# Part 1 – I: Topological Network

## I.1 – Centrality Measures

### Degree Centrality

Degree centrality quantifies the number of direct connections a node has within a network. It is typically calculated as follows:

where is the degree of vertex , and is the total number of nodes in the graph (Freeman, 1978). This measure is critical in identifying nodes with substantial direct interactions within their networks.

In the context of the London Underground, degree centrality aids in identifying stations with numerous direct routes to others, suggesting these are critical hubs or transfer points (Barthélemy, 2011). Such stations enhance system efficiency by providing various routing options and facilitating easier transfers between lines.

|  |  |  |
| --- | --- | --- |
| node | station\_name | degree |
| 940GZZLUKSX | King's Cross St. Pancras | 0.016018 |
| 940GZZLUBST | Baker Street | 0.016018 |
| 940GZZLUOXC | Oxford Circus | 0.013730 |
| 940GZZLUGPK | Green Park | 0.013730 |
| 940GZZLUBNK | Bank | 0.013730 |
| 940GZZLUECT | Earl's Court | 0.013730 |
| 940GZZLUWLO | Waterloo | 0.013730 |
| 940GZZLUTNG | Turnham Green | 0.011442 |
| 940GZZDLCGT | Canning Town (DLR) | 0.011442 |
| 940GZZLULVT | Liverpool Street | 0.011442 |

### Betweenness Centrality

Betweenness centrality gauges a node’s role in acting as a bridge along the shortest paths between other nodes. It is defined as:

​

where ​ represents the total number of shortest paths from node to node and is the number of those paths passing through (Freeman, 1977). This centrality measure is indicative of a node's influence over information or passenger flow within the network.

For the London Underground, betweenness centrality points to stations that are critical for maintaining network connectivity and efficiency. Stations with high betweenness centrality are often unavoidable on many journeys across the network, indicating their strategic importance for overall network cohesion (Crucitti et al., 2006). Disruptions at these stations can have far-reaching impacts on the network, underscoring their operational significance.

|  |  |  |
| --- | --- | --- |
| node | station\_name | betweenness\_t |
| 940GZZLUBST | Baker Street | 36297.775794 |
| 940GZZLUBLG | Bethnal Green | 33670.108333 |
| 940GZZLUFYR | Finchley Road | 32064.800397 |
| 940GZZLUBNK | Bank | 30443.441667 |
| 940GZZLUGPK | Green Park | 30442.438095 |
| 940GZZLUWLO | Waterloo | 30219.900000 |
| 940GZZLULVT | Liverpool Street | 29820.741667 |
| 940GZZLUWSM | Westminster | 27623.541667 |
| 940GZZLUBND | Bond Street | 24635.653175 |
| 910GWHMDSTD | West Hampstead | 22536.658333 |

### Closeness Centrality

Closeness centrality measures a node’s mean distance to all other nodes, providing insight into how quickly a node can reach the rest of the network. It is computed as:

​

where denotes the shortest path distance between and (Bavelas, 1950). This centrality measure is crucial for understanding how effectively a node disseminates or gathers information or resources across a network.

In the London Underground, stations with high closeness centrality can be reached quickly and efficiently from other parts of the network, making them central to minimizing travel times and improving service accessibility (Beauchamp, 1965). These stations are vital for ensuring operational efficiency and enhancing user satisfaction by reducing journey times.

|  |  |  |
| --- | --- | --- |
| node | station\_name | closeness\_t |
| 940GZZLUGPK | Green Park | 0.094897 |
| 940GZZLUBND | Bond Street | 0.093737 |
| 940GZZLUWSM | Westminster | 0.093197 |
| 940GZZLUBST | Baker Street | 0.092900 |
| 940GZZLUWLO | Waterloo | 0.092389 |
| 940GZZLUBNK | Bank | 0.092000 |
| 940GZZLUOXC | Oxford Circus | 0.091614 |
| 940GZZLUTNG | Turnham Green | 0.011442 |
| 940GZZDLCGT | Canning Town (DLR) | 0.011442 |
| 940GZZLULVT | Liverpool Street | 0.011442 |

## I.2 – Impact Measurements

Global impact measures are used in network analysis to evaluate the overall health and resilience of a network upon changes, such as the removal of nodes or links. These measures are not specific to the London Underground; they can be applied to any network, including transportation systems, social networks, communication networks, among others.

**Robustness Index**

The network robustness index measures the network's ability to maintain connectivity under node or link removal, typically evaluated by the size of the LLC as a proportion of the original network's size. A higher robustness index indicates a more resilient network.

where is the size of the largest connected component after node removal, and is the original number of nodes in the network. This index helps to understand how a network might be vulnerable to disruptions or attacks by understanding how critical certain nodes are for overall connectivity.

**Global Efficiency**

Global efficiency is a measure of the network's overall efficiency in transmitting information or resources. It is calculated as the average inverse shortest path length between all pairs of nodes. It reflects the network's performance as a function of its connectivity, where higher efficiency indicates that the network can handle interactions more effectively across its entirety.

where is the total number of nodes in the network, and is the shortest path distance between nodes and . It is particularly useful in evaluating operational performance post-disruptions, which well-suits it as a metric for understanding network resilience.

## I.3 – Node Removal (Full results in appendix)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Degree | | | | Betweenness | | | | Closeness | | | |
|  | Non-seq. | | Seq. | | Non-seq. | | Seq. | | Non-seq. | | Seq. | |
| Rem. | GE | RI | GE | RI | GE | RI | GE | RI | GE | RI | GE | RI |
| 1 | 0.081 | 0.961 | 0.081 | 0.961 | 0.064 | 1.000 | 0.073 | 1.000 | 0.079 | 1.000 | 0.079 | 1.000 |
| 2 | 0.073 | 0.956 | 0.073 | 0.956 | 0.063 | 0.995 | 0.064 | 1.000 | 0.078 | 1.000 | 0.072 | 1.000 |
| 3 | 0.072 | 0.956 | 0.072 | 0.956 | 0.061 | 0.995 | 0.062 | 0.979 | 0.069 | 1.000 | 0.068 | 1.000 |
| 4 | 0.071 | 0.954 | 0.070 | 0.935 | 0.060 | 0.995 | 0.072 | 0.525 | 0.067 | 1.000 | 0.066 | 1.000 |
| 5 | 0.069 | 0.933 | 0.066 | 0.935 | 0.059 | 0.995 | 0.120 | 0.455 | 0.064 | 1.000 | 0.065 | 1.000 |
| 6 | 0.065 | 0.933 | 0.065 | 0.933 | 0.058 | 0.995 | 0.113 | 0.454 | 0.064 | 0.998 | 0.060 | 0.970 |
| 7 | 0.063 | 0.933 | 0.064 | 0.926 | 0.057 | 0.995 | 0.109 | 0.425 | 0.060 | 0.998 | 0.103 | 0.485 |
| 8 | 0.063 | 0.926 | 0.061 | 0.898 | 0.053 | 0.995 | 0.102 | 0.423 | 0.060 | 1.000 | 0.099 | 0.484 |
| 9 | 0.062 | 0.914 | 0.062 | 0.844 | 0.070 | 0.713 | 0.093 | 0.282 | 0.059 | 0.995 | 0.084 | 0.459 |
| 10 | 0.058 | 0.886 | 0.064 | 0.790 | 0.073 | 1.000 | 0.145 | 0.264 | 0.079 | 1.000 | 0.095 | 0.432 |

### Analysing Centrality Measures

Degree centrality maintains a larger component size for more removals compared to closeness and betweenness, suggesting that removing nodes with high degrees tends to preserve network connectivity better. This could indicate that degree centrality is a stronger indicator of a node's importance to maintaining the structural integrity of the network.

Nodes removed based on betweenness centrality led to a quicker decline in the robustness index, indicating that such nodes are crucial for keeping the network united. Closeness centrality also shows a significant drop but not as drastic as betweenness, while degree centrality maintains a higher robustness for longer.

Betweenness centrality shows a higher global efficiency even as more nodes are removed, reflecting that paths remaining in the network are still relatively efficient. This suggests that betweenness centrality identifies nodes that, when removed, severely impact the shortest paths across the network.

### Analysing Resilience Strategies (for Node Removal)

The largest component size and robustness index are maintained longer under the non-sequential strategy, suggesting that random removals potentially distribute the impact more uniformly across the network, preserving its structure better compared to targeted sequential removals.

Similar trends are observed with closeness centrality; however, the differences between the strategies are less pronounced. This might imply that closeness centrality identifies crucial nodes whose removal consistently impacts the network, regardless of the order of removal.

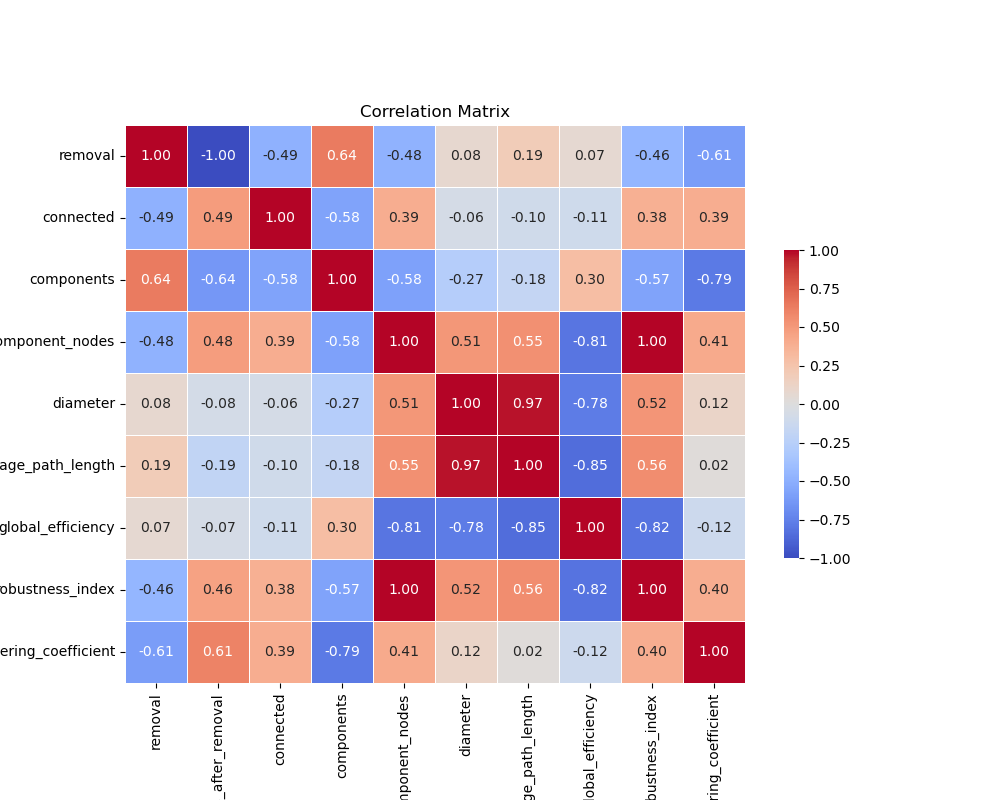
The impact of node removal is more severe with betweenness centrality, as indicated by a sharper decline in both the largest component size and robustness index. This centrality appears to be critical for network cohesion, and the sequential strategy seems to cause slightly quicker disintegration compared to the non-sequential strategy.

### Impact Measures for Assessing Damage

There's a strong negative correlation (-0.82) between global efficiency and the robustness index. This indicates that as the network becomes less robust (i.e., fewer nodes remain interconnected), the efficiency of the network also decreases, reflecting longer paths and reduced communication or transportation efficiency.

The diameter has a moderate positive correlation (0.52) with the robustness index, suggesting that as the network becomes less robust, the diameter tends to increase. This makes sense because as more nodes are removed, the remaining network paths become longer, leading to an increase in the maximum path length between any two nodes.

There is a moderate positive correlation (0.40) between the clustering coefficient and robustness index. This relationship highlights that a more robust network tends to have more tightly-knit clusters of nodes, which could help in maintaining connectivity and redundancy even as nodes are removed.



### Summary of Analysis

Degree centrality appears to better reflect the importance of a station for maintaining network integrity initially. However, betweenness centrality also shows significant impacts, especially on the network's efficiency and connectivity, suggesting it's crucial for strategic network analysis focused on critical links.

The non-sequential strategy generally shows a slower degradation of network properties (robustness and efficiency), suggesting it might be more effective at modelling resilience by simulating random failures, which could better represent real-world scenarios like random station closures.

Both the robustness index and global efficiency provide useful insights into the network's performance. The robustness index is more indicative of the network's overall cohesion and connectivity, while global efficiency provides a clear picture of operational efficiency. Consequently, the choice between them may depend on whether the focus is on connectivity or performance. Due to the imperative of the London Underground to provide connectivity to as large a population as possible, the robustness index is seen here to provide the greatest insights into the networks’s resilience.

# Part 2 – III: Models and Calibration

## III.1 – Spatial Interaction Models

Spatial interaction models are used in geographical, urban planning, and regional science research. These models, built on the theoretical framework of Newton's law of gravitation, quantify the flow of goods, services, people, or information between spatial units based on the attributes of these units and the distances separating them (Wilson, 1971).

### Unconstrained Gravity Model

The Unconstrained Gravity Model is foundational for spatial interaction theory and is the simplest form of the gravity model. It estimates the flow, ,​ between an origin, , and a destination, :

* : The estimated flow from origin to destination .
* : A constant to ensure that the predicted flows fit the total observed flows.
* ​: The magnitude of the flow-generating factor at the origin .
* ​: The attractiveness or flow-receiving factor at the destination .
* : A deterrence function, which is typically an exponential function of the cost , such as distance or time. The parameter modulates the influence of cost.

This equation represents the prediction of flows, such as migration or trade, between two points based on their 'masses' and the 'distance' between them. The work of Isard (1956) in "Location and Space-Economy" first formalised the analogy between physical gravity and spatial interaction, setting the stage for future models. The constant acts as a scaling factor to adjust the predicted flows to the observed total flows; it scales the model output to real-world data.

The unconstrained model posits that the flow between two locations is proportional to some measure of the locations' economic mass, such as population or GDP, and decreases with increasing separation, such as distance or travel cost. This model's primary use is in cases where we are looking at aggregate flow patterns without specific constraints at either the origins or destinations. For example, Stewart (1948) worked on the concept of social physics applied this model to understand urban population patterns, while Zipf (1946) applied a similar concept to rank-size distribution of cities.

### Singly-Constrained Models

#### Origin-Constrained Model (Production-Constrained):

The production-constrained model adjusts the unconstrained model by including a constraint that the sum of flows from an origin, , equals some observed total, :

* ​: A balancing factor to ensure that the constraint is satisfied.

Wilson's (1967) "A statistical theory of spatial distribution models" introduced the concept of entropy-maximizing models, where the origin-constrained gravity model was derived. Here ​ ensures that the sum of all flows from an origin matches the observed outflows.

The origin-constrained model reflects scenarios where the total outflow from any origin is fixed. Wilson's (1967) foundational work on entropy-maximizing spatial interaction models introduced constraints into the gravity model, which led to the development of this origin-constrained variant. This model is particularly useful for understanding commuting patterns, where the number of people leaving a residential area for work is known.

#### Destination-Constrained Model (Attraction-Constrained):

The attraction-constrained model includes a constraint that the sum of flows to a destination, , equals some observed total, ​.

* ​: A balancing factor to ensure that the constraint for the destination is satisfied.

The balancing factor ​ ensures the predicted inflows to a destination do not exceed its observed capacity. This model variant is central to the study of urban retail patterns as discussed in "Models of the Retail Location Process" by Lakshmanan and Hansen (1965), which analyses the spatial distribution of retail facilities.

The destination-constrained model is used when the total inflow to any destination is fixed. It is commonly applied in retail studies to predict consumer flows to shopping centres, assuming the capacity of these centres is known.

### Doubly-Constrained Model

The most sophisticated is the Doubly-Constrained Model. This model is a combination of the origin and destination constrained models, including constraints on both the sum of flows from each origin and the sum of flows to each destination.

* Both ​ and ​ are balancing factors computed such that both sets of constraints are satisfied.

This model is thoroughly discussed in Fotheringham and O'Kelly's (1989) "Spatial Interaction Models: Formulations and Applications." It requires known constraints on both origins and destinations and is widely used in transportation and urban planning to estimate traffic flows and service usage.

The doubly constrained model is the most comprehensive, incorporating both origin and destination constraints. This model is particularly well-suited for transportation planning, where both the number of trips from each origin and the capacity of each destination (such as the number of parking spaces in urban centres) are known. Fotheringham and O'Kelly (1989) in their work "Spatial Interaction Models: Formulations and Applications" have extensively discussed the calibration and application of this model.

To find the values of ​​ and ​, we need to solve an iterative process where each is adjusted to satisfy the origin constraints, and then each is adjusted to satisfy the destination constraints; this is repeated until the system converges.

The parameter is particularly important, as it modulates the impact of distance (or cost) on the flows in the model. It is effectively a measure of 'friction' (Hanson, 1980). The iterative computation of ​ and ​ for the doubly constrained model is an adaptation of the method proposed by Evans (1971). A higher implies that distance or cost has a strong deterrent effect, which might represent situations where transportation costs are high or the infrastructure is poor.

## III.2 – Calibrate a model for cost function of pop., jobs, and flows

Given the availability of both "population" data (at origins) and "jobs" data (at destinations), alongside the flows between these, a doubly constrained model would be the most accurate and realistic of the gravitational attraction spatial interaction models. This accounts for the capacity of origins to produce flows and the capacity of destinations to attract flows, adjusting for both. The equation, as seen before, is:

Where:

And:

* ​ is the flow from station to station ,
* ​ is the balancing factor for the origin ,
* ​ is the balancing factor for the destination ,
* ​ is the population at the origin station ,
* ​ is the number of jobs at the destination station ,
* is the deterrence function that negatively relates the distance to the flow, with being the parameter to be calibrated,
* ​ is the cost of travel, which in this case, is represented by the distance.

The calculation of  relies on knowing  and the calculation of  relies on knowing . Senior (1979) describes a useful algorithm for iteratively calculating the values for  and  by equalling each to ‘1’ initially and then continuing to calculate each in turn until the difference between each value is small enough that it becomes (effectively) irrelevant and the likelihood of the observed flows given the model has been maximised.

The negative exponential model is identified as having the best modelling capacity with an R-squared value of 0.4766 and RMSE value of 95.196. This compares to the inverse power model which achieves an R-squared value of 0.3099 and RMSE value of 109.597; therefore the negative exponential model can better model the data with a higher R-square and lower RMSE demonstrating it explains more of the variance in the data.

The calibrated value of beta is . This value effectively reflects the sensitivity of the flow to the cost of travel, which in this dataset is represented by the distance between stations.

(See appendix for calculations.)

# Part 2 – IV: Scenarios

## IV.1 – Scenario A: Canary Wharf has a 50% decrease in jobs

To compute the new flows, we use the calibrated , adjust the jobs for Canary Wharf, and ensure that the total number of commuters remains conserved by recalculating the ​ and ​ balancing factors for the doubly constrained gravity model.​

In this scenario, if Canary Wharf sees a 50% reduction in jobs post-Brexit, it would directly impact the destination attractiveness of Canary Wharf in the model. The reduction in ​ for Canary Wharf implies that fewer people would choose or need to travel there. The adjustment in the model would need to redistribute these flows to other destinations where jobs are still available, maintaining the overall commuting volume consistent.

Ensuring the number of commuters is conserved the the destination jobs (Dj) values were adjusted by recalculating flows with the new ​ values for Canary Wharf (multiplying destination jobs in Canary Wharf by 0.5) while keeping the total outflow from origins, , unchanged. This adjustment redistributes the total commuting volume to other destinations proportionately, as the total outflows from each origin remain fixed.

These values can then be inputted to the aforementioned equation for the doubly-constrained model to calculate the new flows distribution.

## IV.2 – Scenario B: Significant increase in cost of transport

Scenario B - Increase in Transport Costs: This scenario involves two variations with different increases in the cost function parameter . Increasing enhances the decay effect in the exponential cost function, thereby significantly reducing the interaction (or flow) as the distance increases. This effectively simulates an increase in perceived transport costs.

Beta Increase 1 (moderate increase) and Beta Increase 2 (high increase) likely show a progressive reduction in longer-distance commuting, with more significant impacts visible under the higher value. This could lead to a shift towards more localized commuting patterns, with more significant reductions in long-distance flows.

## IV.3 – Discussion of Scenarios; Greatest Impact on Flow Distribution

The total commuter flows under normal conditions are approximately 1,541,917.

Under Scenario A the total flows decrease to 359,312, indicating a substantial reduction due to the decrease in jobs at Canary Wharf. This scenario significantly impacts flows to Canary Wharf, effectively halving the number of commuters relative to the baseline, suggesting a strong localised impact.

Under Scenario B1 (Moderate Increase in Transport Costs) the total flows are 364,555, showing a significant reduction compared to the baseline but not as severe as in Scenario A. This suggests that while transport costs influence commuting patterns, the effect is less drastic than a major employment shift.

Under Scenario B2 (High Increase in Transport Costs) the total flows slightly increase to 364,788 compared to B1, indicating that the further increase in transport costs has a minimal additional impact on reducing commuting volumes beyond the moderate increase scenario.

**Analysis and Discussion:**

The large drop in total commuting flows under Scenario A confirms the significant impact of job reductions at a major employment hub like Canary Wharf. This dramatic reduction showcases the dependency of the network on Canary Wharf as a key destination for commuters.

The B scenario with increased transport costs shows that commuters may be somewhat resilient to changes in transport costs up to a certain threshold. The minimal difference between the moderate and high increases in transport costs suggests a diminishing return effect on further discouraging high-cost commuting. It also demonstrates that the threshold at which commuting becomes to expensive for most people is below the 50% increase scenario. In other words, at a threshold value below a 50% increase in commuting costs, significant numbers of commuters would cease travelling.

Scenario A has a more profound impact on the network, drastically reducing the total commuter flows due to the loss of jobs at Canary Wharf. In contrast, Scenario B, even with increases in transport costs, does not curtail overall commuting volumes as significantly.

This analysis could guide urban planners and policy-makers to focus on economic strategies that could mitigate the potential negative impacts of job losses at major employment centres, such as Canary Wharf. It also suggests that moderate adjustments to transport cost policies may significantly discourage commuting, requiring a more nuanced approach to managing transport economics and urban mobility – often cost increases are seen as the ideal mean through which to raise capital, however if costs are increased to a level that is too high, commuters will cease travelling.

# Appendix

Github link to code and other outputs:

# References:

Barthélemy, M. (2011). Spatial networks. *Physics Reports*, [online] 499(1-3), pp.1–101. doi:https://doi.org/10.1016/j.physrep.2010.11.002.

Beauchamp, M.A. (1965). An improved index of centrality. *Behavioral Science*, 10(2), pp.161–163. doi:https://doi.org/10.1002/bs.3830100205.

Burnett, P. (1980). Spatial Constraints-Oriented Modeling as an Alternative Approach to Movement, Microeconomic Theory, and Urban Policy. *Urban Geography*, 1(1), pp.53–67. doi:https://doi.org/10.2747/0272-3638.1.1.53.

Crucitti, P., Latora, V. and Porta, S. (2006). Centrality measures in spatial networks of urban streets. *Physical Review E*, 73(3). doi:https://doi.org/10.1103/physreve.73.036125.

Evans, A.W. (1971). The calibration of trip distribution models with exponential or similar cost functions. *Transportation Research*, 5(1), pp.15–38. doi:https://doi.org/10.1016/0041-1647(71)90004-9.

Fotheringham, S. and O’Kelly, M.E. (1989). *Spatial Interaction Models: Formulations and Applications*. [online] *University of St Andrews Research Portal*. Dordrecht/Boston/London: Kluwer Academic Publishers. Available at: https://research-portal.st-andrews.ac.uk/en/publications/spatial-interaction-models-formulations-and-applications [Accessed 23 Apr. 2024].

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

Freeman, L.C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, 1(3), pp.215–239. doi:https://doi.org/10.1016/0378-8733(78)90021-7.

Isard, W. (1956). *Location and Space-economy: A General Theory Relating to Industrial Location, Market Areas, Land Use, Trade, and Urban Structure*. [online] *Google Books*. =Technology Press of Massachusetts nstitute of Technology and Wiley, New York. Available at: https://books.google.co.uk/books/about/Location\_and\_Space\_economy.html?id=vccmAAAAMAAJ&redir\_esc=y [Accessed 23 Apr. 2024].

Lakshmanan, J.R. and Hansen, W.G. (1965). A RETAIL MARKET POTENTIAL MODEL. *Journal of the American Institute of Planners*, 31(2), pp.134–143. doi:https://doi.org/10.1080/01944366508978155.

Senior, M.L. (1979). From gravity modelling to entropy maximizing. *Progress in Human Geography*, 3(2), pp.175–210. doi:https://doi.org/10.1177/030913257900300218.

Stewart, J.Q. (1948). Demographic Gravitation: Evidence and Applications. *Sociometry*, 11(1/2), p.31. doi:https://doi.org/10.2307/2785468.

Wilson, A. (2010). Entropy in Urban and Regional Modelling: Retrospect and Prospect. 城市和区域建模中的熵：回顾与展望. *Geographical Analysis*, 42(4), pp.364–394. doi:https://doi.org/10.1111/j.1538-4632.2010.00799.x.

Wilson, A.G. (1967). A statistical theory of spatial distribution models. *Transportation Research*, [online] 1(3), pp.253–269. doi:https://doi.org/10.1016/0041-1647(67)90035-4.

Zipf, G.K. (1946). The P 1 P 2 D Hypothesis: On the Intercity Movement of Persons. *American Sociological Review*, 11(6), p.677. doi:https://doi.org/10.2307/2087063.