Problem Chosen	2025 MCM/ICM Summary Sheet	Team Control Number
D		2517674

The Graph-Based Traffic Optimization Model for Baltimore City

Summary

Changes in the transportation system can have diverse impacts on different stakeholders. Aiming to enhance residents' living experiences and improve their quality of life, it is necessary to optimize the transportation network in Baltimore by improving infrastructure and strengthening public transportation.

Initially, we pre-process and clean the data in the table. **Data processing** encompasses handling missing values, outliers, as well as matching, simplifying, transforming, and supplementing relevant data. Subsequently, with the cleaned data, we utilize relevant graph - theory and topological - structure knowledge to construct the overall road network, the bus network, and the associated data flows. However, in the initially constructed road network, numerous edges were fragmented, and the nodes were not integrated. By leveraging the proximity analysis and simple clustering among points through the Python interpreter in **ArcGIS Pro**, we obtained **a regularized route network**.

For Problem 1, we simulate the collapse and reconstruction of the Francis Scott Key Bridge as the presence or absence of relevant nodes and edges in the graph. The impacts on the transportation system are specifically manifested as changes in public transportation (such as traffic flow, congestion, and commuting time) and the interests of different stakeholders (pedestrians, commuters, the government, drivers, and tourists). Specifically, **Dijkstra's Algorithm** is employed to determine the shortest path. **The l-covering number method** is adopted to indirectly determine the traffic flow of roads with missing flow data, thereby visualizing the traffic - flow heat map. The AHP is used to determine the weight coefficients and then establish the impact indicators.

For Problem 2, we select the viaduct as a project that affects the Baltimore bus system. First, **the KNN clustering algorithm** is applied to obtain representative key nodes, and a sub - graph containing only these key nodes is constructed. Then, considering the construction cost, under the constraint that the route length is less than 12 kilometers, we aim to maximize the sum of the weights of the key nodes that the viaduct route passes through. Thus, we use **the back-tracking algorithm based on depth-first search** to obtain the viaduct construction route. Subsequently, a step-distribution strategy is adopted to re-allocate the road traffic flow, and the model and relevant indicators from Problem 1 are applied to analyze the impacts before and after the viaduct construction.

For Problem 3, regarding how to demonstrate the "most" characteristic of the project, our approach : First, we identify seven candidate projects. Then, **AHP** is used to evaluate these projects based on indicators related to improving residents' lives. Eventually, the project of filling and renovating Highway 40 yields the best results. By referring to literature containing public - opinion surveys, we find that our project indeed meets the needs of the majority of residents. Next, the model in Problem 1 is used to quantify the impact on other stakeholders. Then, **the minimum-cut analysis** of the overall traffic road network is conducted, and the clustering algorithm is used to analyze the public - transportation road network to obtain the interference of traffic demand. Finally, we have written a memorandum to summarize the proposed project and analyze its impacts.

Keywords: Transportation system; Graph theory; Traffic flow network; KNN clustering; Depth-First Search with Backtracking;

Contents

1 Introduction

1.1 Problem Background

Nowadays, transportation[1] system has two sides. A well-developed transportation system can bring benefits in many ways, such as the business economy, population movement, tourism, etc., but at the same time, it can also cause different benefits and losses to various stakeholders, because different groups of people have different needs for the transportation system, and naturally the benefits that the transportation system brings are different. For example, urban sidewalks and traffic signals benefit pedestrians but delay motorists, while highways are the opposite. In addition, transportation systems are also affected by natural conditions such as the topography of the city, the watershed, and weather conditions[2].

1.2 Restatement of the Problem

With the goal of improving the experience and quality of life for residents, we need to optimize the transportation network in Baltimore City by improving infrastructure and enhancing public transportation. Thus, in order to analyze the problems in a clear and systematic way, we restate the problem as follows:

- Based on the network model developed, analyze the impact that the collapse and reconstruction of the Francis Scott Key Bridge will have on the transportation system.
- Select a project that would impact transit or sidewalks and analyze the impact.
- Recommend a project that would best improve the lives of Baltimore residents and analyze the impact it would have on other transportation needs and people's lives.

The results of all of the above problems need to highlight the impacts on residents and other stakeholders.

1.3 Problem Analysis

To address the three traffic - related issues above, the primary task is to pre-process and clean the data in the table first. Data processing involves handling missing values, outliers, as well as matching, simplifying, and transforming relevant data. This can be achieved using Python and Excel for analysis. For more detailed data processing, refer to the data processing section. Additionally, some relevant data should be supplemented using the official website of Baltimore City. Given the large amount of data in this problem, data processing is of utmost importance for model building. Therefore, this part is extremely crucial.

Construction of Graph - Theoretic Models. By using the cleaned data, relevant graph - theoretic and topological structure knowledge is applied to construct the overall road network, the bus network, and the associated data flow. The node - related information and attributes in the nodes_all.csv file are used to construct the nodes in the network model, while the edge - related information and attributes in the edges_all.csv file are used to construct the edges of the network model. Certain relevant indicators in the tables are employed to define the weights of the edges in a weighted manner. Subsequently, the Bus_Routes.csv and

Bus_Stops.csv files are utilized to construct the bus network, preparing for the analysis and solution of specific problems. For the construction of relevant graph - theoretic models, refer to the "graph construction" section below.

For Problem 1, we simulate the collapse and reconstruction of the Francis Scott Key Bridge as the presence or absence of relevant nodes and edges in the graph. The impact of this on the transportation system is specifically manifested as changes in public transportation (such as traffic flow, congestion, and commute time) and the interests of different stakeholder (pedestrians, office workers, the government, drivers, and tourists). When the transportation system changes, pedestrians consider the length of the route and walking time, drivers consider the route length, traffic volume, and congestion, commuters consider the commute time, the government considers the impact on economic benefits, and tourists consider the route length between different attractions and accessibility. Among them, Dijkstra's Algorithm is used to find the shortest path, and the adjacency matrix and reachability matrix of the graph are used to analyze the accessibility. The maximum bearing capacity of the network traffic flow can be analyzed by solving the maximum - flow and minimum - cut problems. The definition and implementation of relevant indicators are detailed in the construction of Problem 1 below.

For Problem 2, considering the impact on buses and pedestrian walkways, the project we selected is to build a viaduct with bus lanes and pedestrian walkways. Since constructing a viaduct requires a large amount of capital, it is necessary to achieve a better traffic - diversion effect within a certain route length. Therefore, we first use K - Nearest Neighbors (KNN) clustering based on the traffic flow of all routes to obtain a certain number of main nodes, and visualize them. It is specified that the viaduct route needs to pass through several main nodes to achieve more effective traffic diversion and thus reduce traffic congestion. However, the length limit of the viaduct also needs to be considered. Thus, a graph - topological structure is constructed with the conditions of maximizing the node weights and restricting the route length. The specific location for viaduct construction is obtained through a backtracking algorithm based on depth - first search. After that, a step - distribution strategy is used to re - allocate the traffic flow caused by the viaduct, and then the relevant indicators are evaluated using the model from Problem 1 for measurement.

For Problem 3, we are motivated to recommend a project that can best improve the lives of Baltimore residents. To identify the project with the characteristic of "being the best for improvement", we determine seven candidate projects and analyze the indicators related to improving residents' lives to select the optimal one. These indicators are still based on the relevant data of the transportation network, and then the Analytic Hierarchy Process (AHP) is used to determine the relevant weights. Eventually, we find that the project we proposed is more advantageous. Subsequently, we analyze the convenience of this project for residents' travel and its impact on other stakeholders, and solve these problems using the model constructed in Problem 1. Then, we analyze the traffic demand and the interference with people's lives. The traffic demand is analyzed using the public transportation network and the overall road network, while the interference with people's lives is analyzed based on the specific content of the project.

1.4 Our Work

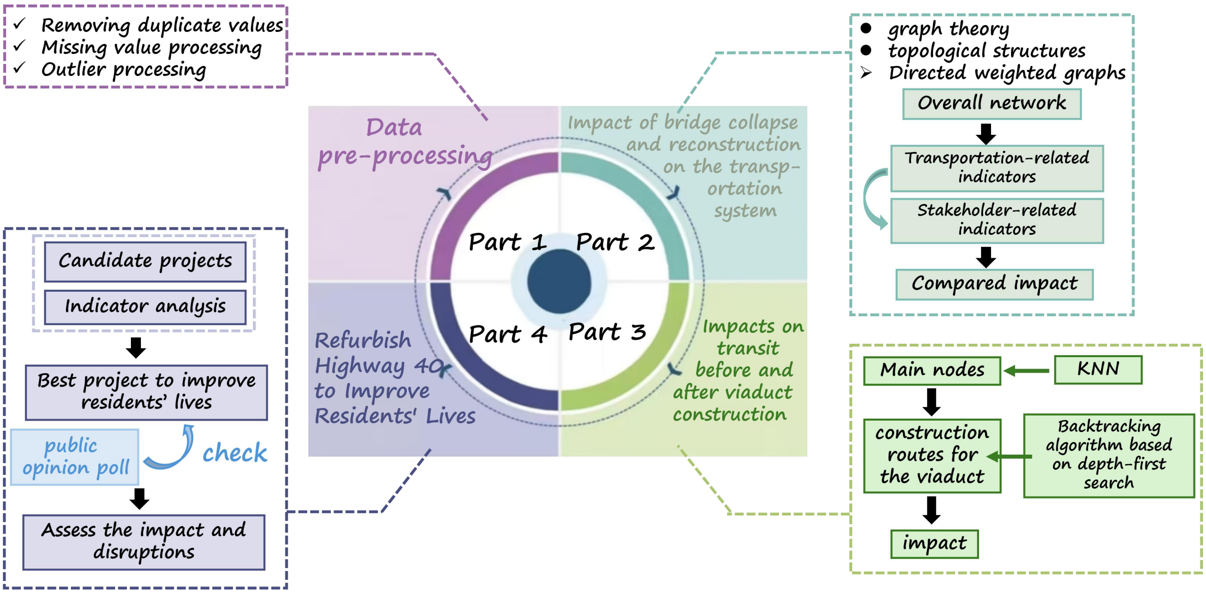


Figure 1: Our Work

2 Assumptions and Justifications

▼**Assumption 1:** The impact on residents' lives in this issue is largely influenced by traffic.

▲**Explanations:** This is a transportation-related issue, and indicators related to residents' lives can only be constructed using transportation data.

▼ **Assumption 2:** This issue is related to transportation, and the indicators related to residents' lives can only be constructed using transportation data.

▲**Explanations:** The model we built must be based on accuracy data. For some identifiable missing values and duplicate values, they can be handled, but outliers cannot be recognized.

▼ **Assumption 3:** The commuting time, congestion levels, traffic flow, and other related factors are solely dependent on the available or retrieved data, with no consideration for other influencing factors.

▲ **Explanations:** Only by disregarding other unknown factors can the model be correctly constructed.

3 Notations

Table 1: Notations used in this paper

Symbol	Description
L	the length of roads
F	traffic flow
J	congestion level (esp. in traffic jam)
t	passing time
Impact_walk	impact of transportation system benefits on pedestrians
impact_driver	impact of transportation system benefits on drivers
impact_tourist	impact of transportation system benefits on tourists
impact_gvm	impact of transportation system benefits on government
impact_office	impact of transportation system benefits on office workers
α、β、μ、θ、γ、a、b、c、d	the weights of jam/t/impact_office/impact_driver

4 Data pre-processing

4.1 Removing duplicate values

Edge_Names_With_Nodes.csv: For the same street, the nodes on the street are the same, so only the first occurrence of the street is retained, and the later recurrences are deleted.

nodes_all.csv &nodes_drive.csv: The street identifiers that appear in the first "osmid" column should be included in Edge_Names_With_Nodes.csv, so delete all rows in the table that are not included.

4.2 Missing value processing

nodes_all.csv & nodes_drive.csv: For some indicators, since there are an excessive number of missing values and they are insignificant columns, such as the "highway" column, "ref" column, "junction" column, etc., these indicators are directly removed.

edges_all.csv & edges_drive.csv: For some indicators, due to an excessive number of missing values and their insignificance, such as the "ref", "lanes", and "area" columns, these indicators are directly removed from consideration in the model. However, for indicators like "max_speed" and "width", although they have missing values, they are crucial for the model analysis. Therefore, we supplement these values by querying data from the official website. Regarding the "access" column, if a value exists, it indicates that the road is inaccessible, so the corresponding paths are deleted. The "bridge", "junction", and "tunnel" columns can be used to describe the road types and can serve as visualization elements for ArcGIS Pro.

MDOT_SHA_Annual_Average_Daily_Traffic_Baltimore.csv: A significant amount of data is missing in the "node start" and "node end" columns. After examining all the indicators contained in the csv files, it is found that these missing values cannot be filled based on the existing data. Therefore, all rows with missing node data are deleted.

Bus_Stops.csv&Bus_Routes.csv: After screening, it is found that there are no missing values.

5 General Modeling

5.1 graph construction

We plan to use a directed weighted graph model based on graph theory and relevant topological structures to describe the transportation network. The graph consists of the following three main components:

Node Set: Nodes represent locations in the transportation network, such as bus stops, road intersections, starting and ending points of a road segment, etc. The determination of nodes is based on nodes_all.csv.

Edge Set: Edges represent the connection relationships between nodes, indicating the connectivity and accessibility between different locations in the transportation network. An edge represents a possible traffic route, such as a road, a sidewalk, or a bus route. The determination of edges is based on edges_all.csv. The direction of an edge can be obtained by integrating the “oneway” column and the “reversed” column.

Edge Weight Set: Weights are used to quantify the attributes of edges. The attributes of an edge include the length of the road.

First, we utilized ArcGIS Pro in combination with the Python interpreter to visualize all the nodes, thereby identifying the regions where the nodes are located. Through an investigation of the information regarding Baltimore, we found that the areas of Baltimore include urban and suburban regions. The visualizations of the node - related information for the urban and suburban areas are as follows.

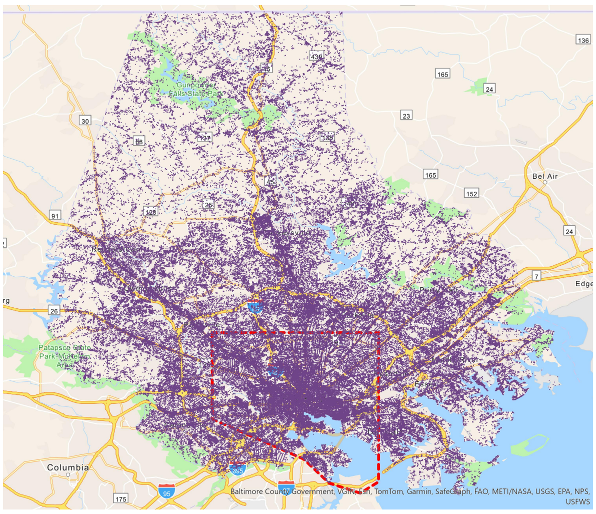


Figure 2: Node distribution map

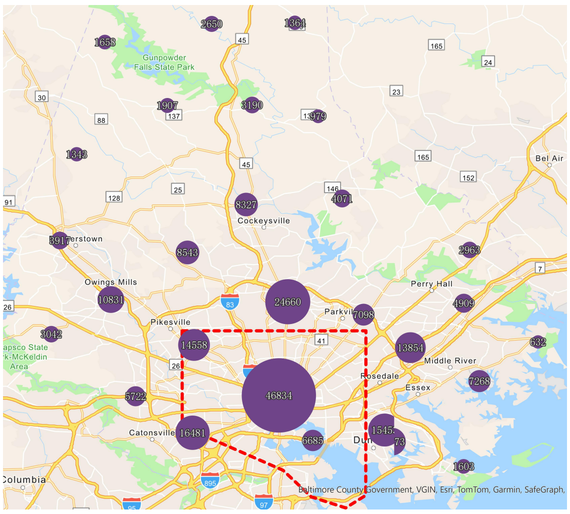


Figure 3: Node clustering

To make the distribution of points clearer, enabling us to observe the density of different points and edges, and thus understand the traffic distribution, we conducted a K - Nearest Neighbors (KNN) clustering analysis on the above - mentioned points and edges. The results are shown in Figure 3.

Regarding the relevant edge information, we indexed the edges using their unique identifiers and constructed a multi - dimensional matrix representing the various attributes of the edges. This matrix is intended for use in the subsequent construction of the network model. First, we visualized the edges.

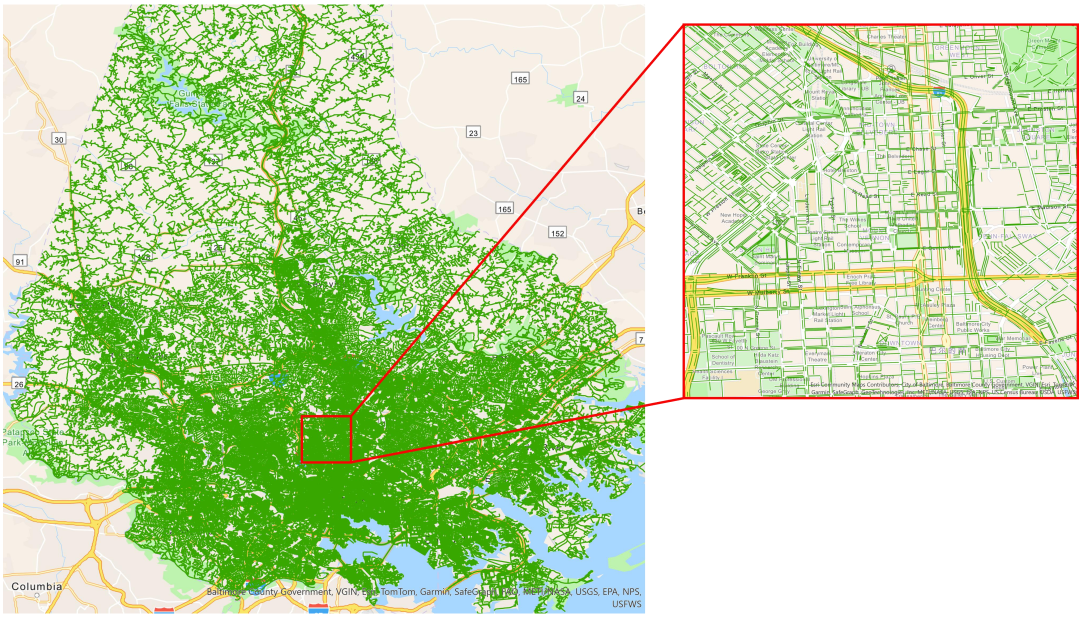


Figure 4: Local route magnification

The above is the visualization using edge information. However, due to the large number of edges, upon magnification, we found that many edges split, and the points did not merge together, making it impossible to analyze the subsequent model. With the help of ArcGIS Pro and the Python interpreter, we used proximity analysis between points combined with simple clustering to obtain a regular route network. Since we need to analyze the network - related changes in bridge collapse and reconstruction, we only performed such processing on the urban area of Baltimore and the county on the right, obtaining the following results:

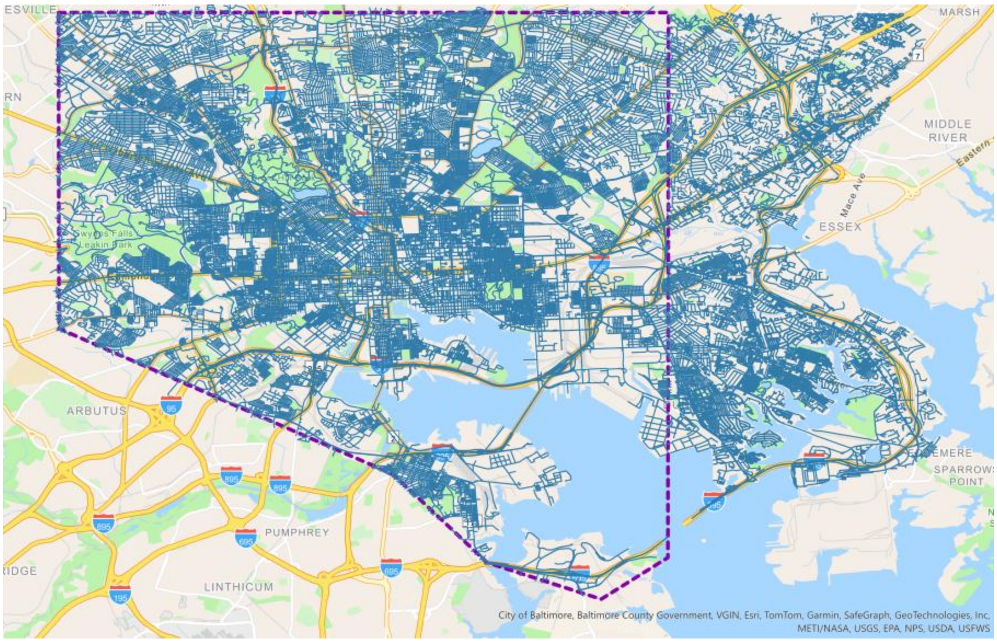


Figure 5: Filtered urban road network

Conduct a comparison between the pre - regularization routes and the post - regularization routes:

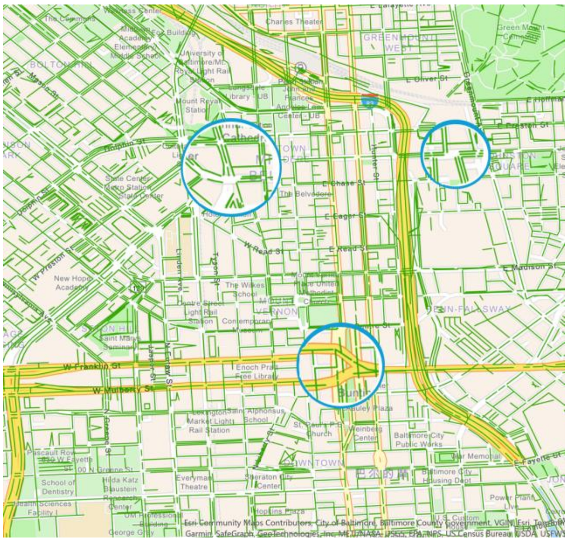


Figure 6: A discontinuous route map

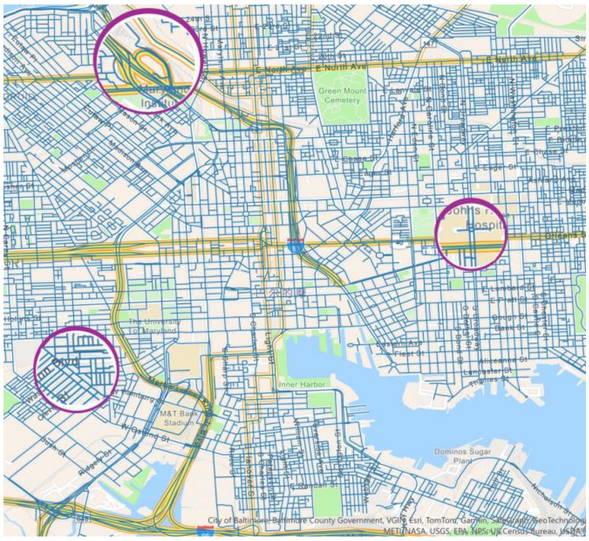


Figure 7: A continuous route map

Up to this point, the relevant road graph has been successfully constructed, and model analysis can be carried out.

Next, we will construct a public transportation route map. This is accomplished using the Bus_stops.csv and Bus_Routes.csv files. First, we use the Bus_stops.csv file to construct the bus stop nodes.

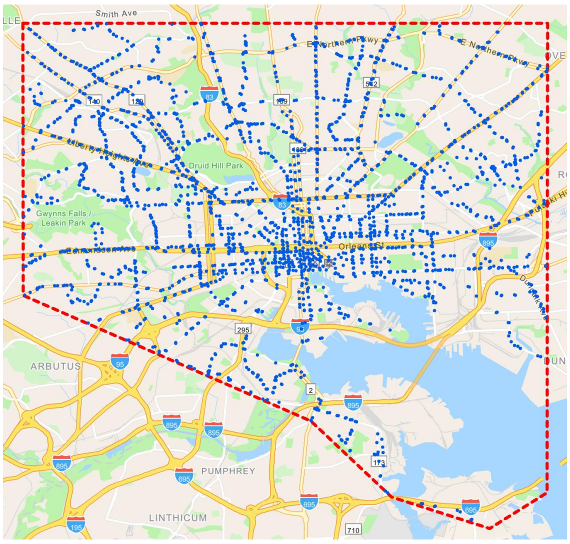


Figure 8: Distribution map of bus stops

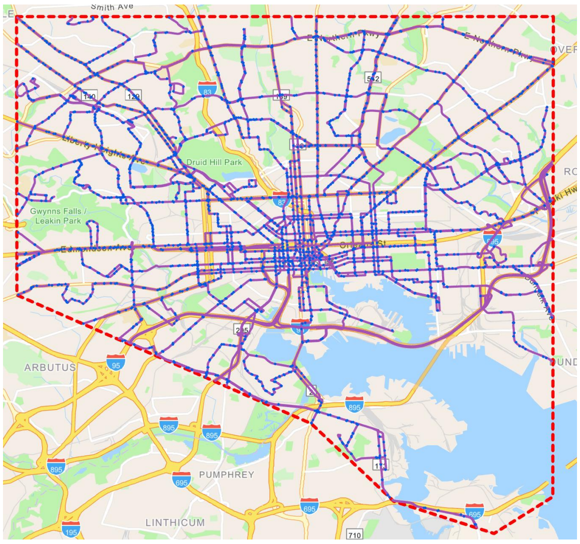


Figure 9: Distribution map of bus routes

Then, the Bus_Routes.csv file, which contains information about transportation routes, is utilized. The data is imported into ArcGIS Pro and visualized using Python code. This visualization facilitates subsequent indicator analysis of the traffic flow of public transportation routes and the passenger volume of buses.

6 Problem 1

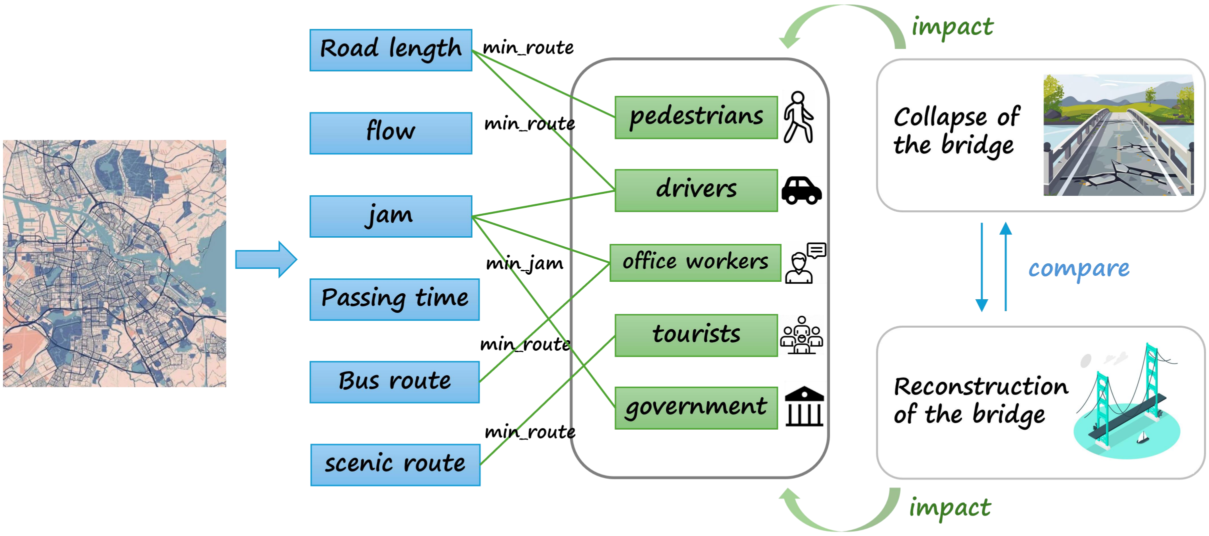


Figure 10: Mindmap of problem 1

6.1 Definition of transportation-related concepts

- **The length of Road:** It is directly determined by the "length" column in edges_all.csv.
- **Flow:** It is divided into two cases: all days and only weekdays, represented by AADT and AADWT in MDOT_SHA_Annual_Average_Daily_Traffic_Baltimore.csv respectively. We found that there is a large amount of missing data in this table, and the traffic flow of some roads cannot be obtained directly. Therefore, we use the l - covering number method to indirectly determine the traffic flow of these roads. The following introduces the l - covering number method:

ℓ-covering number. Let X be a finite set. A graph is a pair (X, E), where E is a symmetric subset of $X \times X \setminus \{(x, x) : x \in X\}$. We call $x \in X$ a vertex (or a node) and $(x, y) \in E$ an edge.Denote d(x, y) the minimal length of paths from x to y if x and y can be connected by a path; d(x, y) = 0 if x = y; and $d(x, y) = +\infty$ if x and y cannot be connected by a path. Then it is clear that d(x, y) is an integer value metric. Recall that the degree of a vertex $x \in X$ is the total number of edges connected to x, i.e., $\deg(x) = \#\{(x, y) : (x, y) \in E\}$ (#A is the total number of elements in A).

Let $\ell > 0$ be a fixed natural number, a subset $X^* \subset X$ is called an ℓ-cover set of X, if $X \subset S\, \{B(x', \ell) : x' \in X^*\}$, where $B(x', \ell) = \{x \in X : d(x', x) \le \ell\}$ is the ball centered at x′ with radius ℓ. This means X is covered by the balls $\{B(x', \ell) : x' \in X^*\}$. We call the following quantity the ℓ-covering number of X:

$$N_l^c = (N_l^c(X) =) \min\{\# X^*\}$$

In order to apply the above - mentioned method to our model, a variant of the algorithm is developed. The pseudo - code of the algorithm is as follows:

Algorithm 1: Calculate Traffic Flow

Input : Edges set E (each edge has attributes: coordinates, trafficFlow)

```
Output : Updated edges set E with estimated trafficFlow
for edge ∈ E do
    if edge.trafficFlow is None then            // Check if traffic flow is unknown
        node ← randomly Select Node(edge)
        (latitude, longitude) ← node.coordinates
        range Latitude ← initial Latitude Range
        range Longitude ← initial Longitude Range
        containedEdges ← empty set
        while True do
            bBox ← getBoundingBox(latitude, longitude, rangeLatitude, rangeLongitude)
            containedEdges ← findEdgesInBoundingBox(bBox, E)
            if length(containedEdges) > 3 then
                break                            // Stop if more than 3 edges are found
            rangeLatitude ← rangeLatitude + increaseLatitudeRange
            rangeLongitude ← rangeLongitude + increaseLongitudeRange
        knownTF ← [edge.trafficFlow for edge∈containedEdges where edge.trafficFlow is
                   not None]
      if length(knownTF) > 0 then
          edge.trafficFlow ← average(knownTrafficFlows)
      else
          edge.trafficFlow ← 0                  // Set traffic flow to 0 if no known flows
      return E
```

First, by adding the traffic data of relevant detection points provided on the official website to MDOT_SHA_Annual_Average_Daily_Traffic_Baltimore.csv, and then visualizing it using ArcGIS Pro, we can obtain:

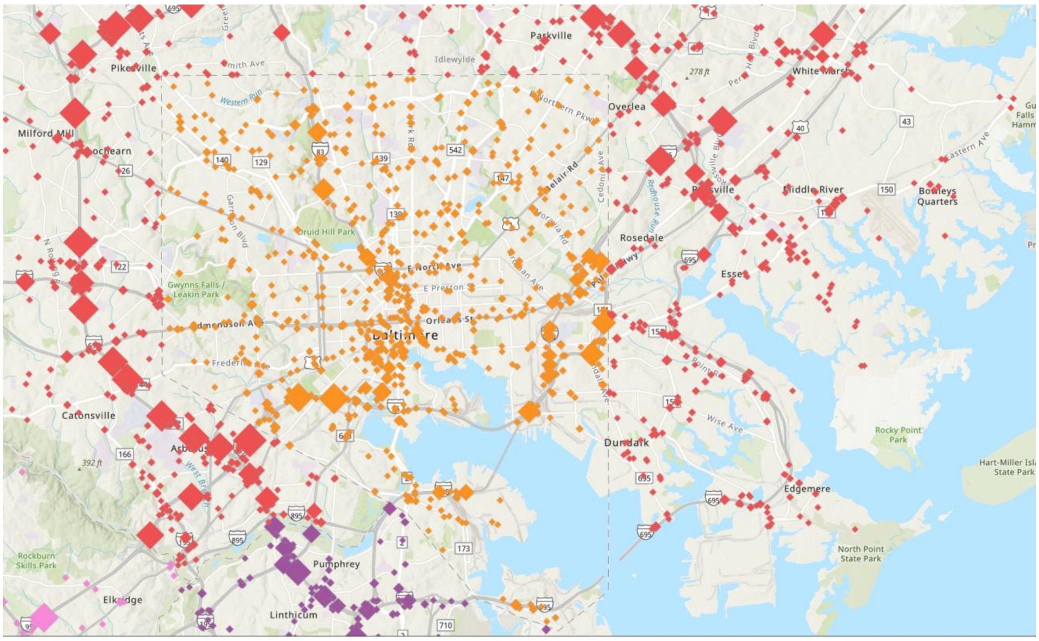


Figure 11: Distribution map of traffic flow detection points

Among them, the marked points on the graph represent the traffic flow detection locations. Each marked point contains AADT and AADWT data from 2014 to 2023. Since the bridge broke in 2023, we use the data from 2014 to 2022 as a benchmark. A logistic regression model is employed to predict the AADT and AADWT data in 2023 under the condition that the bridge did not break. This predicted data is used as the traffic flow data after reconstruction. Consequently, a traffic network traffic - flow model comparison can be established. The modified l - covering method is used to calculate the traffic flow of each road, and a heat map is created.

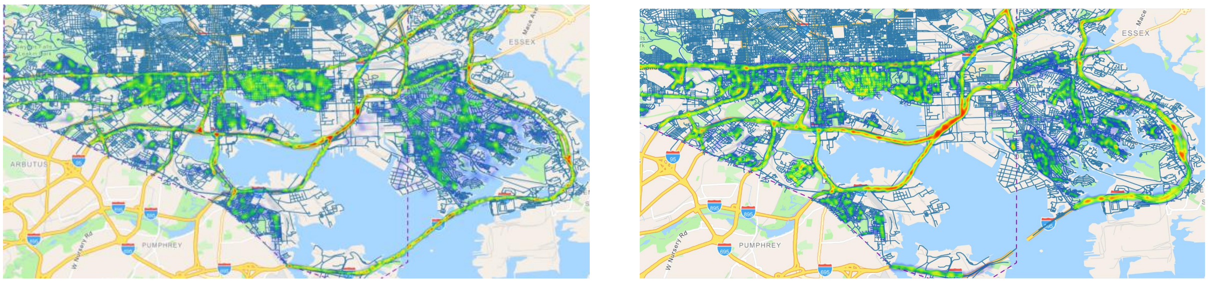


Figure 12: Traffic flow heat map, Left: Reconstruction, Right: Collapse

Through the heat map, it can be directly observed that the collapse and reconstruction of the bridge evidently lead to significant differences in the average traffic flow. The collapse significantly increases the traffic pressure on the middle part of the bridge, while after reconstruction, the original traffic pressure is evidently diverted, alleviating the traffic - flow pressure in the urban transportation system.

Among them, for the traffic flow used in the comparison, that is, the traffic flow of the reconstructed bridge, the data before the bridge broke is utilized, and the exponential smoothing method is employed for data prediction:

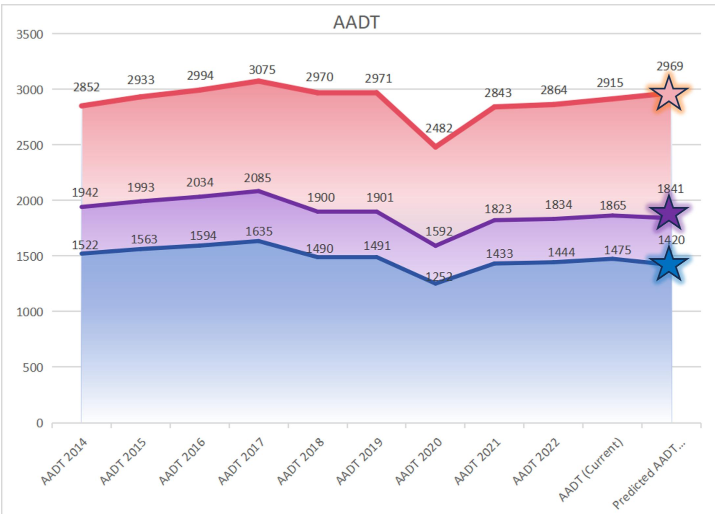


Figure 13: Traffic flow prediction for three detection points

The last data point is the total traffic flow of the reconstructed roads, which is what we are looking for. Here, we only present the predictions of traffic flow at three selected points.

- **Degree of congestion--jam:** First, obtain the lane width from the "width" column in edges_all.csv, and the number of lanes from the "lanes" column. Then, obtain the number of roads connected to the starting and ending points of the road from the "street_count" column in nodes_all.csv, denoted as count1 and count2 respectively. Subsequently, define the congestion degree "jam" as follows, where a, b, c, and d are weight coefficients.

$$jam = a * flow + b * width + c * lanes + d * (count1 + count2)$$

- **Travel time--t:** First, obtain the maximum speed vm from the "maxspeed" column in edges_all.csv. Define t as follows:

$$t = \frac{len}{vm} + \alpha * jam$$

6.2 Definition of concepts relevant to different stakeholders

- **Pedestrians:** Define the impact on pedestrians' interests when there are traffic changes as follows:

$$impact\_walk = \min\_route$$

Here, $\min\_route$ represents the average value of the shortest - path lengths. The start and end points are randomly selected from the entire route, and the shortest paths are obtained using Dijkstra's Algorithm.

- **Drivers:** Define the impact on drivers' interests when traffic conditions change as follows:

$$impact\_driver = \beta * \min\_route + \gamma * jam$$

- **Office workers:** To simplify the problem, we only consider office workers commuting by bus. Define the impact on the interests of office workers when traffic conditions change as follows:

$$impact\_office = \theta * l\_bus + \mu * jam$$

Where l_bus is the average value of the shortest - path lengths obtained by randomly selecting starting and ending points within the bus routes.

- **Tourists:** Define the impact on tourists' interests when traffic changes as follows:

$$impact\_tourist = l\_\text{scenic}$$

where l_scenic is the average value of the shortest -path lengths obtained by randomly selecting the starting and ending points from all the scenic spots on the entire route.

- **Government:** Considering that the government mainly hopes to minimize the congestion level in transportation, the impact on the government's interests when traffic conditions change is defined as follows:

$$impact\_gvm = sum\_jam$$

where sum_jam represents the degree of congestion obtained by adding up that of all roads.

For the analysis of scenic areas, the data provided by the official website of Baltimore City is used for visualization. In the subsequent establishment of model indicators, only the nodes of scenic areas need to be considered.

6.3 Analytic hierarchy process to calculate weights

The weight coefficients in the above definitions are determined by first reading relevant literature, using prior knowledge to determine the importance comparison between each pair, and then applying the Analytic Hierarchy Process (AHP)[4]. Consider each set of factors involved in each definition as a system, and the steps of using AHP are as follows:

1) Analyze the relationships among the factors in the system and establish a hierarchical structure of the system.
2) Compare the importance of elements at the same level pairwise with respect to a certain criterion in the upper level, and construct a judgment matrix.
3) Calculate the relative weights of the compared elements with respect to this criterion from the judgment matrix. To ensure the robustness of the results, the arithmetic mean method, geometric mean method, and eigenvalue method are used for calculation respectively, and then the average value is taken. This can avoid the errors caused by a single method and make the conclusion more comprehensive and effective.
4) Conduct a consistency check, and it is found that all pass the consistency check.

After obtaining all the weight coefficients, the above formulas are summarized as follows.

$$jam = 0.556 * flow + 0.076 * width + 0.126 * lanes + 0.242 * (count1 + count2)$$

$$t = \frac{len}{vm} + 0.5 * jam$$

$$impact\_walk = \min\_route$$

$$impact\_driver = 0.17 * \min\_route + 0.83 * jam$$

$$impact\_office = 0.125 * l\_bus + 0.875 * jam$$

$$impact\_tourist = l\_\mathrm{scenic}$$

$$impact\_gvm = sum\_jam$$

Conduct 100 experiments for each of the two scenarios: bridge collapse and bridge reconstruction. In each experiment, randomly select one point from each of the two regions connected by the bridge, and calculate the shortest path. This enables us to obtain the specific values of min_route, l_bus, and l_scenic, thereby determining each impact indicator. For public transportation, utilize the public transportation route map. For tourists, only select scenic area nodes in the two regions connected by the bridge. The specific data is as follows[3].

Comparison of results after 100 experiments		
	Collapse	Reconstruction
Flow	5346.67	5467.83
Jam	3453.36	2756.73
Pedestrians	1543.68	653.78
Drivers	3128.71	2399.23
Office workers	2956.35	2293.46
Tourists	890.35	687.46
Government	534667.33	546783.45

1. **Increased Traffic Flow:**The reconstruction of the bridge typically improves traffic capacity by providing more lanes or optimizing traffic flow, which may lead to more vehicles choosing to use the newly built bridge. This is particularly true when the bridge connects busy traffic areas, which can result in a noticeable increase in traffic volume.

2. **Reduced Congestion:**The new bridge offers a shorter route, relieving traffic pressure on other bridges, thereby reducing congestion for passing vehicles. This leads to smoother traffic flow and reduced bottlenecks.
3. **Improved Pedestrian Access:** A wider bridge design can increase pedestrian space, especially when the original bridge was not suitable for pedestrians. The rebuilt bridge can provide better walking paths, offering safer and more convenient access for people on foot.
4. **Shortened Commute Time for Drivers:** For drivers, a redesigned bridge can more effectively organize traffic flow, allowing them to pass through more quickly and reduce commute times. This improvement enhances driving efficiency and reduces delays.
5. **Benefits for Commuters:** Commuters often face heavy traffic during rush hours. After the bridge reconstruction, the increased traffic capacity can significantly shorten commute times, boosting work efficiency and making daily travel more predictable.
6. **Benefits for Tourists:** For tourists, the reconstructed bridge makes travel within the city more convenient and efficient, especially when the bridge connects tourist attractions or transport hubs. This enhances the overall travel experience for visitors.
7. **Impact on the Government:** By easing traffic congestion and improving traffic flow, the government can reduce the frequency of traffic accidents, thus improving public safety. Additionally, higher traffic efficiency contributes to reduced vehicle emissions, which has a positive impact on the environment.

7 Problem 2

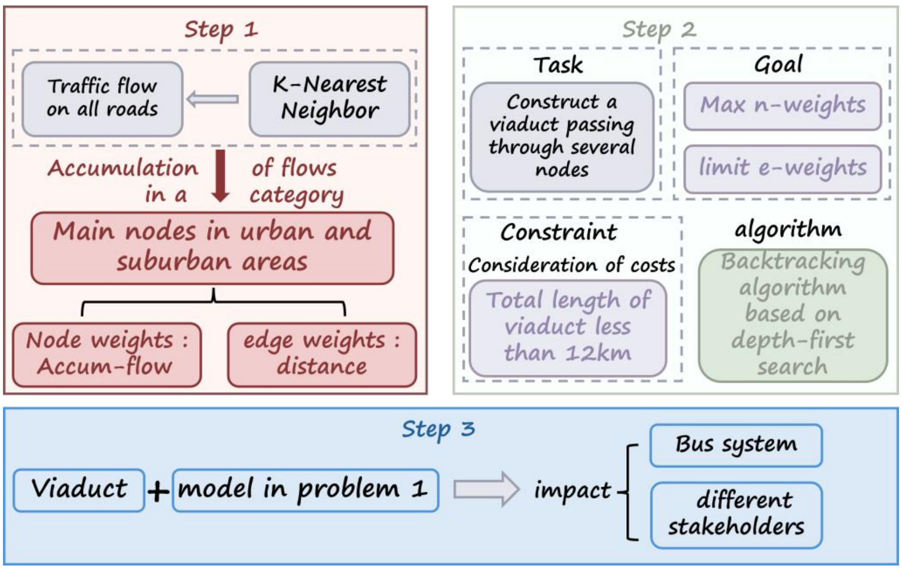


Figure 14: Mindmap of problem 2

We have chosen the viaduct as a project that has an impact on the Baltimore bus system for the following reasons:

- Enhancement of bus operation efficiency: Viaducts generally reduce interference from ground - level traffic. When buses travel on viaducts, they can avoid frequent red - light stops and traffic congestion on ground roads. Dedicated bus lanes on viaducts enable buses to pass quickly, significantly shortening the running time of buses and improving their punctuality rate. This, in turn, attracts more residents to choose bus travel[5].

- Diversion of traffic flow: Viaducts provide new travel routes for vehicles, especially in areas where ground - level roads are severely congested. They can divert the traffic flow that was originally concentrated on the main ground - level roads onto the viaducts, thus alleviating the traffic volume pressure on ground - level roads.
- Improvement of overall traffic operation efficiency[6]: Viaducts usually have a relatively high designed speed and fewer traffic interferences. Vehicles can maintain a relatively stable driving speed on viaducts. This contributes to improving the operation efficiency of the entire transportation system. By reducing the number of vehicle accelerations, decelerations, and waiting times at stops, traffic congestion can be mitigated.

7.1 Construct sub-graph containing major nodes

Apply the K - Nearest Neighbors (KNN) clustering algorithm to the nodes on all routes in both urban and suburban areas of Baltimore. Nodes with similar distances are regarded as belonging to the same category. In this way, representative major nodes can be obtained. The steps to implement the KNN algorithm are as follows:

1) Calculate distances: Given an unclassified node, calculate its distance to each node in the classified nodes.
2) Find neighbors: Identify the K classified nodes that are closest to the unclassified node as the neighbors of the unclassified node.
3) Perform classification: Determine the category to which the unclassified node belongs based on the category to which the majority of these K neighbors belong.

The major nodes obtained after applying the KNN clustering algorithm are shown in Figure 15.

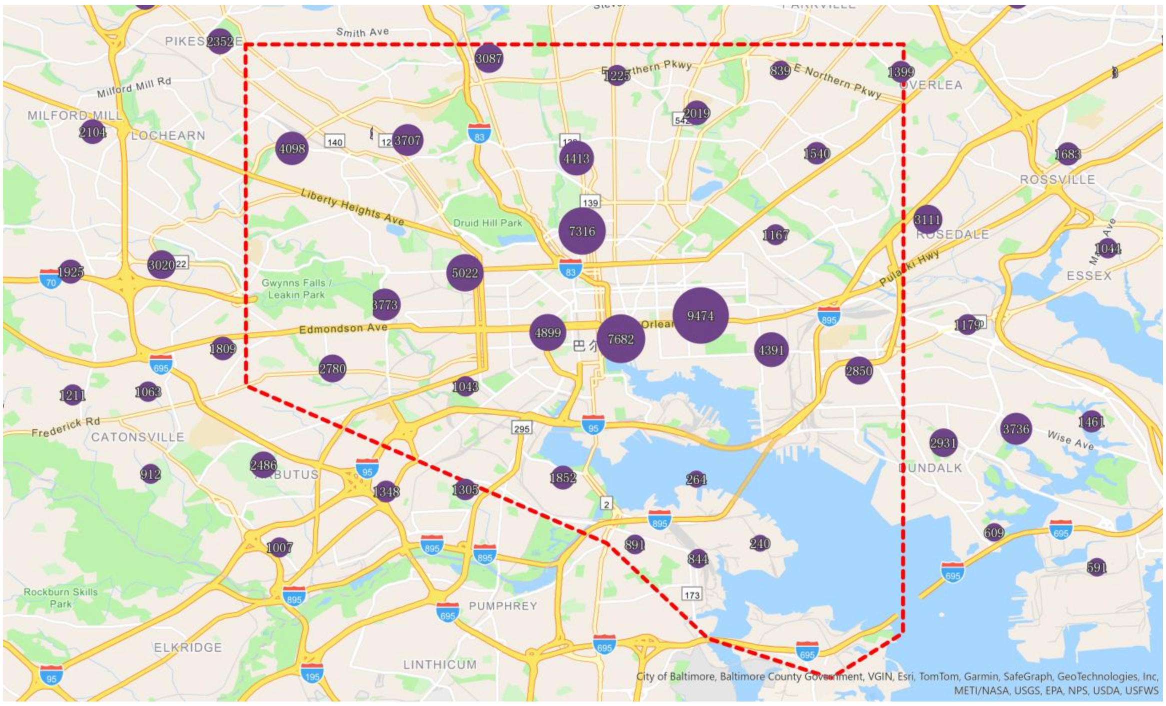


Figure 15: Sub - graph of main nodes

The weight of each major node is the sum of the traffic flow of all nodes in its corresponding category. The edge weight between major nodes is the straight - line distance between the nodes. Thus, the sub - graph containing only major nodes is constructed.

7.2 Get the viaduct route

Our task is to construct a viaduct that passes through several major nodes. Considering the construction cost, under the constraint that the route length is less than 12 kilometers, we aim to maximize the sum of the weights of the major nodes along the viaduct route. The purpose of this is to divert traffic from areas with high traffic volume, thereby alleviating traffic congestion in those areas.

Let x_a and x_b be the starting and ending points. Suppose there are n nodes on this route. The weight of the i- th node is NW_i, the edge weight between the i- th node and the j - th node is EW_{ij}, and the longitude of the d- th node is d_i (only considering routes across urban areas).

$$\max \sum_{i=1}^{n} NW_i \quad \text{st.} \begin{cases} \sum\limits_{i=1, j=i+1}^{n-1} EW_{ij} \le \alpha, \quad \alpha = 12km \\ d_1 < d_2 < \ldots < d_n \end{cases}$$

While ensuring the above - mentioned constraints, we adopt a backtracking algorithm based on depth - first search to solve this optimization problem. The following is the pseudocode of our algorithm[7]:

Algorithm 2: Max nodes weights with Distance Constraint

Input : Graph G = (V, E) where V is nodes with weights and E is edges with distances
Distance limit D = 12km

Output : Maximum sum of weights for a valid path

```
maxWeightSum ← 0
path ← empty list                          // To store current path of nodes

function DFS(currentNode, currentWeight, currentDistance):
  if currentDistance >= D then
     return                                // Backtrack
  if currentWeight > maxWeightSum then
     maxWeightSum ← currentWeight
  path.append(currentNode)

  for each neighbor ∈ G.neighbors(currentNode) do
     nextWeight ← currentWeight + neighbor.weight
     nextDistance ← currentDistance + G.distance(currentNode, neighbor)
     DFS(neighbor, nextWeight, nextDistance)          // Recursive DFS call
  path.remove(currentNode)        // Remove current node from path after exploring

start the algorithm by selecting each node in the graph as the starting point:
```

```
for each node ∈ V do
    DFS(node, node.weight, 0)
return maxWeightSum
```

After implementing the algorithm, the route for constructing the viaduct is shown in the figure below.

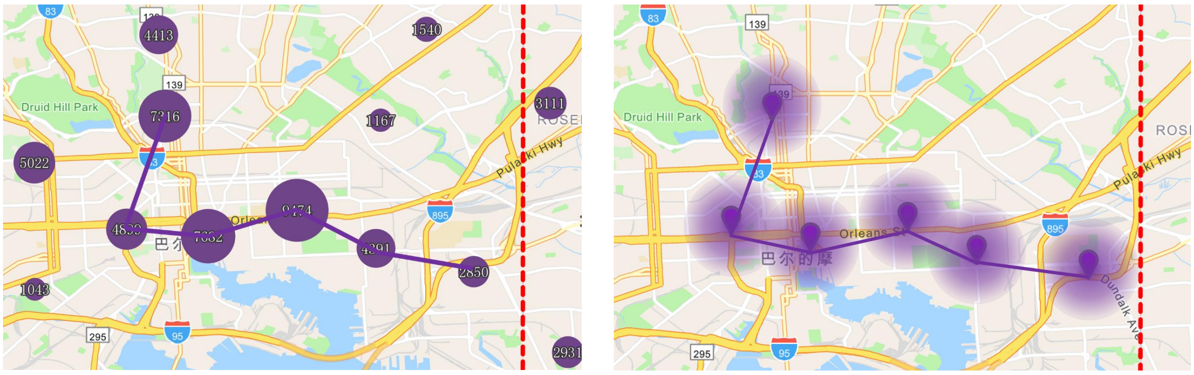


Figure 16: Viaduct path (left) and radiation of the impact of traffic flow distribution (right)

Subsequently, the traffic flow re - allocation is carried out. Due to the large - scale nature of the transportation network we have constructed, when re - allocating the traffic flow, we utilize the greedy principle for calculation. A step - like distribution approach is adopted to re - allocate the traffic flow at the entrances of the viaduct. Specifically, for the roads in the radiation area around the viaduct entrances, the traffic flow is assumed in a step - by - step manner. By leveraging the Euclidean distance from the entrance point, the farther a road is from the entrance point, the smaller the percentage of its traffic flow diverted to the viaduct.

$$v_i = \begin{cases} \theta(1-\beta a_{ij})v_{ij}, & 1-\beta a_{ij} \geq 0 \\ 0, & 1-\beta a_{ij} < 0 \end{cases}$$

Where v_i represents the traffic flow allocated to the entrance of the viaduct, v_{ij} represents the traffic flow at a point on the road near the viaduct, θ represents the percentage of traffic flow expected to be distributed in a step -wise manner at the entrance of the viaduct, a_{ij} represents the Euclidean distance between the traffic - flow point on the road near the viaduct and the entrance of the viaduct, and β represents the mapping to a percentage.

After the traffic - flow allocation is completed, the indicators of Model in problem 1 can be used to evaluate relevant requirements.

7.3 Observing the impact of the viaduct

After determining the construction route of the viaduct, the subsequent problem is solved using the approach from Problem 1. Conduct 100 experiments for both the pre - viaduct - construction and post - viaduct - construction scenarios. In each experiment, randomly select two nodes from all nodes, calculate the shortest path, and re - allocate the road traffic flow based on the shortest path. Thereby, determine the traffic volume, congestion level, and various impact indicators. The specific data is as follows:

Comparison of results after 100 experiments		
	before	after
Flow	4336.67	4657.63
Jam	5453.36	4756.73
Pedestrians	1343.68	853.68
Drivers	2128.71	1893.23
Office workers	2856.35	2493.46
Tourists	560.35	534.46
Government	545336.32	475673.41

As can be seen from the table, after the construction of the viaduct:

- The construction of the viaduct can generally divert a portion of the traffic from ground - level roads to the dedicated upper - level passages. Especially for some traffic - congested intersections, it can effectively reduce road congestion and improve traffic flow.
- It reduces road congestion. Particularly during peak hours, the traffic pressure on ground - level roads is alleviated, thereby improving traffic efficiency. This is of great help in relieving the traffic pressure in the city center area.
- For pedestrians: Convenient pedestrian passages are set up, enabling pedestrians to cross the street more safely and conveniently and avoiding congestion at intersections. The travel indicators for pedestrians are improved to a certain extent.
- For public transportation[8]: The efficiency of bus services is further enhanced. Roads with priority for buses are provided. Especially during peak hours, buses can avoid congestion on ground - level roads, shorten travel times, and improve punctuality.

8 Problem 3

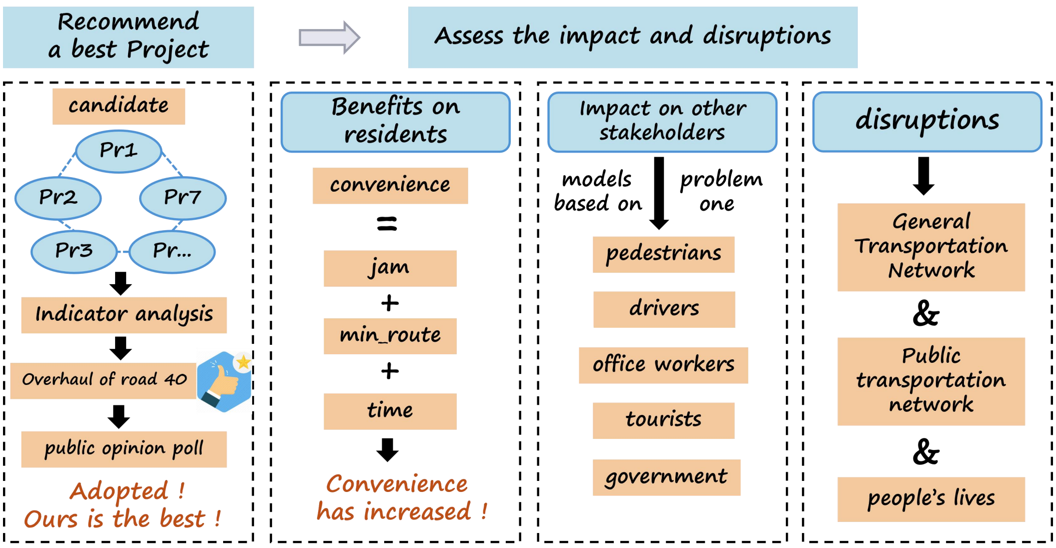


Figure 17: Mindmap of problem 3

8.1 Choose the best project

Figure 18: Seven candidate project

By referring to information related to traffic problems in Baltimore and some future projects on the official website of Baltimore, we have extracted the above six projects in total. In response to the issues raised by residents and based on public opinion surveys, we have recommended one additional project ourselves. The project involves completely filling the ditch on Route 40, widening the north - south roads, and adding bus routes, bus stops, and sidewalks. In total, there are seven projects, namely:

- Baltimore City added two projects aimed at improving pedestrian and bicycle safety. The 2022 and 2023 Pedestrian & Roadway Safety Improvements will develop a variety of safety improvements at over 50 intersections citywide.
- Project in Harford County has received multiple federal grants to provide ADA improvements and better access to the Harford Transit LINK and MTA bus services.
- Project on Snowden River Parkway from Broken Land Parkway to Oakland Mills Road which adds a third lane along Snowden River Parkway in each direction and will provide new 10-foot shared-use paths in an effort to increase transportation alternatives to public transit.
- The Maryland Port Administration continues construction on the Howard Street Tunnel.
- MDOT MTA proposes funding for the Red Line project which will provide an east-west, high frequency, high-capacity transit line for the Baltimore region.
- The Maryland Transportation Authority continues construction on the I-95 Express Toll Lanes between White Marsh and Bel Air.
- Completely fill the ditch on Route 40, widen the north - south roads, add bus routes, bus stops, and sidewalks. Additionally, allocate some area in the filled - up region to serve as a construction area (such as residential buildings, supermarkets, etc.) to facilitate the daily lives of residents.

Establish relevant indicators:

Travel convenience (convenience): Considering that residents' travel convenience is related to road congestion levels, the shortest path to the destination, and travel time, it is defined as follows.

$$convenience = a * jam + b * \min\_route + c * t$$

Then, similar to the determination of the weight coefficients of each indicator in Problem 1, the Analytic Hierarchy Process (AHP) is used. After passing the consistency check, we obtain a = 0.11, b = 0.31, and c = 0.58. The formula is rewritten as follows.

$$convenience = 0.11 * jam + 0.31 * \min\_route + 0.58 * t$$

Then, select multiple points around the project and conduct 100 random simulation experiments. Take the average value to obtain the final result.

	P1	P2	P3	P4	P5	P6	**Ours**
convenience	1214.35	3758.64	134.45	5246.98	8645.46	7648.54	**9754.32**

Thus, it can be determined that our project - the filling and construction of Route 40 can best improve the lives of residents. The following specifically presents the details of our project：

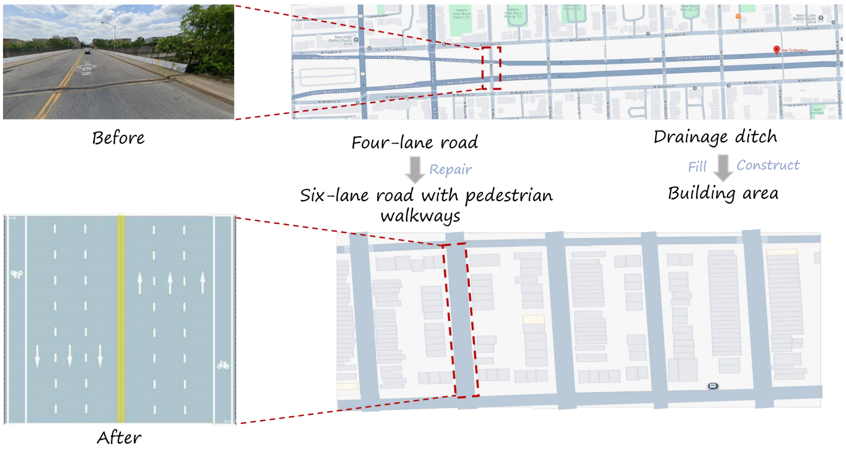


Figure 19: Explanation of our project

Moreover, through our investigation, we found that the renovation work of Route 40 has been ongoing. Currently, the most prominent issue is the problem of the ditch on the selected section dividing the community. Therefore, based on relevant literature[9][10], it can be confirmed that this project indeed improves the living conditions of residents.

8.2 Observing the impact of the viaduct

We use the model from Problem 1 to address the impact on various stakeholders. We still select starting and ending points above and below the renovated road, and conduct 100 experiments using both the overall road network and the traffic network. The average values are taken for measurement, resulting in the following table.

Comparison of results after 100 experiments					
	before	after		before	after
Flow	3425.42	3326.35	Drivers	465.35	435.32
Jam	546.24	335.68	Office workers	567.54	456.98
Pedestrians	368.56	243.32	Tourists	124.24	123.43
Government	1745.56	1342.34			

Through the results, it is found that after the refilling and construction of Route 40, the traffic flow on the selected section of Route 40 does not change significantly, while the congestion level is greatly reduced. Due to the addition of pedestrian lanes, it also becomes more convenient for pedestrians to travel. With the widening of lanes and the addition of bus lanes and relevant bus stops, the convenience for drivers, tourists, and commuters has increased. The overall traffic situation has been significantly improved, thus enhancing the government - related indicators. However, after referring to relevant literature, it is known that this work requires a high budget, so comprehensive consideration is still needed.

8.3 Disruption of transportation needs and people's lives

We simulated the traffic flow after widening the road width of this section in the overall road network and found that the traffic flow on the renovated section did not increase. This is because the vehicle road capacity was not fully utilized either before or after the renovation. Therefore, for vehicles, the improvement of this project is relatively redundant. This can be illustrated as shown in the following figure (i.e., the minimum cut in the figure remains the same).

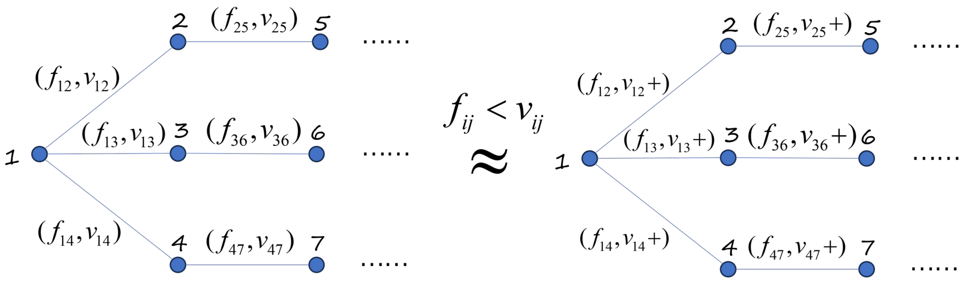
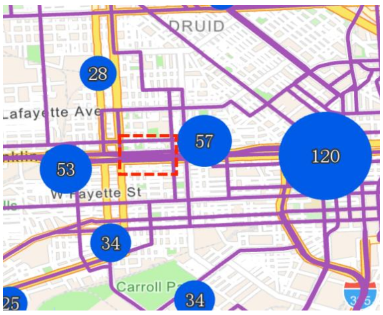


Figure 20: Equivalent Graph of Network Minimum Cut



By adding public transportation routes, the coverage rate of public transportation has increased. As can be seen from Figure after clustering the bus stops, it was found that the small number of public transportation services between the two communities, north and south of the project area, has led to difficulties in commu - nication between the two communities. Therefore, our project aims to increase the demand for public transportation by adding bus routes and stops. At the same time, this also reduces the use of private vehicles.

Impact on People's Lives

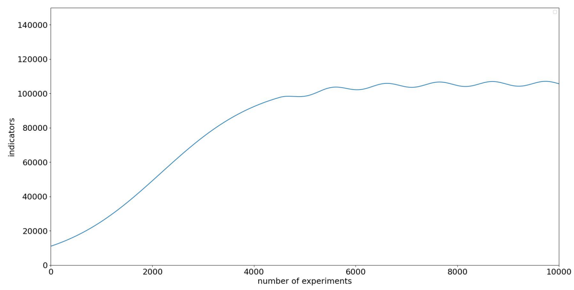
- After filling the ditches and widening the roads, we build residential and commercial areas in the remaining space. The newly - built commercial supermarkets and facilities can bring more convenience to residents, reducing their travel needs. Daily shopping, entertainment, etc. become more accessible, thus improving the quality of life of residents.
- Increasing lanes can ease traffic congestion, especially during peak hours, reducing people's commuting time and improving efficiency. However, without effective traffic management measures, more lanes and higher traffic volume may pose potential safety hazards, especially in areas close to residential communities.
- The addition of public transportation stops and routes means that more people can choose public transportation for travel. This has a positive impact on reducing dependence on private vehicles, alleviating traffic pressure, and reducing pollution.
- The increase in commercial and residential areas may bring about public security issues. In areas with high - volume human traffic and frequent commercial activities, more

safety measures may be required to ensure the safety of residents.

9 Sensitivity Analysis

The main components for conducting sensitivity analysis on our model are as follows:

- ➢ Sensitivity analysis of experimental randomness. This part mainly demonstrates the generalization ability of the project's advantages. We use Monte Carlo random simulations to conduct this sensitivity analysis. The Monte Carlo simulations were performed for 10 to 10,000 experiments, and the results are as follows:



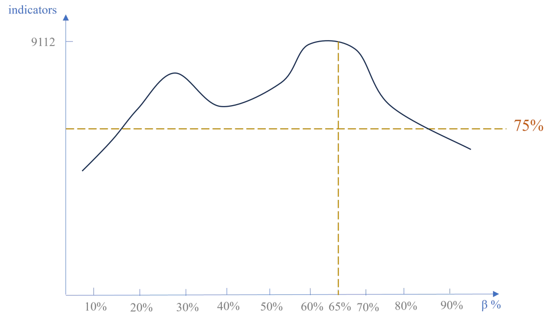


Figure 21: Monte Carlo simulation chart(left) and Flow distribution sensitivity analysis result chart(right)

- ➢ Sensitivity analysis of the traffic flow distribution model. This part is mainly used to test whether the threshold values we selected are reasonable and whether the conclusions derived from the greedy algorithm are stable.

As shown on the right side of Figure 21, we test the β parameter of our traffic flow distribution model. It is specified that if the value is greater than 75% of the maximum value, the parameter is considered representative. Therefore, the value of 65% that we selected meets the sensitivity requirements.

10 Model Evaluation

10.1 Strengths

1. Our model incorporates a multitude of evaluation indicators, which to a large extent represent the merits and demerits of the model.
2. Our model features rich visualizations and graphical representations. These not only facilitate understanding and implementation but also enable analysis in combination with graph theory and relevant topological information.
3. The algorithms in our model are innovative. By modifying common algorithms, we have made them more suitable for this particular model, endowing it with strong generalization capabilities.
4. We have integrated the Python interpreter with relevant software, significantly enhancing the code inference speed.

10.2 Weaknesses

1. Our model is highly data - dependent. In the event of data being unauthentic or having a large number of missing values, it becomes challenging to conduct in - depth analysis effectively.
2. Although we have constructed a relatively comprehensive set of indicators, numerous factors can still influence the actual results. Therefore, a comprehensive consideration is necessary.

Memo

To： Mayor of Baltimore
From: Team # 2517674
Date: January 28, 2025
Subject: Transportation Projects to Improve Safety in Baltimore

Safety is of utmost importance for the well - being of our city and its residents.How to optimize the transportation network of Baltimore City and improve the sense of security of residents is an urgent problem to be solved.

Based on this, we designed the following two projects.

Project 1: Strategic High - Viaduct Construction for Traffic Optimization

We found that the flow in the urban center was significantly large, so we used KNN clustering to get the subgraph of the main nodes, and planned to build viaducts between these nodes to achieve the purpose of diversion. After pre-processing the data and constructing the road network and bus network, using the above subgraph, the backtracking algorithm based on depth-first search is used to determine the specific construction route of the viaduct.

Benefits

Ensure the safety of residents' travel: the viaduct reduces the interference of ground traffic, sets up bus lanes, attracts more residents to choose public transport, reduces the number of private cars on the road, reduces the risk of collision between vehicles.

Improved Emergency Response Times: With a more efficient traffic flow, emergency vehicles such as ambulances and fire trucks will be able to reach the destinations more quickly.

Project 2: Fill and Reconstruct Road 40

After analyzing six potential projects and combining the results of public opinion surveys in the literature, we proposed a new project: fill in the ditch on Highway 40, widen the north-south road, and add bus routes, bus stops, and sidewalks. Then we use the model from the first project and the concept of maximum flow-minimum cut to get the result and its impact.

Benefits

Enhanced Pedestrian Safety: By adding sidewalks along Route 40, pedestrians are provided with a dedicated and safer space to walk.

Improved Traffic Safety for Motorists: The widened roads and addition of bus lanes enable vehicles to move more freely, reducing the likelihood of traffic jams and the associated risks of rear - end collisions, sudden lane - changes, and road rage incidents.

Drawbacks

High Costs and Vehicle Traffic Redundancy: Simulation shows road capacity wasn't fully used pre - renovation, so vehicle traffic flow improvement may be redundant.

In conclusion, these two projects provide transportation level solutions to solve the safety problems of Baltimore city from different angles. We hope our projects will contribute to a safer and more efficient transportation environment in Baltimore.

Yours sincerely
Team # 2517674

References

[1] Rodrigue J P. The geography of transport systems[M]. Routledge, 2020.

[2] Burns L D. A vision of our transport future[J]. Nature, 2013, 497(7448): 181-182.

[3] Markovich J, Lucas K. The social and distributional impacts of transport: a literature review[J]. Transport Studies Unit, School of Geography and the Environment Working Paper, 2011, 1055.

[4] Al-Harbi K M A S. Application of the AHP in project management[J]. International journal of project management, 2001, 19(1): 19-27.

[5] Hagerstrand T. Impact of Transport on the Quality of Life[C]//This paper was presented at the 5th International Symposium on Theory and Practice in Transport Economics, Transport in the 1980-1990 Decade, Athens, October, 1973. 1974, 1(Topic 5).

[6] Nijkamp P, Blaas E W. Impact assessment and evaluation in transportation planning[M]. Springer Science & Business Media, 2012.

[7] Tarjan R. Depth-first search and linear graph algorithms[J]. SIAM journal on computing, 1972, 1(2): 146-160.

[8] Reynolds C C O, Harris M A, Teschke K, et al. The impact of transportation infrastructure on bicycling injuries and crashes: a review of the literature[J]. Environmental health, 2009, 8: 1-19.

[9] Manville M, Cummins B. Why do voters support public transportation? Public choices and private behavior[J]. Transportation, 2015, 42: 303-332.

[10] Van Leuven N, Smith N. Collaboratively Connecting: Public Polling as a Foundation for Integrated Transportation Decision-Making Networks[J]. 2019.