open Code available:

- Yes
- No

T08

Towards Mobile Social Crowd-Sensing for Transport Info. Management

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



- Georgantas, Nikoloas, Sr. Res. Scientist (lead Author)
- Bouloukakis, Georgios, Phd. Student
- Pathak, Animesh, Res. Scientist
- Issarny, Valerie, Res. Director
- Agarwal, Rachit, Post Doc

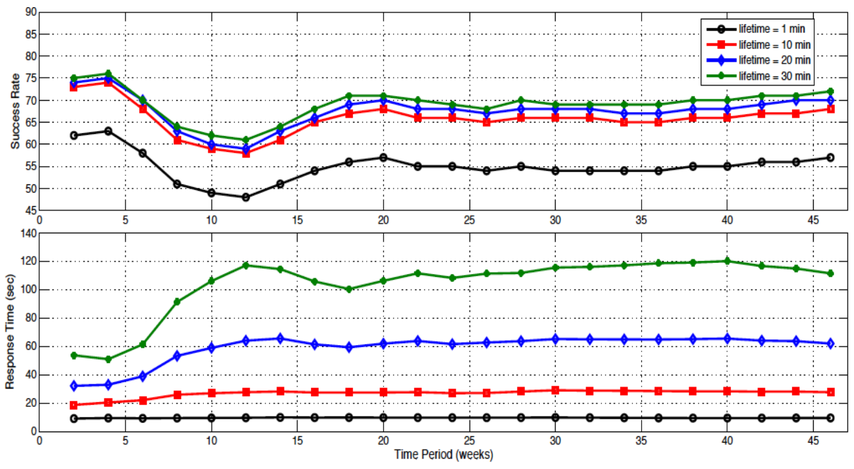
Project Summary:

Transport in Senegal is chaotic and large, especially in main cities. Additionally, although most people have mobile phones, large part them still rely on SMS. Considering this, we propose the development of an application platform for large-scale transport information management relying on 'mobile social crowd-sensing'.

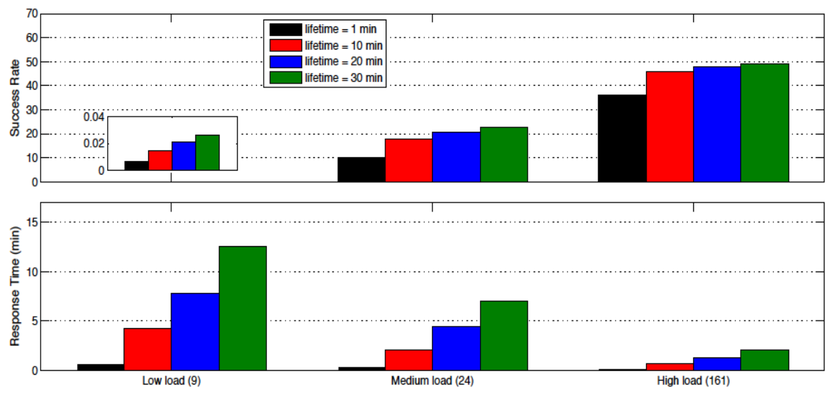
To support this platform, we model a large-scale mobile publish/subscribe system using queuing theory. We developed the MobileJINQS simulator with the realistic load for the analysis.

Possible use for development:

The project provides telecom providers inputs to better tune the communication backbone, and application developers a platform for transportation application.



Success rates and response times for network traffic from low load Antenna 9 to high load Antenna 161 with varying message lifetime periods



Mean end-to-end transaction success rates and response times

Main results:

- High load observed in antennas near Dakar.
- Varying incoming loads and service delays has a significant effect on response time.
- Success rate and response time are directly proportional to Message lifetime with proportionality constant greater for response time.
- Response time is dependent on subscribers behavior.

By properly setting event lifetime spans, system or application designers can best deal with the tradeoff between freshness of information and information delivery success rates. Still, both of these properties are highly dependent on the dynamic correlation of the event input flow and delivery flow processes, which are intrinsically decoupled.

Method:

- Let N^t_i be the number of people in an antenna i at a given time t over the period of the trace (50 weeks)
  - Let λ_{in} be the input process at the input access point associated to the antenna i, then λ_{in} is a non-homogeneous Poisson process with rate $\lambda(t)=N^t_i/|t|$. Similary μ_{out} is a non-homogeneous Poisson process with rate $\mu(t)=N^t_j/|t|$ at the output access point associated to the antenna j.
  - μ_{out} is equivalent to service time that follows an exponential distribution with mean equal to $1/\mu(t)$.

Full paper is here:
http://xsb.inria.fr/docs/d4d2015.pdf

DataViz or video are here:
http://xsb.inria.fr/d4d#visualization

Data sources used for this project:
- D4D data set 1, communication between antennas [ ]
- D4D data set 2, high resolution movement routes [x]
- D4D data set 3, low resolution movement routes [ ]
- D4D synthetic data set [ ]

Other data sets used in this project:
- None

Main Tools used:
- XSB
- MobileJINQS
- Queueing Theory
- Python mpl_toolkit

Open Code available:
- Yes [x]
- No [ ]

T09

Operational Planning of Transport & Infrastructure Using Mobile Data Analysis

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network

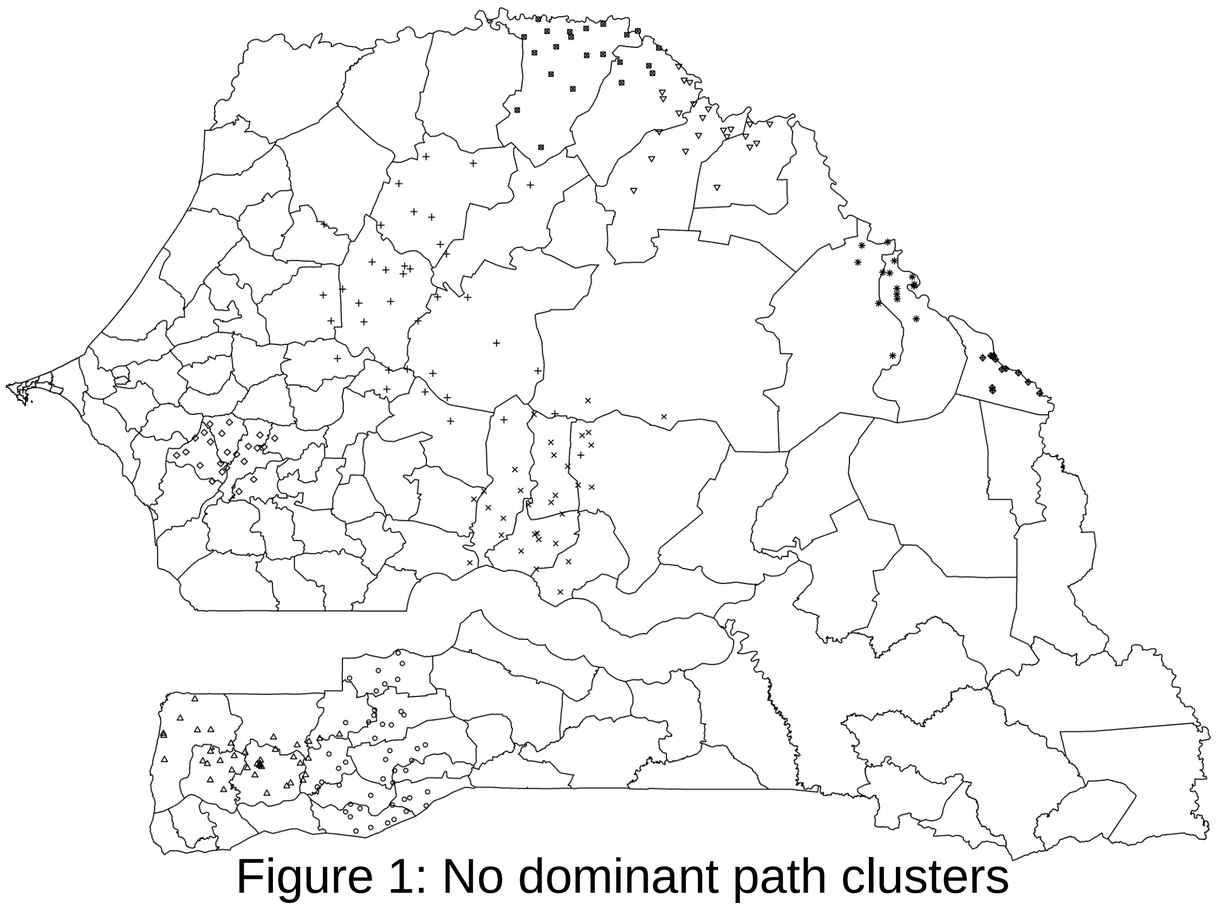


Figure 1: No dominant path clusters

- Singla, Ankit, G1/PJ-DM, RBEI
- Agrawal, Shivendra, G1/PJ-DM, RBEI
- Aatre, Kiran R, G1/PJ-DM, RBEI
- Iyer, Parameshwaran S, G1/PJ-DM, RBEI

Project Summary:

In the following work we present a short description of data mining methods & techniques applied to creating an infrastructure development and planning methodology using the data cleaved from cell phone usage patterns. The tower-to-tower communication data was utilized to derive clusters in which people commute across Senegal. In these clusters dominant user trails were first indentified. Time variant statistics on these trails across the year, highlighting congestion, and inferences for one communication cluster are presented in detail.

Possible use for development:

The findings showcase that with further granularity in the data along with the GPS information will enable the development of infrastructure by determining the high traffic regions. The information may be used to determine the optimal location of emergency services like fire trucks, ambulances etc.

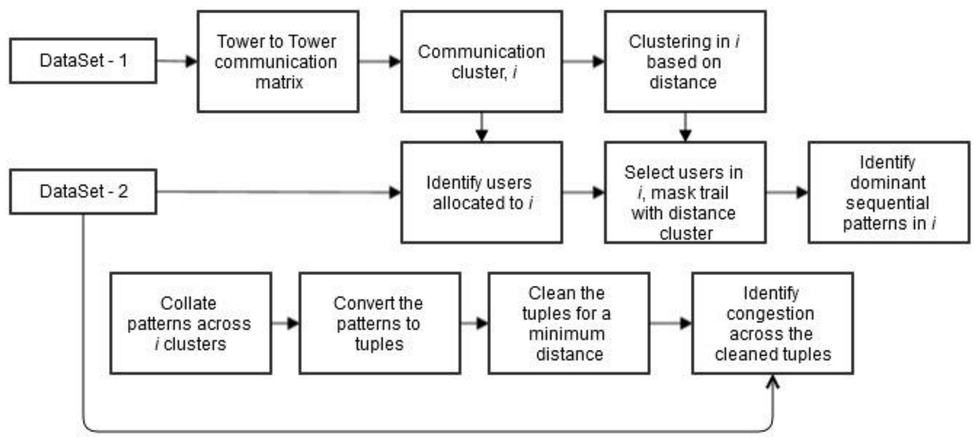


Figure 2: Architecture of OPTIMA

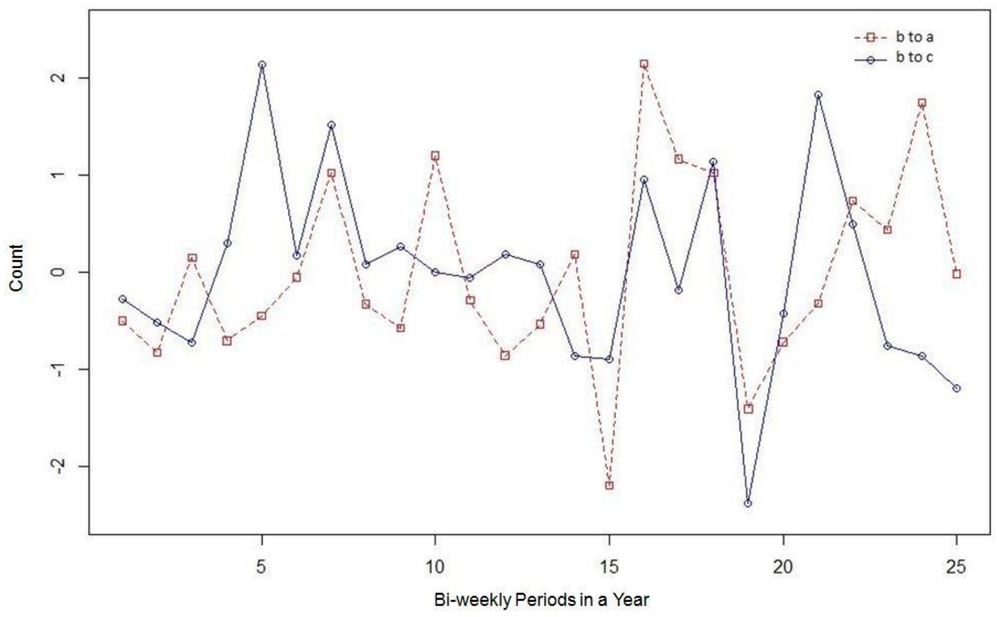


Figure 3: Scaled volume flow on paths

Main results:

- A total of 62 communication clusters using a graph clustering algorithm and 60 tuple paths were isolated.
- In 9 communication clusters, no dominant trails or tuple paths were observed. We present these 9 communication clusters on the Senegal map in Figure 1.
- 27 tuple paths were observed within clusters and the rest across clusters. The dominant paths covered about 23 communication clusters (43%); indicating a good coverage of the rural areas in Senegal.
- Figure 2 depicts the working of our methodology for the Fatick region in Senegal.
- The bi-weekly time series plot of total users travelling (scaled) from b to a (b-a) and b to c (b-c) in Figure 3. shows traffic volumes from b-a are significantly high compared to b-c throughout the year. Further in Figure 6b we also find that the instantaneous point wise correlation between both the series are also high.
- The median velocity observed on b-a & b-c over the biweekly periods of the year, infrastructure can be improved over b-c to decrease the travel time.
- Improvements on trail path b-a are also warranted.

Methods:

- A communication matrix for each pair of tower over the year created.)
- MCL clustering algorithm used to infer 62 clusters on the basis of communication b/w towers.
- Hierarchical clustering within these 62 clusters, we cluster the towers within 5km radius and treat them as same identity.
- Location info of ~3L*25 = ~79L users, a trail of each and every user was generated and assigned to a cluster depending upon his maximum presence.
- Sequential patterns in each of these 62 clusters identified.

For access to the full paper, please contact the authors:

Ankit.Singla@in.bosch.com
KiranRangaswamy.Aatre@in.bosch.com

Data sources used for this project:

- D4D data set 1, communication between antenne ☑
- D4D data set 2, movement routes (high resolution) ☑
- D4D data set 3, movement routes (low resolution) x
- D4D synthetic data set x

Main Tools used:

- SPMF
- R, Python
- MCL, h-clustering
- Sequential Pattern Mining

Open Code available:

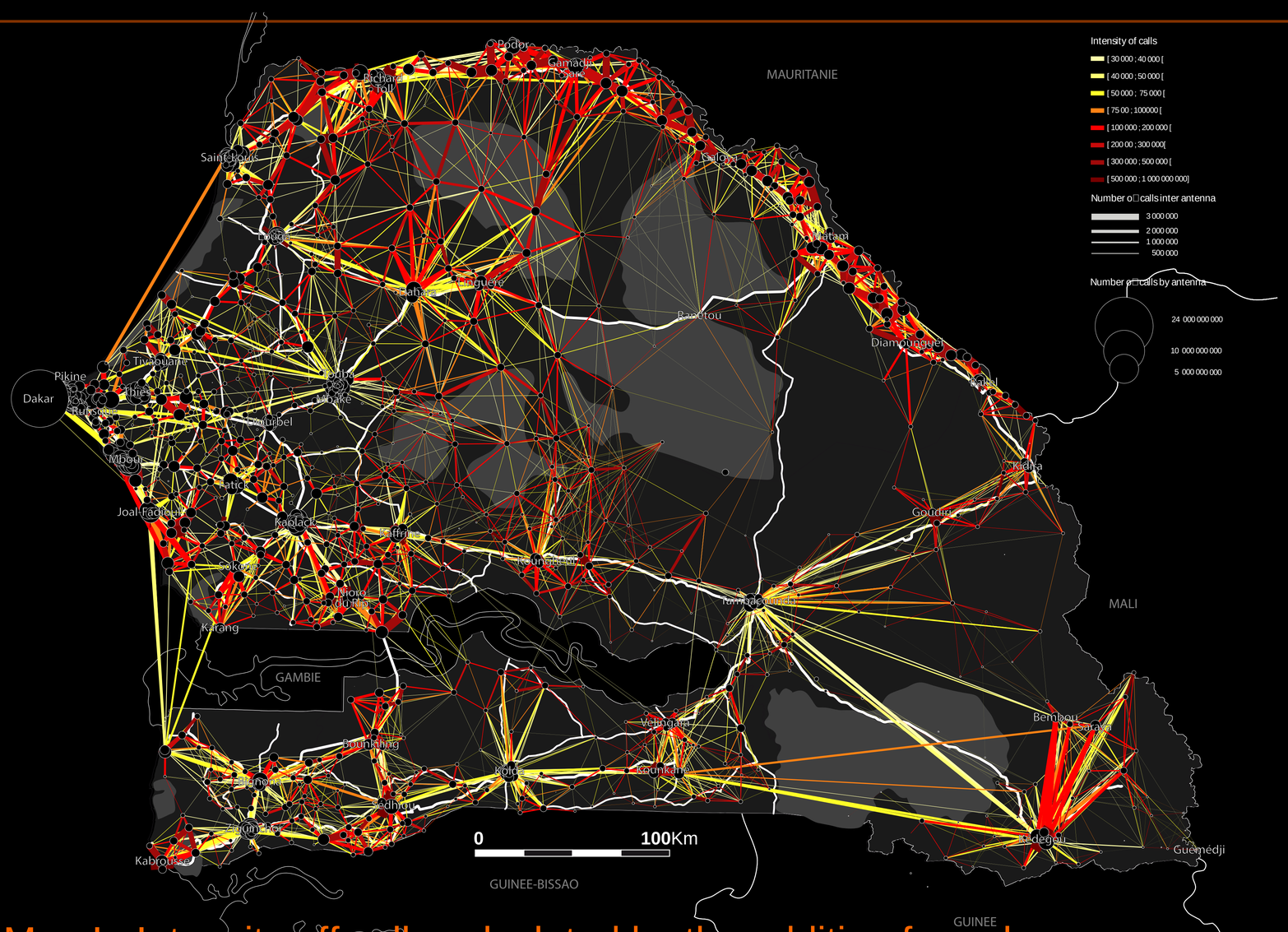
- Yes ☑
- No ☐

T10

Mobile Data For Land Management

D4D challenge data for development

Health	Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



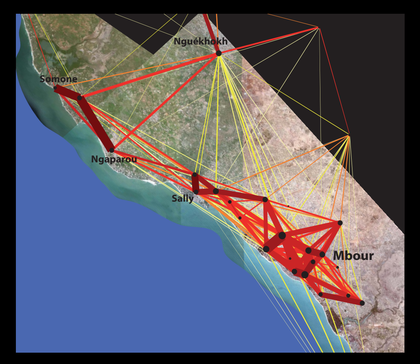
Map I : Intensity off calls calculated by the addition formula

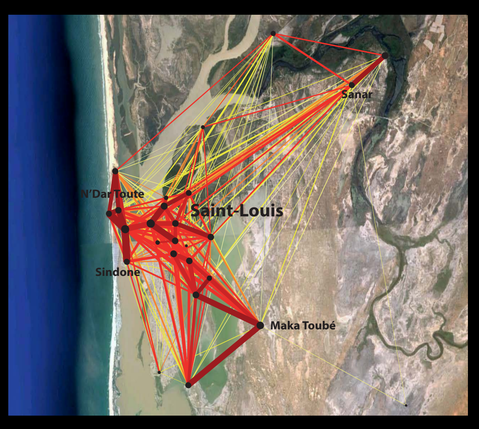
Project summary:

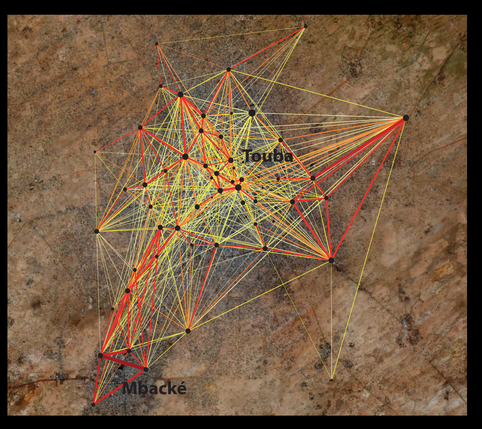
Working on phone calls and their intensities, it is possible to show the network of cities of Senegal and therefore allow a diagnosis of the situation of the urban system (how it works?). Should result a series of recommendations for land planning, public policy development, development of future infrastructures and thus tender to a sustainable land use.

Possible use for development:

The project aims to develop new public policies and provide a diagnosis for the development of a sustainable land use. The hierarchy of cities provides guidance on the planning of networks (roads, electricity, water supply, …). It shows the imbalances between the territories (Dakar and the rest of the country) and give keys to correct it.



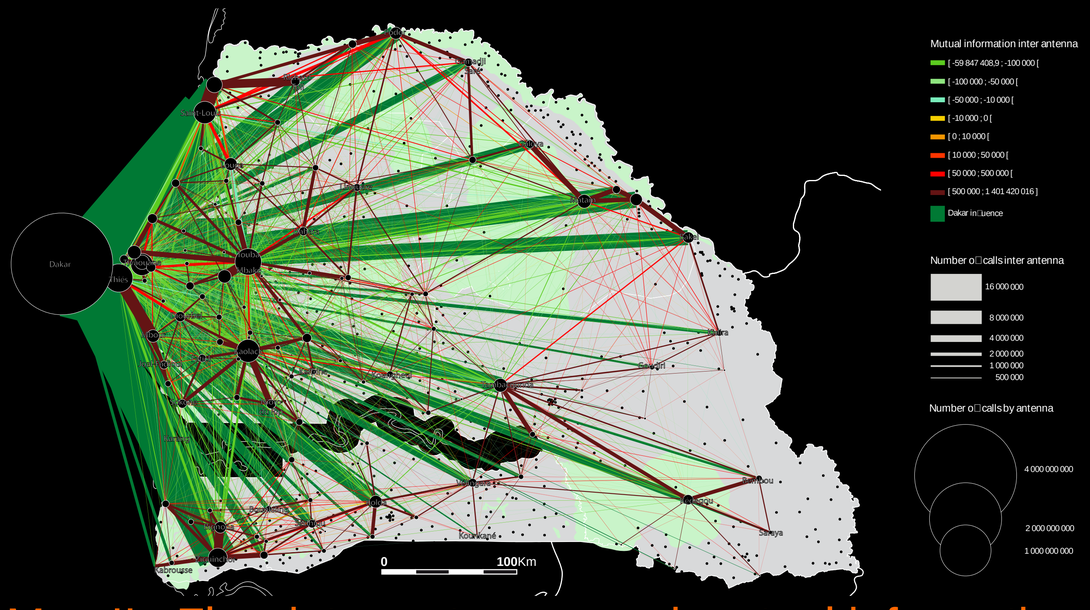




Ecole Politechnique Fédérale de Lausane (EPFL)
www.epfl.ch/

Groupe Huit
www.groupehuit.com

Prédiccis
www.predicsis.com/



Map II : Flow inter-antennas and mutual information

Main results:

1. The calls and their intensities clearly show how the urban system works, they also show the economic role of the different regions, cities or municipalities.

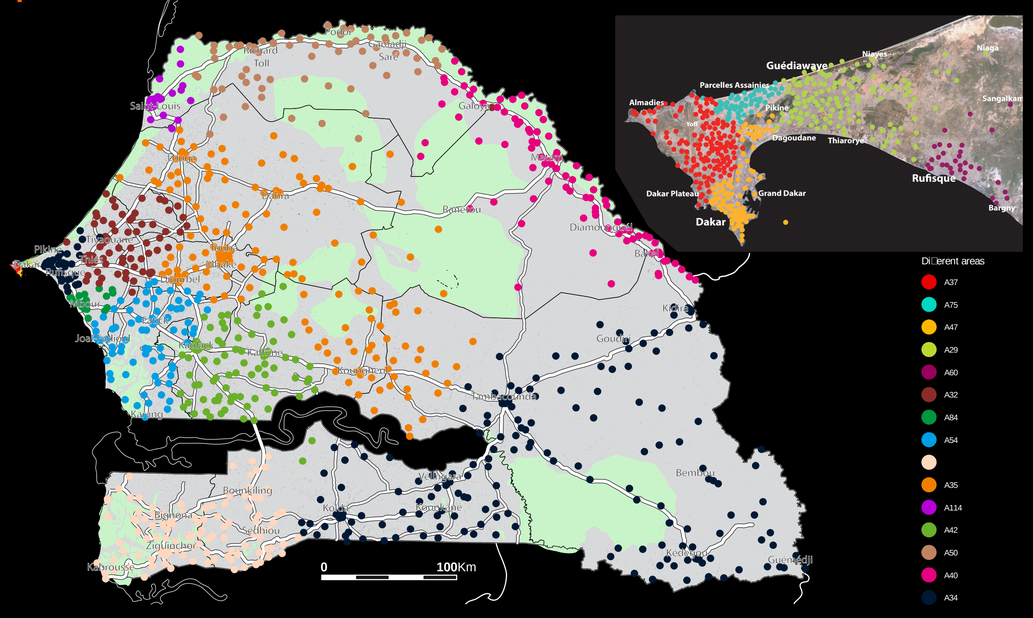
2. Dakar unbalance the network, in terms of land use that would argue for a greater balance between Dakar and the "desert" of Senegal.

3. The important role of natural barriers, like the Gambia, but also the areas of "fossil valleys".

4. Calls are close to relatives, from one city to another with the exception of Dakar.

La formule est :
Inf Mutuelle = Pij *Ln(Pij / (Pi. * P.j) * total

Pij correspond à la probabilité du flux i vers j donc flux i vers j divisé par le nombre de flux total.
Pi. correpsond à la probabilité du flux i vers X donc la somme des flux i vers X divisé par le nombre de flux total.
P.j pareil que Pi. mais dans l'autre sens.



Map III : Co-Clustering

Methods:

The three maps were produced with the same set of data, but by three different methods:

Map 1: Calls and intensity.
The thickness of the lines is the number of calls between antenna A and B. The colour corresponds to the intensity (intensity = (calls between A and B) / (total calls A + total calls B)) on the basis of 1666 antennas

Map 2: Mutual information
Map of mutual information is based on probability to receive calls using the following formula: Inf Mutual Pij = * Ln (P ij / (Pi Pj *) * Total. Pij is the probability of ux i to j ux, so ux i to j divided by the number of total ux. Pi. correspond for the probability of ux i to X therefore the sum of ux i to X divided by the number of total ux.

Map 3: Behaviour of antennas
The colours show the antennas, which are called as each other and have the same behaviour (cluster)

Full paper is here :
Link

Data sources used for this project:

- D4D data set 1, com between antenna [X]
- D4D data set 2, movement routes high res [ ]
- D4D data set 3, movement routes low res [ ]
- D4D synthetic data set [ ]

Other data sets used in this project:

- Type of data: Data Map — Source: INC Sénégal
- Type of data: — Source:
- Type of data: — Source:

Main Tools used:

- Tool 1 [ ]
- Tool 2 [ ]
- Algorithm [X]

Open Code available:

- Yes [ ]
- No [ ]

T11

Sanitization of Call Detail Records via Differentially-private Summaries

Health	Transport Urban	National Statistics	Privacy
Agriculture	Energy	DataViz	Network

- Source: http://www.howtogeek.com/189831/6-reasons-why-you-cant-move-your-cell-phone-to-any-carrier-you-want/

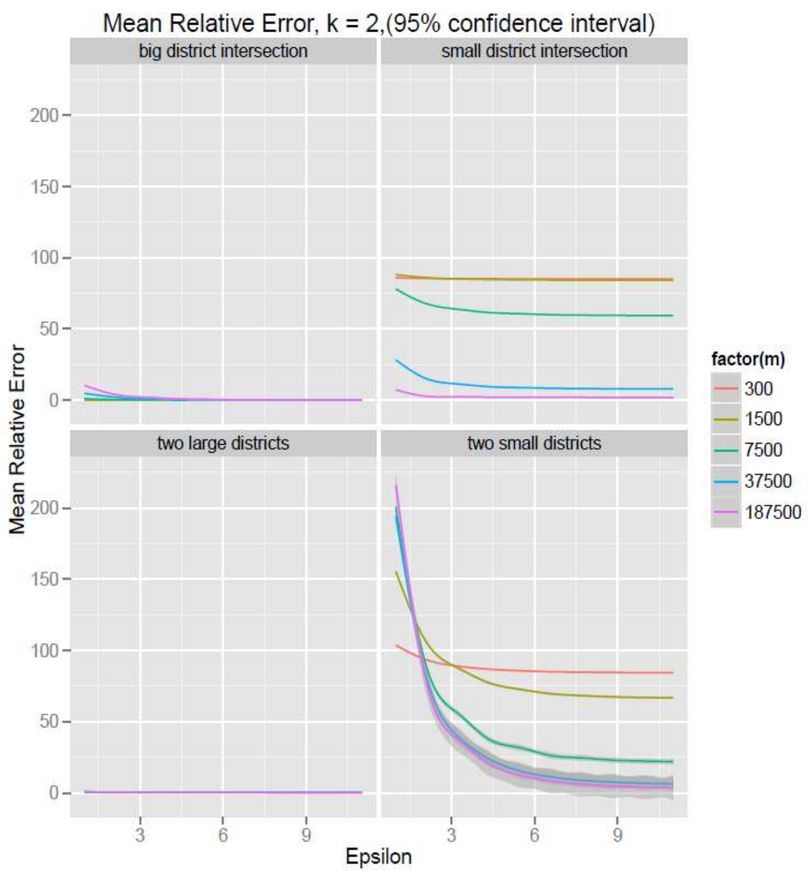
- Alaggan, Mohammad, Helwan University, Egypt.
- Gambs, Sébastien, Université de Rennes 1 – Inria, France.
- Matwin, Stan, Dalhousie University, Halifax, Canada.
- Tuhin, Mohammed, Postdoc, Dalhousie University, Canada. (lead author)

Project Summary:

We propose a novel, privacy-driven approach for estimating the number of users moving from one area to another within a given period, based on the analysis of call detail records (CDRs). Our method generates a summary of CDRs maintaining high utility while ensuring strong privacy guarantees. The proposed approach provides the ability to release global, high-value mobility data unlinkable to an individual, opening the way for using CDR data while fully addressing privacy concerns.

Possible use for development:

The study of human mobility, while respecting privacy, is important for evidence-based solutions to developmental issues. Our approach could be highly valuable for improvement in urban transportation or understanding the spread of diseases.



Mean relative error (MRE) of the estimation of the size of the intersection between two geographic areas for different levels of the privacy parameter ε (number of hash functions $k = 2$)

Main results:

- We propose a data sanitization method that learns a privacy-preserving summary out of CDRs. A summary can be used to identify global mobility patterns in the population while hiding the individual movements of users. We also explore the achievable trade-off between utility (measured in terms of the accuracy of the derived mobility patterns) and privacy for the sanitized location data.

Methods:

- The proposed method estimates the size of the intersection of two or more sets ***non-interactively*** as follows:
  - Each cellular tower publishes a differentially-private data structure summarizing information about users observed in that tower (*e.g.*, those who have generated a CDR record through a voice call or SMS). This is achieved by using the privacy-preserving Bloom filter (Alaggan, Gambs and Kermarrec 12).
  - After releasing the differentially-private summary, a tower can erase from its memory all the information recorded during the given time period.
  - Using probabilistic inference methods, the private summaries of two towers can be combined to estimate the number of users they have in common without being able to retrieve their precise identities.
  - Similarly, the size of the intersection of three towers can be estimated by training a random forest classifier.

Full paper is here:
Full paper submitted to D4D committee, we choose not to put online for the moment until we receive the approvam from the D4D committee
Additional Author: Erico N de Souza

Data sources used for this project:
- D4D data set 1, com between antenna [X]
- D4D data set 2, movement routes high res [ ]
- D4D data set 3, movement routes low res [X]
- D4D synthetic data set [ ]

Other data sets used in this project:
- Type of data: location data of mosques in cities of Senegal
  Source:http://www.islamicfinder.org/prayerDetail.php?city=Dakar&state=01&country=senegal&lang=

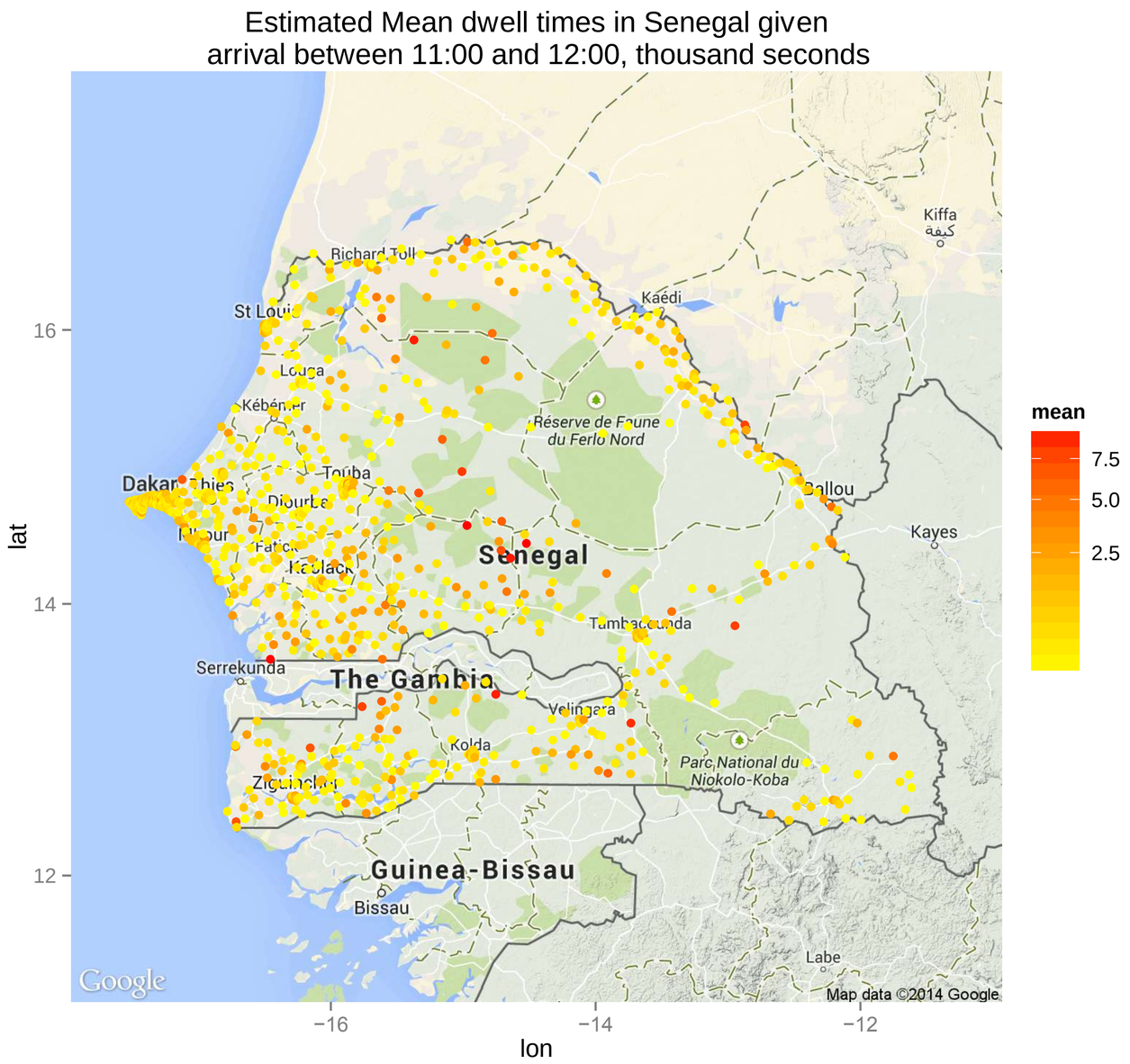
Main Tools used:
- Bloom filter
- Differential privacy
- Random forests

Open Code available:
- Yes [ ]
- No [X]

T12

A Cell Dwell Time Model for Synthetic Population Generation from Call Detail Records

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



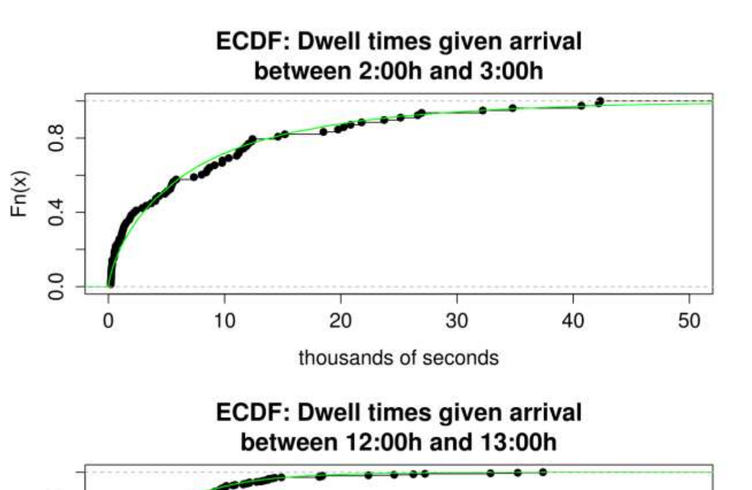
- Derrmann, Thierry, Dipl-Inf, University of Luxembourg
- Frank, Raphaël, Dr., University of Luxembourg
- Melakessou, Foued, Dr., University of Luxembourg
- Castignani, German, Dr., University of Luxembourg
- Engel, Thomas, Prof. Dr., University of Luxembourg

Project Summary:

In this work we propose a novel Cell Dwell Time Model that can be used to generate a synthetic population. The resulting dwell time model can be used to compute accurate user trajectories even with partial information. This work represents a first step towards the generation of a synthetic population that can be used to perform a wide range of simulative studies to evaluate and optimize transportation networks.

Possible use for development:

The Cell Dwell Time Model can be combined with the imputed trajectories to create a synthetic mobility model of the Senegalese population for different applications. For example, policy makers can rely on such studies to improve the transportation network on a country-wide scale.



Main results:

Cell Dwell Time Model:

Enables the creation of **full mobility models** from the truncated and anonymized datasets.

- Allows estimating the amount of activity and movement between cell sites
- Function linking Sets 1 and 2 to the dwell times

A Weibull distribution has been used to fit the dwell times for each hour and cell site.

Methods:

- Dijkstra Shortest Road Network Path Routing
- KS-Test for the evaluation of different attractivity functions
- A Weibull distribution fit for each hour and site

Full paper is here:
http://goo.gl/PlQD3d

Data sources used for this project:
- D4D data set 1, com between antenna [X]
- D4D data set 2, movement routes high res [X]
- D4D data set 3, movement routes low res [ ]
- D4D synthetic data set [ ]

Other data sets used in this project:
- Type of data: OpenStreetMap
- Source: http://www.openstreetmap.org

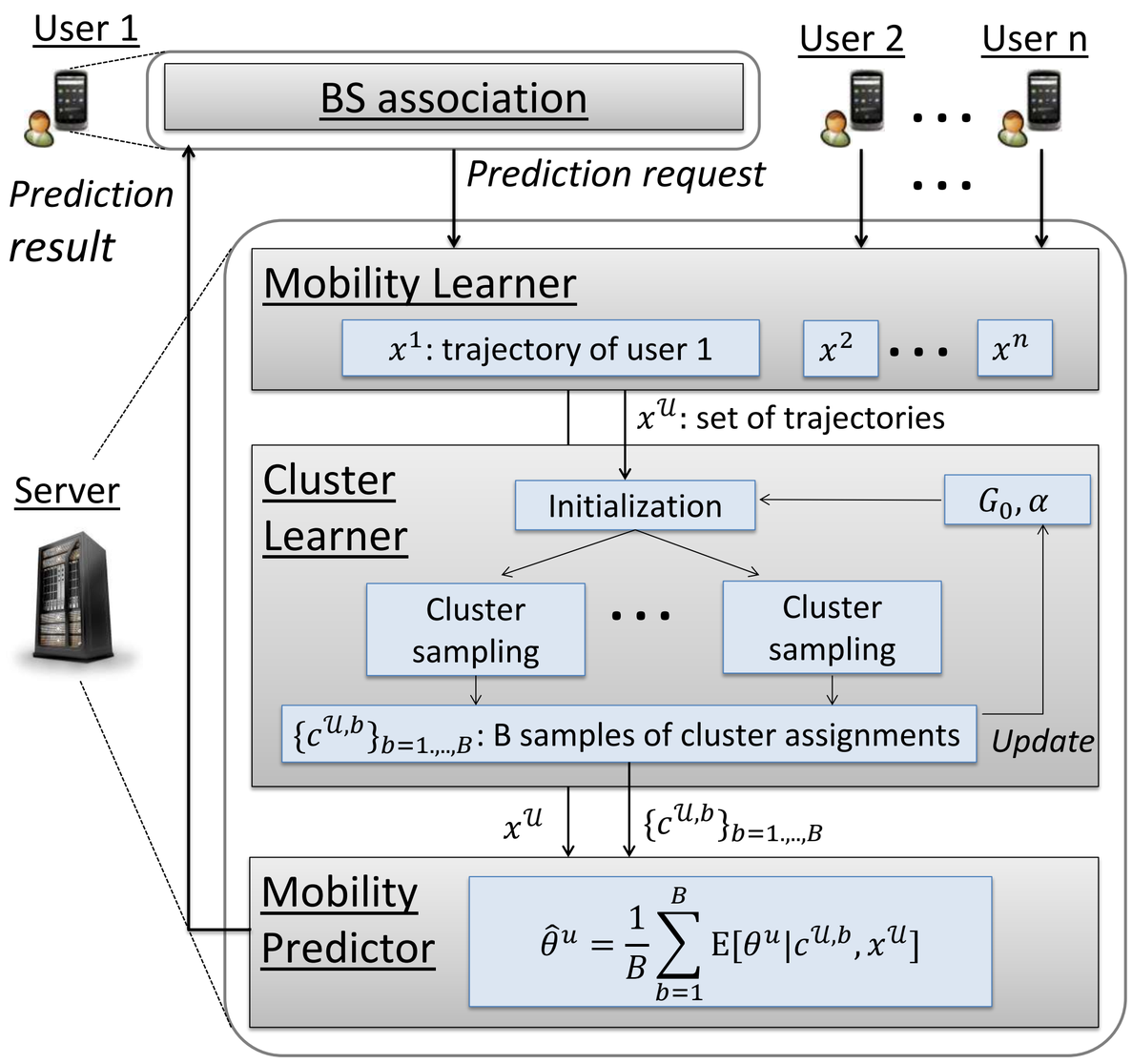
Main Tools used:
- PostGIS, pgRouting
- OpenStreetMap
- R, Python and PANDAS

Open Code available:
- Yes [ ]
- No [X]

T13

Improving Mobility Predictability based on User Clustering on D4D Dataset

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



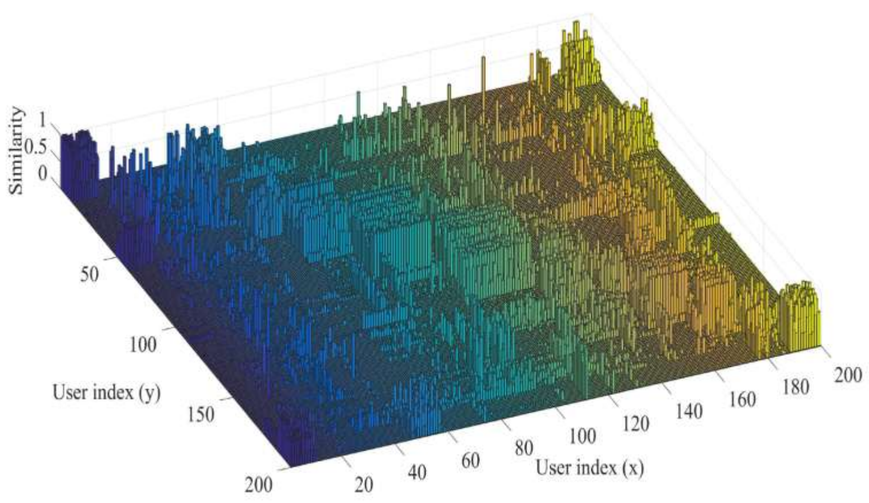
- Jaeseong Jeong, Post-doc, KTH, Sweden (lead author)
- Mathieu Leconte, Post-doc, KTH, Sweden
- Alexandre Proutiere, Associate professor, KTH, Sweden

Project Summary:

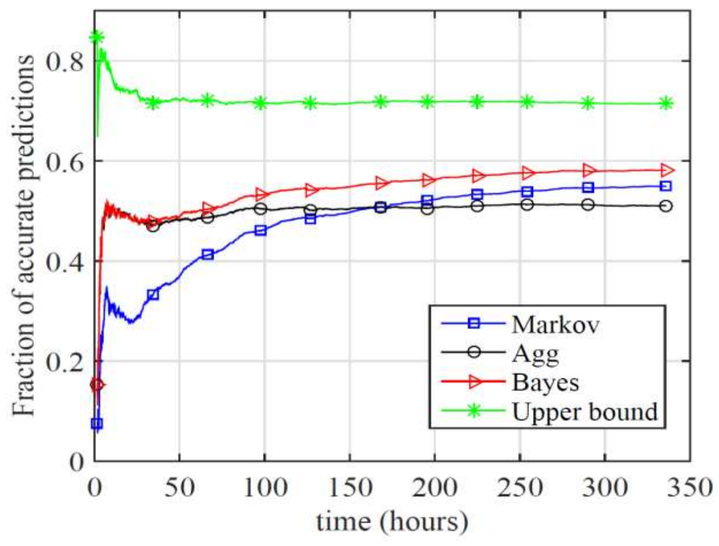
In practice, the mobility predictors can suffer from the lack of training data. Motivated by the empirical study on the large-scale D4D dataset, where the similarities between mobility patterns of users are exhibited, we develop an advanced mobility predictor which can improve predictability by replenishing its training data with the training data of other users selected by clustering.

Possible use for development:

1) Advanced location-based services
2) Urban traffic engineering, forecasting
3) Protocols in wireless mobile networks (scheduling, handover management, data prefetching)



Similarity between mobility patterns of 200 users



Predictability

Main results:

- The observation of similarities between mobility patterns of users
  - Definition of similarity between user u and v: The normalized predictability of user u achieved by the ideal predictor of user v
  - Many pairs of users are observed to have high similarity (almost one).
  - This motivates the predictor with less training data to merge with the training data of other similar users
- High prediction performance of the proposed predictor (Bayes)
  - Tested predictors
    - Bayes: the proposed predictor
    - Markov, Agg: the existing predictors
  - Bayes outperforms other predictors over all time span
  - During the first and second days at which the gathered training data is insufficient, Bayes outperforms Markov by up to 65%

Methods:

- Kernel ($\hat{\theta}^u$) estimation based on Dirichlet Process Mixture model
  - Cluster sampling : sampling the cluster assignment c^u from $p(c^u|x^u)$ based on Gibbs sampling algorithm
  - Using the multiple cluster samples, we can compute

$$\hat{\theta}^u = \int_{c^u} \mathrm{E}[\theta^u | c^u, x^u]\, dp(c^u|x^u) \cong \frac{1}{B}\sum_{b=1}^{B} \mathrm{E}[\theta^u | c^{u,b}, x^u]$$

- Prediction of the next location
  - Predict the most likely location in terms of $\hat{\theta}^u$

Full paper is here:
put your link here

DataViz or video are here:
put your link here
Login:
Pw:

Data sources used for this project:

- D4D data set 1, com between antenna ☐
- D4D data set 2, movement routes high res ☒
- D4D data set 3, movement routes low res ☐
- D4D synthetic data set ☐

Other data sets used in this project:

- Type of data: Source:
- Type of data: Source:
- Type of data: Source:

Main Tools used:

- Matlab
- Gibbs sampling algorithm

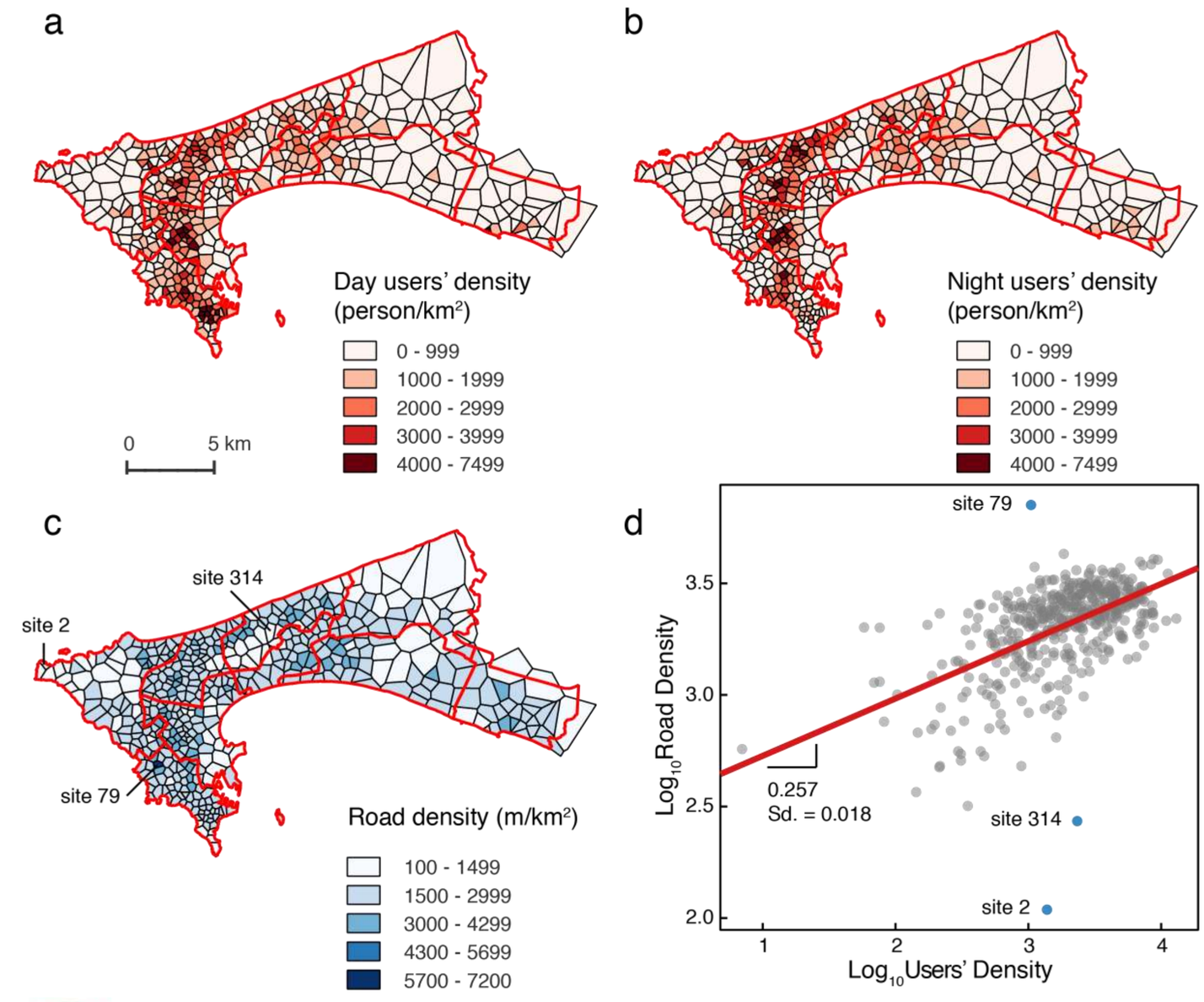
Open Code available:

- Yes ☐
- No ☒

T14

From digital footprints to the dynamic population distribution and road network efficiency

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



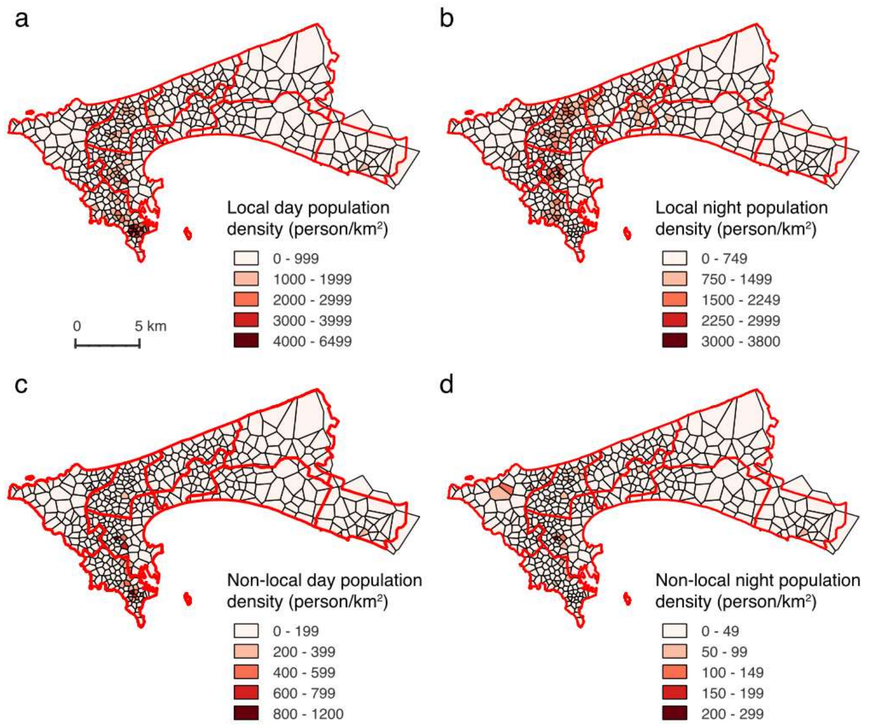
- Dong, Lei, Ph.D., School of Architecture, THU (lead author)
- Li, Ruiqi, Ph.D., SSS, BNU (co-lead author)
- Zhang, Jiang, A. P., School of System Sciences (SSS), BNU

Project Summary:

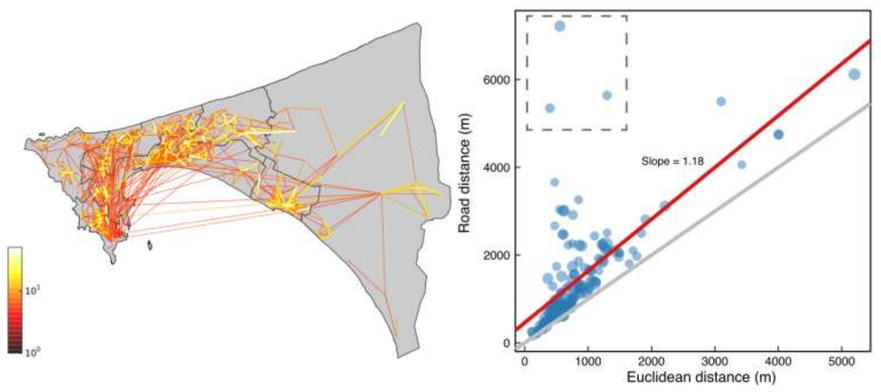
In this paper, we present a method to detect users' home and work places (i.e. OD) from mobile phone data, and we analysis the day and night, local and non-local population distribution in Dakar. These results reveal not only the dynamic process of human mobility, but also could be used to discover "hotspots" in cities. Combined with road data, we get the regression coefficient and un-matched region between population and road in Dakar. We also introduced a road network efficiency indicator, which provides a quantitative measure to guide future transportation infrastructure development.

Possible use for development:

The identification of hotspots can be used to guide construction, for example, better location based services, hotels. The unmatched regions and inefficient routes could be scientific basis for the government to make future planning improvement.



(a) Local-users day density, (b) Local-users night density, (c) Non-local users day density, (d) Non-local users night density



(a) OD pairs (b) Scaling relation between real route and Euclidean distance, with slope equals 1.18

Main results:

- Obtain day and night population distribution (ODs detection) by our method, and find the matching indicator and unmatched regions between road and population.
- Detect local and nonlocal users from the users' data, and show the different aggregation and OD mobility patterns. Nonlocal users' activities are more likely to be at specific areas, such as tourism attractions, airports and so on.
- We also use OD matrix and Google Map API to analysis the road network structural efficiency, which could be an important indicator for cities to evaluate their road network construction.

Methods:

- Integrating all the records of one users during two week, we count the locations he/she appears at night and day, and regard the place appeared with highest frequency as home and workplace.
  - The duration of day and night is determinted by the local usage habit, there are two peaks in phone call volume, which, we assume,correspond to work and entertainment, respectively. So after the second peak till next morning, we assume it's time for staying at home. The duration of day is the work hour without commuting time.
  - With Google map's API, we can get the travel time and distance between two locations.

Full paper is here:
http://www.idonglei.github.io/d4d
Video are here:
http://www.idonglei.github.io/d4d
Login: (able to access freely)

Data sources used for this project:
- D4D data set 1, com between antenna [√]
- D4D data set 2, movement routes high res [√]
- D4D data set 3, movement routes low res [X]
- D4D synthetic data set [X]

Other data sets used in this project:
- Type of data: road networks Source: http://metro.teczno.com/

http://www.openstreetmap.org/

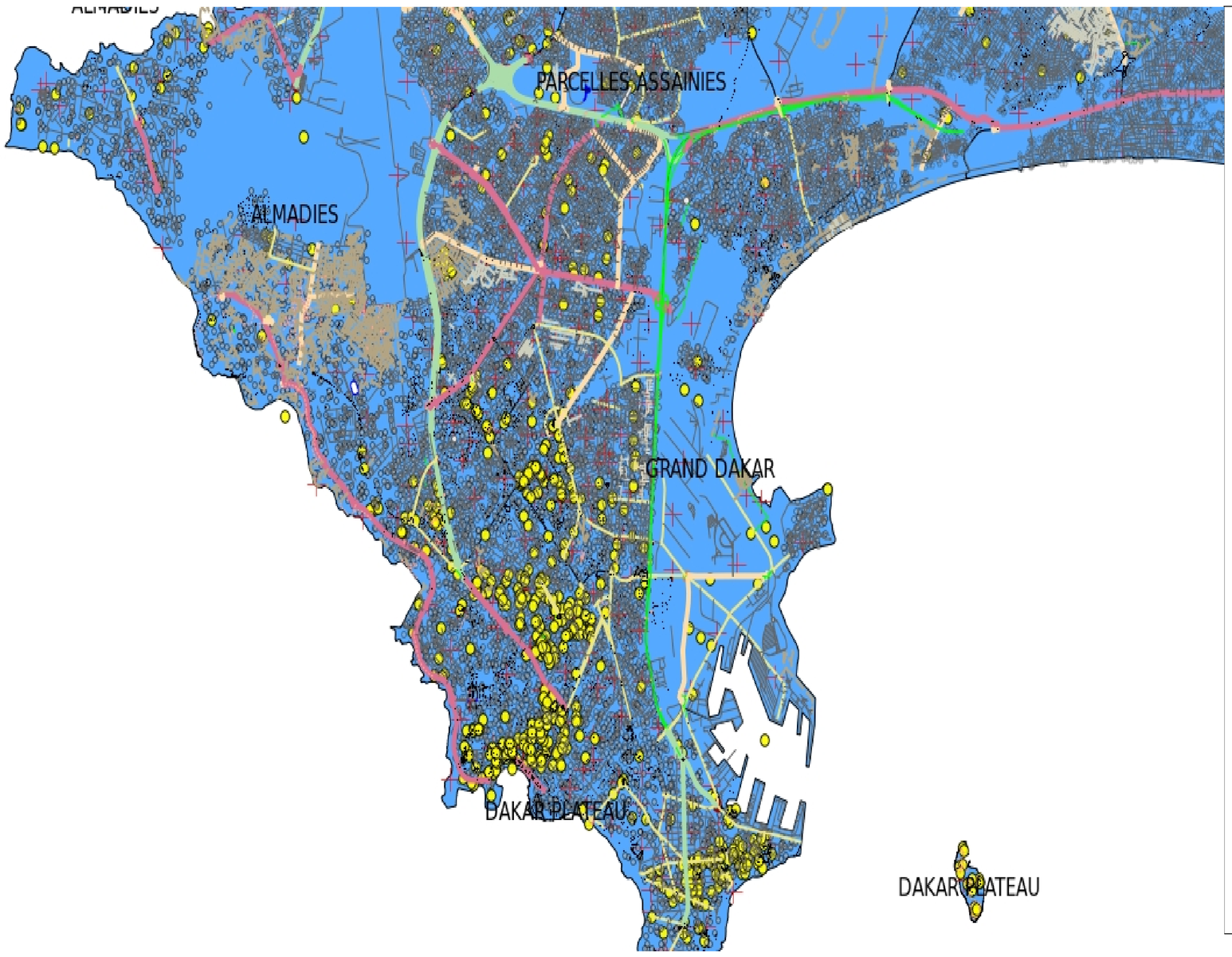
Main Tools used:
- Tool 1 Python
- Tool 2 GIS
- Algorithm: Fast detection of local/nonlocal users.

Open Code available:
- Yes [X]
- No

T15

Spatial Planning simulation and Optimization Technologies (SPOT)

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network

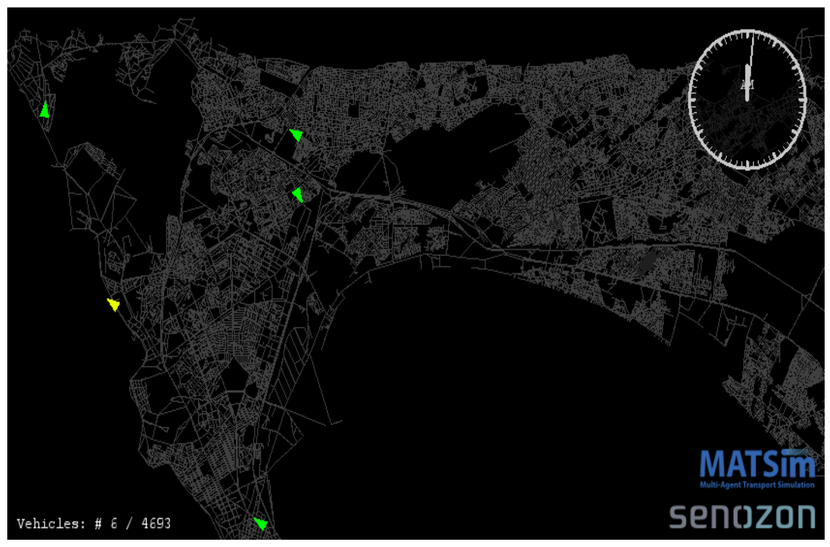


Project Summary:

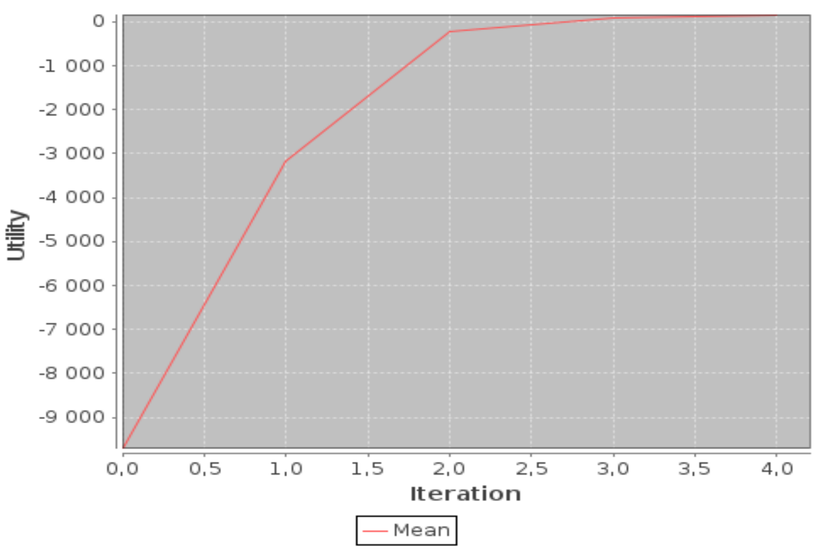
We propose a methodology of location and relocation of amenities (home, shop, work, leisure places) for urban planning decision. Our methodology exploits mobile phone data to propose optimal amenity locations to reduce the overall travel time or travel distance. We mix multi-agent simulation with combinatorial optimization techniques. MATSim is used as a multi-agent simulator system based on geographical data within a GIS.

Possible use for development :

Urban Planning, Amenity Locations, Transportation Network Design, Spatial Marketing



Spatial flows of simulated agents



Agent Utility Mean Variation

Main results:

- A methodology for generating individual plans from mobile phone and geographical data
- Automatic computation of suitable amenity relocations
- A Prototype Software called SPOT
- Tests on Dakar Region : Multi-Agents Simulation, optimization of amenity locations

Methods:

- Combinatorial Optimization Approach
- Clusterings
- Local Search method
- Greedy Heuristic method
- Queries in Spatial Data Base Management System (GIS)

Full paper is here:
https://www.dropbox.com/s/lxa9fe4dx0kkpmc/SPOT_D4D_SENEGAL_REPORT_V1.pdf?dl=0

DataViz or video are here :
https://www.dropbox.com/s/syxgkn9w7w69q12/SPOT-SENOZON.mov?dl=0

Data sources used for this project:
- D4D data set 2
- Open Street Map data, Osmosis web site
- D4D synthetic data set

Other data sets used in this project:
- Type of data : Household surveys
  Source : CETUD,GMAT

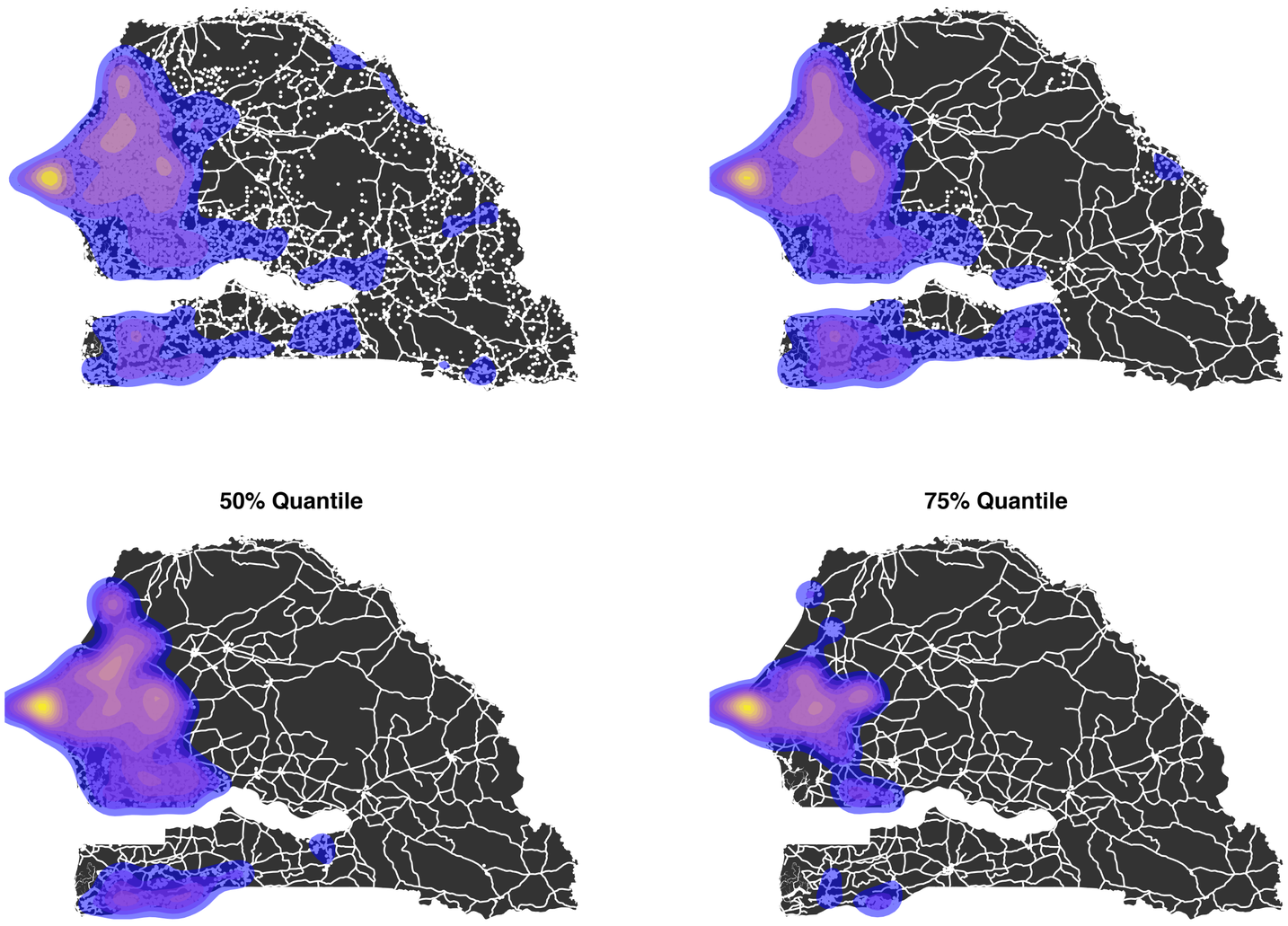
Main Tools used:
- MATSim, Senozon
- QGIS
- Eclipse, JAVA

Open Code available:
- Yes ☑
- No ☐

T16

Where do we Develop? Discovering Regions for Urban Investment in Senegal

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network

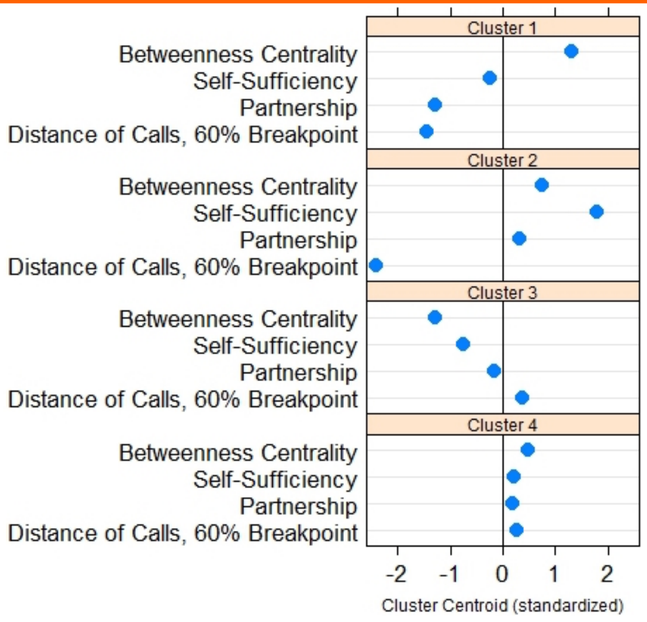


- **Derek Doran**, Dept. Of Computer Science & Engineering, Kno.e.sis Research Center, Wright State University
- **Andrew Fox**, Dept. Of Industrial Engineering And Management Science, Northwestern University
- **Veena Mendiratta**, Bell Laboratories, Alcatel-Lucent

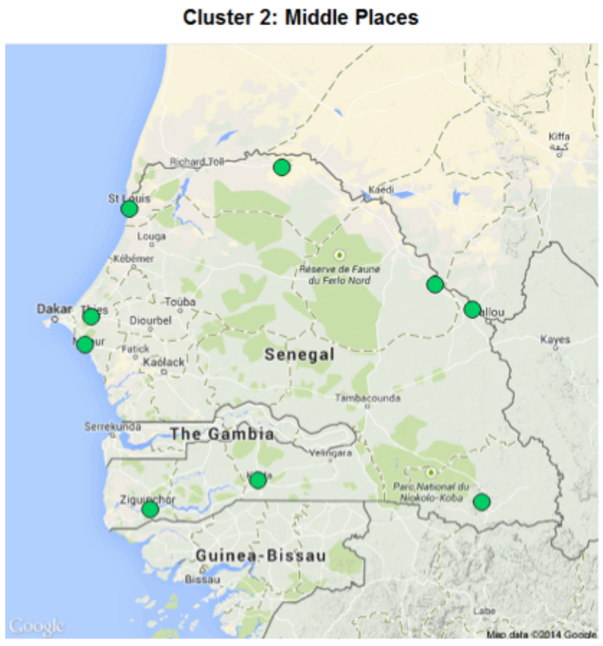
Project Summary:

Rapid growth of existing urban areas is a typical phenomenon in developing nations and leads to poor living conditions, weak industrialization, and struggling economies. Urban planning studies agree that the best means to mitigate these effects are to develop other locations and cities in the country, so that citizens are encouraged to emigrate out of the country's central hubs.

This work operationalizes two critical geographic and urban planning theories – Central Place Theory (CPT) and Central Flow Theory (CFT) - to identify arrondissements whose cities represent promising locations for rapid investment that may alleviate urbanization. We mine features related to CPT and CFT through mobile phone data to identify the most promising arrondissements for urban development using an unsupervised learner.



Centroids of clusters in best FMM solution



Clustering solution: Best investment opportunities

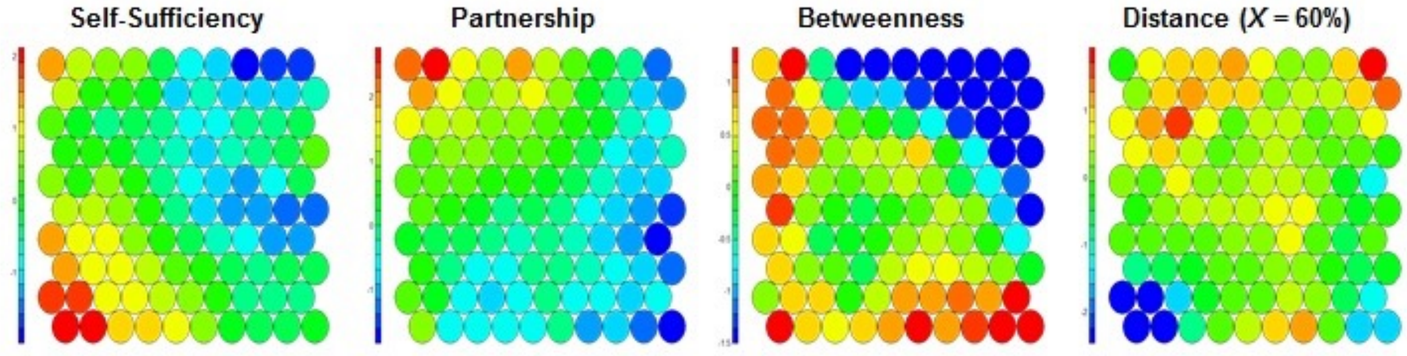
Main results:

- We capture concepts of CPT and CFT within mobile phone calling patterns across arrondiseements in Senegal and apply an unsupervised learner (clustering) to classify them by their ability to support Low, Middle, and Central Places (CPT concepts)
  - CPT Related Features: Distance of Calls, Self-Sufficiency, Partnership
  - CFT Related Features: Centrality, Total Call Volume

- We compared numerous Finite Mixture Model solutions (different FMM hyper-parameter settings and features). The best solution is identified by the **multicollinearity** of features considered, **actionability** of the results, **BIC,** and **Pseudo-F**.

- Following CPT/CFT, the cluster with extreme centroid component values (cluster 2) represent the best locations for urban investment. They include:

1. **Thies:** Largest transportation hub and provides countries top exports
2. **St Louis**: Tourism based economy, high rates of sugar prod., farming, and fishing
3. **Mbour:** South of Dakar and among one of the fastest growing cities in the country



Features used in best solution: Middle Places exhibit high self-sufficiency, partnership, and betweenness, and extremely high and low call distance.

Full paper is here:
put your link here

Data sources used for this project:

- D4D data set 1, communication between antennas
- Global Gazetteer Version 2.2 http://www.fallingrain.com/world/index.html

Main Tools used:

- R
- Packages: igraph, ggmaps, ggplot2, mclust, kohonen

Open Code available:

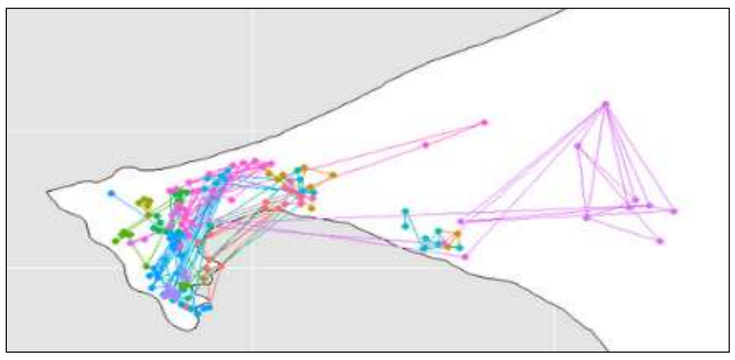
- Yes (please contact authors)

T17

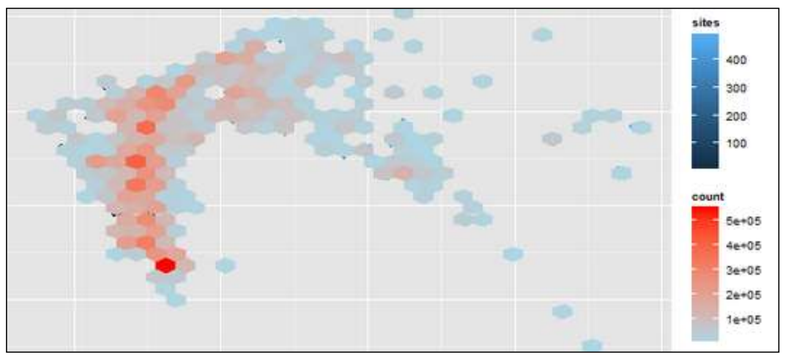
Understanding Traffic Matrix for Transportation Planning with Interregional Connectivity

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network

(a) Inter-regional traffic changes.



(b) User displacement patterns (14 users)



(c) Density graph of traffic

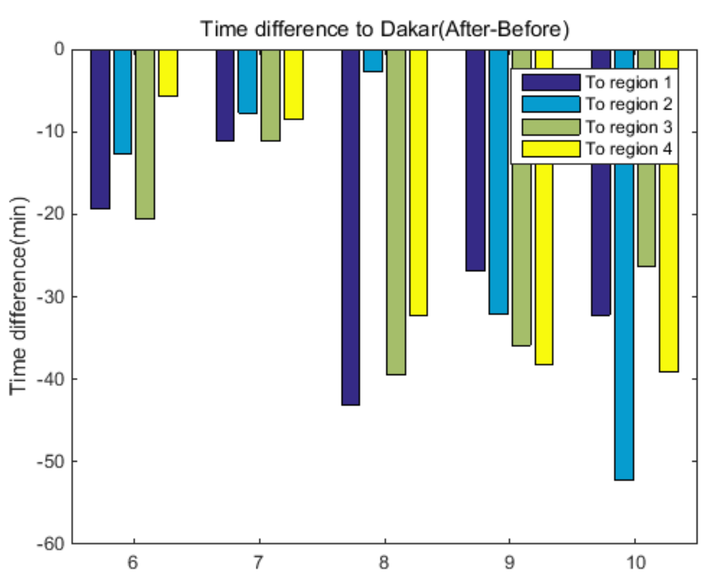
- Eunji Im, M.S. cnadidate, KAIST.
- Seungho Kim, M.S. canddiate, KAIST
- Jongin Lee, PhD candidate., KAIST
- John Kim, Professor, KAIST

Project Summary:

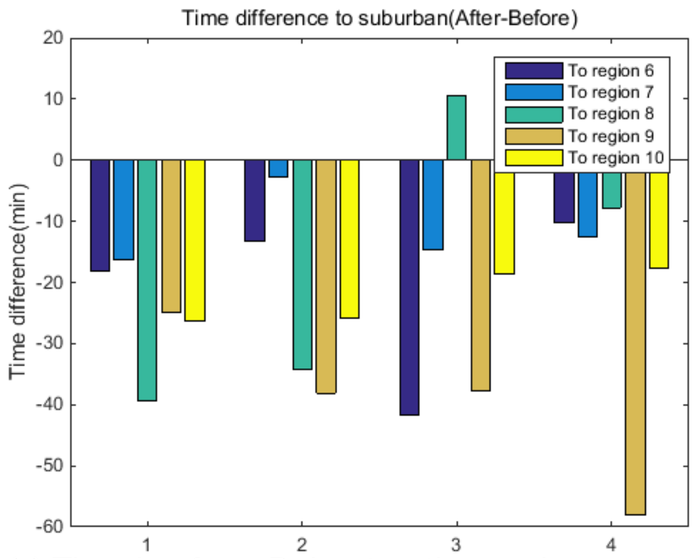
In this paper, we evaluate the effect of constructing new "Dakar Diamniadio Toll Highway" by showing connectivity between regions in nearby the road and user movement patterns. Obviously, changes of traffic volume and time exist after constructing new highway among 10 arrondissements in Darkar. Users are also affected by the new highway in detail analysis based on weekend and weekday.

Possible use for development:

The results could be used for reevaluation of the new highway for supplement. Other developing countries refer to the result with considering country-specific characteristics of Senegal while planning for new highway in the future.



(d) Time lag from suburban to Dakar before and after the construction of highway.



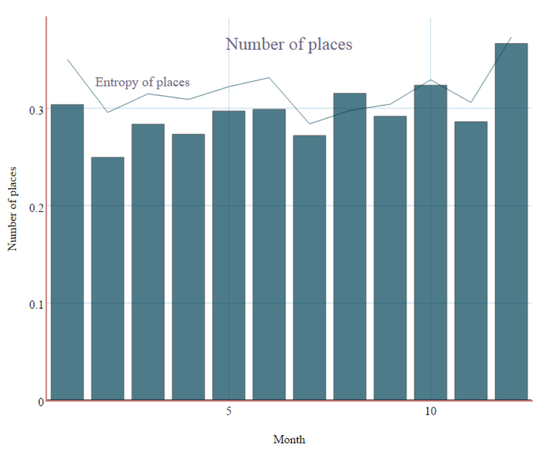
(e) Time lag from Dakar to other regions before and after the construction of highway.

Main results:

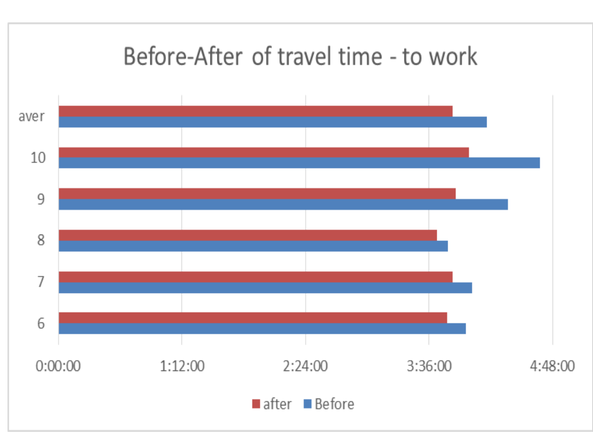
- (a) The red arrows show the traffic increases, and blue ones show traffic decreases. The orange line is the new highway constructed on August 1st. The tendency of increase and decrease are along the road.
- (b) The displacement of 14 individual patterns in Dakar with SET2
- (c) Density of CDR data has availability to understand population density with most frequent visits.
- (d) Time lag from suburb to Darkar, (e) Time lag from Darkar to other regions
- (f) Similar tendency in number of places.
- (g) Time lag in commute time
- (h) Dakar versus Suburban regions

Methods:

- Step 1 : selecting 10 arrondissements of Darkar in nearby DDTH and divide 2 groups (one group is left side of new highway and another group is right side)
- Step 2 : finding changes in traffic volume and time between the 2 groups.
- Step 3 : finding changes of user movement in weekend and weeekday.



(f)



(g)

Table 1. Dakar versus Suburban regions

ID	Department	ARR	Division
1	DAKAR	PARCELLES ASSAINIES	Left side of the new highway
2	DAKAR	ALMADIES	
3	DAKAR	GRAND DAKAR	
4	DAKAR	DAKAR PLATEAU	
5	GUEDIAWAYE	GUEDIAWAYE	
6	PIKINE	PIKINE DAGOUDANE	Right side of the new highway
7	PIKINE	THIAROYE	
8	PIKINE	NIAYES	
9	RUFISQUE	RUFISQUE	
10	RUFISQUE	BAMBILOR	

(h)

Full paper is here:
put your link here

DataViz or video are here:
http://cdb.io/1t8vuAf
Login:
Pw:

Data sources used for this project:

D4D data set 1, com between antenna	X
D4D data set 2, movement routes high res	X
D4D data set 3, movement routes low res	X
D4D synthetic data set	

Other data sets used in this project:

- Type of data: Source:
- Type of data: Source:
- Type of data: Source:

Main Tools used:

- Tool 1 Matlab
- Tool 2 R studio
- Languages: python, matlab, R
- Algorithm

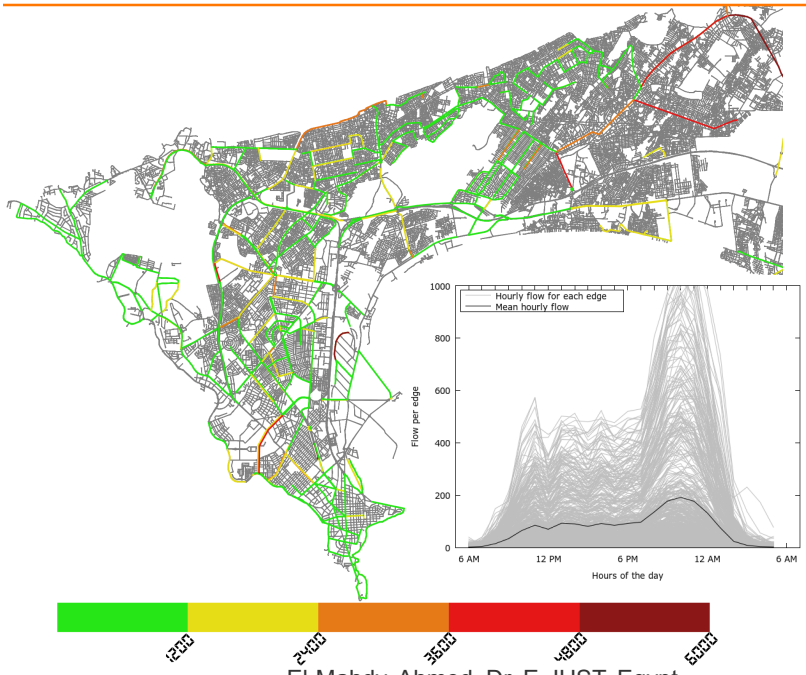
Open Code available:

- Yes X
- No

T18

High Resolution Traffic Maps Generation Using Cellular Big Data

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



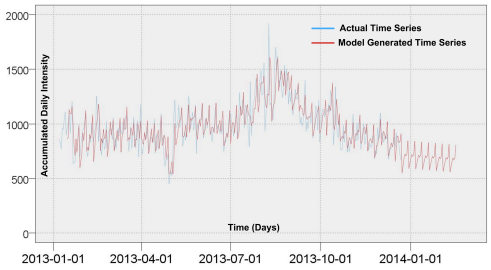
- El-Mahdy, Ahmed, Dr, E-JUST, Egypt
- Algizawi, Essam, Eng., E-JUST, Egypt
- Ogawa, Tetsuji, Dr, E-JUST, Egypt
- Shishiny, Hisham, Dr, IBM, Egypt
- Baddar, Mohamed, Eng., IBM, Egypt
- Kimura, Keiji, Prof. Waseda University, Japan

Project Summary:

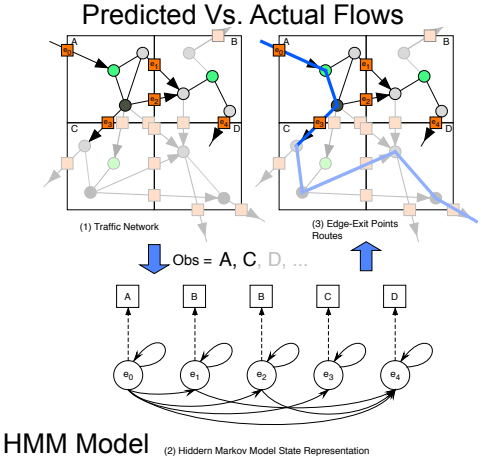
We consider, for the first time, utilising the mobile big data for 'microscopic' level traffic analysis. The project develops a HMM formulation, and uses Viterbi decoding to discover actual road segments of trips. This facilitates road-level analysis without the need for high cost traffic on-road traditional sensors. We then generate Dakar traffic intensity maps for main roads, for every hour in the covered 50 weeks, of the year 2013. Moreover, we develop and apply a traffic prediction model to the data and identify significant traffic seasonality patterns.

Possible use for development:

Generating traffic intensity map for Senegal's major roads. Together with the provided traffic analysis and predictions component, the project can supply accurate information to both public users and decision makers, helping in improving a traffic metric and support 'new road' planning, thereby decreasing road congestions.



Predicted Vs. Actual Flows



HMM Model

Main results:

- A real-time, model-based, traffic monitoring system that yields fine-grained traffic intensity maps from sparse CDR observations.
- Estimated trip trajectories, for the covered 50 weeks, that are highly correlated with criteria satisfied by the transportation phenomena, e.g., gravity and typical city mobility pattern.
- Identifying hourly and daily road traffic seasonality patterns
- Forecasting road traffic for the next coming hours and days

Methodology:

- The traffic phenomena on the road network is modelled using Hidden Markov Model (HMM) that is designed to cope with missing observations and nearby antennas associations. It has states defined as the exit edges of Voronoi regions defined by the antenna locations. The proposed system consists of following components:
  - Trip detection: extracts BTS fingerprints for each trip.
  - Trip trajectory mapping: maps BTS fingerprints onto road segments using Viterbi algorithm in which "cutting-branch"-like techniques are exploited to achieve efficient processing.
  - Accurate route visualization: estimates accurate routes using both Viterbi outputs and other road information such as junctions.
  - Traffic intensity map generation: accumulated statistics of trip trajectories using segmental K-means clustering, yielding intensity of road segments.
  - Road traffic Time Series analysis, modelling and forecasting.

Full paper is here:
http://tinyurl.com/nnq8ego

DataViz or video are here:
http://tinyurl.com/nqvudcv
Login: ejust-pcl
Pw: KjRs4!As

Data sources used for this project:

- D4D data set 1, com between antenna ☐
- D4D data set 2, movement routes high res ☒
- D4D data set 3, movement routes low res ☐
- D4D synthetic data set ☐

Other data sets used in this project:

- Type of data: Senegal OSM Map Source: D4D
- Type of data: Source:
- Type of data: Source:

Main Tools Used:

- Voronoi Partitioning Algorithm
- Viterbi Decoding Algorithm
- IBM SPSS Modeler 16
- IBM SPSS Statistics 22

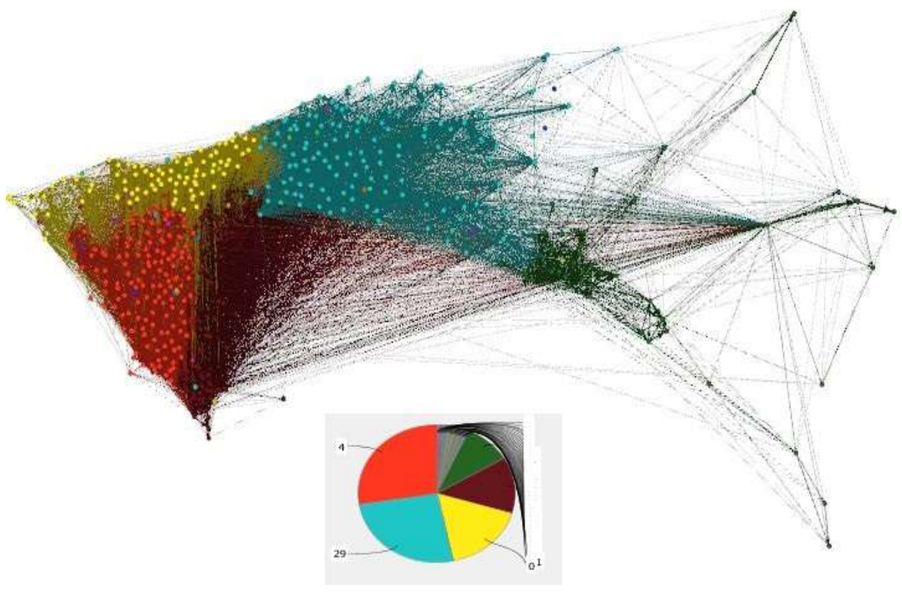
Open Code Available:

- Yes ☒
- No ☐

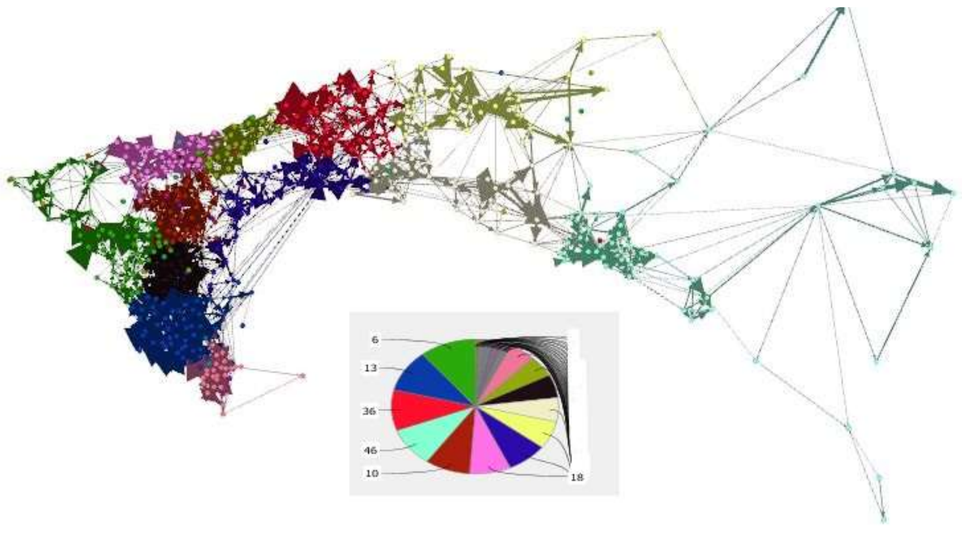
T19

Structure and resilience of call networks in Senegal

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



Community structure in the network induced by calls for the month of January. Colors indicate different communities.



Community structure in the network induced by mobility traces for the month of January. Colors indicate different communties.

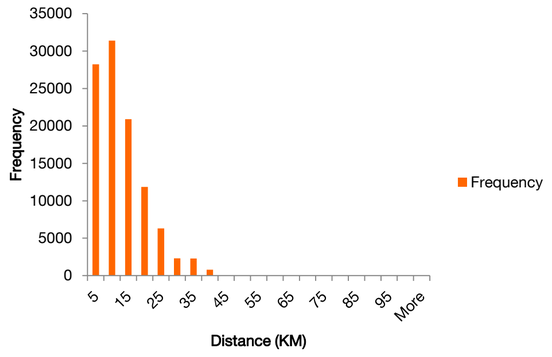
- Jose Cadena, Virginia Bioinformatics Institute and Dept. Of Computer Science, Virginia Tech
- Tania Hamid, Virginia Bioinformatics Institute and Dept. Of Computer Science, Virginia Tech
- Achla Marathe, Virginia Bioinformatics Institute and Dept. Of Agricultural Economics, Virginia Tech
- Anil Vullikanti, Virginia Bioinformatics Institute and Dept. Of Computer Science, Virginia Tech

Project Summary:

We study the structural properties and resilience of cellular networks for Senegal, with the goals of understanding: (1) The underlying connectivity patterns, (2) Effect of geographically localized failures, and (3) Regions of significant variation in mobility and call patterns. We focus on the Dakar metropolitan region, and analyze the network induced on the cell towers by calls and mobility traces.

Possible use for development:

The community structure can help understand call and mobility patterns within the social network, which need to be maintained for social resilience. The cell towers in regions of significant variation in call patterns can help in identifying critical regions.



Frequency distribution of the distance between the end points of edges in the call network for the month of January.

Cell towers with significant variation in daily call volumes, compared to the baseline over a 28 day period, and over a disk of radius 1km.

Main results:

- The call network has a small number of communities, and is quite resilient to spatially localized failures. In contrast, the mobility network has many more communities, and is much less resilient.
- The communities have a strong spatial assortativity structure. Most high weight edges in both networks span close-by regions.
- There is a small set of cell towers which often have significant variations in call volume and mobility. Some of these lie in regions of high flood risk.

Methods:

We analyze two networks on the cell towers induced by calls and mobility traces for different periods of time. We use the following methods

- Community detection: we run the Louvain clustering method of [Blondel et al., 2008] to identify the community structure.
- Spatial scan statistics: we use the approach of non-parametric spatial scan statistics [Neill, 2012] to identify regions with anomalous call patterns and mobility counts.

Full paper is here:

DataViz or video are here:
Login:
Pw:

Data sources used for this project:

- D4D data set 1, com between antenna [X]
- D4D data set 2, movement routes high res [X]
- D4D data set 3, movement routes low res [X]
- D4D synthetic data set [ ]

Other data sets used in this project: Report by [Montoliu-Munoz, M. and Wang, H.G., 2009]

Main Tools used:

- Community detection algorithm by Blondel et al., 2008
- Fast spatial Scan Statistics by Neill, 2012

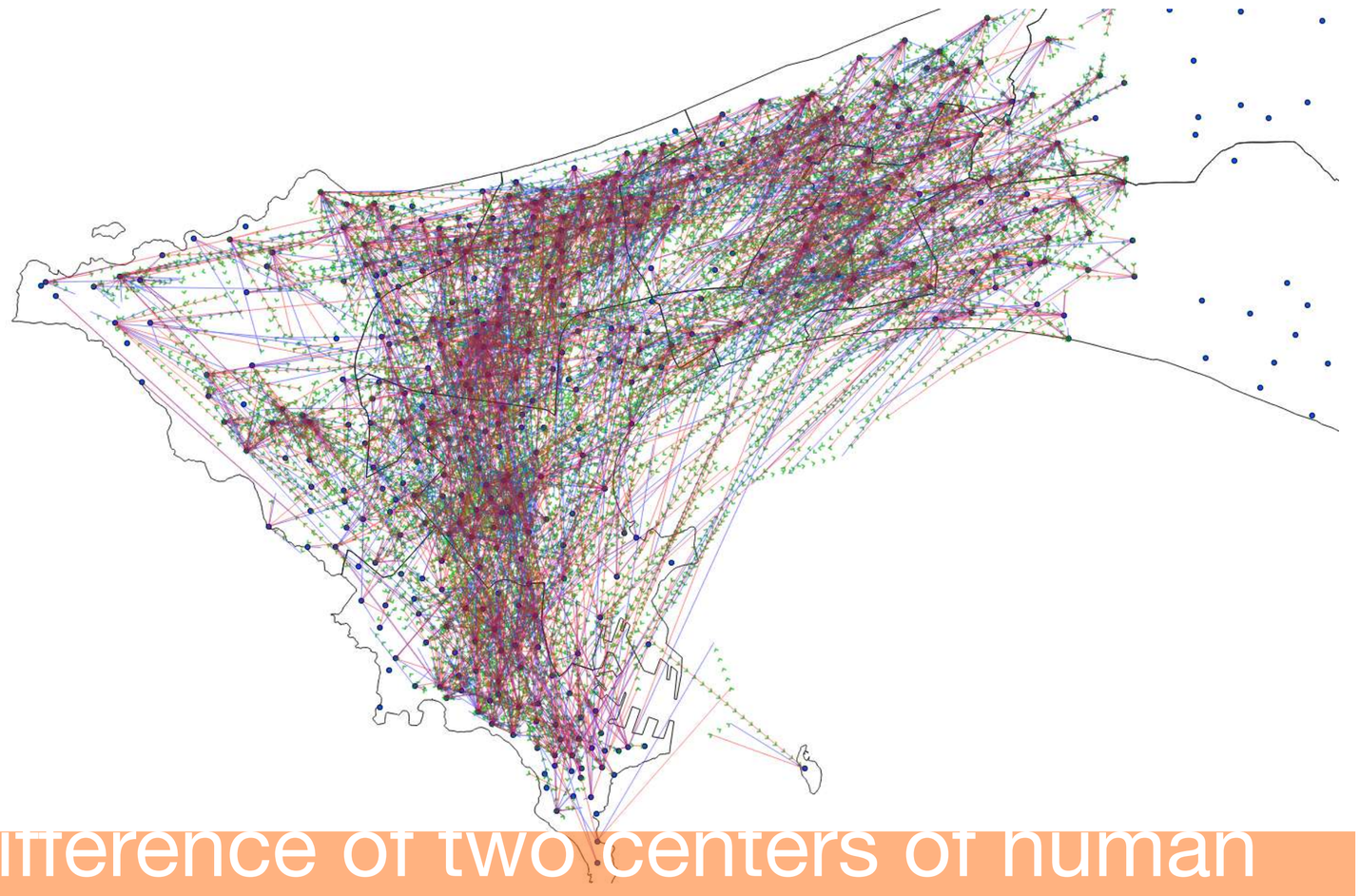
Open Code available:

- Yes [X]
- No [ ]

T20

Exploring relationships between human mobility motifs and rainfall

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



Difference of two centers of human mobility motifs between dry and rainy days in Dakar

Fukui University of Technology

- Akihiro Fujihara (lead author),
- Daisuke Nogiwa,
- Toshihiro Kasai,
- Shota Maegawa

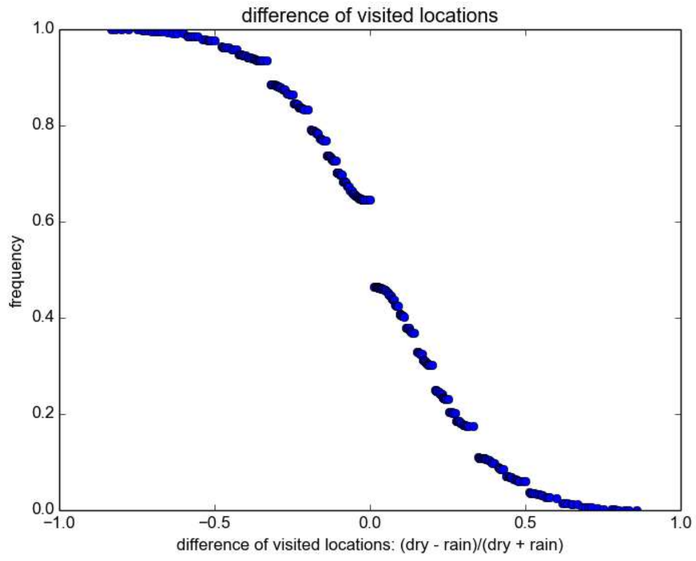
Project Summary:

We investigate difference of daily human mobility motifs between rainy and dry (no rain) days. We use the high-resolution movement routes dataset and precipitation dataset in Dakar and analyze size distribution of visited locations and distribution of center of human mobility motifs for detecting clear quantitative and spatial difference of the patterns.

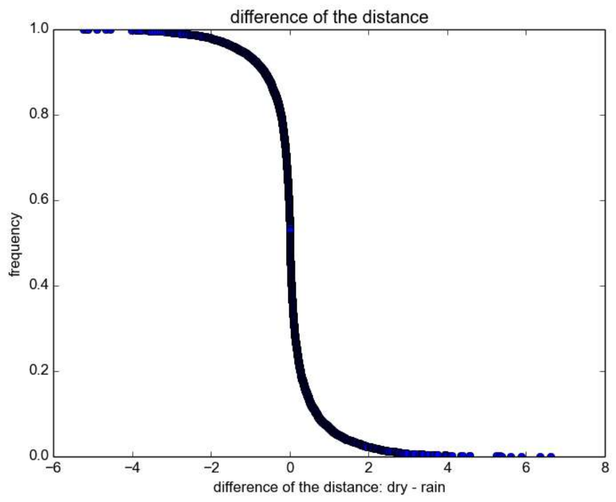
As a result, we find the evidence that human mobility motifs change by rainfall.

Possible use for development

At this time, it is difficult to apply our result as it is, but we successfully observe the difference of human mobility motifs using some statistical values. For future work, we wish to continue this study using Synthetic CDR to develop an autonomous rainfall detection system using human mobility motifs.



Distribution of difference of size distribution of visited locations



Distribution of difference of center of human mobility motifs

Main results:

- We find that both the distributions of the difference weakly shift to positive (dry day) side. This means that the majority of human mobility motifs shrinks around the most visited location by rainfall.

  - The first figure on the left shows that the distribution of $(N_{dry} - N_{rain}) / (N_{dry} + N_{rain})$ is weakly biased to the positive side although there are many samples with zero values and negative values.
  - The second figure on the left shows that the distribution of $D = C_{dry} - C_{rain}$ is also weakly biased to the positive side.

Methods:

- We separate the high-resolution movement routes dataset into dry and rainy days based on precipitation data in Dakar. Then we consider two statistical values: difference of size distribution of visited locations *N* and that of center of human mobility motifs *C*. We show complementary cumulative distribution functions of these statistical values.

  - The first figure on the left shows the distribution of the difference of size distribution of visited locations.
  - The second figure on the left shows the distribution of the difference of center of human mobility motifs (which is calculated similarly by doing the center of gravity). The top fiture shows the spatial patterns in Dakar.

Full paper is here:
https://drive.google.com/file/d/0B7XkKn30vxbUNnZNelpkU0lPZ2c/view

DataViz or video are here:
https://drive.google.com/file/d/0B7XkKn30vxbUV3ZhS2V2ODNLUFE/view

Login:

Data sources used for this project:

- D4D data set 1, com between antenna
- D4D data set 2, movement routes high res [x]
- D4D data set 3, movement routes low res
- D4D synthetic data set

Other data sets used in this project:

- Type of data: weather data (precipitation)
  Source: http://www.meteociel.fr/temps-reel/obs_villes.php?code2=61641

Main Tools used:

- Python Programming Language
- SAS
- QGIS

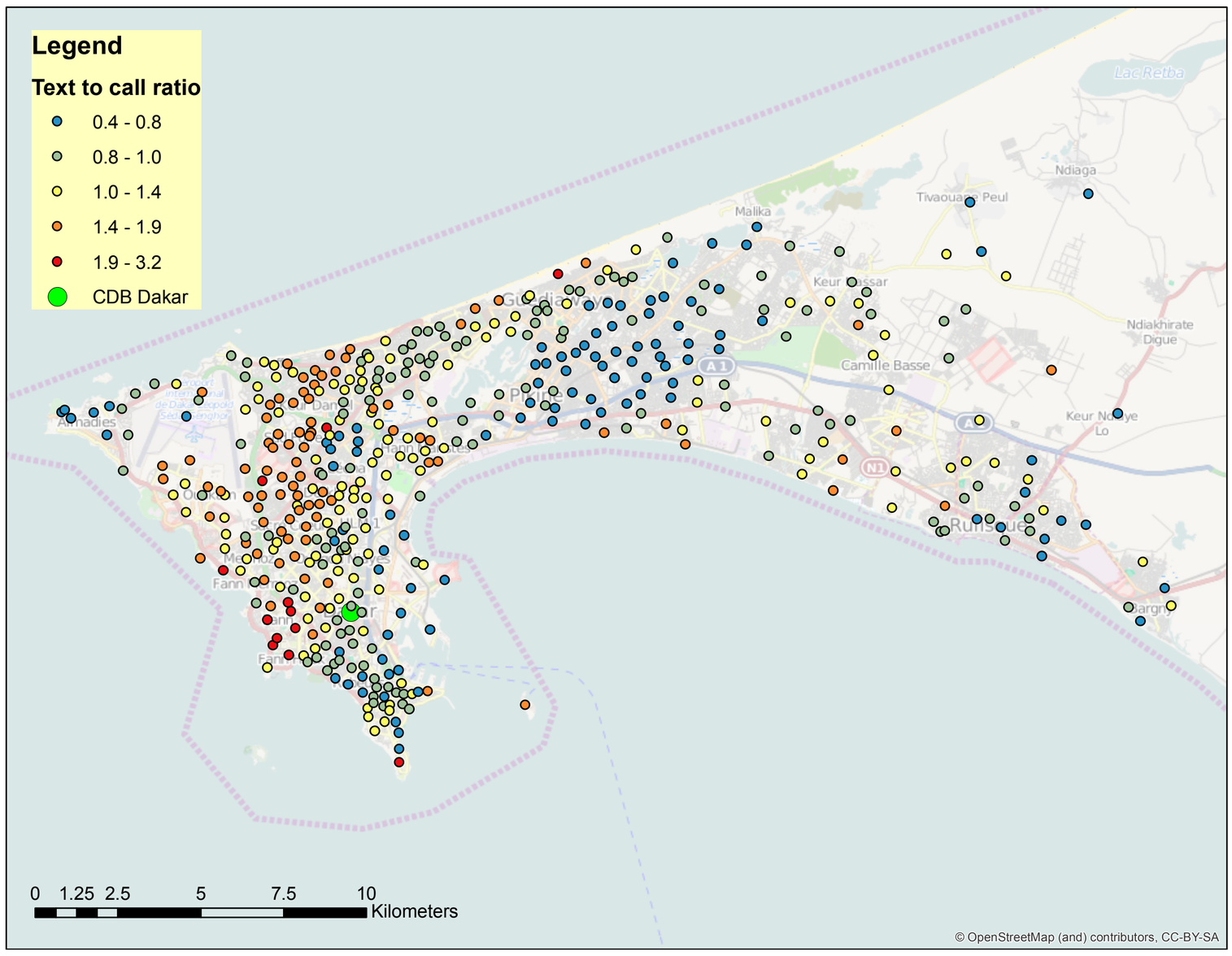
Open Code available:

- Yes [x]
- No [ ]

T21

Text me - if you can: Literacy, Networks and Mobility in Dakar

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

- Kirchberger, Martina, Post-Doctoral Research Fellow, Columbia University
- Small, Christopher, Professor, Columbia University

Project Summary:

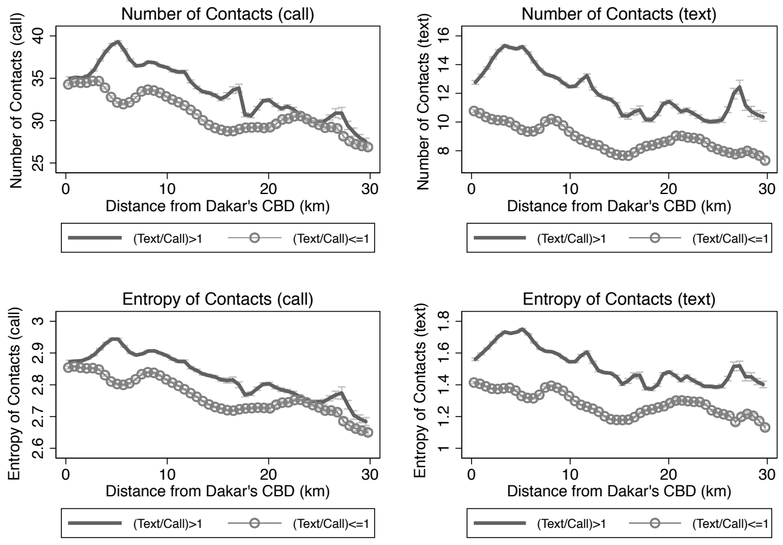
This project illustrates the use of call and text volumes, mobility patterns and bandicoot behavioral indicators to understand the level of integration of individuals into networks. We find that individuals residing closer to the central business district of Dakar are not only geographically more centrally located but also have a higher number of contacts, communicate from a larger number of towers, and have a higher entropy of contacts (relatively more weaker ties).

We then investigate the use of the text to call ratio as a proxy for literacy. We show that there is a gap for individuals residing in areas where calls dominate relative to texts: these individuals have fewer contacts they text and call with, and they have fewer weak ties.

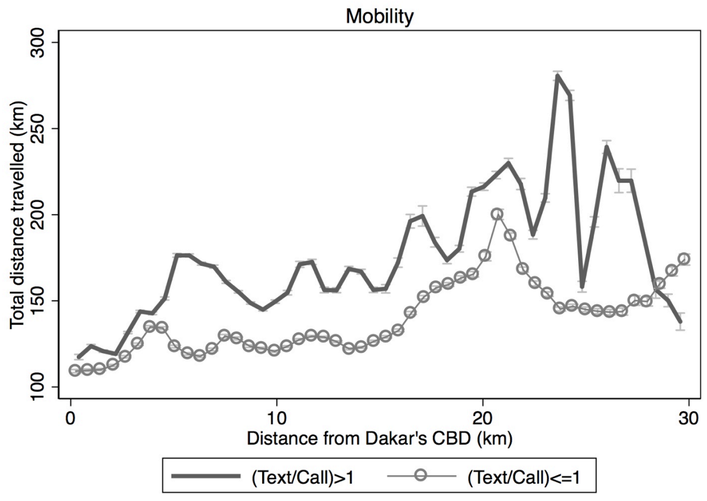
Finally, we do not find evidence that individuals who are residing in these areas compensate the lack of mobile phone activity with a higher level of mobility.

Possible use for development:

Identification of areas with low literacy and integration into networks could help governments target interventions such as the dissemination of information on jobs and facilitating matching of workers and jobs. The relationship between texts and calls could also inform policy makers which way of disseminating (i.e. sending text messages vs calling) information is most effective in the particular context.



Text to call ratio and behavior



Text to call ratio and mobility

Main results:

- There is a remarkable drop-off in all the variables measuring mobile phone activity and network size as one moves away from the city center. Individuals contact fewer numbers, have fewer communications, and have lower entropy of calls and texts.
- Individuals residing in areas where calls dominate over texts, over a substantial spatial distance from Dakar, are in touch with a significantly lower number of contacts. So even though they call *relatively more* compared to texting, they call *fewer* people and have fewer weak ties over a substantial range of the distribution.
- Individuals residing in areas where calls dominate texts are not compensating for the lower level of mobile phone activity by travelling longer distances.

Methods:

- We define an individual's home location as the tower that is recorded most often between 7pm and 7am in the two week window. We then compute the travel distance of the individuals' ``home tower'' to the Central Business District (CBD) of Dakar. To do this, we extract the Open StreetMap extents for Dakar, build a network dataset and use the Origin Destination Cost Matrix tool in ArcGIS to compute the network distances between these two locations. As features are snapped to the nearest location on the network, we define the total distance to the CBD as the travel distance on the network to the center of Dakar plus the straight line distance of the tower to the nearest link on the network.
- From SET2 we aggregate tower level texts and calls to compute the tower level text to call ratio.
- We overlay the Voronoi cells with 5km buffer around the survey clusters of the 2010 Demographic and Health Survey to compute the average literacy rate for each of the Voronoi cells.

Data sources used for this project:

- D4D data set 1, com between antenna [x]
- D4D data set 2, movement routes high res [x]
- D4D data set 3, movement routes low res [ ]
- D4D synthetic data set [ ]

Other data sets used in this project:

- Type of data: Demographic and Health Survey 2010 *source: www.measuredhs.com*
- Type of data: Open Street Map; *source: www.openstreetmap.com*

Main Tools used:

- Tool 1: ArcGIS 10.2.2
- Tool 2: R
- Tool 3: Stata

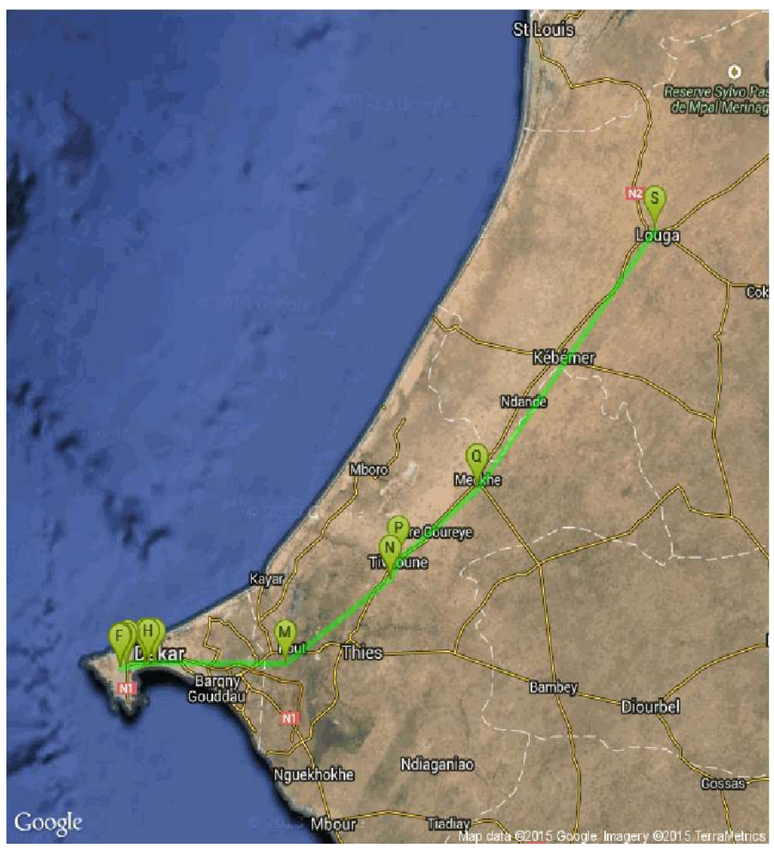
Open Code available:

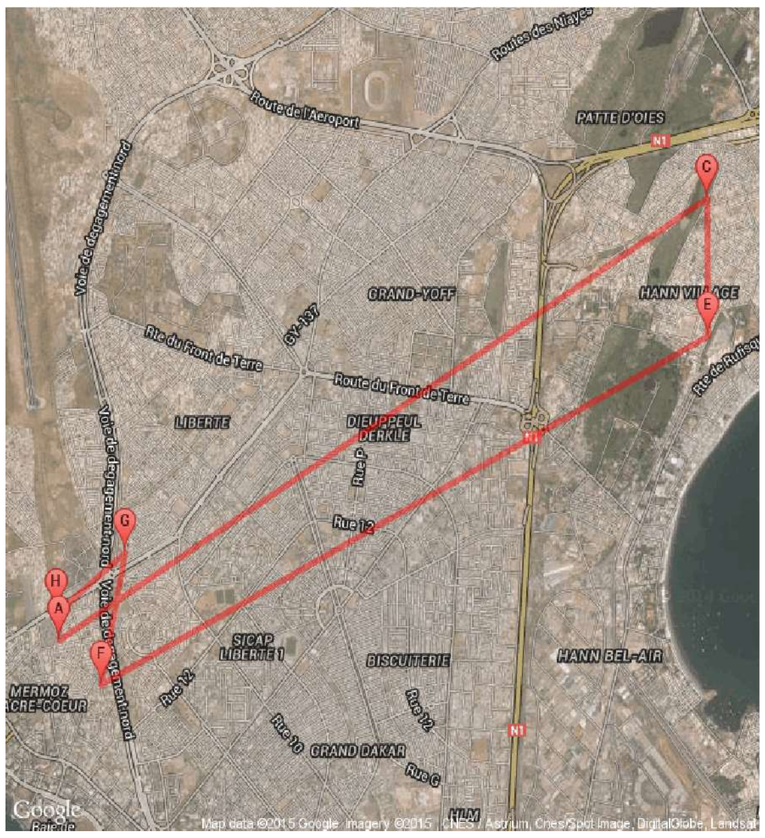
- Yes [x]
- No [ ]

T22

Cellphone's Users Trajectories Clustering from Call Detail Records

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network





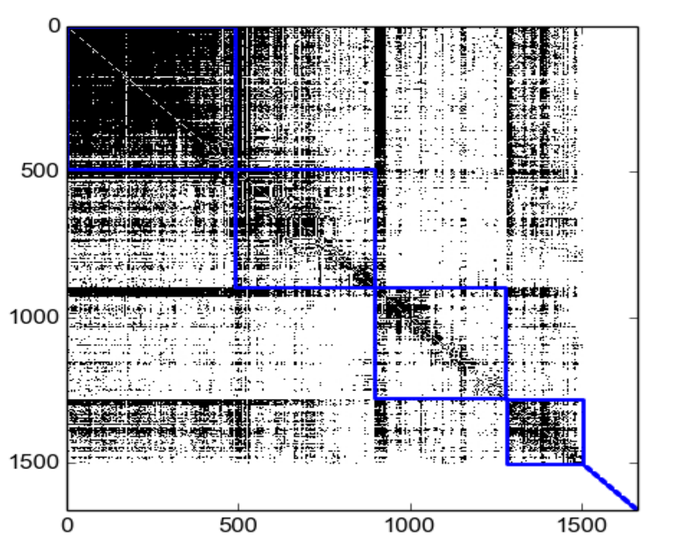
- Paulin Melatagia, Dr., LIRIMA-IDASCO
- Blaise, Ngonmang, Dr., LIRIMA-IDASCO
- Romaric Meleu, Ph.D. Student, LIRIMA-IDASCO
- Vanessa Kamga, Ph.D. Student, LIRIMA-IDASCO
- Armel Nzekon, Master Student, LIRIMA-IDASCO
- Claude Tinku, Master Student, LIRIMA-IDASCO

Project Summary:

We propose a k-medoid like algorithm to cluster the trajectories of cellphone's users. The trajectories are deduced from Call Detail Records and are clustered to generate spatio-temporal trajectories which allows to browse the maximum locations of each cluster in accordance with their timestamps. The clustering algorithm used an new distance function based on sequences alignment and called Alignment distance with Threshold Penalty. We have conducted some experimentations on D4D Senegal Challenge datasets which highlighted three classes of cluster's centroids : daily trip centroids, circular trajectories centroids and short trajectories centroids. These results, combined to the the spatial-temporal representation of the centroids can be used to improve transportation management in cities and villages.

Possible use for development:

The goal of the project is to deduce the movements patterns of cellphone's users in order to provide to transportation companies and government some tools to improve the scheduling of resources deployed for the traffic management.



Communities in Movements Networs

A circular trajectories centroid

Main results:

We found that degree distribution of the movements networks are bimodal with high transitivity (a mean of 0.705) and high average clustering coefficient (a mean of 0.732). Bimodal degree distribution is characteristic of complex networks which contain two types of nodes. Nodes in the distribution with the higher mode are called hubs, those in the other distribution peers.

The experimentations conducted highlighted three classes of cluster's centroids : daily trip centroids, circular trajectories centroids and short trajectories centroids.

Methods:

- We transformed the dataset to generatd compac trajectories representation
- We used a PAM like clustering algoritm
- We define the Aligment distance with Threshold Penality to measure the dissimilarity between trajectories
- We propose an algorithm to agregate the trajectories of a cluster to obtain a representaive one

Full paper is here:
put your link here

DataViz or video are here:
put your link here
Login:
Pw:

Data sources used for this project:
- **D4D data set 1, com between antenna** ☐
- **D4D data set 2, movement routes high res** ☐
- D4D data set 3, movement routes low res ☒
- D4D synthetic data set ☐

Other data sets used in this project:

Main Tools used:
- Python Language
- Google Maps API
- Louvain Algorithm

Open Code available:
- **Yes** ☐
- No ☐

O01

Generating anonymous sequence preserving datasets

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network

Ben-Gurion University of the Negev

Deutsche Telekom

- Rokach, Lior, Associate Professor of Information Systems and Software Engineering, Ben-Gurion University of the Negev
- Shaked, Sigal, Doctoral Student, Ben-Gurion University of the Negev

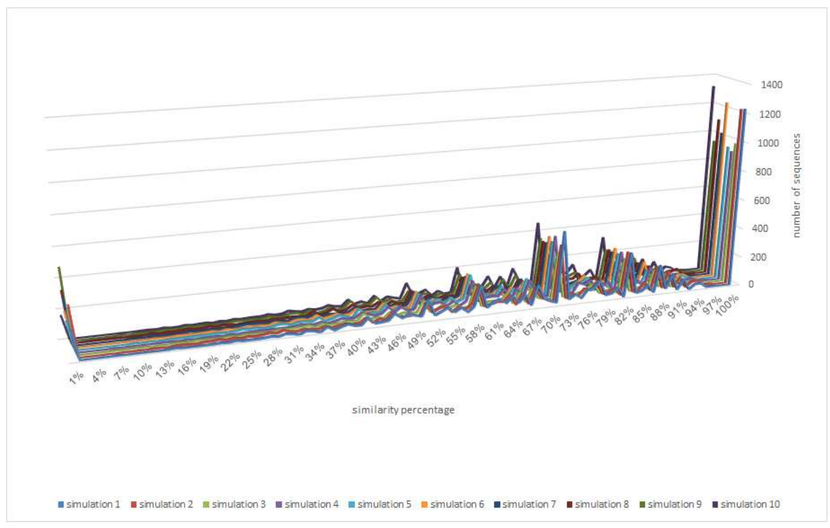
Project Summary:

We suggest a novel technique for generating synthetic anonymous datasets that preserve sequence patterns within the origin dataset.

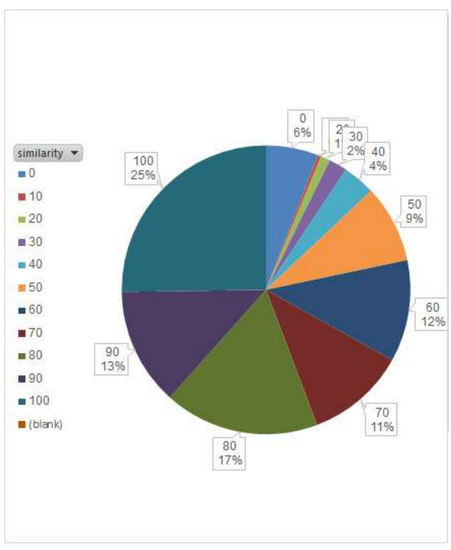
Sequences are first partitioned into groups with similar behavior. Then a Markov model is trained for each group and a synthetic dataset is generated according to the ensemble of models.

Possible use for development:

The generation of anonymous datasets that preserve sequences, or movement patterns in our context, will facilitate the release of sensitive data like call records or GPS data for research community. This will enable analyzing populations and better fitting transportation and infrastructures to their needs.



Preserved similarity within sequences



Sequences partitioned by similarity

Main results:

- Most of the sequences in the origin dataset preserve high similarity to their equivalent in the synthetic dataset. The averaged similarity between sequences in the origin and fabricated data is 69%. This score is damaged mainly for sequences that were suppressed from the synthetic dataset due to k-anonymity requirements.

Methods:

- Sequence within the data are clustered using an agglomerative hierarchical clustering approach, using the single-linkage clustering algorithm. Similarity between sequences is measured using the Longest Common Subsequence (LCSS) measure.
- Clusters where k-anonymity is not preserved are suppressed.
- A Markov model is trained for each group.
  - Statistics are kept regarding starting states and transitions between states.
  - Influencing factors such as hour in the day are combined in the collected statistics.
- An algorithm for generating a synthetic dataset according to the trained model is provided.

Full paper is here:
put your link here

DataViz or video are here:
put your link here
Login:
Pw:

Data sources used for this project:

- D4D data set 1, com between antenna [ ]
- D4D data set 2, movement routes high res [X]
- D4D data set 3, movement routes low res [ ]
- D4D synthetic data set [ ]

Other data sets used in this project:

- Type of data: Source:
- Type of data: Source:
- Type of data: Source:

Main Tools used:

- R TraMineR package
- R cluster package
- Algorithm

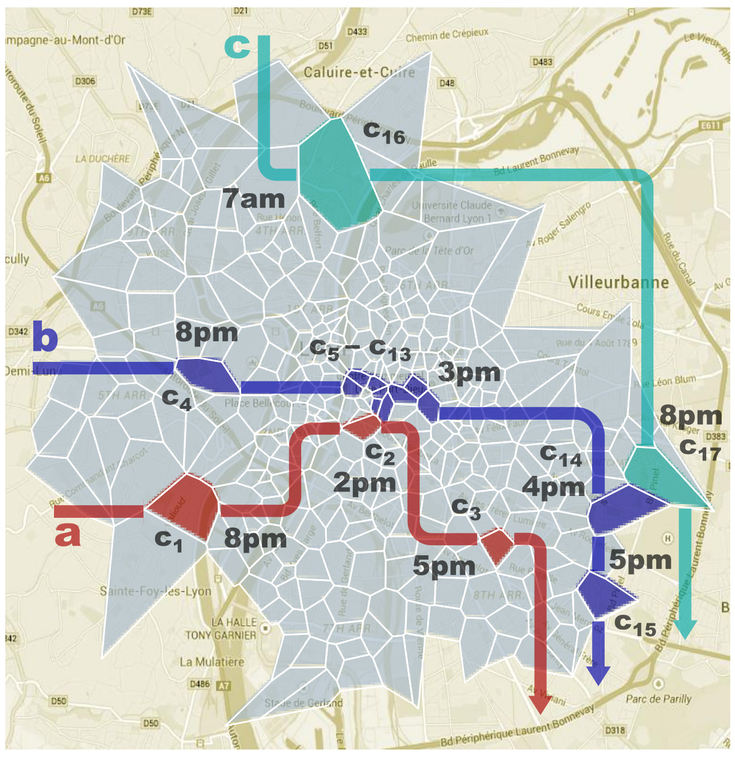
Open Code available:

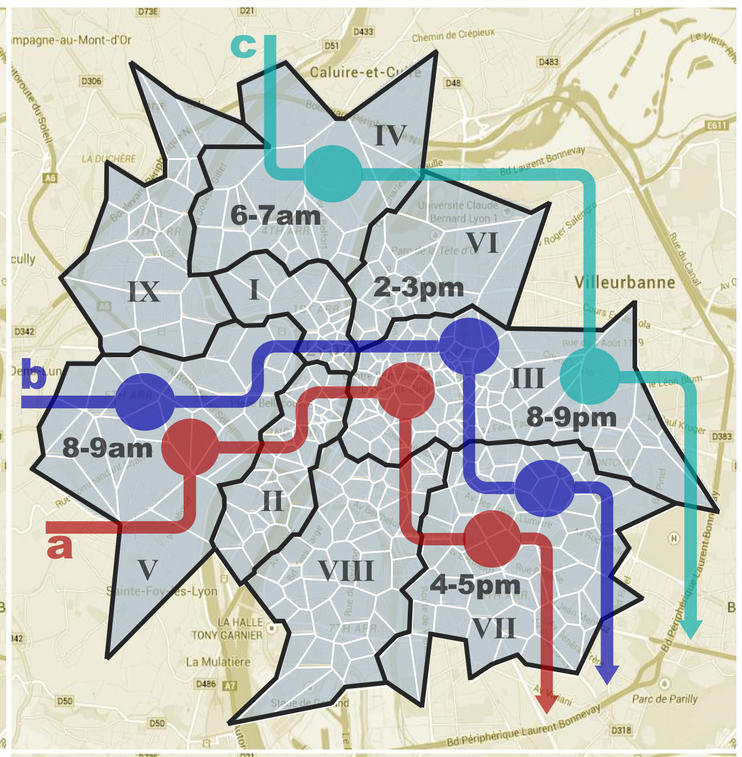
- Yes [X]
- No [ ]

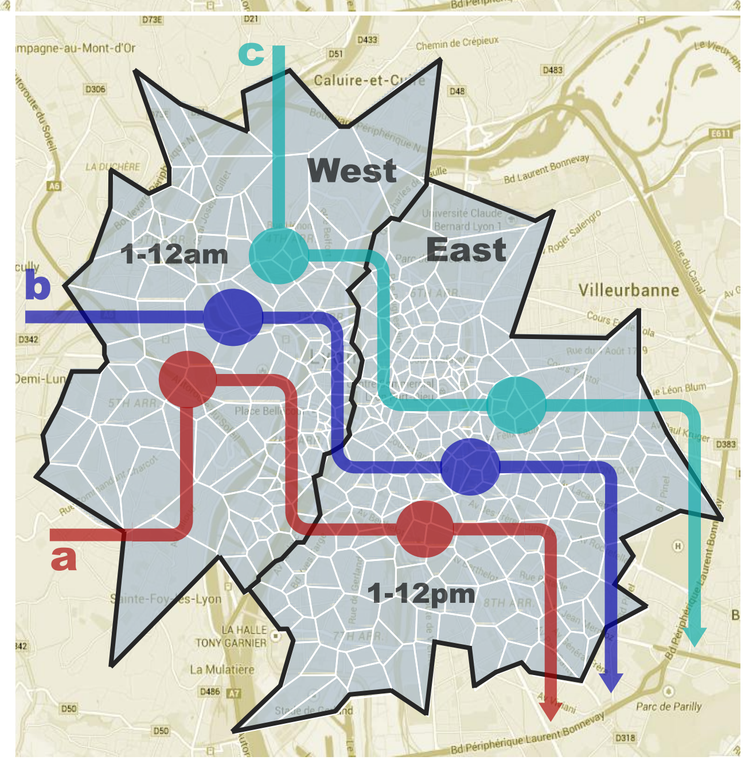
O02

Anonymizability of mobile traffic datasets

Health	Transport Urban	National Statistics	Privacy
Agriculture	Energy	DataViz	Network







Consiglio Nazionale delle Ricerche
Istituto di Elettronica e di Ingegneria dell'Informazione e delle Telecomunicazioni

- **Gramaglia**, Marco, Research Fellow, PhD — CNR-IEIIT Torino, Italy
- **Fiore**, Marco, PhD, , Researcher, PhD HDR — CNR-IEIIT Torino, Italy

Project Summary:

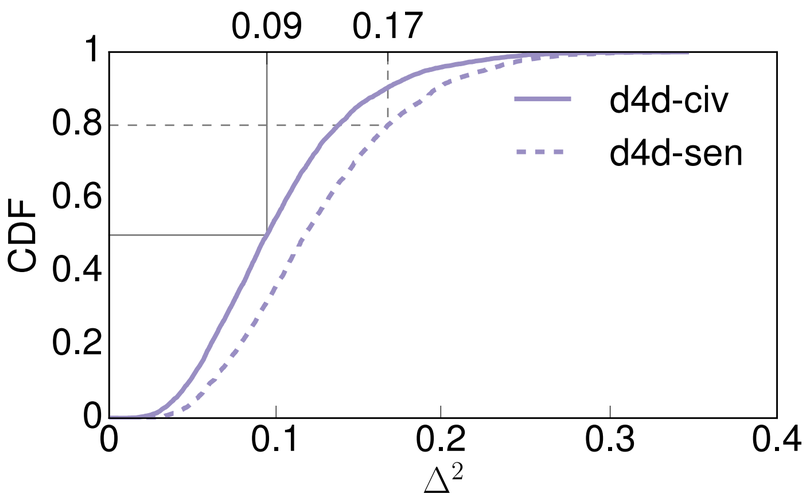
Public disclosure of datasets containing information on precise individuals is an increasingly frequent practice. They yield fine-grained data about large populations that has proven critical to seminal study in a number of research field.

Preserving user privacy in publicly accessible dataset is currently an open problem, as incorrect anonymised dataset may disclose sensible information about specific users. E.g., the European Union set up a Data Protection Working Party, specifically operating on the protection of individuals with regard to the processing and free movement of personal data.

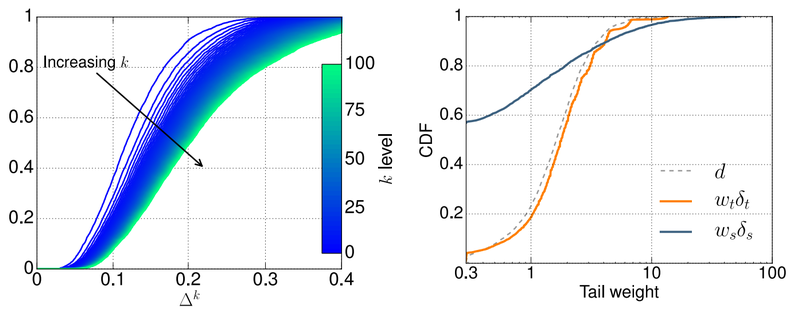
We study the *anonymizability* of mobile traffic datasets, highlighting why they are challenging from a privacy point of view.

Possible use for development:

Distributing better anonymized datasets that do not raise privacy concerns is a key enabler in future applied research. Our study is a step forward in that direction.



CDF of anonymizability of the reference datasets, under the 2-anonymity criterion



Anonymizability for different k-anonymity requirements (left).
Long tail of hard-to-anonymize spatiotemporal samples (right)

Main results:

We analyze user re-identifiability and anonymizability of high-resolution mobile traffic datasets collected in Senegal and Ivory Coast.

We confirm previous finding that mobile traffic datasets feature

•**elevate re-identifiability** – mobile traffic datasets do not satisfy *k*-anonymity for any individual, not even for the minimum $k = 2$

•**poor anonymizability** – mobile traffic datasets require high spatial and temporal generalisation in order to slightly improve the user privacy.

We unveil key causes of the poor anonymizability of the datasets:

•mobile traffic **fingerprints tend to have long tails of privacy-resistant elements** that are much more difficult to anonymize than the average spatiotemporal sample

•these privacy-resistant elements require a huge reduction of granularity in order to be hidden in the crowd, which makes spatiotemporal aggregation inefficient.

Methods:

We propose a measure of anonymizability of subscriber fingerprints in a mobile traffic dataset, based on the privacy criterion of *k*-anonymity.

Keywords – k-anonymity, statistical analysis, distance measure, cumulative distribution function, Gini coefficient, Tail weight index.

Full paper is here:
http://perso.citi.insa-lyon.fr/mfiore/data/gramaglia_netmob15.pdf

Data sources used for this project:
- D4D data set 1, com between antenna ☐
- D4D data set 2, movement routes high res ☒
- D4D data set 3, movement routes low res ☐
- D4D synthetic data set ☐

Other data sets used in this project:
- Type of data: High-resolution CDR Source: D4D'12

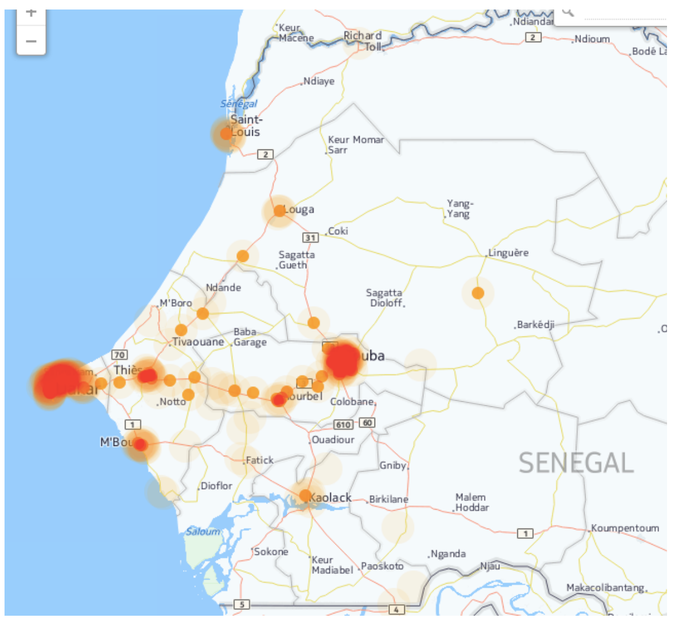
Main Tools used:
- ✦Python
- ✦Pandas
- ✦Matplotlib
- ✦Cuda

Open Code available:
- Yes ☐
- No ☒

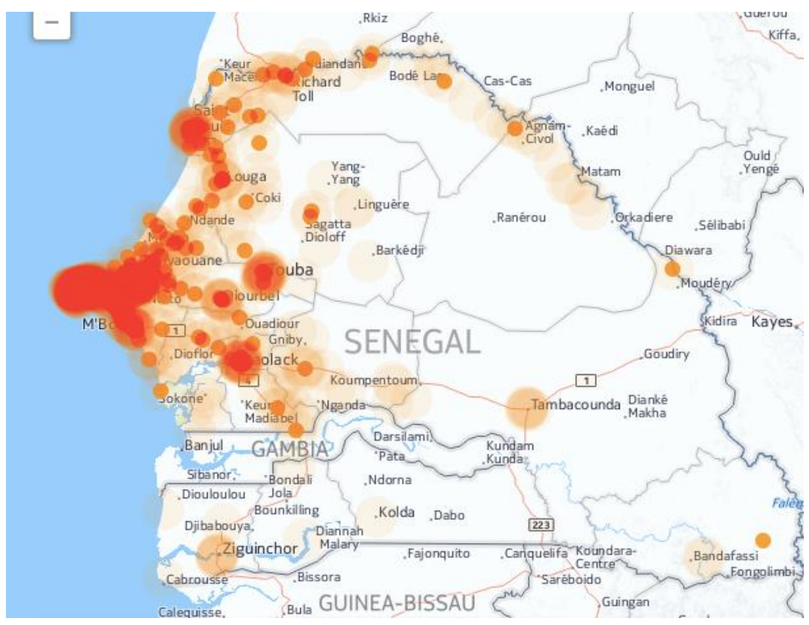
O04

Human Mobility during Religious Festivals in Senegal

Health	**Transport Urban**	National Statistics	Other
Agriculture	Energy	**DataViz**	Network



Human Mobility during Magal of Touba (12/22/2013)
* Outgoing calls



Human Mobility during Gamou of Tivaouane (01/23/2013)
* Outgoing calls

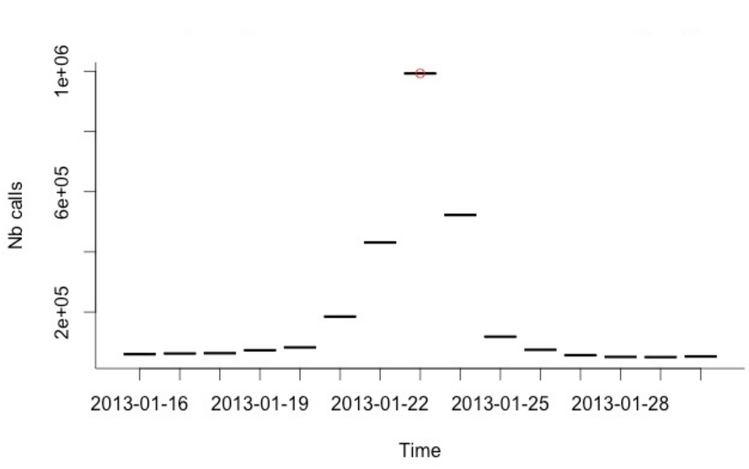
- Dr. Christelle Scharff, Megan Jordan, Briana Vecchione, Pace University, NYC, NY, USA, {cscharff,megan.jordan, bv26460n}@pace.edu
- Dr. Khadidiatou Ndiaye, George Washington University, Washington DC, USA, kndiaye@email.gwu.edu
- Dr. Aminata Niang Diene, University of Cheikh Anta Diop of Dakar (UCAD), Dakar, Senegal, aminaniang@orange.sn
- Dr. Fatou Maria Drame, Gaston Berger University, St Louis, Senegal, fatou-maria.drame@ugb.edu.sn

Project Summary:

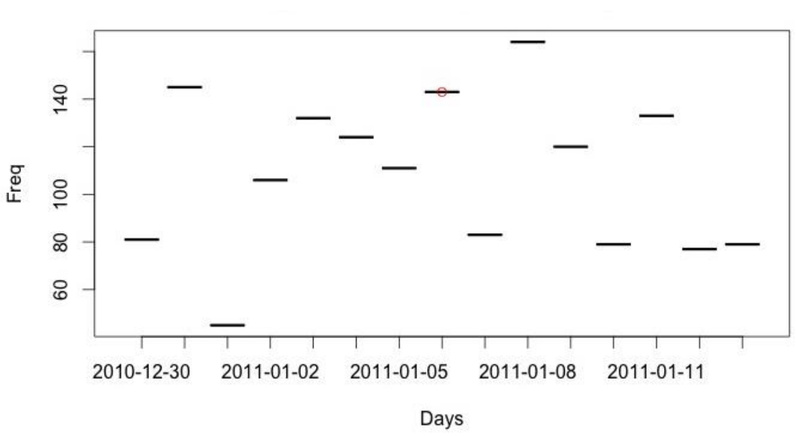
We used the 2013 Orange CDRs to model human mobility before, during and after two key Muslim religious festivals in Senegal that are attended by millions of people: Gamou of Tivaouane and Magal of Touba. We studied the correlation between human mobility and health issues using data on health collected through a toll free hotline service (Numéro Vert) of the Ministry of Health.

Possible use for development:

Festival are often subject to health issues (e.g., cholera outbreaks in 2008). Roads are crowded and accidents are frequent. As pilgrims come from different regions in Senegal, it is important to understand human mobility to control population flow and stop the spread of infectious diseases. Mobile phone use provides opportunities for public safety and mobile health campaigns (e.g., SMS and voice).



Frequency of (Orange) calls to Tivaouane (-/+ 7 days of the Gamou of Tivaouane on 01/23/2013)



Frequency of calls to the Numéro Vert (-/+7 days of the Magal of Touba on 12/22/2013)

Main results:

- Communications (calls and SMS) showed human mobility during two of the key religious festivals in Senegal. SMS were not used as much as calls.
- We found interesting structures in the human mobility patterns showing that these festivals imply massive movements of population from different parts of Senegal depending on the festival and permitting to identify the interconnectedness of communities: Gamou of Tivaouane for followers of the Tijaniyya brotherhood and Magal of Toubal for the followers of the Mouridiyya brotherhood. We could also observe the main routes used by pilgrims.
- Callers to the Numéro Vert hotline were mainly from Dakar and vicinity, Mbacke, Diourbel and Touba. We did not extract significant patterns from the calls and need to investigate further.
- These findings have important implications ranging from resource management to service allocation and awareness campaigns during religious festivals in Senegal.

Methods:

- Subset of the Orange data based on the cities (Tivaouane and Touba) and dates of the festivals: Gamou of Tivaouane (01/23/2013) and Magal of Touba (12/22/2013).
- Use of 4 (out of 4) antennas for Tivaouane and 6 (out of >40) antennas for Touba.
- Model human mobility based on volumes of incoming and outgoing communications in Tivaouane and Touba.
- Plots to show the frequency of calls and SMS.
- Animated map visualizations to show human mobility.
- Map of the phones calls to the Numéro Vert, frequency of calls, and summary of the reasons of the calls.

Full paper is here:

http://bit.ly/d4d2014pgu

DataViz are here:

http://bit.ly/d4d2014pgu

Data sources used for this project:

- D4D data set 1 for 2013, com between antennas
- SET1/SET1V_01.CSV / SET1/SET1V_12.CSV
- SET1/SET1S_01.CSV / SET1/SET1S_12.CSV

Other data sets used in this project:

- Type of data: 2011 & 2012 Health Data / Numéro Vert
- Source: Senegal Ministry of Health / SNEIPS

Main Tools used:

- R / R Studio
- CartoDB
- Amazon EC2
- GitHub

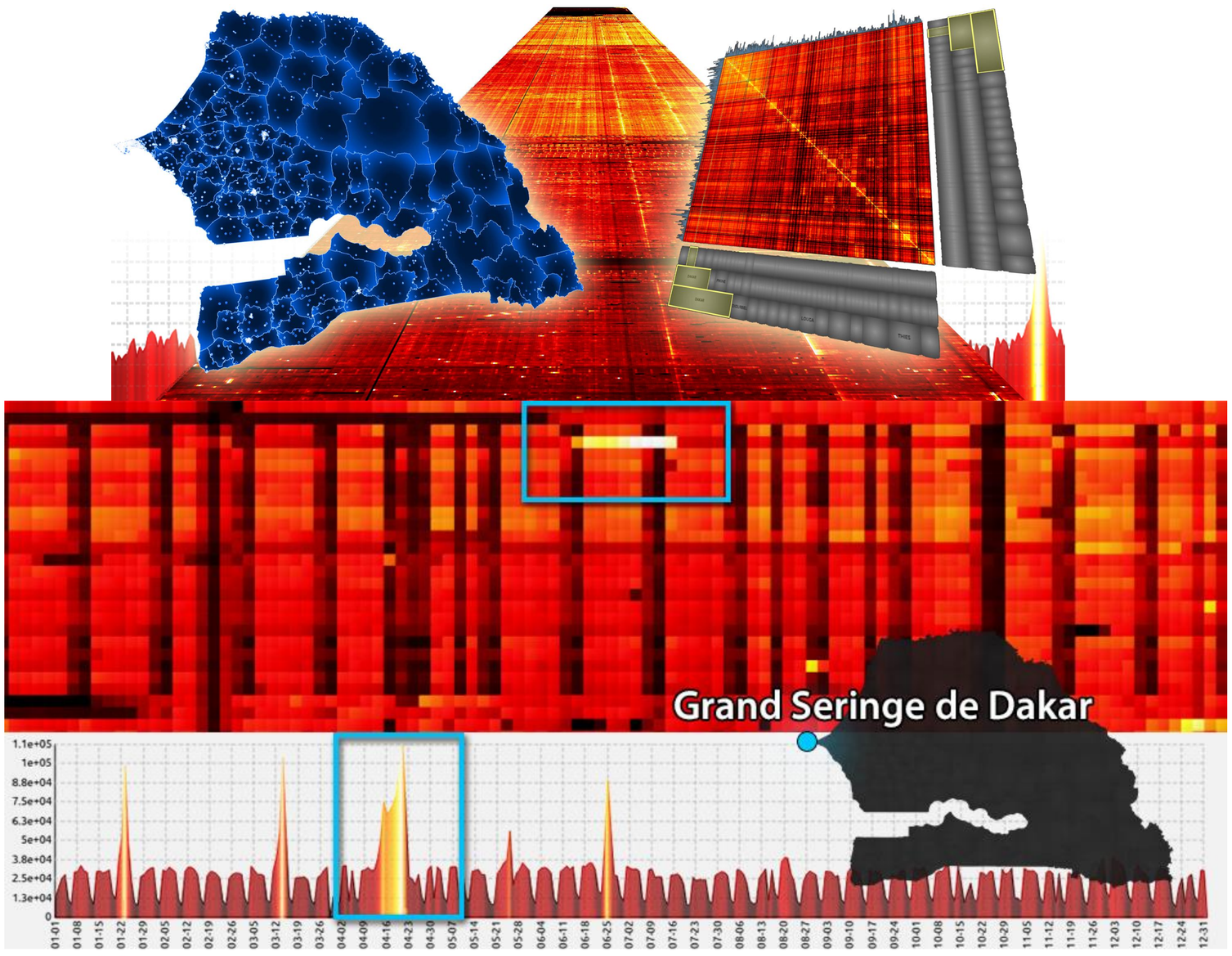
Open Code available:

- Yes ■
- No □

O05

Data for Development Reloaded: Visual Matrix Techniques for the Exploration and Analysis of Massive Mobile Phone Data

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



Stef van den Elzen[12], Martijn van Dortmont[12], Jorik Blaas[2], Danny Holten[2], Willem van Hage[2], Jan-Kees Buenen[2], Jarke J. van Wijk[1], Robert Spousta[3], Simone Sala[3], Steve Chan[3], Alison Kuzmickas[3]

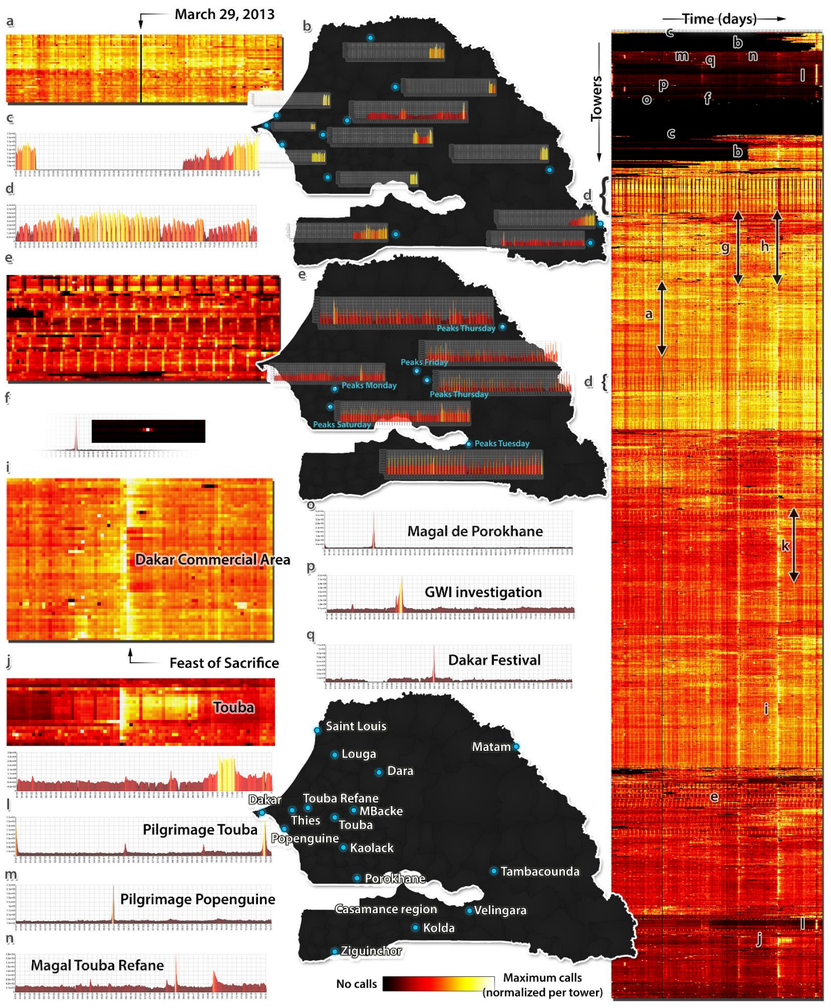
(1) Eindhoven University of Technology (2) SynerScope BV (3) Sensemaking Fellowship

Project Summary:

We present visual analytics techniques for the exploration and analysis of massive mobile phone data. We use a multiple coordinated view approach with a scalable and flexible visual matrix as central element to our solution. Users are enabled to identify both temporal and structural patterns such as normal behavior, outliers, anomalies, periodicity, trends and counter-trends. From this data we extract and discuss different patterns such as global events, weekly recurring events, regional patterns and outlier events.

Possible use for development:

The visual analytics methods are implemented in a prototype and applied to the provided data to enable and support users in the discovery of global and local patterns, outliers, trends, counter-trends, periodicity and anomalies. The insights gained in the exploration and analysis process can be used for better policy decision making.



Temporal matrix (right) with annotated events (left). A more detailed description can be found in the paper.

Main results:

- We developed a highly interactive prototype system for the exploration of massive mobile phone data in context of the D4D challenge. Using a visual matrix we provide and discuss techniques for the discovery of patterns. We found amongst others:
  - Increase in number of calls correlating with local and global religious events, such as Pilgrimage to *Touba* and *Popenguine* and the end of Ramadan and the Feast of Sacrifice.
  - Towers activated or deactivated throughout the year.
  - Week-weekend patterns for the identification of commercial areas.
  - Identification of Islamic and Christian areas.
  - Correlations of increased call intensity with the harvesting season and weather conditions influencing call intensity such as thunderstorms.

Methods:

- We choose a visual matrix as starting point for the exploration process due to its flexibility and scalability. Furthermore, we provide a multiple coordinated view solution with linked geographic and temporal views. The most important features are:
  - Providing flexibility of attribute projection on both axes.
  - Color-mapping.
  - Hierarchical aggregation
  - Normalization and clustering
  - Summarizing histograms
  - Interaction
  - Coupling with other visualizations.

Full paper is here:

http://tinyurl.com/d4d2014reloaded

Data sources used for this project:

- D4D data set 1, com between antenna [X]
- D4D data set 2, movement routes high res [ ]
- D4D data set 3, movement routes low res [ ]
- D4D synthetic data set [ ]

Other data sets used in this project:

- Type of data: International Disaster Database — Source: http://www.emdat.be/database
- Type of data: historical weather data — Source: http://wunderground.com/history
- Type of data: various online news media — Source: internet

Main Tools used:

- Qt/C++
- Various algorithms (clustering, normalization)

Open Code available:

- Yes [ ]
- No [X]

O06

Adaptive Power Load Balancing in Cellular Networks

Timothy Larock, Xiaojun Feng, Mariya Zheleva, Petko Bogdanov
Department of Computer Science, University at Albany SUNY

UNIVERSITY AT ALBANY
State University of New York

Motivation

- Projections on Internet traffic demand indicate that the traffic in 2018 alone will be as much as the sum traffic from 1984 to 2013.
  - Majority will be handled by mobile networks.
- Mobile networks demand is often unevenly distributed in space.
  - Spatial regularization through load balancing.
  - Current methods are *reactive*, risking deteriorated user experience
- We design *proactive* load balancing techniques based on manipulation of emitted power.
  - Model the network coverage as a power diagram, a generalization of voronoi diagrams.
  - Evaluate the feasibility of this approach on a real-world network: Orange™ in Senegal.

Contributions

- Using real-world data from Senegal, we demonstrate the existence of **significant disparity of load across towers** in a cellular network over time.
- We introduce a novel approach based on power diagrams for **proactive re-distribution of users**, designed to minimize cell tower overload and maximize utilization.
- We perform extensive evaluation of our spatial load balancing approach and demonstrate that it has the **potential of improving the operation of the existing network** in Senegal as a concrete example.
- We provide an **extensive discussion** of the implication of our approach to both commercial cellular networks, but also ones deployed in remote areas and for the purposes of disaster relief.

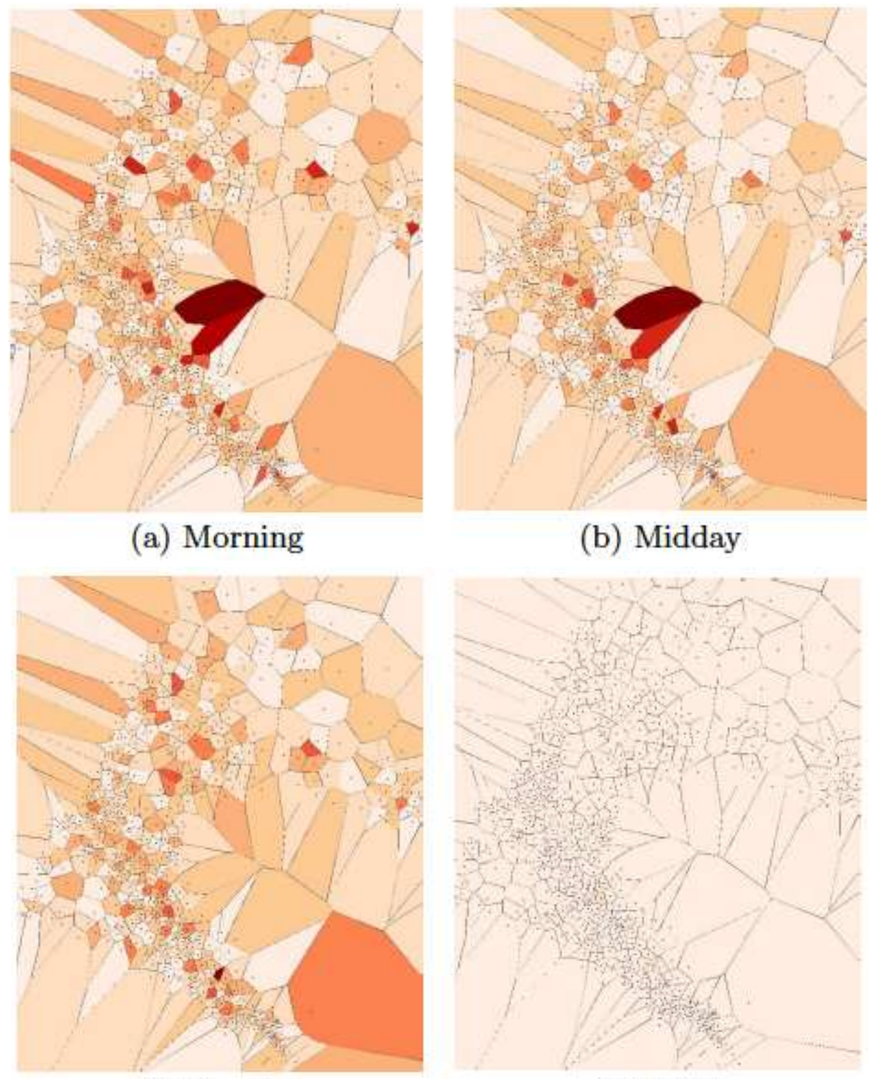
Discussion

- Implications on commercial cellular network deployments.
- Opportunities in community cellular networks in rural or disaster areas.

Data

Our analysis is based on cellular network traces provided by Orange™ in Senegal for the D4D Challenge.

- **Orange™ dataset**
  - High level antenna information collected over the month of January 2013.
  - Perturbed antenna location information.



(a) Morning (b) Midday

(c) Evening (d) Night

Voronoi diagrams showing cells in the region of Dakar during four periods of the day. The diagrams are colored based on the average number of calls in each cell during each time period compared to the maximum average call value observed thus far. Darker red color in a cell corresponds to a higher number of calls.

It is evident that there are multiple instances of neighboring cells of different loads.

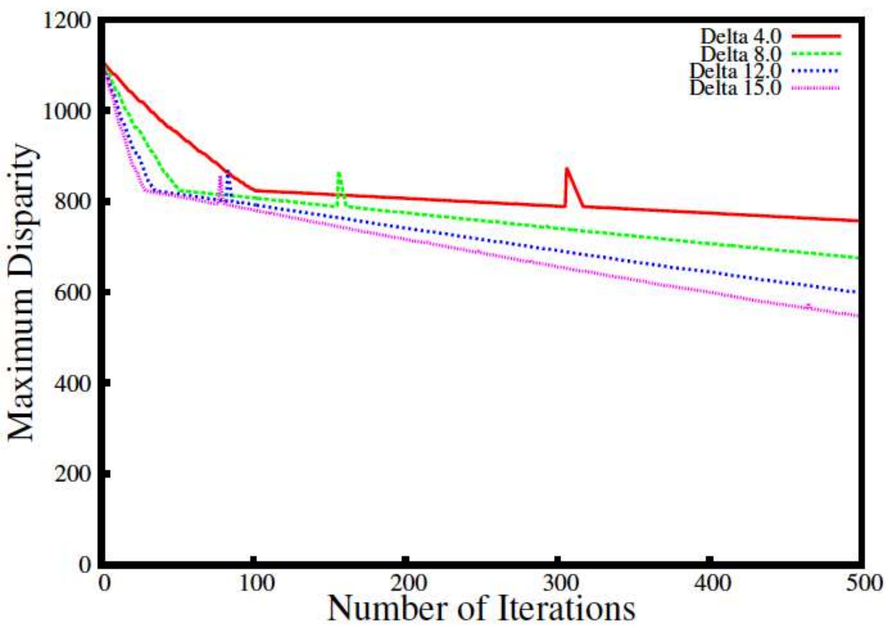
Adaptive Power Assignment

Intuitively, our Adaptive Power Assignment (APA) algorithm identifies cells whose load is very different than that of its neighbors (dubbed high **discrepancy** cells) and updates their power so that the load spreads within the neighborhood.

Discrepancy: $$\delta_i = \frac{\sum_{j \in N_i} |V_i - V_j|}{|N_i|}$$

Algorithm:
Input: Set of N sites, site loads V, atomic unit of power increment Δ
Output: Optimal power distribution

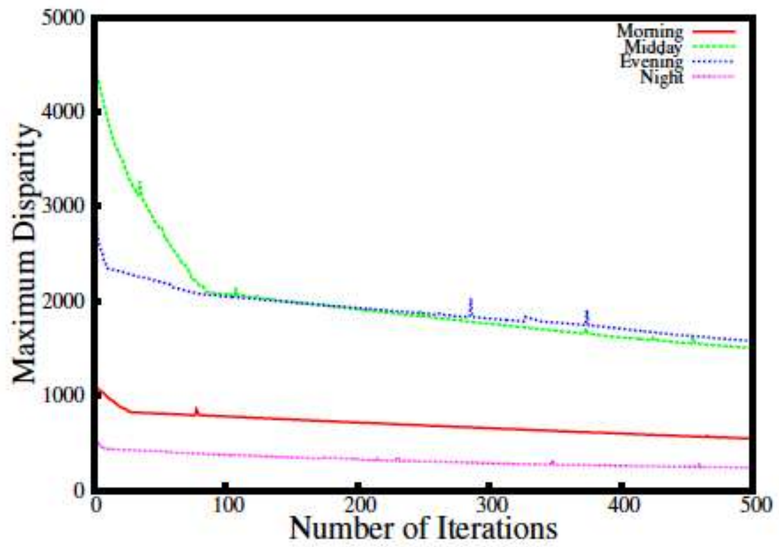
- Find the highest discrepancy site
- Compute an updated power diagram
- Continue up to *maxit* iterations.

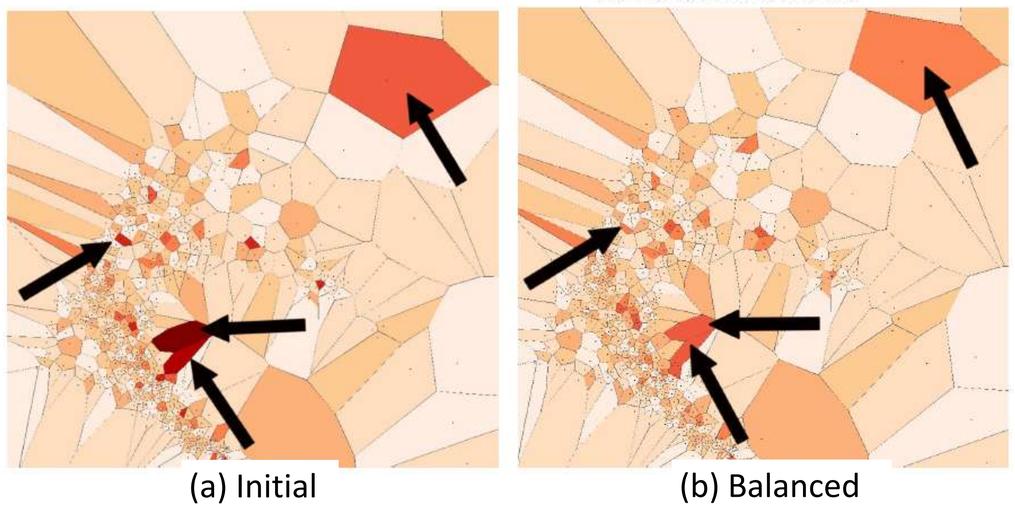


Change in maximum disparity in average calls over 500 iterations of the APA algorithm.

Results

Change in maximum disparity in average calls over 500 iterations at different times of the day, each using delta 15.0.





(a) Initial (b) Balanced

(a) Initial Voronoi diagram representing the average call load of sites in the Dakar region in January 2013. (b) Power diagram representing the same load and sites after 500 iterations of the greedy APA algorithm with Δ=8.0

O07

Senegal Time-Space 2013

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network

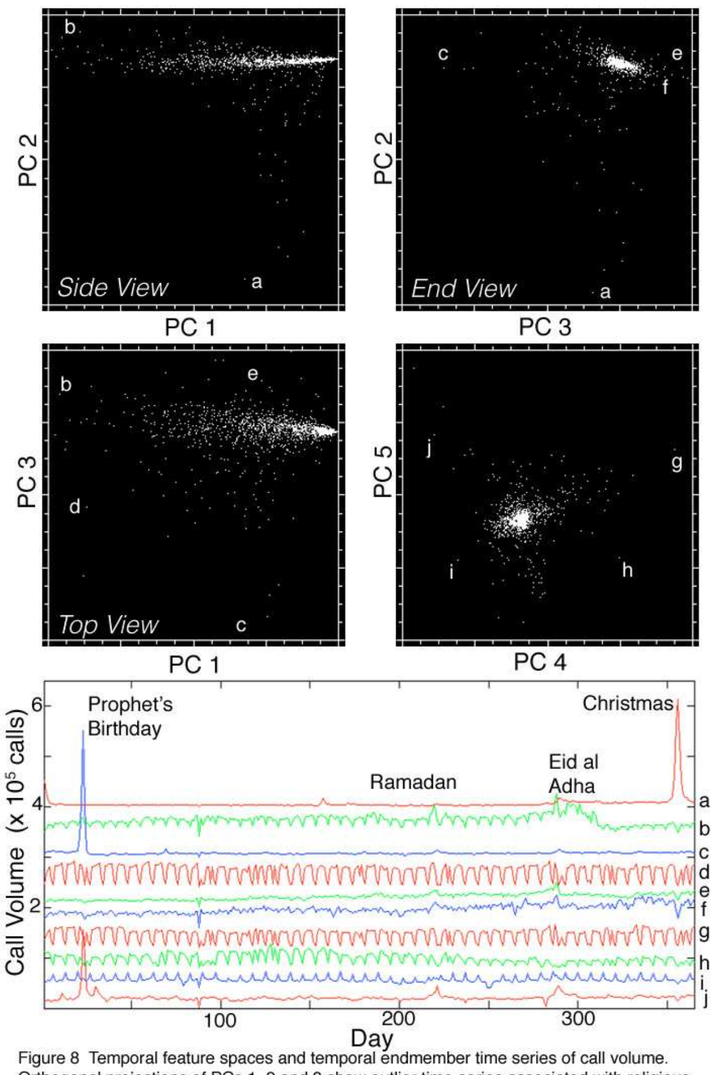


Figure 8 Temporal feature spaces and temporal endmember time series of call volume. Orthogonal projections of PCs 1, 2 and 3 show outlier time series associated with religious holidays (a,b,c). The space of PCs 4 and 5 shows strong weekly periodicitiies.

- Small, Christopher, Professor, Columbia University
- Kirchberger, Martina, Post-Doctoral Research Fellow, Columbia University

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

Project Summary:

We use spatial and temporal correlation matrices and Empirical Orthogonal Function (EOF) analysis to characterize spatiotemporal dynamics of communication to infer collective behaviour from call and text volume data. The spatial and temporal correlation matrices identify temporal and spatial disruptions to normal communication patterns. The EOF analysis identifies the spatial and temporal patterns that most concisely represent the dominant features in the data. The temporal feature space of the low order Principal Components provides a simple representation of the diversity of temporal patterns of call and text communication. We illustrate the use of these analytical tools to highlight spatiotemporal differences between calling and texting. The topology of the feature space clearly distinguishes between high volume weekly periodicities associated with developed urban areas and low volume aperiodic communication associated with rural populations.

Possible use for development:

Objective identification of normal spatial and temporal patterns of communication provides a basis for identification of abnormal patterns and disruptions. Our analysis quantifies the magnitude of spatial and temporal extent of travel patterns on holidays and night time activity during Ramadan. We also map high volume weekly periodicities associated with commuter destinations.This may facilitate transportation planning.

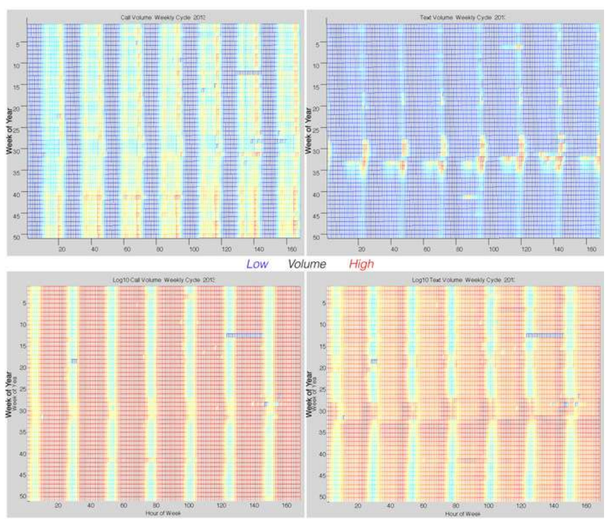


Figure 3 Calendar plots of call and text volumes. Daily and weekly cycles dominate both call and text volumes. Call volumes drop during Ramadan (weeks 27 to 30) but night text volumes increase and persist for weeks after the end of Ramadan.

Temporal Calendar Plots

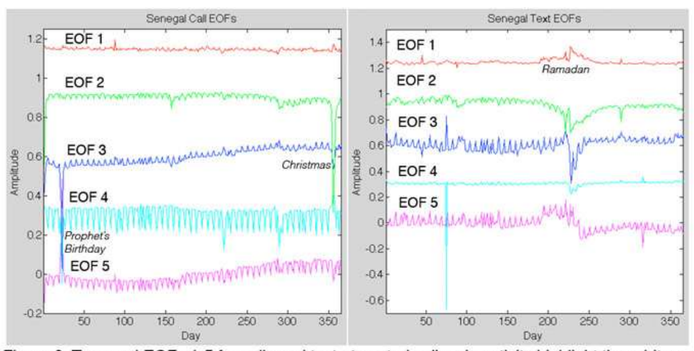


Figure 6 Temporal EOFs 1-5 for calls and text. Inverted spikes in activity highlight the arbitrary polarity (sign) of individual EOFs and corresponding PCs. Mixed polarity of same event spikes in different EOFs illustrates how constructive and destructive interference among dimensions modulates individual features. Call EOFs show strong influence of Christmas and the Prophet's birthday while text EOFs show strong influence of Ramadan and holidays in the weeks following. Both calls and texts show weekly cycles distributed over multiple EOFs, illustrating the difficulty of interpreting individual EOFs independently.

Temporal EOFs for calls and texts

Main results:

- Spatial correlation matrices identify the Prophet's Birthday and Christmas as the two largest distruptions to normal call & text communications patterns in 2013. Strong increase in night text volume during Ramadan.
- Temporal correlation matrices identify Dakar and Touba as having distinct temporal patterns of call and text volumes from rural areas and each other.
- EOF analysis & temporal feature spaces identify high volume weekly cycles in Grand Dakar – possibly resulting from large transient commuter populations.

Methods:

- We use spatiotemporal correlation and eigenstructure to characterize normal and anomalous patterns of call and text volumes.
- Temporal feature space of spatial PCs shows continuum of temporal EOFs contributing to observed temporal patterns.
- Subspaces of temporal feature spaces identify specific types of temporal patterns resulting from constructive and destrutive interference of temporal EOFs.

Full paper is here:
put your link here

Data sources used for this project:
- D4D data set 1, call/text between antenna [X]

Other data sets used in this project:
- Type of data: VIIRS+OLS night light [X]

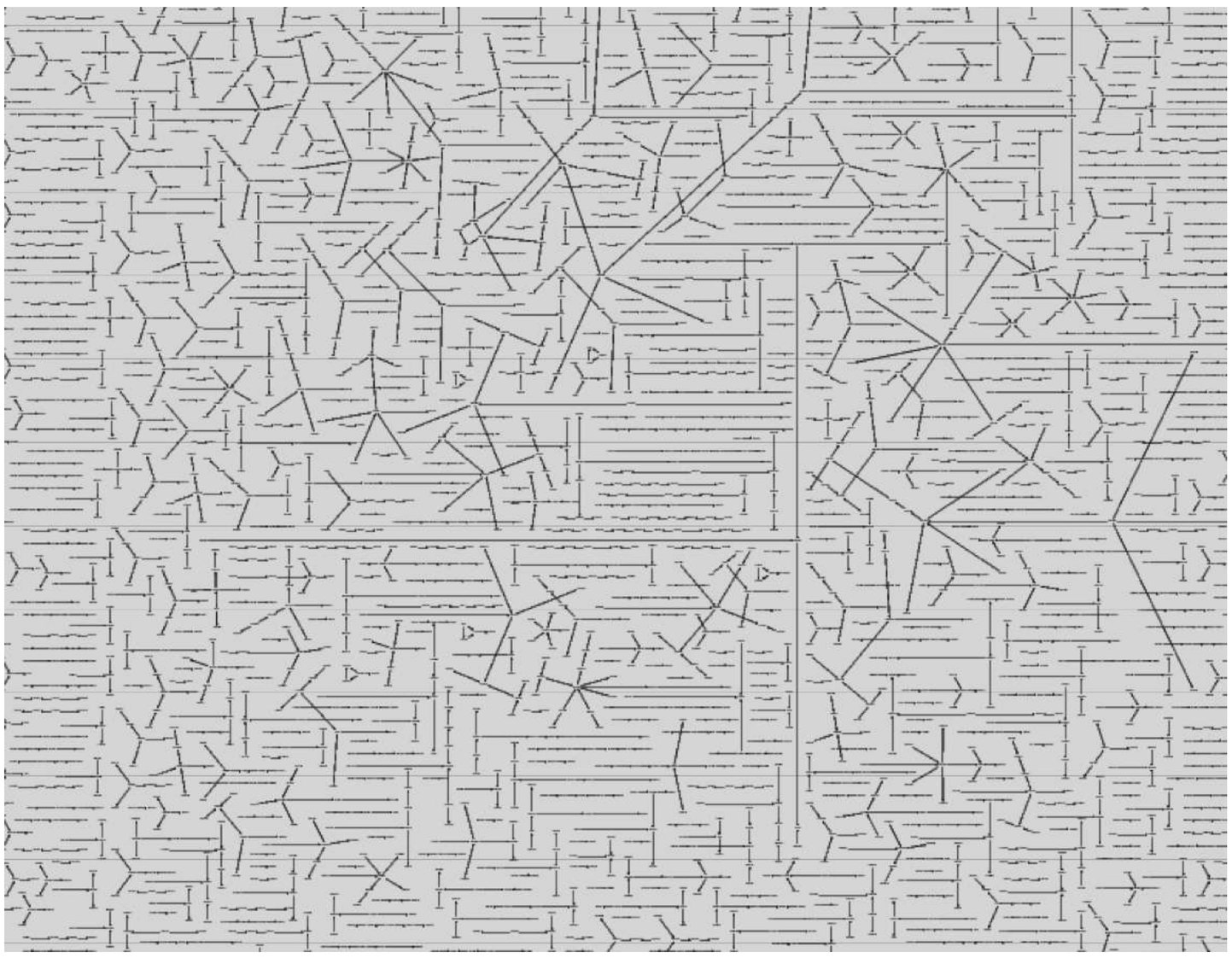
Main Tools used:
- Tool 1: Matlab
- Tool 2: ENVI/IDL
- Tool 3: EOF analysis

Open Code available:
- Yes [X]
- No

O08a
O08b
O08c
O08d

Call data network structure, revealed mobility & social networks

Health	Transport Urban	National Statistics	Other
Agriculture	Energy	DataViz	Network



UC San Diego

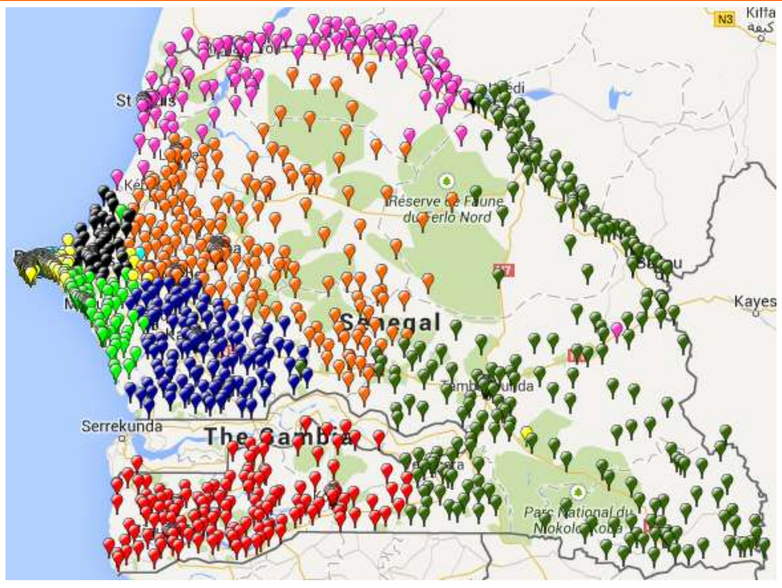
- D. A. Meyer, UC San Diego
- O. Bucicovschi, L. Fedus, D. Song, M. Gartner, A. Georges, D. Rideout, R. Douglass, T. Yu, UC San Diego

Project Summary:

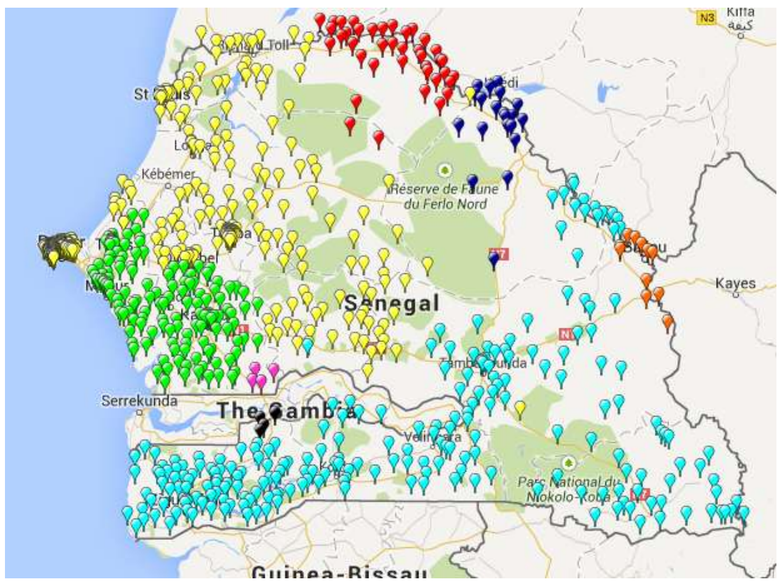
We apply topological data analysis to Sénégal's call data network; we develop a new method for extracting high resolution mobility information; and we show how to combine the three sets of data provided to obtain glimpses of the social network among Sonatel's users.

Possible Development Uses:

The global structure of the call data network reflects local communities, but also the transportation infrastructure, potentially identifying infrastructural gaps. High resolution mobility estimates can support such identifications. They can also parametrize epidemiological models and provide information about internal migration. Social networks provide more detailed information about local communities.



Clusters from modularity maximization



Clusters from zeroth homology

Main results:

- Persistent homology captures transportation infrastructure features.
- Homological clustering is less useful than communities obtained by modularity maximization.
  - Neither changes significantly with opening of Dakar Motorway.
- Population flow can be estimated at 10 min or greater time intervals.
  - Use to parameterize epidemiological models.
- User IDs have been cross-identified between datasets
  - Potential check for synthetic data sets.
- Partial reconstruction of social network of Sonatel users
  - Provides more detail than community detection.
- Call volumes can be predicted from previous three days' data.

Methods:

- Topological data analysis, specifically persistent homology using javaPlex
- Modularity maximization via Louvain algorithm
- Parameter estimation for time dependent Markov process using an expectation maximization algorithm
- Timeseries prediction with genetic algorithms

All reports are here:
http://tinyurl.com/n2389st

Data sources used for this project:

- D4D data set 1, com between antenna [X]
- D4D data set 2, movement routes high res [X]
- D4D data set 3, movement routes low res [X]
- D4D synthetic data set [ ]

Other data sets used in this project:

- GIS data on arrondissements Source: D4D
- Cultural/historical information Source: various
  bookssenegal_jul.png

Main Tools used:

- javaPlex
- Graphviz
- Persistent homology
- Expectation maximization

Open Code available:

- Yes [ ]
- No [X]