Urban Simulation Final Assessment

Word Count: 3243

1 London's underground resilience

1.1 Topological Network

1.1.1 Centrality Measures

Among all centrality measures comparatively illustrated on **Table 1**, the 'closeness centrality', 'betweenness centrality' and 'information centrality' have been selected as the measures for the following study.

Table 1 Definition and Equation of different Centrality Measures(Oldham et al., 2019)

Centrality Measure	Characteristics of a central node	Equation
Degree Centrality	Connected to many other nodes	$DC_i = d_i = \sum_{j \neq i} A_{ij}$
Closeness Centrality	**Minimal total cost to access to all other nodes**	$CC_i = \frac{N}{\sum_j l_{ij}}$
Betweenness Centrality	**Most frequent stop-by in all potential routes**	$BC_i = \sum_{p \neq q, q \neq i, p \neq i} \frac{\sigma_{pq}(i)}{\sigma_{pq}}$
Eigenvector Centrality	Connected to many other high-degree nodes	$EC_i = \frac{1}{\lambda_1} \sum_j A_{ji} V_j$
Katz Centrality	Connected to many other nodes from global network accessibility	$KC_i = \alpha \sum_j A_{ji} V_j + \beta$
Information Centrality	**Can be easily reached by paths from other nodes**	$IC_i = \left(C_{ii} + \frac{\sum_j C_{jj} - 2\sum_j C_{ij}}{N}\right)^{-1}$ or $IC_i = \frac{E(G) - E(G-i)}{E(G)}$
Laplacian Centrality	Removal of this node would greatly impair the network	$LAPC_i = d_i^2 + d_i + 2\sum_{j \in \mathcal{N}(i)} d_j$

A = adjacent matrix; d_i = degree of node i; N = number of nodes in a network; l_{ij} = length of the mean geodesic between nodes i and j; λ_1 = leading eigenvalue of A; v = leading eigenvector of A; $\sigma_{pq}(i)$ = the number of shortest-paths between any nodes pair p and q; $\sigma_{pq}(i)$ = the number of shortest-paths between any nodes pair p and q which pass through i; α = penalty on distant connections to a node's centrality score; β = preassigned centrality constant; $C = (L + J)^{-1}$ where L is the Laplacian of A and J is a $N * N$ with all elements equal to 1; $E(G)$ is the network's global efficiency and $E(G-i)$ is the network's global efficiency after removing node i.

The reasons for selecting these measures are:

- As the London tube network is undirected, some measures like 'PageRank Centrality', specifically for directed one is not appropriate for this question, same for some local centrality measures(Wan *et al.*, 2021) like degree or H-index.
- In terms of the robustness and resilience of network, betweenness centrality and harmonic closeness centrality will be more indicative(Wan *et al.*, 2021) on each node's functional support for the whole system, while degree centrality and Katz centrality only consider the neighbors rather than global factors.
- Honestly, Laplacian centrality(Qi *et al.*, 2012) might be the best choice for measuring nodes' significance of maintaining network's functional structure, which perfectly aligns with our research question on resilience. Yet, the networkX does not have corresponding functions for convenient calculation.

- As the commuting flow could be generalized as information flow, the information centrality could be also useful for measuring the efficiency of whole system transferring information or people.

In general, closeness centrality could quantify the accessibility of a station, indicating its efficiency in serving as a departure point for reaching all other stations in the network with minimal travel distance. It underscores the station's role in facilitating relatively short and convenient trips across the network. Betweenness centrality identifies the key transit hubs, highlighting the stations serving as critical junctions or bridges in the system. And information centrality represents station’s importance as boosting network’s operating efficiency on transferring passengers.

The calculation results are shown as **Table 2** and **Figure 1**.

Table 2 Top 10 Stations for 3 Centrality Measures

RANK	Closeness Centrality		Betweenness Centrality		Information Centrality	
	Station Name	Score	Station Name	Score	Station Name	Score
1	Green Park	0.1148	Stratford	0.298	Bank and Monument	0.000598
2	Bank and Monument	0.1136	Bank and Monument	0.290	King’s Cross St. Pancras	0.000591
3	King's Cross St. Pancras	0.1134	Liverpool Street	0.271	Liverpool Street	0.000587
4	Westminster	0.1125	King's Cross St. Pancras	0.255	Oxford Circus	0.000586
5	Waterloo	0.1123	Waterloo	0.244	Green Park	0.000585
6	Oxford Circus	0.1112	Green Park	0.216	Waterloo	0.000579
7	Bond Street	0.1110	Euston	0.208	Baker Street	0.000578
8	Farringdon	0.1107	Westminster	0.203	Bond Street	0.000573
9	Angel	0.1107	Baker Street	0.192	Stratford	0.000572
10	Moorgate	0.1103	Finchley Road	0.165	Moorgate	0.000570

While the first two metrics align with our expectations, indicating that stations in the city center exhibit higher centrality, but the information centrality appears a little weird. This is possibly due to unweighted network, where the distances between stations are not considered. Consequently, stations on suburban lines, characterized by longer distances between fewer stations (steps), tend to stand out more easily. For example, Caledonian Road would be recognized as same closer to King’s Cross as Russell Square because they are both one step to King’s Cross.

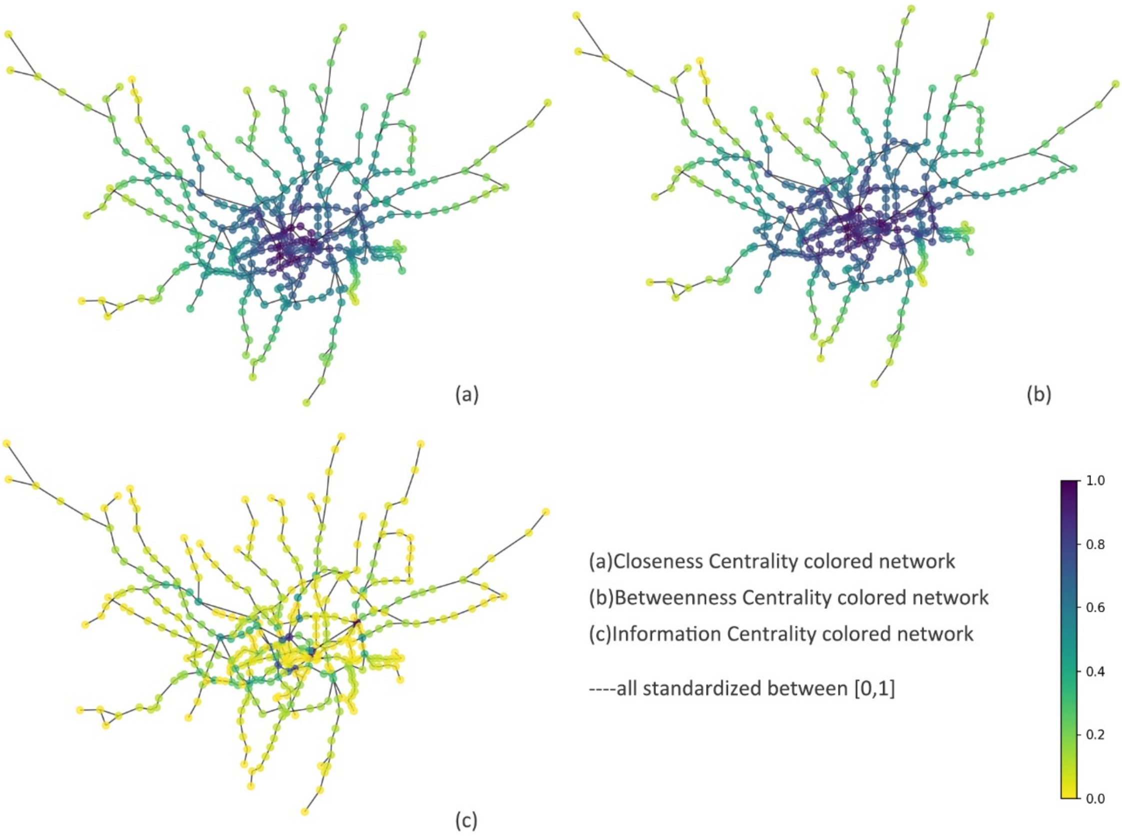

Figure 1 Three Centrality Measures for London Tube Network

1.1.2 Impact Measures for Node Removal

In order to assess the resilience of London's underground network, especially in scenarios of node removal, two global measures to evaluate are "Global Efficiency" and "Modularity".

$$E_{glob} = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{d_{(i,j)}} \tag{1}$$

Global Efficiency, equation (1), is an indicator of the overall communication efficiency within a network. For the London Underground, a decline in global efficiency would translate to longer travel times between stations, directly impacting commuter convenience and system service levels. A minimal change in global efficiency after node removal would suggest that the network maintains the efficiency of information flow, or in this case, passenger flow.

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \tag{2}$$

Modularity, equation (2), refers to the extent to which the network can be divided into modules or communities with dense interconnections within modules. A significant increase in modularity upon node removal might indicate the loss of critical bridges which connected different communities and bond different regions together.

To summarize, Global Efficiency stands from the perspective of flowing efficiency, while modularity is seeking insights in terms of structural stability, both helping to identify vulnerabilities and inform strategies to enhance the network's robustness. The computing results of impact measures is shown in **Table 3**.

Table 3 Global Measures of London Tube Network

Global Efficiency	**Modularity**
0.10125619359721513	0.8302138117924331

1.1.3 Node Removal

1.1.3.1 Non-sequential Removal

As required in assignment, the codes of removal process could be accessed here. The results are shown in **Figure 2** and **Figure 3** ('*GE*': Global Efficiency; '*Mod*': Modularity; And the difference is compared to the previous removal step instead of the initial measure value). The more detailed numeric table, including every node's name and centrality measure, is in tables in ***Appendix***.

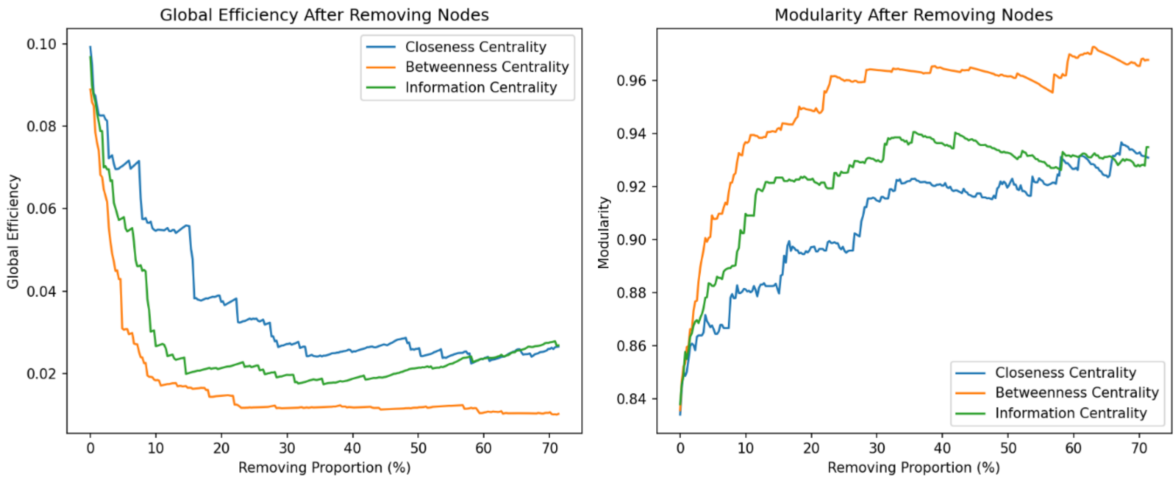

Figure 2 Impact Measures Trend of Removing 70% of all nodes

As we expected in 1.1.2 Impact Measures, removing high-centrality nodes would generally increase the modularity and decrease the Global Efficiency as shown on **Figure 2**, indicating that network system is more likely divorced into several separated communities and weakened on its flow efficiency.

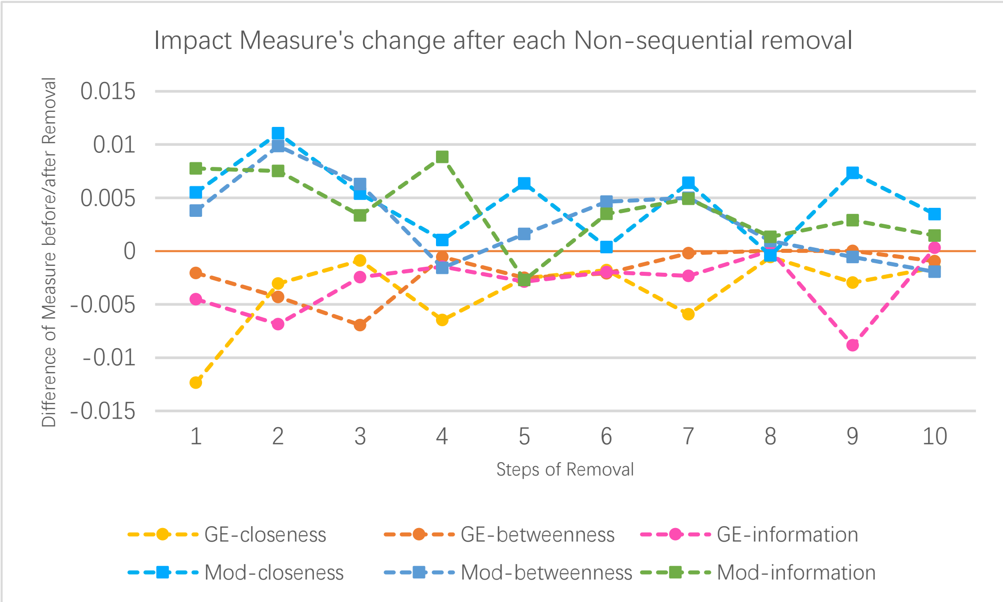

Figure 3 Impact Measure difference after each Non-sequential Removal.

However, following certain node removals, the corresponding measurements did not continue to rise or fall as expected, but rather oscillated in the opposite direction (**Figure 3**). Surprisingly, some of the values even crossed the zero threshold, indicating a complex and possibly non-linear response in the system dynamics.

1.1.3.2 Sequential Removal

In the Sequential Removal process, the weird oscillation did not show again. All the Global Efficient measures consistently decrease as nodes are being removed no matter which centrality measures are used as guidance. Same situation for Modularity.

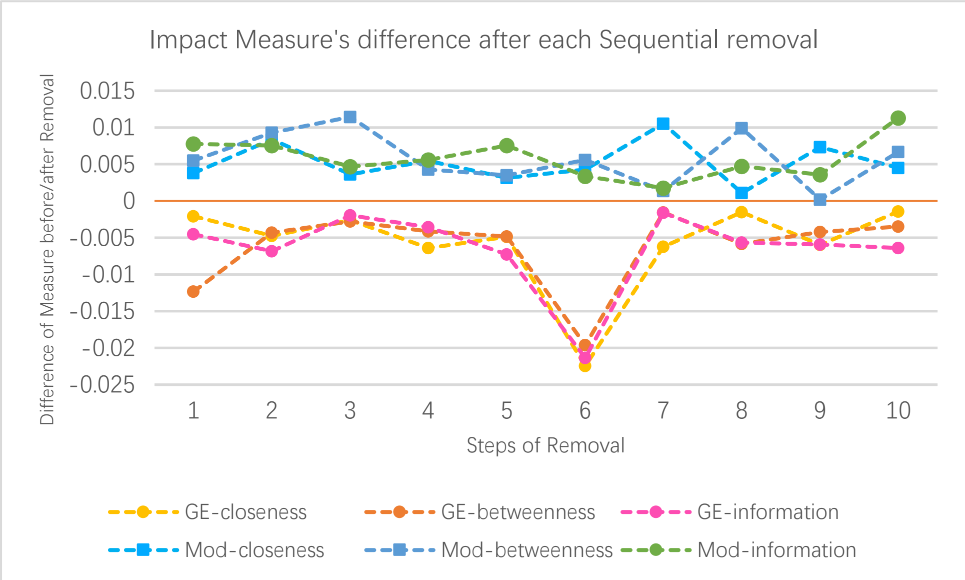

Figure 4 Impact Measure difference after each Sequential removal

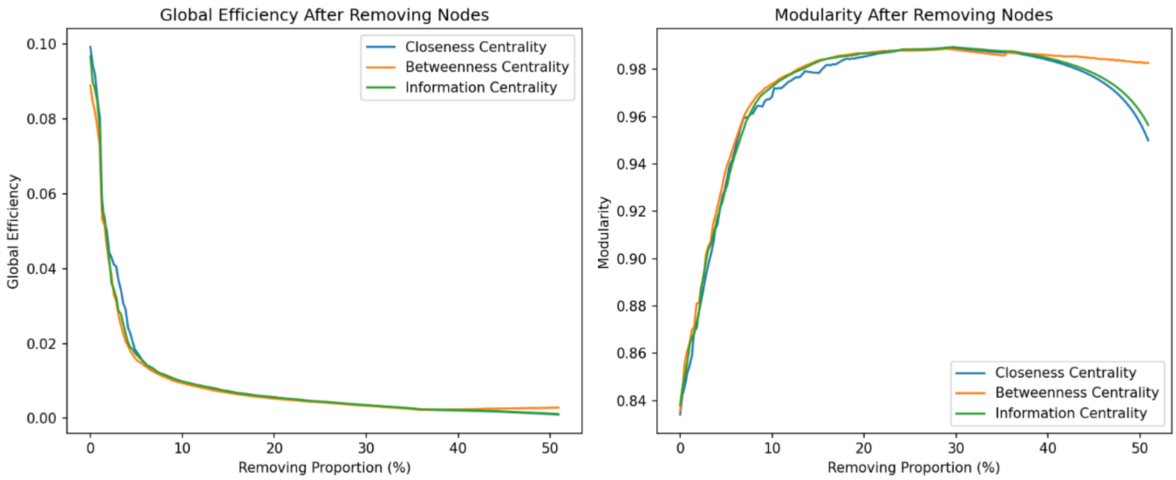

Figure 5 Impact Measure difference after each Sequential Removal

Following the sequential removal principles, the entire curve could be smoother, the reason for only removing 50 percent of all nodes is that there will be no more meaningful or valuable edges anymore, which makes the global efficiency less indicative. For example, we could see that once beyond 40%, modularity will decrease sharply (**Figure 5**), indicating that the whole network has been broken into several parts and lost its functional structure as public transport system.

1.1.3.3 Comparative Analysis

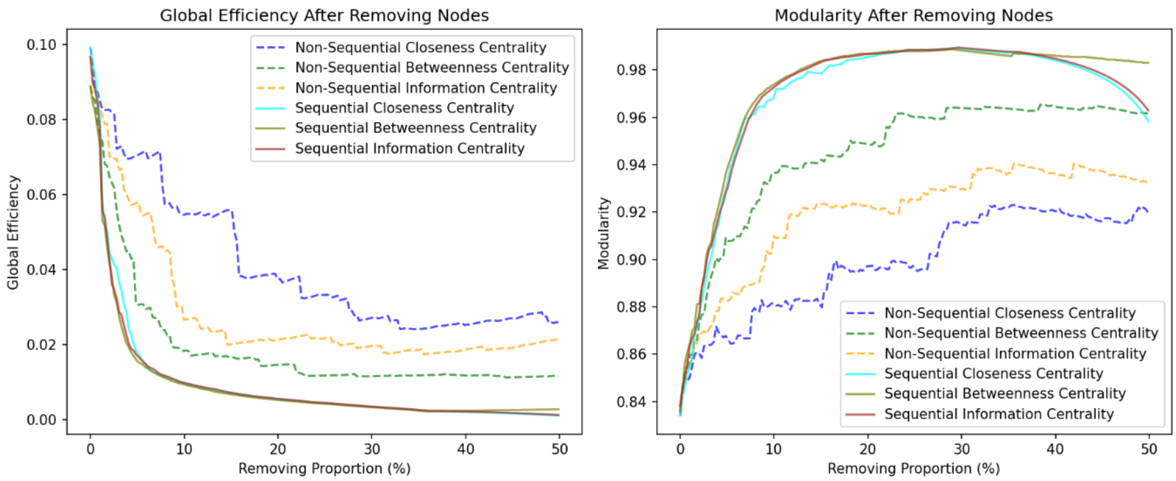

Figure 6 Comparison between Two Removal Strategies Guided by Three Centrality Measures

By comparing two different removal strategies (**Figure 6**), we can conclude that sequential removal always has better performance than non-sequential removal in following aspects:

a) Convergence rate: sequential removal will reach stability in relatively less steps, indicating that this strategy will negatively impact on the system more prominently.

b) Monotonic Behavior: sequential one exhibits a consistent monotonic increasing or decreasing trend without significant fluctuations, while non-sequential removal, without a real-time updated indicator's guidance, is more unpredictable.

c) Robustness towards measures: sequential strategy does not differ significantly towards different centrality measures. And for non-sequential removal, betweenness centrality has prominent advantages over the other two indicators, with closeness centrality be the worst one.

In conclusion, betweenness centrality reflects more indicative importance of a station for functioning of the underground. And sequential strategy is more effective in terms of studying resilience. Additionally, global efficiency is the better one when aiming to assess the damage after node removal because modularity will illustrate weird bounce-back after 40 percent removal.

1.2 Flows: weighted network

The weighted network added distance attributes to every edge in the network, also with the flow population between any neighbor stations. We could re-compute all centrality measures and global impact measures by taking distance or flow population into consideration.

1.2.1 Weighted Centrality Measures

In centrality measures computation, the weight of edges could be interpreted as various representations according to research question context. For example, the distance can represent ‘commuting cost time’ or ‘possibility of maintenance’, and the flow could refer to ‘connection intensity’ or ‘construction cost for larger space’.

Table 4 Top 10 Stations for 3 Weighted Centrality Measures

RANK	Closeness Centrality		Betweenness Centrality		Information Centrality	
	Station Name	Score	Station Name	Score	Station Name	Score
1	Holborn	7.923E-05	Bank and Monument	0.2215	Bank and Monument	0.7070
2	King's Cross St. Pancras	7.901E-05	King’s Cross St. Pancras	0.2097	Liverpool Street	0.6947
3	Tottenham Court Road	7.886 E-05	Stratford	0.1825	Stratford	0.6939
4	Oxford Circus	7.871 E-05	Baker Street	0.1642	Kings Cross St. Pancras	0.6934
5	Leicester Square	7.834 E-05	Oxford Circus	0.1573	Waterloo	0.6910
6	Piccadilly Circus	7.830 E-05	Euston	0.1551	Green Park	0.6880
7	Charing Cross	7.829 E-05	Earl’s Court	0.1435	Baker Street	0.6827
8	Chancery Lane	7.822 E-05	Shadwell	0.1394	Oxford Circus	0.6780
9	Covert Garden	7.805 E-05	Waterloo	0.1302	Highbury & Islington	0.6775
10	Embankment	7.798 E-05	South Kensington	0.1291	Bond Street	0.6726

When calculating these centrality measures (closeness centrality, betweenness centrality, and information centrality) in **Table 4**, they are generally based on the concept of shortest paths, where "cost" can be understood as the total weight of the path. Therefore, we set the argument “weight=’distance’” to re-calculate all above. The results are shown in **Figure 7**.

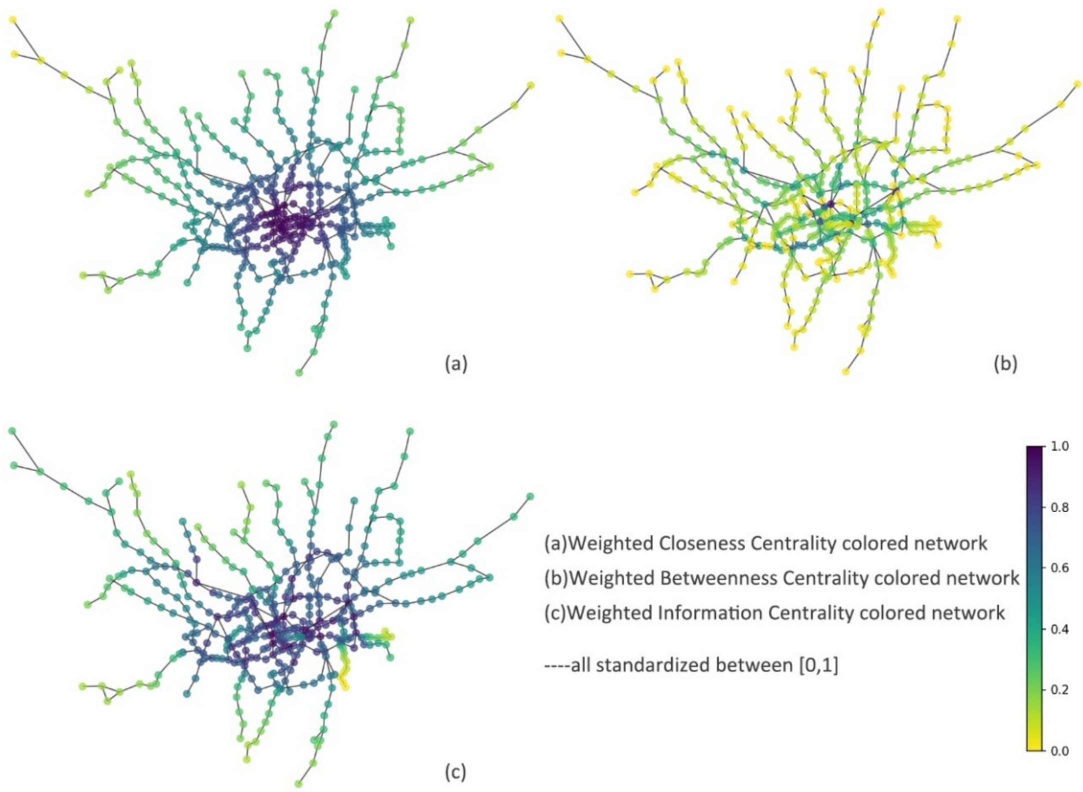

Figure 7 Three Weighted Centrality Measures for London Tube Network

Generally, the top 10 list for centrality measures does not change much compared with the unweighted one. However, some stations such as Tottenham Court Road and Piccadilly Circus are more likely to account for more importance after weight taken into consideration, especially for the closeness centrality.

1.2.2 Weighted Impact Measures for Node Removal

After distance and flow between stations have been included, both global efficiency and modularity can be calculated for weighted networks, so we assume that it is unnecessary to introduce a new measure because the original two are indicative enough.

In weighted networks, the weights of the edges are often interpreted as the inverse of the connection strength or the cost of the connection, so global efficiency considers the cost or length of the transmission paths. The calculation of modularity also considers the weights of the edges, meaning that not only the presence or absence of connections but also the strength of those connections play a crucial role in assessing modularity.

Since we have two attributes ’length’ and ‘flow’ for edges, in global efficiency, edge length should be used as the weight, representing lower costs and higher transmission efficiency, thus the calculation of global efficiency will be based on spatial distances to assess the efficiency of information or resource transfer in the network. And in modularity, then edge flow as the weight might be more suitable, reflecting the intensity of interactions between nodes, where edges with high flow might indicate the presence of communities or modules, as they represent more frequent interactions between nodes.

Hence, the results are shown in **Table 5**, due to the calculation equation difference, the direct comparison between weighted and unweighted measure would be less meaningful.

Table 5 Global Measures of London Tube Network

Global Efficiency (weight=’length’)	**Modularity (weight=’flow’)**
0.20251238719443027	0.7321343579048248
Global Efficiency (unweighted)	**Modularity (unweighted)**
0.10125619359721513	0.8302138117924331

1.2.3 Analysis and Discussion

As illustrated in **Figure 6**, Betweenness Centrality with the best performance will be utilized to assess the node removal impact **Table 6**.

Table 6 Top-3 Node Removal based on Betweenness Centrality

	Non-sequential Removal Node	Non-sequential Removal Global Efficiency	Non-sequential Removal Modularity	Sequential Removal Node	Sequential Removal Global Efficiency	Sequential Removal Modularity
1	**Bank and Monument**	**0.192504577**	**0.768687027**	**Bank and Monument**	**0.192504577**	**0.768687027**
2	**King's Cross St. Pancras**	**0.177975636**	**0.787755348**	**King's Cross St. Pancras**	**0.177975636**	**0.787755348**
3	**Stratford**	**0.15817742**	**0.798174503**	**Canada Water**	**0.163636298**	**0.790404175**
4	Baker Street	0.14839165	0.812908708	West Hampstead	0.118086273	0.791648762
5	Oxford Circus	0.14510062	0.835005856	Earl's Court	0.114552712	0.801013081
6	Euston	0.133036731	0.8364893	Oxford Circus	0.111933462	0.822412272
7	Earl's Court	0.129083012	0.848909598	Shepherd's Bush	0.100026102	0.825738617
8	Shadwell	0.125362495	0.850763879	Baker Street	0.092569934	0.841841261
9	Waterloo	0.121726901	0.875143913	Acton Town	0.080628221	0.842309325
10	South Kensington	0.12038975	0.879513971	Stratford	0.068053389	0.859556271
..	……	......	......	......	......	……

Something interesting is that, in the 3rd step, non-sequential strategy is seemingly more influential than sequential way. Even though they share the same first two options on ‘Bank and Monument’ and ‘King’s Cross’, some diversity shows up in 3rd step. However, in a long-term perspective, sequential strategy still exhibits better performance (**Figure 8**) on assessing node removal's impact on network system.

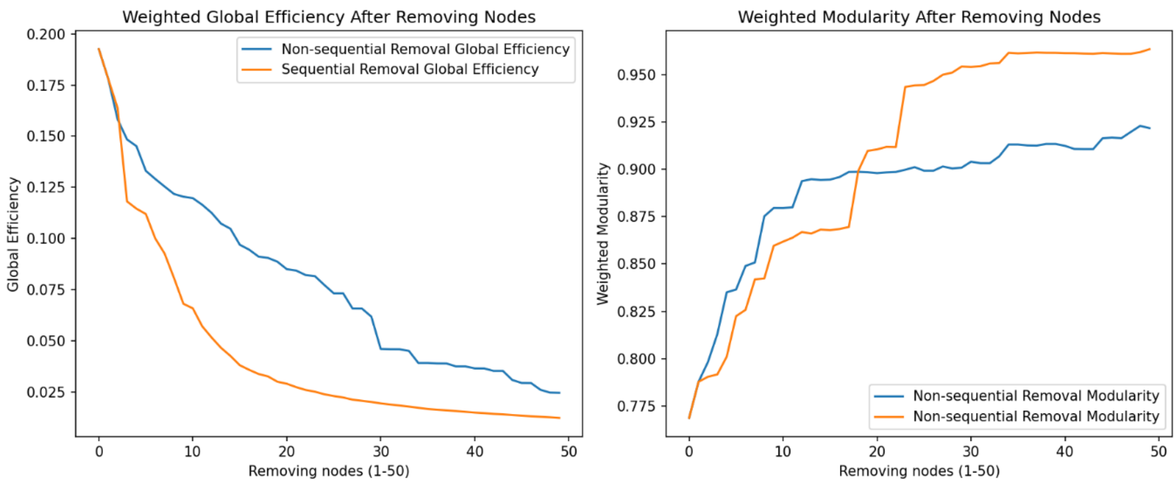

Figure 8 Global Efficiency and Modularity after 50 nodes removal by non-sequential/sequential

In general, we could conclude that 'Bank and Monument' should be the most crucial stations for the functional and structural operation of London underground system. Moreover, 'Bank and Monument' station also plays top roles if taking other centrality measures, like information centrality, into consideration.

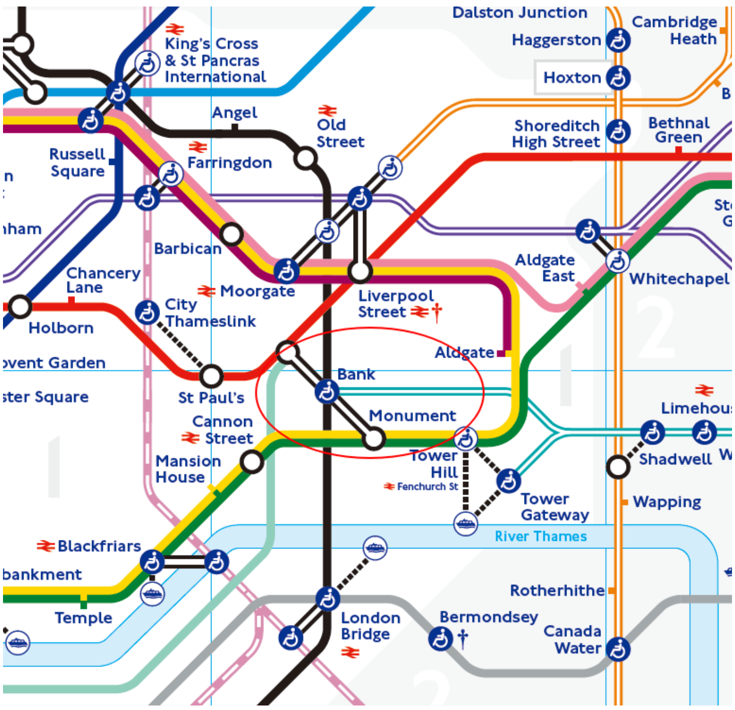

Figure 9 London Tube Map Published by TFL (Full Big Map HERE)

After checking the tube map of London, we could see that there are totally four lines in the central London, especially City of London, helping passengers across the Thames River. They are 'Northern', 'Waterloo & City', 'Thameslink' and 'Overground', and 'Bank and Monument' accounts for half of them, which are the more crowded two. Additionally, it is the crucial junction of London's east-west (along the Thames River direction) and northsouth (from King's Cross to Elephant Castle) axes (which could be inspected from London's population flow here). Also, 'Bank and Monument' also in the first and second busiest tube lines in London. We believe this could possibly explain its significance. Therefore, 'Bank and Monument' station's closure will have the largest impact on passengers.

However, we must be cautious that station's closure cannot be perfectly interpreted as the breakpoint or removed node in network. In the topological graph, if we removed certain nodes, it means that the whole route cannot be accessible anymore through that node, while in tube network, station's closure only means that we cannot use the station's entrance or exit but still could go through under it. Therefore, I believe more compatible analysis need be operated which aligns with our research goal in a better way rather than nodes removal.

---

2 Spatial Interaction Models

Spatial interaction models, originated from gravity model and introduced from physics to social research area by Stewart (Stewart, 1941), improved by Zipf emphasizing the scaling and distance's significance (Zipf, 1946), and systematically illustrated by Walter for economic geography and urban regional research (Isard, 1949), has evolved several versions and models such as intervening opportunities model (Stouffer, 1940) and competing destination model (Fotheringham, 1983).

2.1 Models and Calibration

2.1.1 Introduction to Models

Based on the Newton's universal gravitation theory, the gravity model assumes that the force of attraction between two sites is proportional to their scaling and inversely proportional to the distance separating them. As introduced in the lecture, the family of gravity models (Wilson, 1971) consists of several different models aligning with different spatial interaction patterns.

2.1.1.1 The Unconstrained Model

The equation of gravity models could be demonstrated as equation (3) (Some parameters are denoted differently for better aligning with our question context).

$$T_{ij} = K\frac{P_i W_j}{f(d_{ij})} \quad \text{subject to} \quad \sum_{i=1}^{n}\sum_{j=1}^{m} T_{ij} = T \tag{3}$$

T_{ij} : The transportation flow from station i to j.

P_i : The population of the origin station i.

W_j : The working opportunities/jobs of the destination station j.

d_{ij} : The distance or travelling cost from station i to j.

K : The calibration constant.

$f(d_{ij})$: The cost function, usually is $e^{-\beta * d_{ij}}$, and β is the decay parameter representing significance degree of d_{ij}.

n is the total number of origins i and m is the total number of destinations j.

Unconstrained gravity model will be utilized when only the total amount of the flow is accessible to the researcher, such like calculating the trading volume between countries or regions after certain amount of liquidity has been injected into the market(Anderson, 2010).

2.1.1.2 The Singly-Constrained Model

Singly-constrained Model elaborated in equation (4) will set either the origin (4-1) or the destination (4-2) fixed, meaning the total number of trips emanating from an origin or the total number of trips attracted to a destination is known and does not change.

$$T_{ij} = A_i \frac{P_i W_j}{f(d_{ij})} \quad \text{subject to} \quad \sum_{j=1}^{m} T_{ij} = P_i \tag{4-1}$$

$$T_{ij} = B_j \frac{P_i W_j}{f(d_{ij})} \quad \text{subject to} \quad \sum_{i=1}^{n} T_{ij} = W_j \tag{4-2}$$

A_i : often represents the "attractiveness" or "pull" factor of the destination zone j, which might be based on variables such as retail space, number of outlets, or employment, which quantify the shopping opportunities available.

B_j : could represent a balancing factor or a scaling parameter to ensure that the estimated trips align with observed data for the origin zone i.

In an applied context, the singly-constrained model can be used to analyze consumer behavior for shopping trips, taking into account the various sizes and numbers of stores within a zone. However, a strategically located shopping center may draw an unevenly large share of consumer visits due to its accessibility and offerings (Cochrane, 1975).

2.1.1.3 The Doubly-Constrained Model

The doubly-constrained model extends the principles of the singly-constrained models by incorporating both origins and destinations as fixed constraints. This model is employed when both the number of trips emanating from each origin and the number of trips attracted to each destination are predetermined and must match known data. Such constraints often come from comprehensive surveys or historical traffic data, ensuring that the model's predictions are aligned with actual transportation patterns.

$$T_{ij} = A_i B_j \frac{P_i W_j}{f(d_{ij})} \quad \text{subject to} \quad \sum_{j=1}^{m} T_{ij} = P_i \ \& \ \sum_{i=1}^{n} T_{ij} = W_j \qquad (4\text{-}2)$$

The formulation ensures that the interaction between origins and destinations reflects the realistic distribution of flows, taking into account both the pull of destinations and the push of origins. One practical scenario for the model is especially relevant in urban planning and population flow prediction.

2.1.2 Parameters Calibration

Considering the current data about population sizes P_i, number of jobs at each station W_i, and the actual flows from an Origin-Destination (OD) matrix, the Gravity Model is well-suited for scenarios where transport flows depend heavily on population and job availability, which directly aligns with the context of underground station commuters. Furthermore, the singly-constrained gravity model is selected for calibration because of its straightforward consistency to following scenarios where production(population) needs to be fixed and some changes will happen to destination(attraction).

Next, the model's parameters, particularly the decay parameter, need to be calibrated to fit the observed data. The objective of calibration is to minimize the difference between observed flows and those predicted by the model.

2.1.2.1 Dependent Variables Distribution

First, we need to determine the distribution patterns between 'flows' and 'distance' data so that we could choose the best method for transforming the data to meet the criteria for certain fitting models (Poisson or Negative Binomial). As shown in **Figure 10**, the histograms indicate that overall flows (after removing all zeros value) is more likely a power-law distribution with a Heavy-tailed distribution.

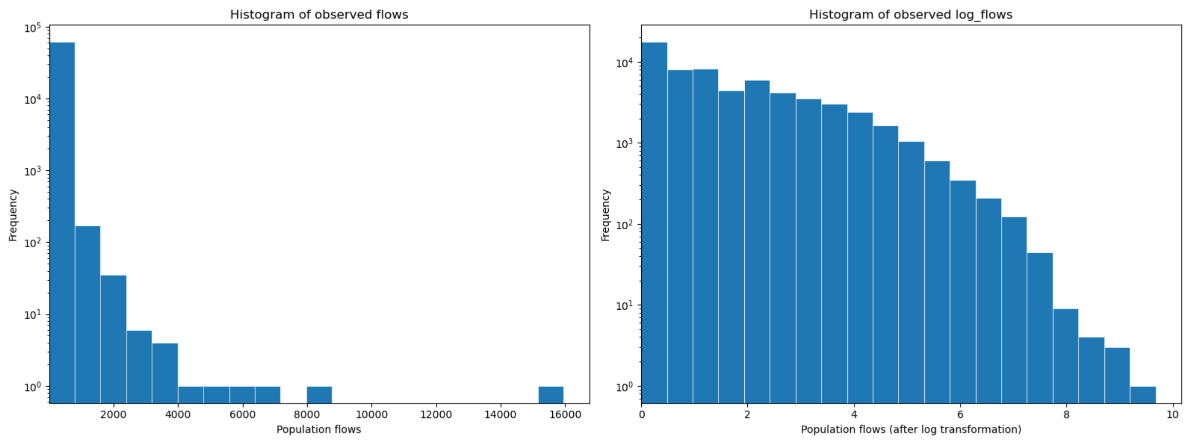

Figure 10 Histogram of traffic flows

And if we plot the log-log relationship between 'flows' and 'distance', the definitive conclusion is hard to be clearly illustrated while the Poisson distribution correlation is seemingly uncertainty. Therefore, we need to compare more than one models including the negative binomial apart from Poisson regression in GLR model.

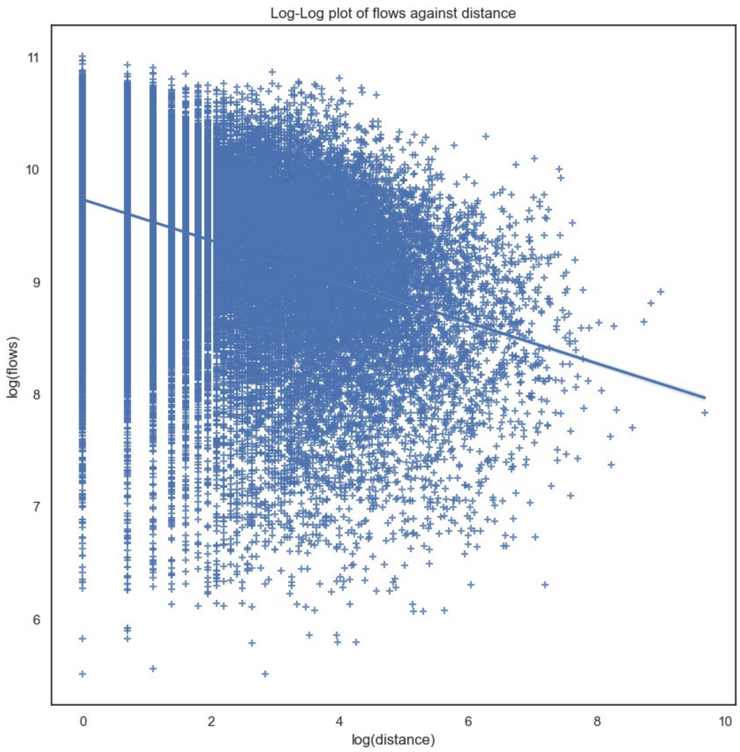

Figure 11 Log-log plot of flows against distance

Another argument against the Poisson regression is that it is based on the assumption that the variable share the similar variance and mean value, which is not perfectly matched even after operating the log transformation in **Table 7**.

Table 7 Mean and Variance for variable 'flows' and 'log_flows'

	'flows'	'log_flows'
Mean	35.09262	**2.027877**
Variance	23804.5416	**2.550937**

2.1.2.2 Comparison

Additionally, the 'distance' variable might also need log-transformation, and whether power or exponential function as the distance delay still need to the determined. The best strategy is just preparing both of them and iteratively fitting the models on various variables and comparing the performance. As for the calibration methods, maximum likelihood estimation (MLE) can be used to estimate parameters (The processing code can be accessed here).

The model's performance is shown in **Table 8**.

Table 8 Summary of GLM model

Models			Negative Binomial		Poisson	
Metrics			R-square	RMSE	R-square	RMSE
flows	Power Decay	Distance	5e-08	9e+04	0.16838*	152.287*
		Logged Dist.	7e-07	2e+67	0.38039	122.17
	Exponential Decay	Distance	0.3795	125.374	0.45590[#]	114.37[#]
		Logged Dist.	5e-08	9e+04	0.16838*	152.287*
Logged flows	Power Decay	Distance	0.0197*	7.201*	0.30449*	1.44*
		Logged Dist.	2e-06	1e+24	0.39358	1.245
	Exponential Decay	Distance	0.51175	1.158	**0.53872[#]**	**1.088[#]**
		Logged Dist.	0.0197*	7.201*	0.30449*	1.44*

*: The Logged distance after exponential decay would be equal to power decay, so they are the same one model.

**: The model's fitting process contains infinity value error or singularity matrix errors.

#: Even though taking logged flows into formula will get best R^2, but it will be difficult to apply the formula backwards to calculate explicit flow value in what-if scenario. Therefore, the traditional one with R^2=0.46 will be utilized in following section.

In conclusion, the exponential decay of distance along with the logged flows in Poisson regression model has the best performance, with the R^2=0.539 and RMSE=1.088 (Direct numeric comparing among RMSEs would be less meaningful due to the re-scaling and logarithm of flow values).

2.2 What-if Scenarios

2.2.1 Scenarios A: Job Decrease at Canary Wharf

Assuming a 50% decrease in jobs at Canary Wharf post-Brexit, which means the attractiveness of Canary Wharf should be halved. In production-constrained model, we could change the jobs of Canary Wharf in dataset from 58772 to 29386. This is based on a valid assumption that there will still be same number of people needing for commuting, while their destination will possibly change according to attractiveness decreasing.

And then, the production constrained model in python will automatically recalculate the flows while conserving the total number of commuters. Since we have constrained the origin station, so the total number of outgoing flows for each origin station is unchanged, thus the sum of all outgoing flows from each origin equals the total population.

Since we need to analysis the whole system's feedback towards Canary Wharf's falling down, so it would be a better way to focus more on the general distribution patterns instead of certain station's changes (there are 398 stations in the dataset, the one-by-one comparison plot would be messy).

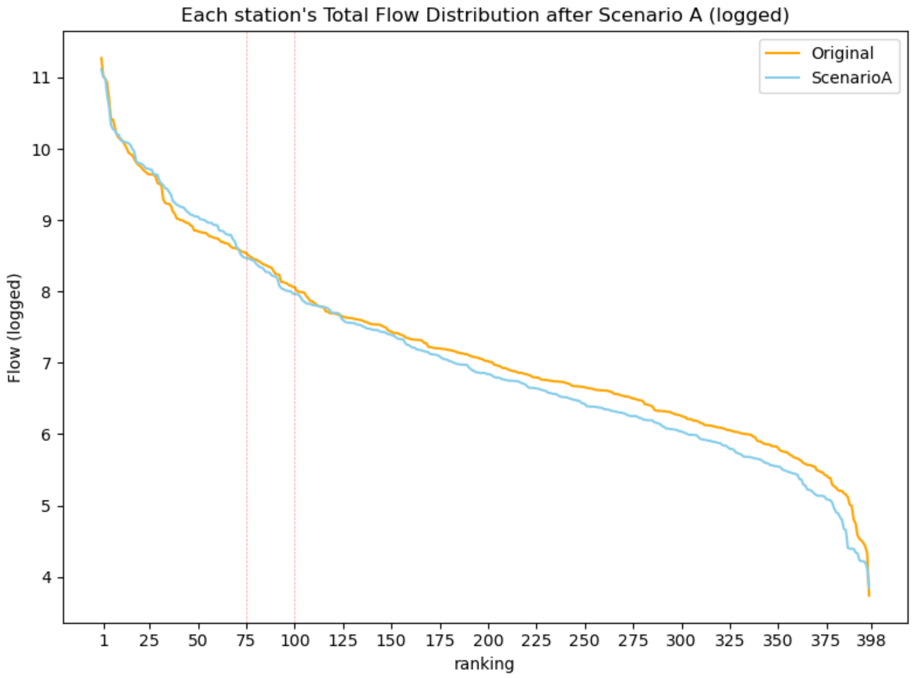

Figure 12 Flowing Distribution Comparison After Scenario A

Therefore, in **Figure 12** we just ignore the names of stations and ranked them in descending order. We could see that the population flow lost by Canary Wharf does not be distributed evenly into all stations. Generally, the busier stations are more likely to pick up those population flow. And contrarily, the less important stations are becoming more deserted with fewer people. Even though the model's prediction might not be reliable, yet it generally aligns with our common sense about the laws in monopolistic market, where the rich get richer and the poor the poorer if any large company bankcrupted. A more interesting observation is that the intersection of the orange and blue lines coincidentally occurs around the 75-100 range, which aligns perfectly with the 'Pareto Principle'.

2.2.2 Scenarios B: Increased Transport Costs

If there is a significant increase in the cost of transport, there are two ways for interpreting that in models. Firstly, the transport cost increasing will discourage people to take public transportation especially for tube, turning to cheaper way for commuting such as bus or e-scooter or bicycles, resulting in less production for all origin stations. Secondly, people will be less likely to commuting in long distance especially for cross-zone higher fare, resulting in being more sensitive to distance parameter.

Therefore, we can interpret the increase in the cost of transport as the increase in the distance parameter or the decrease in the production parameter. The flow's redistribution is illustrated in **Figure 13**, which is generally aligned with the assumption that increase on the distance decay parameters will lead to total population flow's decline.

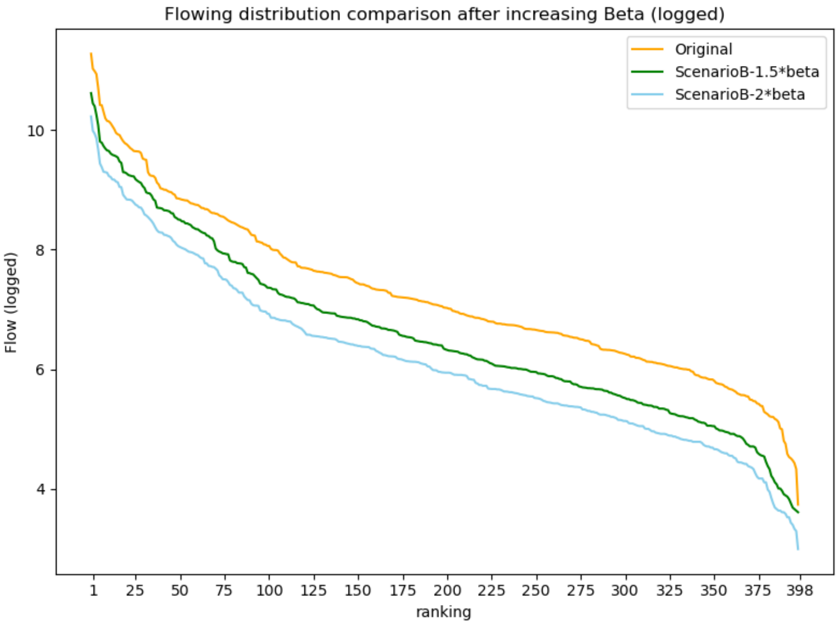

Figure 13 Flow's Redistribution after Increasing Beta (Distance Decay)

Additionally, we could also recalculate the flow by adjusting the alpha parameters, which represents the production ability for each origin station in **Figure 14**. The $\alpha + \ln 0.9$ means production parameter will be decreased to 90 percent of its original value after the exponential function in the formula.

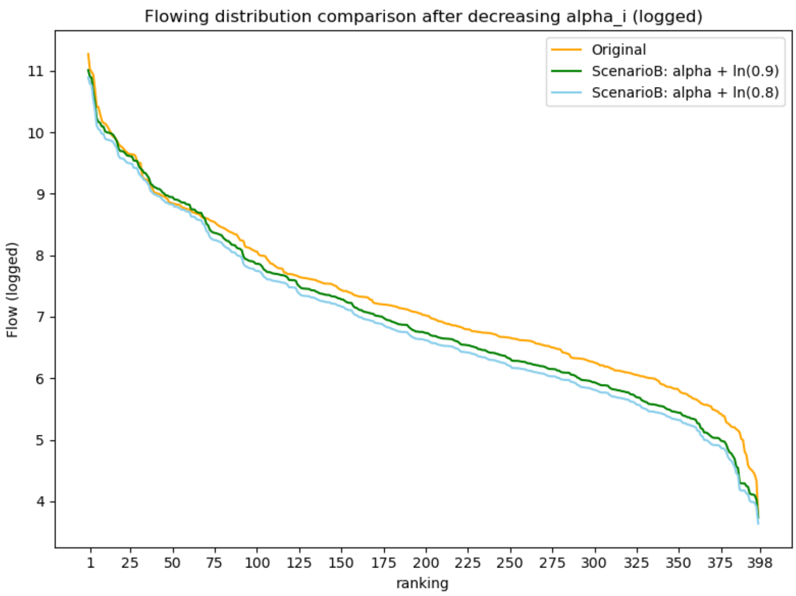

Figure 14 Flow's Redistribution after Decreasing Alpha (Production Ability)

2.2.3 Comparative Analysis

The first pre-assumption must be clarified is that all parameters in Scenario B are difficult to be quantified specifically into the empirical meaning but only for qualitative analysis. For example, the 90 percent of alpha parameter in the formula does not mean the 90 percent of production population for origin stations. Therefore, we could only observe the effect patterns for each scenario in **Figure 15**.

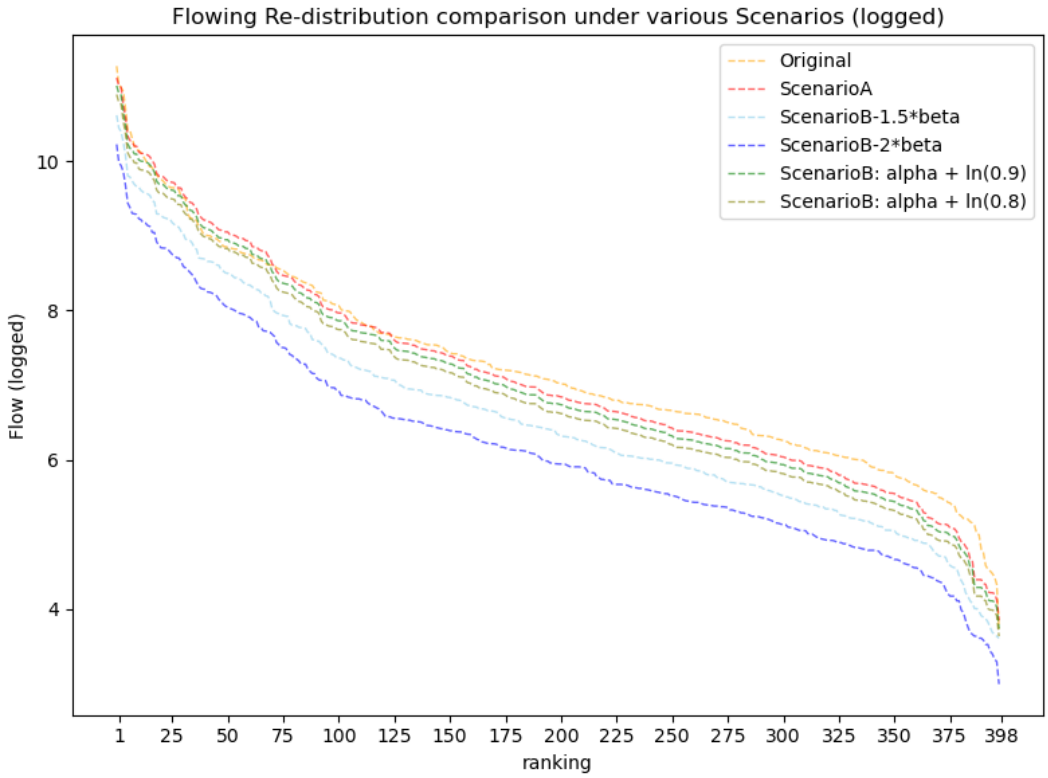

Figure 15 Flow's Redistribution comparison under various Scenarios

In conclusion, among all three scenarios, distance decay (increasing Beta) has more impact in the redistribution of flows, and there will be a significantly reduction of the overall flow whether for a high-ranking station or a low-ranking station (also in **Figure 13**).

And production decline (decreasing alpha) will impact more on low-ranking stations while high-ranking stations remain relatively same level even at 80% as shown in **Figure 14**. Single station's closure might accelerate the polarization between high-ranking and low-ranking stations, potentially along with 80/20 rule.

Reference

Anderson, J. E. (2010). 'The Gravity Model'. National Bureau of Economic Research (Working Paper Series). doi: 10.3386/w16576.

Cochrane, R. A. (1975). 'A Possible Economic Basis for the Gravity Model'. *Journal of Transport Economics and Policy*. [London School of Economics, University of Bath, London School of Economics and University of Bath, London School of Economics and Political Science], 9 (1), pp. 34–49.

Fotheringham, A. S. (1983). 'A New Set of Spatial-Interaction Models: The Theory of Competing Destinations'. *Environment and Planning A: Economy and Space*. SAGE Publications Ltd, 15 (1), pp. 15–36. doi: 10.1177/0308518X8301500103.

Isard, W. (1949). 'The General Theory of Location and Space-Economy'. *The Quarterly Journal of Economics*. Oxford University Press, 63 (4), pp. 476–506. doi: 10.2307/1882135.

Oldham, S., Fulcher, B., Parkes, L., Arnatkevičiūtė, A., Suo, C. and Fornito, A. (2019). 'Consistency and differences between centrality measures across distinct classes of networks'. *PLOS ONE*. Public Library of Science, 14 (7), p. e0220061. doi: 10.1371/journal.pone.0220061.

Qi, X., Fuller, E., Wu, Q., Wu, Y. and Zhang, C.-Q. (2012). 'Laplacian centrality: A new centrality measure for weighted networks'. *Information Sciences*. (Intelligent Knowledge-Based Models and Methodologies for Complex Information Systems), 194, pp. 240–253. doi: 10.1016/j.ins.2011.12.027.

Stewart, J. Q. (1941). 'An Inverse Distance Variation for Certain Social Influences'. *Science*. American Association for the Advancement of Science, 93 (2404), pp. 89–90. doi: 10.1126/science.93.2404.89.

Stouffer, S. A. (1940). 'Intervening Opportunities: A Theory Relating Mobility and Distance'. *American Sociological Review*. [American Sociological Association, Sage Publications, Inc.], 5 (6), pp. 845–867. doi: 10.2307/2084520.

Wan, Z., Mahajan, Y., Kang, B. W., Moore, T. J. and Cho, J.-H. (2021). 'A Survey on Centrality Metrics and Their Network Resilience Analysis'. *IEEE Access*, 9, pp. 104773–104819. doi: 10.1109/ACCESS.2021.3094196.

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. doi: 10.1068/a030001.

Zipf, G. K. (1946). 'The P1 P2/D Hypothesis: On the Intercity Movement of Persons'. *American Sociological Review*. [American Sociological Association, Sage Publications, Inc.], 11 (6), pp. 677–686. doi: 10.2307/2087063.

Appendix

Table 9 Word Count for Sections

<table>
<tr><td rowspan="6">1 London's Underground Resilience</td><td rowspan="3">1.1 Topological Network</td><td>1.1.1 Centrality Measures</td><td>313</td><td rowspan="3">946</td><td rowspan="6">1652</td><td rowspan="11">3143</td></tr>
<tr><td>1.1.2 Impact Measures</td><td>197</td></tr>
<tr><td>1.1.3 Node Removal</td><td>436</td></tr>
<tr><td rowspan="3">1.2 Flows: Weighted Network</td><td>1.2.1 Centrality Measures</td><td>146</td><td rowspan="3">706</td></tr>
<tr><td>1.2.2 Impact Measures</td><td>229</td></tr>
<tr><td>1.2.3 Node Removal Analysis</td><td>331</td></tr>
<tr><td rowspan="5">2 Spatial Interaction Models</td><td rowspan="2">2.1 Models and Calibration</td><td>2.1.1 Introduction to Models</td><td>512</td><td rowspan="2">900</td><td rowspan="5">1491</td></tr>
<tr><td>2.1.2 Parameters Calibration</td><td>388</td></tr>
<tr><td rowspan="3">2.2 What-if Scenarios</td><td>2.2.1 Scenarios A</td><td>260</td><td rowspan="3">591</td></tr>
<tr><td>2.2.2 Scenarios B</td><td>180</td></tr>
<tr><td>2.2.3 Comparative Analysis</td><td>151</td></tr>
</table>

Table 10 Non-sequential Removal Top-10 nodes by Closeness Centrality

	Removal Node	Node's Centrality	Global Efficiency	Change on Global Efficiency	Modularity	Change on Modularity
0			**0.10126**	**NA**	**0.83021**	**NA**
1	Green Park	0.1148	0.09919	-0.00207	0.83402	+0.00381
2	Bank&Monument	0.1136	0.09487	-0.00432	0.84388	+0.00986
3	King's Cross St. Pancras	0.1134	0.08793	-0.00694	0.85017	+0.00629
4	Westminster	0.1125	0.08737	-0.00056	0.84858	-0.00159
5	Waterloo	0.1123	0.08486	-0.00251	0.8502	+0.00162
6	Oxford Circus	0.1112	0.08278	-0.00208	0.85484	+0.00464
7	Bond Street	0.111	0.08258	-0.0002	0.85983	+0.00499
8	Farringdon	0.1107	0.08261	+0.00003	0.86079	+0.00096
9	Angel	0.1107	0.08262	+0.00001	0.86025	-0.00054

10	Moorgate	0.1103	0.08167	-0.00095	0.8583	-0.00195

Table 11 Non-sequential Removal Top-10 nodes by Betweenness Centrality

	Removal Node	Node's Centrality	Global Efficiency	Change on Global Efficiency	Modularity	Change on Modularity
0			**0.10126**	**NA**	**0.83021**	**NA**
1	Stratford	0.298	0.08892	-0.01234	0.83571	+0.0055
2	Bank and Monument	0.290	0.08586	-0.00306	0.84676	+0.01105
3	Liverpool Street	0.271	0.08496	-0.0009	0.85214	+0.00538
4	King's Cross St. Pancras	0.255	0.07849	-0.00647	0.85317	+0.00103
5	Waterloo	0.244	0.07594	-0.00255	0.85951	+0.00634
6	Green Park	0.216	0.07415	-0.00179	0.85988	+0.00037
7	Euston	0.208	0.06821	-0.00594	0.86629	+0.00641
8	Westminster	0.203	0.06766	-0.00055	0.86589	-0.0004
9	Baker Street	0.192	0.06470	-0.00296	0.87323	+0.00734
10	Finchley Road	0.165	0.06314	-0.00156	0.87671	+0.00348

Table 12 Non-sequential Removal Top-10 nodes by Information Centrality

	Removal Node	Node's Centrality	Global Efficiency	Change on Global Efficiency	Modularity	Change on Modularity
0			**0.10126**	**NA**	**0.83021**	**NA**
1	Bank&Monument	0.0005983	0.09673	-0.00453	0.83798	+0.00777
2	King's Cross	0.0005906	0.08988	-0.00685	0.84551	+0.00753
3	Liverpool Street	0.0005865	0.08744	-0.00244	0.84885	+0.00334
4	Oxford Circus	0.0005862	0.08598	-0.00146	0.85769	+0.00884
5	Greed Park	0.0005847	0.0831	-0.00288	0.85497	-0.00272
6	Waterloo	0.0005788	0.08113	-0.00197	0.85848	+0.00351
7	Baker Street	0.0005786	0.0788	-0.00233	0.86338	+0.0049

8	Bond Street	0.0005726	0.07884	+0.00004	0.86471	+0.00133
9	Stratford	0.0005722	0.06999	-0.00885	0.86762	+0.00291
10	Moorgate	0.0005705	0.0703	0.00031	0.86909	+0.00147

Table 13 Sequential Removal Top-10 nodes by Closeness Centrality

	Removal Node	Node's Centrality	Global Efficiency	Change on Global Efficiency	Modularity	Change on Modularity
0			**0.10126**	**NA**	**0.83021**	**NA**
1	Green Park	0.11478	0.09919	-0.00207	0.83403	0.00382
2	King's Cross St. Pancras	0.11236	0.09443	-0.00476	0.84241	0.00838
3	Waterloo	0.10465	0.09182	-0.00261	0.84603	0.00362
4	Bank & Monument	0.09742	0.08543	-0.00639	0.85148	0.00545
5	West Hampstead	0.08173	0.08054	-0.00489	0.85461	0.00313
6	Canada Water	0.07598	0.0581	-0.02244	0.85891	0.0043
7	Stratford	0.06373	0.05188	-0.00622	0.86943	0.01052
8	Earl's Court	0.06364	0.05035	-0.00153	0.87051	0.00108
9	Sheperd's Bush	0.06043	0.04439	-0.00596	0.87785	0.00734
10	Oxford Circus	0.0505	0.04296	-0.00143	0.88236	0.00451

Table 14 Sequential Removal Top-10 nodes by Betweenness Centrality

	Removal Node	Node's Centrality	Global Efficiency	Change on Global Efficiency	Modularity	Change on Modularity
0			**0.10126**	**NA**	**0.83021**	**NA**
1	Stratford	0.29785	0.08892	-0.01234	0.83571	0.0055
2	King's Cross St. Pancras	0.24726	0.0846	-0.00432	0.84501	0.0093
3	Waterloo	0.25418	0.08183	-0.00277	0.85645	0.01144
4	Bank and Monument	0.21465	0.07768	-0.00415	0.86075	0.0043
5	Canada Water	0.2449	0.07283	-0.00485	0.86425	0.0035

6	West Hampstead	0.45683	0.05321	-0.01962	0.86982	0.00557
7	Earl's Court	0.09618	0.5166	0.46339	0.87119	0.00137
8	Sheperd's Bush	0.12885	0.04584	-0.47076	0.88107	0.00988
9	Euston	0.08708	0.04163	-0.00421	0.88125	0.00018
10	Baker Street	0.09844	0.03816	-0.00347	0.88793	0.00668

Table 15 Sequential Removal Top-10 nodes by Information Centrality

	Removal Node	Node’s Centrality	Global Efficiency	Change on Global Efficiency	Modularity	Change on Modularity
0			**0.10126**	**NA**	**0.83021**	**NA**
1	Bank and Monument	0.0005983	0.09673	-0.00453	0.83798	0.00777
2	King's Cross St. Pancras	0.0005693	0.08988	-0.00685	0.84551	0.00753
3	Green Park	0.0004761	0.08793	-0.00195	0.85017	0.00466
4	Baker Street	0.000456	0.08436	-0.00357	0.85577	0.0056
5	Canada Water	0.0004407	0.07711	-0.00725	0.8633	0.00753
6	West Hampstead	0.0002965	0.05575	-0.02136	0.86665	0.00335
7	Earl's Court	0.0006428	0.05413	-0.00162	0.86842	0.00177
8	Shepherd's Bush	0.0005907	0.04849	-0.00564	0.87313	0.00471
9	Turnham Green	0.0005748	0.04256	-0.00593	0.87672	0.00359
10	Stratford	0.0014937	0.03616	-0.0064	0.88803	0.01131