Simulating the London Underground with Network Science and Spatial Interaction Techniques

Gavin Rolls

April 11, 2024

This Paper was submitted as a part of assessment for CASA 0002 (Urban Simulation) at the Centre for Spatial Analysis, University College London. A link to the project GitHub is attached here.

1 London's Underground Resilience

In this section, we'll investigate the resilience of London's transport system by identifying the stations whose removal poses the highest risk to its operation. We'll consider both the topology of the network and a weighted network with travel volumes.

1.1 Topological Networks

1.1.1 Centrality Measures

To identify the most topologically important stations in London's transport network, we'll evaluate three different measures of centrality for each station, or node. These measures attempt to identify the prominence or importance of a given node to the network as a whole (Bloch et al. 2023). There exist a number of metrics to capture centrality, but (Stamos 2023) write that for transport networks, wherein nodes represent stations on the network, *degree centrality*, *betweenness centrality*, and *eigenvector centrality* are particularly useful metrics. We will detail these below[1]:

1. Degree Centrality:
Degree centrality measures the centrality of a node with respect to its quantity of neighbours (Bloch et al. 2023). To that end, a node's degree and thus its centrality is measured as the count of its edges. We can represent this mathematically as:

$$k_i = \sum_{j=1}^{N} A_{ij}$$

Where k_i is the *degree centrality* of node i, for a network of N nodes and A_{ij} is the element of the adjacency matrix representing this network. An adjacency matrix A is an N-by-N matrix, where the element in position ij represents whether or not there is an edge in the network connecting nodes i and j (Stamos 2023).

In the context of a transport network, higher degree will generally indicate a station

[1] All descriptions of networks and centrality measures are in the context of an undirected network. For instance, in and out-degree are not the same, but we will conceptualise them as such in this section

with multiple lines which may act as a hub for transfers (Stamos 2023). Although *degree centrality* is less useful when traffic flow is considered, its is a simple, but powerful metric when evaluating topology alone (Stamos 2023).

The top ten stations in the underground network, ranked by *degree centrality*[2] are:

Station	**Degree**
King's Cross St. Pancras	12
Liverpool Street	10
Bank and Monument	10
Baker Street	10
Stratford	9
Paddington	9
West Ham	8
Moorgate	8
Embankment	8
Euston	7

Table 1: Top Ten Stations by *Degree Centrality*

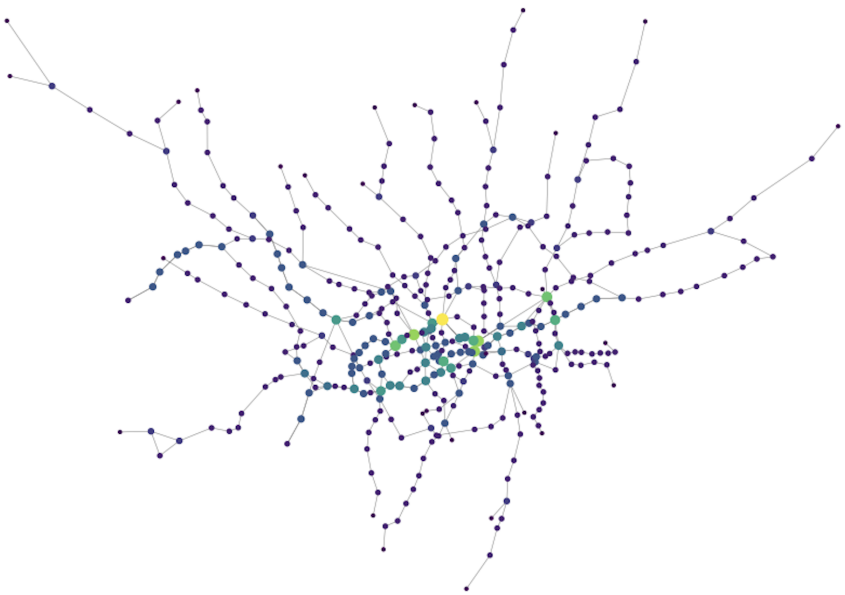

Figure 1: *Degree Centrality* of Underground Stations

2. Betweenness Centrality

Betweenness centrality is a measure of the frequency with which a given node lies on the geodesic (shortest path) between all other pairs of nodes in the network (Bloch et al. 2023). This measure attempts to find the nodes that operate as an intermediary for flows between other parts of the network. Multiple geodesics may exist for a given pair of nodes so this centrality also divides by the number of geodesics between a particular pair of nodes i, j to consider what proportion of the shortest paths pass through each

[2]Because we are representing the Underground as a MultiGraph where different edges form different lines, degree will be the number of adjacent edges but not always adjacent nodes

node. This centrality is represented as:[3]

$$B_i = \sum_{s \neq i \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

Where B_i is the betweeness centrality of node i, $\sigma_{st}(i)$ is the number of geodesics between s and t through i and σ_{st} is the total number of geodesics between s and t (Stamos 2023). This can also be normalised to a measure between the minimum betweeness centrality 0 and maximum possible *betweenness centrality* 1 through the following equation, where N represents the total number of nodes in the network.

$$B_i = \frac{1}{N^2} \sum_{s \neq i \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

Betweenness centrality is a valuable metric in a transport context because it can help reveal the main arteries of a given network, even without flow information (Stamos 2023). It's thus useful in identifying bottlenecks and locations which pose the greatest risk to a system's overall resilience.

The top ten stations in the underground network by normalised *betweenness centrality* are:

Station	**Normalised Betweenness Centrality**
Bank and Monument	0.2206
King's Cross St. Pancras	0.2103
Stratford	0.1823
Baker Street	0.1654
Oxford Circus	0.1577
Euston	0.1554
Earl's Court	0.1435
Shadwell	0.1394
Waterloo	0.1304
South Kensington	0.1291

Table 2: Top Ten Stations by *Betweenness Centrality*

[3]This format is adapted slightly from Stamos (2023) in order to maintain consistency with the formulas for other centrality measures

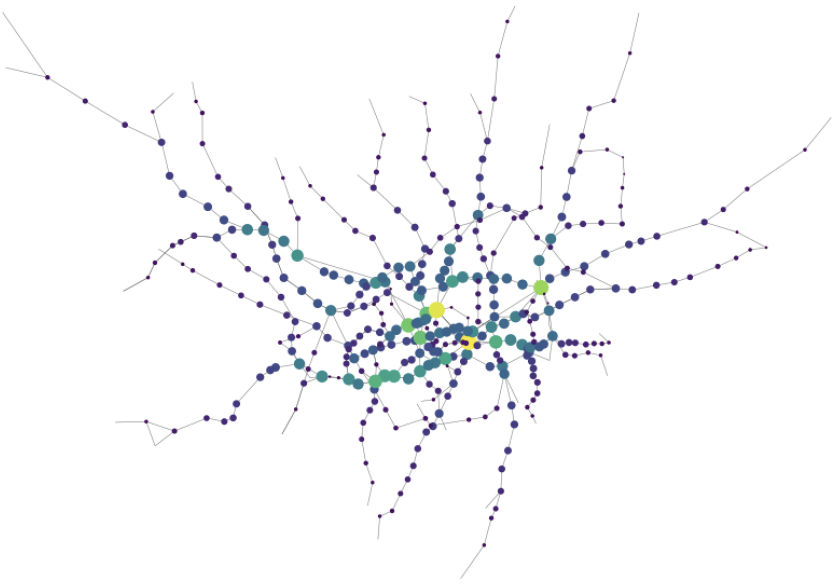

Figure 2: *Betweenness Centrality* of Underground Stations

3. Eigenvector Centrality

Eigenvector centrality considers not only the number of neighbours a given node has, but how well-connected those neighbours are. The centrality of a node is thus proportional to the sum of the centrality of all of its neighbours (Bloch et al. 2023). *Eigenvector centrality* for a node i, written x_i, can be expressed as:

$$x_i = \lambda_1^{-1} \sum_j A_{ij} x_j$$

Where λ_1^{-1} is the first eigenvalue of the adjacency matrix A (Stamos 2023). An eigenvector v is a non-zero vector that, when multiplied with by the adjacency matrix A produces a vector which is a scaled version of v. An eigenvalue, then, is the scalar λ which satisfies the equation $Av = \lambda v$.

Eigenvector centrality is useful for identifying nodes that function as hubs and most influence the entire network's structure and connectivity. It has been used extensively in literature on transport network analysis as a result (Stamos 2023).

The top ten stations in the underground network, ranked by *eigenvector centrality* are[4]:

[4]Because the adjacency matrix of the MultiGraph isn't always diagonalizable, we are using a simple graph representation to compute eigenvector centrality

Station	Eigenvector Centrality
Bank and Monument	0.3834
Liverpool Street	0.3288
Stratford	0.2692
Waterloo	0.2497
Moorgate	0.2151
Green Park	0.1976
Oxford Circus	0.1840
Tower Hill	0.1717
Westminster	0.1686
Shadwell	0.1591

Table 3: Top Ten Stations by *Eigenvector Centrality*

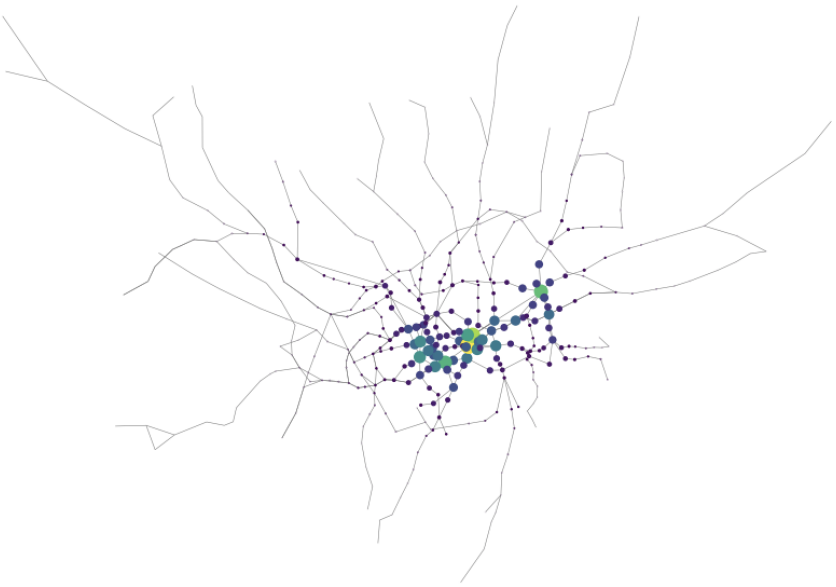

Figure 3: *Eigenvector Centrality* of Underground Stations

1.1.2 Impact Measures

We'll evaluate the impact of removing nodes from the Underground through two metrics:

1. Characteristic Path Length
Characteristic Path Length $L(p)$ refers to the average length of a shortest path in the network (Watts & Strogatz 1998). On the underground, this could correspond either to the average distance travelled from origins to destinations, or the number of stations traversed. Both metrics are valuable in that they both represent different elements of the time cost of travelling on the underground (the time it takes to traverse physical distance on the network and the dwell time in each station).

This measure, weighted or unweighted, is a valuable measure of network strength not just for transport networks, but for any network. In fact, Latora & Marchiori (2001) define the efficiency of a network through these terms, writing that the efficiency of a

graph G with N vertices, written $E(G)$ can be defined as the inverse of the average shortest path:

$$E(G) = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{d_i j}$$

2. Average Number of Transfers Transferring between trains is also a costly endeavour and thus evaluating the average number of transfers required for travel between stations is important for network efficiency. With that said, *transfer count* cannot be generalised as an impact metric as it is specific to the characteristics of transport networks.

It's also important, then, to ensure that computed shortest paths minimise transfers to reflect real-world trip choices. Given that this minimisation is almost universally advantageous for route efficiency (Zhao & Ubaka 2004), we will equate a transfer to 10km of added distance, which we later find to be longer than the average trip distance itself (Table 13). This distance penalty was incorporated into a modified version of Dijkstra's shortest path algorithm to ensure that selected travel routes appropriately minimised transfers and that computed flows reflect this.

1.1.3 Node Removal

We will now remove the most important stations from the Underground network (as determined by each the above measures) by applying two different strategies:

1. **Non-Sequential removal:** Remove stations from the top ten rankings at random.

2. **Sequential removal:** Remove stations in order of centrality. After each station is removed, centrality metrics are recalculated.

After each removal, the impact is assessed by computing the average number of transfers and the *characteristic path length* $L(p)$, as well as marking the percent of the network which remains in the largest connected component to contextualise the other two measures. If the network is split into multiple components, the largest connected component is used. The results of these removals are reported below[5]

Removals	Rem. Station	Avg. Path Length	Avg. Transfer	% Connected
0	N/a	13.80	1.03	100%
1	West Ham	14.059	1.10	99.75%
2	Baker Street	15.27	1.23	99.50%
3	Embankment	15.049	1.32	99.25%
4	Paddington	15.05	1.34	98.50%
5	King's X St. Pancras	15.54	1.55	97.76%
6	Earl's Court	16.10	1.77	97.51%
7	Waterloo	16.30	1.77	97.25%
8	Moorgate	16.26	1.74	96.01%
9	Willesden Junction	18.05	2.07	92.52%
10	Stratford	18.16	2.24	86.53%

Table 4: Impact of Non-sequential Node Removal (*Degree Centrality*)

[5]Because stations in the top ten might be entirely disconnected from the graph, the top fifteen stations by centrality are considered for removal in non-sequential approaches

Removals	Rem. Station	Avg. Path Length	Avg. Transfer	% Connected
0	N/a	13.80	1.03	100%
1	King's X St. Pancras	14.73	1.23	99.75%
2	Bank and Monument	15.33	1.53	99.50%
3	Baker Street	16.08	1.60	98.75%
4	Liverpool Street	15.88	1.53	97.26%
5	Paddington	15.83	1.56	96.51%
6	Embankment	15.70	1.60	95.26%
7	Stratford	15.21	1.59	89.77%
8	Willesden Junction	16.61	1.87	86.28%
9	West Ham	17.61	1.93	85.54
10	Earl's Court	18.79	2.31	85.29

Table 5: Impact of Sequential Node Removal (*Degree Centrality*)

Removals	Rem. Station	Avg. Path Length	Avg. Transfer	% Connected
0	N/a	13.80	1.03	100%
1	King's X St. Pancras	14.73	1.23	99.75%
2	Earl's Court	14.98	1.37	99.50%
3	Shadwell	15.023	1.39	99.00%
4	South Kensington	14.75	1.48	98.75%
5	Hammersmith	14.72	1.50	98.00%
6	Waterloo	14.79	1.59	97.76%
7	Wembley Park	15.82	1.74	96.51%
8	Oxford Circus	16.11	1.92	96.26%
9	Gloucester Road	16.12	1.91	96.01%
10	Highbury & Islington	16.13	2.06	95.76%

Table 6: Impact of Non-sequential Node Removal (*Betweenness Centrality*)

Removals	Rem. Station	Avg. Path Length	Avg. Transfer	% Connected
0	N/a	13.80	1.03	100%
1	Bank and Monument	13.99	1.19	99.75%
2	King's X St. Pancras	15.33	1.53	99.50%
3	Canada Water	16.76	1.56	99.25%
4	West Hampstead	12.06	1.20	86.86%
5	Earl's Court	12.56	1.40	56.61%
6	Oxford Circus	12.97	1.46	56.36%
7	Shepherd's Bush	13.82	1.52	48.88%
8	Baker Street	17.15	1.81	47.88%
9	Acton Town	11.53	1.21	31.42%
10	Euston	11.18	1.56	25.68%

Table 7: Impact of Sequential Node Removal (*Betweenness Centrality*)

Removals	Rem. Station	Avg. Path Length	Avg. Transfer	% Connected
0	N/a	13.80	1.03	100%
1	Westminster	13.80	1.09	99.75%
2	Moorgate	13.96	1.15	99.50%
3	Tower Hill	13.94	1.17	99.25%
4	Liverpool Street	14.28	1.27	98.75%
5	Stratford	13.70	1.20	93.26%
6	Aldgate East	13.70	1.20	93.02%
7	Piccadilly Circus	13.87	1.30	92.77%
8	Bank and Monument	13.91	1.32	92.52%
9	West Ham	14.11	1.35	91.77%
10	Embankment	13.93	1.40	90.52%

Table 8: Impact of Non-sequential Node Removal (*Eigenvector Centrality*)

Removals	Rem. Station	Avg. Path Length	Avg. Transfer	% Connected
0	N/a	13.80	1.03	100%
1	Bank and Monument	13.99	1.16	99.75%
2	Oxford Circus	14.12	1.19	99.50%
3	Stratford	13.71	1.13	94.01%
4	Earl's Court	13.97	1.29	93.76%
5	Westminster	14.28	1.37	93.51%
6	Baker Street	15.52	1.65	93.01%
7	King's X St. Pancras	16.64	1.86	92.27%
8	Canning Town	16.85	1.91	88.78%
9	Turnham Green	17.07	2.13	88.53%
10	Leicester Square	18.02	2.34	88.28%

Table 9: Impact of Sequential Node Removal (*Eigenvector Centrality*)

1.1.4 Node Removal Analysis

The above results for each scenario and impact are displayed graphically below:

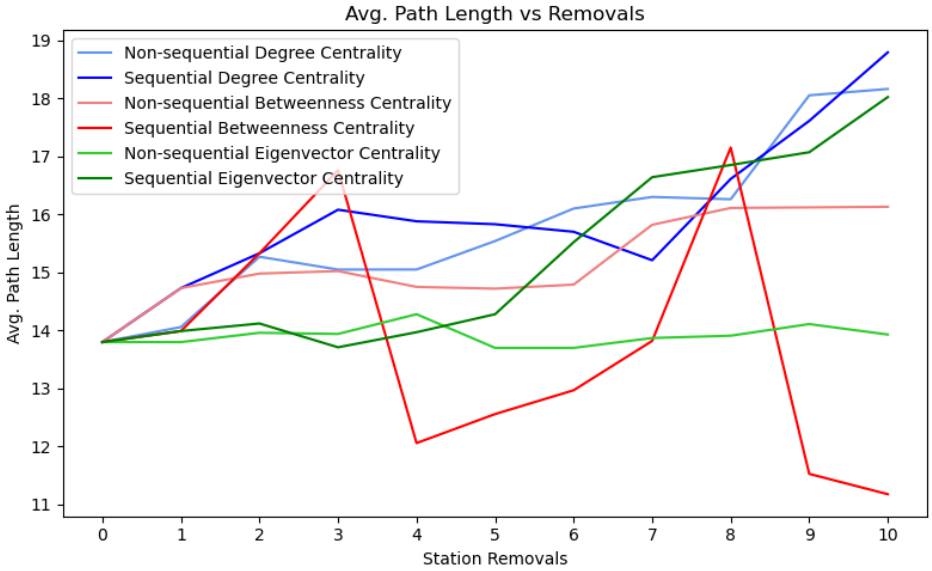

Figure 4: Average Path Length by Number of Stations Removed

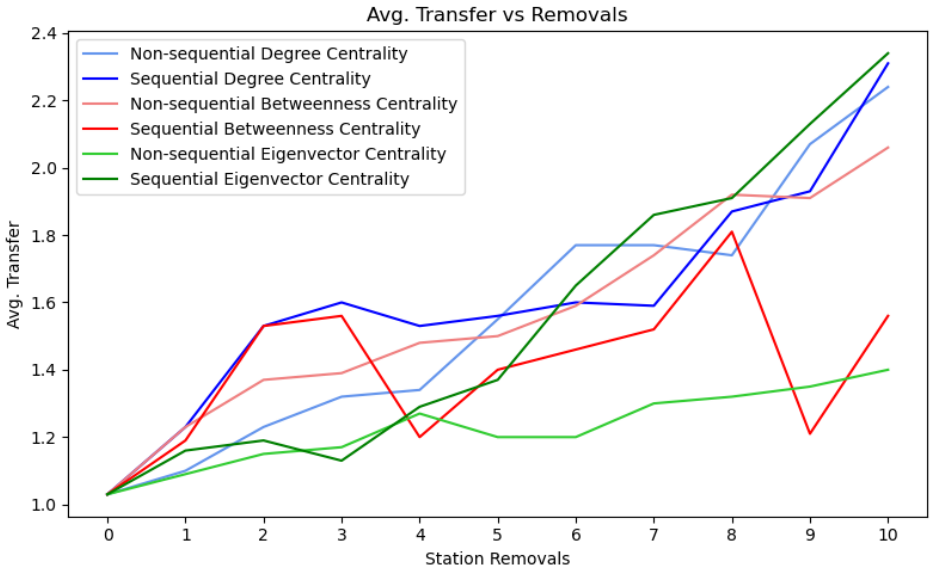

Figure 5: Average *Transfer Count* (per path) by Number of Stations Removed

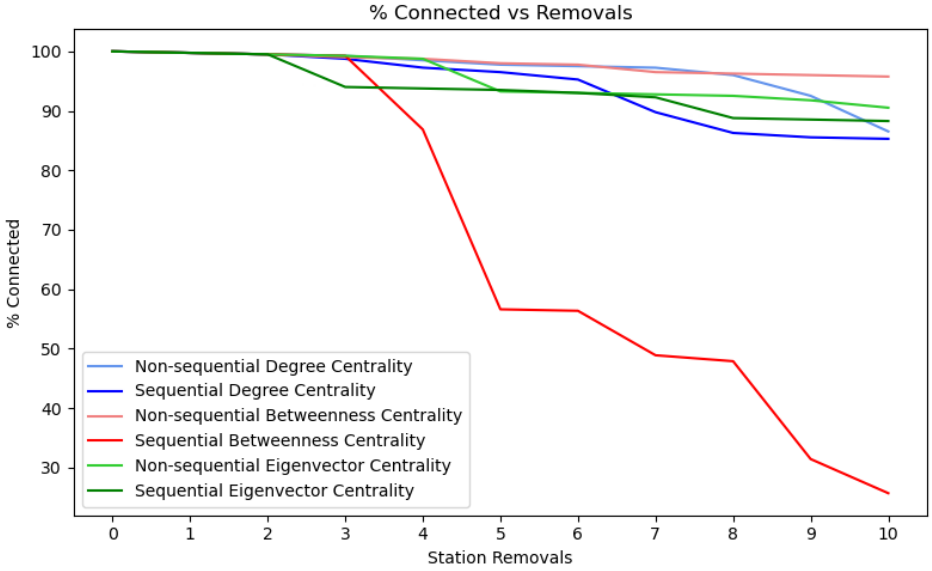

Figure 6: Size of Largest Component (Relative to Full Network) by Number of Stations Removed

As Figure 6 clearly demonstrates, sequential removal of stations by their *betweenness centrality* was by far the most effective technique to disconnect and otherwise impact the network: while all other scenarios featured a connected component of at least 85% the size of the original network, sequential *betweenness centrality* removal brought the remaining network below 30%. Because the network became substantially smaller as stations were removed, this also impacted the analysis of this scenario with respect to both $L(p)$ and *average transfer count*; this is why the scenario's metrics are erratic for both impact measures.

Although the sequential removal technique outperformed its non-sequential counter-

part (Figure 4, 5) on both impact metrics,[6] non-sequential removal is probably the most effective strategy to study resilience. This is because the quasi-random removal of nodes is a better reflection of real-world disruptions, wherein stations are not likely to be disabled in neat order.

The effectiveness of removing the station with the highest *betweenness centrality* also indicates that betweenness is likely the best metric for assessing station centrality. The nature of a transport network, as being driven primarily by flows, makes *betweenness centrality* useful to roughly approximate how many trips depend on the efficient operation of a particular station (Derrible 2012). In contrast, a high degree isn't necessarily an indicator of high importance (Derrible 2012) and the above analysis demonstrates that *eigenvector centrality* is at best inconsistent, especially in the non-sequential case.

Average transfer count is a preferable impact measure to path length because, in addition to better tracking with broad disruption with less sensitivity to scale (Figure 5), high *average transfer count* also reflects a high system fragmentation and an inefficient network. In other words, A trip with a short $L(p)$ but lots of transfers will remain a long trip.

1.2 Flows in a Weighted Network

Although only network topology has been considered thus far, we will now introduce passenger flows to the network. These flows were computed using TfL data on trip counts between station pairs during the morning commute, and paths were computed using the same modified version of Dijkstra's algorithm as described above.

[6] With the exception of sequential *betweenness centrality*, which disconnected the network the most effectively of any method

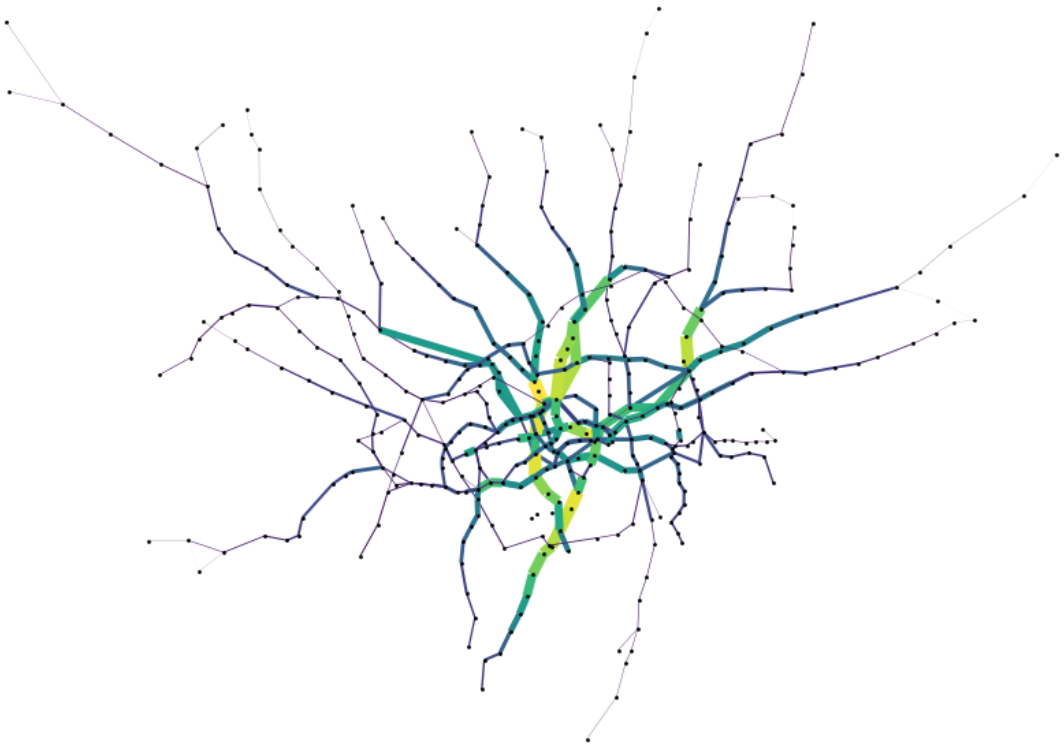

Figure 7: Computed Passenger Flows on the Underground

1.2.1 Centrality in a Weighted Network

To adjust betweeness centrality for use in a flow-sensitive network, we will weight the measure such that betweenness is determined not by the number of geodesics σ passing through a given node i but instead the number of flows. Whereas *betweenness centrality* was previously calculated as:

$$B_i = \sum_{s \neq i \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

We express flow *betweenness centrality* as:

$$B_i = \sum_{s \neq i \neq t} \frac{q_{st}(i)}{q_{st}}$$

Where $q_{st}(i)$ is the number of flows from s to t through i and q_{st} is the total flow between s and t (Stamos 2023).
The top ten nodes in the network by flow *betweenness centrality* are:

Station	Flow Betweenness Centrality
Bank and Monument	0.4388
King's Cross St. Pancras	0.4210
Holborn	0.3650
Chancery Lane	0.3584
St. Paul's	0.3575
Liverpool Street	0.3026
Stratford	0.2999
Baker Street	0.2816
Oxford Circus	0.2615
Russell Square	0.2546

Table 10: Top Ten Stations by Flow *Betweenness Centrality*

Although both measures of *betweenness centrality* identified similar stations, almost all of the stations in the unweighted *betweenness centrality* model sat at the intersection of multiple underground lines (reflecting a high degree of path betweenness). Flow betweeness centrality also identified stations like Chancery Lane and St. Paul's, which are not transfer points but instead capture a high number of flows due to their position at the centre of a high-throughput line (the Central line).

1.2.2 Node Removal Impact in a Weighted Network

To adapt *average transfers* to a network with flows, we'll have to consider the average number of transfers per passenger trip instead of *average transfers* between station pairs. I'll call this *per capita transfers*.

Passenger flow information will also allow for the busiest section of the underground and its flow volume to be identified. Overcrowding is an issue of key concern in transport networks (Li & Hensher 2013) and this metric will thus capture the maximum severity of overcrowding due to station closures.

1.2.3 Recomputing Vulnerable Nodes

We will remove nodes in order of their unweighted *betweenness centrality* and *flow betweenness centrality* using the impact metrics above.

Removals	Rem. Station	Busiest Section	Max. Section Flow	Per Capita Transfers
0	N/a	Camden Town - Euston	55970	0.34
1	Bank and Monument	King's X St. Pancras - Russell Square	66654	0.48
2	King's X St. Pancras	Bermondsey - London Bridge	129362	0.90
3	Canada Water	Gospel Oak - Hampstead Heath	233677	0.96

Table 11: Flow-Sensitive Impact of Sequential Node Removal (Unweighted *Betweenness Centrality*)

Removals	Rem. Station	Busiest Section	Max. Section Flow	Per Capita Transfers
0	N/a	Camden Town - Euston	55970	0.34
1	Bank and Monument	King's X St. Pancras - Russell Square	66654	0.48
2	King's X St. Pancras	Bermondsey - London Bridge	129362	0.90
3	Canada Water	Gospel Oak - Hampstead Heath	233677	0.96

Table 12: Flow-Sensitive Impact of Sequential Node Removal (*Flow Betweenness Centrality*)

At each removal and reevaluation of centrality, flow betweenness and unweighted *betweenness centrality* measures identified the same station as being most central, despite

the fact that below the top ranked station, results differed substantially.

1.2.4 Highest Impact Station Closure

Given the relative consistency of both unweighted and flow-weighted *betweenness centrality*, it becomes easy to identify Bank and Monument as the station whose impact would cause the highest degree of disruption to the underground network. This is an intuitive result as in addition to being the top ranked station by both metrics above, Bank and Monument also ranks as the station with the highest *eigenvector centrality* and is second only to King's Cross when ranked by *degree centrality* (but is ranked first if degree is defined by adjacent nodes instead of adjacent edges). We can quantify this impact in a number of ways but at a glance, the removal of Bank and Monument would:

a) Increase transfers by 41%, from 0.34 to 0.48 for the average journey.

b) Increase crowds on the most crowded underground segment by 19% (from 55970 to 66654).

c) Likely overcrowd the Piccadilly, Jubilee, and Circle/Hammersmith and City/Metropolitan lines as seen in Figure 8.

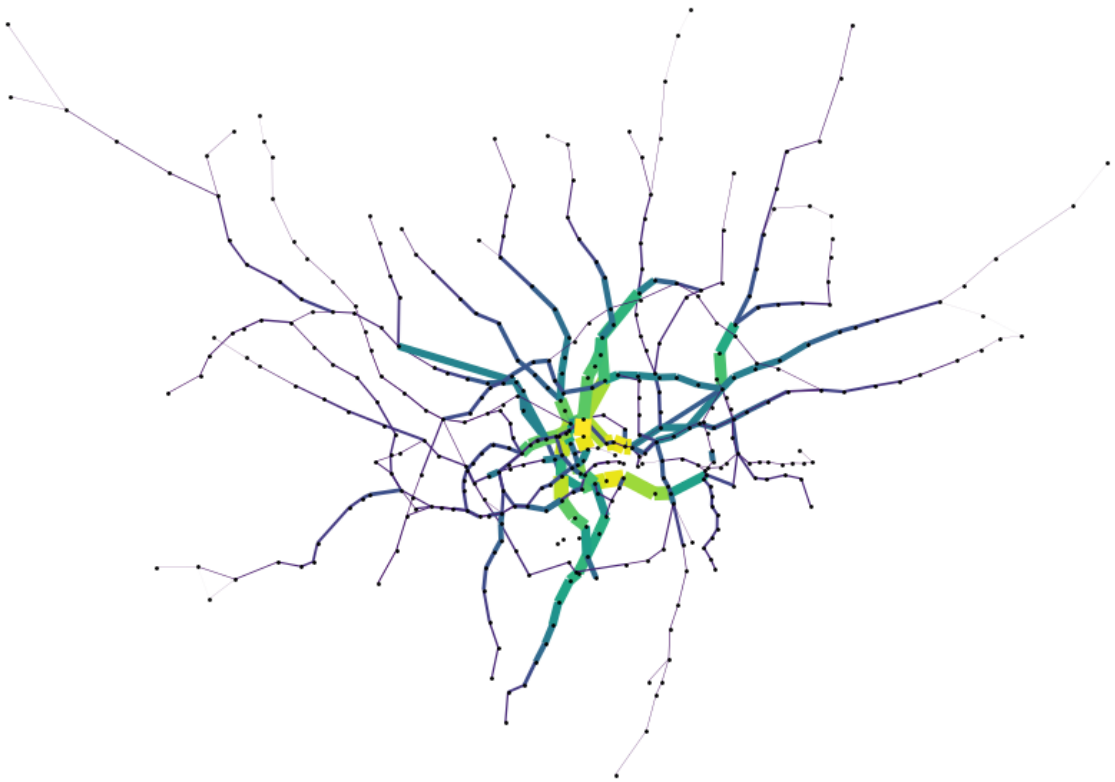

Figure 8: Underground Flows after removal of Bank and Monument

2 Spatial Interaction Models

Spatial Interaction models are models concerned with 'Movements of goods, people, and information between different spatial locations' (Batty 2009). The foundational model in this category is the basic gravity model, named such because it is derived from the same principles dictating gravitational attraction, namely that the size of planets and the distance between them dictate the resulting attractive force (Fotheringham & O'Kelly 1989). The most generic version of the gravity model is:

$$T_{ij} = \frac{P_i P_j}{d_{ij}^2}$$

Where T_{ij} describes the total interaction, or flow between places i and j, P_i and P_j refer to their populations, and d_{ij}^2 is the square of the distance between them (Fotheringham & O'Kelly 1989). This equation can be modified to reach the form broadly accepted as the basic gravity model of spatial interaction:

$$T_{ij} = K\frac{P_i^\alpha P_j^\gamma}{d_{ij}^\beta} = KO_i^\alpha D_j^\gamma d_{ij}^{-\beta}$$

Here, we've added K, a scaling constant that relates flows to the total volume of the system. P_i and P_j have been changed to O_i and D_j to reflect that one place is the origin and the other the destination. Additionally, we've added scaling parameters α, β, and γ so the effect of distance or the importance of origins and destinations can be adjusted depending on the particular context. Fotheringham & O'Kelly (1989) explain that, for example, poorer countries may be more sensitive to traveling large distances and thus a larger β will be more appropriate. Similarly, certain types of interactions may be more sensitive to origins and/or destinations. To that end, these parameters can be calibrated with a set of 'ground truth' observed flows to create a model that best represents a particular context using regression. Flowerdew & Aitkin (1982) propose the use of a Poisson regression model to fit (calibrate) the gravity model to a set of real world observations.

Geographer Alan Wilson proposed a 'family of spatial interaction models' based on the gravity model, which incorporate a number of constraints or other considerations (Wilson 1971). Perhaps most importantly, he defines a set of models based on the knowledge of flows from or to particular locations. We represent the knowledge of total flows from each origin i, as:

$$\sum_j T_{ij} = O_i$$

And for destinations j,

$$\sum_i T_{ij} = D_j$$

Wilson derives three spatial interaction models based on this knowledge, in addition to our original model, which we will now refer to as the unconstrained model. It is unconstrained in that we do not have knowledge about the flows originating from or to any particular zone. These models replace the generic scaling constant K with A_i or B_j, which are 'set[s] of balancing factors' (Wilson 1971) which ensure that flows from each origin (A_i) and/or destination (B_j) sum to their known values. They are[7]:

[7] We have adapted a different notation than Wilson uses for terms in the equation such as the 'mass terms' for consistency with earlier formatting

The production (origin) constrained model, where we know the total flow from each zone:

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

The attraction (destination) constrained model, where we know the total flow to each zone:

$$T_{ij} = O_i^{\alpha} B_j D_j f(d_{ij})$$

The doubly constrained model, where we know both total flow to and from each zone:

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

Note that Wilson also extends our original model by defining the distance decay effect not with a power β but by a general decay function $f(d_{ij})$.

2.1 Models and Calibration

We will use a combination of population and employment counts to model the same morning commute rush used earlier. We will thus set our mass terms for each origin station to the population at that station and the mass term for each destination station as the job count at that station.[8]

One of the scenarios being evaluated by this model concerns a decrease in employee counts at a particular destination. It's imperative that the model is sensitive to changes in total destination counts and a doubly constrained model is thus inappropriate for this analysis. As we can still use total flows from each origin station, we will instead use an origin constrained model.

In addition to selecting the data will be used for mass terms of interest, a distance decay function $f(d_{ij})$ must also be selected. The two main classes of distance decay functions are power decay, which can be expressed as $d_{ij}^{-\beta}$, and negative exponential decay, expressed as $exp(-\beta D_{ij})$ (de Vries et al. 2009). Fotheringham & O'Kelly (1989) write that there exists 'a reasonably widespread consensus that the exponential function is more appropriate for analysing short distance interactions such as those that take place within an urban area.' With that said, de Vries et al. (2009) demonstrate that in Copenhagen, commuting flows are better described by a power decay function except at the shortest and longest intra-urban distances. Given the relatively wide belief that the exponential decay function better models intra-urban flow, then, we choose to use it in our model. When taken together, then, the following spatial interaction model is generated:

$$T_{ij} = A_i O_i D_j^{\gamma} exp(-\beta d_{ij})$$

Where $A_i O_i$ is the known set of origin balancing factors and mass terms, respectively, D_j is the set of destination mass terms and γ its scaling parameter. Finally, $exp(-\beta D_{ij})$ is our negative exponential distance decay factor, where β is the distance scaling parameter and d_{ij} is the distance between origin and destination

To calibrate this model and determine the most appropriate values for parameters

[8] Battersea Park station was removed from the data as it had not yet opened when flows were collected

γ and β , we'll apply a Poisson regression model, as per Flowerdew & Aitkin (1982), by taking the log of the above equation. This yields:

$$\lambda_{ij} = \exp(\alpha_i + \gamma \ln D_j - \beta \ln d_{ij})$$

The results of this calibration against known flows yields a value for γ of 0.7509 and for β of 0.0015, suggesting that our model is relatively insensitive to the distance between origin and destination, but quite sensitive to employment count at destination. Our final spatial interaction model for London is thus:

$$T_{ij} = A_i O_i D_j^{0.7509} exp(-0.0015 d_{ij})$$

2.2 Scenarios

We will now apply the above gravity model to the following scenarios:

2.2.1 Scenario A: Decrease in Jobs at Canary Wharf

In this scenario, Brexit has caused jobs around the station at Canary Wharf to decrease by half, dropping from 58,772 to 29,386. With this change, estimated flows will be recomputed whilst holding the number of commuters constant. To achieve this, we can recalculate our set of origin balancing factors A_i to ensure that total commuter counts remain consistent. Because $O_i D_j^{0.7509} exp(-0.0015 d_{ij})$, from our interaction model consists entirely of known values, we can recompute A_i with ease whilst holding our sum total of flows constant.

2.2.2 Scenario B: Increase in Transport Cost

In Scenario B, we will consider the effects of a system-wide increase in transport cost. Given that our data captures no true measure of 'transport cost', we will use distance as a proxy for cost and thus make modifications to our distance scaling parameter β to reflect these changes.

Given that our default β value of 0.00015 is quite small, especially relative to the destination scaling parameter ($\gamma = 0.7509$) we will consider the impact of increasing β by 50% to 0.00023 and doubling it to 0.00030 to reflect large changes to distance sensitivity. We will name these scenarios B1 and B2, respectively. As with Scenario A, we will ensure the total commuter count is preserved by recalculating our origin balancing factors A_i.

2.2.3 Changes in Flows from Scenarios

In Scenario A, changes in most flows relative to the default model were relatively minor. Most station pairs saw no change (roughly 85%) or a slight increase (roughly 14%) in flows at the expense of every pair with Canary Wharf as a destination. Total flows terminating at Canary Wharf decreased by 38%, from 48,514 to 30,075. This decrease is only slightly smaller than the 50% decrease in overall employment at Canary Wharf, as dictated by the scenario parameters.

In the case of both scenarios B1 and B2, however, changes to even a small β value to begin with massively shift the resulting flow estimates. Whereas over 85% of flows

remained identical in scenario A, this figure dropped to just 23.5% for scenario B1, and 17.6% for scenario B2. In fact, the flow between pairs changed, on average, 38.61% for scenario B1 and 63.15% for scenario B2 as opposed to just 1.33% for scenario A (excluding flows which 'increased infinitely' from zero). It's thus extremely clear that scenarios B1 and B2 both impact flow redistribution far more than scenario A.

When considering how flows changed specifically in the case of scenarios B1/B2, station pairs with the lowest relative distances between them saw their flows increase the most. Indeed, this effect was magnified in scenario B2, relative to B1. Indeed, the average distances for each scenario are shown in Table 13 (with A as a baseline).

Scenario	β	**Average Trip Distance (km)**
Scenario A	0.00015	8.59
Scenario B1	0.00023	7.23
Scenario B2	0.00030	6.04

Table 13: β and Average Trip Distance per Scenario

We can see this effect visually by plotting the distance between station pairs against the total number of flows between them (Figure 9)

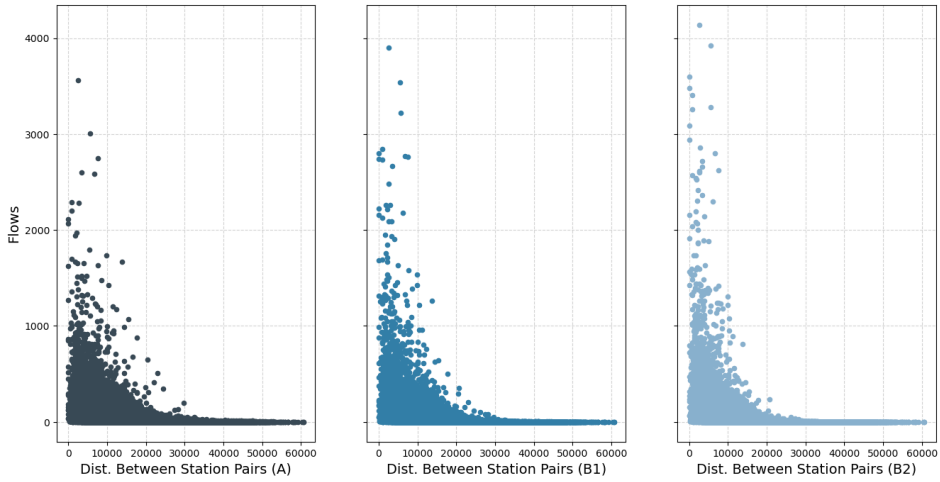

Figure 9: Station Pair Distance v.s. Flows for all Scenarios

To better visualise the importance of distance in each scenario, we can regenerate the above figure using the log of the flows between each station pair, as seen in Figure 10.

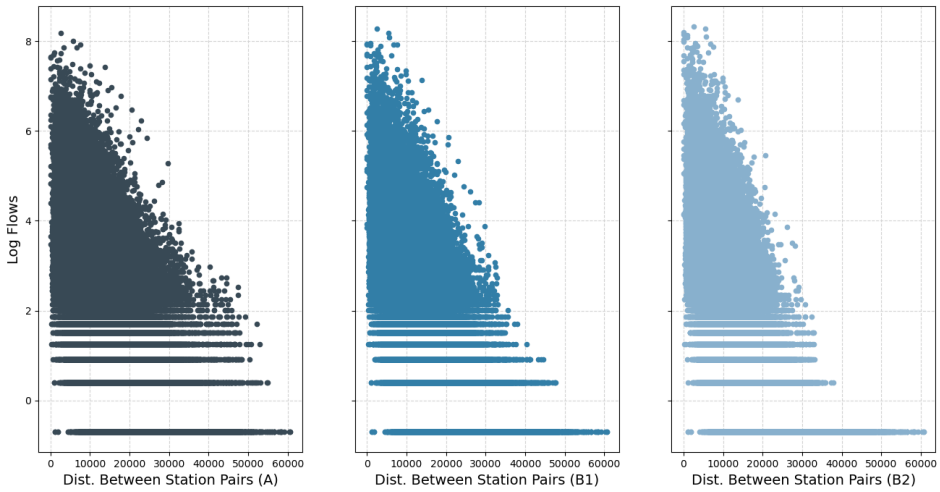

Figure 10: Station Pair Distance v.s. Log of Flows for all Scenarios

Note that 0.5 commuters have been added to each flow count in order to allow the logarithm of each flow to be computed without dividing by zero.

To demonstrate the increasing importance of distance in determining the extent of flows between stations, least squares regression was used to determine the percent of variance in flow count which can be explained by station pair distance in each of our three models. The equation to represent these regressions are:

$$\ln Flows = \alpha + \beta_1 Distance$$

Where β_1 represents the coefficient of the least squares regression line. The results of this regression are as follows:

Scenario	R^2	β_1
Scenario A	0.371	$-1.06 * 10^-4$
Scenario B1	0.456	$-1.28 * 10^-4$
Scenario B2	0.504	$-1.41 * 10^-4$

Table 14: Least Squares Regression Results for Scenarios

Although the use of linear regression in this situation is not rigorous[9], these results still succinctly demonstrate the increasing importance of distance between scenarios A, B1, and B2. Higher sensitivity to distance in B1 and B2, as modelled with a higher β is reflected by a higher coefficient of determination with respect to inter-station distance.

Resources

References

Batty, M. (2009), Urban Modeling, *in* 'International Encyclopedia of Human Geography', Elsevier, pp. 51–58.
URL: *https://linkinghub.elsevier.com/retrieve/pii/B9780080449104010920*

[9]Our regression results violate the assumption of homoscedasticity

Bloch, F., Jackson, M. O. & Tebaldi, P. (2023), 'Centrality measures in networks', *Social Choice and Welfare* **61**(2), 413–453.
URL: *https://doi.org/10.1007/s00355-023-01456-4*

de Vries, J. J., Nijkamp, P. & Rietveld, P. (2009), 'Exponential or Power Distance-Decay for Commuting? An Alternative Specification', *Environment and Planning A: Economy and Space* **41**(2), 461–480. Publisher: SAGE Publications Ltd.
URL: *https://doi.org/10.1068/a39369*

Derrible, S. (2012), 'Network Centrality of Metro Systems', *PLoS ONE* **7**(7), e40575.
URL: *https://dx.plos.org/10.1371/journal.pone.0040575*

Flowerdew, R. & Aitkin, M. (1982), 'A Method of Fitting the Gravity Model Based on the Poisson Distribution*', *Journal of Regional Science* **22**(2), 191–202. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9787.1982.tb00744.x.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9787.1982.tb00744.x*

Fotheringham, A. S. & O'Kelly, M. (1989), *Spatial interaction models: formulations and applications*, Studies in operational regional science, Kluwer Academic, Dordrecht. OCLC: 781941244.

Latora, V. & Marchiori, M. (2001), 'Efficient Behavior of Small-World Networks', *Physical Review Letters* **87**(19), 198701. Publisher: American Physical Society.
URL: *https://link.aps.org/doi/10.1103/PhysRevLett.87.198701*

Li, Z. & Hensher, D. A. (2013), 'Crowding in Public Transport: A Review of Objective and Subjective Measures', *Journal of Public Transportation* **16**(2), 107–134.
URL: *https://www.sciencedirect.com/science/article/pii/S1077291X22012486*

Stamos, I. (2023), 'Transportation Networks in the Face of Climate Change Adaptation: A Review of Centrality Measures', *Future Transportation* **3**(3), 878–900. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.
URL: *https://www.mdpi.com/2673-7590/3/3/49*

Watts, D. J. & Strogatz, S. H. (1998), 'Collective dynamics of 'small-world' networks', *Nature* **393**(6684), 440–442. Publisher: Nature Publishing Group.
URL: *https://www.nature.com/articles/30918*

Wilson, A. G. (1971), 'A Family of Spatial Interaction Models, and Associated Developments', *Environment and Planning A: Economy and Space* **3**(1), 1–32. Publisher: SAGE Publications Ltd.
URL: *https://doi.org/10.1068/a030001*

Zhao, F. & Ubaka, I. (2004), 'Transit Network Optimization – Minimizing Transfers and Optimizing Route Directness', *Journal of Public Transportation* **7**(1), 63–82.
URL: *https://www.sciencedirect.com/science/article/pii/S1077291X22003824*