LONDON'S GLOBAL UNIVERSITY
Department of Computer Science

Enhancing Urban Transport Resilience: A Weighted Network Model Informed by Metro Congestion and Station Removal Vulnerability

Yijie Lu

Supervisors: Lopane, Fulvio

MSc Urban Spatial Science

Centre for Advanced Spatial Analysis

University College London

This dissertation is submitted in partial fulfillment of the requirements for the degree of MSc in the Centre for Advanced Spatial Analysis

Word count: 10,183 words

August 2023

Abstract

1 | Abstract

Urban transportation networks, particularly metro systems, play a pivotal role in supporting daily commuting and travel activities in cities. Disruptions to these networks, such as subway strikes, can lead to serious impacts ranging from increased congestion and longer commute times to significant economic impacts. Through empirical research, this study identifies the wide-ranging impacts of subway strikes, emphasizing the need for preparedness and resilience. Around this context, this study explores the concept of a weighted network to assess the impact of a hypothetical subway station removal. Through this lens, the objectives of this study include integrating the principles of weighted networks to assess the impact of station demolition on urban traffic flows and to provide alternative route recommendations for urban planners. This study, while not directly modeling a strike scenario, lays the groundwork for cities to assess the potential vulnerability of their subway systems with the goal of improving the resilience and adaptability of urban transportation networks.

Keywords— Urban transportation networks - Tube strikes - weighted network analysis - data-driven decision making - resilience

Declaration

I, Yijie Lu, I declare that the thesis has been composed by myself and that the work has not be submitted for any other degree or professional qualification. I confirm that the work submitted is my own, except where work which has formed part of jointly-authored publications has been included and referenced. The report may be freely copied and distributed provided the source is explicitly acknowledged.

Yijie Lu 20/08/23

Signature *Date*

Table of Contents

List of Figures

List of Tables

2 | Introduction and Background

2.1 Motivation and Description of Problems

A metro system serves as a lifeline for countless urban dwellers, facilitating their daily commuting and travel needs. Its efficiency and rapid transit capabilities make it indispensable. However, the system's seamless functioning can be jeopardized by two main factors: metro strikes and extreme congestion due to high passenger traffic.

Metro strikes, precipitated when staff members halt their services, can wreak havoc on an urban transport framework. The immediate consequence is a substantial disruption to the city's transport dynamics. Commuters, left with no choice, often resort to alternate transportation modes such as buses, taxis, or private cars (Lin, 2017)[1]. This sudden shift can overburden these alternative systems, escalating congestion and diminishing their operational efficiency.

Moreover, strikes lead to extended travel durations. Given the metro's reputation for punctuality and swiftness, its alternatives often fall short. Tsapakis et al. (2012) discovered that during the Tube strikes in London between 2009 and 2010, travel delays were rampant, with some days witnessing travel times swelling by up to 74%[2]. The economic ramifications are equally dire. Businesses in proximity to metro stations suffer as employees face tardiness or absence. Exel and Rietveld (2001) found that during a thirteen-day subway strike in New York City on New Year's Day, 1966, various sectors experienced significant declines[3]. Church attendance decreased by 30%, restaurant business by 20-30%, cultural and entertainment events by 50-90%, and retailers received only 20-25% of their expected business. Another study indicated a 1% seasonally adjusted revenue loss for retailers six months after service was restored (Ferguson, 1992)[4].

On the congestion front, Haywood et al. (2018) highlighted that in-vehicle crowding in

Paris subways led to economic losses. Their findings suggested that the gap between current and optimal public transport (PT) patronage stood at 9%, translating to an economic cost of congestion of about 64.6 million euros annually, roughly 0.9% of the total user costs[5]. Another perspective by Schmöcker et al. (2002) posits that, unlike private transport users, public transport users may face significant uncertainty in waiting time during peak periods. Congestion can result in passengers being unable to board the first arriving vehicle due to a lack of space [6].

In essence, both metro strikes and high passenger traffic present multifaceted challenges to urban transit systems. Addressing these challenges is pivotal to bolster the resilience and sustainability of urban transport infrastructures.

2.2 Research Question and Objectives

Research Question: How can the integration of metro congestion levels and network vulnerability from potential subway station removals inform a weighted network model to enhance the resilience of urban transport networks and propose alternative routes?

2.2.1 Research Objectives

1. **Integration of Weighted Network Concepts Based on Congestion and Vulnerability:** Delve into the interplay of urban transport networks' dynamics, highlighting the importance of each station in terms of passenger traffic congestion and network fragility upon its removal.

2. **Assessing Network Vulnerability through Subway Station Removal:** Quantify the potential ripple effects of subway station removals on the overall network connectivity and traffic flow, and embed these outcomes into the weighted network paradigm.

3. **Formulation of Resilient Route Proposals via Weighted Network Analysis:** Utilize the weighted network model's insights to conceptualize alternative routes, aiming to guide urban planners especially when confronting unforeseen network disruptions.

4. **Crafting a Decision-Support Framework:** Develop a decision-aiding mechanism aimed at urban planners, leveraging the weighted network analysis, to grant instantaneous feedback, strategic direction, and optimal route choices.

2.3 Description of the Work to Address the Challenges

Given the potential disruptions to urban transportation networks caused by factors such as subway station removal or congestion events, this study aims to provide detailed understanding and strategic solutions through weighted network analysis. The overall goal is to provide urban planners with nuanced insights and strategies to understand and manage the impact of these disruptions on transportation infrastructure.

- **Weighted Network Analysis:** At the heart of this project is the use of weighted network analysis to provide insight into the resilience and vulnerability within the metro system. Through this approach, quantitative measures of the importance of subway stations and their routes will be generated, allowing planners to identify key nodes and connections that are vulnerable to disruptions.

- **Impact Assessment of Subway Station Removal:** Understanding the ripple effects of removing subway stations is critical. This project will assess how the loss of certain stations affects the efficiency and connectivity of the entire network, especially during potential disruptions.

- **Optimal Route Planning Based on Weighted Network Analysis:** Based on the insights gained from weighted network analysis and impact assessment, this study will propose route planning tools. By taking into account passenger volumes, expected travel times, and likely areas of congestion, these tools will propose optimized alternative routes that will provide planners with practical recommendations for managing disruptions.

- **Decision Support System Design and Application:** This project combines the analytical power of weighted network analysis with a decision support framework designed to provide urban planners with real-time insight, strategic direction, and recommendations for optimal routes before or during disruptions.

By weaving these elements into a unified approach, this research hopes to provide urban planners with an empirically based perspective on the analysis of urban transportation dynamics, especially in the face of disruptions. The expectation is to combine rigorous academic methods with practical applications to ensure that planners are able to make informed decisions when faced with transportation challenges.

3 | Literature Review

3.1 Background

A considerable fraction of commuters fail to optimize their daily travel routes due to the prevalence of noisy information and inefficient search behavior, as suggested by a study by Larcom et al.(2017)[7]. The study demonstrated the significant impact of exogenous factors, like subway strikes, on the commuting patterns and the importance of effective decision-support tools in improving urban transport networks' efficiency.

It is based on these findings that this project builds a decision support system based on weighted network analysis to help urban planners understand and cope with the impacts of the removal of key metro stations. The system aims to provide insights into potential alternative routes and optimise the resilience of urban transport networks in the face of such disruptions.

3.2 Prior Studies and Analysis

Investigating the resilience of subway systems through weighted network analysis is an area of research where numerous scholars have delved into a variety of methods to deepen our understanding of this intricate relationship.

3.2.1 Spatial Traffic Network Models

Scott et al. (2006) introduced an innovative approach known as the Network Robustness Index (NRI) for identifying key network links and assessing traffic network performance. Their method transcends the simple origin-destination (OD) demand, placing importance on network connectivity to avoid being overly sensitive to partial disruptions within the system[8]. This concept underscores the significance of transportation flexibility and reliability, providing a fresh perspective for traffic network planning. By testing with three hypothetical networks, they demonstrated that the Network Robustness Index (NRI) can generate road planning solutions distinct from traditional Volume/Capacity (V/C) ratios, and these solutions bring system-wide benefits (measured in terms of saved travel time) significantly greater than solutions determined by V/C ratios.

Similarly, Serdar et al. (2022) focused on spatial traffic network models aiming to assess the resilience of urban traffic networks[9]. Their methodologies are highly effective in identifying traffic bottlenecks and analyzing disruption factors, providing system-wide advantages.

While both the works of Scott et al. and Serdar et al. mainly concentrate on road traffic, their foundational principles and analytic methods still provide inspiration and reference for our subway network model. They emphasize the importance of network connectivity and offer powerful tools for dealing with traffic bottlenecks and disruption factors. The research can draw on these insights to identify and address similar issues within the subway network. For instance, the research might study which subway stations or routes experience the most pressure during sudden events such as subway strikes, and which congested routes have the most significant impact on the overall network performance. This would help us to better understand and improve the resilience of the subway network.

However, there are distinct differences between road traffic and subway systems. Therefore, applying these methodologies directly to subway systems might encounter challenges, as subway systems have their unique attributes. To address this issue, it needs to modify and extend these existing studies to better adapt to the characteristics and demands of the subway network.

3.2.2 Impact Analysis and Mitigation Strategies

Ghose et al. (2006) proposed the GIS Optimal Route Model for determining efficient collection routes with minimum cost/distance when transporting solid waste to landfills[10]. Ghose's model takes into account numerous parameters including population density, waste generation rate, road network, road type, storage bins, and collection vehicle characteristics. The strength of the model is that it provides municipalities with decision support to optimize their daily waste management operations. This includes transportation of solid waste, load balancing within vehicles, fuel consumption management, and carefully developed schedules for workers and vehicles. The basis for developing these collection routes is based on a GIS approach. Notably, the network module of the Arc/Info GIS software (ESRI, 1995) used in conjunction with the planned infrastructure was designed to find shortest paths or minimum impedance paths.

Complementing Ghose's work, Cadarso et al. (2017) not only focused on studying the impact of disturbances due to service disruptions, but also developed mitigation strategies for these disturbances[11]. Their approach deals with the uncertainty associated with disruptions in passenger service demand by introducing risk aversion measures, in particular the concept of VaR. These approaches not only provide new perspectives on network design, but also provide effective strategies to cope with various potential disruptions.

The studies of Ghose et al. and Cadarso et al. provide insights in dealing with transportation network disturbances and their impacts on urban infrastructure. In particular, Ghose's GIS optimal path model provides a structured approach to consider numerous relevant parameters such as population density and road network. This is particularly useful when developing a simulation strategy for subway strikes, as these parameters can help us better understand the pressure that certain subway stops may be subjected to under contingencies.

However, it is worth noting that our project focuses primarily on the metro network, whereas Ghose's model considers road traffic and solid waste management more. This means that the research needs to make careful adjustments to the model to ensure its applicability in simulating subway strikes and assessing network resilience. For example, it needs to consider metro-specific factors such as the interconnectivity between metro lines, the capacity of metro stations in terms of foot traffic and frequency of trips.

3.2.3 Network Design

The design and optimization of transport and network systems are essential considerations for modern cities and industries. Various researches have explored the principles governing such networks, with some drawing inspiration from natural systems.

Studies by Wey and Huang (2018) and Tero et al. (2010) have delved into network design and the potential of nature-inspired network formation[12][13]. Wey and Huang (2018) focused on urban sustainability and livability, particularly in Taipei City[12]. They used big data to study sustainable and livable transportation strategies, introducing a dynamic approach that accounts for temporal and spatial changes. By employing methods like the Fuzzy Delphi and ARIMA, they managed to predict dynamic trends for key indices, which subsequently informed decision-making strategies. Their research underscores the importance of integrating modern data techniques with urban planning, aiming not just for sustainability but also for a more transparent decision-making process that enhances urban life in the near future.

On the other hand, Tero et al. (2010) explored a biological approach to network development. Drawing inspiration from the slime mold Physarum polycephalum, they studied the organism's adaptive network development and foraging strategies. Physarum is known for its ability to find the shortest paths in mazes and create efficient networks when connecting food sources[13]. Tero and colleagues replicated this behavior by allowing the organism to connect a template of food sources, representing cities around the Tokyo area. The result was then compared with Japan's actual rail network. Their research suggests that capturing the essence of such biological systems in simple rules might guide decentralized network development in various domains.

These studies provide beneficial theoretical frameworks and inspiration for network design. However, as Wey and Huang (2018) noted, practical application may face challenges due to constraints such as geography, economy, and technology inherent to subway network design. Nevertheless, these research works offer insights into enhancing subway system resilience through strategic network design, provided the real-world limitations are considered[12].

3.2.4 Dynamic Weighted Network

Within the context of our project, the research is primarily focused on route optimization during subway strikes. Ensuring that passengers can still find the most efficient routes during a strike requires an understanding of dynamic weighted networks and how community structures are captured and analyzed within them.

The 2014 paper "Evolutionary Community Structure Discovery in Dynamic Weighted Networks" by Guo, Wang, and Zhang offers valuable insights in this regard. In their work, they explore how community structures can be identified and optimized within dynamic weighted networks(Guo et al., 2014)[14]. This aligns with the objectives of our project because, in the backdrop of a subway strike, the subway network will undergo dynamic changes; certain routes may be suspended, while others may witness increased congestion. By understanding how these dynamic networks function, it can better offer optimized route recommendations to passengers.

However, there are limitations to the approach proposed by Guo, Wang, and Zhang. While they introduced an Evolutionary Community Structure Discovery algorithm (ECSD) based on node analysis, this method struggles with overlapping and hierarchical community structures. In a system as intricate as a subway network, such overlapping and hierarchical community structures are likely the norm, and as such, it requires a methodology capable of handling this complexity.

Moreover, their take on link community analysis provides an interesting perspective. Within the subway network, certain links (or routes) between stations might be more critical or congested than others. Through link community analysis, it can better identify these critical links and, based on this, optimize the travel paths for passengers. However, this method too has its shortcomings, as it necessitates access to the entire network structure information, which might be a challenge during strikes.

In conclusion, the research by Guo, Wang, and Zhang offers invaluable insights and methodologies for our subway strike route optimization project. Nonetheless, further adaptation and refinement of these methodologies are needed to better fit our specific application.

3.2.5 Network Resilience

Building on the understanding of network resilience, Minette and Francesca(2015) delved deeper into the quantification of this concept in their study [15]. Drawing inspiration from Pimm's (1991) perspective [16], which emphasizes the speed at which a system returns to equilibrium after disruption, DâLima and Medda introduced an innovative measurement method based on mean reversion random models. Their research underscores the relevance of this method by demonstrating its ability to capture the characteristics of systems with diverse behaviors.

An in-depth literary investigation was the starting point of their research, focusing on resilience concepts across various system types. The broad span of their review, covering disciplines from ecology to disaster management, is instrumental. It illustrates the ubiquity of the quest to measure or quantify system resilience. By introducing some of these frameworks and measurement techniques, DâLima and Medda highlighted potential shortcomings in existing methodologies.

A major strength of their study lies in the practical application of their resilience measurement approach. Advocating the idea that systems constantly undergo minor random shocks and perturbations, they propose the system's resilience can be gauged by the speed at which its state returns to historical normal levels post disturbance. This behavior is encapsulated using a mean reversion random model. Taking a hands-on approach, they applied the model to the London Underground data, using passenger numbers as a proxy for the system's state.

In terms of the London Underground transportation system, DâLima and Medda's model offers the ability to assess the resilience of various subway lines to disturbances. This allows for a comparative study of all subway lines, pinpointing which ones exhibit higher or lower levels of resilience. By further employing a mean reversion model with jumps, the resilience of specific subway lines to both minor and major shocks can be evaluated. Such studies serve as critical tools for making informed investment decisions aimed at enhancing subway line robustness.

However, it is crucial to understand the nuances and limitations of DâLima and Medda's model in relation to our specific project. First, while their model is rooted in the London Underground system, our project might encompass different metro systems with unique characteristics, thereby demanding adjustments in the model's application. For instance, if our project was looking into an entirely different metro system with varying infrastructural and operational dynamics, it would have to ensure that the model can accommodate these variations. Furthermore, the reliance on detailed historical data, especially in a setting different from the London Underground, may present challenges if such data is not readily available or differs in structure and granularity.

Lastly, while DâLima and Medda's model stresses minor random shocks, our project may focus on larger strike. This distinction necessitates further exploration to ensure the model's applicability in our context. In conclusion, while DâLima and Medda's research offers a compelling perspective on resilience measurement, its application to our project necessitates thoughtful customization and keen awareness of its limitations.

3.2.6 Path Optimization

Path optimization involves finding the best solution among many network problems. While there are multiple algorithms available, it is imperative to choose the most suitable method for a particular scenario. This paper compares and analyzes Dijkstra's algorithm, A* algorithm and genetic algorithm based on Noto and Sato (2000)[17].

3.2.6.1 Algorithm Comparisons

1. A* Algorithm

Principle: Originating in the realm of heuristic searching algorithms, the A* algorithm is an extension of Dijkstra's, where it employs an evaluation function to guide its search through graph nodes. It leverages the evaluation function $f(n) = g(n)+h(n)$ to order nodes' exploration, where $g(n)$ signifies the actual cost from the start node to the node n and $h(n)$ is the estimated cost from node n to the target, based on the problem's heuristic information(Yao et al., 2010)[18].

Advantages: The A* algorithm, due to its heuristic nature, can avoid traversing the entire map or all nodes. By effectively choosing the evaluation function, the algorithm guides the search towards promising directions, thereby enhancing efficiency and often obtaining the optimal path(Zhen et al., 2023)[19].

Limitations: The efficiency and accuracy of the A* largely hinge on the selection of the heuristic function $h(n)$. If not chosen appropriately, the A* algorithm may not deliver the optimal path or could even end up being less efficient than Dijkstra's method. Furthermore, ensuring that $h(n)$ remains a lower bound of the actual cost to reach the goal from node n is crucial for the algorithm's optimality(Yao et al., 2010)[18].

2. Genetic Algorithm

Principle: Genetic Algorithms (GAs) emulate the natural selection process, treating paths or solutions as "genes". They use evolutionary operations like crossover and mutation to evolve towards optimal solutions (Mathew, 2012)[20].

Advantages: GAs are versatile, capable of addressing a wide variety of problems ranging from optimization to scheduling and layout challenges. Their global search capabilities make them well-suited for complex networks. Additionally, the inclusion of the crossover operation not only speeds up the search process but also aids in navigating out of local maxima, a prevalent issue with hill-climbing algorithms (Mathew, 2012)[20]. This is especially significant given that GAs generally face fewer challenges with local maxima compared to other algorithms like back-propagation neural networks.

Limitations: Despite their potential, GAs do come with challenges. They often require multiple iterations and may exhibit instability, particularly if reliant purely on mutation (Mathew, 2012)[20]. The efficacy can vary, producing outstanding results for some problems and underwhelming ones for others. When dealing with large-scale networks, GAs might not be the most efficient choice, especially if the crossover operation isn't effectively utilized (Srinivas and Patnaik, 1994)[21]. Furthermore, the correct selection of parameters and operators is crucial to the success of GAs, influencing their overall efficacy (Srinivas and Patnaik, 1994)[21].

3. Dijkstra Algorithm

Principle: The Dijkstra algorithm is fundamentally designed to find the shortest path between a starting point (source) and all other nodes in a weighted directed network. The weight on each edge might represent parameters like distance or travel cost between adjacent nodes (Liu et al., 1994)[22].

Advantages: Dijkstra's algorithm is renowned for its effectiveness, and it is the only method capable of working autonomously to address the shortest path challenge in weighted directed networks. It's a classic approach that has established its reliability over time (Liu et al., 1994)[22].

Limitations: Even though the Dijkstra algorithm is proficient, applying it directly, especially in applications like Embedded-GIS, can be problematic due to its high space and time complexities. Several current shortest path algorithms in Embedded-GIS base themselves on Dijkstra, yet they frequently lack in terms of efficiency and practicality. A significant limitation is that these algorithms often do not consider the characteristics of GIS data and the nature of roads when loading GIS data (Zhang et al., 2010)[23]. Furthermore, while the algorithm is effective, it can be computationally wasteful to scan an entire road network, especially when the network comprises thousands of routes and junctions. In many scenarios, leveraging commonsense or geographical knowledge can considerably speed up the process, negating the need to search the whole road network (Liu et al., 1994)[22].

3.3 Research Gaps

Current research on the impact of subway strikes on transportation networks suffers from a number of significant shortcomings.

First, the depth of literature focusing on passenger behavior and decision-making during subway strikes is insufficient. A deeper understanding of how riders adapt during these disruptions is critical to developing effective mitigation strategies.

Second, there is a distinct lack of comprehensive, real-time, data-driven tools specifically designed to predict and measure the impacts of subway strikes. This underscores the need for forward-looking solutions that emphasize preemptive monitoring and prediction, not just after-the-fact assessments.

Third, the current emergency response paradigm related to subway strikes tends to rely heavily on empirical strategies, often at the expense of empirical data-driven analysis. Adopting strategies based on real-time data and sophisticated traffic modeling could lead to more rapid and accurate interventions in subway strikes.

Finally, post-strike recovery efforts remain relatively lacking. Mitigating passenger inconvenience and restoring network efficiency requires insight into the optimal strategies and the time span required to restore normal subway function.

To bridge these gaps, this study suggests the use of weighted network analysis to enhance the resilience of urban transportation networks. This innovative approach not only aims to enhance current transportation management practices, but also adds credibility to future research exploring this area.

4 | Methodology

4.1 Resilience assessment of the London Underground

4.1.1 centrality metric

Centrality metrics play a crucial role in network analysis by highlighting the importance or influence of nodes within the network. In transport systems like the London Underground, identifying stations with higher centrality can offer insights into potential congestion points, key interchanges, or vulnerabilities in the system.

4.1.1.1 Degree Centrality

Definition Degree centrality is a fundamental measure in a network graph that quantifies the number of direct links a node has with other nodes(Scott, 1988)[24]. Mathematically, Freeman(2002) suggests that it represents the number of edges connected to a given node[25].

Significance

- Stations with higher degree centrality are likely interchange stations serving multiple lines, acting as pivotal nodes for passengers to switch between different routes.
- This measure directly correlates with the traffic volume of a station; more connections typically indicate more passenger footfall(Brandes, 2001)[26]. High degree centrality might also hint at potential bottlenecks during peak hours.

Mathematical Formulation of Degree Centrality The degree centrality of a node can be calculated using various formulations. The most commonly used equations include:

1.
$$C_d(k) = \sum_i^n a_{ij} \tag{4.1}$$

- Where:
- a_{ij} is an element of the adjacency matrix A, which is 1 if there's a link between nodes i and j and 0 otherwise.

2.
$$k_i = \sum_{j=1}^n A_{ij} \tag{4.2}$$

- This represents the original degree of nodes, counting the number of connections node i has in the network.

3.
$$C_j^d = \frac{k_i}{N-1} \tag{4.3}$$

- This normalized version of degree centrality takes into account the size of the network. Here, N is the total number of nodes.

Applications and Implications for the London Underground Utilizing degree centrality in the context of the London Underground can aid in:

- Identifying potential chokepoints which may require additional resources during peak travel times.
- Informing decisions related to expansion or modification of routes, ensuring efficient passenger dispersal.
- Highlighting key interchange stations which might benefit from improved facilities or additional services due to their importance in the network.

4.1.1.2 Closeness Centrality

Definition Closeness centrality quantifies a node's average shortest path length to all other nodes in the network(Zhang and Luo, 2017)[27]. Essentially, it measures the average "distance" from a particular node to every other node in the system.

Significance

- Stations with high closeness centrality values represent key connectivity points, ensuring faster and more efficient routes for passengers.
- This metric can also highlight potential weak points in a network, where disruptions might have significant repercussions.
- In transportation systems, such as the London Underground, it can indicate stations with higher accessibility, offering alternative routes and thus, resilience to potential disruptions.

Mathematical Formulation of Closeness Centrality Various formulations represent closeness centrality. Commonly used equations include:

1.
$$C_c(k) = \frac{n-1}{\sum_{i \neq j} d_{ij}} \tag{4.4}$$

   - Where d_{ij} represents the shortest path between nodes i and j, and n is the total number of nodes.

2.
$$l_i = \frac{1}{n} \sum_j d_{ij} \tag{4.5}$$

   - l_i is the average geodesic distance from one node to all others.

3.

$$C_i = \frac{1}{l_i} = \frac{n}{\sum_j d_{ij}} \tag{4.6}$$

- A normalized version of closeness centrality, emphasizing the inverse relationship of the average shortest path length (Evans and Chen, 2022)[28].

Applications and Implications for the London Underground By using closeness centrality:

- Transport authorities can prioritize stations requiring infrastructure upgrades or modifications to improve accessibility.
- Service schedules might be adjusted to optimize traffic flow through stations with high closeness centrality, ensuring smoother transitions for passengers.
- Potential vulnerabilities in the network can be identified and fortified, promoting resilience against disruptions.

4.1.1.3 Betweenness Centrality

Definition Betweenness centrality measures the extent to which a node (or vertex) lies on the paths between other nodes in a network. It reflects a node's influence over the spread of information or flow within the network(Newman, 2005)[29].

Significance

- According to Brandes(2001), nodes with high betweenness centrality can act as gatekeepers, controlling the flow of information, making them pivotal in determining how information or resources spread in the network[26].
- From the example given by Barthelemy(2004), the removal of nodes with high betweenness can cause significant disruptions to the network[30]. These nodes often

serve as critical junctions or bridges in the network, connecting various regions or clusters.

- Nodes with high betweenness, even with low degree centrality, might play a critical role in connecting disparate parts of a network(Barthelemy, 2004)[30].

Mathematical Formulation of Betweenness Centrality The general formula for betweenness centrality is given by:

$$C_b(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{4.7}$$

Where:

- σ_{st} is the total number of shortest paths from node s to node t.
- $\sigma_{st}(v)$ is the number of those paths that pass through node v.

Applications and Implications for the London Underground

- Stations with high betweenness centrality act as critical transit points for passengers. Their disruption can have cascading effects, leading to delays in multiple lines and routes.
- These stations might need higher maintenance attention and might be considered as critical points for emergency evacuation plans.
- While stations with high footfall (degree centrality) are essential for logistical reasons, stations with high betweenness centrality are vital for the smooth functioning and robustness of the entire underground network.

4.1.2 Node Removal Impact

In assessing the resilience of the London Underground system, it is vital to recognize the consequences of station closures, particularly in the event of a potential strike. This study simulates the impact of disrupting key nodes (stations) within the network to understand the knock-on effects that such decisions may have on the system as a whole. This assessment focuses on the impacts of stations, which are an important part of our study, and we aim to calculate the impacts of stations to be applied to the subsequent weighted network.

The Impact from Station Elimination (R) is calculated as

$$R = \frac{\Sigma_{(u,v)\in E} F_{uv} - \Sigma_{(a,b)\in E'} F'_{ab}}{Sigma_{(u,v)\in E} F_{uv}}$$

Where R is the effect from elimination, E denotes the edges of the original network, F_{uv} denotes the flow of a given edge, E' is the set of edges after the node elimination, and F'_{ab} denotes the flow of edges that change the network after elimination of a predetermined node.

The R provides the effect of node elimination on the overall network traffic. The metric calculates the percentage decrease in overall transit traffic after the removal of a site from the initial system. The formula compares the total traffic of the original network to the total traffic of the modified network (after cancellation), emphasizing the difference between the two.

4.1.3 Calculating the Resilience Factor

The resilience of each station in the London Underground network can be quantified using a measure which terms as ‘Resilience Factor‘. This factor is a weighted combination of various centrality measures and the results from simulated node removal.

4.1.3.1 Centrality Measures

The centrality measures considered are:

- Degree Centrality (C_d)
- Closeness Centrality (C_c)
- Betweenness Centrality (C_b)
- Results from Simulated Node Removal (R)

4.1.3.2 Weight Assignments

Weights are assigned to each of the above measures based on their perceived importance in determining network resilience. Let the weights be represented as w_d, w_c, w_b, and w_r respectively. For the purpose of this analysis, and as a starting point, equal weights are assigned i.e., $w_d = w_c = w_b = w_r = 0.25$.

4.1.3.3 Resilience_Factor Computation

The 'Resilience Factor' for a station i can then be computed as:

$$RF_i = w_d \times C_{d,i} + w_c \times C_{c,i} + w_b \times C_{b,i} + w_r \times R_i \tag{4.8}$$

This factor provides a quantifiable measure of the resilience of each station in the network, taking into consideration its topological importance and the impact of its removal on the overall network functionality.

4.2 Resilience Evaluation Considering Flow

To achieve a comprehensive understanding of the intrinsic vulnerabilities within metro stations, it is imperative to incorporate centrality measures combined with flow data. Such an amalgamation provides insights into the structural and functional significance of stations within the network.

4.2.1 Flow-adjusted Degree Centrality

By integrating passenger flow with connectivity, it can derive the following degree centrality:

$$C_d = \frac{\text{Peak weighted degree}}{\sum(\text{adjacent nodes} \times \text{flow factor})} \tag{4.9}$$

4.2.2 Flow-adjusted Closeness Centrality Redefined

Flow-adjusted closeness centrality helps in identifying pivotal stations within the network:

$$C_c = \frac{1}{\sum \text{swift paths considering inverse flows}} \tag{4.10}$$

4.2.3 Flow-adjusted Betweenness Centrality Redefined

Incorporating passenger flow, it can derive the following betweenness centrality:

$$C_b = \frac{\text{Aggregate shortest paths}}{\text{Paths through node}} \tag{4.11}$$

4.2.4 Computation of Vulnerability Factor

Utilizing these adjusted centrality measures, the Vulnerability Factor (VF) for a station can be computed as:

$$VF = w_d \times C_d + w_c \times C_c + w_b \times C_b + w_r \times (1 - R) \tag{4.12}$$

Where:

- C_d, C_c, and C_b are the flow-adjusted degree centrality, closeness centrality, and betweenness centrality, respectively.
- R represents the resilience factor of the station.In contrast to Resilience_Factor, here is the use of the (1 - R) is used because when R has a higher value, vulnerability is actually lower.
- w_d, w_c, w_b, and w_r are weight coefficients, reflecting the relative importance of each centrality measure and resilience in the overall vulnerability computation.

For the purpose of this analysis, and as a starting point, equal weights are assigned i.e., $w_d = w_c = w_b = w_r = 0.25$.

4.3 Intelligent Path Optimization for Planning Applications

In emergency situations such as subway strikes or other emergencies, it becomes crucial to provide stable, fast and safe transportation proposals for passengers. To this end, the study proposes to develop an intelligent path optimization system. The system aims to combine network resilience assessment with passenger flow data to provide transportation planners with more scientific decision support.

Path optimization will go beyond mere distance or time metrics. Instead, it will combine network resilience with actual passenger flows at each station. By assigning weights to each station or connection, planners can factor in the actual conditions and potential risks of the subway system.

4.3.1 Foundation:A* Algorithm for Path Optimization

Principle: A* algorithm is a heuristic-based method that seeks to find the shortest path from a starting point to a destination by evaluating the cost of paths(Chunyu et al., 2020)[31]. The cost function in the A* algorithm combines both the actual cost from the starting point to the current node and an estimated cost from the current node to the destination. This heuristic nature of the algorithm helps in making informed decisions, thereby reducing unnecessary explorations and enhancing the search efficiency.

Operational Steps:

1. Define the cost function for each node n:

$$f(n) = g(n) + h(n) \tag{4.13}$$

   where:

   - $f(n)$ represents the total cost function.
   - $g(n)$ denotes the actual path cost from the starting point to the current node n.
   - $h(n)$ is the heuristic estimated cost from node n to the destination. This is typically calculated using the Euclidean distance when the actual spatial coordinates of the nodes are known.

2. Calculate $h(n)$ using the Euclidean distance formula:

$$h(n) = \sqrt{(x_n - x_{\text{goal}})^2 + (y_n - y_{\text{goal}})^2} \tag{4.14}$$

   where x_n and y_n are the coordinates of the current node, and x_{goal} and y_{goal} are the coordinates of the destination.

3. Begin at the starting node. Add it to a list of nodes to be explored.

4. For each node, calculate $f(n)$ and select the node with the lowest value of $f(n)$.

5. Expand the chosen node by considering its neighbors. Calculate their $f(n)$ values and add them to the list of nodes to be explored. Mark the current node as explored.

6. Repeat the process until the destination node is explored or the list of nodes to be explored is empty.

7. Trace back from the destination node to the start to get the optimal path.

8. If the list of nodes to be explored is empty and the destination has not been reached, no path exists.

4.3.2 Advanced Weight Computation

Incorporating the insights from the flow-based resilience and vulnerability evaluations, it can enhance the traditional distance-based approach by introducing a composite weight for each station, taking inspiration from the concepts of weighted network(Barrat et al., 2004)[32]. For a given station i, its weight w_i can be computed as:

$$w_i = \alpha \times RF_i + \beta \times VF + \gamma \times D_i \tag{4.15}$$

where,

- $RF_i = w_d \times C_{d,i} + w_c \times C_{c,i} + w_b \times C_{b,i} + w_r \times R_i$ is the resilience factor of the station i, which integrates various centrality measures and resilience scores.
- $VF = w_d \times C_d + w_c \times C_c + w_b \times C_b + w_r \times (1 - R)$ is the vulnerability factor that provides an inverse measure of a station's robustness.
- D_i represents the traditional distance or time score for the station i.
- α, β, and γ are weight coefficients that determine the relative importance of the resilience factor, vulnerability factor, and distance/time score, respectively.

In order to establish the basis of the weight distribution among the stations, I determined initial values for the coefficients α, β, and γ of 0.1, 0.1, and 0.8, respectively, which I chose to make sure that the introduction of a brand new variable would not drastically skew the results from the expected. My aim is to provide a balanced and informed starting point for the weight distribution, ensuring that it not only makes the

4.4 Data Sources

The primary dataset used in this project is secondary data provided from school programs. Although the original source of this dataset is not explicitly stated, it has been widely adopted and recognized by the academic community. The dataset provided by the school has been rigorously screened and validated to ensure its accuracy and reliability.

The database contains detailed information about each site, such as the name of the site, traffic data for each site, and information about the distance between sites. This dataset provided me with a comprehensive perspective that allowed me to gain insight and analyze the resilience of the city's subway system.

In order to maintain the integrity and objectivity of my analysis, I further cleaned and preprocessed the data to meet the specific requirements of this study. However, due to the lack of clarity in the initial data sources, there may be certain limitations that will be taken into account when interpreting and applying the results.

4.5 Technologies and Tools

In this project, the primary technologies the research employed include Docker and Python.

- **Docker**: Docker is an open-source application container engine, streamlining the process of development, deployment, and running applications. By packaging an application and its dependencies into a standardized unit called a container, Docker ensures consistent software performance across diverse environments(Rad et al., 2017)[33]. Although container technology has been around for years, Docker introduced innovations that garnered widespread attention and use. It offers a significant degree of isolation and resource limitation for its containers, proving effective even in default configurations(Bui, 2015)[34]. Further security enhancements are recommended, such as deploying AppArmor or SELinux.
- **Python**: Python, a versatile high-level programming language, is cherished for its concise and readable syntax. For this project, Python stands out due to its powerful data processing and numerical computation libraries like Pandas and NumPy. These libraries, coupled with the robust database tools and vast Python ecosystem, have substantially simplified the workflow for data handling, analysis, and visualization, ensuring seamless and efficient project execution(Python, 2021)(Stančin and Jović, 2019)[35, 36].

By synergizing these technologies and tools, the project achieved both efficiency and accuracy, meeting the high standards required for the research.

5 | Results

5.1 Introduction to the Results

The primary research question this study seeks to answer is: How can the integration of metro congestion levels and network vulnerability from potential subway station removals inform a weighted network model to enhance the resilience of urban transport networks and propose alternative routes?

To investigate this, it employed a methodological framework encompassing:

1. A comprehensive assessment of network resilience through centrality metrics.
2. An in-depth analysis of the impact of node removal on the overall network integrity.
3. The computation of a unique resilience factor based on centrality measures and associated weight assignments.
4. An exploration of intelligent path optimization techniques, with a specific focus on Dijkstraâs algorithm and the advanced computation of weights for optimal route determination.

In the following results section, the research will delve into the findings of each aforementioned aspect. This will offer a detailed understanding of the networkâs resilience, its vulnerable components, and strategies for optimal route planning during disruptions.

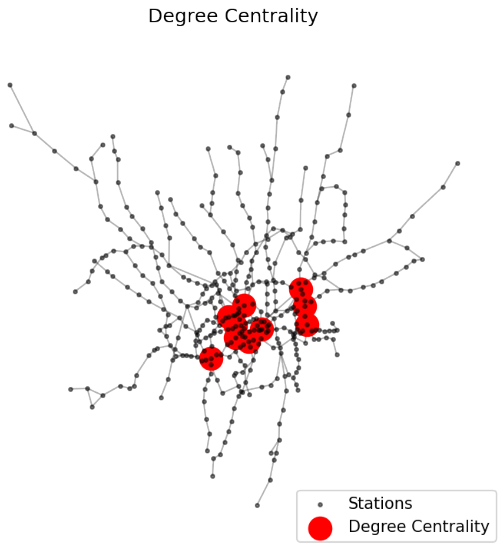
5.2 Resilience Assessment of the London Underground

5.2.1 Centrality Metric

5.2.1.1 Degree Centrality

Station Name	Degree Centrality
Stratford	0.0225
Bank and Monument	0.02
Baker Street	0.0175
King's Cross St. Pancras	0.0175
Green Park	0.015
Canning Town	0.015
Earl's Court	0.015
West Ham	0.015
Waterloo	0.015
Oxford Circus	0.015

(a) The First 10 Ranked Nodes



(b) Spatial Distribution

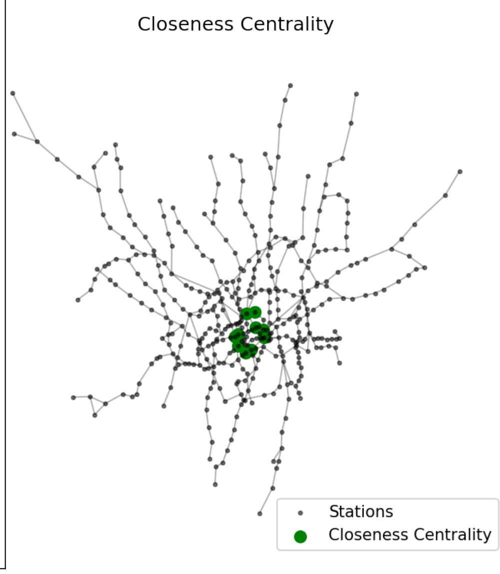
Figure 5.1: Degree Centrality for 10 Stations

An in-depth study of the degree centrality of the London Underground reveals that there are several key nodes that play a pivotal role in the transportation matrix. Stratford station tops the list with a degree centrality value of 0.0225, highlighting its inherent importance in the network. In addition, Stratford Station is closely followed by Bank, Monument, and Baker Street stations, highlighting the interconnectedness of these nodes. The visualization of the spatial distribution helps present a clear geographic pulse. The presence of these well-connected stations may suggest potential weaknesses or priorities for strategic improvements.

5.2.1.2 Closeness Centrality

Station Name	Closeness Centrality
Green Park	0.1148
Bank and Monument	0.1136
King's Cross St. Pancras	0.1134
Westminster	0.1125
Waterloo	0.1123
Oxford Circus	0.1112
Bond Street	0.111
Farringdon	0.1107
Angel	0.1107
Moorgate	0.1103

(a) The First 10 Ranked Nodes



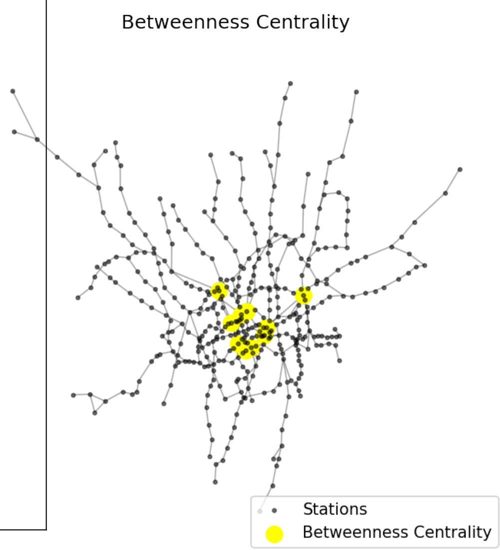
(b) Spatial Distribution

Figure 5.2: Closeness Centrality for 10 Stations

Delving deeper into the network efficiency of the London Underground, the proximity center metric depicts the accessibility of each station relative to the others. The leader of the pack is Green Park with a score of 0.1148, the station with the best accessibility. Notably, Bank and Monument, King's Cross St Pancras and Westminster stations also boast impressive proximity scores, emphasizing their role as central hubs. These stations play a vital role in ensuring passengers make timely and efficient interchanges, potentially reducing journey times across the network. The accompanying spatial distribution map shows the location of these hub stations, suggesting areas of potential congestion or highlighting areas of efficient transportation.

Station Name	Betweenness Centrality
Stratford	0.2978
Bank and Monument	0.2905
Liverpool Street	0.2708
King's Cross St. Pancras	0.2553
Waterloo	0.2439
Green Park	0.2158
Euston	0.2083
Westminster	0.2033
Baker Street	0.1916
Finchley Road	0.1651

(a) The First 10 Ranked Nodes for Betweenness



(b) Spatial Distribution

Figure 5.3: Betweenness Centrality for 10 Stations

5.2.1.3 Betweenness Centrality

In assessing the role of individual stations as passages or bridges in the network, the "zone centrality" metric brings out fascinating findings. Stratford appears again, this time with a value of 0.2978, emphasizing its role as a major conduit or bridge between different parts of the network. Neighboring stations such as Bank and Monument, Liverpool Street and King's Cross St Pancras also play an integral role as interchange nodes. This narrative highlights the profound impact these stations have on passenger flows and the potential bottlenecks or pressure points they may represent, particularly during peak periods or traffic disruptions. The spatial visualization shows the strategic location of these nodes, illustrating the backbone of the London Underground system.

5.2.2 Node Removal Impact

The concept of 'impact reduction' quantifies the impact of extracting a specific node on the cumulative traffic within the network. This metric essentially describes the proportional decrease in total network traffic after a node has been removed. The basic equation shows the difference between the cumulative traffic of the original network and the cumulative traffic of the modified version of the network after the nodes have been removed. Evaluating this difference is crucial because it emphasizes the importance of individual nodes in maintaining optimal traffic.

Using this metric, our findings reveal significant differences in the ability of networks to adapt to node removal. For example, removing "Banks and Monuments" leads to a 7.14 % reduction in total traffic. This highlights its critical role in ensuring network traffic mobility. Similarly, nodes such as 'Greenfield Park' and 'Waterloo' also had a significant impact with reductions of 6.43 % and 5.82 % respectively. These figures highlight the importance of these sites in maintaining network integrity and performance.

Table 5.1: Impact of Stations on the Network

Rank	Station	Impact
1	Bank and Monument	7.14%
2	Green Park	6.43%
3	Waterloo	5.82%
4	King's Cross St. Pancras	4.82%
5	Westminster	4.35%
6	Liverpool Street	4.25%
7	Stratford	3.71%
8	Euston	3.66%
9	Baker Street	3.21%
10	Oxford Circus	3.05%

5.2.3 Calculating the Resilience_Factor

For the analysis, it employed a uniform weighting strategy across the centrality measures and the reduction impact. Specifically, a weight of $w = 0.25$ was chosen for each factor: $C_{d,i}$, $C_{c,i}$, $C_{b,i}$, and R_i. This equal weighting approach ensures that no single factor dominates the computation of RF_i, reflecting an assumption that each measure contributes equally to a station's significance within the network.

To ascertain the significance of each station in the network, the research employed the RF_i measurement, defined as:

$$RF_i = 0.25 \times C_{d,i} + 0.25 \times C_{c,i} + 0.25 \times C_{b,i} + 0.25 \times R_i$$

Upon applying this formula across stations, it tabulated the results, revealing a comprehensive insight into the relative importance of each station within the network.

Table 5.2: Computed RF_i for Each Station

Station Name	RF_i Value
Bank and Monument	0.123865
Stratford	0.115457
Liverpool Street	0.109640
King's Cross St. Pancras	0.108613
Waterloo	0.107346
Green Park	0.102478
Westminster	0.092346
Euston	0.091814
Baker Street	0.087532
Oxford Circus	0.052637

Upon analyzing the RF_i values, the "Bank and Monument" station emerged as the most pivotal within the network, boasting an RF_i value of 0.123865. This was closely trailed by "Stratford" with a score of 0.115457. A subsequent cluster of stationsâcomprising "Liverpool Street", "King's Cross St. Pancras", and "Waterloo"âshowed comparable prominence, with RF_i values ranging between 0.10 and 0.11. Conversely, "Oxford Circus" was identified as having a markedly lesser impact with an RF_i value of 0.052637, roughly half in comparison to the leading stations. This data accentuates the indispensable role specific stations hold in the network's structure, providing vital insights for infrastructure development and emergency response strategies.

5.3 Resilience Evaluation Considering Flow

5.3.1 Flow-adjusted Degree Centrality

Table 5.3: The First 10 Ranked Nodes for Flow-adjusted Degree Centrality

Station Name	Flow-adjusted Degree Centrality
Bank and Monument	1.0
Green Park	0.9001
Waterloo	0.8151
King's Cross St. Pancras	0.6749
Westminster	0.6096
Liverpool Street	0.5953
Stratford	0.5201
Euston	0.513
Baker Street	0.4499
Oxford Circus	0.4277

The Degree Centrality results previously pinpointed Stratford as the paramount node in

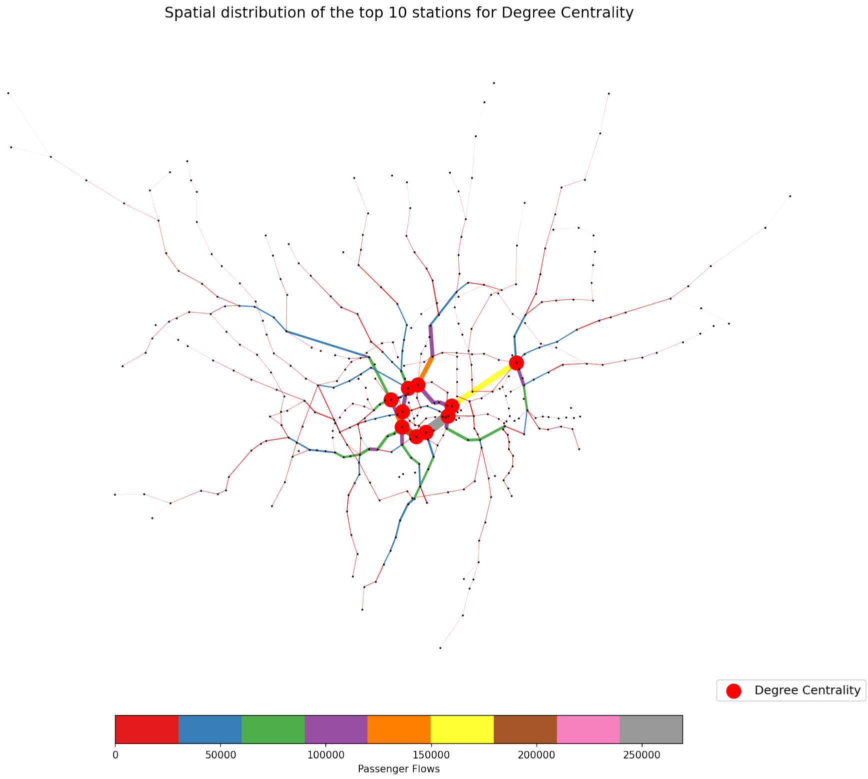


Figure 5.4: Spatial Distribution for Flow-adjusted Degree Centrality

terms of sheer connectivity, with a centrality value of 0.0225. However, when adjusting for flow dynamics in the network, the landscape shifts noticeably. Bank and Monument took the crown in the Flow-adjusted Degree Centrality results, emphasizing its importance in handling actual commuter traffic. Green Park's surge to the second position, in contrast to its placement in the unadjusted measurements, is also particularly telling. It suggests that while some stations might be key in terms of simple connections, their roles can differ vastly when real-world flow is considered. This distinction between theoretical and practical centrality reinforces the importance of understanding the actual utilization of these stations for effective transport planning.

Table 5.4: The First 10 Ranked Nodes for Flow-adjusted Closeness Centrality

Station Name	Flow-adjusted Closeness Centrality
Green Park	0.0001
Westminster	0.0001
Bank and Monument	0.0001
Oxford Circus	0.0001
Liverpool Street	0.0001
Bond Street	0.0001
Warren Street	0.0001
Hyde Park Corner	0.0001
Moorgate	0.0001

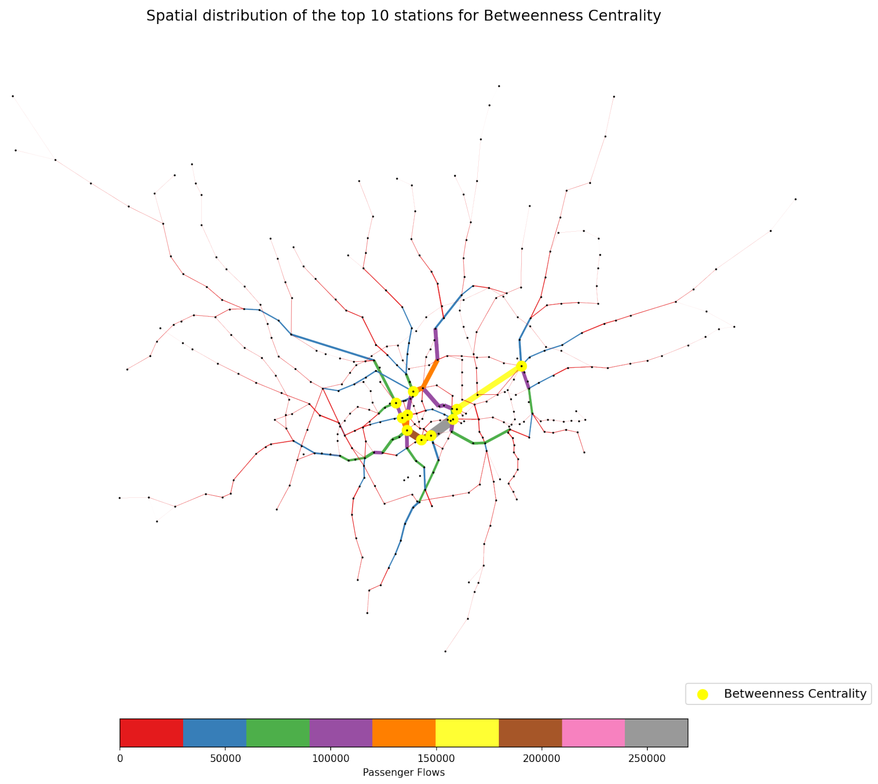


Figure 5.5: Spatial Distribution for Flow-adjusted Closeness Centrality

5.3.2 Flow-adjusted Closeness Centrality

Delving into the Flow-adjusted Closeness Centrality, the research witnesses Green Park retaining its pivotal status, aligning with its position in the Flow-adjusted Degree Centrality. Yet, what's intriguing is the pronounced centrality of Westminster and Bank and Monument, both registering a closeness value of 0.0001. Such findings indicate their central position not only in terms of traffic flow but also in enabling efficient and swift movement throughout the network. This duality in their importanceâboth in handling traffic and facilitating connectivityâunderscores their role as key transportation hubs.

5.3.3 Flow-adjusted Betweenness Centrality

Table 5.5: The First 10 Ranked Nodes for Flow-adjusted Betweenness Centrality

Station Name	Flow-adjusted Betweenness Centrality
Green Park	0.5492
Bank and Monument	0.5267
Waterloo	0.4256
Westminster	0.3743
Liverpool Street	0.3441
Stratford	0.3375
Euston	0.2722
Oxford Circus	0.2472
Bond Street	0.2447
Baker Street	0.2404

The initial Betweenness Centrality score positioned Stratford as the most important transit point in the network, with a score of 0.2978, bridging many of the shortest paths across the system. However, the introduction of flow adjustment changes this pattern. Greenbelt Park is a distant second in the flow-adjusted centrality metric, indicating its

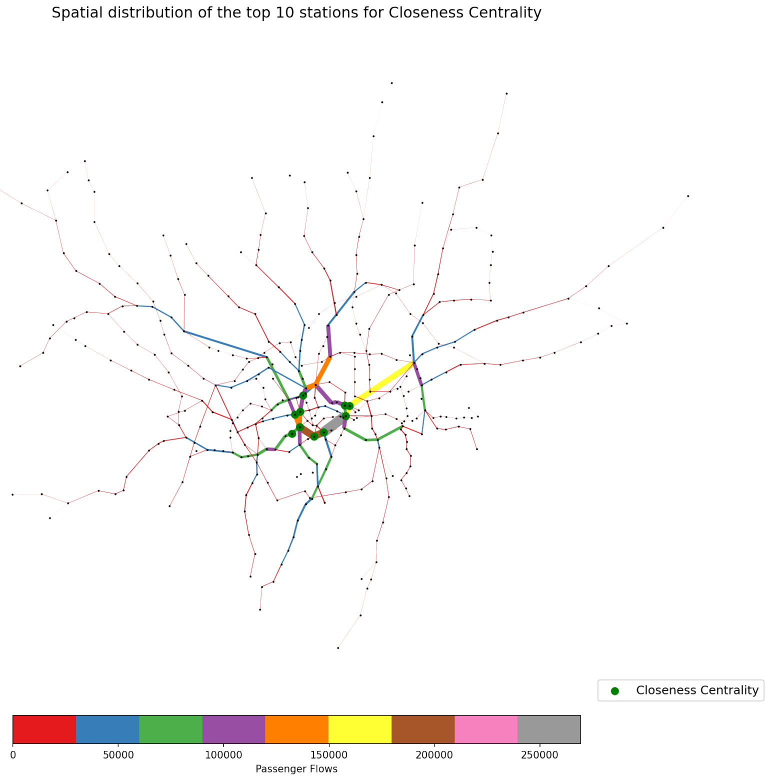


Figure 5.6: Spatial Distribution for Flow-adjusted Betweenness Centrality

importance in real-world traffic scenarios. It is a focal point for the convergence and dispersion of numerous paths, especially when passenger traffic is considered. The dominance of Bank, Monument, and Waterloo in the passenger flow-adjusted scenario further highlights the difference between actual traffic patterns and mere structural connectivity.

5.3.4 Computation of Vulnerability Factor

To compute the vulnerability factor VF_i for each station, it employed weighted centrality measures, including weighted degree centrality, weighted betweenness centrality, and weighted closeness centrality. These were combined with considerations for diminished impacts to holistically assess the vulnerability. Here, "weight" refers to the traffic between stations, i.e., the number of passengers moving from one station to another. A higher

VF_i value indicates a higher vulnerability of the station within the London Underground network.

Below are the top 10 stations sorted by their VF_i values:

Station	**VF_i**
Bank and Monument	0.613848
Green Park	0.596268
Waterloo	0.545670
Westminster	0.485131
Liverpool Street	0.474240
King's Cross St. Pancras	0.462157
Stratford	0.455149
Euston	0.437174
Baker Street	0.414562
Oxford Circus	0.411124

As can be seen from the results, the "Bank and Monument" station has the highest vulnerability factor. This suggests that it may be the most vulnerable station to disruption on the London Underground. The prominent location of this station can be attributed to several factors:

1. Interconnectivity: Bank and Monument station is a connection point for a number of Tube lines and is therefore an important interchange station for passengers.
2. Location: The downtown location is a gathering point for daily commuters and tourists, and therefore has a high volume of passenger traffic throughout the day.
3. Economic Impact:mDue to its proximity to the financial district, any disruption would have a significant economic impact by causing delays for professionals working in the neighborhood.

5.4 Intelligent Path Optimization for Planning Applications

In this analysis, the research evaluates two primary metrics for assessing the significance and impact of stations in the London Underground network: Risk Factor (RF) and Value Factor (VF).

5.5 Metric Definitions

5.5.1 Risk Factor (RF)

The Risk Factor (RF) is derived from four distinct indicators: *degree_ centrality*, *closeness_ centrality*, *betweenness_ centrality*, and *impact*. Notably, *impact* is a measure that indicates the repercussions of a station's removal on the overall network, resembling scenarios like strikes. Thus, a station with a high *impact* value plays a pivotal role in the network, where its failure could lead to significant disruptions.

5.5.2 Value Factor (VF)

The Value Factor (VF) comprises weighted versions of *degree_ centrality*, *closeness_ centrality*, and *betweenness_ centrality*. Additionally, it factors in $(1 - impact)$, indicating that while it has elements opposing RF, it's not entirely inverse. Essentially, VF gauges a station's overall significance or value, taking into account more than its mere centrality.

5.6 Test Scenarios

To evaluate the performance of the refined model, several test scenarios are utilizedï¼

5.6.1 Adjacent Stations

A baseline test to verify the algorithm's capability to identify the shortest and most direct route.

Example: From "Holborn" to "Russell Square".

Table 5.6: Comparison of Paths: Standard A* vs Weighted A* for Holborn to Russell Square

Original Astar	Weighted Astar
Holborn → Russell Square	Holborn → Russell Square

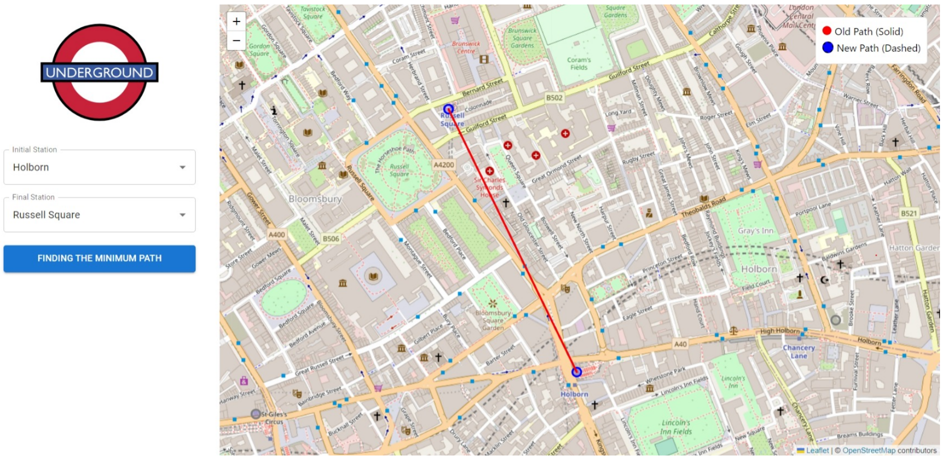


Figure 5.7: Path Visualization between Holborn and Russell Square

The data presented is a basic validation of the functionality of the algorithm, especially when the start and end points are directly adjacent to each other.

When analyzing the trip between Holborn and Russell Square, it is evident that the weighted version of the A* algorithm produces exactly the same results as the traditional A* method. This means that for direct routes between neighboring stations, the weights introduced have little or no effect on the path determined.

Given that the nature of the A* algorithm is to seek the most efficient path, the most direct path is already the optimal path for directly adjacent stations. Therefore, in this case, the introduction of additional weights does not change the path selection.

This test fundamentally emphasizes an important point: the weighted version of the algorithm fundamentally follows the principles of the traditional A* algorithm. By confirming the consistency of the weighted algorithm in determining the most direct routes for neighboring sites, it verified that the heart of the weighted algorithm is the A* routing logic. Therefore, any enhancements or modifications introduced by weighting are based on a solid and reliable foundation.

5.6.2 Cross-city Long-distance Stations

To evaluate the algorithm's global path-finding ability and its adeptness in locating the optimal path as per weights.

Example: From "Heathrow Terminal 4" to "Stratford".

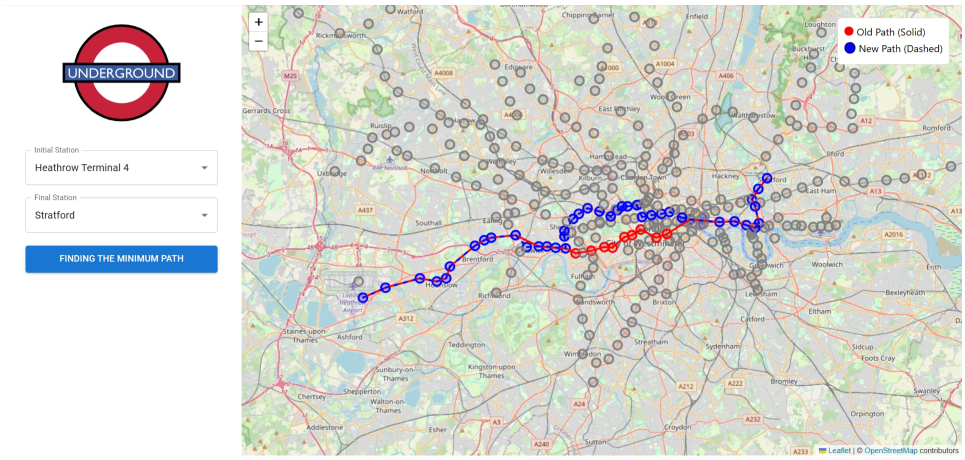


Figure 5.8: Path Visualization between Heathrow Terminal 4 and Stratford

The impact of applying weights in the A* algorithm becomes apparent in the extended route example from "Heathrow Terminal 5" to "Stratford". Whereas the basic A* algorithm usually determines the most direct route, our weighted version adjusts dynamically according to the priority set by the weights. The resulting path is sometimes longer, but

Table 5.7: Comparison of Paths: Standard A* vs Weighted A* for Heathrow Terminal 4 to Stratford

Standard A* Path	**Weighted A* Path**
Heathrow Terminal 4 → Hatton Cross → Hounslow West → Hounslow Central → Hounslow East → Osterley → Boston Manor → Northfields → South Ealing → Acton Town → Turnham Green → Hammersmith → Barons Court → Earl's Court → Gloucester Road → South Kensington → Knightsbridge → Hyde Park Corner → Green Park → Westminster → Waterloo → Bank → Shadwell → Limehouse → Westferry → Poplar → All Saints → Devons Road → Bow Church → Pudding Mill Lane → Stratford	Heathrow Terminal 4 → Hatton Cross → Hounslow West → Hounslow Central → Hounslow East → Osterley → Boston Manor → Northfields → South Ealing → Acton Town → Chiswick Park → Turnham Green → Stamford Brook → Ravenscourt Park → Hammersmith → Goldhawk Road → Shepherd's Bush (H) → Latimer Road → Ladbroke Grove → Westbourne Park → Royal Oak → Paddington → Edgware Road (B) → Marylebone → Baker Street → Regent's Park → Oxford Circus → Tottenham Court Road → Holborn → Chancery Lane → St. Paul's → Bank → Shadwell → Limehouse → Westferry → Poplar → All Saints → Devons Road → Bow Church → Pudding Mill Lane → Stratford

it meticulously avoids sections of the network that might be considered undesirable based on set criteria - whether due to congestion, maintenance work, or other factors.

The comparison of the paths selected in the standard and weighted A* versions demonstrates how weighting can significantly alter path selection. The weighted route, although potentially longer, provides a tailored journey experience that considers more than just the shortest route. This personalization is invaluable to commuters who may seek a more comfortable, less congested, or simply different commuting experience based on their unique preferences.

5.6.3 High RF and VF to Low RF and VF

This scenario tests if the weights can deter routes from high RF (crowded) and high VF (indicating high value or significance) stations in favor of those with low RF and VF.

Example: From "Baker Street" to "Convent Garden".

Table 5.8: Comparison of Paths: Standard A* vs Weighted A* for Baker Street to Covent Garden

Standard A* Path	**Weighted A* Path**
Baker Street → Bond Street → Oxford Circus → Tottenham Court Road → Leicester Square → Covent Garden	Baker Street → Regent's Park → Oxford Circus → Tottenham Court Road → Leicester Square → Covent Garden

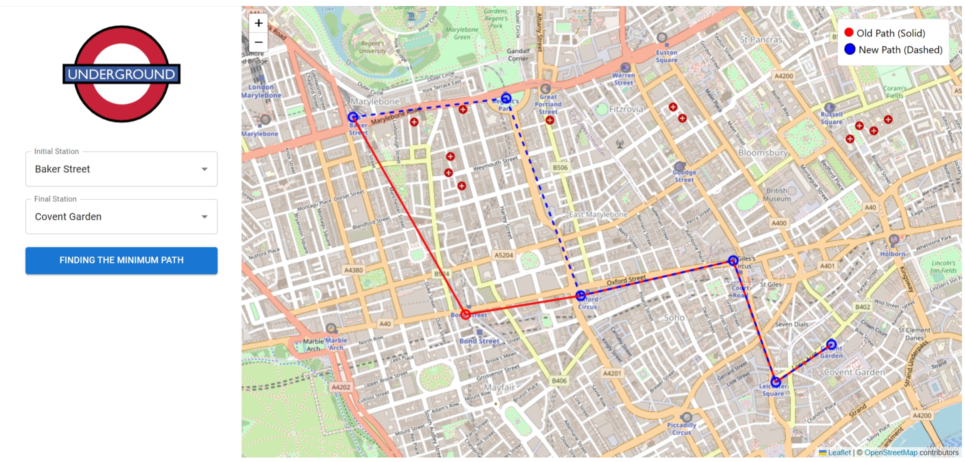


Figure 5.9: Path Visualization between Baker Street and Convent Garden

This test emphasizes the effect of station importance and congestion on route choice. Starting from "Baker Street" and ending at "Covent Garden", the traditional A* algorithm's choice is more straightforward, with the route passing through "Bond Street", which is one of the top ten stations in terms of RF (crowd frequency) and VF (value or importance).

However, the algorithm shows an interesting bias when weighting factors are taken into account. It bypassed "Bond Street" altogether in favor of "Regent's Park" - a station that ranked much lower on both the RF and VF metrics, around the 300th percentile. This rerouting demonstrates the effectiveness of weighting algorithms to proactively bypass major hubs or crowded stations in path selection. This decision was likely influenced by the goals of reducing exposure to high-traffic or critical stations and providing a less congested and potentially more efficient route, especially during peak hours or events.

5.6.4 Low RF and High VF to High RF and Low VF

This assesses the balance between the metrics, probing if the algorithm leans towards either VF or RF.

Example: From "Oval" to "Old Street".

Table 5.9: Comparison of Paths: Standard A* vs Weighted A* for Oval to Old Street

Standard A* Path	Weighted A* Path
Oval → Kennington → Waterloo → Bank → Moorgate → Old Street	Oval → Kennington → Elephant & Castle → Borough → London Bridge → Bank → Moorgate → Old Street

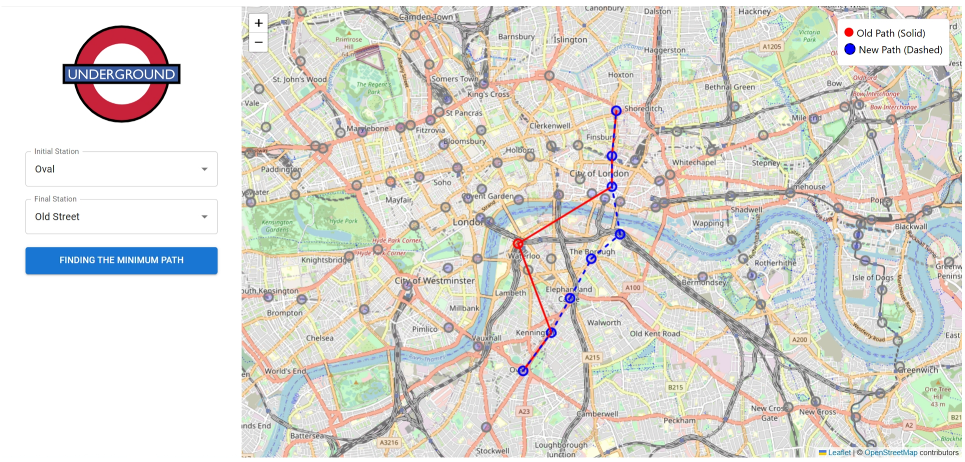


Figure 5.10: Path Visualization between Oval and Old Street

From "Oval" to "Old Street," the path options show a clear conflict between the station's high VF (value factor) and high RF (crowding).

Starting at the "Oval" station, it has a low RF (less crowded) and a high VF (indicating its value or importance).

It passes through major hub stations such as "Waterloo", which is known for its high RF due to its high foot traffic. The algorithm prioritizes directness and connectivity without taking into account the station's congestion. The terminus is the "Old Street" station, which has a high RF (indicating congestion) and a low VF.

Starting at 'Oval' station, the route detours around 'Waterloo' station and passes through 'Elephant Castle' station, 'Borough' station and 'London Bridge' station before ending at 'Bank' station. This longer route shows that the algorithm consciously avoids heavily congested areas. The final arrival at "Old Street" maintains the final objective despite the higher RF.

The deviation of the weighted algorithm's path around "Waterloo" emphasizes its attempt to balance VF and RF. By selecting stations with moderate or even low VFs but also low RFs, the algorithm prioritizes passenger comfort and ensures less congestion while reaching important destinations.

In short, while the standard path passes through high-value but crowded hubs like Waterloo, the weighted algorithm aims to balance the journey. It starts at an important but less crowded station, avoids overcrowded hubs, and then arrives at a less important but crowded destination. This balanced approach ensures that passengers experience value and comfort during the transfer.

5.6.5 Via Major Interchange Stations

This examines if the algorithm gravitates towards major transportation hubs or circumvents them.
Example: From "Victoria" to "Liverpool Street", observing if the route traverses "Waterloo" or other significant interchange stations.

In this evaluation, the focus was on determining how the algorithm, with the added weighting, would work through or around major transportation hubs, especially stations like "Waterloo" that are known for their high passenger volumes and connectivity.

Table 5.10: Comparison of Paths: Standard A* vs Weighted A* for Victoria to Liverpool Street

Standard A* Path	**Weighted A* Path**
Victoria → St. James's Park → Westminster → Waterloo → Bank → Liverpool Street	Victoria → St. James's Park → Westminster → Embankment → Temple → Blackfriars → Mansion House → Cannon Street → Monument → Tower Hill → Aldgate → Liverpool Street

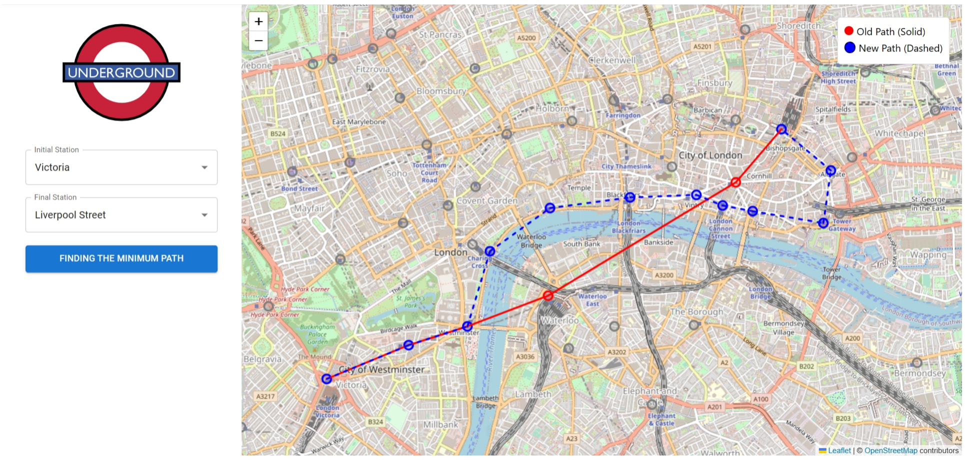


Figure 5.11: Path Visualization between Victoria and Liverpool Street

The traditional A* algorithm, without any weighting, would naturally choose the shortest route directly through "Waterloo" station. This is an expected behavior as it tends to prefer the shortest path based on geographical distance and direct connectivity.

However, the algorithm's behavior changes significantly in the case of weighted considerations. Instead of choosing to go through the seemingly direct route via "Waterloo", the weighted A* algorithm bypasses this major hub and goes through other stations such as "Embankment", "Temple" and "Blackfriars". This deviation from the direct route is very serious. The algorithm's tendency to bypass the "Waterloo" station, given its high visitor frequency (VF), which is among the top ten stations, demonstrates the algorithm's responsiveness to weighting. This behavior underscores the potential of weighting methods to divert passengers away from overcrowded hub stations, potentially reducing congestion and improving overall transit efficiency.

5.7 Conclusion

Incorporating RF and VF in routing decisions provides a comprehensive understanding of the London Underground system. It not only factors in the geographical layout but also the strategic importance and potential impact of each station. Such insights are crucial for effective urban planning, especially when anticipating disruptions or devising improvements.

6 | Discussion

6.1 Integration with Existing Literature:

6.1.1 Centrality Measures:

Echoing Scott et al. (2016) and Serdar et al. (2022) who underscore the significance of connectivity in urban traffic networks, our analyses of the London Underground's centrality reaffirm this principle[8, 9].

6.1.1.1 Degree Centrality:

Stratford's pronounced centrality in our study quantifies and deepens the understanding from prior qualitative discussions. Our data-driven approach accentuates Stratford's prominence relative to other stations.

6.1.1.2 Closeness Centrality:

Existing literature often highlights the significance of certain nodes using qualitative assessments. Our empirical findings corroborate these, with Green Park's centrality confirming its notable status due to its strategic location.

6.1.1.3 Betweenness Centrality:

While Stratford's pivotal role has been alluded to in past studies, our analysis provides a granular understanding, spotlighting its function as a nexus in the network and its extensive influence on transit journeys.

6.1.2 Resilience and Vulnerability Assessments:

Previous works by Ghose et al. (2006) and Cadarso et al. (2017) paved the way for comprehending disruptions in transportation networks[10, 11]. Our examination of the effects of station removal and the subsequent RF and VF evaluations reflect similar resilience and vulnerability metrics in urban transit systems.

6.1.2.1 RF Values:

The critical role of the "Bank and Monument" station, as demonstrated by its RF value in juxtaposition with significant nodes like "Stratford" and "Liverpool Street", delivers nuanced insights into the network's resilience. These empirical findings have implications for urban planning and crisis management strategies.

6.1.2.2 VF Values:

By assessing the vulnerability of the "Bank and Monument" station, our research underlines its indispensable role within the network. This quantitative emphasis complements the existing literature's focus on the strategic importance of certain transit points.

6.1.3 Pathfinding Algorithms:

Employing a weighted A* algorithm for route identification, our approach offers a rigorous, quantitative methodology to urban transit routing, akin to Ghose et al.'s GIS-based model. Our foundational tests vouch for our algorithm's precision and reliability, aligning with the academic quest to decipher optimal routes in urban systems.

Our study reinforced the general principles articulated by Yao et al. (2010) regarding the A* algorithm's evaluation function $f(n) = g(n) + h(n)$. The use of heuristics indeed guides the search in the desired directions, facilitating efficiency. Notably, our findings further emphasize the significance of the heuristic function $h(n)$. While Yao et al. (2010)

noted potential inefficiencies if $h(n)$ isnât appropriately chosen, our research offers a more in-depth insight into the criteria for selecting suitable heuristic functions. Interestingly, the Zhen et al. (2023) study highlighted the A* algorithm's ability to obtain optimal paths due to its heuristic nature. Our results echo this sentiment.

In summation, our findings not only resonate with previous academic endeavors but also extend their insights. While the research harnesses established paradigms, it further refines and adapts them, delivering a holistic analysis of the London Underground. This amalgamation of prior insights with our innovative methodologies provides a substantive addition to urban transportation scholarship.

6.2 Advancement of Current Ideas:

Our study delves into areas that existing research has touched upon, offering our perspective on urban transportation networks. Here's how the research see our contribution in relation to prior works in three main areas:

6.2.1 Dynamic Weighted Network:

The disruptions caused by subway strikes and other unforeseen events significantly impact commuter travel. Building upon the research by Guo, Wang, and Zhang, the research focused intently on two key factors: the level of congestion in the subway and service interruptions due to strikes. Using this information, the research assigned weights to each route, reflecting its practical availability and efficiency at any given time.

More crucially, by considering track closures and passenger congestion data, the research constructed a novel weighted network, offering commuters more timely and accurate route recommendations. This approach ensures that passengers receive optimal travel advice even during strikes or other disturbances.

6.2.2 A Nuanced Look at Resilience:

Previous research underscored the importance of resilience, but our work has presented a more nuanced view. By focusing on specific stations such as "Bank and Monument," the research highlighted how individual nodes can have a disproportionate impact on a network's overall resilience. This challenges the prevailing notion of treating all nodes with an equal lens and calls for a more differentiated, granular approach to urban planning and crisis management[10, 11].

6.2.3 Algorithmic Refinement:

The use of the weighted A* algorithm in our study not only validates the principles discussed by Yao et al. (2010), but also amplifies the need for the appropriate selection of heuristic functions. Our findings, especially on certain network topologies, raise questions on the universal applicability of the heuristic function $h(n)$, suggesting that a one-size-fits-all approach might not always yield optimal results[18, 19]. This beckons researchers to continually refine algorithms tailored to specific network configurations.

In short, while the research build on existing work, the research tried to add new insights and solutions to challenges the research identified.

6.3 Limitations of the Study:

Our study's approach to constructing a dynamic weighted network based on subway congestion and service disruptions, while innovative, does present certain limitations:

- **Lack of Visualization:** the research relied on data representation rather than providing a complete visual website. A visual approach, using tools like heat maps or bar graphs, would potentially have provided planners with a more intuitive understanding of the impact of service disruptions.

- **Absence of Real Strike Data:** Our study did not incorporate actual strike data. Instead, the research simulated the potential effects of service disruptions by removing specific stations from our dataset. This might not entirely replicate the complexities of a real-world strike scenario.

- **Route Recommendations:** While the research proposed new routes based on station congestion and vulnerabilities, some of these routes might be less direct, requiring passengers to stop at more stations and potentially increasing travel time.

6.4 Recommendations for Future Work:

Given our findings and limitations, the research sees potential paths for future research:

- **Visual Representation:** An immediate recommendation would be the development of an interactive website or platform, utilizing visual tools like heat maps or bar graphs. This would not only enhance data comprehension but also facilitate actionable insights for urban planners.

- **Incorporate Actual Strike Scenarios:** A more realistic representation could be achieved by incorporating real strike data. This would allow for a better understanding of the actual dynamics and repercussions of a service disruption.

- **Trade-off Analysis:** A deeper investigation into the trade-offs between route directness and congestion mitigation is warranted. This would provide commuters and planners with a clearer picture of potential compromises.

7 | Conclusion

In this study, we set out to explore the construction of a dynamically weighted network for the subway system, focusing primarily on the level of subway congestion and possible service disruptions caused by strikes. Our methodology draws on previous basic research, but turns to an innovative approach that combines station vulnerability and congestion data to propose new routes.

Our findings reveal important implications for urban transportation. While we succeeded in outlining the potential impacts on the network in the event of service disruptions at certain metro stations, our data-centric approach lacked an intuitive visual presentation that might be more intuitive to urban planners. The lack of real strike data also highlights the speculative nature of our simulated disruptions, emphasizing the importance of further real-world validation.

Despite these limitations, the implications of our study are clear. In an era of increasing urbanization, understanding the vulnerability and capacity of transportation networks is critical. By mapping these dynamics, we provide a foundational tool for urban planners to make informed decisions.

As a forward-looking proposal, the development of a visualization platform, the incorporation of real-world strike scenarios, and an in-depth analysis of route trade-offs can further refine our approach, ultimately benefiting urban planners and daily commuters.

Overall, our study breaks new ground in urban transportation research, while also highlighting the evolving nature of this challenge and the need for further research and collaboration.

References

[1] T. Y.-T. Lin, *Transit user mode choice behaviour in response to TTC rapid transit service disruption.* University of Toronto (Canada), 2017.

[2] I. Tsapakis, J. Turner, T. Cheng, B. G. Heydecker, A. Emmonds, and A. Bolbol, "Effects of tube strikes on journey times in transport network of london," *Transportation Research Record*, vol. 2274, no. 1, pp. 84–92, 2012.

[3] N. A. van Exel and P. Rietveld, "Public transport strikes and traveller behaviour," *Transport Policy*, vol. 8, no. 4, pp. 237–246, 2001.

[4] E. Ferguson, "Transit ridership, incident effects and public policy," *Transportation Research Part A: Policy and Practice*, vol. 26, no. 5, pp. 393–407, 1992.

[5] L. Haywood, M. Koning, and R. Prud'homme, "The economic cost of subway congestion: Estimates from paris," *Economics of Transportation*, vol. 14, pp. 1–8, 2018.

[6] J.-D. Schmöcker, M. Bell, and C. Lee, "An application of congested transit network loading with the markov chain approach," in *9th Meeting of the EURO Working Group on Transportation âIntermobility, Sustainability and Intelligent Transport Systemsâ, Bari*, 2002.

[7] S. Larcom, F. Rauch, and T. Willems, "The Benefits of Forced Experimentation: Striking Evidence from the London Underground Network*," *The Quarterly Journal of Economics*, vol. 132, pp. 2019–2055, 05 2017.

[8] D. M. Scott, D. C. Novak, L. Aultman-Hall, and F. Guo, "Network robustness index: A new method for identifying critical links and evaluating the performance of transportation networks," *Journal of Transport Geography*, vol. 14, no. 3, pp. 215–227, 2006.

[9] M. Z. Serdar, M. Koç, and S. G. Al-Ghamdi, "Urban transportation networks resilience: indicators, disturbances, and assessment methods," *Sustainable Cities and Society*, vol. 76, p. 103452, 2022.

[10] M. Ghose, A. K. Dikshit, and S. Sharma, "A gis based transportation model for solid waste disposal–a case study on asansol municipality," *Waste management*, vol. 26, no. 11, pp. 1287–1293, 2006.

[11] L. Cadarso, E. Codina, L. F. Escudero, and A. Marín, "Rapid transit network design: considering recovery robustness and risk aversion measures," *Transportation research procedia*, vol. 22, pp. 255–264, 2017.

[12] W.-M. Wey and J.-Y. Huang, "Urban sustainable transportation planning strategies for livable city's quality of life," *Habitat International*, vol. 82, pp. 9–27, 2018.

[13] A. Tero, S. Takagi, T. Saigusa, K. Ito, D. P. Bebber, M. D. Fricker, K. Yumiki, R. Kobayashi, and T. Nakagaki, "Rules for biologically inspired adaptive network design," *Science*, vol. 327, no. 5964, pp. 439–442, 2010.

[14] C. Guo, J. Wang, and Z. Zhang, "Evolutionary community structure discovery in dynamic weighted networks," *Physica A: Statistical Mechanics and its Applications*, vol. 413, pp. 565–576, 2014.

[15] M. DâLima and F. Medda, "A new measure of resilience: An application to the london underground," *Transportation Research Part A: Policy and Practice*, vol. 81, pp. 35–46, 2015. Resilience of Networks.

[16] S. L. Pimm, *The balance of nature?: ecological issues in the conservation of species and communities.* University of Chicago Press, 1991.

[17] M. Noto and H. Sato, "A method for the shortest path search by extended dijkstra algorithm," in *Smc 2000 conference proceedings. 2000 ieee international conference on systems, man and cybernetics.'cybernetics evolving to systems, humans, orga-*

nizations, and their complex interactions'(cat. no. 0, vol. 3, pp. 2316–2320, IEEE, 2000.

[18] J. Yao, C. Lin, X. Xie, A. J. Wang, and C.-C. Hung, "Path planning for virtual human motion using improved a* star algorithm," in *2010 Seventh International Conference on Information Technology: New Generations*, pp. 1154–1158, 2010.

[19] R. Zhen, Q. Gu, Z. Shi, and Y. Suo, "An improved a-star ship path-planning algorithm considering current, water depth, and traffic separation rules," *Journal of Marine Science and Engineering*, vol. 11, no. 7, p. 1439, 2023.

[20] T. V. Mathew, "Genetic algorithm," *Report submitted at IIT Bombay*, p. 53, 2012.

[21] M. Srinivas and L. Patnaik, "Genetic algorithms: a survey," *Computer*, vol. 27, no. 6, pp. 17–26, 1994.

[22] B. Liu, S.-H. Choo, S.-L. Lok, S.-M. Leong, S.-C. Lee, F.-P. Poon, and H.-H. Tan, "Integrating case-based reasoning, knowledge-based approach and dijkstra algorithm for route finding," in *Proceedings of the tenth conference on artificial intelligence for applications*, pp. 149–155, IEEE, 1994.

[23] D. Zhang, Z. Wei, J.-H. Kim, and S. Tang, "An optimized dijkstra algorithm for embedded-gis," in *2010 International Conference On Computer Design and Applications*, vol. 1, pp. V1–147–V1–150, 2010.

[24] J. Scott, "Trend report social network analysis," *Sociology*, pp. 109–127, 1988.

[25] L. C. Freeman *et al.*, "Centrality in social networks: Conceptual clarification," *Social network: critical concepts in sociology. Londres: Routledge*, vol. 1, pp. 238–263, 2002.

[26] U. Brandes, "A faster algorithm for betweenness centrality," *Journal of mathematical sociology*, vol. 25, no. 2, pp. 163–177, 2001.

[27] J. Zhang and Y. Luo, "Degree centrality, betweenness centrality, and closeness centrality in social network," in *2017 2nd international conference on modelling, simulation and applied mathematics (MSAM2017)*, pp. 300–303, Atlantis press, 2017.

[28] T. S. Evans and B. Chen, "Linking the network centrality measures closeness and degree," *Communications Physics*, vol. 5, no. 1, p. 172, 2022.

[29] M. J. Newman, "A measure of betweenness centrality based on random walks," *Social Networks*, vol. 27, no. 1, pp. 39–54, 2005.

[30] M. Barthelemy, "Betweenness centrality in large complex networks," *The European physical journal B*, vol. 38, no. 2, pp. 163–168, 2004.

[31] C. Ju, Q. Luo, and X. Yan, "Path planning using an improved a-star algorithm," in *2020 11th International Conference on Prognostics and System Health Management (PHM-2020 Jinan)*, pp. 23–26, IEEE, 2020.

[32] A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani, "The architecture of complex weighted networks," *Proceedings of the national academy of sciences*, vol. 101, no. 11, pp. 3747–3752, 2004.

[33] B. B. Rad, H. J. Bhatti, and M. Ahmadi, "An introduction to docker and analysis of its performance," *International Journal of Computer Science and Network Security (IJCSNS)*, vol. 17, no. 3, p. 228, 2017.

[34] T. Bui, "Analysis of docker security," *arXiv preprint arXiv:1501.02967*, 2015.

[35] W. Python, "Python," *Python Releases for Windows*, vol. 24, 2021.

[36] I. Stančin and A. Jović, "An overview and comparison of free python libraries for data mining and big data analysis," in *2019 42nd International convention on information and communication technology, electronics and microelectronics (MIPRO)*, pp. 977–982, IEEE, 2019.

A | Source Code

Source code for all of the methods implemented in Chap. 4 for the project can be found in the GitHub repository:
https://github.com/ucfnlui/Dissertation.