Understanding ethnical interactions on Ivory Coast

A. J. Morales [2], W. Creixell [1,2,3], J. Borondo [2], J. C. Losada [2], and R. M. Benito [2]
[1] *Departamento de Electrónica. Universidad Técnica Federico Santa María. Chile.*
[2] *Grupo de Sistemas Complejos, Universidad Politécnica de Madrid, ETSI Agrónomos, 28040, Madrid, Spain and*
[3] *Visiting researcher at Center for Spatial Information Science (CSIS), University of Tokyo, Japan.*

Towards the consolidation of peace and national development, the Ivorian society must overcome the lack of cohesion reported by several cooperation agents. In this sense, the present work provides insights on the regional interactions, so efforts may be planned and deployed with more intelligence. We characterize the communication patterns from a social perspective, in order to understand the factors that influence its emergence. We found that in a subregional scale, the ethnical identity plays an important role in the communication patterns, while at the interregional scale, other factors arise like economical interests and available infrastructure.

I. INTRODUCTION

The EU cooperation program for 2008-2013 on Ivory Coast [1], establishes the promotion of social cohesion as indispensable for the strengthening of peace, governance and national stability. In order to promote this cohesion, several improvements on the social services and infrastructures were deployed during this period, in which violence erupted once again among the social groups of the country. Therefore, in order for future actions to be planned and deployed with more intelligence, we aim to provide useful information on the way that the diverse society of Ivory Coast is structured, according to their interactions.

In particular, we intend to characterize and quantify interactions among the geographical and ethnical regions on Ivory Coast to understand the factors that influence the emergent social structure. We accomplish our scope by means of the tools provided by the complex networks theory [2]. For this matter, we used data provided by the *D4D Challenge on Mobile Phone Data*, as well as metadata like ethno-linguistic families, population or infrastructures, to get a better characterization of the analyzed regions.

In summary, on a local and regional scale, the Ivorian communication patterns seem highly influenced by social facts like the ethno-linguistic identity of the habitants, who tend to relate mostly to their own locality and others with similar cultural features. However, between the regions on a wider scale, the underlying infrastructure and the economical interests, seem to play a major influence in the social interacting structure, that end up rupturing the country into two large regions, located at the east and west side of the map.

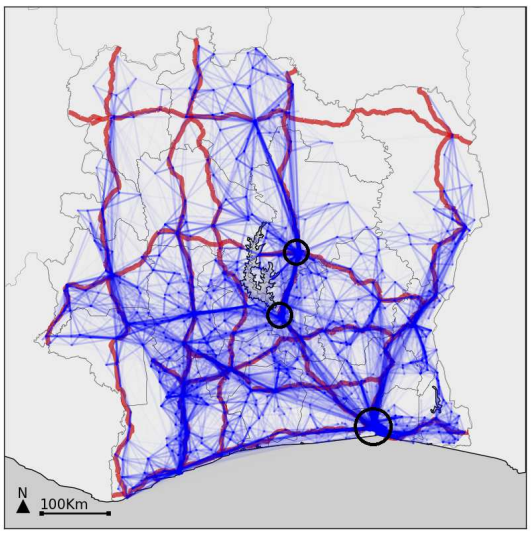

Figure 1. Trajectories network. Snapshot of the trajectories network at the end of the observation period. The blue lines represent the edges of the network and their intensity is proportional to the edge weight. The red lines represent the main roads of the country. Black circles indicate the location of the major cities.

II. DATA SET

The present work is based on the data provided by the *D4D Challenge on Mobile Phone Data*. These datasets were preprocessed in order to construct the complex networks necessary to represent the recorded interactions. We specifically used the antenna-to-antenna traffic dataset (*SET1*), as well as the individual trajectories for 50,000 customers dataset (*SET2*). At both cases, we aggregated the data available from all the observation period. Besides we used the geographical location of the antennas, to find spatial correlation on the Ivory Coast map.

Moreover, in the spirit of the project and to get a better understanding of the analyzed regions, we also used other sources of data available on Internet, like the language map from Lewis, M. Paul [3], and other spatial information files like the main roads location, taken from the African Development Bank [4].

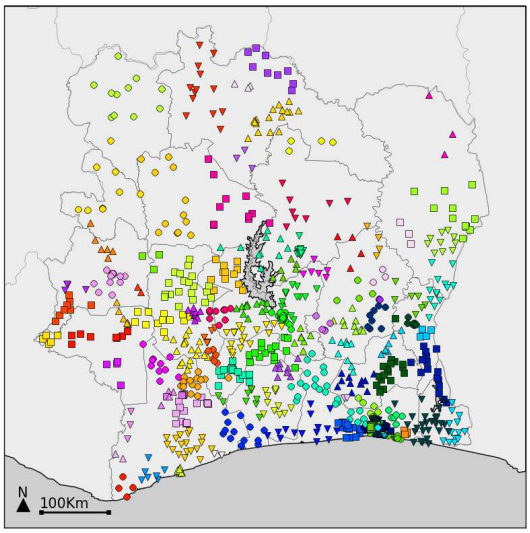

Figure 2. Trajectories network communities. The antennas have been represented by color and shape according to the community detection algorithm.

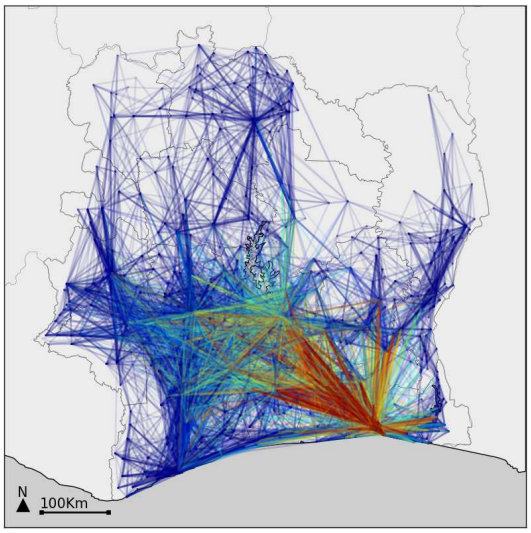

Figure 3. Trajectories network centrality. The edges have been colored according to the closeness centrality mean value of the connected nodes. The red regions indicate higher closeness-centrality, the yellow and pale blue regions indicate medium centrality, and the dark blue regions indicate lower closeness-centrality.

III. ETHINCAL INTERACTIONS

To begin to understand the Ivorian Society geographical structure and interactions, we first analyzed the mobility patterns of the country. For this matter, we built the Trajectories Network from the aggregation of all the individual trajectories found in the dataset *SET2*. An individual trajectory is defined as the sequential set of antennas that served a given user through time, and provides information on the mobility path of the user. In this network, antennas represent nodes, and an edge is created between two antennas, i and j, if a user makes two consecutive calls: first from the antenna i and after from the antenna j. The edges are directed, from i to j, and weighted, according to the number of times that all users performed the trajectory from i to j. The resulting network has 1215 nodes and 187102 edges. A visualization about the dynamical growth of this graph during an arbitrary day is presented on an animation in the supplementary video *VS1* (http://www.gsc.upm.es/materiales/videos/).

A snapshot of the trajectories network is presented in Fig. 1, where the graph has been plotted it in the map of Ivory Coast, using the location coordinates of the antennas. The edges are colored in blue, and the intensity of the edge is proportional to its weight. This network unveils the collective mobility pattern nationwide. The main city (largest black circle), is easily located in the southeast coast where a large amount of edges are concentrated, as well as other major cities in the center of the country (smaller black circles). In the figure we have also superimposed in color red the main roads of the country, to show the impact that the infrastructure has on the mobility patterns, since a large part of the trajectories keep correspondence to them. This network provides information about the flux of people along the roads. Some roads seem to be more frequently used, like the ones linking the north with the south of the country, in contrast to others less frequently used, like the transversal road up in the north. In addition, there are other edges, between the roads, that display the populated areas of the country, and possible roads of lower order. These populated areas are identified after using the modularity optimization method [5]. We found 100 communities, as can be seen in Figure 2, where each antenna has been plotted according to the community it belongs to. It may be noticed that communities comprehend a limited territorial area, which supports that communities emerged from the people displacements around villages and urban areas.

Therefore the patterns found in the trajectories network are a reflection of the country infrastructure and demography, but also of other social factors, like the economical activity, due to underlying reasons that justify the efforts for such displacements. To infer the significance of country regions in the economical activities we studied the closeness-centrality of the antennas in the

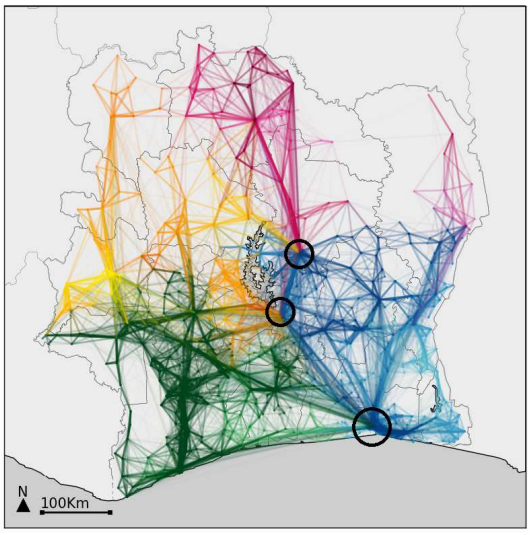

Figure 4. Trajectories network. The edges have been colored according to the linguistic group to which the most connected antenna at each community belongs to. There are four major linguistic families represented in yellow (northwest), purple (norteast), green (southwest) and blue (southeast). Black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

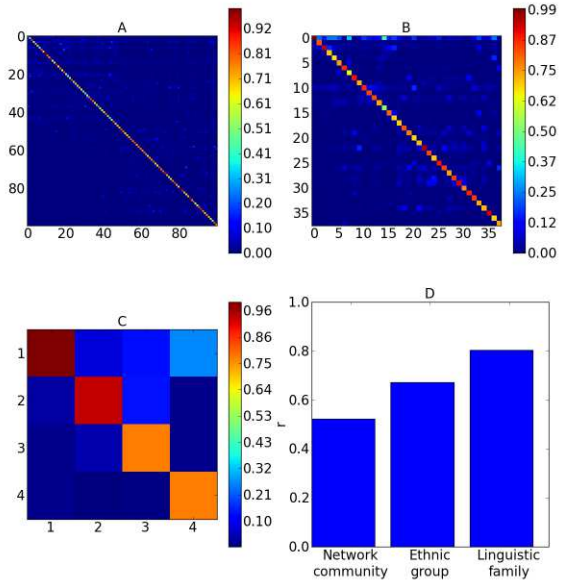

Figure 5. Row normalized adjacency matrices by (A) Network community, (B) Ethnic group and (C) Linguistic family. (D) Assortativity by characteristics coefficient on local scale (community), subregional scale (ethnic group) and regional scale (linguistic family).

trajectories network. This network property is inversely proportional to the average distance, in terms of network connections, that a node presents respect to the rest of the network, providing insights of how central or peripheral a node may be. In Figure 3, we present the trajectories network, coloring the edges according to the mean value of closeness-centrality of the antennas it connect. The red links connect antennas with a high closeness-centrality value, the yellow and pale blue links connect antennas with medium centrality value, and the dark blue links connect peripheral antennas with low closeness-centrality. It can be noticed, that the most central area of the graph (red) corresponds to the main city and the regions it adjoins, and as we get further from it, antennas get more peripheral. However, a medium centrality is noticed around the other major cities in the center of the country. This is in correspondence to the EU cooperation program [1], which identifies the north and the west of the country as the less developed areas.

To further understand the social composition of this graph, we have also taken into account the ethnical and linguistic component of each community. As communities are identified with given localities on the country, we used the ethnical map from [3], to identify the ethnical group that compounds them. To do this, we first identified the antennas with the highest degree at each community and them mapped them to the geographically closest ethnical group, using the location coordinates. The resulting ethnical assignment of the rest of antennas in the community, corresponded quite well with the four large linguistic families to which the ethnical groups belong (see Appendix A). This is shown in Fig. 4, where the trajectories network is presented by coloring the edges according to the linguistic family the network community belongs to. It can be seen that as 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 other regions mainly present trajectories from their own linguistic family. Besides it shows that the trajectories in the northern families, occur more frequently with the southern families, than between them.

Although, the trajectories network provides some information to characterize the social structure on Ivory Coast, it does not provide a clear vision of the interactions taking place among the groups found. To further understand it, we have constructed a second network, taking also into account the mobile calls information provided in the dataset *SET1*. In this network, the nodes are the 100 communities found in the trajectories network from the dataset *SET2*, and the edges correspond to the number of calls made from one community to the other, extracted from the dataset *SET1*. The edge direction goes from the emitter community to the receiver community and the weight is equal to the number of oc-

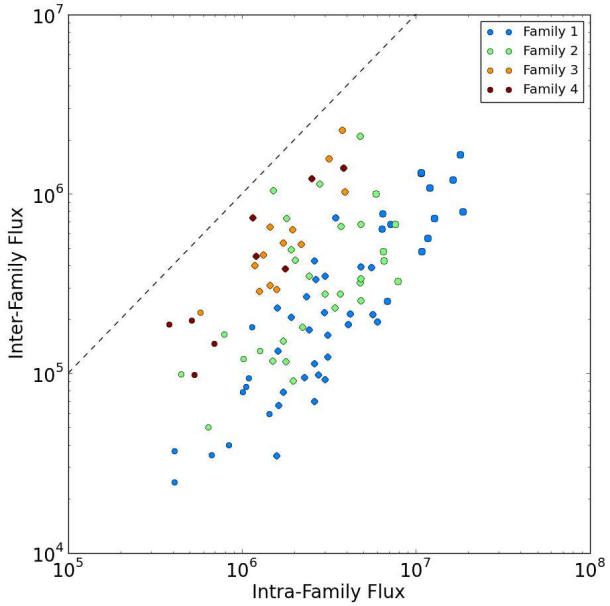

Figure 6. Intra linguistic family flux (calls directed to an antenna in the same linguistic family as the emitter antenna) versus Inter linguistic family flux (calls directed to an antenna in a different linguistic family than the emitter antenna).

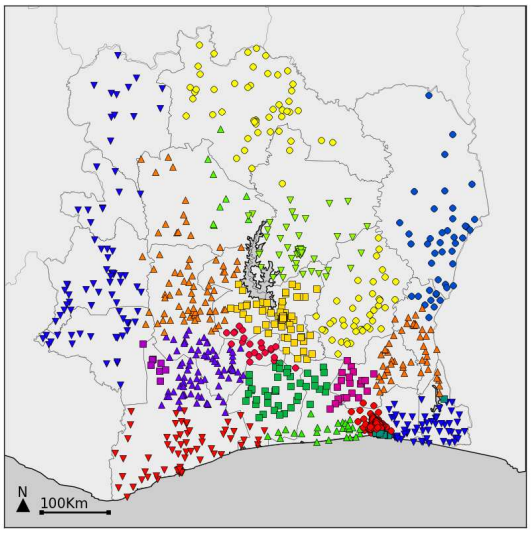

Figure 7. Calls network. The antennas have been represented by color and shape according to the community detection algorithm.

currences found in the data sets. This network represents a second level of interaction among the antennas.

To get a clearer view of the way that these trajectory communities called each other, in Fig. 5A we present the weighted adjacency matrix of the constructed network, normalized by row for a better visual comprehension. This normalization avoids the masking effect that the larger groups have over the small ones, and provides relative information about the destination of the outgoing calls and origin of the incoming ones. It can be noticed, that the diagonal is quite strong, indicating that most of outgoing calls of a community remains in the same community. This effect is also noticeable, when we group these communities using the ethnical metadata, like the ethnical group in Fig. 5B, and the linguistic family in Fig. 5C.

In fact, the preference of people to communicate to similar ones gets stronger as we increase the scale of the network, in terms of the same community, ethnical group or linguistic families. This is shown in Fig. 5D, where we plot the resulting assortative coefficient by characteristic [6] of each matrix. It can be noticed that it increases from 0.5 up to 0.8, when evaluating from network community to language family, which means that at the lower level, the people of a community call more often people in other communities, yet these other mostly belong to the same family language.

Moreover, not all families behave the same way. The south linguistic families (number 1 and 2 in Fig. 5C) present a larger proportion of calls directed to the same linguistic family, in comparison to the north linguistic families (number 3 and 4 in Fig. 5C), since the diagonal values of the north families are redder in the figure than the south families, which are more yellowy. This difference between the northern and southern regions are noticed in Fig. 6, where we represent the communities studied according to the intra-family traffic (calls directed to the same linguistic family) versus the inter-family traffic (calls directed to a different linguistic family). In the figure the symbols represent the communities found in the trajectory 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 is the intra-family traffic in comparison to the inter-family traffic. We see that the northern families are more keen to call different families, than the southern ones.

However, the communication between the northern families to the southern families are found to be quite selective. In Fig. 5C, we see that the family 1 (southeast) has a stronger connection with family 4 (northeast), than with family 3 (northwest), which resulted to have a stronger connection to the family 4 (southwest). This means, that each northern family tends to communicate more with their adjoin southern family, resulting in a larger density of calls from the northeast to the southeast regions, as well as from the northwest to the southwest regions. This observation is in good agreement with

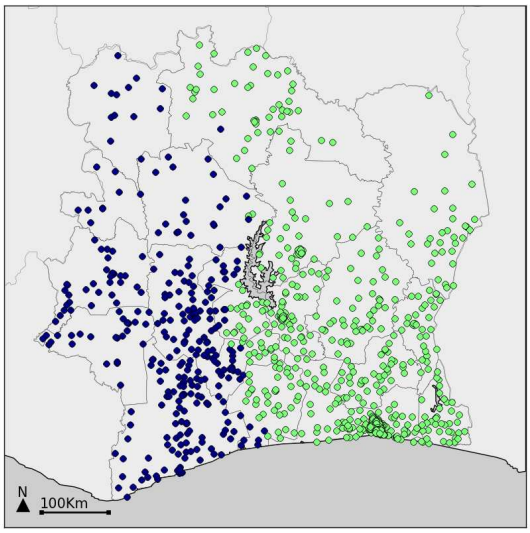

Figure 8. Antennas classification results by the way the calls network communities are related, using k-means clustering classifier.

the mobility patterns shown in Fig. 1, where the vertical roads seam to have a higher significance than the horizontal ones, and in Fig. 4, where we showed that the mobility of the northern families to the south are stronger with the adjoin regions.

To further understand this rupture on the communication patterns between the east and west side of the country, we built a third network, to analyze the calling behavior at the microscale, extracting only information from the dataset *SET1*. The nodes in this network also represent 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. It is a directed and weighted network, where the weight of the edges represents the number of calls made from i to j. This calls network, presented 19 communities, according to the modularity optimization algorithm [5], distributed along the geography of Ivory Coast as may be seen in Fig. 7. In the supplementary video *VS2* (http://www.gsc.upm.es/materiales/videos/), we present an animation with the growth of this network and a visualization of the influence that each of the 19 communities have in the network.

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, 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. A clear division between the east side and west side of the country is appreciated, which corroborates the influence that the underlying infrastructure and other human factors, like cultural bonds or economical interest, may have in the structure of the social interactions, within a country.

IV. CONCLUSIONS

By means of the analysis of the resulting patterns of the trajectories and calling network, we have characterized the interactions and resulting structure of the diverse geographical and social areas of Ivory Coast.

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 of the same linguistic family. Yet this preferences is not equal to all linguistic families, since the peripheral regions of the north, seem to communicate with their adjoin southern regions significantly, which ruptures the map into two interacting regions located at the east and west side of the country. This division of the country patterns, seems to be highly influenced by the underlying infrastructure and economical factors.

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 Grand No. MTM2012-39101.

Special thanks to the "Beca Iberoamérica Santander Universidades 2012" for its travel financing.

Appendix A: Ethno-linguistic groups in Ivory Coast

Ivory Coast presents a complex society compound by more than 60 different ethnic groups. These ethnic groups are classified into four large linguistic families, as can be seen in Fig. 9, where we present the ethnic map realized by Lewis, M. Paul [3]. Although each ethnic group has its own language, French is the official language and it is broadly spoken along the country.

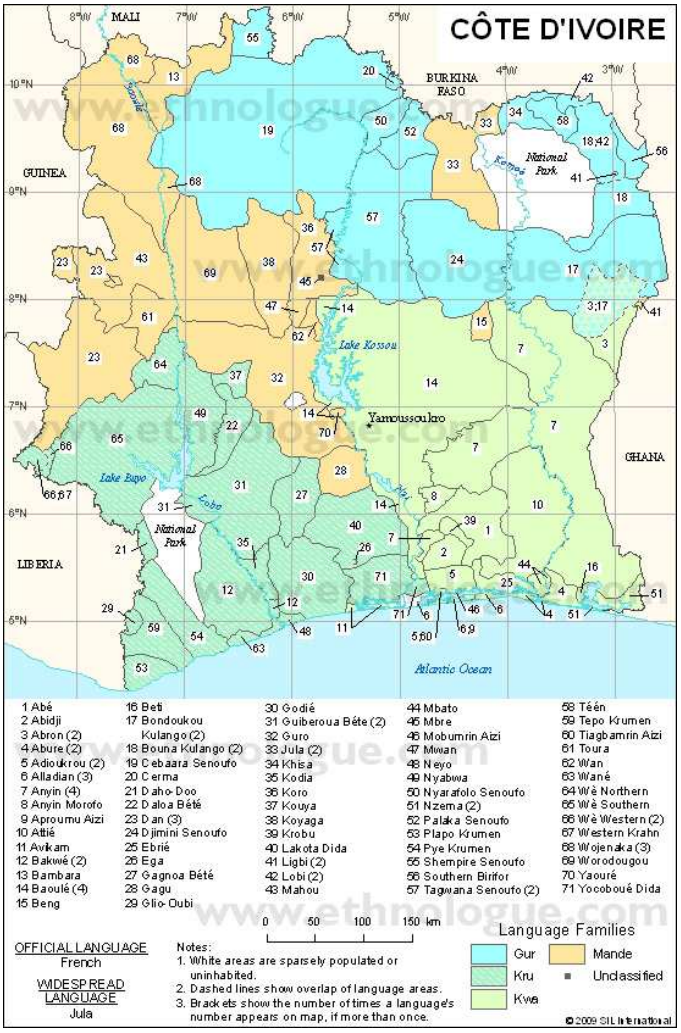

Figure 9. Ethno-linguistic groups in Ivory Coast. Image realized by Lewis, M. Paul [3]

In summary, the Kwa group is located in the southeast side of the country, where the capital city and other major cities are situated, as well as the main Ivorian Airport and Port. The Kru group is located in the southwest side, also in the Atlantic coast. The Mande group is found in the northeast side of the country, and the northwest region is occupied by the Gur family. These last are the least populated regions of the country.

[1] European Commission (2008) Communaute europeenne - Republique de Cote d'Ivore. Document de strategie pays et programe indicatif national pour la periode 2008-2013 http://ec.europa.eu/development/icenter/repository/scanned_ci_csp10_fr.pdf

[2] S. Bocaletti, V. Latora, Y. Moreno, M. Chavez and D.-U. Hwang Complex networks: Structure and dynamics *Phys. Rep.* 424, 175 (2006)

[3] Lewis, M. Paul (ed.), 2009. Ethnologue: Languages of the World, Sixteenth edition. Dallas, Tex.: SIL International. Online version: http://www.ethnologue.com/.

[4] African Development Bank Cte d'Ivoire Roads http://www.infrastructureafrica.org/library/doc/986/cote-divoire-roads

[5] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, (2008) Fast unfolding of communities in large networks, *J.STAT.MECH.*

[6] Newman M.E.J., 2003. Mixing patterns in networks. *Physical Review E* 67, 026126.