Assessment for CASA0002 – Urban Simulation

Part I: London's underground resilience

1. Topological network

1.1. Centrality measures:

The London Underground can be modeled as an undirected topological network, where stations are treated as nodes and lines as edges. This framework allows for the assessment of network resilience under potential node failures. Three centrality measures are selected for analysis: Degree Centrality, Closeness Centrality, and Betweenness Centrality.

1.1.1 Degree Centrality

Degree centrality measures how many stations are directly connected to a given station (Equation 1). It reflects the importance of a node in terms of local connectivity. A higher degree value indicates stronger direct interaction capability with surrounding stations.

$$C_D(i) = \frac{k_i}{n-1} \tag{1}$$

k_i: the degree of node iii, representing the number of directly connected neighboring nodes
n: the total number of nodes in the network

In the context of a metro network, degree centrality reflects the local hub status of a station. The more directly connected stations it has, the more likely it is to serve as a primary access point into the network for surrounding areas. If such a station fails, it can disconnect multiple peripheral branches. This is especially critical in suburban areas, where branch stations tend to have lower degrees and rely heavily on central nodes for network connectivity.

For example, in the Table 1, Stratford and Bank & Monument exhibit significantly higher degree values than most other stations. This is due to their connections with multiple lines in different directions (e.g., east–west and north–south), highlighting their roles as local hubs within the network.

Table 1 Top 10 by degree centrality

	station	degree	num_links
1	Stratford	0.0225	9
2	Bank and Monument	0.0200	8
3	Baker Street	0.0175	7
4	King's Cross St. Pancras	0.0175	7
5	Liverpool Street	0.0150	6
6	Canning Town	0.0150	6
7	Green Park	0.0150	6
8	Oxford Circus	0.0150	6
9	West Ham	0.0150	6
10	Earl's Court	0.0150	6

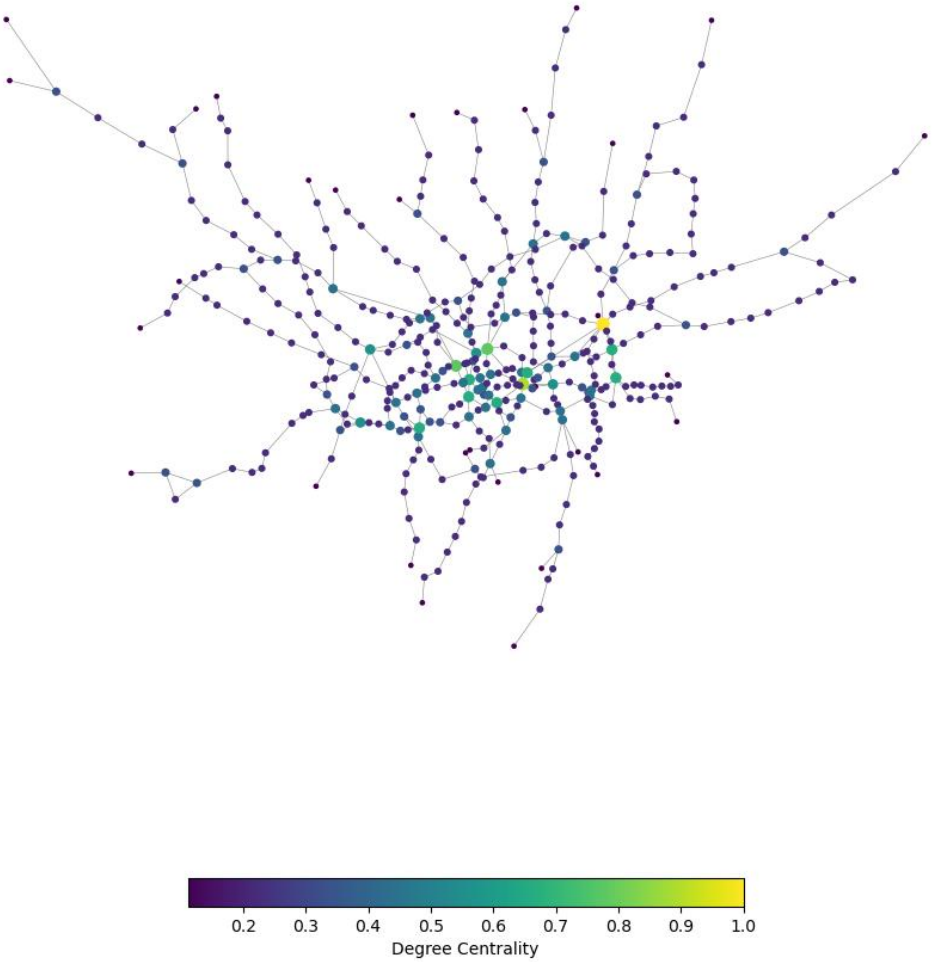

Figure 1 London tube degree centrality network

1.1. 2 Closeness Centrality

Closeness centrality is defined as the reciprocal of the average shortest path length from a given node to all other nodes in the network (Equation 2). It identifies stations that can reach all others most efficiently(Crucitti, Latora and Porta, 2006). A higher closeness value indicates that a station can access the rest of the network more quickly.

$$C_C(v) = \frac{n-1}{\sum_{u \neq v} d(u,v)} \quad (2)$$

d(u,v): the shortest path distance between node u and node v

In a metro network, closeness centrality identifies stations that serve as "time-efficient centers." It reflects a station's role as a navigational hub within the overall network structure, rather than within a single line. If a station with high closeness centrality fails, passengers are forced to take longer and more complex routes, increasing travel time and the number of transfers.

For example, stations such as Green Park has high closeness centrality because they allow passengers to reach most other stations within just a few transfers. The failure of such stations does not disconnect the network, but it significantly increases the average travel time, reducing the overall efficiency of the system and leading to a noticeable decline in service quality.

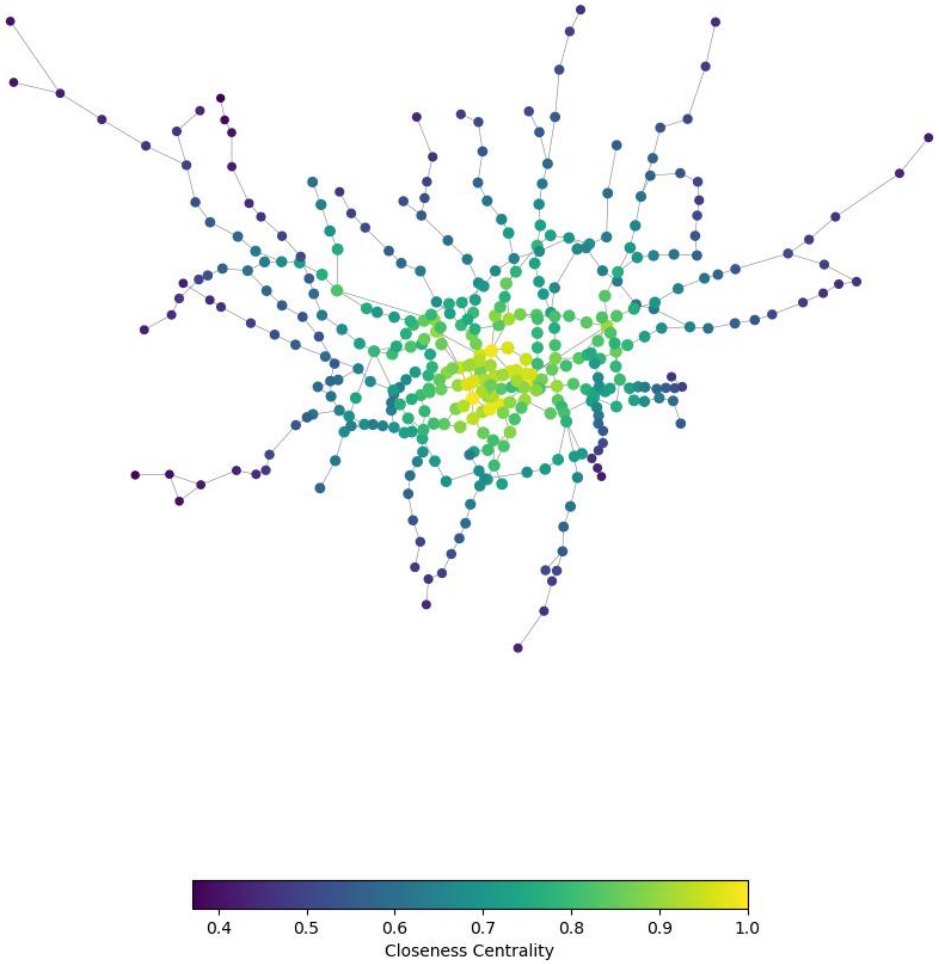

Figure 2 London tube closeness centrality network

Table 2 Top 10 by closeness centrality

	station	closeness	num_links
1	Green Park	0.114778	6
2	Bank and Monument	0.113572	8
3	King's Cross St. Pancras	0.113443	7
4	Westminster	0.112549	4
5	Waterloo	0.112265	6
6	Oxford Circus	0.111204	6
7	Bond Street	0.110988	4
8	Angel	0.110742	2
9	Farringdon	0.110742	2
10	Moorgate	0.110314	4

1.1.3 Betweenness Centrality

Betweenness centrality measures how often a given node lies on the shortest paths between other pairs of nodes (Equation 3). A higher betweenness value indicates that a large number of shortest paths pass through the node. In this sense, the node functions as a "traffic flow controller," acting as a broker or intermediary within the network.

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (3)$$

σ_{st}: *the total number of shortest paths between nodes s and t*
$\sigma_{st}(i)$: *the number of those paths that pass through node i*

n a metro network, betweenness centrality corresponds to key transfer stations that connect different regions. When such a station is disrupted, it can sever a large number of east–west or north–south travel paths, effectively breaking the "backbone" of the system. This metric reflects the probability that a station lies on a passenger's typical shortest route across the network.

If a high-betweenness station fails, it may lead to severe rerouting or even disconnection between districts. For example, Bank connects the City of London (a major financial district) with several residential zones via the Central and Northern lines. Stratford links East London suburbs to central areas through the Central, Jubilee, and DLR lines. Their high betweenness values indicate their critical roles as inter-district transfer hubs in the daily passenger flow.

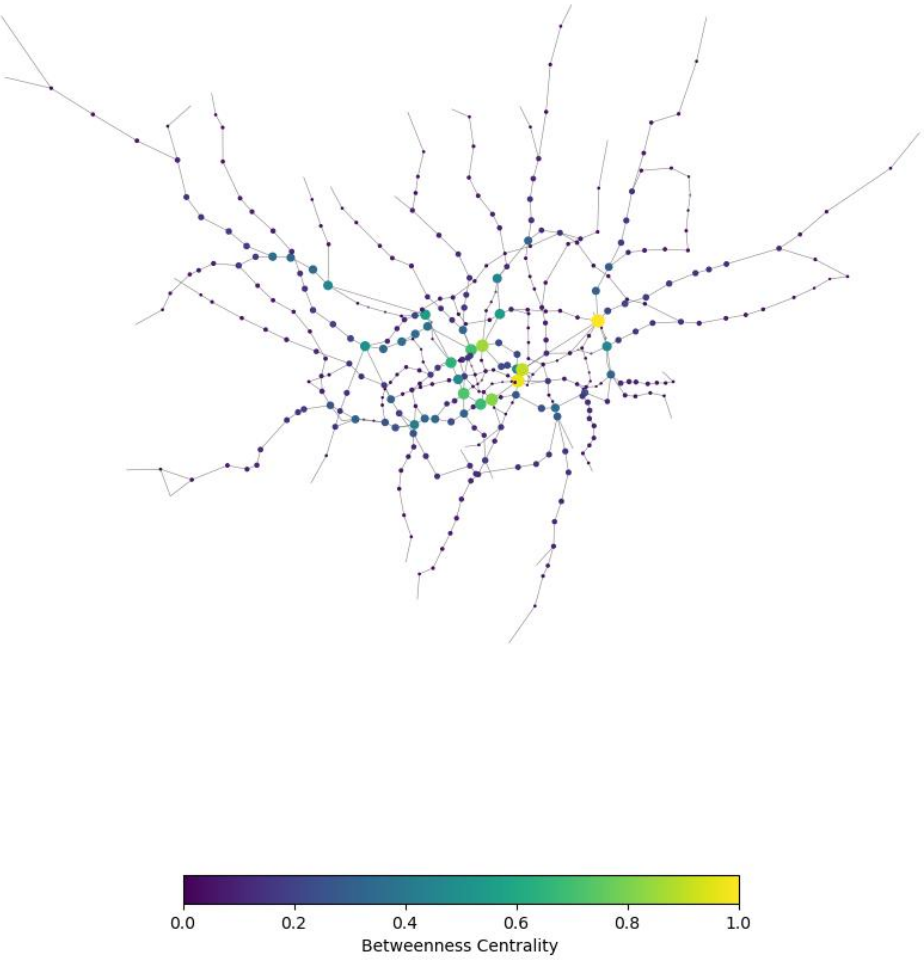

Figure 3 London tube betweenness centrality network

Table 3 Top 10 by betweenness centrality

	station	betweenness	num_links
1	Stratford	0.297846	9
2	Bank and Monument	0.290489	8
3	Liverpool Street	0.270807	6
4	King's Cross St. Pancras	0.255307	7
5	Waterloo	0.243921	6
6	Green Park	0.215835	6
7	Euston	0.208324	5
8	Westminster	0.203335	4
9	Baker Street	0.191568	7

10	Finchley Road	0.165085	4

1.2. Node removal:

In this analysis, the top-ranked nodes for each of the three centrality measures were identified and sequentially removed from the network. After each removal, centrality values were recalculated based on the updated network structure. This process was repeated until a total of 10 nodes had been removed. The impact of each removal was assessed by observing changes in the size of the Largest Connected Component (LCC)(Tables 4, 5, and 6).

Table 4 LCC reduction by removing top-degree nodes

nodes_removed	station_removed	centrality_value	degree	lcc_ratio
1	Stratford	0.0225	9	0.945137
2	Bank and Monument	0.0200	8	0.942643
3	King's Cross St. Pancras	0.0175	7	0.940150
4	Baker Street	0.0175	7	0.932668
5	Waterloo	0.0150	6	0.930175
6	Liverpool Street	0.0150	6	0.915212
7	West Ham	0.0150	6	0.907731
8	Oxford Circus	0.0150	6	0.902743
9	Green Park	0.0150	6	0.900249
10	Canning Town	0.0150	6	0.862843

Table 5 LCC reduction by removing top-closeness nodes

nodes_removed	station_removed	centrality_value	degree	lcc_ratio
1	Green Park	0.114778	6	0.997506
2	Bank and Monument	0.113572	8	0.995012
3	King's Cross St. Pancras	0.113443	7	0.992519
4	Westminster	0.112549	4	0.990025
5	Waterloo	0.112265	6	0.987531
6	Oxford Circus	0.111204	6	0.985037
7	Bond Street	0.110988	4	0.982544
8	Angel	0.110742	2	0.980050
9	Farringdon	0.110742	2	0.977556
10	Moorgate	0.110314	4	0.970075

Table 6 LCC reduction by removing top- betweenness nodes

nodes_removed	station_removed	centrality_value	degree	lcc_ratio
1	Stratford	0.297846	9	0.945137
2	Bank and Monument	0.290489	8	0.942643
3	Liverpool Street	0.270807	6	0.940150
4	King's Cross St. Pancras	0.255307	7	0.925187
5	Waterloo	0.243921	6	0.922693
6	Green Park	0.215835	6	0.920200
7	Euston	0.208324	5	0.862843
8	Westminster	0.203335	4	0.860349
9	Baker Street	0.191568	7	0.852868
10	Finchley Road	0.165085	4	0.845387

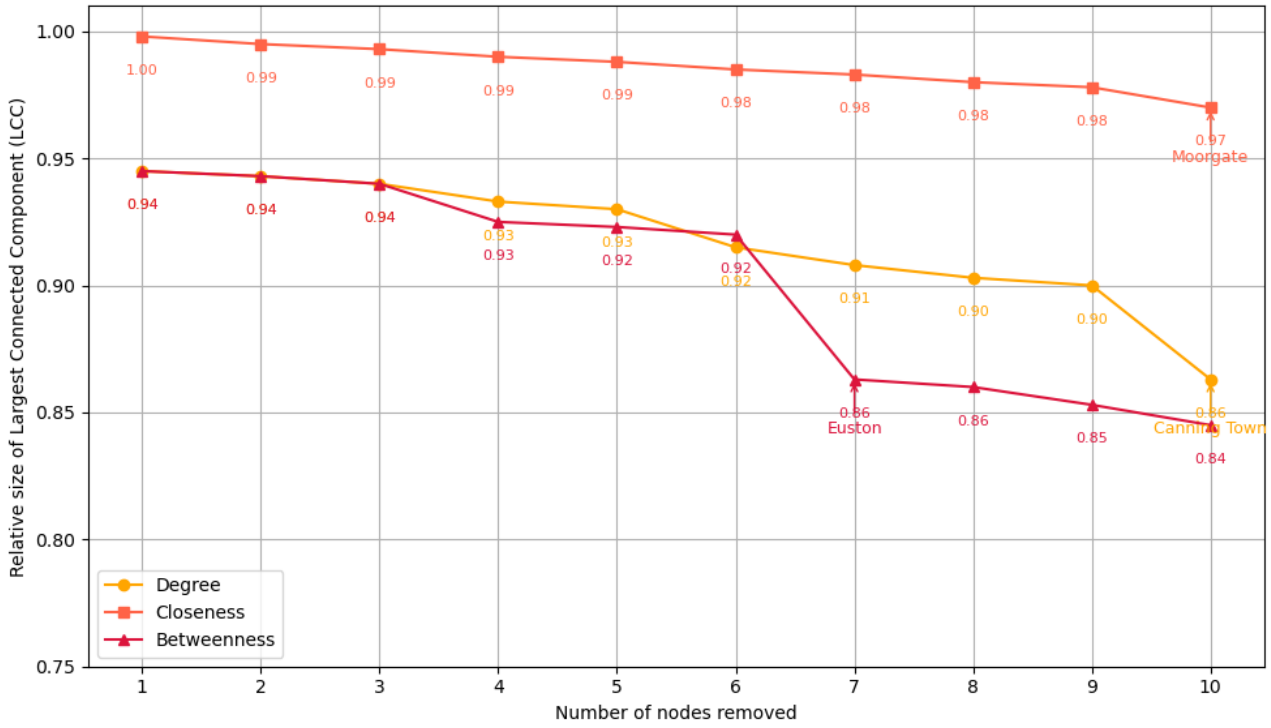

Figure 4 Comparison of Network Vulnerability by Centrality-Based Node Removal

Among the three centrality-based removal strategies, the betweenness centrality method led to the most significant decline in the size of the LCC, indicating that it effectively identifies inter-regional critical paths. As shown in Figure 4, the LCC under betweenness-based removal dropped sharply from the 6th iteration onward, decreasing from 0.94 to 0.84. In particular, the removal of Euston at iteration 7 caused a structural break, demonstrating its role as a key bridging node.

In contrast, the degree centrality strategy resulted in a more gradual decline. After removing 10 nodes, the LCC remained around 0.86, suggesting that high-degree nodes primarily influence local connectivity rather than overall network integrity.

For the closeness centrality method, the LCC exhibited the smallest change, declining only from 0.94 to 0.87. This indicates that although closeness-central nodes are efficient in terms of access, they do not necessarily function as structural bottlenecks. For example, Green Park, which has a degree of 6, connects six different stations. Removing it breaks six edges; however, these neighboring nodes are likely still connected to one another. As a result, the removal of such a node does not immediately cause significant fragmentation in the network, explaining why the LCC remained close to 400 out of 401.

Betweenness centrality is the most effective measure for identifying critical nodes whose removal leads to network fragmentation. It provides better recognition of road class(Crucitti, Latora and Porta, 2006). Specifically, the removal of bridging stations such as Euston resulted in a noticeable drop in the LCC to 0.86, indicating a structural break. Such a sharp decline was not observed under the degree or closeness centrality strategies.

2. Flows: weighted network

2.1. Centrality in the Weighted Network

A passenger flow network of the London Underground was constructed based on the NUMBAT dataset, and the flow map is presented in Figure 5. In this network, edge weights represent the number of OD passengers. Brighter colors and thicker edges indicate higher passenger volumes. Compared to the earlier topological network, this visualization reveals the vulnerability of the system in terms of actual usage, rather than structural layout.

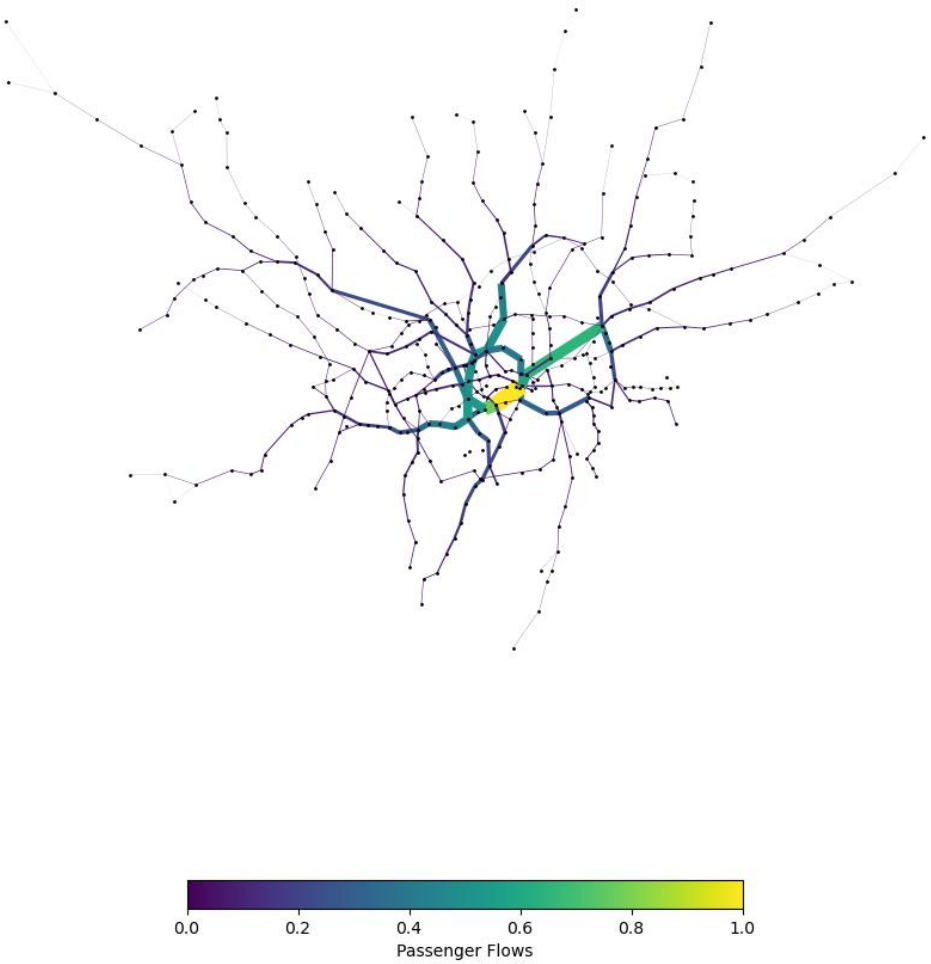

Figure 5 London network Passenger Flows

In the previous topological network, the existence of an edge only indicated whether two stations were connected. In contrast, the weighted network incorporates factors such as physical distance, travel time, or cost as measures of path efficiency. In this analysis, the physical length of each subway segment is used as the travel cost.

(1) Weighted Degree Centrality:
As in the topological network, this measure reflects the number of adjacent connections for each station. In the weighted network, the most physically connected station—i.e., the one with the highest degree centrality—is Stratford (Table 7), indicating its significance as a major structural hub.

Table 7 Top 5 by degree centrality (Weighted)

	station	degree
1	Stratford	0.0225
2	Bank and Monument	0.0200
3	King's Cross St. Pancras	0.0175
4	Baker Street	0.0175
5	Green Park	0.0150

(2) Weighted Closeness Centrality:

This measure incorporates the sum of shortest path distances using distance=length to account for the physical length of connections. The top three stations have similar closeness values, indicating that they are the most centrally located in terms of network geometry or functional accessibility (Table 8).

Table 8 Top 5 by closeness centrality (Weighted)

	station	closeness
1	Holborn	0.000079
2	King's Cross St. Pancras	0.000079
3	Tottenham Court Road	0.000079
4	Oxford Circus	0.000079
5	Leicester Square	0.000078

(3) Weighted Betweenness Centrality:

This measure uses the inverse of edge length (1 / length) as the weight, representing the probability of a station being part of a preferred, shorter path. It simulates passenger behavior that favors distance-efficient routes. Bank emerges as a critical transfer station in this analysis. Its removal would severely disrupt path availability, leading to widespread rerouting and service interruptions (Table 9).

Table 9 Top 5 by betweenness centrality (Weighted)

	station	betweenness
1	Bank and Monument	24228.0
2	Stratford	23682.0
3	Waterloo	20677.0
4	King's Cross St. Pancras	19793.0
5	Liverpool Street	19253.0

Using weighted centrality measures allows for the identification of stations that play a critical role in actual accessibility and passenger transfers, rather than simply having many structural connections. Notably, Stratford, Bank, and King's Cross consistently rank among the top stations across all three measures. This indicates that these stations are not only structurally well-connected and geographically central, but also control a substantial portion of passenger flow, functioning as core nodes within the network.

As shown in the tables, in cases where centrality values are identical, the ranking of stations may vary slightly across runs. For example, stations with equal degree values may appear in different positions due to the absence of strict ordering among ties.

2.2. OD Flow-Based Failure Simulation

To identify the most critical stations for passenger movement, this analysis focuses on actual origin–destination (OD) flows derived from the NUMBAT dataset. The computation is based on observed passenger volumes between departure and arrival stations, reflecting real patterns of network usage.

Table 10 Top 5 Passenger Flow

	Origin	Destination	Passenger Flow
1	Bank and Monument	Waterloo	252173
2	Westminster	Waterloo	195749
3	Liverpool Street	Stratford	164294
4	Euston	King's Cross St. Pancras	149606
5	Bank and Monument	Liverpool Street	148328

According to the results, Bank and Monument is the station with the highest number of originating trips, making it a major source of passenger flows in the network. It also serves as a key interchange for multiple lines, including the Central, District, and Northern lines. The calculated OD flow volume shows that Bank and Monument handles 669,127 trips as an origin station and 16,126 as a destination, resulting in a combined total of 685,253 affected trips in the event of a full closure. This figure is significantly

higher than that of any other single station in the OD dataset.

This finding aligns with the previous analysis, where Bank and Monument also ranked highest in weighted betweenness centrality. This supports the conclusion that stations with high betweenness in the weighted network tend to have the greatest impact on overall system usage when disrupted, because it best identifies the importance of connecting path structures (Derrible and Kennedy, 2010).

2.3. Alternative Walking Station Analysis

In the previous section, Bank & Monument was identified as one of the most critical stations in terms of both OD passenger flow and network structure. In this section, we simulate the station's closure and assess nearby alternative stations that are accessible on foot.

The analysis begins by creating a combined central point, calculated as the average of the coordinates of Bank and Monument stations. This point serves as the origin for the walking accessibility analysis. It is matched to the nearest street network node, along with each underground station node. Using Dijkstra's algorithm, the shortest walking distance from the origin to each station is computed.

Walking time is then calculated based on the shortest path distance, assuming an average walking speed of 1.33 m/s (equivalent to 4.8 km/h)(Department for Transport, 2015). The top accessible alternative stations and their distances and estimated walking times are summarized in Table 11.

Table 11 Top 3 Nearest Stations

	name	distance_	Walk time (sec)	Walk time (min)
1	Mansion House	804.372357	604.791246	10.079854
2	London Fenchurch Street	880.099353	661.728837	11.028814
3	London Liverpool Street	939.771527	706.595133	11.776586

Figure 6 Shortest Walking Path: Bank & Monument (right)→ Mansion House(left)

Part II: Spatial Interaction models

1. Models and calibration

1.1. Spatial Interaction Models

(1) Unconstrained Gravity Model

The core idea of spatial interaction models is that the strength of interaction between two locations is directly proportional to the "masses" or characteristics of the origin and destination, and inversely proportional to the distance or travel cost between them. In this framework, the interaction flow from origin i to destination j, denoted as T_{ij}, is assumed to be proportional to the production potential of the origin (O_i) and the attractiveness of the destination (D_j), and inversely proportional to the separation or impedance (d_{ij}) between them.

This type of model is typically used for exploratory analysis or theoretical modeling when actual OD constraints are unavailable. Common applications include modeling inter-city migration trends, estimating the spatial influence of commercial centers, analyzing long-distance commuting patterns, or simulating catchment areas.

$$T_{ij} = k \cdot O_i^{\alpha} D_j^{\gamma} d_{ij}^{-\beta} \tag{4}$$

T_{ij}: Flow from origin iii to destination j
O_i: Origin mass (e.g., population, jobs)
D_j: Destination attractiveness (e.g., services, jobs)
d_{ij}: Distance or generalized travel cost between iii and j
α: Origin elasticity — how strongly origin mass affects flow
γ: Destination elasticity — how strongly destination mass affects flow
β: Distance decay parameter — higher values mean stronger resistance to distance
k: Scaling constant

(2) Production-Constrained Model

This model constrains the total outflow from each origin by introducing a balancing factor A_i , which adjusts the distribution so that the total flow originating from location i matches a known value. For example, this may represent the total number of trips generated from each residential area or origin zone.

The production-constrained model is well-suited for systems where the total number of departures is fixed, such as daily commuting patterns in which the number of outbound trips from households remains stable. It is commonly applied to predict the commuting destinations of residents from different neighborhoods, or to estimate the outbound flow for supermarket delivery routes.

$$T_{ij} = A_i \cdot O_i D_j^{\gamma} d_{ij}^{-\beta} \tag{5}$$

(3) Attraction-Constrained Model

This model constrains the total inflow to each destination by applying a balancing factor B_j , which ensures that the sum of flows arriving at destination j matches a known value D_j. For example, this may represent the maximum number of students a school can enroll or the number of available hospital beds.

The attraction-constrained model is appropriate for scenarios where receiving capacity is known or limited, such as allocating students to schools, estimating hospital catchment areas, or modeling service distribution for constrained facilities.

$$T_{ij} = B_j \cdot D_j O_i^{\alpha} d_{ij}^{-\beta} \tag{6}$$

(4) Doubly Constrained Model

This model simultaneously constrains both the total outflow from each origin O_i and the total inflow to each destination D_j . To satisfy both constraints, it requires solving for balancing factors $A_i B_j$ through an iterative algorithm.

The doubly-constrained model is typically used for high-accuracy prediction of actual OD flows and for calibrating migration or trip matrices. It is widely applied in real-world transportation modeling, such as estimating passenger flows from smartcard data in bus or subway systems, and is one of the most commonly used models for empirical calibration.

$$T_{ij} = A_i B_j O_i D_j d_{ij}^{-\beta} \tag{7}$$

1.2. Select Model

In this study, the Unconstrained Gravity Model is selected. This choice is motivated by the objective of estimating relative attractiveness, rather than replicating actual OD flows, and the absence of complete OD volume data. The model is computationally efficient and well-suited for predicting the potential flow of residents from various origin zones to different candidate supermarket locations.

The model integrates population size, retail attractiveness (measured by store area), and travel distance to quantify the "gravitational pull" that each supermarket exerts on surrounding residential areas. This makes it a practical tool for informing location-based retail planning.

The parameters used in the model were derived from a template-based calibration:

- α =1.7558: Since α >1, it suggests that areas with larger populations contribute disproportionately higher trip volumes.
- γ =1.6472: A value greater than 1 indicates that larger supermarkets have stronger attraction and can draw more customers.
- β =1.4079: This implies strong distance decay; residents are sensitive to travel distance, and long-distance trips are much less likely.

Since the true scaling constant k is unknown, the model outputs are normalized to align the total predicted flow with the overall population size. This enhances the interpretability and comparability of the estimated flows.

Figure 7 shows the distribution of log-transformed gravity model flows. The distribution is strongly right-skewed, which is typical for spatial interaction models due to the presence of high-attraction, short-distance OD pairs. This pattern also suggests that the calibrated parameters are broadly reasonable, as they produce results consistent with expected spatial flow behavior.

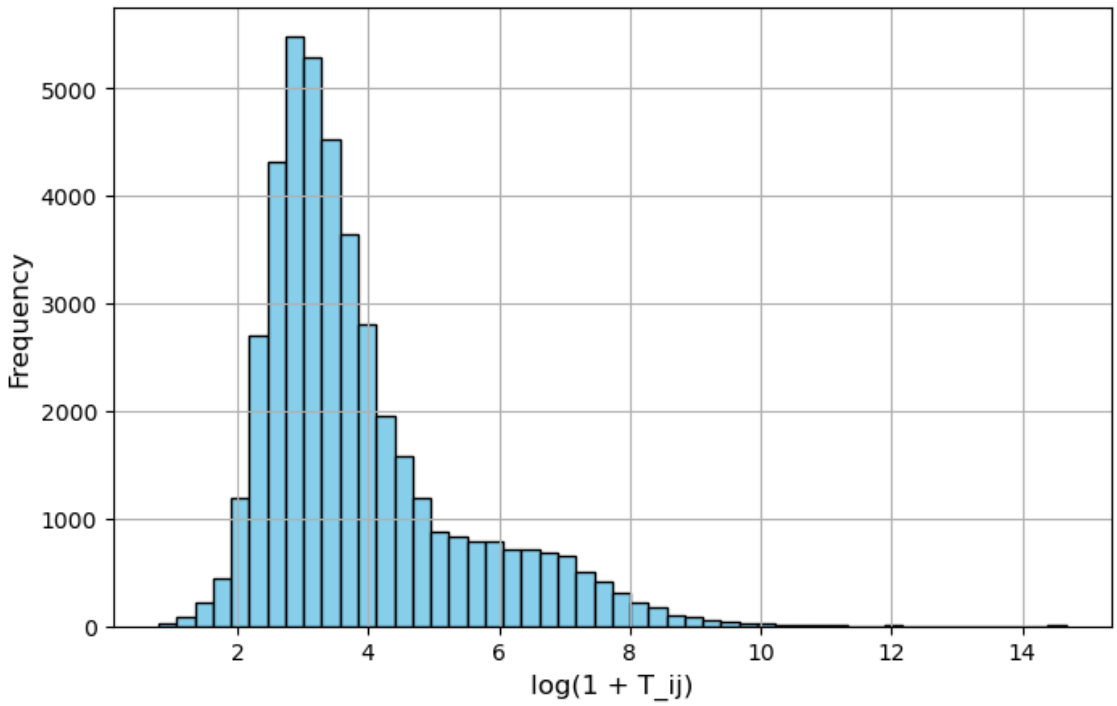

Figure 7 Log-Transformed Gravity Flow Distribution

2. Scenarios

2.1. The retail site selection

The retail site selection analysis focuses on comparing the potential customer flows attracted by two candidate locations—Option 1 and Option 2—across the study area, using a gravity model framework to support the decision-making process.

In the computation, each candidate location was assigned a centroid, and two separate distance matrices were constructed accordingly. These were used to calculate shortest-path distances between each origin OA zone and the respective candidate location. The gravity model was then applied using fixed parameters , while the scaling constant *k* was normalized to ensure comparability. Self-loop flows—i.e., zones where the origin and destination are the same—were excluded to avoid illogical results.

The results show that Option 1 has a total predicted flow of approximately 0.000383, while Option 2 has a flow of about 0.000237. This indicates that Option 1 is expected to attract 61.8% more potential customers than Option 2. Considering factors such as distance, population, and attractiveness, Option 1 appears to have stronger accessibility and greater market potential.

Figure 8 further confirms the stronger spatial influence of Option 1 across the study area. Its gravitational reach is particularly pronounced in the northern and central urban zones, where the station exhibits denser and more widespread connections.

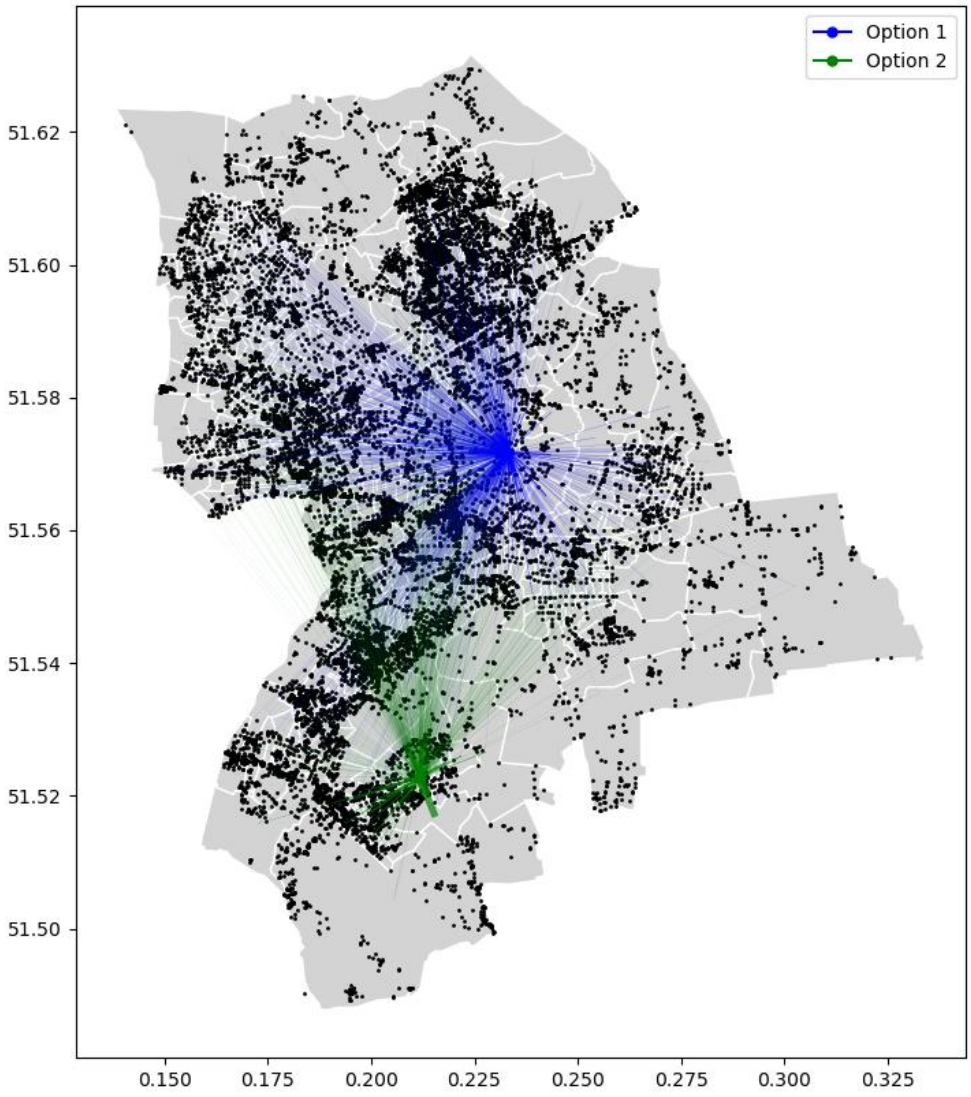

Figure 8 Flows to Candidate Locations

2.2. Sharp Increase In Transport

This scenario examines how the attractiveness of each candidate location changes under increased travel resistance. To simulate a high-cost, high-friction environment, the distance decay parameter β in the gravity model was artificially increased from 1.4079 to 5.0.

Using the updated β value, flows were recalculated for both Option 1 and Option 2. The total predicted flows under this high-cost scenario were then compared to those in the original baseline scenario. The results show that Option 1's total predicted flow dropped to approximately 5.83×10^{-36}, while Option 2 retained a slightly higher total of about 6.16×10^{-36}. Although both flows decreased to near-zero values, Option 2 preserved a larger proportion of its original flow, with only a 5.3% reduction. In contrast, Option 1, which initially had a broader attraction range and relied more heavily on longer-distance flows, experienced an almost complete collapse in predicted demand.

As shown in Figure 9, Option 1 has lost nearly all of its connections under the high-cost scenario, with only a few short-distance links remaining. In contrast, Option 2 retains a limited but stable set of local flows, maintaining access from several nearby residential zones. This suggests that Option 2 offers stronger local accessibility and greater resilience when travel costs increase. Therefore, in contexts where future transport conditions are uncertain, locations with higher walkability and local reach—such as Option 2—may be preferable, or a multi-site strategy may be considered to ensure broader coverage.

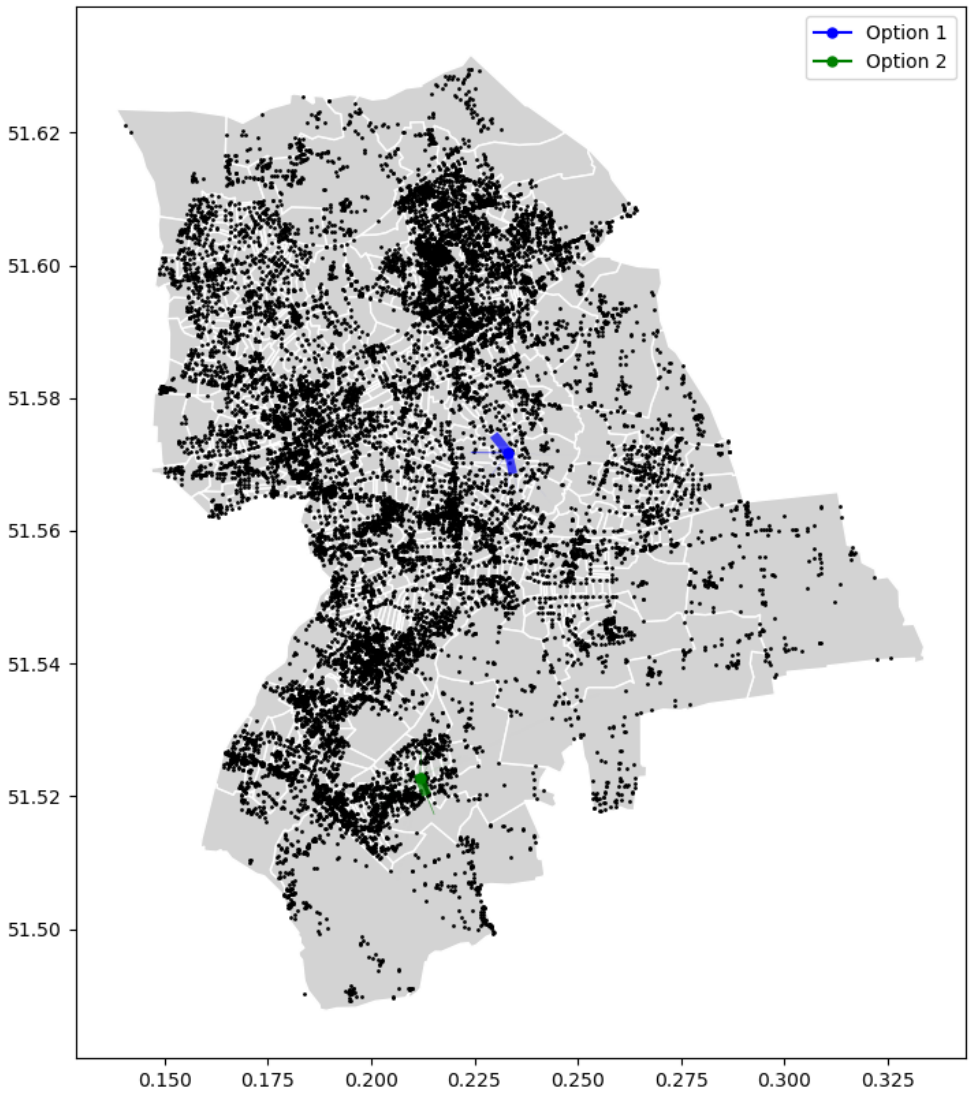

Figure 9 (High cost) Flows to Candidate Locations

Extra 5 points:

To estimate the number of residents that can access each candidate site within a 15-minute walking distance, the analysis adopts a walking speed of 1.33 m/s(Department for Transport, 2015), resulting in a maximum walkable distance of 1,197 meters (1.33 × 60 × 15). Using the distance matrix computed in Section 2.1, the total population living in origin zones within this threshold was identified for each candidate location.

The results show that Option 1 has 4,517 residents within walking range, while Option 2 reaches 6,589 residents—approximately 46% more. This indicates that Option 2 benefits from a denser residential environment, greater geographic centrality, or a more accessible street network. From the perspective of local accessibility, Option 2 holds a clear advantage. In the context of increasing interest in walkability, low-carbon travel, and accessibility for non-driving or elderly populations, such metrics are becoming essential in retail site selection.

Reference

Crucitti, P., Latora, V. and Porta, S. (2006) 'Centrality Measures in Spatial Networks of Urban Streets', *Physical Review E*, 73(3), p. 036125. Available at: https://doi.org/10.1103/PhysRevE.73.036125.

Department for Transport (2015) 'Journey Time Statistics: Notes and Definitions'. Available at: https://assets.publishing.service.gov.uk/media/5dfa46f2ed915d54ab87c859/notes-and-definitions.pdf.

Derrible, S. and Kennedy, C. (2010) 'The complexity and robustness of metro networks', *Physica A: Statistical Mechanics and its Applications*, 389(17), pp. 3678–3691. Available at: https://doi.org/10.1016/j.physa.2010.04.008.

Github link: https://github.com/YULI61/Urbsim_Assessment-