Urban Simulation Coursework

Part 1: London's underground resilience

I. Topological network

I.1 Centrality measures

We chose degree centrality, betweenness centrality, and closeness centrality as measures to characterize nodes in the London Underground network due to their unique properties and relevance to the context of transportation networks:

1. **Degree Centrality:**

- **Definition**: Degree centrality is defined as the number of edges (connections) that a node has with other nodes in the network. In simpler terms, it represents the count of how many neighbours a node has.
- **Equation**:

$$C_d = \frac{d}{N-1} \tag{1}$$

where:

- C_d represents the degree centrality of a node
- d is the degree of the node (number of connections)
- N is the total number of nodes in the network

In the context of the London Underground, the degree centrality of a station represents the number of direct connections it has to other stations. A station with higher degree centrality may be more crucial for the network's functioning since it directly connects to more stations, potentially serving as a significant transfer point for passengers.

2. **Betweenness Centrality:**

- **Definition**: Betweenness centrality is a measure of the importance or influence of a node within a network based on the number of shortest paths that pass through it. It captures the extent to which a node serves as a bridge or intermediary between other nodes in the network.
- **Equation**:

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{2}$$

where:

- $C_B(v)$ represents the betweenness centrality of a node
- v represents the node for which betweenness centrality is being calculated
- σ_{st} is the total number of shortest paths between nodes s and t
- $\sigma_{st}(v)$ is the number of shortest paths between nodes s and t that pass through node v

In the context of the London Underground, a station with high betweenness centrality serves as an important intermediary connecting different parts of the network. Removing a station with high betweenness centrality may significantly increase travel times and congestion, impacting the overall network efficiency.

3. **Closeness Centrality:**

- **Definition**: Closeness centrality is a measure of the importance or influence of a node within a network based on the inverse of the average distance (shortest path length) between the node and all other nodes in the network. It captures the extent to which a node is close to, or easily reachable from, other nodes in the network.
- **Equation**:

$$C_C(v) = \frac{N-1}{\sum_{t \neq v} d(v,t)} \tag{3}$$

where:

- $C_C(v)$ represents the closeness centrality of a node
- v represents the node for which closeness centrality is being calculated
- N is the total number of nodes in the network
- d(v, t) is the shortest path distance between node v and node t

In the context of the London Underground, a station with high closeness centrality is, on average, closer to all other stations in the network, making it an efficient transfer point. A station with high closeness centrality can facilitate shorter travel times and better overall connectivity.

Table 1: The First 10 Ranked Nodes for Each of The Three Measures

	Degree Centrality Station	Degree Centrality Value	Betweenness Centrality Station	Betweenness Centrality Value	Closeness Centrality Station	Closeness Centrality Value
1	Stratford	0.922111	Stratford	0.0985528	Stratford	0.927739
2	Highbury & Islington	0.806533	Liverpool Street	0.0343073	Highbury & Islington	0.836134
3	Whitechapel	0.781407	Bank and Monument	0.0279563	Whitechapel	0.820619
4	West Brompton	0.776382	Canary Wharf	0.0279563	West Brompton	0.817248
5	Canary Wharf	0.771357	Canning Town	0.0277565	Canada Water	0.813906
6	Canada Water	0.771357	West Ham	0.0245515	Bank and Monument	0.810591

	Degree Centrality Station	Degree Centrality Value	Betweenness Centrality Station	Betweenness Centrality Value	Closeness Centrality Station	Closeness Centrality Value
7	Liverpool Street	0.768844	Highbury & Islington	0.0230232	Canary Wharf	0.810591
8	Bank and Monument	0.766332	Whitechapel	0.0196824	Richmond	0.810591
9	Richmond	0.766332	Canada Water	0.0178978	Canning Town	0.808943
10	Canning Town	0.763819	Shadwell	0.0170701	Liverpool Street	0.808943

I.2 Impact measures

In this part, we chose the average shortest path length and network diameter as measures to evaluate the impact of node removal on the network due to their ability to provide insights into the overall connectivity and efficiency of the network.

1. **Average shortest path length:**

- **Definition**: This metric represents the average of the shortest path lengths between all pairs of nodes in a network. It indicates how efficiently information or resources can be transmitted across the network. A lower average shortest path length suggests that nodes are more closely connected and can interact more quickly with each other.
- **Equation**:

$$L = \frac{1}{N(N-1)} \sum_{s \neq t} d(s,t) \tag{4}$$

  where:

- N is the total number of nodes in the network
- d(s, t) is the shortest path distance between nodes s and t

The average shortest path length is not specific to the London Underground; it can be used to evaluate the resilience of any network. In the context of the London Underground, an increase in the average shortest path length after node removal indicates decreased efficiency and accessibility, making it a useful measure for assessing the impact of station closures.

2. **Network diameter:**

- **Definition**: This metric represents the longest shortest path length between any pair of nodes in a network. It indicates the maximum distance between any two nodes in the network. A smaller network diameter implies that the network is more tightly connected, while a larger diameter suggests that the network is more dispersed.
- **Equation**:

$$D = \max_{s \neq t} d(s,t) \tag{5}$$

where:

- d(s, t) is the shortest path distance between nodes s and t
- The maximum (max) is taken over all pairs of nodes (s, t) in the network

The network diameter is not specific to the London Underground and can be applied to any network. In the context of the London Underground, an increase in the network diameter after node removal implies that the maximum distance between any two stations has increased, signalling a decrease in network resilience and efficiency.

I.3 Node removal

In this part, we removed ten nodes from the network using two different methods:

1. Non-sequential removal: From the most significant node to the tenth most essential node, we eliminate 1 node at a time in accordance with the order in Table 1. We proceed until at least 10 nodes have been eliminated.
2. Sequential removal: We eliminate the node with the highest ranking. Recalculate the centrality measure after removal. In the new network, remove the node with the highest ranking.

According to Figure 1, the betweenness centrality removal resulted in a greater change in the average shortest path than the degree and closeness centrality. This suggests that it is more effective in identifying critical nodes. In addition to this, betweenness centrality measures the extent to which a node acts as a bridge for the shortest paths between other nodes in the network, illustrating the role of nodes in connecting different parts of the network. When stations are removed or added, the paths between different stations may change, thus affecting the overall connectivity of the network. In the context of the London Underground, this means that stations with high betweenness centrality are important hubs through which many passenger lines pass.

The two removal strategies differ only in the betweenness centrality, with Sequential Removal leading to stronger fluctuations in the mean shortest path and diameter, suggesting that it is more appropriate for exploring the resilience of the network. In addition, the Sequential Removal strategy takes into account changes in network structure and connectivity, whereas Non-Sequential Removal relies only on the initial ranking, in contrast to the Sequential Removal strategy, which provides a more accurate representation of the importance of nodes in the modified network. Cascading effects, also known as cascading failures, are a chain of events or failures that occur in an interconnected system or network. Understanding and mitigating cascading effects is essential to designing and maintaining resilient and robust networks (Motter and Lai, 2002), and a sequential removal approach makes it easier to help identify the cascading effects that can occur when critical nodes are removed from a network.

Both the average shortest path length and the network diameter provide valuable insights in assessing damage after node removal. However, according to Figure 1, the average shortest path length is more sensitive to changes in the network and provides a better representation of the overall impact on the network following node removal. Therefore, relying on the average shortest path metric to assess the impact of node removal may provide additional information.

II. Flows: weighted network

II.1 Adjusted Centrality Measure for Weighted Networks

In this part, we adjusted the centrality measures for a weighted network, taking into account the passenger flows assigned to the links between stations. For each centrality measure, we modified the calculations to account for the weights associated with the edges.

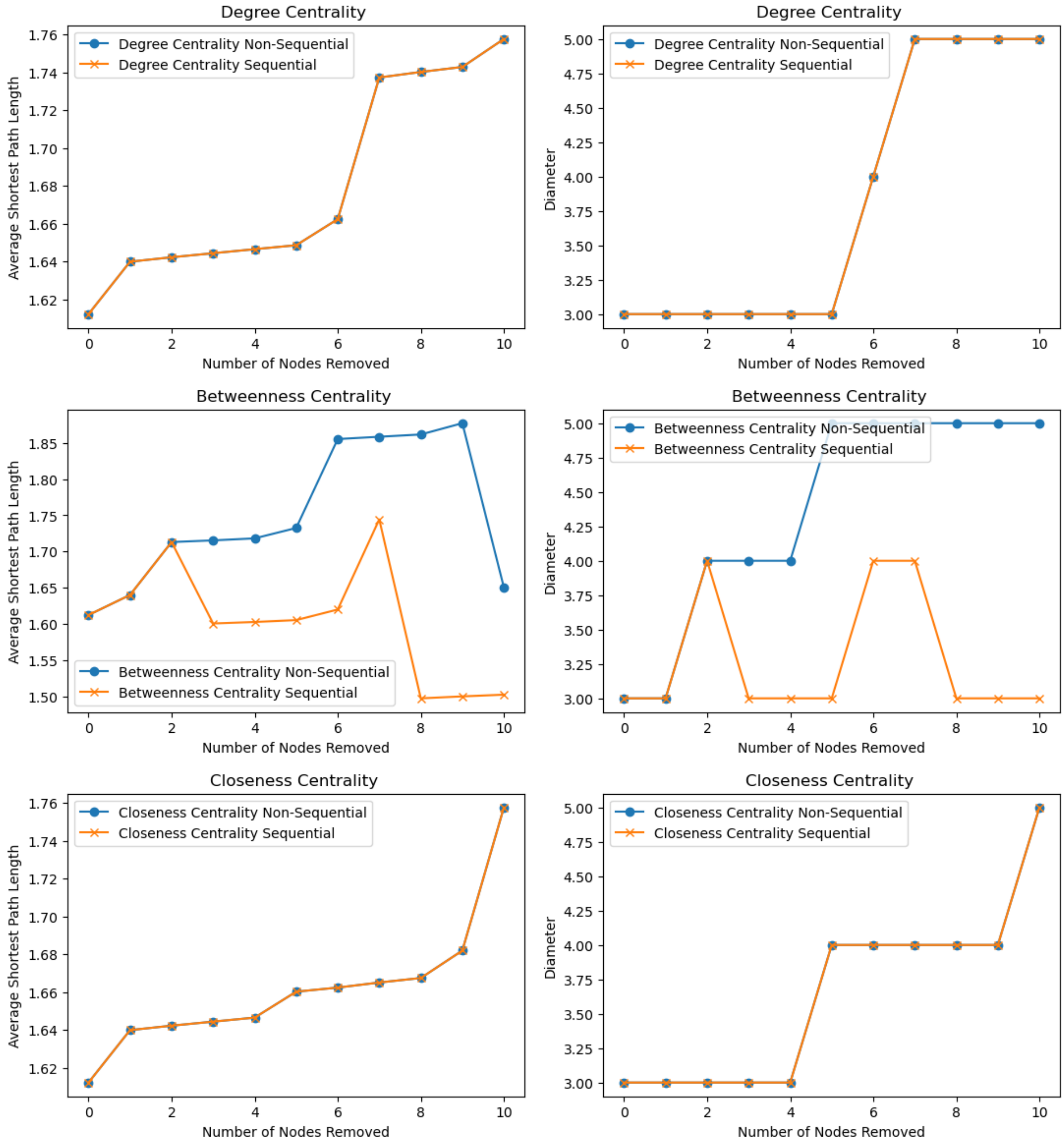

Figure 1: Combined Plots of Average Shortest Path Length and Network Diameter for Different Centralities.

1. **Degree Centrality**: In a weighted network, we can calculate the strength of a node instead of its degree. The strength of a node is the sum of the weights of its edges, representing the total number of passengers passing through a station. In this case, nodes with higher strength values are more important in terms of the total passenger volume they handle.
2. **Betweenness Centrality**: When calculating betweenness centrality for a weighted network, we need to consider the shortest paths based on the weights instead of just counting the number of edges. This means that the shortest paths will be determined based on the minimum total weight, considering the passenger flows, not the minimum number of edges.
3. **Closeness Centrality**: When calculating closeness centrality for a weighted network, we should use the total weight of the shortest paths between the node and all other nodes in the network.

Table 2: The First 10 Ranked Nodes for Each of The Three Adjusted Measures

	Degree Centrality Station	Degree Centrality Value	Betweenness Centrality Station	Betweenness Centrality Value	Closeness Centrality Station	Closeness Centrality Value
1	Stratford	115265	West Ham	2.10088e+62	Stratford	0.927739
2	Bank and Monument	108043	Bank and Monument	1.69137e+61	Highbury & Islington	0.820619
3	Liverpool Street	92595	Stratford	6.48695e+60	Whitechapel	0.812245
4	Waterloo	90838	North Wembley	5.3126e+58	Canada Water	0.810591
5	Canary Wharf	73404	Kensington	1.43776e+58	Canary Wharf	0.808943
6	Victoria	70768	Harlesden	1.35598e+58	Bank and Monument	0.807302
7	London Bridge	62527	Custom House	3.3925e+57	Canning Town	0.807302
8	King's Cross St. Pancras	61637	Prince Regent	1.62296e+57	Liverpool Street	0.805668
9	Highbury & Islington	49099	Gallions Reach	1.62296e+57	West Brompton	0.799197
10	Canada Water	47469	Kensal Green	3.72659e+56	Richmond	0.796000

According to the Table 1 and 2, we can observe that for Degree Centrality, some stations appear in both lists, such as Stratford, Bank and Monument, Liverpool Street, Canary Wharf, and Highbury & Islington.

For Betweenness Centrality, a few stations appear in both lists, including Stratford, Bank and Monument, Canary Wharf, and West Ham.

For Closeness Centrality, the results are largely the same for the unweighted and weighted networks. The sites appearing in both lists are the same, with only slight differences in ranking.

II.2 Impact Measures for Weighted Networks

In this part, we should adjust the measure for a weighted network. In an unweighted network, the shortest path is calculated based on the number of edges, while in a weighted network, the shortest path should be calculated based on the total weight (passenger flow) of the path. The adjusted measure is the weighted average shortest path length:

$$WASPL = \frac{1}{N(N-1)} \sum d_w(i,j) \tag{7}$$

where:

- N is the number of nodes in the network
- $d_w(i,j)$ is the weighted shortest path length between node i and node j.

A measure that could better assess the impact of closing a station, considering passenger flows, is Weighted Network Efficiency (WNE).

- WNE's Equation:

$$WNE = \frac{1}{N(N-1)} \sum (1/d_w(i,j)) \tag{8}$$

where:

- N is the number of nodes in the network
- $d_w(i,j)$ is the weighted shortest path length between node i and node j.

Rubinov and Sporns (2010) applied network efficiency to weighted brain networks to investigate the connectivity and efficiency of information transfer in the human brain, which indicate that the application of network efficiency to weighted networks demonstrates its versatility and adaptability to a variety of research areas, including transport networks. In this part, the weighted network efficiency measures the overall efficiency of passenger flows in the weighted network and is well suited to assessing the impact of station closures.

II.3 Evaluating the Impact of Node Removal in Weighted Networks

In this part, we removed the three stations with the highest centrality values by using the sequential removal strategy based on the betweenness centrality measure, and we evaluate the impact of eliminating a station on the network by the WASPL and WNE.

Table 3: Impact of Sequential Station Closures on Weighted Average Shortest Path Length (WASPL) and Weighted Network Efficiency (WNE)

Removed Node	WASPL After Removal	WNE after removal
-	0.289366	0.00808017
West Ham	0.307148	0.0171796
Abbey Road	0.308196	0.0179712
Stratford	0.205751	0.0150606

According to Table 3, the removal of West Ham station resulted in a longer average path length for passengers but, surprisingly, a significant increase in network efficiency. One possible reason for the increase in WNE could be that the removal of West Ham station resulted in a redistribution of passenger flows, resulting in more efficient use of the remaining network connections. The closure of Abbey Road will have less impact on passengers than the closure of West Ham. The smaller changes in WASPL and WNE suggest that Abbey Road may be less important to overall network efficiency and connectivity. In contrast, the removal of Stratford Station would result in a shorter average path length but would reduce network efficiency. This may be because the station is a major hub connecting multiple lines and its removal forces passengers to use alternative routes which may be more direct but less efficient in distributing passenger flows through the network.

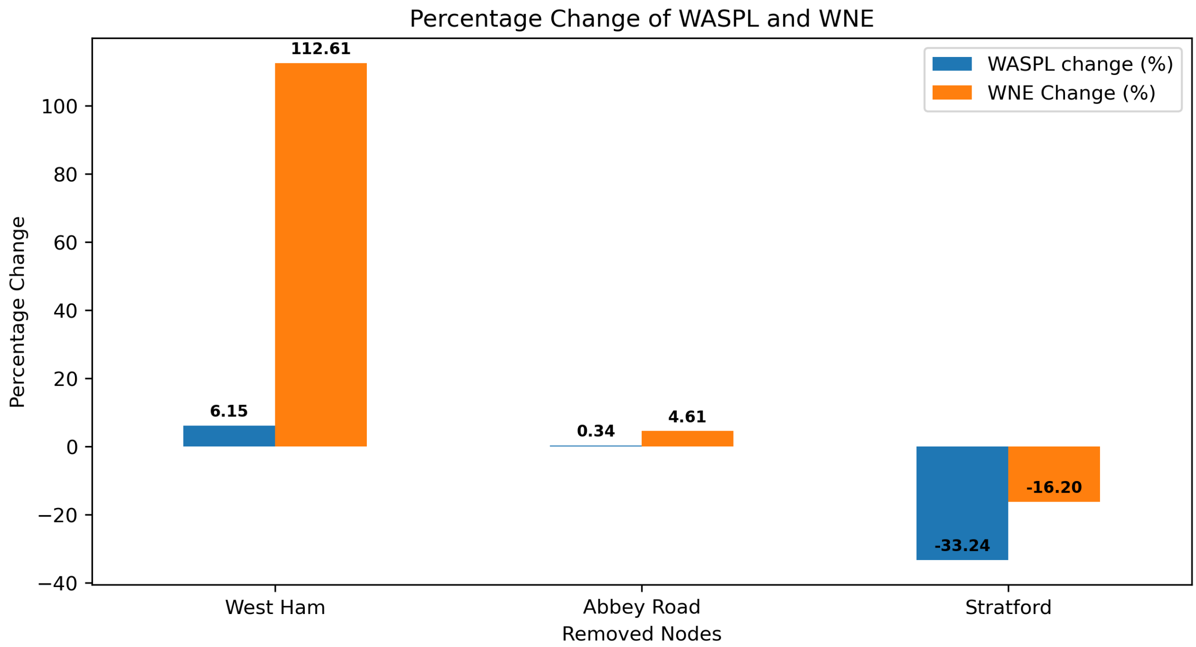

Figure 2: Combined Plots of Average Shortest Path Length and Network Diameter for Different Centralities.

From Figure 2, we can observed more visually that the closure of West Ham has a much greater impact on passengers than Stratford and Abbey Road.

In summary, the analysis of the sequential node removal scenario reveals that West Ham station's closure has the most significant impact on passengers, followed by Stratford and Abbey Road.

Part 2: Spatial Interaction models

III: Models and calibration

III.1 Introduction to Spatial Interaction Models and Their Parameters

In this part, we introduced four types of Spatial Interaction Models: Unconstrained, Origin-Constrained, Destination-Constrained, and Doubly Constrained models, and discuss their parameters and equations.

1. **The Unconstrained Model:**

The Unconstrained Model predicts the flow (T_{ij}) between origin i and destination j using the following equation:

$$T_{ij} = k \cdot O_i \cdot D_j \cdot C_{ij}^{-\beta} \tag{9}$$

$$\sum_{i,j} T_{ij} = T \tag{10}$$

where:

- O_i represents the total number of outflows (e.g., commuters) from origin i.
- D_j represents the total number of inflows (e.g., job opportunities) to destination j.
- C_{ij} represents the cost (e.g., travel time or distance) between origin i and destination j.
- β is a parameter that needs to be estimated, which represents the friction of distance.
- k is a constant that ensures the sum of the total flow to a specific value.
- T is the sum of all flows.

2. **The Origin-Constrained Model:**

The Origin-Constrained Model introduces constraints on the total outflows from each origin. The equation for this model is:

$$T_{ij} = A_i \cdot O_i \cdot D_j \cdot C_{ij}^{-\beta} \tag{11}$$

$$\sum_{j} T_{ij} = O_i \tag{12}$$

where:

- A_i is origin-specific balancing factors that ensure origin constraint is satisfied.

3. **The Destination-Constrained Model:**

The Destination-Constrained Model introduces constraints on the total inflows to each destination. The equation for this model is:

$$T_{ij} = O_i \cdot B_j \cdot D_j \cdot C_{ij}^{-\beta} \tag{13}$$

$$\sum_{i} T_{ij} = D_j \tag{14}$$

where:

- B_j is estination-specific balancing factors that ensure destination constraint is satisfied.

4. **The Doubly Constrained Model:**

The Doubly Constrained Model introduces constraints on both the total outflows from each origin and the total inflows to each destination. The equation for this model is:

$$T_{ij} = A_i \cdot O_i \cdot B_j \cdot D_j \cdot C_{ij}^{-\beta} \tag{15}$$

$$\sum_j T_{ij} = O_i \tag{16}$$

$$\sum_i T_{ij} = D_j \tag{17}$$

where:

- A_i and B_j are origin and destination-specific balancing factors that ensure both origin and destination constraints are satisfied.

III.2 Model Selection and Parameter Calibration for the London Underground Scenario

Fotheringham and O'Kelly (1989) provide a comprehensive overview of spatial interaction models and their applications and discuss various formulations of spatial interaction models, including those that start with population and end with work. so we choose the origin-constrained spatial interaction model with a negative exponential cost function here to analyze the relationship between population (as the origin) and jobs (as the destination) in the context of the London Underground network.

The origin-constrained model is appropriate in this context because it assumes that the total flow originating from each zone (i.e., population) is fixed and known, which is a reasonable assumption when examining commuting patterns. People residing in a particular area are constrained by their residential location when commuting to work, and the total number of commuters from that area is determined by the population.

The negative exponential function implies that the interaction between two zones decreases exponentially with the increasing cost (e.g., distance) between them. This is a reasonable assumption in the context of commuting, as people generally prefer to work closer to their homes, and the propensity to commute declines as the distance or travel time between the origin and destination increases.

Here we used a Poisson regression model by the equation (18) to calibrate the parameters, which has as its central hypothesis that counts of non-negative integers are related to the various flows that the spatial interaction model deals with (commuting flows).

$$\lambda_{ij} = \exp(\alpha_i + \gamma \ln D_j - \beta \ d_{ij}) \tag{18}$$

where:

- λ_{ij} is the estimate of T_{ij}
- α_i is the vector of balancing factors
- D_j represents the total number of inflows to destination j
- d_{ij} represents the cost between origin i and destination j.
- γ and β are paramters.

Table 4: Comparison with Different Spatial Interaction Models with Negative Exponential Function

Model	R2	RMSE	Alpha	Gamma	Beta
Unconstrained	0.29648	124.895	0.698578	0.733996	8.9e-05
population_ constrained	0.448275	97.879	nan	0.750908	0.000150817
jobs_ constrained	0.384959	103.316	0.709926	nan	9.7909e-05
doubly_ constrained	0.476617	95.23	nan	nan	0.000151847

According to Table 4, we found that the origin-constrained spatial interaction model with a negative exponential cost function performs better than all other models except the double-constrained negative exponential model. Although it slightly underperforms the double-constrained negative exponential model, we chose the origin-constrained model for its simpler operationalization and better interpretability. Furthermore, the differences in performance metrics (R2 and RMSE) between the two models are relatively small, indicating that the origin-constrained model provides a sufficiently accurate representation of the commuting patterns in the London Underground network. The calibrated Beta parameter for the cost function is 0.000150817.

IV.Scenarios

IV.1 Scenario A: Assessing the Impact of a 50% Decrease in Jobs at Canary Wharf Post-Brexit

In our origin-constrained spatial interaction model, we conserved the number of commuters by constraining the origin side and using origin-specific factors (A_i). These factors are calculated by considering destination attractiveness (D_j^γ) and the cost function ($d_{ij}^{-\beta}$). The A_i values ensure that predicted commuter flows between origin-destination pairs match observed data. By adjusting predicted flows with A_i factors, we maintain the total number of commuters, aligning the model with real-world observations.

Figure 3 shows the as expected, there is a significant decrease in the number of commuters (-18,439) in Canary Wharf due to the reduction in jobs. This indicates that many commuters previously travelling to Canary Wharf have now been redistributed to other stations, as fewer job opportunities are available in this area

IV.2 Scenario B: Assessing the Impact of Increased Transportation Costs on Commuter Flows

In this scenario, we explored the potential impact of increased transportation costs on the London Underground commuting patterns. Increased costs could result from factors such as inflation, policy changes, or fuel price increases. To reflect the increased cost of transport in the model, we select two different beta values that are larger than the original calibrated beta (0.000150817). A larger beta value implies that distance has a more significant impact on the decision to commute between two stations due to the increased cost of transport.

- Beta value 1: beta(0.000150817) * 2

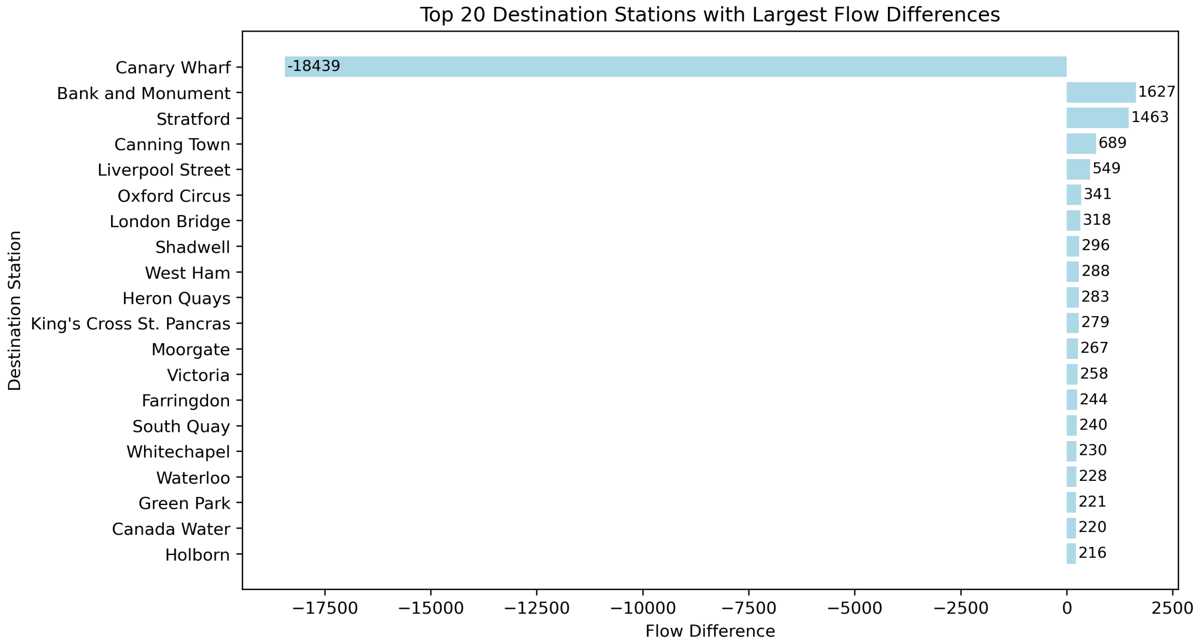

Figure 3: Top 20 Destination Stations with Largest Flow Differences in Scenario A.

With this beta value, the sensitivity of commuter flows to distance doubles compared to the original calibrated beta. This represents a scenario where the cost of transport has increased substantially, making commuters more reluctant to travel long distances.

- Beta value 2: beta(0.000150817) * 3

In this case, the sensitivity of commuter flows to distance triples compared to the original calibrated beta. This represents an even more drastic increase in the cost of transport, making distance a significant factor in determining commuter flows. This scenario could be a result of extreme increases in transportation costs, or other factors that severely discourage long-distance commuting (e.g., fuel shortages or major disruptions to the transport network).

The top 20 destination stations with the largest changes in flows for the different beta values are shown in Figures 4 and 5

IV.3 Comparison of the Impact on Commuter Flows in Different Scenarios

In terms of total absolute change from Figure 6, Scenario B (beta * 3) has a value of 1,234,620, which is significantly larger than Scenario B (beta * 2) and Scenario A. This indicates that flows are redistributed to a greater extent in Scenario B (beta * 3).

Furthermore, we evaluated the traffic changes at Abbey Road, Stratford, and West Ham these three stations in Figure 7 because, in part II.3, we found they have the greatest impact on the network when assessing the percentage change in total job flow for the critical stations, the impact of Scenario B with beta * 3 is also more pronounced. In Scenario B (beta * 2), the percentage changes in flow are 56.27% for Abbey Road, -7.35% for Stratford, and 19.89% for West Ham. In Scenario B (beta * 3), the respective changes are 141.94%, -15.48%, and 45.89%. In contrast, the percentage changes for Scenario A are notably smaller, with values of 5.12%, 2.25%, and 3.7% for the corresponding stations.

In conclusion, the analysis demonstrates that Scenario B with beta * 3 has a more substantial impact on the redistribution of flows. The results show that sharp increases in transport costs,

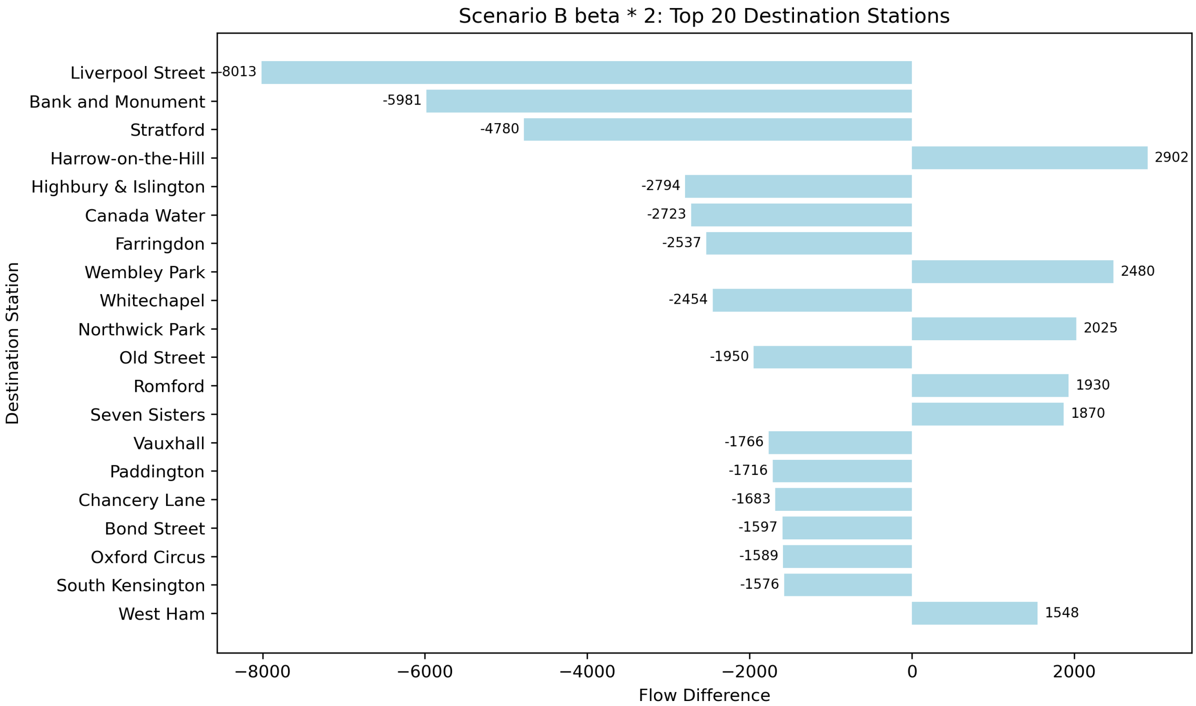

Figure 4: Top 20 Destination Stations with Largest Flow Differences in Scenario B beta * 2.

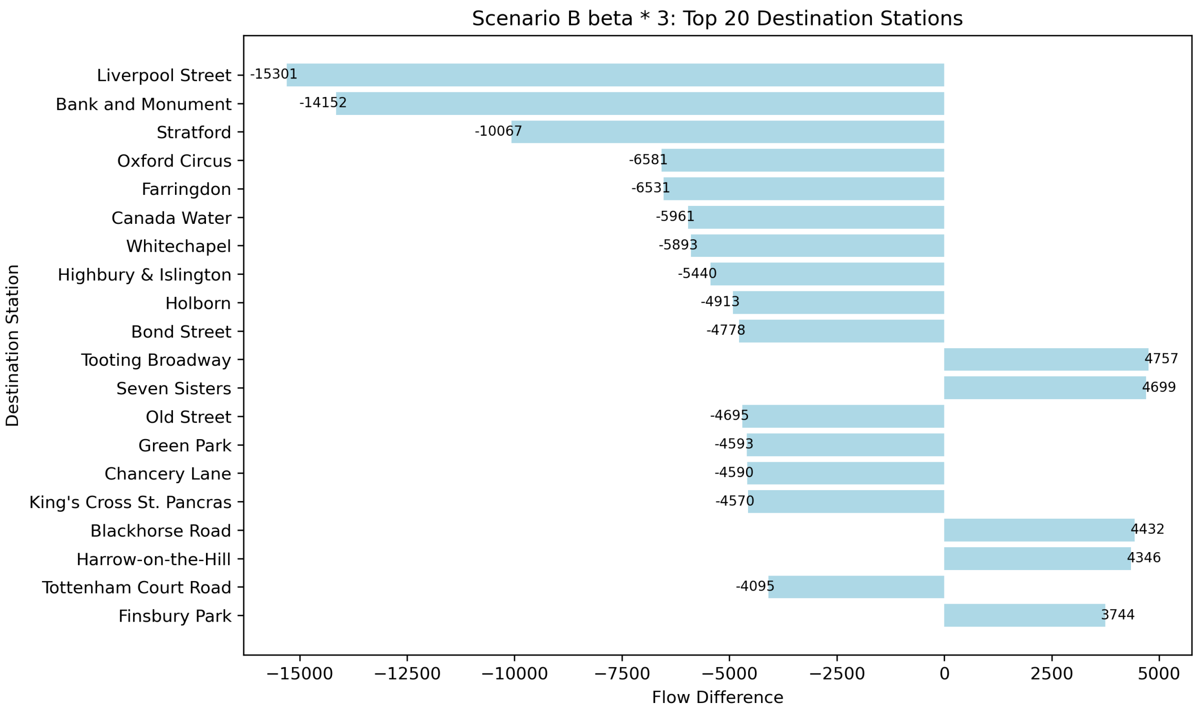

Figure 5: Top 20 Destination Stations with Largest Flow Differences in Scenario B beta * 3.

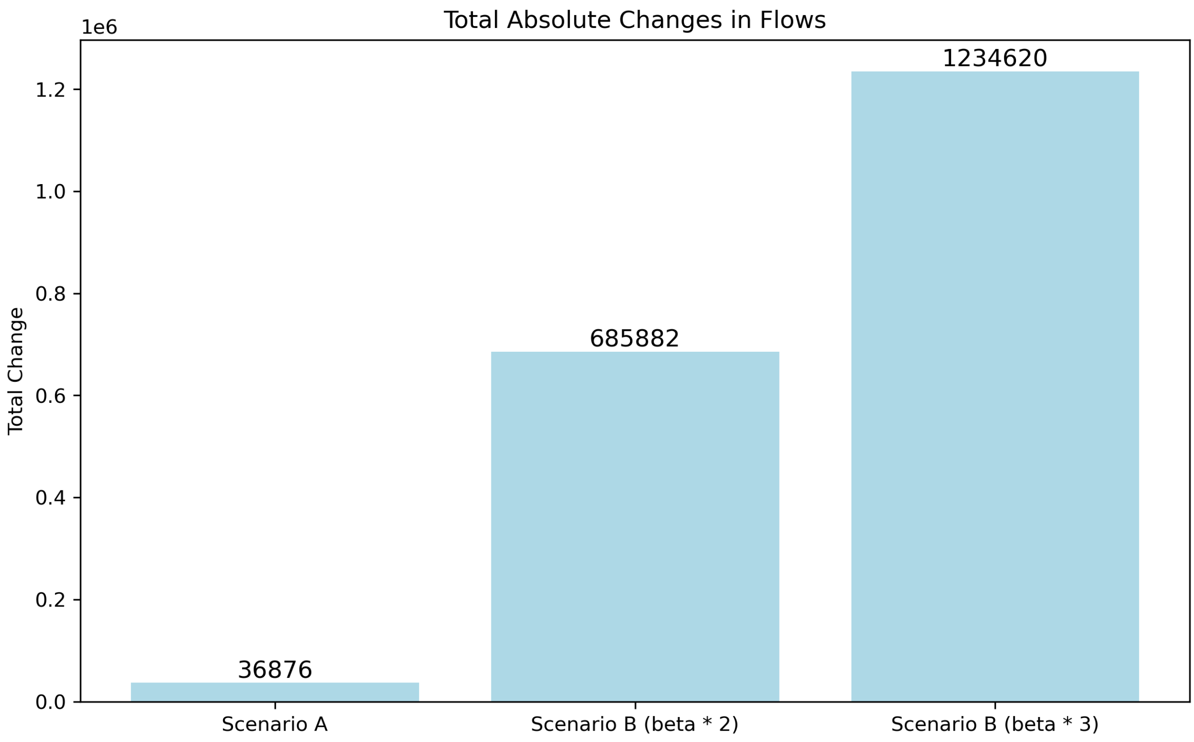

Figure 6: Comparison of Total Absolute Changes of Flows in different Scenario.

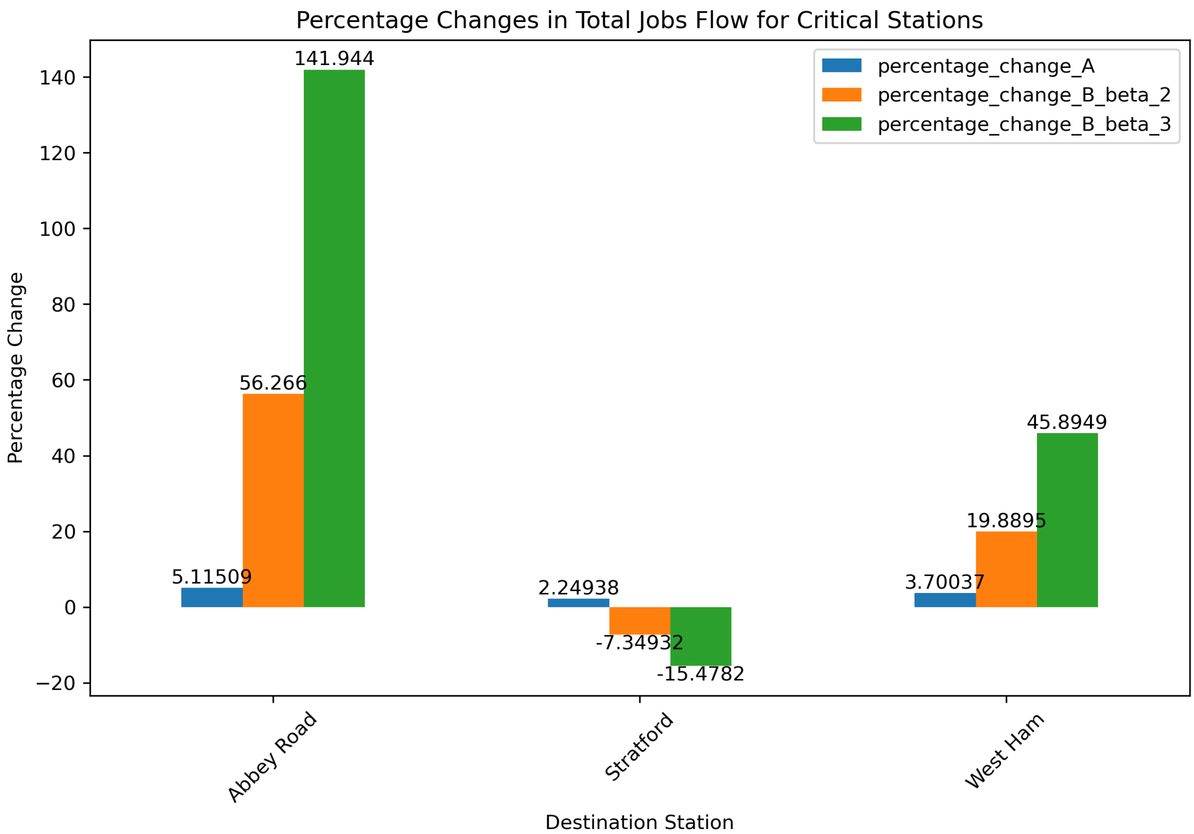

Figure 7: Comparison of Total Percentage Changes of flows for Critical Stations

fuel shortages or severe disruptions in the transport network have the greatest impact on the redistribution of flows.

Word Counts: 3089

Code Link

Code for the coursework

Reference

Fotheringham, A. S., & O'Kelly, M. E. (1989). *Spatial Interaction Models: Formulations and Applications.* Kluwer Academic Publishers.

Motter, A. E., & Lai, Y. C. (2002). *Cascading failures in complex networks.* Physical Review E, 66(6), 065102.

Rubinov, M., & Sporns, O. (2010). *Complex network measures of brain connectivity: Uses and interpretations.* Neuroimage, 52(3), 1059-1069.