Urban Simulation

Word count: 2909

Part 1 London's underground resilience

I. Topological network

I.1 Centrality Measures

1. Degree Centrality: The degree centrality $C_D(i)$ for a node i is given by Equation 1, where k_i is the degree of node i, and N is the total number of nodes in the network. It measures the number of direct connections a node has. In the context of the London Underground network, stations with high degree centrality are important local hubs that are directly connected to many other stations.

$$C_D(i) = \frac{k_i}{N-1} \tag{1}$$

where k_i is the degree of node i, and N is the total number of nodes in the network.

2. Closeness Centrality: The closeness centrality $C_C(i)$ for a node i is given by Equation 2, where $d(i,j)$ is the shortest path distance between nodes i and j, and N is the total number of nodes in the network. It measures the reciprocal of the sum of the shortest path distances from a node to all other nodes. In the context of the London Underground network, stations with high closeness centrality are more central because they minimize the average shortest path distance to all other stations.

$$C_C(i) = \frac{N-1}{\sum_{j \neq i} d(i,j)} \tag{2}$$

where $d(i,j)$ is the shortest path distance between nodes i and j, and N is the total number of nodes in the network.

3. Betweenness Centrality: The betweenness centrality $C_B(i)$ for a node i is defined by Equation 3, where $\sigma_{st}(i)$ is the number of shortest paths between nodes s and t that pass through node i, and σ_{st} is the total number of shortest paths between nodes s and t. It quantifies the importance of a node in acting as a bridge or intermediary between other nodes. In the London Underground network, stations with high betweenness centrality are crucial for maintaining efficient passenger flow and network functionality, as they play a vital role in connecting different parts of the network.

Betweenness Centrality:

$$C_B(i) = \sum_{s \neq i \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}} \tag{3}$$

where $\sigma_{st}(i)$ is the number of shortest paths between nodes s and t that pass through node i, and σ_{st} is the total number of shortest paths between nodes s and t.

By computing these centrality measures for the London Underground network, we can identify the stations that are most crucial for maintaining connectivity, efficiency, and minimizing disruptions. Removing the stations with the highest centrality scores according to each measure would allow us to assess the network's vulnerability and resilience to targeted attacks or failures. Table 1 shows the first 10 ranked nodes for each of the 3 measures.

Rank	Degree Centrality	Closeness Centrality	Betweenness Centrality
1	Stratford	Green Park	Stratford
2	Bank and Monument	Bank and Monument	Bank and Monument
3	King's Cross St. Pancras	King's Cross St. Pancras	Liverpool Street
4	Baker Street	Westminster	King's Cross St. Pancras
5	Earl's Court	Waterloo	Waterloo
6	Oxford Circus	Oxford Circus	Green Park
7	Liverpool Street	Bond Street	Euston
8	Waterloo	Farringdon	Westminster
9	Green Park	Angel	Baker Street
10	Canning Town	Moorgate	Finchley Road

Table 1: Top 10 Ranked Nodes for each Centrality Measure

I.2 Measures to assess impact

To assess the consequences of node removal on the London Underground, we employ two global measures.:

Average Shortest Path Length (ASPL) ASPL represents the mean value of the minimum path distances among all node pairs within the network, reflecting the network's overall connectivity and efficiency of the network. The calculation formula for ASPL is:

$$ASPL = \frac{1}{N(N-1)} \sum_{i \neq j} d(i, j) \tag{4}$$

where N signifies the total count of nodes in the network, and $d(i, j)$ denotes the shortest path distance between nodes i and j.

In the context of the London Underground network, ASPL symbolizes the average minimum number of stations a commuter needs to travel between any two given stations. If removing a station causes a significant increase in the ASPL, it indicates that the station was crucial for maintaining short and efficient travel routes throughout the network (Latora and Marchiori, 2001).

Global Efficiency Global efficiency measures the efficiency of information transmission in the network and is the average of the reciprocals of the shortest path distances between all pairs of nodes in the network. The calculation formula for global efficiency is:

$$E_{glob} = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d(i,j)} \tag{5}$$

where N and $d(i, j)$ are defined the same as ASPL.

In the London Underground network, global efficiency quantifies the efficiency of passengers navigating the system. The higher the global efficiency, the faster and easier passengers can reach their destinations. If removing a station leads to a significant decrease in global efficiency, it indicates that the station was essential for maintaining efficient passenger flow and connectivity across the network (Latora and Marchiori, 2001).

Both ASPL and global efficiency are not specific to the London Underground network and can be used to evaluate the resilience of any other connected network. These measures capture the overall connectivity and efficiency of a network, which are essential aspects of resilience. In any connected network, an increase in ASPL or a decrease in global efficiency after node removal indicates that the network's performance and resilience have been compromised. Therefore, these measures are suitable for assessing the resilience of various networks, such as transportation systems, communication networks, or social networks, where maintaining connectivity and efficiency is crucial.

I.3 Node Removal

From Figures 1 and 2, regardless of sequential or non-sequential removal, all three Centrality Measures maintain a decline trend in Global Efficiency. However, Closeness Centrality and Betweenness Centrality in the sequential removal show a significant decrease in ASPL after the removal of the 5th node, followed by a fluctuating increase. This may be because after the removal of this node, the connectivity of the network decreases, leading to the disappearance of some shortest paths. As more nodes are removed, the shortest paths between the remaining nodes become longer, and ASPL rises again. At the same time, Degree Centrality also maintains a stable increase in ASPL, with the largest increase, indicating that Degree Centrality is more sensitive to node removal in the network. As a result, its stable performance in two different impact measures reflects its higher fault tolerance, and the significant changes in them after node removal can well reflect the significance of a station for the functioning of the underground.

However, although Degree Centrality is the best choice, it also has limitations. Due to the simplicity of the calculation method, it is easy to have stations with the same value, which makes it difficult for us to rank stations with the same Degree Centrality value. This is a major flaw when considering using it for non-sequential removal.

The figures clearly illustrate that Global Efficiency serves as a superior measure for evaluating the extent of damage following node removal. First, all three Centrality Measures show stable trends in Global Efficiency, making it easier for them to capture the impact, while the complex patterns shown in ASPL make it more complex to assess the damage. Second, Global Efficiency is more sensitive to the degree distribution of nodes in the network. Nodes with higher degrees can maintain network connectivity to a certain extent, meaning that passengers can travel between most stations with fewer transfers. Global Efficiency can capture this transmission characteristic of the network. As a distance-based topological indicator, ASPL cannot well reflect the actual transmission performance of the network.

Sequential removal is a more effective strategy for studying resilience. First, after station removal, the redistribution of network flow is more in line with reality. Moreover, the figures for Global Efficiency show that regardless of the Centrality Measures used, the decline of sequential removal is more significant than that of non-sequential removal, indicating that sequential removal is more sensitive and, therefore, a better removal strategy.

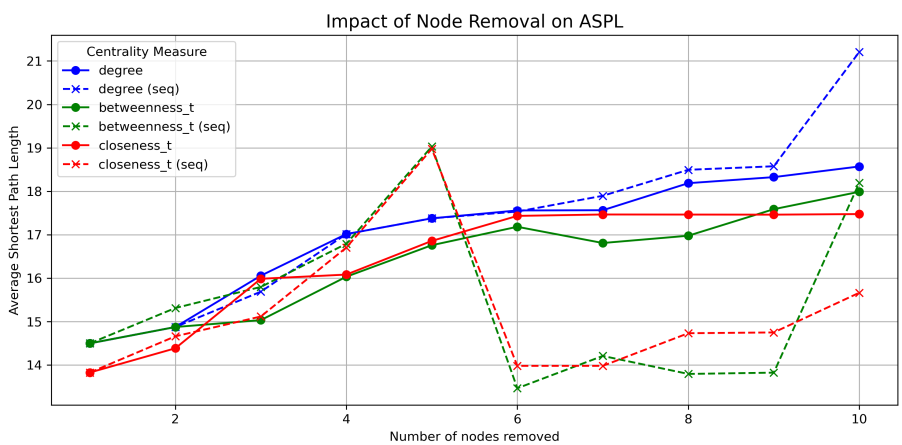

Figure 1: Node Removal on ASPL

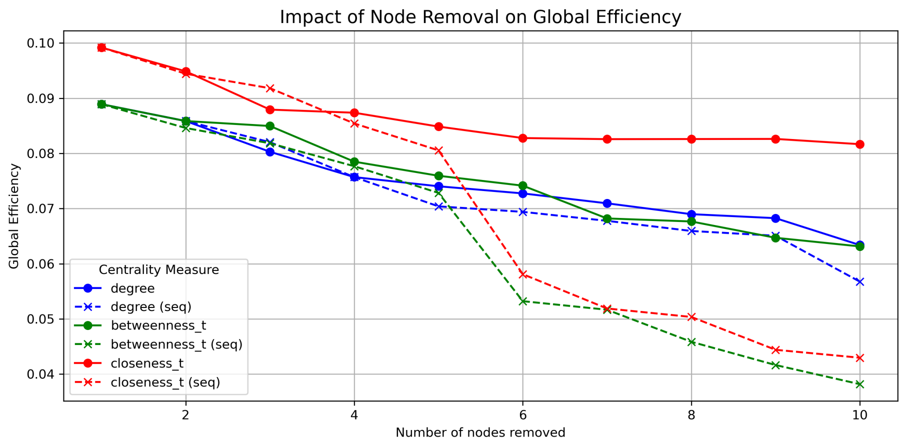

Figure 2: Node Removal on Global Efficiency

II. Flows: weighted network

II.1 Old vs new measure

Considering passenger flow, the calculation methods of centrality metrics in the weighted subway network need to be adjusted accordingly.

For Degree Centrality, since it only focuses on the number of direct connections of a node and is not related to the passenger flow weights on the edges, the calculation method is consistent with the unweighted network.

However, Betweenness Centrality and Closeness Centrality need to consider the impact of flow. Since the larger the flow, the higher the importance of two stations, we first use the inverse of the flow as the edge weight to correctly reflect the importance of stations, and then normalize the reciprocal weights to improve numerical stability.

The ranking obtained by recomputation is shown in Table 2. Compared with Table 1, Bank and Monument remain unchanged in the ranking of Betweenness Centrality, and Green Park and Moorgate remain unchanged in the ranking of Closeness Centrality.

Rank	Degree Centrality	Betweenness Centrality	Closeness Centrality
1	Stratford	Green Park	Green Park
2	Bank and Monument	Bank and Monument	Westminster
3	King's Cross St. Pancras	Waterloo	Waterloo
4	Baker Street	Westminster	Bank and Monument
5	Earl's Court	Liverpool Street	Oxford Circus
6	Oxford Circus	Stratford	Bond Street
7	Liverpool Street	Bond Street	Victoria
8	Waterloo	Euston	Liverpool Street
9	Green Park	Oxford Circus	Warren Street
10	Canning Town	Warren Street	Moorgate

Table 2: Top 10 Nodes of Adjusted Measures

II.2 Impact measure with flows

Based on the research of Latora and Marchiori (2001), in order to better assess the impact of station closure when considering passenger flow, we use Weighted Global Efficiency as a measure, taking into account the edge weights, i.e., the passenger flow between stations.

The calculation formula for Weighted Global Efficiency is:

$$E_{glob}^{w} = \frac{1}{N(N-1)} \sum_{i,j \in G, i \neq j} \frac{1}{d_{ij}^{w}} \tag{6}$$

where:

1. N represents the total number of nodes in the network.
2. i and j represent any two different nodes in the network.
3. G represents the network.
4. d_{ij}^{w} represents the shortest weighted path length between node i and node j. It is the minimum value of the sum of the weights of all edges on the path connecting nodes i and j. If there is no path connecting nodes i and j, then $d_{ij}^{w} = \infty$.

Similarly, we can also add passenger flow as edge weights to ASPL.

II.3 Experiment with flows

According to the requirements, Degree Centrality performs the best in I.1, so we remove the top three according to its ranking order and recalculate the Weighted Global Efficiency and Weighted ASPL for each removal. According to Figure 3, after removing the Bank and Monument nodes, the Weighted Global Efficiency decreases the fastest and the Weighted ASPL increases the fastest, indicating that its closure will have the greatest impact on passengers.

However, although we have made a preliminary judgment through the changes of Weighted Global Efficiency and Weighted ASPL, the cumulative effect of node removal is still worth considering. For example, does the decline in network efficiency intensify after removing the second or third node? This can help us understand the cascading effects of node failures (Crucitti et al., 2004). In addition, we can also consider using more network efficiency indicators for comprehensive evaluation, such as network vulnerability (Latora and Marchiori, 2005), to make our conclusions about the impact of node removal more robust.

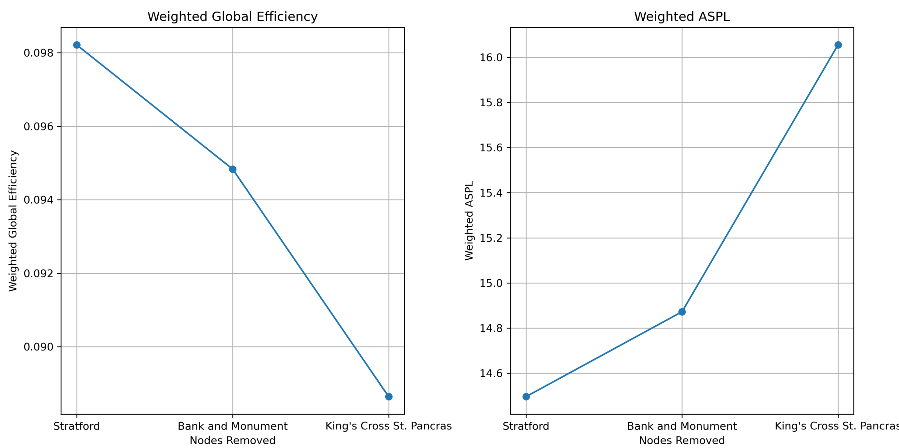

Figure 3: Node removal on Weighted Global Efficiency and Weighted ASPL

III. Models and calibration

III.1. Spatial interaction models

Newton's law of universal gravitation laid the foundation for spatial interaction models. The law of universal gravitation reveals the universal interaction forces between objects in space and provides a mathematical expression of the quantitative relationship between gravitational force, mass, and distance. This idea inspired later generations to use similar methods to study the spatial interaction patterns of human activities. On this basis, geographer Wilson (1971) proposed the standard form of the spatial interaction model to predict the scale of interaction between two regions.

Unconstrained Model The Unconstrained Model is the simplest form of the spatial interaction model. It assumes that the scale of interaction depends only on the size of the origin and destination and the distance between the two points, without considering the supply constraints of the origin or the capacity constraints of the destination. This model is suitable for exploring the general patterns of spatial interaction.

$$T_{ij} = k\frac{O_i^{\alpha} D_j^{\gamma}}{d_{ij}^{\beta}} \tag{7}$$

Or

$$T_{ij} = kO_i^{\alpha} D_j^{\gamma} d_{ij}^{-\beta} \tag{8}$$

T_{ij}: The interaction flow or scale from origin i to destination j.
O_i: The size or importance of origin i.
D_j: The size or attractiveness of destination j.
d_{ij}: The distance or cost between origin i and destination j.
k: This is a proportionality constant used to adjust the overall scale of the model.
α: The degree of influence of the origin size on the interaction scale.
β: The degree of influence of distance on the interaction scale.
γ: The degree of influence of the destination size on the interaction scale.

Singly-Constrained Model The Singly-Constrained Model introduces constraints on the total outflow from the origin (Production-Constrained) or the total inflow to the destination (Attraction-Constrained) based on the Unconstrained Model. In the Production-Constrained Model, each origin has a fixed number of people, goods, or information that need to be allocated to different destinations, and the attractiveness of the destination is determined by its size and distance. This model is suitable for describing scenarios where the origin supply is limited, such as labor migration and commodity exports. Conversely, in the Attraction-Constrained Model, each destination has a fixed receiving capacity, and the size and distance of the origin determine its influence on the destination.

Production-Constrained:

$$T_{ij} = A_i O_i D_j^{\gamma} d_{ij}^{-\beta} \tag{9}$$

A_i: The balancing factor of origin i, which adjusts the interaction intensity of origin i to meet the constraint conditions.

O_i represents the fixed total interaction amount or supply of origin i. It is no longer a measure of the origin size but a known constant.

Other parameters are the same as in the Unconstrained Model. Attraction-Constrained:

$$T_{ij} = D_j B_j O_i^{\alpha} d_{ij}^{-\beta} \tag{10}$$

In the Attraction-Constrained Model, each destination j has a fixed total interaction amount or demand D_j, and the model introduces the balancing factor B_j of the destination to ensure that the sum of all interaction flows into destination j is equal to D_j.

Doubly-Constrained Model The Doubly-Constrained Model simultaneously considers the supply constraints of the origin and the capacity constraints of the destination. The model assumes that the origin has a fixed supply and the destination also has a fixed receiving capacity, and spatial interaction must be carried out under these two constraints.

$$T_{ij} = A_i O_i B_j D_j^{\gamma} d_{ij}^{-\beta} \tag{11}$$

The Doubly-Constrained Model achieves double constraints on spatial interaction by introducing the balancing factors of the origin and destination (A_i and B_j) and the fixed supply and demand (O_i and D_j). This enables the model to more accurately describe the interaction patterns of supply and demand balance in the real world. However, due to the existence of dual constraints, iterative algorithms are needed to solve A_i and B_j simultaneously (Senior, 1979).

III.2. Calibration of model

We choose the Production-Constrained model for the following reasons:

1. In this problem, we know the population (production) of each station and the commuting flow matrix, but we do not know the number of jobs (attraction) at each station. The Production-Constrained model is suitable for situations where production is known and attraction is unknown.

2. The Production-Constrained model assumes that the commuting flow generated by each station is limited by the population of that station, and the attracted commuting flow is proportional to the number of jobs. This is consistent with the actual situation.

In summary, the Production-Constrained model can make good use of the known conditions and reflect the relationship between population and commuting flow. On this basis, since commuters pay more attention to the total travel time and accessibility, the distance between stations has a relatively small impact on travel choices (Guti'errez et al., 2011). Therefore, we use a negative exponential function for distance. At the same time, since the flow follows a Poisson distribution, we use Poisson regression. The calculated calibrated β is 0.00015, and γ is 0.7509.

The calibrated distance decay parameter $\beta = 0.00015$ indicates that the distance between stations has a relatively small impact on subway travel choices, which is consistent with the research conclusion that passengers pay more attention to travel time and accessibility (Guti'errez et al., 2011). The estimated value of the job attraction factor $\gamma = 0.7509$ indicates that for every 1% increase in the number of jobs, the increase in commuting flow is less than 1%, showing an inelastic characteristic. The calculated R^2 is 0.448 and the RMSE is 97.845, indicating a moderate fit of the model.

IV. Scenarios

IV.1. Scenario A

In this scenario, we assume that after Brexit, jobs in Canary Wharf are reduced by 50%. To analyze the new flows, we keep β and γ unchanged in the Production-Constrained model, which ensures that the number of commuters remains constant when jobs decrease. While keeping β and γ constant, we reduce the number of jobs (D_j) in Canary Wharf by 50%, from 58,772 to 29,386, and then recalculate the balancing factor A_i to obtain the new commuting flow distribution T_{ij}.

IV.2. Scenario B

In the scenario where transportation costs are assumed to increase significantly, we still use the Production-Constrained model. To reflect this scenario, we increase β by 50% and 100% in the cost function as the β values for Scenarios B1 and B2, respectively. Therefore, β_1 is (0.00015×1.5), and β_2 is (0.00015×2). Based on this, we calculate A_i and then compute the flow estimates for the two scenarios.

IV.3. Analysis

We compare the three scenarios with the original model estimates. From Figure 4, it can be seen that in Scenario A, 14.44% of the station-to-station flows increase, while in Scenarios B1 and B2, the percentages are 15.24% and 13.63%, respectively. At the same time, only 0.5% of the station-to-station flows decrease in Scenario A, while in the two scenarios with significantly increased transport costs, the percentage of station pairs with decreased flows exceeds 60%. Overall, Scenario A has the highest percentage of unchanged station pairs at 85.06%, while Scenario B2 has the lowest at only 17.61%.

In addition, as shown in Table 3, by comparing the average absolute change and total absolute change, it can be seen that Scenario A has the lowest values for both indicators, indicating the smallest impact, while Scenario B2 has the largest impact, with an average absolute change of 14.38 and a total absolute change of 895,204.

The Frobenius norm is a commonly used matrix norm that is suitable for measuring the overall magnitude of matrix elements, making it particularly appropriate for comparing differences between matrices (Golub and Loan, 2013). This norm accumulates differences over all elements and is sensitive to larger matrix changes, effectively reflecting the overall similarity or difference between matrices. For a matrix A, its Frobenius norm is defined as the square root of the sum of squares of all matrix elements:

$$|A|F = \sqrt{\sum i, j a_{ij}^2} \tag{12}$$

where a_{ij} represents the element in the i-th row and j-th column of matrix A.

We compare the matrix changes by calculating the Frobenius norm differences between the Production Constrained models. The results are shown in Table 3, with Scenario B1 having the smallest Frobenius norm difference and Scenario B2 having the largest, indicating that Scenario B2 has the greatest change.

Considering all the above results, Scenario B2, which corresponds to a 100% increase in the cost of transport, will exert the most substantial influence on the redistribution of network flows. This impact surpasses the effects of a 50% decrease in employment opportunities at Canary Wharf and a 50% rise in transportation expenses.

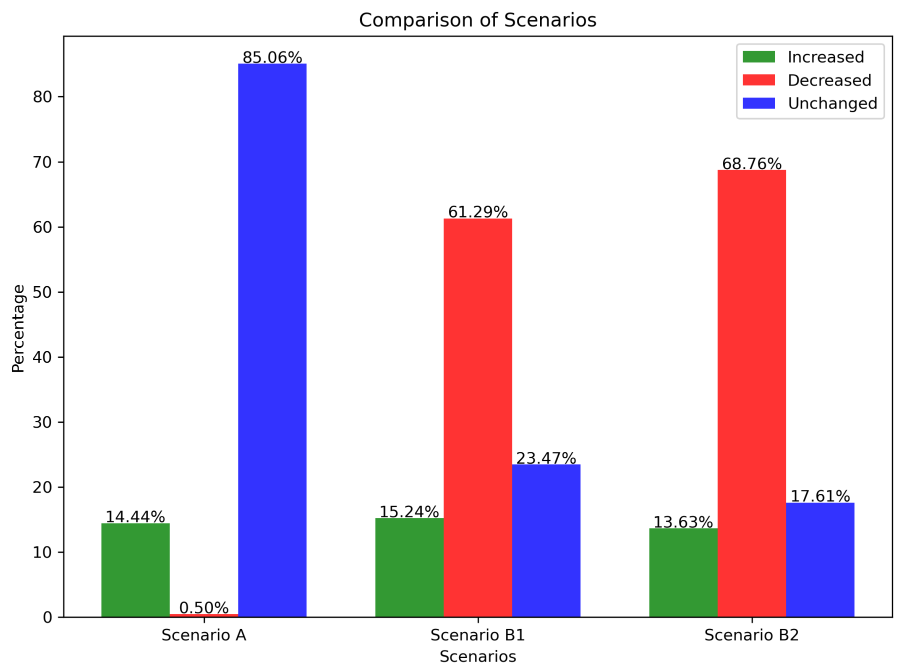

Figure 4: Comparison of flow changes of scenarios.

Table 3: Comparison of scenarios with different measures.

	Average absolute change	Total absolute change	Frobenius norm difference
Scenario A	1.20	74612	18789.38
Scenario B1	7.09	441252	9072.54
Scenario B2	14.38	895204	20271.03

References

Crucitti, P., Latora, V., Marchiori, M., 2004. Model for cascading failures in complex networks. Physical Review E 69, 045104. doi:10.1103/PhysRevE.69.045104.

Golub, G.H., Loan, C.F.V., 2013. Matrix Computations. JHU Press.

Guti'errez, J., Cardozo, O.D., Garc'ıa-Palomares, J.C., 2011. Transit ridership forecasting at station level: an approach based on distance-decay weighted regression. Journal of Transport Geography 19, 1081–1092. doi:10.1016/j.jtrangeo.2011.05.004. special section on Alternative Travel futures.

Latora, V., Marchiori, M., 2001. Efficient behavior of small-world networks. Physical Review Letters 87, 198701. doi:10.1103/PhysRevLett.87.198701.

Latora, V., Marchiori, M., 2005. Vulnerability and protection of infrastructure networks. Physical Review E 71, 015103. doi:10.1103/PhysRevE.71.015103.

Senior, M.L., 1979. From gravity modelling to entropy maximizing: a pedagogic guide. Progress in Human Geography 3, 175–210. doi:10.1177/030913257900300218.

Wilson, A.G., 1971. A family of spatial interaction models, and associated developments. Environment and Planning A: Economy and Space 3, 1–32. doi:10.1068/a030001.

GitHub Repository

The complete **CODE** for this project can be found on GitHub.