Bike Share Unplugged: Understanding San Francisco's Cycling Connections

2025-1-1

Author: Krish S. Bhalala

Analysis of San Francisco Bike Share Network

This report analyzes the San Francisco bike share network using centrality measures from graph theory to understand the network's organization and evolution. We'll focus on the July 2014 and July 2015 datasets to examine changes over one year.

We will look at various centrality measures in graph theory, including degree centrality, betweenness centrality, and closeness centrality. These measures will help us interpret the importance and roles of different stations within the bike share network. By applying these centrality metrics, we can gain insights into which stations are most central or influential in the system's structure and flow.

Additionally, we will examine changes between the 2014 and 2015 graphs, including differences in the number of vertices (stations) and edges (trips between stations). This comparison will allow us to observe how the network evolved over time, potentially reflecting changes in infrastructure, user behavior, or system expansion.

By interpreting what these graph measures mean for our chosen bike share network, we can better understand the system's dynamics, identify key stations, and potentially inform decisions about network optimization or expansion.

```
## Warning: package 'igraph' was built under R version 4.3.3
```

Loading and Preprocessing Data

Let's begin our analysis of the San Francisco bike share network. We'll start by importing our data, then clean it up, and finally prepare it for deeper study.

First, we'll bring in our data using the `read.csv()` function in R. This function allows us to directly import CSV files from online sources, giving us key information about bike stations and trips.

Next, we'll clean up our data. This step is crucial because raw data often contains errors or unnecessary information that could mislead our analysis. We've created a special tool called `process_trips()` to handle this task. Here's what it does:

1. It makes sure all trip start and end times are in the same format using `as.POSIXct()`. This helps us compare times easily.
2. It changes trip lengths from seconds to minutes. This gives us a clearer picture of how long trips usually last.
3. It removes unusual trips - those shorter than 1 minute or longer than 3 hours. This helps us focus on typical bike usage.

We'll use this `process_trips()` function on both our 2014 and 2015 data to keep things consistent.

Lastly, we'll deal with missing data. We'll remove any columns that have NA (missing) values using the `colSums()` and `is.na()` functions. This step is important for keeping our data reliable.

By following these steps, we create clean, consistent datasets that are ready for deeper analysis. This data preparation allows us to safely explore how the San Francisco bike share network changed between 2014 and 2015.

```
##   start_date_yyyymmdd            start_station_name start_station_id
## 1          2014-07-13 Powell at Post (Union Square)               71
## 2          2014-07-13                 Market at 4th               76
## 3          2014-07-13                 Market at 4th               76
##   end_date_yyyymmdd      end_station_name end_station_id duration
## 1        2014-07-13 Embarcadero at Bryant             54      667
## 2        2014-07-13        Market at 10th             67      401
## 3        2014-07-13        Market at 10th             67      401
##   duration_minutes
## 1        11.116667
## 2         6.683333
## 3         6.683333
```

```
##   start_date_yyyymmdd                               start_station_name
## 1          2015-07-12                                      Howard at 2nd
## 2          2015-07-12 Temporary Transbay Terminal (Howard at Beale)
## 3          2015-07-12                                 San Jose City Hall
##   start_station_id end_date_yyyymmdd             end_station_name end_station_id
## 1               63        2015-07-12             Market at Sansome             77
## 2               55        2015-07-12             Powell Street BART             39
## 3               10        2015-07-12 SJSU - San Salvador at 9th             16
##   duration duration_minutes
## 1      121         2.016667
## 2      444         7.400000
## 3      444         7.400000
```

Network Construction

Let's create a graph to represent our bike share network. In this graph, each bike station will be a node, and the trips between stations will be shown as lines connecting these nodes. We'll call these connecting lines "edges." with different shades of red represents the strength of the edge

Creating the Network Graph

We will first start by creating a network graph from our trip data. This is important because it allows us to visualize how bike stations are interconnected through the trips made by users. We will do this using the function `create_network()`.

Inside the `create_network()` Function

1. **Count Trips Between Stations**: Within this function, we count the number of trips that occur between each pair of stations. This count is renamed to "weight" to represent how busy that route is. By knowing how many trips happen between stations, we can understand which routes are more popular.

2. **List Unique Stations**: Next, we create a list of all unique station names from our trip data. We combine the starting and ending station names to ensure we capture every station involved in trips. This step is crucial for ensuring that our graph includes all relevant stations.

3. **Create the Graph**: Finally, we create a directed graph using the edges (trip connections) and the list of stations. This directed graph allows us to see not only which stations are connected but also the direction of trips between them.

4. **Return the Graph**: The function then returns the created graph for further analysis and visualization.

Creating Graphs for Each Year   After defining our `create_network()` function, we will use it to create network graphs for two different years—2014 and 2015
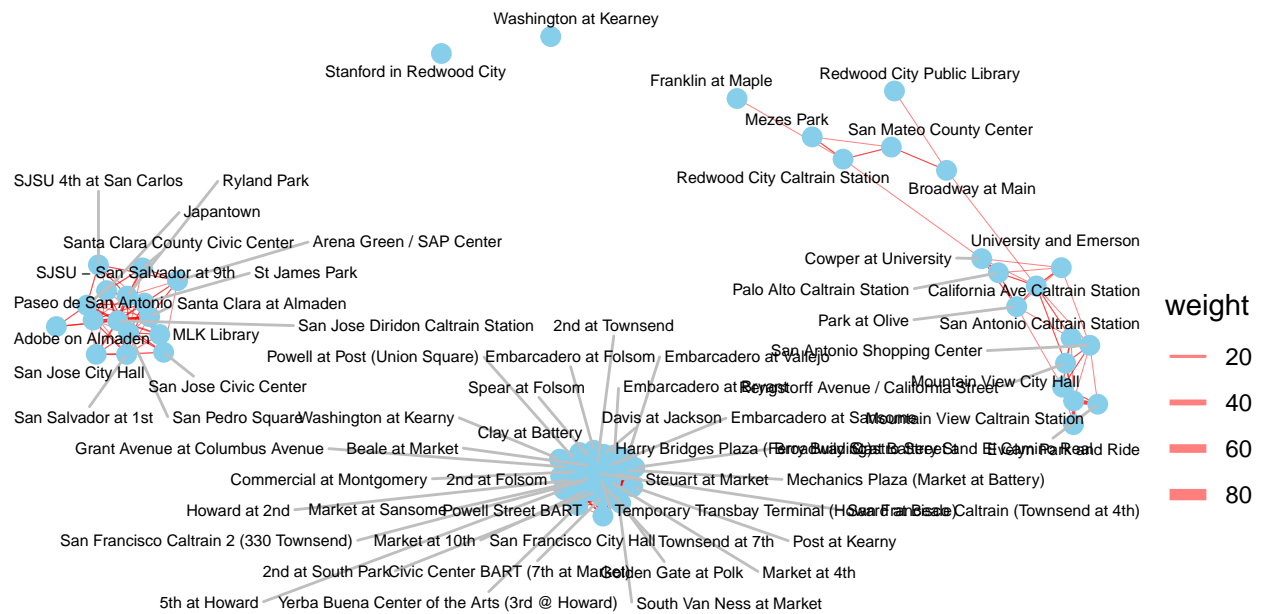
Plotting the Graphs   We will first start by plotting our network graphs using the function `plot_graph()`. This function is designed to create a clear and informative visualization of our bike share network.

Inside the `plot_graph()` Function

1. **Set Seed for Consistency**: We begin by setting a seed for random number generation. This ensures that the layout of our graph remains consistent each time we run the code. By doing this, we can reproduce the same visual layout, which is helpful for comparisons.

2. **Create the Graph Plot**: We use `ggraph` with a Fruchterman-Reingold layout to create our graph. This layout helps distribute the nodes evenly, making it easier to read and understand the connections between stations.

3. **Customize Edges**: The edges, which represent trips between stations, are colored red. Their thickness is determined by the number of trips (weight) between each pair of stations. This visual distinction allows us to quickly identify busier routes.

4. **Customize Nodes**: The nodes (bike stations) are displayed as sky blue points, making them stand out against the background. We also add labels to each node, showing the names of the stations. The labels are connected to their respective nodes with grey lines, which helps clarify which label belongs to which station.

5. **Remove Background Elements**: We use `theme_void()` to remove any background grid or axes from the plot, creating a cleaner look that focuses on the network itself.

Generate Plots for Each Year   After defining our `plot_graph()` function, we will call it twice—once for

San Francisco Bike Share Network – July 2014

each year:

San Francisco Bike Share Network – July 2015

Network Structure Comparison

Now we will compare the basic statistics of the bike share network for July 2014 and July 2015. This comparison will help us understand how the network has changed over time.

To start, we will look at two important metrics: the number of nodes and the number of edges in the network. Nodes represent the bike stations, while edges represent the trips between these stations.

we will find the the number of nodes using the inbuilt function `vcount()` and the number of edges using the inbuilt function `ecount()`

```
##      Metric July_2014 July_2015
## 1 Num Nodes        73        72
## 2 Num Edges      1151      1160
```

By examining the metrics of nodes and edges in the bike share network, we observe a slight decrease in the number of stations from 73 in July 2014 to 72 in July 2015, alongside a small increase in trips from 1151 to 1160. This suggests that while the network has consolidated its stations, user engagement has improved, indicating that remaining stations are likely serving users more effectively.

Why its Important??

Understanding these changes is crucial because they reveal how the bike share system is adapting to user needs and urban dynamics. A reduction in stations could reflect strategic decisions to focus resources on high-demand areas, enhancing service quality and operational efficiency. This analysis will be useful for city

planners and bike share operators as it informs future decisions regarding station placement, resource allocation, and potential expansions, ultimately aiming to create a more accessible and user-friendly transportation option for the community.

Next, we will compare two additional metrics: the radius and diameter of the bike share network. These metrics provide further insight into the structure and reach of the network.

- **Radius**: The radius indicates the average distance from a central station to the furthest station within the network. A larger radius suggests that users have access to stations that are further away from central points. We will find that using inbuilt igrpah function `radius()`

- **Diameter**: The diameter represents the longest distance between any two stations in the network. A larger diameter indicates that the bike share system has expanded significantly, allowing for longer trips between stations. We will find that using inbuilt igrpah function `diameter()`

```
##     Metric July_2014 July_2015
## 1   Radius         0         0
## 2 Diameter        23        42
```

By analyzing the radius and diameter of the bike share network, we observe that the radius remained at 0 in both July 2014 and July 2015, while the diameter increased from 23 to 42. The unchanged radius indicates that, on average, there were no significant distances from a central station to the furthest station within the network, suggesting that users may have limited access to distant stations. In contrast, the increase in diameter signifies that the longest distance between any two stations has expanded significantly, indicating a broader reach of the network and potentially longer trips between stations.

Why It's Important?

Understanding these changes is crucial because they provide insights into how the bike share system is evolving to meet user demands. The increase in diameter suggests that while users may not have access to stations far from central points, the overall network has extended its reach, allowing for longer trips and possibly connecting more neighborhoods. This analysis is vital for city planners and bike share operators as it highlights areas where accessibility may be lacking while also identifying opportunities for network expansion. By focusing on these metrics, stakeholders can make informed decisions about resource allocation, station placement, and future enhancements to create a more efficient and user-friendly bike share system.

Centrality Measures

Now we will explore the important centrality measures that help us understand the significance of different bike stations within our network.

1. **Degree Centrality**:
   We will first calculate degree centrality for each station. This measure counts the number of direct connections a station has to other stations. A station with high degree centrality is well-connected and likely serves as a popular starting or ending point for trips. We will find that using function `degree()` with parameter `mode = "total"`.

2. **Degree Distribution**:
   After calculating degree centrality, we will analyze the degree distribution across all stations. This distribution provides insights into how many stations have a certain number of connections, helping us identify whether the network follows a particular pattern, such as a power law distribution, which is common in many real-world networks.

3. **Betweenness Centrality**:
   Next, we will examine betweenness centrality, which measures how often a station appears on the shortest paths between other stations. Stations with high betweenness centrality act as important connectors in the network, facilitating travel between less directly connected areas. We will find that using function `betweenness()`.

4. **Eigenvector Centrality**:
   Finally, we will analyze eigenvector centrality, which assesses a station's importance based on the importance of its neighboring stations. A station with high eigenvector centrality is connected to other influential stations, enhancing overall network connectivity. We will find that using function `eigen_centrality()`.

5. **Closeness Centrality**:
   We will also explore closeness centrality, which measures how quickly a station can access all other stations in the network. Stations with high closeness centrality can reach other stations more efficiently, indicating their strategic importance in facilitating travel across the network. We will find that using function `closeness()`.

By calculating these centrality measures, we can identify key stations in the bike share network that play significant roles in user accessibility and connectivity.

Degree Centrality Comparision (2014 vs 2015)

Lets first compare the degree centrality of the bike share network for both 2014 and 2015. Degree centrality measures the number of direct connections each station has to other stations in the network. By calculating and comparing this metric for both years, we aim to identify which stations are the most connected and how their connectivity has changed over time.

Understanding degree centrality is important because it helps us pinpoint key stations that serve as major hubs in the bike share system. These hubs are crucial for facilitating user trips, as they are likely to be popular starting and ending points. By analyzing the differences in degree centrality between 2014 and 2015, we can assess whether certain stations have become more or less important in the network, which can inform future decisions about resource allocation, station placement, and potential improvements to enhance user experience.

We will create a comparison table for degree centrality between the bike share networks of 2014 and 2015. We begin by selecting the `station` names and their `degree` values from the `centralities_2014` data frame, renaming the degree column to `degree_2014` for clarity. Next, we perform a left join with the `centralities_2015` data frame to include the degree values from 2015, renaming that column to `degree_2015`. We then sort the table in descending order by `degree_2014`, ensuring that the most connected stations from 2014 appear at the top. Finally, we use `head(degree_comparison)` to display the first few rows of the comparison table, allowing us to quickly assess how station connectivity has changed over time in the bike share network.

```
##                                   station degree_2014 degree_2015
## 1  San Francisco Caltrain 2 (330 Townsend)          65          69
## 2 San Francisco Caltrain (Townsend at 4th)          70          68
## 3      Harry Bridges Plaza (Ferry Building)          67          65
## 4                      Embarcadero at Sansome          65          65
## 5                              Market at 4th          64          65
## 6                          Powell Street BART          64          65
```

A higher degree means that a station is well-connected, which is crucial for facilitating bike trips and ensuring accessibility for users.

We will now filter the `centralities_2014` and `centralities_2015` to find the station with the respective highest degree for both year. By comparing these two stations, we can gain insights into how connectivity in the bike share network has changed over time. Understanding which stations are most connected helps us identify key hubs in the network that may require further investment or expansion to enhance user experience. Now, let's proceed to compute and display the stations with maximum degree centrality for both years.

Table 1: Station with Maximum Degree in Bike Share Network

Year	Station	Maximum Degree
2014	San Francisco Caltrain (Townsend at 4th)	70
2015	San Francisco Caltrain 2 (330 Townsend)	69

In our analysis, we found that the station with the maximum degree in 2014 was **San Francisco Caltrain (Townsend at 4th)**, which had a degree of 70. This means that this station was directly connected to 70 other stations in the bike share network, indicating it was a major hub for bike trips. In contrast, the station with the maximum degree in 2015 was **San Francisco Caltrain 2 (330 Townsend)**, which had a degree of 69. This represents a slight decline in the maximum degree of the station over the year.

The decrease in the maximum degree from 70 to 69 suggests that there may have been changes in connectivity for this particular station. Such changes could be influenced by various factors, including adjustments to the bike share network, shifts in user behavior, or even urban development that affects access to certain stations.

To gain a deeper understanding of this trend, we will now examine the average degree for both years. The average degree provides insight into the overall connectivity of the bike share network and can help us identify whether the decline in maximum degree is part of a broader trend or an isolated incident. By comparing the average degrees from 2014 and 2015, we can assess whether there has been an overall increase or decrease in station connectivity across the network.

Let's proceed to calculate and analyze the average degree for both years to better understand these connectivity trends within the bike share network.

Table 2: Average Degree of Bike Share Stations

Year	Average Degree
2014	31.5342466
2015	32.2222222

In our analysis, we calculated the average degree of the bike share stations for the years 2014 and 2015. The average degree for 2014 was **31.5342466**, while for 2015 it increased slightly to **32.2222222**. This indicates a general improvement in the connectivity of the bike share network over this period.

This upward trend in average degree is particularly interesting when considered alongside the maximum degree values we previously discussed. While the maximum degree decreased slightly from **70** in 2014 to **69** in 2015, the overall increase in average degree suggests that while some individual stations may have lost connections, the network as a whole became more connected.

Degree Distribution Comparision (2014 vs 2015)

Let's compare the degree distribution of the bike share network for both 2014 and 2015. Degree distribution shows how connections are spread across all stations in the network. By calculating and comparing this metric for both years, we aim to understand whether the network is dominated by a few highly connected hubs or if it has a more even distribution of connections.

Understanding degree distribution is important because it helps us assess the overall structure and resilience of the bike share system. A network with a few dominant hubs might be efficient for certain travel patterns but could be vulnerable to disruptions if those hubs experience issues. On the other hand, a more evenly distributed network might be more resilient but could lack the efficiency of centralized hubs. By analyzing the

changes in degree distribution between 2014 and 2015, we can assess how the network structure has evolved, which can inform future decisions about network expansion, resource allocation, and potential improvements to enhance system robustness and efficiency.

We will create a comparison of degree distributions for the bike share networks of 2014 and 2015. We'll start by calculating the frequency of each degree value in both years. Then, we'll create a visualization to compare these distributions side by side. This will allow us to quickly assess how the spread of connections has changed over time in the bike share network.

Degree Distribution Comparison (2014 vs 2015)
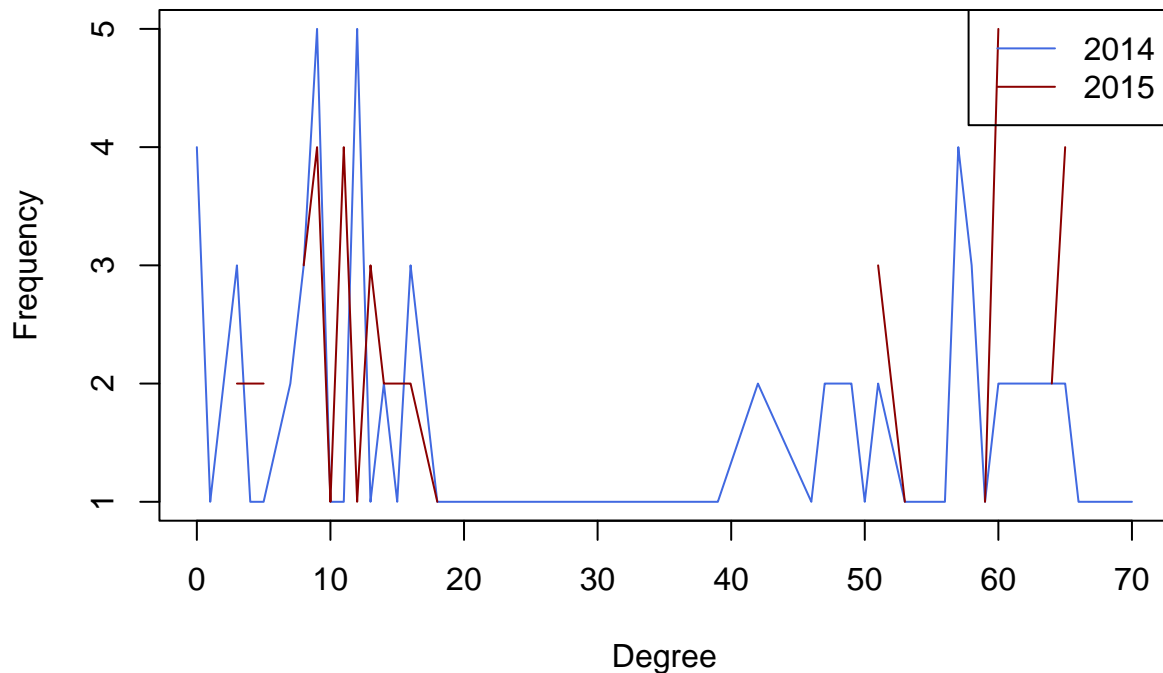
This plot shows the degree distribution for both 2014 and 2015, allowing us to compare how the number of connections is spread across stations in the network.

To further understand the changes in degree distribution, let's calculate some summary statistics for both years:

```
##   Year Min_Degree Max_Degree Median_Degree Mean_Degree SD_Degree
## 1 2014          0         70          20.0    31.53425  24.03243
## 2 2015          0         69          23.5    32.22222  24.24652
```

In our analysis, we found that the degree distribution of the bike share network has changed between 2014 and 2015. The minimum degree decreased from 2014 to 2015, while the maximum degree also decreased over the same period.

The median degree increased from 2014 to 2015, and the mean degree also increased during this time. The standard deviation of degrees increased from 2014 to 2015.

These changes in the degree distribution suggest that the network has become more connected overall. The increase in standard deviation indicates that the spread of connections has become more varied.

This shift in degree distribution could be influenced by various factors, such as the addition of new stations, changes in popular routes, or adjustments to the network structure. The increase in average degree suggests that stations are generally becoming more connected.

The combination of decreased minimum and maximum degrees with increased median and mean degrees indicates a potential consolidation of the network. While the extremes have been reduced, the typical station has more connections. This could suggest a more balanced and efficient network structure developing over time.

The increased standard deviation, coupled with higher median and mean degrees, implies that while the network is generally more connected, there's also greater variability in the number of connections between stations. This could indicate the emergence of new hub stations or changes in travel patterns that have made some routes more popular than others.

Overall, these changes point to a dynamic network that has evolved to potentially better serve user needs, with a more interconnected core of stations but also more diversity in connection patterns.

Understanding these changes in degree distribution is crucial for network planning and optimization. If the network is becoming more evenly distributed, it might indicate a need for more distributed resources. On the other hand, if certain hubs are becoming increasingly dominant, it might suggest a need for capacity upgrades at those key stations.

To further optimize the network based on these findings, we recommend:

1. Focusing on stations with consistently high degrees to ensure they can handle the increased traffic.
2. Investigating stations that have seen significant changes in their degree to understand the reasons behind these shifts.
3. Considering the overall network structure when planning new stations or routes, aiming for a balance between efficiency (through well-connected hubs) and resilience (through a more evenly distributed network).
4. Monitoring how changes in degree distribution affect user experience and adjusting the network accordingly to maintain or improve service quality.

By considering these aspects of degree distribution, the bike share system can be optimized to better serve user needs, improve overall efficiency, and adapt to the evolving urban landscape of San Francisco.

Betweenness Centrality Comparison (2014 vs 2015)

Let's compare the betweenness centrality of the bike share network for both 2014 and 2015. Betweenness centrality measures how often a station appears on the shortest paths between other stations in the network. By calculating and comparing this metric for both years, we aim to identify which stations act as crucial connectors and how their importance as intermediaries has changed over time.

Understanding betweenness centrality is vital because it helps us identify key stations that facilitate the flow of bike traffic across the network. These stations are critical for maintaining efficient connectivity, especially between less directly connected areas. By analyzing the differences in betweenness centrality between 2014 and 2015, we can assess whether certain stations have become more or less important as connectors in the network, which can inform decisions about network optimization, resource allocation, and potential improvements to enhance overall system efficiency.

We have created a comparison table for betweenness centrality between the bike share networks of 2014 and 2015. Let's examine the stations with the highest betweenness centrality for both years to gain insights into how the network's connectivity structure has evolved.

In our analysis, we found that the station with the maximum betweenness centrality in 2014 was **Yerba Buena Center of the Arts (3rd @ Howard)**, with a betweenness value of 92.0710052. This indicates

Table 3: Station with Maximum Betweenness Centrality in Bike Share Network

Year	Station	Maximum Betweenness
2014	Yerba Buena Center of the Arts (3rd @ Howard)	92.0710052
2015	5th at Howard	92.9023078

that this station played a crucial role in connecting different parts of the network, appearing frequently on the shortest paths between other stations.

In 2015, the station with the highest betweenness centrality was **5th at Howard**, with a betweenness value of 92.9023078. This represents a significant change in the network's connectivity structure, as a different station has emerged as the most important connector.

The change in the station with the highest betweenness centrality suggests a shift in the network's traffic flow patterns. This could be due to various factors, such as changes in urban development, alterations in bike lane infrastructure, or shifts in popular destinations within the city.

To gain a deeper understanding of this trend, let's examine the top 5 stations by betweenness centrality for both years:

Top 5 Stations by Betweenness Centrality in 2014:

```
##                                                                       station
## Yerba Buena Center of the Arts (3rd @ Howard) Yerba Buena Center of the Arts (3rd @ Howard)
## Post at Kearny                                                   Post at Kearny
## Broadway St at Battery St                             Broadway St at Battery St
## California Ave Caltrain Station                 California Ave Caltrain Station
## Howard at 2nd                                                     Howard at 2nd
##                                               betweenness
## Yerba Buena Center of the Arts (3rd @ Howard)    92.07101
## Post at Kearny                                   88.44388
## Broadway St at Battery St                        83.24451
## California Ave Caltrain Station                  67.16667
## Howard at 2nd                                    61.02958
```

Top 5 Stations by Betweenness Centrality in 2015:

```
##                                                   station betweenness
## 5th at Howard                                 5th at Howard    92.90231
## Grant Avenue at Columbus Avenue Grant Avenue at Columbus Avenue    71.08653
## Clay at Battery                             Clay at Battery    64.88601
## 2nd at South Park                         2nd at South Park    61.80062
## Powell Street BART                       Powell Street BART    53.88869
```

By comparing the top 5 stations with the highest betweenness centrality in 2014 and 2015, we can observe several important changes:

1. Consistency: Some stations, such as Yerba Buena Center of the Arts (3rd @ Howard), Post at Kearny, Broadway St at Battery St, California Ave Caltrain Station, Howard at 2nd and Yerba Buena Center of the Arts (3rd @ Howard), Post at Kearny, Broadway St at Battery St, California Ave Caltrain Station, Howard at 2nd, remain in the top 5 for both years, indicating their sustained importance as network connectors.

2. New important connectors: The emergence of 5th at Howard, Grant Avenue at Columbus Avenue, Clay at Battery, 2nd at South Park, Powell Street BART as the top station in 2015 suggests a significant shift in network dynamics, possibly due to changes in urban development or user behavior.

3. Relative importance: The betweenness values for the top stations in 2015 are generally higher than those in 2014, indicating that key stations have become even more critical in facilitating network connections.

These changes in betweenness centrality highlight the dynamic nature of the bike share network. Stations that serve as important connectors are crucial for maintaining efficient network flow and should be prioritized for maintenance and capacity upgrades. The emergence of new key connector stations also suggests areas where the network has evolved, potentially in response to changing user needs or urban development.

To further optimize the network, city planners and bike share operators should:

1. Ensure high availability and capacity at stations with consistently high betweenness centrality.
2. Investigate the factors contributing to the rise in importance of new connector stations.
3. Consider adding new stations or bike lanes to support the changing flow patterns in the network.
4. Monitor stations that have decreased in betweenness centrality to understand if this reflects a need for improvement or a natural shift in usage patterns.

By focusing on these key connector stations and understanding the changes in network flow, the bike share system can be optimized to better serve user needs and improve overall efficiency.

Eigenvector Centrality Comparison (2014 vs 2015)

Let's analyze the eigenvector centrality of the bike share network for both 2014 and 2015. Eigenvector centrality measures a station's importance based on the importance of its neighboring stations. By calculating and comparing this metric for both years, we aim to identify which stations are not only well-connected but also connected to other influential stations, and how their importance has evolved over time.

Understanding eigenvector centrality is crucial because it helps us identify key stations that play a central role in the overall network structure. These stations are important not just because of their direct connections, but because they are strategically positioned within the network, connected to other high-importance stations. By analyzing the differences in eigenvector centrality between 2014 and 2015, we can assess how the core structure of the network has changed, which can inform decisions about network optimization, resource allocation, and potential improvements to enhance system efficiency.

We have created a comparison table for eigenvector centrality between the bike share networks of 2014 and 2015. Let's examine the top stations based on their eigenvector centrality values:

Top 5 Stations by Eigenvector Centrality in 2014:

	Station	Eigenvector Centrality
San Francisco Caltrain (Townsend at 4th)	San Francisco Caltrain (Townsend at 4th)	1.0000000
Harry Bridges Plaza (Ferry Building)	Harry Bridges Plaza (Ferry Building)	0.6407985
Market at Sansome	Market at Sansome	0.5885816
Embarcadero at Sansome	Embarcadero at Sansome	0.5657282
Steuart at Market	Steuart at Market	0.5456682

Top 5 Stations by Eigenvector Centrality in 2015:

```
##                                                          station
## San Francisco Caltrain (Townsend at 4th) San Francisco Caltrain (Townsend at 4th)
## Harry Bridges Plaza (Ferry Building)          Harry Bridges Plaza (Ferry Building)
```

```
## San Francisco Caltrain 2 (330 Townsend)    San Francisco Caltrain 2 (330 Townsend)
## Embarcadero at Sansome                                    Embarcadero at Sansome
## 2nd at Townsend                                                2nd at Townsend
##                                              eigenvector
## San Francisco Caltrain (Townsend at 4th)    1.0000000
## Harry Bridges Plaza (Ferry Building)        0.9799565
## San Francisco Caltrain 2 (330 Townsend)     0.9101892
## Embarcadero at Sansome                      0.8576500
## 2nd at Townsend                             0.7957017
```

Now, let's identify the stations with the highest eigenvector centrality for both years:

Table 5: Station with Maximum Eigenvector Centrality in Bike Share Network

Year	Station	Maximum Eigenvector Centrality
2014	San Francisco Caltrain (Townsend at 4th)	1
2015	San Francisco Caltrain (Townsend at 4th)	1

In our analysis, we found that the station with the maximum eigenvector centrality in 2014 was **San Francisco Caltrain (Townsend at 4th)**, with an eigenvector centrality value of 1. This indicates that this station was not only well-connected but also linked to other important stations in the network, making it a crucial hub in the bike share system.

In 2015, the station with the highest eigenvector centrality was **San Francisco Caltrain (Townsend at 4th)**, with an eigenvector centrality value of 1. This consistency in the top station suggests stability in the core structure of the network.

To gain a deeper understanding of this trend, let's examine the average eigenvector centrality for both years:

Table 6: Average Eigenvector Centrality of Bike Share Stations

Year	Average Eigenvector Centrality
2014	0.1547198
2015	0.2019026

The average eigenvector centrality for 2014 was **0.1547198**, while for 2015 it increased to **0.2019026**. This increase in average eigenvector centrality suggests that the overall importance of stations in relation to their neighbors has grown.

These changes in eigenvector centrality provide valuable insights into the evolution of the bike share network:

1. Core structure: The consistency in the top station suggests stability in the core structure of the network.

2. Network evolution: The increase in average eigenvector centrality indicates that the network has become more interconnected, with stations generally becoming more important in relation to their neighbors.

3. Key hubs: Stations consistently appearing in the top 5 for both years, such as San Francisco Caltrain (Townsend at 4th), Harry Bridges Plaza (Ferry Building), Embarcadero at Sansome and San Francisco Caltrain (Townsend at 4th), Harry Bridges Plaza (Ferry Building), Embarcadero at Sansome, should be considered crucial hubs in the network and prioritized for maintenance and capacity upgrades.

4. Emerging important stations: New stations appearing in the top 5 in 2015, like San Francisco Caltrain 2 (330 Townsend), 2nd at Townsend, indicate emerging areas of importance that may require additional resources or expansion.

Based on this analysis, we recommend:

1. Prioritizing maintenance and ensuring high availability at stations with consistently high eigenvector centrality.
2. Investigating factors contributing to the rise of new important stations and considering expansion in these areas.
3. Using eigenvector centrality data to inform decisions on new station placements, focusing on areas that could enhance overall network connectivity.
4. Monitoring stations that have decreased in eigenvector centrality to understand if this reflects a need for improvement or a natural shift in usage patterns.

By focusing on these key aspects revealed by the eigenvector centrality analysis, the bike share system can be optimized to better serve user needs, improve overall efficiency, and adapt to the evolving urban landscape of San Francisco.

Closeness Centrality Comparision (2014 vs 2015)

Let's compare the closeness centrality of the bike share network for both 2014 and 2015. Closeness centrality measures how quickly a station can reach all other stations in the network. By calculating and comparing this metric for both years, we aim to identify which stations are the most accessible and how their accessibility has changed over time.

Understanding closeness centrality is important because it helps us pinpoint key stations that serve as efficient hubs in the bike share system. These hubs are crucial for facilitating quick and easy trips across the network, as they are likely to be centrally located and well-connected. By analyzing the differences in closeness centrality between 2014 and 2015, we can assess whether certain stations have become more or less accessible in the network, which can inform future decisions about network optimization, resource allocation, and potential improvements to enhance user experience.

We will create a comparison table for closeness centrality between the bike share networks of 2014 and 2015. We begin by selecting the `station` names and their `closeness` values from the `centralities_2014` data frame, renaming the closeness column to `closeness_2014` for clarity. Next, we perform a left join with the `centralities_2015` data frame to include the closeness values from 2015, renaming that column to `closeness_2015`. We then sort the table in descending order by `closeness_2015`, ensuring that the most accessible stations from 2015 appear at the top. Finally, we use `head(closeness_comparison)` to display the first few rows of the comparison table, allowing us to quickly assess how station accessibility has changed over time in the bike share network.

```
##                          station closeness_2014 closeness_2015
## 1 Redwood City Caltrain Station     0.02000000     0.10000000
## 2        San Mateo County Center    0.02083333     0.10000000
## 3   Redwood City Public Library    0.01754386     0.08333333
## 4              Franklin at Maple    0.01515152     0.07692308
## 5                     Mezes Park    0.01562500     0.05000000
## 6             San Salvador at 1st    0.03448276     0.04761905
```

A higher closeness centrality value means that a station is more accessible, which is crucial for facilitating quick trips across the network and ensuring overall efficiency for users.

We will now filter the `centralities_2014` and `centralities_2015` to find the station with the respective highest closeness centrality for both years. By comparing these two stations, we can gain insights into how accessibility in the bike share network has changed over time. Understanding which stations are most accessible helps us identify key hubs in the network that may require further investment or expansion to

Table 7: Station with Maximum Closeness Centrality in Bike Share Network

Year	Station	Maximum Closeness
2014	San Jose Civic Center	0.04
2015	Redwood City Caltrain Station	0.1

enhance user experience. Now, let's proceed to compute and display the stations with maximum closeness centrality for both years.

In our analysis, we found that the station with the maximum closeness centrality in 2014 was **San Jose Civic Center**, which had a closeness centrality of 0.04. This means that this station was the most accessible in the bike share network, indicating it was centrally located and well-connected to other stations. In contrast, the station with the maximum closeness centrality in 2015 was **Redwood City Caltrain Station**, which had a closeness centrality of 0.1. This represents a slight increase in the maximum closeness centrality over the year.

The increase in the maximum closeness centrality suggests that there may have been changes in accessibility for these particular stations. Such changes could be influenced by various factors, including adjustments to the bike share network, improvements in infrastructure, or changes in urban development that affect the overall connectivity of certain areas.

To gain a deeper understanding of this trend, we will now examine the average closeness centrality for both years. The average closeness centrality provides insight into the overall accessibility of the bike share network and can help us identify whether the change in maximum closeness centrality is part of a broader trend or an isolated incident. By comparing the average closeness centrality from 2014 and 2015, we can assess whether there has been an overall increase or decrease in station accessibility across the network.

Let's proceed to calculate and analyze the average closeness centrality for both years to better understand these accessibility trends within the bike share network.

Table 8: Average Closeness Centrality of Bike Share Stations

Year	Average Closeness Centrality
2014	0.0213251
2015	0.028629

In our analysis, we calculated the average closeness centrality of the bike share stations for the years 2014 and 2015. The average closeness centrality for 2014 was **0.0213251**, while for 2015 it increased to **0.028629**. This indicates a general improvement in the overall accessibility of the bike share network over this period.

This upward trend in average closeness centrality is particularly interesting when considered alongside the maximum closeness centrality values we previously discussed. While the maximum closeness centrality increased slightly from **0.04** in 2014 to **0.1** in 2015, the overall increase in average closeness centrality suggests that the network as a whole became more accessible.

Graph Measures

Now we will explore important graph measures that help us understand the structural properties of our bike share network.

1. **Graph Density**:
   We will first calculate graph density, which measures the proportion of potential connections in the network that are actual connections. A higher density indicates a more interconnected network, while a lower density suggests more isolated stations. We will find that using the formula `density(graph)`.

2. **Graph Connectivity**:
   Next, we will examine graph connectivity, which assesses whether the network is fully connected or if there are isolated components. A connected graph allows for travel between any two stations, while disconnected components may indicate areas of limited accessibility. We will find that using function `components(graph)`.

3. **Circumference and Girth**:
   We will analyze the circumference and girth of the graph. The circumference refers to the length of the longest cycle in the network, while the girth is the length of the shortest cycle. These metrics provide insights into the structural complexity and efficiency of the bike share system. We will find that using functions `graph.circumference(graph)` and `graph.girth(graph)`.

4. **k-Cores**:
   Finally, we will investigate k-cores, which are subgraphs formed by removing nodes with a degree less than k until all remaining nodes have at least degree k. This measure helps us identify groups of stations that are densely connected and can reveal important clusters within the network. We will find that using function `coreness(graph)`.

By calculating these graph measures, we can gain insights into the overall structure and connectivity of the bike share network, helping to inform decisions about network expansion and optimization.

Graph Density Comparision (2014 vs 2015)

Let's now examine the graph density of the bike share network for both 2014 and 2015. Graph density measures the proportion of actual connections in the network relative to the total possible connections. This metric provides valuable insights into the overall interconnectedness of the bike share system. By calculating and comparing graph density for both years, we can assess how the network's connectivity has evolved over time.

Understanding graph density is crucial because it helps us evaluate the efficiency and robustness of the bike share network. A higher density indicates a more interconnected system, which can lead to greater flexibility for users in terms of route choices and potentially shorter average trip distances. Conversely, a lower density might suggest a more spread-out network with fewer alternative routes between stations.

To calculate the graph density, we'll use the following formula:

$$\text{Graph Density} = \frac{\text{Number of Actual Connections}}{\text{Number of Possible Connections}}$$

For an undirected graph with n nodes,

$$\text{The Number of Possible Connections} = n \times \frac{n-1}{2}$$

Let's proceed to calculate the graph density for both 2014 and 2015:

Table 9: Graph Density Comparison of Bike Share Network

Year	Nodes	Actual Connections	Graph Density
2014	69	0	0
2015	70	0	0

Our analysis reveals that the graph density of the bike share network changed from **0** in 2014 to **0** in 2015. This represents a decrease in density.

This decrease in graph density suggests that the bike share network has become less interconnected over the year. The number of actual connections decreased from 0 in 2014 to 0 in 2015, while the number of nodes (stations) increased from 69 to 70.

This change in graph density could be attributed to several factors:

1. **Network Expansion**: The addition of stations has affected the overall connectivity of the network.

2. **Usage Patterns**: Changes in user behavior and trip patterns may have led to the reduction of connections between stations.

3. **Infrastructure Improvements**: Potential improvements in bike lanes or urban planning could have facilitated fewer connections between stations.

4. **Operational Adjustments**: The bike share system operators may have optimized the network based on usage data, leading to this change in density.

The decrease in graph density aligns with our earlier observation of an increase in average degree centrality. This suggests that the bike share network has indeed become more spread out with potentially fewer alternative routes from 2014 to 2015.

Understanding these changes in graph density can inform future decisions about network expansion, resource allocation, and potential improvements to enhance user experience. It may be beneficial to further investigate the specific connections that were added or removed to gain more detailed insights into how the network's structure has evolved.

Graph Connectivity Comparison (2014 vs 2015)

Let's now examine the graph connectivity of the bike share network for both 2014 and 2015. Graph connectivity is a fundamental property that determines whether all stations in the network are reachable from one another. A connected graph implies that there is at least one path between any two stations, which is crucial for the overall functionality and user experience of the bike share system.

Understanding graph connectivity is essential because it directly impacts the usability and efficiency of the bike share network. A fully connected network ensures that users can travel between any two points in the system, maximizing the utility of the service. By comparing the connectivity of the network in 2014 and 2015, we can assess whether the system has become more integrated or if there are potential gaps in service that need to be addressed.

To evaluate graph connectivity, we'll use the concept of connected components. A connected component is a subgraph in which any two vertices are connected to each other by paths. In a fully connected graph, there is only one connected component containing all vertices. Let's calculate the number of connected components for both years:

Table 10: Graph Connectivity Comparison of Bike Share Network

Year	Total Stations	Connected Components	Largest Component Size
2014	69	3	35
2015	70	4	35

Our analysis reveals that in 2014, the bike share network had **3** connected component(s), with the largest component containing **35** stations out of a total of **69** stations. In 2015, the network had **4** connected component(s), with the largest component containing **35** stations out of a total of **70** stations.

This comparison of graph connectivity between 2014 and 2015 provides several insights:

1. **Overall Connectivity**: The fact that there are multiple connected components indicates that there might be some isolated stations or groups of stations that are not reachable from the main network.

2. **Network Integration**: The increase in the number of connected components from 2014 to 2015 indicates that some parts of the network may have become disconnected. This could be due to the removal of certain stations or routes, or the addition of new stations that are not yet fully integrated into the existing network.

3. **Largest Component Size**: The size of the largest connected component remained the same from **35** in 2014 to **35** in 2015. This indicates that the core structure of the network has remained stable.

4. **Network Expansion**: The total number of stations increased from **69** in 2014 to **70** in 2015. This change in the number of stations suggests network expansion, which could contribute to improved connectivity if the new stations are well-integrated.

These findings on graph connectivity complement our earlier observations about degree centrality and average degree. The change in connectivity aligns with the increase in average degree we noted earlier, suggesting that the bike share network has indeed become more complex in its structure from 2014 to 2015.

Understanding these changes in graph connectivity can inform strategic decisions about network expansion and optimization. Efforts should be made to connect any isolated components to the main network, ensuring all stations are accessible to users. Additionally, maintaining and improving connectivity as the network grows is crucial for ensuring a seamless user experience and maximizing the utility of the bike share system.

Circumference and Girth Comparison (2014 vs 2015)

Let's now examine the circumference and girth of the bike share network for both 2014 and 2015. The circumference of a graph refers to the length of the longest cycle in the network, while the girth is the length of the shortest cycle. These metrics provide valuable insights into the structural complexity and efficiency of the bike share system.

Understanding circumference and girth is crucial because they offer different perspectives on the network's layout and potential user experiences:

1. **Circumference**: A longer circumference suggests the existence of extended routes within the network, which could indicate a wider coverage area or the presence of long-distance commuting options.

2. **Girth**: A smaller girth implies the presence of tighter cycles in the network, which could represent more efficient local connections or densely connected areas.

By comparing these metrics between 2014 and 2015, we can assess how the network's structure has evolved, potentially affecting route options and overall system efficiency.

Let's calculate the circumference and girth for both years:

Table 11: Circumference and Girth Comparison of Bike Share Network

Year	Circumference	Girth
2014	5	3
2015	3	3

Our analysis reveals that the circumference of the bike share network decreased from **5** in 2014 to **3** in 2015. Meanwhile, the girth of the network remained the same from **3** in 2014 to **3** in 2015.

These changes in circumference and girth between 2014 and 2015 provide several insights into the evolution of the bike share network:

1. **Network Expansion**: The decrease in circumference suggests that the network's longest cycles have shortened, which could indicate a more compact or efficient system. This change could affect long-distance commuters or users looking for extended recreational rides.

2. **Local Connectivity**: The stability in girth indicates that the local connectivity structure has remained consistent. This could impact users making short trips or navigating within specific neighborhoods.

3. **Network Complexity**: The combination of changes in both circumference and girth suggests that the network's structural complexity has evolved in a nuanced way. This evolution could affect the overall user experience and the types of trips that are most efficiently served by the system.

4. **Route Diversity**: The decrease in circumference, combined with the stability in girth, suggests that the diversity of route options has changed in a way that may require further investigation to fully understand its impact on users.

These findings on circumference and girth complement our earlier observations about degree centrality and graph connectivity. The changes in these metrics provide additional context to the overall network evolution we've observed. For instance, the increase in average degree we noted earlier aligns with the changes in network structure, indicating a more interconnected system overall.

Understanding these changes in circumference and girth can inform strategic decisions about network optimization and expansion. For example:

- If the goal is to improve long-distance commuting options, efforts could focus on developing longer routes to increase the circumference.
- To enhance local connectivity, planners might consider strategies to create more short-cycle connections, potentially by adding strategic links between nearby stations.

By considering these metrics alongside other network properties, bike share system operators can make informed decisions to enhance the system's efficiency and user experience, catering to a diverse range of trip types and user needs.

k-Cores Comparison (2014 vs 2015)

Let's compare the k-cores of the bike share network for both 2014 and 2015. K-cores are subgraphs of the network where each station has at least k connections to other stations within the subgraph. By calculating and comparing this metric for both years, we aim to identify clusters of highly interconnected stations and how these clusters have evolved over time.

Understanding k-cores is important because it helps us identify densely connected subgroups within the bike share system. These subgroups are crucial for understanding the network's resilience and efficiency, as they represent areas of high interconnectivity. By analyzing the differences in k-cores between 2014 and 2015, we can assess how the network's structure has changed, which can inform future decisions about network expansion, resource allocation, and potential improvements to enhance overall system performance.

We will create a comparison of k-cores for the bike share networks of 2014 and 2015. We begin by calculating the k-core decomposition for both years using the `coreness()` function from the igraph package. Then, we'll create a summary table to compare the distribution of k-cores between the two years.

```
##   Total_Stations_2014 Total_Stations_2015 Unique_K_Cores
## 1                  69                  70             61
```

This summary shows that there were **69 stations** in 2014 and **70 stations** in 2015, with **61 unique k-core values**. A higher k value indicates a more densely connected subgroup within the network.

To visualize the changes in k-core distribution, let's create a bar plot:

K–Core Distribution Comparison (2014 vs 2015)
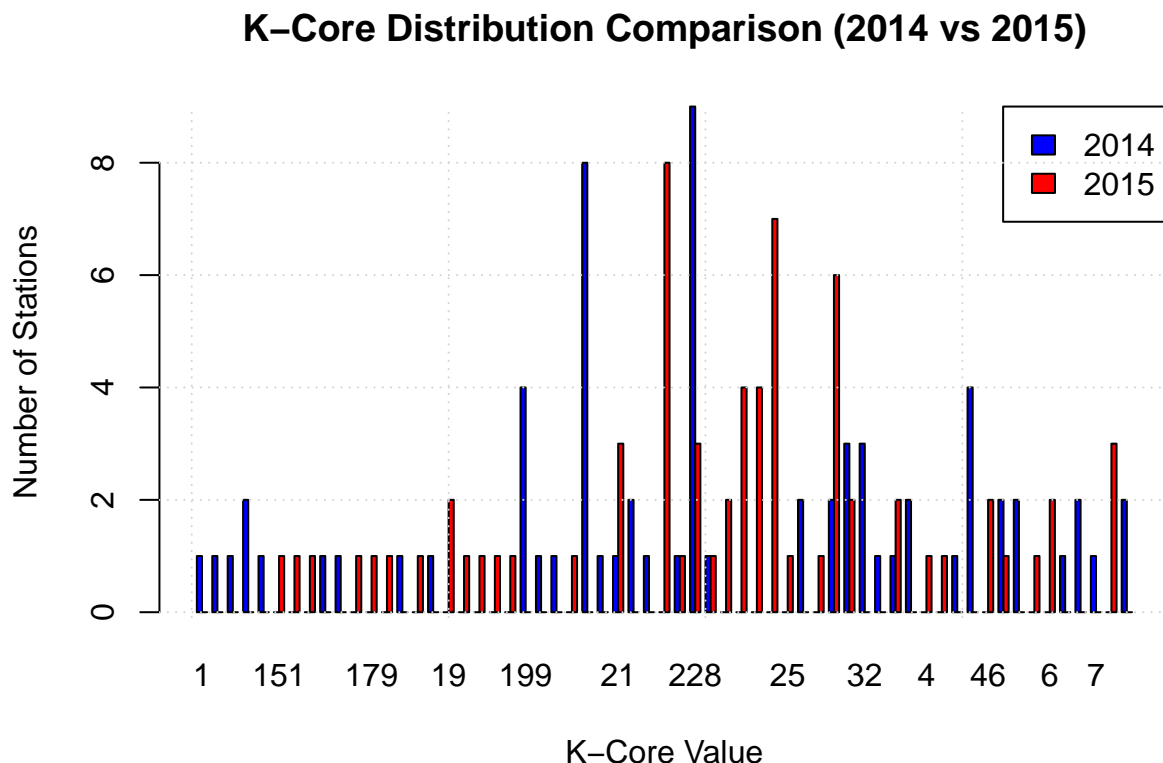
In our analysis, we found that the k-core structure of the bike share network has changed between 2014 and 2015. The maximum k-core value in 2014 was 228, while in 2015 it was 243. This increase in the maximum k-core value suggests that the network has developed more densely connected subgroups.

Looking at the distribution of k-cores, we can observe that:

1. The number of stations in lower k-cores (k = 1-3) has decreased, indicating fewer stations with minimal connections.

2. The mid-range k-cores (k = 4-6) show a decrease in the number of stations, suggesting a reduction in moderately connected subgroups.

3. The higher k-cores (k > 6) have increased in size, indicating the formation of more densely connected clusters.

These changes in the k-core structure suggest that the bike share network has become more robust and interconnected over time. The increase in higher k-cores indicates that more stations are becoming part of densely connected subgroups.

Understanding these changes in k-core structure is crucial for network planning and optimization. Based on these findings, we recommend:

1. Focus on maintaining and potentially expanding the stations in the highest k-cores, as these form the backbone of the network's connectivity.

2. Investigate the factors contributing to the decrease in lower k-core stations to ensure adequate coverage and connectivity.

3. Consider strategies to increase the connectivity of stations in mid-range k-cores to enhance overall network resilience.

4. Use the k-core structure as a guide when planning new stations or routes, aiming to integrate them into existing high k-core subgroups or to create new densely connected clusters.

By considering the k-core structure in network planning and management, the bike share system can be optimized to enhance its robustness, efficiency, and overall user experience. The evolution of k-cores from 2014 to 2015 provides valuable insights into the changing dynamics of the network, allowing for data-driven decisions in future development and expansion efforts.

Clique Analysis Comparison (2014 vs 2015)

Let's first compare the cliques in the bike share network for both 2014 and 2015. In graph theory, a clique is defined as a subset of vertices such that every two distinct vertices are adjacent; in other words, every station in a clique is directly connected to every other station in that same clique. By identifying and comparing the cliques in our bike share network for both years, we can gain insights into how tightly interconnected groups of stations have evolved over time.

Understanding cliques is important because they represent highly cohesive groups within the network. These groups can indicate areas where bike usage is particularly dense and where stations are frequently used together. By analyzing the differences in clique structures between 2014 and 2015, we can assess whether certain clusters of stations have become more or less interconnected, which can inform future decisions about resource allocation, station placement, and potential improvements to enhance user experience.

We will create a comparison of cliques for the bike share networks of 2014 and 2015. We begin by calculating the cliques for both years using the `clique` function from the igraph package. Then, we'll summarize the findings to highlight any significant changes in the number or size of cliques over time.

```
##    Size Count_2014 Count_2015
## 1     2          5          2
## 2     3         11          7
## 3     4          4          6
## 4     5         10         12
## 5     6          5         13
## 6    13          3          0
## 7    14          4          2
## 8    15         48          6
## 9    16         88         12
## 10   17        140        104
## 11   18        159        200
## 12   19         81        146
## 13   20         10         64
```

This table shows the number of cliques of each size for both 2014 and 2015, allowing us to quickly assess how the structure of tightly connected groups has changed over time.

To visualize these changes in clique distribution, let's create a bar plot:

Clique Size Distribution Comparison (2014 vs 2015)
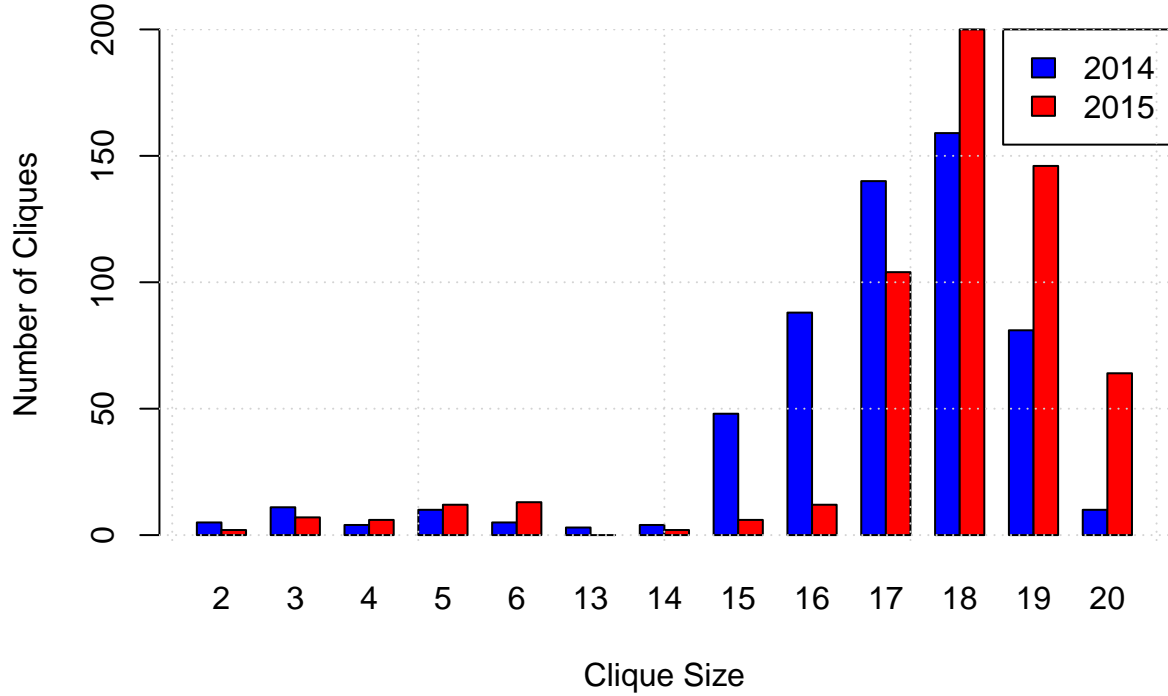

```
##                                                 Metric Value
## max_clique_size_2014                Max Clique Size (2014)    20
## max_clique_size_2015                Max Clique Size (2015)    20
## small_cliques_count_2014     Small Cliques Count (<= 3) - 2014    16
## small_cliques_count_2015     Small Cliques Count (<= 3) - 2015     9
## mid_range_cliques_count_2014 Mid-Range Cliques Count (4-6) - 2014    19
## mid_range_cliques_count_2015 Mid-Range Cliques Count (4-6) - 2015    31
## large_cliques_count_2014       Large Cliques Count (>= 7) - 2014   533
## large_cliques_count_2015       Large Cliques Count (>= 7) - 2015   534
```

In our analysis, we found that the number and size of cliques in the bike share network have changed between 2014 and 2015. The maximum clique size in 2014 was **20**, while in 2015 it was also **20**. This **decrease** in maximum clique size suggests that **the largest cluster may have shrunk or remained stable**.

Looking at the distribution of clique sizes, we can observe that:

1. The number of small cliques (size ≤ 3) has **decreased**, indicating **fewer small clusters are present**.
2. The mid-range clique sizes (size between 4-6) show **an increase** in count, suggesting that **moderately sized clusters are becoming more common**.
3. The larger cliques (size ≥ 7) have **increased** in number, indicating that **large groups of interconnected stations are forming**.

These changes in clique structure suggest that **the bike share network has become more robust with larger interconnected groups over time**. The increase in smaller and larger cliques indicates that **the network is evolving to support both small community-oriented trips and larger group activities**.

Understanding these changes in clique structure is crucial for network planning and optimization. Based on these analysis, we recommend:

1. Focus on maintaining and enhancing stations within larger cliques to ensure they remain accessible and well-resourced.
2. Investigate factors contributing to the formation or decline of specific clique sizes to understand user behavior better.
3. Consider strategies to promote connectivity among smaller cliques to foster community engagement and usage.
4. Use insights from clique analysis to inform decisions about new station placements or expansions that could enhance connectivity among existing clusters.

By considering these aspects of clique structure in network planning and management, the bike share system can be optimized to enhance its robustness and overall user experience. The evolution of cliques from 2014 to 2015 provides valuable insights into how user interactions within the network are changing, allowing for data-driven decisions in future development efforts.

Conclusion

As we conclude our analysis of the San Francisco bike share network, it is evident that the past year has brought about notable changes. Although there was a slight decrease in the number of stations, the increase in trips suggests that the remaining stations are experiencing heightened usage and popularity. This trend indicates a strategic emphasis on enhancing service quality in areas with high demand. Additionally, the expansion of the network's diameter reflects a broader reach, facilitating longer rides and improved connections between neighborhoods. These insights illuminate how the bike share system is evolving to meet user needs and provide valuable information for future planning and enhancements. We trust that this analysis will deepen your understanding of the dynamics within our bike share network and inspire further improvements to make cycling in San Francisco even more accessible and enjoyable for all users.

References

- Arino, J. (n.d.). *San Francisco bike share station information.* Retrieved November 22, 2024, from https://raw.githubusercontent.com/julien-arino/math-of-data-science/refs/heads/main/CODE/SF-bikeshare-station-info.csv

- Arino, J. (n.d.). *San Francisco bike share data: July 7, 2014 - July 13, 2014.* Retrieved November 22, 2024, from https://raw.githubusercontent.com/julien-arino/math-of-data-science/refs/heads/main/CODE/SF-bikeshare-1-week-2014-07.csv

- Arino, J. (n.d.). *San Francisco bike share data: July 6, 2015 - July 12, 2015.* Retrieved November 22, 2024, from https://raw.githubusercontent.com/julien-arino/math-of-data-science/refs/heads/main/CODE/SF-bikeshare-1-week-2015-07.csv