NETWORKS AND HETEROGENEOUS MEDIA doi:10.3934/nhm.2015.10.xx
©American Institute of Mathematical Sciences
Volume **10**, Number **1**, March **2015** pp. **X–XX**

CHARACTERIZING ETHNIC INTERACTIONS FROM HUMAN COMMUNICATION PATTERNS IN IVORY COAST

Alfredo Jose Morales

Grupo de Sistemas Complejos, Universidad Politécnica de Madrid
ETSI Agrónomos, 28040, Madrid, Spain

Werner Creixell[1,2,3], Javier Borondo[1]
Juan Carlos Losada[1] and Rosa Maria Benito[1]

[1] Grupo de Sistemas Complejos, Universidad Politécnica de Madrid
ETSI Agrónomos, 28040, Madrid, Spain
[2] Departamento de Electrónica. Universidad Técnica Federico Santa María
Valparaiso, Chile
[3] Visiting researcher at Center for Spatial Information Science (CSIS)
University of Tokyo. Japan

Abstract. Towards the consolidation of peace and national development, Ivory Coast must overcome the lack of cohesion, responsible for the emergence of two civil wars in the last years. As in many African countries, ethnic violence is a result of the way territories are organized and the prevalence of some groups over others. Nowadays the increasing availability of electronic data allows to quantify and unveil societal relationships in an unprecedented way. In this sense, the present work analyzes mobile phone data in order to provide information about the regional and ethnic interactions in Ivory Coast. We accomplish so by means of the construction and analysis of complex social networks with several types of interactions, such as calling activity and human mobility. We found that in a subregional scale, the ethnic identity plays an important role in the communication patterns, while at the interregional scale, other factors arise like economical interests and available infrastructure.

1. Introduction.

In the recent decades, African countries have gone through several armed conflicts among different ethnic and religious groups. The borders arbitrarily traced by Europeans for administrative convenience of the former colonial order split and joined ethnic groups into new countries, forcing them to coexist within previously nonexistent frontiers. Asymmetries in economical and geographical benefits between different ethnic groups have led some countries to different levels of social polarization, which have eventually resulted in civil wars. Recent studies have shown that violence emerges between ethnic groups when their territories are not well defined [13], or when a group is large enough in order to prevail among others, but not as strong as to maintain order. Ivory Coast is not an exception of this context. In less than two decades the Ivorians have engaged in two internal armed conflicts, due to profound asymmetries between their inhabitants.

2010 *Mathematics Subject Classification.* 91D30, 05C82, 91C20.

Key words and phrases. Data for development, ethnic interactions, call detail records, complex networks, mobility, community structure.

This research was supported by the Ministry of Economy and Competitiveness-Spain under Grand No. MTM2012-39101.

Therefore, the characterization and understanding of their ethnic relationships is crucial to consolidate peace and to strengthen the social cohesion needed for any further economical development.

The increasing availability of user generated data is changing the way scientists understand, analyze and model the human behavioral patterns and societal phenomena [11, 18]. Such data result from the accumulation of individual traces produced by people as they interact either with electronic devices or Internet social networks. Recently, the analysis of electronic traces has enhanced our knowledge about how people behave, the resulting structure of their social networks and the way people influence each other [14]. In fact, these studies are unveiling the characteristics of societies as a whole physical system, rather than a collection of isolated individuals. For instance, the diversity of connections can explain the economical development of cities [9], as well as the emotional state of individuals [19]. Also, topics popularity can explain the economical value of stocks [6] or locate earthquake epicenters [20]. Finally, patterns of mobility can predict the propagation of infectious diseases [22] or evaluate urban land use [21]. This knowledge represents an opportunity for governments and companies to gain intelligence about their social systems without the need of deploying unnecessary fieldwork.

The analysis of user generated data to improve social well-being is a very timely subject that has attracted the attention of several researchers, as well as governmental and international organizations over the last years. During the 2010 earthquake in Haiti, mobile phone activity resulted to be an accurate source of data to estimate human migrations after the natural disaster and the cholera outbreaks [2]. In Kenya, a similar approach may remarkably reduce the spread of contagious diseases like Malaria [22]. Furthermore, in Mexico, the most affected antennas presented a dramatic change in their activity during the large floods occurred in 2009, allowing scientists to reconstruct the population reaction to the catastrophe [17]. These studies are very important, since their results may benefit a large amount of human population, by improving and enhancing the efficacy and efficiency of governmental processes of strategic planning.

In this work we characterize and quantify ethnic and regional interactions in Ivory Coast by means of the study of their mobile phone activity patterns. We construct and analyze complex social networks [5] from human trajectories and calling activity to infer some of the properties of the social system structure [11]. In summary, we found that on a local and regional scale the linguistic identity strongly influences the ethnic interactions, while on a national and wider scale the underlying infrastructure and economical motivations play a major influential role. In fact, we have measured asymmetrical interactions between the regions of economical differences. The poorer and less developed regions of the north seem to preferentially communicate and migrate to the wealthier adjoined regions in the south, and not otherwise. The fact that these two regions engaged in war a couple of years ago, highlights the importance of quantifying interactions to understand to which extent conditions are set for violence emergence before it actually occurs.

The present paper is organized as follows. In section 2 we explain the datasets that we have analyzed. In section 3 we describe the networks that we have built to represent the social structure of Ivory Coast and we also explore their characteristics. In sections 4 and 5 we analyze the ethnic composition of the networks and describe the way that the ethnic groups have interacted among each other.

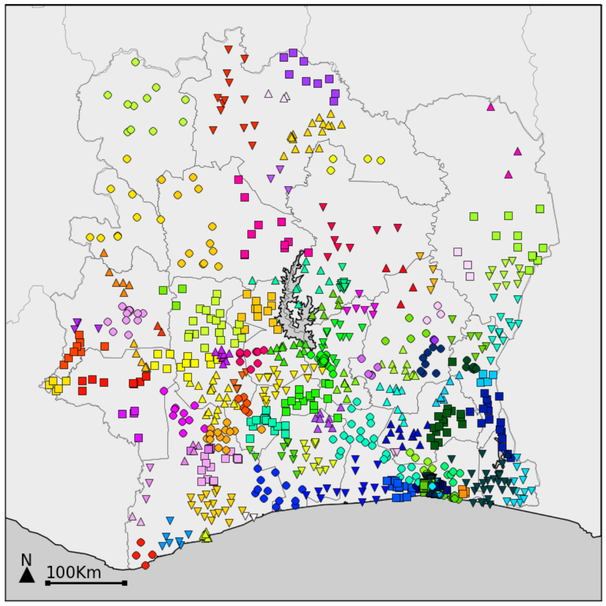

FIGURE 1. Mapping the community structure of the trajectories network of Ivory Coast. Antennas represent nodes and are plotted in different colors and shapes, indicating the network community they belong to.

2. **Data set.** Mobile phones datasets are made out of Call Detail Records (CDR). These are produced by any phone call or SMS in the communication provider data bases. A CDR usually contains information about the origin and destination phone numbers, starting time of each call and its duration. The antennas that served the subscriber are also given. For telephone service providers, CDRs are critical for the production of the monthly bill. For researchers CDRs represent opportunities to understand nationwide calling and mobility patterns.

The present work is based on data from CDRs provided by France Telecom / Orange Côte d'Ivoire within the framework of the Data for Development D4D Challenge [3]. The data was collected for 150 days, from December 1, 2011 until April 28, 2012. The set of collected CDRs contains 2.5 billion calls and SMS exchanges between around 5 million anonymized users.

The first dataset contains the number and duration of all calls between any pair of antennas on hourly basis. Antennas are uniquely identified and their geographical location is known. The second dataset contains individual user trajectories through cell phone antennas during calls. This dataset contains the trajectory of 50,000 individuals during lapses of two weeks. The information provided regards the random identifier, the time-stamp when the call was made and the antenna from which it was made. The geographical location of all antennas are also known.

3. **Characterizing populated areas.** In order to characterize populated areas in Ivory Coast we studied the structure of the human trajectories network at the meso-scale level. This network displays the people's mobility patterns within a given territory. It is built out of the aggregation of individual trajectories. Each trajectory is defined as the sequential set of antennas that served a particular user in time. Antennas represent nodes and an edge is created between two antennas, i and j, if a user makes two consecutive calls, first from antenna i and later from antenna j. The edges are directed, from i to j, and weighted according to the number of times that all users performed the same trajectory. The resulting network has 1,215 nodes and 187,102 edges. A visualization of the dynamical growth of this graph during an arbitrary day is presented in [23].

The modularity quantifies the fraction of edges between groups of nodes in comparison to the expected number if edges were randomly distributed [15]. By applying the community detection algorithm based on modularity optimization [4], we found that the trajectories network could be classified in 100 network communities, which are shown in Fig. 1 together with the map of Ivory Coast. Communities comprehend a limited territorial area, not necessarily contained within the same regional borders, and are related to urban and rural settlements. It can be noticed that there is a larger density of antennas and communities in the south side of the country, while in the north side scattered antennas conform a few communities. Such difference in the density of antennas and communities is consistent with demographic information that reports the south side of Ivory Coast as more densely populated.

The density of edges also display the same structure. A snapshot of the trajectories network is presented in Fig. 2. The nodes are located at the antennas' geographical coordinates and edges are colored in blue. The intensity of the edge is proportional to the edge's weight, which means that the most intense edges represent the trajectories more frequently used. The main cities (black circles) and southern regions concentrate a larger amount of edges than the north side, indicating a remarkable difference in the amount of human displacements between the two regions. Apart from demographic density, these patterns also result from the underlying infrastructure and economical activity. In Fig. 2 we have overlapped in red color the main roads of the country [7]. Most trajectories keep a remarkable correspondence to available roads. Some of them seem to be more frequently used, like the ones linking the north with the south of the country; while others are less frequently used, like the traversal road up in the north. The fact that some infrastructures are more frequently used than others can be a consequence of the fact that the region with more activity showed in Fig. 2 corresponds to the zone of cocoa plantations. Ivory Coast is the largest cocoa producer in the world with 36% of the global share [1].

It has been stated that the economical development of large regions can be characterized and understood by means of cellphone activity patterns [9]. Accordingly, in this study we have analyzed the closeness-centrality property of the antennas in the trajectories network. This network property is inversely proportional to the average distance from a node to the rest of the network in terms of connections. It provides information about the central or peripheral behavior of nodes or regions according to all human displacements. In the top panel of Fig. 3, we present the trajectories network coloring the edges according to the mean value of antennas' closeness-centrality. Red regions are highly central, yellow and pale blue regions are intermediate, and dark blue regions are peripheral. It can be noticed, that the most

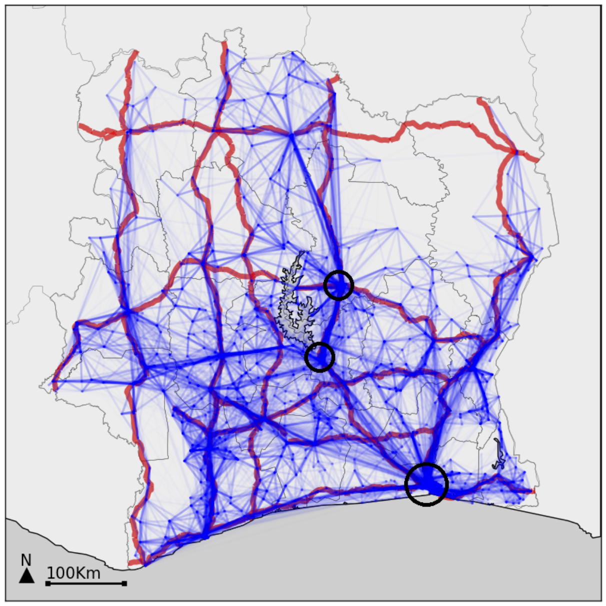

FIGURE 2. Mapping the structure of the trajectories network on the Ivory Coast geographical map. The blue lines represent the edges of the network and their intensity is proportional to the edge weight. Superimposed the main roads of Ivory Coast have been plotted as red lines. The location of the country's main cities are marked with black circles.

central area (red) corresponds to the main city and the regions it adjoins, while the most peripheral regions are located in the north and west sides (blue). The same effect can be seen if we color nodes according to their closeness-centrality value, as shown in the bottom panel of Fig. 3. This is in agreement to international reports [8] that identify the north and the west side of the country as the less developed areas.

4. **Ethnic interactions.** In order to understand the ethnic composition of this graph, we have taken into account the ethnic and linguistic identity of each network community. For this purpose, we mapped each community to its geographically closest ethnic group, according to the location of the communities' most connected antenna and the ethno-linguistic information explained in the Appendix A. Although there are other methods to map ethnic groups, such as the majority rule, we have chosen this criterion because the most connected antenna indicates where most of social activity takes place. In Fig. 4 we present the trajectories network by coloring edges according to the linguistic family. It can be seen how the most densely connected areas, like the capital city or the cities in the center of the country (black circles), concentrate links from different linguistic areas, while most of regions mainly present trajectories within their own linguistic family.

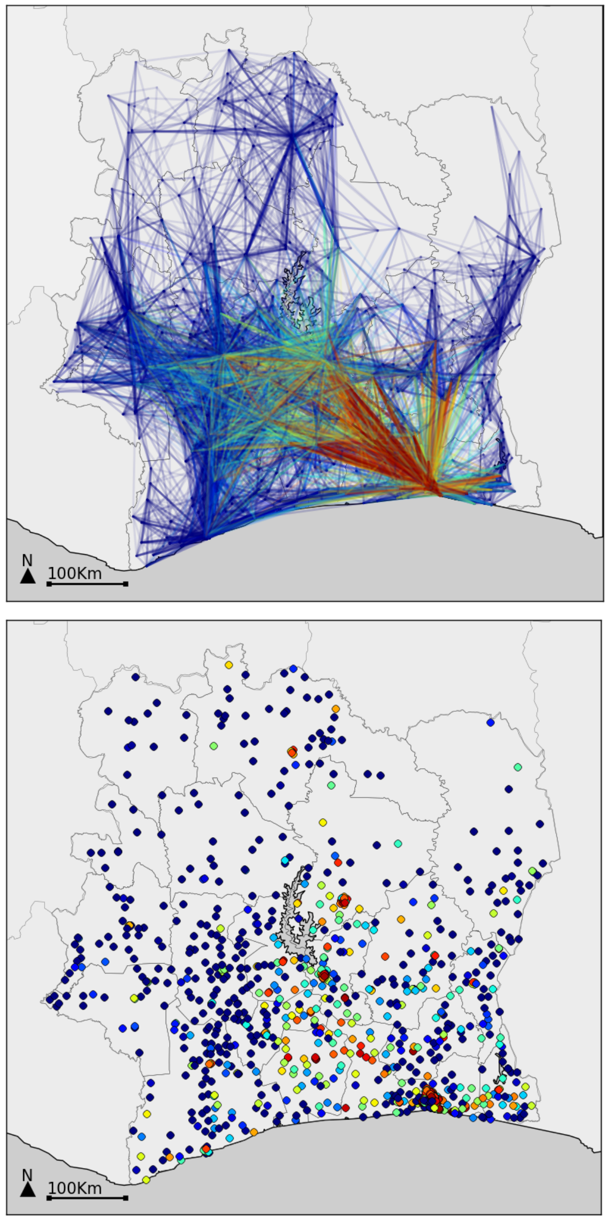

FIGURE 3. Mapping the closeness-centrality property of the trajectories network in Ivory Coast. In the top panel, edges have been colored according to the closeness centrality mean value of the two connected nodes. In the bottom panel, nodes have been colored according to their corresponding closeness centrality value. At both panels, the red regions indicate higher closeness-centrality, the yellow and pale blue regions indicate medium centrality, and the dark blue regions indicate lower closeness-centrality.

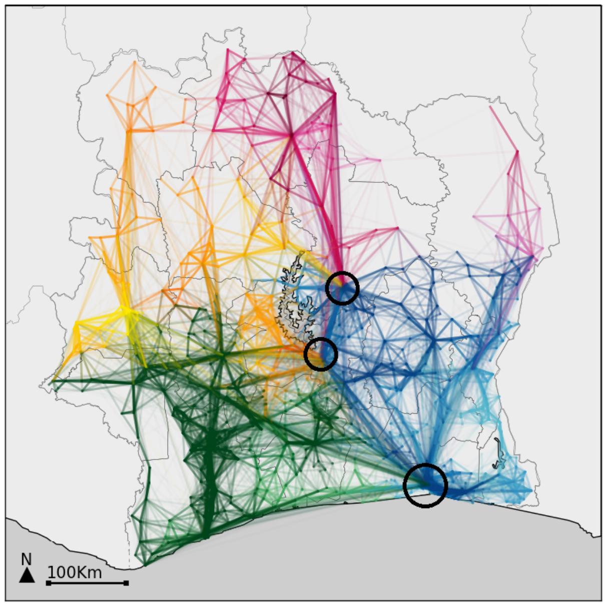

FIGURE 4. Mapping the linguistic identity of the trajectories network of Ivory Coast. The edges have been colored according to the linguistic group to which the most connected antenna at each community belongs to. There are four major linguistic families represented in yellow (northwest), purple (northeast), green (southwest) and blue (southeast). Black circles indicate the location of the major cities.

After mapping the ethnic groups, we have constructed a second network taking into account a new type of interaction, such as the antenna-to-antenna calling information (see section 2). In this network, the nodes represent the 100 communities found in the trajectories network (see section 3) whose ethnic identity is already known. The edges correspond to the number of calls made from one community to the other. The edge direction goes from the emitter community to the receiver community and the weight is equal to the number of occurrences found in the datasets.

In order to get a clearer view of the way that ethnic groups communicate with each other, we present in Fig. 5 A the weighted adjacency matrix of the ethnic groups calling network normalized by row. This normalization provides relative information about the destination and origin of outgoing and incoming calls by group. The diagonal entries of the matrix are higher than the other elements indicating that most of outgoing calls remain in the same community. In fact, the preference of people to communicate with similar ones increases with the social scale of the network. When we aggregate the communities by ethnic group and linguistic family (Fig. 5 B and C), the assortative coefficient [16] of each matrix increases from $r \sim 0.5$ to $r \sim 0.8$ (Fig. 5 D), being $r = 1$ the case of absolute segregation. Such

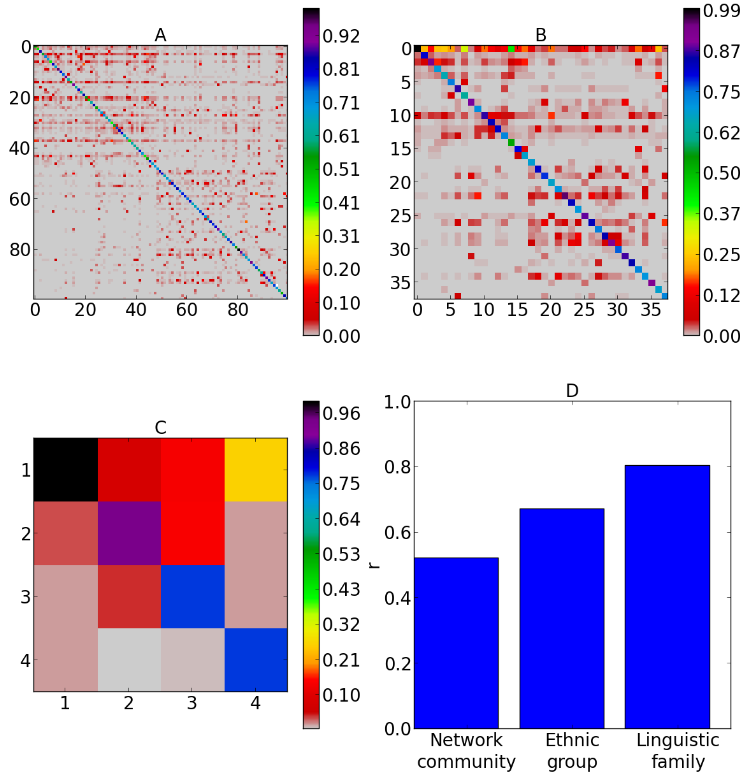

FIGURE 5. Normalized adjacency matrices of the calls network corresponding to the community structure from the trajectories network (A), ethnic group aggregation (B) and linguistic family aggregation (C). Assortativity coefficient of selectiveness to call on local scale (community), subregional scale (ethnic group) and regional scale (linguistic family) (D).

increase indicates that there is a higher segregation between ethnic groups when we consider their linguistic family.

Moreover, not all families behave the same way. The southern families (number 1 and 2 in Fig. 5 C) present a larger proportion of calls directed to their own linguistic family, in comparison to the northern families (number 3 and 4 in Fig. 5 C), whose activity directed to other linguistic families is relatively larger. In Fig. 6, we present the intra-family flux (calls directed to the same linguistic family) and inter-family flux (calls directed to a different linguistic family) of calls. In the figure the symbols represent communities from the trajectories network and the color corresponds to the linguistic family they belong to. The further the community is located below the dashed line of slope 1, the higher the family internal traffic in comparison to the external traffic. Most of the southern ethnic groups (blue and green dots) are farther from the diagonal line than the northern ones (yellow and red dots). This means that the internal traffic in southern ethnic groups is much higher than their external one, while on northern families the external traffic is comparable with the internal one.

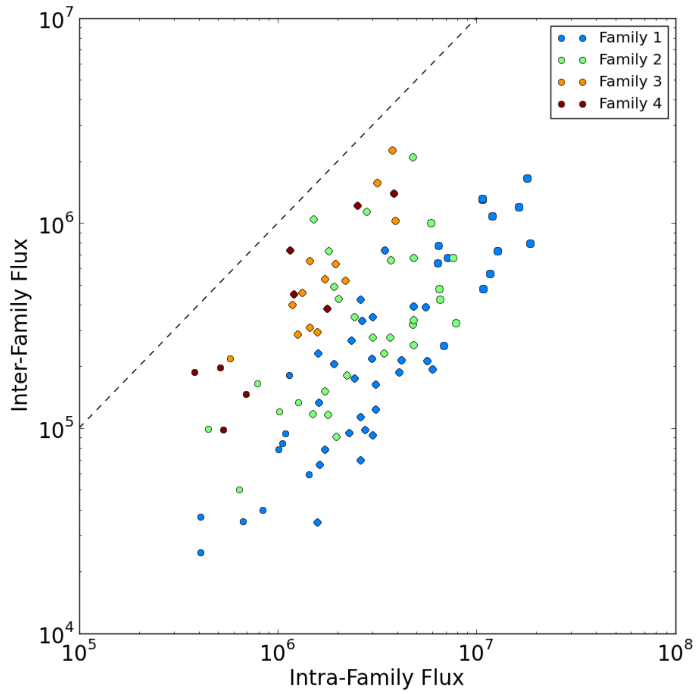

FIGURE 6. Scatter plot of intra linguistic family flux (calls directed to an antenna in the same linguistic family as the emitter antenna) versus inter linguistic family flux (calls directed to an antenna in a different linguistic family than the emitter antenna). Symbols represent communities from the trajectories network and the color indicates the linguistic family to which the community belongs. The dashed line has slope 1.

The external calling traffic from the northern ethnic groups is directed selectively towards their adjoin southern families. In Fig. 5 C, we see that the family 1 and 4 are more connected between them than with the rest of families; while family 2 and 3 behave the same. Such observation is in good agreement with the mobility patterns shown in Fig. 2, where the vertical roads seem to have a higher significance than the horizontal ones; as well as with the patterns shown in Fig. 4, where we showed that the mobility of the northern families to the south are stronger with the adjoin regions.

5. **Effects of selectiveness in the calling behavior.** To further understand the selectiveness in the communication patterns between the east and west side of the country, we built a third network taking into account another type of social interactions. Specifically we built a network from the calling behavior at the tower level, extracting only information from the first dataset described in section 2. The nodes in this network also represent single antennas, and an edge is created from the antenna i to the antenna j, when a user that is being served by the antenna i makes a call to another user who is served by the antenna j. The resulting is a directed

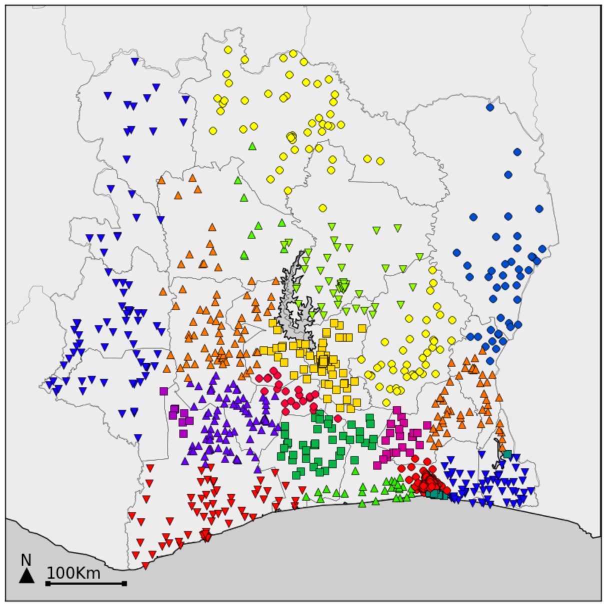

FIGURE 7. Mapping the community structure of the calls network of Ivory Coast. Antennas represent nodes and are plotted in different colors and shapes, according to the community they belong gotten from the community detection algorithm.

and weighted network, where the weight of the edges represents the total number of calls made from the antenna i to the antenna j along the whole observation period.

This calls network is composed by over 19 communities of antennas, according to the modularity optimization algorithm [4]. The distribution of this communities along the geography of Ivory Coast is shown in Fig. 7. The communities show a relationship with administrative areas marked with gray lines, although at some cases these human borders are not in correspondence to the political ones. In [23], we present an animation with the dynamics of this network and a visualization of the influence that each of the 19 communities have among each other.

In order to capture how communities influence the rest of the network, we analyzed the density of calls directed to the given communities from the rest of antennas. To quantify such preference, we have classified these communities using a k-means clustering algorithm [10], according to the density of calls to the rest of communities.

The results are presented in Fig. 8, where we have plotted the antennas with different colors, according to the classifier results. We found that the country is divided between the east side and west side of the map, as was previously intuited in the Fig. 5 C.

6. **Conclusions.** By means of the analysis of the emergent patterns from human trajectories and calling networks, we have characterized the interactions and resulting structure of the diverse geographical and social areas of Ivory Coast.

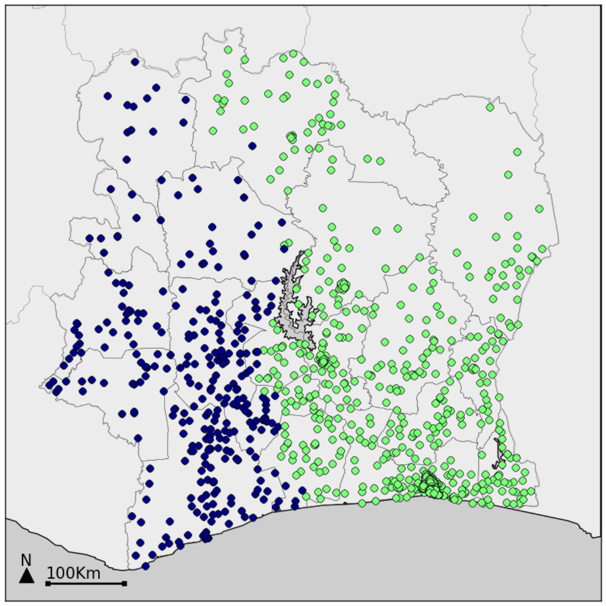

FIGURE 8. Mapping the classification results of antennas according to the way the calls network communities are related. A k-means clustering classifier has been applied to the community structure of the calls network.

From a social and ethnic perspective, we found that the linguistic identity plays a fundamental role in the communication patterns of this country. The Ivorian people, seem to preferentially communicate to those that belong to the same local community, but more drastically to those that share the same linguistic family. Yet these preferences are not equal to all linguistic families. The peripheral regions of the north seem to communicate with their adjoin southern regions more significantly than otherwise. This behavior may be explained due to economical reasons. The north side has to trade with the south and therefore that is reflected in the communication patterns observed. Another reason could be the result of domestic migrations from the less developed areas in the north to the southern and more economically developed areas. Therefore, it could be that the communication patterns reflected in Fig. 6 can just be an artifact of intra-family communications. More evidence is needed in order to clearly distinguish the real influence of both cases.

As a result, the Ivorian communication map is organized in two interacting regions located at the east and west side of the country. On this basis, we conclude that the geographical and social factors, whether cultural or economical, determine the structural features of the social interchange. In the sense that on a local and subregional scale, the ethno-linguistic factor determines the interaction patterns, while on a wider scale, the available infrastructure and economic facts play a major influence in the social dynamics.

Acknowledgments. This research was supported by the Ministry of Economy and Competitiveness-Spain under Gran No. MTM2012-39101. Special thanks to Basal

Centre FB0821 CCTVAL and the “Beca Iberoamérica Santander Universidades 2012” for its travel financing (W.C.). Also special recognition to France Telecom / Orange Côte d'Ivoire for organizing the Data for Development D4D Challenge in 2013.

Appendix A. Ethno-linguistic groups in Ivory Coast. Ivory Coast presents a complex society composed by more than 60 different ethnic groups. Although French is the official language and it is broadly spoken along the country, each ethnic group has its own language. Such many and diverse languages are classified into four large linguistic families: Kwa, Kru, Mande and Gur [12]. The territories of these four linguistic families are well defined in the four coordinates of the country.

In summary, the Kwa group is located in the southeast side of the country. This is the most economically developed region where the capital city and other major cities are located, as well as the main Ivorian airport and seaport. The Kru group is located in the southwest side, also in the Atlantic coast. The second seaport in Ivory Coast is located at this region, which brings economical benefits to these people. The Mande group is found in the northeast side of the country, and the northwest region is occupied by the Gur family. The northern regions occupied by the Mande and Gur groups are the least populated regions of the country and less economically developed areas.

REFERENCES

[1] Annual Report 2011/2012, Report from International Cocoa Organization (ICCO).
[2] L. Bengtsson, X. Lu, A. Thorson, R. Garfield and J. von Schreeb, Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in Haiti, *PLoS Medicine*, **8** (2011), e1001083.
[3] V. D. Blondel et al. Data for development: The D4D challenge on mobile phone data, arXiv:1210.0137, 2012.
[4] V. D. Blondel, J.-L. Guillaume, R. Lambiotte and E. Lefebvre, Fast unfolding of communities in large networks, *J. Stat. Mech.*, **10** (2008), P10008.
[5] S. Bocaletti, V. Latora, Y. Moreno, M. Chavez and D.-U. Hwang, Complex networks: Structure and dynamics, *Phys. Rep.,* **424** (2006), 175–308.
[6] J. Bollen, H. Mao and X.-J. Zeng, Twitter mood predicts the stock market, *J. Comput. Science,* **2** (2011), 1–8.
[7] Cote d'Ivoire Roads Shapefiles, Service from African Development Bank. Available from: http://www.infrastructureafrica.org/library/doc/986/cote-divoire-roads.
[8] Document de strategie pays et programe indicatif national pour la periode 2008-2013, Report from Communaute europeenne - Republique de Côte d'Ivoire, 2008. Available from: http://ec.europa.eu/development/icenter/repository/scanned_ci_csp10_fr.pdf.
[9] N. Eagle, M. Macy and R. Claxton, Network diversity and economic development, *Science,* **328** (2010), 1029–1031.
[10] T. Hastie, et al., *The Elements of Statistical Learning*, Springer, New York, 2009.
[11] D. Lazer, et al., Life in the network: The coming age of computational social science, *Science,* **323** (2009), 721–723.
[12] M. P. Lewis, *Ethnologue: Languages of the World*, 16^{th} edition, SIL International, Dallas, Tex., 2009.
[13] M. Lim, R. Metzler and Y. Bar-Yam, Global pattern formation and ethnic/cultural violence, *Science,* **317** (2007), 1540–1544.
[14] A. J. Morales, J. Borondo, J. C. Losada and R. M. Benito, Efficiency of human activity on information spreading on Twitter, *Social Networks,* **39** (2014), 1–11.
[15] M .E. J. Newman, Modularity and community structure in networks, *Phys. Rev. E,* **103** (2006), 8577–8582.
[16] M. E. J. Newman Mixing patterns in networks, *Physical Review E,* **67** (2003), 026126, 13pp.

[17] D. Pastor, A. J. Morales, Y. Torres, J. Bauer, A. Wadhwa, C. Castro-Correa, A. Calderón-Mariscal, L. Romanoff, J. Lee, A. Rutherford, V. Frias-Martinez, N. Oliver, E. Frias-Martinez and M. Luengo-Oroz, Flooding through the lens of mobile phone activity, in *IEEE Global Humanitarian Technology Conference (GHTC)*, IEEE, 2014, 279–286.

[18] A. Pentland, *Social Physics: How Good Ideas Spread. The Lessons From a New Science*, Penguin Group (USA) Incorporated, 2014.

[19] K. K. Rachuri, M. Musolesi, C. Mascolo, P. J. Rentfrow, C. Longworth and A. Aucinas, EmotionSense: A mobile phones based adaptive platform for experimental social psychology research, *Proceedings of the 12th ACM International Conference on Ubiquitous Computing* (New York, NY, USA), Ubicomp '10, ACM, 2010, 281–290.

[20] T. Sakaki, M. Okazaki and Y. Matsuo, Earthquake shakes twitter users: Real-time event detection by social sensors, *Proceedings of the 19th International Conference on World Wide Web* (New York, NY, USA), WWW '10, ACM, 2010, 851–860.

[21] J. L. Toole, M. Ulm, M. C González and D. Bauer, Inferring land use from mobile phone activity, in *Proceedings of the ACM SIGKDD International Workshop on Urban Computing,* ACM, 2012, 1–8.

[22] A. Wesolowski, et al., Quantifying the impact of human mobility on malaria, *Science,* **338** (2012), 267–270.

[23] http://www.gsc.upm.es/materiales/videos/

Received July 2014; revised December 2014.

E-mail address: alfredo.moralesg@alumnos.upm.es
E-mail address: werner.creixell@usm.cl
E-mail address: fj.borondo@upm.es
E-mail address: juancarlos.losada@upm.es
E-mail address: rosamaria.benito@upm.es