Assessment
CASA0002 – Urban Simulation

SN21125032

1. PART 1: LONDON'S UNDERGROUND RESILIENCE

1.1. Topological Network

Centrality Measures

Aiming to evaluate the resilience of the Tube Network and the impact of node removals, the three centralities measures chosen to identify the most important nodes in the Tube Network are Degree, Eigenvector, and Betweenness centrality. The first experiment of this analysis was also considered to use closeness centrality. However, the application in the context of node removal in this centrality measures has not presented good results.

Degree Centrality:

The number of links associated with a node related to the tube network(Figure 1.1). This centrality reflects the number of connections with a station, which can be linked with another station or tube line that starts and finish in the same station, as a loop, for example (Arcaute and Marin 2022). Regarding this analysis, the degree centrality values are normalised by dividing them by the maximum possible degree, which is n-1. For this analysis, degree centrality can detect a vulnerability of a station, mainly those with a high degree, because removing it can interrupt the commute to an important location through the city. In the map below(Figure 1.1), the stations with high degrees are predominantly located in central London in the north and east region of the city.

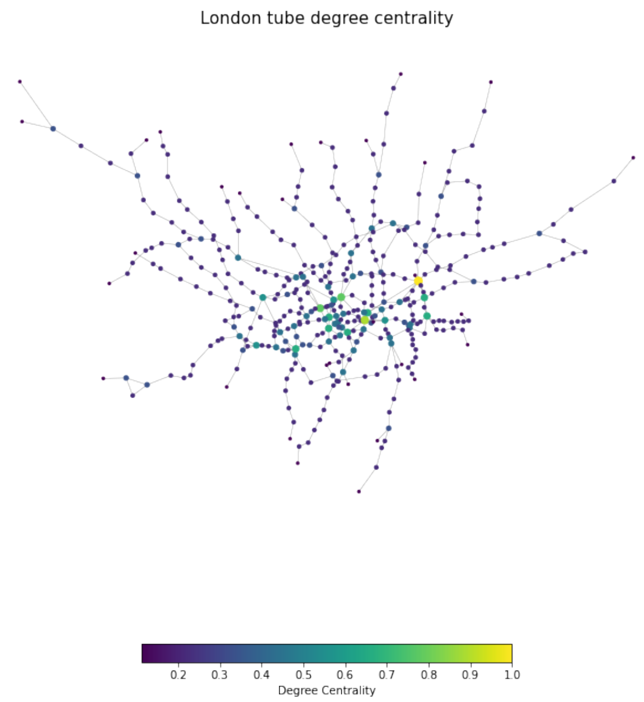

Figure 1.1: Topological Tube network map - Degree centrality

Eigenvector centrality: The transitive influence of nodes(Figure 1.2). Eigenvector centrality measures the influence of a node in a network based on a relative score for all nodes. Thus, nodes with a high-score mean that these nodes are also connected to other nodes with a high score(NetworkX 2023). For tube network, these measure is useful because it is possible to understand the influence of a tube station not only focused on its good connectivity with a range of tube lines. This centrality allows us to understand how good is the connectivity of the neighbour's stations. Therefore, when it is considered a station disruption, a user who is close to other tube stations that are also well connected to the system allows that user to access the transport system without adding large displacements.

$$x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t \in V} a_{v,t} x_t$$

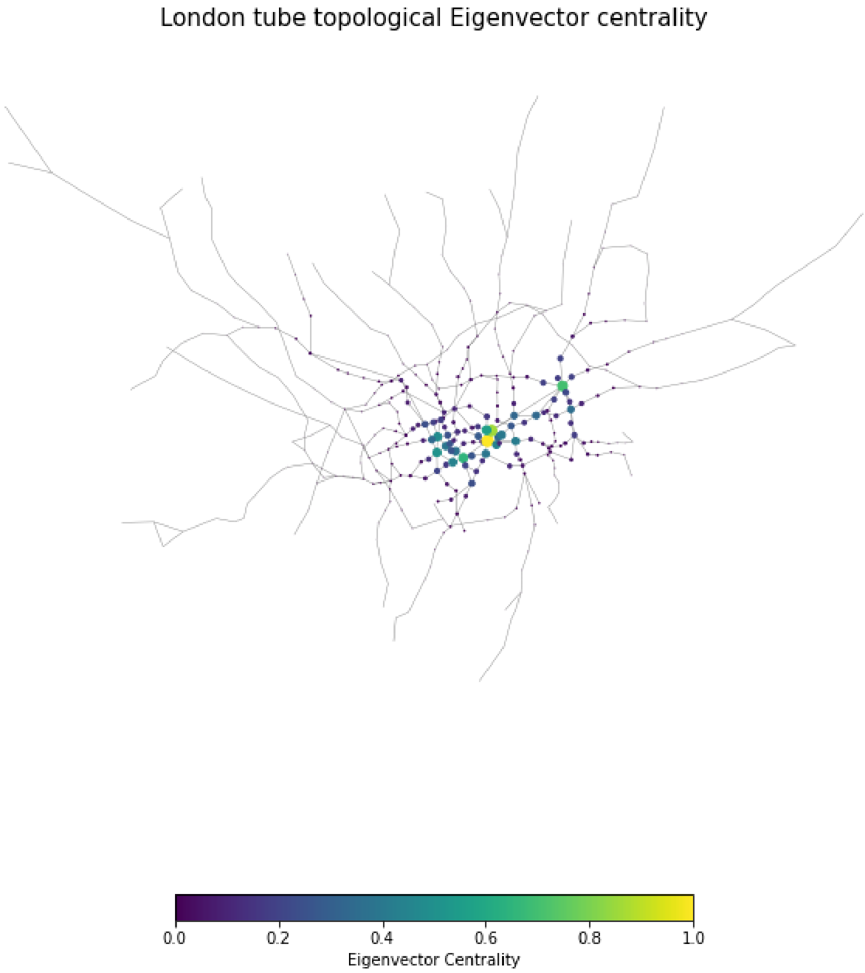

Figure 1.2: Topological Tube network map - Eigenvector centrality

Betweenness centrality: The shortest path passing through a node(Figure 1.3). It is possible to detect the level of influence a node has on the flow of information(Arcaute and Marin 2022). Thus, it can find the nodes that serve as bridges for a different part of the network. Related to the tube network, this measure is an essential outcome for understanding which tube station has the characteristic to allow users to commute easily and fast into the system.

To calculate the betweenness centrality of a node is the sum of all shortest paths that pass through the vertex, where

$$x_i = \sum_{st} \frac{n^i_{st}}{g_{st}} \tag{1}$$

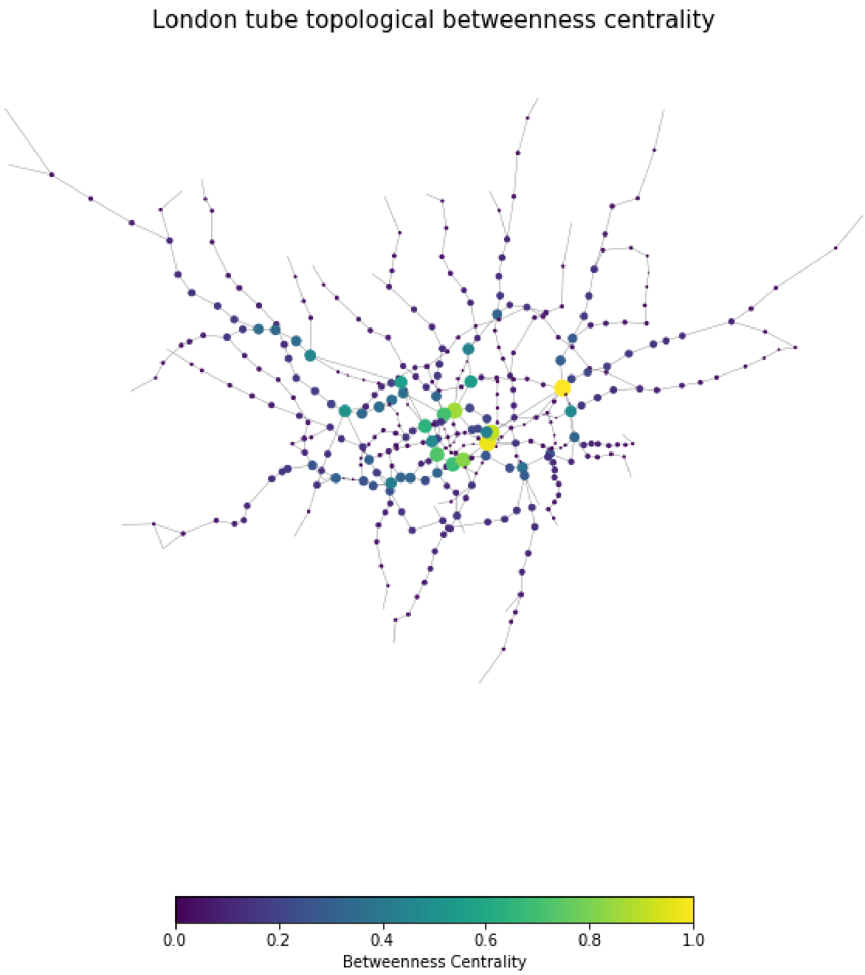

Figure 1.3: Topological Tube network map - Betweenness centrality

To summarise(Figura 1.4), Stratford is the station with the highest degree and betweenness topological centrality in London's Tube Network, while Bank and Monument is the node with the highest eigenvector centrality. Overall, Stratford, Bank and Monument and Liverpool Street represent a significant position in the three rankings, showing a concentration of important nodes in London's central/east region. All the values are normalised to allow a comparison among measures.

rank	station_degree	degree	station_eigenvector	eigen	station_betweness	betweenness
1	Stratford	0.0225	Bank and Monument	0.383725	Stratford	0.297846
2	Bank and Monument	0.0200	Liverpool Street	0.329191	Bank and Monument	0.290489
3	Baker Street	0.0175	Stratford	0.269574	Liverpool Street	0.270807
4	King's Cross St. Pancras	0.0175	Waterloo	0.249708	King's Cross St. Pancras	0.255307
5	Waterloo	0.0150	Moorgate	0.215343	Waterloo	0.243921
6	Earl's Court	0.0150	Green Park	0.197023	Green Park	0.215835
7	Liverpool Street	0.0150	Oxford Circus	0.183441	Euston	0.208324
8	Oxford Circus	0.0150	Tower Hill	0.171839	Westminster	0.203335
9	Green Park	0.0150	Westminster	0.168368	Baker Street	0.191568
10	Canning Town	0.0150	Shadwell	0.159233	Finchley Road	0.165085

Figure 1.4: Centrality measures

Impact Measures

In order to analyze the impact on the entire network and observe its behaviour when nodes are removed, two measures were selected to evaluate the network components. However, a review conducted by Jamakovic and Uhlig 2008 found that most topological measures discussed only focus on the largest components. In this study, the analysis specifically considers measures such as distance and clustering to better understand the network's dynamics and characteristics.

Thus, this analysis uses the Size Largest Component and the Number of Components. The current network has only 1 component. However, when the stations start to be removed, some links will be disconnected, and the number of components will increase. Therefore, it is crucial to choose measures that can detect the impact of removing that station, considering the rearrangement of the topological network.

Node Removal

In order to measure the impact of nodes removal in sequential and non-sequential order, a was developed a loop through values of n from 0 to 10 and then stored the results of the number of nodes removed, the impact measures, the total o nodes and consequently, which station was removed from the system. Thus, this process was applied to the three centrality measures and merged into a unique dataset. Therefore, it is possible to analyse the impact of node removal in the graphs below (Figure 1.5).

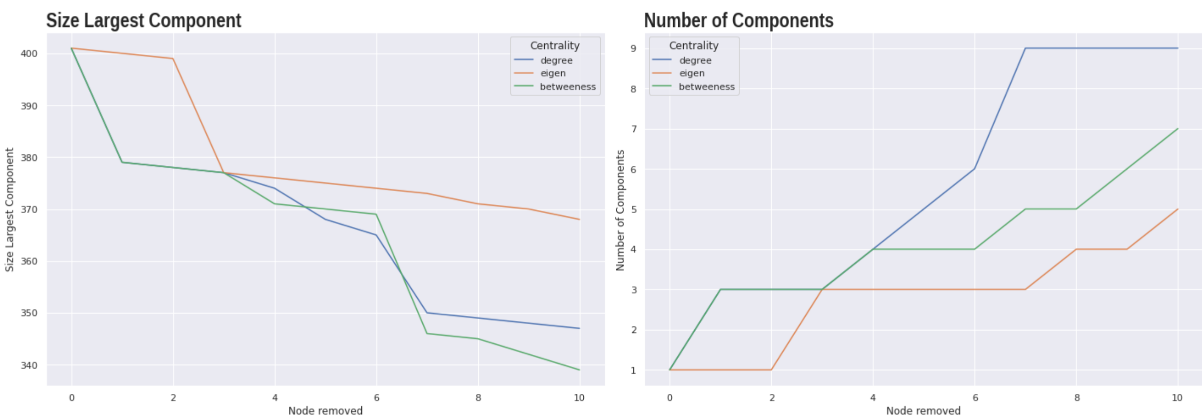

Figure 1.5: Non-sequential node removal

For the sequential removal nodes(Figure 1.6), betweenness centrality had a sharp drop in the sizer of the largest component after removing the 6th node in the system(Green Park). The largest component's size variation remained around the initial value in the other centralities without drastic drops or increases.

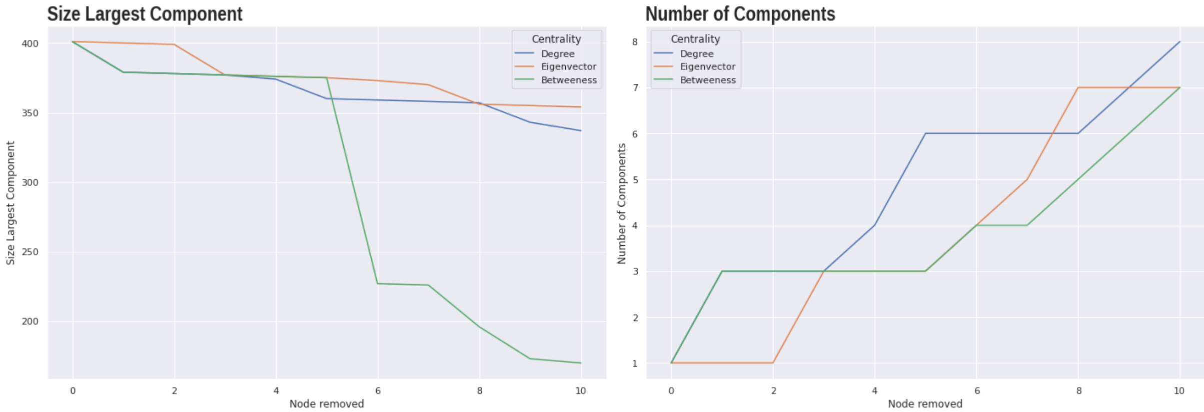

Figure 1.6: Sequential node removal

1.2. Flows: weighted network

Centrality Measures

The weighted network outcomes performed a different ranking behaviour, mainly for the betweenness centrality. *West Hampstead*, *Gospel Oak* and *Finchley Road & Frognal* are stations that were not present in the topological network ranking(Figure 1.7). This situation occurred because the only centrality measure considering the flows in its equation was betweenness centrality. As a result, other stations exhibited greater importance when exclusively considering this centrality measure. However, it is crucial to determine whether the removal of nodes will also significantly impact these newly identified stations within the network.

	rank	station_degree	degree	station_eigenvector	eigen	station_betweness	betweenness_w
0	1	Stratford	0.0225	Bank and Monument	0.383725	West Hampstead	0.367857
1	2	Bank and Monument	0.0200	Liverpool Street	0.329191	Gospel Oak	0.297428
2	3	Baker Street	0.0175	Stratford	0.269574	Finchley Road & Frognal	0.288036
3	4	King's Cross St. Pancras	0.0175	Waterloo	0.249708	Hampstead Heath	0.287105
4	5	Waterloo	0.0150	Moorgate	0.215343	Willesden Junction	0.261078
5	6	Earl's Court	0.0150	Green Park	0.197023	Stratford	0.259261
6	7	Liverpool Street	0.0150	Oxford Circus	0.183441	Brondesbury	0.236494
7	8	Oxford Circus	0.0150	Tower Hill	0.171839	Brondesbury Park	0.235056
8	9	Green Park	0.0150	Westminster	0.168368	Kensal Rise	0.233628
9	10	Canning Town	0.0150	Shadwell	0.159233	Turnham Green	0.175182

Figure 1.7: Flows - Centrality measures

Impact Measures

For this analysis, it will remain the number of component and size largest components, in addition to the Flows of the largest component. Aiming to understand the impact in the flows, removing a tube station.

Node Removal

The Degree centrality ranking, the measure number of components and the size largest component did not have different values. However, the flows of the largest component presented crucial information about the importance of removed stations in the networking flows(Figure 1.8).

	Node removed	Number of Components	Size Largest Component	Flows Largest Component	Total Nodes	Station
0	0	1	401	9830695	401	[]
1	1	3	379	9274809	401	[Stratford]
2	2	3	378	8605744	400	[Bank and Monument]
3	3	3	377	8140045	399	[King's Cross St. Pancras]

Figure 1.8: Weighted Network - Degree Centrality

For eigenvector centrality, Stratford station caused one of the highest impacts on the flow in the network, removing a considerable amount of flow in addition to separating the network into three components(Figure 1.9).

	Node removed	Number of Components	Size Largest Component	Flows Largest Component	Total Nodes	Station
0	0	1	401	9830695	401	[]
1	1	1	400	9161630	401	[Bank and Monument]
2	2	1	399	8887734	400	[Liverpool Street]
3	3	3	377	8496352	399	[Stratford]

Figure 1.9: Weighted Network - Eigenvector Centrality

Although the betweenness centrality measure for Tube networking includes stations that differ from other centrality measures, removing these stations does not significantly impact the flow of network traffic or the size of the largest component. Consequently, when evaluating the resilience of the tube system by incorporating flows, betweenness centrality may not be a suitable centrality measure.

(Figure 1.10).

	Node removed	Number of Components	Size Largest Component	Flows Largest Component	Total Nodes	Station
0	0	1	401	9900993	401	[]
1	1	1	400	9848574	401	[West Hampstead]
2	2	2	397	9826168	400	[Gospel Oak]
3	3	2	397	9826168	399	[Finchley Road & Frognal]

Figure 1.10: Weighted Network - Betweenness Centrality

Considering the significant impact observed with degree centrality and eigenvector centrality, it becomes apparent that Stratford, Bank and Monument, Liverpool Street, and King’s Cross stations play crucial roles in facilitating user flows within the system during the specified period. Conversely, the size of the largest component and the number of components did not prove to be relevant measures for evaluating the flows in the system. However, these measures are still valuable for assessing the behaviour of other metrics, such as the flows within the largest component.

2. PART 2: SPATIAL INTERACTION MODELS

2.1. Models and calibration

Gravity Model(Unconstrained model), and the constrained models(Production-constrained Model, Attraction-Constrained Model and Doubly Constrained model)

- Gravity model: The gravity model assumes the flows between the origin and destination are proportional to the mass of the origin and destination and inversely proportional to the distance between them. That means if the distance decreases, the flow increase and if the flows decrease, the origin and destination mass decrease(Wilson 1971; Senior 1979).

$$T_{ij} = kO_i^{\alpha} D_j^{\gamma} d_{ij}^{-\beta} \tag{1}$$

In unconstrained models, the constant K is employed to ensure the inclusion of all values. However, in constrained models, we can add the sums of origin and destination separately, providing greater flexibility to test the model in more specific scenarios. This allows for a more nuanced evaluation of the model's performance.

- Production-Constrained Model:

The Production(origin) Constrained Spatial Interaction Model is a powerful framework for understanding how origins behave and commute within a city to reach points of interest. These estimated flows can be connected to different variables related to these attractive points, such as budget and store size. In the context of London's tube network, the flows can be estimated using the equation below, which incorporates the logarithm of job values and the distance(cost)(Pooler 1994).

$$T_{ij} = A_i O_i D_j^{\gamma} d_{ij}^{-\beta} \tag{2}$$

- Attraction-Constrained Model

The Attraction(destination) Constrained model plays a crucial role in assessing the impact of a significant increase in demand within a specific region, such as establishing a new university or creating many job opportunities. By applying the value of this increase to the model, we can estimate how this new pattern will affect commuting flows and gain insights into the origins of these flows towards the destination. When applied to the tube network, the model considers the logarithm values of population, distance (cost), and destination stations, resulting and the relationship of all these values with the flows in the system(ibid.).

$$T_i j = D_j B_j O_i^{\alpha} d_{ij}^{-\beta} \tag{3}$$

- Doubly Constrained Model In the doubly constrained model only considered the flows and the distance, in addition, to understanding the origin and destination in the same equation. First, it is crucial to detect a reasonable value of beta. Thus, the beta is used to calculate the new origin and destination flows. Beta values do not change unless the mode of transport changes.

$$T_{ij} = A_i B_j O_i D_j d_{ij}^{-\beta} \tag{4}$$

Thus, the Production-constrained model was selected to analyse London's tube network. In order to determine the most suitable model, both the attraction-constrained and production-constrained models were tested to address the specific scenarios involving a 50% reduction in jobs in Canary Wharf and an increase in transportation costs(Flowerdew and Lovett 2010). While the attraction-constrained model is primarily concerned with the demand generation for a destination like work in Canary Wharf, it was found that the production-constrained model produced a higher R-squared value and could be validated for both scenarios(Figure 2.1, Figure2.2, and Table 2.1).

In the context of the Production-constrained model(Pooler 1994), even though it is primarily associated with the origin of flows, such as the tube stations, the reduction in job opportunities in Canary Wharf necessitates the redistribution of flows from these areas to other city regions. Therefore, this model becomes essential for understanding the new dynamics of flow origins and reallocating attraction flows towards available job opportunities.

station_destination	Abbey Road	Acton Central	Acton Town	Aldgate	Aldgate East	...	Woodford	Woodgrange Park	Woodside Park	Woolwich Arsenal	All
station_origin											
Finsbury Park	0	0	19	156	39	...	0	0	2	0	24735
Canada Water	0	0	17	0	0	...	3	0	4	0	27026
King's Cross St. Pancras	0	0	22	842	160	...	3	0	12	0	28307
Bank and Monument	0	0	6	31	250	...	10	0	9	509	29494
Liverpool Street	0	0	12	384	150	...	12	0	1	0	31473
London Bridge	0	0	24	20	47	...	5	0	9	0	32593
Victoria	0	0	30	59	153	...	7	0	26	0	37517
Stratford	223	25	7	18	186	...	133	0	8	1106	59311
Waterloo	0	0	14	108	381	...	12	0	18	0	67314
All	345	750	2202	7782	7932	...	706	242	745	4428	1542283

Figure 2.1: Model before calibration

In order to build the Production-constrained model, the following variables were applied in the equation: Flows, Jobs, Population and Distance as the cost(Neira 2022).

station_destination	Abbey Road	Acton Central	Acton Town	Aldgate	Aldgate East	...	Woodford	Woodgrange Park	Woodside Park	Woolwich Arsenal	All
station_origin											
Abbey Road	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	5.0	599.0
Acton Central	NaN	NaN	NaN	NaN	NaN	...	NaN	1.0	NaN	NaN	1223.0
Acton Town	NaN	NaN	NaN	18.0	18.0	...	2.0	NaN	2.0	NaN	3749.0
Aldgate	NaN	NaN	2.0	NaN	47.0	...	1.0	NaN	1.0	NaN	2882.0
Aldgate East	NaN	NaN	2.0	52.0	NaN	...	1.0	NaN	1.0	NaN	3169.0
...	...	...	...	...	...	...	...	...	...	...	...
Woodford	NaN	NaN	7.0	35.0	39.0	...	NaN	NaN	NaN	NaN	4868.0
Woodgrange Park	NaN	4.0	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	532.0
Woodside Park	NaN	NaN	5.0	20.0	20.0	...	NaN	NaN	NaN	NaN	3092.0
Woolwich Arsenal	29.0	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	7890.0
All	445.0	391.0	2154.0	8752.0	9305.0	...	654.0	156.0	603.0	1009.0	1541811.0

Figure 2.2: Model after calibration

Measure	Value
R-squared	0.38
RMSE	102.82

Table 2.1: Model after calibration

2.2. Scenarios

Scenario A

Assuming the Canary Wharf decreased in 50% the total number of jobs, the flows were recalculated, reducing from 58,772 to 29,386. Thus, a new value for gamma and beta was calculated and then generated Ai value. As a result, the total number of flows only had a small variation. Furthermore, it is possible to detect that there was a rebalancing of the flows in which the source flows increased from 572 to 601 while the destination flows increased from 445 to 470(Figure 2.3).

station_destination station_origin	Abbey Road	Acton Central	Acton Town	Aldgate	Aldgate East	...	Woodford	Woodgrange Park	Woodside Park	Woolwich Arsenal	All
Abbey Road	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	6.0	601.0
Acton Central	NaN	NaN	NaN	NaN	NaN	...	NaN	1.0	NaN	NaN	1223.0
Acton Town	NaN	NaN	NaN	18.0	18.0	...	2.0	NaN	2.0	NaN	3744.0
Aldgate	NaN	NaN	2.0	NaN	47.0	...	1.0	NaN	1.0	NaN	2885.0
Aldgate East	NaN	NaN	2.0	53.0	NaN	...	1.0	NaN	1.0	NaN	3168.0
...	...	...	...	...	...	...	...	...	...	...	...
Woodford	NaN	NaN	7.0	36.0	39.0	...	NaN	NaN	NaN	NaN	4863.0
Woodgrange Park	NaN	4.0	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	532.0
Woodside Park	NaN	NaN	5.0	21.0	21.0	...	NaN	NaN	NaN	NaN	3091.0
Woolwich Arsenal	32.0	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	7895.0
All	470.0	393.0	2181.0	8836.0	9398.0	...	666.0	156.0	609.0	1081.0	1541964.0

Figure 2.3: Scenario A

Scenario B

Measure 1

Assuming an increase in cost, which in this model refers to distance, two scenarios were simulated. The first scenario involved a 30% increase, while the second involved a 50% increase. Simulations were also conducted with a 10% increase, but there were limitations in validating the results. Therefore, more significant increases were chosen to evaluate the system's behaviour. In the simulation assuming a 30% increase in cost, an equilibrium was achieved in the estimated flows of origin and destination, with flow values around 600, without impacting the total number of flows(Figure 2.4).

station_destination station_origin	Abbey Road	Acton Central	Acton Town	Aldgate	Aldgate East	...	Woodford	Woodgrange Park	Woodside Park	Woolwich Arsenal	All
Abbey Road	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	3.0	600.0
Acton Central	NaN	NaN	NaN	NaN	NaN	...	NaN	1.0	NaN	NaN	1228.0
Acton Town	NaN	NaN	NaN	16.0	16.0	...	1.0	NaN	2.0	NaN	3746.0
Aldgate	NaN	NaN	1.0	NaN	54.0	...	1.0	NaN	1.0	NaN	2880.0
Aldgate East	NaN	NaN	1.0	63.0	NaN	...	1.0	NaN	1.0	NaN	3166.0
...	...	...	...	...	...	...	...	...	...	...	...
Woodford	NaN	NaN	6.0	35.0	40.0	...	NaN	NaN	NaN	NaN	4872.0
Woodgrange Park	NaN	3.0	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	531.0
Woodside Park	NaN	NaN	5.0	20.0	20.0	...	NaN	NaN	NaN	NaN	3094.0
Woolwich Arsenal	29.0	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	7890.0
All	599.0	371.0	2124.0	8848.0	9360.0	...	628.0	176.0	554.0	870.0	1541797.0

Figure 2.4: Measure 1 - Increasing 30%

Measure 2

The second measure produced more notable outcomes, revealing a discernible impact on the origin and destination flows. Specifically, when transport costs were increased by 50%, resulting in longer distances for commuting to work in the city, the origin flows increased to 746 while the destination flows remained nearly unchanged. As a result, many users discontinued their use of the system due to a 50% increase in travel costs(Figure 2.5). Thus, scenario B presented the most significant flow change, especially when applying a 50% cost increase in the model.

station_destination	Abbey Road	Acton Central	Acton Town	Aldgate	Aldgate East	...	Woodford	Woodgrange Park	Woodside Park	Woolwich Arsenal	All
station_origin											
Abbey Road	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	3.0	597.0
Acton Central	NaN	NaN	NaN	NaN	NaN	...	NaN	1.0	NaN	NaN	1226.0
Acton Town	NaN	NaN	NaN	14.0	14.0	...	1.0	NaN	1.0	NaN	3740.0
Aldgate	NaN	NaN	1.0	NaN	58.0	...	0.0	NaN	0.0	NaN	2873.0
Aldgate East	NaN	NaN	1.0	70.0	NaN	...	1.0	NaN	0.0	NaN	3158.0
...	...	...	...	...	...	...	...	...	...	...	...
Woodford	NaN	NaN	5.0	35.0	40.0	...	NaN	NaN	NaN	NaN	4862.0
Woodgrange Park	NaN	2.0	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	527.0
Woodside Park	NaN	NaN	4.0	19.0	19.0	...	NaN	NaN	NaN	NaN	3092.0
Woolwich Arsenal	29.0	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	7893.0
All	746.0	368.0	2162.0	8805.0	9292.0	...	628.0	194.0	579.0	792.0	1541547.0

Figure 2.5: Measure 1 - Increasing 50%

REFERENCES

Arcaute, Elsa and Valentina Marin (2022). *Review of centrality measures in weighted networks. Jupyter Notebook*.

Flowerdew, Robin and Andrew Lovett (2010). “Fitting Constrained Poisson Regression Models to Interurban Migration Flows”. en. In: *Geographical Analysis* 20.4, pp. 297–307. ISSN: 00167363. DOI: 10.1111/j.1538-4632.1988.tb00184.x.

Jamakovic, Almerima and Uhlig (2008). “On the relationships between topological measures in real-world networks”. en. In: *Networks & Heterogeneous Media* 3.2, pp. 345–359. ISSN: 1556-181X. DOI: 10.3934/nhm.2008.3.345.

Neira, Mateo (2022). *Working with real-world data. Jupyter Notebook*.

NetworkX (2023). *Centrality — NetworkX 3.1 documentation*. URL: https://networkx.org/documentation/stable/reference/algorithms/centrality.html (visited on 05/07/2023).

Pooler, Jim (1994). “An extended family of spatial interaction models”. en. In: *Progress in Human Geography* 18.1, pp. 17–39. ISSN: 0309-1325, 1477-0288. DOI: 10.1177/030913259401800102.

Senior, Martyn L. (1979). “From gravity modelling to entropy maximizing: a pedagogic guide”. en. In: *Progress in Human Geography* 3.2, pp. 175–210. ISSN: 0309-1325, 1477-0288. DOI: 10.1177/030913257900300218.

Wilson, A G (1971). “A Family of Spatial Interaction Models, and Associated Developments”. en. In: *Environment and Planning A: Economy and Space* 3.1, pp. 1–32. ISSN: 0308-518X, 1472-3409. DOI: 10.1068/a030001.

To access GitHub Repository Link