Uncovering Digital Divide and Physical Divide in Senegal Using Mobile Phone Data

Song Gao, Bo Yan, Li Gong, Blake Regalia, Yiting Ju, Yingjie Hu

STKO Lab, Department of Geography, University of California, Santa Barbara, CA 93106

1. Introduction

The mobile phone call detailed records (CDRs) distributed within the framework of "Data for Development" Senegal challenge were run through several processes that intended to anonymize all source users' information while still providing sufficient and meaningful data to researchers (de Montjoye et al. 2014). For instance, the hourly site-to-site traffic data for cellphone sites is beneficial for analyzing the dynamic digital communication patterns at the spatial resolutions of a cellphone tower's coverage or at other aggregated regional scales. In addition, the individual-based records provide opportunities to study the human mobility patterns at both the individual level and the geographically aggregated level, which has been a hot topic in the existing literature (Gonzalez et al., Song et al. 2010, Kang et al. 2012, de Montjoye et al. 2013). For this research, we first aim at developing data analytics that can derive insights on how people from different regions communicate and connect via phone calls and physical movements. While there are many studies applying the community detection techniques based on Graph theory to identify the spatial connectivity and characteristics of regions, social segregation, or functional zones of a city using mobile phone data (Ratti et al. 2010, Gao et al. 2013, Amini et al. 2014, Chi et al. 2014), few research have addressed the spatiotemporal resolution issue (Cheng and Adepeju 2014). The chosen spatial unit (e.g., cell-based, region-based) or temporal scale (e,g., by hour, day, week, month) might affect the results of analyzing human mobility and urban dynamics in the mobile age (Gao 2014). To this end, we will discuss the impact of changing the spatial analysis unit and the temporal resolution when detecting community patterns of spatial interactions in both cyberspace and physical space extracted from one-year CDRs in Senegal.

2. Methods

Two types of weighted graphs can be built based on the given CDRs. Let *G_CallFlow (V, E)* denote a weighted undirected graph of phone call flows among different spatial units (*S*) where cellular sites or administrate places (e.g., regions, departments, arrondissement) are transformed into graph nodes (*V*) while communication flows among places are represented as weighted edges (*E*). Let W_{ijt} represent the total phones calls between a spatial unit *i* and another spatial unit *j* during the time interval *t* (by hour, day, or month). As an example of one selected node accompanied by its links in the graph, fig. 1 shows the monthly phone call flows that connect the capital city Dakar to other arrondissements in Senegal.

Similarly, let *G_MobilityFlow (V, E)* be a weighted undirected network graph of human movement flows in physical space and let M_{ijt} represent the total volume of movement flow between a spatial unit *i* and another spatial unit *j* during time interval *t*, including the movement flows both from *i* to *j* and from *j* to *i*. Note that we can also build the weighted directed graph of spatial interactions by adding the direction of flows but it is not required for community detection operations.

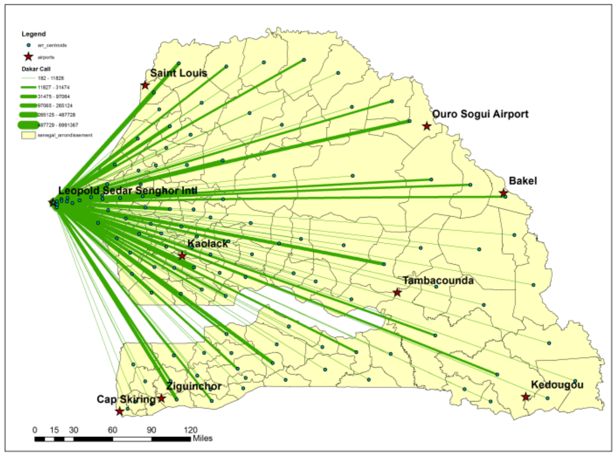


Figure 1. Visualizing the phone call interactions between the capital Dakar and other arrondissements in Senegal.

In the study of complex networks, a community is defined as a subset of the whole graph where nodes within the same community are densely connected and grouped together. The identification of such divisions in a graph is called community detection. Newman and Girvan (2004) proposed a modularity metric to evaluate the quality of a particular division within a graph into communities. Modularity compares a proposed partition to a null model in which connections between nodes are random. The larger the modularity value is, the more robust (stable) the detected community structure is. We apply two popular techniques for community detection in our work: (1) fast-greedy modularity maximization algorithm (FG) (Clauset et al. 2004); and (2) multi-level algorithm (ML) (Blondel et al. 2008).

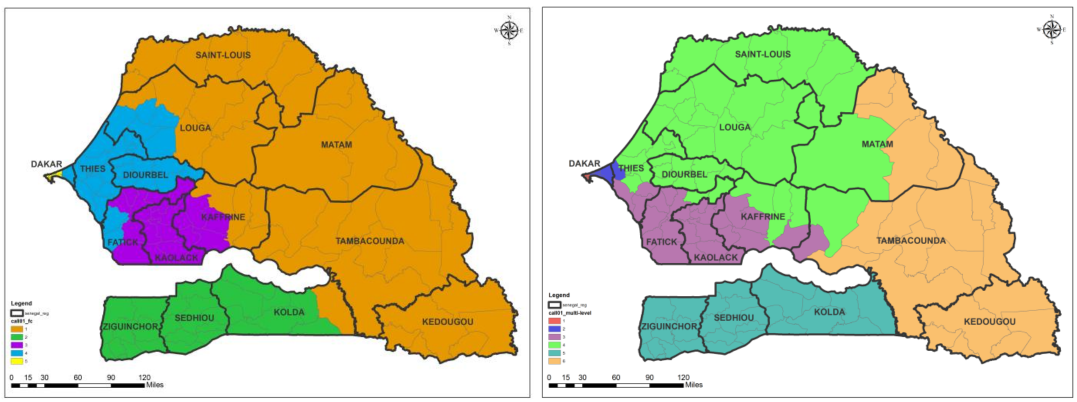
For each type of the weighted graphs (*G_CallFlow* or *G_MobilityFlow*), we will process the data in different spatial and temporal resolutions then identify the communities for each graph by maximizing the modularity value. In order to compare the similarity of different scenarios of the community detection results, we calculate the normalized mutual information index (NMI) proposed by Danon et al. (2005) to measure the similarity between different partitions. The range of NMI value is between 0 and 1. The higher the value is, the more similar the graph partitions are.

3. Results

3.1 Digital Divide

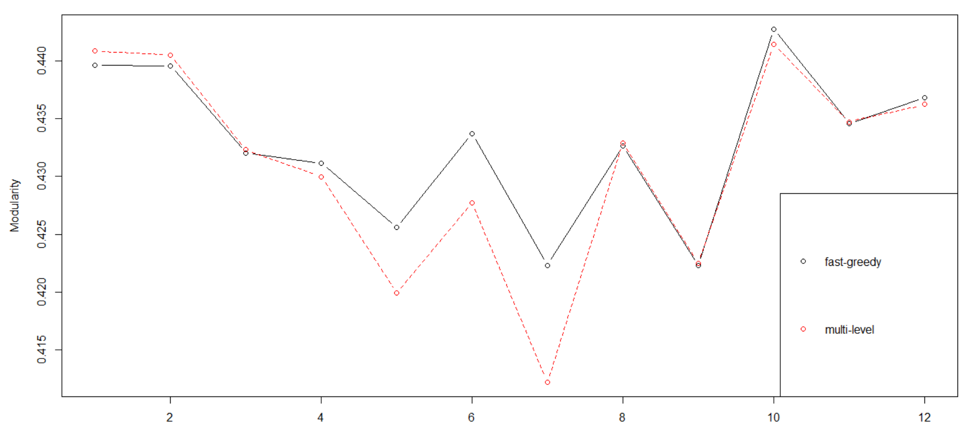
Fig. 2 shows the spatial distributions of community detection results for the phone call flow graph *G_CallFlow* in January using FG and ML algorithms. We can identify the digital divide in Senegal, i.e., the connected arrondissements within the same community have more intensive call communications than those of inter-communities, and they tend to spatially cluster. For instance, it is clear that the DAKAR region itself has more intra communications while arrondissements of TAMBACOUNDA, MATAM, SAINT-LOUIS and KEDOUGOU tend to group together. The modularity values of two detection algorithms are similar 0.4396 (FG) and 0.4408 (ML), while the partition structures also have a high similarity value (NMI=0.84). Fig. 3 demonstrates the temporal changes of modularity values and the similarity of community detection results of phone call flows among arrondissements from different months.

Considering the temporal resolution effect (fig. 4), we also apply these two detection algorithms to the hourly and daily aggregated phone call flow graphs and compare the modularity values as well as the partition structures.

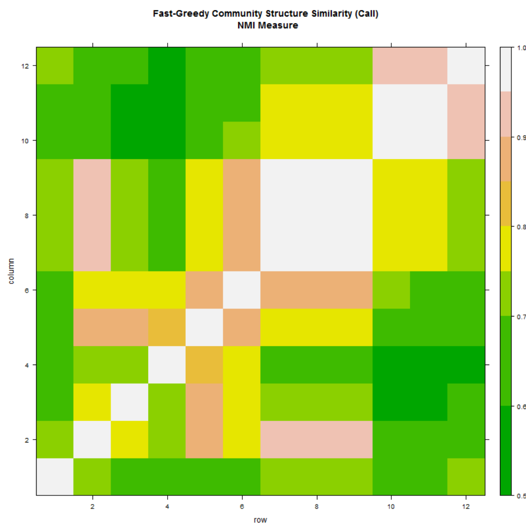


(a) (b)

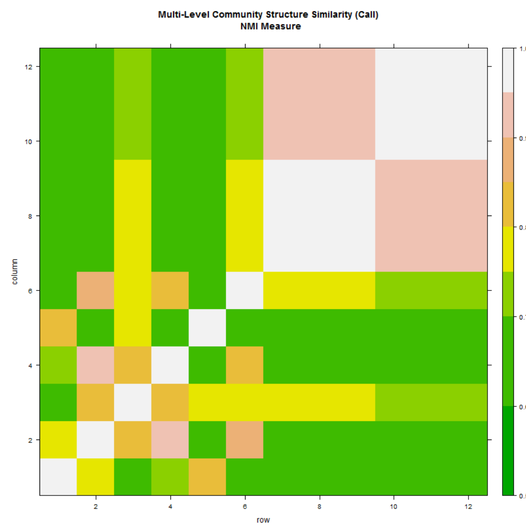
Figure 2. Community detection of phone call flows at the arrondissement level in January using two algorithms: (a) FG ; (b) ML.



(a)

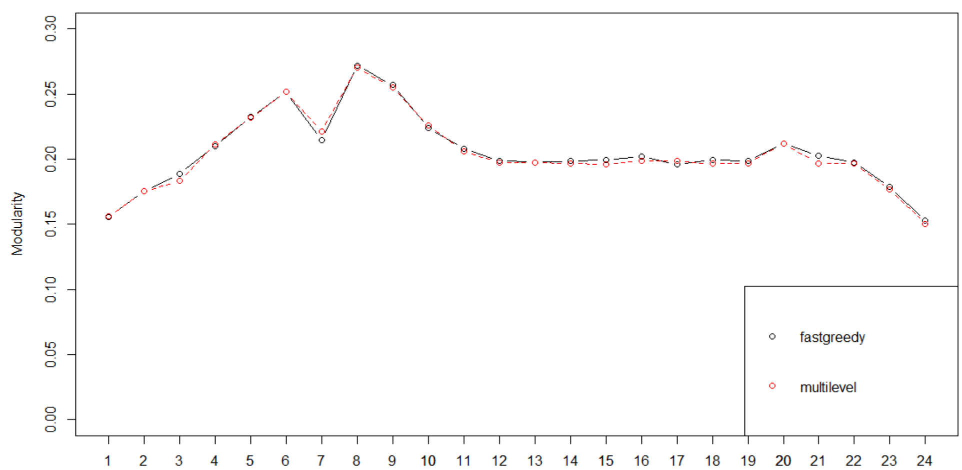


(b)

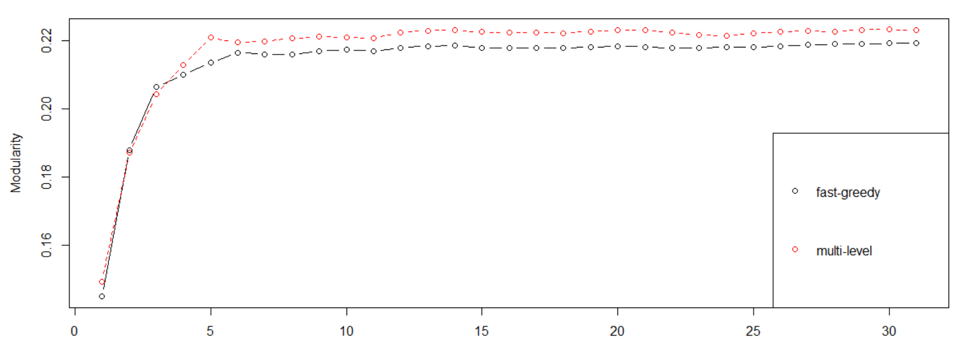


(c)

Figure 3. (a) The modularity values for the community detection results of *G_CallFlow* in different months; (b) similarity matrix of FG partition results; (c) similarity matrix of ML partition results.



(a)

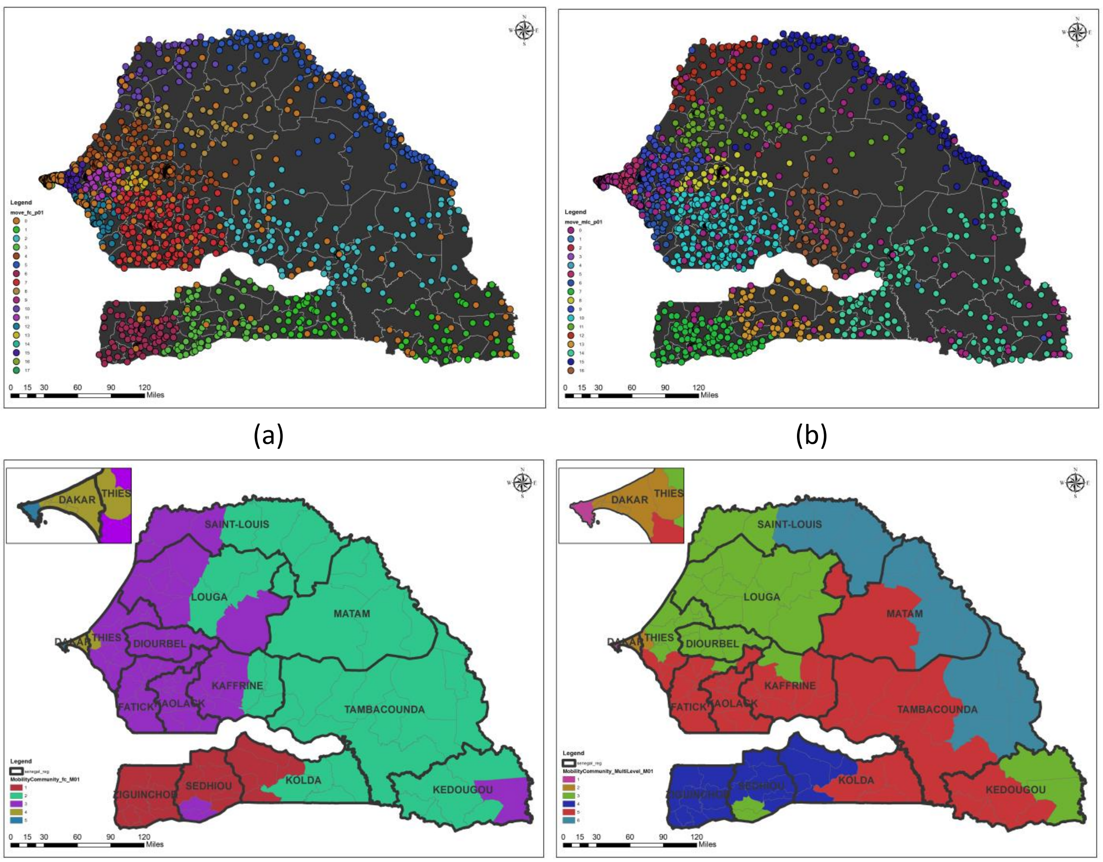


(b)

Figure 4. The temporal variability of modularity by (a) hour; (b) day.

3.2 Physical Divide

Fig. 5a and 5b show the spatial segregation of monthly human mobility flow graph *G_MobilityFlow* at the cellular site scale by applying two community detection algorithms. Not surprisingly, the spatially adjacent cellular sites are more likely to be grouped together although there are several abnormal grouping patterns. For example, the northeastern blue community along the Country boundary tends to have more cross-site mobility flows. In addition, we found that the modularity values (M_{FG} = 0.7260, M_{ML}= 0.7248) based on site-to-site mobility graph are larger than that of the partition results of the arrondissement-to-arrondissement mobility graph (M'_{FG} = 0.4396, M'_{ML}= 0.4408) as shown in fig. 5c and 5d. From the geographical context perspective, such physical divide patterns might be associated with terrain barriers (fig. 5e), streets network centrality (fig. 5f) (Gao et al. 2013b) or other potential socioeconomic factors. The temporal changes and similarity of partition results of the monthly mobility flow graphs have also been studied in this work (fig. 5g and 5h).



(a) (b)

(c) (d)

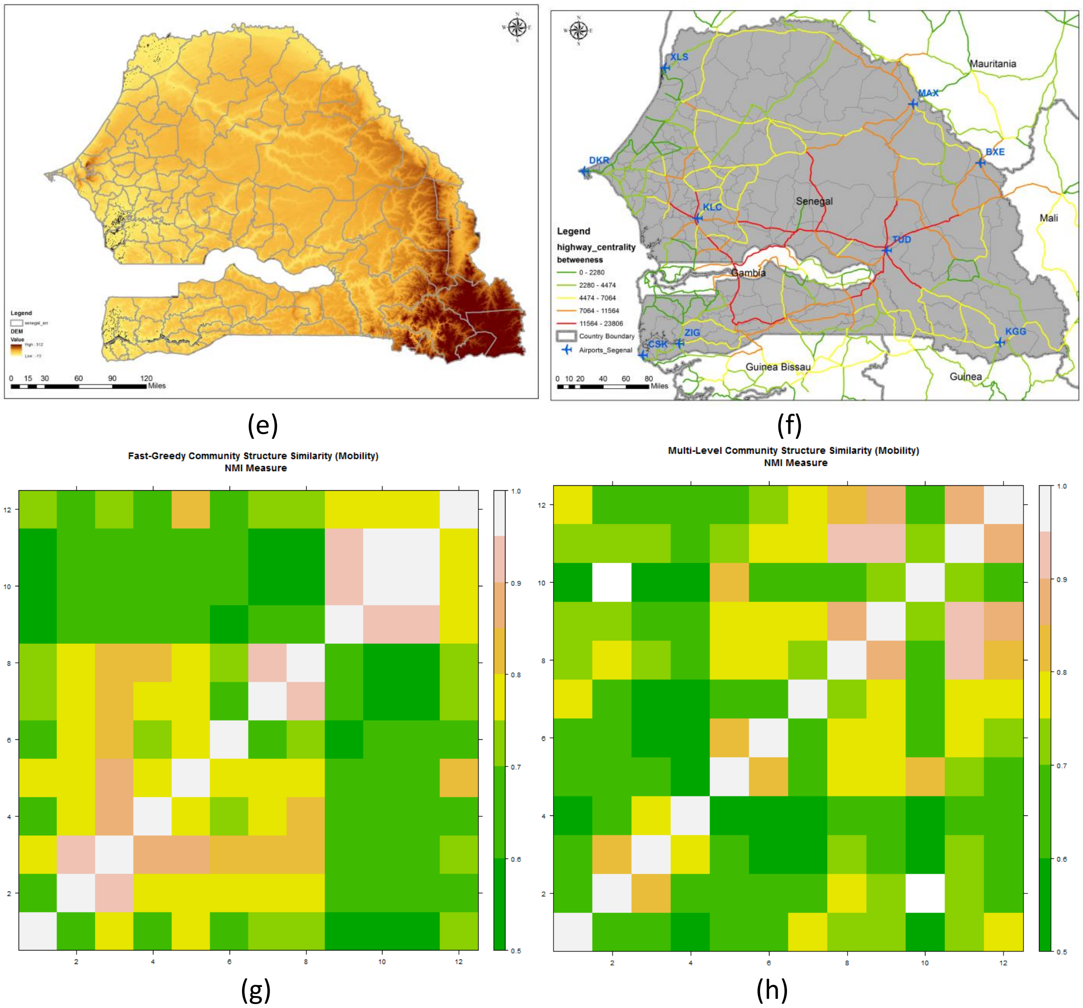


Figure 5. Community detection of monthly mobility flows at (a) site-to-site scale using FG ; (b) site-to-site scale using ML; (c) arrondissement-to-arrondissement scale using FG; (d) arrondissement-to-arrondissement using ML; (e) terrain elevation in Senegal; (f) highway streets network centrality; (g) similarity matrix of monthly FG partition results; (h) similarity matrix of monthly ML partition results.

4. Conclusions

In this work, we were trying to uncover the digital divide (phone communication patterns) and physical divide (human mobility patterns) in Senegal based on the large-scale mobile phone data. The research justifies that the chosen spatial unit and the temporal resolution can affect the community detection results for the two types of spatial interaction graphs. We also found that the daily-based detection has generated a more stable partition structure than the hourly-based partition, while there are also monthly changes across a year. The presented framework can help identify patterns of spatial interactions in both cyberspace and physical space with call detailed records in some regions where census data acquisition is difficult, especially in African countries.

5. References

Amini A, Kung K, Kang C, Sobolevsky S and Ratti C, 2014, The impact of social segregation on human mobility in developing and industrialized regions. *EPJ Data Science*, 3(1), 6.

Blondel VD, Guillaume JL, Lambiotte R and Lefebvre E, 2008, Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008.

Cheng T and Adepeju M, 2014, Detecting emerging space-time crime patterns by prospective STSS. *Proceedings of the 12th International Conference on GeoComputation*. Available: http://www. geocomputation. org/2013/papers/77. pdf.

Chi G, Thill JC, Tong D, Shi L, and Liu, Y, 2014, Uncovering regional characteristics from mobile phone data: A network science approach. *Papers in Regional Science*. DOI: 10.1111/pirs.12149

Clauset A, Newman ME and Moore C, 2004, Finding community structure in very large networks. *Physical Review E*, 70(6), 066111.

Danon L, Diaz-Guilera A, Duch J and Arenas A, 2005, Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(09), P09008.

de Montjoye YA, Hidalgo CA, Verleysen M and Blondel V D, 2013, Unique in the Crowd: The privacy bounds of human mobility. *Scientific reports*, 3.

de Montjoye YA, Smoreda Z, Trinquart R, Ziemlicki C and Blondel VD, 2014, D4D-Senegal: The Second Mobile Phone Data for Development Challenge. *arXiv preprint*:1407.4885.

Gao S, Liu Y, Wang Y and Ma X, 2013a, Discovering spatial interaction communities from mobile phone data. *Transactions in GIS*, 17(3), 463-481.

Gao S, Wang Y, Gao Y and Liu Y, 2013b, Understanding urban traffic-flow characteristics: a rethinking of betweenness centrality. *Environment and Planning B: Planning and Design*, 40(1), 135-153.

Gao S, 2014. Spatio-Temporal Analytics for Exploring Human Mobility Patterns and Urban Dynamics in the Mobile Age. *Spatial Cognition and Computation*, DOI:10.1080/13875868.2014.984300.

Gonzalez MC, Hidalgo CA and Barabasi AL, 2008, Understanding individual human mobility patterns. *Nature*, 453(7196), 779-782.

Kang C, Ma X, Tong D and Liu, Y. 2012, Intra-urban human mobility patterns: An urban morphology perspective. *Physica A: Statistical Mechanics and its Applications*, 391(4), 1702-1717.

Newman ME and Girvan M, 2004, Finding and evaluating community structure in networks. *Physical Review E*, 69(2), 026113.

Ratti C, Sobolevsky S, Calabrese F, Andris C, Reades J, Martino M, Claxton R and Strogatz SH, 2010, Redrawing the map of Great Britain from a network of human interactions. *PloS one*, 5(12), e14248.

Song C, Qu Z, Blumm N and Barabási AL, 2010, Limits of predictability in human mobility. *Science*, 327(5968), 1018-1021.