Urban Simulation Project

Word Count: 2997

23 April 2024

Github LINK

1 Topological Network

1.1 Centrality Measures

Centrality measures are essential metrics employed to identify the most important nodes in a network (Rodrigues, 2019). Regarding the underground network, these matrices can identify stations (hubs) that are essential in the functioning of the underground network.

1.1.1 Degree Centrality

Degree centrality is a local measure that considers the immediate connections that a node has making it a straightforward measure but powerful way to assess the potential influence or activity level of a node (Freeman, 2002).

$$C_D(v) = \frac{deg(v)}{n-1} \tag{1}$$

Regarding the underground network, stations with a higher degree of centrality are those having many connections with other stations depicting the station's importance in terms of connectivity and accessibility. They also represent stations where many passengers switch lines in the underground network indicating their essence in facilitating movement, travel time, and functionality in the overall network.

Table 1: Degree Centrality

Serial No.	Stations	Degree Centrality
1	Stratford	0.0225
2	Bank and Monument	0.0200
3	Baker Street	0.0175
4	King's Cross St. Pancras	0.0175
5	Liverpool Street	0.0150
6	Canning Town	0.0150
7	Waterloo	0.0150
8	Green Park	0.0150
9	Oxford Circus	0.0150
10	West Ham	0.0150

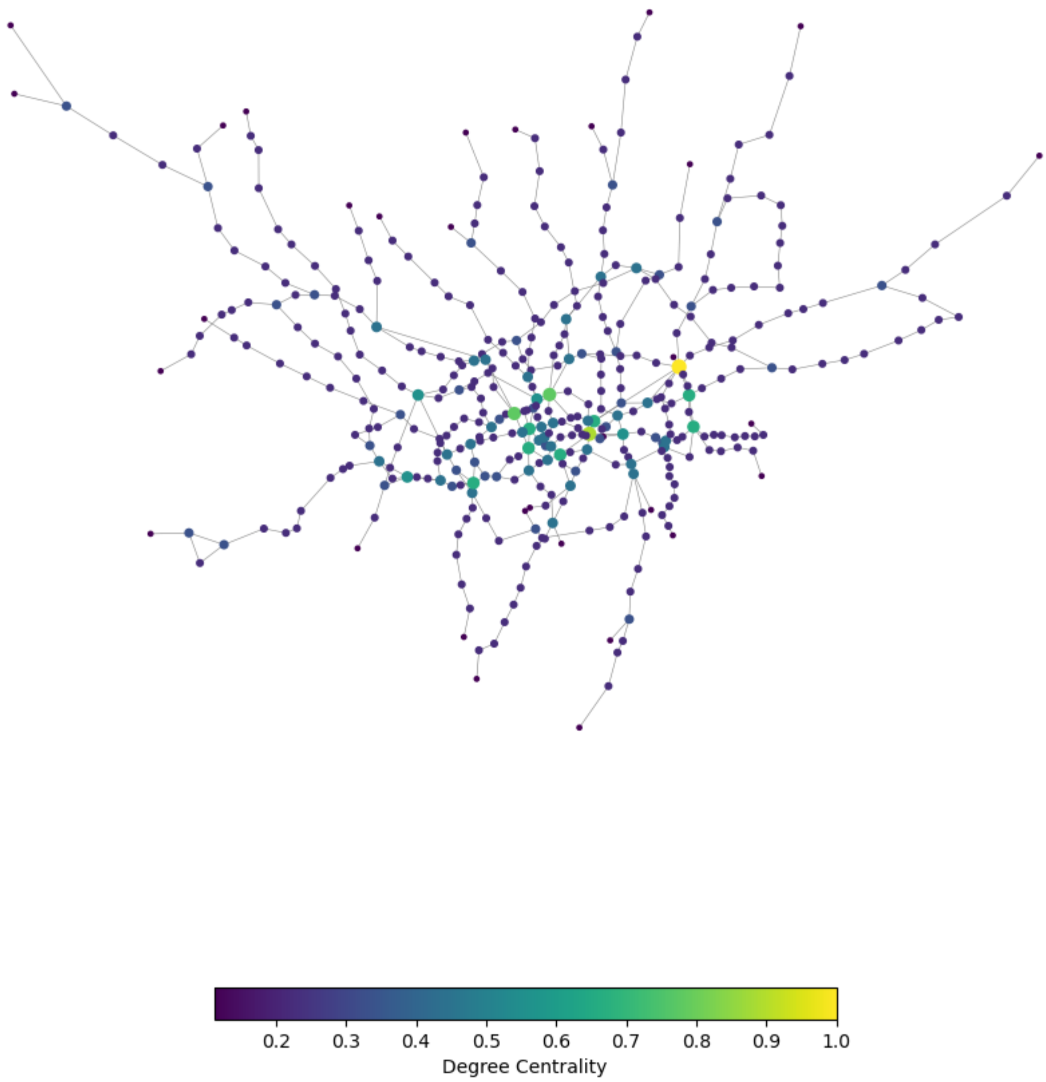
Table 1, shows top ten ranked degree centrality of the topological network in ascending order with Stratford Station recording the highest.

1.1.2 Betweenness Centrality

As Freeman (1977) stated, betweenness centrality is a measure used to calculate how much a node is positioned on the shortest paths between other nodes within the network. It is essentially useful in identifying important nodes that, if removed, would significantly affect the interactions in the network.

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

Figure 1: London Tube Degree Centrality Plot



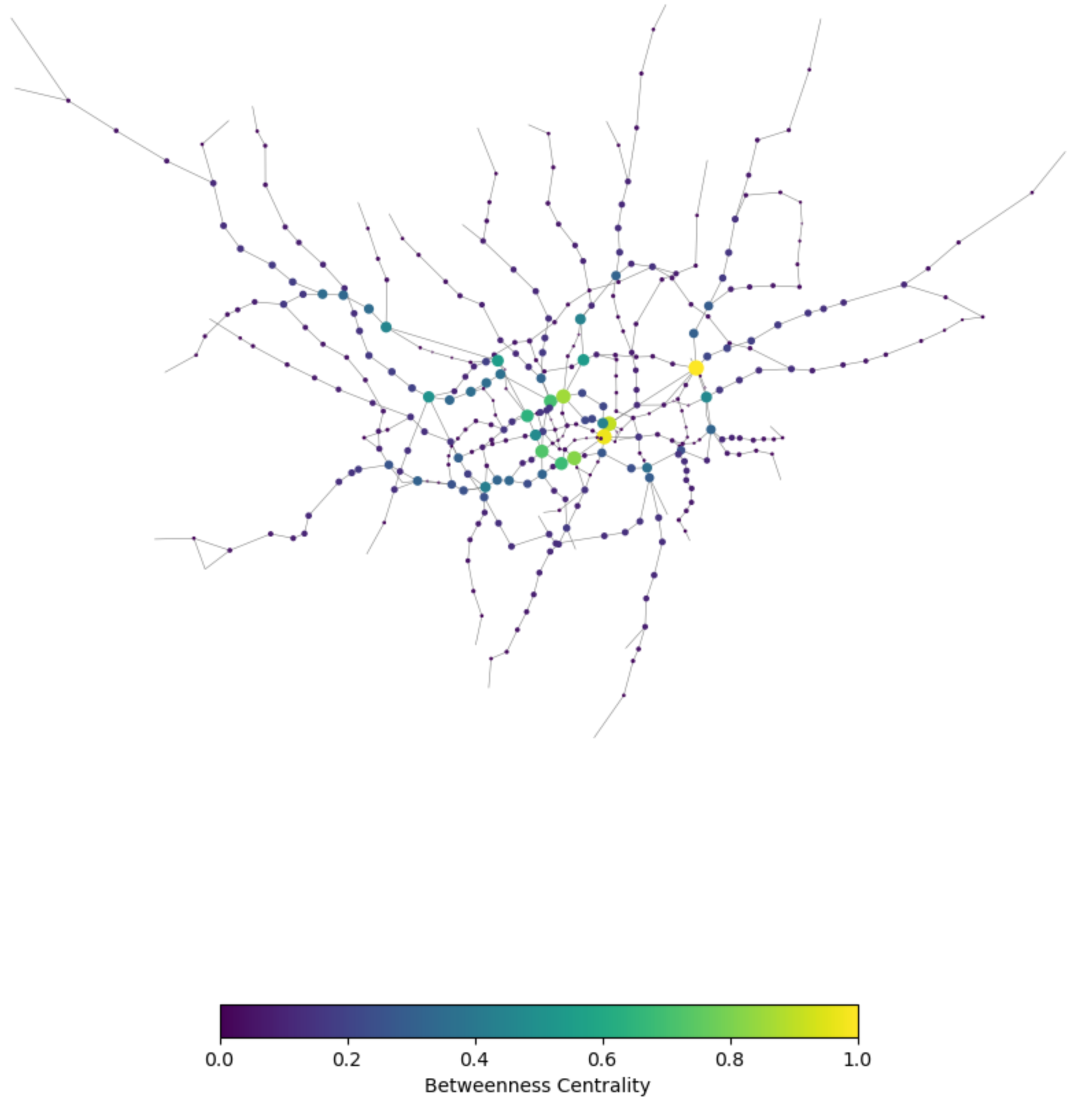
Relating to the underground network, stations with high betweenness centrality lie on many shortest paths between other stations, which means they are essential in connecting different parts and facilitating passenger flow in the network. They also promote the movement of passengers between different parts of the network.

Table 2: Betweenness Centrality

Serial No.	Stations	Betweenness Centrality
1	Stratford	23768.09
2	Bank and Monument	23181.06
3	Liverpool Street	21610.39
4	King's Cross St. Pancras	20373.52
5	Waterloo	19464.88
6	Green Park	17223.62
7	Euston	16624.28
8	Westminster	16226.16
9	Baker Street	15287.11
10	Finchley Road	13173.76

Table 2, shows the top ten ranked Betweenness centrality of the topological network in ascending order with Stratford Station recording the highest.

Figure 2: London Tube Betweenness Centrality Plot



1.1.3 Closeness Centrality

It evaluates a node's proximity to other nodes within the network, showing its effectiveness in accessing or interacting with the network (Freeman, 2002). Nodes with high closeness centrality are those that can reach other nodes quickly.

$$C_i = \frac{1}{l_i} = \frac{n}{\sum_j d_{ij}} \tag{3}$$

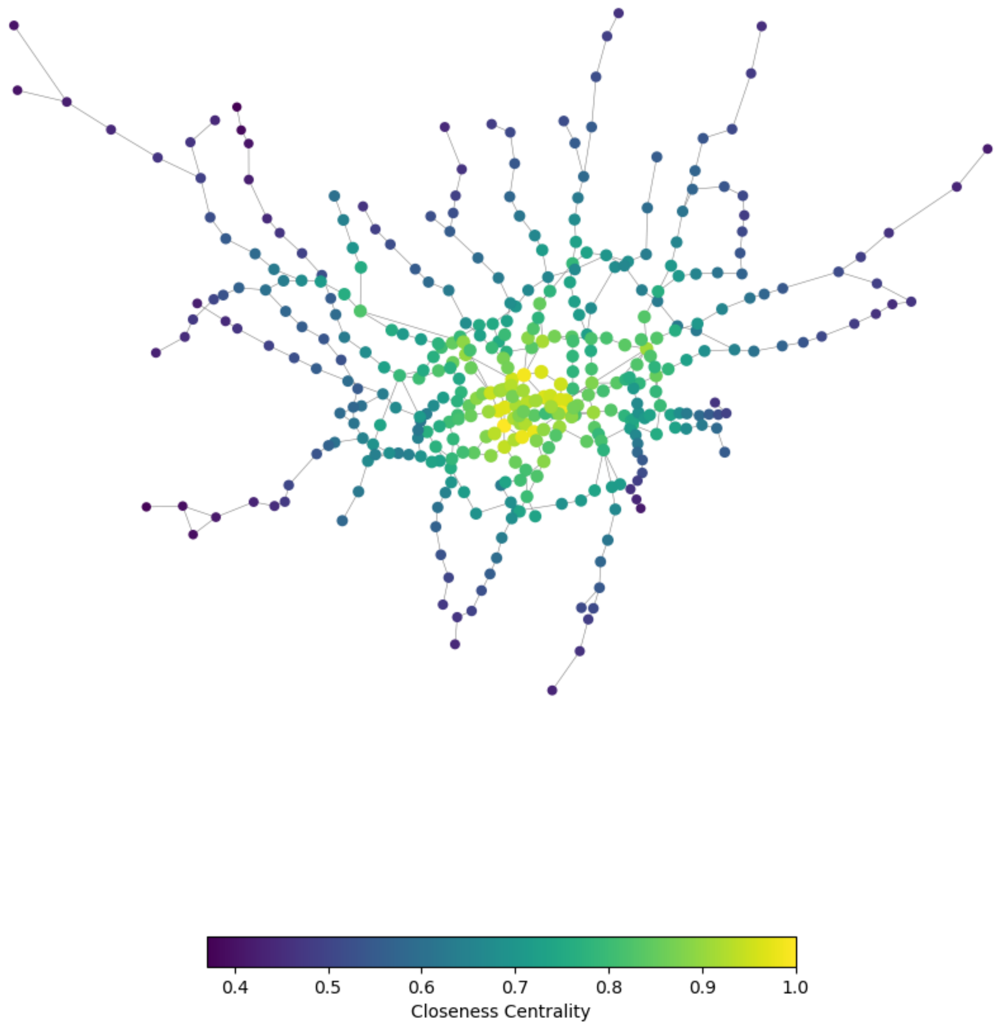
Regarding the underground network, closeness centrality represents how easily passengers can reach a station from other stations making them vital hubs for passenger movement.

Table 3: Closeness Centrality

Serial No.	Stations	Closeness Centrality
1	Green Park	0.114778
2	Bank and Monument	0.113572
3	King's Cross St. Pancras	0.113443
4	Westminster	0.112549
5	Waterloo	0.112265
6	Oxford Circus	0.111204
7	Bond Street	0.110988
8	Angel	0.110742
9	Farringdon	0.110742
10	Moorgate	0.110314

Table 3, shows the top ten ranked closeness centrality of the topological network in ascending order with Stratford Station recording the highest.

Figure 3: London Tube Closeness Centrality Plot



1.2 Impact Measures

Assessing the impact of node removal in the transportation network requires global measures like Average Shortest Path Length and Global Efficiency. These metrics are vital for understanding a network's resilience and connectivity, applicable to scenarios from the London Underground to broader network models, and provide a clear picture of the structural efficiency and robustness across various types of networks (Mao and Zhang, 2013; Mouronte-López, 2021).

1.2.1 Average Shortest Path Length (ASPL)

According to Mao and Zhang (2013), ASPL is the mean distance of the shortest paths connecting every pair of nodes in a network topology. It evaluates how efficient information or mass transportation functions in a network.

$$ASPL(G) = \frac{1}{N(N-1)} \sum_{i \neq j} d(n_i, n_j) \tag{4}$$

Where: G is the graph, N is the total number of nodes in the graph, $d(n_i, n_j)$ is the shortest path distance between nodes n_i and n_j, The summation is over all pairs of nodes $i \neq j$.

If a node is removed, ASPL typically increases particularly when the node is highly connected which shows that the network becomes less efficient and more dispersed. The rise in path length after removing nodes emphasizes key nodes that greatly affect network performance. This measurement applies to all types of networks, such as the London Underground, social networks, etc., regardless of their specific characteristics.

1.2.2 Global Efficiency (GE)

Global Efficiency is a metric that assesses the effectiveness of information exchange across the network. It represents the network's functional efficiency by calculating the average inverse shortest path length (Mouronte-López, 2021; Ek et al., 2015).

$$GE = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d(n_i, n_j)} \tag{5}$$

Where: $d(n_i, n_j)$ is the shortest path length between nodes n_i and n_j.

Removing key nodes will cause a decrease in global efficiency, signaling a reduction in the network's overall ability to support effective interactions. Reduction in GE after removing a node helps to identify possible weak hubs in the network infrastructure. Both ASPL and GE applicability are not limited to just the London Underground. They offer vital understandings of the resilience of different network types by measuring how the removal of nodes can impact network operations globally.

1.3 Node Removal

1.3.1 (A) Non-Sequential Removal of Nodes: Degree Centrality

Table 4: Topological Degree Centrality

Serial	Station Removed	No.Components	ASPL	GE	Component Size
0	Liverpool Street	1	14.1003	0.0979	400
1	Waterloo	1	14.4150	0.0957	399
2	Oxford Circus	1	14.5106	0.0948	398
3	Euston	2	14.6897	0.0953	375
4	Shadwell	3	14.8173	0.0943	373
5	Stratford	5	15.3491	0.0933	351
6	Green Park	5	15.5908	0.0914	350
7	Earl's Court	5	15.9755	0.0894	349
8	Baker Street	6	16.9779	0.0850	347
9	West Ham	7	17.6003	0.0830	344

1.3.2 Non-Sequential Removal of Nodes: Betweenness Centrality

Table 5: Topological Betweenness Centrality

Serial	Station Removed	No.Components	ASPL	GE	Component Size
0	King's Cross St. Pancras	1	14.2508	0.0970	400
1	Finchley Road	1	15.3150	0.0929	399
2	Liverpool Street	1	15.9102	0.0896	398
3	Whitechapel	1	16.5962	0.0866	397
4	Willesden Junction	2	17.4491	0.0850	383
5	Bermondsey	2	17.7373	0.0838	382
6	Shadwell	3	19.9658	0.0782	380
7	West Brompton	4	27.0280	0.0713	372
8	Willesden Green	5	14.3254	0.1156	188
9	Canning Town	7	14.3253	0.1156	188

1.3.3 Non-Sequential Removal of Nodes: Closeness Centrality

Table 6: Topological Closeness Centrality

Serial	Station Removed	No.Components	ASPL	GE	Component Size
0	King's Cross St. Pancras	35	9.6140	0.1929	39
1	Park Royal	36	8.4506	0.2224	30
2	Stockwell	39	8.4506	0.2224	30
3	Leicester Square	40	8.4506	0.2224	30
4	Paddington	43	8.4506	0.2224	30
5	South Woodford	43	8.4506	0.2224	30
6	Royal Victoria	44	8.4506	0.2224	30
7	Finchley Road	45	8.4506	0.2224	30
8	Seven Sisters	47	8.4506	0.2224	30
9	Ealing Common	49	8.4506	0.2224	30

The tables indicate that non-sequential removal of nodes based on different centrality measures affects the network's resilience differently. Degree centrality, suggests that stations like Liverpool Street and Waterloo are critical, as their removal significantly impacts both GE and ASPL. However, when considering Betweenness centrality, removing a station like King's Cross St. Pancras causes an increase in ASPL, signifying its importance for the functioning of the underground. Closeness centrality results also revealed an increase in ASPL and significant drops in GE, indicating its sensitivity in reflecting the stations' reachability. In studying resilience, the strategy focusing on Betweenness centrality appears more effective as it captures the impact of node removal on network flow more distinctly. For assessing the effect after node removal, GE seems to provide a clearer, quantitative measure of the network's functional efficiency, while ASPL shows how accessible parts of the network are from one another, with the latter being better at capturing the increased travel or transmission costs post-removal.

1.3.4 (B) Sequential Removal of Nodes: Degree Centrality

Table 7: Topological Degree Centrality

Serial	Station Removed	No.Components	ASPL	GE	Component Size
1	Stratford	3	14.4964	0.0982	379
2	Bank and Monument	3	14.8725	0.0948	378
3	King's Cross St. Pancras	3	16.0549	0.0886	377
4	Baker Street	4	17.0134	0.0844	374
5	Green Park	4	17.1753	0.0833	373
6	Canning Town	6	17.5311	0.0828	359
7	Earl's Court	6	17.8947	0.0808	358
78	Oxford Circus	7	18.0391	0.0800	356
9	Willesden Junction	8	20.0615	0.0754	342
10	Waterloo	8	21.2073	0.0727	341

Table 8: Topological Betweenness Centrality

Serial	Station Removed	No.Components	ASPL	GE	Component Size
1	Stratford	3	14.49	0.0982	379
2	King's Cross St. Pancras	3	15.31	0.0934	378
3	Waterloo	3	15.80	0.0904	377
4	Bank and Monument	3	16.79	0.0858	376
5	Canada Water	3	19.03	0.0804	375
6	West Hampstead	4	13.46	0.1086	227
7	Earl's Court	4	14.21	0.1041	226
8	Shepherd's Bush	5	13.79	0.1091	196
9	Euston	6	13.82	0.1131	173
10	Baker Street	7	18.19	0.0976	170

Table 9: Topological Closeness Centrality

Serial	Station Removed	No.Components	ASPL	GE	Component Size
1	Green Park	1	13.82	0.0992	400
2	King's Cross St. Pancras	1	14.66	0.0944	399
3	Waterloo	1	15.11	0.0918	398
4	Bank and Monument	1	16.70	0.0854	397
5	West Hampstead	1	18.97	0.0805	396
6	Canada Water	2	13.98	0.1047	226
7	Stratford	4	13.98	0.1047	226
8	Earl's Court	4	14.73	0.1002	225
9	Shepherd's Bush	5	14.75	0.1034	195
10	Oxford Circus	5	15.66	0.0977	194

In the above tables, Betweenness Centrality seems to most accurately indicate a station's significance in the operation of the underground system. The removal of important stations like Stratford and King's Cross St. Pancras causes a significant disruption in the network, as evidenced by the changes in the GE and ASPL. Removing nodes one by one seems more efficient in studying resilience because it demonstrates the combined impact of removing multiple nodes. GE is more sensitive for assessing immediate functional damage after node removal, while ASPL helps understand the broader implications for network navigability in the long run.

1.4 (C) Visualisation of the Results of the Two Global Measures

1.4.1 Degree Centrality

Figure 4: Non-Sequential Node Removal Degree Centrality Plot

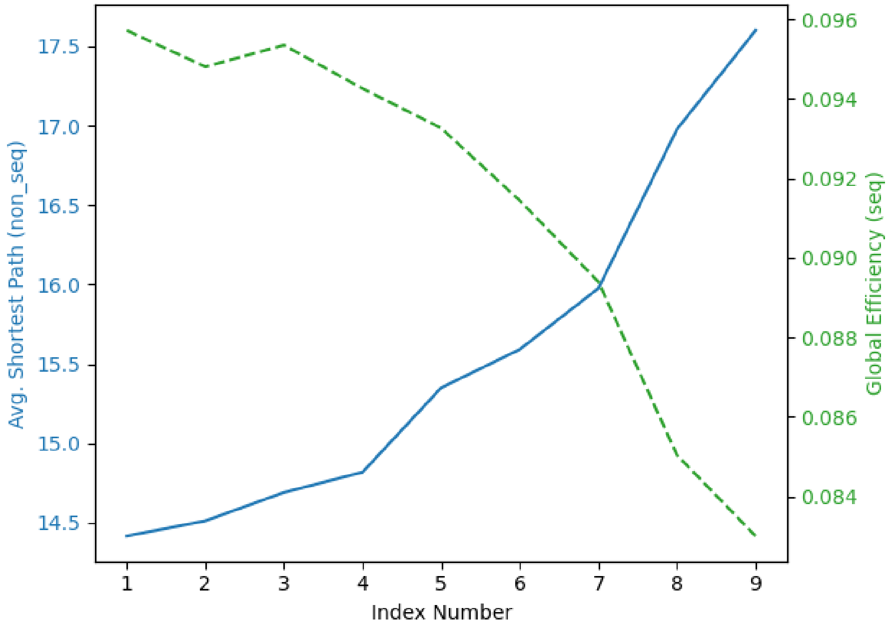
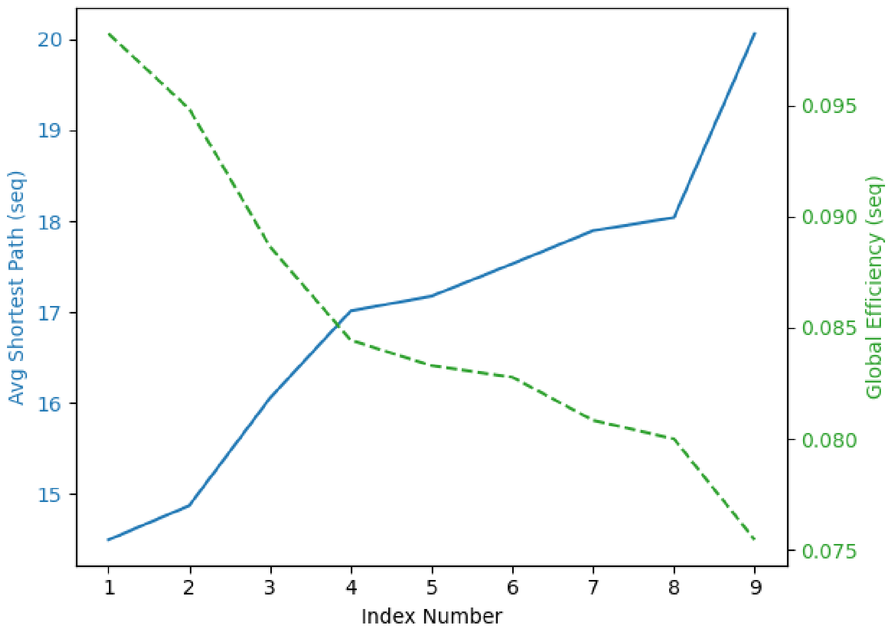


Figure 5: Sequential Node Removal Degree Centrality Plot



In the above figures, the ASPL increases, indicating that the overall connectivity of the network is worsening, making travel between stations less direct. Concurrently, GE decreases, suggesting the network's decreasing capacity for effectively facilitating travel or information flow. The more pronounced effects in the sequential removal plot (Figure 5), suggest that the network's integrity is more significantly compromised when nodes are removed in a sequence rather than randomly. This implies that the sequential removal of high-degree nodes quickly deteriorates the network's efficiency and connectivity. GE stands out as it directly relates to the network's operational capacity, making it a clear and immediate indicator of network performance post-node removal.

1.4.2 Betweenness Centrality

Figure 6: Non-Sequential Node Removal Betweenness Centrality Plot

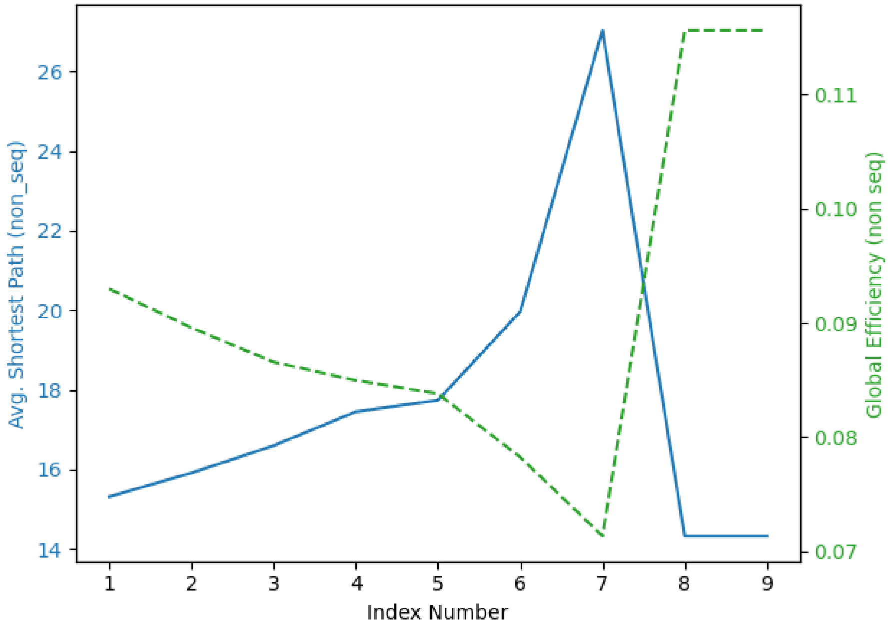
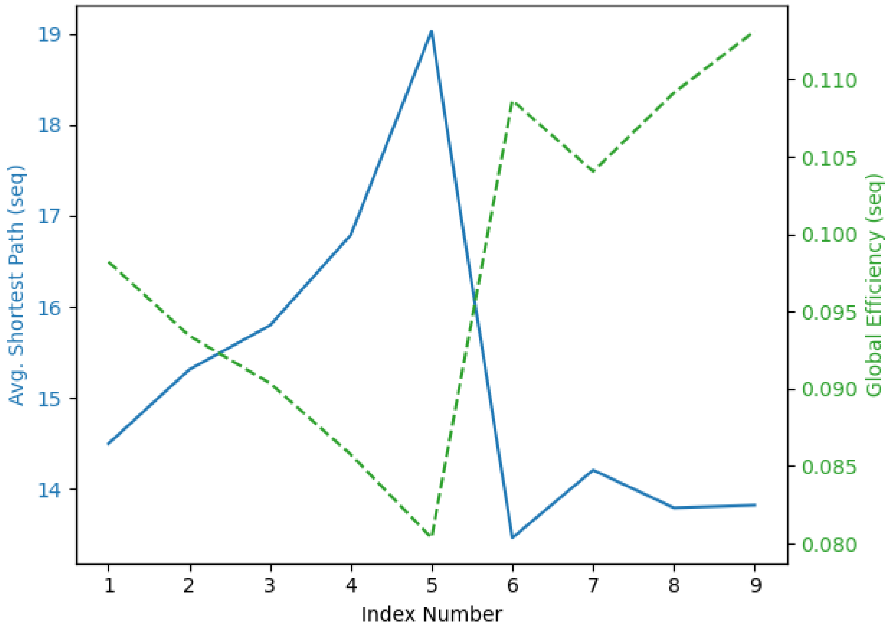


Figure 7: Sequential Node Removal Betweenness Centrality Plot



In Figure 6, ASPL experiences a sharp increase at specific points, suggesting that removing specific nodes greatly distorts connections in the network. The GE shows a similar pattern, experiencing significant drops at specific removals, highlighting a decrease in overall network efficiency. Figure 7 shows that sequential removal impacts the network negatively, evident in the consistent rise in ASPL and a drop in GE. The variations in GE imply that certain nodes are more vital than others in ensuring network efficiency.

1.4.3 Closeness Centrality

Figure 8: Non-Sequential Node Removal Closeness Centrality Plot

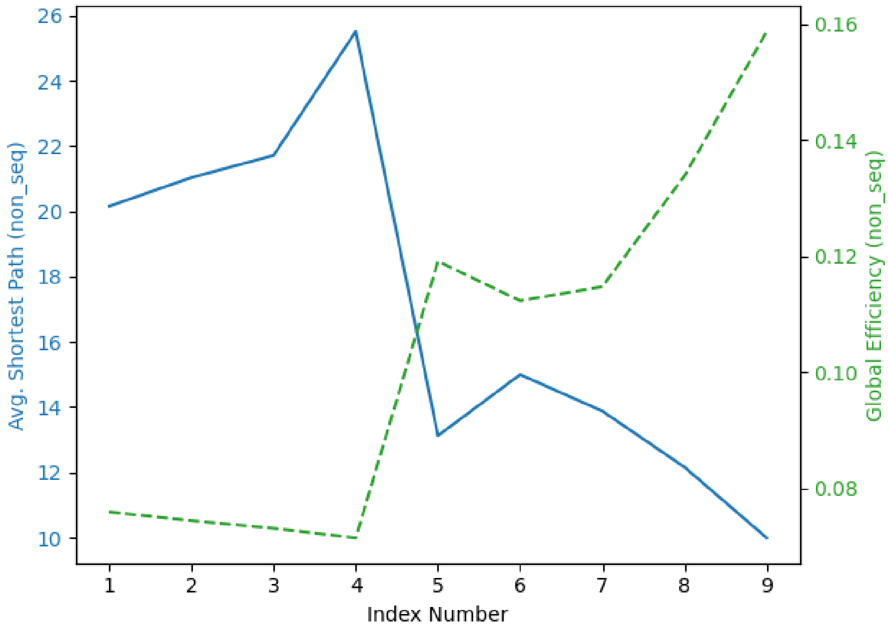
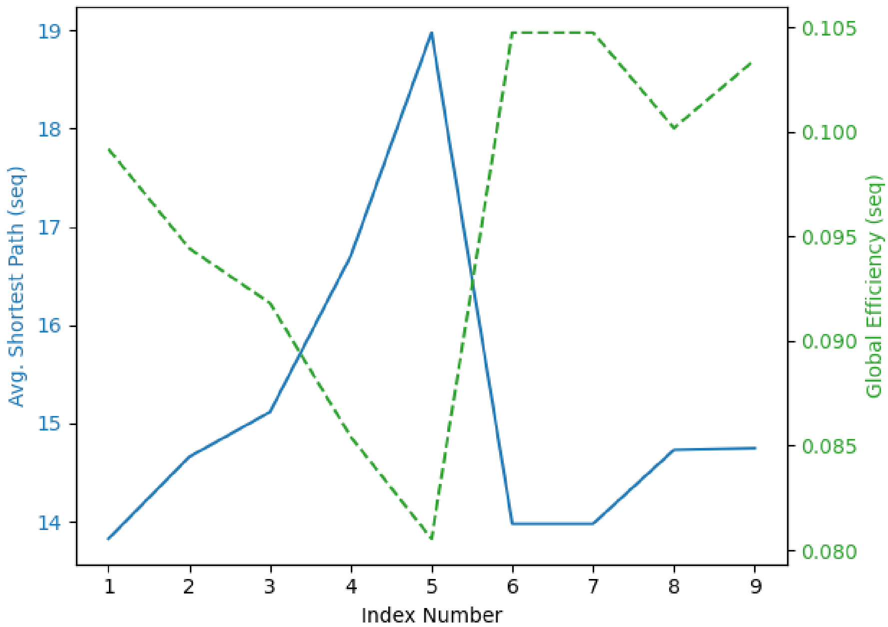


Figure 9: Sequential Node Removal Closeness Centrality Plot



Figures 8 and 9, show the network effect after nodes are removed based on Closeness Centrality. The plots indicate that removing items in non-sequential order led to decreased efficiency and longer connectivity while removing them in a sequential order results in significant variations. After removal, GE proves to be a more significant global measure of network resilience.

1.5 Flows: Weighted Network

1.5.1 Centrality Measure with Weight: New vs Old

Table 10: New Betweenness Centrality Including Flows

Serial No.	Stations	Betweenness Flows
1	Green Park	0.5772
2	Bank and Monument	0.4552
3	Waterloo	0.3749
4	Westminster	0.3518
5	Liverpool Street	0.3439
6	Stratford	0.3281
7	Euston	0.2862
8	Victoria	0.2762
9	Oxford Circus	0.2744
10	Warren Street	0.2551

Table 11: New Betweeness Centrality Including Flows and Edge Length

Serial No.	Stations	Betweenness Length
1	Bank and Monument	0.2214
2	King's Cross St. Pancras	0.2118
3	Stratford	0.1839
4	Baker Street	0.1653
5	Oxford Circus	0.1567
6	Euston	0.1562
7	Earl's Court	0.1423
8	Shadwell	0.1400
9	South Kensington	0.1287
10	Gloucester Road	0.1268

Table 12: Old Betweenness Centrality Measure

Serial No.	Stations	Betweenness Centrality
1	Stratford	23768.09
2	Bank and Monument	23181.06
3	Liverpool Street	21610.39
4	King's Cross St. Pancras	20373.52
5	Waterloo	19464.88
6	Green Park	17223.62
7	Euston	16624.28
8	Westminster	16226.16
9	Baker Street	15287.11
10	Finchley Road	13173.76

Table 10 Green Park emerges as the highest centrality while Table 11 accounts for the actual physical distances with Bank and Monument having the highest centrality. When compared to the old (Table 12), it is noticed that the rankings of stations differ significantly. For example, Stratford which was a central station initially, has now moved down in both the new centrality rankings. This shift indicates that when flow or distance is considered, the relative importance of stations changes, suggesting that the raw number of connections may not fully capture the stations' roles in the network.

1.5.2 Impact measure with flows

Using ASPL and GE for a weighted network requires adjustment to reflect the true impact of node removal, especially in the context of passenger flow in a transportation system. ASPL could be adjusted to compute the shortest paths based on the weights (distances) rather than merely counting the steps between nodes. For GE, the weights would adjust the shortest paths to prioritize those that are more efficient in terms of passenger travel time (distance) and service frequency, rather than those with fewer stops.

Weighted Global Efficiency (WGE) can be used to measure the efficiency of the network by averaging the inverse of the shortest weighted paths between all pairs of nodes.

1.5.3 Experiment with flows

1.5.4 Sequential Removal of Betweenness Centrality for a Weighted Network

Table 13: Betweenness Centrality for a Weighted Network

Serial	Station Removed	No.Components	ASPL	WGE	Component Size
1	Green Park	1	13.85	0.0992	397
2	King's Cross St. Pancras	1	14.69	0.0944	396
3	Bank and Monument	1	16.01	0.0879	395
4	Canada Water	1	18.49	0.0815	394
5	West Hampstead	2	13.56	0.1077	224
6	Stratford	4	13.56	0.1077	224
7	SEarl's Court	4	14.29	0.1032	223
8	Shepherd's Bush	5	14.02	0.1074	194
9	Oxford Circus	5	15.18	0.1006	193
10	Queen's Park	5	18.79	0.0916	192

1.5.5 Sequential Removal of Betweenness Centrality for Weighted Network including Flows

Table 14: Betweenness Centrality: Weighted Network Including Flows

Serial	Station Removed	No.Components	ASPL	WGE	Component Size
1	Bank and Monument	1	18.79	0.0968	397
2	King's Cross St. Pancras	1	18.79	0.0899	396
3	Canada Water	1	18.79	0.0831	395
4	West Hampstead	2	18.79	0.1114	225
5	Stratford	4	18.78	0.1114	225
6	Earl's Court	4	18.79	0.1068	224
7	Shepherd's Bush	5	18.79	0.1126	195
8	Euston	6	18.79	0.1171	172
9	Baker Street	7	18.79	0.1007	169
10	Acton Town	9	18.79	0.1205	147

Removing Bank and Monument, then King's Cross St. Pancras, and Canada Water, each station caused a slight decrease in the size of the largest connected component in the network, it became slightly smaller after each removal. Surprisingly, the ASPL remained constant within the main component following each removal. This might imply that eliminating these high-betweenness stations has no much impact on the path lengths between the remaining stations in the main component of the network. Nevertheless, the network's overall effectiveness declined as each station was removed. The significance of these stations is highlighted by the decline in WGE upon their removal. Therefore, London Tube's overall efficiency relies heavily on these stations, which are essential because of their locations and the volume of passengers they serve. Therefore, closing down Bank and Monument would have the biggest effect on commuters in the London Tube system, as shown by the drop in overall efficiency after its closure. The station's high score in "betweenness centrality length" indicates its crucial role in aiding passenger movement throughout the network.

1.6 Spatial Interaction models

1.6.1 Models and calibration

Spatial interaction refers to the phenomenon where entities located at different points within physical space engage in various forms of engagement (Roy and Thill, 2004) while Spatial interaction modelling entails the examination of movements from an origin to a destination, either across tangible space such as migration or within conceptual space (Oshan, 2016). Wilson (1971), develops a 'family' of four spatial interaction models to provide more information and flexibility.

1.6.2 Unconstrained Model

This model indicates that the flow between the origin and destinations is proportional to the product of the start and end point's masses, and inversely with the distance separating them. In simple terms, the flow or interaction between the origin and destination depends on the size of both origin and destination. However, the distance between them plays a role; the farther they are, the less likely they are to interact frequently.

$$T_{ij} = k\frac{O_i^{\alpha} D_j^{\gamma}}{d_{ij}^{\beta}} \tag{6}$$

According to Wilson (1971), the Gravity Model can also be expressed in a more recognizable or familiar form:

$$T_{ij} = kO_i^{\alpha} D_j^{\gamma} d_{ij}^{-\beta} \tag{7}$$

Where: T_{ij} symbolizes the flows, T, from the origin as i to the destination as j. O is origin characteristics relating to how well an origin point can distribute flows to other locations. D represents destination characteristics that evaluate the appeal of each destination (j). d represents a matrix that includes various costs related to the connections between i and j, such as transportation fees, travel duration, and distance. k, α, γ, and β are the parameters of the model. α consists of parameters that depict how origin attributes impact flows. γ is a set of parameters that show how destination attributes influence flows. β is a parameter that represents the impact of cost (transportation expenses) on flows between locations. k is a scaling factor that maintains consistency between the total observed and predicted flows.

$$k = \frac{T}{\sum_i \sum_j O_i^{\alpha} D_j^{\gamma} d_{ij}^{-\beta}} \tag{8}$$

and T is the sum of the matrix of observed flows:

$$T = \sum_i \sum_j T_{ij} \tag{9}$$

1.6.3 Production-constrained Model

$$T_{ij} = A_i O_i D_j^{\gamma} d_{ij}^{-\beta} \tag{10}$$

Where

$$O_i = \sum_j T_{ij} \tag{11}$$

and

$$A_i = \frac{1}{\sum_j D_j^{\gamma} d_{ij}^{-\beta}} \tag{12}$$

In the production-constrained model, O_i is fixed and does not change because it is a known constraint for each origin i. A_i is a "balancing factor," acting like k in the unconstrained model. In this case, A_i ensures that the predicted flows from each origin add up to the specific known total O_i instead of only the overall total. In other words, it ensures the total outflows are preserved in the predicted flows.

1.6.4 Attraction-Constrained Model

$$T_ij = D_j B_j O_i^{\alpha} d_{ij}^{-\beta} \tag{13}$$

Where

$$D_j = \sum_i T_{ij} \tag{14}$$

and

$$B_j = \frac{1}{\sum_i O_i^{\alpha} d_{ij}^{-\beta}} \tag{15}$$

Dj: vector indicating the total number of flows ending at destination j.

Bj: vector of destination balancing factors that ensure the preservation of the total in-flows in the predicted flows.

1.6.5 Doubly Constrained Model

$$T_{ij} = A_i B_j O_i D_j d_{ij}^{-\beta} \tag{16}$$

Where

$$O_i = \sum_j T_{ij} \tag{17}$$

$$D_j = \sum_i T_{ij} \tag{18}$$

and

$$A_i = \frac{1}{\sum_j B_j D_j d_{ij}^{-\beta}} \tag{19}$$

$$B_j = \frac{1}{\sum_i A_i O_i d_{ij}^{-\beta}} \tag{20}$$

the calculation of A_i relies on knowing B_j and the calculation of B_j relies on knowing A_i. $f(dij)$ is a function of cost or distance or travel time called the distance-decay function. This is often an exponential or power function (Oshan, 2016),

Power

$$f(d_{ij}) = d_{ij}^{\beta} \tag{21}$$

Exponential

$$f(d_{ij}) = exp(\beta d_{ij}) \tag{22}$$

1.7 Model Selection and calibration

The unconstrained model is used due to its simplicity and flexibility. The rationale behind my choice of method is that I wish to capture the overall impact of population and jobs on the interaction between different locations, without necessarily implementing any form of specific limitations like policy restrictions and the unconstrained gravity model best fits this.

Before calibration, the $r^2 = 0.0346$ and Root Mean Squared Error (RMSE) = 485.364 which means the model outcome is poor. So we calibrated the model to improve its predictive performance. Total flow data = 1542283.

Taking logarithms

$$\ln T_{ij} = K + \alpha \ln O_i + \gamma \ln D_j - \beta \ln d_{ij} \tag{23}$$

where $K = \ln k$ transforming it into a linear regression model.

1.7.1 After Calibration

Table 15: Generalized Linear Model Regression Results

Variables	Coefficients	Std Error	Z-score	P-Values
Intercept	-3.0646	0.014	-223.280	0.000
log-population	0.6973	0.001	987.428	0.000
log-jobs	0.7037	0.001	1054.578	0.000
log-distance	-0.5961	0.001	-647.450	0.000

In Table 15, it can be observed that the Poisson regression model has calibrated all four parameters.

K (intercept) = -3.0646. $K = \ln k$, so k can never be negative!

$\alpha = 0.6973$

$\gamma = 0.7037$
and $\beta = 0.5961$ (in equation (7) there is a negative sign in front of β). It can also be seen from the other outputs that all variables are highly significant (P-values less than 0.01), with the z-scores revealing that the job has the most influence on the model.

After calibration, the r^2 has increased from 0.0346 to 0.3114, and Root Mean Squared Error (RMSE) dropped from 485.364 to 128.176, indicating improvement in the model outcome.

1.7.2 Sectional Comparison of Actual and Predicted Flows

Figure 10: Unconstrained Predicted Flows

station_destination	Abbey Road	Acton Central	Acton Town	Aldgate	Aldgate East	All Saints	Alperton	Amersham	Anerley
station_origin									
Abbey Road	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Acton Central	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Acton Town	0.000000	0.000000	0.000000	23.866634	23.985767	0.000000	9.473315	0.000000	0.000000
Aldgate	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Aldgate East	0.000000	0.000000	8.669822	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
...	...	...	...	...	...	...	...	...	...
Woodford	0.000000	0.000000	8.139883	29.880398	32.055753	0.000000	0.000000	0.000000	0.000000
Woodgrange Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Woodside Park	0.000000	0.000000	6.638777	20.594905	20.700542	0.000000	0.000000	0.000000	0.000000
Woolwich Arsenal	6.432026	0.000000	0.000000	0.000000	0.000000	7.313614	0.000000	0.000000	0.000000
All	210.033437	237.898658	2670.625134	9988.054975	10818.337330	295.023947	615.356713	80.391197	47.318523

Figure 11: Actual Flows

station_destination	Abbey Road	Acton Central	Acton Town	Aldgate	Aldgate East	All Saints	Alperton	Amersham	Anerley	Angel	...	Wimbledon	Wim
station_origin													
Abbey Road	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
Acton Central	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
Acton Town	0.0	0.0	0.0	3.0	17.0	0.0	35.0	0.0	0.0	11.0	...	77.0	
Aldgate	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	17.0	...	0.0	
Aldgate East	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	20.0	...	24.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	
Woodford	0.0	0.0	2.0	5.0	47.0	0.0	0.0	0.0	0.0	22.0	...	2.0	
Woodgrange Park	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
Woodside Park	0.0	0.0	1.0	26.0	11.0	0.0	0.0	0.0	0.0	59.0	...	0.0	
Woolwich Arsenal	20.0	0.0	0.0	0.0	0.0	7.0	0.0	0.0	0.0	0.0	...	0.0	
All	345.0	750.0	2202.0	7782.0	7932.0	444.0	741.0	256.0	173.0	8103.0	...	6295.0	

Noticeably, the model predicted about 60 percent of the total at the destination (Abbey Road) and over-predicted at Aldgate (destination).

1.8 Scenarios

1.8.1 Scenario A

With a fifty percent job reduction in Canary Wharf, commuters will move to other destinations with high jobs, eventually changing the flow pattern. And since commuters must be conserved at the origin (known constrained), the Origin Constrained Model best fits this scenario. The model ensures that the predicted flows from each origin add up to the specific known total.

1.8.2 Origin Constrained Model Calibration

Table 16: Linear Model Regression Results

Station-Origin	Coefficients	Std Error	Z-score	P-Values
Abbey Road	3.2580	0.042	77.238	0.000
Acton Central	4.9485	0.031	160.277	0.000
Acton Town	4.3416	0.020	213.123	0.000
Aldgate	3.3038	0.022	152.570	0.000
Aldgate East	3.3538	0.021	159.844	0.000
...	...	...	...	...
Woodgrange Park	5.2652	0.045	117.496	0.000
Woodside Park	4.4410	0.022	202.179	0.000
Woolwich Arsenal	6.7087	0.017	406.253	0.000
log-jobs	0.7359	0.001	1141.131	0.000
log-distance	-0.8139	0.001	-704.756	0.000

The value of the γ parameter for job destination attractiveness in Table 16 is 0.7359. The decay parameter for distance, denoted as β, equals 0.8139 (remember to include the negative sign in the equation).

The model's outputs show that all explanatory variables are statistically significant (P-values less than 0.01) and the z-scores reveal that jobs have the biggest impact on the model, closely followed by distance. Next, several parameters consist of the set of α_i values linked to our initial constraints to guarantee that the projected flows from each origin sum up to the given total amount.

The model's fit quality has significantly increased: from 0.3114 in the previous model to 0.3805 in the current one. There has been a significant decrease in the RMSE, from 128.178 to 121.441.

1.8.3 Sectional Comparison of Actual and Predicted Flows

Figure 12: Origin Constrained Predicted Flows

station_destination	Abbey Road	Acton Central	Acton Town	Aldgate	Aldgate East	All Saints	Alperton	Amersham	Anerley
station_origin									
Abbey Road	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Acton Central	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Acton Town	0.000000	0.000000	0.000000	20.270944	20.321786	0.000000	9.741848	0.000000	0.000000
Aldgate	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Aldgate East	0.000000	0.000000	2.947131	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
...	...	...	...	...	...	...	...	...	...
Woodford	0.000000	0.000000	8.482367	37.697902	41.316341	0.000000	0.000000	0.000000	0.000000
Woodgrange Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Woodside Park	0.000000	0.000000	6.218865	21.963962	22.023169	0.000000	0.000000	0.000000	0.000000
Woolwich Arsenal	35.772274	0.000000	0.000000	0.000000	0.000000	40.278395	0.000000	0.000000	0.000000
All	415.199870	360.677263	2282.048678	8907.895665	9701.516864	554.558442	541.677372	65.720088	151.282656

Figure 13: Actual Flows

station_destination	Abbey Road	Acton Central	Acton Town	Aldgate	Aldgate East	All Saints	Alperton	Amersham	Anerley	Angel	...	Wimbledon	Wim
station_origin													
Abbey Road	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
Acton Central	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
Acton Town	0.0	0.0	0.0	3.0	17.0	0.0	35.0	0.0	0.0	11.0	...	77.0	
Aldgate	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	17.0	...	0.0	
Aldgate East	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	20.0	...	24.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	
Woodford	0.0	0.0	2.0	5.0	47.0	0.0	0.0	0.0	0.0	22.0	...	2.0	
Woodgrange Park	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
Woodside Park	0.0	0.0	1.0	26.0	11.0	0.0	0.0	0.0	0.0	59.0	...	0.0	
Woolwich Arsenal	20.0	0.0	0.0	0.0	0.0	7.0	0.0	0.0	0.0	0.0	...	0.0	
All	345.0	750.0	2202.0	7782.0	7932.0	444.0	741.0	256.0	173.0	8103.0	...	6295.0	

The over-predicted total at the destination by 20 percent at Abbey Road and under-predicted at Acton Central (destination). Perhaps this is the cause of the change in the pattern of commuters to other destinations due to the reduction in Canary Wharf jobs.

1.9 Scenario B

Using the original flow data without a decrease in jobs of Canary Wharf to analyze scenario B. The Production Constrained Model is utilized.

Table 17: Linear Model Regression Results

Station-Origin	Coefficients	Std Error	Z-score	P-Values
Abbey Road	3.2704	0.042	77.519	0.000
Acton Central	5.0089	0.031	162.363	0.000
Acton Town	4.3974	0.020	216.218	0.000
Aldgate	3.3611	0.022	155.436	0.000
Aldgate East	3.4087	0.021	162.703	0.000
...	...	...	...	...
Woodgrange Park	5.3202	0.045	118.758	0.000
Woodside Park	4.4967	0.022	205.016	0.000
Woolwich Arsenal	6.7019	0.017	406.033	0.000
log-jobs	0.7302	0.001	1151.002	0.000
log-distance	-0.8152	0.001	-706.669	0.000

In Table 17, the γ parameter related to the destination attractiveness (jobs) is 0.7302. The β distance decay parameter is 0.8152 (not forgetting the negative sign in the equation) without Canary Wharf jobs reduction.

The model's goodness of fit has $r^2 = 0.3937$ and RMSE = 120.146, indicating a better model based on these matrices.

1.9.1 Significant Increase in Cost of Transport

In the cost function, when $\beta = 0$, it means cost will not affect the flows. If β is large, cost will have less impact hence commuters would love to move to these areas thereby increasing the flows between the origin and destination. However, if β is small, cost will be high or increase which will cause people not move to these places thereby decreasing the the flows between that particular origin and destination. Therefore, the type of β to adopt in the cost function largely depends on what effect you want the cost function to have on the flows.

To reflect the significant increase in the cost of transport (distance), the selected βs have to be below the actual. The original or actual beta = -0.8152. Therefore, β = -0.65 and β = -0.75 will be adopted (not forgetting the negative in the formula).

1.9.2 When beta is 0.72

Figure 14: Predicted Flows

station_destination	Abbey Road	Acton Central	Acton Town	Aldgate	Aldgate East	All Saints	Alperton	Amersham	Anerley
station_origin									
Abbey Road	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Acton Central	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Acton Town	0.000000	0.000000	0.000000	50.819389	51.009187	0.000000	22.062169	0.000000	0.000000
Aldgate	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Aldgate East	0.000000	0.000000	7.445412	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
...	...	...	...	...	...	...	...	...	...
Woodford	0.000000	0.000000	22.587860	93.396702	101.437608	0.000000	0.000000	0.000000	0.000000
Woodgrange Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Woodside Park	0.000000	0.000000	16.353526	55.180444	55.395693	0.000000	0.000000	0.000000	0.000000
Woolwich Arsenal	81.002563	0.000000	0.000000	0.000000	0.000000	91.781321	0.000000	0.000000	0.000000
All	841.457636	857.520131	5530.150327	20226.513726	22119.972347	1122.264145	1346.413437	173.465132	343.202456

The model predicted a total flow of 3,576,447. The model's fitness has $r^2 = 0.4118$ which has improved, however, the RMSE increased to 166.454 compared to scenario A.

1.9.3 When beta is 0.78

Figure 15: Predicted Flows

station_destination	Abbey Road	Acton Central	Acton Town	Aldgate	Aldgate East	All Saints	Alperton	Amersham	Anerley
station_origin									
Abbey Road	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000
Acton Central	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000
Acton Town	0.000000	0.000000	0.000000	28.332729	28.414352	0.000000	13.238861	0.000000	0.00000
Aldgate	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000
Aldgate East	0.000000	0.000000	4.147421	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000
...	...	...	...	...	...	...	...	...	...
Woodford	0.000000	0.000000	12.132579	52.290146	57.117838	0.000000	0.000000	0.000000	0.00000
Woodgrange Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000
Woodside Park	0.000000	0.000000	8.883873	30.720586	30.814611	0.000000	0.000000	0.000000	0.00000
Woolwich Arsenal	46.835128	0.000000	0.000000	0.000000	0.000000	52.805829	0.000000	0.000000	0.00000
All	534.258397	498.952378	3161.069646	11997.594941	13082.927123	693.457919	760.024520	94.689227	206.97635

The model predicted a total flow of 2,102,813. The model's fitness has $r^2 = 0.4016$ which has improved, however, the RMSE increased to 123.517 Compared to Scenario A.

Table 18: First Ten Production Constrained Estimates

Serial	prodsimest1	prodsimest2	prodsimest3
1	1.4069	3.3293	1.9344
2	3.6774	8.0541	4.9136
3	76.0629	171.3870	102.7052
4	56.1239	116.9081	73.6138
5	3.5684	8.2525	4.8649
6	3.8417	9.0903	5.2820
7	2.8138	6.4928	3.8329
8	3.3261	7.3176	4.4516
9	7.3593	15.9513	9.7956
10	2.1139	4.9553	2.8964

In Table 18, prodsimest1 is the actual estimate without jobs reduction. Prodsimest2 is when $\beta = 0.72$, and prodsimest3 is when $\beta = 0.78$ reflects the cost of transport increment.

1.10 Discussion of Scenario Flows

Comparing the scenarios, Scenario B without the job reduction and with the original β value represents a network closer to current conditions and thus might be expected to have a lower impact on the redistribution of flows. However, altering β significantly changes the predicted flows: lower β (0.72) value increase predicted flows and higher β (0.78) value relatively decrease the total flow. The scenario that would have the most impact on the redistribution of flows is not solely determined by the parameter β, but also by external factors such as job availability at Canary Wharf. Therefore, the job reduction in Scenario A had more significant impact by altering commuting patterns and redistributing of flows across the network which is noticeable in Figure 12. The changes in predicted flows due to parameter adjustments emphasize the sensitivity of the spatial interaction models to the cost function parameter, reflecting the importance of accurately calibrating models to predict commuter behaviour in response to network and economic changes.

References

[1] Freeman, L.C., (2002). Centrality in social networks: Conceptual clarification. Social network: critical concepts in sociology. Londres: Routledge, 1, pp.238-263.

[2] Freeman, L.C., (1977). A set of measures of centrality based on betweenness. Sociometry, pp.35-41.

[3] Mouronte-López, M.L., (2021). Modeling the public transport networks: A study of their efficiency. Complexity, 2021, pp.1-19.

[4] Mao, G. and Zhang, N., (2013). Analysis of average shortest-path length of scale-free network. Journal of Applied Mathematics, 2013, pp.1-5.

[5] Ek, B., VerSchneider, C. and Narayan, D.A., (2015). Global efficiency of graphs. AKCE International Journal of Graphs and Combinatorics, 12(1), pp.1-13.

[6] Rodrigues, F.A., (2019). Network centrality: an introduction. A mathematical modeling approach from nonlinear dynamics to complex systems, pp.177-196.

[7] Roy, J.R. and Thill, J.C., (2004). Spatial interaction modelling. Papers in regional science, 83(1), pp.339-361.

[8] Oshan, T.M., (2016). A primer for working with the Spatial Interaction modeling (SpInt) module in the python spatial analysis library (PySAL). Region, 3(2), pp.R11-R23.

[9] Wilson, A.G., (1971). A Family of Spatial Interaction Models, and Associated Developments. Environment and Planning A: Economy and Space. SAGE Publications Ltd, 3(1), pp.1-32.