The Resilience of London's Underground

May.2023

Words Count	2989

Part 1: London's Underground Resilience

I. Topological Network

An undirected topological network of London underground is created.

I.1 Centrality Measures

a. Degree Centrality

Degree centrality is a simple centrality measure based on the number of links attached to it (Equation 1). Nodes with the higher degree centrality value mean that they have more connections. Sometimes, degree centrality will be converted to a 0-1 scale, also known as normalised version C_j^d (Equation 2).

$$k_i = \sum_{j=1}^{n} A_{ij} \quad (1)$$

$$C_j^d = \frac{k_i}{N-1} \quad (2)$$

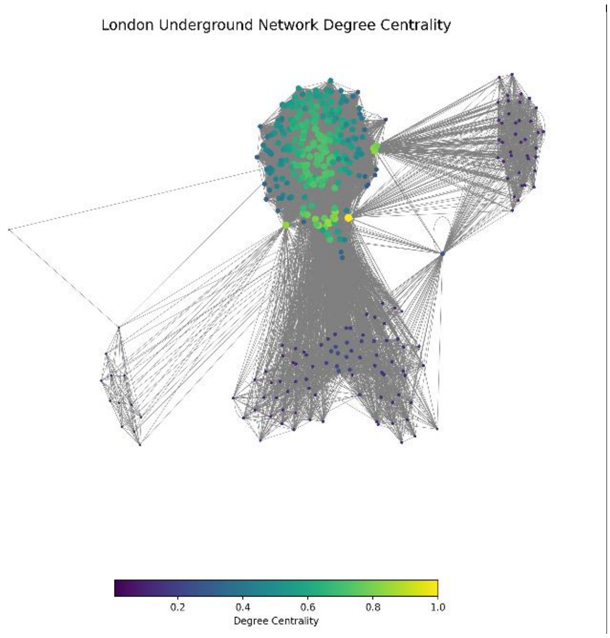
Where: k_i is the original degree of nodes, N is the total number of nodes.

(k_i and N/n have the same notation in the equation of full text, so their annotations will be omitted in the subsequence.)

In underground network, stations with higher degree centrality indicates that they are likely to be the interchange station for different lines, which plays an important role in getting passengers to their destinations efficiently.

Station Name	Degree	Degree Centrality
Stratford	367	0.922111
Highbury & Islington	321	0.806533
Whitechapel	311	0.781407
West Brompton	309	0.776382
Canada Water	307	0.771357
Canary Wharf	307	0.771357
Liverpool Street	306	0.768844
Bank and Monument	305	0.766332
Richmond	305	0.766332
Canning Town	304	0.763819

1 (a) The First 10 Ranked Nodes



1 (b) Degree Centrality Plot

Figure 1 Degree Centrality of Topological Network

b. Closeness Centrality

Closeness centrality (C_i) is based on the average geodesic distance (l_i) from one node to all other nodes in the network, in which the geodesic distance is the minimum number of edges or shortest path between two nodes. Nodes with high closeness centrality mean better and more efficient connectivity to all other nodes in the network.

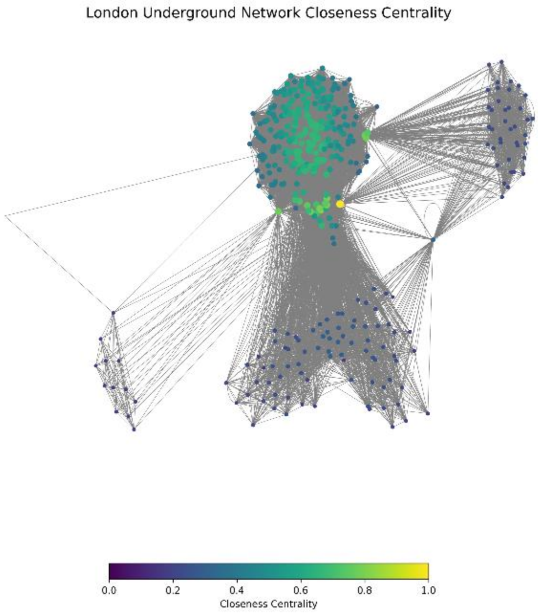
$$l_i = \frac{1}{n}\sum_j d_{ij} \quad (3)$$

$$C_i = \frac{1}{l_i} = \frac{n}{\sum_j d_{ij}} \quad (4)$$

Where: d_{ij} is the geodesic distance between i and j.

In underground network, stations with higher closeness centrality indicate the higher accessibility and more alternative routes to other stations, which makes them more resilient to disruptions and crucial for network.

Station Name	Closeness Centrality
Stratford	0.927739
Highbury & Islington	0.836134
Whitechapel	0.820619
West Brompton	0.817248
Canada Water	0.813906
Richmond	0.810591
Canary Wharf	0.810591
Bank and Monument	0.810591
Liverpool Street	0.808943
Canning Town	0.808943



2 (a) The First 10 Ranked Nodes **2 (b) Closeness Centrality Plot**

Figure 2 Closeness Centrality of Topological Network

c. Betweenness Centrality

Betweenness centrality (x_i) is based on whether the node is on the geodesic path from other nodes pairs in the network, which using the total number as the quantitative criteria. Nodes with high betweenness centrality indicate that they are important broker and bridge for the flow between other nodes, which are crucial for maintaining network's connectivity.

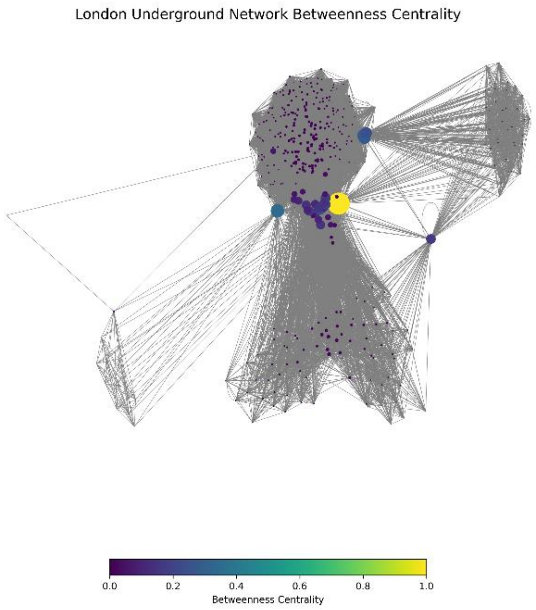
$$n_{st}^{i} = \begin{cases} 1, if\ node\ i\ lies\ on\ the\ geodesic\ path\ from\ s\ to\ t. \\ 0, otherwise. \end{cases} \quad (5)$$

$$x_i = \begin{cases} \sum_{st} n_{st}^{i}, single\ geodesics\ from\ s\ to\ t \\ \sum_{st} \frac{n_{st}^{i}}{g_{st}}, multiple\ geodesics\ from\ s\ to\ t \end{cases} \quad (6)$$

Where: g_{st} is the total number of geodesics from s to t.

In underground network, stations with high betweenness centrality always facilitate passenger movement between various parts of the network and a serious impact may occur if they are out of service or disrupted. Therefore, identifying and maintaining these stations is crucial for network system.

Station Name	Betweenness Centrality
Stratford	7785.967073
Liverpool Street	2710.37716
Canary Wharf	2208.627931
Bank and Monument	2208.627931
Canning Town	2192.848026
West Ham	1939.641831
Highbury & Islington	1818.904025
Whitechapel	1554.965946
Canada Water	1413.977663
Shadwell	1348.588902



3 (a) The First 10 Ranked Nodes **3 (b) Betweenness Centrality Plot**

Figure 3 Betweenness Centrality of Topological Network

I.2 Impact Measures

Average clustering coefficient and global efficiency are used here to reflect the resilience and robustness of whole network, in which the former focuses on tightness and the latter quantifies the efficiency of information or resource flow. Both of them are general measure which not specific to London underground but can applicable to social network (Hansen et al., 2020), biological network (Sahoo et al., 2016) and other transportation networks (Zhang et al., 2021). Besides, they are suitable for the measure of both connected and disconnected networks.

a. Average Clustering Coefficient

Clustering coefficient reflects how well the neighbours of a given node associated to each other. There are two computed methods for global clustering coefficient, one takes the average of local clustering coefficient C_i (Average clustering coefficient C) (Watts and Strogatz, 1998)and another considers the transitivity (Wasserman and Faust, 1994). Here, the first method is selected.

$$C_i = \frac{2L_i}{k_i(k_i-1)} \quad (7)$$

$$C = \frac{1}{N}\sum_{i=1}^{n} C_i \quad (8)$$

Where: L_i is the number of links between the k_i neighbours of node i, C in [0,1].

b. Global efficiency

Global efficiency was first introduced by Latora and Marchiori (2001), which the underlying idea is that more far away two nodes are, the lower the communication efficiency (Ek et al., 2015). This metrics can be applied to both connected and disconnected networks, in which d_{ij} is infinitely large and $\epsilon_{ij} = 0$ when there is no path from i to j.

$$E_{glob}(G) = \frac{\sum_{i \neq j \in G} \epsilon_{ij}}{N(N-1)} = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{d_{ij}} \quad (9)$$

Where: d_{ij} is the smallest number of links from i to j.

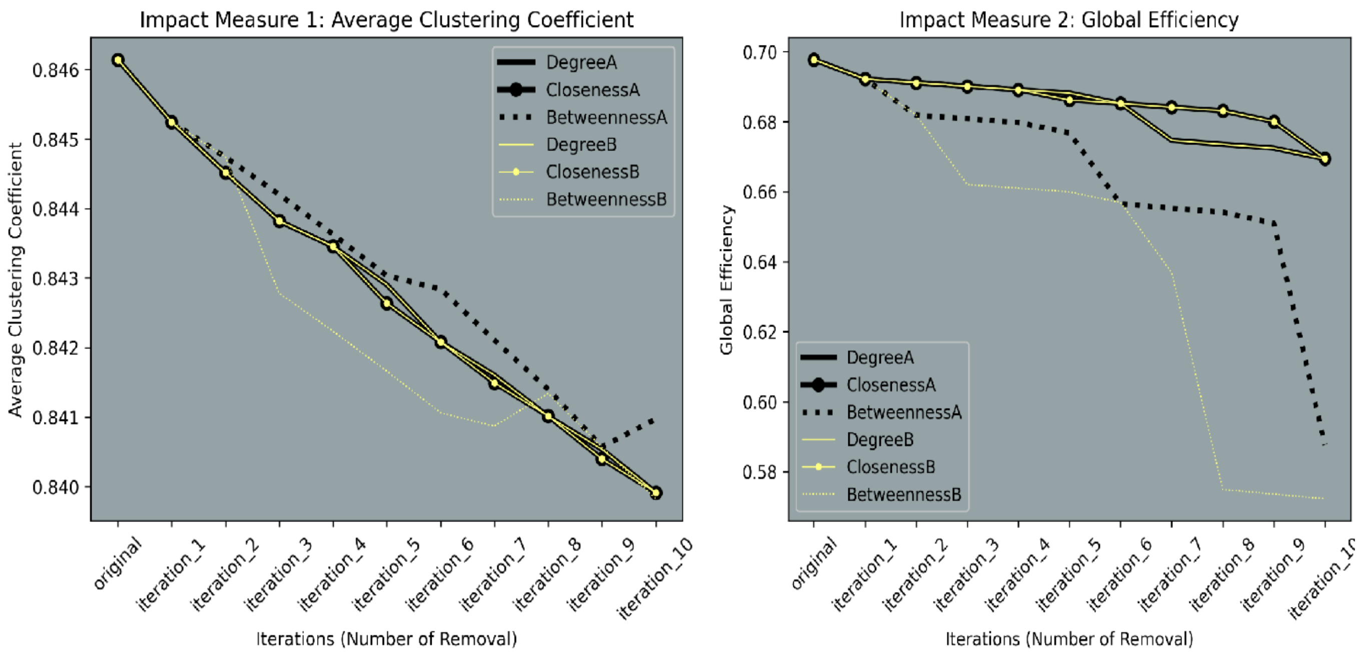
I.3 Node Removal

To explore the resilience and robustness of London underground network, two node-attacking strategies are used to simulate the disruption. For non-sequential strategy (A), removal is performed following the sequence of three centralities calculated in I.1, and the network will be measured by two measures proposed in I.2 after each removal. Sequential strategy (B) is similar to strategy A, but it recalculates the centrality after each removal and takes the highest value for the next iteration. For better comparison, the number of removed nodes for both strategies are 10. It can be seen from Table 1 that the removed nodes of degree centrality and closeness centrality are almost the same in both strategies, while betweenness centrality is more different.

Table 1 Removed Nodes in I.3

iteration	Strategy A: Non-Sequential						Strategy B: Sequential					
	Degree Centrality		Closeness Centrality		Betweenness Centrality		Degree Centrality		Closeness Centrality		Betweenness Centrality	
1	Stratford	0.922111	Stratford	0.927739	Stratford	7785.967073	Stratford	0.922111	Stratford	0.927739	Stratford	7785.967073
2	Highbury & Islington	0.806533	Highbury & Islington	0.836134	Liverpool Street	2710.37716	Highbury & Islington	0.806045	Highbury & Islington	0.835789	Liverpool Street	5535.077442
3	Whitechapel	0.781407	Whitechapel	0.820619	Canary Wharf	2208.627931	Whitechapel	0.780303	Whitechapel	0.819876	Upminster	4985.163205
4	West Brompton	0.776382	West Brompton	0.817248	Bank and Monument	2208.627931	West Brompton	0.774684	West Brompton	0.816116	Bank and Monument	2963.750509
5	Canada Water	0.771357	Canada Water	0.813906	Canning Town	2192.848026	Canary Wharf	0.769036	Canada Water	0.812371	Canary Wharf	3956.423715
6	Canary Wharf	0.771357	Richmond	0.810591	West Ham	1939.641831	Canada Water	0.768448	Bank and Monument	0.808642	Canning Town	5991.716826
7	Liverpool Street	0.768844	Canary Wharf	0.810591	Highbury & Islington	1818.904025	Liverpool Street	0.765306	Canary Wharf	0.808247	West Ham	10391.83241
8	Bank and Monument	0.766332	Bank and Monument	0.810591	Whitechapel	1554.965946	Bank and Monument	0.762148	Richmond	0.807851	Shadwell	13236.8022
9	Richmond	0.766332	Liverpool Street	0.808943	Canada Water	1413.977663	Richmond	0.761538	Canning Town	0.805785	Highbury & Islington	1939.282673
10	Canning Town	0.763819	Canning Town	0.808943	Shadwell	1348.588902	Canning Town	0.758355	Liverpool Street	0.802062	Whitechapel	1842.320411

Applying two global measures to two strategies and three centrality measure, there will be twelve combinations. The twelve results are integrated into two plots, in which the black line and yellow line represent strategies A and B respectively and the different types of lines represent different centrality measures (Figure 4).



4 (a) Average Clustering Coefficient **4(b) Global Efficiency**

Figure 4 Results of Impact Measure in I.3

In general, average clustering coefficient and global efficiency decrease with the removal of key nodes, which indicates that the robustness and the resilience of the network is becoming weaker. And the sharper the decrease, the more important the nodes are for the network. Based on this conclusion, centrality measures, global measures and strategies are specifically discussed:

a. Centrality Measure

The greater the overall decline and fluctuations of the line, the more important the removed nodes are to the network's robustness, which indicates that the centrality measure is better. In both two plots, the line of closeness centrality and degree centrality are very similar and overlap in several iterations. In 4(b), two betweenness centrality lines have a significantly higher degree of decline than other four lines, and in 4(a), one of the betweenness centrality lines has a clear advantage while another has a similar performance to degree centrality and closeness centrality. Overall, betweenness centrality reflects better.

b. Strategy

The aim of removal strategies is to identify the nodes that make network vulnerable, and the faster the strategy can identify the vulnerability, the better it is. Given that degree centrality and closeness centrality have the same result in both strategies (Note that there are nodes with duplicate values of closeness centrality), betweenness centrality under strategy A (line ‘BetweennessA’) and strategy B (line ‘BetweennessB’) are used to compare. In 4(a), line ‘betweennessB’ drops sharply at iteration 3 and ‘distances’ itself from ‘betweennessA’, and in4(b), ‘betweennessB’ decreases to a greater extent after iteration2 but ‘betweennessA’ does not, suggesting that strategy B is more effective on finding out the nodes that cause network vulnerable. The overall lower values of strategy B for both global measures than strategy A also indicate that strategy B is more sensitive. Therefore, sequential strategy (B) is better.

c. Global Measure

The greater the decline rate of the measure, the more sensitive it is, indicating a stronger reflection of the network's robustness and vulnerability. The average decline rate of six lines for average clustering coefficient and global efficiency are 0.72% and 8.3% respectively. Therefore, the global efficiency measure is more suitable for this case.

II. Flows: weighted network

Passenger flows is considered as the weight to create an undirected weighted network.

II.1 Weighted Centrality Measures

The degree and degree centrality of weighted network are the same as unweighted network because the only consideration is the number of edges connected to nodes. For calculating closeness centrality and betweenness centrality, the weight of edges (flows) should be considered. It is worth noting that the weight is flow instead of distance, so two extra steps should be done:

1. Change the 0 value of flows to 0.001 (It could be any very small positive number).

2. Invert flow value.

The result shows that the ranking of closeness centrality and betweenness centrality have obvious differences from topological network.

Table 2 Comparison of Centrality Measures for Topological and Weighted Networks

Old: Topological Network							**New:** Weighted Network						
Degree Centrality			**Closeness Centrality**		**Betweenness Centrality**		**Degree Centrality**			**Closeness Centrality**		**Betweenness Centrality**	
Stratford	367	0.922111	Stratford	0.927739	Stratford	7785.967073	Stratford	367	0.922111	Waterloo	0.396916	Waterloo	38082
Highbury & Islington	321	0.806533	Highbury & Islington	0.836134	Liverpool Street	2710.37716	Highbury & Islington	321	0.806533	Canary Wharf	0.396913	Bank and Monument	33506
Whitechapel	311	0.781407	Whitechapel	0.820619	Canary Wharf	2208.627931	Whitechapel	311	0.781407	Bank and Monument	0.396912	Canary Wharf	32176
West Brompton	309	0.776382	West Brompton	0.817248	Bank and Monument	2208.627931	West Brompton	309	0.776382	Stratford	0.396906	Stratford	31474
Canada Water	307	0.771357	Canada Water	0.813906	Canning Town	2192.848026	Canary Wharf	307	0.771357	Liverpool Street	0.396898	Liverpool Street	20611
Canary Wharf	307	0.771357	Richmond	0.810591	West Ham	1939.641831	Canada Water	307	0.771357	London Bridge	0.396889	Oxford Circus	12832
Liverpool Street	306	0.768844	Canary Wharf	0.810591	Highbury & Islington	1818.904025	Liverpool Street	306	0.768844	Oxford Circus	0.396877	Hammersmith	9379
Bank and Monument	305	0.766332	Bank and Monument	0.810591	Whitechapel	1554.965946	Bank and Monument	305	0.766332	Farringdon	0.396870	Farringdon	9080
Richmond	305	0.766332	Liverpool Street	0.808943	Canada Water	1413.977663	Richmond	305	0.766332	Victoria	0.396858	Victoria	8629
Canning Town	304	0.763819	Canning Town	0.808943	Shadwell	1348.588902	Canning Town	304	0.763819	King's Cross St. Pancras	0.396858	Canada Water	6851

II.2 Impact Measures with Flows

In I.2, average clustering coefficient and global efficiency are selected to measure the performance of topological network. For weighted network, clustering coefficient will be adjusted and continue to be used, while global efficiency will be abandoned and replaced with a better method-weighted modularity. The reason is that global efficiency measures efficiency by distance, so it is not very meaningful when the weights are flows but not actual distance.

a. Weighted Clustering Coefficient

There are several definitions of weighted clustering coefficient (Saramäki et al., 2007), and the one used here is proposed by Onnela et al. (2005) defining as “the geometric average of the subgraph edge weights”. The larger weights represent stronger connections, so there is no need to invert ‘flows’ when calculating.

$$C_i = \frac{1}{k_i(k_i-1)} \sum_{v\omega} (\widehat{\omega}_{iv}\widehat{\omega}_{i\omega}\widehat{\omega}_{v\omega})^{\frac{1}{3}} \quad (10)$$

$$C_w = \frac{1}{N}\sum_{i=1}^{n} C_i \quad (11)$$

Where: $\widehat{\omega}_{iv}$ *is the edge weight.*

b. Weighted Modularity

Modularity measures the extent of community division, in which higher positive modularity values indicating a stronger community structure. For weighted measure, the adjacency matrix will be replaced of edge weight matrix in euqaiton (Newman, 2010). In NetworkX package, the calculation is reduced as Eq.13 (Clauset et al., 2004).

$$Q = \frac{1}{2m}\sum_{ij}(A_{ij} - \gamma\frac{k_i k_j}{2m})\delta(c_i, c_j) \quad (12)$$

$$Q = \sum_{c=1}^{n}\left[\frac{Lc}{m} - \gamma\left(\frac{kc}{2m}\right)^2\right] \quad (13)$$

Where: A_{ij} *is the edge weight matrix,* γ *is the resolution parameter,* $\delta(c_i, c_j)$ *is 1 if* i *and* j *are in the same community else 0,* m *is the number of links.*

II.3 Impact Measures with Flows

According to part I, the best performing centrality measure of the best strategy is the betweenness centrality of sequential strategy (B). Therefore, it is selected as the first experimental set. As the question requires, the rest of the experimental sets are from II.1, so there are 4 experimental sets.

Table 3 Removed Nodes in II.3

Experimental Sets	Statement	3 Highest Ranked Nodes			Whether considering weight?
Experimental Set 1	Best-performing in I	Stratford 7785.967073	Liverpool Street 5535.077442	Upminster 4985.163205	NO
Experimental Set 2	Degree Centrality Measure in II	Stratford 0.922110553	Highbury & Islington 0.806532663	Whitechapel 0.781407035	NO
Experimental Set 3	Closeness Centrality Measure in II	Waterloo 0.396915774	Canary Wharf 0.396913326	Bank and 0.396912418	YES
Experimental Set 4	Betweenness Centrality Measure in II	Waterloo 38082	Bank and Monument 33506	Canary Wharf 32176	YES

The results of two measures are shown in plots, which black and yellow line denote the unweighted and weighted centrality sets respectively. An interesting finding is that the trends of sets considering weights (E3, E4) and not considering weights (E1, E2) are almost opposite performance. As important nodes are removed, average clustering coefficient rises and modularity first decrease and then increase for weighted experimental sets.

What the plots show is actually reasonable for this case, which the weight of network represent flows. Important stations tend to have high flows, the removal of them disperses the flows to other originally tightly-knit clusters, causing the increasing clustering coefficient. In terms of modularity, the sudden removal of the most important node fragments the overall community and causing a sudden drop, but the remaining nodes may reorganize to new communities as the network adapting, which causing an increase.

Based on the analysis, it can be concluded that the closure of Waterloo station will have the largest impact on passengers, which is the first removed nodes for experimental set 3 and 4.

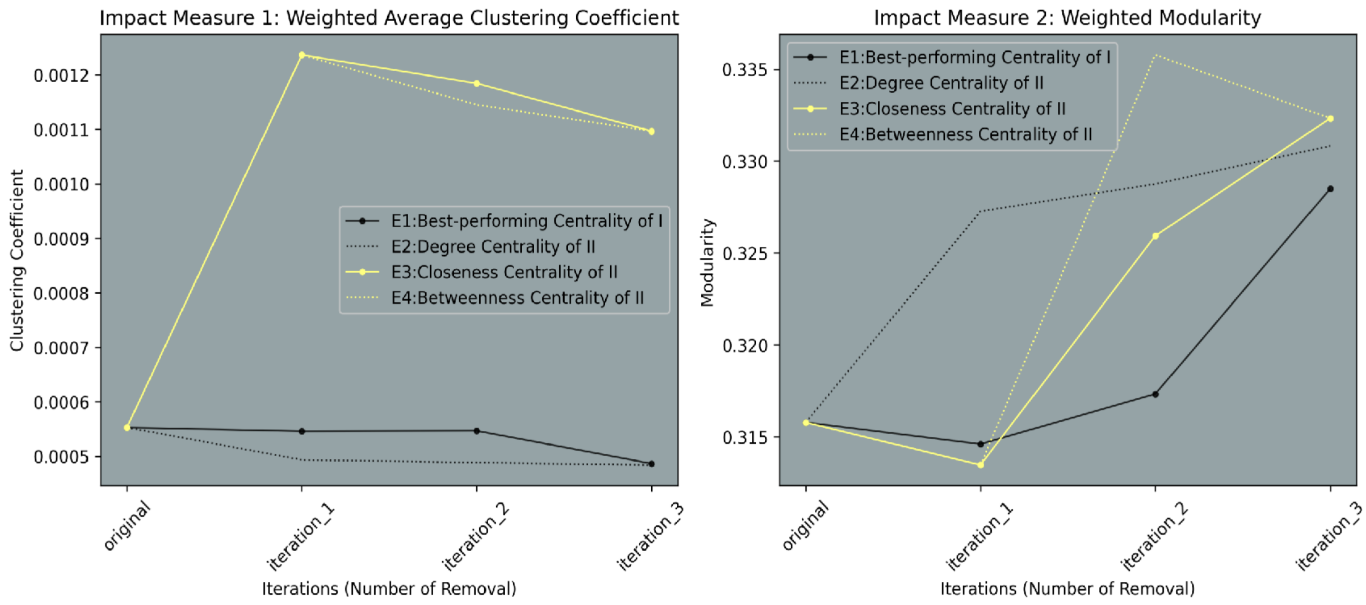


Figure 5 Results of Impact Measure in II.3

Part 2: Spatial Interaction models

III. Models and calibration

III.1 Spatial Interaction Models

Among various forms of spatial interaction models, the gravity model is most commonly used. The basic gravitational model was firstly proposed in 18th and later developed by Stewart (1948), Hansen (1959), Haynes and Fotheringham (1984) etc., eventually becoming the traditional model that used today, which is more familiar from Wilson (1971)'s paper.

$$T_{ij} \propto \frac{O_i D_j}{d_{ij}^{\beta}} = K \frac{O_i D_j}{d_{ij}^{\beta}} = K O_i D_j d_{ij}^{-\beta} \quad (14)$$

In research analysis, the parameters for O_i and D_j are always added to the equation:

$$T_{ij} \propto \frac{O_i^{\alpha} D_j^{\gamma}}{d_{ij}^{\beta}} = K \frac{O_i^{\alpha} D_j^{\gamma}}{d_{ij}^{\beta}} = K O_i^{\alpha} D_j^{\gamma} d_{ij}^{-\beta} \quad (15)$$

The central idea of this model is that the mass of origin and destination is directly proportional to flows, while the distance between them is inversely proportional. Specifically:

- T_{ij}: Measure of the interaction between zones i and j, which is generally considered as the total flows.
- O_i: 'Mass term' of origin i, also considered attractor, which can be denoted by many properties such as population (P_i) and salary (S_i).
- D_j: 'Mass term' of destination j, as with O_i , can be denoted by attractive properties such as population (P_j), salary (S_j), etc.
- d_{ij}: Measure of the distance or generalised cost between i and j, also considered deterrence.
- K: A constant of proportionality or 'scaling' parameter, which preserving the equation when estimating.

- α, γ, β: Calibration parameter, which can represent the impact extent.

In fact, the equation can be written in a more general form with distance decay parameter $f(d_{ij})$. The $f(d_{ij})$ could be any function that indicates the negative correlation between distance and flow, such as inverse power law $d_{ij}^{-\beta}$ in the traditional model and negative exponential function $\exp(-\beta d_{ij})$.

$$T_{ij} = KO_i^{\alpha} D_j^{\gamma} f(d_{ij}) \quad (16)$$

In Wilson (1971)'s paper, he further developed it to some constrained models, which is 'the Family of Spatial Interaction Models'.

a. Unconstrained Model

$$T_{ij} = KO_i^{\alpha} D_j^{\gamma} f(d_{ij}) \;\; subject\ to \sum_{i=1}^{n} \sum_{j=1}^{m} T_{ij} = T \quad (16)$$

b. Singly-Constrained Model

b1. Origin-Constrained Model

$$T_{ij} = A_i O_i D_j^{\gamma} f(d_{ij}) \; subject\ to \sum_{j=1}^{m} T_{ij} = O_i \quad (17)$$

b2. Destination-Constrained Model

$$T_{ij} = B_j D_j O_i^{\alpha} f(d_{ij}) \; subject\ to \sum_{i=1}^{n} T_{ij} = D_J \quad (18)$$

c. Doubly Constrained Model

$$T_{ij} = A_i O_i B_j D_j f(d_{ij}) \; subject\ to \sum_{j=1}^{m} T_{ij} = O_i \;\; \& \sum_{i=1}^{n} T_{ij} = D_J \quad (19)$$

III.2 Model Calibration

In this case, population and jobs are the mass of origin and destination respectively, which can be extended to the number of people who live near each station (origin) that are willing to find work and the job opportunities at each destination. Generally, there is always little variability of resident population as people are unwilling to change the residence easily, but the attractiveness of job opportunities easily draws flows. Based on this social reality consideration, origin-constrained model is selected.

For the choice of distance decay function, there is no hard and fast rule. In this case, given that the inverse power function has a more rapid decaying effect compared to negative exponential function, conjecturing the negative exponential would be more suitable for the 'residence-work' commuting model. But to ensure the calibration more accurate, two functions will be calibrated and the β of the better fitting one will be selected.

$$Inverse\ Power: \; T_{ij} = A_i O_i D_j^{\gamma} d_{ij}^{-\beta} \;\; subject\ to \sum_{j=1}^{m} T_{ij} = O_i \quad (20)$$

$$Negative\ Exponential: T_{ij} = A_i O_i D_j^{\gamma} \exp(-\beta d_{ij}) \;\; subject\ to \sum_{j=1}^{m} T_{ij} = O_i \quad (21)$$

Before calibrating the β, three preparing steps are done:

1. Detecting if there is internal flow and removing: 18 rows are removed.

2. Removing the rows with population or jobs equal to 0: 43 rows of Battersea Park station are removed.

3. Replacing the flow values of 0 with 0.001. (Not necessarily required in this case).

Then two models are calibrated and the parameter β for inverse power and negative exponential are 0.8781 and 0.0002 respectively.

Table 4 Calibration Result of 'Inverse Power Law' Function

Generalized Linear Model Regression Results (Inverse Power Law)			
Dep. Variable:	flows_c	No. Observations:	61413
Model:	GLM	Df Residuals:	61013
Model Family:	Poisson	Df Model:	399
Link Function:	Log	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-1.0169e+06
Date:	Sat, 06 May 2023	Deviance:	1.8612e+06
Time:	14:19:44	Pearson chi2:	2.78e+06
No. Iterations:	8	Pseudo R-squ. (CS):	1.000
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
station_origin[Abbey Road]	3.2502	0.042	76.939	0	3.167	3.333
station_origin[Acton Central]	5.0169	0.031	162.461	0	4.956	5.077
station_origin[Acton Town]	4.5629	0.02	224.183	0	4.523	4.603
station_origin[Aldgate]	3.3238	0.022	153.542	0	3.281	3.366
station_origin[Aldgate East]	3.4577	0.021	164.94	0	3.417	3.499
station_origin[All Saints]	3.3806	0.038	88.25	0	3.306	3.456
station_origin[Alperton]	4.1759	0.028	150.076	0	4.121	4.23
...	...	...	...	...	...	...
station_origin[Woodside Park]	4.6986	0.022	213.954	0	4.656	4.742
station_origin[Woolwich Arsenal]	6.9056	0.017	417.01	0	6.873	6.938
log_jobs	**0.7686**	0.001	1229.506	0	0.767	0.77
log_distance	**-0.8781**	0.001	-767.044	0	-0.88	-0.876

Table 5 Calibration Result of 'Negative Exponential' Function

Generalized Linear Model Regression Results (Negative Exponential)			
Dep. Variable:	flows_c	No. Observations:	61413
Model:	GLM	Df Residuals:	61013
Model Family:	Poisson	Df Model:	399
Link Function:	Log	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-9.0994e+05
Date:	Sat, 06 May 2023	Deviance:	1.6474e+06
Time:	14:19:44	Pearson chi2:	2.40e+06
No. Iterations:	8	Pseudo R-squ. (CS):	1.000
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
station_origin[Abbey Road]	-2.9142	0.041	-70.505	0	-2.995	-2.833
station_origin[Acton Central]	-1.162	0.029	-39.957	0	-1.219	-1.105
station_origin[Acton Town]	-1.613	0.017	-92.795	0	-1.647	-1.579
station_origin[Aldgate]	-2.9429	0.02	-150.131	0	-2.981	-2.904
station_origin[Aldgate East]	-2.8546	0.019	-151.952	0	-2.891	-2.818
station_origin[All Saints]	-2.8782	0.037	-77.214	0	-2.951	-2.805
station_origin[Alperton]	-1.6542	0.026	-64.728	0	-1.704	-1.604
...	...	...	...	...	...	...
station_origin[Woodside Park]	-1.149	0.019	-60.218	0	-1.186	-1.112
station_origin[Woolwich Arsenal]	0.5181	0.013	40.462	0	0.493	0.543
log_jobs	**0.7552**	0.001	1184.99	0	0.754	0.756
distance	**-0.0002**	1.88E-07	-814.151	0	0	0

To test the fitting performance of two models, R-squared (R^2) and Square Root of Mean Squared Error (RMSE) are used, which higher R^2 and lower RMSE indicate the better "goodness-of-fit".

Table 5 “Goodness-of-fit” Results

	R^2	RMSE
Inverse Power	0.388269921	102.893
Negative Exponential	0.468063256	96.264

Therefore, the model with negative exponential function is selected and the β is calibrated as 0.0002 (If more decimal places are retained, it will be 0.00015315869998…), with the corresponding γ is 0.7552.

IV. Scenarios

IV.1 Scenario A

Assuming that Canary Wharf has a 50% decrease in jobs and using the calibrated parameter β and γ to recalculate the flow. The flows into Canary Wharf drop from 55954.03 to 29493.89 and the ranking drops from 3 to 7, indicating a strong impact. But among the top 10 stations in terms of inflows, most station are not increase as expected, suggesting that economies of urban central areas are mutually influential.

It is worth noting that the number of commuters should be conserved after recompute. To ensure that, the A_i should be recalculate after using the updated ‘jobs’ data. After calculating the new A_i, the original ‘jobs’ data can be substituted in the calculation to check whether it is consistent with the original flow. Finally, checking if the total flow of recalculation is equal to the original.

Table 6 Flows of Original Scenario

station_destination / station_origin	Bank and Monument	Liverpool Street	Canary Wharf	Stratford	Oxford Circus	King's Cross St. Pancras	Victoria	London Bridge	Green Park	Farringdon	...	Roding Valley	Grange Hill	Emerson Park	All
Abbey Road	0.00	NaN	1.00	285.00	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	**599.00**
Acton Central	NaN	NaN	NaN	11.00	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	**1224.02**
Acton Town	66.00	55.00	57.00	10.00	38.00	64.00	53.00	39.00	79.00	19.00	...	NaN	NaN	NaN	**3745.05**
Aldgate	5.00	67.00	1.00	6.00	15.00	351.00	5.00	2.00	3.00	677.00	...	NaN	NaN	NaN	**2886.11**
Aldgate East	95.00	47.00	3.00	82.00	50.00	111.00	31.00	25.00	46.00	140.00	...	0.00	0.00	NaN	**3172.09**
All Saints	93.00	NaN	67.00	258.00	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	**740.01**
Alperton	7.00	4.00	7.00	2.00	3.00	15.00	19.00	7.00	22.00	1.00	...	NaN	NaN	NaN	**1624.04**
Amersham	14.00	57.00	54.00	4.00	15.00	42.00	27.00	26.00	15.00	97.00	...	NaN	NaN	NaN	**1190.06**
Anerley	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	**643.00**
Angel	591.00	19.00	96.00	29.00	131.00	95.00	189.00	436.00	71.00	14.00	...	0.00	1.00	NaN	**4199.07**
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	**...**
Woodgrange Park	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	**530.01**
Woodside Park	240.00	27.00	42.00	18.00	69.00	92.00	90.00	90.00	48.00	32.00	...	NaN	NaN	NaN	**3093.03**
Woolwich Arsenal	1340.00	NaN	82.00	1642.00	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	**7892.00**
All	**78549.01**	**61122.01**	**58772.01**	**55954.03**	**44368.00**	**33330.01**	**33251.00**	**29926.01**	**26754.00**	**25592.03**	**...**	**84.04**	**76.08**	**42.00**	**1542300.47**

Table 7 Flows of Scenario A

station_destination station_origin	Bank and Monument	Stratford	Liverpool Street	Oxford Circus	King's Cross St. Pancras	Highbury & Islington	Canary Wharf	Victoria	London Bridge	Green Park	...	West Ruislip	Cheshunt	Emerson Park	All
Abbey Road	83.33	187.19	NaN	NaN	NaN	NaN	63.22	NaN	NaN	NaN	...	NaN	NaN	NaN	**599.02**
Acton Central	NaN	39.90	NaN	NaN	NaN	68.27	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	**1224.05**
Acton Town	92.58	22.58	67.34	101.16	55.82	30.60	20.83	99.83	50.40	80.59	...	NaN	NaN	NaN	**3745.11**
Aldgate	212.78	66.95	196.43	72.85	73.14	40.10	46.17	53.86	89.64	47.54	...	0.03	NaN	NaN	**2886.10**
Aldgate East	224.55	91.36	207.24	76.88	77.16	42.30	63.01	56.84	94.60	50.17	...	0.04	NaN	NaN	**3172.07**
All Saints	125.71	127.71	NaN	NaN	NaN	NaN	120.61	NaN	NaN	NaN	...	NaN	NaN	NaN	**740.01**
Alperton	41.48	10.12	30.17	45.33	25.01	13.71	9.33	44.73	22.59	36.11	...	NaN	NaN	NaN	**1623.98**
Amersham	14.82	5.51	12.28	18.27	13.20	9.42	3.09	10.46	6.92	10.68	...	NaN	NaN	NaN	**1190.09**
Anerley	NaN	NaN	NaN	NaN	NaN	39.52	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	**642.98**
Angel	232.81	68.08	203.01	138.20	164.78	90.34	48.55	79.11	98.07	80.79	...	0.06	NaN	NaN	**4199.04**
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	**...**
Woodgrange Park	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	**529.99**
Woodside Park	117.88	35.40	102.80	120.70	110.47	60.56	24.58	69.09	49.66	70.56	...	NaN	NaN	NaN	**3093.01**
Woolwich Arsenal	990.40	1429.09	NaN	NaN	NaN	NaN	950.22	NaN	NaN	NaN	...	NaN	NaN	NaN	**7891.99**
All	**70019.15**	**67142.79**	**59367.45**	**40065.21**	**30802.99**	**29980.58**	**29493.89**	**28934.48**	**27424.71**	**26728.79**	**...**	**69.93**	**64.88**	**41.81**	**1542301.48**

(Inflow for Table 5 & 6 Are Sorted in Descending Order)

IV.2 Scenario B

As the distance is conserved, β is changed to denote the increase of transportation cost, which the value of double (β_1) and quadruple (β_2) of the original β are selected. The larger the β value, the more rapidly the distance decay effect will appear, which indicates a higher travel cost (Figure 5).

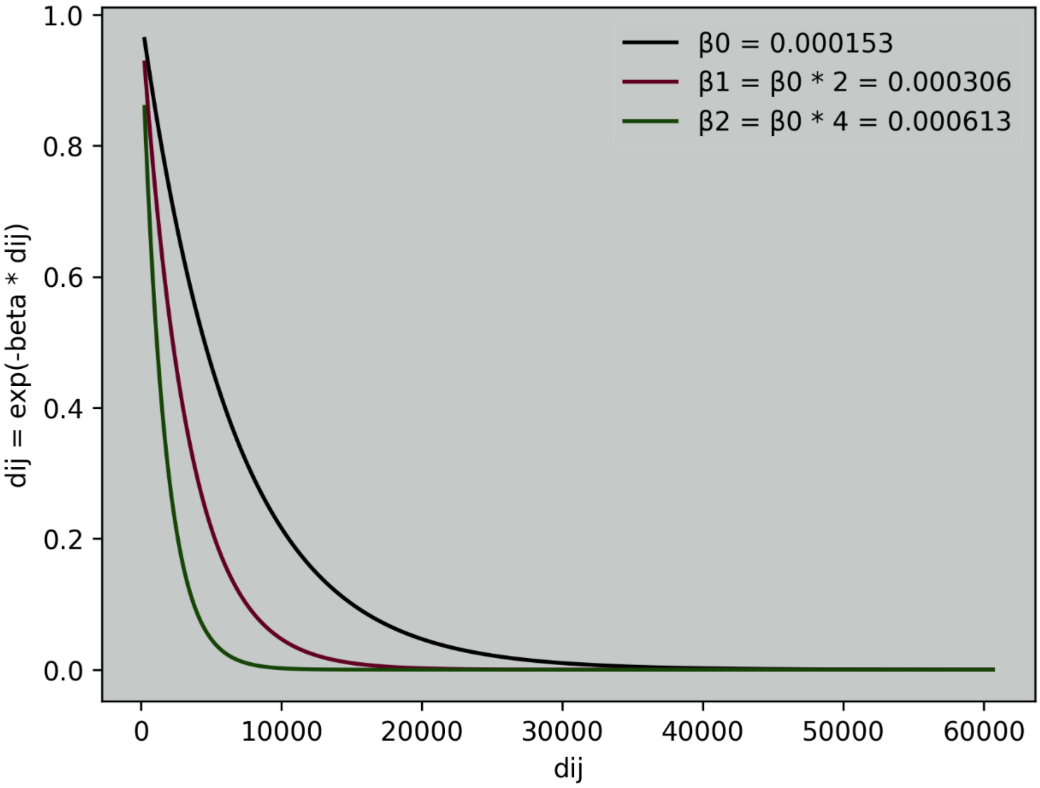


Figure 6 Decay Curves of Different Betas

Using β_1 and β_2 to recompute the flow, the results are as Table 7 and Table 8.

Table 8 Flows of Scenario B (β_1)

station_destination / station_origin	Bank and Monument	Stratford	Liverpool Street	Canary Wharf	Oxford Circus	King's Cross St. Pancras	Highbury & Islington	Victoria	London Bridge	Green Park	...	Emerson Park	South Hampstead	Beckton Park	All
Abbey Road	37.15	242.24	NaN	75.85	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	**599.00**
Acton Central	NaN	3.69	NaN	NaN	NaN	NaN	19.61	NaN	NaN	NaN	...	NaN	4.89	NaN	**1224.02**
Acton Town	35.03	2.69	22.40	6.29	64.39	24.33	8.98	77.97	21.52	59.89	...	NaN	NaN	NaN	**3745.05**
Aldgate	313.81	40.14	323.21	52.41	56.63	70.84	26.15	38.49	115.41	35.34	...	NaN	NaN	NaN	**2886.11**
Aldgate East	323.24	69.13	332.74	90.27	58.33	72.93	26.92	39.64	118.88	36.40	...	NaN	NaN	NaN	**3172.09**
All Saints	73.27	97.69	NaN	239.21	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	**740.01**
Alperton	11.26	0.87	7.20	2.02	20.69	7.82	2.89	25.06	6.92	19.25	...	NaN	NaN	NaN	**1624.04**
Amersham	0.09	0.02	0.08	0.01	0.22	0.14	0.09	0.09	0.04	0.11	...	NaN	NaN	NaN	**1190.06**
Anerley	NaN	NaN	NaN	NaN	NaN	NaN	7.28	NaN	NaN	NaN	...	NaN	NaN	NaN	**643.00**
Angel	278.89	30.81	256.30	43.01	151.28	266.95	98.55	61.64	102.57	75.76	...	NaN	NaN	NaN	**4199.07**
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Woodgrange Park	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	**530.01**
Woodside Park	68.58	7.99	63.03	10.58	110.67	115.07	42.48	45.09	25.22	55.42	...	NaN	NaN	NaN	**3093.03**
Woolwich Arsenal	401.10	1078.93	NaN	1309.44	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	7.50	**7892.00**
All	**62987.96**	**60980.33**	**49484.01**	**45260.01**	**38386.00**	**29495.32**	**27104.82**	**26359.56**	**25821.05**	**25369.97**	...	**113.01**	**95.49**	**93.72**	**1542300.47**

Table 9 Flows of Scenario B (β_2)

station_destination / station_origin	Bank and Monument	Stratford	Canary Wharf	Liverpool Street	Oxford Circus	King's Cross St. Pancras	Highbury & Islington	London Bridge	Green Park	Moorgate	...	South Hampstead	Beckton Park	Chesham	All
Abbey Road	6.02	330.78	31.25	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	**599.00**
Acton Central	NaN	0.01	NaN	NaN	NaN	NaN	0.66	NaN	NaN	NaN	...	1.41	NaN	NaN	**1224.02**
Acton Town	1.45	0.01	0.06	0.71	7.52	1.33	0.22	1.13	9.53	0.37	...	NaN	NaN	NaN	**3745.05**
Aldgate	464.38	9.82	16.12	595.37	23.28	45.21	7.57	130.16	13.28	229.32	...	NaN	NaN	0.00	**2886.11**
Aldgate East	464.99	27.48	45.15	595.52	23.31	45.22	7.57	130.33	13.30	229.38	...	NaN	NaN	NaN	**3172.09**
All Saints	25.05	57.53	332.38	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	**740.01**
Alperton	0.13	0.00	0.01	0.07	0.70	0.12	0.02	0.10	0.88	0.03	...	NaN	NaN	NaN	**1624.04**
Amersham	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	NaN	NaN	11.37	**1190.06**
Anerley	NaN	NaN	NaN	NaN	NaN	NaN	0.07	NaN	NaN	NaN	...	NaN	NaN	NaN	**643.00**
Angel	306.14	4.83	9.07	312.49	138.66	535.89	89.71	85.81	50.95	204.87	...	NaN	NaN	0.00	**4199.07**
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Woodgrange Park	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	**530.01**
Woodside Park	2.55	0.04	0.08	2.60	10.23	13.73	2.30	0.72	3.76	1.71	...	NaN	NaN	NaN	**3093.03**
Woolwich Arsenal	39.16	366.04	519.52	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	2.32	NaN	**7892.00**
All	**51204.21**	**49788.16**	**40669.77**	**35311.05**	**29506.08**	**23807.04**	**23449.51**	**22149.65**	**19990.05**	**19984.23**	...	**140.26**	**120.04**	**58.96**	**1542300.47**

(Inflow for Table 7 & 8 Are Sorted in Descending Order)

Comparing top 10 stations and bottom 3 stations of inflows between scenario B and the original. it shows that the top 10 stations all experienced a decrease in inflows, while the bottom 3 stations saw an increase. Increasing transportation costs under constant distance conditions cause central areas to experience decreased flows, while remote or unpopular areas experience increased flows, and this effect becomes more pronounced with greater increases in costs. There are probably two main reasons, one is that the high demand and density of transportation in urban centers lead people to travel less or seek a lower-cost trip mode when the cost rises, another is that the increasing travel cost leads some people to move to the suburbs, reducing their living costs to cope with economic issues.

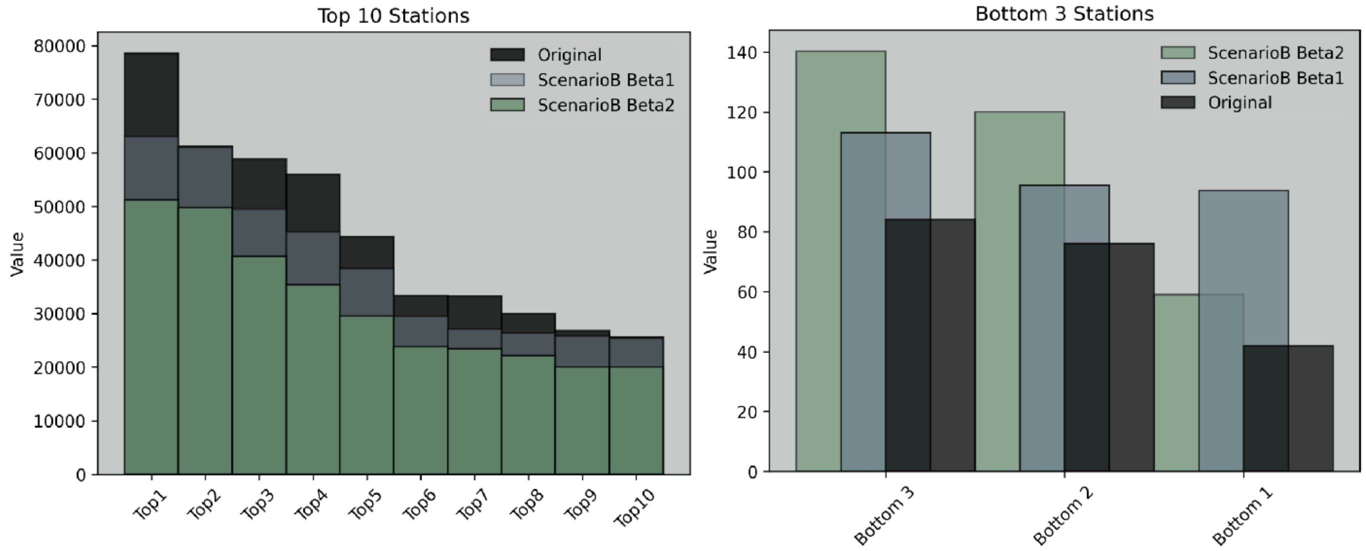


Figure 7 Inflow Comparison of Original and Scenario B

IV.3 Comparison of Three Situations

To explore the redistribution, three-step process is implemented following a research logic, transitioning from local to global, and from intuitive understanding to empirical data analysis. Considering that the application of origin-constrained model, all the investigation are based on the flows to destination (inflows).

a. Top 10 Stations Comparison

Top 10 stations in terms of inflow are extracted for comparison. Result shows that three situations are very similar to the original, in which the numbers of discrepancies are only 1, 1, 2 for scenario A, scenario B (β_1), scenario B (β_2) respectively. From this partial (local) perspective, it cannot clearly show which situation has the greatest impact on the redistribution.

Table 10 Top 10 Stations for Different Scenarios

Ranking	Top 1	Top 2	Top 3	Top 4	Top 5	Top 6	Top 7	Top 8	Top 9	Top 10
Original	Bank and Monument	Liverpool Street	Canary Wharf	Stratford	Oxford Circus	King's Cross St. Pancras	Victoria	London Bridge	Green Park	Farringdon
	78549.01	61122.01	58772.01	55954.03	44368.00	33330.01	33251.00	29926.01	26754.00	25592.03
Scenario A	Bank and Monument	Stratford	Liverpool Street	Oxford Circus	King's Cross St. Pancras	Highbury & Islington	Canary Wharf	Victoria	London Bridge	Green Park
	70019.15	67142.79	59367.45	40065.21	30802.99	29980.58	29493.89	28934.48	27424.71	26728.79
Scenario B Beta1	Bank and Monument	Stratford	Liverpool Street	Canary Wharf	Oxford Circus	King's Cross St. Pancras	Highbury & Islington	Victoria	London Bridge	Green Park
	62987.96	60980.33	49484.01	45260.01	38386.00	29495.32	27104.82	26359.56	25821.05	25369.97
Scenario B Beta2	Bank and Monument	Stratford	Canary Wharf	Liverpool Street	Oxford Circus	King's Cross St. Pancras	Highbury & Islington	London Bridge	Green Park	Moorgate
	51204.21	49788.16	40669.77	35311.05	29506.08	23807.04	23449.51	22149.65	19990.05	19984.23

b. Distribution Comparison

Inflows of all stations are considered through frequency distribution plots, which each figure represents a comparison between scenario and the origin. All data have been log-transformed to prevent skewness. The three plots show an increase in the concentration of data in the middle, so speculating a decrease of the dispersion across scenarios.

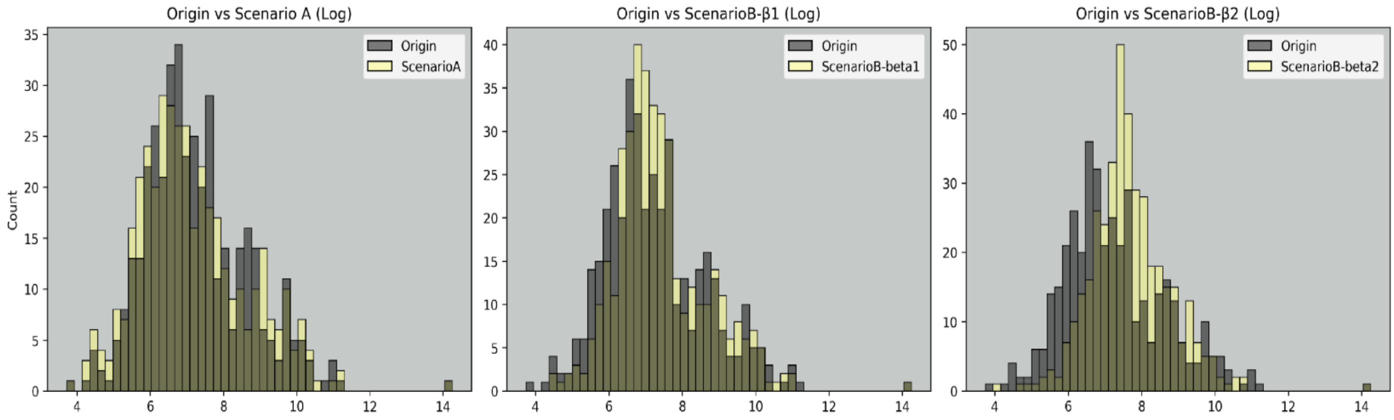


Figure 8 Redistribution Comparison

c. Empirical data analysis

To compare inflow redistribution, three performance indicators are used: variability (dispersion), asymmetry, and "tailedness", which are measured by standard deviation, skewness, and kurtosis, respectively. As conjecture, the dispersion of data decreases from scenario A to scenario B (β_2), which indicates that more values close to the mean and extremums reduce. The increasing skewness indicates that the extent of data skew is increasing slightly. While the decreasing kurtosis indicates that, although the shape of distribution may show more asymmetry, the frequency of extreme values is relatively low.

Table 11 Distribution Metrics

	Standard Deviation	Skewness	Kurtosis
Origin	77464.62374	19.63560	389.80330
ScenarioA	77435.08246	19.65720	390.40380
ScenarioB-beta1	77374.10896	19.70271	391.64338
ScenarioB-beta2	77242.87388	19.80181	394.33126

In conclusion, decreasing job opportunities and increasing travel costs all reduce the inflows of central areas but increase the flow into unpopular areas, leading a more balanced flow throughout London city. Among three situations, travel cost increasing has greater impact on redistribution than job reduction, and the impact increases as the β grows. Therefore, the scenario B with the larger β has the most impact of redistribution.

Bibliography

Clauset, A., Newman, M.E.J., Moore, C., 2004. Finding community structure in very large networks. Phys. Rev. E 70, 066111. https://doi.org/10.1103/PhysRevE.70.066111

Ek, B., VerSchneider, C., Narayan, D.A., 2015. Global efficiency of graphs. AKCE International Journal of Graphs and Combinatorics 12, 1–13. https://doi.org/10.1016/j.akcej.2015.06.001

Hansen, D.L., Shneiderman, B., Smith, M.A., Himelboim, I., 2020. Chapter 3 - Social network analysis: Measuring, mapping, and modeling collections of connections, in: Hansen, D.L., Shneiderman, B., Smith, M.A., Himelboim, I. (Eds.), Analyzing Social Media Networks with NodeXL (Second Edition). Morgan Kaufmann, pp. 31–51. https://doi.org/10.1016/B978-0-12-817756-3.00003-0

Hansen, W.G., 1959. How Accessibility Shapes Land Use. Journal of the American Institute of Planners 25, 73–76. https://doi.org/10.1080/01944365908978307

Haynes, K., Fotheringham, A., 1984. Gravity and Spatial Interaction Models. 1985.

Latora, V., Marchiori, M., 2001. Efficient behavior of small-world networks. Phys Rev Lett 87, 198701. https://doi.org/10.1103/PhysRevLett.87.198701

Newman, M.E.J., 2010. Networks: an introduction,page 224. Oxford University Press, Oxford.

Onnela, J.-P., Saramäki, J., Kertész, J., Kaski, K., 2005. Intensity and coherence of motifs in weighted complex networks. Phys. Rev. E 71, 065103. https://doi.org/10.1103/PhysRevE.71.065103

Sahoo, R., Rani, T.S., Bhavani, S.D., 2016. Chapter 17 - Differentiating Cancer From Normal Protein-Protein Interactions Through Network Analysis, in: Tran, Q.N., Arabnia, H.R. (Eds.), Emerging Trends in Applications and Infrastructures for Computational Biology, Bioinformatics, and Systems Biology, Emerging Trends in Computer Science and Applied Computing. Morgan Kaufmann, Boston, pp. 253–269. https://doi.org/10.1016/B978-0-12-804203-8.00017-1

Saramäki, J., Kivelä, M., Onnela, J.-P., Kaski, K., Kertész, J., 2007. Generalizations of the clustering coefficient to weighted complex networks. Phys. Rev. E 75, 027105. https://doi.org/10.1103/PhysRevE.75.027105

Stewart, J.Q., 1948. Demographic Gravitation: Evidence and Applications. Sociometry 11, 31–58. https://doi.org/10.2307/2785468

Wasserman, S., Faust, K., 1994. Social Network Analysis: Methods and Applications. Cambridge University Press.

Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of ‘small-world’ networks. Nature 393, 440–442. https://doi.org/10.1038/30918

Wilson, A.G., 1971. A Family of Spatial Interaction Models, and Associated Developments. Environ Plan A 3, 1–32. https://doi.org/10.1068/a030001

Zhang, L., Zhao, Y., Chen, D., Zhang, X., 2021. Analysis of Network Robustness in Weighted and Unweighted Approaches: A Case Study of the Air Transport Network in the Belt and Road Region. Journal of Advanced Transportation 2021, e8810254. https://doi.org/10.1155/2021/8810254