Social Network Analysis

CS/SDP 361/352

HW1: Homework 01 (Network Analysis and Visualization)

Group Members : Hammad Malik and Ahtisham Uddin

1. Introduction

This study presents a detailed network analysis of the Karachi bus system, including the structural characteristics and connectivity patterns of bus stops across the city. The network has 572 bus stops (vertices) that are linked by 1,139 bus routes (edges), representing one of the major modes of public transport in Karachi, which is Pakistan's largest city.

Figure 1 : Presents the geographic distribution of all 572 bus stops across Karachi

2. Global-Level Network Analysis

2.1 Basic Network Properties

The Karachi bus network shows characteristics that represent a sparse transportation network with the following network characteristics properties:

- **Network Size:** The network comprises 572 vertices (bus stops) and 1,139 edges (direct bus routes between stops)
- **Network Density**: 0.0035 (0.35%)
- **Average Path Length**: 8.09 stops
- **Network Diameter**: 25 stops

- **Mean Degree**: 3.98
- **Max Degree** : 31

The low density of **0.35%** indicates that the network is highly sparse, with each bus stop connected to only a small fraction of all possible stops. This is expected for a transportation network where direct connections between all pairs of stops would be impractical. The average degree of approximately **4** connections per stop suggests that most bus stops serve as intermediate points rather than major hubs.

2.2 Connectivity and Reachability

The network's connectivity analysis reveals:

- **Number of Components**: 2
- **Largest Component Size**: 568 nodes (99.3% of the network)
- **Disconnected Component**: 4 nodes

The presence of two components indicates that the network is nearly, but not completely, connected. The main component encompasses **99.3%** of all bus stops, suggesting good overall connectivity for the vast majority of the city. However, 4 bus stops (Police Training School, Khayaban-e-Shamsheer**,** Society Hospital**,** Moach Mor) remain isolated from the main network, potentially representing the peripheral areas with limited public transportation access. The disconnected stops were identified by their closeness centrality value of 1.0, indicating isolation from the main network which we will discuss later.

The average path length of **8.09 stops** means that, on average, passengers need to traverse through approximately 8 intermediate stops to reach their destination.. The network diameter of **25** stops represents the longest shortest path between any two stops in the network, indicating that even the most distant locations can be connected within **25 stops.**

2.3 Clustering and Community Structure

The clustering analysis reveals interesting patterns in the network's local structure:

- **Global Clustering Coefficient**: 0.142
- **Local Clustering Coefficient**: 0.153

These moderate values of clustering coefficients indicate that the bus stops are forming **localized clusters** where the neighboring stops are connected. This behavior follows the usual planning of the bus network in which the routes generally overlap some parts in order to add redundancy as well as flexibility for the commuters.

Community detection using two different algorithms yielded:

- Louvain Method: 16 communities with modularity of 0.7373
- Walktrap Method: 64 communities with modularity of 0.6747

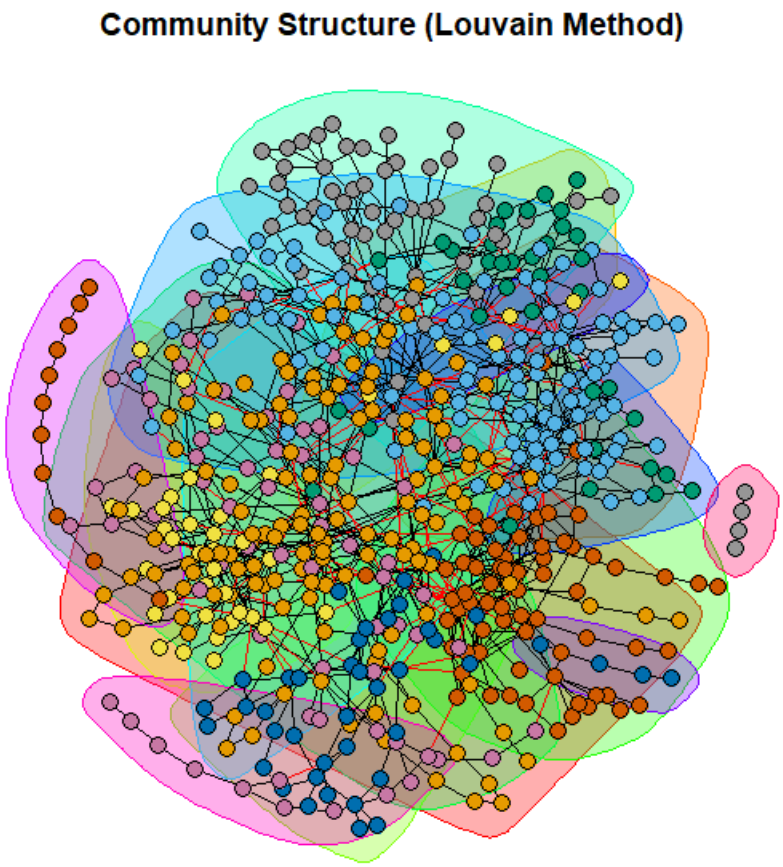

Figure 2 : Presents the community structure using Louvain Method

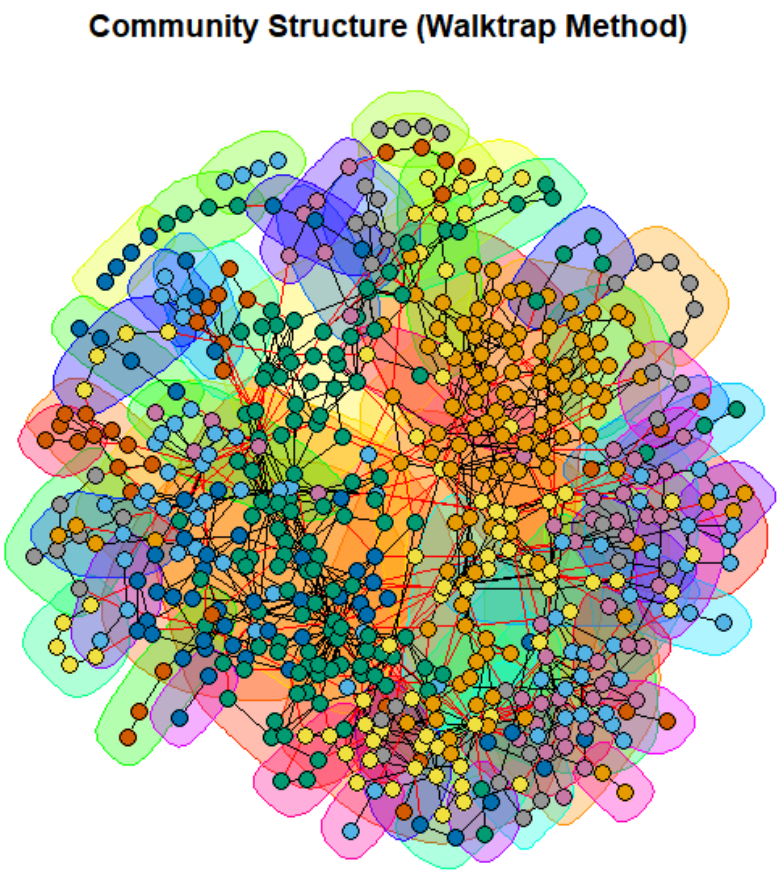

Figure 3 : Presents the community structure using WalkTrap Method

High modularity values (>0.6) suggest the presence of a robust community structure in the network. It is likely that the Louvain algorithm's identification of **16 large** communities consist of different geographical areas or Karachi neighborhoods, whereas the denser network resulting from the Walktrap algorithm (in 64 communities) reflects local areas of small size or **isolated** clusters of bus routes.

2.4 Degree Distribution

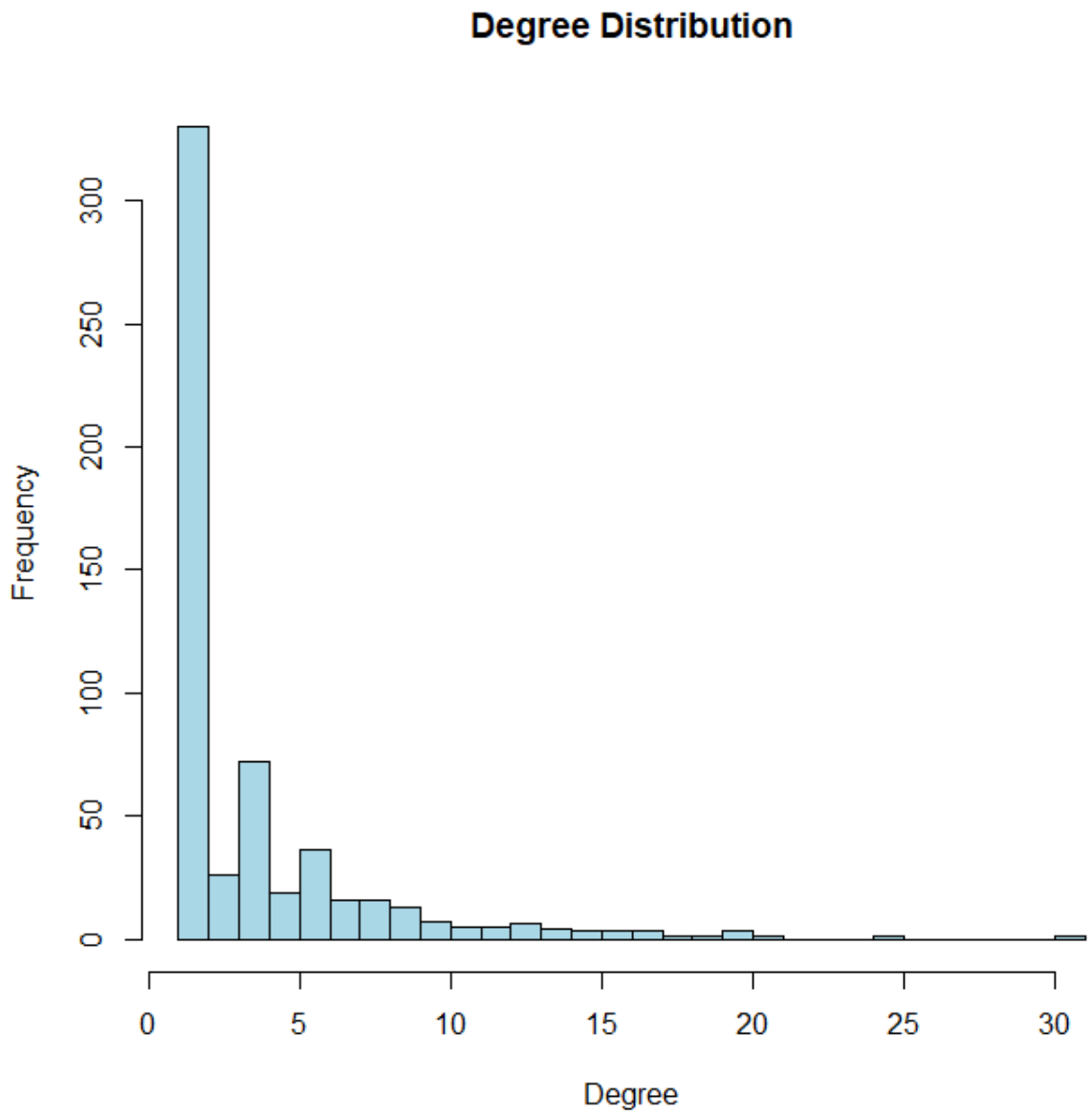

Figure 4 : Degree Distribution

The degree distribution analysis reveals a heavily right-skewed pattern characteristic of scale-free networks:

- Mean Degree: 3.98
- Median Degree: 2

- Degree Range: 1 to 31

The degree distribution histogram clearly demonstrates this skewness, with approximately 300 stops **(52% of the network)** having only 1-2 connections, while a small number of hubs have 15+ connections. Specifically:

- 300+ stops have degree 1-2 (minimal connectivity)
- 70 stops have degree 3
- 40 stops have degree 4-5
- Only 15 stops have degree >10 (major hubs)

This gross disparity in connectivity (mean 3.98 vs median 2) verifies a hub-and-spoke system whereby most stops come as endpoints or form basic connectors, but only a few **super-connected** hubs absorb the majority of the route intersections. This pattern, while creating efficiency in routing and reducing the total number of routes, also shows vulnerability for the hubs to collapse as well as possible **clogging** at the crucial and important nodes.

3. Centrality Analysis

3.1 Degree Centrality

Degree centrality identifies bus stops with the **most direct connections**. The top 10 stops by degree centrality are:

1. NIPA Chowrangi (31 connections) - The most connected hub
2. Tower (25 connections)
3. Sohrab Goth (21 connections)
4. M.A. Jinnah Road (20 connections)
5. Gurumandir (20 connections)
6. Numaish (20 connections)
7. Saddar (19 connections)
8. Petrol Pump (18 connections)
9. Liaquatabad No. 10 (17 connections)
10. Teen Hatti (17 connections)

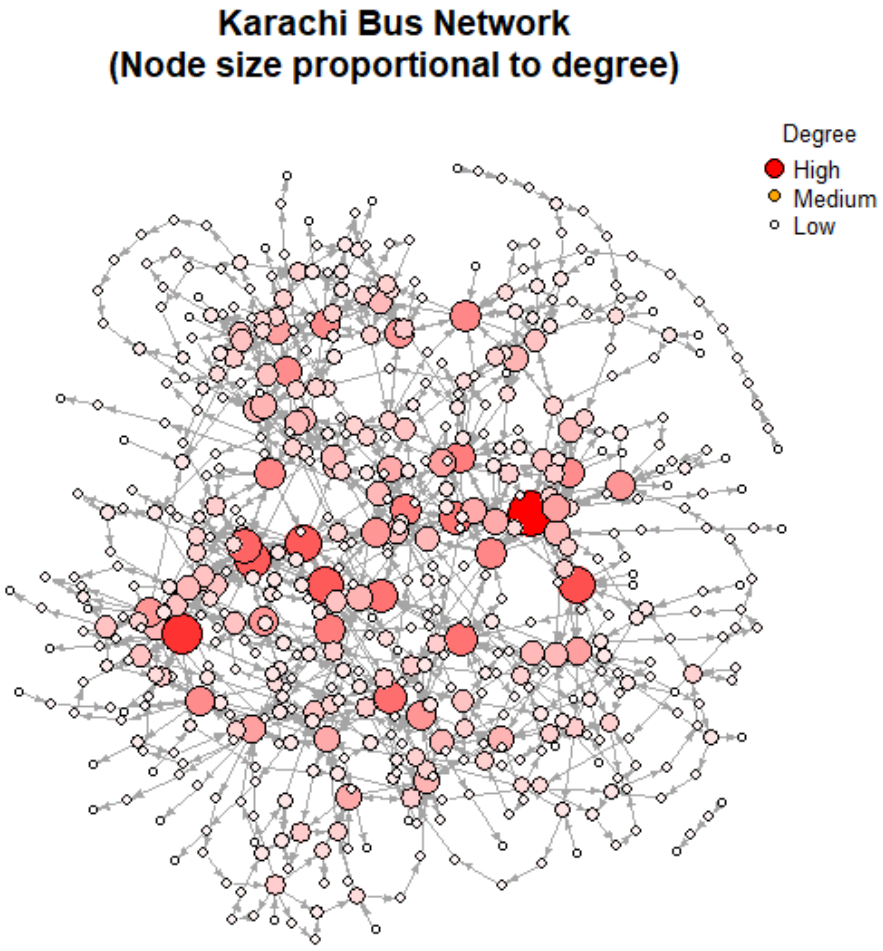

Figure 5 : Degree Centrality Visualization

NIPA Chowrangi becomes the network's super-hub with **31 direct links**, almost 8 times the average. Such highly interconnected stops act as large interchange hubs where several bus routes facilitate the transfers among various city parts. Sites such as **Tower, Saddar, and M.A. Jinnah Road** testifies their value as commercial as well as administrative hubs.

3.2 Betweenness Centrality

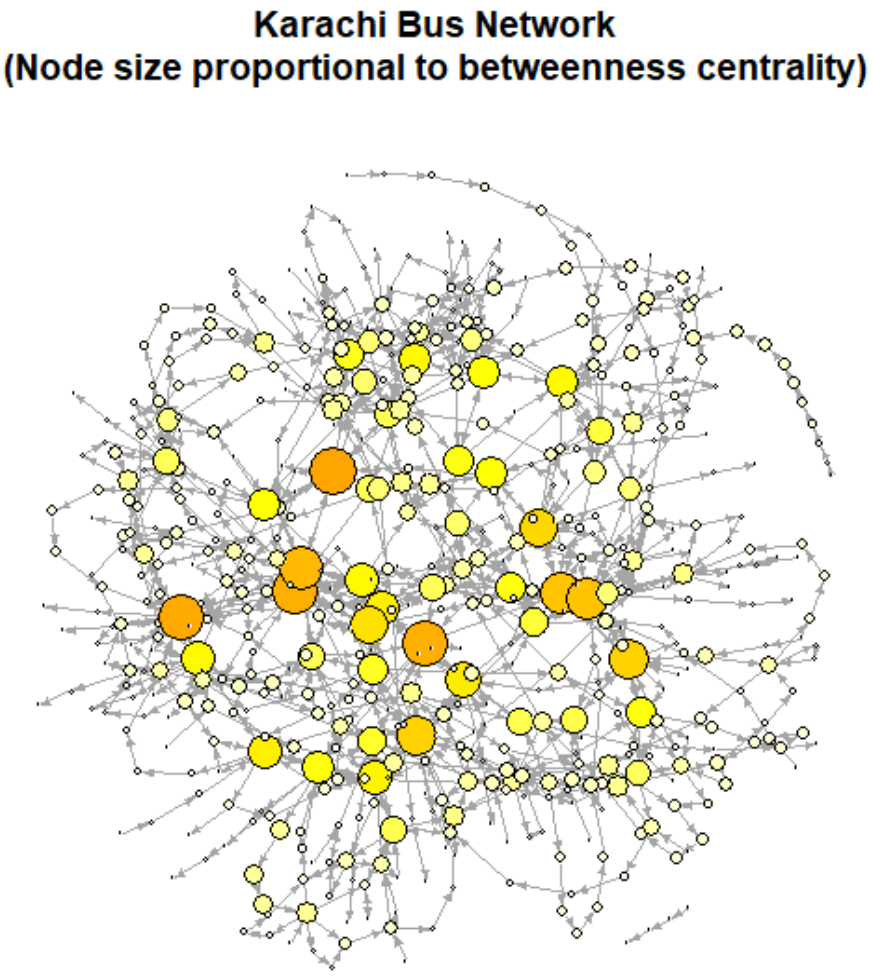

Figure 6 : Between Centrality Visualization

Betweenness centrality reveals stops that serve as critical bridges in the network:

1. Jinnah Hospital (37,008.87) - The most critical bridge node
2. Tower (35,801.30)

3. M.A. Jinnah Road (34,663.06)
4. Labour Square (34,576.15)
5. Saddar (31,844.01)
6. NIPA Chowrangi (30,336.94)
7. Rashid Minhas Road (29,660.03)
8. Sohrab Goth (27,094.97)
9. Petrol Pump (27,022.02)
10. Drigh Road (26,121.52)

Interestingly, the highest betweenness centrality lies in **Jinnah Hospital** but it is not the most centrally connected node itself, making it centrally located as an interim bridge between the network's different parts. Such stops are important for the preservation of network connectivity and their **collapse** would greatly affect travel time throughout the city.

3.3 Eigenvector Centrality

Eigenvector centrality identifies stops that are connected to other well-connected stops:

1. NIPA Chowrangi (1.0) - Maximum eigenvector centrality
2. Purani Sabzi Mandi (0.65)
3. Hassan Square (0.60)
4. University Road (0.49)
5. Gulshan Chowrangi (0.47)
6. Sohrab Goth (0.45)
7. Safari Park (0.42)
8. Jail Chowrangi (0.37)
9. Civic Centre (0.37)
10. Islamia College (0.37)

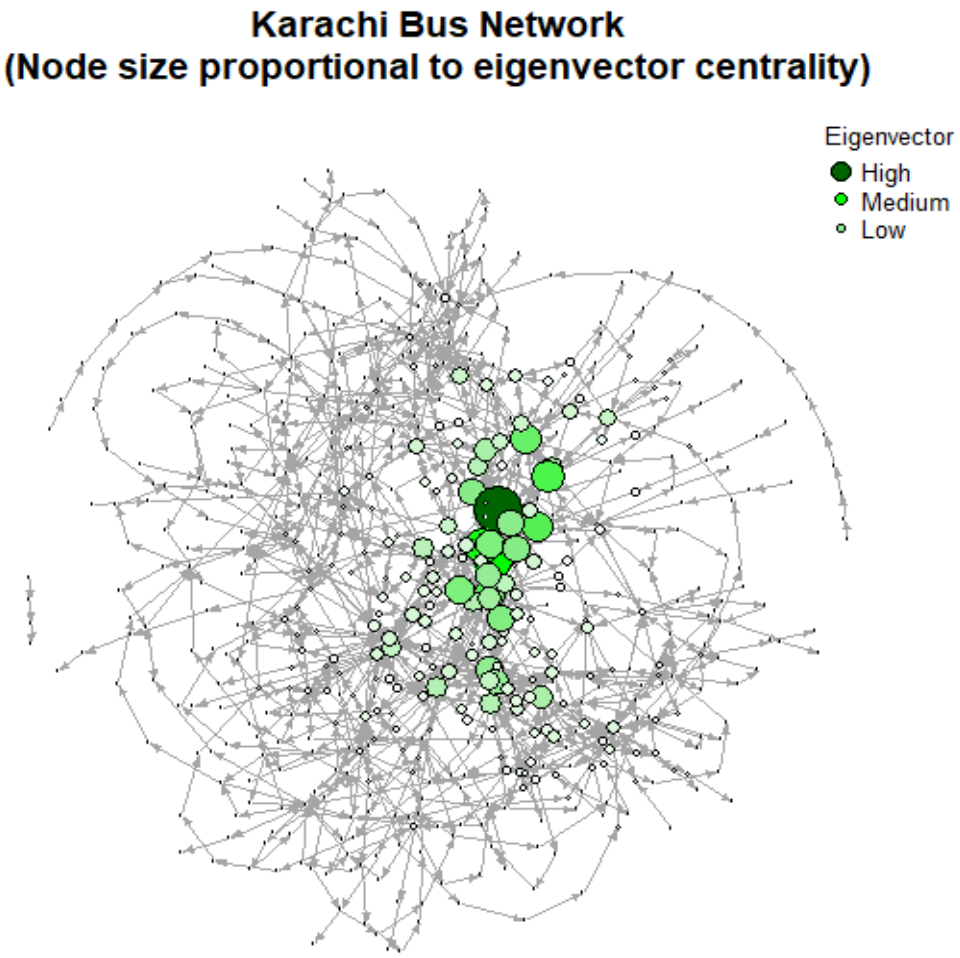

Figure 7 : Eigen Vector Centrality Visualization

The perfect eigenvector centrality score by **NIPA Chowrangi** validates the ranking of the network as not only the highly connected node but as the node also connecting to the other influential hubs forming hierarchical configuration in the network.

3.4 Closeness Centrality

To avoid NaN (Not a Number) values that would arise from disconnected components, closeness centrality was calculated only for the main connected component containing 568 nodes (99.3% of the network). The analysis revealed:

- Mean Closeness: Calculated for main component only
- Top-ranking stops: Include Police Training School, Khayaban-e-Shamsheer, Society Hospital, and various residential areas

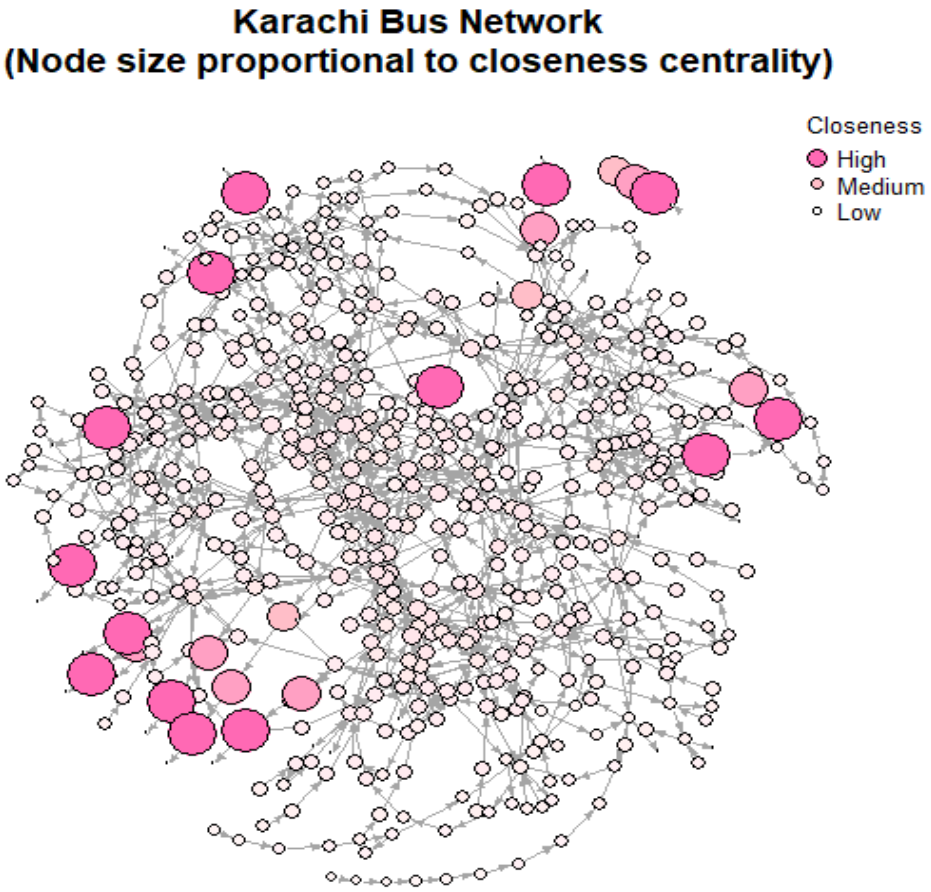

Figure 8 : Closeness Centrality Visualization

The closeness centrality measure identifies stops that can reach all other stops in the network with the shortest average distance. While the specific values require careful interpretation due to the component-based calculation, this metric helps identify stops that are optimally positioned for accessibility across the entire network. The presence of residential areas and institutional locations like hospitals, schools and training centers among high-closeness nodes suggests these locations have strategic positions that minimize travel distances to other parts of the city, even if they don't necessarily have the most direct connections.

3.5 PageRank Centrality

The PageRank visualization reveals a more distributed pattern of importance compared to other centrality measures:

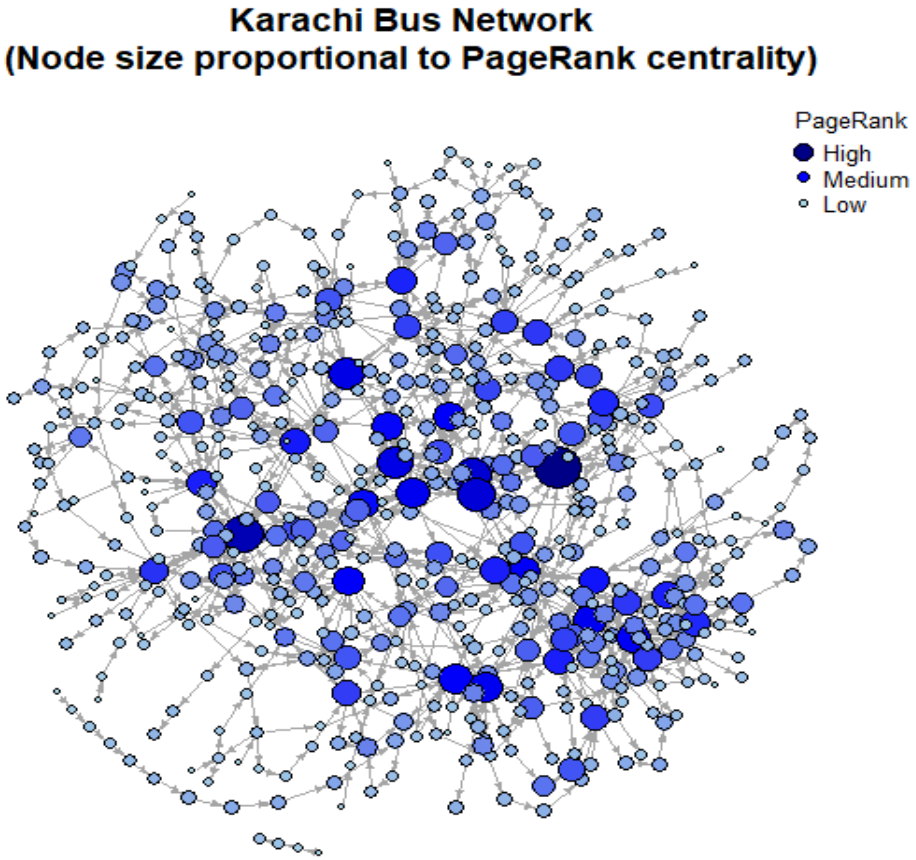

Figure 9 : Page Rank Centrality Visualization

Key Observations:

- The visualization shows **multiple clusters** of high PageRank nodes (dark blue) rather than a single dominant hub
- Unlike degree centrality which favored **NIPA Chowrangi** heavily, PageRank identifies several equally important zones
- The central region maintains high importance but shares influence with peripheral clusters
- Medium PageRank nodes (medium blue) form connective corridors between high-importance clusters

Although **NIPA Chowrangi** has the highest number of direct routes (31 connections), PageRank shows that additional stops achieve equal or larger significance due to their advantageous position in the actual traffic flows.

Lets say a stop which has 3 connections may possess a large PageRank due to the connections being made towards the large hubs such as Tower or Saddar. While traveling through the network the users tend to go through this stop repeatedly although having fewer numbers of direct routes.

This creates a different hierarchy than simple connectivity suggests:

- Some low-degree stops become critical because they're the only way to reach certain areas
- Traffic naturally flows through multiple zones rather than everything going through NIPA Chowrangi
- Peripheral stops connected to major hubs inherit importance from their well-connected neighbors

For example, if Stop A has 5 connections to minor peripheral stops, it might have a lower PageRank than Stop B with only 2 connections to NIPA Chowrangi and Tower. This is because passengers randomly traveling through the network would more frequently pass through Stop B due to its connections to high-traffic hubs. This is in accordance with the Page Rank definition we studied in class.

6. Conclusion

Analysis of the Karachi bus network identifies an intricate transport network featuring high hierarchical order and distinct community structure. It achieves connectivity for 99.3% of its 572 bus stops using 1,139 routes and shows evidence of thorough metropolitan area coverage despite being highly fragmented in overall connectivity (0.35%) density.

The network demonstrates several collaboration patterns:

1. Hierarchical Collaboration: The extreme degree inequality (52% of stops with only 1-2 connections vs. 3% with 10+ connections) creates a tiered collaboration system where super-hubs like NIPA Chowrangi (31 connections) coordinate traffic flow between multiple regions while local stops serve specific neighborhoods.

2. Community-Based Collaboration: The 16 identified communities (Louvain modularity: 0.7373) represent localized collaboration clusters where stops within each community maintain dense internal connections, enabling efficient local transportation while relying on bridge nodes for inter-community travel.

3. Critical Bridge Collaboration: Stops like Jinnah Hospital (highest betweenness centrality: 37,008) serve as essential collaboration points that maintain network cohesion by connecting otherwise distant communities, though this creates vulnerability to single-point failures.

The network analysis presents both the weaknesses and the strengths of the Karachi bus system's network. Positively, the **high modularity** and the **robust community** structure allow for redundancy on the local level, so neighborhoods also have good internal connectivity when the connections with other areas go down. But the network's extreme dependency on some **critical hubs** which we saw like NIPA Chowrangi, Tower, and Jinnah Hospital—poses large **vulnerabilities.** Those stops act as single points of failure where any disruption would propagate throughout the network.

In regards to patterns of accessibility, the network presents satisfactory total connectivity at an average route length of 8 stops, indicating the majority of destinations could be accessed without an excessive number of transfers. Nevertheless, the existence of isolated components downplays transportation deserts impacting about 0.7% of stops which fully isolates them from the network as a whole. While the overarching hub-and-spoke configuration makes for efficient routing through the concentration of links at key interchange sites, it creates **congestion bottlenecks** at these same critical nodes during peak travel times.

The service distribution across the network exhibits clear inequality, as evidenced by the right-skewed degree distribution. Central commercial and administrative areas enjoy robust service with multiple route options, providing residents and workers with flexibility and redundancy in their travel choices. Whereas we can see that peripheral areas suffer from **limited connectivity**, with many stops having only one or two connections. This disparity in transportation access potentially affects socioeconomic mobility, as residents in poorly connected areas face greater challenges accessing employment, education, and services in the city center.