Metropolitan City Transportation Analysis & Optimal Route Suggestions

MSc Research Project
MSc Artificial Intelligence

M Anis Mumtaz Chaudhary
Student ID: x23215011

School of Computing
National College of Ireland

Supervisor: Mayank Jain

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	M Anis Mumtaz Chaudhary
Student ID:	x23215011
Programme:	MSc Artificial Intelligence
Year:	2024
Module:	MSc Research Project
Supervisor:	Mayank Jain
Submission Due Date:	16/09/2024
Project Title:	Metropolitan City Transportation Analysis & Optimal Route Suggestions
Word Count:	7169
Page Count:	20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	12th September 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	■
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	■
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	■

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Contents

List of Figures

Metropolitan City Transportation Analysis & Optimal Route Suggestions

M Anis Mumtaz Chaudhary
x23215011

Abstract

Urban center's economic prosperity is largely dependent on the development of strong transit systems that enable easy access to key locations such as places of employment, recreation areas, and educational institutions. It will help the environment and reduce the need for personal automobiles. The reliability of transport networks has been studied in the past using stops density. The primary focus of this study is to do a comprehensive analysis of the Dublin transportation system using GIS with spatial analysis and suggest the optimal route for the city. The dataset for this problem is taken from a public source that contains the information on the bus routes. I have formulated the problem into a graph optimization problem by considering bus stops as nodes and the value of the edges as the distance between the nodes(bus stops). I have implemented Dijkstra and A* algorithms to find the optimal route. The Dijkstra algorithm shows a low-cost value of 422.87 compared to the A* algorithm which shows 430.74. On the other hand, Dijkstra visited 152 nodes to find the optimal route while A* visited just 59 nodes. The cost-value difference is quite nominal compared to the node visited which makes the A* algorithm the best performer.

1 Introduction

Building a robust transport infrastructure is essential to the growth of metropolitan areas because it allows for easy access to important locations including places of employment, entertainment centers, and educational institutions. Examining the relationship between urban economic expansion and effective transport networks will help to clarify the significance of accessibility in social welfare and sustainable development. As cities grow, careful planning is necessary to determine the best sites for important services and to create a seamless network of roadways, public transportation, and parking facilities. Urban environments are significantly shaped by the transport framework, which also has an impact on regional growth, economic sectors, and environmental dynamics. The main objective of this study is to investigate the specifics and opportunities associated with developing and implementing transport planning strategies designed for large cities. In the contemporary world, which hails globalization and rapid advancement in urban surroundings, appropriate public transportation systems are essential for the 'smooth-running' of metropolitan cities. With an aim of improving its transportation system, this project entails the designing of efficient bus routes in Dublin City implying some of the problems that may be common in most big cities all over the world. The need for means of transport rises due to the increasing size of cities and affects not only the transportation between home and workplace but the city's livability as well. To tackle these issues, this project tries to reformulate the problem of route optimization in terms of graphs. Dublin, as one of the most important European cities, has a powerful and developed public transportation system; however, the lines and routes of buses operate constantly, hence we will try to improve the existing plan by dividing the bus stops as nodes and determining the best paths. The advancement resulting from this project is essential as it enhances the local transport services while also helping in attaining the long-term objective of constructing smarter urban transport systems. Finally, the project evolves to meet this urgent need by enhancing the general road mapping which affects traffic jams, environmental issues, and the economic development of the city. Dublin has rapidly expanded over the past 20 years, housing a population of over 1.25 million people today. It is anticipated that the population will continue to grow, putting further strain on the current transit system. According to Perveen et al. (2020), the transportation sector is a large contributor to total environmental pollution, accounting for 23% of CO emissions in 2017. This conclusion is drawn from an analysis of worldwide fuel consumption patterns in the transportation sector. The study's objectives are to conduct a thorough review of Dublin's bus transport system and provide the best possible route recommendations for the city's. Using GIS and spatial analysis, all routes will be mapped, and bus stops rate according to socio-economic factor, and an ideal route will be generated based on analysis results.

1.1 Key Contributions

- Developed specialized software to visualize bus stops and their connections, enhancing the understanding of the transit network and aiding in the accurate representation of spatial relationships within the city.

- Created a comprehensive graph-based model that represents bus stops as nodes and calculates optimal paths using Dijkstra's and A* algorithms, effectively translating a complex transportation problem into a solvable optimization framework.

1.2 Research Questions

- How can effectively model and visualize bus stop locations to accurately represent the real-world transit network in Dublin City. This intends to establish effective ways of displaying a bus stop and its location in relation to other bus stops so that an accurate representation of the city transport system for operational reasons can be made.

- What are the strengths and limitations of Dijkstra's and A* algorithms in finding the optimal bus routes within a large metropolitan network. This aims to assess and compare the performance of Dijkstra's as well as A* algorithms when it comes to finding the shortest path with reference to features like distance, time, and flexibility of the algorithms with different stages of processing.

- How does the performance of the optimized bus route model vary with different algorithmic parameters, such as cost factor and the number of nodes explored, and what are the trade-offs between accuracy and computational efficiency. This question aims to study how the change in parameters of the Dijkstra's and A* algorithms affects the model's performance regarding cost factor and the number of nodes explored and to understand the trade-offs between getting accurate route optimization and maintaining computational efficiency at the same time.

The second section of this study provides in-depth assessments of public transportation accessibility, emphasizing the key elements influencing mobility in urban settings. These variables, which provide important insights into the variables impacting the availability of public transportation systems, include population density, size of the population, regional area, and proximity to the closest public transit stops. The fourth part of this study provides a detailed explanation of the design using research resources such as datasets. It also addresses ethical issues for those datasets in depth and discusses the overall research roadmap. In section 6 of this study, several evaluation measures will be utilized for the assessment of the models used for finding the optimal route suggestions. Section 7 of this study will conclude the overall outcomes of the research and the future work directions.

2 Related Work

After a thorough analysis of the literature, Saif et al. (2019) came to the crucial conclusion that accessibility is important when designing the infrastructure for public transit. The authors underscored the noteworthy correlation between several facets of transportation networks and the accessibility of public transit, including mobility and sustainability. They also emphasized how it affects a number of facets of human existence, including integration, employment prospects, and communal well-being. The study emphasizes how important accessibility is in shaping infrastructure design and transportation policy. They draw attention to the various advantages of giving accessibility priority when developing public transit, which helps to create a more robust and desirable urban environment. A new method for evaluating public transport accessibility in urban settings was presented by Saghapour et al. (2016), who also emphasized the significance of population density incorporation as a critical component of public transport accessibility assessment. The Statistical Area 1 (SA1) in Melbourne was the subject of their study's modeling, which showed that 0.4% of the population lacked a connection to public transit. For trains, walking distances of 1200 meters and 800 meters were included of the modelling parameters. Planners and legislators attempting to enhance public transport systems in densely populated places may learn a great deal from this research, which emphasizes the significance of population size when evaluating accessibility.

Oliveira and Garcia (2021) have carried out the performance comparison of several route generation methodologies employed for the collection of primary wastes with the help of human-powered vehicles. They set out to reduce waste pickers' energy expenditure by evaluating and comparing various heuristics and search algorithms with regard to route planning. The study used geographical context information obtained in Salvador, Brazil, and two kinds of experiments with 15 and 30 stops. The study data was collected from the Open Street Map and Geomorphometric Database of Brazil. These are heuristics, Nearest Neighbor, Nearest Insertion, and Farthest Insertion, and the search algorithms include, SPFA, Bidirectional Dijkstra, and A-star. Evaluation criteria involved in the process were based on the time taken to complete processing as well as the distance completed by the vehicle and decided by physical work input utilized in pushing the selected vehicle. The investigation showed that, for the shortest processing time, the Nearest Neighbor used in combination with Bidirectional Dijkstra was optimal whereas the exploration of minimal physical effort was achieved using Nearest Insertion. More so the study gives a lead on routing the waste collection in a way that can actually minimize the efforts that are taken in the process. Bringing up the discussion, the research done by REDDY and HARIKA (2024) examines a proficient recommendation and suggestion system for travel routes through POIs. Thus, based on the rising volume of location-based social network (LBSN) data available on sites like Facebook and Flickr, the authors introduce the Keyword-aware Representative Travel Route framework. This framework utilizes the data derived from users' previous movement history and PIIA to categorize POI-related tags into groups relevant to the user query keywords. They also present a route reconstruction algorithm and discuss the Representative Skyline idea to illustrate trade-offs among various POIs' characteristics. The authors also compare the efficacy of their approach on real LBSN datasets with what was obtained by previous methods. However, it owes its challenges concerning the variety of input and capacity of receiving user-generated content, which results in potential issues with detecting user preferences and the feature of the dynamic character of travel routes.

According to Lerin et al. (2012), there is a proposal for a web map system for drawing opportunistic travel modes with the use of a mouse-sensitive following path indication. Their interaction model enables the users to learn the flow of procedures that are necessary to draw a route in a highly comprehensible manner and minimizes the number of user interactions. It is important to note that the system offers the user fixed and changing suggestions to help them without demanding their pledges. To reduce the effect of such communication delay, the architectural design has preloaded the road network data of the map server upon connection with the web client. Modeled and empirical assessments showed that the proposed system allows the users to create routes in less number of clicks, within a shorter amount of time, and with less amount of perceived exertion than the previously used systems. The findings of this research provide suggestions for the enhancement of user interaction and productivity when applying web-based route drawing applications.

Based on the problem, Tostes et al. (2015) introduced a multicommodity flow model to optimize the distribution of lanes in a vehicular network to reduce traffic congestion. They assessed four new heuristics with the help of two constructive heuristics such as Random Routes and Sorted Routes and also with the help of another local search and metaheuristic of reactive GRASP. The study used a large-scale realistic mobility data-set of TAPAS Cologne that is cropped at two hours of movement of the vehicles within a region of 400 km^2 of Cologne in Germany. The results proved that actual savings ranged from 18% of CO2 emission reduction, 18% of fuel reduction, to 56% reduction in time travel for the route. Nonetheless, the complexity of solving the problem to find the optimal solution and the necessity to maintain the search for the balance between good solutions and reasonable computation time are seen as major drawbacks of this approach. The study underscores possible enhancements of the vehicular communication network and the environmental effects as a result of optimal suggested routes. Brodje et al. (2015) focused on examining the effectiveness of a specific functionality called 'route suggestion' in naval VTS. In their study, an actual and controlled integrated environment which includes full mission bridge simulators and E-navigation Prototype Display systems was used to assess this new aspect. In doing so, the research sought to evaluate the usability and operation efficiency of the feature to implement the envisaged improvement in communication between VTS operators and navigators. The study used video protocols, observations, and questionnaires To collect data the main research areas include technical aspects such as the ease of sending the route suggestion and the readability of the text while operational the study looked at aspects such as the additional workload of the navigator's in adding further verification on the suggested routes. Some features such as the "route suggestion" can be potentially effective in enhancing on-road safety and information exchange; however, proactive problems such as usability concerns and increased workload must be considered. These challenges assert the importance of addressing the concerns to maximize the efficiency of route suggestions in the overall optimization of traffic flow in the maritime sector.

Rajarajeswari et al. (2013) et al have proposed methods for recommending the best routes for vehicles through real-time traffic information in automotive navigation systems. Navigation has been enhanced by choosing less congested paths using other methods like vehicular traffic prediction with the help of GPS, speed, and the car's accelerometer telemetry. Research has revealed that the utilization of real-time data helps improve the accuracy of the chosen routes; however, there is a necessity for increasing the effectiveness of real-time updates and the employed algorithms. Some improvements include clustering of the vehicle based on behavior and, traffic forecast by means of exponential moving average. The typical approaches include graphic models of road networks and heuristic procedures, such as A* search using current GPS data and the history of traffic movements. From the equation, it is clear that while headway has been made, there are still issues of computational complexity and more so, the need to continuously update the data fed to the formula.

Currently, Cardoso et al. (2018) have advanced notable progress within the field of route planning with the aid of a computer, namely, optimizing visits to cultural points of interest. Their work couples RS with route optimization algorithms to design a personal and optimized tour experience. The one they suggest, in the framework of the M5AR project, is to establish an augmented reality system in museums extending a mobile application of cultural tour guides. This application employs an RS for anticipating the user tastes and an NSGA-II for the routing problem. The used datasets involve the virtual network of the museum that contains 108 Points of Interest (PoI) and the network of Faro City in Portugal. The RS uses methods such as KNNcb and CF. Is for filtering while the optimization uses NSGA-II to get multiple objectives such preference of a user and the feasibility of the route. However, the approach has some shortcomings in accommodating more than limited numbers of PoIs without burdening the users, guaranteeing feasible routes within the maximum permitted visiting time, and diversifying options while meeting customers' demands. The incorporation of RS and genetic algorithms has the potential to improve the experiences of cultural sightseeing; Nevertheless, there is still space for improvement in other terminals if certain aspects such as handling large dumps of data work, improving the interface to reduce data overload, and getting the right balance between users' preference and route optimization.

A GIS-based indexing approach was created by Yigitcanlar et al. (2007) to determine which locations of Australia's Gold Coast Local Government Area (LGA) had different degrees of access to public transit. As an origin-centric accessibility model, LUPTAI included network analysis as a fundamental component and employed a variety of GIS-based spatial analytical tools. Based on the travel time to a public transportation stop and its regularity of operation, this technique categorized accessibility into four categories: bad, low, medium, and high. In order to support decision-making for growth and transportation strategy within the Gold Coast LGA, they sought to provide a thorough knowledge of accessibility for public transport patterns. A GIS-based tool designed for assessing transit accessibility in London, UK, was presented by (Ford et al., 2015). This tool was developed to make use of an aggregate cost strategy that takes distance as a crucial component and incorporates transport costs across several networks. According to their research, there were significant accessibility issues in London's less populated districts and the periphery of suburban neighborhoods. Through the identification of these disparities, the GIS-based technology offered insightful information to urban planners and politicians, enabling focused actions to improve transportation efficiency and fairness in particular areas.

Using the measure of spatial blank spots, Alamri et al. (2023) assessed the effectiveness and connectivity of Melbourne, Australia's public transport system, identifying residential areas that were inaccessible to public transport stations. In light of the Local Government Area's (LGA) size and population density, the research also investigated service variety throughout various time periods (weekdays, weekends, evenings, nights, or 24-hour periods). According to the findings, the LGA with higher population densities had outstanding accessibility and availability of transportation services, while the bigger LGA tended to have a larger geographical blank region. Insights critical for optimizing service provision and infrastructure development within metropolitan regions are provided by this research, which emphasizes the impact of population growth and LGA extent on transportation accessibility.

A dearth of research has compared public transport systems in various cities, especially with regard to accessibility. An extensive investigation comprising six European cities was carried out by Ingvardson and Nielsen (2019). According to the study, customer satisfaction scores were positively correlated with greater accessibility and lower trip expenses. While comparing the satisfaction levels with public transport use among cities was the main objective of this study, it did not evaluate any particular demographic or area that did not have access to public transport. However, the results underscore the importance of accessibility and cost factors in determining public transport customer happiness, providing insightful information to transportation planners and politicians to improve the standard of service and customer satisfaction in metropolitan settings. Similar to this, Osman et al. (2020) compared the public transit systems of two comparable European cities, focusing primarily on urban area's time as well as temporal procedures for chrono-urbanism. These policies include travel schedules, commute times, and temporal trends in activities, all of which are intended to lessen the difficulties associated with long commutes. The study looked at how public transport links are distributed between two cities for various travel scenarios, days of the week, and moments of the day. It is noteworthy, nonetheless, that the cities were viewed as single entities in the analysis, as opposed to smaller grid units or Local Government Areas (LGA).

Ceder and Teh (2010) compared the accessibility of public transportation amongst the main cities in New Zealand, evaluating eleven characteristics that influence an individual's preference for public transportation over alternative forms of mobility. According to their analysis, Wellington and Christchurch have lower connectivity scores than Auckland's central business center. Interestingly, the analysis focused only on the cities' main business districts and assessed connection rather than accessibility. In their study, Biswas et al. (2023) compare the effectiveness and convenience of public transit in Sydney and Melbourne, two cities with similar populations, sizes, and economic attributes. The methodology uses accessibility-focused spatial PostGIS queries to analyze every domestic mesh block in order to identify uncovered regions and assess service availability at the Local Government Area (LGA) level according to duration and time of day. They discover a pattern: the ratio of blank spaces rises as LGA gets away from the city center, while service availability and weekend/night service stops decline. Despite its longer distance from the city center, Melbourne's LGA performs well in terms of accessibility, particularly service frequency and punctuality, whereas Sydney has a lower number of blank spots and wider coverage.

The related work highlights the importance of public transport design, route planing and role of technology to improve transportation in various urban cities. The research gap is to identify the underserved areas and improve service delivery based on existing infrastructure to provide quick solutions.

3 Methodology

In this section, I am going to describe the methodology of the project. Firstly I gathered the data from TFI website for all the bus routes operating under Dublin bus network in Dublin county including each bus stop comes across the whole route. Then I identify the unique bus stops across the whole of Dublin county. I got the city council polygon from the Administrative Area DCC and then divided the whole polygon into 5 sub-parts named 'Central', 'North Central', 'North West', 'South Central', and 'South East'. Now discuss the distribution of bus stops in these sub-parts of the city council by using heat maps and clusters using folium to identify the busy areas.
Then I generate our custom polygon and identify the bus stops in that specific region and convert this data into a graph optimization by treating bus stops as a node and assign value to each node based on the socio-economic importance of bus stops. Socio-economic ratings are given by determining the closeness of bus stops to the city centre. I use O'Connell Bridge as a base point and each bus stop within a radius of 300m of the centre point assigned rating 1 we used 300m as a base value for rating calculation, as move away from the centre your rating will be increased accordingly. I use this rating and distance between stops to assign weight to edges. Now our data is translated into a graph optimization problem and applied A* and Dijkestra's algorithms to identify the optimal path between two bus stops in our specific predefined area and compare the results of the algorithms.

4 Design Specification

In this section, I will describe the complete design, details of the dataset, and all the analysis performed on the dataset. Firstly the data has been taken from the TFI (Transport for Ireland) and several preprocessing has been implemented to prepare the dataset for analysis and implementation of algorithms. I have used the route data to find the unique stops in Dublin City. From the unique stops, I have also extracted the stops according to a polygon that represents a specific area of Dublin city. After performing preprocessing steps, I have converted the problem into graph optimization by using bus stops as nodes and the distance between the stops as edges. The advantage of converting this problem into graph optimization is that I can implement various optimization algorithms to find the optimal path. I have implemented two algorithms namely Dijkstra and A* search to find the optimal route for the buses in Dublin City. The overall methodology diagram is shown in Figure 1.

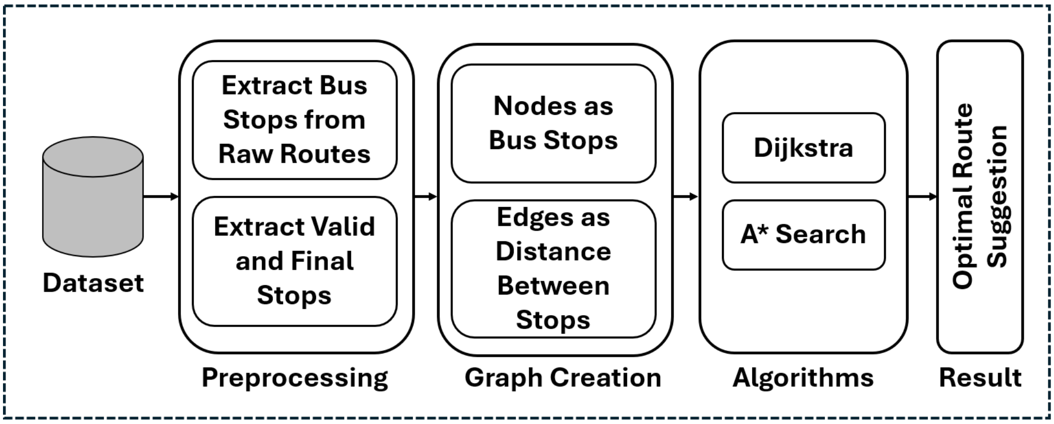

Figure 1: Design Specification.

4.1 Dataset

The data for this research has been taken from the TFI (Transport for Ireland). The dataset consists of raw bus routes which are in the form of a CSV file that has a total of 57170 records and five features namely 'route number', 'Stop', 'longitude', latitude', and 'stop number'. The longitude and latitude provide information for each bus stop, while the stop column contains the name of the bus stop and the stop number represents each stop with a unique number.

4.2 Data Preprocessing

I have performed the data preprocessing to ready the data for visualization and for later methodology implementations. To find all the bus stops from the routes data, I have merged the data of the Dublin city council inbounds and outbounds that contain all the bus routes, but still, this contains several stop numbers multiple times, so I have removed all the duplicates stop number from the data to get the unique bus stops. I have got the Dublin City Council polygon from the Administrative Area DCC. This polygon contains several sub-distributions namely 'Central', 'North Central', 'North West', 'South Central', and 'South East'. So, this is mainly used for the separation of each sub-division/area as mentioned above. After deriving the polygon of each area, I have used this polygon mechanism to extract the distribution of the bus stops from each region within the Dublin city council. I also use the same polygon to identify bus stops distribution in the city council and outside of Dublin city council, now I can perform several analyses like how many bus stops are there in any of the regions of the Dublin city council and outside of city council boundaries.

4.3 Data Analysis

After performing the preprocessing of the data, now I am able to perform any kind of analysis of the data. The overall area has been divided into two portions that include the inside of the Dublin City Council which is further divided into 5 regions as mentioned in the previous section and the outside of the Dublin City Council which is also known as the Dublin County. The distribution of the bus stops among these two regions is visually shown in Figure 2. The pie chart shows that about 60% of the bus stops are outside of city council boundary while about 40% of the bus stops are in the city council boundary. On the other hand, the bar chart shows the exact count of the bus stops in each region. This shows that outside of city council boundary has a rich number of bus stops compared to the city council.

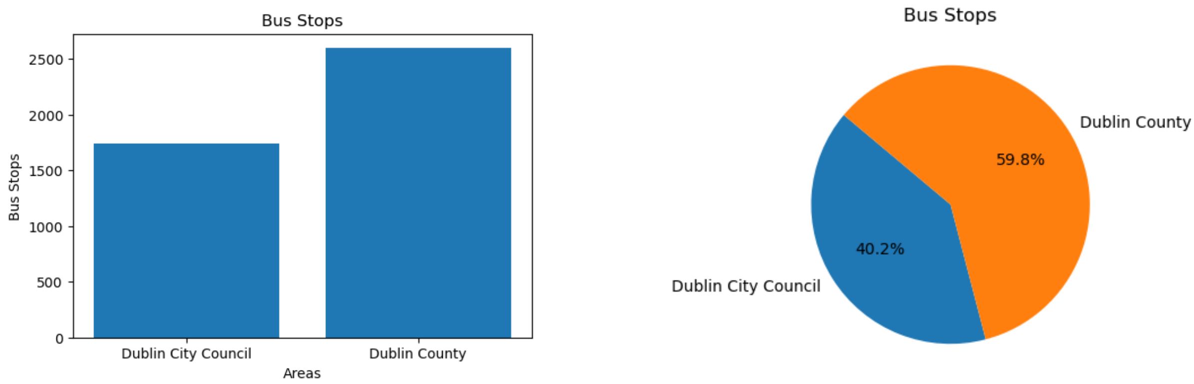

Figure 2: Comparison of City Council and County.

The Dublin City Council consists of 5 sub-regions and the distribution of bus stops in each region is different from others based on the population size and requirements of that region. The distribution of the bus stops in all five regions is graphically shown in Figure 3. The bar graph shows the count of the bus stops in each region while the pie chart shows the proportions of bus stops in each region. South East has more than 500 bus stops while the North West has less than 200 bus stops, while the rest of the region has values in between the maximum and minimum. Similarly, the South East has 29.4% of the bus stops which is the maximum proportion, while the North West has the lower proportion of bus stops which is 10.0%.

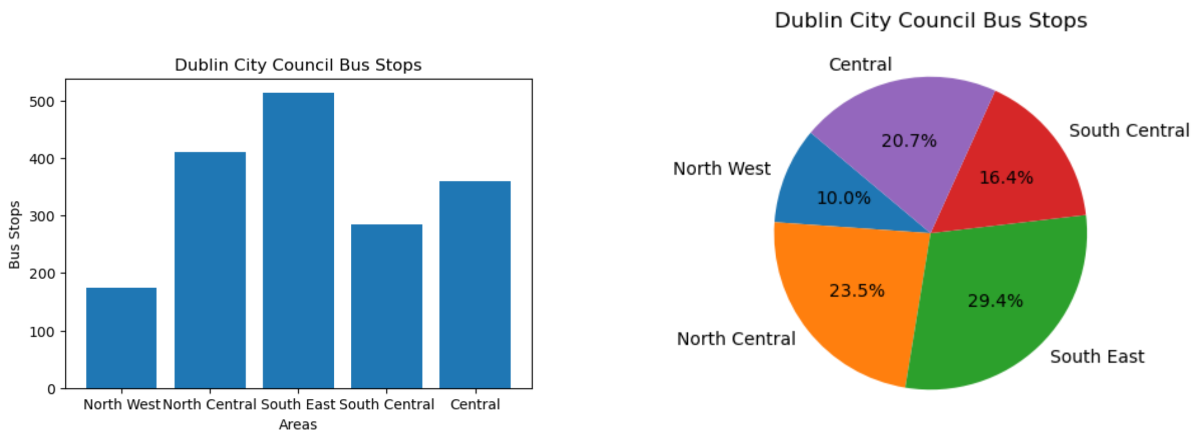

Figure 3: Comparison of City Council.

I have performed a further analysis in which I have plotted the density of bus stops in the Central region in the form of a heat map and cluster. The visual representation of the graph is shown in Figure 4.

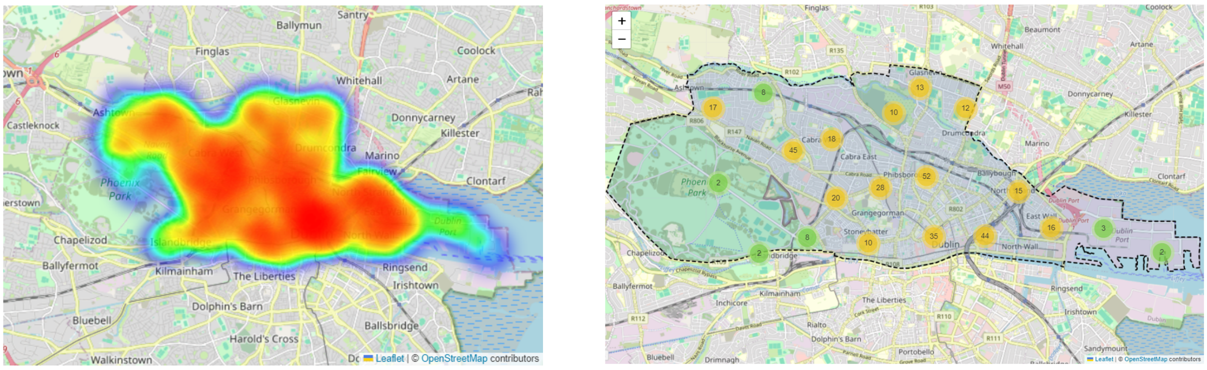

Figure 4: Central Region.

For the North West region, I have also plotted the cluster and heat map graphs for a better understanding of the distribution of bus stops in each region. The graphical representation of the cluster and heat map for the 'North West' region is shown in Figure 5 respectively.

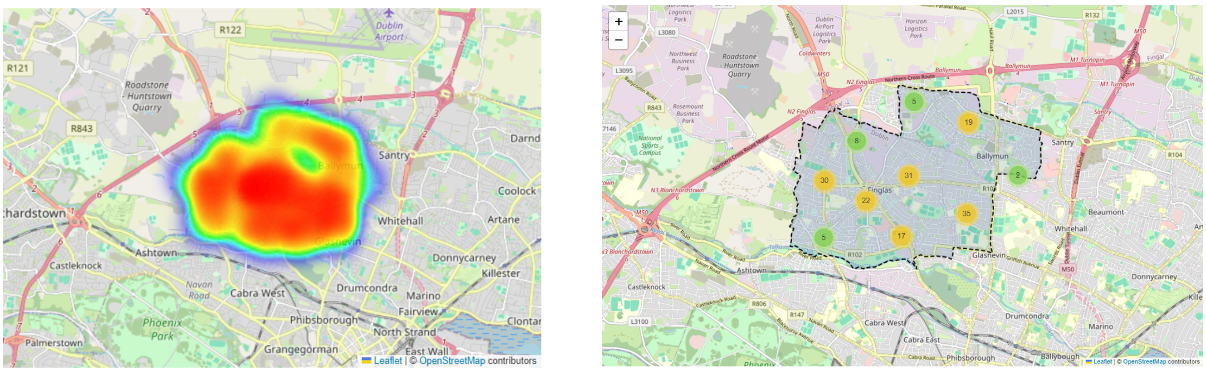

Figure 5: North West Region.

To understand the North Central region, I have also plotted the cluster and heat map for this region. The cluster and heat map for the North Central region are displayed in Figure 6 respectively.

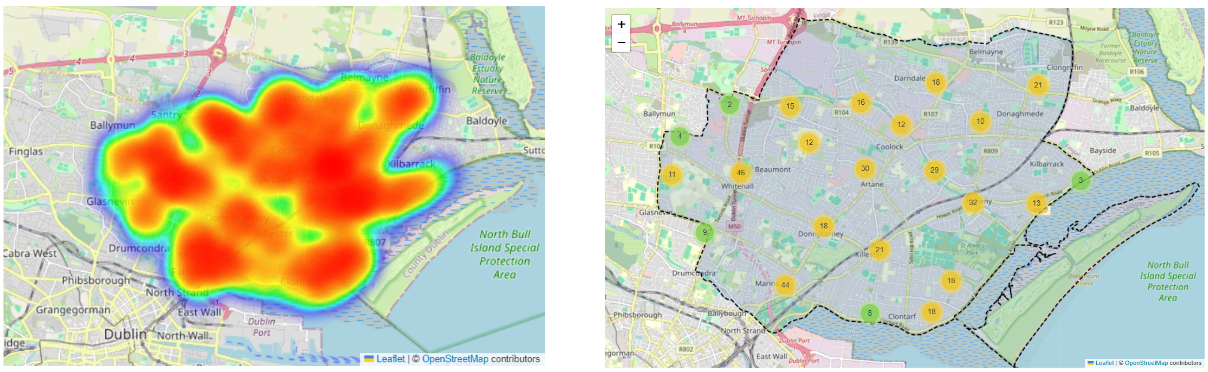

Figure 6: North Central Region.

For the estimation of bus stops for the Southeast region, I have plotted the bus stops cluster and heat map graph for this region. The visual representation of these graphs is shown in Figure 7 respectively.

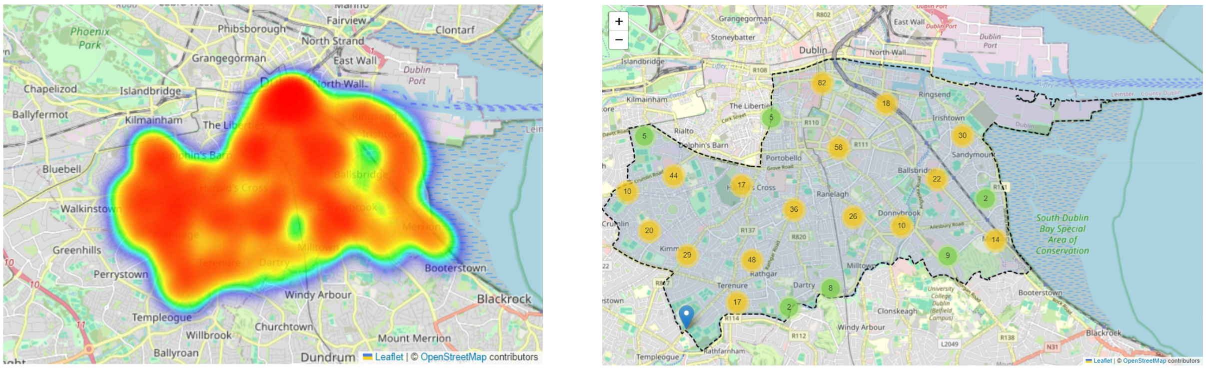

Figure 7: South East Region.

Finally, to understand the distribution of the bus stops in the region of South Central, I have analyzed this region by plotting the cluster graph for the bus stops and heat map for this region. The graphical representation of these plots is shown in Figure 8 respectively.

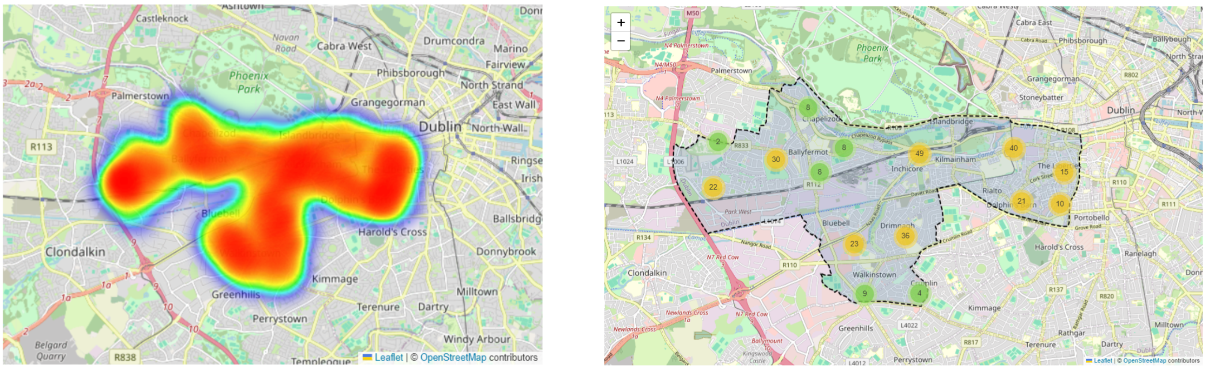

Figure 8: South Central Region.

I have drawn a custom polygon within a Dublin City Council to work on this specific polygon as the original polygon was very spread. I have also extracted the bus stops that fall under this custom polygon. Figure 9 shows the custom polygon that was used to create the graph.

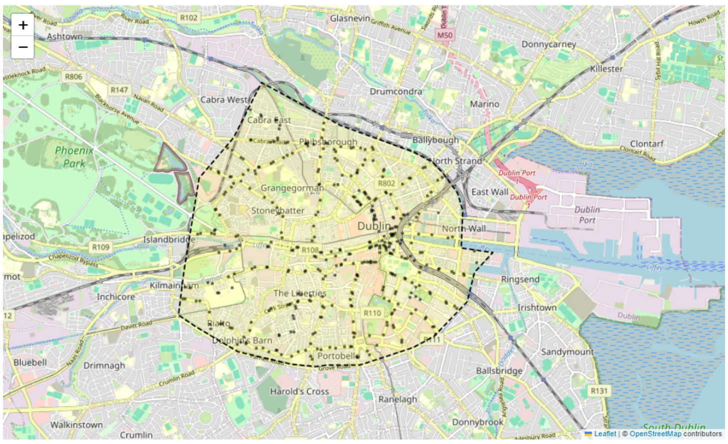

Figure 9: Custom Polygon Bus Stops.

Similarly, to understand the distribution of the bus stops in the custom polygon, I have plotted the cluster graph and heat map for this polygon. The graphical representation of these plots is shown in Figure 10 respectively.

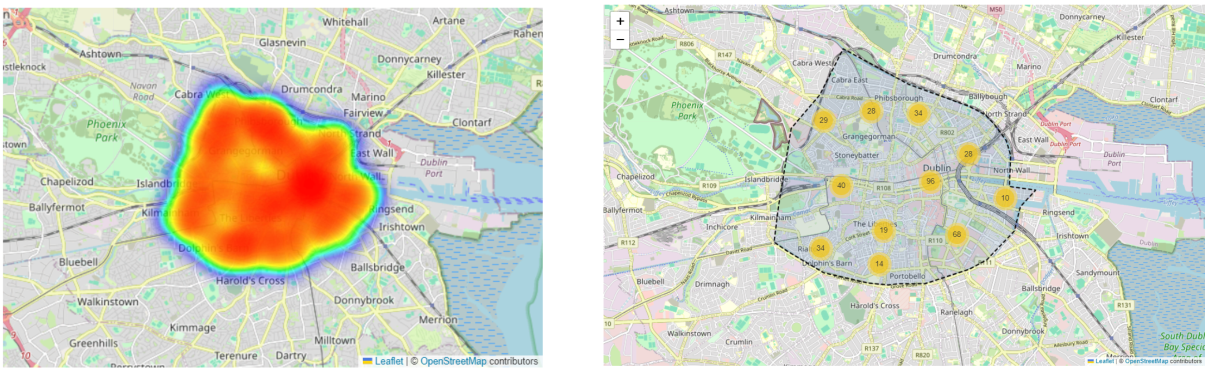

Figure 10: Custom Polygon Region.

5 Implementation

To convert the problem of finding the shortest or optimal path for buses running in Dublin City into a graph optimization problem, I have performed the following steps:

5.1 Modelling Bus Stops as Node

In graph theory, a graph is formed by nodes or vertices and edges or the connections between the nodes or vertices. To solve our problem I will represent each bus stop in Dublin City as a node say 'N'. This means that I have a unique point for each bus stop in the case of the graph that I am drawing. Graph theory shall assume bus stops as nodes for identifying the best routes between bus stops by implementing several algorithms. This abstraction allows to pay attention to the linkages and distances between bus stops rather than their physical appearance in the city.

5.2 Assigning Values to Nodes

After identifying the nodes, the next step is to assign values to these nodes. I have done this based on their closeness to a central point in the city, such as I have selected the central point as O'Connell Bridge. The assignment of values has followed the following criteria:

5.2.1 Distance from City Center

I have measured the distance from O'Connell Bridge to each bus stop.

5.2.2 Socio-economic base value assignment

I have assigned a value of 1 to nodes (bus stops) within a 300-meter radius of O'Connell Bridge. This means that all bus stops within this radius have the same value, indicating they are very close to the city centre. Then the value of the node increases as it moves further away from the central spot. All bus stops within a radius of 300 meters to 600 meters have a value of 2, all bus stops within a radius of 600 to 900 meters have a value of 3 and so on, I have assigned values proportional to their distance from the city centre. The further a bus stop is from O'Connell Bridge, the higher its assigned value.

Assigning values to nodes in the way that described above will help in determining the socio-economic importance of each bus stop based on its closeness to the city centre. Bus stops closer to the centre might be considered more central or significant in terms of connectivity and accessibility.

5.3 Assigning Weight to Edges

To create edges or connections between nodes or bus stops, firstly I have calculated the distance between each pair of bus stops. For this task, I have used the geopolitical coordinates namely the longitude and latitude of each bus stop. I have used the Euclidean distance formula to find the straight line distance between the stops. This method provides a simplified measure to calculate the straight-line distance but real-world routes may not always follow straight lines due to roads and obstacles, this method is useful for initial modeling and analysis. Then I use node value and distance to assign value to the edge between the nodes, for this purpose we use the following formula:

$$\text{NodeValue} = \frac{\text{SocioEconomicRating}}{100}$$

$$\text{EdgeWeight} = \text{Distance} \times (\text{StartNodeValue} + \text{EndNodeValue})$$

5.4 Final Graph Representation and Optimization

By implementing all the steps mentioned above I have converted the problem of finding optimal bus routes in Dublin City into a graph optimization problem where nodes are bus stops and edges are the distance between these stops with socio-economic rating of the bus stop. Now I can implement graph algorithms like Dijkstra and A* search to find the optimal route.

5.5 Directional Graph

This directional graph represents the bus stops and routes in Dublin City as shown in Figure 11, where each node (blue circle) corresponds to a bus stop, and each directed edge (line with arrows) represents a connection between stops. The direction of the edges indicates the flow of bus routes from one stop to another. The numbers on the edges denote weights, which represents distances between stops and socio-economic rating of the bus stop. This visualization aids in understanding the connectivity and optimal bus paths within the city, helping to analyze and improve the transportation network.

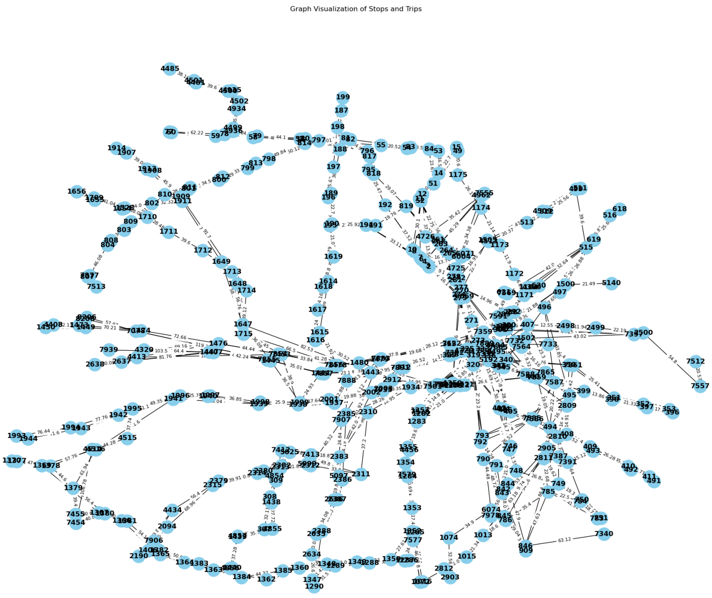

Figure 11: Graph

5.6 Algorithms

I have implemented two algorithms on the created graph to find the optimal route between the selected bus stops. The details of the implemented algorithm are mentioned below.

5.6.1 Dijkstra's Algorithm

Dijkstra's Algorithm (discovered by E.W. Dijkstra) is used to determine the shortest distance from the source node to the destination node in a graph. It is a greedy algorithm that works on both directed and undirected graphs. Every node in this algorithm corresponds to edges, which are equipped with weights. These weights (or edge costs) are used to find the cost of an edge from one node to another. This is done by adding these weights thereby allowing us to work out the most efficient way of going from the source node to the destination node. Initially, zero is assigned to the source vertex and infinity to the rest of the vertices. For every unvisited neighbor node, edge costs are determined and undefined labels are updated to the cost from the source vertex. The vertex with the smallest possible cost is then explored further in case it's not already visited. This continues until the destination vertex is visited and there are no more vertices left. Once the smallest costs are determined, we can trace the shortest path back from the destination node, by determining its predecessors until the source vertex is reached.

5.6.2 A* Algorithm

A* Algorithm is an extension of Dijkstra's algorithm which consists of some characteristics of breadth-first search. It is a greedy algorithm, that makes use of a heuristic function which calculates the estimated distance between the current node and the target node. Based on the sum of the heuristic estimate and edge cost, it determines which vertex to explore next. It allows the algorithm to optimally search for the target node without diverting from the goal. Much like Dijkstra's algorithm, a value of zero is assigned to the source vertex, and infinity to the rest of the vertices. For each explored node with the smallest total estimated cost, a heuristic function will determine the distance between the destination vertex and the current node. The total estimated cost is the sum of the cost it took to reach the current vertex from the source vertex and the heuristic estimate. This total cost is labeled on the explored vertex. The unvisited neighboring vertices with the lowest total estimated cost are explored next, for which the procedure is then repeated until the destination vertex is explored. To find the shortest path from the destination vertex, we can trace back its ancestors until we reach the source vertex.

In the case of our research, bus stops are denoted by nodes and the distance between the nodes is denoted by edges in a graphical representation. For each bus stop in the radius of 300 meters from the specfic point in city center named O'Connell Bridge, the value of 1 is assigned. For bus stops beyond the subsequent range of 300 meters, this value is increased in the same proportion. In this way, I have converted the problem into the graph optimization problem. I have selected the node whose value is 2154 as the starting node and the node whose value is 54 as the destination node. I have used the Dijkstra and A* algorithm to find the shortest path between these nodes or more precisely between these bus stops.

6 Evaluation

In this section, I will describe the experimental setup used for running the algorithm, Evaluation measures used for assessing the algorithms, results obtained from the implemented algorithm, and a brief discussion of the obtained results.

6.1 Experimental Setup

For the experimental setup, all computations were performed on a laptop equipped with an Intel Core i7 5th generation processor and 16GB of RAM. This setup was chosen to ensure a balance between computational power and accessibility, allowing for efficient processing of the route optimization tasks. The hardware configuration facilitated the smooth execution of the algorithms and timely analysis of the transportation data.

6.2 Evaluation Method

I have used two methods for the assessment of the implemented model namely cost and number of nodes explored. The details of these evaluation measures are as follows:

6.2.1 Cost Factor

Cost is a type of computational complexity that describes the sum of weights to find the optimal path. The cost of an algorithm is the sum of edge weights between the suggested path. As a result, it is highly dependent on the size of the processed data. Cost provides a theoretical estimation of the algorithm's performance in terms of the suggested path. It plays a significant role in know-how the of the algorithm with large graphs. By comparing costs, we can choose the most appropriate algorithm for a given problem size and constraints. For example, knowing that A* can be faster in practice than Dijkstra's under certain conditions helps in making decisions. Cost estimation is essential for theoretical analysis and offers insights, particularly for very large datasets or high-density graphs.

6.2.2 Number of Nodes Explored

Node (Vertex) is considered as a point or a junction in a graph where edges meet. Every node can store information in its members and could be potentially linked to other nodes via edges to create the network that constitutes a graph. The number of nodes explored reflects how efficiently the algorithm travels the search space. It directly impacts the computational resources used, for example, memory and processing power. For algorithms like A*, the number of nodes explored indicates the effectiveness of the heuristic. Fewer nodes explored generally imply that the heuristic is guiding the search more efficiently. This metric is very important in practical applications since the greater distance between nodes would mean more computation is required, and for large graphs, this could be a real issue. That way, the efficiency of algorithms on these structures can be determined and if the algorithm needs to be fine-tuned or optimized.

Thus, these evaluation metrics prove that they are very useful when determining the behaviour and applicability of path-finding algorithms to different situations to help in decision-making and to devise better optimization methods.

The first algorithm used to find the optimal route was the Dijkstra which utilized 152 nodes but the cost was 422.87. On the other hand, the second algorithm that is utilized is the A* search algorithm which is the extension of the Dijkstra algorithm. This algorithm utilizes less number of nodes 59 but the cost is 430.74 which is greater than the Dijkstra algorithm. The Dijkstra algorithm shows a slightly cheaper path with a cost of 422.87 compared to the A* algorithm's cost of 430.74. This indicates that Dijkstra's path is marginally more optimal in terms of cost. A* explored significantly fewer nodes (59) compared to Dijkstra (152), demonstrating that A* is more efficient in terms of node exploration.

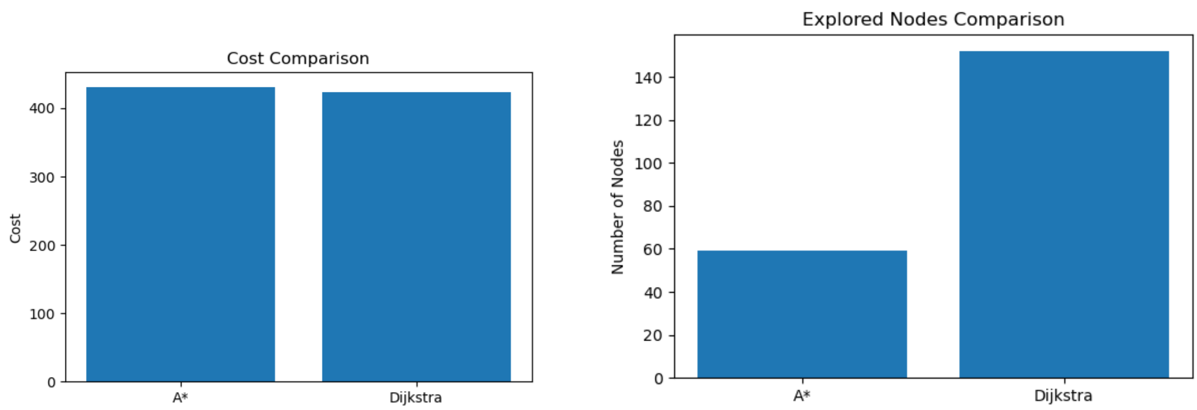

Figure 12: Cost and Expored Nodes Comparison.

Figure 13 shows the result of the A* algorithm for finding the optimal path.

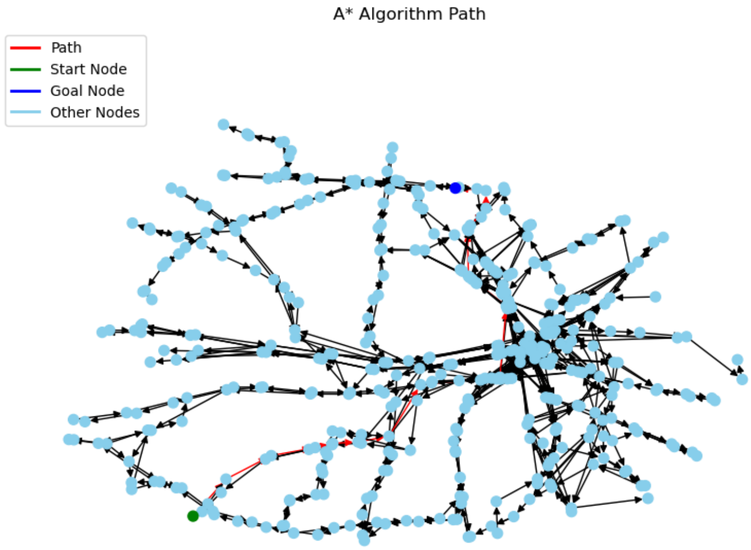

Figure 13: A* Optimal Path

Figure 14 shows the result of Dijkstra's algorithm for finding the optimal path.

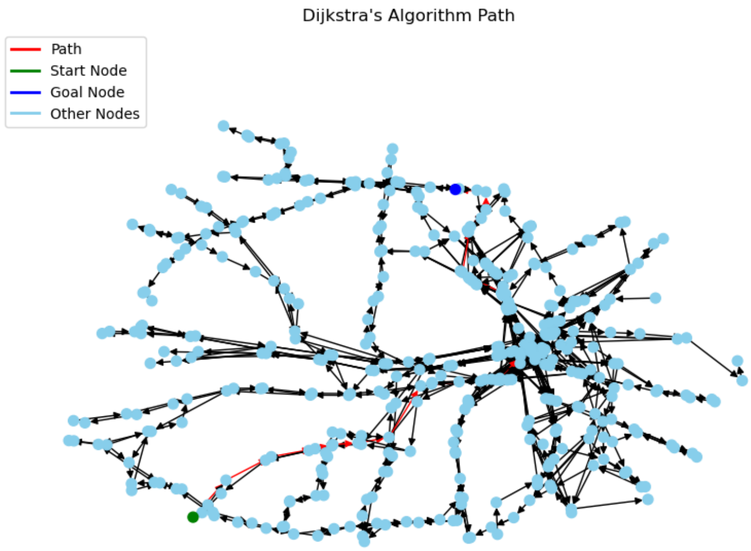

Figure 14: Dijkstra's Optimal Path

Figure 15 shows the number of bus stops in each suggested optimal route by both the A* algorithm and Dijkstra's algorithm

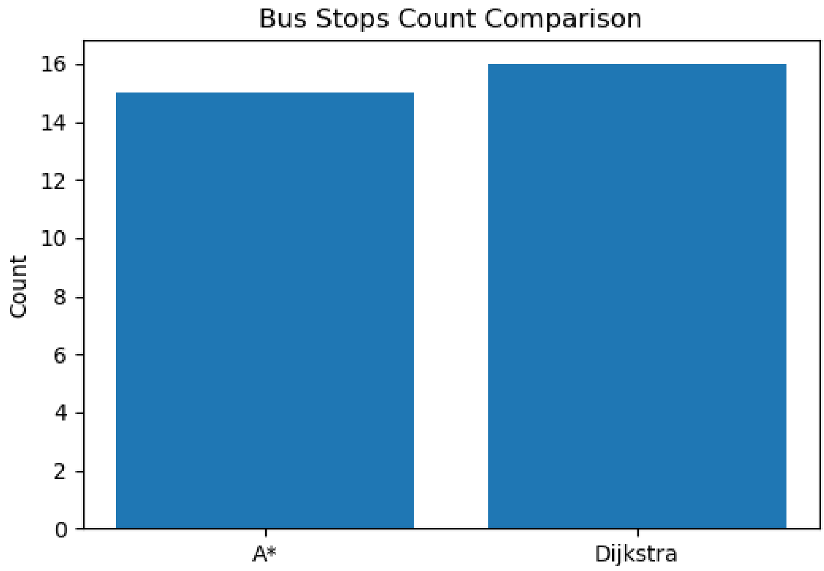

Figure 15: Number of Bus Stops

A* is generally preferred for its efficiency in node exploration and the number of suggested nodes to visit, especially when the difference in path cost is marginal. In this case, since the path cost difference is small (8 units), the A* algorithm will be considered better due to its efficiency.

7 Conclusion and Future Work

The main goal of this study is to use GIS and spatial analysis to undertake a thorough examination of Dublin's transport system and recommend the best path for any areas of the city. The dataset has been taken from a public website that contains information of the bus routes in a tabular format. The unique bus stops have been extracted from this data and the problem has been converted into a graph optimization problem. Bus stops have been considered as nodes and the distance between them is considered as edges of the graph. Dijkstra and the A* algorithm have been utilized to find the optimal route.

A* visited 59 nodes while the Dijkstra visited 152 nodes to find the optimal path between the source and destination node. The cost to reach the destination node by suggesting the optimal path for A* was 430.74 and the Dijkstra was 422.87.

A* is generally preferred for its efficiency in node exploration, especially when the difference in path cost is marginal. In this case, since the path cost difference is small (8 units), the A* algorithm will be considered better due to its efficiency. The Straight line (Euclidean) distance is employed to estimate the distance between each bus stop. This elimination does not consider the real road artery and any hindrances to the routes, and can thus reduce the efficiency of the recommended optimal routes. The model does not include dynamic factors that characterize traffic conditions such as changes within the traffic flow as well as events within the road network like construction, an accident, or congestion during rush hour that can affect bus routes and travel time. This analysis has been made in relation to the existing available bus stops and routes. Adjustments in future plans of the transportation system as witnessed by the incorporation of new bus stops or new bus routes would mean a later modification of the model and analysis of the outcome.

The limitation of the implemented work does not include real-time data like traffic peak hours, congestion, and passenger flow that vary throughout the day. Without this, the recommendations could not be optimal during peak hours. The implemented work does not integrate passengers' demands such as which areas are hot spots, and where and when passengers are more likely to be onboard. This information is crucial for effective bus route suggestions. This implemented work does not consider bus capacity constraints and bus frequency constraints which are important for the operational efficiency of transport system.

In the future, I will incorporate further features including real-time traffic data, population, optimized socio-economic factors, and other dynamic factors to enhance the accuracy and reliability of the models. In this way, the models will be able to find the best routes according to the situation. I will also extend the model to incorporate multiple criteria other than distance, such as travel time, passenger convenience, and operational cost. We will Adopt predictive analytic approaches to anticipate conditions of traffic flow and passenger demand in the near future. This could assist in changing routes and timings in advance so that they suit the expected conditions. There are alternative methods that can also be used for problems like heuristic approaches including genetic algorithm or ant colony optimization and will be helpful when the problem space is too large for exact algorithms to solve efficiently. Predictive models can be used to predict the traffic pattern or passenger demands and use these predictions to optimize the route accordingly. I will test the applicability of the methodology to other metropolitan regions of the world to check the reliability of the developed method. Analyzing the results obtained in various cities, it will be possible to evaluate the effectiveness of the model in relation to other areas and assess the prospects for expanding its application.

References

Alamri, S., Adhinugraha, K., Allheeib, N. and Taniar, D. (2023). Gis analysis of adequate accessibility to public transportation in metropolitan areas, *ISPRS International Journal of Geo-Information* **12**(5): 180.

Biswas, A., Adhinugraha, K. and Taniar, D. (2023). Comparative gis analysis of public transport accessibility in metropolitan areas, *Computers* **12**(12): 260.

Brodje, A., Weber, R., Camre, D., Borup, O. and Porathe, T. (2015). Supporting voice communication between navigator and vts by visual solutions-exploring the use of the" route suggestion" functionality within vts.

Cardoso, P. J., Guerreiro, P., Pereira, J. A. and Veiga, R. J. (2018). A route planner supported on recommender systems suggestions: enhancing visits to cultural heritage places, *Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion*, pp. 144–151.

Ceder, A. and Teh, C. S. (2010). Comparing public transport connectivity measures of major new zealand cities, *Transportation Research Record* **2143**(1): 24–33.

Ingvardson, J. B. and Nielsen, O. A. (2019). The relationship between norms, satisfaction and public transport use: A comparison across six european cities using structural equation modelling, *Transportation research part A: policy and practice* **126**: 37–57.

Lerin, P. M., Yamamoto, D., Takahashi, N. et al. (2012). Mouse-sensitive following path suggestion for drawing travel routes in web map systems, *Journal of Geographic Information System* **4**(05): 393.

Oliveira, M. V. R. and Garcia, I. C. (2021). Performance evaluation of route suggestion approaches for primary waste collection, *2021 17th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, IEEE, pp. 181–188.

Osman, R., Ira, V. and Trojan, J. (2020). A tale of two cities: The comparative chrono-urbanism of brno and bratislava public transport systems, *Moravian Geographical Reports* **28**(4): 269–282.

Perveen, S., Yigitcanlar, T., Kamruzzaman, M. and Agdas, D. (2020). How can transport impacts of urban growth be modelled? an approach to consider spatial and temporal scales, *Sustainable Cities and Society* **55**: 102031.

Rajarajeswari, S., Reddy, S. R. P. and Venkatesh, A. (2013). Improving real-time route suggestions in automotive navigation systems using vehicle cluster behavior, *International Journal of Computer Applications* **76**(5).

REDDY, C. R. K. and HARIKA, M. (2024). An efficient recommendation and suggestion system for travel route using places of interest implementation, *IInternational Journal of Engineering and Techniques* .

Saghapour, T., Moridpour, S. and Thompson, R. G. (2016). Public transport accessibility in metropolitan areas: A new approach incorporating population density, *Journal of Transport Geography* **54**: 273–285.

Saif, M. A., Zefreh, M. M. and Torok, A. (2019). Public transport accessibility: A literature review, *Periodica Polytechnica Transportation Engineering* **47**(1): 36–43.

Tostes, A. I. J., Maia, G., de LP Duarte-Figueiredo, F. and Loureiro, A. A. (2015). Suggestion of routes for vehicles in vehicular networks using the multicommodity flow model, *2015 IEEE Symposium on Computers and Communication (ISCC)*, IEEE, pp. 451–456.

Yigitcanlar, T., Sipe, N., Evans, R. and Pitot, M. (2007). A gis-based land use and public transport accessibility indexing model, *Australian planner* **44**(3): 30–37.