Optimizing Mobi Vancouver's Bike Share Network

Angelina Dubrule (82317231), Sumer Mann (65751364),
Richard Pillaca (96474556), and Sawneet Sood (29247657)

University of British Columbia, Okanagan Campus, 3333 University Way,
Kelowna, BC V1V 1V7, Canada
contact@ubc.ca

Abstract: Bike-sharing systems are integral to promoting sustainable urban mobility, yet they face significant challenges in resource allocation and network efficiency. This research focuses on optimizing Mobi Vancouver's bike-share network by analyzing ridership data to address inefficiencies in bike distribution and rebalancing strategies. The study aims to identify major hubs, evaluate the impact of potential disruptions, and propose data-driven recommendations to enhance system operations and connectivity.

A graph-theory-based approach was employed to analyze network metrics, including in/out-degree distributions, betweenness centrality, and closeness centrality, to pinpoint critical stations and key traffic flows. Hierarchical clustering techniques were applied to reveal functional station groups, simplifying the network's complexity for targeted interventions. The analysis used one month of data from a seven-year period to capture seasonal patterns while ensuring computational efficiency.

Preliminary results indicate significant variability in station centrality, with key hubs exhibiting high betweenness and closeness centrality. Core clusters of 84, 80, 74, and 26 stations were identified as essential network components. These findings underscore the potential for tailored rebalancing strategies and improved resource allocation across the system. The study provides actionable recommendations for optimizing bike-share operations and establishes a scalable framework applicable to other urban networks. By identifying stations with variable usage patterns and areas prone to congestion, this research supports predictive modeling to anticipate demand and identify underutilized stations. Future work will explore additional metrics, such as eigenvector centrality, to further enhance system scalability and sustainability.

Keywords: Bike-sharing, Urban mobility, Network optimization, Graph theory, Traffic flow analysis, Mobi Vancouver, Data analysis

1 INTRODUCTION

1.1 Background and Context

As cities grow, sustainable urban transportation becomes a critical challenge, and bike-sharing systems play a pivotal role in addressing this need. These systems offer a low-cost, eco-friendly alternative to traditional transportation, promoting health and reducing traffic congestion. Vancouver's Mobi bike-share network exemplifies this trend, serving as an integral component of urban mobility. However, like many bike-share systems, it faces challenges in resource allocation, network efficiency, and operational sustainability. According to a recent report, global bike-share usage increased by over 20% in the past decade, yet many systems struggle with uneven demand and station imbalances, leading to inefficiencies and user dissatisfaction. Addressing these challenges is crucial to maximizing the potential of bike-sharing systems in modern cities.

1.2 Problem Statement

Despite the widespread adoption of bike-sharing systems, current research has focused predominantly on high-level analyses, leaving a gap in localized, data-driven strategies for optimizing network performance. Specifically, there is limited work on integrating graph theory and clustering to identify critical stations and traffic flow patterns, which are essential for efficient resource allocation and rebalancing strategies. This research aims to bridge this gap by focusing on Vancouver's Mobi bike-share network to provide actionable insights.

1.3 Research Questions

1. Which stations serve as major hubs, and how does their connectivity influence overall network efficiency?
2. How can data-driven strategies optimize bike distribution and reduce congestion?
3. What is the impact of transit frequency variations on resource allocation?
4. How can clustering and predictive modelling enhance the system's scalability and operational performance?

1.4 Objectives and Goals

The primary objective of this research is to optimize the Mobi Vancouver bike-share network by leveraging graph theory and clustering methods. This involves:

1. *Analyzing major hubs*: Identify critical stations using centrality metrics (e.g., degree, betweenness, closeness) and evaluate their impact on network performance.
2. *Optimizing resource allocation*: Develop rebalancing strategies to address congestion and underutilized stations based on transit frequency and demand patterns.
3. *Improving operational efficiency*: Propose data-driven recommendations to enhance urban mobility and align the network with user needs.
4. *Building predictive models*: Lay the groundwork for forecasting demand changes and suggesting infrastructure adjustments for future growth.

1.5 Significant Contribution

This study offers significant advancements by integrating graph-theory analysis, clustering, and predictive modelling to enhance the efficiency and sustainability of Vancouver's Mobi bike-share system. Utilizing centrality metrics, it identifies critical stations that function as major hubs within the network, enabling targeted resource allocation and the development of rebalancing strategies to alleviate station congestion and underutilization. This data-driven approach not only improves operational efficiency and user experience but also provides actionable insights for urban planners and bike-share operators, facilitating informed decision-making and strategic infrastructure enhancements. Additionally, the research establishes a scalable framework adaptable to other urban bike-share networks, thereby advancing academic knowledge in network optimization and offering practical solutions for sustainable urban mobility on a global scale. Addressing the lack of analysis, modelling, and optimization research on Vancouver's Mobi bike-share system, this study supports the city's efforts to promote eco-friendly transportation, fosters interdisciplinary collaboration, and sets a precedent for future investigations aimed at optimizing bike-share systems in diverse urban environments.

2 Related Work

Although research on optimizing bike-sharing systems exists, Mobi Vancouver's network lacks comprehensive data-driven optimization. This study addresses that gap, offering tailored insights and a scalable framework for similar networks in British Columbia.

Chiariotti et al. (2018) developed a dynamic rebalancing approach for bike-sharing systems using predictive models and graph theory. Tested on New York City's network, their framework optimized redistribution routes based on historical data, enhancing efficiency over static methods. This aligns with our focus on optimizing Vancouver's Mobi system, emphasizing centrality metrics and clustering to manage congestion and improve resource allocation.

In their study on inventory rebalancing and vehicle routing for bike-sharing systems, Schuijbroek, Hampshire, and van Hoeve (2017) address the operational challenge of efficiently redistributing bikes by integrating inventory management with optimized vehicle routing. Their model adapts to dynamic demand and station capacities, reducing imbalances and operational costs. This approach underscores the value of data-driven strategies, offering insights applicable to enhancing Vancouver's Mobi network through improved resource allocation and service reliability.

Adham et. al.(2016) evaluated various clustering methods within the AEA(Artificial Ecosystem Algorithm) to enhance the bike redistribution in London's Bike Sharing Scheme. This study provided a dynamic framework that improved upon traditional methods by redistributing bikes to meet user demands. This approach was well aligned with our focus on optimizing Vancouver's Mobi Network, focusing on utilizing community detection and optimizing resource allocation.

Soto and Ceballos(2024) present a stochastic agent-based model whose goal is to optimize bike share systems. They use agent behaviour to investigate the search space by employing four behavioural options. There is collaboration among agents and it was tested on real-world data. The model demonstrated significant potential in improving system efficiency and handling any uncertainties. The study focused on the importance of agent-based techniques and stochastic optimization to address the difficulties of bike share rebalancing.

Salih-Elamin and Al-Deek (2020) developed a methodology that integrates the Maximal Covering Location Problem (MCLP) with centrality measures—closeness, betweenness, and degree—to optimize bike station locations for maximum demand coverage. Applied to Washington, D.C.'s Capital Bike Share System, their approach significantly improved network efficiency by prioritizing key stations based on centrality. This methodology parallels our research on Vancouver's Mobi network, where we utilize centrality and clustering to enhance station connectivity and optimize resource allocation.

Wang and Cullinane (2016) adapted Freeman's centrality measures to evaluate port importance in maritime container transportation, integrating factors such as shipping capacity and foreland market coverage. Their findings emphasize the significant role of market coverage in enhancing closeness centrality, which parallels our focus on utilizing centrality metrics to improve station connectivity and accessibility within Vancouver's Mobi network. This study provides a robust framework for understanding how external factors influence centrality, thereby informing our strategies for optimizing bike-share station placement.

3 Methodology

3.1 Data Filtering and Collection

To start analyzing the mobi-cycle network, data is first extracted and refined from the Mobi by Rogers Website for September 2024. Unique station identifiers were collected and standardized to 4-digit codes to unify the station representation and minimize discrepancies. Columns 'Departure Station' and 'Return Station' are extracted and filtered to remove non-informative and invalid entries such as empty fields and placeholder values like '0000' to ensure data accuracy for analysis. This resulted in a clean dataset of distinct and consistent nodes. The edges dataset contained the data for trips between a pair of stations. Each trip is a directed edge from the 'departure station' to the 'return station'. For edges, a similar procedure is followed to extract consistent station codes. Finally, this filtered dataset is used to make a weighted edges dataset by counting the frequency of trips between each unique pair of stations. This provided observations on connections and routes that were frequently used for further analysis.

3.2 Hierarchical Clustering

With the filtered dataset of nodes and weighted edges, hierarchical clustering can be used to identify natural grouping within the whole network. This further simplified the complexity of the entire network into more manageable directed, weighted sub-networks. First, the graph of the network is made using the igraph library in R. An adjacency matrix is generated from the graph which is further converted into a distance matrix using the inverse of weights, assigning smaller distances to the edges with more frequency of trips. Ward's method was used for clustering which minimized the variance within the clusters, providing well-defined groups in a complex network(Adham & Bently, 2016). Nodes and edges for each cluster were saved in separate CSV files for independent analysis. For analysis, graphs and dendrograms were generated for each cluster and the whole network. This helped in further understanding the overall network structure, and cluster relationships, and identifying major stations in each cluster.

3.3 Betweenness Centrality

The betweenness centrality analysis of the network focused on identifying the key stations that serve as major intermediaries in the system. The data was imported as CSV files and using igraph in Rstudio, graphs were created. A graph of the whole network was made which showed the betweenness centrality. The results for betweenness centrality highlighted the main stations that acted as bridges. The top 5 stations along with the average of the whole network was computed. Stations with high betweenness values indicated areas with importance for maintaining connectivity. Nodes with high centrality values are mostly points that have higher congestion and hence they can be seen as stations that could use changes for improvement. The in depth analysis of betweenness centrality using Rstudio helped in identifying areas that could use improved connectivity to make the network flow much smoother and strengthen the network.

3.4 Eigenvector Centrality

The analysis of eigenvector centrality was conducted to assess the influence that individual stations had in the network focusing on their connections. Eigenvector centrality looks at a station's connectivity and compares it with other stations that are highly connected to identify influential hubs in the network. After the data was imported as CSV files, igraph was used to create graphs in Rstudio. A graph of the whole network was created and it showed the eigenvector centrality. The average and the top 5 stations of the entire network were computed. Stations with the highest scores were identified as influential hubs and were critical to the network. Stations with low scores were isolated highlighting underutilized resources. Analyzing eigenvector centrality using Rstudio helped in understanding the differences in influence across the entire network and because of this it helps to strengthen the network.

3.5 Degree Centrality

The degree centrality analysis of the bike share network evaluated station connectivity using five datasets: one representing the entire network and four representing the main clusters identified through hierarchical clustering. As previously stated, node and edge data were imported from CSV files containing station and trip information, and a directed graph was constructed using the igraph library in RStudio, where nodes represented bike stations and edges indicated trip volume. In-degree centrality, representing the number of incoming trips to each station, was calculated to identify frequently used destination nodes, with the top five nodes highlighted for each dataset. The in-degree distribution was visualized using histograms overlaid with normal distribution curves to assess the spread and density of the data, providing insights into deviations from expected patterns.

Similarly, out-degree centrality was analyzed to capture the number of trips originating from each station, identifying the top five nodes with the highest out-degree values as key origins. Out-degree distributions were also visualized for comparative insights. The average degree of the network was calculated to determine the mean connectivity per station, and the degree ratio (in-degree divided by out-degree) was computed to evaluate the balance of incoming and outgoing trips. Stations with the highest and lowest degree ratios were highlighted to identify potential sink nodes or sources, crucial for optimizing bike redistribution strategies. Additionally, comparing degree centrality metrics across clusters provided a deeper understanding of local station dynamics within different regions of the network. This comprehensive analysis, conducted in RStudio, not only identified high-traffic nodes but also revealed patterns of underutilization, informing operational strategies such as bike redistribution and future network expansion planning.

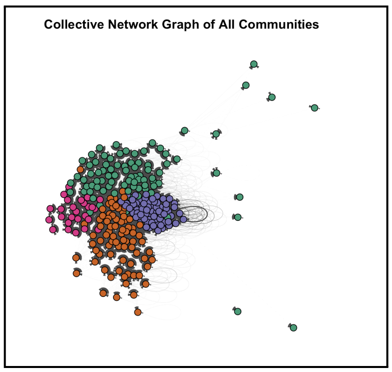
3.6 Closeness Centrality

The closeness centrality analysis was conducted to evaluate station accessibility and strategic importance within the system. Closeness centrality was calculated using the closeness() function with weights parameter set to edge weights and normalization enabled. This weighted approach treated trip frequencies as indicative of shorter "distances," allowing the centrality measure to account for the intensity of connections between stations. As a result, nodes with higher closeness centrality values were identified as more central, demonstrating their ability to efficiently interact with all other nodes in the network. Centrality scores were compiled into a data frame and ranked to highlight the top five most accessible and bottom five least accessible stations, providing insights into key hubs and peripheral nodes both overall and within each of the four clusters identified through hierarchical clustering. Cluster-specific analyses identified major hubs that were well-connected and essential for maintaining efficient flow within their respective clusters. Conversely, some stations exhibited low closeness centrality, indicating limited connectivity and highlighting areas that may require improvement.

4 Results

4.1 Hierarchical Clustering

The average cluster height for the whole graph was 0.428, with four major clusters identified: Cluster 1, consisting of 84 stations with an average cluster height of 0.529 (represented in green); Cluster 2, with 80 stations and an average cluster height of 0.396 (represented in orange); Cluster 3, containing 74 stations with an average cluster height of 0.201 (represented in lilac); and Cluster 4, consisting of 26 stations with an average cluster height of 0.373 (represented in pink). These clusters represent areas with distinct connectivity patterns which can be targeted for rebalancing or station expansions.



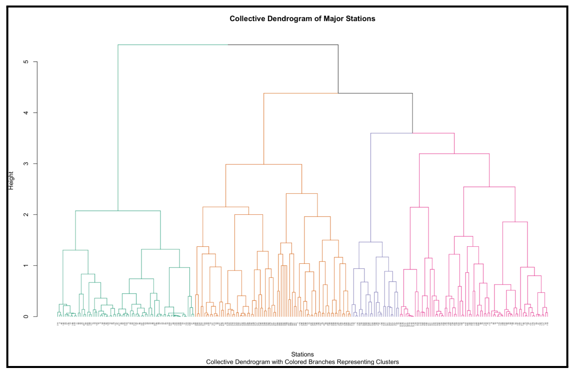


Fig. 4.1.1. Directed Network for the whole network **Fig. 4.1.2.** Dendrogram for the whole network

The dendrogram and graph clearly show four clusters. We can observe a dense core of stations in clusters 2 and 3. The dense core represents high intra-cluster connectivity and high cohesion. The nodes in the periphery represent stations with fewer connections, potentially requiring a better integration with the network. The height of the branches in the dendrogram represents dissimilarity between clusters and nodes. The tall branches splitting clusters 1 and 2 indicate less similarity among them in comparison to the rest of the network. This is because of different connectivity structures or usage patterns.

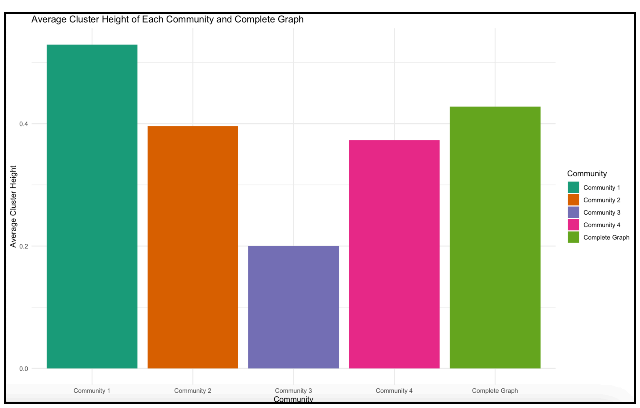


Fig. 4.1.3. Average cluster height for each community and whole graph

The bar plot here shows average cluster heights where cluster 1 has the highest average cluster height in the whole graph. This shows that this cluster has the maximum internal cohesion among all the clusters. On the other hand, cluster 3 has the lowest average cluster height indicating weak internal connectivity between stations. To increase connectivity while keeping the same number of stations, some of the stations can be moved to more accessible areas or places visited by people more often than the already existing station areas.

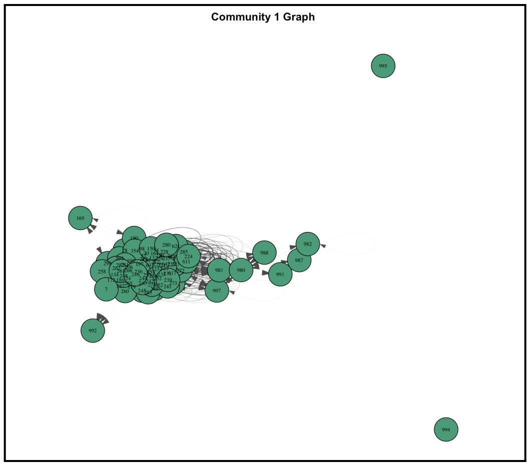


Fig. 4.1.4. Directed network for Cluster 1

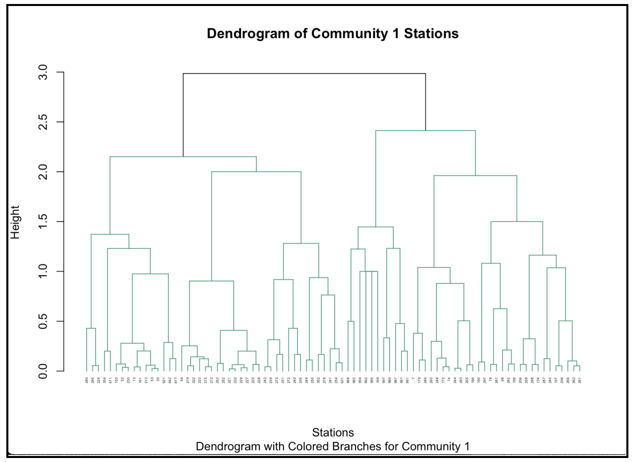


Fig. 4.1.5. Dendrogram for Cluster 1

Cluster 1 shows the majority of nodes connected very closely, with a few of the stations spread far from the main cluster. This strong interconnectivity results in a higher average cluster height of 0.529. This also shows that some nodes in this cluster could be more congested with the traffic. The probable congestion can be reduced by evening the distribution of route traffic across the cluster involving the nodes scattered far.

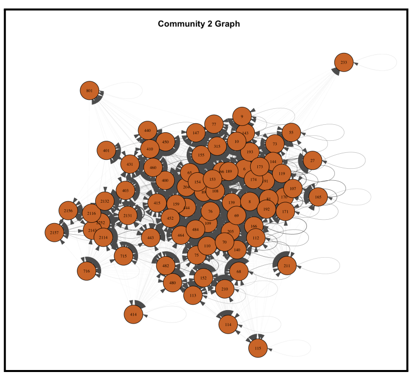


Fig. 4.1.6. Directed network for Cluster 2

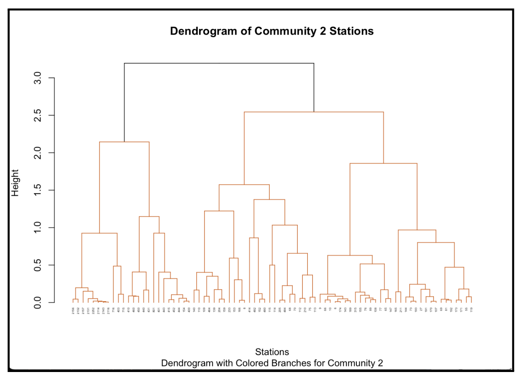


Fig. 4.1.7. Dendrogram for Cluster 2

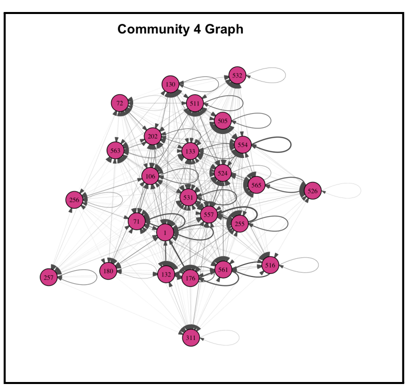


Fig. 4.1.8. Directed network for Cluster 4

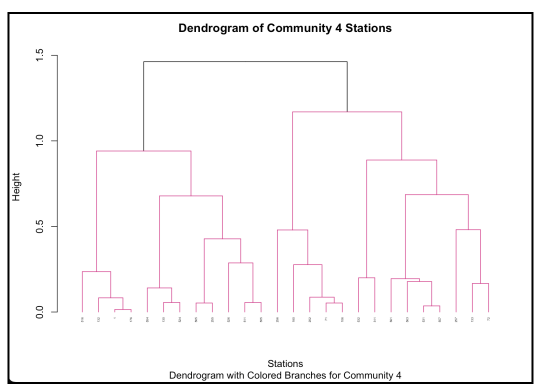


Fig. 4.1.9. Dendrogram for Cluster 4

Communities 2 and 4 are relatively more balanced in terms of connectivity among nodes. The average cluster heights of 0.396 and 0.373 respectively show good interconnectivity among stations.

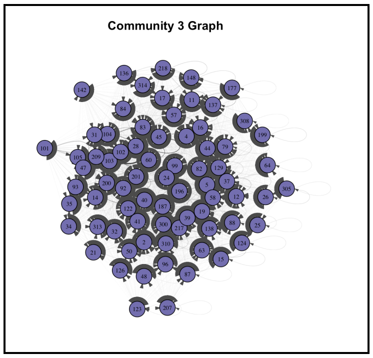


Fig. 4.1.10. Directed network for Cluster 3

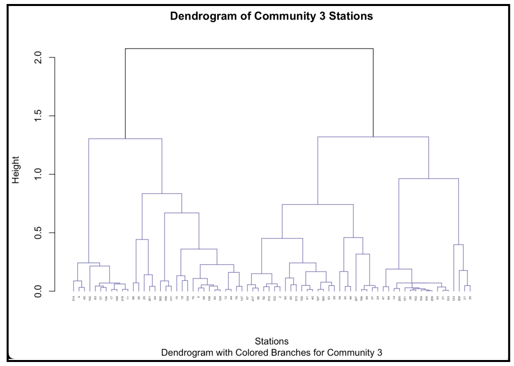


Fig. 4.1.11. Dendrogram for Cluster 3

Cluster 3 shows a dense network of 74 stations with connections throughout. Visually, the connectivity seems consistent but overall connections are relatively less, suggesting some underutilized nodes in the cluster(Zhang et al., 2016). This results in an average cluster height of 0.201. This is the lowest among all the cluster heights which can also be seen from the bar plot above. Introducing new routes between stations and increasing infrastructure would increase the average cluster height. Increasing cohesion in such clusters will help reduce congestion at other heavily utilized stations and provide balanced options for the users.

4.2 Betweenness Centrality

The average betweenness centrality was 239.2, with stations 222, 76, 223, 187, and 77 identified as the top hubs.The stations act as bridges to connect other parts of the network that are less accessible. High betweenness indicates that these stations are critical intermediaries, playing a key role in maintaining connectivity within the network.

Looking at the graph we can see that large circles indicate that nodes near the center (in blue) have high betweenness centrality, acting as crucial intermediaries in the network. Nodes near the center act as bridges connecting other parts of the network and facilitate bikes. Nodes further away in the outer rings have lower betweenness centrality and are less significant for information flow.

4.3 Eigenvector Centrality

The average eigenvector centrality was 0.0188, with the top stations being 209, 105, 103, 102, and 28. Eigenvector centrality measures the influence of a station inside of the network. These stations are influential, serving as hubs in densely connected areas, and enhancing local and global network efficiency. Stations with lower eigenvector centrality often have fewer connections to higher traffic areas.

Analyzing the graph for eigenvector centrality we see that the dense central cluster suggests a highly interconnected core of influential nodes. Peripheral nodes (near the edges) are less central and have lower influence. The top stations in the graph are influential as they sustain significant global and local traffic because of their connection with other nodes. The graph focuses on identifying densely connected clusters or sub-networks.

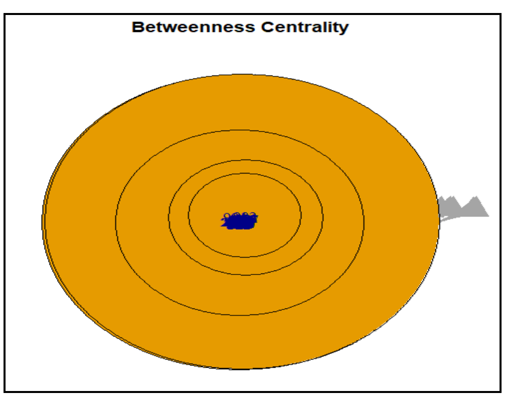


Fig 4.2.1. Betweenness Centrality of Network

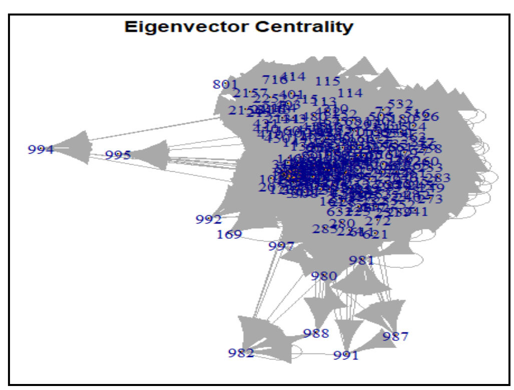


Fig 4.3.1. Eigenvector Centrality of Network

In cluster 1 for betweenness centrality, node 994 is a major hub for information flow in the cluster. It has high betweenness centrality and is a critical connector for paths within the cluster. Once again looking at the eigenvector centrality for cluster 1, node 994 is connected to various other nodes indicating it is influential and has the highest eigenvector centrality. Cluster 4 has a different betweenness compared to cluster 1 because many nodes in the cluster have high betweenness not a single node like in cluster 1. For cluster 4, looking at eigenvector centrality it is easy to distinguish that nodes 621 and 611 have higher eigenvector centrality and are more influential than other nodes.

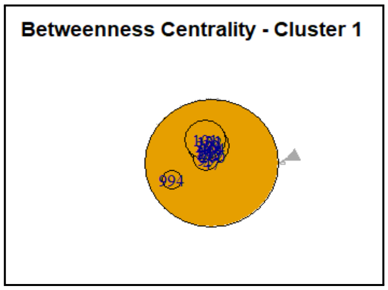


Fig 4.2.2. Cluster 1 Betweenness Centrality

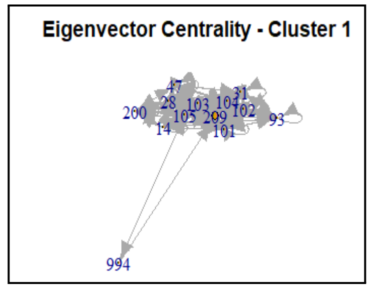


Fig 4.3.2. Cluster 1 Eigenvector Centrality

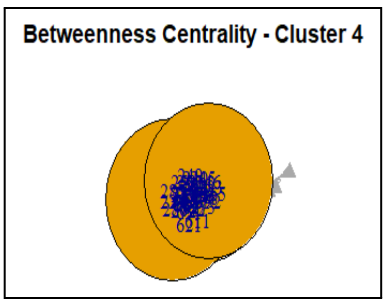


Fig 4.2.3. Cluster 4 Betweenness Centrality

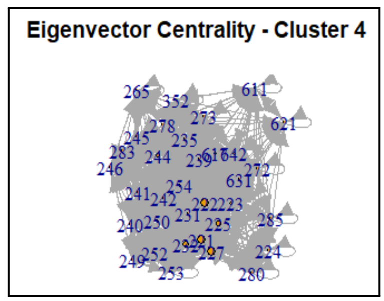


Fig 4.3.3. Cluster 4 Eigenvector Centrality

4.4 Degree Centrality

The in-degree and out-degree distributions of Mobi Vancouver's Bike Share Network provide key insights into the network's structure. The in-degree distribution (Fig. 4.1) is right-skewed, with the highest in-degrees concentrated in a few central nodes (e.g., Node 11 with 202 in-degrees), indicating the presence of highly connected hubs. Similarly, the out-degree distribution (Fig. 4.2) follows a right-skewed pattern, with the highest out-degrees found in a few nodes (e.g., Node 174 with 192 out-degrees), suggesting that while most stations contribute fewer outgoing rides, a small group of nodes are more active in initiating trips. Notably, no sink nodes (nodes with an out-degree of zero) were identified, indicating that all stations actively participate in outbound traffic.

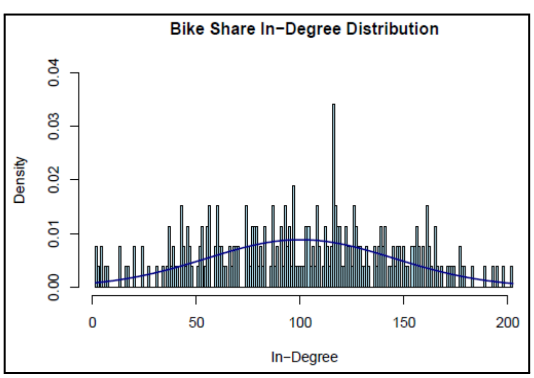


Fig 4.4.1. In-Degree Distribution for Whole Network

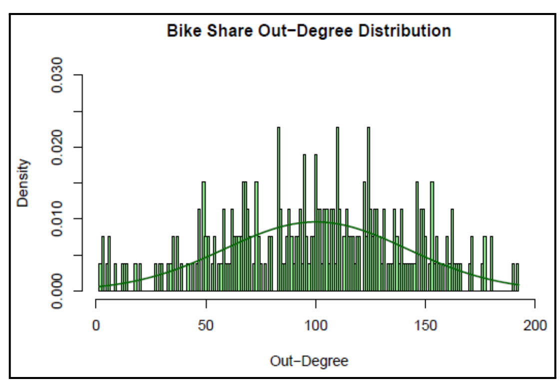


Fig 4.4.2. Out-Degree Distribution for Whole Network

The overall network has an average degree of 200, with an average of 100 incoming and 100 outgoing connections per node, reflecting a highly interconnected structure. Degree ratio analysis further highlights varying balances between incoming and outgoing connections. For example, Node 982 has the highest degree ratio (3.5), indicating a significantly higher number of incoming than outgoing trips, while Node 991 exhibits the lowest ratio (0.333), reflecting an inverse relationship. These findings underscore the presence of specialized nodes functioning as either primary receivers or initiators of bike traffic.

Cluster-level analysis reveals notable variations in connectivity. In Figure 4.3, the in-degree distributions for the four clusters show distinct patterns. Cluster 1, for instance, has nodes with in-degrees ranging from 19 to 60, with a concentration of central nodes (e.g., Node 222 with 60 in-degrees), reflecting a pattern similar to the overall network but with lower average in-degrees. Cluster 2 displays a more evenly spread distribution, with in-degrees ranging from 18 to 71, indicating a moderate concentration of central nodes. Cluster 3 shows a narrower range (62 to 74), suggesting more evenly distributed node connectivity. Cluster 4 exhibits a smaller range (13 to 26), indicating fewer central nodes and a less interconnected structure. The out-degree distributions (Fig. 4.4) reflect similar trends, with Cluster 1 showing more skewed results (15 to 59), indicating a few active nodes initiating trips. Clusters 2 and 3 have more evenly distributed out-degrees (values range from 16 to 69 in Cluster 2 and 58 to 74 in Cluster 3), while Cluster 4 shows minimal variation (8 to 25), suggesting more uniform activity across its nodes.

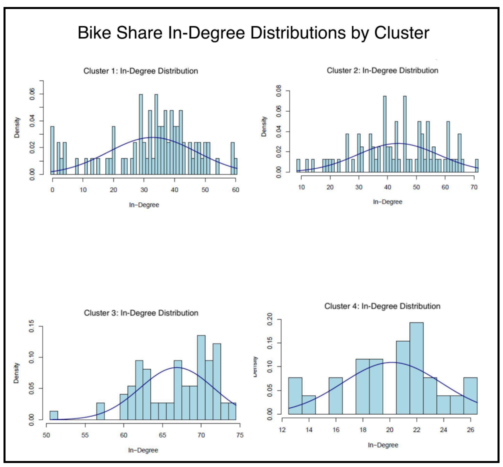


Fig. 4.4.3. In-Degree Distributions by Cluster

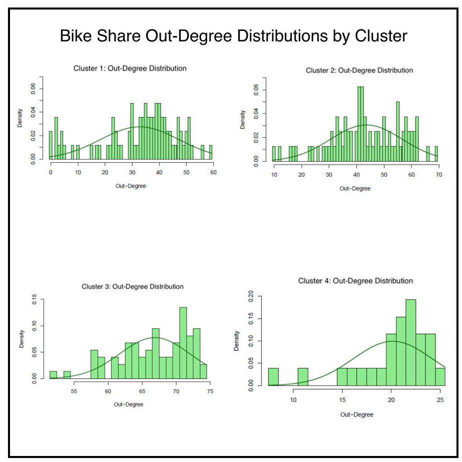


Fig. 4.4.4. Out-Degree Distributions by Cluster

Degree ratio analysis within the clusters reveals further differences in traffic patterns. In Cluster 1, nodes show significant imbalances between in-degree and out-degree. For example, Node 980, with a degree ratio of 2.2, primarily receives trips, likely functioning as a popular destination. Conversely, Node 992, with a degree ratio of 0, does not receive or initiate any trips. In Cluster 2, the ratios are more balanced, with Node 401 having a ratio of 1.46, suggesting relatively equal incoming and outgoing traffic. Node 2114, with a ratio of 0.512, is more outbound-focused but still maintains a balanced flow. In Clusters 3 and 4, connectivity remains more balanced, with nodes like Node 308 in Cluster 3 (ratio of 1.17) preferring incoming trips, while Node 256 in Cluster 4 (ratio of 2.38) is more inbound-focused.

The average degree across the clusters also reveals differences in connectivity. Cluster 1 has an average degree of 65.5, the lowest among the clusters, suggesting that while it contains highly specialized nodes, it is less interconnected overall. Cluster 2, with an average degree of 87.2, shows a more moderate level of connectivity. Cluster 3 has the highest average degree of 134, indicating the most interconnected structure among the clusters, with nodes exhibiting a more balanced flow of traffic. Cluster 4, with an average degree of 40.3, is the least connected, reflecting fewer interactions between stations within the cluster.

The individual clusters reveal varying levels of station centralization and activity. Cluster 1 is more centralized, with specialized hubs acting as key nodes in the network, while Clusters 2, 3, and 4 exhibit more evenly distributed connectivity, suggesting a more balanced traffic flow across these areas. Notably, no sink nodes were found in any of the clusters, ensuring that all stations in the network with inbound traffic contribute to outbound traffic.

4.5 Closeness Centrality

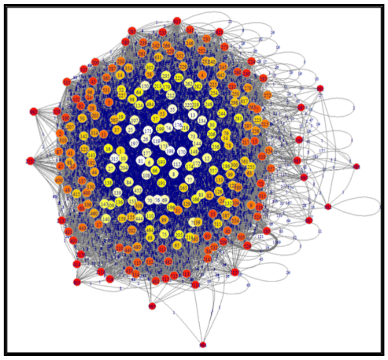


Fig 4.5.1. Closeness Centrality Network

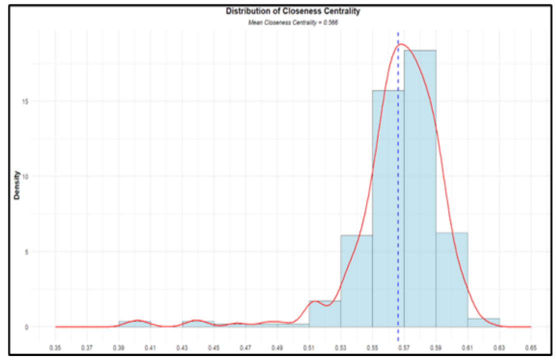


Fig 4.5.2. Distribution of Closeness Centrality

Closeness centrality analysis reveals a **mean of 0.566**, indicating that most stations are relatively well-connected with short paths to others, while centrality scores range from **0.398 to 0.620**, highlighting both peripheral and highly central stations. The **Fig 4.5.1** effectively employs a color gradient from white to red to represent closeness centrality: **red nodes** (e.g., stations **988, 994, 982**) denote the least accessible stations with the lowest centrality, **orange nodes** serve as bridges with moderate centrality, **yellow** represent medium-high centrality as influential secondary hubs, and **white nodes** (stations **176, 81, 198**) signify the most accessible and central stations with the highest closeness centrality. Larger nodes correspond to higher closeness values, making central hubs more prominent. The layout strategically positions white nodes near the center, emphasizing their pivotal role in ensuring efficient communication and resource flow across the network; red nodes are situated towards the periphery, indicating their limited connectivity and reliance on longer paths for accessibility.

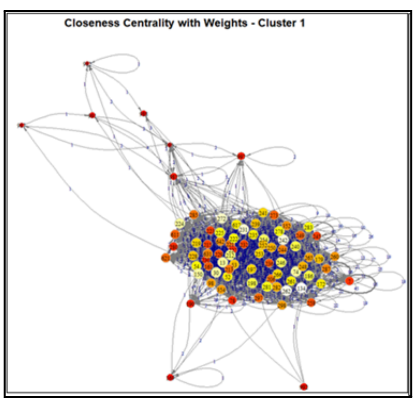


Fig 4.5.3. Closeness Centrality Cluster 1

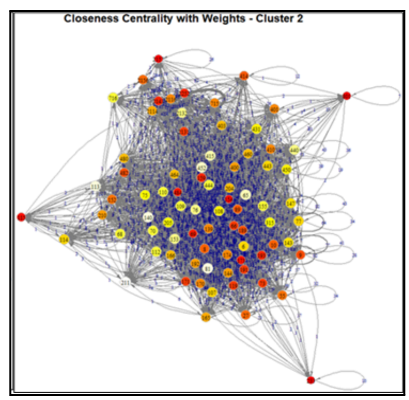


Fig 4.5.4. Closeness Centrality Cluster 2

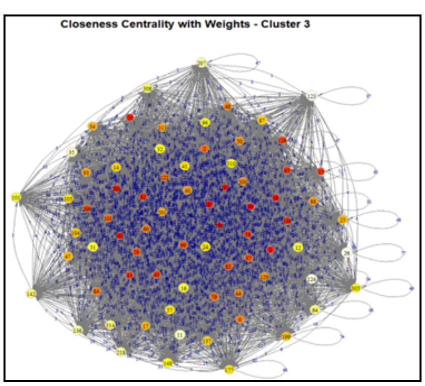


Fig 4.5.5. Closeness Centrality Cluster 3

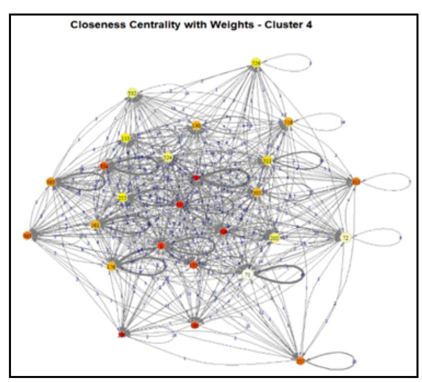


Fig 4.5.6. Closeness Centrality Cluster 4

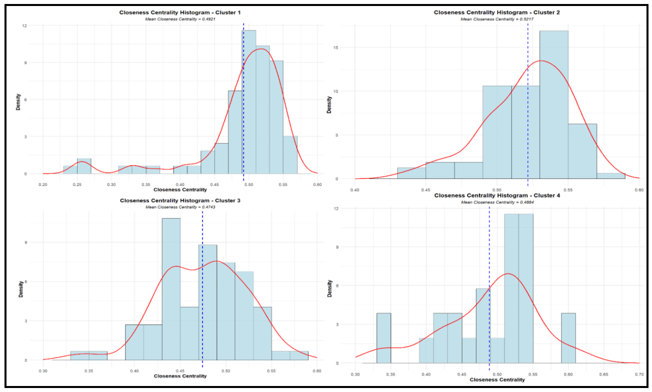


Fig 4.5.7. Closeness Centrality distribution by cluster

Figures 4.5.3–4.5.6 reveal distinct localized dynamics and variations in station connectivity. Prominent hubs, such as Station 134 in Cluster 1 and Station 71 in Cluster 4, exhibit high closeness centrality (0.567 and 0.610), highlighting their essential roles in maintaining efficient intra-cluster flows and overall network connectivity. The overall network has a mean closeness centrality of 0.566, ranging from 0.398 to 0.620 (Figure 4.5.7), with a distribution skewed toward higher values, indicating key nodes that facilitate seamless resource flow. In contrast, peripheral stations like Station 982 in Cluster 1 and Station 557 in Cluster 4 show lower centrality (0.248 and 0.333), suggesting potential inefficiencies and limited integration within their clusters.
These insights guide targeted optimization strategies to enhance both local and global network efficiency. High-centrality stations should be prioritized for maintenance and bike redistribution, while low-centrality nodes need interventions such as infrastructure improvements or promotional campaigns to improve connectivity and accessibility. Addressing these areas through strategic enhancements will boost overall system performance and user experience. Additionally, the histograms for each cluster confirm varying levels of station accessibility, offering actionable insights for tailored interventions that support the sustainable growth and efficiency of the bike-share network.For example, Stations 196 in Cluster 3 and 557 in Cluster 4 are potential bottlenecks that may cause longer travel times and reduced accessibility.

5 Discussion

This research analyzed Mobi Vancouver's bike-share network using graph-theory-based approaches and clustering techniques, revealing valuable insights into its structure and operations. Key findings include the identification of critical hubs with high betweenness and eigenvector centralities, such as stations 222, 76, and 209, which facilitate connectivity and traffic flow, playing a pivotal role in maintaining network efficiency. Cluster analysis uncovered four distinct clusters, each with unique characteristics in station centralization and activity patterns. Cluster 1 features specialized hubs with imbalanced traffic flows, while Cluster 3 demonstrates the highest connectivity. The degree distribution, exhibiting a right-skewed pattern, confirms the presence of major hubs that dominate traffic flow, although no sink nodes were found, indicating that all stations are actively involved in traffic. Variations in degree ratios highlighted stations functioning as primary sources or sinks. For instance, Node 982, with a high degree ratio (3.5), serves as a significant receiver, while Node 991, with a low degree ratio (0.333), primarily acts as an origin. These findings suggest that the Mobi Vancouver network is highly connected, but experiences variability in station activity. Tailored strategies could address congestion at hubs and improve underused stations.

The findings have several implications for optimizing the Mobi Vancouver bike-share network. Operationally, the identification of key hubs provides a basis for prioritizing rebalancing efforts, ensuring bike availability during peak hours and reducing downtime at less-used stations. From an infrastructure planning perspective, insights into cluster dynamics can inform the placement of new stations or the expansion of existing ones, enhancing overall accessibility and connectivity.

The study has several limitations. Firstly, the data scope was limited to just one month, which may not fully capture long-term trends or account for natural events influenced by weather or policy changes. Additionally, the analysis assumes a static network, while bike-share systems are inherently dynamic, with factors such as weather continuously affecting operations.

There were several challenges in the data cleaning and cluster interpretation processes. Initially, the raw data contained inconsistencies, which were addressed by filtering to ensure the necessary data was obtained. Additionally, understanding the relationships between clusters proved difficult due to variability in connectivity. However, dendrograms and graphical representations were utilized to help clarify these relationships.

Future directions for this research include incorporating additional metrics for a more comprehensive analysis. Expanding the dataset to cover multiple months or years would also provide a broader perspective and potentially yield different results, enhancing the understanding of long-term trends and network dynamics.

In reflection, data enrichment could enhance the analysis by incorporating external datasets, such as weather patterns, traffic data, and other relevant variables, providing a more in-depth understanding of network dynamics. Additionally, adopting a user-centric approach, such as conducting user surveys or integrating app usage data, would help validate the findings and ensure that recommendations align with the actual needs and experiences of users.

6 Conclusion

This study aimed to analyze the mobi-cycle network to identify critical stations, evaluate connectivity patterns, and optimize network flow. The research problem focused on understanding station significance and areas for improvement in connectivity. Raw trip data was refined and filtered, creating a weighted edges dataset that helped reveal frequently used routes.

Key methodologies included hierarchical clustering to identify sub-networks and centrality metrics (betweenness, eigenvector, degree, and closeness) to assess station importance. Findings highlighted key stations with high betweenness and eigenvector centrality, indicating congestion points and hubs of influence. Degree and closeness centrality further revealed stations that could benefit from improved bike redistribution and better accessibility.

The implications of this study contribute to optimizing the mobi-cycle network's design and operations, informing strategies for better resource allocation and future expansion. However, the study was limited by the lack of dynamic data, such as weather or real-time usage patterns, and does not account for external factors.

Future research should incorporate real-time data and predictive models to enhance network adaptability. By expanding the dataset and exploring machine learning, the network can become more responsive to changing demands, making it more efficient and resilient.

7 References

1. Chiariotti, F., Pielli, C., Zanella, A., & Zorzi, M. (2018). A Dynamic Approach to Rebalancing. Bike-Sharing Systems. Sensors, 18(2). https://doi.org/10.3390/s18020512
2. Schuijbroek, J., Hampshire, R. C., & van Hoeve, W.-J. (2017). Inventory rebalancing and vehicle routing in bike sharing systems. European Journal of Operational Research, 257(3), 992–1004. https://doi.org/10.1016/j.ejor.2016.08.029
3. Soto, D. A., & Ceballos, Y. (2024). Stochastic agent-based models optimization applied to the problem of rebalancing bike-share systems. International Journal of Electrical and Computer Engineering (IJECE), 14(5), 5641. https://doi.org/10.11591/ijece.v14i5.pp5641-5651
4. Adham, Manal T., and Peter J. Bentley. "Evaluating clustering methods within the Artificial Ecosystem Algorithm and their application to bike redistribution in London." BioSystems, vol. 146, 2016, pp. 43–59, https://doi.org/10.1016/j.biosystems.2016.04.008.
5. Salih-Elamin, R., & Al-Deek, H. (2020). A new method for determining optimal locations of bike stations to maximize coverage in a bike share system network. Canadian Journal of Civil Engineering. https://doi.org/10.1139/cjce-2020-0143
6. Wang, Y., & Cullinane, K. (2016). Determinants of port centrality in maritime container transportation. Transportation Research Part E: Logistics and Transportation Review, 89, 1-17. https://doi.org/10.1016/j.tre.2016.04.002

❖ GitHub Repository: https://github.com/Sumer26/COSC_421_Project_Newtork_Science.git