Urban Simulation Report

Student number: 23083053

1 Part 1: London's Underground Resilience

The first part of this report addresses the resilience of the London Underground using network analysis.

1.1 Topological Network

In this section, the topological network of the London Underground is analysed, without considering the passenger flows. The weight of each edge is the geographic distance of lines between station pairs.

1.1.1 Centrality Measures

Centrality measures are characteristics of nodes indicating their significance in various aspects. In this report, the measures used to identify significant stations are: degree centrality, betweenness centrality, and closeness centrality. The number of nodes within the network is denoted by n, and the number of links between nodes i and j as A_{ij}.

Degree centrality is the number of links that are connected to each node. The degree centrality k_i for node i is calculated as

$$k_i = \sum_j A_{ij} \tag{1}$$

In the context of the underground network, the degree corresponds to the number of directions that serves the station. Since our dataset does not double count two lines serving the same two station pairs, this is equivalent to the number of adjacent stations.

Table 1 shows stations with the highest degree centrality. A high degree centrality indicates there are many lines serving the station, quantifying the importance as a transit hub allowing for transfers.

Table 1: Stations with the highest degree centrality. All stations tied at a degree of 6 are listed.

Rank	Station	Degree
1	Stratford	9

Rank	Station	Degree
2	Bank and Monument	8
3	Baker Street	7
3	King's Cross St. Pancras	7
5	Earl's Court	6
5	Canning Town	6
5	Liverpool Street	6
5	Green Park	6
5	Oxford Circus	6
5	Waterloo	6
5	West Ham	6

Betweenness centrality of a node is the number of shortest paths that run through the node. The betweenness centrality x_i for node i is calculated as

$$x_i = \sum_{s,t} \frac{n^i_{st}}{g_{st}} \tag{2}$$

where

$$n^i_{st} = \begin{cases} 1 & (\text{if node } i \text{ is on geodesic from } s \text{ to } t) \\ 0 & (\text{otherwise}) \end{cases}$$

and g_{st} is the total number of geodesic paths from s to t. This can be normalised by dividing by $(n-1)(n-2)/2$ (combinations between remaining nodes) to obtain a range between 0 and 1. A station with high betweenness centrality is on the shortest path between many origin-destination pairs of stations, therefore many people are expected to pass it, indicating its importance. The stations with the highest betweenness centrality are shown in Table 2.

Table 2: Stations with the highest betweenness centrality.

Rank	Station	Betweenness Centrality
1	Bank and Monument	17,625
2	King's Cross St. Pancras	16,716
3	Stratford	14,563
4	Baker Street	13,180
5	Oxford Circus	12,573
6	Euston	12,345
7	Earl's Court	11,452
8	Shadwell	11,127
9	Waterloo	10,425
10	South Kensington	10,302

Closeness centrality is the inverse of the mean geodesic distance from one node to other reachable n_0 nodes (l_i) in the same connected component. This value is scaled by the ratio of reachable nodes to account for connectedness, as proposed by Wasserman and Faust (1994). The closeness centrality C_i is calculated as

$$C_i = \frac{n_0 - 1}{n - 1} \cdot \frac{1}{l_i} = \frac{(n_0 - 1)^2}{(n - 1)\sum_j d_{ij}} \tag{3}$$

where d_{ij} is the geodesic distance between nodes i and j.

A high closeness centrality indicates the station is within a short distance from all the other stations. These stations are located within the city centre, as the high-ranked stations clustered in central London shown in Table 3.

Table 3: Stations with the highest closeness centrality. Charing Cross is where the 'distance to London' is measured from (Leatherdale, 2016) - supporting that these stations are located in the city centre.

Rank	Station	Closeness Centrality [$\times 10^{-5}$]
1	Holborn	7.924
2	King's Cross St. Pancras	7.907
3	Tottenham Court Road	7.888
4	Oxford Circus	7.873
5	Leicester Square	7.835
6	Picadilly Circus	7.831
7	Charing Cross	7.830
8	Chancery Lane	7.823
9	Covent Garden	7.806
10	Embankment	7.799

1.1.2 Impact Measures

For measuring the impact of node removal, we have used the delta centrality measure proposed by Latora and Marchiori (2007). This quantifies the change in performance of the network when a node is removed. For a performance measure P of the network G, the delta centrality C_i^{Δ} of node i is defined as

$$C_i^{\Delta} = \frac{\Delta P}{P} = \frac{P(G) - P(G')}{P(G)} \quad (G' : G \text{ after removing node } i)$$

The performance measures considered are summarised in Table 4.

Table 4: Performance measures considered for delta centrality.

Performance Measure P	Explanation
Size of Largest Connected Component (LCC)	The size of LCC shows the connectedness of the network. A smaller number indicates many stations are disconnected from the main network, resulting in less connectess.
Network Efficiency $E(G) = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{d_{ij}}$	Efficiency measure proposed by Latora and Marchiori (2001). A high value indicates lower shortest path lengths between nodes. This measure is applicable for disconnected networks as well, unlike the average shortest path length. Since the `nx.global_efficiency()` function cannot incorporate edge length, a custom function was defined to calculate this value.

Applications can be found within and beyond transport networks - closures of NYC subway stations after 9/11 (Wyatt, 2002; Paaswell, 2012) is one example of analysing transport networks. In other contexts, Broder *et al.* (2000) analysed the early worldwide web from the size of (strongly) connected components, and Achard and Bullmore (2007) is an example of applying the network efficiency measure to the human brain. These examples indicate the global applicability of impact measures.

1.1.3 Node Removal

The measures in Table 4 is used to analyse the impact of node removal. Two strategies shown in Table 5 are used to remove 10 nodes from the network. The stations removed are shown in Table 6.

Table 5: Strategies for removing nodes.

Strategy	Explanation
Non-sequential Removal	Removes nodes in order that appears in the original rank table (Table 1, Table 2, and Table 3).
Sequential Removal	Recalculates centrality on the deformed network after each node removal, and determines the next node to remove based on the recalculated centrality.

The change in performance measures caused by the node removal is illustrated in Figure 1, and the geographic locations of removed stations are mapped in Figure 2.

* Discussions

The sequential removal is more effective to study resilience of the network compared to non-sequential removal. Since the characteristics of the network change upon node removal, the crucial node may also change when re-evaluating the remaining network. By assessing the importance of node at each step, the sequential removal is capable of identifying nodes that gain importance after the initial deformation.

Table 6: Stations removed for each centrality measure and strategy.

(a) Degree Centrality

Rank	Non-sequential	Sequential
1	Stratford	Stratford
2	Bank and Monument	Bank and Monument
3	Baker Street	King's Cross St. Pancras
4	King's Cross St. Pancras	Baker Street
5	Earl's Court	Green Park
6	Canning Town	Earl's Court
7	Liverpool Street	Canning Town
8	Green Park	Oxford Circus
9	Oxford Circus	Willesden Junction
10	Waterloo	Waterloo

(b) Betweenness Centrality

Rank	Non-sequential	Sequential
1	Bank and Monument	Bank and Monument
2	King's Cross St. Pancras	King's Cross St. Pancras
3	Stratford	Canada Water
4	Baker Street	West Hampstead
5	Oxford Circus	Earl's Court
6	Euston	Oxford Circus
7	Earl's Court	Shepherd's Bush
8	Shadwell	Bakes Street
9	Waterloo	Acton Town
10	South Kensington	Stratford

(c) Closeness Centrality

Rank	Non-sequential	Sequential
1	Holborn	Holborn
2	King's Cross St. Pancras	King's Cross St. Pancras
3	Tottenham Court Road	Embankment
4	Oxford Circus	Waterloo
5	Leicester Square	London Bridge
6	Picadilly Circus	West Hampstead
7	Charing Cross	Clapham Junction
8	Chancery Lane	Mile End
9	Covent Garden	Stratford
10	Embankment	Notting Hill Gate

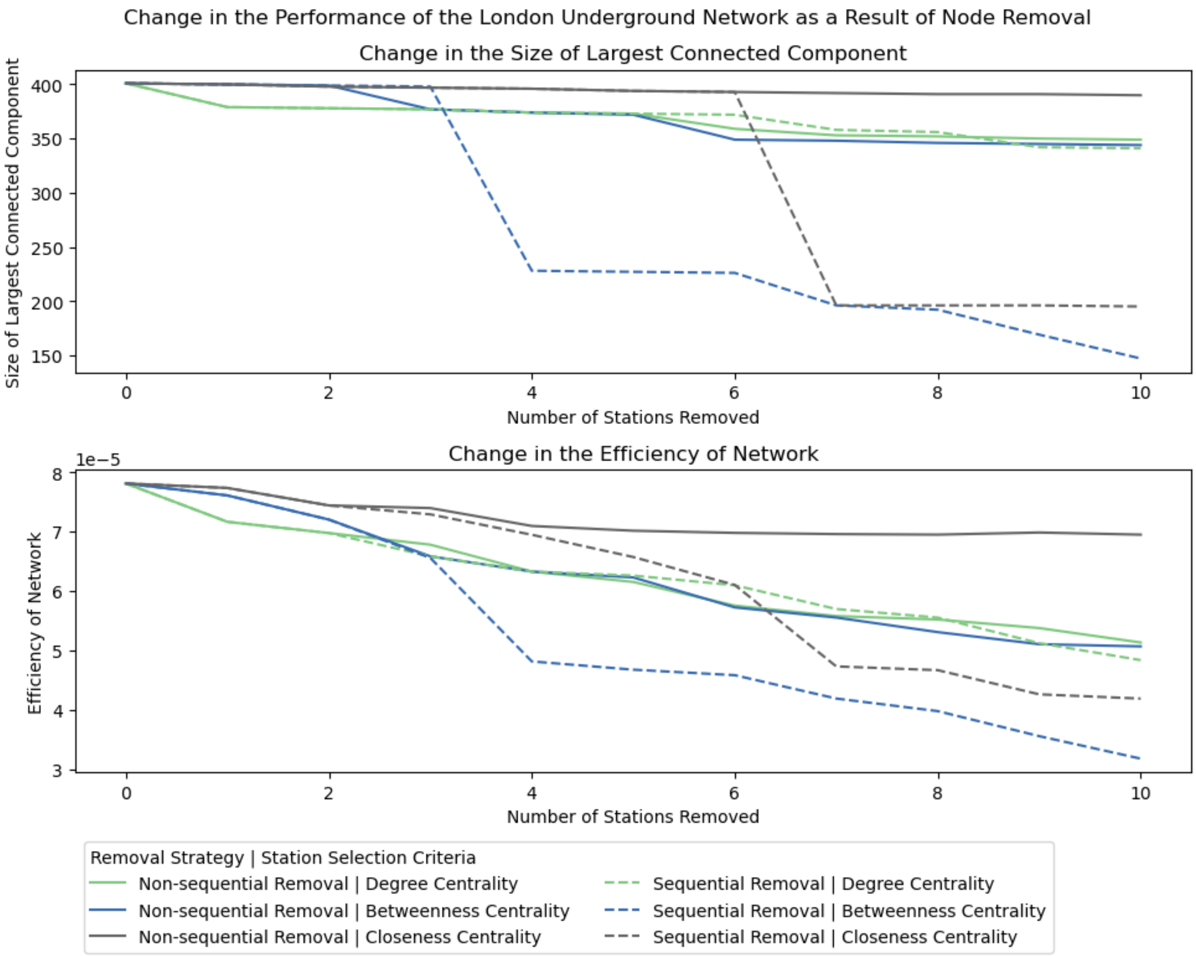

Figure 1: The change in the performance measures as a result of node removal on the London Underground Network.

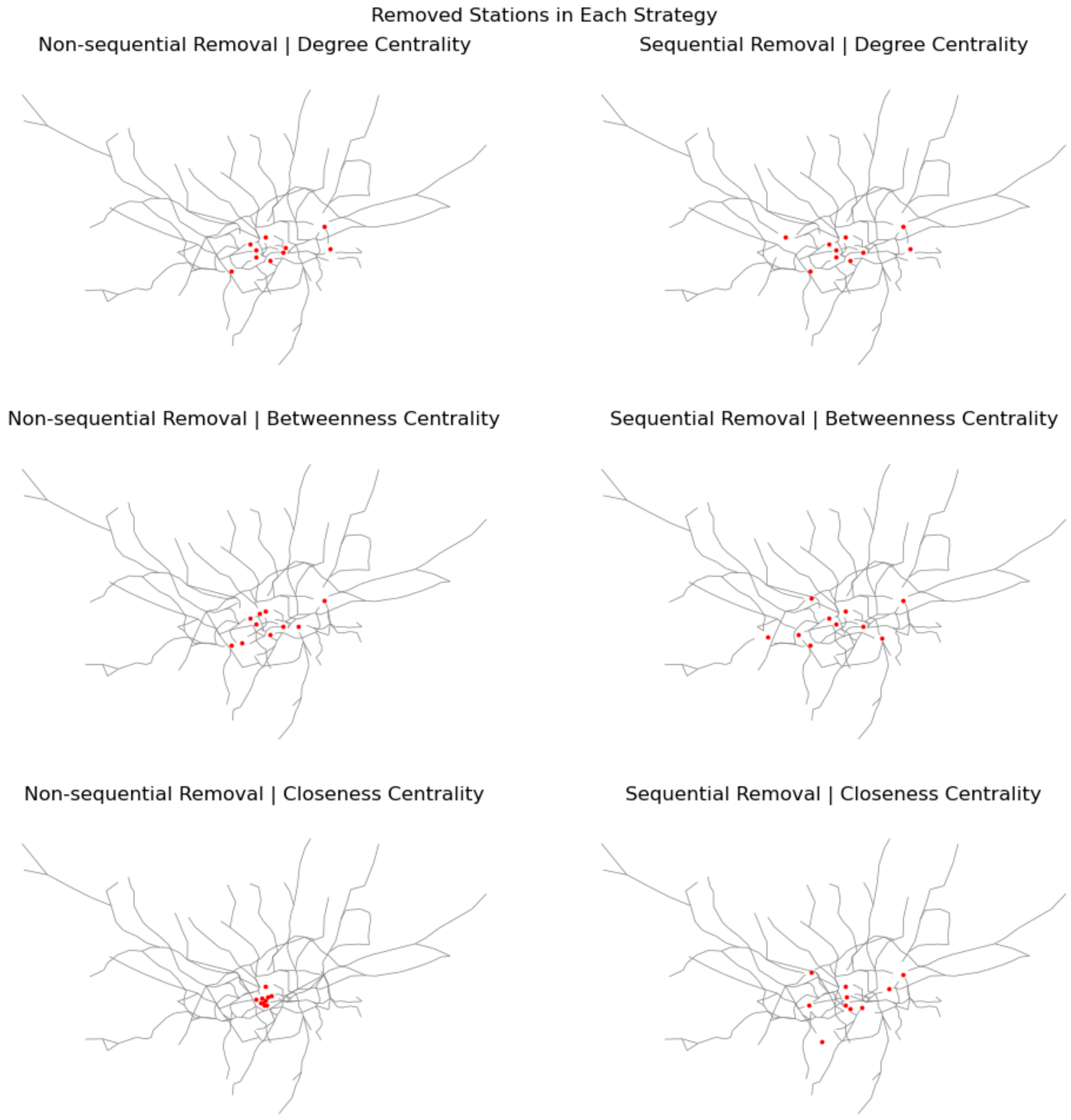

Figure 2: The map of the London Underground Network after each node removal strategy.

Node removal based on the betweenness centrality has the largest impact on performance. Since a high betweenness centrality indicates many geodesics passing the node, removing this node will force many paths to divert to different routes. Interestingly, the degree centrality has identified the first few stations with larger impact. Further investigation is required to identify whether this is a general trend or specific to the Underground network.

When comparing between impact measures, the efficiency is a better evaluation of the performance. The size of LCC is unable to assess the impact of station removal unless it disconnects nodes from the network, whereas the efficiency can assess the change within a connected network, thus the impact is more sensitively reported. Comparing the observations, the LCC shows little change except when removing hub stations that disconnect branches, where efficiency gives a continuous evaluation of performance.

1.2 Flows: Weighted Network

In this section, a weighted network considering the passenger flows is analysed. Based on the morning peak origin-destination data from Transport for London (2021), flows have been assigned to the shortest path between nodes considering the length of each edge. The flow is either used to weight origin-destination pairs or used directly as the weight of edges between stations.

1.2.1 Centrality Measures

First, we have considered whether centrality measures in Section 1.1 need adjustments to be applied to the weighted network. The adjustments made for the weighted measures are summarised in Table 7.

Table 7: Adjustments made for each centrality measure on the weighted network.

Centrality measure	Weighted measure	Explanation
Degree Centrality	Strength $s_i = \sum_j A_{ij} w_{ij}$	w_{ij} is the flow on link between nodes i and j. A measure discussed as the strength s_i in Lee *et al.* (2008), quantifying the amount of flow on edges connected to the node. This measure is highly related to the betweenness centrality, since they both address the flows involving the node. The unweighted degree centrality is a valid measure in a weighted network as well.
Betweenness Centrality	Weighted Betweenness Centrality $x'_i = \sum_{st} \frac{n^i_{st}}{g_{st}} T_{st}$	T_{ij} as the number of flows from i to j. The betweenness centrality, weighting each origin-destination pair by the number of flows.

Centrality measure	Weighted measure	Explanation
Closenese Centrality	Harmonic Weighted Closeness Centrality $C''_i = \frac{1}{n-1}\sum_j \frac{1}{u_{ij}}$	When considering the inverse of flows as the **social distance** between nodes, u_{ij} is defined as the smallest social distance between nodes i and j. A harmonic mean was used to account for edges with zero flow (infinite social distance). Simply giving the weight T_{ij} in Equation 3 as $C'_i = \frac{\sum_j T_{ij}}{\sum_j T_{ij} d_{ij}}$ gives the average distance travelled by the user of the station, which is difficult to interpret from the perspective of importantness. (C'_i may be larger in the city centre attracting people from longer distances, or at the outskirts where passengers need to travel long distances in general.)

Comparing the weighted degree and betweenness centralities in Table 7, their difference lies in whether it counts journeys that originate or terminate at the node in question, and whether it double-counts flows that pass through the node. Thus, their relationship can be expressed as

$$s_i = 2x'_i + (O_i + D_i - T_{ii}) \tag{4}$$

where O_i, D_i is the number of flows originating and terminating at node i.

Based on the above measures, we have examined the stations with high weighted centrality measures.

The stations with the highest weighted degree centrality is shown in Table 8.

Table 8: Stations with the highest weighted degree centrality.

Rank	Station	Weighted Degree (Strength) $(\times 10^5)$
1	Bank and Monument	5.188
2	Oxford Circus	4.889
3	King's Cross St. Pancras	4.671
4	Baker Street	3.358
5	Waterloo	3.344
6	Euston	3.164
7	Green Park	3.143
8	Victoria	2.947
9	Liverpool Street	2.770
10	Embankment	2.539

6 out of 10 stations in this table also appear in Table 1, while terminal stations such as Euston and Victoria are unique to this list.

The **betweenness centrality** considering flows, given T_{ij} as the number of flows from node i to j, can be proposed as follows:

$$x_i' = \sum_{st} \frac{n_{st}^i}{g_{st}} T_{st} \tag{5}$$

The stations with the highest weighted betweenness centrality values are calculated as shown in Table 9.

Table 9: Stations with the highest weighted betweenness centrality.

Rank	Station	Weighted Betweenness Centrality $(\times 10^5)$
1	Oxford Circus	4.415
2	Bank and Monument	4.104
3	King's Cross St. Pancras	4.052
4	Baker Street	3.131
5	Green Park	2.852
6	Euston	2.816
7	Waterloo	2.433
8	Embankment	2.393
9	Charing Cross	2.277
10	Victoria	2.236

6 out of 10 stations are also highly ranked in the unweighted measure in Table 2. Because of the similarity between the weighted degree centrality, 9 out of 10 stations in Table 9 appear in Table 8 as well.

The stations with high harmonic weighted closeness centrality values are shown in Table 10.

Table 10: Stations with the highest weighted harmonic closeness centrality.

Rank	Station	Harmonic Weighted Closeness Centrality
1	Bank and Monument	68.816
2	Liverpool Street	66.250
3	Stratford	65.282
4	Waterloo	64.690
5	Green Park	63.928
6	King's Cross St. Pancras	63.876
7	Moorgate	62.571
8	Oxford Circus	61.793
9	Westminster	61.636
10	Baker Street	61.169

These stations are not geographically concentrated compared to Table 3. Terminal stations with high originating or terminating journeys, or stations connected by high-flowing edges are highly ranked in this measure. King's Cross St. Pancras and Oxford Circus are the only stations that appear in both Table 3 and Table 10.

1.2.2 Impact Measures

We have considered the need for adjusting the impact measures in Table 4 for the weighted network. When considering the weight, each measure should be adjusted as summarised in Table 11.

Table 11: Adjustments for the impact measures for a weighted network.

Original Performance Measure	Weighted Equivalent	Explanation
Size of LCC	Ratio of journeys within same connected component $J(G)$	Drawing the principle of LCC by considering whether two nodes are within the same connected component. Weighting this by the flows between each node pair, resulting in measuring the ratio of journeys that can be completed within a connected component. Note that the size of LCC is also valid in a weighted network, although unable to consider flows.
Network Efficiency	Weighted Network Efficiency $E'(G)$	Each distance between node pairs weighted according to the number of passengers travelling between them, calculating the average inverted distance of journeys.

$J(G)$ and $E'(G)$ are calculated as shown in Equation 6 and Equation 7 respectively.

$$J(G) = \frac{\sum_{i \neq j}(T_{ij}\delta_{ij})}{T} \quad \left(\delta_{ij} = \begin{cases} 1 \text{ if nodes } i \text{ and } j \text{ are connected} \\ 0 \text{ otherwise} \end{cases} \right) \tag{6}$$

$$E(G) = \frac{1}{T} \sum_{i \neq j} \frac{T_{ij}}{d_{ij}} \tag{7}$$

1.2.3 Node Removal

The best performing methodology in Section 1.2 was the sequential removal using betweenness centrality. The unweighted and weighted variants removes stations shown in Table 12.

Table 12: Stations removed using the highest unweighted and weighted betweenness centrality. Stations for the unweighted centrality is extracted from Table 6b.

Rank	Unweighted	Weighted
1	Bank and Monument	Oxford Circus
2	King's Cross St. Pancras	King's Cross St. Pancras
3	Canada Water	Bank and Monument

The changes in the performance using $J(G), E(G)$ for each strategy are shown in Figure 3. Comparing these results, removing stations based on the unweighted measure has the most impact on the network. The weighted efficiency has dropped significantly, indicating journeys are taking longer or disconnected in the nodes-removed network. Thus, the stations that have the most impact on the performance of the network are: Bank and Monument, King's Cross St. Pancras, and Stratford.

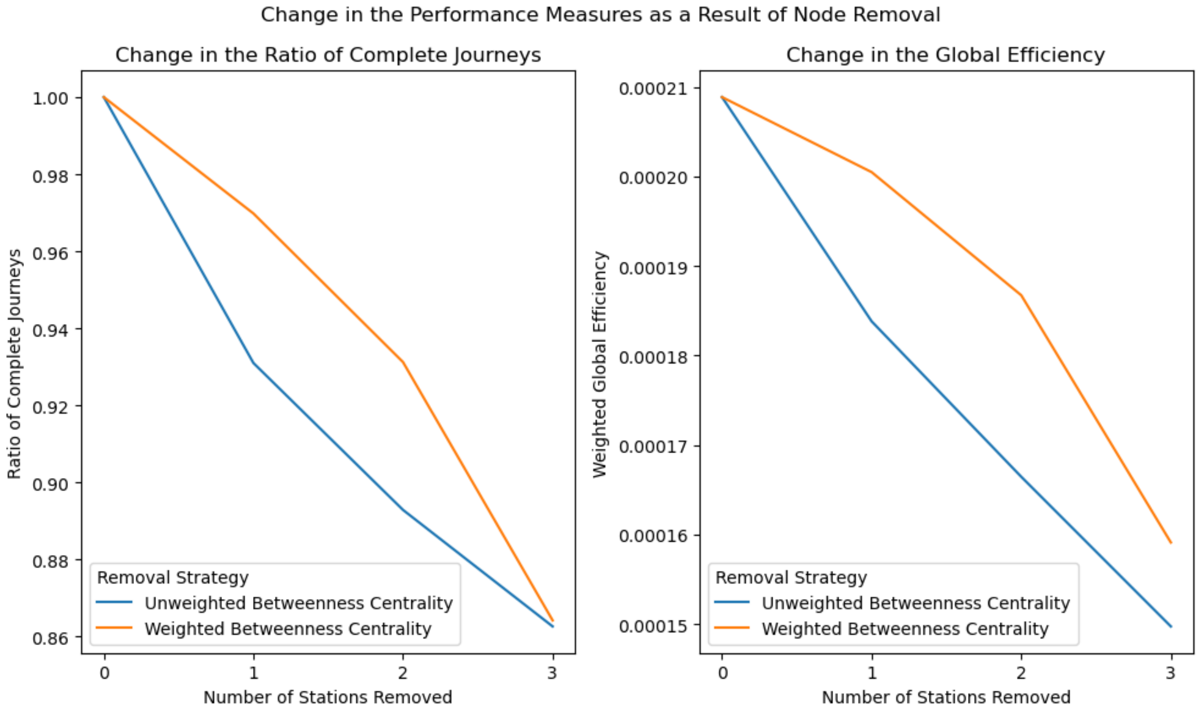

Figure 3: Change in the performance measures as a result of node removal according to the weighted and unweighted betweenness centrality.

The weighted strategy takes into account the flows of passengers, but did not outperform the unweighted strategy. One possible explanation is due to the characteristics of the stations removed: the weighted strategy removes stations near the city centre with dense networks allowing for minimal diversions, while the unweighted strategy removes the suburban transit hub of Stratford that immediately disconnects branches, highly impacting performance measures. A further investigation is necessary for a more accurate comparison; continuing the removal may give additional insights.

2 Part 2: Spatial Interaction Models

In the second part, we have analysed the morning-peak passenger flows of the London Underground, following a brief introduction of the spatial interaction model.

2.1 Models and Calibration

2.1.1 Family of Spatial Interaction Models

The spatial interaction model predicts the flow between origin and destination from the amount of activity at both ends and the distance between them, drawing an analogy from Newton's law of gravity (Waddell, 2002). Spatial interaction models are classified by the constraints cast on them deriving from actual observations; whether the amount of originating and terminating journeys of nodes are preserved.

In this section, the variables will be denoted as follows:

- T_{ij}: estimated flow from station i to j
- T: total flow within a network, thus $T = \sum_{i,j} T_{ij}$
- d_{ij}: distance between stations i and j - the cost function is a function of d_{ij} hence denoted as $f(d_{ij})$
- O_i, D_j: activity at the origin and destination that impacts the originating and terminating flows

In this model, O_i and D_j are the number of total journeys from / to the node. As we are analysing the data for the morning peak, O_i corresponds to the population that uses the station and D_j to the number of jobs around the node.

The **unconstrained model** only constrains the model with matching the total flows with the observed value, calculated as:

$$T_{ij} = KO_i^{\alpha} D_j^{\gamma} f(c_{ij}) \quad \left(K = \frac{T}{\sum_i \sum_j O_i^{\alpha} D_j^{\gamma} f(d_{ij})} \right) \tag{8}$$

K is determined so that the total flow T is preserved. The unconstrained model is used when there is information on the total amount of journeys but no detailed information is available.

The **singly constrained model** constrains the total number of observations for each component at the origin or the destination. The **origin constrained model**, shown in Equation 9, is a singly constrained model that fixes the total number of journeys at the origin.

$$T_{ij} = A_i O_i D_j^{\gamma} f(d_{ij}) \tag{9}$$

The parameter A_i is determined so that the total at the origin is constrained.

$$\sum_j T_{ij} = O_i \therefore A_i = \frac{1}{\sum_j D_j^{\gamma} f(d_{ij})} \tag{10}$$

This model is widely used for flow prediction in various use-cases, including deciding the location and size for retail developments based on the residential distribution within a certain area (Haynes and Fotheringham, 2020).

Similarly, the **destination constrained model** constrains the total at the destination, where the flows are calculated by Equation 11.

$$T_{ij} = O_i^{\alpha} B_j D_j f(c_{ij}) \quad \text{where} \quad \left(B_j = \frac{1}{\sum_i O_i^{\alpha} f(c_{ij})} \right) \tag{11}$$

This can be used for predicting flows from origins given a particular destination, such as estimating the impact of a new development within a city (Haynes and Fotheringham, 2020).

The **doubly constrained model** constrains both the total at the origin and destination.

$$T_{ij} = A_i O_i B_j D_j \exp(-\beta d_{ij}) \tag{12}$$

This is used in the transport planning context as a method of trip distribution in the 4-step method, estimating flows based on activity of both ends (Robinson, 2011).

2.1.2 Calibration of Parameters

In this report, an origin-constrained model is used to simulate the different scenarios. This is justified in Section 2.2. The dataset used for analysis have the following data for every London Underground Station.

- population of origin
- jobs at the destination
- distance between origin and destination
- flow from the origin to the destination

For the calibration process, we will first use the doubly constrained model to estimate the best cost function and parameters by comparing with the observed flow. Then, the origin-constrained model required for the scenarios are used to calibrate the γ parameter. For the cost function, constraining both the total journeys for both the origin and destination will enable the full utilisation of observed data, enabling the most accurate calibration. The γ parameter does not appear in the doubly constrained model, and is required to calibrate in the origin-constrained model.

* Calibration of the cost function

The Poisson model (Flowerdew and Aitkin, 1982) of the doubly constrained model Equation 12 is written as:

$$\ln(T_{ij}) = \ln A_i + \ln O_i + \ln B_j + \ln D_j + \ln f(d_{ij})$$

For the cost function $f(d_{ij})$, the negative exponential and inverse power functions are compared (Equation 13) to determine the better performing relationship and the optimal parameter β. The results are shown in Table 13.

$$f(d_{ij}) = \begin{cases} \text{exp}(-\beta d_{ij}) \text{ (Negative Exponential)} \\ d_{ij}^{-\beta} \text{ (Inverse Power)} \end{cases} \tag{13}$$

Table 13: Comparison of results for the doubly constrained model using inverse power and negative exponential cost functions.

Cost Function	Parameter β	R^2 value
Negative exponential $f(d_{ij}) = \text{exp}(-\beta d_{ij})$	$\beta = 1.543 \times 10^{-4}$	$R^2 = 0.4979$
Inverse Power $f(d_{ij}) = d_{ij}^{-\beta}$	$\beta = 9.096 \times 10^{-1}$	$R^2 = 0.4077$

Since the negative exponential model (Equation 14) has a better fit to the observed flows, this is used in the rest of the analysis.

$$f(d_{ij}) = \text{exp}(-\beta d_{ij}) \quad (\beta = -1.543 \times 10^{-4}) \tag{14}$$

* Calibration of γ

Since the origin-constrained model was used for the analysis of the scenarios, the gamma variable must be calibrated before applying to the new scenarios. Equation 9 is transformed into a Poisson Model as follows:

$$\ln(T_{ij}) = \ln A_i + \ln O_i + \gamma \ln D_j + \ln f(d_{ij}) \tag{15}$$

γ with the highest R-squared value was used, resulting as follows:

$$\gamma = 7.556 \times 10^{-1} \quad (R^2 = 0.4680) \tag{16}$$

2.2 Scenarios

The scenarios we have considered are summarised in Table 14.

Table 14: The scenarios explored in this report

Scenario	Explanation
Scenario A	Jobs at Canary Wharf decrease by 50 %
Scenario B	Increase in cost of transport - considering 2 parameters

Since scenario A involves the change in the characteristics of the destination, the origin constrained model is used for the analysis to preserve the number of commuters starting their journeys in each area. The same model was used for scenario B and the original state for a fair comparison between scenarios.

2.2.1 Scenario A

We have first decreased the number of jobs at Canary Wharf by 50% from D_j to D'_j (Equation 17). We have observed how the destination of commuters changed in reaction to this decrease using the origin-constrained model (Equation 15).

$$D'_j = \frac{D_j}{2} \tag{17}$$

A_i needs to be adjusted so that the new parameter A'_i fulfills Equation 18, ensuring the total flows that originate from each station is conserved.

$$A'_i = \frac{1}{\sum_j D'^{\gamma}_j f(d_{ij})} \tag{18}$$

Finally, the new flows T'_{ij} is estimated as shown below, derived from Equation 9.

$$T'_{ij} = A'_i O_i D'^{\gamma}_j \, \mathsf{exp}(-\beta d_{ij})$$

As a result, a significant drop in the number of journeys terminating at Canary Wharf was observed, from 47,690 in total to 29,496 (61.9 % of original flow). As observable from Figure 4, the decrease in the flows to Canary Wharf occured evenly among all origins, and were redistributed into other destinations.

2.2.2 Scenario B

We have considered the increase of cost in this scenario, using two scenarios B1 and B2 each modifying the cost functions β_1, β_2 as shown in Equation 19 and illstrated in Figure 5.

$$\begin{cases} \beta_1 = 2\beta \\ \beta_2 = 10\beta \end{cases} \qquad (\beta = -1.543 \times 10^{-4}) \tag{19}$$

The new flows for this scenario was calcuated following the same parameter-recalibration process in scenario A. The flows between origin-destination pairs plotted by distance of journeys are shown in Figure 6. As the β increases, the longer distance journeys have lower frequency compared to the original scenario.

2.2.3 Discussion

In order to examine flow change and to compare between scenarios, measurements in Table 15 were calculated.

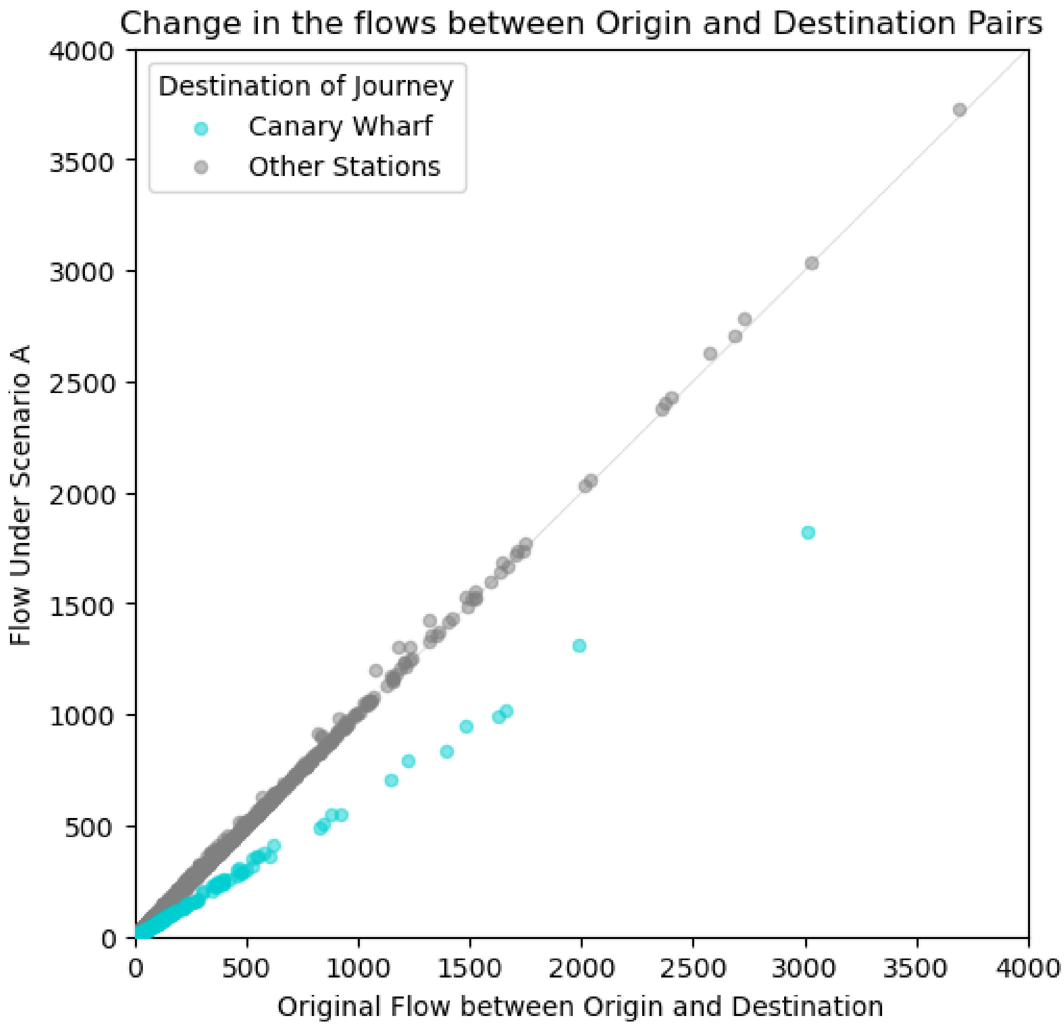

Figure 4: Comparison of the flows of OD pairs the original simulation and the flows under scenario A.

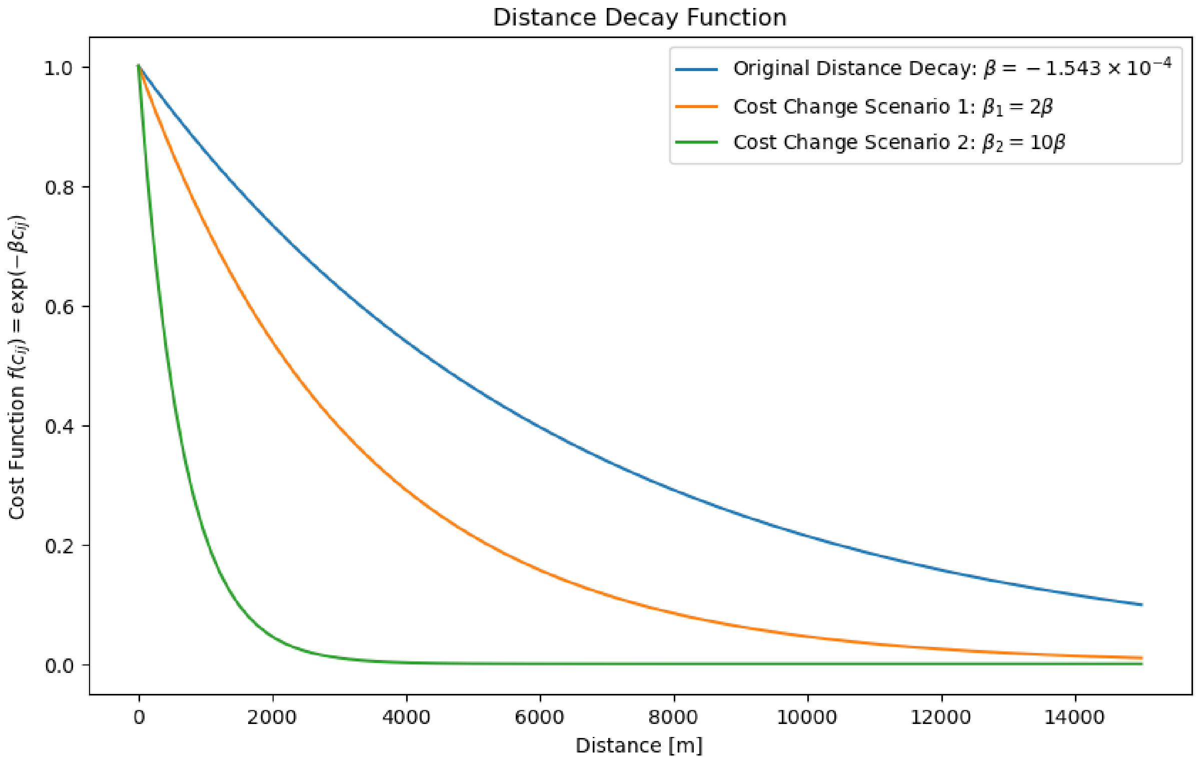

Figure 5: Distance decay function for the original and modified scenarios. The cost is equivalent to travelling twice (B1) or 10 times (B2) the distance in the original model.

Table 15: Measurements considered to quantify the change in flows for each scenario.

Measurement	Explanation
Quantity of change in destination $T_{\text{diff}} = \frac{\sum_{i,j} \lvert T'_{ij} - T_{ij} \rvert}{2}$	The number of flows that have different destinations than the original scenario. (T'_{ij} is the number of journeys from station i to j under the scenario.) A larger change in destinations indicate larger impact on the movement of people within the underground network.
Mean distance of journeys $\bar{\delta} = \frac{\sum_{i,j}(T_{ij} d_{ij})}{T}$	The mean distance of all journeys simulated in each scenario. A larger change in the distance compared to the original simulation indicate larger impact.

For each scenario, values for these measurements are calculated as shown in Table 16.

Table 16: Impact measurements for each scenario.

Scenario	T_{diff}	T_{diff}/T [%]	$\bar{\delta}$ [m]
Original Simulation	-	*(Total flows: 1,542,283)*	*8,583*
Scenario A	18,193	1.2 % of total flows	8,579
Scenario B1	346,503	22.47 % of total flows	6,030
Scenario B2	1,222,191	79.25 % of total flows	1,613

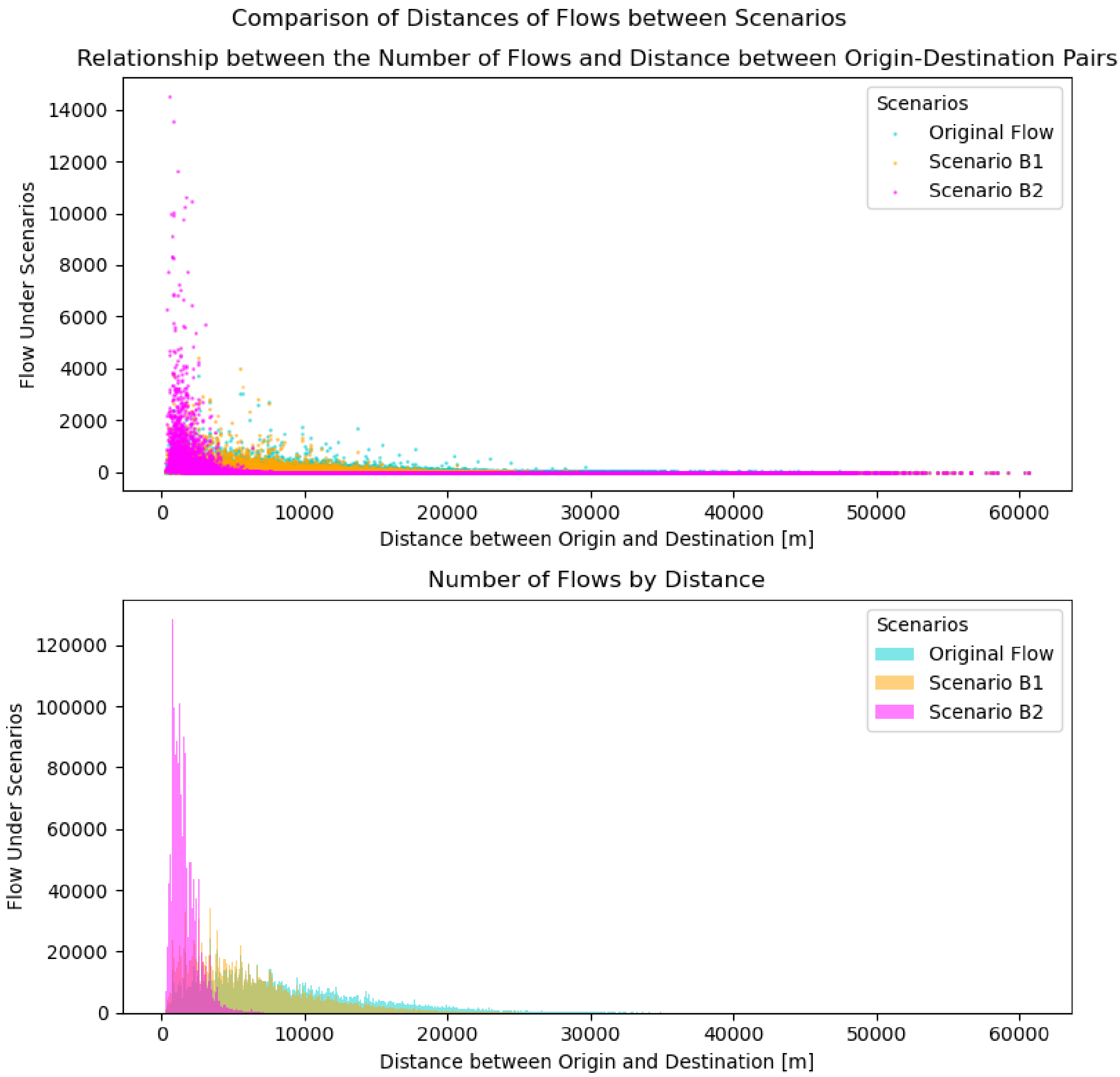

Figure 6: Comparison of the relationship between the distance and the number of flows for scenario B and the original model. Raising the cost function results in a negatively skewed distance distribution.

T_{diff} and $\bar{\delta}$ both indicate that scenario B2 had the largest impact on the number of flows, and scenario A the least impactful.

The intervention to the travel cost (scenario B) directly impacts all origin-destination pairs, while the reduction of jobs in one area (scenario A) only impacts journeys that end at the station affected. This different nature of the interventions dictate the magnitude of change it causes; thus scenario B has a larger impact on the flows on the network as a whole.

Recalling the impact of scenario A on Canary Wharf was a 39 % decrease in the total inflows, which falls between scenarios B1 and B2 in terms of T_{diff}. Even focusing on one station, it is rational to conclude scenario B2 has the largest impact.

In conclusion, a drastic impact on the cost function has more impact throughout the network compared to a large change in the demand in a particular area. Further analysis can be done using the actual changes in the underground fares as a natural experiment, comparing with predicted results in a replicated scenario.

Word count: 2,994 words

GitHub repository (as hyperlink): Urban_Simulation_Report

References

Achard, S. and Bullmore, E. (2007) 'Efficiency and Cost of Economical Brain Functional Networks', *PLOS Computational Biology*, 3(2), p. e17. doi: 10.1371/journal.pcbi.0030017.

Broder, A. *et al.* (2000) 'Graph structure in the Web', *Computer Networks*, 33(1), pp. 309–320. doi: 10.1016/S1389-1286(00)00083-9.

Flowerdew, R. and Aitkin, M. (1982) 'A Method of Fitting the Gravity Model Based on the Poisson Distribution*', *Journal of Regional Science*, 22(2), pp. 191–202. doi: 10.1111/j.1467-9787.1982.tb00744.x.

Haynes, K. and Fotheringham, A. (2020) 'Gravity and Spatial Interaction Models', *1985*.

Latora, V. and Marchiori, M. (2001) 'Efficient Behavior of Small-World Networks', *Physical Review Letters*, 87(19), p. 198701. doi: 10.1103/PhysRevLett.87.198701.

Latora, V. and Marchiori, M. (2007) 'A measure of centrality based on network efficiency', *New Journal of Physics*, 9(6), p. 188. doi: 10.1088/1367-2630/9/6/188.

Leatherdale, D. (2016) 'Are we nearly there yet? Finding out where 'there' is', *BBC News*.

Lee, K. *et al.* (2008) 'Statistical analysis of the Metropolitan Seoul Subway System: Network structure and passenger flows', *Physica A: Statistical Mechanics and its Applications*, 387(24), pp. 6231–6234. doi: 10.1016/j.physa.2008.06.035.

Paaswell, R. E. (2012) 'Approaches to Infrastructure Redevelopment, WTC, Lower NYC and the Region', pp. 215–226. doi: 10.1061/40717(148)18.

Robinson, D. (ed.) (2011) *Computer Modelling for Sustainable Urban Design: Physical Principles, Methods and Applications*. London: Routledge. doi: 10.4324/9781849775403.

Transport for London (2021) 'Project NUMBAT'.

Waddell, P. (2002) 'UrbanSim: Modeling Urban Development for Land Use, Transportation, and Environmental Planning', *Journal of the American Planning Association*, 68(3), pp. 297–314. doi: 10.1080/01944360208976274.

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

Wyatt, E. (2002) 'Subway Service to Resume On Routes Closed After 9/11', *The New York Times*.