ES404 Networks and Complex Systems

Assignment 2

Student: Shashank Ghosh, Jainam Chajjer
Lecturer: Prof. Udit Bhatia

Abstract

A "dark store" is a specialized retail space that is designed exclusively for the fulfillment of online orders. Unlike conventional stores that welcome walk-in customers, dark stores operate primarily as warehouses or distribution centers where products are stored, organized, and picked to prepare orders for delivery or pickup. Their layout is optimized not for customer browsing but for rapid picking and packing to meet the demands of e-commerce—especially in environments where speed is essential, **often promising delivery times as short as 10 to 15 minutes.**

Location: Typically situated in urban or suburban areas near dense populations to facilitate quick delivery, dark stores often occupy repurposed retail spaces or warehouse facilities.

1 Problem Statement

In today's rapidly evolving e-commerce environment, the strategic positioning of dark stores is paramount for ensuring efficient last-mile delivery and meeting surging customer demands. This project seeks to develop an advanced optimization framework based on network science to determine optimal dark store locations within an urban road network. By leveraging comprehensive road network data and applying dynamic edge weighting to account for temporal traffic variations—particularly along major thoroughfares—the proposed model aims to minimize travel times and bolster logistical efficiency.

The framework will integrate network centrality measures and community detection algorithms to identify critical nodes and delineate neighborhoods, thereby ensuring equitable service coverage across diverse urban areas. Moreover, the model will incorporate real-world constraints such as zoning regulations, infrastructure availability, and budgetary limitations. This multi-objective, adaptive approach is designed to enhance the resilience and scalability of last-mile logistics, ensuring that dark store deployments effectively serve local demand, reduce delivery times, and contribute positively to urban retail dynamics.

Problem 1: Network Construction & Preliminaries

(a) **Dataset Description: Provide context about your network**:

- (i) What does each Node represent?
- (ii) What do the edges signify?
- (iii) Where/how did you acquire this data?

(b) **Construct Your Network:** Describe any data cleaning/preprocessing steps before building the adjacency list/matrix

(c) **Initial Observations:**

- (i) Report the number of nodes (N) and edges (E)

(ii) Plot your network's degree distribution with proper binning. Comment on any notable patterns (e.g., skewness, heavy tail).

Solution 1(a).

(i) **Node Interpretation:** - In the road network, each node typically represents an intersection (or a point on a road) in your city.

(ii) **Edge Interpretation** - Each edge connects two nodes and represents a road segment. In our case, we also plan to assign weights reflecting travel time, adjusted by factors (such as traffic on main roads during peak hours)

(iii) **Data** The dataset is extracted from OpenStreetMap using the OSMnx package.

Solution 1(b).

For Data Cleaning and Pre-processing we use the Inbuilt support functions from OSMnx and NetworkX, also road networks dont require a lot of preprocessing and cleaning. Optionally, we have removed isolated nodes, or performed edge weight adjustments (for instance, assigning higher weights to primary roads during rush hour)

Solution 1(c).

Number of nodes: 4217, Number of edges: 10576
(Code attached separately in ipynb notebook to calculate this.)

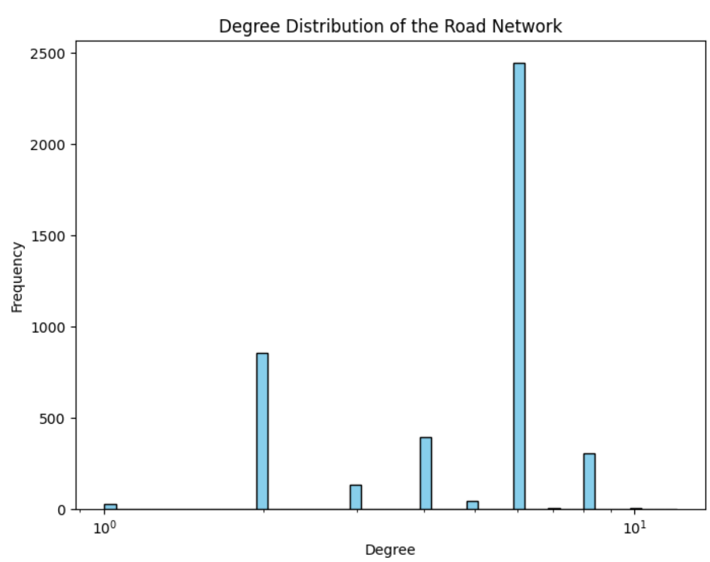

Figure 1: Degree Distribution with Binning

Problem 2: Centrality Analysis

Compute and discuss:

(a) Eigenvector Centrality

(b) Katz Centrality.

(c) Betweenness Centrality.

(d) Closeness Centrality

(e) HITS (Authority and Hub Scores)

All the source codes and the files used are separately attached in an .ipynb file.

Solution 2(a). Eigenvector Centrality reflects how well-connected a node is, considering the connectivity of its neighbors.

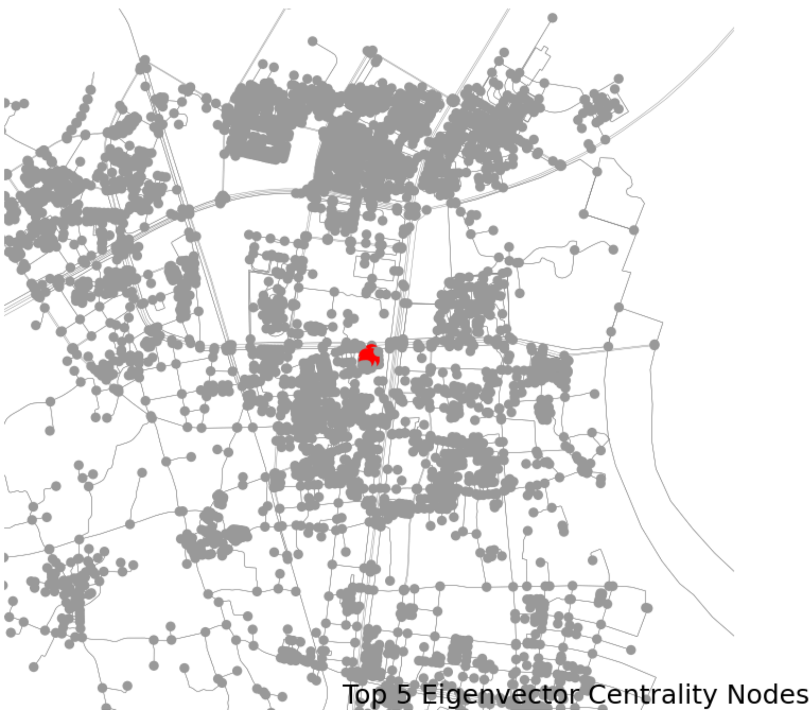

Figure 2: Top 5 Eigenvector Nodes, Map of Kudasan, Gandhinagar

The Red nodes are the top 5 eigenvector nodes.

Solution 2(b). Katz centrality is a measure of a node's influence in a network, considering both direct and indirect connections. Unlike simpler measures like degree centrality (which counts immediate connections) or betweenness centrality (which focuses on shortest paths), Katz centrality accounts for all possible paths to a node, with a damping factor α) that weights shorter paths more heavily than longer ones. This makes it a holistic measure of a node's integration into the network.

The top 5 katz nodes are central to the network's connectivity, not just because they have many direct roads (high degree), but because they lie on numerous short paths connecting other parts of the network. They act as junctions where traffic can flow through multiple efficient routes.

These intersections likely experience high traffic volumes, as they serve as convergence points for trips originating from or destined to various parts of the network. They might be prone to congestion or strategic for traffic management.

Figure 3: Top 5 Katz Centrality Nodes(in Red), Map of Kudasan, Gandhinagar

Solution 2(c). Betweenness Centrality is a measure in network analysis that quantifies a node's importance based on how often it lies on the shortest path between pairs of other nodes. In a road network, nodes represent intersections or key points, and edges represent roads.

Figure 4: Top 5 Betweenness Centrality Nodes(in red)

These nodes act as major conduits for the flow of traffic. They are frequently used in the shortest routes between many origin-destination pairs, making them essential for efficient travel. Disruption at these nodes can cause traffic bottlenecks and jams.

Solution 2(d). Closeness Centrality measures how centrally located a node is within a network, based on the **average shortest path** distance to all other nodes. It is calculated as the reciprocal of the sum of the shortest path distances from a node to every other node. Higher values indicate a node is more "central" (i.e., it can reach all other nodes more quickly).

Figure 5: Top 5 Closeness Centrality nodes(in red)

These nodes are ideal for placing facilities like hospitals, fire stations, or logistics hubs, as they minimize average travel time/distance to all other locations. High closeness nodes are likely to experience heavy traffic, as they lie on efficient routes connecting many parts of the network. This could highlight congestion risks.

Solution 2(e). HITS (Hyperlink-Induced Topic Search) Centrality is an algorithm that assigns two scores to nodes in a directed graph: authority and hub. Authorities are nodes linked by many hubs (important destinations), while hubs are nodes that link to many authorities (key connectors).

Top 5 Authority Nodes:
Interpretation: These intersections act as critical destinations or convergence points.
Why They Matter:

1. They receive traffic from many high hub-score nodes (e.g., major highways, connectors).
2. Likely central business districts, hospitals, or busy commercial areas where routes terminate or merge.
3. Useful for identifying congestion hotspots or prioritizing infrastructure upgrades.

Top 5 Hub Nodes:
Interpretation: These intersections serve as major connectors or distribution points.
Why They Matter:

1. They channel traffic to numerous high-authority nodes (e.g., highways, bridges, or roundabouts).
2. Act as critical junctions for efficient route planning and emergency response logistics.
3. Vulnerable to bottlenecks if disrupted, making them key targets for redundancy planning.

Figure 6: HITS nodes

Problem 3: Modularity and Community Detection

(a) Use a community detection algorithm (e.g., Louvain or Girvan–Newman).

(b) Compute the **modularity score**.

Solution 3(a). The Louvain Community Detection Algorithm is a widely-used, efficient method for uncovering communities (densely connected groups of nodes) in large networks. It optimizes modularity, a metric quantifying how well a network is divided into communities, favoring groupings where nodes have more connections within their community than expected by chance.

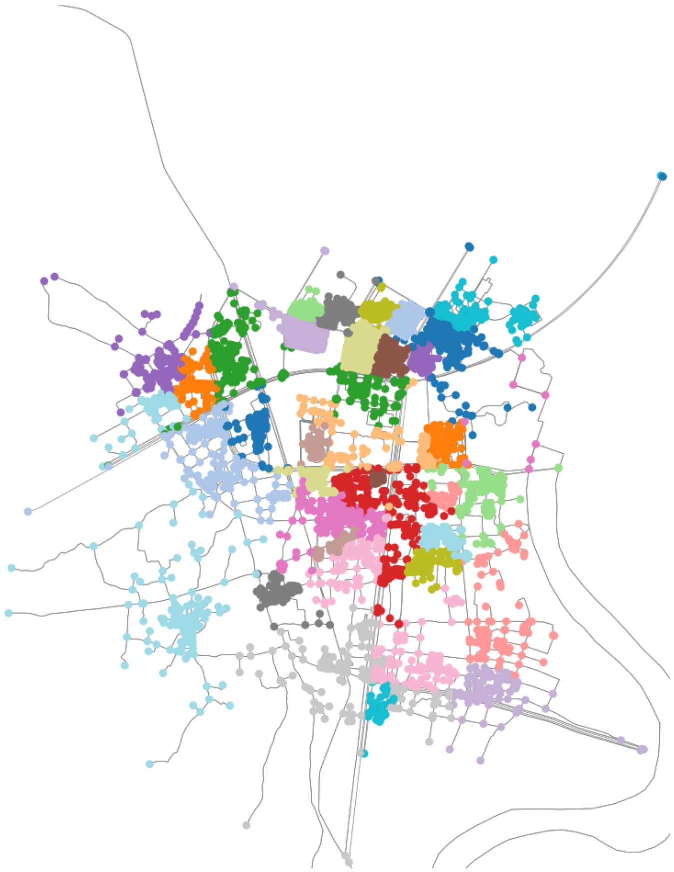

Figure 7: Louvain Community Detection, Modularity = 0.94

The **Modularity score** came out to be **0.94**.

> **Problem 4: Assortativity & Degree-Degree Correlations**
>
> (a) **Degree Assortativity:** Do high-degree nodes connect preferentially to other highdegree nodes?
>
> (b) **Degree-Degree Correlation Plot:** Plot average neighbor degree vs. node degree.

Solution 4(a). Assortativity and disassortativity describe how nodes in a network connect based on their attributes.
In an Urban Road network, there is no preferential connection between high degree nodes. This is also evident via the **Pearson Coefficient** which comes out to be **0.007** which indicates there is no significant assortative/disassortative mixing.

Solution 4(b). The plot is shown as follows.

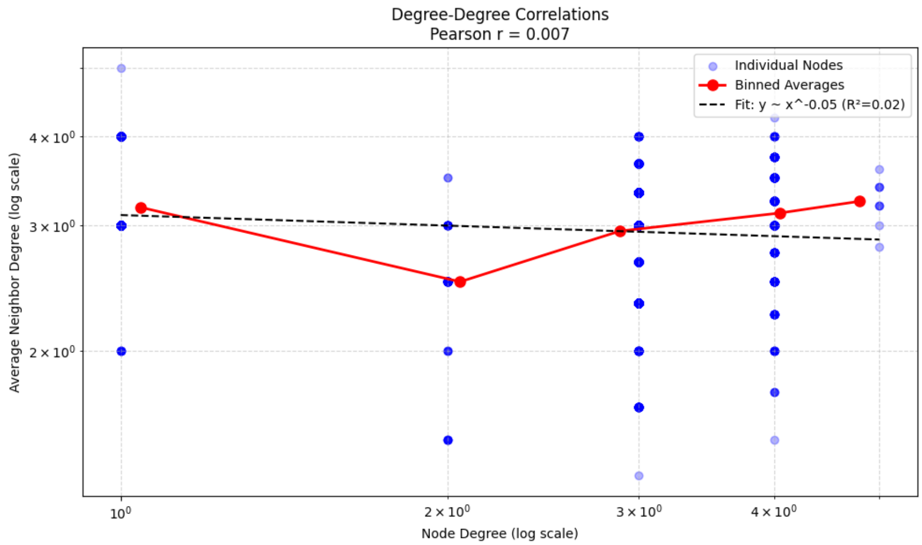

Figure 8: Degree-Degree Correlation Plot

Since r **(Pearson Degree Assortativity Coefficient)=0.007** and R^2 = **0.02**, the network exhibits **Neutral Assortativity** which means neither assortative nor disassortative.

Node Connections are random with respect to degree in this network.

> **Problem 5: Clustering Coefficients**
>
> **Analysis:**
>
> (a) Compute the global clustering coefficient (transitivity).
>
> (b) Compute local clustering coefficients for each node (or for a subset)

Solution 5(a). The global clustering coefficient (GCC) is a measure of how interconnected nodes are in a network, specifically quantifying the tendency for nodes to form triangles (sets of three nodes where each node is connected to the other two). It answers the question: "How likely is it that two neighbors of a node are also neighbors of each other?"

GCC = 1: Every triplet forms a triangle (maximally clustered, e.g., a fully connected network).
GCC = 0: No triangles exist (e.g., a tree-like structure or purely random network with no clustering).

Comparison to Local Clustering Coefficient:

Global Clustering Coefficient: Measures clustering for the entire network.

Comparison to Local Clustering Coefficient: Global Clustering Coefficient: Measures clustering for the entire network.

Local Clustering Coefficient: Measures clustering for individual nodes (the average of these gives the average local clustering coefficient, which is not the same as the GCC) Measures clustering for individual nodes (the average of these gives the average local clustering coefficient, which is not the same as the GCC)

The Global clustering coefficient comes out to be **0.0437** as roads rarely form a connected triangle triplet. Roads are built hierarchically (e.g., highways → arterials → local streets) to avoid unnecessary loops.

Table 1: Top 10 Nodes by Degree - Local Clustering Coefficients

Node ID	**Degree**	**Local Clustering Coefficient**
3469607001	5	0.0000
4038755388	5	0.0000
4038723555	5	0.0000
2476545271	5	0.0000
6621517666	5	0.1000
6390642238	5	0.1000
8395811570	5	0.1000
6390642252	5	0.2000
6390642260	5	0.2000
3832369436	4	0.0000

For a ER graph with nodes = 4217 and edges = 10576 the Global Clustering Coefficient (Transitivity): **0.001525** Average Local Clustering Coefficient: **0.001425**.

The lower values of clustering coefficients suggest that Graph is random.

1.1 Codes and Results:

The link to the Google Colab files are attached below.

1. Problem 1
2. Problem 2
3. Problem 3
4. Problem 4
5. Problem 5