Chaos
An Interdisciplinary Journal of Nonlinear Science

Agricultural activity shapes the communication and migration patterns in Senegal

S. Martin-Gutierrez, J. Borondo, A. J. Morales, J. C. Losada, A. M. Tarquis, and R. M. Benito

Citation: Chaos **26**, 065305 (2016); doi: 10.1063/1.4952961
View online: http://dx.doi.org/10.1063/1.4952961
View Table of Contents: http://scitation.aip.org/content/aip/journal/chaos/26/6?ver=pdfcov
Published by the AIP Publishing

CHAOS 26, 065305 (2016)

Agricultural activity shapes the communication and migration patterns in Senegal

S. Martin-Gutierrez,[1] J. Borondo,[1] A. J. Morales,[1] J. C. Losada,[1] A. M. Tarquis,[1,2] and R. M. Benito[1]
[1]*Grupo de Sistemas Complejos, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid, Avda. Complutense s/n 28040 Madrid, Spain*
[2]*CEIGRAM, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid, Calle Senda del Rey 13—Campus Sur de prácticas E.T.S. Ingenieros Agrónomos, Spain*

(Received 8 March 2016; accepted 7 April 2016; published online 7 June 2016)

The communication and migration patterns of a country are shaped by its socioeconomic processes. The economy of Senegal is predominantly rural, as agriculture employs over 70% of the labor force. In this paper, we use mobile phone records to explore the impact of agricultural activity on the communication and mobility patterns of the inhabitants of Senegal. We find two peaks of phone calls activity emerging during the growing season. Moreover, during the harvest period, we detect an increase in the migration flows throughout the country. However, religious holidays also shape the mobility patterns of the Senegalese people. Hence, in the light of our results, agricultural activity and religious holidays are the primary drivers of mobility inside the country. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4952961]

The increasing availability of user generated data from interactions in a wide spectrum of communication networks has enabled human behavior research. While many studies focus on industrialized societies, analysis on developing countries is also being carried out. Despite the lack of information availability in these countries, the high pervasiveness of mobile phones and other information technologies provide rich sources of data for this kind of experimental works. In this paper, we combine information extracted from mobile phone datasets with satellite images of Senegal. We find that agricultural activity and other relevant local events trigger detectable changes in user behavior.

I. INTRODUCTION

During the last few years, there has been an explosion of human behavioral studies using data generated by different kinds of Information and Communication Technologies (ICT). One of the main causes can be found in the increasing availability of data, which is stimulated by the pervasiveness of mobile phones.[1,2] Many studies are focused on the developed world.[3–8] However, there are fewer studies focused on developing countries. Some of them are devoted to human mobility,[9,10] to epidemiology,[11] to the detection of unusual events,[12] or to ethnic interactions.[13] In this work, we study the communication patterns of the country of Senegal and explore their correlations with underlying real world structures and dynamics. Our findings show that agricultural activities have a key role in shaping the phenomena recorded in the data. Another important factor would be the traditional religious festivities.

It is important to remark some relevant aspects of Senegal that will help to interpret the results of these analyses. In this small West-African country, there is a very broad cellular coverage that allows for access in rural areas. In every rural household, there is at least one mobile phone, which is often shared by many people.[14] Its economy is predominantly rural. Agriculture workers in Senegal represent over 70% of its labor force,[15] which means that weather drives the actions of a large part of the population. Senegal is a part of the Sahelian climate zone, which is characterized by cycles of rainy and dry seasons. The rainy season in Senegal spans approximately from June to October and is a strong conditioning for the growing season (see top panel of Figure 1). The primary cash crops are groundnuts, cotton, and horticulture. Groundnuts are grown in the central region, particularly in the Peanut Basin, and horticulture is concentrated in coastal regions. Sorghum and millet are grown in the northern and central regions, and rice is grown in the southern Casamance and in the Senegal River Valley, along the north and north-east border.[15,16] The seasonal nature of an agriculture-based economy implies the alternation of periods of higher and lower activity as well as the seasonal migration of workers.

Another important matter in Senegal is religion. Although several beliefs coexist in the country, including Christian and traditional animism, the predominant religion is Islam. Within the Islamic collectivities, one of the most prominent is the Mouride brotherhood, which is a large Islamic Sufi order founded in 1883 and characterized for considering work as a form of adoration.[17,18] This philosophy has catalyzed the emergence of rural communities dedicated to agriculture, especially, to the cultivation of groundnuts, contributing to the conformation of the region known today as the Peanut Basin.[19] Every year several Mouride celebrations are held. They usually imply a pilgrimage or *Magal* to some holy place. The most important is the Grand Magal of Touba that marks the exile of the founder of Mouridisme, Cheikh Ahmadou Bamba, and gathers around 3 $\times$ 10^6 people every year.

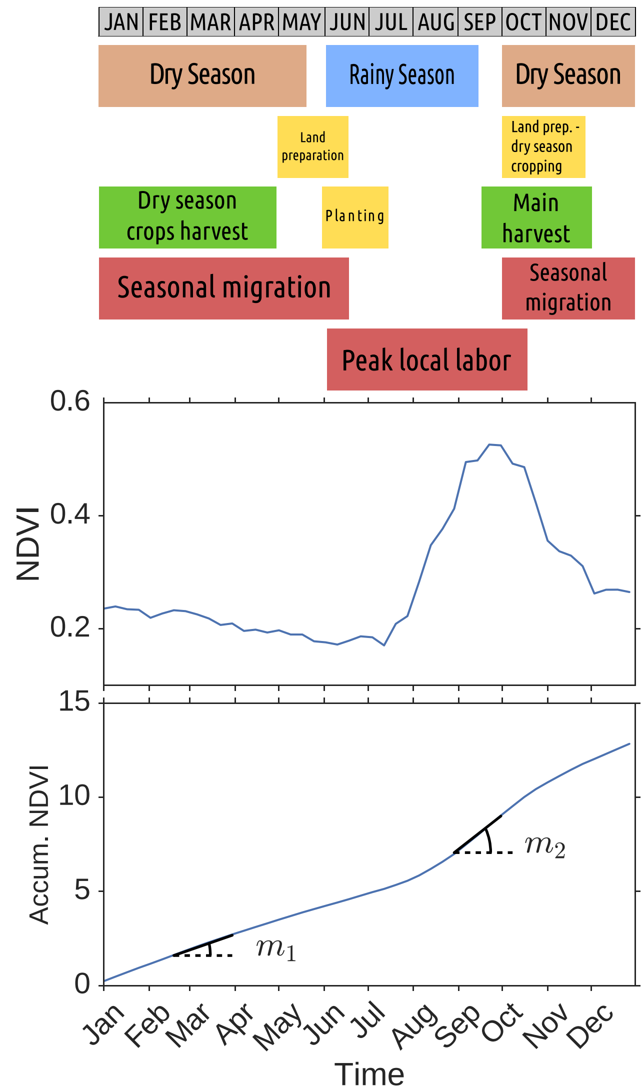

FIG. 1. Agricultural season calendar of Senegal (top), NDVI time series (middle), and its accumulated counterpart (bottom), associated to a randomly chosen antenna.

Several previous researches have shown the potential of mobile phone data for improving our knowledge about human behavior. For instance, the study of mobility patterns[3] has revealed an underlying universal regularity in human motion. Another relevant discovery is the dependence of the searchability of a social network on its geographical scale.[8] Other studies have proven that the division of a country based on the communities of its telecommunications networks may be more realistic than its administrative boundaries.[20] The characterization of the communication routines of the users and the measure of deviations from those routines can also help us infer relevant socioeconomic indicators like unemployment incidence.[21]

In this work, we combine information extracted from mobile phone datasets with satellite images in order to obtain insight about the socioeconomic characteristics of Senegal. To this end, we analyze the temporal evolution of communication patterns and vegetation indices. This enables us to find correlations between deviations of the regular user behavior and periods of higher agricultural activity. We are also able to detect other relevant local events. Additionally, we study the seasonal migration of workers with the aid of mobility networks, showing the influence of the harvesting season on the migration flows.

This article is organized as follows: in Sec. II, we describe the datasets used in this study. In Sec. III, we introduce the methods adopted to extract information from the data. Then, we present and discuss the results; and finally, we summarize the main conclusions.

II. DESCRIPTION OF THE DATASETS

Call Detail Records (CDRs) are metadata produced by phone interactions (either calls or SMS). We have worked with three mobile phone datasets extracted from CDRs. The datasets are based on phone calls exchanged between 9×10^6 users during the year 2013 in Senegal. All these data have been properly anonymized.[22]

Dataset 1 contains information about the number of calls exchanged between any pair of mobile phone antennas on an hourly basis. It is composed of 1666 antennas whose approximate geographical location is known. Dataset 2 contains fine-grained mobility data of 300 000 individuals during lapses of two weeks. The set of users for dataset 2 changes every two weeks. This dataset has been built as follows: when a given user makes a call, the information associated to that call is stored. The information provided includes the antenna, time (with a precision of 10 min), and user identifier. Dataset 3 consists of one year of coarse-grain mobility data of about 150 000 randomly sampled users. This dataset has been built in an analogous way to dataset 2. The information included regards the municipality (not the antenna), the time (also with a precision of 10 min), and user identifier associated to every call.

Besides the CDR datasets, we have used georeferenced images that include the whole country. These images have been retrieved from the MODIS (Moderate-resolution Imaging Spectroradiometer) database, available at The Land Processes Distributed Active Archive Center website.[23] MODIS is a scientific instrument mounted on Terra and Aqua satellites that measures surface reflectance in 36 spectral bands with high spatial resolution. The dataset we have worked with represents the whole 2013 with a set of 8 days composed images with a spatial resolution of $500\,\mathrm{m} \times 500\,\mathrm{m}$. From these 46 images, Normalized Difference Vegetation Index (NDVI) was calculated, as it is one of the most common indices used to monitor live green vegetation. NDVI is defined by the following formula:

$$NDVI = \frac{NIR - VIS}{NIR + VIS}, \tag{1}$$

where VIS and NIR stand for the spectral reflectance measurements acquired in the visible (red) and near-infrared regions, respectively. For vegetation, the NDVI typically ranges between 0 and 1, although by design it can vary between -1 and $+1$, as it is the case in our data. The negative values correspond to water bodies.

III. METHODS

A. Phone calls and NDVI time series

We have built time series with the number of outgoing calls for each one of the 1666 antennas using data from dataset 1, which gives information about the number of calls exchanged between any pair of antennas. Different sampling frequencies have been used depending on the temporal scale considered. In order to detect peaks of calling activity, we have used a peak-enhancing algorithm. It consists in computing a moving Z-Score: k values are taken from each side of every point x_i of the time series $\{x_i\}$ (without including the point itself) and the mean μ_i^k and the standard deviation σ_i^k of these $2k$ values are computed. The score associated to the point x_i is given by

$$ZS_k(x_i) = \frac{x_i - \mu_i^k}{\sigma_i^k}. \quad (2)$$

In order to identify growing areas, we have also built the NDVI time-series for the whole country. Next, we have associated an NDVI time series to each antenna by taking a three-by-three pixels lattice centered at the antenna coordinates and computing the average NDVI for those nine pixels. Each pixel corresponds to $500\,\mathrm{m} \times 500\,\mathrm{m}$ of land and the sampling frequency of the data is 8 days. In order to improve the signal-noise ratio, we have worked with the accumulated NDVI.

B. Phone calls and mobility networks

We have built phone calls networks from dataset 1 described in Section II. In order to do so, we have taken the 1666 antennas as nodes and linked antenna i to antenna j when a phone call is made from i to j. Hence, we have a weighted directed network. The weights w_{ij} of the links correspond to the total number of calls made from i to j during a given time period, which we will arbitrarily choose depending on the temporal scale of the phenomena we want to analyze.

We have also built trajectories networks with two different spatial scales using information from datasets 2 and 3. When considering dataset 2, which contains information of displacements of users between antennas, the nodes of the networks correspond to the antennas. In the case of dataset 3, where the data refer to displacements of users between municipalities, the nodes correspond to the municipalities. Two nodes i and j are linked to each other if a user that has made a call at time t from i makes a call at a posterior time $t + \Delta t$ from j. We treat this as a memoryless process. If a given user makes a call from A at time t_1, from B at time t_2 and from C at time t_3 such that $t_1 < t_2 < t_3$, we make a link from A to B and from B to C, but not from A to C. The result is a weighted directed network, where the weights w_{ij} are the number of displacements from location i to location j during a given time interval.

In order to detect and characterize longer lasting events like seasonal migrations, we have built another kind of networks. The nodes of the migration networks are the municipalities. The first step is to assign home municipalities to users. In order to do that we have considered that a user is more likely to be at home from 8 P.M. to 7 A.M.[6] For each day, we have extracted the vector of municipalities from which a user has made a call within that time interval. We have defined her *daily home* as the mode of that vector. Then, we have defined her *regular home* as the mode of the vector of 365 *daily homes*. We have also computed *temporary homes* for other time periods. Particularly, we have computed the *monthly home* as the mode of the vector of *daily home* municipalities for each month. Comparing the *regular home* to the *monthly home* we can get insight of the migration flows.

When performing these calculations, we have filtered out certain users whose home was not stable enough: we have only kept users whose *regular home* was the *daily home* for more than 182 days (half a year).

The migration networks are then built as follows: the municipalities are the nodes. Two municipalities i and j are joined with a directed link from i to j if municipality i is the *regular home* of a given user and municipality j is the *temporary home* of the same user during a given time interval (for example, during a month or a day). In this weighted directed network the weights w_{ij} correspond to the number of users that have migrated from node i to node j during a given time period.

The community structure of the networks described above has been computed by means of the Louvain algorithm[24] implemented in the NetworkX python module.[25] This algorithm performs a partition of the network in several communities aiming for the maximization of its modularity.[26]

C. Flows of users

We have studied migration patterns through the evolution of the mobility networks. Depending on the spatial and temporal scales we consider, different kinds of phenomena can be detected, from celebrations of religious festivals to seasonal migration of workers. The flows of users are measured by the displacements of people between different locations. Given a mobility network where the nodes are locations, we define the flow of users from location i to location j as the weight of the link w_{ij}. Thus, in order to determine the net incoming flow to location j, we compute the difference between its weighted in degree and its weighted out degree: $F_j^{in} = k_j^{in} - k_j^{out}$. Different spatial scales have been considered by aggregating the nodes of the networks at different levels.

IV. RESULTS

A. Phone calls and NDVI time series

We have built the time series corresponding to the number of outgoing calls per hour for each one of the 1666 antennas using the information provided in dataset 1, which contains the number of calls made between pairs of antennas. In Figure 2, we present the time series of the number of calls per hour for the average week and the average day in a given antenna. The regular behavior of this time series reflects the routines of the users. We can see daily cycles similar to the

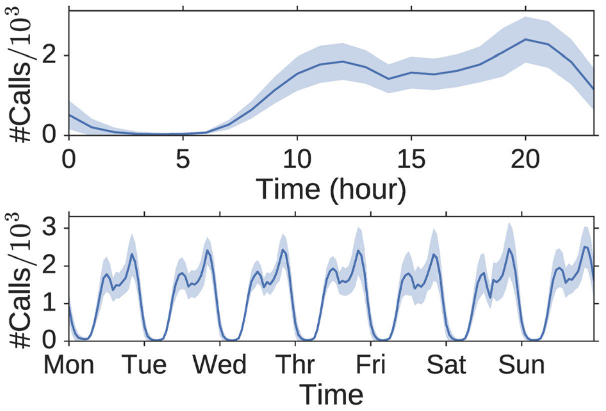

FIG. 2. Time series of the number of outgoing calls per hour in a given antenna for the average day (top) and the average week (bottom) during the year 2013 in Senegal. The shadowed area around the mean corresponds to the standard deviation.

ones detected in previous studies,[7] with maxima and minima repeated each day at the same times. Each daily cycle starts at 12 A.M. with activity decreasing until reaching a low activity plateau at around 3 A.M. The number of calls per hour starts to rise at 6 A.M. and reaches a local maximum at 12 P.M. After that point, activity slightly decreases until 2 P.M., which coincides with lunch time. There is a second plateau that lasts until around 6 P.M. Then, the number of calls increases again until reaching its absolute maximum at 8–9 P.M. Finally, the activity decreases again until it hits its minimum after midnight.

Let us consider now the evolution of the total number of calls per day accumulating the number of calls each 24 h. In Figure 3, we present examples of time series of daily activity for two different antennas. There is a significant number of antennas whose time series, like in the top panel of Figure 3, exhibit abrupt increments of the number of outgoing calls on two particular dates: one on the 9th of August and the other on the 16th of October. The origin of these two peaks will be discussed further on. There are however other antennas in which this effect is not noticeable, as is the case of the time series shown at the bottom panel of Figure 3.

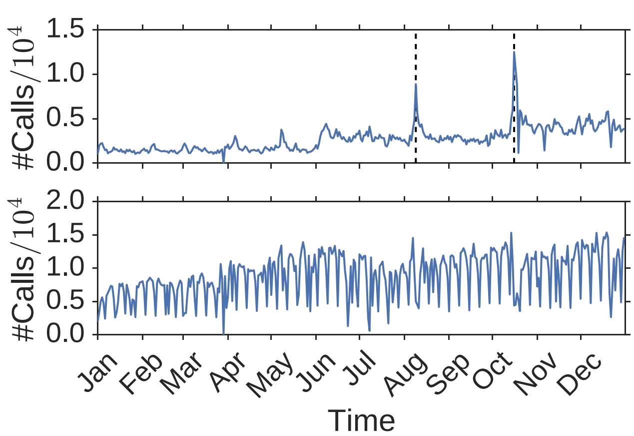

FIG. 3. Time series of outgoing calls per day of two antennas. One presents two peaks of activity on the 9th of August and the 16th of October (top) and one does not (bottom).

In order to establish a connection between the above mentioned communication patterns and the real world events that may have triggered them, we have researched the evolution of the Normalized Difference Vegetation Index (NDVI). The motivation to take an agriculture-based approach is the fact that the most relevant economic sector in Senegal when it comes to proportion of work force involved is agriculture (over 70%). The NDVI measures the quantity and quality of vegetation within a given area (see Section III) and is bounded between -1 and $+1$. Figure 4 shows the temporal evolution of this index during the year 2013 in Senegal. Red pixels indicate high NDVI, whereas green pixels mean low NDVI values. As can be seen, NDVI increases during the rainy season from south to north and then decreases from north to south.

We have associated an NDVI time series to each antenna as described in Sec. III A. In order to improve the signal-noise ratio, we have computed the accumulated NDVI time series. For crop fields, the NDVI should increase during the growing season, which spans from planting to harvesting, and decay after harvesting. As it can be seen in the bottom panel of Figure 1, this pattern is reflected in the accumulated NDVI time series as an increase in the slope during the growing period. In fact, the accumulated NDVI time series is roughly composed of three long intervals of linear growth separated by two inflection points corresponding to short intervals when the slope changes rapidly.

We have carried out a comparison of the outgoing phone calls time series of each antenna and its associated accumulated NDVI time series. In order to do that, we have used the time series of the number of outgoing phone calls per period of 8 days. We found that the inflection points of the accumulated NDVI time series happen roughly at the same dates as the two characteristic peaks found in the phone calls time series (see Figure 5). This correlation may suggest that the abrupt increments of phone calls may be related to periods of high agricultural activity (peak of local labor). However, these two peaks also coincide with the celebration of two important holidays of Islam: Eid al-Fitr (Festival of Breaking the Fast, the end of Ramadan)[27] and Eid al-Adha (Festival of the Sacrifice).[28] Hence, the origin of the two peaks in the time series of outgoing calls probably lies in a combination of both factors.

In order to further research the possible relationship between phone calls and agriculture, we have carried out a classification of the antennas attending to criteria involving the characteristics of both the phone calls and the NDVI time series.

We have taken the following considerations into account. First, we have filtered the NDVI data associated to cities (which we suppose to be lower than in crop fields) and forests (which we suppose to be higher than in crop fields). If a time series had a yearly accumulated NDVI lower than 10.5, it was labeled as city type; whereas if it presented a yearly accumulated NDVI higher than 18, it was labeled as forest type. Then, we performed a linear regression for the first interval of linear growth of the accumulated NDVI time series (from the 1st of January to the 25th of May), obtaining

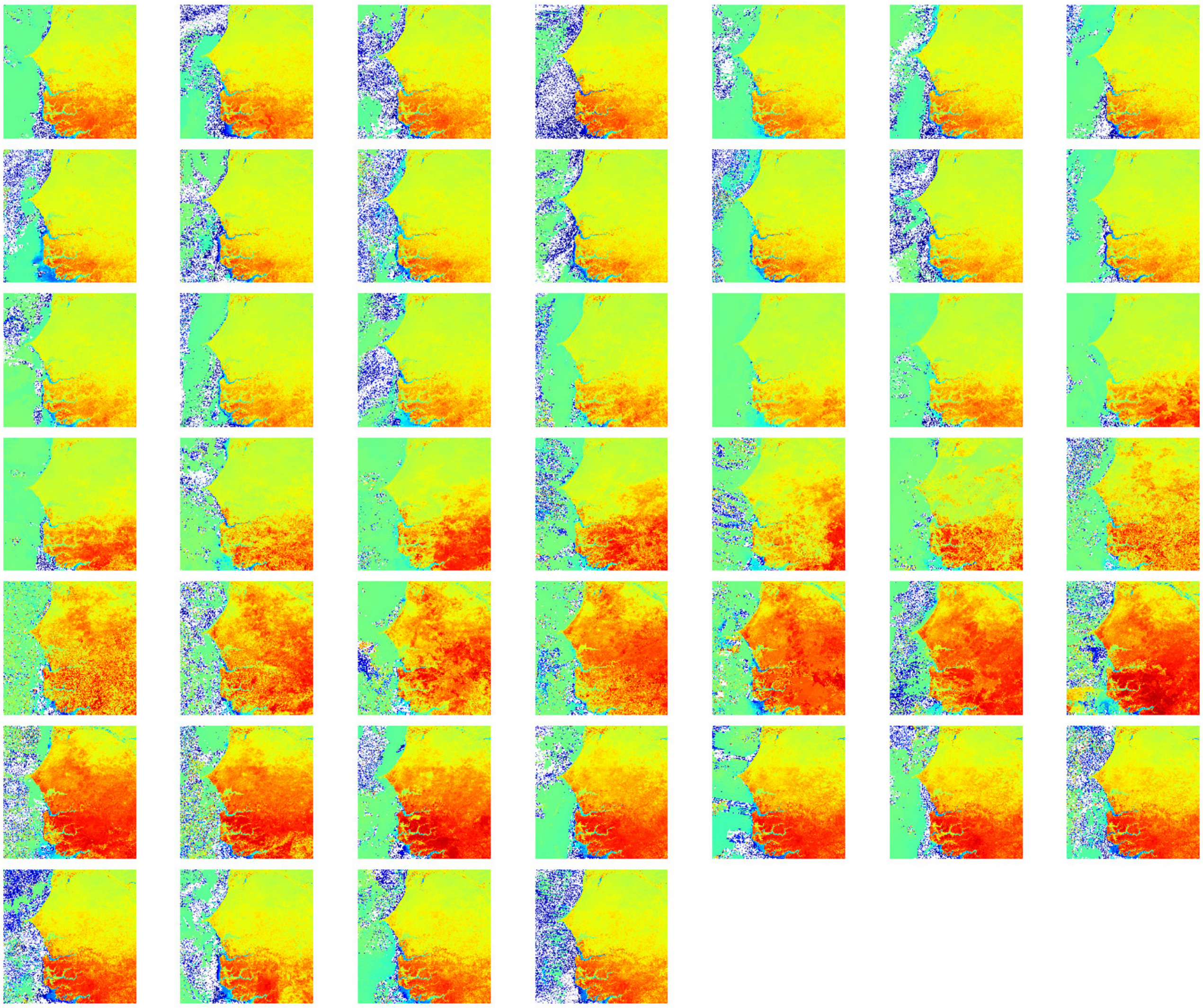

FIG. 4. NDVI evolution during 2013 in Senegal. Each image corresponds to a period of 8 days, starting on 1–8 of January (top left panel). Each pixel is 500 m × 500 m. Red pixels correspond to high NDVI, whereas green ones correspond to low NDVI.

its slope m_1. If the regression had an $R^2 < 0.99$ that time series was labeled as noisy. Otherwise, we performed a second linear regression for the second interval of linear growth, which approximately coincides with the growing season period (from the beginning of August to the end of October) and obtained the slope m_2. Next, we computed the ratio of the second slope m_2 with respect to the first one m_1. The measure m_2/m_1 gives a notion of the ratio of average NDVI during the growing season with respect to the average of NDVI during the rest of the year. Thus, if $m_2/m_1 > 1.65$ the time series was tentatively labeled as crop field type. A final filtering was done by computing the first inflection point of the accumulated NDVI time series as the intersection of the first and second regressions. If it was located earlier than expected (the 24th point of the NDVI time series, which corresponds to the interval between 12 and 20 of July), the NDVI was classified as forest type.

In a second step, we classified each antenna according to its calling patterns. In this case, we have used the peak-enhancing algorithm described in Section III to look for those time series that show two or more peaks during the growing season (roughly between the 20th of July and the 9th of November). As explained in Section III, this algorithm assigns a score $ZS_k(x_i)$ to each point x_i of the time series. The points whose scores are above a threshold of $2.5\sigma_i^k$ are considered peaks, but, in order to avoid noisy data, they are accounted for only if there is a reasonable separation between them (7 points).

The classified antennas are plotted over the map of Senegal shown in Figure 6. Each antenna has been assigned a color and shape according to its class. A red dot corresponds to an antenna whose phone calls time series present two or more peaks of activity between 20th of July and 9th of November and an NDVI time series that we can associate with a crop field. Blue squares are antennas whose phone calls time series present two or more peaks of activity but an NDVI that cannot be associated to crop fields. They are located mainly in cities. Green diamonds are antennas without activity peaks or just one peak during the period of interest and an NDVI that can be associated to crop fields. The population of red dots seems to concentrate in the so-called Peanut Basin, pointing towards a correlation between agriculture and communication patterns.

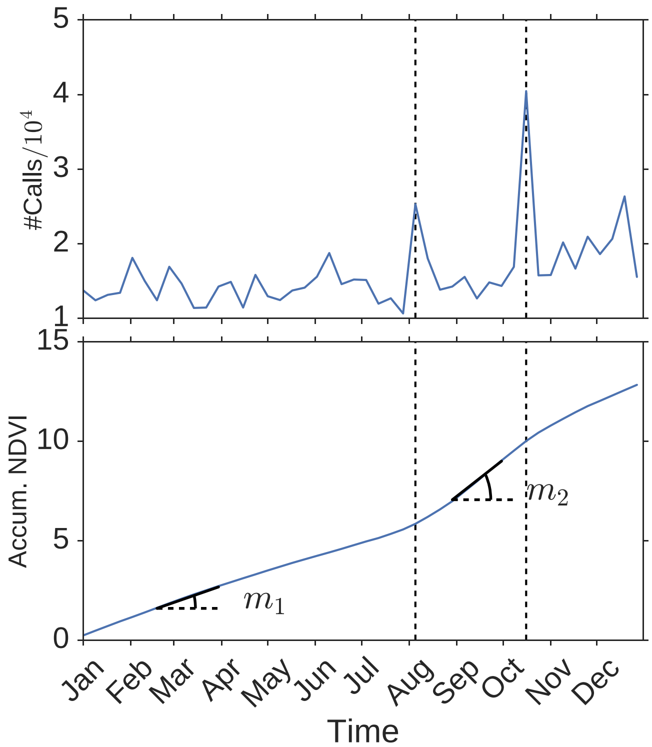

FIG. 5. Time series of outgoing calls per period of 8 days (top) and time series of accumulated NDVI (bottom) of one of the antennas. The intervals that contain the 9th of August and the 16th of October are marked by vertical dashed lines.

B. Phone calls and trajectories networks

We have built the phone calls networks of the mobile phone antennas by considering each antenna as a node and making a directed link from antenna i to antenna j if a call is made from i to j. The weight of a given link w_{ij} corresponds to the number of calls made from i to j during a given time interval. In Figure 7(a) we present the phone calls network for the whole year of 2013. Link size and color are proportional to $\log_{10}(w_{ij})$. Geographically, the network is composed of subnetworks with radial structure. There are hubs located on cities and radii that connect them to the neighboring towns. The strongest links are located within cities (like Dakar), implying, as one would expect, that the connectivity within a city is denser than between towns. Notice the dense subnetwork that borders the north and northeast boundary of the country coinciding with the course of the Senegal river. It is also worth pointing out the strong link between the cities of Tivaouane and Touba, in the mid-west part of the country.

In Figure 7(b), the community structure of the yearly calling network is shown. Node color corresponds to the community to which the node has been assigned by the Louvain algorithm,[24,25] which computes the community structure of a network by optimizing its modularity. As we can see, the geographic boundaries of the communities found in this network resemble pretty well the areas associated to different land uses,[16] which are shown in Figure 7(c). For example, community 8 occupies approximately the same area as the Peanuts zone and community 9 roughly corresponds to the Rice and gardening zone. There are some communities and land use zones that are mixed, like communities 7 and 10, which span areas from the Cassava, Cowpea, and Sylvo-Pastoral zones, or community 6, which is composed of two land use zones.

We have also built trajectories networks where, again, antennas are considered as nodes and there is a directed link from antenna i to antenna j when a user that has made a call from i at a given moment t, makes a call from j at a later time $t + \Delta t$. The weight w_{ij} of a link is given by the sum of displacements from i to j during the time interval under analysis.

We have carried out a community analysis of the daily trajectories networks using the Louvain method. In Figure 8, we present the temporal evolution of the number of communities found and the values of modularity for each day

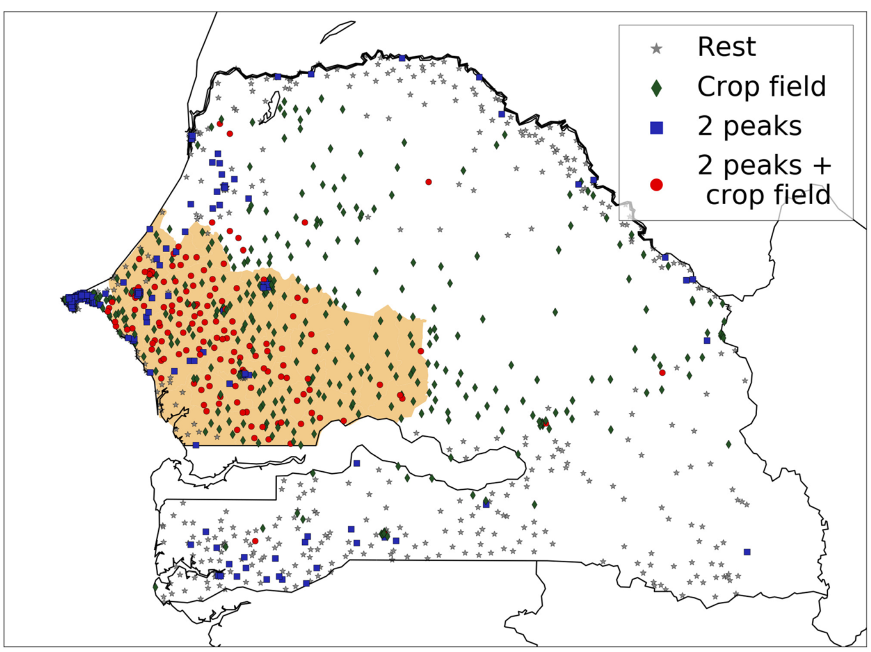

FIG. 6. Geolocalization of the antennas used in this study on the map of Senegal, with an approximate delimitation of the Peanut Basin colored in light brown. Shapes and colors correspond to the different classes of antennas of the classification carried out in Section IV A. Red dots: Two or more peaks in the calling activity time series during the growing season and crop field type NDVI. Blue squares: Two or more peaks in the calling activity time series but not crop field type NDVI. Green diamonds: One or less peaks in the calling activity time series and crop field type NDVI. Grey stars: The rest.

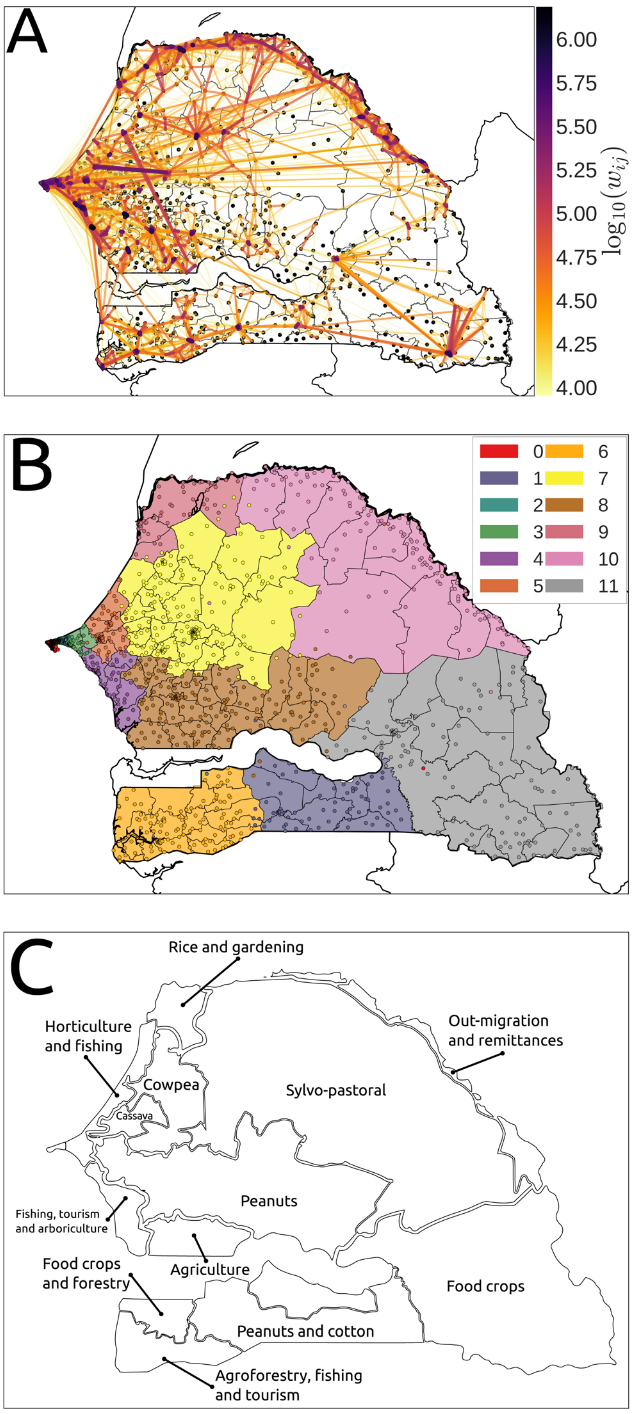

FIG. 7. (a) Phone calls network of Senegal aggregated for the whole year of 2013. Link weights correspond to the total number of calls made between two antennas during 2013. (b) Corresponding communities of the phone calls network shown in (a). (c) Map of land use of Senegal.

during 2013. Although in general modularity lies around 0.74–0.78, an abrupt increase is found on the 9th of August and on the 16th of October, when modularity reaches values of up to 0.82, implying that, on those dates, the communities found by the algorithm are statistically more significant. Concerning the number of communities detected by the Louvain algorithm, an increase with respect to the average value of 20 is also found on these two dates, when rises to around 30. It is worth noticing that the modularity-based

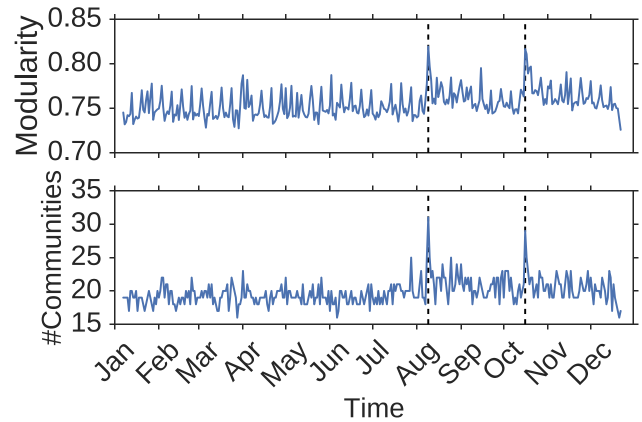

FIG. 8. Time series of modularity (top) and number of communities (bottom) of the daily trajectories networks. The 9th of August and the 16th of October are marked by vertical dashed lines.

methods have a known resolution issue that makes the algorithm merge small communities into larger ones when it may not be appropriate.[29] That means that obtaining more and smaller communities is, in principle, harder, which adds significance to the previous result.

The increment of the modularity and the number of communities in the trajectories network can be interpreted as a change of user behavior: users move more locally than during the rest of the year. This interpretation is supported by the fact that, as can be seen in the top panel of Figure 9, on those two dates there are local minima of the number of edges. The higher value of modularity is also favored by an improvement of the statistics caused by the increase of the total weight of the network (see middle panel of Figure 9). It may correspond to an increase in the total number of displacements and implies, either that users have traveled more than in a regular day, or that they have called more than in a regular day or both. The second explanation would imply the possibility that the communities have always been there but we have only detected them on those dates because of the increase of the number of calls and the consequent improvement of statistics. Nevertheless, this possibility should be rejected given that, as it is shown in Figure 10, the

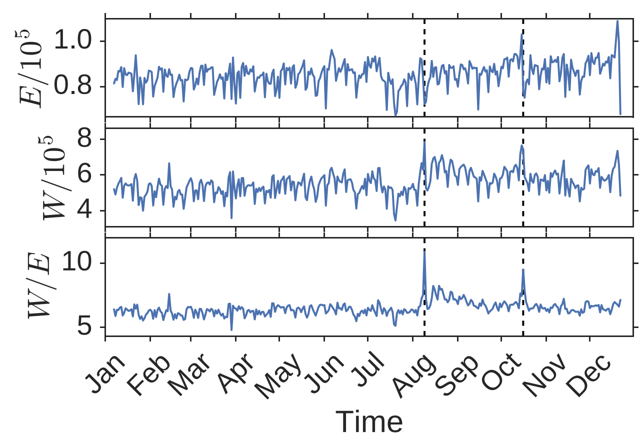

FIG. 9. Temporal evolution of several properties of the daily trajectories networks: Number of links E (top), total weight $W = \sum_{ij} w_{ij}$ (middle), and the ratio W/E (bottom). The 9th of August and the 16th of October are marked by vertical dashed lines.

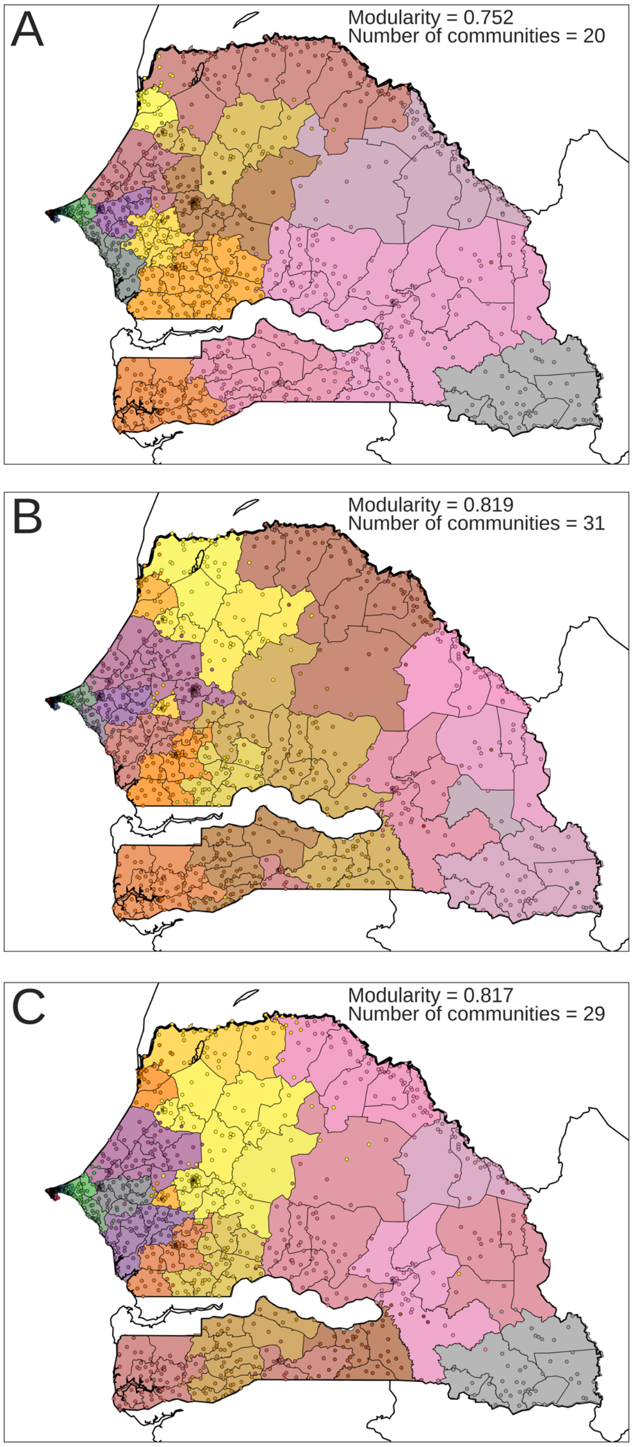

FIG. 10. Community structure of the trajectories network accumulated for the whole year (a), for the 9th of August (b) and for the 16th of October (c). Dots represent the geolocation of the antennas considered. Color corresponds to the communities assigned by the Louvain algorithm.

accumulated trajectories network for the whole year presents fewer and larger communities than the networks of the 9th of August and the 16th of October. To sum up, the most plausible explanation is that users travel more and make more calls than in a regular day, although their displacements tend to be local.

A preference for local displacements and the growth of the number of displacements can be associated both to the trips people make to visit their families and friends for the holidays and the intensification of agricultural labor.

C. Flows of users

To further analyze the mobility patterns of users in Senegal, we have studied the flows of users at different scales of time and space, from days to months and from municipalities to administrative regions. Depending on the scale at which we look at the data, we are able to detect different kinds of phenomena.

As we previously pointed out, the relevance of religion in Senegal is manifested in the mobility behavior of the users. By analyzing the motion of users between municipalities at the daily scale, we are able to detect several religious holidays. For example, on the 23rd of January takes place the celebration of the *mawlid*, the birth of the prophet Muhammad, during which people travel to the city of Tivaouane. This festivity causes an increment of the net incoming flow to the municipality of Pambal, where Tivaouane lies. The flow rises from an average value of 24 users per day to 10 125 users on the 23rd of January. In Figure 11(a) the trajectories network of that day shows the flow of people to that city. Another holiday that can be detected this way is the Magal of Porokhane, a tribute to the wife of the prophet of a local Islamic brotherhood, the Mouride, which is celebrated on the 14th of March in Porokhane. On this date, the net incoming flow of users to the Paoskoto municipality, where Porokhane is located, rises from an average value of 5 to 1720. The trajectories network of the 14th of March, which is presented in Figure 11(b), shows a convergence of links to this locality. On the 25th of June, when the Magal of Darou Mousty is celebrated, we also detect an increment of users flow to its municipality. The net incoming flow grows from the average of 9 users to 1885. The effect on the trajectories network can be appreciated in Figure 11(c). The most important festivity of Mouridisme is the Grand Magal of Touba. It should be noticed that 2013 was a peculiar year, because there were two of these holidays: on the 1st of January and on the 20th of December. For the Grand Magal of the 20th of December, the number of net incoming users to the municipality of Ndame, where Touba is located, experimented an increase from an average value of 64 per day to 19 054. In Figure 11(d), the effects of this festivity on the trajectories network can be observed.

Looking at a larger spatial scale, we have computed the time series of net incoming flow to the different areas defined by the communities of the phone calls network described in Section IV B. As previous work shows,[8] this approach can give a more reasonable partition of a country than its administrative divisions. The resulting time series for communities 6, 7, 8 and 11 are presented in Figure 12. There, we can notice the presence of some peaks. Some of them appear in couples, with one positive peak and one negative peak, like the two couples of peaks at the beginning of the year in Figure 12C or the one that happens around June in Figure 12B. These peaks imply a large displacement of users that travel to a given area one day and leave approximately the next one. On the other hand, we have isolated positive peaks caused by users that travel to a given area on a particular

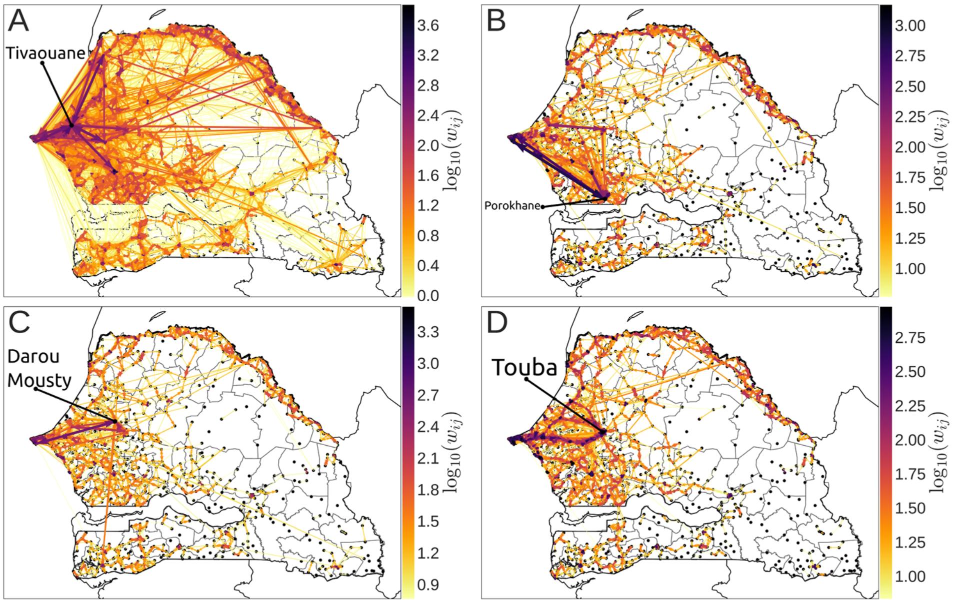

FIG. 11. Daily trajectories network displaying user displacements during some local religious holidays: The *mawlid* or birth of Mohammed, which is mainly celebrated in the city of Tivaouane (a); the Magal of Porokhane (b); the Magal of Darou Mousty (c); and the Grand Magal of Touba (d).

date and stay during a longer period of time, like those that appear on Figures 12B and 12C in August and October.

The first kind of peaks can be attributed to religious festivities involving some kind of pilgrimage or *Magal*, as the ones described above. The second kind of peaks could imply seasonal migration of workers, since they appear mainly at the same dates as the ones at which we detected activity increments in the time series of phone calls as well as a fragmentation in the community structure of the trajectories networks.

Finally, in order to further analyze the effects of seasonal migration on the mobile phone network, we have used dataset 3, which contains information about user mobility at the municipality level. Following the procedure described

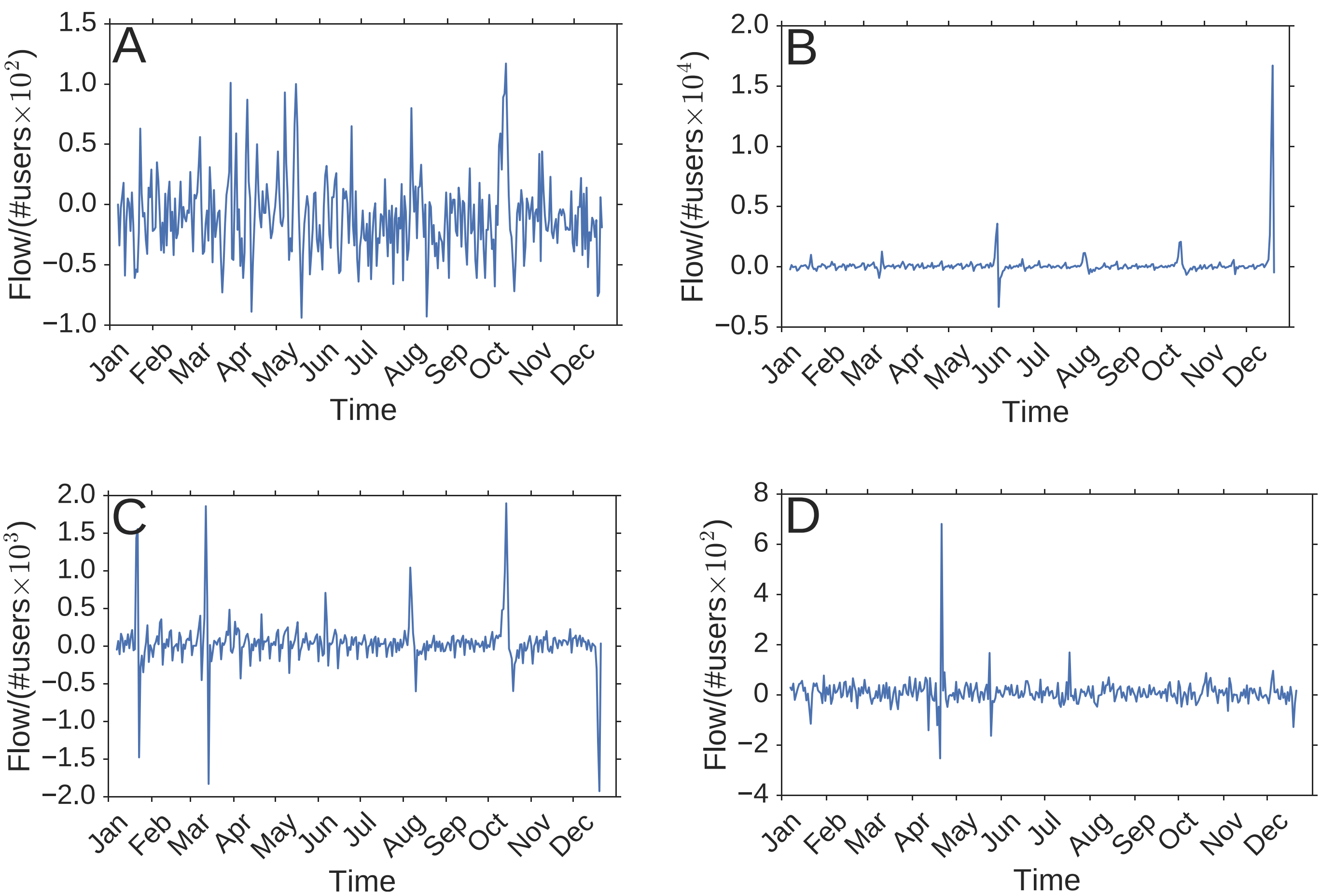

FIG. 12. Time series of net incoming flow of users to some of the communities found in the phone calls network as described in Figure 7(b). Panels (a)–(d) correspond respectively to communities 6, 7, 8, and 11.

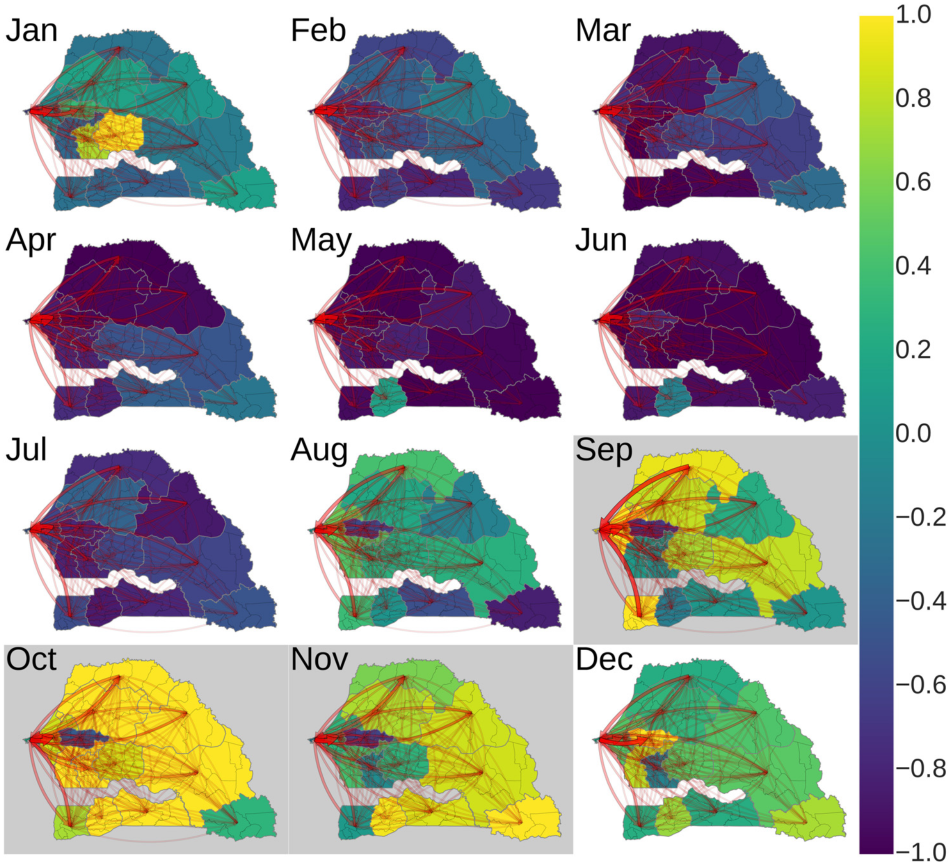

FIG. 13. Maps that show the monthly migration networks at the administrative region level. Link width corresponds to link weight. Arrows are traced from the *regular home* to the *monthly home*. The colors of each region give a notion of the number of *foreign* users within each region; that is, users whose home during a given month is not their regular home. The shadowed panels correspond to the harvest season.

in Section III, we have assigned a *regular home* as well as *temporary homes* to each user. Then, we have built a migration network for different temporal scales, where the nodes are the municipalities of Senegal and two municipalities i and j are joined through a directed link if a user that has her *regular home* in municipality i, lives temporarily in municipality j.

We have aggregated the results at the administrative regions level and at a monthly temporal scale. In order to do that, we have counted the number of users whose *monthly home* was not their *regular home* inside a given region. However, we have applied a distance filter, such that we only take into account those users whose *monthly home* is farther than 15 km from their *regular home*. By doing this, we avoid getting false positives. Notice that, since the regions are the nodes of the aggregated network built from the municipalities network, this aggregated network has self loops: a given region is linked to itself and the link weight w_{ii} is the number of users that have migrated between municipalities within that region.

In order to better visualize the migration networks, we have built time series of the number of *foreign* users per month within each region. Then, we have normalized them between -1 and $+1$ such that a point in the time series represents a relative distance to the maximum and the minimum of the time series. This way, all the time series are comparable point by point. The results are shown in Figure 13, where the color associated to each region corresponds to the value of the normalized time series of *foreign* users for a particular month. We can see that, whereas in the first half of the year most of the regions remain quite inactive, there is a global increment of *foreign* users throughout the country between September and November. Notice that these months correspond roughly to the time when the main harvest takes place (see top panel of Figure 1). The only region whose values seem to remain closer to the minimum than to the maximum during that period is the region of Diourbel, which includes the holy city of Touba, where the Grand Magal is celebrated each year. This religious festival gathers up to 3×10^6 people, and in 2013 it was celebrated the 20th of December. This is the reason for the maximum of the number of *foreign* users to be shifted to the month of December.

V. CONCLUSIONS

By analyzing the communication and mobility patterns of the population of Senegal through mobile phone data, we have unraveled correlations with economic processes like the agricultural activity. We have also found connections with cultural events like religious festivities.

The analysis of time series of outgoing phone calls for 1666 antennas located throughout Senegal has enabled us to determine the calling habits of the Senegalese people. By computing the time series of the number of outgoing calls per day and associating each antenna to a time series of Normalized Difference Vegetation Index (NDVI), we have shown that there exists a correlation between phone calls and agricultural activity. The structure of the time series of phone calls can also be related to the celebration of two Islamic holidays: Eid al-Fitr (Festival of Breaking the Fast, the end of Ramadan) and Eid al-Adha (Festival of the Sacrifice).

By building phone calls networks between the antennas and computing their community structure, we have revealed the communication patterns between the people of Senegal. The communities that emerge from this arrangement have clear geographical boundaries, implying that users have a preference for short distance calls. Moreover, the geographical distribution of the communities resembles approximately the delimitations of the map of land use of Senegal, which constitutes another correlation between agriculture and the mobile phone network.

The mobility patterns that we have detected through the study of the community structure of trajectories networks indicate an increase in the number of communities during periods that are closely related to agriculture. This fragmentation of the community structure of the trajectories network can be explained by an alteration of the user behavior: users move more locally on those periods than on a regular day. The preference for short range displacements can be attributed mainly to the organization of agricultural labors.

Looking at the flows of users at municipality level and daily temporal scale, we have been able to detect several movements of large masses of people. These phenomena occur when religious holidays, like the Grand Magal of the city of Touba, take place in a particular location. In order to detect seasonal migrations, we have aggregated our data at a monthly scale and at the administrative regions level. Our results clearly show a general increment of *migrants* (users that temporary live in a place that is not their regular home) throughout Senegal during September, October, and November, a period that coincide with the harvest season in Senegal.

Finally, we can conclude that our study shows a clear correlation between an increase of phone calls and migratory flows with the agricultural activity of Senegal.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the Ministerio de Economía y Competitividad (MINECO) under Contract Nos. MTM2012-39101 and MTM2015-63914-P. The data used in this work were made available by ORANGE/SONATEL within the framework of the D4D Challenge: Senegal.

[1]International Telecommunication Union, "World telecommunication development conference (WTDC-2014): Final report," in *World Telecommunication Developement Conference* (2014).
[2]GSMA Intelligence, "The mobile economy: Sub-Saharan Africa 2014," Technical Report No. 2014, GSMA, 2014.
[3]M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," Nature **453**, 779–782 (2008).
[4]J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, "Structure and tie strengths in mobile communication networks," Proc. Natl. Acad. Sci. **104**, 7332–7336 (2007).
[5]J. Giles, "Making the links," Nature **488**, 448–450 (2012).
[6]Y.-A. Montjoye, J. Quoidbach, F. Robic, and A. S. Pentland, "Predicting personality using novel mobile phone-based metrics," in *Social Computing, Behavioral-Cultural Modeling and Prediction: Proceedings of the 6th International Conference, SBP 2013, Washington, DC, USA, April 2–5, 2013* (Springer, Berlin, Heidelberg, 2013), pp. 48–55.
[7]V. Frias-Martinez and E. Frias-Martinez, "Spectral clustering for sensing urban land use using twitter activity," Eng. Appl. Artif. Intell. **35**, 237–245 (2014).
[8]C. Herrera-Yagüe, C. M. Schneider, T. Couronné, Z. Smoreda, R. M. Benito, P. J. Zufiria, and M. C. González, "The anatomy of urban social networks and its implications in the searchability problem," Sci. Rep. **5**, 10265 (2015).
[9]N. Xiao, S. Cai, M. Moritz, R. Garabed, and L. W. Pomeroy, "Spatial and temporal characteristics of pastoral mobility in the far north region, Cameroon: Data analysis and modeling," PLoS One **10**, 1–30 (2015).
[10]A. Wesolowski, C. O. Buckee, D. K. Pindolia, N. Eagle, D. L. Smith, A. J. Garcia, and A. J. Tatem, "The use of census migration data to approximate human movement patterns across temporal scales," PLoS One **8**, e52971 (2013).
[11]C. O. Buckee, A. Wesolowski, N. N. Eagle, E. Hansen, and R. W. Snow, "Mobile phones and malaria: Modeling human and parasite travel," Travel Med. Infect. Dis. **11**, 15–22 (2013).
[12]Y. Dong, F. Pinelli, Y. Gkoufas, Z. Nabi, F. Calabrese, and N. V. Chawla, "Inferring unusual crowd events from mobile phone call detail records," in *Machine Learning and Knowledge Discovery in Databases, European Conference, ECML PKDD 2015, Porto, Portugal, September 7–11, 2015, Proceedings* (Springer International Publishing, Cham, 2015) pp. 474–492.
[13]A. J. Morales, W. Creixell, J. Borondo, J. C. Losada, and R. M. Benito, "Characterizing ethnic interactions from human communication patterns in Ivory Coast," Networks Heterog. Media **10**, 87–99 (2015).
[14]Research Program on Climate Change, Agriculture and Food Security, "The impact of climate information services in Senegal," Technical Report No. 3, Consultative Group for International Agricultural Research, 2015.
[15]M. Ndiaye, "Senegal agricultural situation country report," Technical Report No. SG7001, USDA Foreign Agricultural Servie, 2007.
[16]"Comprehensive food security and vulnerability analysis. Livelihood zone descriptions," Technical Report No. 2011, World Food Program, United Nations Food and Agriculture Organization, Centre de Suivi Ecologique, Commissariat National á la Sécurité Alimentaire and FEWS NET, 2011.
[17]T. Judah, "Senegal's Mourides: Islam's mystical entrepreneurs," see http://www.bbc.com/news/world-africa-14344082, 2011 (last accessed May 27, 2016).
[18]A. Thomson, "Senegal's powerful Mourides have global reach," see http://www.reuters.com/article/us-senegal-caliph-mourides-idUSL2919142320071229, 2007 (last accessed May 27, 2016).
[19]C. Boone, *Merchant Capital and the Roots of State Power in Senegal: 1930–1985* (Cambridge University Press, 2006), pp. 40–47.
[20]C. Ratti, S. Sobolevsky, F. Calabrese, C. Andris, J. Reades, M. Martino, R. Claxton, and S. H. Strogatz, "Redrawing the map of Great Britain from a network of human interactions," PLoS One **5**, e14248 (2010).
[21]A. Llorente, M. Garcia-Herranz, M. Cebrian, and E. Moro, "Social media fingerprints of unemployment," PLoS One **10**, e0128692 (2015).
[22]Y. de Montjoye, Z. Smoreda, R. Trinquart, C. Ziemlicki, and V. D. Blondel, "D4D-Senegal: The second mobile phone data for development challenge," e-print arXiv:1407.4885 [cs.CY].
[23]See https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod09a1 (2002–2014) for MOD09A1 product: Surface reflectance 8-day l3 global 500m (last accessed May 27, 2015).
[24]V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," J. Stat. Mech.: Theory Exp. **2008**, P10008.
[25]A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring network structure, dynamics, and function using NetworkX," in *Proceedings of the 7th Python in Science Conference*, edited by G. Varoquaux, T. Vaught, and J. Millman (Pasadena, CA, USA, 2008), pp. 11–15.
[26]M. E. J. Newman, "Modularity and community structure in networks," Proc. Natl. Acad. Sci. **103**, 8577–8582 (2006).
[27]See http://www.saphirnews.com/Aid-al-Fitr-2013-vendredi-9-aout-au-Maroc-et-au-Senegal_a17362.html for Aïd al-Fitr 2013: vendredi 9 août au Maroc et au Sénégal (2013) (last accessed May 27, 2016).
[28]See http://www.saphirnews.com/Aid-el-Kebir-2013-le-16-octobre-au-Maroc-deux-dates-au-Senegal_a17707.html for Aïd el-Kébir 2013: le 16 octobre au Maroc, deux dates au Sénégal (2013) (last accessed May 27, 2016).
[29]S. Fortunato and M. Barthélemy, "Resolution limit in community detection," Proc. Natl. Acad. Sci. **104**, 36–41 (2007).