Freetown - Characterizing a City in a Data-Scarce Environment

CY PLAN 257

Team:

Jeremy Halpern (MS CEE/MCP CED)
Marie Pelagie Elimbi Moudio (IEOR Phd)
Andrew Nelson (MS CEE/MCP CED)
Yuan Tian (MS Geosystems)

Introduction

Motivation

In August 2017, a devastating landslide struck the Freetown Peninsula in Sierra Leone. After combining with high-volume river flow, the resulting mudslide caused severe and lasting impacts. While over 1,100 people were declared dead or missing and 6,000 people lost home or property, the mudslide also greatly affected the city's transportation infrastructure. Eight pedestrian bridges, two road bridges and 5.5 kilometers of roadways were destroyed.

As a result, the World Bank (WB) is seeking to make targeted investments in the surface transportation infrastructure. However, Freetown does not have a clear hierarchy of roads or pedestrian facilities in most high pedestrian traffic areas. Data publicly available in the United States like origin-destination (OD) data and travel diaries are not available. Strategies to help guide investment must be creative and take advantage of freely available or inexpensive data.

Previously, with only topographical street network data available through OSMnx, limited shapefiles for river polylines and digital elevation maps, a group study sought to identify parts of the street network which were both vulnerable to disruption events and important to the network using the network statistic betweenness centrality (BC). BC is a measured by calculating the shortest paths between every origin and destination node in the system and finding the percentage of those shortest paths which pass through the node or edge of interest. For example, a node with a BC of 0.35 means that 35% of all the shortest paths between every combination of origins and destinations passes through that node. By finding nodes which had both high vulnerability and high BC, the study tested scenarios in which those nodes were disabled, and how centrality changed throughout the system.

However, there were serious drawbacks to this method. Using only the BC measure to determine the importance of the road infrastructure only focused on the physical structure of the network. It failed to consider demand in the measure for importance, and therefore failed to consider roads which may be heavily trafficked while existing in the physical periphery of the network. The utility of future study is dependent upon the ability to estimate network demand in addition to structure.

This study seeks to begin characterizing the surface transportation network in Freetown by demand. Using data provided by the WB in concert with publicly available data, we first seek to describe the population spatially by clustering areas by population density thresholds. We then compare population density distributions to other cities. By incorporating point of interest (POI) data from Google Places, we then use an extended radiation model to estimate trips between zones. Finally, we assign these trips to the physical networks as weights and characterize roads by both demand and BC.

Data

Data for this study came from a mixture of private and public sources. Broadly, they can be categorized by their points of origin: the WB, Google, and OSMnx.

World Bank

Population hexagons

The WB had access to population data, which were created as a result of a survey with consultant Arup. This survey estimated population based on expected average number of people per story of formal and informal structures and was validated by Sierra Leone's 2015 Population and Housing Census. These data were provided in shapefile format, where population was aggregated to 100m-diameter hexagons. Population and population density per hexagon were provided. These population estimates provided an aggregate population total of 1,084,320, which corresponds approximately to the 1,055,964 total reported in the 2015 census.

Administrative boundary

The administrative boundary of the city of Freetown was provided in shapefile format. Limited edits to the borders of this polygon were made to encompass the entirety of the provided population data.

Orange cell tower positions

The WB additionally provided geo-coordinates of Orange cell towers. While the World Bank recently purchased anonymized cell phone data from two service providers which tracks the cell towers used for calls, texts, web browsing, etc., this anonymized data was not yet available. In anticipation of the availability of this data in the future, we used the location of the cell phone towers as a reference to create zones of analysis. This process is detailed in the Methods and Results: Trip Distribution section below. While there were 330 cell phone tower locations provided for Sierra Leone, only 111 fell within the administrative boundary of Freetown.

Google

POI were scraped from Google using the Google Places API, an iterative process which requested the closest locations on Google to single points in space. This was repeated several hundred times across a wide area to get coverage for the whole of Freetown Peninsula. A total of 1,099 POI were found within the administrative boundary of Freetown.

OSMnx

OSMnx is an Python library which allows street networks from OpenStreetMap to be downloaded. It displays these street networks as graphs, with nodes representing intersections and edges representing streets. Since the physical structures of street networks are interpreted in the mathematical language of nodes and edges, several statistics can be calculated to describe the system. An additional benefit to OSMnx is that the edges are assigned polyline geometries, which allow them to be displayed realistically. OSMnx represents the Freetown surface transportation network within the administrative boundary as a graph with 7,118 nodes and 8,747 edges.

Literature Review

Population Clusters and Percolation

In "Modelling urban growth patterns," Makse et al. propose a method to model urban growth in cities that establishes a correlation between the existing units and the development of new units. This model is an improvement to the Diffusion Limited Aggregation (DLA), used to predict a large cluster, with most of the cluster growth taking place at the tips of the cluster branches. In this method, the population density is assumed to follow an exponential distribution where the population decays radially away from the highest density area. In addition, the method accounts for the fact that the development of new units is not random but rather correlated to existing units. This is because if a unit is occupied, the probability of neighbouring units becoming occupied increases. The authors provide a correlated percolation model to quantify the above interactions.

To introduce the correlation between variables, the authors convolutated the uncorrelated variables with a suitably chosen power law kernel. They defined a new set of random variables with long-range power law correlations that decayed as $r^{-\alpha}$. The power law assumption is used because decision for unit development decays gradually with distance from occupied region, and α is determined from empirical observations. The authors simulate correlated clusters from a fixed urban population density, λ.

The correlations cause the units to aggregate around an urban area. The correlated clusters created are compact around their centers and become less compact as distance from their centers increases. The perimeter of the connected cluster to the Central Business District (CBD), previously defined as the central core, was denoted the urban boundary of the largest city. They argue that correlation between occupancy probability can account for irregular morphologies in urban centers. In addition, the study finds that the town size distribution in 2 cities, N(A) follows the power law, where A is the size of the cluster..

The main findings in the paper included the following:

- The mean distance of the perimeter from the center depends on the percolation threshold, r_f
- The small clusters are situated at distance r such that $r > r_f$
- The maximum Area occupied by a cluster situated at distance r from the CBD

The authors check their model by plotting the model result versus the actual data from cities of Berlin and London on a log-log plot and observe a power law for both cities. In addition, a semi-log plot of density of occupied area versus radius for Berlin is plotted. Least square fits for various λ values show decreased λ in with time.

In Methods and Results below, a similar analysis will be performed on the city of Freetown as described in the "Modelling urban growth patterns" paper. Percolation theory and fractal analysis will be used to deduce the distribution clusters with the city.

Population Density Distribution

In Alain Bertaud's working paper on the spatial organization of cities paper (2004), he attempts to define cities' urban spatial structure along three primary characteristics: daily trip patterns, average built-up density and density gradient. He posits that city formation and growth are fundamentally based around labor markets. Initially, cities tend to be monocentric around a central business district (CBD) with most workers traveling to the center. As cities grow, they tend to become more polycentric with multiple strong centers of activity. No city is completely monocentric or polycentric but generally somewhere in between.

The speed of transition from monocentric to polycentric depends on characteristics of the city such as amenities in the CBD, topography, layout of the road network. Two key changes that come with a polycentric city are that trips tend to be longer and the density gradient should emanate from the center of mass rather than the highest density point.

Comparing different cities, the density of built up areas was not found to be correlated with income or population size. Rather they tend to group by geographic region. Density profiles reflect the density as a function of distance from the CBD. Given a functioning land market, there tends to be a negative density gradient reflecting the decreased value and lower density farther from the CBD. Exceptions to this, resulting in a positive density gradient, can occur with certain regulatory regimes such as South African apartheid, Moscow from communist government policy and Brasilia. This density gradient reflects where people live and sleep at night, not necessarily where they travel during the day.

Trip Distribution - To Be done by Yuan

Network Analysis and Route Assignment

In "The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles," Guimera et al. use the global air transportation network to evaluate the relationship between centrality and degree of network nodes.

Guimera et al. characterized the air transportation network as following a small world network model. Small world networks describe pairs of nodes that are connected by short paths and exhibit a high degree of cliquishness. Cliquishness, quantified by the clustering coefficient, is defined as the probability that two cities connected to the same third city are also connected to each other.

The study also contrasts the network to a random graph. In a random graph, all nodes have similar degrees, while in real world graphs, some nodes are significantly more connected than others. Many complex networks have degree distributions that decay as a power law. This can be explained by preferential attachment, or the tendency to connect preferentially with nodes that already have high degrees. BC in this case is also likely to be a scale-free distribution.

Guimera et al. explained the distribution of both degree and BC by the function:

$$P(> y) \approx y^{-\alpha} f(y/y_{\times})$$

Where:
y = betweenness centrality b, or degree k
α = power law exponent
$f(u)$ = a truncation function, and
$y_{\times}$ = a crossover value that depends on the size of the network

Generally, degree provides information on a node's importance to the network, but does not provide information on the role in the network. The BC provides this information, describing how central it is to the structure of the network.

By comparing the real air network with a similar randomized network, Guimera et al. found that while degree and BC are strongly correlated in both networks, the BC decays much faster in the randomized network, indicating anomalously large BC than expected for degree. In other words, many airport hubs had small degrees with large BC, or large degree with small BC. In comparing BC to degree, they found that while the relationship was well described by a quadratic function, the network is still comprised of many cities that have high BC but low degree and conversely many cities with low BC and high degree.

This is explained by the existence of distinct communities within the transportation network. Because the network was subdivided into more intra-connected communities, the role of each city hub had to be re-examined based on its relationship to its community and the broader network as a whole.

As is pertains to Freetown, our primary interest is in using the Guimera et al. paper as inspiration to relate edge BC to weighted degree and categorize surface transportation facilities based on their conformity or nonconformity to expectations of correlation. Demand-based weight of edges are a strong analogy to degree of air transportation network nodes because it describes how important the facility is to the network. Edge BC additionally serves as a strong analogy to BC in the air transportation network because it describes how central actual road facilities are to the structure of the network, and is described in more detail in Methods and Results, below.

While the Guimera et al. study focused additionally on identification of communities, there is not a strong analogy to the Freetown road network to recreate this portion of the study. Communities in the Guimera study were created due to complex socio-political, cultural and geographic constraints that made identification of the evolutionary pressures that formed the communities difficult. Freetown, operating under a single municipality, is unlikely to have such unidentifiable characteristics. Any factors brought into the formation of "communities" in the street network could likely be more easily explained by examining geography or interviewing local government officials. Additionally, the air transportation network is explained as a small world network, which does not explain street transportation networks well since, due to their structure, they are unlikely to exhibit cliquishness.

Methods and Results

Population Clusters and Percolation

To assess the population spread within Freetown, we used methods similar to those described in the "Modelling urban growth patterns" by Makse et al. The shapefiles containing the population hexagons were processed in python. The total population considered from our 100m hexagons was 1,084,320 people. To facilitate analysis, the centroids of the hexagons were used as points instead of the hexagons themselves.

Freetown Mapping and Resolution

Figure 1a shows a map of Freetown which is the region considered for the analysis. To carry out our analysis, the x and y scales of the plot (latitude versus longitude axes) are converted to pixels. This processing is done in matlab via the function. (Details in Code - Freetown Percolation) The resolution of the newly converted image is very dependent on a variable, dl. The optimal dl value was obtained by finding the average difference between latitudes or longitudes of adjacent data points. Based on our data, the dl value was set to 0.001. Figure 1b shows a similar image of Freetown with axes converted to pixels.

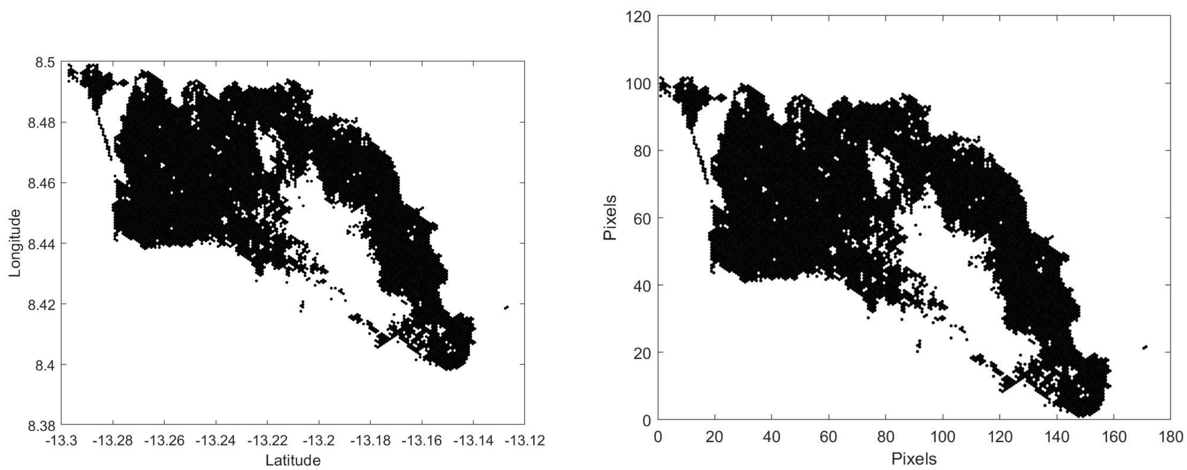

Figure 1: 1a - To the left, latitude versus longitude plot for the city of Freetown. 1b - To the right, Freetown plot with axes converted to pixels

The point with the largest population density is identified and noted as the CBD. A heat map of the distribution the log of population density is presented in Figure 2. The regions with highest population density marked in yellow are mainly located around the arc in the top right corner of Freetown, close the CBD. Figure 3a shows the population density plot for the city of Freetown plotted with the region with highest population density (within 0.3km) highlighted in green. In Figure 3b, a heat map representing the distance from the CBD is presented with the furthest distances marked in yellow and the closest distance in blue.

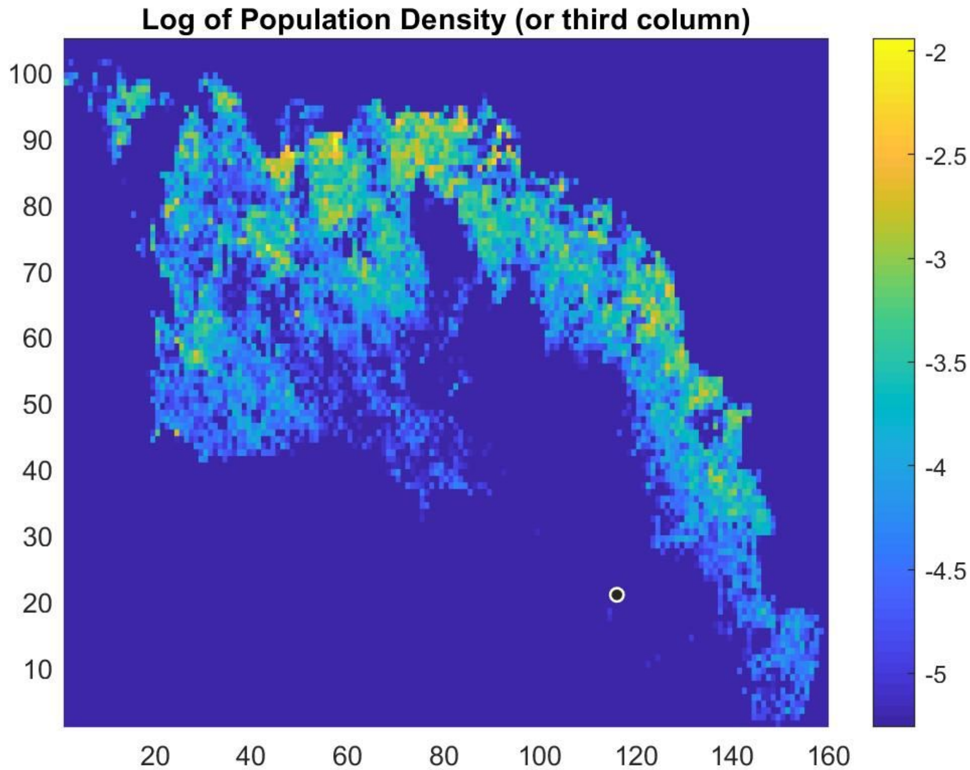

Figure 2: Heat map depicting log of Population density for the city of Freetown

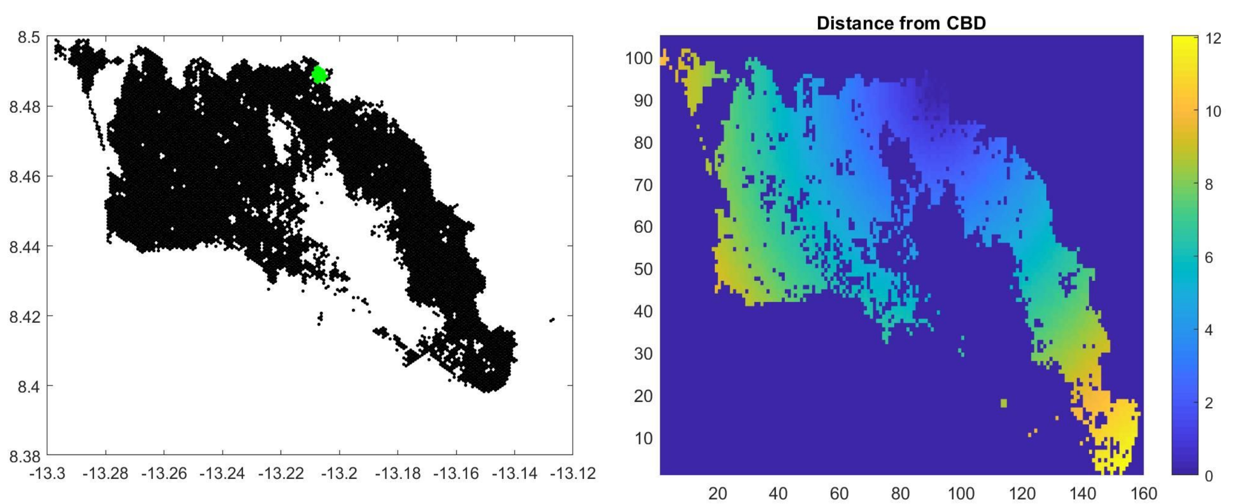

Figure 3: 3a - To the left, map of Freetown plotted with the region with highest population density (within 0.3km) highlighted in green. 3b - To the right heat map representing the distance from the CBD is presented with the furthest distances marked in yellow and the closest distance in blue

Percolation and Fractal Analysis

The goal of percolation analysis in the context of urban cities is to describe the distribution of population clusters spatially. From "Modelling urban growth patterns," there is a known relation between the the distribution of clusters sizes, N(A) and the fractal dimension of the perimeter described by equation 1.

$$N(A) \equiv \int_0^{p_c} n(A,p)\, dp \sim A^{-(\tau + 1/d_f \nu)} \quad (1)$$

Where p_c = percolation threshold

p(r) = occupation probability

ξ(r) = the average linear size of clusters at distance r

df = Fractal dimension

Equations (2) and (3) provide values for some parameters in equation 1.

$$\tau = 1 + 2/df \quad (2)$$

$$\xi(r) \sim |p(r) - pc|^{-\nu} \quad (3)$$

The Box counting algorithm was used to estimate the fractal dimension (df) of Freetown. This algorithm works by dividing the space containing the object under consideration into hypercubic cells of linear size, ε. Then one counts the number of boxes, N(ε). which that contain the object. This process is repeated multiple times and the fractal dimension, df, is obtained from the equation:

$$N(\varepsilon) \sim \varepsilon^{df} \quad (4)$$

We used a τ value of 187/91 based on theory.

Results and Discussion

From the Box counting algorithm, we obtained df ≈ 1.621 for the case of Freetown. This value, as expected, falls between the range of 1 and 2 because the map is an incomplete 2d object. The percolation threshold for population density was set to 0.02 people/m^2 after experimentation to find the minimum value needed for a visible connected cluster to appear. Figure 4a shows the occupied regions in Freetown based on the percolation threshold of 0.02 people/m2 with the giant component (largest connected cluster) labeled in red and remaining occupied regions labelled in blue. In Figure 4b, the log-log plot of number of boxes versus magnifying factor is used to estimate fractal dimension, df. We observe that the giant component labelled in red in figure 4a is found along the arc in the top right corner of our plot.

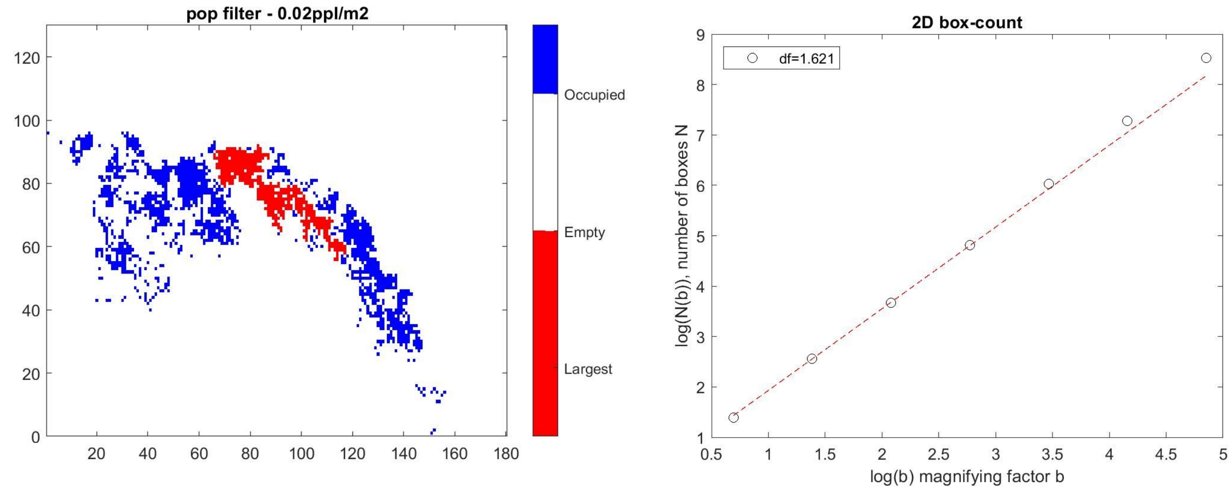

Figure 4: 4a - To the left, giant component (largest connected cluster) labeled in red and remaining occupied regions labeled in blue. 4b - To the right, log-log plot of number of boxes versus magnifying factor to estimate fractal dimension, df

Figure 5 provides a spatial visualization for different population clusters. The labeled plot of clusters with threshold = 0.02 people/m^2 is presented in Figure 5a. Figure 5b highlights the distribution cluster sizes.

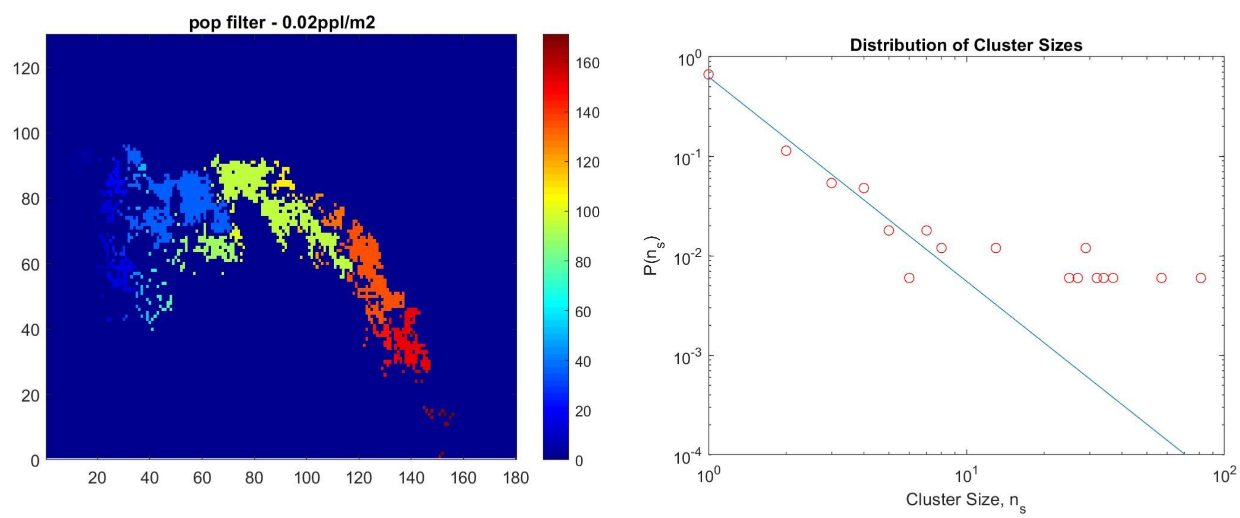

Figure 5: 5a - To the left, labeled percolation clusters with threshold = 0.02 people/m^2 . Each cluster is labelled with a different color on the plot. 5b - To the right, log-log plot showing distribution of cluster sizes.

This percolation analysis provides a basis for future analysis and decision making for the city of Freetown. The clusters denote regions of high population density and can inform policy makers and investors about regions of high need. Some of these decisions could involve locating regions that require more infrastructure investment such as schools, hospitals etc. per square meter because the region is more densely populated.

Population Density Distribution

As shown in Figure 6, the population density in Freetown is highest near the northern coast while the center of mass is closer to the interior surrounded by a ring of high density. Given the lack of data on trip flows, it is difficult to fit a model to assess which is a more likely characterization of Freetown.

Center of mass was calculated using the following formula:

$$C_{lat} = \sum_{i=1}^{n} lat_i \frac{pop_i}{pop_{total}} \qquad (5)$$

$$C_{long} = \sum_{i=1}^{n} long_i \frac{pop_i}{pop_{total}} \qquad (6)$$

Where the center of mass has coordinates (C_{lat}, C_{long}), lat_i and $long_i$ are the latitude and longitude of hexagon i respectively. pop_i is the population of hexagon i while pop_{total} is the total population for all hexagons.

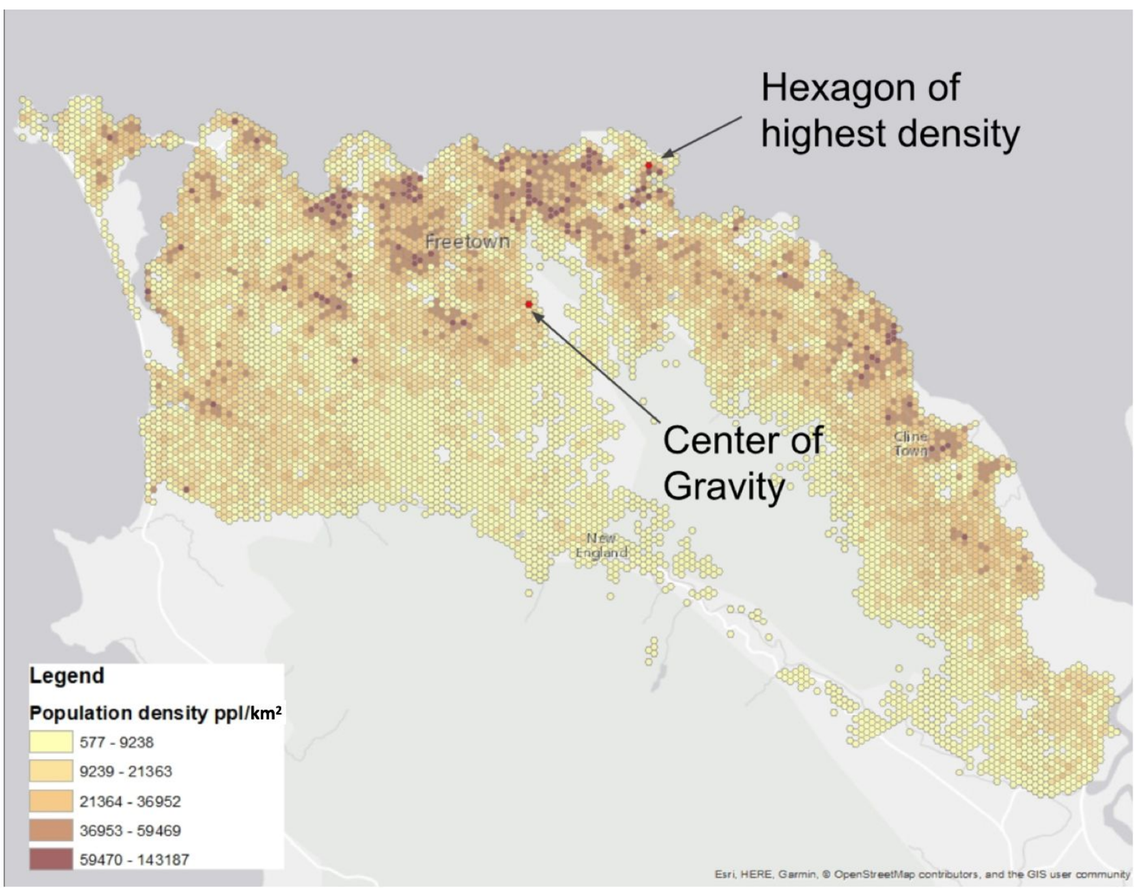

Figure 6: Population density in Freetown

The next category was evaluation the density gradient as shown in Figure 7. Of the city density gradients presented in Bertaud 2004, the closest fit to Freetown was New York City. New York City has topological constraints as does Freetown. New York City, like most American cities, is also quite polycentric which may indicate that Freetown is more polycentric, with population densities that do not decay regularly with distance from a single central CBD.

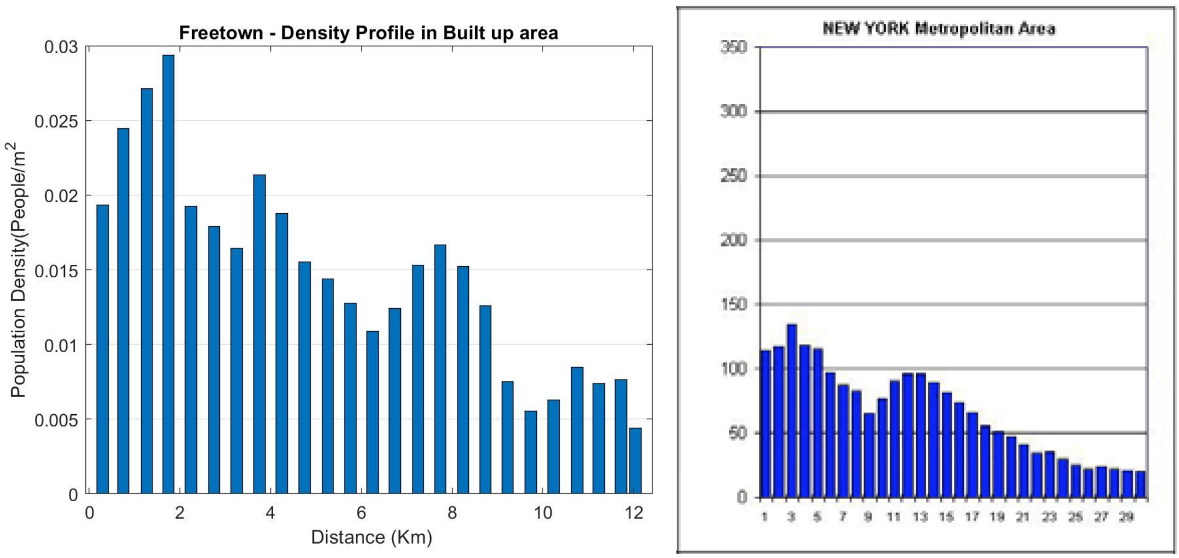

Figure 7: Comparison of Freetown and New York City density profiles

Trip Distribution

An extended radiation model estimates flows without parameters, using only population distribution as an input. By aggregating population within zones as trip generators, and aggregating POI as a proxy for trip attractors, flows between zones can be estimated.

Determining Zones of Analysis

The first step in aggregating population and POI within zones of analysis was determining what the zones would be. As mentioned in the Data section above, the nature of the zones of analysis was determined by likely future study to be performed with the WB. Using the location of the cell phone towers as a reference to create zones of analysis will allow for future comparison between the radiation model and aggregate individual trip flows extrapolated from cell phone data.

Using the position of the cell towers as a base, a Voronoi diagram was calculated using the SciPy and numpy libraries for Python. Within the borders of each polygon of the Voronoi diagram, all locations are closest to the cell tower within the region, rather than any other cell tower. Since these calculated polygons can extend to infinity, they must be clipped, in this case by the provided administrative border.

Rather than use the position of the tower as point to aggregate population and POI, the centroid of the tower was calculated. This served as a better proxy for trip origin and destination than tower location, since it is more representative as an average of actual user positions within the polygon. Figure 8 shows the Voronoi diagram with cell tower position in orange and centroid in blue.

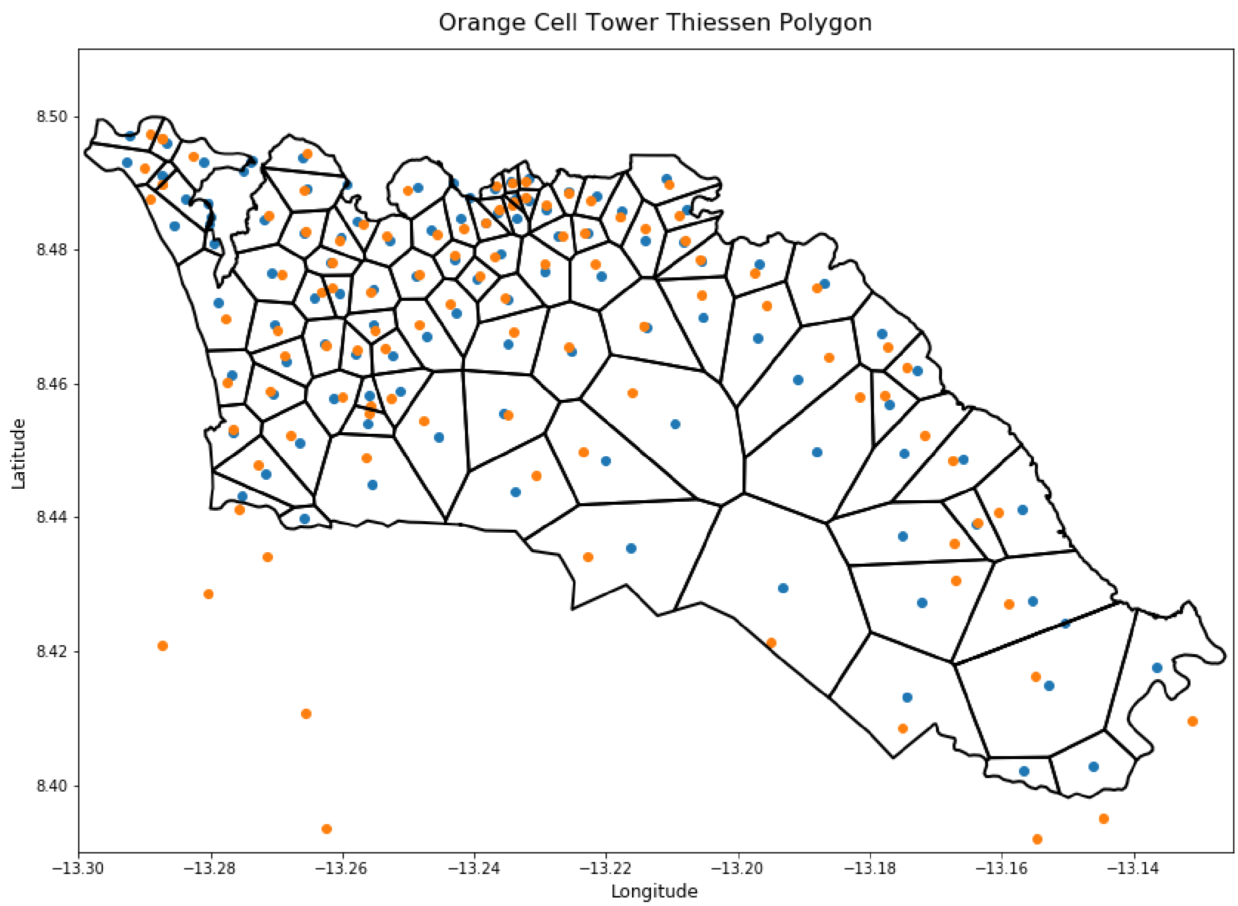

Figure 8: Voronoi diagram with cell phone tower positions (orange) and zone centroids (blue)

Aggregating Population and POI

The geopandas library of Python was used to represent zone polygons, population polygons and POIs as geolocated dataframes. Centroids of each hexagon were calculated so population could be assigned to single points in space in the same format as POI. Population centroids and POI were assigned to indexed zones based on spatial location and aggregated to each zone via spatial merge.

The zones geodataframe, which contained zone polygon geometry, centroid geometry, aggregated population, and aggregated POI, was further analyzed using MATLAB.

Generating Trip distribution Matrix between Voronoi cells based on POI and population data

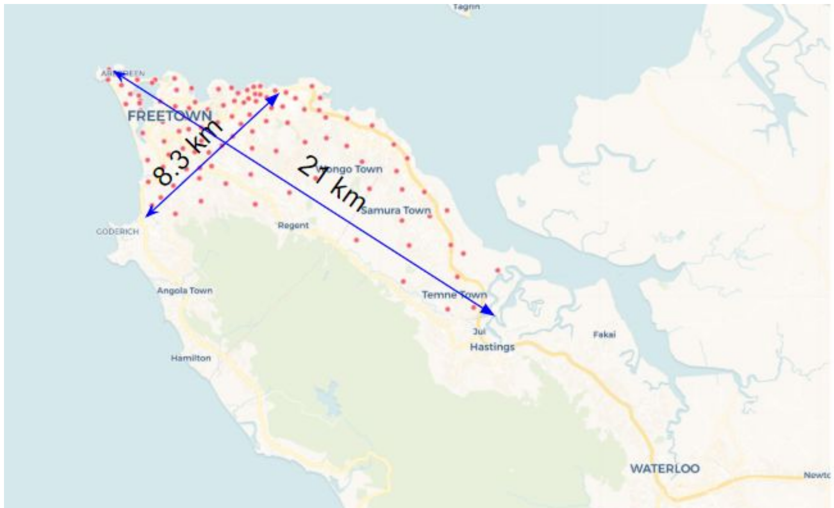

Figure 9: Zone centroids (red) and calculation parameters for l

Extended radiation model is used to generate the origin-destination trip depending on the population in Voronoi cell.. By using this model, trips between the zone centroids (see Figure 9) were generated in the area. There were two input parameters, α and norm_sum. For parameter α:

$$\alpha = (l/36[km])^{1.33} \quad (7)$$

Where:

l is the mean distance between neighboring Voronoi cell

As shown in Figure 9, the length of Freetown area is 21 km and width is 8.3 km. The small size of the area results in a relatively small l here, and the α value is small as well. For Freetown, $\alpha = 0.053$ was calculated. For norm_sum, we take norm_sum as 75% of the total population, found to be 1,052,625 after data processing.

Trip distance distribution

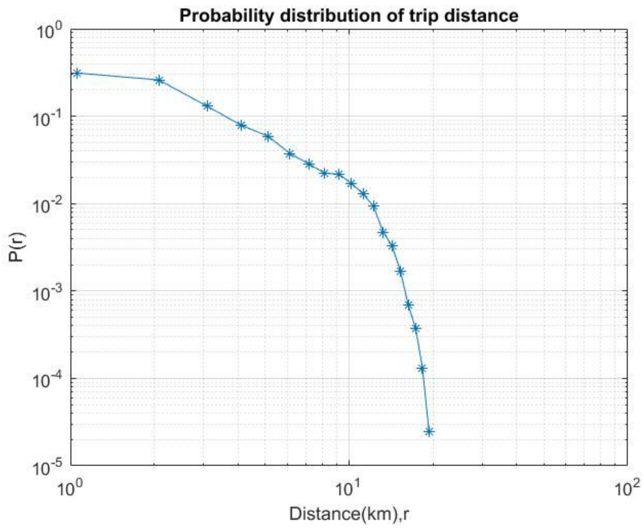

Figure 10: PDF of trip distance

As shown in Figure 10 above, the probability is the highest when the trip distance is within 3km, and will keep a constant value. As trip distance increases, the corresponding probability will decrease. After the trip distance exceeds 10 km, the probability of such trips will decrease dramatically.

Relationships between trips and given data (POI and Population)

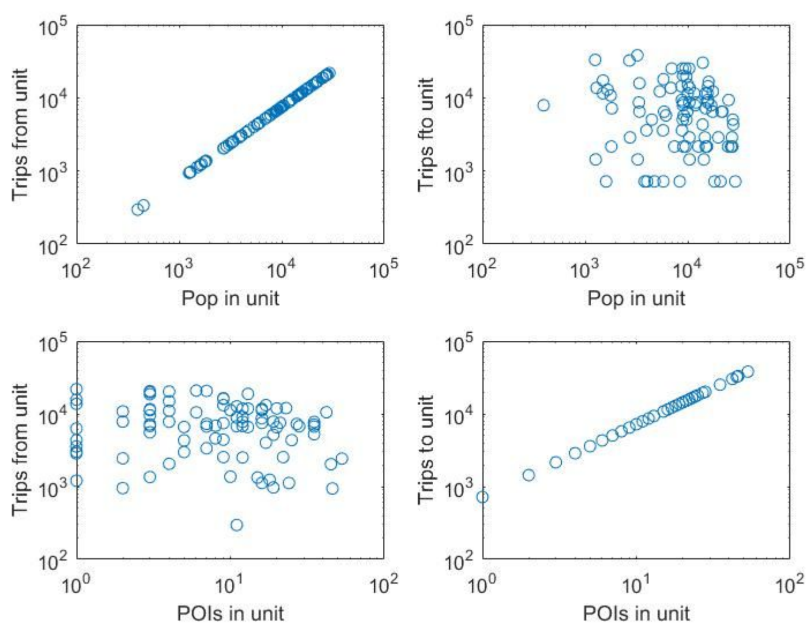

Figure 11: 11a - To the top left, trips from unit compared to population, 11b - To the top right, trips to unit compared to population, 11c - To the bottom left, trips from unit compared to POI, 11d - To the bottom right, trips to unit compared to POI

Figures 11a and 11b show the relationship between population and trips. With the increase of population, the number of trips generated from the area becomes larger. Figure 11a shows that there exists a log-linear relationship between the population and trips from the area. Figure 11d shows a similar log-linear also exists for POI and trips to the unit.

So, the population in the unit can be seen as a reflection of trip production. The more people in the area, the demand to take trips will be bigger and there will be more trip production. In addition, the POI in the area reflects the attraction of the place. Areas with more POI will absorb more trips.

Places receiving and generating the most trips

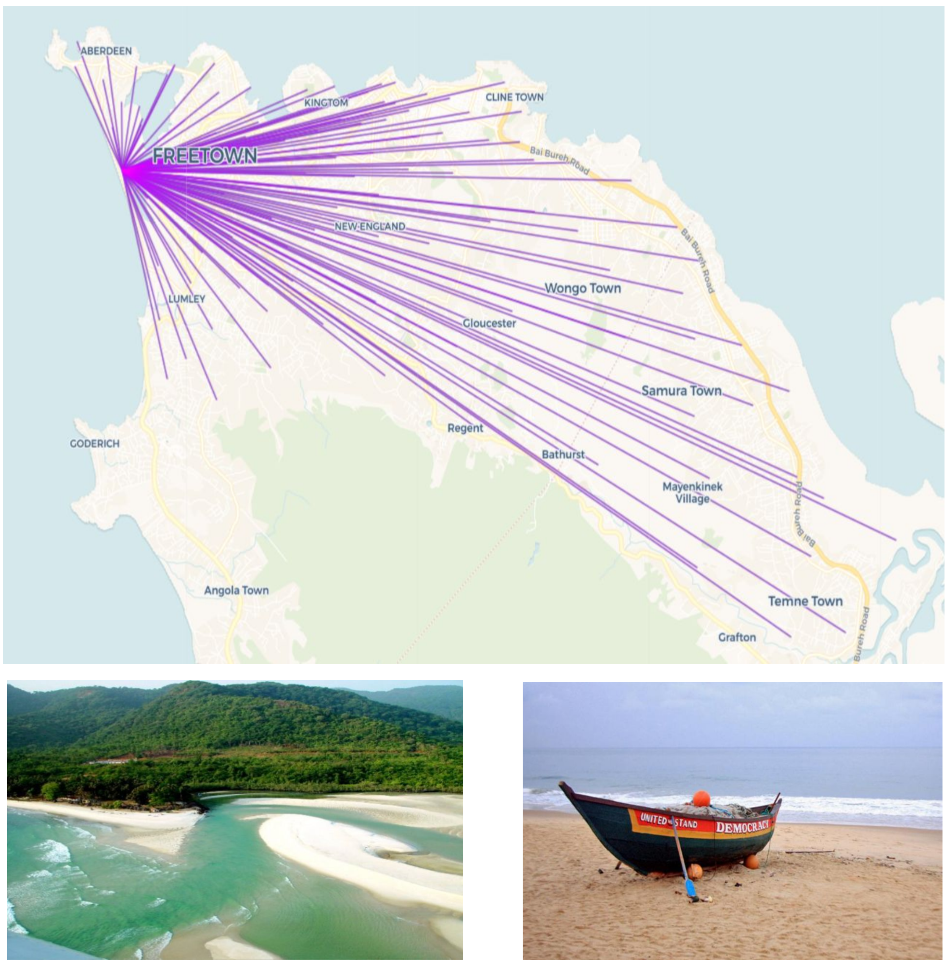

Figure 12: Visualization of place receiving the most trips - Lumeley Beach (-8.476458,13.283681) and points of origin for trips.

Lumeley Beach is one of the most famous attractions in Freetown. The reason why it will attract most trips is the dense distribution of tourist destinations like restaurants or clubs around, which results in the high POI around it.

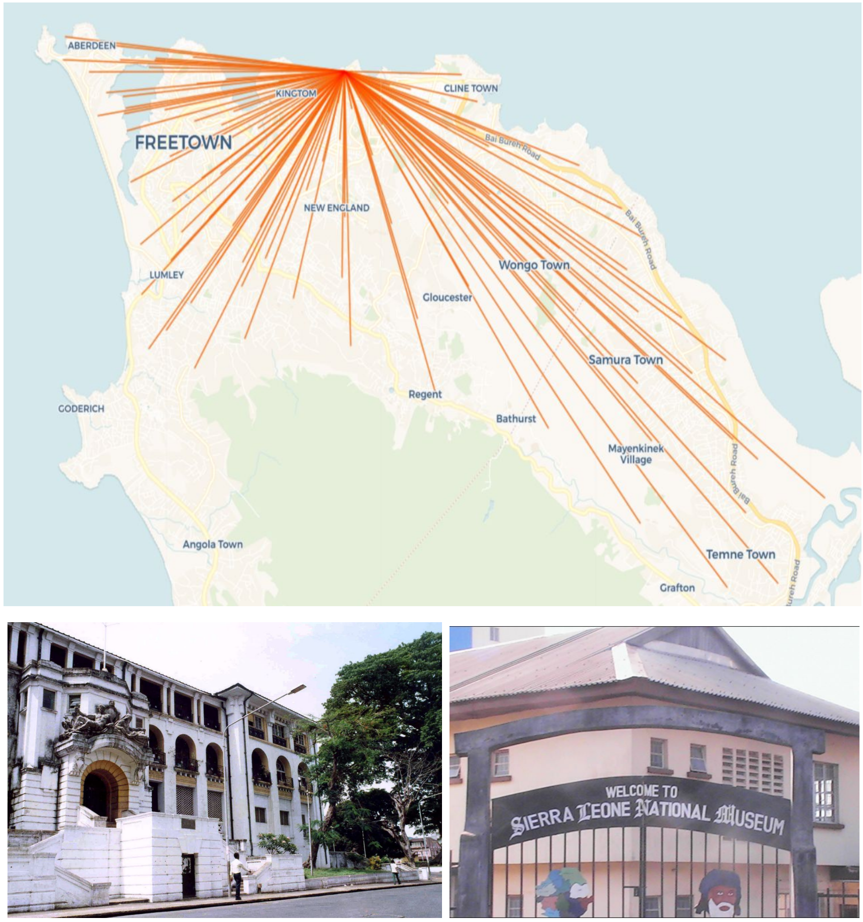

Figure 13: Visualization of place generating the most trips - Cotton Tree (-8.476458,13.283681) and points of destination for trips.

Cotton Tree is in the CBD of Freetown. It is near the Supreme Court of Freetown and the National Museum. As the population density is high there, it becomes the place where most trips depart from.

Network Analysis and Route Assignment

OSMnx was used to generate a simplified graph for the street network within the administrative boundary (see Figure 14). This graph was used to calculate nearest nodes to each centroid position and routes for OD pairs based on distance-weighted dijkstra shortest paths (see Figure 15). Routes are defined as a list of all the nodes that the shortest path passes through, as well as the origin and destination nodes.

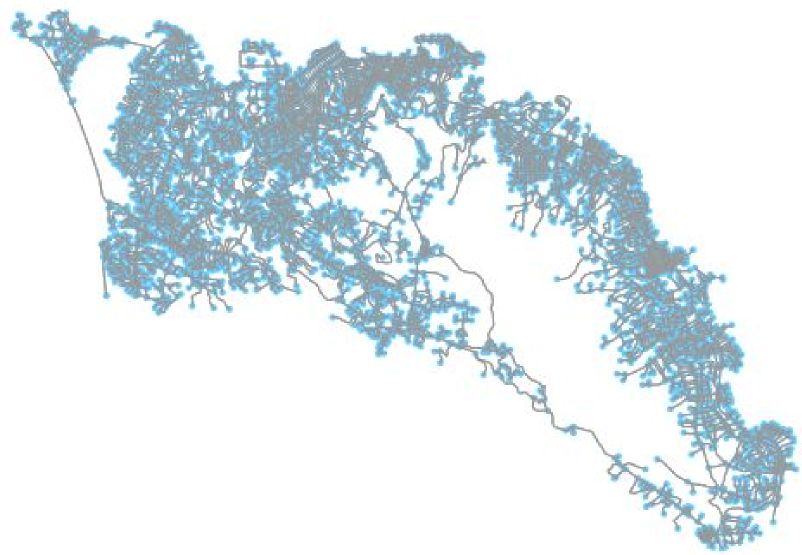

Figure 14: Visualization of OSMnx-generated graph with nodes (in blue) and edges (in grey)

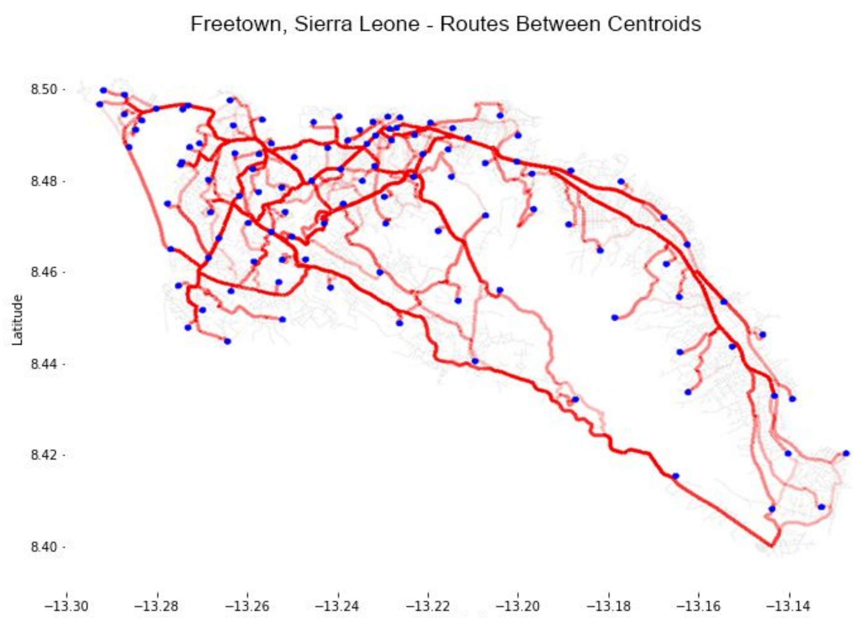

Figure 15: Routes generated by dijkstra distance-weighted shortest path between all OD pairs

The route list was assigned a demand weight based on whether the first and last nodes in the route matched the origin and destination nodes listed in the extended radiation model's origin-destination matrix. Then, each pair of nodes in each path in the path list was located in the edges list of the OSMnx graph and the demand was added as a weight (see Figure 16).

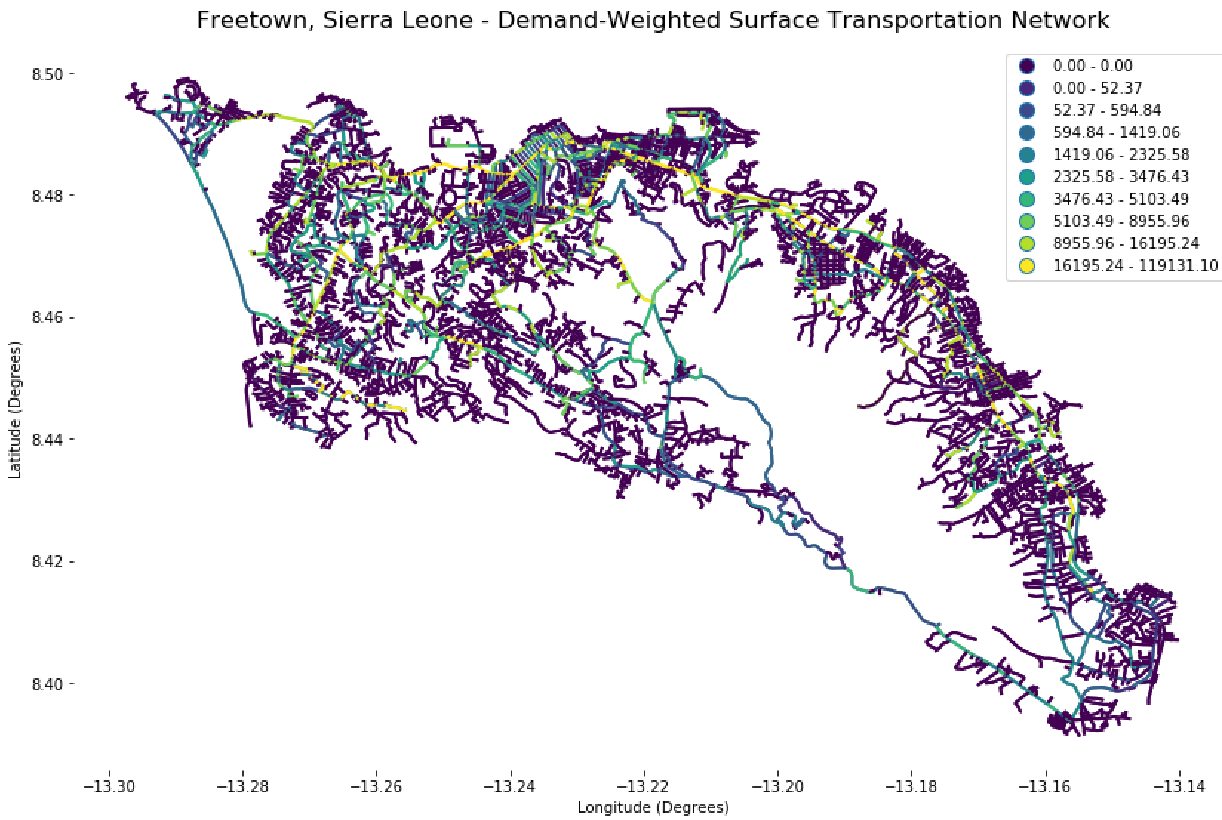

Figure 16: Surface transportation facilities weighted by demand

Edge BC was additionally calculated for the graph using the networkx library of Python (see Figure 17). Similar to BC, edge BC quantifies the number of times an edge lies on all shortest paths in the graph from every possible pair of origin and destination points. Its calculation is given by equation:

$$BC(e_{u,v}) = \sum_{i \neq j} \frac{\sigma_{i,j}(e_{u,v})}{\sigma_{i,j}} \quad (6)$$

Where:
e is any edge lying between nodes u and v,
$\sigma_{i,j}$ is the total number of shortest paths between unordered node pairs i and j, and
$\sigma_{i,j}(e_{u,v})$ is the number of those shortest paths which pass through e.

It thus represents the percent of total shortest paths that pass through any given edge, ranging from 0 to 1 in value. BC serves as a strong measure of how important each edge is for all origin-destination node pairs within the transportation network, especially for identifying critical transportation junctures which lie on many shortest paths and have low redundancy in nearby links. Edges with high BC represent bridge-like connectors between parts of the network.

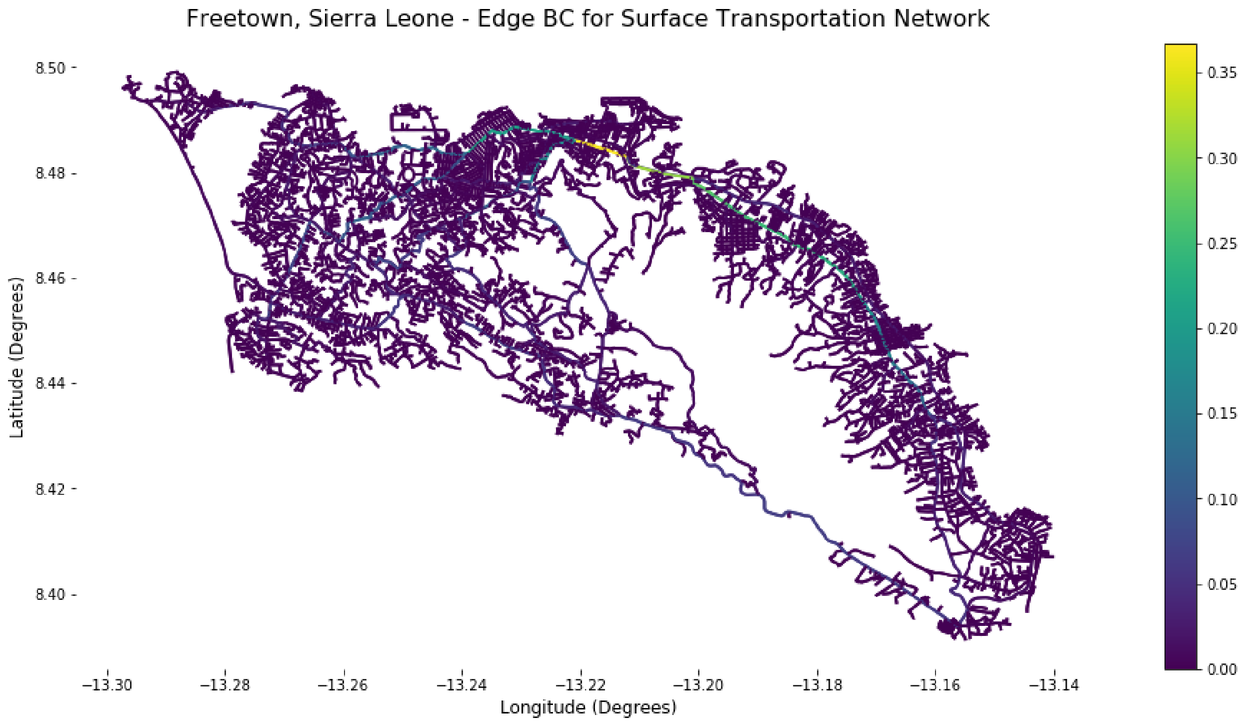

Figure 17: Surface transportation facilities by edge betweenness centrality

Reinforcing Guimera et al., the uneven distribution of both edge betweenness centrality and weighted degree was immediately apparent. Figure 18 shows scatter plot of both statistics. The 85th percentile of each component are highlighted in blue for betweenness centrality and yellow for weighted degree. This shows a widely varying spread of edges high in both statistics while remaining edges are heavily focused around low betweenness centrality and weight of 0. This is to be expected based upon the method with which demand was assigned, which neglects to add a demand penalty where traffic along high demand routes may be diverted due to resulting increases in travel time. Using cell phone data to calculate demand weights would likely add weight to many of these 0 edges.

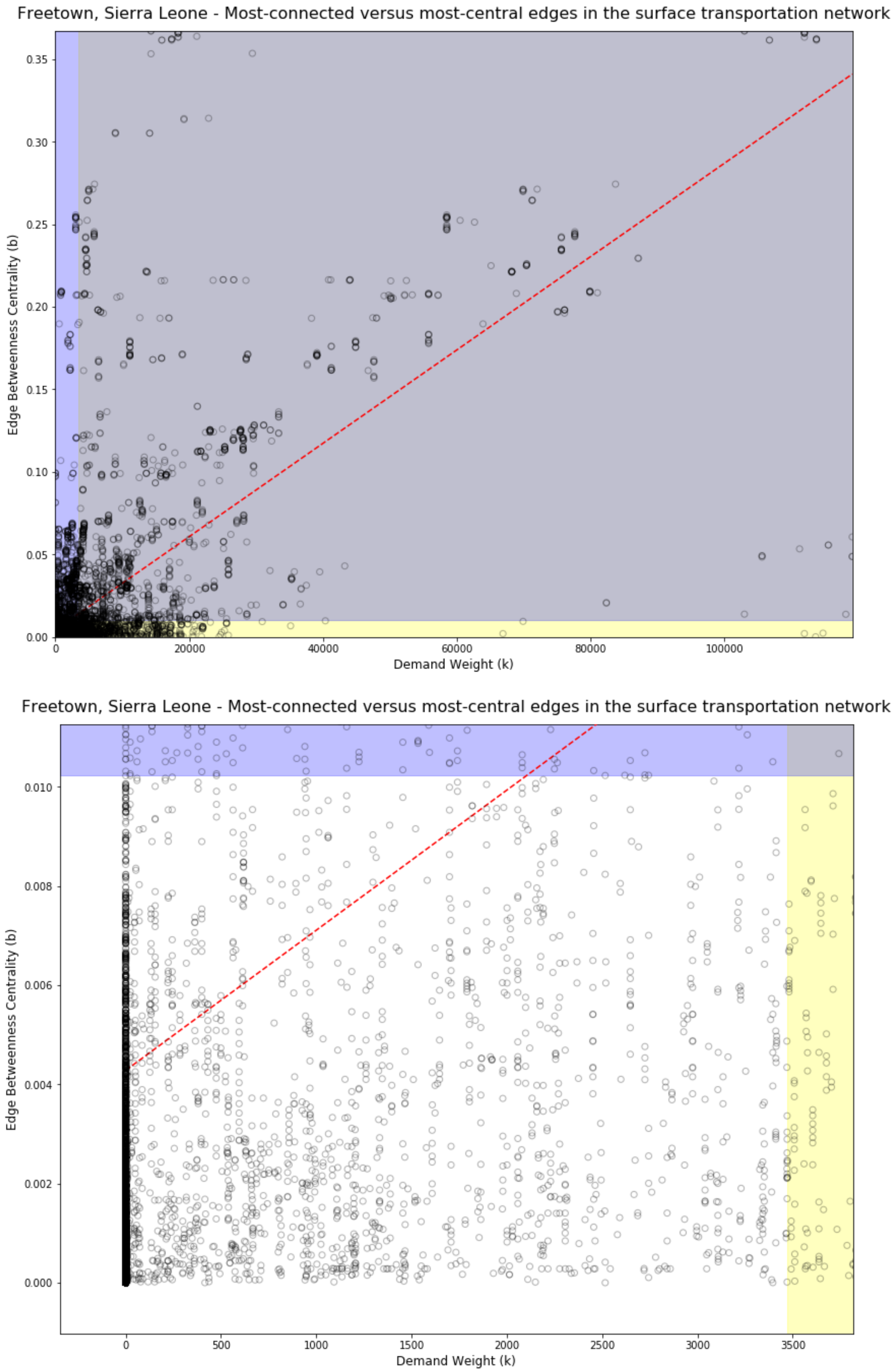

Figure 18: 18a - To the top, demand weight plotted against edge betweenness centrality, 18b - To the bottom, zoom focused on the lower 85% of data points.

When corrected to discount zero values, edges with low betweenness centrality and demand are still highly represented (see Figure 19). Distributions are well categorized by truncated power law distributions (see Figure 20).

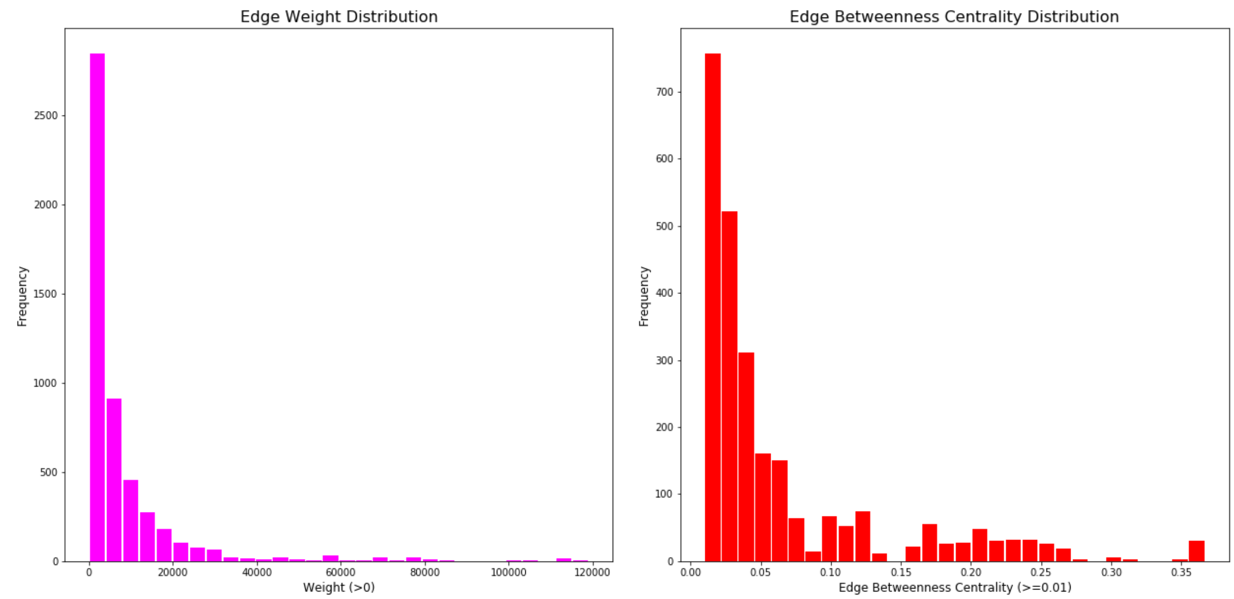

Figure 19: 19a - To the left, zero-corrected weight distribution of edges, 19b - To the right, zero-corrected edge betweenness centrality distribution of edges

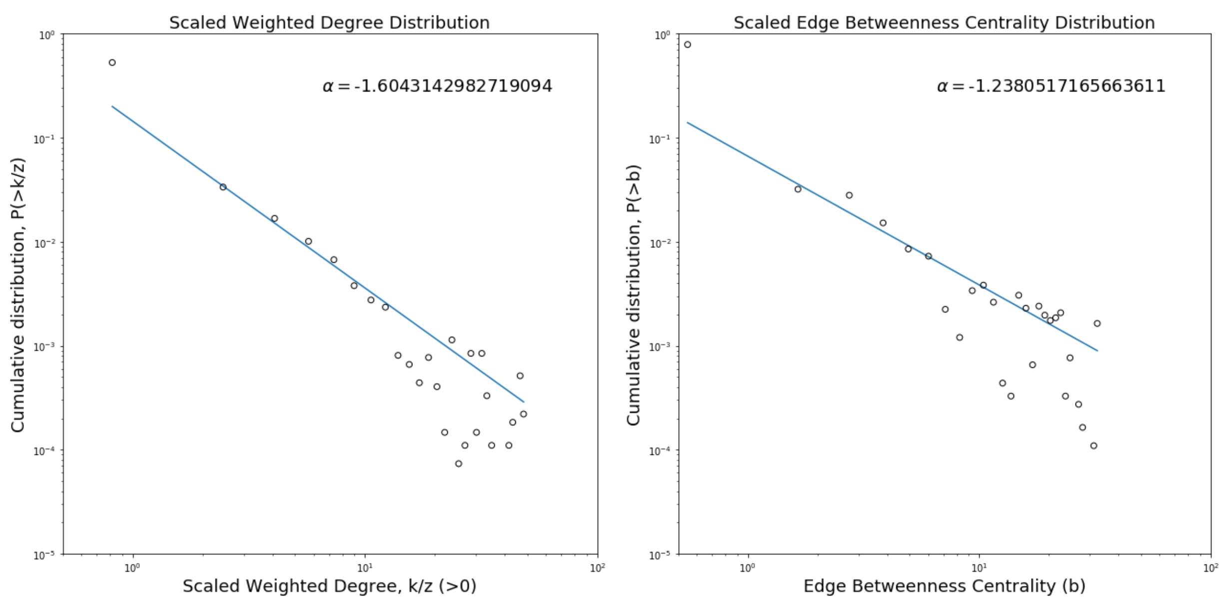

Figure 20: 20a - To the left, log log cumulative distribution of scaled edge weight with best fit power function, 20b - To the right, log log cumulative distribution of scaled edge betweenness centrality with best fit power function

Using the 85th percentile as a basis for categorization, roads can be assigned into four groups:

- Connectors (top 15% *b* and *k*): Both topologically important and diversely used by travellers.
- Peripheral connectors (top 15% *b* only): Topologically important, but less diversely used.
- Attractors (top 15% *k* only): Diversely used, but at the physical periphery of the network.
- Locals (low *b* and *k*): Not topologically important and only locally used.

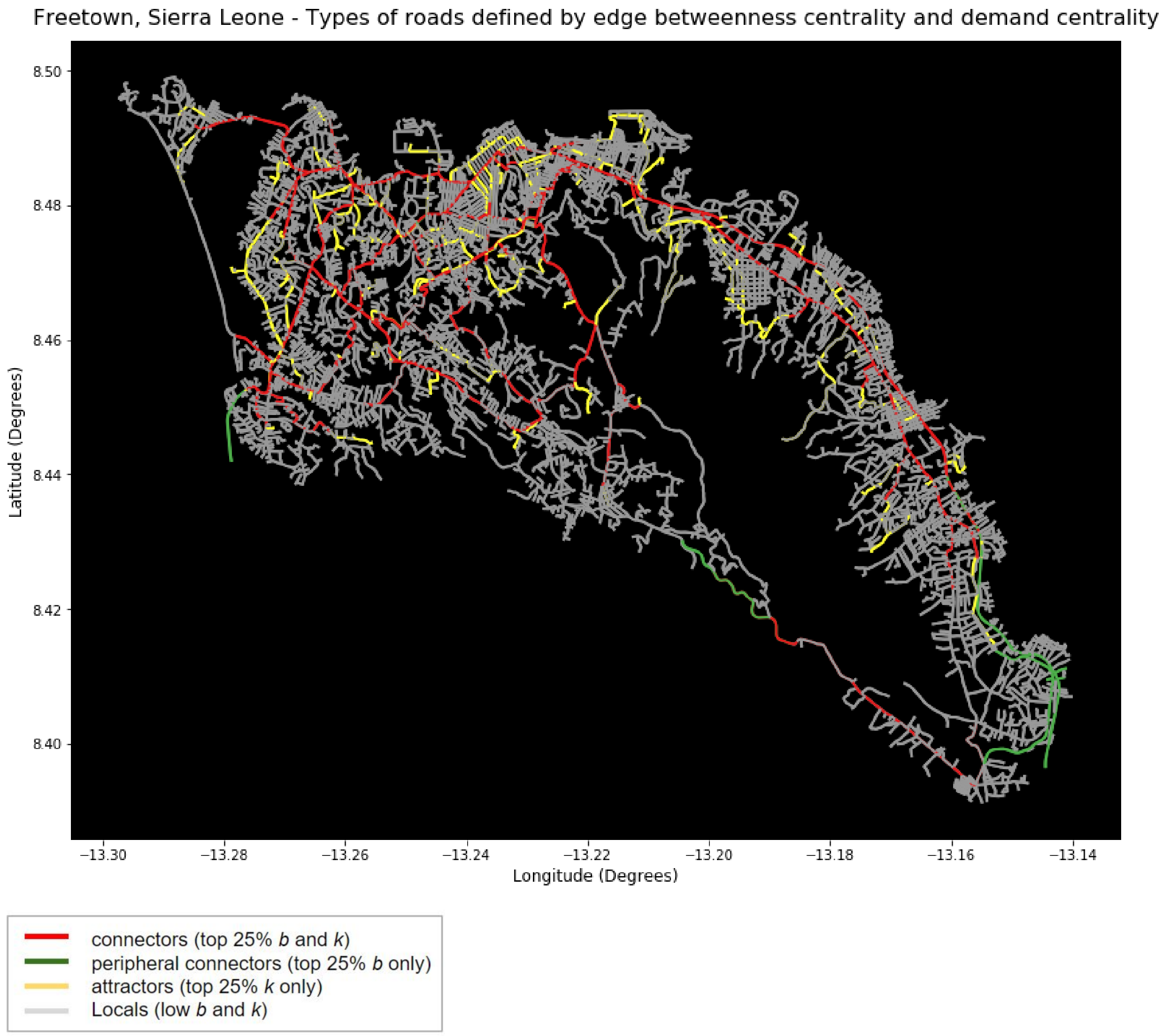

Figure 21: Categorization of surface transportation facilities by edge betweenness centrality and edge degree

The categories assigned to roads follow expectations. Connectors tend to describe linear pathways that link areas. Attractors appear in densely gridded downtown areas which contain the majority of POI, but are located at the outskirts of the network. Finally, peripheral connectors appear in the less used central valley roads and at the southeastern periphery.

Conclusion and Future Work

Relating Population Clusters to the Road Network

Earlier in the report we analysed the population cluster distribution for the city of Freetown and it is interesting to note that the population clusters form around regions that have high road network density. Our hypothesis for this match between road networks and population cluster is because Freetown is constrained by geography (mountains and the ocean). Normally, regions with high road density will be flat lands which will most likely also attract higher population density because the presence of roads provide access to jobs and opportunities.

Population Clusters and Percolation

The information from the percolation analysis allowed us to create a bare bones structure to characterise the city growth using population density. This analysis is an important preliminary step in order to better understand how the population is distributed spatially in Freetown. These results would serve as a basis for further research that involves providing resources to the population and understanding the market in Freetown.

It would be interesting to fit our model to historical data to observe whether or not our growth model captures the city spreading accurately. Freetown is very constrained by geography (mountains and ocean) and as such it is possible that the current model does not fully capture the city growth. Retrofitting of models to historical data could allow us to validate our results but also enable use to develop a more suitable city growth model for regions constrained by natural topography.

In addition, with some additional data, we could get information about the health and education of the population. The health of the population could be evaluated by carrying out a spatial analysis for the spreading of different non-communicable diseases such as diabetes, high blood pressure, obesity. Moreso, a similar analysis can be carried out for the spread to information via different outlets such as news, social media and word of mouth. This analysis will be inspired on the Collective behavior in the spatial spreading of obesity by Gallos et. al.

Validation of Trip Distribution Model, Network Analysis and Route Assignment

This study weights the network based upon an extended radiation model. While these trip weights are a good estimate to begin analyzing the network from a demand perspective, they also require validation. Using aggregated cell phone data from the World Bank to build actual origin-destination matrices would create more realistic demand weights, and provide data to validate the radiation model against using common part of commuters (CPC) indicators. CPC can be calculated to estimate the fitting degree between model and real data. In addition, different models such as gravity model and radiation model can also be used and comparison between different results can be made. In this way, appropriate model and its parameters can be found, which can reflect the real condition in a better way.

After creating a true or verifiable estimate of trip weights, a mode choice model would be applied to separate trips by mode. Care should be taken when applying a mode choice model to Freetown, particularly with respect to the low income population in informal settlements. As highlighted by Salon and Gulyani (2010) working in Nairobi, those living in slums do not have modal 'choice' in the same way and barriers to access should be taken into account when evaluating choice sets. For example, people may walk distances that another model may remove from the choice set determining it to be too far. Following the mode choice model, a route assignment method would be applied to distribute the trips along routes taking congestion into account.

The network could then be analyzed similarly to the original group study, which sought to recalculate centralities after eliminating nodes vulnerable to disaster events. For this study, routes could be recalculated with certain nodes eliminated from the graph, showing how demands would be redistributed throughout the network, rather than just routing. This would give organizations like the WB an idea of the type of demand expectations a network might face in the event of a variety of natural disaster scenarios.

Risk assessment of natural hazard in Freetown

Freetown is one of Africa's picturesque cities but it is also a place which suffers from many natural hazards. With a mountain range located in the middle of the city, the probability for hazards like landslides or mudflow is high. In addition, being one of the wettest places in Africa, flooding will be a usual case for Freetown in rainy seasons.

For the risk assessment, the main risk sources need to be determined at first. For Freetown, the risk source comes from flooding, landslide and mudflow. Then, for each risk source, the likelihood and consequence of it will need to be considered. After that, grading will be done according to the combination of the two parts. Areas with worse grading are the places where we should pay more attention to. The robustness of the road network can be tested by trying different failing roads risk assessments.

References

Bertaud, Alain. “The spatial organization of cities.” Deliberate Outcome or Unforeseen Consequence, Background Paper to World Development Report (2003).

Gallos, Lazaros K., et al. "Collective behavior in the spatial spreading of obesity." *Scientific reports* 2 (2012): 454.

Guimera, Roger, et al. “The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles.” Proceedings of the National Academy of Sciences 102.22 (2005): 7794-7799.

Makse, Hernán A., Shlomo Havlin, and H. Eugene Stanley. “Modelling urban growth patterns.” Nature 377.6550 (1995): 608.

Salon, Deborah and Sumila Gulyani. “Mobility, Poverty, and Gender: Travel ‘Choices’ of Slum Residents in Nairobi, Kenya.” Transport Reviews. Vol 30 Issue 5 (2010).