

MATH 70103 - Assessment 2, CID: 06026467

I have worked independently on this assignment.

London Underground Network Vulnerabilities: Comparing the Effects of Targeted and Random Station Closures

1 Problem Statement and Motivation

In this paper, I analyse whether the London Underground network (the Tube) suffers more from planned station closures or random failures. I do so by comparing the effects of removing targeted and random stations for both the full network and for central London.

The Tube is a vital part of the transportation system in London, hosting up to 5 million passengers per day (Transport for London (2025c)). As the world's first underground railway (Transport for London (2025a)), the Tube is especially prone to station closures, causing massive disruptions across London. These closures may arise from planned maintenance closures. They may equally stem from random failures; for example, according to the BBC, in May 2025, power failures resulted in the closure of multiple stops in the Tube (BBC News (2025)). While the planned closures are meant to reduce the number of random failures (for example, for upgrades and renewal work (BBC News (2010))), the planned failures themselves can cause massive disruptions. The natural question therefore arises about whether it is better to continue with planned closures for maintenance, despite any disruption caused, or whether to keep important stations open and allow for random failures.

It is important to consider the effects of random and targeted station closures on the entire Tube network and compare against central London, which generally has more Tube stations with better connection. Specifically, this paper addresses the following questions:

1. Which are the five most important Tube stations by their connectivity?
2. By how much does the average time to traverse the Tube change if these five stations were closed?
3. By how much does the average time to traverse the Tube change if five random stations were closed?
4. Are there any parts of the Tube that are isolated if the top five Tube stations close? What about if five random stations close?
5. Do these results change if we consider only central London?
6. Is it better for the Tube to suffer from random closures or targeted closures?

By answering these questions, this paper finds that for the entire Tube network, random failures are more detrimental than planned closures, as entire parts of the Tube network can be disconnected from the map due to random failures, but with

planned closures one is still able to traverse a less straightforward path to their destination. However, for central London only, planned closures are more detrimental than random failures, as closing important stations means that one cannot get to destinations by staying only in central London; in other words, they will be forced to go outside of central London if they wish to reach their destination.

This paper also finds that relying on how much the average time to traverse the Tube network changes after removing nodes can be misleading, as the measures do not take into account parts of the network that are fully disconnected after removing nodes.

In section 2, I discuss the data I selected and how it was prepared for analysis; in section 3, I describe and justify the methods used for the analysis; in section 4, I interpret and reflecting on the output of the analysis; in section 5, I conclude with policy recommendations and further steps for the research.

2 Data Selection

The data used in this paper are from the Transport for London’s (TfL) Unified API (Transport for London (2025b)). Specifically, by querying this API, I was able to obtain a list of Tube stations and geographical coordinates in order to recreate the Tube map of London faithfully. The data are free and fair to use.

Each stop in the API is represented as a node in an undirected network, and the station name and latitude and longitude associated with the node is also stored. Adjacent nodes are connected by an undirected edge, such that the edge represents the track between two stations for (at least) a Tube line.

The coordinates of each Tube station are used to calculate the distance between the two stations. This paper uses Euclidean distance for its simplicity, but a Haversine distance (scikit-learn developers (2025)) may have also been used to incorporate the angular distance caused by the Earth’s curve.

The resultant graph, shown in Figure 1, is a recreation of the Tube network such that each node is a Tube station and each edge represents the track between the stations. To note is that all further analysis measures connectivity between stations by distance rather than time it takes to traverse along the edges.

All analysis is done with this spatial map of the Tube. In the next section, I describe and justify the analysis and methods used.

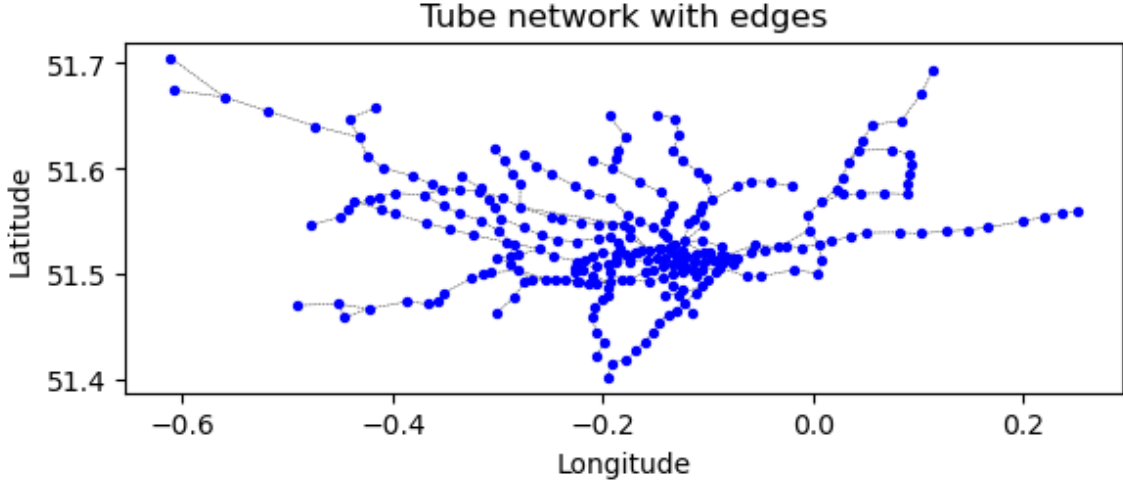


Figure 1: Map of the London Underground

3 Description and Justification of Methods and Analysis

3.1 Description and Justification of Methods

To answer the six questions outlined in section 1, I needed to compute the following metrics: average shortest path length, betweenness centrality, largest connected components, and community detection.

Average shortest path length: The average shortest path length indicates, for any two random nodes (in this case, stations), the expected length (in this case, in kilometers) between the two nodes. It is a proxy for how quickly one can traverse the network, making it a benchmarking statistic when measuring how station shut-downs would affect the efficiency of the Tube network. Note that I calculate the shortest path by distance rather than number of nodes, as, for a user, it matters less how many stops one passes by and more how long it takes to get to one's destination.

To calculate the average shortest path, this paper utilises Dijkstra's algorithm, which is a greedy shortest-path algorithm that takes a source node and calculates the minimum distance to get to every other node in the network. The algorithm is easy to implement and intuitive to understand, while still being effective in calculating average shortest paths, making it useful for the purpose of the Tube network.

Betweenness centrality: Given that we know shortest paths in a network, a node (or edge) has high betweenness centrality if it lies on a large number of shortest paths. In the context of the Tube network, these nodes are the stations that are crucial to getting between locations in London quickly; consequently, the closure of these stations would result in larger inefficiencies than the closure of stations with low betweenness centrality.

This paper specifically considers geodesic betweenness centrality, which considers strictly shortest paths, as opposed to random walk betweenness centrality, which utilises densities of random walks.

Largest connected components: Once nodes are removed from a network (in this case, due to station closures), certain sections of the network may be completely disconnected from the rest. The largest connected component measures, from the remaining disjointed network), what the maximum number of connected nodes (or edges) are in the largest subset of the network. In the case of the Tube network, this will indicate how isolated parts of the Tube network are due to station closures.

Community detection: Within networks, community detection aims to identify the presence of node clusters that are more connected to each other than they are to other parts of the network. In the Tube network example, one may expect that stations in north London would be more connected to each other than to stations in west London, even though it would be possible to traverse the network to go between the two areas. In this paper, community detection is an important tool to identify which communities, if any, are isolated by station closures.

This paper uses the Girvan-Newman algorithm, which is a divisive algorithm where *edges* with high betweenness-centrality are removed, leaving behind smaller connected components that are communities. By doing so, I can examine whether *nodes* with high betweenness centrality sits at the border to different communities. To note is that edges with high betweenness centrality do not always result in their nodes having high betweenness centrality, and vice versa.

3.2 Description and Justification of Analysis

Using the methods outlined in section 3.1, I answer the questions outlined in section 1 as follows:

- 1. Which are the five most stations by their connectivity?** The top five stations are the ones with the highest betweenness-centrality.
- 2. By how much does the average time to traverse the Tube change if these five stations were closed?** Before making any changes to the network, I calculate the average shortest path using Dijkstra's algorithm, Once the top five stations by betweenness centrality are removed from the network, I recalculate the average shortest path for comparison.
- 3. By how much does the average time to traverse the Tube change if five random stations are closed?** I randomly remove any five nodes and recalculate Dijkstra's algorithm to compare by how much the average shortest path has changed.
- 4. Are there any parts of the Tube that are isolated if the top five Tube stations close? What about if five random stations close?** I calculate the largest connected component of the network after the removal of five stations, and visualise the network by their disjointed components. Furthermore, I visualise whether the five stations lie between communities, as detected by the Girvan-Newman algorithm, to see whether those stations' closures would isolate the detected communities.
- 5. Do these results change if we only consider central London?** I bound the network by a geographical range to focus on the geographical centre of London, and repeat questions 1 to 5 for this area. By doing so, I can compare whether the

results differ to the centre of London, which has the highest population density in all of London (Trust for London (2024)); thereby, station closures may affect more people in central London.

6. Is it better for the Tube to suffer from random closures or targeted closures? I use the findings from the previous five questions to analyse whether random closures are more disruptive than planned closures. Specifically, I check whether communities are isolated, how large the largest connected component is, by how much station closures change the average shortest path measure, and whether results differ between all of London and central London.

In the following section, I display and interpret the results using the methods and analyses outlined in this section.

4 Interpretation and Reflection of Output

In this section, I discuss the results of my findings in relation to the questions outlined in section 1, using the methods and analyses outlined in section 3. In each section, I compare the results from the entire Tube map to central London only.

Base case

The Tube network, pictured in Figure 1, has 272 nodes and 5 edges. In comparison, the central London Tube network, pictured below in Figure 2, has 133 nodes and 166 edges.

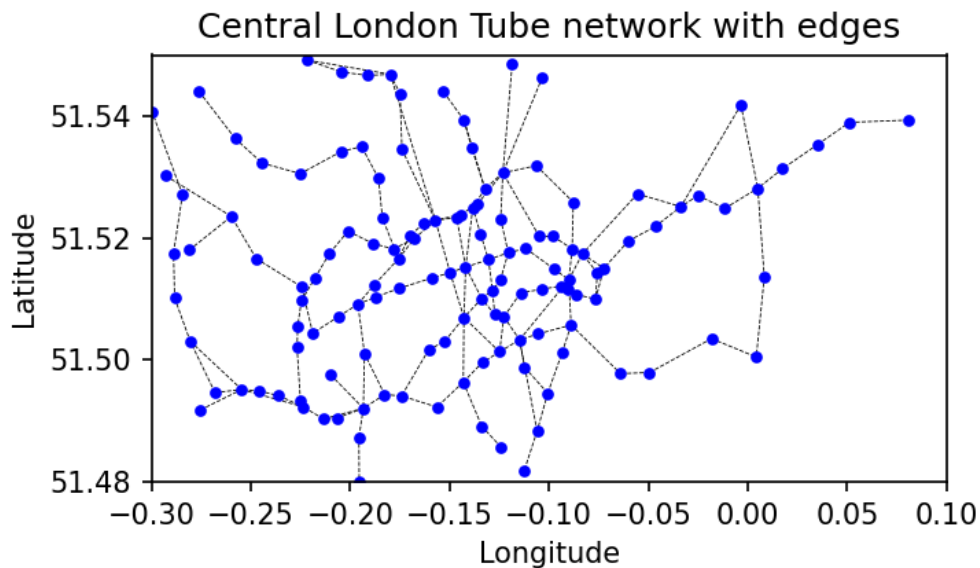


Figure 2: Map of the central London Underground

In this paper, central London is defined geographically, i.e. central London is defined as all nodes that lie between a latitude of -0.3 to 0.1 and a longitude of 51.48 and 51.55 , which was decided using visual inspection.

1. Which are the five most important Tube stations by their connectivity?

In the entire Tube network, the top five stations with the highest betweenness centrality are Baker Street, King's Cross St.

Pancras, Liverpool Street, Aldgate East, and Whitechapel.

In central London, the top five stations with the highest betweenness centrality are Baker Street, Edgware Road, Oxford Circus, Bond Street, and Earl's Court.

These stations represent the nodes that lie between the greatest number of shortest paths in their respective networks.

2. By how much does the average time to traverse the Tube change if these five stations were closed?

The table below summarises the change in average shortest path before and after removing the five nodes corresponding to the highest betweenness centrality, as outlined in question 1.

	Before Nodes Removal	After Nodes Removal	Change
All of London	26.18 km	32.44 km	23.91%
Central London	12.41 km	12.4 km	-0.08%

Table 1: Average Shortest Path after Top 5 Nodes Removal (Dijkstra's Algorithm)

At first, it appears as though the removal of the top five nodes disproportionately disadvantages the entire Tube network over central London's network. Specifically, the average shortest path increases by around 24% for the entire Tube map, but remains roughly the same (and even becomes shorter). One might erroneously surmise that targeted station closures (corresponding to targeted node removals) have little effect on central London, even though they may negatively impact the rest of the network.

In fact, Dijkstra's algorithm only calculates average shortest path for nodes that are connected by edges; that is, parts of the network that are now unreachable after the closure of stations is not included in the calculation. As seen in the results for question 4 below, the closure of these five stations isolates significant parts of the network, leaving a much smaller section remaining, which consequently reduces the average time it takes to traverse the smaller network.

3. By how much does the average time to traverse the Tube change if five random stations are closed?

The randomly chosen five nodes that are removed from the entire Tube network are Northolt, Warwick Avenue, Southfields, Hammersmith, and Upton Park. The randomly chosen five nodes that are removed from the central London Tube network are St. Paul's, Blackfriars, Shepherd's Bush (Central), Liverpool Street, and Bow Road.

The table below summarises the change in average shortest path before and after removing these five nodes.

	Before Nodes Removal	After Nodes Removal	Change
All of London	26.18 km	25.38 km	-3.06%
Central London	12.41 km	13.57 km	9.35%

Table 2: Average Shortest Path after Random 5 Nodes Removal (Dijkstra's Algorithm)

The data suggest that the average time to traverse all of London decreases by roughly 3%, but the average time to traverse

central London increases by roughly 9%. This is exactly the opposite to what Dijkstra's algorithm suggests when removing the top 5 nodes by betweenness centrality. Overall, one might surmise that it is beneficial to have random failures (corresponding to random node removals) in central London and beneficial to have targeted station closures (corresponding to targeted node removals) for all of London.

However, this would once again be an erroneous conclusion. Aside from the fact that Dijkstra's algorithm does not take into account parts of the network that are isolated by the random removal of nodes, it is worth remembering that these are five randomly removed stations; another set of randomly removed stations may change the results entirely.

Therefore, examining whether targeted closures or random failures are more detrimental to the Tube network cannot be done solely using average shortest path calculations; further calculations (as shown in questions 4 and onwards) are needed to form a robust conclusion.

4. Are there any parts of the Tube that are isolated if the top five Tube stations close? What about if five random stations close?

a. Map visualisation

The figure below compares, for the entire London Tube network, which parts of the network have splintered off from the rest as a result of removing nodes.

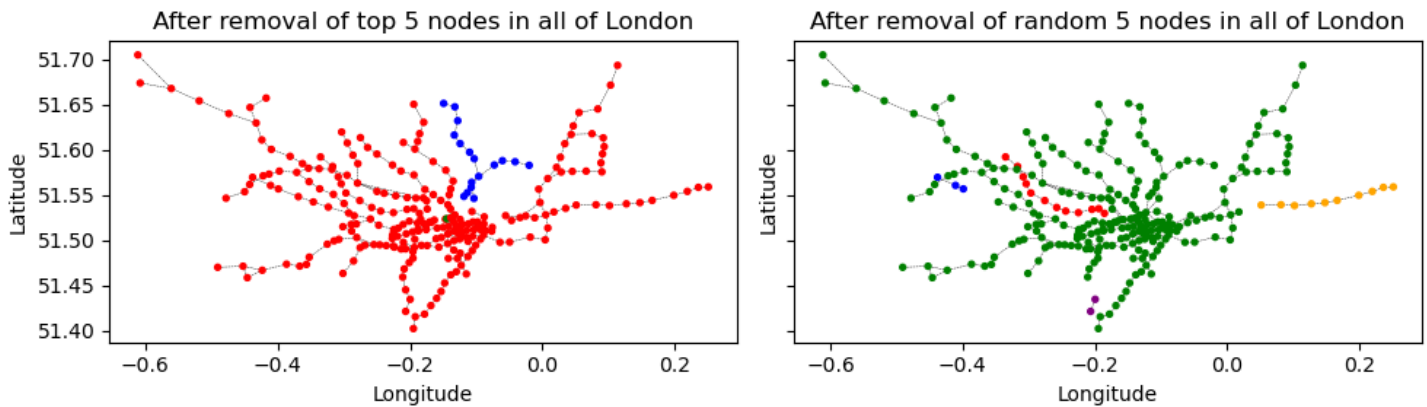


Figure 3: The splintered network for the entire Tube network after node removals

After removing the top five nodes, the largest component size (pictured in red on the left in figure 3), has 248 nodes, in comparison to the 272 nodes for the original Tube map shown in figure 1. After removing the random five nodes, the largest component size (pictured in green on the right in figure 3) has 240 nodes, compared to 272 for the entire Tube map.

As seen in figure 3, removing targeted nodes (top five nodes, corresponding to planned closures of stations) can leave parts of the map isolated, but much fewer than random node removals (corresponding to random failures). This suggests that, regardless of average shortest path, there are ways to traverse most parts of the Tube despite planned closures, likely through indirect routes and Tube changes.

In comparison, random closures can isolate larger parts of the map, as random node removals are less likely to be solely focussed in parts of London that have alternate routes (e.g. central London). This suggests that there are no alternative paths to get to large parts of outer London. Therefore, in the entire Tube network, planned closures are less detrimental than random failures.

The figure below compares, for only central London, which parts of the network have splintered off from the rest as a result of removing nodes.

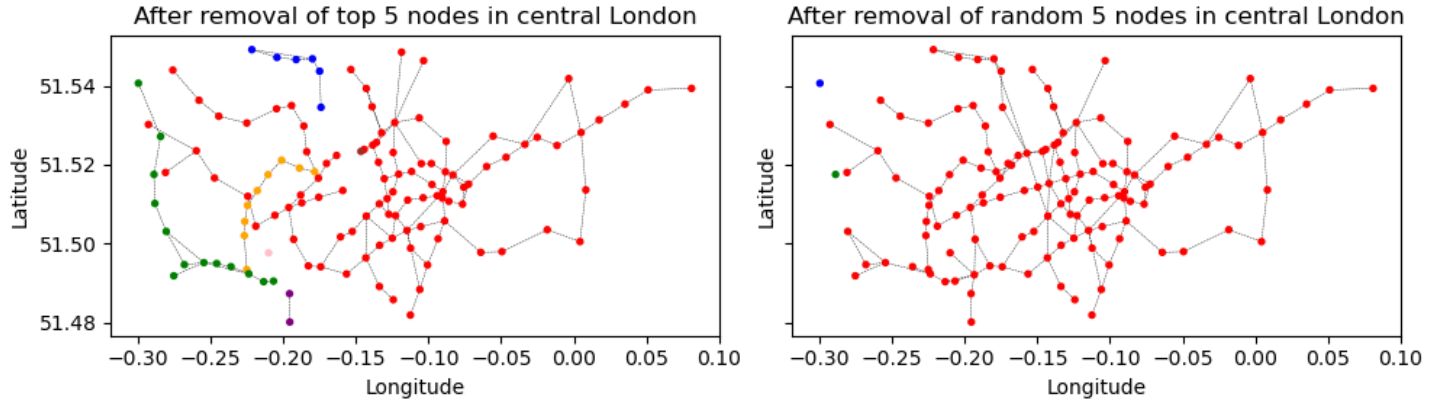


Figure 4: The splintered network for the central London tube network after node removals

After removing the top five nodes, the largest component size (pictured in red on the left in figure 4), has 96 nodes, in comparison to the 133 nodes from the original central London tube map shown in figure 2. After removing the random five nodes, the largest component size (pictured in red on the left in figure 4) has 123 nodes, compared to 133 for the entire Tube map.

The figure suggests that removing targeted nodes (top five nodes, corresponding to planned closure of stations) leaves larger sections of the map isolated than randomly removing five nodes (corresponding to random failures). This suggests that, if important stations are closed (for example, King's Cross), it will be impossible to get to certain stations from central London alone - one would have to leave the centre to get to another part of central London, if it is at all possible.

In comparison, while random failures may isolate smaller sections of the central London network, so long as the most important stations (measured by betweenness centrality) remain open, it is more likely that there will be a way to traverse the central London network, even if it takes longer.

Note that this conclusion is in direct contrast to the conclusion drawn for the entire Tube network: while, for central London, random failures are preferred to planned closures, for the entire Tube, planned closures are preferred to random failures. This is explained in more depth when answering question 5.

b. Girvan-Newman

In this paper, I partition the map into four communities using the Girvan-Newman algorithm. The figure below compares,

for the entire Tube network, where the communities lie in comparison to the five closed stations.

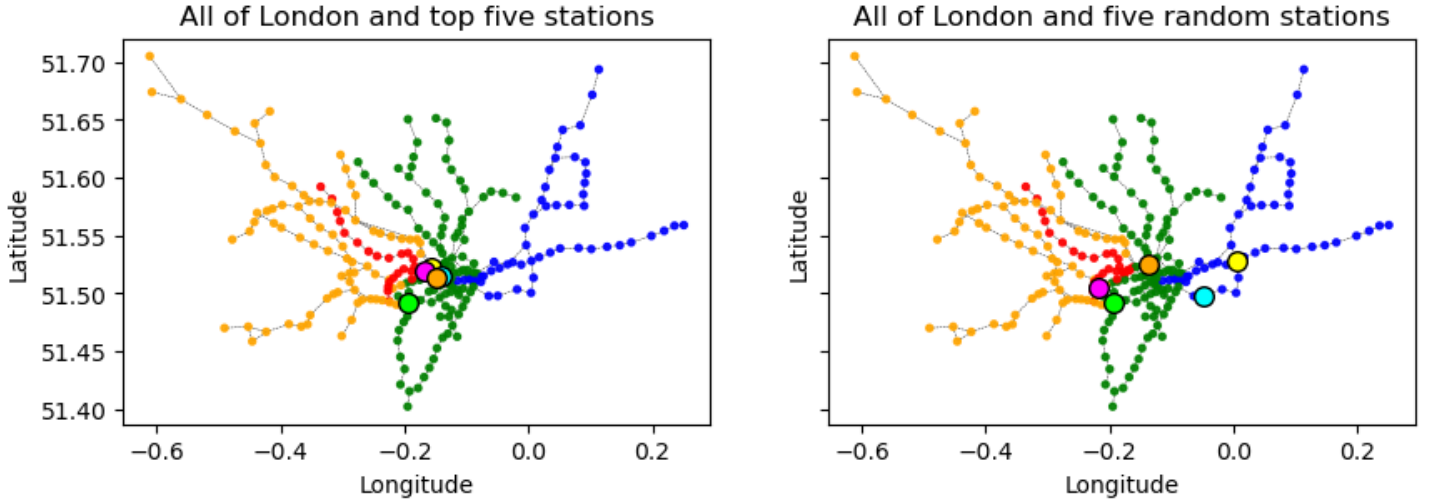


Figure 5: Community detection (Girvan-Newman) in relation to closed stations for the entire Tube network

Note that each community has a different colour and each of the closed stations is indicated by a coloured dot. At the level of four communities, there are two communities in west London, one community in north, south, and central London, and one community in east London. The top five stations by betweenness centrality are all clustered in the centre of London, whereas the five random stations are spread out.

The figure above suggests that the top five stations lie between communities more so than five random stations do. Combined with the results shown in figure 3, this implies that communities are not fully isolated should the top five stations close. However, if five random stations close, they are likely to close mid-community, which can isolate the community from the rest of the network. In other words, within a community, there is a heavier reliance on a set of nodes (or, equivalently, a set of stations) to traverse London.

The figure below considers the four communities within central London using the Girvan-Newman algorithm, along with the five closed stations.

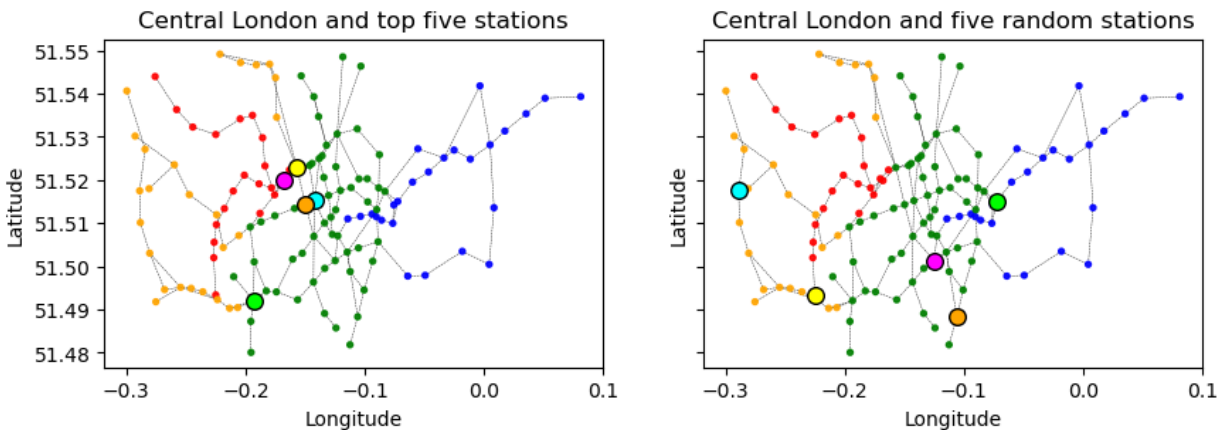


Figure 6: Community detection (Girvan-Newman) in relation to closed stations for central London

The communities are demarcated by different colours, and the five closed stations are demarcated by coloured dots. The communities in central London are stratified as such: two communities in west London, one in central London, and one in east London. The top five stations (pictured on the left of figure 6) are closer to (but do not coincide exactly with) the borders between communities, and the random five stations (pictured on the right of figure 6) are more spread out.

Combined with the findings from figure 4, it suggests that the Tube stations that help connect communities are more likely the only way to get between communities in central London, making their planned closure much more difficult for users.

6. Is it better for the Tube to suffer from random closures or targeted closures?

The findings in their entirety suggest that, for a user who enters the Tube in central London, in order to traverse the Tube during planned closures (targeted node removal), they must leave central London to get to their stop. Staying within central London will be much more difficult as larger sections of the Tube are isolated compared to station closures due to random failures (random node removal). Therefore, from the point of view of convenience for someone traversing central London, random closures are less detrimental than targeted closures.

For a user who enters the Tube outside of central London, random closures may well isolate them from the rest of the network, as large communities in London rely on lines that have no alternative (equivalently, they rely on traversing through a specific set of nodes and edges). Targeted closures, however, can be navigated by taking a longer detour to get to their desired stop, and are therefore preferred to random closures.

In section 5, I conclude by elaborating on this point from a policy perspective, and discuss further steps in future research.

5 Conclusion

5.1 Policy Recommendations

On one hand, central London is home to large amounts of business activity and jobs (City of London Corporation (2025)), as well as having high population density (Trust for London (2024)). Navigating the area with ease and speed is therefore important for the economic activity and urbanisation of London. From this point of view, requiring users to navigate outside of central London to get back into central London during planned Tube closures is a detriment to not only the users but also to the foot traffic in alternate lines during planned closures. It would therefore imply that reducing planned maintenance (*ceteris paribus*; i.e. not taking into account safety measures or transport ordinances) and allowing for random failures is ultimately better for central London.

On the other hand, allowing for random failures will unfairly isolate those who do not live in central London. Planned closures, at least, will afford them mobility, even if it is at the detriment of speed of travel for central London users. It would therefore imply that planned maintenance, which would reduce the likelihood of random closures (BBC News (2010)), cannot be sacrificed for central London.

The policy recommendation put forth in this paper is to favour planned closures over allowing random failures, but to mitigate the effects of planned closures on central London by building in more redundancy into the network. Specifically, stations in central London that lie between communities, or that otherwise isolate parts of the central London network, cannot be the only way to traverse between stops in central London. More lines built around these important areas would ensure that planned closures will not isolate central London, but will allow them to take place so as to reduce the chance of random failures, to the benefit of all of London.

5.2 Further Steps

The two main areas for further steps are in the definition of central London, and in the modes of transport that are considered in this paper.

Firstly, in this paper, I approximated central London by a geographical bounding box. However, as London is split into zones and the data are available in the TfL United API, it would be a natural next step to merge the zone tags and filter by zones 1 and 2 to define central London. The advantage of this would be a more precise definition of central London.

Secondly, I considered only the Tube network as a means of transportation around London, but including bus lines (which are also available in the TfL United API) would present a richer picture of navigating London, especially as not everyone in London lives near a Tube station and therefore rely on buses.

With these two extensions, definitions of communities, calculations of average shortest paths, and visualisations of isolated parts of the network due to station closures would present a richer, more accurate view of the London transport network.

With these additions, one can more precisely ascertain whether planned station closures or random failures are better for the Tube, allowing us to shape policy recommendations to serve all of London.

References

BBC News (2010), ‘Three week tube closures ‘considered’, lu director says’. Published: 13 December 2010; Accessed: 7 January 2026.

URL: <https://www.bbc.com/news/uk-england-london-11982645>

BBC News (2025), ‘Power failure knocks out london underground lines’. 12 May 2025; Accessed: 7 January 2026.

URL: <https://www.bbc.com/news/articles/cd7gnzyd7nro>

City of London Corporation (2025), ‘City statistics briefing’. Updated: 7 November 2025; Accessed: 7 January 2026.

URL: <https://www.cityoflondon.gov.uk/supporting-businesses/economic-research/research-publications/city-statistics-briefing>

scikit-learn developers (2025), ‘sklearn.metrics.pairwise.haversine_distances’. Accessed : 7 January 2026.

URL: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.pairwise.haversine_distances.html

Transport for London (2025a), ‘A brief history of the underground’. Accessed: 7 January 2026.

URL: <https://tfl.gov.uk/corporate/about-tfl/culture-and-heritage/londons-transport-a-history/london-underground/a-brief-history-of-the-underground>

Transport for London (2025b), ‘Unified api’. Accessed: 7 January 2026.

URL: <https://tfl.gov.uk/info-for-open-data-users/unified-api>

Transport for London (2025c), ‘What we do’. Accessed: 7 January 2026.

URL: <https://tfl.gov.uk/corporate/about-tfl/what-we-do-on-this-page-1>

Trust for London (2024), ‘London’s geography and population’. Last updated: November 2024; Accessed: 7 January 2026.

URL: <https://trustforlondon.org.uk/data/geography-population/>