A Spatial Analysis of Cellular Tower Placement Along Cities and Highways to Determine Optimal Tower Placement Criteria Using Geographic Information Science (GIS)

Patrick McGregor
Department of Resource Analysis, Saint Mary's University of Minnesota, Minneapolis, MN 55404

Keywords: Geographic Information Systems (GIS), Telecom, Cellular Tower Placement, Python Programming, Spatial Analyst

Abstract

There is a constant increase in cell phone subscriptions and cellular data usage across the world. The data and signals are transmitted and received by a network of cellular towers utilizing antennas. To power this network, additional cell towers and antennas need to be added to meet the demand. Selecting the correct placement of these towers is crucial to providing the correct coverage and capacity to areas that need it while optimizing spacing not waste money and resources. Using Geographic Information Systems (GIS) and an accompanying Python computer program, existing cellular tower location data were analyzed based on proximity to nearby cities and roads. The results of the analysis were used to predict future tower placement with a weighted criteria system. This process can provide the foundation for cellular towers placement approximations as well as further cellular studies.

Introduction

Background

In the early 1980's the first generation of cellular mobile phones was created. Earlier mobile communications existed, but had very low capacities due to being able to hold a limited number of people on a frequency without interference. These new cellular systems divided large areas into smaller "Cells" that are able to use the same frequency repeatedly to increase the system capacity exponentially (Korhonen, 2003).

Today there are almost 7 billion cell phone subscriptions worldwide which is approximately 96% of the world's population (Sanou, 2014). This massive industry was estimated to be worth more than 750 billion dollars by the end of 2014 (Franco, 2011). The newest cellular technology, called Long Term Evolution (LTE) has a standard peak download and upload rate of 300 Mbit/s and 75 Mbit/s respectively. The next step in technology, called LTE Advanced or "True 4G" standards will require devices to have a standard peak download and upload rate of 1000 Mbit/s and 500 Mbit/s (Keely, 2012).

Because of this explosion of cellular usage and bandwidth needs, the radio frequency engineers in charge of the cellular networks must work with their marketing departments to determine where to place new cellular base stations. These new transmitting locations will expand the systems, and increase quality and capacity within them (Schmidt, 2013).

Purpose

McGregor, Patrick. 2016. A Spatial Analysis of Cellular Tower Placement Along Cities and Highways to Determine Optimal Tower Placement Criteria Using Geographic Information Science (GIS). Volume 19, Papers in Resource Analysis. 10 pp. Saint Mary's University of Minnesota University Central Services Press. Winona, MN. Retrieved (date) http://www.gis.smumn.edu

Radio frequency engineers analyze coverage maps to determine where the placement of new towers will accomplish 3 main goals: expansion, capacity, and quality (Schmidt, 2013). This study will focus on the first of those criteria, expansion. The study has three main goals.

The first goal is to use GIS to analyze existing tower data and its relationship to nearby cities and roads. While significant research has focused on the topic of cellular tower placement and cite these factors as important criteria for site selection, none that were publically available had actually studied the distances that the towers were away from cities and highways. The primary driver for cell tower locations is the service delivery needs of the wireless carriers. Simply put, they want to be sure they are investing in infrastructure where it is needed most. When considering a tower placement, they will evaluate population and demographic data, plus the profiles of nearby businesses, pedestrian traffic, and the proximity of roads and highways (Harris, 2011).

The second goal is to use data obtained from the analysis to determine weighted criteria for the creating search rings, which are geographical rings where a new tower location can be placed to have the coverage that is needed (Schmidt, 2013). The third goal was to automate this process as much as possible with python programming. GIS studies are often time consuming using several different tools building on top of one and other and take a large amount of time and large amounts of information. Automating a large portion of this process both saves time and makes increasing the size and scope of the study area much easier.

The program created as part of this project was able to input a spreadsheet of site data and derive weighted distributions maps showing cellular placement. In addition to saving time for the user this also makes the process easy to use for people who are not well versed in the GIS industry.

Methods

Four data sets were used for this study. The first two, a major city shapefile showing populations, and a major highway shapefile showing all Federal, State and County highways were obtained from the National Tower database. A tower location database was obtained from SBA Communications Corporation which is a tower owning and leasing company. Relevant data from the SBA database included tower locations (LAT/LONG). An attempt was made to find additional data on other cellular tower owners; however they were unwilling to release their tower location records. Statistical analysis and weighted overlay models were completed using Environmental Systems Research Institute (ESRI) ArcGIS system version 10.2 with the Spatial Analyst Extension. The study area boundary was projected into the NAD 1983 UTM Zone 15 coordinate system and clipped to cover only the study area.

Overview of Study Area

The areas of study for this work included the states of Tennessee, Mississippi, Alabama and Arkansas USA (Figure 1). The reason for choosing this area was twofold. First, the area is mostly flat land with no incredibly large geological barriers or anomalies that could skew the data for a large number of tower locations. Secondly, this area had the most obtainable data. Though there are thousands of towers across the United States, they are owned by multiple tower companies, many of which would not release data for this studies purpose. The states with the highest percentage of the main data sources towers (SBA communications) were used for the study. State borders were also a convenient cut off for tower locations since the cellular carriers are required to only transmit in licensed frequencies which have been auctioned to them by the FCC (Schmidt, 2013). These frequencies are generally different between states and cut off at the state borders.

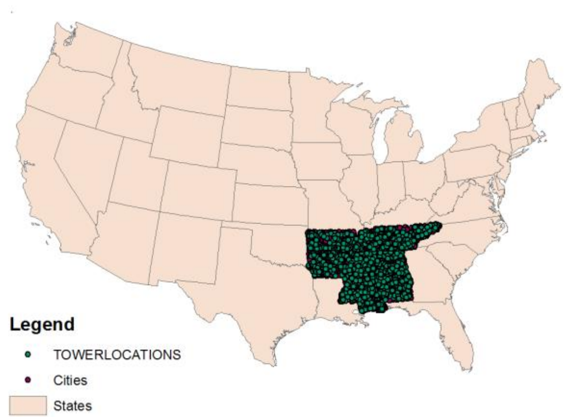

Figure 1. Map of the shapefiles showing the States of Tennessee, Mississippi, Alabama and Arkansas. Included is United States highway map, city map and tower locations map, cut to fit the study area.

Procedure

Initial Data Creation and Conformity

In order to analyze existing tower spatial data, all data sources had to be converted to shapefiles. This was followed by projecting all data to the same coordinate system so they could be analyzed with Esri software tools.

The initial data, obtained from the SBA website contained information for all of their towers across the world. It was parsed to show only the towers within the four states study area and formatted into a spreadsheet showing the site names, latitude and longitude as well as tower height.

A python program was created to read this spreadsheet, and create a shapefile mapping the points (Figure 2).

```
while dataRow:
    dataList=dataRow.split(',')
    readingSITENAME= dataList [0]
    readingLat= dataList [1]
    readingLon= dataList [2]
    readingTWRHEIGHT= dataList [3]
    pnt.X = readingLon
    pnt.Y = readingLat
```

Figure 2. The code above reads a CSV file created by the user and creates a shapefile of the tower information in the shapefile.

The City and Roads shapefiles were also parsed to show only the study states (Figure 3). Esri software was then used to project each shapefile into the UTM NAD 1983 Zone 15 coordinate system to be able to measure the results in meters instead of decimal degrees.

Figure 3. The three shapefiles in the study area were combined into a single map for analysis.

Near Analysis and Statistical Analysis

Once all of the shapefiles were in the same coordinate system, a ‘near analysis’ was conducted to determine the proximity of the existing towers to the closest city/road/ and other adjacent towers. The python program ran a ‘near analysis’ of the tower location shapefile, creating three new near analysis shapefiles showing the distance of each tower to the nearest city, major road and closest other tower respectively (Figures 4, 5, and 6).

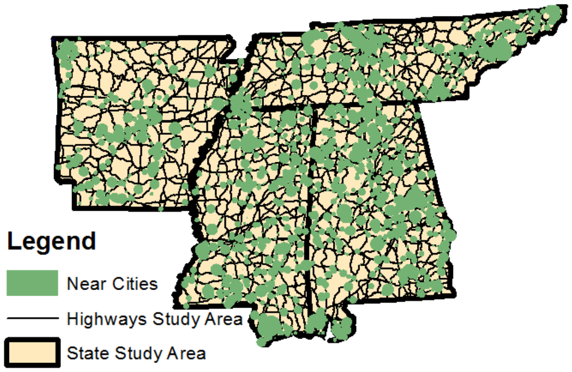

Figure 4. A near analysis of the existing tower locations to cities. Each green circle radius is the distance from a tower to the nearest city.

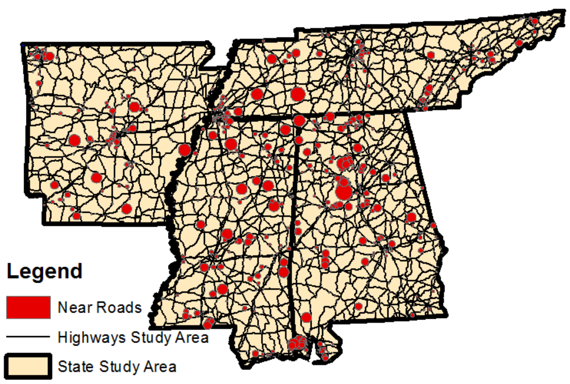

Figure 5. A near analysis of the existing tower locations to Highways. Each red circle radius is the distance from a tower to the nearest highway.

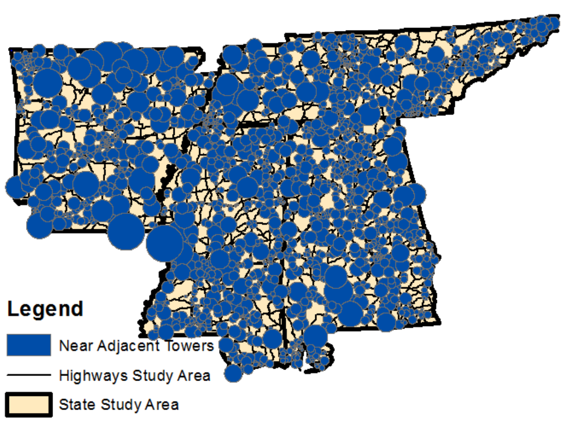

Figure 6. A near analysis of the existing tower locations to cities. Each blue circle radius is the distance from a tower to the nearest adjacent tower.

Using the attribute table of each shapefile, trends in the near distances can be seen. Each data set was analyzed and separated into one of four major groups; Tower transmitting to a city, Tower transmitting to a road, Tower transmitting to both a road and a city and other.

Each data set was statistically analyzed using an interquartile analysis to review the statistics and look for outliers. Strong outliers were identified and removed from their corresponding datasets. This method of outlier removal was chosen due to the multivariable nature of the study so not to remove outliers of one of the variables from all the data sets. Once the outliers were removed a new mean was obtained for each data set.

With the optimal distances to roads, cities and adjacent sites determined, a weighted analysis map showing predictions of optimal site locations could be created.

Buffer and Conversion to Rasters

Using the buffer analysis function the python program created several new buffered shapefiles based on the distances obtained from the analysis above.

A road shapefile was created was a multi-buffer around it correlating with the statistical analysis. The analysis was triple tiered with the closest buffer ending at the first quartile. The second buffer spanned from the first quartile to the third quartile. The third buffer spanned from the third quartile to the 90^{th} percentile of the previous analysis (Figure 7).

Another buffered shapefile was created for the cities. Once again, a three tiered system was used the closest buffer ending at the first quartile. The second buffer spanned from the first quartile to the third quartile. The third buffer spanned from the third quartile to the 90^{th} percentile of the previous analysis (Figure 8).

Finally, buffers were created around the Tower Locations shapefile. This distance was determined from the analysis to be the optimal distance for the tower to be

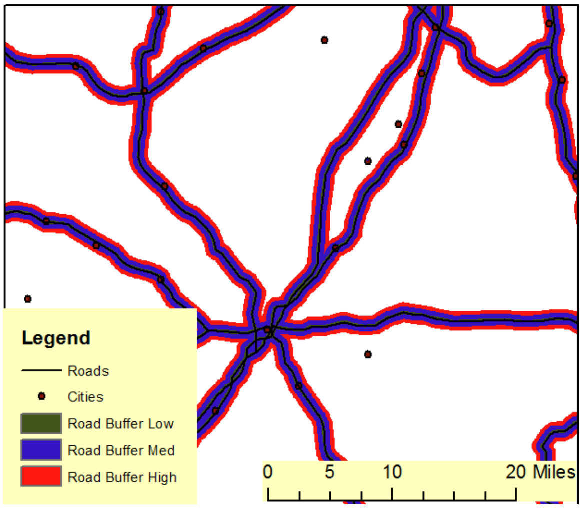

Figure 7. A sample of the road shapefile in the picture above illustrates the triple tiered buffer surrounding the roads for the analysis.

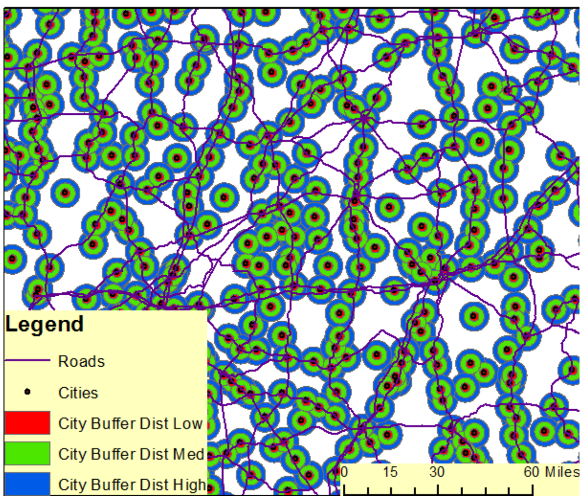

Figure 8. A sample of the city shapefile in the picture above illustrates the triple tiered buffer surrounding the cities used in the analysis.

away from other towers. The same three tiered system was used, however this time the nearest buffer represents where towers should not be located, with the smallest buffer being the areas least in need of coverage (Figure 9).

Once these buffer shapefiles were created, each needed to be turned into a raster for weighted analysis. A python program was used to convert the shape files into Rasters with a 100-meter cell size

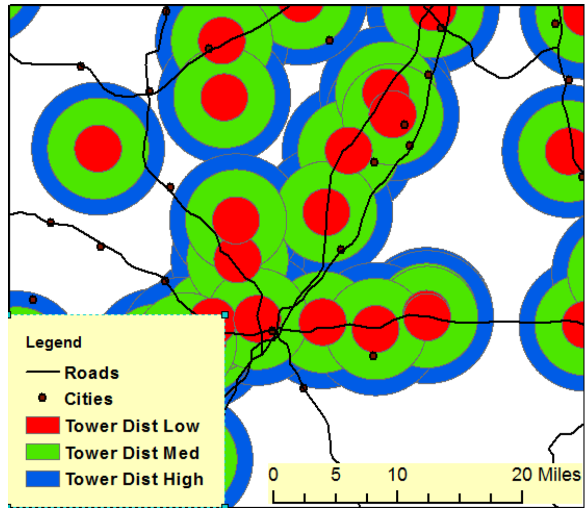

Figure 9. A sample of the existing Tower shapefile in the picture above shows the triple tiered buffer surrounding the existing towers used in the analysis. Note: unlike the city and road shapefiles, new towers should be located away from existing towers, so the buffers indicated are places where towers SHOULD NOT be placed.

Weighted Raster Analysis

With each buffer turned into a raster, a weighted raster analysis was conducted to show the locations where coverage was needed based on proximity to highways, cities and other towers. This was done with the weighted overlay table tool, using the results from the aforementioned analysis to create weights and values based on what logically were the most important for existing tower placement.

The purpose of the weighted table was to identify different areas of need for towers. The lowest numbers showed areas that already have tower coverage while the highest show areas that need coverage.

The weighted buffer uses a scale of 0-9 where 0 is the least desirable place for new tower coverage and 9 was the most desirable. Because existing literature on placement of towers did not indicate if any of the variables were more important than the others, they were equally weighted. The three variables (distance to road, distance from city, and distance from other towers) were given weights of 33%, 33%, and 34% respectively.

City buffers were weighted at 33% of the total weight. Since the city buffers were split up into three separate buffer files the total sum of the three weights added up to the sum of the weight put on the cities with the center (where the most population is located) having the most weight and expanding outward with less weight. The value of each city was based on which buffer raster was used. Nine (needing towers the most) locations were based close to the centers of the cities (where the population or coverage needed was the greatest and the lowest having a value of 7 (least needing towers). The areas with no data were given a value of 1 to signify that towers were not needed as no cities were in the area (Figure 10).

Raster	% Influence	Field	Scale Value
cityrastlow	11	VALUE	
		2390	9
		NODATA	1
cityrastmed	11	VALUE	
		5233	8
		NODATA	1
cityrasthigh	11	VALUE	
		7542	7
		NODATA	1
roadrast	33	VALUE	
siterastlow	11	VALUE	
siterastmed	11	VALUE	
siterasthigh	12	VALUE	

Figure 10. Chart of city weights and how they affect the total weighted raster.

The second criteria were proximity to roads. This was also given a weight at 33%, divided into three separate buffered layers. The rasters closest to the roads were given the most weight, with decreasing value going away from roads (Figure 11).

Raster	% Influence	Field	Scale Value
cityrastlow	11	VALUE	
cityrastmed	11	VALUE	
cityrasthigh	11	VALUE	
roadrast	33	VALUE	
		188	9
		780	8
		1277	7
		NODATA	1
siterastlow	11	VALUE	
siterastmed	11	VALUE	
siterasthigh	12	VALUE	

Figure 11. Chart of road weights and how they affect the total weighted raster.

Finally, the adjacent tower proximities were given a 34% total weight. The low and medium range city buffers were given a restricted value to indicate the area was already covered by cellular coverage and additional towers were not needed (and would override the city and road criteria showing that a tower may be needed in the area). The largest tower buffer was given a value of 1 indicating that a tower was most likely not necessary but could be overridden by road and city needs. The percentage weight on each tower buffer was also decreased as the buffer size expanded. The areas with NO DATA were given a value of 4 to take into account areas that have no existing tower coverage that may need it for expansion sake. The value of 4 was chosen so that areas that had no city or roads near them would end up with a value of 2, to show that there was no coverage, but not a significant population that required it Figure 12).

Raster	% Influence	Field	Scale Value
cityrastlow	11	VALUE	
cityrastmed	11	VALUE	
cityrasthigh	11	VALUE	
roadrast	33	VALUE	
siterastlow	11	VALUE	
		3059	Restricted
		NODATA	4
siterastmed	11	VALUE	
		6571	Restricted
		NODATA	4
siterasthigh	12	VALUE	
		8606	1
		NODATA	4

Figure 12. Chart of tower weights and how they affect the total weighted raster.

Results

The purpose of this study was to analyze existing tower location information in a four state region, to create a prediction model for future tower placement in a similar geographic region. A secondary purpose was to create a python program able to aid in the study, automating much of the data processing so additional information can be added easily to increase the accuracy of both the analysis and the prediction model.

Initial Analysis

After the near analysis was completed, data from the attribute table were statistically analyzed. Statistical analysis indicated a normal distribution of the data across all three near analysis studied. 92% of the existing cellular sites were within the 3rd quartile of either the near roads or cities analysis. The initial mean distances of towers to attributes were calculated as shown in Table 1.

Table 1. Distances (m) of towers to analysis attributes of roads, cities, and towers.

Data	Mean Distance (m)
Road	1283
City	5241
Tower	6723

The data was then analyzed to find outliers. Using the interquartile method, strong outliers were calculated (Table 2).

Table 2. Number of outliers for each analysis attribute.

Data	# Strong Outliers
Road	163
City	1
Tower	14

Outliers were removed from the data sets and the means were recalculated as shown in Table 3.

Table 3. Outliers removed and distances modified for for each analysis attribute.

Data	Mean Distance (m)
Road	780
City	5233
Tower	6571

Each outlier was only removed from the dataset that it was an outlier in. This was done to take into account the multivariable nature of the study. When evaluating the outliers, it was often easy to determine outliers were in a location to transmit only to one geographic group. Examples of this are sites that are near a long stretch of highway and no cities, or sites that are near a city but only county roads.

Final Coverage Map

Figures 13 and 14 illustrate the final result of the weighted raster analysis. Red areas indicate there is already coverage in the area, while lighter shades indicate

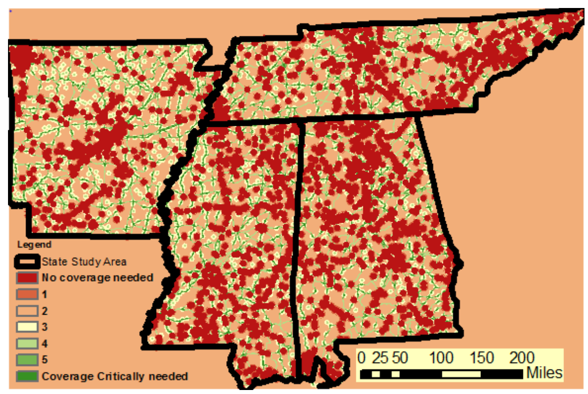

Figure 13. The three raster maps combined into a weighted overlay. The weighted raster correlates with the analysis showing bubbles around cities where coverage is needed and along the major highways. The intersecting red and pink areas in the city and roads indicate that there are already adjacent towers covering the area and that additional towers are most not likely needed.

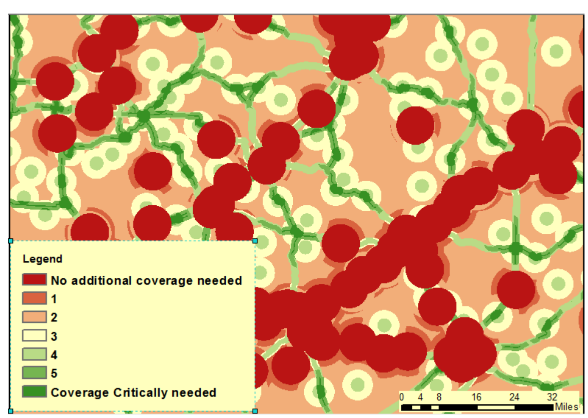

Figure 14. A closer scale snapshot of the final weighted raster map.

additional coverage may be needed. The green areas are where coverage is greatly needed and indicates the intersection of a city and a major road with no adjacent cellular coverage.

Python Program

The final Python code requires two user inputs. First, a CSV file of the tower locational data and the names of the city and road shapefiles to check against are needed. Then, based on results of the analysis, the user is then required to input buffer distances for the final weighted analysis. Once these are entered, the program completes the near and buffer analysis and creates a final weighted raster.

Discussion

Interpretation of Results

The purpose of this study was to develop a methodology to determine optimal cellular tower locations based on existing location information. The findings of this study revealed that the three main criteria studied are all important for tower placement. The statistical analysis indicated a normal distribution of the data across all three near analysis studied. This makes a lot of sense given most customers reside in cities, and use major highways, and correlates with the literature on the subject.

After comparing the three variable studied, optimal distances from each feature could be determined. The number of towers that were not near cities and roads were not insignificant and represent total network coverage areas as well as use for non-highway major roads that are not represented in the study. The final weighted raster map gave an accurate representation of where new towers could be placed to have the maximum effect. This study is a great starting point for expanding work on optimal search ring and cellular tower placement.

Sources of Error

There are several sources of error that could be associated with this project as it is only a preliminary study in the methodology of tower placement analysis. One of the biggest sources of error are tower locations owned by other tower owners. This source of error was mitigated by choosing states that had the highest density of towers owned by SBA with the lowest populations, indicating that SBA was the primary tower owner in the area. Another source of error was not considering factors outside of the 3 main criteria tested (distance to nearest, road, city and adjacent tower). Local zoning ordinances, environmental concerns as well as availability to electrical power are also major factors (Harris, 2011). While these factors may skew the results a small amount, the study was done across multiple states with a large amount of data so the effect should be minimal since the main priority for site selection is reaching customers (Harris, 2011). Another source of error was the areas that were covered by multiple cellular towers. The goal of this study was to look at tower coverage and expansion, not capacity. For very large cities the weighted analysis only takes into account single towers coverage. This error should be small considering that only 6% of the cities studied had a population over 10,000, and only 2.9% had a population over 49,000. Still the results of this study should not be used for cities over 100,000 residents where several towers are needed to cover the same areas.

Recommendations for Future Work

There are a number of ways this study/analysis can be expanded. With the methodology provided and the python program created, it would be easy to modify this project with different data or locations.

Initial Tower Analysis

Additional tower data could be added to the original analysis for a more accurate distance analysis. As stated in the sources of error section, not all towers in the area were able to be analyzed so any additional data would be useful in helping to provide a more accurate initial analysis.

Cellular Carrier Analysis

Another angle for this study would be to change the study to a specific carrier (AT&T, Verizon, Sprint, etc) analysis. This would show the areas the carrier supplied signal to and could get as specific as showing the direction each antenna on each tower was pointing to show exactly how many antennas are servicing each area. The difficulty with a study such as this is that the coverage areas are proprietary information that is kept secret to those outside of the companies.

Additional Criteria for Analysis

Some additional criteria could be included for analysis such as adding zoning districts or a geographic map to the study to account for locations that towers cannot be placed.

More Accurate Weighted Criteria

It was assumed equal weights on all three variables. Cellular carriers would obviously weigh these differently depending on the area, so a more accurate distribution of weights would potentially yield a more accurate analysis.

Conclusions

This project provided a methodology that was used to analyze existing tower data and predict future tower needs. The results provide an easy to read map showing where cellular towers are needed the most to increase capacity and expand the network to provide coverage to more areas. The study has the potential to assist cellular companies in placing their towers based on proximity to customers. The analysis method in this paper as well as the final python program offers a cost effective, quick, and user friendly experience that can be used by someone not intimately familiar with GIS.

Acknowledgements

I would like to thank the Staff of the GIS department at Saint Mary's University of Minnesota. I would especially like to thank Dr. John Ebert for his constant help and encouragement through the program. Several things in my life became priority issues during my Grad school experience but talking to John made me want to get through it made me believe I could. I would also like to thank the company I work for Black and Veatch. They helped me with tuition as well as made me interested in the telecom industry which led to this topic. Finally, I would like to thank my parents who helped me out with countless hours of free childcare so I could concentrate on my studies, and my amazing girlfriend, Alyshia who kept me on track and cut through my excuses and doubts. I will be forever grateful to all of you for helping me through this journey.

References

Franco, V. 2011. Telecom Services Market to Reach $1.4 Trillion by 2014. Retrieved May 18, 2015, from http://www.

marketwired. com/press-release/telecom-services-market-to-reach-14-trillion-by-2014-1504230.htm.

Harris, M. 2011. How Towers Work. Retrieved May 18, 2015, from http://www.unisonsite.com/pdf/resource-center/How Towers Work.pdf.

Keely, M. 2012. LTE-Advanced: The Challenges and Opportunities of "True 4G. Retrieved May 18, 2015, from http://web.b.ebscohost.com.xxproxy.smumn.edu/ehost/pdfviewer/pdfviewer?sid=7fef6d6b-424b-4dde-a631-ddd840ebb8ae@sessionmgr110&vid=1&hid=123.

Korhonen, J. 2003. Introduction to 3G Mobile Communications 2nd Edition (2nd ed.). London: Artech House. 568 pp.

Sanou, B. 2014. The World in 2014 ICT Facts and Figures. Retrieved May 18, 2015, from http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2014-e.pdf.

Schmidt, K. 2013. CellTowerInfo.com- How Do Wireless Carriers Choose Where to Provide Coverage. Retrieved May 18, 2015, from http://www.celltowerinfo.com/CellTowerSiteSelection.htm.