Where do we Develop? Discovering Regions for Urban Investment in Senegal

Derek Doran
Wright State University - Main Campus, derek.doran@wright.edu

Andrew Fox

Veena Mendiratta

Follow this and additional works at: http://corescholar.libraries.wright.edu/cse

Part of the Computer Sciences Commons, and the Engineering Commons

Repository Citation

This Conference Proceeding is brought to you for free and open access by Wright State University's CORE Scholar. It has been accepted for inclusion in Computer Science and Engineering Faculty Publications by an authorized administrator of CORE Scholar. For more information, please contact corescholar@www.libraries.wright.edu.
Where do we Develop? Discovering Regions for Urban Investment in Senegal

Derek Doran  
Dept. of Computer Science & Engineering  
Kno.e.sis Research Center  
Wright State University  
derek.doran@wright.edu

Andrew Fox  
Dept. of Industrial Engineering and  
Management Science  
Northwestern University  
andrewfox2014@u.northwestern.edu

Veena Mendiratta  
Bell Laboratories  
Alcatel-Lucent  
veena.mendiratta@alcatel-lucent.com

Abstract

The rate of urbanization in developing countries, defined as the speed with which a population shifts from rural to urban areas, is among the highest in the world. The disproportionate number of citizens that live in a small numbers of cities places incredible pressure on the largest cities in these countries, which may already be faced with limited resources, weak industrialization, and underdeveloped infrastructures. Urban planning researchers as well as policy makers have suggested that governments in developing countries make capital investments within and surrounding smaller cities to attract citizens away from large urban centers, thereby lowering the pressure placed on overpopulated urban centers and making it more attractive for citizens to migrate to the smaller cities. This paper proposes a methodology that maps signals in mobile phone usage data to long-standing urban planning theories. These signals are subsequently combined in an unsupervised learner to discover regions within which city investments should be made. Qualitative evaluations of the selected arrondissements illustrate the promise of our approach.

1 Introduction and Motivation

A virtually universal trait across developing countries is the extraordinarily high rate of urbanization, which is defined as the migration of citizens from traditional, tribal, and rural regions to large city centers [14]. Ever increasing political turmoil in rural or tribal towns, ecological breakdowns, and the romantic (and often unrealistic) notion held by citizens that great opportunity exists in urban areas as compared to rural towns all contribute to such high urbanization rates [23]. However, urbanization is one of the most pressing challenges that faces developing nations. This is because urbanization leads to very poor living conditions as a city’s population exceeds its capacity with respect to infrastructure and available jobs. They also encourage an unstable bazaar economy that is impossible for the country’s government to tax or regulate, high rates of crime, and pollution. Urbanization is also causally related to the fact that developing countries exhibit a distorted and interdependent economy that produces products specifically for developed countries, and has large population growth and widespread poverty. The country of Senegal is no exception to the urbanization phenomenon; over 42.5% of the population lives in urban areas\(^1\) and over 71.9% of citizens living in the country’s 50 most popular cities reside in Dakar and Grand Dakar. To further demonstrate the intensity with which urbanization occurs in Senegal, Figure 1 shows how the majority of Senegal’s population is concentrated on its West coast, and the top quarter of cities with highest population density primarily lying in a region to the east of Dakar and Grand Dakar.\(^2\)

Urban planning researchers and policy makers concur that an effective way to reduce the negative effects of urbanization is to encourage a country’s citizens to migrate out of, rather than into, overpopulated urban centers by investing in the rapid development of promising towns and cities in alternative areas of the country [16]. Doing so simultaneously relieves the pressure applied to large central cities while Investing dollars into the development of new cities that will add to the power of the country’s economy. The ideal town or city for rapid investment is one that already has an established local economy, has a developed infrastructure that supports its present inhabitants, and is self-sustaining; that is, it is located sufficiently far from existing large urban centers so that it does not rely on their economy, people, or ser--

\(^1\)http://www.indexmundi.com/senegal/urbanization.html  
\(^2\)http://en.wikipedia.org/wiki/Template:Largest_cities_of_Senegal
services to thrive [1]. However, the socioeconomic data about towns and cities in developing countries that is required to measure economy, self-sustainability, and infrastructure development is understandably unreliable, dated, and difficult to collect [6]. This makes it all but impossible for researchers and government officials to identify promising locations for investment, and hence reduce urbanization in a developing country. Rather than relying on empirical data, we may alternatively rely on theoretical models of urban development as exhibited by developed and emerging countries, on theories developed by geographers and urban planners that explain how and why cities within a country essentially ‘self-organize’ into predictable patterns according to universally applicable geographic, economic, and social constraints. Central Place Theory (CPT) is a long-standing, hotly debated, and recently more accepted theory explaining such self-organization of cities across a country [2]. It hypothesizes that some cities in a country are ‘Central Places’ that carry a very high population and produce a disproportionately large number of goods. Other types of communities, namely small ‘villages’ and middle-sized ‘towns’, naturally develop at different distances from central places depending on their reliance to its goods, people, and economy, with ‘middle towns’ being self-sustaining yet less developed compared to central places. The more recent Central Flow Theory (CFT) postulates that cities develop in a cooperative manner by sharing information and interests using modern technology so that distance is not a constraining factor. Intriguingly, there is almost no work towards operationalizing these concepts to quantitatively assess the degree to which geographical areas follow this pattern. Such an operationalization would be immensely beneficial; identifying locations that central place or central flow theory identifies as a self-sustaining middle town would strongly suggest that, with appropriate investment, it could one day become a central place that relieves the urbanization effect of closely connected, overpopulated urban centers.

This paper proposes an innovative approach that uses mobile phone data to operationalize concepts from CPT and CFT to identify locations in Senegal where increased investment is most likely to (theoretically) reduce the migration of citizens to the large over-populated urban cities and instead make it more attractive to migrate to the newly emerging urban areas. Our approach is unique in its: (i) ability to identify promising locations for urban development without needing to rely on detailed socioeconomic data; and (ii) quantification of geographic and urban planning theories through the use of mobile phone data. Given the fact that mobile phone data is collected across many of developing countries already [3], our approach may be applicable for any nation facing intense urbanization.

The layout of this paper is as follows: Section 2 introduces Central Place and Central Flow Theory, concepts on which our methodology is based. Section 3 identifies fea-
2 Central Place and Central Flow Theory

Geographers have developed two spatial theories that attempt to explain how and why urban centers are distributed across a geographic space. This section describes these two theories in more detail, and through a preliminary analysis of regional data over Senegal, finds evidence that supports these theories within the country.

2.1 Central place theory

Central Place Theory (CPT) is a method to explain the tendency of villages, towns, and cities to self organize according to a cascading spatial hierarchy [7]. It proposes a spatial organization illustrated in Figure 2 where small villages and towns (low places) and secondary centers (middle places) lie in regions where larger urban centers (central places) carry their influence. The hierarchical structure is centered at an urban center or Central Place - a large population zone able to supply goods and services (low-level outputs) as well as knowledge and culture (high-level outputs) to its surrounding area. Thus, a necessary requirement for a Central Place to thrive is sufficient distance from other Central Places, so that neither offers a redundant and wasteful outputs that a nearby Central Place would already satisfy. Low Places, manifested as towns and and villages, live within the sphere of influence of a Central Place. Due to their strong reliance on the nearby Central Place for low- and high-level outputs, they may have low population, have an underdeveloped local economy, and carry a weak infrastructure. We define communities living on the periphery of a Central Place’s influence as a Middle Place. Middle places are by necessity partially self-reliant due to the larger geographic distance between them and the Central Place. They are able to produce some, but not all of the low-level outputs provided by Central Places and remain reliant on Central Places for high-level outputs. Being located at the periphery of regions of influence, Middle Places are by definition situated between a number of other Central Places and may exert a pressure on all of them simultaneously. Despite their less developed infrastructure, the ability for these self-sustaining Middle Places to agglomerate resources from a number of independent Central Places [28] places them in a unique position to integrate knowledge and resources that would otherwise be separate from each other [27].

While the hierarchical signature of CPT can be seen across many landscapes [8, 20, 13, 32], there has been limited work towards operationalizing or modeling the phenomenon so that it may be applied to geographic datasets. These limited contexts include mathematical models based on CPT to predict city population growth [31], understand the hierarchical organization of cities over a geographic area [18], evaluate the way CPT interplays with economic growth over a spatial area [19], and to help explain geographical factors impacting phenomena such as sports tourism [9]. CPT has undergone a recent resurgence in popularity given its complementary relationship with modern urban economic geography theories, and is accepted as a reasonable model for explaining the spatial patterns of city development [29].

To evaluate the degree to which CPT is exhibited across Senegal, we scraped detailed population and location data across 6,135 cities, towns, and villages over Senegal from...
version 2.2 of the Global Gazetteer. As expected by CPT, major population centers are located far away from each other, as seen in Figure 4 that plots the locations of Dakar, Louga, and Thies, which are among the most populated cities in Senegal. These 3 cities, as expected according to CPT, are located far enough away from each other so that their population, economy, cultures, and support provided to their immediate regions do not interfere with each other.

Figure 3 explores how the mean population and percent change in population among cities that lie within 5km bands radiating from the center of these three cities changes with distance. We identify a pattern where populations quickly drop near a Central Place, and then remain steady or slowly rise for cities ever farther away. Spikes may signal a Middle City that can sustain a larger population. To better identify population increases that may represent a Middle Place, the blue plots on the bottom row of Figure 3 compare the percent change in city populations as a function of distance. The drastic downward spikes seen within 20km from the Central Place, and again at approximately 60km and 170km as we move away from Louga, and at 60km and 80km as we move away from Thies, correspond to the big population declines between the other Central Places within 200km and the Low Places that immediately surround these Central Places. For example, Dakar is approximately 178km away from Louga and 77km from Thies by road, while Louga is 114km away from Thies. These fluctuations in population as a function of distance from a Central Place are a strong signal that CPT may explain the distribution of cities in Senegal.

According to CPT, Middle Places should have a high potential to evolve to become economic and cultural drivers for a country by developing into new Central Places. This is because Middle Places are already positioned in between the influences of existing Central Places, thus minimizing the disturbance of their evolution into a Central Place on the economies of neighboring cities. They are also already self-sustainable, with an infrastructure in place that supports a moderate population and production of goods and services. Finally, Middle Places have the ability, in the future, to create new low- and high-level outputs by agglomerating the outputs provided by nearby Central Places. We therefore hypothesize that such Middle Places are the most promising locations for economic and infrastructure investment in a developing country to mitigate the negative effects of increasing migration to existing large urban centers (Central Places).

2.2 Central flow theory

Central Flow Theory (CFT) is a recently proposed theory for explaining urban development that is complementary to CPT [35]. Whereas CPT is anchored around the spatial influence of Central Places, CFT describes non-local interactions among places without regard for physical distance. It also emphasizes the cooperative aspects of place interactions where information, ideas, specialists and other ‘foreign’ commerce are exchanged for mutual economic benefit rather than an organization of places into a dependency hierarchy. The complementary nature of the two theories have been seen in studies on the historical development of various urban places. Large Central Places interact with their geographic surroundings and nearby cities (CPT) [17] to provide outputs that drive their economy, but their further development hinges on the free exchange of ideas and integration

---

of ‘foreign’ commerce (CFT) [34, 33]. Agent-based simulations further explain the interlocking relationship between CPT and CFT for Central Place development [21]. We hypothesize that places performing such exchanges occurring at a low to moderate rate (compared to the level of exchanges occurring among Central Places) signal a willingness to integrate foreign commerce, and already have the capacity to share new ideas and information with places they may not be dependent on according to the CPT hierarchy. These are all desirable properties that would magnify the effects of economic and infrastructure investments.

To evaluate the degree to which CFT holds across Senegal, we use a (meta) dataset consisting of all mobile phone calls in the time period between January and December 2013 [10]; the data is at the level of cell phone towers. Figure 5 plots the distribution of the total duration of all conversations made between the 1,666 towers in the country over this time. The distribution exhibits a clear power-tailed shape as seen in the distribution of calling activity across many other mobile phone datasets [11, 30, 5]. We seek to use this mobile phone communication data as a proxy for the amount of information or ideas exchanged between individual places. Towards this end, we only consider communication between towers whose cumulative duration of all conversations fall in the top 1.5% of this distribution, which translates to an average of 2,739 minutes of conversations per day. This filtering step leaves 38,613 flows of communication that fall in the tail of the distribution in Figure 5, where statistically significant calling activity occurs.

We subsequently form an undirected graph where nodes represent towers and edges correspond to the flows of activity as described above. To evaluate the popularity of a calling tower (e.g. the extent to which information and ideas are exchanged within places nearest to it) we consider the PageRank centrality of towers in this graph. PageRank considers the popularity of calling tower *i* to be proportional to the popularity of the tower it communicates with. It is defined by:

\[
p_i = \alpha \sum_j A_{ij} \frac{p_j}{g_j} + (1 - \alpha) \frac{1}{N}
\]

where \( A \) is a matrix with \( A_{i,j} \) given as the cumulative length of all conversations between towers *i* and *j*, \( k_j \) is the degree of node *j*, \( g_j = \max(1, k_j) \), \( N \) is the number of towers, and \( \alpha = 0.87 \) is a damping parameter set according to the recommendations based on earlier work [4]. In Figure 6, we compare the location of the 10 most populated cities in Senegal in the Global Gazetteer against the location and PageRank centrality of calling towers (larger vertices correspond to higher PageRank). We identify a strong correlation between the position of the most popular cities (Central Places) and the location of call towers that exhibit the largest amount of activity, as predicted by CFT. We also observe, even though the distribution of PageRank centralities is skewed, many call towers with high PageRank lying between these most populated cities as seen in Figure 6(a). Although these locations may not have a high population, large PageRank centrality suggests that places around these towers are undergoing significant exchange of ideas and information with Central Places. According to CFT, such exchanges are positive indicators for these places becoming Central Places in the future.

### 3 Identifying Middle Places for Development

Recognizing the fact that CPT and CFT may help explain urban development in Senegal thus far, we consider unsupervised methods for identifying areas in the country most likely to correspond to Middle Places. We intentionally decided not to focus on supervised methods for this problem as there is virtually no ground truth data available for what is considered to be the ‘best’ place for urban investment. Instead, our unsupervised approach considers a number of features from a dataset of mobile phone calls that are theoretically linked to CPT and CFT, and combines these in a methodology that classifies arrondissements by the types of places (Central, Middle, or Low) they support. We chose to classify arrondissements rather than individual towns because: (i) high-resolution data expressing the calls made between villages, towns, and cities are unavailable; (ii) government investments in urban development can likely be more easily be budgeted for an administrative unit, rather than for a specific city; and (iii) arrondissements that support Middle Places may be prime areas for infrastructure investment, and for making modern investments such as development of planned communities or technology parks. In this section, we present the features we consider for modeling and the classification methodology.

#### 3.1 Features considered

We consider four different features of a dataset consisting of mobile phone calls made between call towers in each arrondissement of the country. We chose features that, according to CPT and CFT, should take on an extreme value if an arrondissement supports the development of Middle Places over Low or Central Places. These features include:
Figure 6: Spatial comparison of the most populated cities in Senegal and Pagerank centrality of calling towers

- **Total call volume**: This is defined as the total number of calls placed by mobile users in an arrondissement.

- **Distance of Calls**: This feature is defined as the $X^{th}$ percentile of the distribution of the geographic distance calls placed by cities in an arrondissement travel. This feature provides consideration for the geographic component of CPT, where Middle Places tend to find themselves far away from the Central Places they contact for information and knowledge. The best value of $X$ is found during model selection.

- **Demand-weighted distance of calls**: This is defined as the sum of call durations weighted by the physical distance that each call traveled in kilometers.

- **Self-Sufficiency**: This is defined as the percentage of an arrondissement’s calls that occur between mobile cell towers within the same arrondissement. This percentage reflects the “locality” of calls made within an administrative region; areas with strong internal communication suggest a weaker reliance on the information provided by people located in other arrondissements.

- **Partnership**: This is defined by counting the number of unique arrondissements that comprise the top 80% most active connections (in terms of call volume) from an arrondissement. Noting that Central Places combine information from a number of other places in order to create new products and knowledge, it should be the case that the most active communications from an arrondissement supporting Middle Places connect to as many external locations as possible.

- **Centrality**: We represent all calls between arrondissements as a graph, with an edge feature as the total number of calls between two arrondissements. We then consider the eigenvector, PageRank, and betweenness centrality. The betweenness centrality of arrondissements in this graph measures the number of shortest paths that pass through the arrondissement being measured. Betweenness centrality thus reflects the ability of cities in an arrondissement to connect to other locations in Senegal, thus acting as a broker of information and resources, and as a place where ideas and knowledge across the country meet. Eigenvector and PageRank metrics score an arrondissement on the graph based on the scores of other arrondissements it is strongly connected to; thus Middle Places may take on high values due to their (theoretically) strong connections to many Central Places.

### 3.2 Methodology

We classified arrondissements by the degree to which they support Middle Places by clustering over a vector that represents an arrondissement and whose components are defined as the value of the features presented above. $K$-means clustering is a standard algorithm for clustering such vectors, however it is very sensitive to initialization and the distance measure used. Instead, we work with Finite Mixture Models (FMM) that search for a best fitting mixture of probabilistic data distributions that explain the total distribution of values exhibited in the entire dataset. FMM relaxes many of the constraints imposed by $k$-means clustering and is less sensitive to the scale and range of values of the features [15]. Relaxation of these assumptions is suitable to the research objective of operationalizing CPT/CFT because a larger proportion of places should be characterized into a Low Place cluster, followed by Middle Places, and finally Central Places. Cluster sizes should also follow this pattern. We used the mclust Finite Mixture Modeling software package in R to search for clustering solutions where the mixed models were part of the exponential family. The package reports results from many combinations of hyperparameter settings that encode assumptions about the types
### Table 1: Correlation between distance of calls and self-sufficiency features

<table>
<thead>
<tr>
<th>Distance Traveled by X% of Calls</th>
<th>Correlation with Self-Sufficiency</th>
<th>Variance of Distance Traveled</th>
</tr>
</thead>
<tbody>
<tr>
<td>50% (median)</td>
<td>-0.58</td>
<td>454</td>
</tr>
<tr>
<td>60%</td>
<td>-0.37</td>
<td>854</td>
</tr>
<tr>
<td>70%</td>
<td>-0.17</td>
<td>3,581</td>
</tr>
<tr>
<td>80%</td>
<td>0.10</td>
<td>10,872</td>
</tr>
</tbody>
</table>

Table 2: Clustering solutions with different variable settings

<table>
<thead>
<tr>
<th>Solution</th>
<th>Variables</th>
<th>BIC</th>
<th>Pseudo-F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best</td>
<td>Self-Sufficiency, Partnership, Betweenness Distance ((X = 60%))</td>
<td>-1.288</td>
<td>41.6</td>
</tr>
<tr>
<td>Alt. A</td>
<td>Self-Sufficiency, Partnership, Betweenness, Distance ((X = 50%))</td>
<td>-1.297</td>
<td>46.2</td>
</tr>
<tr>
<td>Alt. B</td>
<td>Self-Sufficiency, Partnership, Betweenness, Demand-Weighted Dist.</td>
<td>-1.342</td>
<td>31.0</td>
</tr>
<tr>
<td>Alt. C</td>
<td>Self-Sufficiency, Partnership, PageRank Centrality, Demand-Weighted Dist.</td>
<td>-1.316</td>
<td>38.4</td>
</tr>
</tbody>
</table>

3.3 Model selection

Model selection criteria in unsupervised learning has an inherent level of subjectivity due to the latency of the dependent variable, and no observable outcome exists to compare model validity against [26]. We adopt the following process to evaluate a potential solution that classifies arrondissements by the degree to which they support Middle Places in terms of the following criteria, in order of priority:

1. **Multicollinearity**: prior to introducing independent variables into a clustering model, high correlations between variables inhibit variable selection. Correlations of \(>|r|>0.5\) are considered high, and correlations between 0.3 and 0.5 are monitored as we evaluate the solution using criteria (2) through (4). If two variables are highly correlated then these are not introduced into the models because they overstate the impact of their phenomena on the solution.

2. **Actionability**: In this criteria, we ask if cluster variables and boundary values allow for a governing body to take action on the results. For instance, if Middle Places, as defined by the CPT and CFT features, fall entirely within Grand Dakar or if they comprise a large proportion of Senegal’s cities, the ability for an organization to take action on the results is limited. This is a logical and subjective, yet necessary, criterion.

3. **Bayesian Information Criteria (BIC)**: Finite Mixture Modeling, the primary clustering technique used in our work, utilizes BIC as the key statistic for comparing solutions [12]. It is defined as:

\[
B = 2 \log P(X|M, \Theta) - d \log n
\]

where \(X\) is the set of observed data vectors, \(M\) is the fitted clustering model with maximum likelihood parameters \(\Theta\), \(d = |\Theta|\), and \(n = |X|\). Models with larger \(B\) tend to be better models, since if the data \(X\) fits the model \(M(\Theta)\) well, its log-likelihood should be higher.

4. **Pseudo-F Statistic**: The Pseudo-F statistic is a measure of the efficiency of a clustering result. It is defined as the ratio of the mean sum of squares distance between vectors in different clusters to the mean sum of squares distance between vectors in the same cluster [22]. Larger Psuedo-F scores correspond to ‘tighter’ clusterings where intra-cluster distances between vectors is small and inter-cluster distances are high.

In our analysis we found that total call volume, demand-weighted distance of calls, weighted average distance of calls, Eigenvector centrality, and PageRank centrality were heavily skewed to a very small number of well developed cities including Dakar. This skewness reduces the actionability of results; they would consistently suggest that Dakar and other well developed cities should be further developed, but it is difficult to channel resources into these complex...
urban spaces. Some of these features also caused multicollinearity issues; for example Figure 8 identifies how PageRank centrality exhibits high correlation with call distance and self-sufficiency.

Table 2 enumerates through FMM solutions using features such as self-sufficiency, partnership, betweenness, call distance, PageRank, and demand-weighted distance. We found that the best solution given in the first row identifies 4 clusters using the self-sufficiency, partnership, betweenness centrality, and distance at the $X = 60^{th}$ percentile. Besides exhibiting the highest BIC and nearly highest Psudeo-F, we found that setting the call distance feature using the $60^{th}$ percentile of the distribution minimized the correlation between this feature and self-sufficiency. As seen in Table 1, the $60^{th}$ percentile is an approximate elbow point that reduces correlation while maintaining a small amount of variance that does not heavily skew this feature value to Central Places that almost the entire country contacts (e.g. Dakar).

Figure 7 uses a Self-Organizing Map to visualize the distribution of the features used in the best clustering solution across the arrondissements of Senegal. The colors of the nodes in each map represent the scaled values of the features from low (cool colors) to high (hot colors). The number of nodes of some color is proportional to the number of arrondissements whose value is in the range represented by the color [25, 36]. Note that each map is initialized with a random assignment of arrondissements to nodes. The maps identify how the distance of calls, partnership, and self-sufficiency metric exhibit a small skew towards a small number of arrondissements (those that host Middle Places) while betweenness centrality is better distributed. The more even distribution of betweenness centrality is likely due to the fact that both Middle and Central Places have important brokerage locations for information and communication across the country, hence both types of Places may be represented by the hotter nodes. The large number of cool betweenness centrality and partnership nodes capture the Low Places that do not serve as brokers of any kind of information nor do they communicate with a large number of external places.

Figure 7: Feature values for best clustering solution

4 Cluster results and discussion

Figure 9 uses a dot plot to present the centroid positions of the four clusters in the best FMM solution. We subjectively map these values to being relatively LOW, MODERATE, or HIGH for each cluster in Table 3. We label each of the four clusters as:

- **Cluster 1: Dakar and its Suburbs.** Cluster 1 identifies 8 arrondissements that, as visualized in Figure 10(a), contain Dakar and its suburbs. These arrondissements show high betweenness, meaning they are hubs for calls throughout the country. Yet, their low call distance and partnership implies exclusivity; information flow passes primarily through partners within the same cluster. This cluster is grouped by both geography and numerical values of the features, supporting the theoretical definition of a Central Place.

- **Cluster 2: Middle Places.** The nine arrondissements placed in Cluster 2 quantitatively support the definition
Table 3: Cluster labels and features

<table>
<thead>
<tr>
<th>Cluster Description</th>
<th>Self-Sufficiency</th>
<th>Partnership</th>
<th>Betweenness</th>
<th>Distance</th>
<th>Cluster Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central Places: Dakar and Suburbs</td>
<td>MODERATE</td>
<td>LOW</td>
<td>HIGH</td>
<td>LOW</td>
<td>8</td>
</tr>
<tr>
<td>Middle Places: Emerging Opportunities</td>
<td>HIGH</td>
<td>MODERATE</td>
<td>HIGH</td>
<td>LOW</td>
<td>9</td>
</tr>
<tr>
<td>Low Places: Villages Between</td>
<td>LOW</td>
<td>MODERATE</td>
<td>LOW</td>
<td>MODERATE</td>
<td>37</td>
</tr>
<tr>
<td>Middle-Low Places: Common Towns</td>
<td>MODERATE</td>
<td>MODERATE</td>
<td>MODERATE</td>
<td>MODERATE</td>
<td>69</td>
</tr>
</tbody>
</table>

Figure 10: Cluster assignments (a) and Middle Places (b)

for Middle Places. They have high self-sufficiency, and when calls do leave these areas, they are reaching a large number of other arrondissements. The low distance that calls travel may be due to short connections with other proximate cities, which CPT supports.

- **Cluster 3: Low Places.** The 27 arrondissements in this cluster exhibit a low degree of self-sufficiency and betweenness, and a moderate level of partnership and call distance. Low self-sufficiency is an indicator of a Low Place which needs to strongly rely on nearby other places for resources and information. Similarly, a low betweenness value indicates that the location is not a broker of information, and that these arrondissements are not of interest to most other arrondissements. In Figure 10 (a), the Low Places (blue positions) tend to be surrounded by a number of other nearby arrondissements, further supporting the notion that they rely on nearby Central Places, Middle Places or Low-Middle Places (Cluster 4).

- **Cluster 4: Low-Middle Place.** Finally, the majority of arrondissements fall into a cluster with moderate self-sufficiency and betweenness as well as partnership and distance, suggesting that they support a mixture of Low and Middle places. The positions of such arrondissements in Figure 10 (a) find them to be near Dakar and its suburbs, (b) by the border of the country, (c) in remote regions, or (d) immediately surrounded by arrondissements that only support Low Places.

Because arrondissements in Cluster 2 support Middle Places much more strongly as compared to those in Cluster 4, we further investigate the cities seen in Cluster 2 arrondissements to validate that they exhibit features that make them promising opportunities for urban development: (i) **Thies.** Thies is one of Senegal’s largest cities and sits in an area considered to be a transportation hub that services routes between St Louis, Dakar and Bamako. It is also a major producer of peanuts and fertilizer that are among the country’s top exports, and host reserves of important metals. It thus has the potential to become an even stronger economic hub for the city under further investment. (ii) **St Louis.** St Louis is the capital of Senegal’s St Louis arrondissement and is located in the northwest of the country near the mouth of the Senegal river on the Mauritanian border. It was a capital of Mauritania which at the time was a neighboring colony. It has a heavy tourism based economy, has a high rate of sugar production, fishing irrigated alluvial agriculture, pastoral farming, trading and exportation of peanut skins. The city was listed as a UNESCO World Heritage Site in 2000 and cultural tourism has become an engine

---
4[http://www.aljazeera.com/indepth/features/2012/02/201222695110410730.html](http://www.aljazeera.com/indepth/features/2012/02/201222695110410730.html)

of growth. (iii) Mbour. Mbour is a city in the Thies Region of Senegal. It lies on the Petite Cote 80km south of Dakar. The city’s major industries are tourism, fishing and peanut processing. It is Senegal’s fifth largest city and, by some indicators, is among one of the fastest growing. (iv) Ziguinchor. Ziguinchor is a river-port town in southwestern Senegal lying along the Casamance River. It is one of the largest cities in Senegal, but is largely separated from the north of the country by The Gambia. Ziguinchor remains economically dependent on its role as a cargo port, transport hub and ferry terminal. A primary highway crosses the Casamance River just east of the city, linking the region with Bignona about 25km to the north, and (via The Gambia), the rest of Senegal. It features a large peanut oil factory and is also known for producing great quantities of rice, oranges, mangoes, bananas, cashews, tropical fruits and vegetables, fish, and prawns. Ziguinchor is also home to a new University which opened in 2007.

5 Conclusions and Future Work

In this paper we introduced a data driven methodology to identify the most promising areas in Senegal for economic investment. We identified features, using mobile phone data, that speak to Central Place and Central Flow Theory, which are important geographic and urban planning theories that explain the way cities in a country naturally develop. To the best of our knowledge, this paper is the first attempt made to operationalize these theories for forecasting the places in a country where investments should be made, and to quantify CPT/CFT concepts in a dataset of mobile phone records. Future work will examine alternative clustering methods and distance metrics that define similarity, formulate other data features that are related to CPT and CFT, and reformulate our idea as an optimization problem that ranks arrondissements in order of the ‘best’ places in Senegal for investment.

References


