*Words count: 2649*

# Assessment for CASA0002 – Urban Simulation

**Part 1: London's underground Resilience**

## I. Topological Network

### I.1. Centrality Measures

In this research, degree centrality, closeness centrality and betweenness centrality would be used. First, degree centrality can be described as the number of connections a node has. In other words, it represents how many other stations are connected to (Zhang and Luo, 2017). With degree centrality, it can highlight the importance of a station in the term of connectivity. Formula of degree of centrality:

$$D(i) = \frac{1}{n-1}\sum_{j=1}^{n} A_{ij}$$

where $A_{ij}$ is the ij-th element of the adjacency matrix A and n is the number of nodes in the network. Secondly, closeness centrality measures how close a node to other nodes. In topological network, it is calculated from the average shortest path distance from a node to all others (Zhang and Luo, 2017). In the context of London's underground resilience analysis, it helps to understand how easy a station to be accessed. Formula of closeness centrality:

$$Cb(i) = \frac{1}{\sum_j d(j,i)}$$

Where d is the distance between i and j. Thirdly, betweenness centrality identifies nodes that act as an intermediary or 'bridge' between other nodes. Operationally, it is calculated from how often a node appears on the all-shortest paths possible in the network (Zhang and Luo, 2017). It can be used for searching a hub station that has a function of maintaining travel efficiency. Formula of betweenness centrality:

$$B(i) = \frac{1}{(n-1)(n-2)}\sum\nolimits_{u,v=1,u\neq i\neq v} \frac{n\,\sigma uv\,(i)}{\sigma uv}$$

Based on the calculation of these formula to the London's underground network data, it shows a following result:

*Words count: 2649*

*Table 1 Centrality Rank of London's Underground Station (Unweighted)*

| Rank | Degree Centrality (DC) | DC_Value | Closeness Centrality (CC) | CC_Value | Betweenness Centrality (BC) | BC_Value |
|---|---|---|---|---|---|---|
| 1 | Stratford | 0,023 | Holborn | 0,0000793 | Stratford | 23768,09 |
| 2 | Bank and Monument | 0,020 | King's Cross St, Pancras | 0,0000790 | Bank and Monument | 23181,06 |
| 3 | Baker Street | 0,018 | Tottenham Court Road | 0,0000789 | Liverpool Street | 21610,39 |
| 4 | King's Cross St, Pancras | 0,018 | Oxford Circus | 0,0000788 | King's Cross St, Pancras | 20373,52 |
| 5 | Green Park | 0,015 | Leicester Square | 0,0000784 | Waterloo | 19464,88 |
| 6 | Canning Town | 0,015 | Piccadilly Circus | 0,0000783 | Green Park | 17223,62 |
| 7 | Earl's Court | 0,015 | Charing Cross | 0,0000783 | Euston | 16624,28 |
| 8 | Liverpool Street | 0,015 | Chancery Lane | 0,0000782 | Westminster | 16226,16 |
| 9 | Waterloo | 0,015 | Covent Garden | 0,0000781 | Baker Street | 15287,11 |
| 10 | Oxford Circus | 0,015 | Embankment | 0,0000780 | Finchley Road | 13173,76 |

From the table above, it can be seen that the distribution of rank across the centralities is vary. However, there are several stations that appear twice (Stratford, Bank and Monument, Baker Street, Green Park, Liverpool Street, Waterloo and Oxford Circus) or even thrice (King's Cross St, Pancrass). From this finding, it can be seen that these stations can be seen as the important node from multiple perspective (connectivity, accessibility, and efficiency).

### I.2. Impact Measures

To measure the impact of the node removal, giant component's size and average path length are appropriate methods. Giant component size refers to the size of the largest connected component in the whole network (Dong, Xiong and Hou, 2012). By using this metric, the size of giant component can be observed overtime. A decrease in the size of giant component also indicates that the network becomes more fragmented and less interconnected. In the context of transportation study, it can be used for quantifying the impact on network connectivity and accessibility when a station/hub is being disrupted/attacked or removed (Diop *et al.*, 2022). The formula for giant component size can be seen as follows:

$$\lim_{n\to\infty} \frac{s(n)}{c} = c > 0,$$

where n is the number of nodes and s(n) is the size of connected component. Secondly, average path length is also a good metric for measuring the impact of node removal because of its strength to measure network efficiency and reachability. In principle, it measures the average shortest distance or number of minimum edges required to travel between all pairs of nodes/stations in the network (Dong, Xiong and Hou, 2012). The information about average path length is important because it is directly related with the change of travel cost. Here is the formula for average path length (L):

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

where d is the distance between i and j.

### I.3. Node Removal

The impact's measurement of node removal to the big component size and average path length is being designed under the frame of two removing strategies (A: non-sequential and B: sequential) and the value of three centrality measures with the number iteration of 10. Here are the results of this experiments:

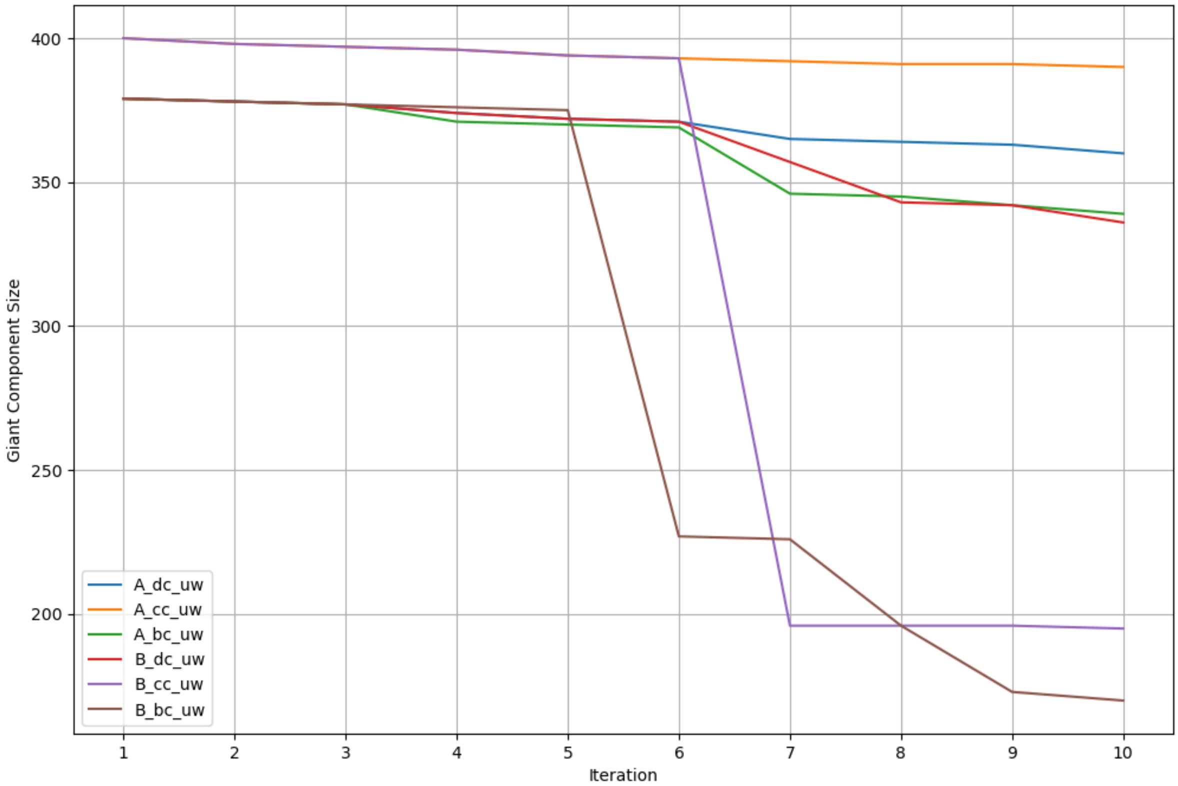



*Figure 1 Big Component Size Experiment across Scenarios*

From Figure1, it seems that removing 10 stations based on their betweenness values in sequential manner is the most effective method in disrupting the resilience of the London underground's network. The graph shows the huge decrease of size (from 400 to 227 nodes) when the 6-th node (West Hampstead station) is being removed. Meanwhile, the other methods show insignificant decrease, except for closeness centrality in sequential manner that shows similar result.

*Words count: 2649*

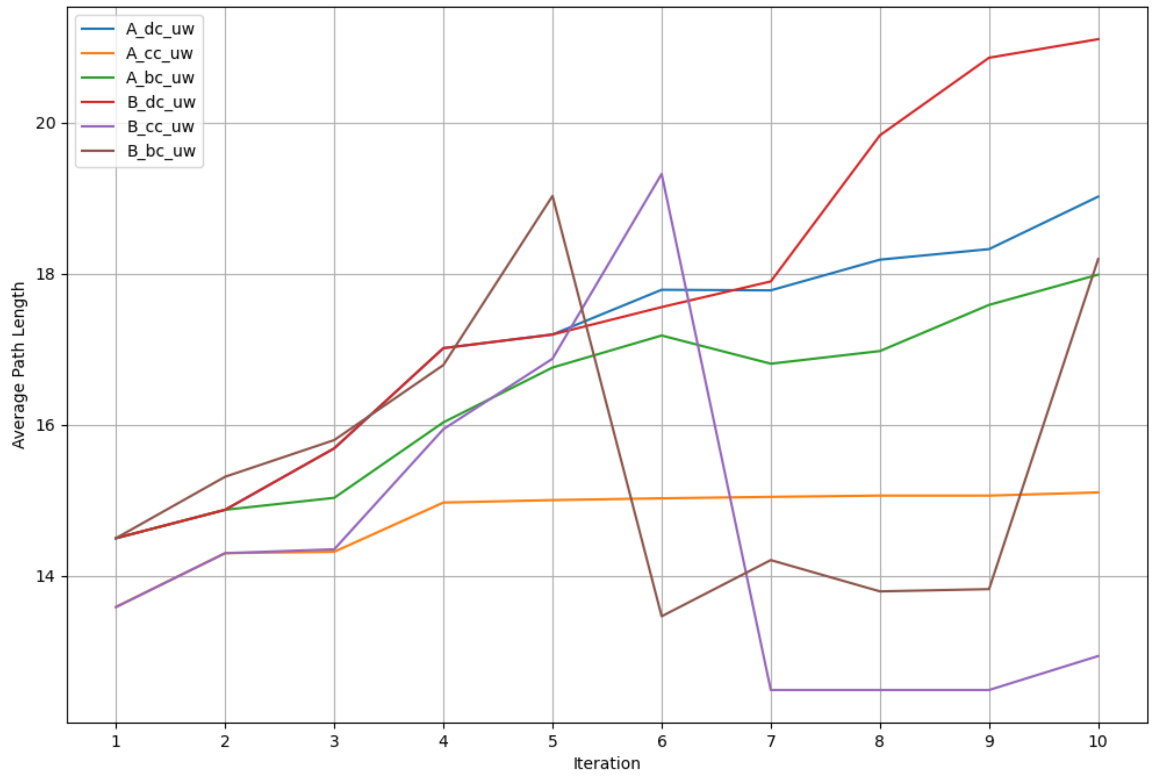
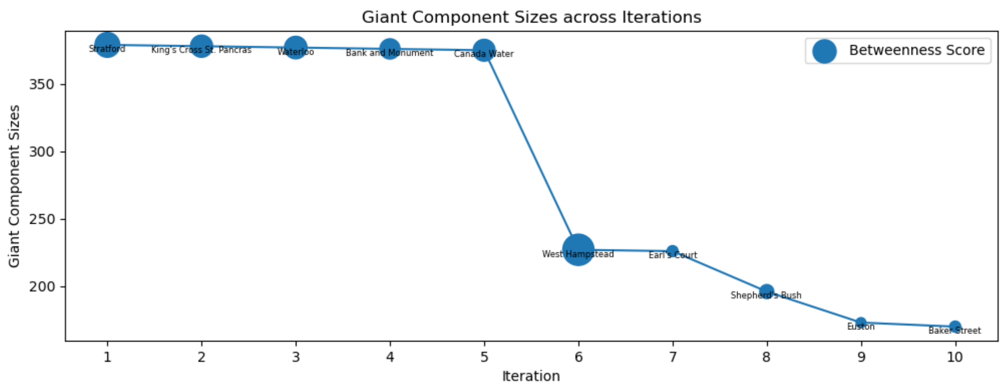



*Figure 2 Average Path Length across Scenarios*

The different result is shown in the average path length (Figure 2), where every method shows a different pattern. It seems that the value of average path length is depended on what the big component size looks like. For instance, the graphs produced by betweenness and closeness centrality in sequential manner have a fluctuated pattern, which fall drastically at 6$^{th}$ and 7$^{th}$ iteration at the same time when the network's giant component is disconnected significantly. Meanwhile, the graphs from the other methods are continuously increasing, align with the continuously decrement of big component size.



*Words count: 2649*

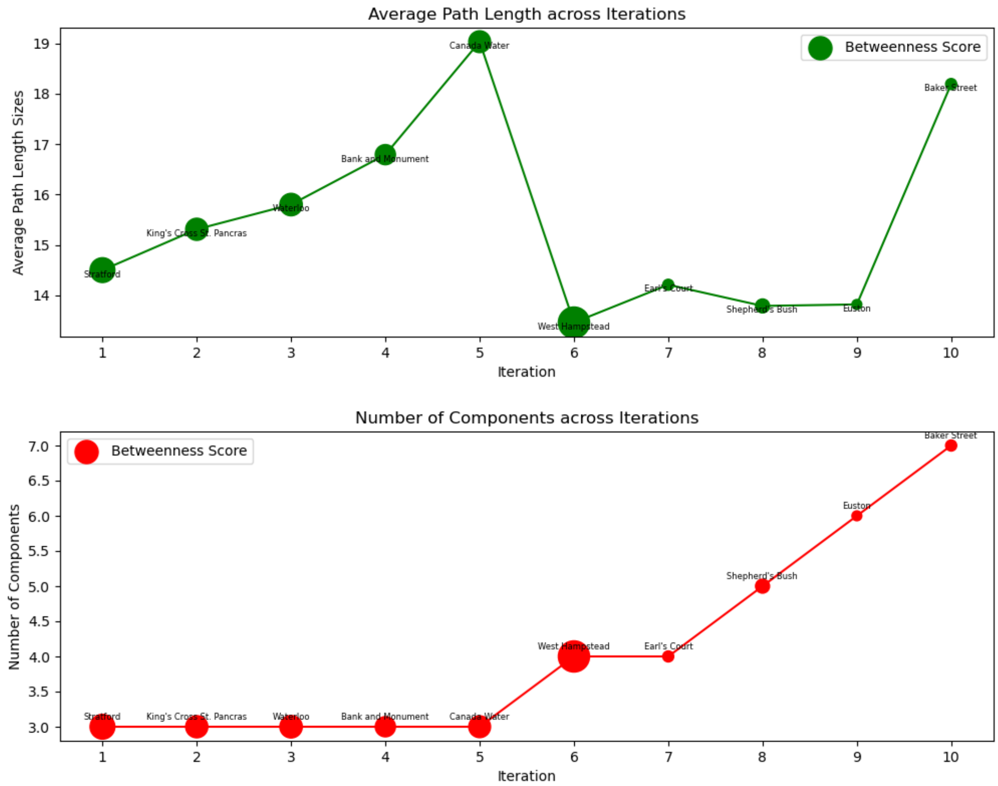



*Figure 3 The Performance of Betweenness Centrality in Disrupting Underground Network's Resilience*

**Up to this point, there is a strong tendency to choose the betweenness centrality in sequential manner in assessing the impact of node removal to the underground resilience.** From the Figure 3, it can be seen that the chain station removals with the highest betweenness centrality in the 1$^{st}$ – 5$^{th}$ iteration (Stratford, King Cross St. Pancras, Waterloo, Bank and Monument and Canada Water) would lead to increase of average path length. Not only that, West Hampstead station is made to be having a highest betweenness centrality because of its chain node removal. At the same time, there is a rise in inefficiency of the network due to the loss of several hubs in this period, but the biggest component is still intact. However, there is a huge disconnection where the West Hampstead is being removed. It can be seen by the fall of big component size and the increase of number of components in the 6$^{th}$. In this moment, the network is being split almost equally like the Figure has shown. Regardless of this loss in the network interconnection, the average path length for the biggest component is becoming low again because it has been split away from the other half which make it inefficient beforehand.

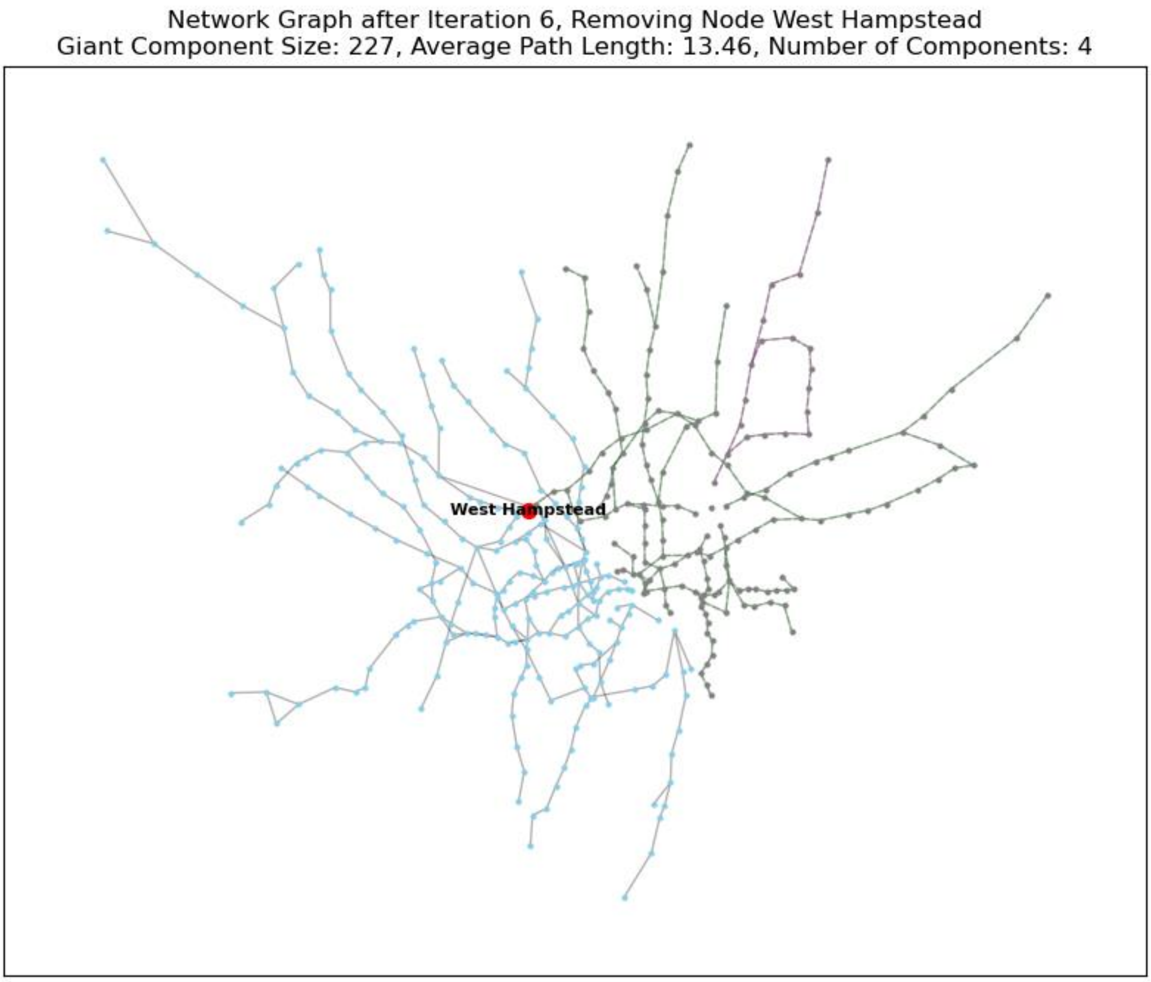



*Figure 4 Huge Disconnection Happened during 6-th Iteration*

## II. Flows: Weighted Network

### II.1. Centrality Measurement in the Weighted Network

When considering the weight of the network, which is flow of passengers in this case, there would be several adjustments to the centrality measure. The adjustments are 1) updating the degree to the eigenvector centrality, 2) changing the weight of closeness centrality from 'length' to the 'flows' and 3) considering weight when calculating the betweenness centrality.

First, eigenvector centrality is basically similar to the degree of centrality. However, instead of measuring the importance/degree of connectivity with calculating how much nodes are connecting to one node, this measurement assigns score to a node based on another importance/degree of connectivity of the other node (Kolios, Panayiotou and Ellinas, 2017). In the context of underground network, this method can be used to identify the influential nodes that act as the backbone in transferring majority of passengers in the network. Here is the explanation of the formula:

$$X_i' = \sum_j A_{ij} X_j \quad \leftrightarrow \quad X^i = AX$$

Where A is adjacency matrix and the bidirectional symbolize the iteration process. Secondly, the weighted closeness centrality is switched from length to the flows. The rationality behind this change is because in this case, the flows represent more comprehensive idea about travel cost, not only about the distance.

Thirdly, betweenness centrality with weight is the same with the topological betweenness centrality, except for the algorithm in considering the shortest paths based on the weight. In this context, the flows that is used as weight can be defined as 'easiness in accessing the station' which affecting the decision in choosing the shortest path between two nodes (Opsahl, Agneessens and Skvoretz, 2010).

With using the adjusted centrality measures, here is the new top ten rank stations for each measurement:

*Table 2 Centrality Rank of London's Underground Station*

| Rank | Eigenvector Centrality | Eig Value | Closeness Centrality | Clo Value | Betweenness Centrality | Bet Value |
|---|---|---|---|---|---|---|
| 1 | Waterloo | 0,53 | Green Park | 2373,5 | Green Park | 44129,75 |
| 2 | Bank and Monument | 0,48 | Westminster | 2366,4 | Bank and Monument | 35695,50 |
| 3 | Westminster | 0,43 | Waterloo | 2361,1 | Waterloo | 29855,75 |
| 4 | Green Park | 0,33 | Bank and Monument | 2357,8 | Westminster | 28385,75 |
| 5 | Liverpool Street | 0,27 | Oxford Circus | 2354,6 | Liverpool Street | 27010,00 |
| 6 | Moorgate | 0,14 | Bond Street | 2342,6 | Stratford | 25715,00 |
| 7 | Oxford Circus | 0,14 | Victoria | 2342,3 | Euston | 22765,75 |
| 8 | Stratford | 0,13 | Liverpool Street | 2338,5 | Oxford Circus | 21063,50 |
| 9 | Bond Street | 0,12 | Warren Street | 2334,1 | Bond Street | 19493,00 |
| 10 | Victoria | 0,11 | Moorgate | 2318,0 | Warren Street | 19358,25 |

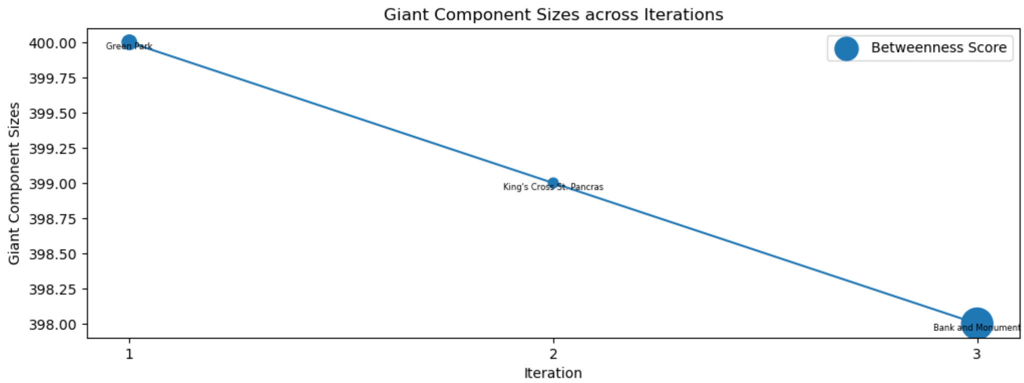
From the table above, it can be seen clearly that by using weight, the rank of each category is more centralized to the same stations. Seven out of ten stations are showing up as the top ten in all metrics (Waterloo, Bank and Monument, Westminster, Green Park, Liverpool Street, Oxford Circus and Bond Street) while the rest has appeared on the two different centralities (except of Euston with only appear once in the betweenness centrality rank).

### II.2. Adjustment for the Global Measurement

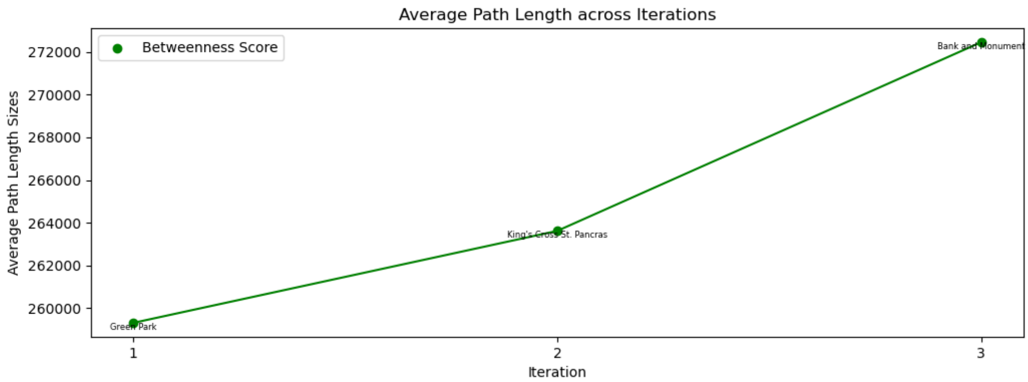
With considering weight into the network, using big component size would be still relevant because the node removal pattern that leads to the disconnection of network is changed due to the change of behaviour in weighted centrality measure. The more relevant global measure that can be used is weighted average path. In this global measure, the method considers weight to redefine what is shortest path. Shortest path in the weighted average path length is the path that gives least resistance which is symbolized mathematically as the sum of weight, which is flow in this case (Opsahl, Agneessens and Skvoretz, 2010).

### II.3. Experiment in the Weighted Network

After reconduct the experiment with adjustment measurements (removing 3 nodes based on the highest weighted betweenness centrality), here is the following the result:




From the graph above, it can be seen that there is no significant change in the component size up to 3-th iteration. The change is just slightly changing from 401 to 398 which is not disconnect any part of the network (except the removed node itself).




However, the weighted average path length across iterations is increasing as the node is removed. Based on the graph, the removal of Bank and Monument is giving a highest impact towards the distribution of flows in the network.

*Words count: 2649*

## Part 2: Spatial Interaction Models

### III. Models and Calibration

### III.1. Introduction to Spatial Interaction Model

Adopted from Newtonian gravity model, spatial interaction model is trying to model the relationship of place A, place B, and the flow between them (Wilson, 1971). There are four models from spatial interaction model which take account 'available knowledge' to constraint the value. In mathematics language, the relationship can be described as follows:

*Unconstrained/Total Constrained Model:*

$$T_{ij} = KO_i^{\propto}D_j^{\gamma}\exp(-\beta c_{ij})$$

*Origin-Constrained Model:*

$$T_{ij} = A_iO_iD_j^{\gamma}\exp(-\beta c_{ij})$$

*Destination-Constrained Model:*

$$T_{ij} = B_iO_i^{\propto}D_j\exp(-\beta c_{ij})$$

*Doubly-Constrained Model:*

$$T_{ij} = A_iB_jO_iD_j\exp(-\beta c_{ij})$$

Where:

- Tij = Flow from origin (i) to destination (j)
- K, A, B = balancing constant
- Oi = emissivity or size of the origin i
- Dj = attractiveness or size of the destination j
- α, γ = parameters that determine the influence of origin and destination to the interaction flow
- cij = the distance function from i to j
- β = parameters that determine the influence of distance

In the London's underground data, four of this spatial interaction model would be used to see which one is better to explain the reality. Not only that, every model would be experimented twice with using different distance decay function: power law and exponential. The reason behind this choice is because there are two possibilities in describing travel cost function, the first one is when the people is more sensitive to the increase of cost (because the reason of commuting is for something that have multiple choices, like go to groceries, entertainments, etc) and the second one is when the people is less sensitive to increase of cost (like commuting for the job where it has one fixed place). To incorporate these two possibilities, power law is used to represent the sensitive distance decay function and exponential is used for the less sensitive one (Chen, 2015).

### III.2. Calibration and Select the Model

After doing the experiment with four spatial interaction models and two distance decay functions stated above to the London's population, jobs, and flows data (in the scale of station), here is the following result in respect to parameters and accuracy:

*Table 3 Performance Result of Spatial Interaction Models to London's Underground Network*

| model | R2 | alpha | gamma | beta |
|---|---|---|---|---|
| Uncosim_pow | 0,32 | 0,74 | 0,77 | 0,612140 |
| Uncosim_exp | 0,36 | 0,71 | 0,74 | 0,000090 |
| Origcosim_pow | 0,39 | vary | 0,77 | 0,863078 |
| Origcosim_exp | 0,47 | vary | 0,76 | 0,000153 |
| Descosim_pow | 0,35 | 0,75 | vary | 0,622229 |
| Descosim_exp | 0,40 | 0,72 | vary | 0,000099 |
| ODcosim_pow | 0,41 | vary | vary | 0,893505 |
| ODcosim_exp | 0,50 | vary | vary | 0,000154 |

From the table above, it can be seen that doubly constrained model with exponential distance decay function has the best performance in representing the observed commuting pattern in London. It can be seen from the $R^2$ which reaching 50%. However, this value is also being competed by the origin constrained model with exponential distance decay function which has $R^2$ = 47%. Based on the result, these two models can be used for the scenario testing. The choice between these two models would be decided on the assumption used for the scenario testing, whether it is only for experimenting the flow distribution with a fixed total origin and destination or it wanted to see the change in the origin or destination as well.

## IV. Scenarios

### IV.1. Scenario A: Decrement of a Job in Canary Wharf for 50%

To compute the new flow where there is a decrease in a job number in Canary Wharf for 50%, it would be expected that there is a reduction in total flow inside the system if the total attraction and destination is conserved (except for Canary Wharf itself). However, if the total flow must be conserved, it means that the change of flow in Canary Wharf would be redistributed to another place. With this information, the scenario A can be conducted by **using origin constrained exponential.**

Based on the recalculation of the flow, there are several destinations that gain a surplus of flow and some of other are opposite. The detail can be seen in the Figure 5. From that image, Stratford, White Chapel, Canada Water, Highbury & Islington and Leicester Square are the places with the highest increase. Meanwhile, Victoria, Wimbledon, Oxford Circus, Hammersmith, and Bank and Monument are the places with highest decrease of flow.

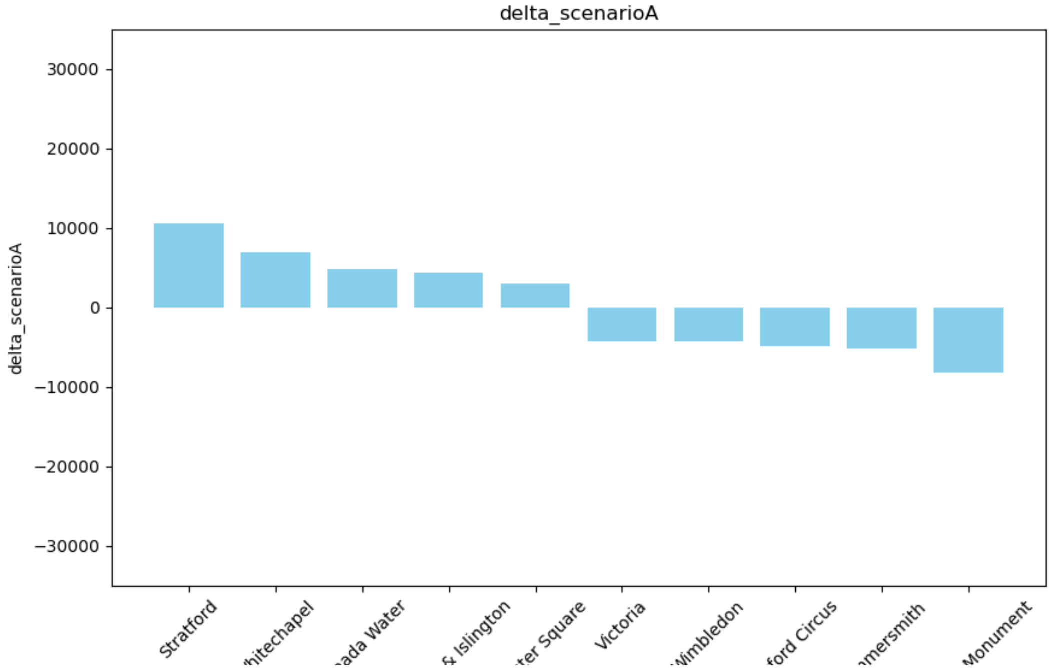



*Figure 5 Change in Flow with Scenario A*

## IV.2. Scenario B: An increase in Beta

In essence, doubly constrained exponential model can be used in calculating the change of flow with a different beta while preserving the number of commuters. However, in order to be able to be compared with scenario A and the performance is quite similar, origin constrained exponential would still be used.

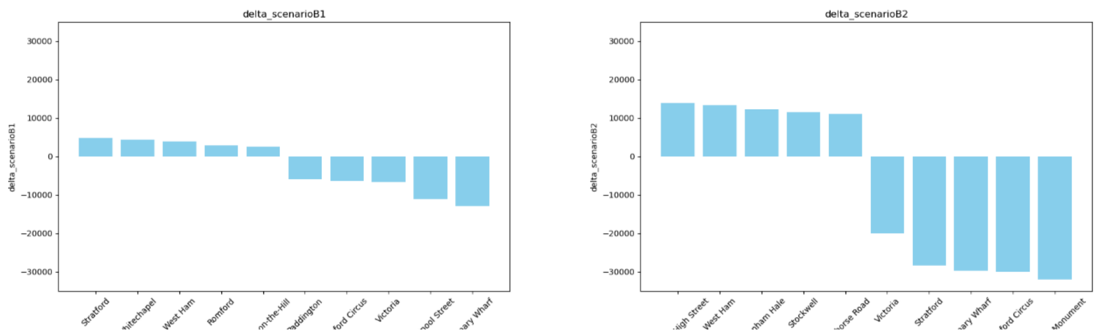



*Figure 6 Change of Flow with B1 (left) and B2 (right)*

After the calculation with two beta values (b1 = twice beta from scenario A, b2 = ten times beta from scenario A), they show a similar pattern with scenario A -there are areas with the increase and decrease in flow- but with different magnitude. The magnitude of scenario A is smaller than scenario B1, while scenario B1 is smaller than B2 (A < B1 < B2).

### IV.3. Comparing the Scenario A and B

From the previous discussion, the reduction of job in Canary Wharf to 50% and the rise of transport cost for 2 and 10 times give a similar result, that is an increase and decrease in total in-flow to the other destinations. However, the magnitude of these results is different, so do the location of distribution. If we deepen more the analysis to see spatial configuration of this phenomenon, the distribution of this change is actually following a certain spatial structure. For instance, let us see the change happened in the Waterloo station.

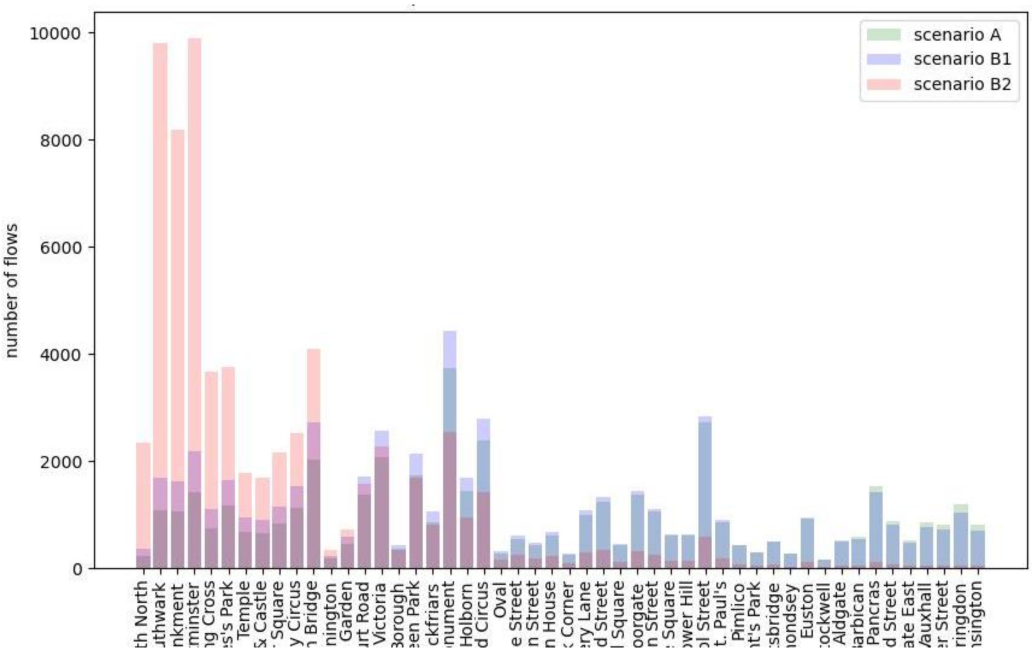



*Figure 7 Comparison of flow change between Scenario A and B in Waterloo Station*

The Figure 7 is sorted based on location proximity, where the most left side station (Lamberth North) is the closest station and the right side is the furthest station (South Kensington) from Waterloo. It can be seen that as the impact is getting bigger (A -> B1 -> B2), the distribution of flow from Waterloo is getting more concentrated to nearest stations to Waterloo.

Surprisingly, this phenomenon is not only happening in Waterloo, but also to all of the places in the network (Figure 8). It can be said that, **as the disincentive increase (the reduction of job opportunity, increase in transport cost, loss of attraction, etc), the distribution of flow would be more concentrated to area close to origin. It is because the people in the origin don’t have an enough pull force (lose of job) and push force (transport fee) to go further.**

*Words count: 2649*

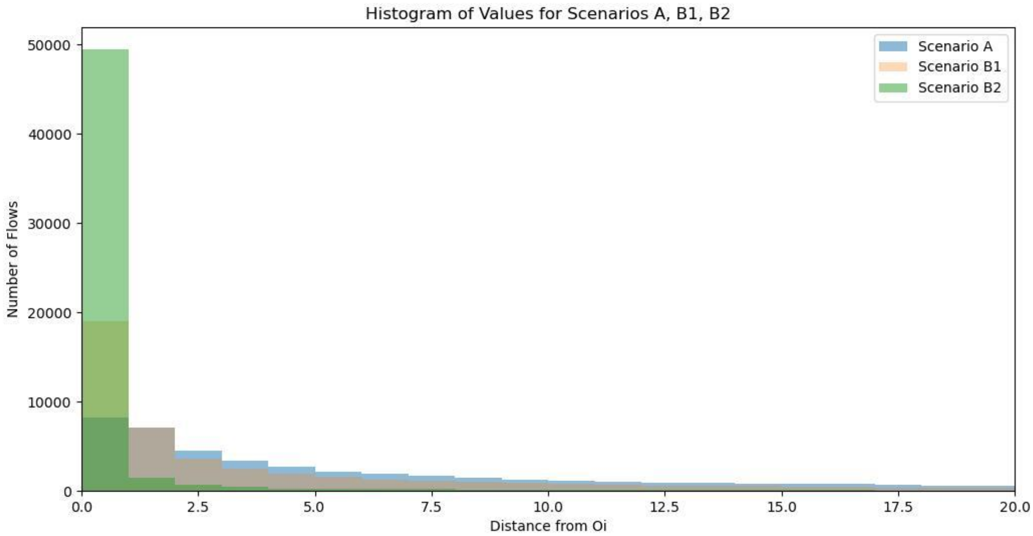



*Figure 8 Comparison of flow change between Scenario A and B in All of London's Underground Area*

## Reference

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.

Diop, I.M. *et al.* (2022) 'Targeted attacks on the world air transportation network: impact on its regional structure'. arXiv. Available at: http://arxiv.org/abs/2209.05617 (Accessed: 25 April 2024).

Dong, X., Xiong, G. and Hou, J. (2012) 'Construction of Artificial Grid Systems Based on ACP Approach', in *Service Science, Management, and Engineering:* Elsevier, pp. 285–304. Available at: https://doi.org/10.1016/B978-0-12-397037-4.00014-4.

Kolios, P., Panayiotou, C. and Ellinas, G. (2017) 'Disaster Resilient Telematics Based on Device-to-Device Communication', in *Wireless Public Safety Networks 3*. Elsevier, pp. 211–230. Available at: https://doi.org/10.1016/B978-1-78548-053-9.50010-X.

Opsahl, T., Agneessens, F. and Skvoretz, J. (2010) 'Node centrality in weighted networks: Generalizing degree and shortest paths', *Social Networks*, 32(3), pp. 245–251. Available at: https://doi.org/10.1016/j.socnet.2010.03.006.

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, J. and Luo, Y. (2017) 'Degree Centrality, Betweenness Centrality, and Closeness Centrality in Social Network', in *Proceedings of the 2017 2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM2017). 2017 2nd International*

*Words count: 2649*

*Conference on Modelling, Simulation and Applied Mathematics (MSAM2017)*, Bankog, Thailand: Atlantis Press. Available at: https://doi.org/10.2991/msam-17.2017.68.