



**HAL**  
open science

## Spatial structure and efficiency of commuting in Senegalese cities

Rémi Louf, Giulia Carra, Hadrien Commenges, Jean-Marie Dembele, Riccardo Gallotti, Maxime Lenormand, Thomas Louail, Marc Barthelemy

► **To cite this version:**

Rémi Louf, Giulia Carra, Hadrien Commenges, Jean-Marie Dembele, Riccardo Gallotti, et al.. Spatial structure and efficiency of commuting in Senegalese cities. NetMob'15 - The main conference on the scientific analysis of mobile phone datasets, MIT Media lab, Apr 2015, Boston, United States. pp.110. hal-03826984

**HAL Id: hal-03826984**

**<https://hal.science/hal-03826984v1>**

Submitted on 24 Oct 2022

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

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

# Spatial structure and efficiency of commuting in Senegalese cities

Rémi Louf<sup>1</sup>, Giulia Carra<sup>1</sup>, Hadrien Commenges<sup>2</sup>, Jean-Marie Dembele<sup>3</sup>,

Riccardo Gallotti<sup>1</sup>, Maxime Lenormand<sup>4</sup>, Thomas Louail<sup>1</sup>, Marc Barthelemy<sup>1,5</sup>

<sup>1</sup> *Institut de Physique Théorique, CEA-CNRS (URA 2306), F-91191, Gif-sur-Yvette, France*

<sup>2</sup> *Laboratoire Eau, Environnement et Systèmes Urbains,  
Ecole des Ponts, F-77455, Marne-la-Vallée, France*

<sup>3</sup> *Université Gaston Berger, UFR SAT, Saint-Louis, BP 234, Senegal*

<sup>4</sup> *IFISC, Instituto de Física Interdisciplinar y Sistemas Complejos (CSIC-UIB),  
Campus Universitat de les Illes Balears, E-07122 Palma de Mallorca, Spain and*

<sup>5</sup> *Centre d'Analyse et de Mathématique Sociales, EHESS-CNRS (UMR 8557),  
190-198 avenue de France, FR-75013 Paris, France*

Senegal is experiencing an unprecedented urban revolution: according to the latest UN projections, its urban population will be multiplied by 3 in the forthcoming decades, to reach 18 million people by 2050. While cities are often lauded as the solution to mankind's socio-economical and environmental issues, when badly managed, they can also be recipes for disasters. In order to propose well-informed transport and planning policies in Senegal, it is crucial to first measure and understand the key spatial dynamics that shape its cities. In this report, we uncover commuting patterns in the 12 largest urban areas of Senegal and their coarse-grained spatial properties using mobile phone data, allowing us to characterize the efficiency of commuting and to compare cities with each other. At the inter-urban level, we show that for most cities, the vast majority of commuters live and work in the same urban area, meaning that Senegalese cities are well-integrated employment markets. We then compute the geographical area of influence of each city. This confirms the importance of large cities such as Dakar, but also highlights smaller cities which play an important economical role such as Tambacounda or Touba. At the intra-urban level, we quantify the spatial mismatch between residence and workplace locations and we propose a measure of the 'optimality' of the commuting structure. We find that Dakar, with a high optimality index, has a coherent spatial structure with nested residential and employment areas which is reflected in the fact that 80% of the residential and activity hotspots overlap. Smaller cities however—such as Louga, Kolda, Mbour—are far from the optimal commuting solution. The methods proposed in this study could help urban planners in identifying locations and areas which are the most penalized by inefficient commuting, a source of economic loss and stress on people's life and the environment.

## Introduction

### Urban transition in Senegal

Senegal, along with a large part of African countries, experiences a rapid economic growth with an average of 3% GDP growth per year [1]. This economic growth goes hand-in-hand with a dramatic increase of its urban population : while Senegal is estimated to currently host 6 millions people in its cities, the UN estimates that this number should be multiplied by 3 by 2050. It is therefore an understatement to say that Senegalese urban areas are changing fast. The current transition carries a lot of promises in terms of development; yet, badly managed, it could as well lead to a socio-economic and environmental disaster. In order to make useful investments in transportation infrastructures and propose well-informed planning policies, policy-maker need to understand which areas are going to develop, at what rate, and quantify the ties between different neighbourhoods (at the intra-urban level) and different urban areas (at the inter-urban level).

In general, rapid urbanization goes along with unplanned settlements, health and sanitary problems, as well as congestion problems. In particular, a key in-

redient of sustainable cities is an efficient organization of commuting. Non-pedestrian commuters in Senegal travel almost exclusively by cars, and the longer the commuting, the larger the pollution and  $CO_2$  emissions are. Dakar urban and industrial areas expanded in direction of Thies, and zones of informal settlements appeared. Poor people are pushed toward the periphery which is usually under-equipped and with much lower level of transport infrastructure. It is therefore important to identify areas and trips with large volumes and strong demand for a better, targeted planning. Such information is however usually difficult to obtain and it has been almost a decade now that scientists have realized that geolocated traces passively generated by individuals' ICT devices could revolutionize the quantitative and theoretical understanding of human spatial dynamics, and urban dynamics [5]. To give a few examples of urban phenomena whose understanding has been enhanced in the recent years through renewed quantitative approaches applied to new sources of ICT data, we can mention the statistical and spatial properties of individuals' mobility in cities [6–10]; the universal structure of subway networks and streets networks [11, 12]; the number and the spatial organisation of centers in urban areas [13, 14];

the spatial properties of social networks in countries and cities [15, 16]; and the scaling of diverse quantities with the population size of the city [17–19].

For a few years now, we can investigate these urban questions in much more detail than during previous decades, by analyzing vast amounts of data available at different spatial and temporal scales. Problems that in the past were addressed through surveys by geographers and transport scientists are nowadays addressed by interdisciplinary teams, with many different data sources, data that are more precise both spatially and temporally. Sixty years after its emergence as an academic research field, spatial analysis and quantitative geography may be living their second quantitative and theoretical revolution.

In this paper, we will present an example of such studies, and by using mobile phone data recorded in Senegal, we will show by proposing new measures how we can extract useful information about the spatial structure of urban areas and commuting.

### Mobile phone data and urban mobility

*Limitations of classic data sources* Traditionally, urban planning has relied on travel surveys and censuses. However these data sources have several limitations:

- They require an important logistic, time, and are expensive.
- For these reasons, they are performed over long time intervals, typically every decade or so. With such data it is then impossible to estimate changes of the urban structure over short time scales.
- Transport surveys are often based on samples of a view thousands of individuals only.

In contrast, individual mobile phone data provide anonymized location information about a large fraction of the population with a temporal resolution below 24h, and with a spatial resolution which depends on the density of antennas. For most cities this resolution is of the order of a few hundred meters in the centers, and 1 to 2 kms in the peripheric neighbourhoods. In Senegal the mobile phone penetration rate was estimated to be about 85% in mid-2013 and projections estimated a 110% figure at the end of 2014 [40]. It thus allows to monitor population displacements, such as the daily journey to work, important events implying many travels such as Touba’s Magal, or long term residential migrations (mostly rural to urban migrations).

Interestingly, a recent study in Madrid and Barcelona – the two largest Spanish urban areas – demonstrated that new sources of mobility data (mobile phone data in the first place, but also geolocated tweets) provide at the city scale a very comparable picture of the commuting structure, when compared to the information obtained with transport surveys [4]. This result opens the door

to a more systematic use of new sources of ICT data to work on mobility issues in cities.

*Previous studies of urban mobility in developping countries with ICT data.* For obvious structural and economical reasons, most of the ICT-related studies of the last decade have focused on cities located in rich and developed countries. Few results have focused on cities of other regions of the world and other continents, notably Eastern Europe, South-America and Africa. Several important papers have claimed to uncover universal mobility and urban patterns and propose general models, but are so far limited to a small number of geographical areas [6, 18, 24]. Continents and countries have however different urbanization histories, exhibit different spatial properties such as densities and spatial organization [21, 22], and the universality of urban patterns is not yet proved to be correct for all regions of the world. To this day, Africa is the less urbanized continent, and is currently experiencing a very fast urban transition [23]. Most large African cities, including Dakar, have no subway network, bike-sharing or car-sharing systems, and their highways systems are much less developped than in US, European or Asian cities of the same size. Apart from the informal collective transport (‘bus rapides’) and the numerous taxis, public collective transport in Dakar is provided by the municipality bus service (‘Dakar Dem Dikk’). It is currently an important issue to obtain a better spatial knowledge of trips in order to develop this service. For these reasons, measuring and comparing the spatial properties and the structure of intra-urban mobility in Senegal is particularly important both for Senegal and Dakar urban planning questions, and also on the scientific side for the elaboration of the emerging science of cities. Such a quantitative knowledge could help to guide planning policies of rapidly urbanizing areas.

The previous edition of Orange’s D4D challenge provided communication datasets in Ivory Coast, and allowed for quantitative studies of mobility patterns in Abidjan and Ivory Coast. Numerous interesting studies related to transport and mobility were performed. For example Kung et al. [25] used the data to test the long-lasting hypothesis of a universal, fixed time-budget for daily mobility (often refered as ‘Zahavi’s law’). They obtained surprisingly large commuting times, and their results questioned the possibility of using individuals mobile phone data to infer travel duration [25]. Berlingerio et al. [26] developed an interactive and modular application ”to optimize the public transport network, with the goal to improve ridership and user satisfaction”. They used individuals’ travel and activity patterns detected in the data to extract origin-destination (OD) matrices and individuals’ travel preferences, to determine optimal design of potential new transit services. A number of studies proposed methods to automatically extract OD matrices from individual data, notably [27]. Wakita et al. analysed the temporal patterns of communication data and could infer the dominant type of land-use of the geographical area covered by each antenna, and proposed

maps of land-use at various scales in the country [29]. Andris and Bettencourt [28] applied network analysis on the communication network at three different scales (individuals, cities, and the whole urban system), drew the communication networks upon natural resource layers and discussed the resulting maps. While providing advices about possible future developments, they didn't characterize the attractiveness of cities, nor their spatial organization and its interplay with mobility patterns of individuals. In addition, none of these 2013 projects proposed a comparison of cities based on the spatial properties of commuting, especially from an efficiency perspective. Such a comparison could help urban planners in identifying cities which are the most penalized by inefficient commuting, a source of economic loss and stress on people's life and the environment.

### Objectives of the study and organization of the report.

The objective of our project is to extract from mobile phone data the intra-urban and inter-urban commuting patterns of individuals, discuss these patterns, and to provide a set of quantitative characterizations of the organisation of journey-to-work mobility in Senegalese cities. Another important motivation is to develop coarse-grained indicators summarizing these large amounts of data, able to provide synthetic and large scale pictures of the structure of individual mobility in cities. Such meso-scale information is also particularly useful for validating synthetic results of urban mobility models (such as [30] for example), for comparing different cities and also for comparing different models. An accurate modeling of mobility is indeed crucial in a large number of applications, including the important case of epidemic spreading which needs to be better understood, especially at the intra-urban level [31, 32]. The recent case of the Ebola virus, that didn't spread in Senegal but probably had serious impact on international migrations and touristic activities in the country, provides a contemporary illustration of the societal usefulness of such an understanding.

In this report, we focus on commuting (journey-to-work trips) in Senegalese cities, which represents everywhere the largest part of the daily mobility. We have extracted the 12 largest cities from the map of Orange's antennas in Senegal (see Table I), and we have computed the origin-destination (OD) matrices from the twenty-five 2-weeks individual activity datasets (see Methods). We focused successively on two different spatial scales. We first studied the inter-urban case and characterized the relations of each city with the others. We evaluate their geographical area of influence ('attraction basin') and the degree of 'integration' of their labour and housing markets. We then focus on the intra-urban scale, and provide several measures of the spatial organisation of these cities. We first focus on residential and activity

hotspots, and discuss their spatial organisation and the shape of the Senegalese cities. We then analyse the structure of commuting in the cities, and propose an original method to characterize the efficiency of the commuting structure, by evaluating the spatial mismatch between residences and workplaces locations. Finally we discuss the results obtained, and the first hints in terms of planning.

## Results

The following analysis is based on origin-destination (OD) matrices extracted from the dataset 2 for the largest Senegalese cities. An OD matrix is a classic object in transport planning and mobility studies, and is a  $n \times n$  matrix  $F_{ij}$  where  $n$  is the number of spatial units that compose the city at the spatial aggregation level considered. The element  $F_{ij}$  represents the number of individuals living in location  $i$  and commuting to location  $j$  where they have their main, regular activity (work or school for most people). The general idea for extracting the OD matrices from individual mobile phone data is to determine for each individual the pair of locations (i.e. antennas) that are the best proxys for their home and main daily activity locations (see Methods for details). In order to reduce the noise as much as possible with the dataset at hand, the results presented here have been obtained by summing the 25 OD matrices of the dataset 2, each covering a 2-weeks period.

### Structure of inter-urban commuting

We first study interactions – in terms of commuting – between each of the major cities identified in the data, and between each city and the rest of the country.

For a given city  $i$ , we can distinguish three types of commuters:

- The *internal* commuters, who live in  $i$  and whose main daily activity is also located in  $i$ ;
- The *divergent* commuters, who live in  $i$ , and whose main daily activity is based outside  $i$ ;
- The *convergent* commuters, who live outside  $i$ , and whose main daily activity is based in  $i$ ;

In the following we denote by  $N_i^o$ ,  $N_i^{\rightarrow}$  and  $N_i^{\leftarrow}$  the numbers of internal, divergent and convergent commuters, respectively. The total commuting population  $C$  of the urban area (for the summed 25 OD matrices) is then given by the total number of internal commuters plus the total number of divergent commuters

$$C = N^o + N^{\rightarrow} \quad (1)$$

These three types of commuting are represented schematically on Fig. 1.

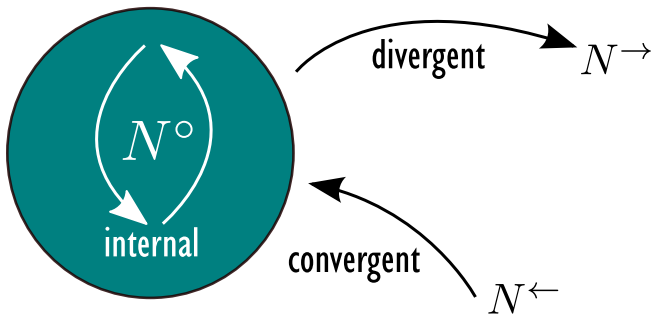


FIG. 1: Schematic representation of the different types of commuters: internal, convergent and divergent commuters.

	$C$	$N^\circ$	$N^\leftarrow$	$N^\rightarrow$	$f_0$	$f_1$
Dakar	3,099,116	3,049,579	47,540	49,537	0.03	0.95
Thies	261,050	248,950	10,718	12,100	0.09	0.88
Touba	230,498	218,987	13,980	11,511	0.11	1.21
Mbour	216,418	204,312	11,111	12,106	0.11	0.91
Kaolack	178,365	170,821	8,097	7,544	0.09	1.07
Saint-Louis	162,875	154,910	5,843	7,965	0.08	0.73
Ziguinchor	135,562	130,264	4,255	5,298	0.07	0.81
Tambacounda	65,577	62,900	3,049	2,677	0.09	1.13
Tivaoune	63,198	53,815	7,532	9,383	0.31	0.80
Diourbel	62,345	59,357	3,585	2,988	0.11	1.19
Louga	59,883	56,964	3,649	2,919	0.11	1.25
Kolda	57,895	55,676	2,304	2,219	0.08	1.03

TABLE I: Partition of commuting flows into internal, divergent and convergent commuting flows, for the 12 largest urban areas identified in Senegal (ordered by total commuting population).

The values measured for the 12 Senegalese cities we identified are given in Table I, ranked in decreasing order of commuting population size.

In order to estimate the relative importance of the different flows, we define the two following ratios

$$f_0 = \frac{N^\rightarrow + N^\leftarrow}{N^\circ}$$

$$f_1 = \frac{N^\leftarrow}{N^\rightarrow}$$

whose values are also given in the two last columns of Table I. The quantity  $f_0$  characterizes the importance of internal commuters, and  $f_1$  simply compares the number of convergent and divergent commuters.

For most cities studied here, convergent and divergent commuters represent roughly 10% of the internal commute, signifying that Senegalese cities are well-integrated employment markets. The smallest fraction of out-commuting is found in Dakar (3%) and the largest in Tivaoune (31%), which is an important religious center in Senegal, thereby explaining this large value.

The ratio  $f_1$  of the number of divergent and convergent

commuters allows us to divide Senegalese cities in three classes:

- Cities for which  $f_1 > 1$  - Diourbel, Louga, Touba, Tambacounda - where the number of ingoing commuters is larger than the number of outgoing commuters. This implies that there are more people present in the city at daytime when compared to nighttime.
- Cities with  $f_1 \approx 1$  - Dakar, Kaolack, Kolda - where the number of outgoing and ingoing commuters is roughly the same, meaning that the total population present in the city is basically the same during day- or nighttime.
- Cities with  $f_1 < 1$  - Tivaoune, Thies, Ziguinchor, Saint-Louis - where the number of outgoing commuters is larger than the number of ingoing commuters, also meaning that there are more individuals present in the city during the night than during the day.

This simple indicator  $f_1$  allows to illustrate the balance between jobs and residences. An ‘attractive’ city ( $f_1 > 1$ ) for example means that the job/activity market is larger than the residential offer. In contrast, a city with  $f_1 < 1$  is in majority residential. We note however that, surprisingly, there are very few differences between cities, and that the ratios are almost all of order 1, suggesting an equilibrium between the number of people that live outside a city and spend the day in it, and the number of people who live in a city and spend the day outside.

We can further characterize the convergent commuters by measuring the average distance they are traveling. We call this distance the *attraction radius*  $r$  of the city, computed as

$$r^\rightarrow = \frac{1}{N^\leftarrow} \sum_n \ell_n \quad (2)$$

where  $\ell_n$  is the distance traveled by the  $n$  convergent commuters. This quantity characterizes the regional influence of cities in terms of commuting and job/activity market. The radius for Senegalese cities are shown on Fig. 2, and their values listed in the Table II.

Dakar has a large zone of influence – as expected – however there are some surprises such as Tambacounda which is a small city but displays a large attraction radius. It is interesting to note that this large value probably reflects the fact that Tambacounda is an important commercial stop on the road to eastern countries such as Mali and to Casamance, and for stock trading.

We believe that this first study of convergent and divergent commuting should be interesting for planning purposes, but we note that the available data here are too sparse to reach very reliable conclusions. Indeed, when extracting journey-to-work OD matrices (see Methods), we realized that on a 2-weeks period, there is an important proportion of individuals whose mobility is not reg-

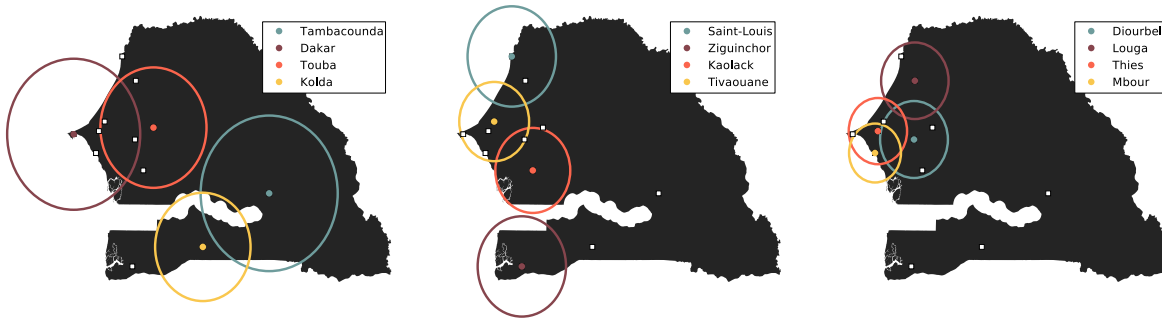


FIG. 2: **Attraction radius for the major Senegalese cities.** We represent the radius of attraction computed using the OD matrices that we extracted from the data. The radius for different cities are shown on three different panels for a matter of clarity.

	$r^{\leftarrow}$ (km)
<i>Dakar</i>	137
Thies	60
<i>Touba</i>	109
Mbour	53
Kaolack	77
Saint-Louis	91
Ziguinchor	91
<i>Tambacounda</i>	141
Tivaouane	72
Diourbel	69
Louga	69
Kolda	98

TABLE II: **Attraction radius  $r^{\leftarrow}$ .** Cities are ordered by total number of commuters, showing that there is no clear correlation between the attraction radius and the total number of commuters.

ular enough to infer with confidence their home and activity locations (see Methods). To inspect these aspects further, we would need the data provided in dataset 2, but available for a longer period of time for each individual. The dataset 3 could allow to solve this problem, but its spatial resolution is much scarcer, and the small number of individuals per city (Dakar excepted) would raise other difficulties. We thus leave this investigations for further studies and additional datasets. In the following we will focus on what the journey-to-work OD matrices can teach us regarding the intra-urban mobility and the spatial structure of Senegalese cities.

### Intra-urban organisation of Senegalese cities

*Spatial structure of hotspots* A first interesting look on a city’s structure is provided by the locations of hotspots, which are local maxima of the density of individuals. Using the extracted OD matrices, we can

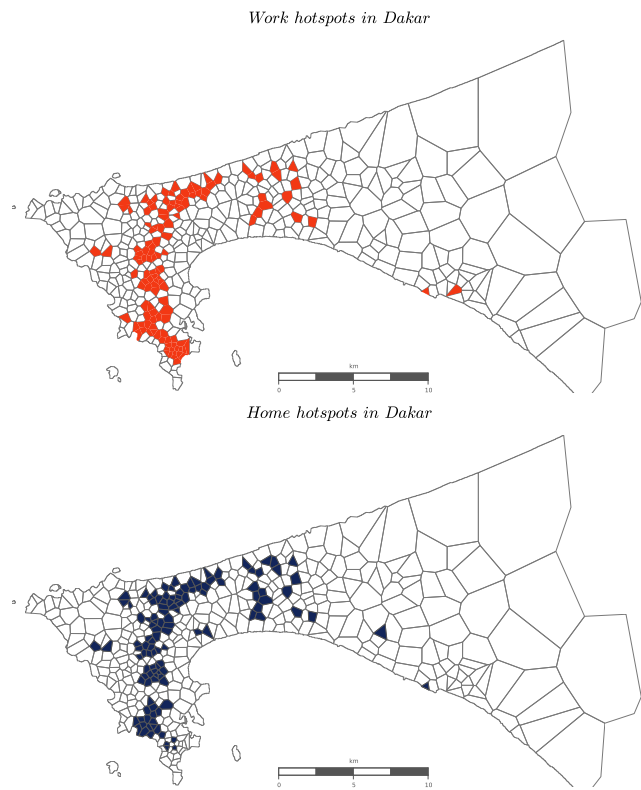


FIG. 3: Location of residential and activity hotspots in Dakar, determined with the method described in [14].

identify the most important residential and daily activity areas. In order to determine hotspots we used the ‘LouBar’ method based on the Lorenz curve of the distribution of densities [14]. For most cities the hotspot spatial structure is simple and monocentric. Only for the city of Dakar we have an interesting organization shown on Fig. 3, where we see a polycentric structure as it is observed for other large cities in the world [13].

We show in this figure 3, the map of residential hotspots (i.e. with large night activity) and job/activity hotspots (i.e. daily activity). The overlap between these

maps is large which means that important residential areas in Dakar are also the important daily activity centers. Our spatial delimitation of the urban area of Dakar (see Methods) includes major surrounding municipalities (Pikine, Rufisque), and some parts of these municipalities are indeed the most important crowded parts of the urban area. The hotspots of Dakar are not all along the coast but are located in internal crowded neighbourhoods (Liberté, Derklé, Sacré-Coeur, Medina, etc.).

The ‘work’ hotspots identified in the urban area of Dakar (26% of all the antennas in the city) contain 37% of the total daily population (as identified by our OD matrices). The ‘home’ hotspots (24% of all the antennas in the city) on the other hand contain 39% of the total population identified by our OD matrices (over the whole year). We note that rich residential neighbourhoods (Les Almadies, Le Plateau), which are also the areas where most of the administration and buildings of the major companies are located, are not residential hotspots. For example one can see on the work hotspots map that some areas of ‘Le Plateau’ are indeed tagged as employment/daily activity hotspots, but not as residential hotspots.

Surprisingly, home and work hotspots overlap then quite well in the city of Dakar: 80% of the work hotspots are also home hotspots. It gives us a first intuition that in Dakar, the commuting distances could be quite short in average, and that the city seems coherently organized, in the sense that many individuals don’t have to travel long distances with cars everyday to travel from their home to their workplace.

*Mobility of individuals within cities: hotspots and organization level* Once we have extracted the commuting network of individuals we can investigate the spatial properties of this network. We first measure the distance daily traveled by individuals to go from their home to their main activity place. These distances for Dakar and Thies are mapped on Figure 4.

These two figures reveals two typical forms of organization of cities. In the case of Dakar, we observe a polycentric structure with individuals traveling longer distances as they live further from the main activity hotspots (see figure 3), but also with secondary activity centers appearing in the suburbs (Rufisque, Pikine), where many people live and work, resulting in shorter commuting distance on average. Still in Dakar, while several secondary activity centers have developed as the city expanded to Rufisque and Pikine, the historical center remains the most influent and attracts more commuters from these secondary centers than the opposite (see Figure 5). In contrast to this polycentric structure, we observe for the smaller city of Thies a clear monocentric structure with a unique central zone that attracts most daily activity. Consequently, individuals that live in the center commute over short distances, while the further people live from the center, the longer their daily commuting distance is.

An intriguing question is if we can characterize the level of organization of cities from the spatial structure

of commuting flows. The first metric that we have is the total commuting distance  $L$ . The value of  $L$  however does not have a clear interpretation by itself, but we can compare it to reference values. We will then compare  $L$  with the two extreme situations:

- A totally *disorganized* mobility structure, where commuting patterns occur at random, while conserving the static structure of the city given by the numbers of inhabitants and employees attached to each antenna. In this case, we obtain a total commuting distance denoted by  $L_R$ .
- An optimally (*organized*) mobility structure, for which mobility patterns are such that the total commuting distance in the city is minimum (while conserving the number of inhabitants and employees attached to each antenna). Given the constraints of numbers of inhabitants and employees, this leads to a minimum total commuting distance denoted by  $L_O$ .

In order to compute the values of  $L_R$  and  $L_O$  for each city, we calculate the corresponding OD matrices of the two extreme cases, optimal and random. The OD matrix for the disorganized state is calculated by adding flows at random, making sure to respect the in- and out- degree of each node in the network (see Methods). The distance  $L_R$  is averaged over 100 random realisations. The OD matrix for the optimal state is obtained by simulated annealing, a local search optimisation method [35]. Once we have calculated the two quantities  $L_R$  and  $L_O$  we can then define an *organization index* of the city

$$O = \frac{L - L_R}{L_O - L_R} \quad (3)$$

This index  $O$  is equal to 0 when the mobility patterns are completely disorganised ( $L = L_R$ ) and equal to 1 when the mobility patterns are completely organised ( $L = L_O$ ). This measure is indeed a measure of the *spatial mismatch* in the city: the larger the value, the more organized the city is, and individuals in this case live very close to their activity location. We give in Table III the values of  $L/L_O$ ,  $L/L_R$  and  $O$  for the twelve Senegalese cities identified by our urban areas detection method.

We observe in this table that for most cities the ratio  $L/L_R$  is small and approximately constant (on average equal to 0.25), the ratio  $L/L_O$  displays larger variation: on average we obtain 3.8 and extremes such as 8.98 (Tivaoune) and 1.92 (Kaolack). Although we would need the corresponding values of cities in other parts of the world, these numbers suggest that there is probably some room for improving the commuting figures in many cities in Senegal. In particular, in the case of Dakar, we obtain the values indicated in Table IV (as a reference we also give the typical distance of the city, taken as the square root of its surface  $\sqrt{A}$ ).

It is interesting to note that in the largest city of the country, the individuals journey-to-work mobility is

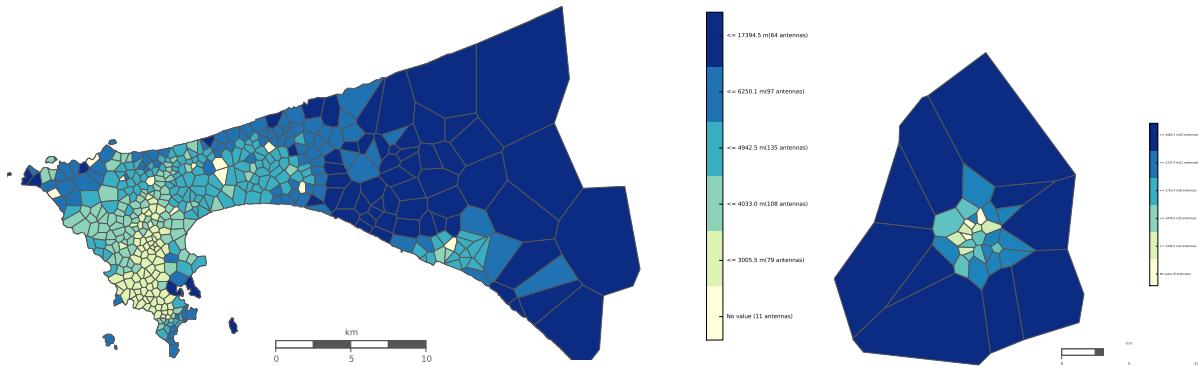


FIG. 4: **Average commuting distance at the antenna level** (Left) Dakar: Average outreach per antenna, i.e. average distance commuted by people living in the area covered by this antenna. When calculating the average we do not take into account the people who are flagged as living and working at the same place. (Right) Same measure for the urban area of Thies. These results are obtained by averaging the OD matrices over the whole year.

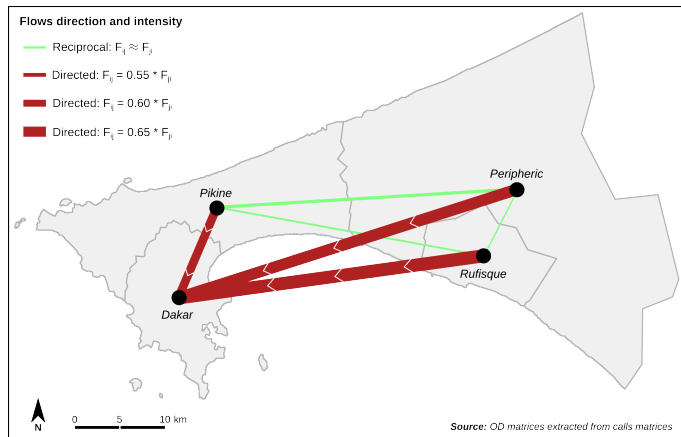


FIG. 5: **Main directed flows of commuters at the municipality scale, in the urban area of Dakar.** As the urban area expanded in direction of the surrounding municipalities, several secondary activity and employment hotspots appeared in the urban area of Dakar. Still the historical activity centers, located in the municipality of Dakar, attract more commuters that live in the surrounding municipalities, than the opposite.

rather short, and closer to the optimal situation than to the random, disorganized situation. This result suggests that there is indeed a good match between locations where people live and the ones where they perform their daily activity. It would be interesting to calculate these values for cities in European and US countries for which mobile phone datasets have been used in many papers in the recent years (France, Spain, Portugal, etc.).

### Discussion

We extracted the OD matrices for the 12 largest Senegalese urban areas from mobile phone data and proposed

	$L/L_O$	$L/L_R$	$O$
Dakar	2.09	0.22	0.87
Thies	3.49	0.23	0.82
Touba	3.70	0.18	0.86
Mbour	3.36	0.21	0.84
Kaolack	1.92	0.30	0.83
Saint-Louis	3.84	0.20	0.84
Ziguinchor	3.08	0.29	0.78
Tambacounda	3.90	0.28	0.77
Tivaoune	8.98	0.20	0.82
Diourbel	2.97	0.28	0.79
Louga	4.47	0.31	0.74
Kolda	3.38	0.27	0.79

TABLE III:

	$L$ (km)	$L_O$ (km)	$L_R$ (km)	$\sqrt{A}$ (km)
Dakar	4.6	1.7	21	20

TABLE IV: Comparison of the average commuting length as measured on the data ( $\ell$ ), on the optimal OD matrix ( $\ell_O$ ), the random OD matrix ( $\ell_R$ ), and the typical size of the city ( $\sqrt{A}$ ).

several measures that can help in characterizing the spatial structure of commuting and its efficiency, and to compare different cities with each other.

At the interurban level, we could show that Senegalese cities display a good integration of labor and housing markets. In addition, the attraction radius of cities allowed us to identify important large cities and also important economical nodes. Cities with the largest attraction radius – such as Dakar, Tambacounda, Touba, Saint-Louis and Ziguinchor – would then naturally benefit from transport infrastructure improvements at the country scale, linking cities.

At the intra-urban level, we characterized the efficiency



of the spatial organization of residential and working areas in terms of commuting. In particular, we showed that for Dakar, commuting distances can be very short and that the city seems to be coherently organized in this respect. However for other cities such Louga, Kolda, Mbour, it seems that we are far from the optimal commuting solution. One could have naively expected that the larger the city, the more 'anarchic' it becomes, but our results prove that this naive representation is wrong. These preliminary conclusions would benefit from further investigations, in order to understand the origin of the spatial mismatch for these cities.

Our results show that mobile phone data can effectively be used to characterize how well an urban area is organized. In this respect, they can help in identifying the more fragile urban areas that deserve a particular attention for future intra-urban transport planning.

## Material and Methods

### Delimitation and selection of cities.

When one wants to compare cities, an important issue is to rely on a common, reasonable spatial definition/delimitation applied to all cities [20, 33]. For example, Dakar as a geographical object cannot be restricted to the municipality of Dakar, both in terms of morphology and function. The spatial layers provided in the project include larger spatial delimitations corresponding to administrative entities. Since we didn't find any documentation explaining the territorial criterion chosen - if any - to construct these spatial objects, we cannot assume that such spatial delimitations are suitable to properly define Senegalese cities.

For this reason, we conceived a method to delimitate cities using the spatial points pattern of Orange's mobile phone antennas. We proposed a simple density-based clustering method based on the hypothesis that the density of antennas reflects the density of population. For each antenna we count the number of its neighbors within a growing distance threshold ranging from 0 up to 20 km, with a fixed step of 0.5 km. We then classify those quasi-linear distributions (see Figure 6) and distinguish between Dakar's antennas with a steep slope, Touba's and other cities with a less steep slope, and non-urban antennas with a slight slope. Our delimitation includes urban cores but also peripheric neighborhood (see Figure 6).

We compare the criterion used to satellite pictures of the corresponding cities [41] and to the the OpenStreetMap boundaries of each city, and find that the definition captures well the built areas of each city (see Figure 7

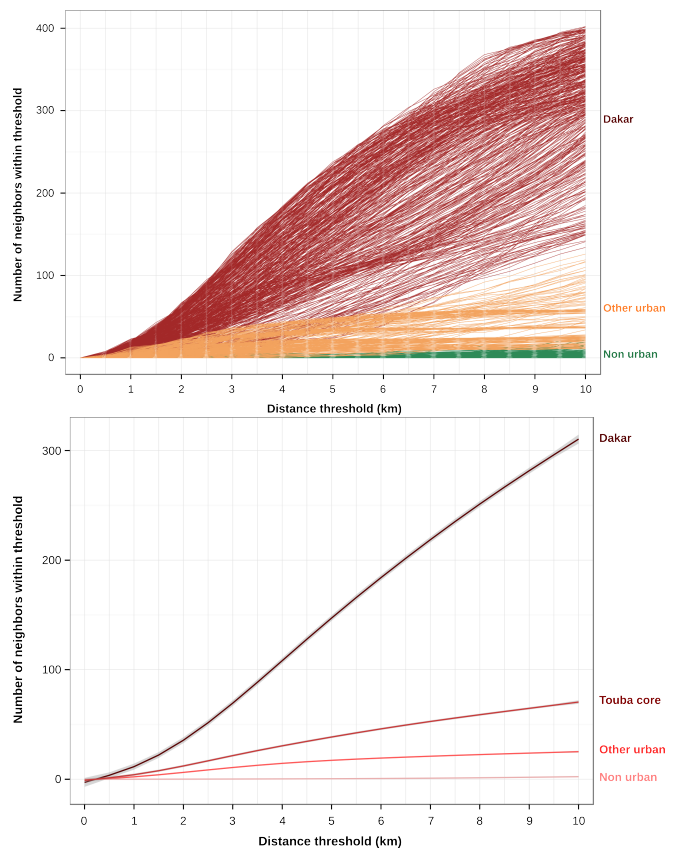


FIG. 6: Illustration of the method used to delimitate cities. For each antenna of the dataset, we count its number of neighbors in a circle of increasing radius (x-axis of the top figure). Each curve represent an antenna. We then apply a hierarchical clustering method on the resulting set of vectors, and represent the average profile of each class on the bottom figure.

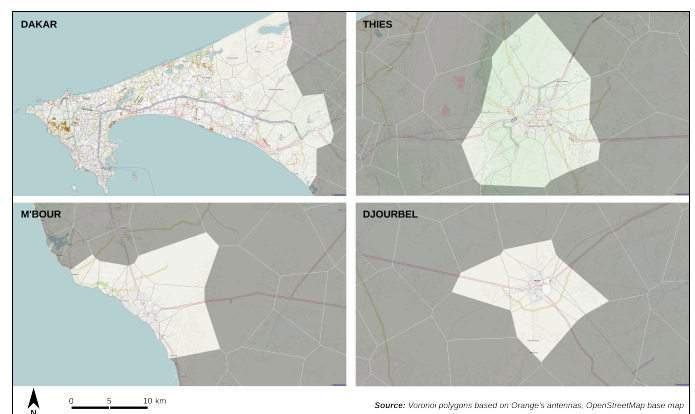


FIG. 7: Comparison of the antennas chosen to delimitate four of the twelve cities selected, following our criterion based on the density of antennas (red), and the corresponding OpenStreetMap layers. We checked each city individually (others not shown here).

## Extraction of the Origin-Destination matrices

In the following, we call commuters the people identified by our algorithm, and commutes the trips as they appear in the resulting Origin-Destination matrix.

The reference files to calculate the OD matrices are *SET2\_PXX.CSV* of Dataset 2, where *XX* varies from 01 to 25. The output of the method is a  $m \times n$  matrix where  $C_{ij}$  is the number of commuters that live in place  $i$  and whose main daily activity is located in  $j$ . In the following we call 'Home' the residence cell of the user and 'Work' the cell of its main daily activity place.

For each user  $u$ , the extraction procedure is the following : for each hour of the two weeks period - weekends excepted - during which the user used her phone at least once, we identify the most visited cell/antenna during this hour. This cell/antenna is the one from which the user has given/received the most calls/sms during this particular hour. Hours are partitionned in two groups: (1) the daily hours that are spent at work/school for most people during weekdays (hours between  $min_W$  and  $max_W$ ) ; (2) the late evening, night and early morning hours, spent at home for most people (hours between  $min_H$  and  $max_H$ ). For both groups of hours, we identify the cell to which the user has been 'attached' the greatest number of hours. We then calculate the proportion of time spent in the cell (number of hours / total number of hours during which the user called). Finally, if in both cases these proportions are greater than a parameter  $prop$ , then the two cells are tagged as the user's work/home cell and the user's home cell. Otherwise the user is not selected because her locations don't show enough regularity to assume than the two most frequent antennas are resp. her workplace and home.

Once we have applied the extraction procedure to all users we end up with an OD matrix of commuting flows for the whole country, for each two weeks period of the dataset 2.

*Sensitivity analysis* We analyzed the influence of the value of  $prop$  on the number of users selected, and also on the proportions of intra-cell flows (i.e. the proportion of individuals who have the same Home and Work cell) for the file *SET2\_P01.CSV*, by using the following parameter values:

- $min_W = 8$
- $max_W = 17$
- $min_H = 19$
- $max_H = 7$

In the figure 8 we can see that the number of users decreases when  $prop$  increases, which is an expected effect. However the proportion of intra-cell flows tends to sharply increase when  $prop$  increases. In order not to remove users and try to keep the network structure we choosed to fix  $prop$  to 1/3.

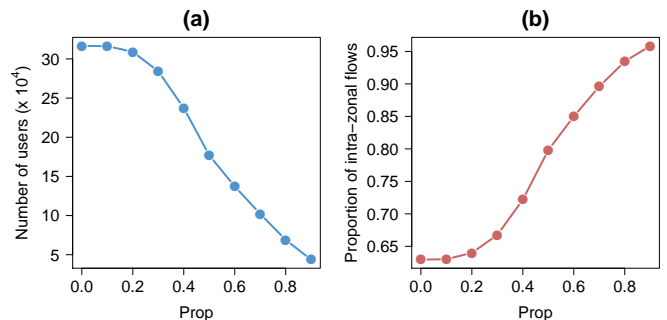


FIG. 8: Sensitivity of (a) the number of individuals selected and (b) the proportion of intra-antenna flows, to the parameter  $prop$  of the OD matrix extraction method. These tests have been performed on the dataset SET2\_P01.

## Computation of the mobility networks to random and optimum networks

### *Random Matrices and optimal matrices*

In the following, we detail the methods used to obtain the random origin-destination matrices—corresponding to a situation where people would choose their working location at random from the existing possibilities— and the optimal origin-destination matrices—corresponding to a situation where people would be assigned to a working location so that the total distance commuted at the city level is optimal. What we are interested in here is the optimality of the commuting patterns given the existing spatial distribution of homes and jobs. We therefore constrain the random and optimal OD matrices in such a way that the number of homes and jobs at each antenna are identical to those observed in data. If we interpret the OD matrix as representing a network between homes and jobs in different locations, then both our random and optimal null model preserve the in- and out- degree sequences.

To generate a random graph that conserves the in- and out- degree of each node of the reference graph, we use the Molloy-Reed algorithm [34] which complexity is in  $O(n)$ , where  $n$  is the sum of the weights of the edges (i.e. the number of individuals in the OD case).

In order to generate the optimal network, we use simulated annealing, a probabilistic method for global optimization problems [35]. At each step, we invert 2 OD pairs  $(a, b)$  and  $(c, d)$  to  $(a, d)$  and  $(c, b)$  (so that the conservation of in- and out- degrees is guaranteed). The move is accepted with probability 1 if the proposed solution is better than the previous one, i.e. if the total commuting distance is smaller than that of the previous situation. If the new situation is worse, however, the move can still be accepted with a probability that depends on temperature  $T$  as

$$P(1 \rightarrow 2) = \exp\left(\frac{L_1 - L_2}{T}\right) \quad (4)$$

and the temperature is decreased as the search progresses. This trick allows to avoid getting stuck in local minima.

### Using calls data to estimate population movements

#### Framework

We define the activity  $\tilde{A}_i$  of an antenna  $i$  as the total number of calls and text messages sent and received from  $i$  over a time window  $\tau$ . We call  $\tau$  the resolution – the minimal resolution is 1 hour, imposed by the dataset (*D1*). In other words, we are theoretically able to follow the change of activity in the city at the one-hour level.

$$\tilde{A}_\tau(i) = \int_\tau A(i, x) dx \quad (5)$$

where  $A(i, x)$  is the measure of activity (in-call, out-calls, number of users) for the antenna  $i$  at the hour interval  $x$ . Then we define the mass of the antenna as

$$M_\tau(i) = \frac{\tilde{A}_\tau(i)}{\sum_{i \in \mathcal{N}} \tilde{A}_\tau(j)} \quad (6)$$

Depending on the area of study  $\mathcal{N}$  and the time-window  $\tau$  we might be able to catch different phenomena. For instance, setting  $\mathcal{N}$  to be a city,  $\tau$  to be of the order of an hour, one can identify the patterns of daily activity in cities. Setting  $\mathcal{N}$  to be an entire country and  $\tau$  to be of the order of a week, a month... one can possibly identify internal migrations.

#### Difference between day and night activity

We first start with comparison between day and night activities in the city. Census traditionally give information on the residential population in cities, sometimes

also on the working population. Mobile phone data allow us to get information with better time resolution and to follow the locations of people within the city during the day. Such information is important to have, for instance for emergency evacuation plans, for understanding epidemic spreading over short time-scales, etc. For each antenna  $i$  we plot the quantity

$$DN(i) = \frac{M_{day}(i) - M_{night}(i)}{M_{night}} \quad (7)$$

which represents the relative difference of activity of the antenna between night and day. With other data to calibrate the relation between mass of antennas and populations, one should be able to estimate the relative differences in population from these differences in antenna mass. Even without calibration, the differences in the mass of antennas give a good idea of the changes in the spatial location of populations over time.

### Authors contributions

TL, RL, GC, RG, HC, ML, J-MB and MB designed the study; RL, TL and MB coordinated the study. RL, ML, HC, GC, TL and RG processed and analysed the data. RL, GC, HC and ML made the figures. TL, RL, and MB wrote the paper. All authors read, commented and validated the final version of the manuscript.

### Acknowledgements

MB is supported by the European Commission through the FET-Proactive project PLEXMATH (Grant No. 317614), and the project EUNOIA (FP7-DG.Connect-318367). TL is supported by the EU commission through project EUNOIA (FP7-DG.Connect-318367). ML acknowledges funding from the PD/004/2013 project, from the Conselleria de Educacin, Cultura y Universidades of the Government of the Balearic Islands and from the European Social Fund through the Balearic Islands ESF operational program for 2013-2017. RG is supported by the European Commission FET-Proactive project PLEXMATH (Grant No. 317614)

- 
- [1] The World Bank, Senegal <http://www.worldbank.org/en/country/senegal> (accessed 22/12/2014).  
 [2] City population website <http://www.citypopulation.de/Senegal.html>  
 [3] Senegal Country Assessment Report - MIT <http://web.mit.edu/urbanupgrading/upgrading/case-examples/overview-africa/country-assessments/reports/Senegal-report.html> (accessed 10 december 2014).  
 [4] Lenormand, M. et al. Cross-checking different sources

- of mobility information. *PLoS ONE* **9**(8): e105184. doi:10.1371/journal.pone.0105184 (2014).  
 [5] Ratti, C., Williams, S., Frenchman, D. & Pulselli, R.M. Mobile landscapes: using location data from cell phones for urban analysis. *Environ. Plann. B* **33**, 727–748 (2006).  
 [6] González M., Hidalgo C. and Barabási A.-L. (2008) Understanding individual human mobility patterns. *Nature* **453**, pp 779-782.

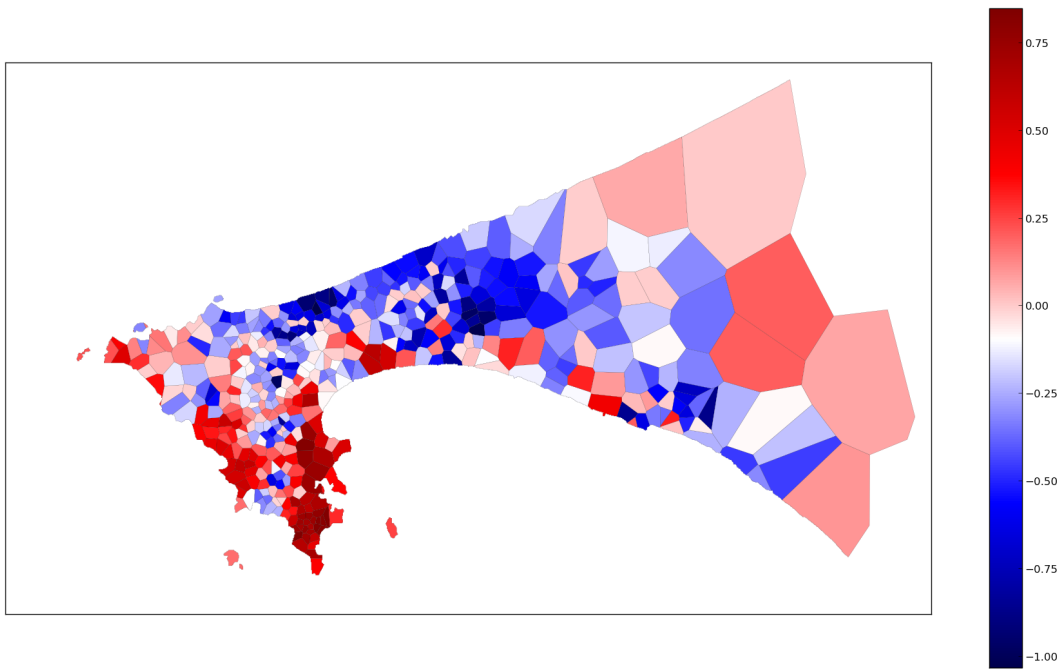


FIG. 9: Relative difference of antenna mass between night and day in Dakar. Each Voronoi cell is colored with respect to the value  $\frac{M_{day}(i) - M_{night}(i)}{M_{day}}$  calculated for the corresponding antenna  $i$ .

- [7] Roth, C., Kang, S.-M., Batty, M. & Barthelemy, M. Structure of urban movements: polycentric activity and entangled hierarchical flows. *Plos ONE* **6**, e15923 (2011).
- [8] Noulas, A., Scellato, S., Lambiotte, R., Pontil, M. & Mascolo, C. A tale of many cities: universal patterns in human urban mobility. *Plos ONE* **7**:e37027 (2012).
- [9] Schneider, C.M., Belik, V., Couronné, T., Smoreda, Z. & González, M.C. Unravelling daily human mobility motifs. *J R Soc Interface* **10**:20130246 (2013).
- [10] Louail, T. et al. Uncovering the spatial structure of mobility networks. *Nature Communications*, in press (2014).
- [11] Roth C., Kang S.M., Batty M. and Barthelemy M. (2012) A long-time limit of world subway networks. *Journal of the Royal Society Interface*, doi:10.1098/rsif.2012.0259.
- [12] Louf R. and Barthelemy M. (2014) A typology of street patterns. *Journal of The Royal Society Interface* **11** (101).
- [13] Louf, Rémi, and Marc Barthelemy. “Modeling the polycentric transition of cities.” *Physical review letters* **111**.19 (2013): 198702.
- [14] Louail, T. et al. From mobile phone data to the spatial structure of cities. *Scientific reports* **4**:5276 (2014).
- [15] Expert P., Evans T.S., Blondel V.D. and Lambiotte R. (2011) Uncovering space-independent communities in spatial networks, *Proceedings of the National Academy of Sciences* **108**, 7663.
- [16] Herrera-Yagüe et al. (2014) The origins of searchability in social networks, online on MIT HumNet’s publications page.
- [17] Pumain D., Paulus F., Vacchiani C. and Lobo J. (2006) An evolutionary theory for interpreting urban scaling laws, *Cybergeo*, 343, 20 p.
- [18] Bettencourt LMA, Lobo J, Helbing D, Kuhnert C and West GB (2007) Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Science*,104:7301–7306.
- [19] Louf R. and Barthelemy M. (2014) Scaling: lost in the smog. *Environment and Planning B* **41** (5), 767-769.
- [20] Arcaute, E. et al. Constructing cities, deconstructing scaling laws. *Journal of the Royal Society Interface*, <http://dx.doi.org/10.1098/rsif.2014.0745> (2014).
- [21] Bertaud, A. & Malpezzi, S. The spatial distribution of population in 48 world cities: implications for economies in transition. *World Bank Report* (2003).
- [22] Bretagnolle, A., Pumain, D. & Vacchiani-Marcuzzo, C. [The organisation of urban systems] in *Complexity perspective in innovation and social change*, [Lane, D., Pumain, D., van der Leeuw, S.E. and West, G. (eds)][197–220], Springer (2009).
- [23] Grataloup C., Fumey, G. and Boucheron, P. L’atlas Global, *Editions les Arènes* (2014).
- [24] Simini F., González M., Maritan A. and Barabási A.-L. A universal model for mobility and migration patterns. *Nature* **484**, pp 96–100 (2012).
- [25] Kung K.S., Sobolevsky, S. & Ratti, C. Exploring universal patterns in human home/work commuting from mobile phone data. *Plos ONE* **9**(6) (2014).
- [26] Berlingerio M. et al. AllAboard: a system for exploring urban mobility and optimizing public transport using cellphone data. *D4D 2013 Challenge book* (2013).
- [27] Mamei M. and Ferrari L. Daily commuting in Ivory Coast: development opportunities. *D4D 2013 Challenge book* (2013).
- [28] Andris C. and Bettencourt, L. Development, Information and Social Connectivity in Côte d’Ivoire. SFI Working paper (2013).

- [29] Wakita, K., Kawasaki, R. and Takami, M. Observation of Human Dynamics in Cote d'Ivoire Through D4D Call Detail Records, *AAAI Technical Report WS-13-04 When the City Meets the Citizen 2*, pp 29–32 (2013).
- [30] Eubank, S. et al. Modelling disease outbreaks in realistic urban social networks. *Nature* **429**, 180-184 (2004)
- [31] Balcan, D. et al. Multiscale mobility networks and the spatial spreading of infectious diseases. *PNAS* **106**, 21484–21489 (2009).
- [32] Dalziel, B.D., Pourbohloul, B. & Ellner, S.P. Human mobility patterns predict divergent epidemic dynamics among cities. *Proc R Soc B* **280**:20130763 (2013).
- [33] Bretagnolle, A., Paulus, F. & Pumain, D. Time and space scales for measuring urban growth. *Cybergeo* 219, <http://cybergeo.revues.org/3790> (2002).
- [34] Molloy, M. and Reed, B. A critical point for random graphs with a given degree sequence. *Random structures and algorithms*, **6**, 161–180 (1995).
- [35] Mézard, M., Parisi, G. and Virasoro, M. A. Spin glass theory and beyond. World Scientific (1987).
- [36] Jiang, S. et al. A Review of Urban Computing for Mobile Phone Traces:Current Methods, Challenges and Opportunities. Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing, doi:10.1145/2505821.2505828 (2013).
- [37] Wu, L., Zhi, Y., Sui, Z. & Liu, Y. Intra-urban human mobility and activity transition: evidence from social media check-in data. *Plos ONE* **9**(5)e97010 (2014).
- [38] Zhong, C., Arisona, S.M., Huang, X., Batty, M. & Schmitt, G. Detecting the dynamics of urban structure through spatial network analysis. *Int J Geo Inf Sc* **28** (2014).
- [39] Deville, Pierre, et al. “Dynamic population mapping using mobile phone data.” *Proceedings of the National Academy of Sciences* **111**(45) : 15888-15893 (2014).
- [40] NB: a penetration rate value superior to 100 means that there are more active phone numbers than individuals living in the country, which happens if for example some people have two numbers, plus professional lines, etc.
- [41] not shown for copyright reasons