Exploring the Divvy (Chicago) Bike Sharing System Network

William O. Agyapong

University of Texas, El Paso (UTEP), Department of Mathematical Sciences

Abstract

In this project I explored a 12-month sample of bike-share data from the Divvy bike-sharing system in Chicago and incorporated network analysis of the data to identify key stations and community structures. As a result of the large size of data obtained for just one year, a 10% stratified random sample was drawn to ease computational burden without compromising the representativeness of the data population. Bike share users in chicago use bike share for both one-way trips and round trips, but mostly one-way trips, and throughout the weekdays and on weekends, with casual users making more rides on weekends than on weekdays and vice versa for members. The analysis revealed that the Streeter Dr & Grand Avenue bike station is the most central station, while the State St & 95th St station appeared to be the most critical to information flow in the bike-share network. there were many Divvy bike-share stations with low usage of public bikes. The bike-share network can be thought of as involving one big giant component which suggests that similar strategies can be adopted to bring improvement in ridership across all or most of the stations.

1 Introduction

This project explores the bike-share data from the Divvy bike-share system in Chicago and incorporates network analysis of the data to identify key stations and community structures.

Divvy is a bike-sharing system that now serves the cities of Chicago and Evanston in the Chicago metropolitan area. The Chicago Department of Transportation owns the system, which Lyft has been operating since 2019. Divvy had 5,800 bicycles and 608 stations as of July 2019, covering almost the whole city (except Pullman, Roseland, Beverly, Belmont Cragin, and Edison Park). In the year 2017, National Association of City Transportation Officials (NACTO) named Divvy among the four largest station-based bike share system in the US. According the Wikipedia (2022), the name **Divvy** is a playful reference to sharing ("divvy it up"). The double Vs in the name is a reference to the shared-lane markers painted on bike lanes around the city, and shows the city's commitment to promoting bike safety while also making it easier for novice riders to get around. Divvy's light-blue color evoke the Chicago flag.

1.1 Background

Bike sharing systems has become one of the major means of transportation in major cities in the U.S. and around the world. It is great for quick trips around town or leisurely rides through parks. People use it to commute to work or school, run errands and explore cities. According to Alissa Walker, urbanism editor at Curbed, Bike share has been one of the major success stories in the US transportation system over the last decade. Annual bike-share data from the NACTO indicates that more than 119 U.S. communities now operate bike-share systems, beginning with Tulsa, Oklahoma's program in 2007. Bike share has been proved in studies to increase transit ridership and may even be safer than using personal bikes, since bicycle share schemes raise biker visibility, making riding safer for everyone.

Regarding how the system operates, users commonly check out a bike with a membership or credit/debit card. They can then ride to their destination and leave the bike at a docking station nearby. Bike-share system bikes are usually comfortable, have integrated locks and cargo baskets and usually includes features that make city biking safe and pleasurable.

Apart from the safety and environmental friendly mode of transport provided by bike-share systems, they bring huge economic gains by attracting more customers to nearby businesses. For instance, in congested places such as downtown regions, bike share customers spend far less time looking for parking and far more time patronizing neighboring businesses. Additionally, Bike sharing is a great way for individuals to choose active transportation for short trips which has a lot of health benefit since bike riding is a form of good exercise.

Due to the usefulness and the many benefits provided by bike share systems, many researchers have taken it upon themselves to conduct analyses of bike-share system data to enhance the understanding of the internal mechanisms within bike sharing systems. What is interesting is that in recent years some scholars have placed emphasis on the use of network analysis methodologies. For example, Yao et al. (2019) employed complex network methods to analyze the relationship between stations within the Nanjing bike-sharing system in China. Rixey (2013) also expanded on prior studies involving the use of station-level ridership data by including the network effects of the size and spatial distribution of the bike sharing station network, resulting in a more robust regression model for predicting station ridership. It is against this background that I sought to embark on this exploratory studies of bike-sharing system networks, and in particular the Chicago Divvy bike-share as a case study.

1.2 Objectives

The project aims to provide insight into the Divvy bike-share system data and specifically address the following research questions:

- Do different user types (members or casual users) and bike rideable types (classic, docked, and electric bikes) influence ridership in general and in terms of network structure?
- What are the central stations within the bike-share system station network?
- Is there some underlying community structure that can be utilized by the operators of the Chicago public bike-share system to help meet operational needs as well as the needs of bike riders?

Identifying differences in ridership for different users and different rideables, and the underlying network community structure is relevant for the successful operation of bike-share systems such as the Divvy. Also, knowing the central stations can lead to further studies of those stations to help understand what makes those stations popular for bike ridership and then transfer what is learned, if applicable, to other stations.

1.3 Setting

Data for this project is a historical trip data maintained by Divvy, the official bike share system in Chicago, and span a 12-month period starting from May 2021 to April 2022. The data is provided according to the Divvy Data License Agreement (2022) and released on a monthly schedule for public use. The trips were made by both members and casual public bike users to and from the over 800 bike-share stations in the Chicago metropolitan area.

1.4 Participants

The main subjects of the study included the over 800 bike-share stations in the Chicago metropolitan area. Bike trips to and from these stations were made by members of Divvy who are most likely to be local residents, and casual users who are probably visitors to the city.

1.5 Source of data and Variables

All the variables used in this project come from one single source, the Divvy system data. The data were obtained directly from the Divvy company's historical trip records available at https://divvy-tripdata.s3.amazonaws.com/index.html. The data are stored in compressed comma separated formats (csv). The source of the data claims that the data has been processed to remove trips that are taken by staff as they service and inspect the system; and any trips that were below 60 seconds in length (potentially false starts or users trying to re-dock a bike to ensure it was secure). Each trip is anonymized and Table 1 provides information about the trips.

Table 1: Available variables and their definitions

Variable Name	Description	Data Type
ride_id	Unique trip identification number	Alphanumeric character
rideable_type	Either a classic bike, docked bike, or an electric bike for the trip	Character/categorical
started_at, ended_at	Trip start and end day and time	Date/Time
start_station_name	Trip start station name	Character/categorical
start_station_id	Trip start station ID	Alphanumeric character/categorical
end_station_name	Trip end station name	Character/Categorical
end_station_id	Trip end station ID	Alphanumeric character/categorical
start_lat, start_lng	Pick-up station location latitude and longitude	Numeric
end_lat, end_lng	Bike return station location latitude and longitude	Numeric
member_casual	Whether the trip involved a subscribed member or a casual rider	Character/Categorical

Except for the ride id and the geographic location variables (longitudes and latitudes), the project utilized all the variables listed in Table 1.

1.6 Bias

It is worthy of mention that the study results or findings may be biased for several reasons. One such source of bias could arise from the sampling procedure. To address sampling bias, a probability sampling scheme called stratified sampling as described in Section 2. The sampling frame used also helped to address any potential class imbalance from the different user types and rideable types. Additionally, Weather conditions are an important influencing factors for the use of bike-share systems but such information were not available from the data repository.

1.7 Study size

The 12 months data were individually imported into the R statistical software and merged together with the help of the `vroom` R package. This resulted in a dataset with **5,757,551** observations measuring the total number of trips recorded over the one year period under review. Due to the huge number of observations, I took a 10% stratified random sample across the user type and rideable type variables. The stratification was done to ensure balanced representative sample. A final study size of **455275** after removing missing data was used. Table 2 presents details of the one year imported data set and the stratified sample.

Table 2: Original Data Versus Sampled Data

Attribute	Full 12 Months Data	10% Stratified Sample
Unique Station IDs	857	854
Unique Station Names	866	862
Number of Observations	5757551	575753

From Table 2, we have **854/862** unique stations in the sampled data compared to the **857/866** unique stations. in the full 12 months worth of data. This means our stratified sampled data is representative of the full data set. One would expect the number of distinct station ids to be equal to the distinct number of station names. However, this is not the case for the trip data under consideration for whatever reason, so I resorted to using the station names as those hold more meaningful information.

All statistical analyses and visualizations were conducted in the R statistical software. Codes for the analysis and report creation is available upon request. The rest of the report is organized as follows. Section 2 presents methods covering graph theory concepts and analysis and the sampling techniques used. Results from various exploratory and network analyses are presented in section 3. I conclude the report with a brief discussion of the results including key findings, interpretation and generalisability of findings in section 4.

2 Methodology

2.1 Stratified Random Sampling

It is often not realistic to work with data from an entire population, so a subset of data is selected through a number of ways. This project utilized the stratified random sampling method. Stratified random sampling is a probability sampling scheme where the target population is divided into distinct classes called strata. Subjects within each strata are similar in terms of some characteristics from the population. It allows researchers to obtain samples that best represents the overall population being studied. Here in this project, the large data set obtained were grouped by the user type and rideable type variables. After that 10% random samples from each resulting group were taken to obtain a balanced reasonable sample size.

2.2 Graph Theory Concepts

Networks were utilized in largely in this project to answer meet the study objectives. In graph theory, a network is defined, in its simplest form, as a collection of points joined together in pairs by lines, where a point is referred to as a node or vertex and a line is referred to as an edge. In the mathematical literature, a network is also called a graph. Nodes are often chosen to represent research objects, while edges connecting any two nodes denote a relationship of some kind. A network could either be directed or undirected, weighted or unweighted. In a directed network each edge has a direction, pointing from one node to another. Edges that connect nodes to themselves are called *self-edges* or *self-loops*. A network that has neither self-edges nor multiedges is called a *simple network*.

To construct the networks for the project, I treated bike-share stations as the nodes and edges as directed links pointing from from station A to station B if there was a trip between the two stations. Each edge is weighted by the number of trip records from one station to the other. The final networks were constructed from edge list data frame that I created consisting the starting station name and the ending station name with weigths, user types and rideable types attributes.

2.2.1 Centrality Measures

Centrality measures quantify how important or central nodes are in a network, where the definition of importance often relies on the particular context from which the network was derived. This report considers the four popular centrality measures which are described in the below.

Degree

Degree is the simplest centrality measure for a node in a network which measures the number of connections (edges) the node has. In directed networks, nodes have both an *in-degree* and an *out-degree*, and both may be useful as measures of centrality in the appropriate circumstances. The out-degree of a node is the number of other nodes to which a vertex has an outgoing edge directed to. The in-degree is the number of edges received from other vertices. A node with the highest degree has the most connections to other nodes in the network. This means degree centrality can help us find very connected individuals, popular individuals, individuals who are likely to hold most information or individuals who can quickly connect with the wider network.

Betweenness

The amount of times a node is on the shortest path between other nodes is measured by betweenness centrality. A path is a series of adjacent nodes (A series of edges that take us from one node to another node). The shortest path between any two nodes is the least amount of total steps (or edges). If a node C is on a shortest path between A and B, then it means C is important to the efficient flow between A and B. Flows would have to take a longer route from A to B without C. As a result, betweenness measures how many shortest pathways each node has. Nodes with high betweenness are key bridges between different parts of a network. The higher a node's betweenness, the more important they are for the efficient flow in a network. Betweenness centrality can be very large, so it is often helpful to normalize it by dividing by the maximum and multiplying by some scalar when plotting.

Closeness

The closeness centrality also makes use of the shortest paths between nodes. The length of the shortest path between two nodes is used to calculate the distance between them. The average distance between a node and all other nodes is called farness. Closeness is then the reciprocal of farness (1/farness).

Eigenvector Centrality

Degree centrality only takes into account the number of edges for each node, but it leaves out information about the relative importance of the neighboring nodes. In many circumstances a node's importance in a network is increased by having connections to other nodes that are themselves important. For instance, If A and B have the same degree centrality, but A has ties with all high degree individuals and B is related to all low degree individuals, the we would want to see A with a higher score than B. This is where the eigenvector centrality comes in as an extension of the degree centrality by also taking into account how well connected a node is, and how many edges their connections have, and so on through the network. In other words, the eigenvector centrality awards each node points proportional to the centrality scores of their neighbors.

2.2.2 Measuring Network Structure

Density of a network is the proportion of edges that actually exist out of the total possible edges that can be formed. This also indicates how interconnected a network is. Another measure of how interconnected a network is **average path length**. This is computed by determining the mean of the lengths of the shortest paths between all pairs of vertices in the network. The longest path length between any pair of vertices is called the **diameter** of the network graph.

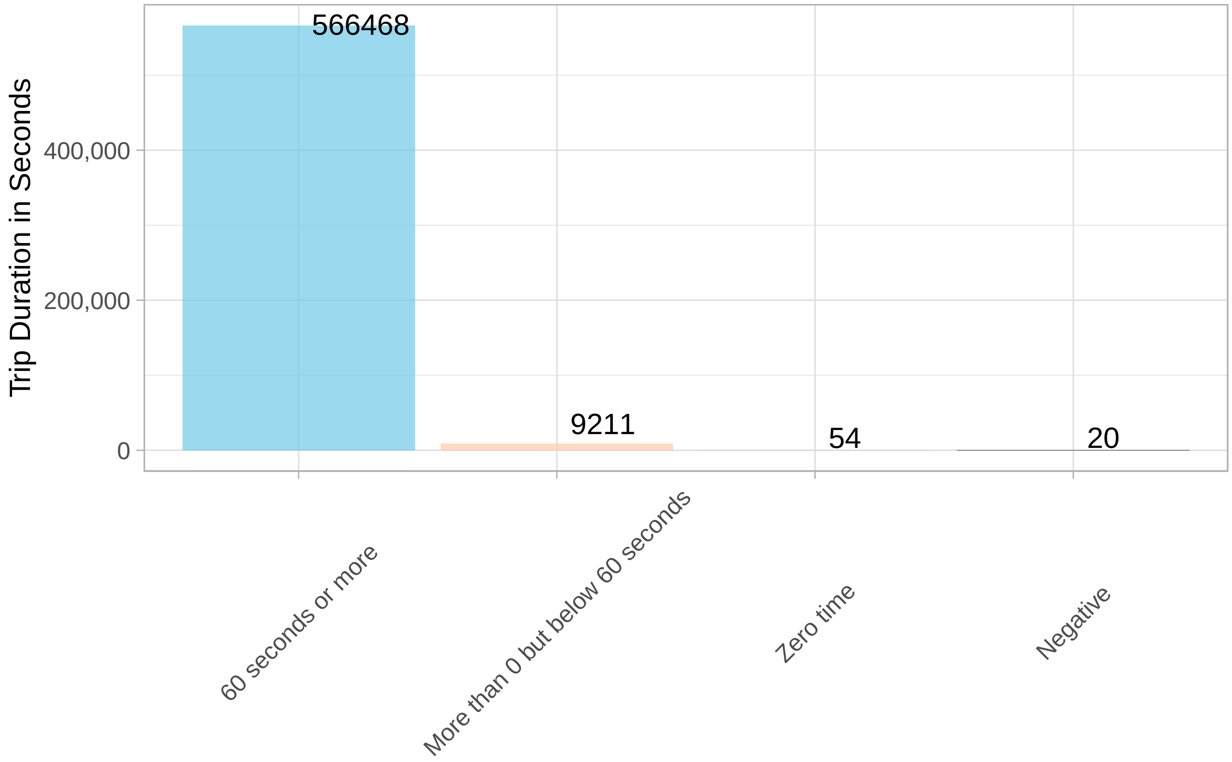
A related concept is **community detection**. Community detection is very useful in understanding and examining the structure of large networks such as the ones considered in this project. A measure known as modularity is used to assess the quality of community division. I used the Louvain and walktrap community detection algorithms because they are often used in practice and have proved to work well with large networks. The task of community detection can be seen as an optimization problem. Louvain is an agglomerative modulatiry optimization algorithm for finding community structure in a network. It is well-known for its speed. At the beginning, each node is assigned to a community on its own. In subsequent steps each node is combined with the community that increases modularity the most. Several rounds are repeated until a configuration where nothing changes or where the modularity cannot be improved any further. The final configuration is then taken asthe community division of the network. The **walktrap** algorithm is also agglomerative but it uses random walks instead of modularity whereby "distance" between nodes are measured through random walks in the network. The basic idea of the algorithm is that random walks on a graph or network tend to get trapped into densely connected parts corresponding to communities. The quality of the divisions can be assessed using either modularity or any other measure.

3 Analysis and Results

3.1 Exploratory Data Analysis

3.1.1 Data anomalies and missing data

Figure 1: Trip Duration Anomalies



According to the data description from the data source (https://ride.divvybikes.com/system-data), the trip data were preprocessed to remove trips that lasted below 60 seconds, meanwhile this initial exploration of the data has revealed a good number of trips having trip duration below 60 seconds and even negative time duration. Therefore, without any further information to point out what could be responsible for the obvious anomalies, I decided to discard trips whose time length were below 60 seconds and such trips amounted to **9,285** as can be seen in the above graph. Next is a table of missing data present after removing the observations with unreasonable trip time length.

Table 3: Missing values in the 10% stratified sample data

variable	n_miss	pct_miss
end_station_name	81550	14.4%
end_station_id	81550	14.4%

variable	n_miss	pct_miss
start_station_name	76834	13.56%
start_station_id	76834	13.56%
end_lat	479	0.08%
end_lng	479	0.08%

Table 3 reveals what I observed from the data set where some of the trips had station name or id and location. Given the limited amount of information about these missingness and the fact that we already have more than large enough sample, I removed the rows of the data with any missing record from the data set resulting in a final sample size of **455,275** for the study.

3.1.2 Types of users and rideable types and the rides per day of week

The figures 2 and 3 below show the distribution of rides in thousands (K) for the two types of users, members and casual users as well as the distribution of rideables, respectively. We can see that there are more members (subscribers) than casual bike riders, with a difference of about 12%. Clearly, members and casual users show different usage patterns throughout the week. Casual users tend to use bikes more during weekends (from about Friday midnight to Sunday midnight) than during the weekdays.

According to figure 3, **Classic bikes are the most patronized bikes with a 69% share of the total ridership**. Looks like Docked bikes are likely to phase out over time. A yearly analysis can be very useful in discovering how the usage patterns has been evolving over time.

Figure 2: Distributions of Rides by day of week and type of bike users

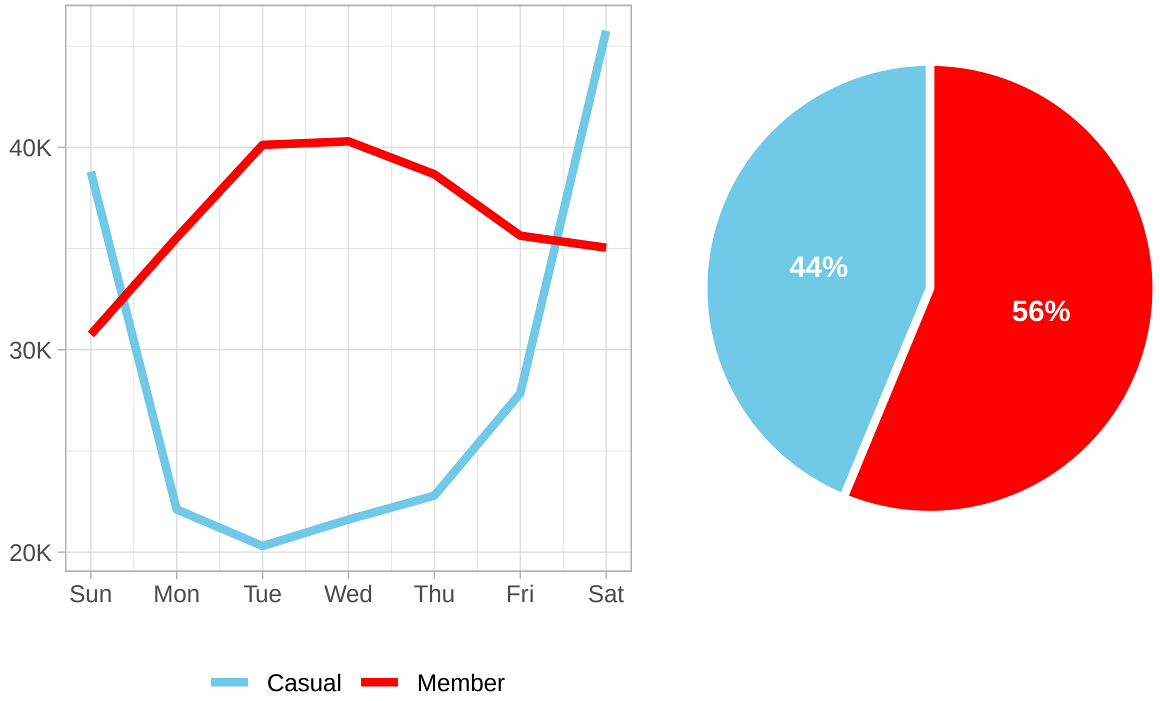
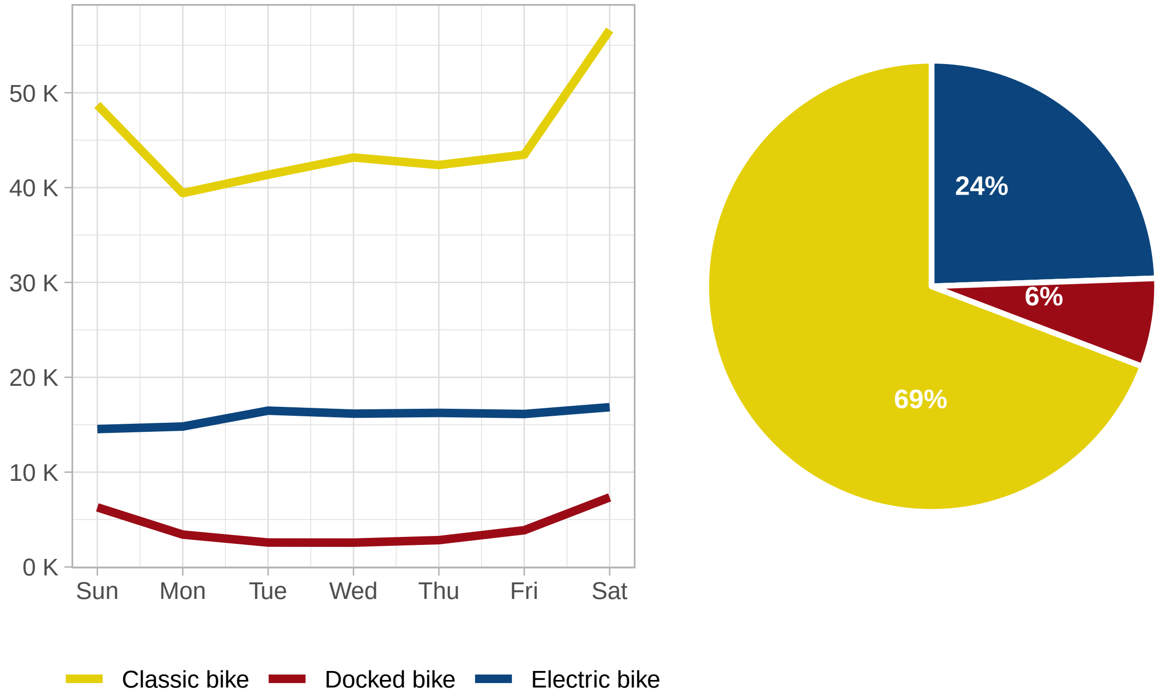


Figure 3: Distributions of Rides by day of week and rideable types



3.1.3 Weekday vs weekend trips and daily trips

Figure 4: Weekday VS Weekend Trips By Time Of Day

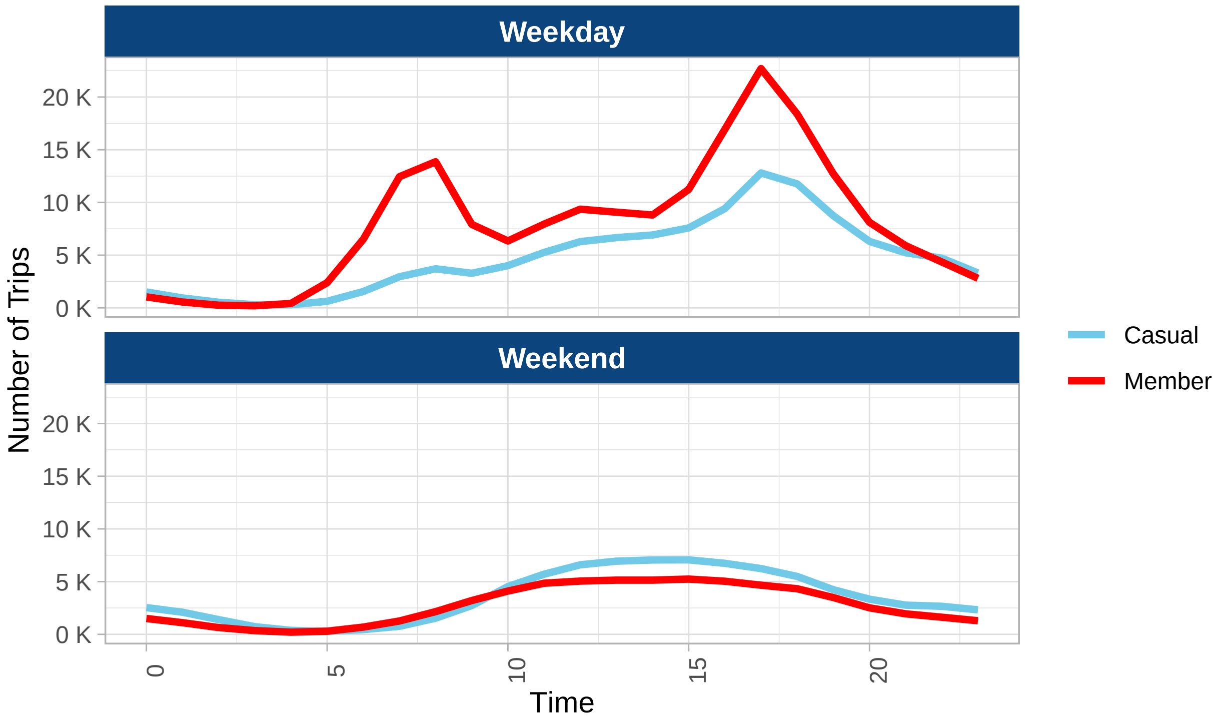
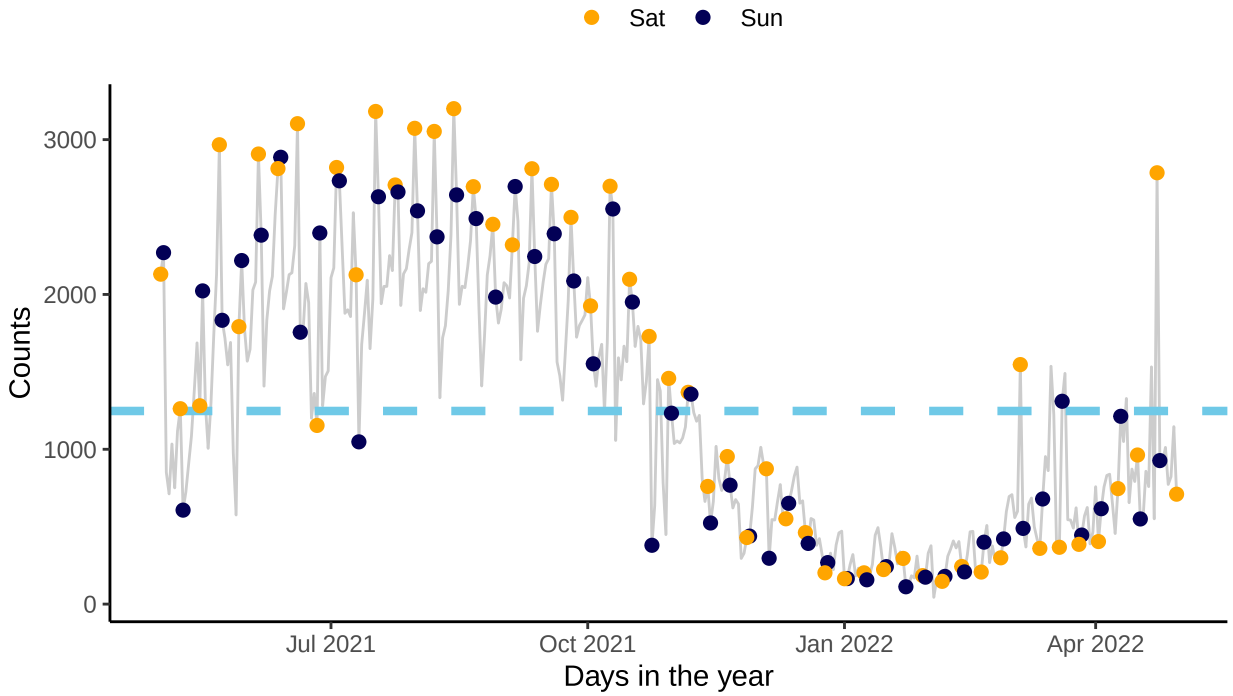


Figure 5: Daily rides (number of trips)



Bike usage fluctuates throughout the day, and the majority of trips occurred between 6

am and 8 pm. On weekdays, members and casual users show different travel patterns. Members show commute activity (spikes in usage from 7:00 – 9:00 and 17:00 – 19:00), and constant usage throughout the day. Casual usage increased throughout the day with a constant peak from noon-7pm. On the weekends, members and casual users exhibit similar behavior with the daytime peak occurring between 2-4pm. - Casual users made a higher proportion of mid-day trips.

The trend line spikes in Figure 5 represent the high usage of bikes, usually on Saturday and Sunday, and on some weekends the bike usage dropped; The drop is likely to have occurred during the Holidays or the winter period. The usage in summer time is above the average (the green dashed line).

3.2 Network Analysis

Now that we have gleaned much information from the Divvy May 2021 - April 2022 bike-share trip data set, it is time to look at the networks constructed from the same data set. In this section I present results obtained from performing various network analysis in an attempt to help answer the research questions motivating the study.

3.2.1 Network construction and characteristics

To create networks, the trips data were grouped by the start and end station names (so as to make them the vertices in the network) and summarized to get all the trips that occurred between any given pair of stations and edge weights. After obtaining an edge list data frame from the previous operation, a giant network having 843 veritices (bike stations) and 160,754 edges was constructed. This network I called full network. A second network data was created by taking another 2% stratified random sample from which sub graphs for the user types and rideable types were constructed for the purpose of network visualization. The same subset of edge list data was used to construct the network for the community structure detection analysis. I left out a plot of this large network since it did not make for a good viewing. However, I computed the centrality measures from this main graph to help identify the key stations in the Divvy bike share system. Numerical characteristics of the resulting networks are presented in Table 4.

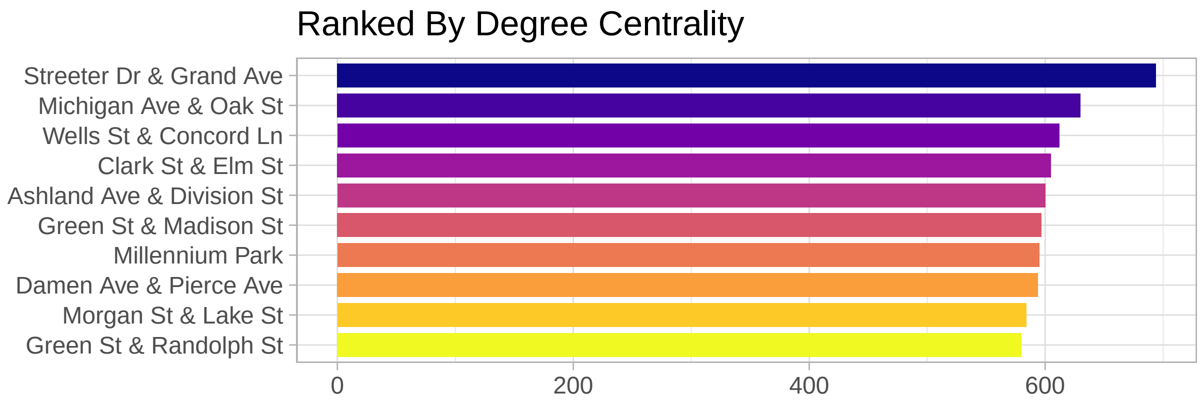
Table 4: Characteristics of constructed networks

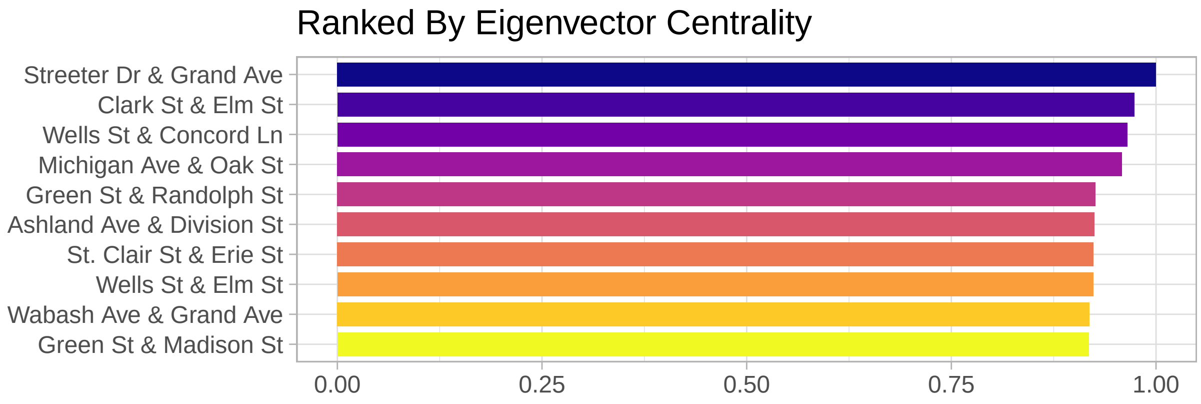
Characteristic	Full Graph	Member Graph	Casual Graph	Classic Graph	Docked Graph	Electric Graph
Number of Vertices	8.430e+02	526.0000	5.52e+02	549.0000	332.0000	506.0000
Number of Edges	7.834e+04	4491.0000	3.49e+03	5337.0000	522.0000	2128.0000
Density (edges)	1.104e-01	0.0163	1.15e-02	0.0177	0.0048	0.0083
Average path length	2.391e+00	3.1987	3.23e+00	3.1136	4.5903	3.5108
Diameter	8.000e+00	11.0000	9.00e+00	9.0000	12.0000	12.0000

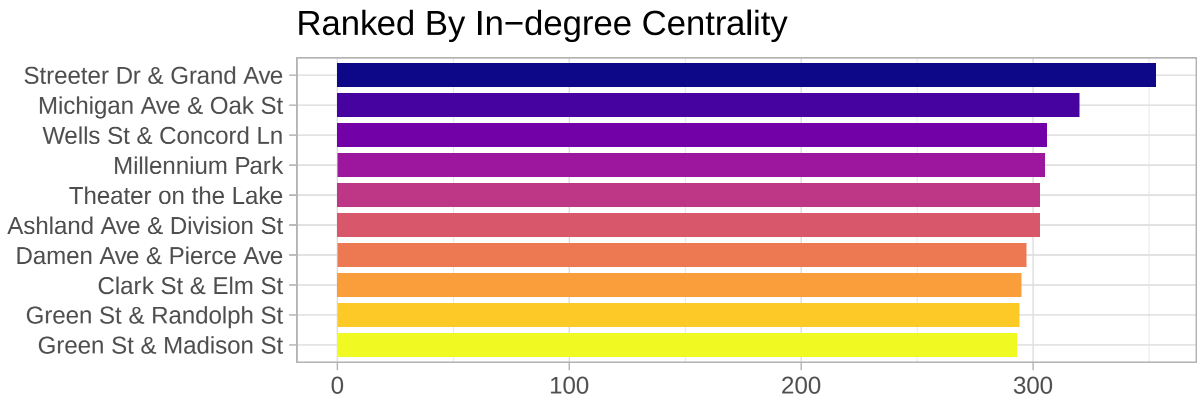
3.2.2 Ranking stations by Centrality Measures computed from the full graph

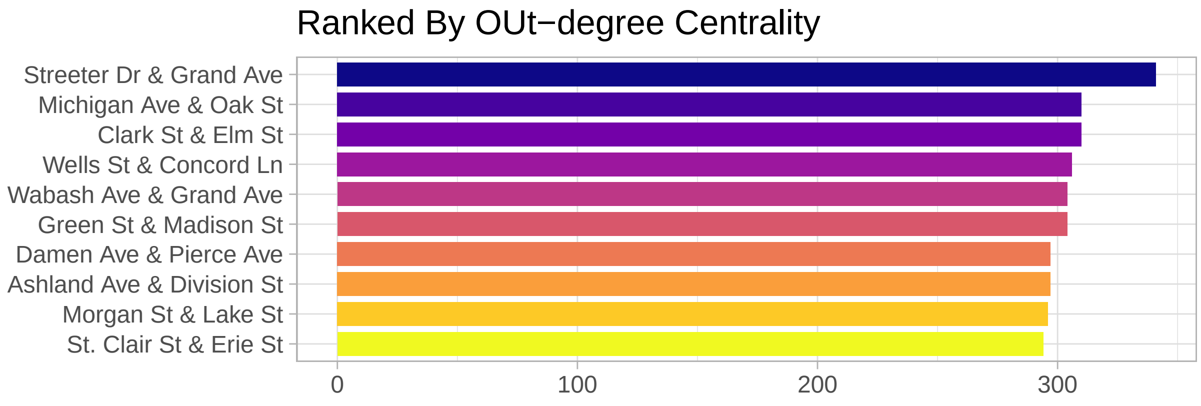
Interestingly, apart from the betweenness centrality the *Streeter Dr. & Grand Avenue* station appears to be the most popular station in the sampled data followed by *Wells St & Concord Ln* and *Michigan Ave & Oak St* in either second or third positions for the most part. In terms of betweeness centrality the *State St & 95th St* station is the most central. Though, closeness centrality ranked the *Streeter Dr. & Grand Avenue* in first place, the differences between the rankings is not that significant as depicted by the roughly equal sizes of the bars.

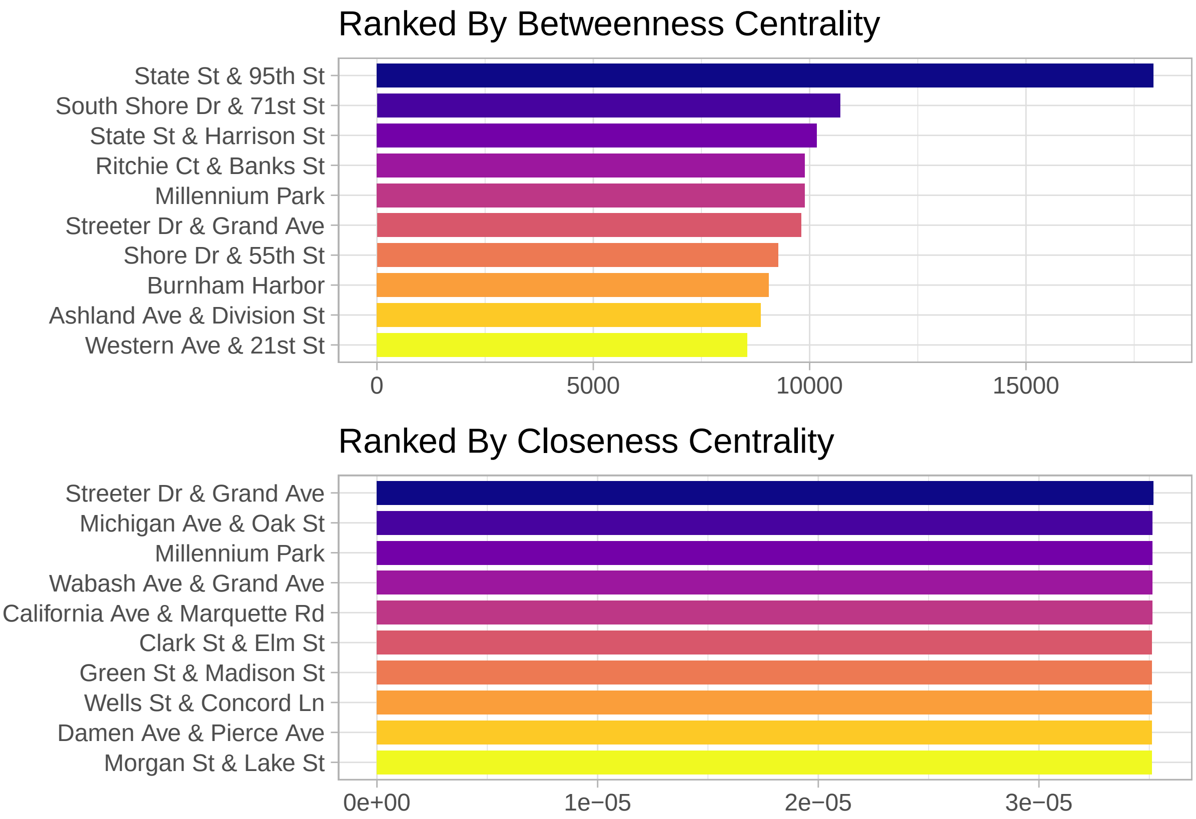
Figure 6: Top 10 Stations According to Centrality Measures


Ranked By Degree Centrality
Streeter Dr & Grand Ave
Michigan Ave & Oak St
Wells St & Concord Ln
Clark St & Elm St
Ashland Ave & Division St
Green St & Madison St
Millennium Park
Damen Ave & Pierce Ave
Morgan St & Lake St
Green St & Randolph St
0
200
400
600


Ranked By Eigenvector Centrality
Streeter Dr & Grand Ave
Clark St & Elm St
Wells St & Concord Ln
Michigan Ave & Oak St
Green St & Randolph St
Ashland Ave & Division St
St. Clair St & Erie St
Wells St & Elm St
Wabash Ave & Grand Ave
Green St & Madison St
0.00
0.25
0.50
0.75
1.00


Ranked By In−degree Centrality
Streeter Dr & Grand Ave
Michigan Ave & Oak St
Wells St & Concord Ln
Millennium Park
Theater on the Lake
Ashland Ave & Division St
Damen Ave & Pierce Ave
Clark St & Elm St
Green St & Randolph St
Green St & Madison St
0
100
200
300


Ranked By OUt−degree Centrality
Streeter Dr & Grand Ave
Michigan Ave & Oak St
Clark St & Elm St
Wells St & Concord Ln
Wabash Ave & Grand Ave
Green St & Madison St
Damen Ave & Pierce Ave
Ashland Ave & Division St
Morgan St & Lake St
St. Clair St & Erie St
0
100
200
300



Network graphs for the sub groups defined by user type and rideable type are presented in Figure 7 and Figure 8. Since the four centrality measures produced fairly similar rankings in terms of which station appeared in the top 10, for the subgraphs I only presented network plots where the size of vertices are normalized by the degree centrality. Generally it can be seen that the vertices are not well separated into the various groups created by the user type and rideable type, depicting a similar underlying network structure.

Figure 7: Sub network by user types

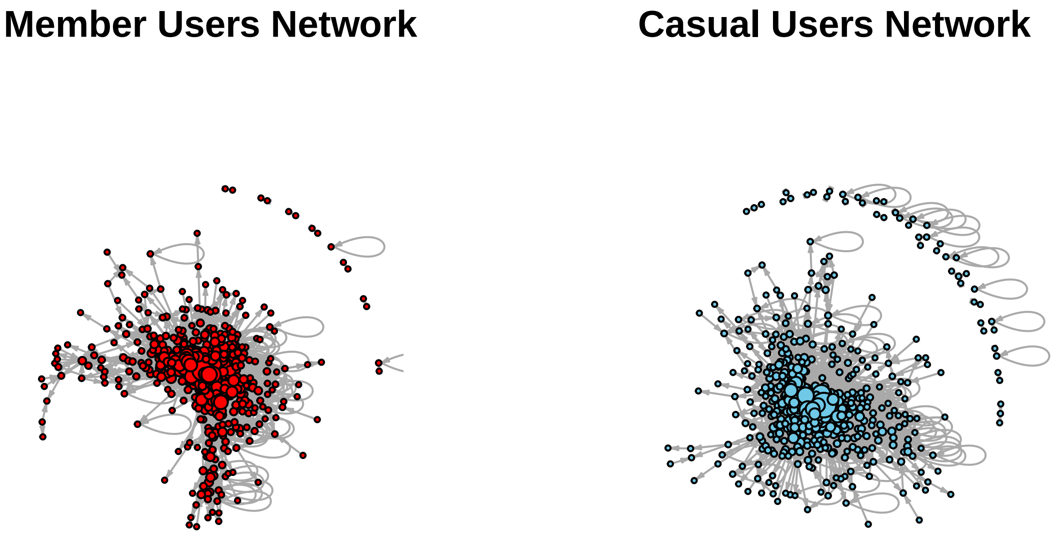
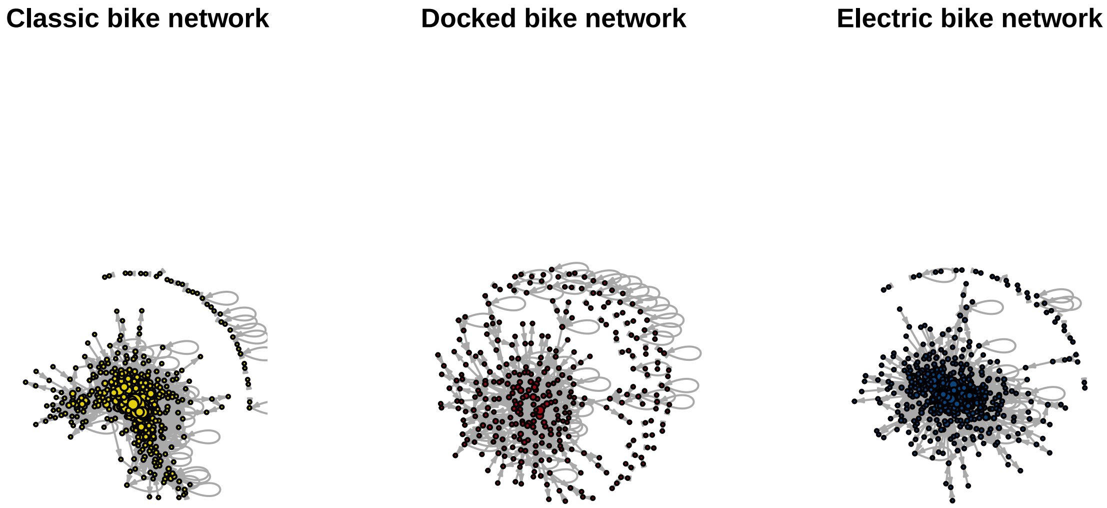


Figure 8: Sub networks by rideable types



Classic bike network

Docked bike network

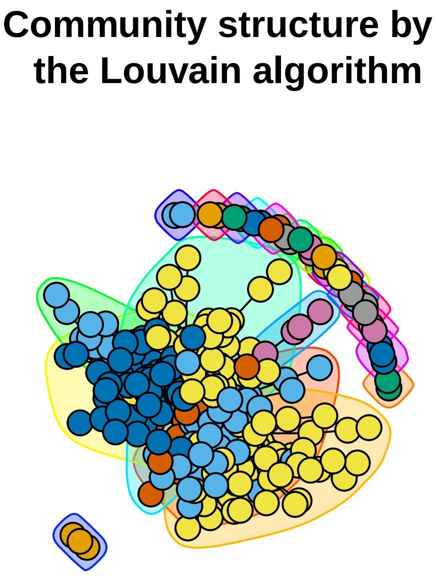
Electric bike network

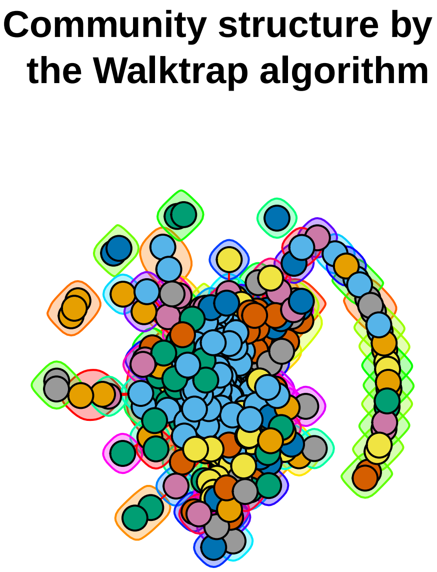
The self-loops represent instances where riders return bikes back to the station where the trip started. This is a most likely phenomenon as depicted by the above networks. Also, there are a lot of isolates in all sub networks.

As can be observed from the graphs there are many self-loops. Here, self-loops indicate situations where a bike rider picks up a bike from station A and returns it to the same station, thus an edge moves from station A back to itself. This is a reasonable and possible phenomenon in a bike-sharing system. Therefore, based on the context of my project, I believe strongly that the self-loops provide very useful information. Hence, the main reason why I maintained them in the above graphs. However, the self-loops appear to be too many to the extent that they somehow conceal the true nature of the networks, so in the subsequent graphs for the community detection I simplified the graphs down to remove self-loops and also make them undirected for simplicity. It must be generally understood from here that whenever a node stands in isolation without any edge (connection) going out or coming in, there is at least one bike trip starting from it and ending at that same node.

3.2.3 Analysis of Communinity Structure

Figure 9:





The network communities structure is not much different from what was already observed from the previous graphs. The Louvain algorithm produced 40 communities while the Walktrap algorithm detected 103 communities. These sub groups appear to be too many to be all that useful, however, the communities detected by the Louvain algorithm appears quite reasonable.

4 Discussion

4.1 Key Findings and Interpretation

The exploratory data analysis revealed that bike ridership differ significantly between members and casual users and also between bike rideable types. Bike usage fluctuates throughout the day over the weekday as well as the weekend, and the majority of trips occurred. There appeared to be more member users, as expected, than casual users, probably because members of the Divvy bike-share system are residents of the various communities around the Chicago region, while the casual users may likely be visitors. Also, Classic bikes were found to be the most patronized bikes with a 69% share out of the total 455,275 ridership sampled, with docked bikes usage likely to phase out over time. Some anomalies including recorded time lengths and missing information were found.

The network centrality measures suggested that stations including the Streeter Dr & Grand Avenue, Michigan Ave & Oak St, Wells St & concord Ln, Millennium Park, Clark St & Elm St, Ashland Ave & Division St, among a few others, appear be the central bike-share stations in the Divvy bike-share system. However, the Streeter Dr & Grand Avenue seems to be the most popular station. The results signifies that receives the most in and out trips (based on degree), the State St & 95th St stations is critical for information inflow in the bike-share system (according to betweenness centrality), and again the Streeter Dr & Grand Avenue station lies closer to most of the stations.

Moreover, the networks produced from the user type subgrops and rideable type subgroups did not show much significant differences in terms of network structure. This was further assentuated by the community detection analysis. According to the Louvain algorithm and the Walktrap algorithm, the Divvy bike-share system can be divided into 40 communities and 103 communities, respectively, based on a random sample of the data. These sub groups appear to be too many to be of any practical significance, however, the size communities detected by the Louvain algorithm appears quite reasonable. Strong overlaps were observed in all communities produced by the two algorithms. By this, I can say that the Divvy bike-share system appears to consist of one big giant component, suggesting that a homogeneous strategies can be adopted to improve ridership across all or most of the stations.

4.2 Limitations

First and foremost, the study was largely plagued by limited time constraint. One major setback to the study has to do with the absence of demographic data such as gender and age of the users and weather condition factors such as temperature in the data set. Such information could have been utilized to explain some of the phenomena observed. Another limitation results from the fact that the study did not make use of the geographic location information provided in the data set, once again due to the limited time for the project. Incorporating such spatial information in this project can help Divvy and other communities interested in adopting bike sharing systems to identify potential station locations that will serve the most amount of number of riders. It is also recommended that the Divvy bike-system data recording be validated to ensure accuracy of information.

4.3 Generalisability

Overall, I believe the results generalizes well to the population of bike-share riders in the Chicago metropolitan area. This is because the sampling framework used ensured balanced and representative samples. In spite of this, an analysis of the full one year data obtained might show slightly different results. However, the one caveat is that the findings may not generalize well to other bike-share systems across the United States of America since the data only came from one city, Chicago. So I will say that external validity of the study is compromised. Therefore, as an extension to this project, future studies can consider samples from the various or the four largest bike-share systems, namely Citi Bike (New York city), Divvy (Chicago), Hubway (Greater Boston Area), and Captital Bikeshare (Washington, DC), as was done by Rixey (2013).

5 Refernces

- Rixey, R. A. (2013). Station-level forecasting of bikesharing ridership: Station network effects in three US systems. Transportation research record, 2387(1), 46-55.
- Yao, Y., Zhang, Y., Tian, L., Zhou, N., Li, Z., & Wang, M. (2019). Analysis of network structure of urban bike-sharing system: A case study based on real-time data of a public bicycle system. Sustainability, 11(19), 5425.
- Source of data: https://divvy-tripdata.s3.amazonaws.com/index.html https://ride.divvybikes.com/system-data. Accessed on May 6, 2022.
- Data Description: https://ride.divvybikes.com/system-data. Accessed on May 6, 2022.
- Data Usage License: https://ride.divvybikes.com/data-license-agreement. Accessed on May 6, 2022.
- The quiet triumph of bike share: https://archive.curbed.com/2019/12/16/20864145/bike-share-citi-bike-jump-uber. Accessed on May 6, 2022.
- Wikipedia (2022). https://en.wikipedia.org/wiki/Divvy. Accessed on May 8, 2022.
- NACTO Bike Share and Shared Micromobility Initiative (2017): https://nacto.org/bike-share-statistics-2017/ Accessed on May 8, 2022.
- Network Analysis in R by Dai Shizuka: https://dshizuka.github.io/networkanalysis/tutorials.html
- Centrality measures: https://cambridge-intelligence.com/keylines-faqs-social-network-analysis/
- https://bookdown.org/markhoff/social_network_analysis/your-first-network.html