Uncovering the Digital Divide and the 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

Abstract

In this research, we first aim at developing data analytics that can derive insights about how people from different regions communicate and connect via mobile phone calls and physical movements. We uncover the digital divide (geographical segregation of phone communication patterns) and the physical divide (geographical limits of human mobility) in Senegal. The research also demonstrates that the chosen spatial unit and temporal resolution can affect the community detection results of spatial interaction graphs when analyzing human mobility patterns and exploring urban dynamics in the mobile age. We find that the daily detection has generated a more stable partition structure than an hourly one, while monthly changes also exist over time. The presented framework can help identify patterns of spatial interaction in both cyberspace and physical space with phone call detailed records in some regions where census data acquisition is difficult, especially in African countries.

Keywords: digital divide, physical divide, community detection, mobile phone data, spatial interaction

1. Introduction

The mobile phone call detail records (CDRs) distributed within the framework of “Data for Development” Senegal challenge were running 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 example, the hourly site-to-site traffic data for cellphone sites is beneficial for analyzing dynamic digital communication patterns at the spatial resolution of a cellphone tower’s coverage or at other aggregated regional scales. In addition, individual-based records provide opportunities to study 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 about how people from different regions communicate and connect via phone calls and physical movements. While many studies exist applying 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. 2013a, Amini et al. 2014, Chi et al. 2014), few researchers have addressed the spatiotemporal resolution issue (Cheng and Adepeju 2014). The chosen spatial unit (e.g., cell-based, region-based) or temporal resolution (e,g., hour, day, week, month) might affect the results of analyzing human mobility and urban dynamics in the mobile age (Gao 2015).

To this end, we discuss the impact of changing the spatial analysis unit and temporal resolution when detecting community patterns of spatial interaction 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, arrondissements[1]) are transformed into graph nodes (*V*) while communication flows among places are represented as weighted edges (*E*). Let W_{ijt} represent the total phone 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 a 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 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 although we can build the weighted directed graph of spatial interaction by adding the direction of flows, it is not required for community detection operations.

In the study of complex networks, a community is defined as a subset of the entire 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) propose a modularity metric to evaluate the quality of a particular division into communities within a graph. 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) a fast-greedy modularity maximization algorithm (FG) (Clauset et al. 2004) that merges pairs of communities iteratively and always chooses the pair that yields the maximum increase in the overall modularity; and (2) a multi-level algorithm (ML) (Blondel et al. 2008) in which nodes are moved between communities such that each node makes a local choice that maximizes its own contribution to the modularity score, and can unfold a complete hierarchical community structure in multiple steps.

For each type of weighted graph (*G_CallFlow* or *G_MobilityFlow*), we process the data for different spatial and temporal resolutions, and 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 NMI value is in the range between 0 and 1. The higher the NMI value is, the more similar the graph partitions are.

[1] Arrondissement is usually a level of administrative division under Department in Francophone countries.

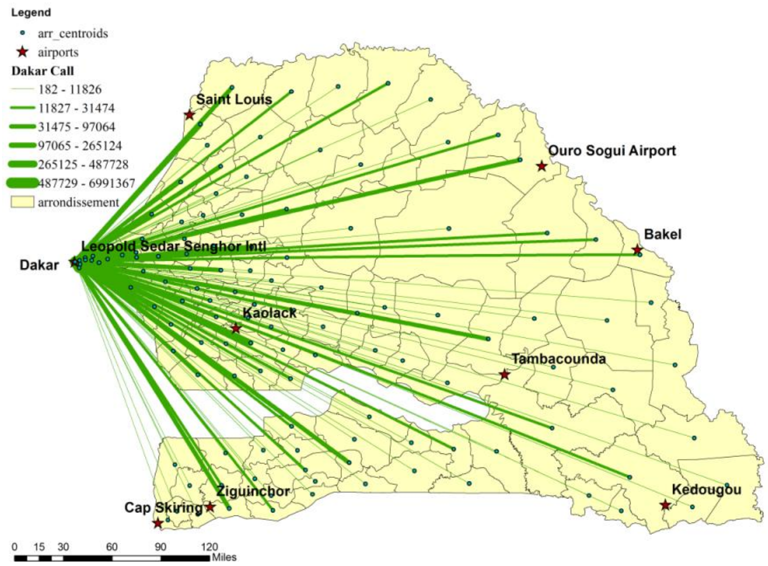


Figure 1. Visualizing the phone call interactions between the capital Dakar and other arrondissements in Senegal.

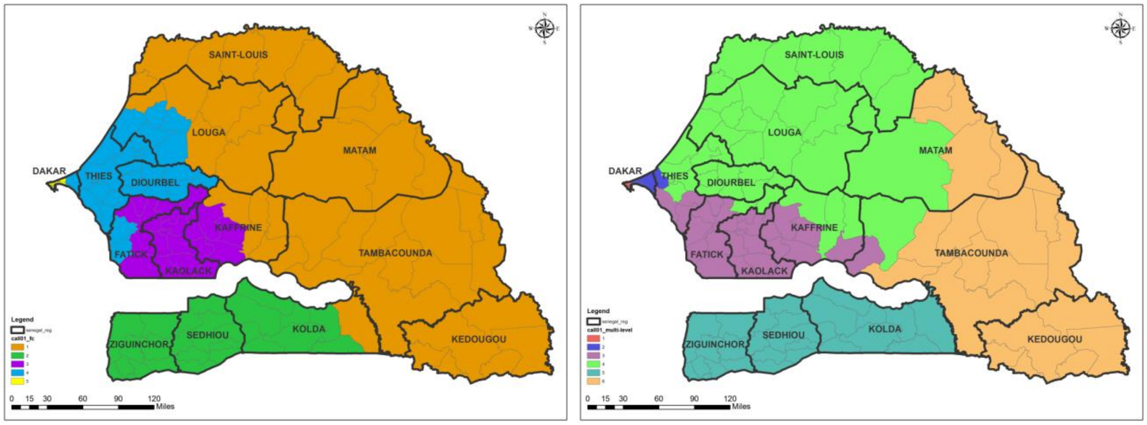
3. Results

In this section, we apply the aforementioned community detection algorithms to two types of spatial interaction graphs and discuss the results.

3.1 The Digital Divide

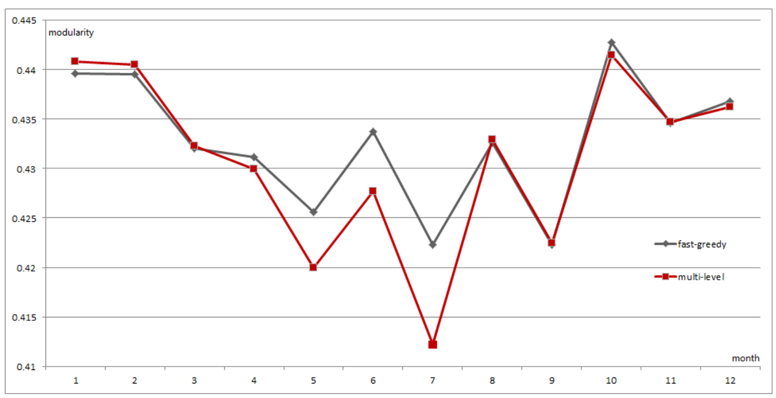
Fig. 2 shows the spatial distributions of community detection results for the phone call flow graph *G_CallFlow* in January using the FG and ML algorithms. We can identify the *digital divide* (geographical segregation of communication patterns) in Senegal; i.e., the geographically adjacent arrondissements within the same community have more intensive call communications than those of inter-communities, and they tend to spatially cluster. For example, the DAKAR region itself has more intra communications, whereas the arrondissements of TAMBACOUNDA, MATAM, SAINT-LOUIS and KEDOUGOU tend to group together. The modularity values of the two detection algorithms are similar 0.4396 (FG) and 0.4408 (ML), while the partition structures have a high similarity value (NMI=0.84). Fig. 3 depicts the temporal changes of modularity values and the structural similarity of community detection results of phone call flows among arrondissements in different months. The month-to-month similarity matrix shows that more similar community structures are detected from October to December (pink and white color grids at the top-right corner of Figs. 3b and 3c).

Considering the temporal resolution effect, 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. Fig. 4 demonstrates the hourly and daily changes of modularity values. The modularity value reaches a maximum during the hour 07~08, which represents the most stable community structure, whereas the value gets lower at night, which indicates a relatively unstable community structure (Fig. 4a). The daily detection has generated a more stable partition structure over time, although we can still identify the unstable community structure in the first days of January, which might result from irregular mobile phone call patterns on New Year's Day (Fig. 4b). No significant difference exists between the temporal changes of community structure using the FG and ML detection algorithms.

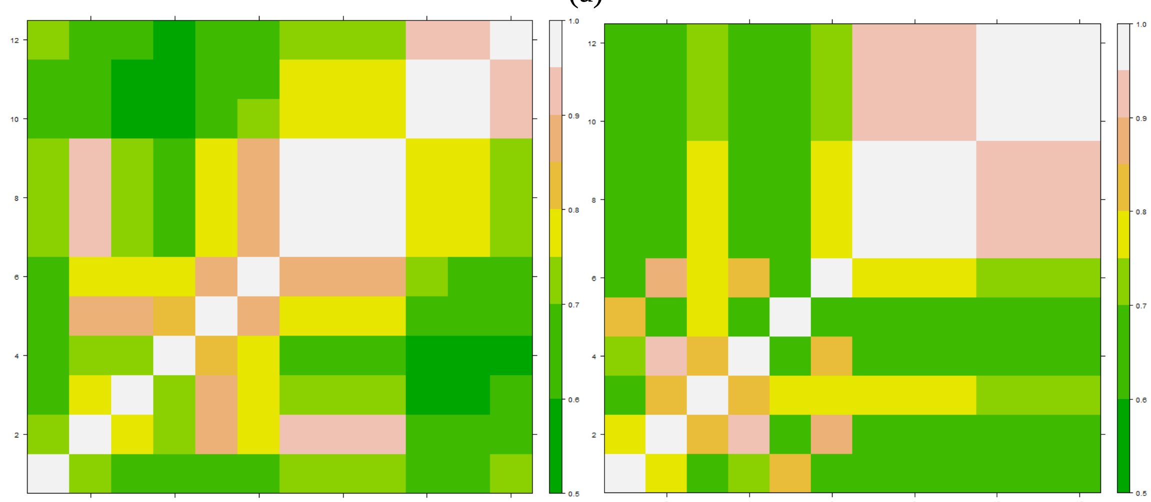


(a) (b)

Figure 2. Community detection of phone call flows at the arrondissement level in January using the 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) a month-to-month similarity matrix for the FG partition results; and, (c) a month-to-month similarity matrix for the ML partition results.

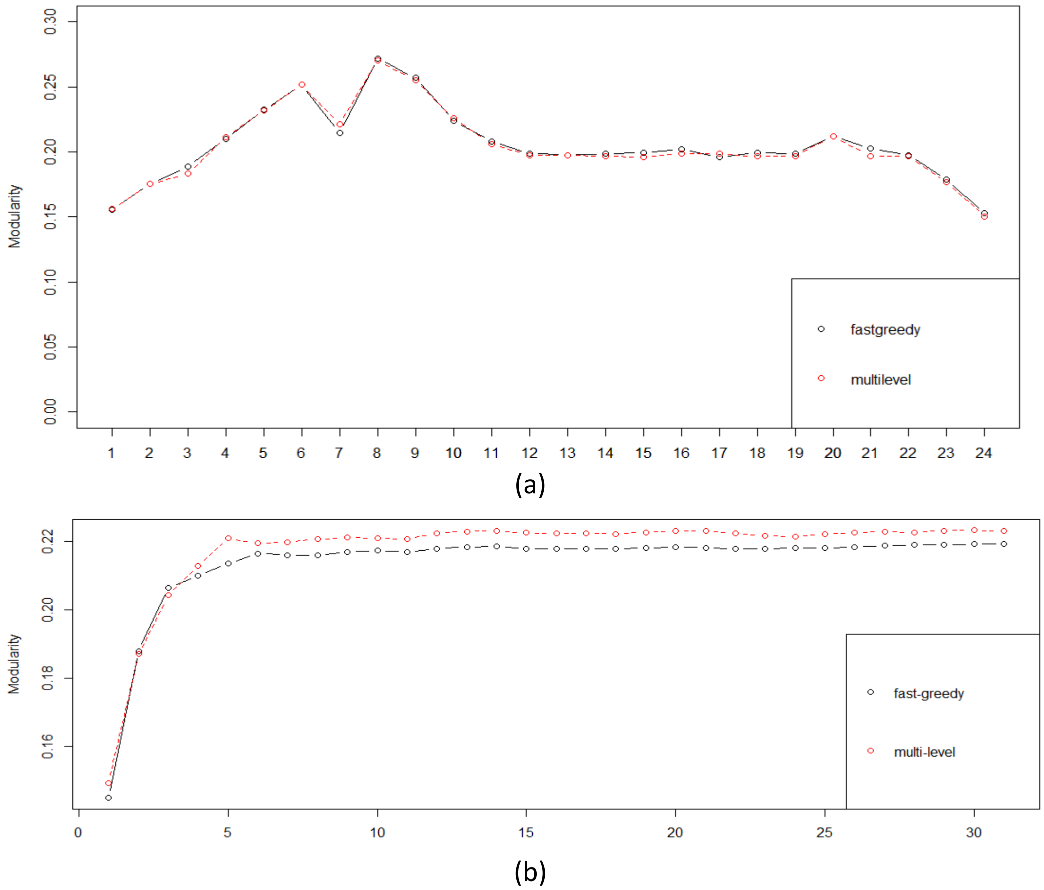


Figure 4. The temporal variability of modularity in January by (a) hour; (b) day.

3.2 The Physical Divide

Another type of spatial interaction network is generated by phone users' physical movement; more detailed discussion can be found in Gao et al. (2013a). Figs. 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 (FG and ML). The term "physical divide" in this context is used to represent such human mobility patterns within limited geographical space. Not surprisingly, the spatially adjacent cellular sites are more likely to be grouped together based on mobility flows, although several abnormal grouping patterns occur across space. For example, the northeastern blue community along the country boundary tends to have more cross-site mobility flows because of highway connections along the border. In addition, we found that the modularity values (M_{FG} = 0.7260, M_{ML}= 0.7248) based on a site-to-site mobility graph are larger than those for the partition results of an arrondissement-to-arrondissement mobility graph (M'_{FG} = 0.4396, M'_{ML}= 0.4408) as shown in Figs. 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 natural environment and socioeconomic factors. The temporal changes and structural similarity of graph partition results for monthly mobility graphs have also been studied in this work (see Fig. 6a and 6b). In general intra-season similarity tends to be higher than inter-season similarity.

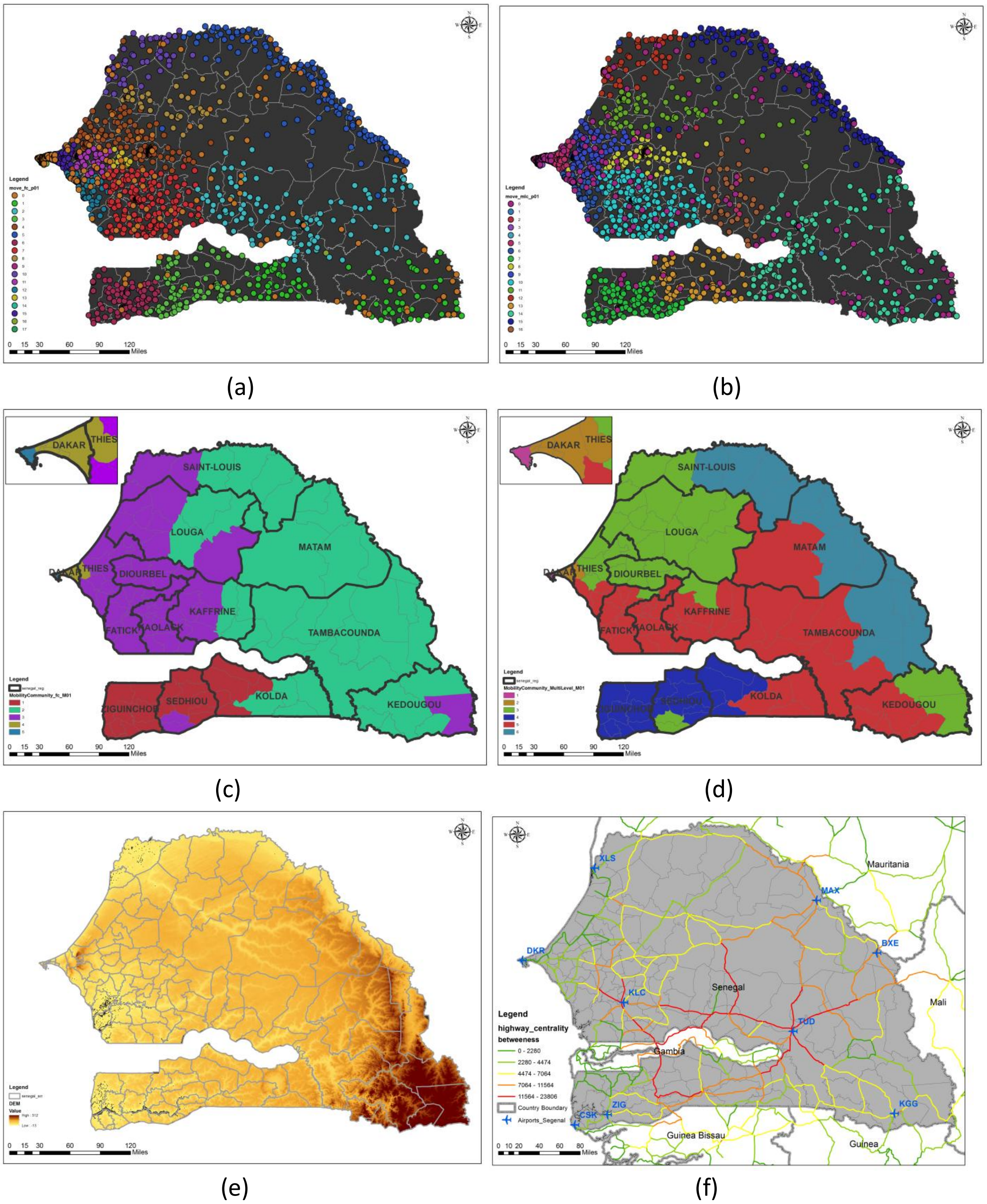
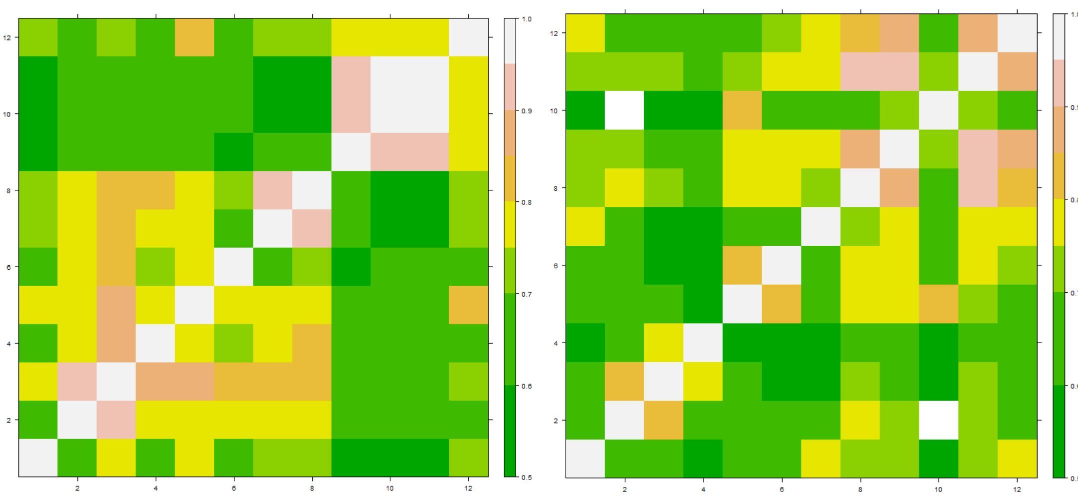


Figure 5. Community detection of monthly mobility flows (a) at the site-to-site scale using FG; (b) at the site-to-site scale using ML; (c) at the arrondissement-to-arrondissement scale using FG; (d) at the arrondissement-to-arrondissement using ML; (e) a terrain elevation map in Senegal; and, (f) a map of highway network centrality.



(a) (b)

Figure 6. Month-to-month comparisons on mobility flow graphs: (a) a similarity matrix for the FG partition results; and, (b) a similarity matrix for the ML partition results.

4. Conclusions

In this work, we seek to uncover the digital divide (geographical segregation of phone communication patterns) and the physical divide (geographical limits of human mobility) in Senegal based on large-scale mobile phone data. The research demonstrates that the chosen spatial unit and temporal resolution can affect the community detection results of two types of spatial interaction graphs: a phone-call communication graph, and a human movement graph. We find that daily detection can generate a more stable partition structure than an hourly one, while monthly changes also exist over time. In addition, the intra-season similarity is generally higher than inter-season similarity. We apply two popular techniques for community detection (i.e., a fast-greedy modularity maximization algorithm and a multi-level algorithm) in this work. However, no significant difference is found between temporal changes of community structure by using these two detection algorithms.

The presented framework can help identify patterns of spatial interaction in both cyberspace and physical space with mobile phone call detail records in some regions where census data acquisition is difficult, especially in African countries. A potential value exists for supporting regional planning and policy making by mining large-scale geospatial datasets, although there has been some debate on whether to use mobile phone data or other big data analytics because of geo-privacy concerns. Related research in this direction (e.g., geospatial data anonymization) might attract more attention from both academic researchers and industry engineers.

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. doi: 10.1111/tgis.12042.

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. doi:10.1068/b38141.

Gao S (2015) Spatio-Temporal Analytics for Exploring Human Mobility Patterns and Urban Dynamics in the Mobile Age. *Spatial Cognition and Computation*, 15(2), 86-114. 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.