Part 1: London's underground resilience

I. Topological network

I.1. Centrality measures

The idea of centrality is fundamental to network analysis (Negre *et al.*, 2018). In this study of London underground network, harmonic centrality, betweenness centrality and eigenvector centrality are used for ranking the most influential stations.

Harmonic centrality was developed on the basis of closeness centrality (Marchiori and Latora, 2000), aiming at overcome the "the cases where an infinite distance outweighs the others". Due to this advantage, it can be used in the disconnected network, and in the following analysis of node removal, such a disconnected network will appear in the context of the London Underground network.

$$C_H = \frac{1}{n-1}\sum_{j\neq i}\frac{1}{d_{ij}} \tag{1}$$

The betweenness centrality measure (Freeman, 1977) evaluates how frequently a particular node serves as a link in the shortest path connecting two other nodes (Bloch, Jackson and Tebaldi, 2023), giving more values to vertex through whom flows are more likely to pass.

$$C_B(g) = \frac{2}{(n-1)(n-2)}\sum_{(j,k),j\neq i,k\neq i}\frac{v_g(i{:}j,k)}{v_g(j,k)} \tag{2}$$

where n is the overall nodes in the graph, j and k are a pair of nodes passing through i in different paths.

Eigenvector centrality captures the influence of the neighbour of the node and add it to the node itself. In the iteration of calculation, a node is self-referential but has a well-defined fixed point (Bloch, Jackson and Tebaldi, 2023). Proximity to a major station in the London tube network raises the significance of an observing station.

$$x_i = \frac{1}{\lambda}\sum x_j = \frac{1}{\lambda}\sum a_{i,j}, x_j \qquad Ax = \lambda \mathrm{x} \tag{3}$$

Table 1. First 10 ranked stations for three centrality measures

rank	Harmonic Centrality	Betweenness Centrality	Eigenvector Centrality
1	Bank and Monument	Stratford	Bank and Monument
2	Liverpool Street	Bank and Monument	Liverpool Street
3	Stratford	Liverpool Street	Stratford
4	Waterloo	King's Cross St. Pancras	Waterloo
5	Green Park	Waterloo	Moorgate
6	King's Cross St. Pancras	Green Park	Green Park
7	Moorgate	Euston	Oxford Circus
8	Oxford Circus	Westminster	Tower Hill
9	Westminster	Baker Street	Westminster
10	Baker Street	Finchley Road	Shadwell

All three measurements identify the same top three stations (table 1). Betweenness centrality assigns a higher ranking to stations with more traffic like King's Cross St. Pancras and Euston.

I.2. Impact measures

When research on the impact of node removal and network resilience, the connectivity is the "core element in the recognition of the evolution of resilience/vulnerability states" (Reggiani, Nijkamp and Lanzi, 2015). To measure the connectivity and quantify the impact of node removal in the topological tube network, this study the following two impact measurement:

Network efficiency (*Ef)* is usually used as an indicator of network connectivity (Zhang *et al.*, 2018). This measurement still works when the graph is disconnected. *Ef* indicates how direct the connections are between all node pairs (Mattsson and Jenelius, 2015).

$$E_f = \frac{1}{N(N-1)} \sum_{j \neq i} \frac{1}{d_{ij}} \tag{4}$$

The normalised size of the largest connected component (S) is an indicator to show how the tube network evolves as it separated into different graphs when deleting the most important nodes from the network (Leskovec and Faloutsos, 2006) (Berche *et al.*, 2009). Normalisation makes it able to make comparison between different scenarios.

$$S = \frac{N_1}{N} \tag{5}$$

I.3. Node removal

To measure the importance of each node, stations are removed follow the rank of nodes in each centrality measures. In the process of non-sequence removal, *Ef* and S are both dropping quickly. In theory, nodes with higher centrality rankings should exert a greater influence on the network as a whole. Figure 1 show that the value decline most rapidly for nodes removed based on betweenness centrality, followed by harmonic centrality and then eigenvector centrality. These findings suggest that betweenness centrality is the most effective at capturing the overall network structure, while eigenvector centrality is the least effective.

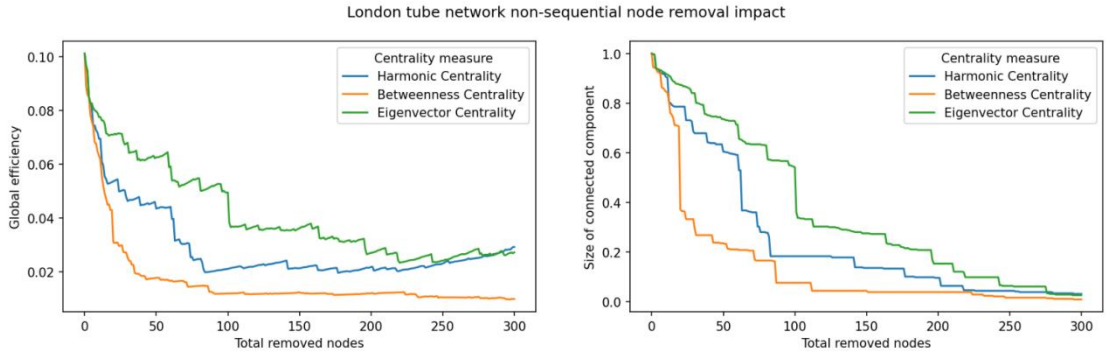

Figure 1. London topological tube network non-sequential removal impact

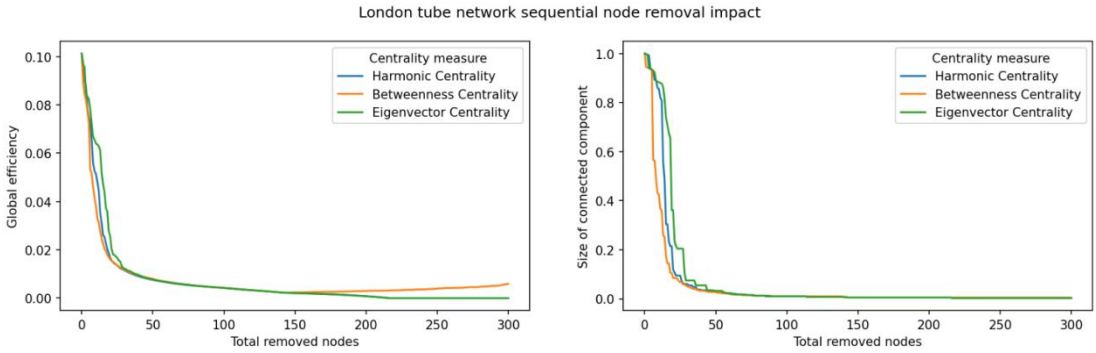

Figure 2. London topological tube network sequential removal impact

While the three centrality measures experience almost the same trend during sequential node removal, it is evident that betweenness centrality declines the fastest when the first 50 nodes are removed. Comparing the two node removal approaches, it is apparent that recalculating and deleting the most significant node in the new graph after each removal results in a more significant attack on the network. Overall, betweenness centrality proves to be the most effective of the three centrality measures in identifying the centrality of the London Underground network.

Comparing the two impact measures, *Ef* and S, they performed consistently in sequential removal for assessing attack; however, S was less sensitive to the identification of damage in non-sequential removal. Overall, *Ef* was the superior way to assess network hazards.

II. Flows: weighted network

II.1 Centrality measure

This section incorporates data on the flows between stations for analysis, where larger flows indicate a stronger connection between the two stations. As such, the inverse flows can serve as a distance parameter in the equation. While betweenness centrality is most effective in identifying node importance in topological networks, in weighted networks, the reciprocal of the flows (1/flows) can be assigned as distance to compute the shortest path between each pair of stations and generate a new centrality ranking.

Table2. Betweenness centrality ranked stations for the weighted underground network

rank	**Centrality rank in topological network**	**Centrality rank in weighted network**
1	Stratford	Green Park
2	Bank and Monument	Bank and Monument
3	Liverpool Street	Waterloo
4	King's Cross St. Pancras	Westminster
5	Waterloo	Liverpool Street
6	Green Park	Stratford
7	Euston	Euston
8	Westminster	Victoria
9	Baker Street	Oxford Circus
10	Finchley Road	Bond Street

With the inclusion of flows, Green Park has emerged as the top-ranking station, with Westminster Station, located nearby, experiencing a significant rise in rank. On the other hand, stations such as Stratford and Liverpool Street, which are vital for transport and connectivity, have experienced a drop in their ranking.

II.2 Impact measures

Efficiency *Ef* is able to caputure the distance in a weighted network (formula 4). In topological network above, every edge is assigned the distance value 1. While in this part, the distance will be the inverse flow.

The S parameter mentioned in I.2 is more suitable for observing global changes after removing a large number of nodes. In order to study local variations, the average shortest path length (L) can be introduced as a parameter to calculate and normalise the shortest path between any pair of nodes i and j. Lower value means a faster network (Mattsson and Jenelius, 2015).

$$L = \frac{1}{N(N-1)} \sum_{j \neq i} d_{ij} \tag{6}$$

II.3 Node removal:

The top three nodes are Green Park, Bank and Monument, and Waterloo after recalculating betweenness centrality. To determine which node has the greatest impact on the weighted network, two node removal methods in I.3 and two impact measures mentioned in section II.2 are used for comparison. During sequential removal, the network's highest betweenness centrality belongs to King's Cross St. Pancras after deleting two nodes. Thus a comparison is made between these four stations.

In the graph, observe the increasing slope of L and decreasing slope of *Ef*. The two slope are greatest when removing Bank and Monument, followed by Green Park and least by Waterloo in non-sequential removal (figure 3). During sequential removal (figure 4), King's Cross St. Pancras is found to be the most significant for the increase in L, while it has the same effect on *Ef* as Bank and Monument. To sum up, closing the King's Cross St. Pancras station would have the greatest impact on both efficiency and commuting distance, making it the station that has the most significant impact on passengers.

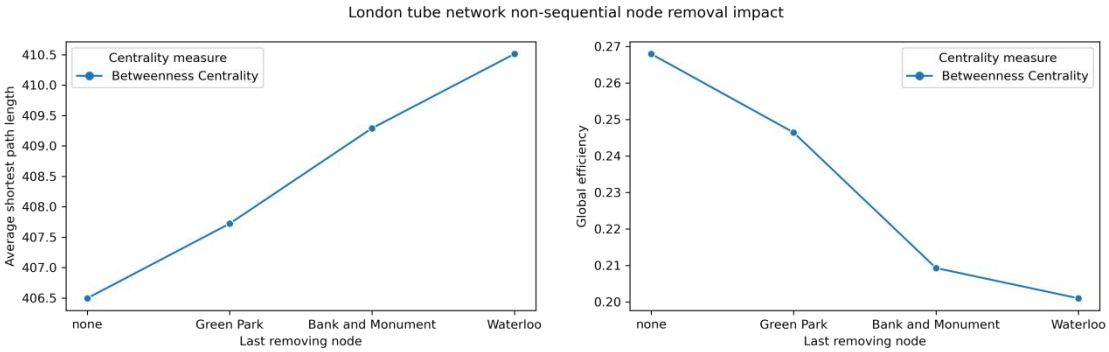

Figure 3. Weighted underground network non-sequential removal impact

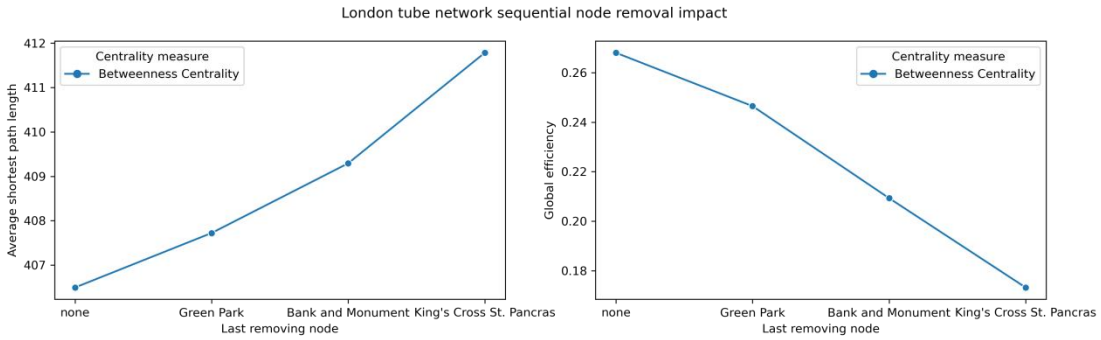

Figure 4. Weighted underground network sequential removal impact

Part 2: Spatial Interaction models

III. Models and calibration

III.1. Introduction

The spatial interaction models, proposed and developed by Wilson in 1970 (Wilson, 1971), quantify the interaction between urban areas using the geographical gravity model, which was first identified through an analogy to Newton's law of universal gravitatio. From the 1950s, social scientists and urban planners became increasingly interested in the gravity model for quantifying urban theories, leading to its expansion from analyzing human group interactions and migration to broader social science applications (Carrothers, 1956). Wilson combined two scenarios in formula (7) and (8) and expanded the interaction model to include a family of spatial interaction models, including:

$$\sum_j T_{ij} = O_i \tag{7}$$

$$\sum_i T_{ij} = D_j \tag{8}$$

(1) Unconstrained model (both unknown):

$$T_{ij} = KW_i^{(1)}W_j^{(2)}f(c_{ij}) \tag{9}$$

(2) production constrained model (total flow of origin is known):

$$T_{ij} = A_iO_iW_j^{(2)}f(c_{ij}) \tag{10}$$

(3) Attraction constrained model (total flow of destination is known):

$$T_{ij} = B_jW_i^{(1)}D_jf(c_{ij}) \tag{11}$$

(4) Doubly constrained model (both known):

$$T_{ij} = A_iO_iB_jD_jf(c_{ij}) \tag{12}$$

Tij measures the interaction between zones i and j, while Wi(1) and Wj(2) represent the 'mass term' for each zone. The distance between the zones is measured by cij, and K is a constant while n is a parameter. The function f(cij) decreases as cij increases.

III.2. Select and calibrate the model

In this study, the travel flows during the morning peak time in London tube network can be specified as a home-work commute problem. The attractiveness of destination is captured by the number of jobs while the production side is the origin tube stations where people live.

In this study, production constrained model (formula 10) will be used because in the next part of job decrease scenario, the flows to the destination (jobs) will change, leaving the flows from the origin (home) be the same as usual.

Model calibration plays an important part in the simulation process. The f(cij) in the spatial interaction model can be usually generalised by the inverse power function or the

negative exponential function in the studies on complexity (Chen, 2015). Based on the formula and concept of entropy-maximization, Wilson developed exponential function of the distance (Wilson, 2013). This study uses inverse power function (formula 13) and negative exponential function (formula 14), compares their effects on model accuracy and calculates different β values in the cost function.

$$T_{ij} = A_i O_i^{\alpha} D_j^{\gamma} d_{ij}^{-\beta} \quad \rightarrow \quad \lambda_{ij} = \exp(\alpha_i + \gamma \ln D_j - \beta \ln d_{ij}) \tag{13}$$

$$T_{ij} = A_i O_i^{\alpha} D_j^{\gamma} \exp(-\beta d_{ij}) \rightarrow \quad \lambda_{ij} = \exp(\alpha_i + \gamma \ln D_j - \beta d_{ij}) \tag{14}$$

The London Underground commuting data contains flows, population, jobs and distance data for origin and destination tube stations. After examining the distribution of the data, it was found that the flows data conformed to a Poisson distribution and were suitable for processing using logarithm. The two cost functions were used in the production constrained model to simulate the London Underground commuting data and the performance of the model was constructed as shown in the table below.

Table 3. Performance of production constrained model applied with different distance decay function

Distance decay function	**R2 of the model**	**RMSE of the model**
Inverse power	0.39	102.89
Negative exponential	0.47	96.26

The results reveal that using the negative exponential function improves the fitness of the OD flow model for the London Underground network, with higher R^2 and lower RMSE values. Commuters in the London tube network prioritize travel time over station distance due to the tube's faster pace, resulting in reduced sensitivity to distance traveled. Therefore, both data and realistic analysis support using the production constrained model with the exponential function (formula 9) for this study.

IV. Scenarios

IV.1. Scenario A:

As previously stated, the production constrained model controls for Oi. To determine the value of Ai for each origin station based on the original flows, a Poisson distribution is employed, with the value of β (1.53e-4). If the number of jobs in Canary Wharf decreases by 50%, then Dj in the formula will change. A change in Dj results in changes in Ai (formula 15). Therefore, simply substituting the new value of Dj into the formula is insufficient. Instead, Ai needs to be recalculated based on the changed value of Dj. In this way, the number of commuters flowing out of each origin station is fixed, and the total number of commuters in the whole system is conserved.

$$A_i = \frac{1}{\sum_j D_j^{\gamma} d_{ij}^{-\beta}} \tag{15}$$

Figure 1 shows the overall change of origin flows in each London tube stations after a sharp job decrease happened in Canary Wharf. Some of the stations as the origin of trips experience flow increase (red colour), most of which rise less than 2000 flows. The four

stations with the largest increases in the number of flows exceeded 4,000, with the highest increase occurring at Stratford, which saw an increase of over 10,000.
Other stations experience flow decrease instead (blue colour), with canary wharf station experiencing the largest decrease of almost 30,000. Other stations experiencing a decrease of around 4,000 were Bank and Monument, Hammersmith, Oxford Circus, Victoria, and Wimbledon.

Figure 5. The change of flows in the destination stations after the scenario 1

The projections suggest that when the number of jobs in canary wharf is significantly reduced, there will be a strong impact on traffic volumes across the network. Passengers who used to commute out to canary wharf will go to other locations with more job opportunities and better transport links, such as Stratford and Whitechapel. Most of this redistributed additional traffic will be concentrated in the centre of the city, which is consistent with people commuting in search of work.

IV.2. Scenario B:
As the distance parameter controlling the effect of the generalized travel cost, β gets larger if the negative effect of the distance gets higher. In the scenario of tube cost rising, commuters will tend to take less tube and travel closer to save the money. The long travel distance will be a bigger obstacle for the cost issue. Thus, to simulate a scenario that there is a significant increase in the travel cost, this study chooses double and five times of the original β value as the new friction parameter.

For better comparison, production constrained model will still be used in this scenario. Controlling the total amount of out flows from the origin stations, the flows will be redistributed to different new destination stations.

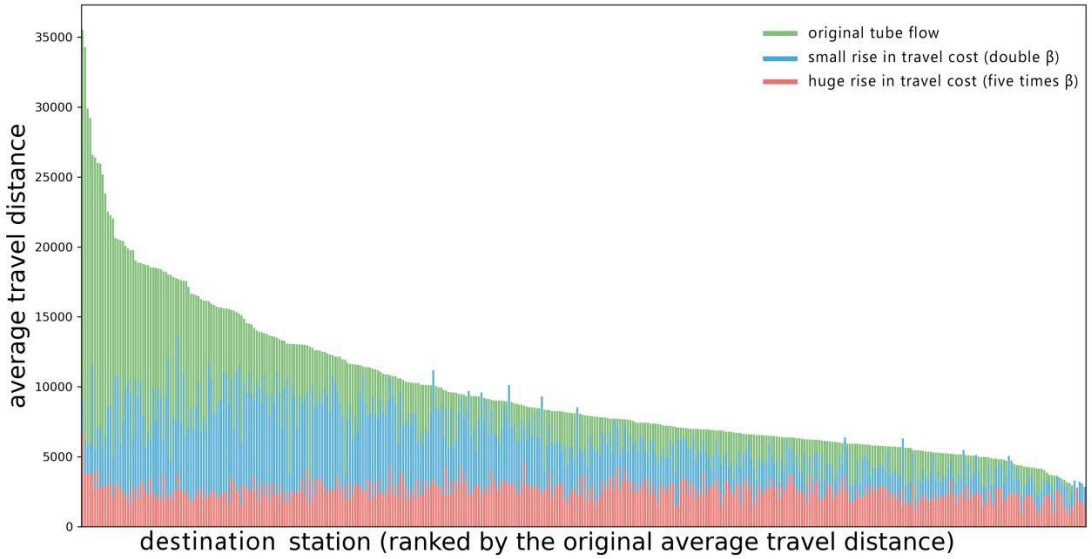

Figure 6. The average travel distances of each destination station after different degree of cost increase

Since the total flow from origin is controlled, the results of the redistribution of the flow show different degrees of increase and decrease in traffic at different destination stations. To make the impact of rising transport costs on the overall commuting network clearer, this study presents average travel distance from each origin station (figure 6). Sorting the data by original average travel distance, the closer to the left station in the figure the further away the commute is. It can be seen that when the price of travel increases, the average travel distance decreases throughout the system. When transport costs increase by a small amount, it has a greater impact on stations with longer commuting distances, compared to the shorter commuting stations on the right side of the graph. When there is a further large increase in transport costs, the average travel distance drops to an even level for all stations, with no more than 5000.

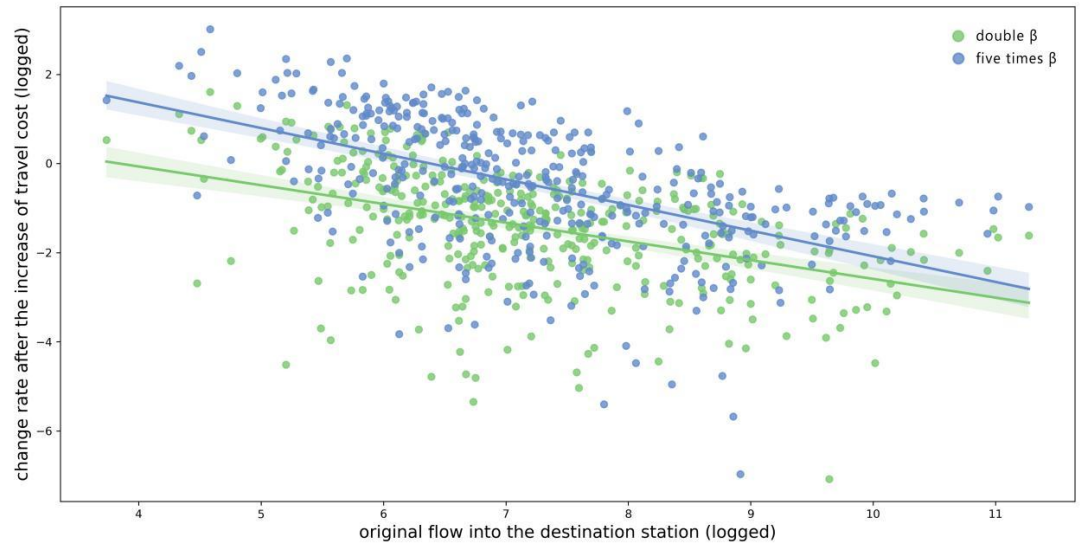

Figure 7. The relationship between the flow change rate and the original flows

Stations with high patronage tend to have more jobs, higher salaries, or have a significant transport hub status. To further investigate whether transport fare increases have different effects on stations with different patronage, this study analysed the relationship between the rate of change in traffic and the original traffic (figure 7).
The results show that the larger the original station, the smaller the reduction in patronage in the presence of a fare increase. This suggests that larger stations are more firmly established and important in commuting. This phenomenon is even more pronounced when the fare increase is changed from two to five times.

IV.3. Discuss
The origin constrained model is used to simulate and calculate how the flows redistributed for all three scenarios: scenario A, scenario B1 and scenario B2. Therefore, in order to compare the impact of these three scenarios on the overall OD network traffic, we only need to observe The flow redistribution at each destination station.

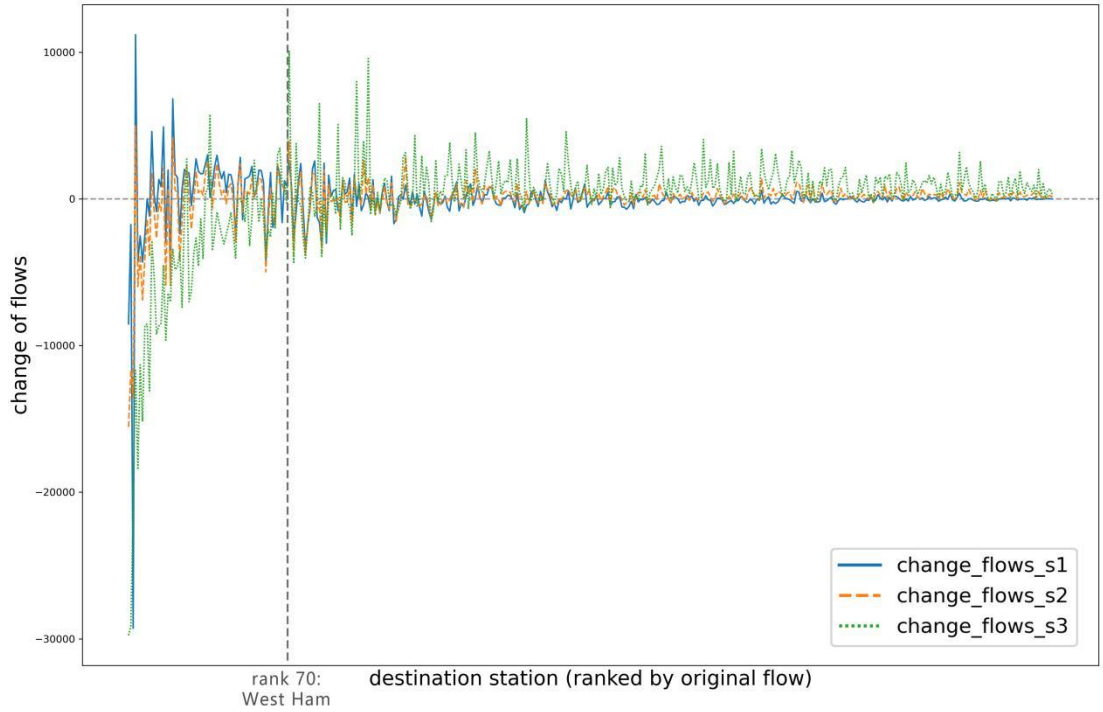

Figure 8. Change of flows in three scenarios

To begin with, we examine the absolute values of flow changes (Figure 8). It is reasonable to observe that all scenarios show fluctuations in flow changes. Overall, scenario 1 displays less volatility compared to scenario 2, and both are less volatile than scenario 3, particularly for stations ranked to the right of 70. Stations with high original flows are also subject to greater changes in flow when conditions shift, which explains the gradual decrease of the pendulum from left to right in the graph.
In Scenario 1, the significant changes in flow are caused by larger stations, and following a sudden drop in flow at Canary Wharf, other large stations experience a notable increase in flow, as commuters seek work elsewhere. In contrast, Scenario 2 and Scenario 3 alter the flows at every station in the system, with both scenarios having a similar impact on

different stations, albeit with Scenario 3 having a greater effect. In both cases, traffic at larger stations decreases significantly, while traffic at smaller stations increases.

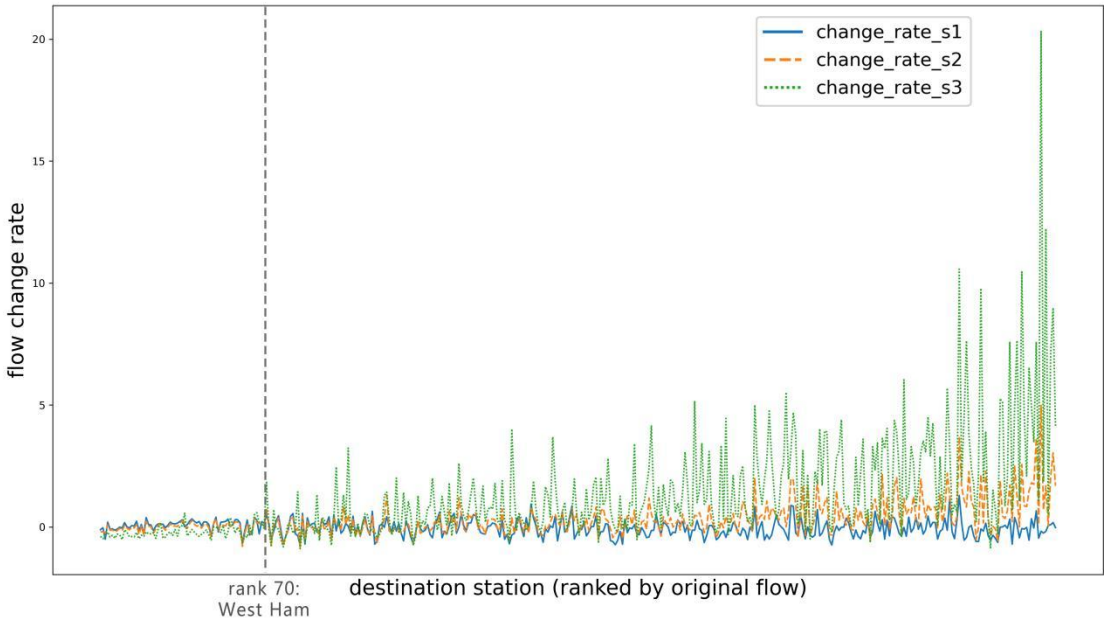

Figure 9. Change rate of flow in three scenarios

Figure 5 displays the traffic growth rate for each station, calculated using formula:

$$R_i = (N_i - O_i)/O_i \quad (16)$$

where i represents the destination station, R is the change rate of the station, N is the predicted new flow in the different scenario, and O is the original flow.

Table 4. Sum of station change rate

	Scenario 1	Scenario 2	Scenario 3
Sum of all station's change rate	-14.50	126.76	503.44

In summary, the change rate of small and medium-sized stations is most affected by scenario3 and scenario2, while scenario1 has a more significant impact on large stations. The analysis of flow changes in absolute and relative values suggests that scenario1 leads to a significant redistribution of traffic at large stations, while scenario2 and scenario3 have a global impact, particularly on the redistribution of traffic at small and medium-sized stations. Overall, scenario3 has the largest impact on all stations combined (table 4).

Word count: 2934

Reference

Berche, B. *et al.* (2009) 'Resilience of public transport networks against attacks', *The European Physical Journal B*, 71(1), pp. 125–137. Available at: https://doi.org/10.1140/epjb/e2009-00291-3.

Bloch, F., Jackson, M.O. and Tebaldi, P. (2023) 'Centrality measures in networks', *Social Choice and Welfare* [Preprint]. Available at: https://doi.org/10.1007/s00355-023-01456-4.

Carrothers, G.A.P. (1956) 'An Historical Review of the Gravity and Potential Concepts of Human Interaction', *Journal of the American Institute of Planners*, 22(2), pp. 94–102. Available at: https://doi.org/10.1080/01944365608979229.

Chen, Y. (2015) 'The distance-decay function of geographical gravity model: Power law or exponential law?', *Chaos, Solitons & Fractals*, 77, pp. 174–189. Available at: https://doi.org/10.1016/j.chaos.2015.05.022.

Freeman, L.C. (1977) 'A Set of Measures of Centrality Based on Betweenness', *Sociometry*, 40(1), p. 35. Available at: https://doi.org/10.2307/3033543.

Leskovec, J. and Faloutsos, C. (2006) 'Sampling from large graphs', in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. *KDD06: The 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Philadelphia PA USA: ACM, pp. 631–636. Available at: https://doi.org/10.1145/1150402.1150479.

Marchiori, M. and Latora, V. (2000) 'Harmony in the small-world', *Physica A: Statistical Mechanics and its Applications*, 285(3–4), pp. 539–546. Available at: https://doi.org/10.1016/S0378-4371(00)00311-3.

Mattsson, L.-G. and Jenelius, E. (2015) 'Vulnerability and resilience of transport systems – A discussion of recent research', *Transportation Research Part A: Policy and Practice*, 81, pp. 16–34. Available at: https://doi.org/10.1016/j.tra.2015.06.002.

Negre, C.F.A. *et al.* (2018) 'Eigenvector centrality for characterization of protein allosteric pathways', *Proceedings of the National Academy of Sciences*, 115(52). Available at: https://doi.org/10.1073/pnas.1810452115.

Reggiani, A., Nijkamp, P. and Lanzi, D. (2015) 'Transport resilience and vulnerability: The role of connectivity', *Transportation Research Part A: Policy and Practice*, 81, pp. 4–15. Available at: https://doi.org/10.1016/j.tra.2014.12.012.

Wilson, A. (2013) *Entropy in Urban and Regional Modelling (Routledge Revivals)*. 0 edn. Routledge. Available at: https://doi.org/10.4324/9780203142608.

Wilson, A.G. (1971) 'A Family of Spatial Interaction Models, and Associated Developments', *Environment and Planning A: Economy and Space*, 3(1), pp. 1–32. Available at: https://doi.org/10.1068/a030001.

Zhang, D. *et al.* (2018) 'Resiliency assessment of urban rail transit networks: Shanghai metro as an example', *Safety Science*, 106, pp. 230–243. Available at: https://doi.org/10.1016/j.ssci.2018.03.023.