E-scooter Trip Analysis and Urban Micro-mobility Construction in City of Chicago

Project Report for MLC Summer 2021

Maia Guo (yg2483@nyu.edu)
Xiaolin Li (xl3389@nyu.edu)
Zhuoqi Niu (zn2034@nyu.edu)

Abstract

E-Scooter, as a kind of micro-mobility, is being piloted. There were pilot programs in Chicago, New York, Kelowna, London, and other cities. We conduct network analysis and cluster analysis on Chicago's e-scooter Pilot Program data in 2019 and 2020, hoping to help E-scooters future development. Network analysis results show the northern region has higher trip volume, central areas have more cross-community trips, and all of the communities can be classified into 8 partitions. After doing the clustering analysis, we find out areas such as West Town have long-distance travel and areas such as Albany Park and Archer Heights with the feature of short-distance travel. Students and commuters generally use E-scooters to facilitate their class and commute instead of choosing to travel at night or use E-scooters as their primary travel tool during weekend entertainment. The team gives suggestions on micro-mobility control to facilitate Chicago people's daily travel in this report.

Introduction

Background

Electric scooters, or e-scooters, are emerging as an alternative mode of transportation in cities across the United States due to the promise that they will enhance mobility, replace short car and ride-hail journeys and bridge the 'last mile' to and from public transit. However, this new mobility option has also brought operational challenges to cities, including safety concerns, sidewalk clutter and impacts on people with disabilities. To evaluate whether e-scooters can provide a sustainable, safe and equitable method of transportation for residents and to analyze the performance of e-scooters in conjunction with riders' characteristics and behaviors, the City of Chicago has hosted a shared E-Scooter Pilot Program in 2019 and 2020.

Objectives

This project will explore the disclosed E-Scooter Pilot Program data to analyze the e-scooter usage in temporal and spatial angles, like the hourly and daily usage trends and trip routes. Then, we can intuitively understand when most people ride electric scooters and where they go most often. This will help the company's vehicle dispatching and government policy making in urban micro-mobility construction, like expansion of non-motor vehicle lanes and parking areas.

Literature Review

Micro-Mobility is a very promising mode of urban transportation, especially for its potential to reduce the use of private cars for short-distance travel. When people gradually realize the problems of traffic congestion, car exhaust emissions and air quality caused by private cars, Micro-Mobility has become more and more popular. Micro-Mobility usually includes shared bicycles, electric skateboards, etc. At the same time, the popularity of Micro-Mobility will also depend on the re-planning of roads and parking spots by urban transportation departments(Abduljabbar, Rusul L., et al, 2021).

There have been two e-scooter pilot program activities in Chicago. The first e-scooter pilot program started in June 2019, lasted for four months, and ended in October 2019. The second e-scooter pilot program started in August 2020 and ended in December. The report "2020 E_SCOOTER PILOT EVALUATION" pointed out that 90% of e-scooter riders said that e-scooter makes their travel more convenient. 60% of e-scooter riders said that shared e-scooter can meet their daily travel needs. At the same time, sidewalk riding and e-scooter parking bring inconvenience to non-rider's life. This will be an urgent problem for the future e-scooter pilot program(Center for Neighborhood Technology, 2020).

Problem Statement

In many cases, regular public transit does not make people's travel more convenient. For example, students need 15 minutes to reach the following class location on campus, but only 10 minutes between classes. Nearby E-scooters can help students to be on time. Therefore, the vehicle control such as Escooters scheduling and parking supervision are essential. By exploring the usage of E-scooter in time and space from 2019 to 2020, and through spatial visualization, network analysis and clustering analysis, we can identify which communities have strong connectivity and what are their common travel patterns. To help the coordination of the vehicle scheduling and policy making of parking spot construction, we will analyze and answer following problems:

1. How about the general trip patterns like counts, distance and duration in temporal and spatial scale?
2. Which communities have strong network connectivities and how about their influences in their partitions?
3. Which communities have strong temporal commonality and how about the features in each of the clusters?

Data and Methods

Data Description

The City of Chicago has conducted an E-scooter Share Pilot Project in 2019 and 2020. The 2019 Pilot was from June 15, 2019 to October 15, 2019 and 10 companies were issued to operate 250 e-scooters each within a specified area on the northwest and west sides of the city. The 2020 e-scooter pilot ran from August 12 to December 12 with three participating vendors: Bird, Lime and Spin. Each e-scooter company that participated in the 2020 pilot was permitted to deploy up to 3,333 scooters for a total citywide fleet of 10,000 devices.

In this project, we will use the two-years datasets which have 1,341,655 records in total. Each of the records has the information of trip ID, distance and duration, start and end time, start and end community and also the vendor. Below table shows the main columns in the dataset.

Main Columns	Trip ID	Trip Distance	Trip Duration	Start Time End Time	Start Community Centroid End Community Centroid	Vendor
Type	Plain Text	Number	Number	Date & Time	Point	Plain Text

Table 1 Data Description

Data Preprocessing

The datasets itself have lots of dirty data: 1) Nan values; 2) YMDHMSw.d 12-hour time format; 3) confusing data types.

	Trip ID	Start Time	End Time	Trip Distance	Trip Duration	Accuracy	Start Census Tract	End Census Tract	Start Community Area Number	End Community Area Number	Start Community Area Name
10291	0dd95046-1bc8-4464-ba39-7f9e92c668b3	08/06/2019 06:00:00 PM	08/06/2019 06:00:00 PM	1004	236	152.0	1.703120e+10	1.703120e+10	20.0	20.0	HERMOSA
10293	6819df6a-9623-4ae0-b1d3-ec617197d229	07/08/2019 04:00:00 PM	07/08/2019 05:00:00 PM	9124	3182	10.0	1.703115e+10	1.703119e+10	15.0	19.0	PORTAGE PARK

Fig : Raw Data Example

Firstly, we concatenated the 2019 dataset and 2020 dataset. The next step is to handle the missing in each row, we drop all rows with Nan values in columns ‘Start Community Area Name’ and ‘End Community Area Name’. As the date time format, we use ‘`pd.to_datetime`’ to convert it to the regular 24-hour datetime format. We want to analyze the dataset in terms of time and spatial, so we extracted Year, Month, Day, Days of Week and Hours of Day from the time data and created 5 new columns.

year	month	day	dow	hour
2019	8	6	1	18
2019	7	8	0	16

Fig : New Columns

As the data type, we convert all "Community Area Number" and "area_num_1" in Chicago geometry dataset to integer to make sure these two dataset can merge successfully.

Exploratory Data Analysis

In order to better understand the characteristics of the data, we analyze the data according to three grouping methods: Days of week, Hours of Day, and Community Area number. The exploratory data analysis graphs are shown in Appendix A.

First, we visualize the data by grouping the data according to Days of a week and Hours of Day (Fig A.1, Fig A.2), using line charts and histograms. As we can see from the figure grouped by Days of a week, E-scooter is used the highest number of times on Saturday, and the trip duration is also the highest. The value of Trip Distance was the highest on Wednesday. It can be seen from the graph using the Hours of day grouping that from 3:00 to 4:00 in the morning, the trip_counts, trip distance and duration are all approaching zero. However, after 3 PM, the trip_counts, trip distance and duration reached the peak.

Then we grouped the data according to the community area number, and merged it with Chicago Geometry Information to plot the spatial distribution of Trip Count, Trip Distance and Trip Counts (Fig A.3). We respectively analyzed the data information of "Start Community Area Number" and "End Community Area Number". From the figure, we can see that the usage of e-scooter varies greatly in different community areas.

According to the information given by the dataset, we believe that we can maximize the benefits of suppliers by adjusting the amount of e-scooter and the number of parking spots at different times and community areas, and at the same time, we can maximize the space utilization of the city through these adjustments. In order to further determine areas with strong connectivity and determine which communities have more robust connectivity to coordinate vehicle scheduling, we will further conduct network analysis and clustering analysis.

Network Analysis

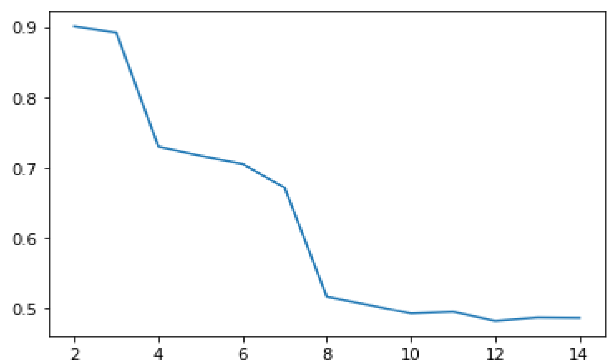
We first use Networkx to visualize the electronic scooter travel routes and the degree centrality of each community node. Degree centrality assigns an importance score based simply on the number of links held by each node, which can tell us how popular a network node is. So, we can learn the general trip distributions and the community scooter usage popularity.

Then, we use Combo partition tools, whose underlying is a stacking classification method, to cluster the community nodes based on the travel routes and get bigger areas with strong connectivities. The partition number is chosen according to the partition modularity score, higher of the score, better model performance.

Based on the partition results, we will further analyze some of the bigger areas which are communities with strong connectivities by visualizing the trip routes and computing Pagerank Centrality. PageRank assigns nodes a score based on their connections, and their connections' directions and can measure uncovered nodes whose influence extends beyond their direct connections into the wider network. So, we can get an intuition of which community has a strong influence within its partition.

K-means Analysis

First, we use silhouette score to find out the most appropriate k, which equals to 4.



The data indexed in the end community is classified by hour of day features, resulting in 4 clusters with more commonality in travel behavior respectively. Then, we performed the temporal analysis of the mean values of each cluster, to identify their travel peak hours and attributes of weekday and weekend volumes. And communities within each cluster can collectively coordinate the delivery and dispatch of e-scooters, the control of illegal parking behaviors and extension of the parking areas according to the commonality in travel time.

Results

Network Analysis

The network result map figures are shown in Appendix B. In the general routes and degree centrality figure, the route line width is weighted by directed trip counts and the node size is weighted by degree centrality. As we can see, lines are more obvious in the north of the city and bigger community nodes are concentrated in the central region, which means the travel volume is greater in the northern region, but cross-community travel is more common in the central region.

Following Figures show the top 5 communities by degree centrality and pagerank centrality respectively of the overall city trips. Since degree centrality is scored by number of links and pagerank centrality also takes link direction and weight into account, communities with higher degree centrality have more cross-community trips, and communities with higher pagerank centrality have more influence over the whole network, not just those directly connected to them. So, e-scooter companies may mobilize vehicles more frequently in higher-degree-centrality communities and governments need to pay more attention to the parking and riding issues in higher-pagerank-centrality communities.

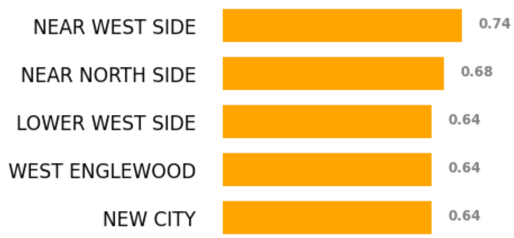


Figure 1 Top 5 Degree Centrality Communities

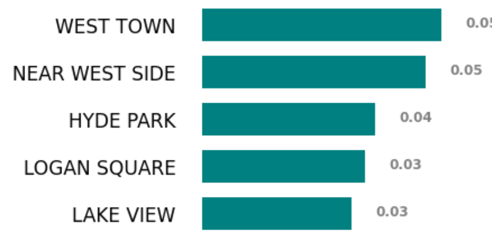


Figure 2 Top 5 Pagerank Centrality Communities

We did larger community detection by Combo partition, and the result is also shown in Appendix B. With the partition modularity scores in Figure 3, we chose the most ideal parameter, 8 partitions. The central and southern regions are very completely integrated into two larger communities, but the popular northern region are partitioned into 1 large area and 5 small areas. And the most of the community nodes are located in three of the partitions, respectively partition 0, partition 4 and partition 7. Horizontal bar plots in Figure 4 are the top 5 communities by pagerank centrality score in each of the three partitions. They are more influential in their whole partition and may be the key development areas for the construction of micro-mobility transportation facilities, like parking lots and non-motor vehicle lanes.

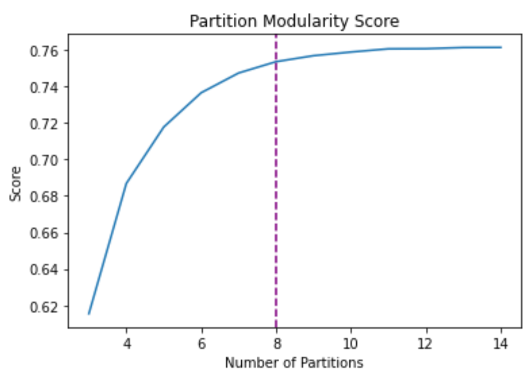


Figure 3 Modularity Score

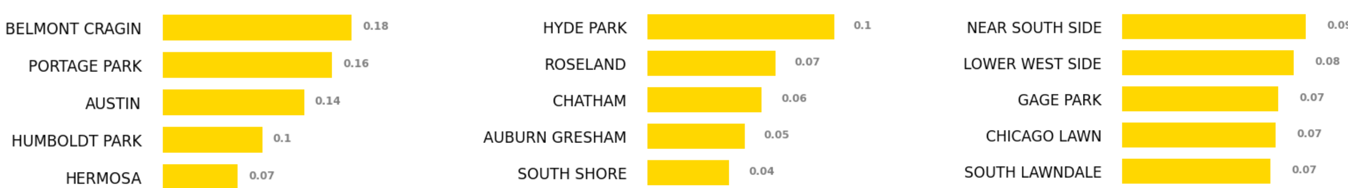
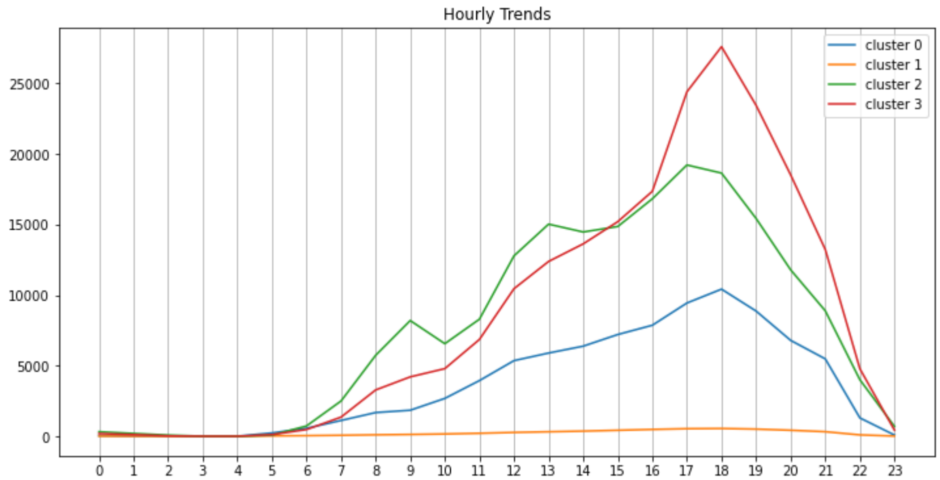


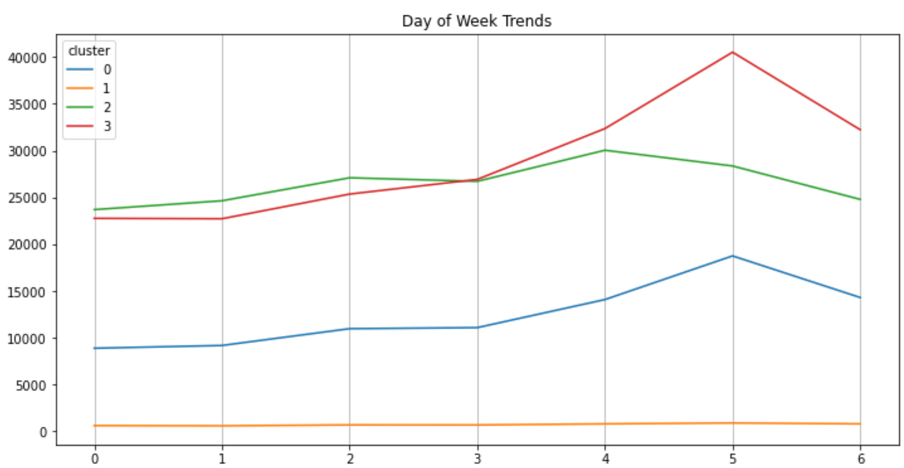
Figure 4 Top 5 Pagerank Centrality Communities of Partitions 0, 4, 7 (left, middle, right)

K-means

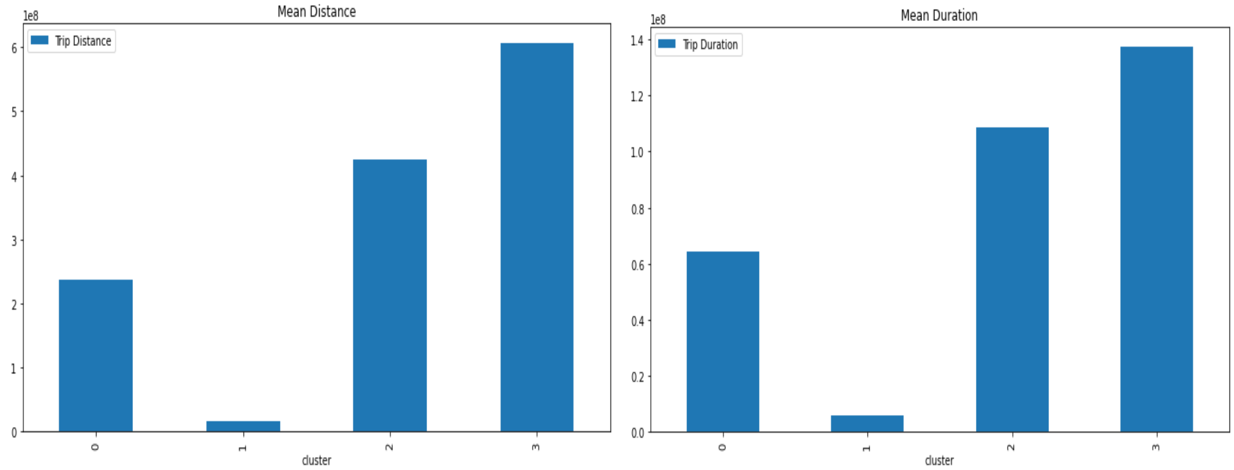


According to the plot, the team noticed that the usage of E-scooters of a different cluster is close to zero from 12 am to 5 am. The usage of e-scooter is gradually increasing from 6 in the morning. This result suggests that we can charge all E-scooters in the early morning because the usage rate continues to be low during this period.

The significant increase in usage is from 6 am to 9 am, which indicates that many commuters and students will use E-scooters as their travel mode; The team also noticed a significant drop after 6 pm which may be because people will not regard E-scooters as their travel mode for their nighttime activities.



From the overall perspective, the usage of E-scooters in clusters 2 and 3 is generally higher than that of clusters 0 and 1. According to the plot of clustering on day of the week, the team noticed an increasing trend presenting from Tuesday to Friday (dow =1,2,3,4) which the usage of E-scooters are higher than other days, and the usage on weekends shows a decreasing trend from Saturday to Monday(dow = 5,6,0) which is because many people do not consider E-scooters as tools for going out for entertainment. E-scooters are more likely to be a means of travel to work than a means of transportation for going out on weekends. However, for cluster 3, the West Town, the usage of E-scooters on Saturday is the highest among other weekdays. We will combine with the characteristics of travel of communities to talk about this point below.



The team utilized the “groupby” method of variable ‘End Community Area Name’ and took a sum to do the k-means clustering. Based on this plot, cluster 1 in both plots is extremely lower than the other three clusters, and cluster 3 contains areas with the highest travel distance and duration. Cluster1 has 72 communities, such as Albany Park and Armour Square, while Cluster 3 includes only one community called West Town; however, trip duration and distance in Cluster 3 are the highest among the other three clusters, reflecting the characteristics of long-distance travel in this community. The usage of E-scooters on Saturday in West town is higher than in other communities. Therefore, we may consider strengthening vehicle scheduling and parking supervision on Saturday in West Town.

Conclusions

Through Exploratory Data Analysis, we can see that the distribution of trip counts, distance, and duration are very different in temporal and spatial scale. From the temporal scale, the usage of e-scooter is high from 3 PM to 7 PM, and almost no one uses e-scooter from three to four o'clock in the morning. At the same time, judging from the distribution of e-scooter in the days of week, the usage from Monday to Thursday fluctuates little, but the usage increases starting from Friday. More people choose to use e-scooter to travel on Friday, Saturday, and Sunday. From the spatial scale, the distribution of e-scooter usage in each community area is also quite different. This is also related to the amount of e-scooter that the supplier puts in each region(Center for Neighborhood Technology, 2020).

Based on the network analysis, the central and southern regions have strong network connectivity respectively, but the northern region with higher trip volume was classified into 6 partitions. Policy making can be coordinated collectively in each partition. Community nodes with higher pagerank centrality are more influential to their network system and should be key areas when making e-scooter dispatch plans and micro-mobility construction policies.

Areas such as Albany Park and Armour Square have the commonality of “short-distance” travel while communities such as West Town and Near West Side have “long-distance” travel characteristics. As for clustering on hourly and weekly usage, students and commuters generally use E-scooters to facilitate their class and commute instead of choosing to travel at night or use E-scooters as their primary travel tool during weekend entertainment. Therefore, More E-scooter control should be done during rush hours. Our advice is to strengthen the parking supervision and E-scooter scheduling during rush hours in West Town on Saturday and in the Near West Side from Monday to Friday.

Team Roles and Contributions

Maia Guo: Network analysis

Xiaolin Li: Data Exploratory Analysis

Zhuoqi Niu: K-means Clustering

References

Abduljabbar, Rusul L., et al. “The Role of Micro-Mobility in Shaping Sustainable Cities: A Systematic Literature Review.” *Transportation Research Part D: Transport and Environment*, Pergamon, 9 Feb. 2021, www.sciencedirect.com/science/article/pii/S1361920921000389.

E-scooter pilot evaluation. Center for Neighborhood Technology. (2020, January 30). https://www.cnt.org/publications/e-scooter-pilot-evaluation.

Appendix A: Exploratory Data Analysis

FigA.1 :Trip Counts, Trip Distance, Trip Duration by DOW

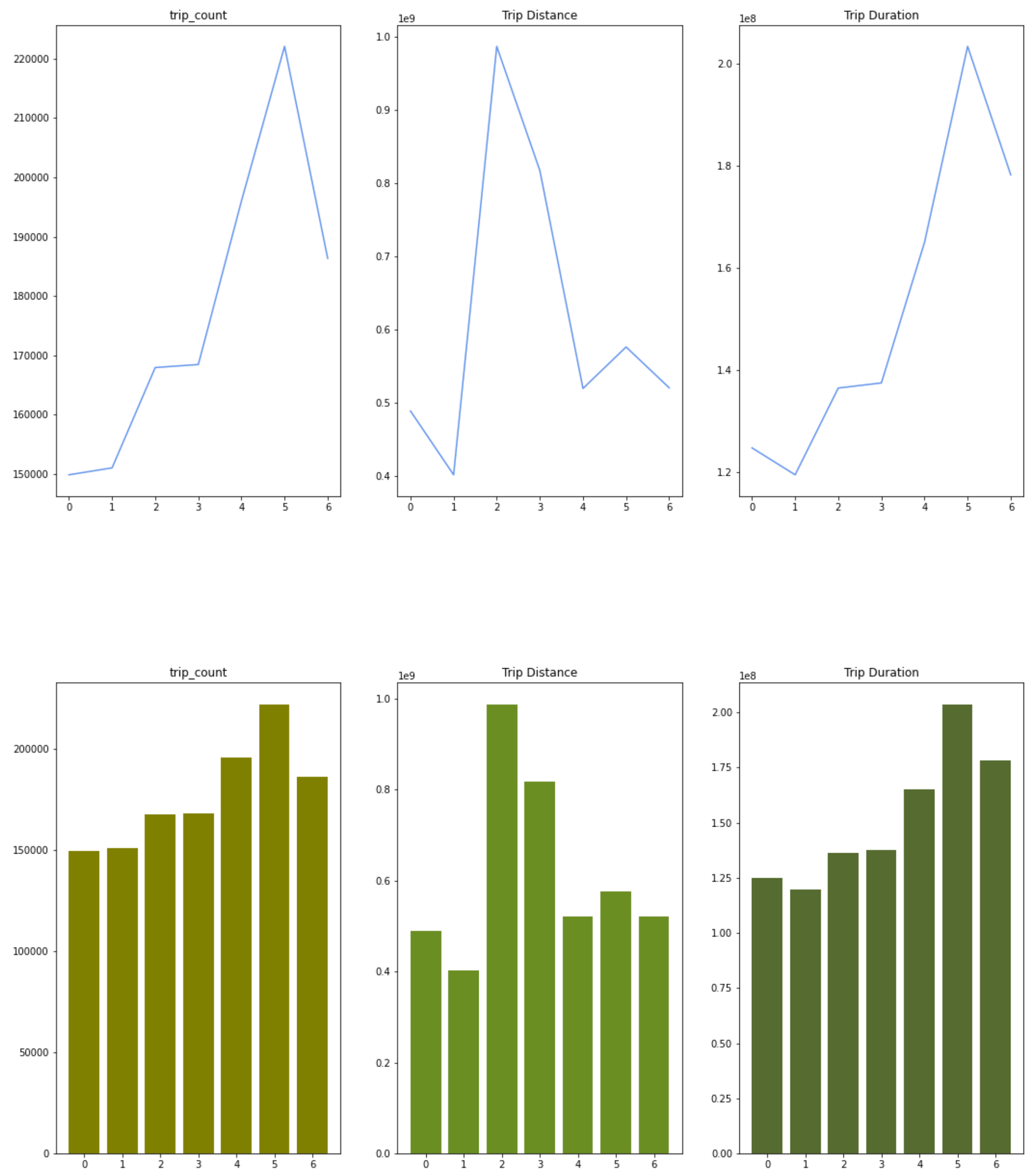
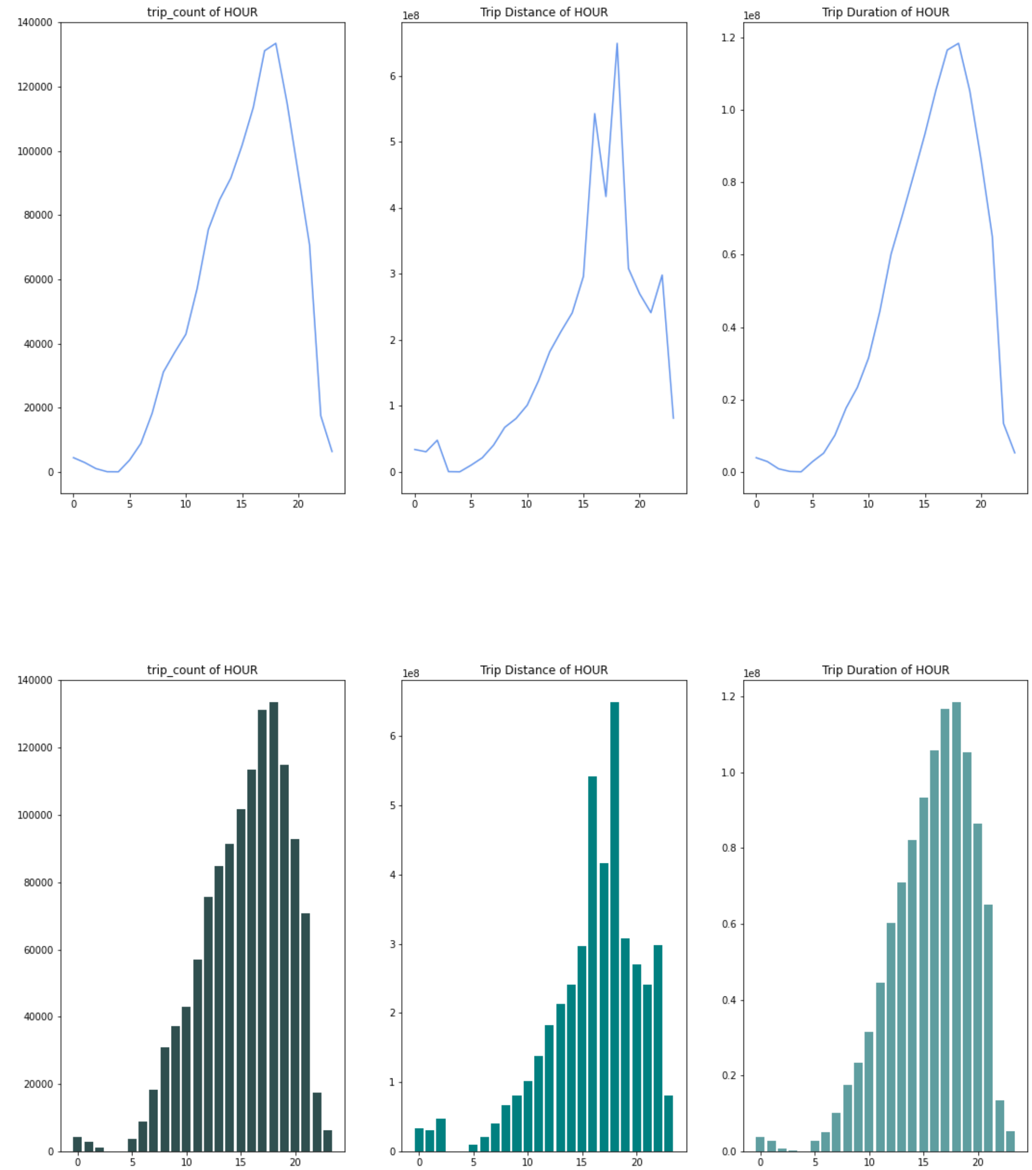
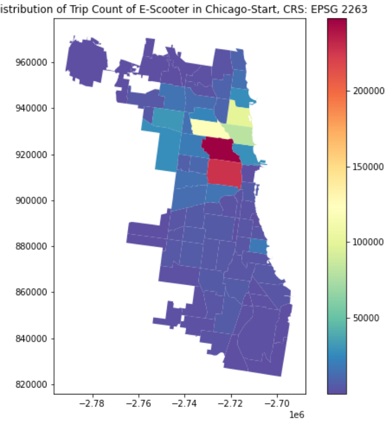


Fig :Trip Counts, Trip Distance, Trip Duration by HOUR

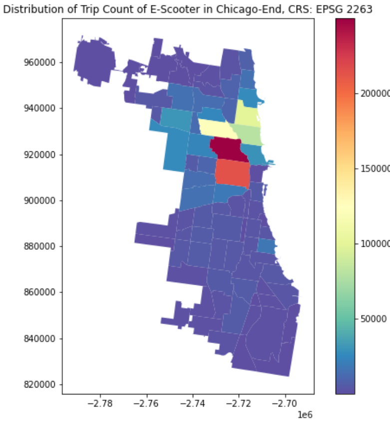


FigA.3 :Trip Counts, Trip Distance, Trip Duration by Community Area

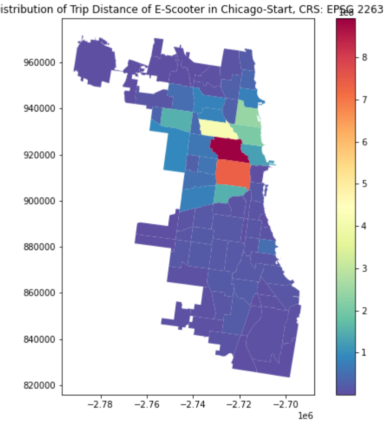
Distribution of Trip Count of E-Scooter in Chicago-Start, CRS: EPSG 2263



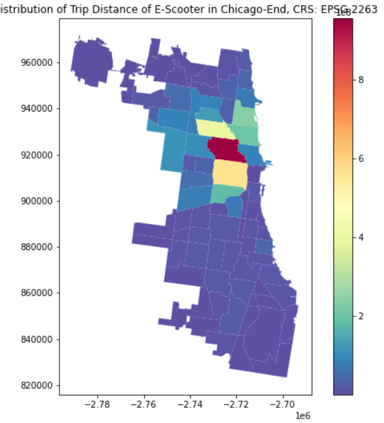
Distribution of Trip Count of E-Scooter in Chicago-End, CRS: EPSG 2263



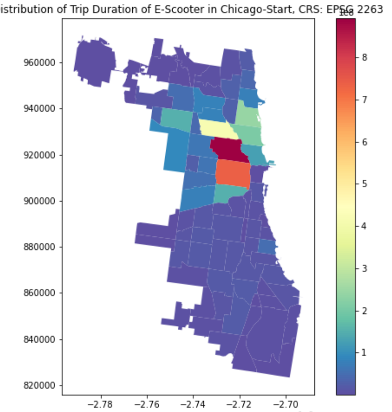
Distribution of Trip Distance of E-Scooter in Chicago-Start, CRS: EPSG 2263



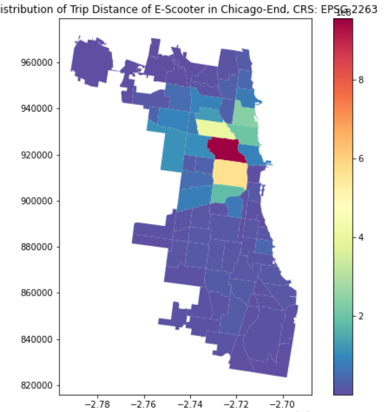
Distribution of Trip Distance of E-Scooter in Chicago-End, CRS: EPSG 2263



Distribution of Trip Duration of E-Scooter in Chicago-Start, CRS: EPSG 2263



Distribution of Trip Distance of E-Scooter in Chicago-End, CRS: EPSG 2263



Appendix B: Network Analysis

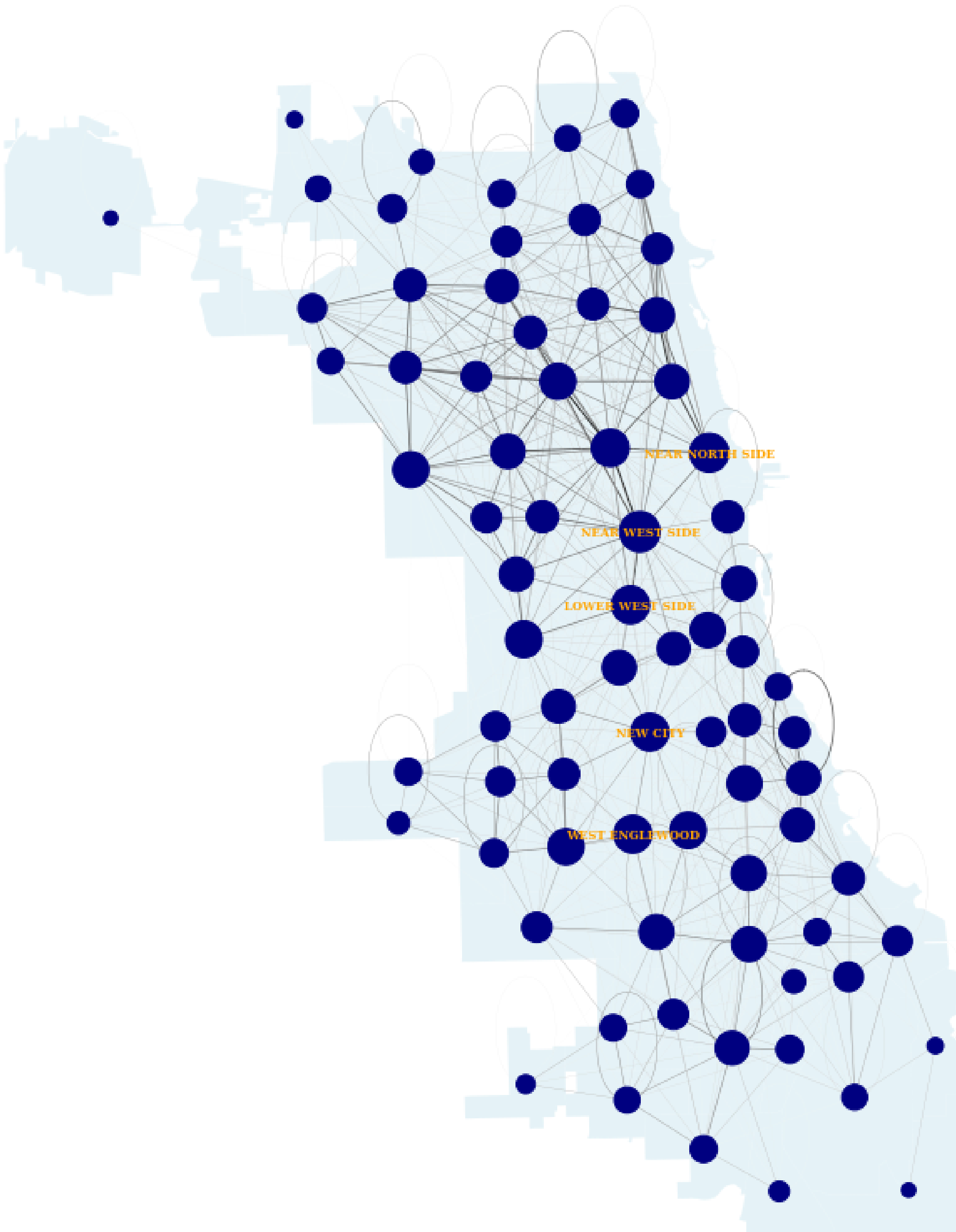


Figure B.1 Travel Routes and Community Nodes by Degree Centrality

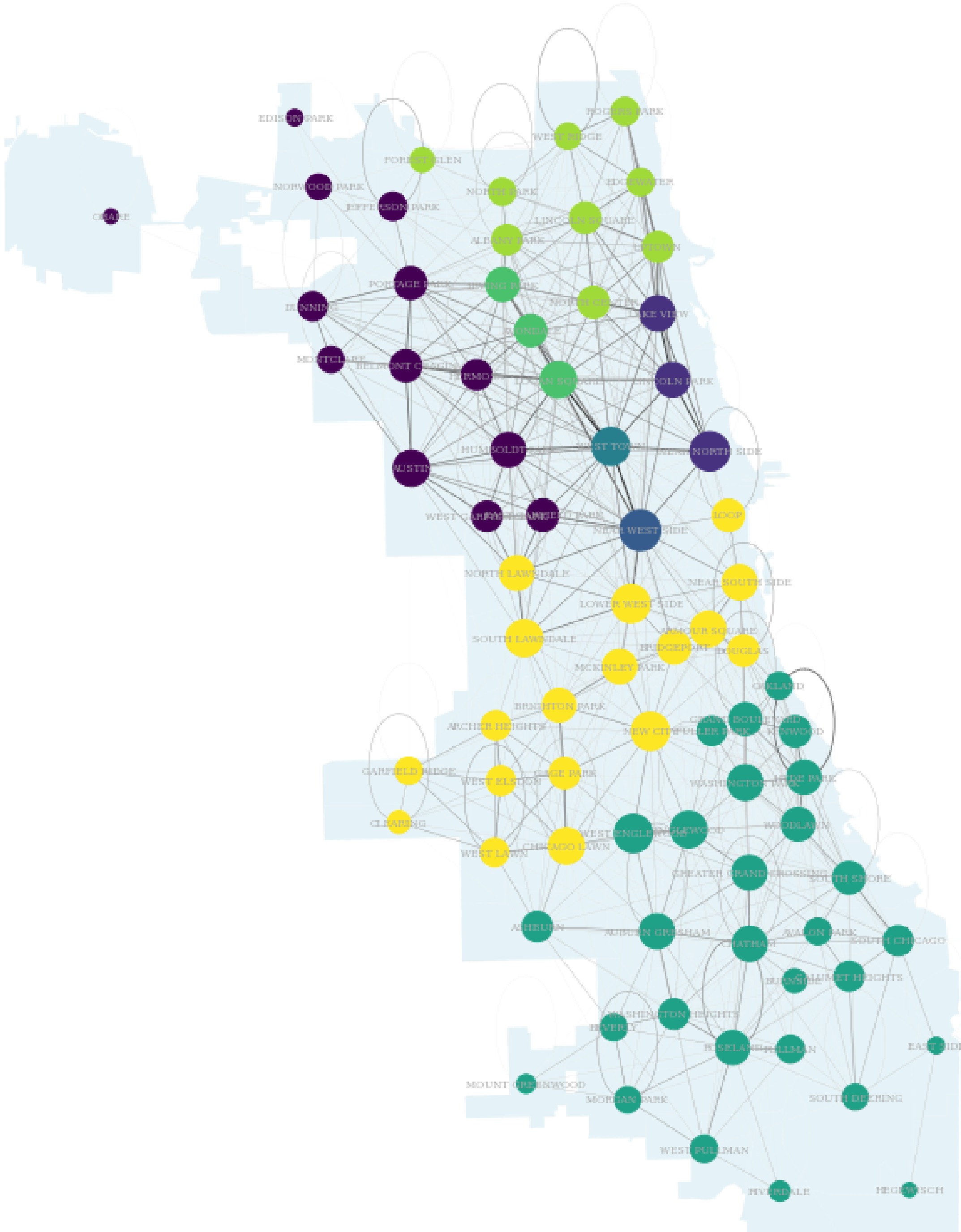


Figure B.2 Partition Results and Community Nodes by Degree Centrality

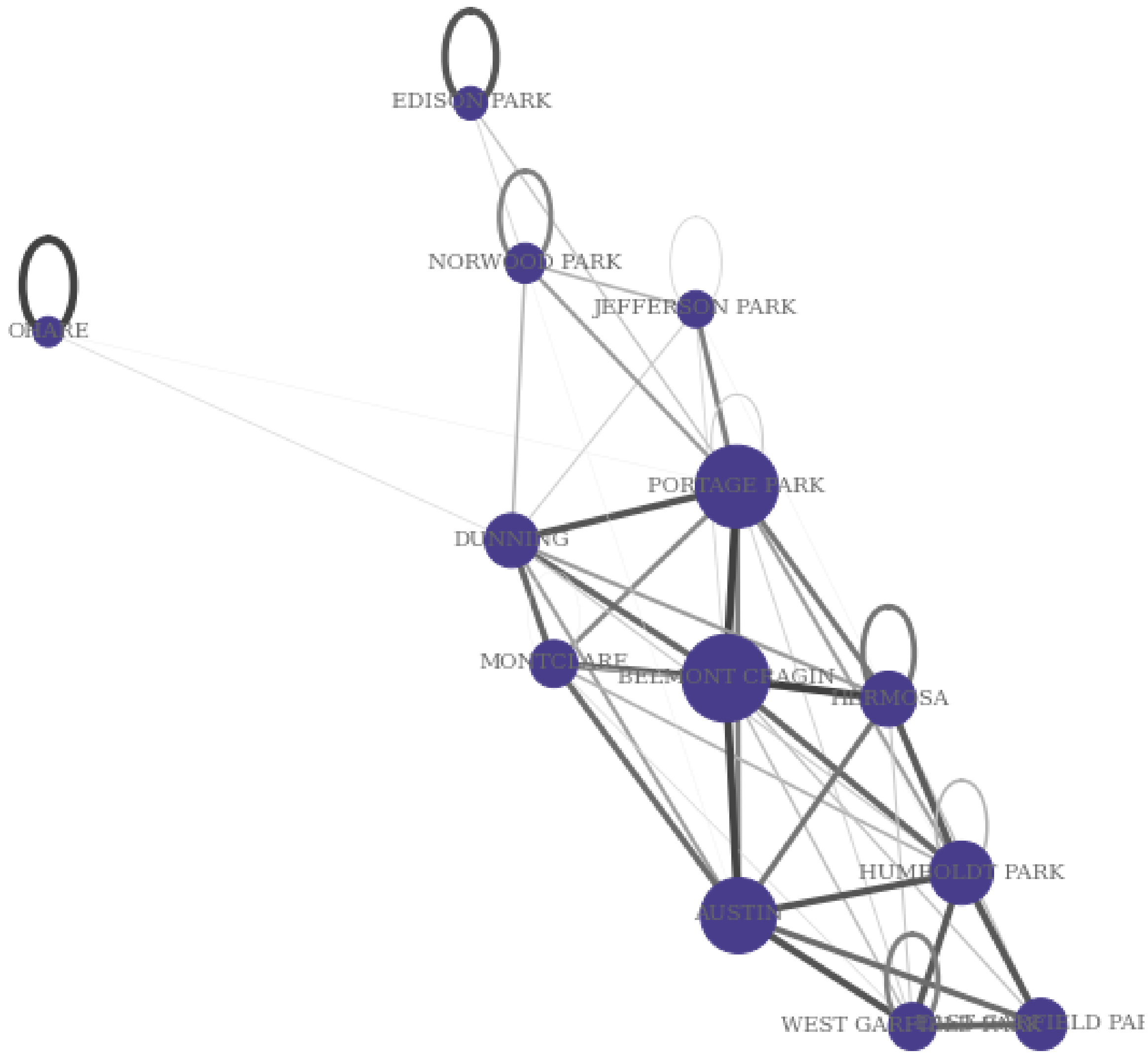


Figure B.3 Partition 0 Routes and Community Nodes by Pagerank Centrality

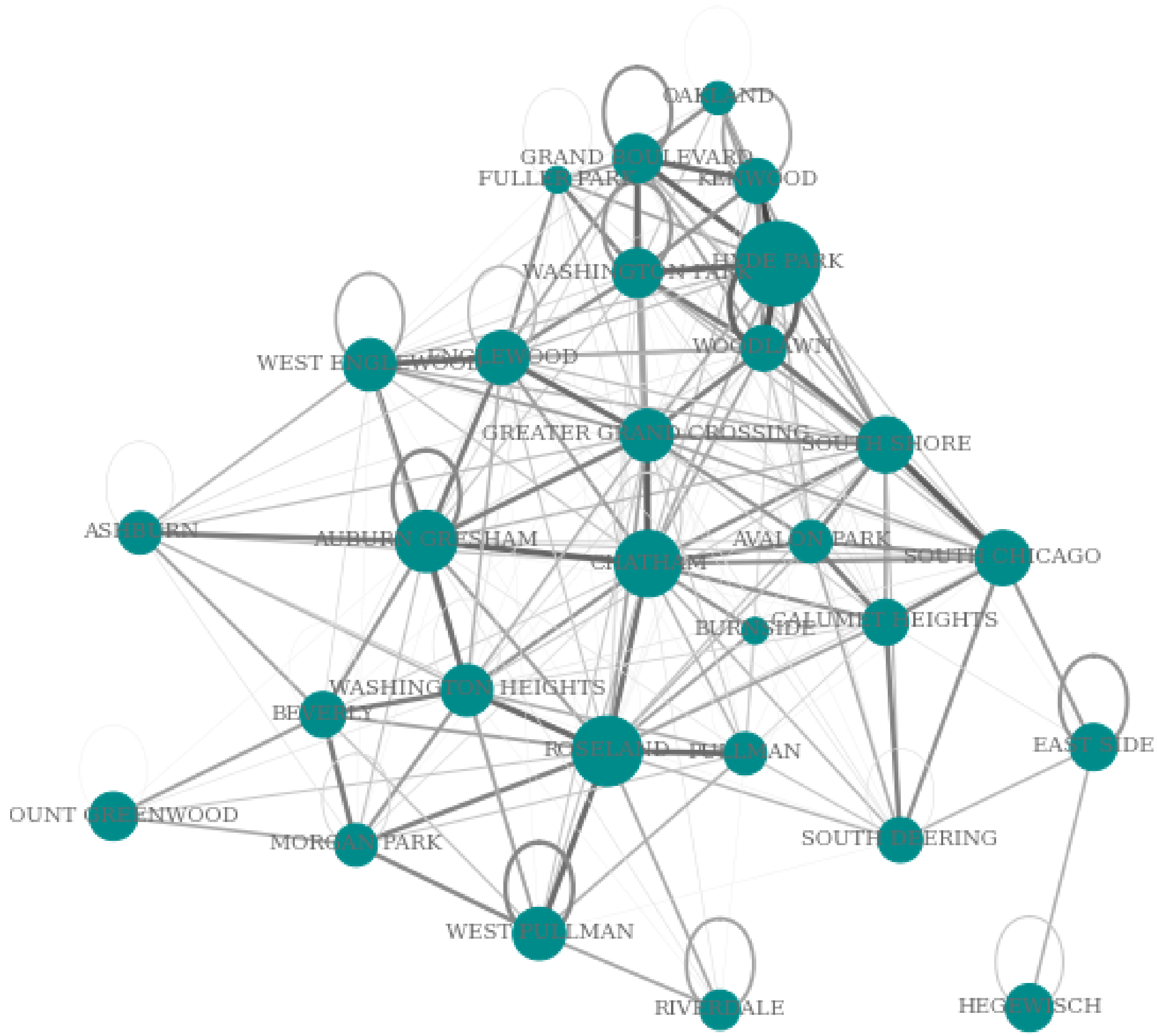


Figure B.4 Partition 4 Routes and Community Nodes by Pagerank Centrality

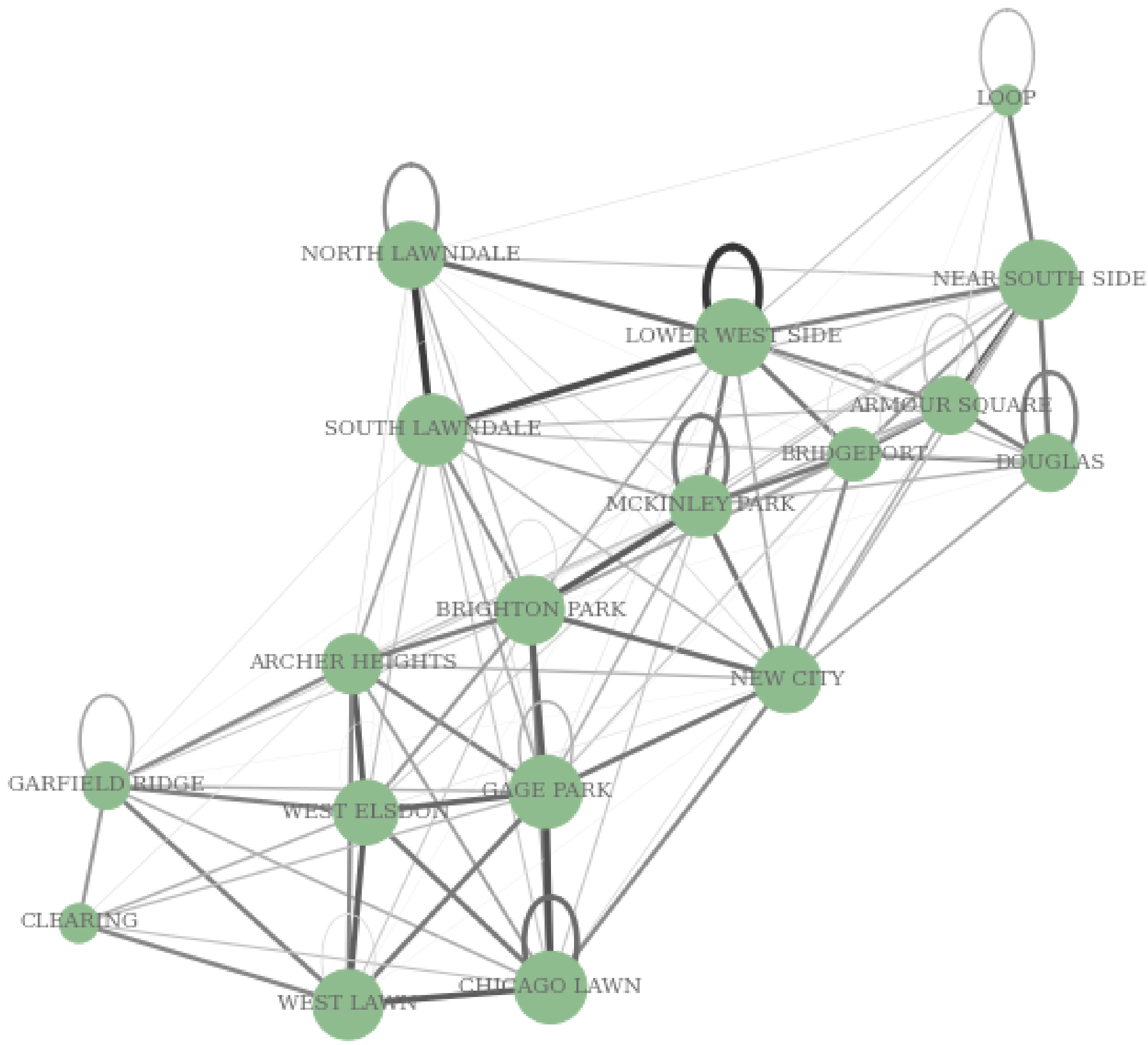


Figure B.5 Partition 7 Routes and Community Nodes by Pagerank Centrality

Appendix C: K-means Clustering Analysis

End Community Area Name	Trip Duration	cluster
WEST TOWN	137413889	3

End Community Area Name	Trip Duration	cluster
NEAR WEST SIDE	108447066	2

End Community Area Name	Trip Duration	cluster
ALBANY PARK	3009804	1
ARCHER HEIGHTS	3461160	1
ARMOUR SQUARE	3843682	1
ASHBURN	2563722	1
AUBURN GRESHAM	9798353	1
...	...	...
WEST GARFIELD PARK	4431895	1
WEST LAWN	5277738	1
WEST PULLMAN	5443752	1
WEST RIDGE	3561134	1
WOODLAWN	5425673	1

72 rows × 2 columns

End Community Area Name	Trip Duration	cluster
LAKE VIEW	66964573	0
LINCOLN PARK	61928698	0
LOGAN SQUARE	63589398	0

Figure C.1 Partition 4 Clusters by Community areas (Cluster 1,2,3,4)