The Wheels of Life

Written Report

Yin Junren Ryan 3i1 (30) (Leader)

Chow Xing Yu 3i1 (4)

Wong Yee Hern 3i1 (27)

Hwa Chong Institution
(High School)

Contents

1. Introduction

1.1 Description of idea

Singapore's transport network has been recognized globally as a leading model in providing a wide range of transport options for its people that are easily accessible from all parts of the island. Singapore has done well in establishing an interconnected bus and MRT network, but with burgeoning private car usage, it continues to face problems of traffic congestion on its roads especially during peak periods, leading to increased air pollution and reduced productivity. This raises the need for Singapore to search for alternative transport options for its people, and one of such solutions is cycling.

Cycling is becoming increasingly popular in Singapore, especially for short trips, as it is not only convenient, but also effective in bringing riders closer to nature and outdoors (Ministry of Transport, 2020). Considering Singapore's land constraints, planning the most effective cycling route is critical. Hence, our group decided to embark on this topic to devise the best cycling routes and positioning of bicycle parking spaces, such that it benefits people while using minimal resources, so as to provide the necessary facilities cyclists need and promote cycling as an alternative to vehicles.

1.2 Rationale

As depicted by the red indicators in Fig. 1 below, cycling paths in Singapore are currently clustered in the Jurong, Yishun, Tampines, Bedok and Pasir Ris areas. On the other hand, planning areas in the West of Singapore are not as sophisticated in cycling infrastructure compared to other parts of Singapore.

Fig. 1 Cycling Paths in Singapore by LTA

(Picture taken from https://www.lta.gov.sg/content/ltagov/en/map/cycle.html)

Due to the lack of connectivity of cycling routes in Singapore, not many people would choose to cycle as a means of transport in their daily lives. Furthermore, the West has always been regarded as the country's largest manufacturing hub. With the development of the 360-hectare Jurong Lake District (JLD) in the West, there will be four MRT lines serving the district by 2035 and new urban infrastructure built to enhance work productivity. This raises the need to provide more interconnected cycling paths and ensure a sufficient number of bicycle parking lots in the district so that travel to work and MRT stations is made more convenient, while also encouraging people to live a healthy lifestyle by cycling (URA, 2019).

1.3 Scope of Study

This project will be focusing on the Bukit Batok and Bukit Panjang planning areas in the West district, as the number of cycling paths there is the fewest compared to other areas like Jurong East, as shown in Fig. 2, and the bicycle parking spaces there are also not evenly spread out, as shown in Fig. 3.

Fig. 2 Cycling paths in the West District, denoted by red solid lines

(Picture taken from https://www.lta.gov.sg/content/ltagov/en/map/cycle.html)

Fig. 3 Bicycle Racks in the West District

(Picture taken from https://data.gov.sg/dataset/lta-bicycle-rack)

1.4 Objectives

The objectives of the research are as follows:

1. To analyse data from past research papers on cyclists' preferences and come up with a model of the best cycling route in Singapore.
2. To ensure the connectivity of the cycling path to make travelling by bicycle more convenient for all users.
3. To optimise the positioning of bicycle parking lots such that none of them become overused, nor abandoned.

1.5 Research Problems

The main questions the research aims to answer are as follows:

1. How to optimise the environmental factors that cyclists find important in the proposed model of the cycling route?
2. How to determine the most efficient placement of cycling paths that meets cyclists' travelling needs while not taking up unnecessary space?
3. How to ensure the most efficient placement of bicycle parking lots such that they are sufficient enough to meet cyclists' needs, while not overly clustered together such that some become unused?

2. Literature Review

Shin and Kai (2018) embarked on a case study of the Woodlands Planning Area (WPA) in Singapore. A Geographic Information Systems (GIS)-based path planning framework was used to decide the location of cycling paths, taking into consideration the perspectives of three key groups of stakeholders, namely the members of the public, transport experts and government planners. A suitability map that combined the three stakeholders' preference maps in equal proportion was derived and extracted to the footpaths in the area, depicting locations in the area suitable for building cycling paths, in comparison to the existing paths. Through these methods of study, Shin and Kai (2018) found that existing cycling paths required big detours to be made as the paths did not reach a few of the destinations directly, and there was a need for more connectivity between MRT stations and amenities. They also came up with an improved model of the cycling path that would meet the needs of the three stakeholders.

Similarly, Koh and Wong (2013) asked members of the public to rate the level of importance of 11 environmental factors in affecting their willingness to select a cycling route, as well as draw on a map, the usual route taken from the station to his/her destination. For each route, the shortest possible route was identified, and points were awarded to each of them based on the 11 factors. For example, the Route Directness Indices (RDI) were calculated. After this, paired t-tests between points of the actual route and shortest route were derived to find out the reasons why a person would choose his/her route over the shorter one, and it was found that cyclists chose routes with lower degrees of slopes and numbers of steps, good scenery and low accident risks.

Oh and Jeong (2001) investigated the usefulness of the GIS - fuzzy set approach as compared with a crisp approach in the evaluation of housing and environmental conditions of a residential area in Seoul, Korea. First, four main objectives - safety, health, efficiency and comfort - were identified, and three specific factors were developed under each objective for evaluating the urban residential environment. A pairwise comparison method was used to judge the relative importance of the factors based on expert opinion, and a crisp approach was then used to evaluate the residential area based on criteria that awards two or three discrete values between 0 and 1 when conditions meet a certain threshold. Next, a fuzzy set approach was used where for the same evaluation criteria, membership functions were developed which give a degree of membership for each input value on a continuous scale. The mean suitability output values for each factor under each approach were taken, and their correlation coefficients were compared. It was concluded that the fuzzy approach determined suitability values under each criteria more precisely, and also displayed more diverse qualitative levels, whereas the crisp approach is more restrictive, proving fuzzy approach to be a more reliable approach in spatial analysis when multiple factors are involved.

There has also been a study into cycling routes that involves graph theory. Kapasny (2016) conducted a study on the shortest cycling route through famous San Francisco movie locations. After determining the places that screened the most famous movies, Google API was used to compute the cycling distance between any two locations, which formed the edges of the graph, and the locations were represented by nodes in the graph. The locations with biggest betweenness were compared, and the Traveling Salesman Problem (TSP) was applied to the dataset, producing a map of the shortest route to all the movie locations.

From these studies, the important factors influencing cyclists' choice of bicycle routes, mainly from the studies of Koh and Wong (2013), can be used in the project to determine the environmental factors the bicycle paths should possess. The approach of Shin and Kai (2018) can also be followed in plotting suitability maps that benefit the most number of people possible. With reference to the approach by Kapasny (2016), graph theory can be incorporated into our project to determine the shortest distance between places and where to position the paths such that the maximum number of places can be reached. The conclusions made by Oh and Jeong (2001) also provide insights into using the more precise fuzzy membership functions, instead of discrete values, to evaluate multiple criteria. In addition, the question of where to position bicycle parking lots has also not been investigated in these studies, and is thus one area this project can explore.

3 Study and Methodology

3.1 Terminology

1. Consistency Ratio

   - Consistency Index for the set of judgments divided by the Index for the corresponding random matrix. If the ratio exceeds 0.1, the set of judgments may be too inconsistent to be reliable. (Saaty, 1990)

2. Fuzzy Membership

   - A method that transforms the input data to a 0 to 1 scale based on the possibility of being a member of a specified set.

3. Graph Theory

   - Mathematical structures that model pairwise relations between objects. A graph is made up of vertices (also called nodes) which are connected by edges (also called lines).

4. Geographic Information Systems (GIS)-based planning

   - A framework that analyzes spatial location and organizes information into visualizations using maps and 3D scenes.

5. Multi - Criteria Decision Analysis (MCDA)

   - A process of determining the optimal alternative, considering multiple, conflicting and interactive criteria (Chen et al., 1992).

6. Paired t-tests
    - A statistical procedure used to determine whether the mean difference between two sets of observations is zero. Common applications of the paired sample t-test include case-control studies or repeated-measures designs.
7. Suitability Analysis
    - The search for locations that are characterized by a combination of certain properties. Often, the result is a suitability map showing the locations suitable for a specific use in the form of a thematic map.
8. Weighted Graph
    - A graph in which each edge is given a numerical weight.
9. Weighted Linear Combination
    - A process of calculating the total score of an alternative by taking the weighted average of the suitability values of all attributes (Drobne and Lisec, 2009).

3.2 Methods applied

The following is a brief outline of the steps in the research. Firstly, the four environmental factors cyclists find most important in their choice of cycling path were identified. With reference to such data of the significance level of each factor, the pairwise comparison method was used to derive a measurement of the relative weighting of each factor from a value of 0 to 1. The consistency ratio was checked to ensure the precision of our measurements.

Next, data for the four environmental conditions in the Bukit Batok and Bukit Panjang planning areas were retrieved from government data sources such as Datamall and data.gov.sg. The data was first visualized on a map in Tableau to ensure its accuracy before being imported into RStudio to be analysed. Criteria for each of the environmental factors was determined and fuzzy membership functions were developed using RStudio that standardize the suitability values between 0 and 1. Suitability maps were plotted, assessing how well environmental conditions meet the criteria.

A Weighted Linear Combination method was then used to combine the suitability maps for each factor into one complete map, and Graph Theory was applied to the final suitability map to determine the most efficient placement of cycling paths that not only fulfill the preferences of cyclists, but also take up minimum space in the planning areas.

Lastly, suitability analysis was conducted for the bicycle parking lots in the two planning areas in terms of proximity to surrounding facilities, and graph theory was applied to assess the connectivity of the network.

3.3 Results and Findings

3.3.1 Research Question 1

The first research question was how to optimize the environmental factors cyclists find important in our cycling route. First, from the paired sample t-tests derived by Koh and Wong (2013), the four top environmental factors generally perceived as the most important by cyclists in Singapore were identified, as shown in Fig. 4 below. Excluding comfort, which is difficult to measure as it is subjective to the cyclists' own tastes, and detour, which will be analysed with graph theory later on in the project, the three most important factors are slope / stairs, Accident Risk and Accessibility to Shops along the route. Considering the need to promote first and last mile transport in the West District, proximity to MRT stations was included as the final environmental factor.

Infrastructural compatibility factors	Actual (mean point)	Shortest (mean point)	Paired *t*-tests		
			t statistics	$P(T \leq t)$ 2 tail	Bonferroni test
Detour, F2	4.08	6.20	*−5.46*	*<0.001*	*S*
Road crossing delays, F3	6.72	6.72	–	–	–
Directional signs, F4	0.29	0.36	−0.32	0.75	NS
Comfort, F5	6.91	6.28	*4.68*	*<0.001*	*S*
Weather protection, F6	1.36	1.77	−1.21	0.24	NS
Stairs/slope, F7	8.76	8.50	1.93	0.07	NS
Accident risk, F8	8.05	8.52	*−2.34*	*0.03*	NS
Crowdedness, F9	8.74	9.64	*−2.74*	*0.01*	NS
Shops along route, F10	0.48	0.33	1.24	0.23	NS
Good scenery, F11	2.15	1.66	1.75	0.09	NS

Italic represents significance at 95% confidence level.

Fig. 4 Paired sample t-tests for comparison of infrastructural compatibility factors (Koh and Wong, 2013)

The Pairwise Comparison method was used to rank the four environmental factors according to level of importance, between values of zero and one. The calculation of the relative weighting of each factor was necessary when combining the suitability maps of each factor later in the project. First, each factor was given a score based on its relative importance to the other three factors, and a relative importance matrix was plotted. Afterwards, each cell value was divided by the sum of its column, producing the corresponding importance for each factor. Next, normalization was carried out to determine each factor's weight with a range of 0 to 1.

F1 - Proximity to Shops

F2 - Proximity to Major Roads

F3 - Proximity to MRT / LRT stations

F4 - Gentle Slopes

	9	7	5	3	1	3	5	7	9	
F1									✓	**F2**
F1								✓		**F3**
F1				✓						**F4**
F2				✓						**F3**
F2	✓									**F4**
F3		✓								**F4**

1 - Equal

3 - Slightly favours

5 - Strongly favours

7 - Very strongly favours

9 - Extremely favours

Fig. 5 Scoring the factors by comparing relative importance

	F1	F2	F3	F4
F1	1	1/9	1/7	3
F2	9	1	3	9
F3	7	1/3	1	7
F4	1/3	1/9	1/7	1
Sum	17.3333	1.5556	4.2857	20.0000

Fig. 6 Relative Importance Matrix

	F1	F2	F3	F4	Weight
F1	0.0577	0.0714	0.0333	0.1500	0.0781
F2	0.5192	0.6429	0.7000	0.4500	0.5780
F3	0.4038	0.2143	0.2333	0.3500	0.3003
F4	0.0192	0.0714	0.0333	0.0500	0.0435

Fig. 7 Normalization and Weight Determination

Next, before plotting the fuzzy membership functions, suitability values and their respective criteria were generated for each of the four environmental factors. For the percentage slope factor, terrain with a percentage slope value of between 2% and 4% is comfortable for cyclists to travel on (American Association of State Highway and Transportation Officials, 2018). For proximity to major roads, research was conducted on guidelines for cycling paths, and it was found that a roadside verge ranging from 3m to 5m is necessary to be left between a development and a road, for tree planting to sustain a pervasive sense of greenery and the safety of cyclers (Urban Redevelopment Authority, 2018). As for the accessibility to MRT Stations and Shops factors, guidelines for the maximum walking distance to bus and LRT services were stated to be 400m, which approximates to a 5 minutes walk (Land Transport Authority, 2019). As such a distance of more than 400m was awarded 0 points for suitability, and a distance of less than 80m, a 1 minute walk, was considered to be the optimum for the placement of the cycling route.

Factor	**Minimum (0)**	**Intermediate (0.5)**	**Optimum (1)**
Amount of slope (%)	> 4	2 - 4	< 2
Proximity to major roads (m)	< 3	3 - 5	> 5
Accessibility to MRT stations (m)	> 400	80 - 400	< 80
Accessibility to shops (m)			

Fig. 8 Criteria for allocation of suitability values

After determining the criteria for allocation of suitability values, fuzzy membership functions were plotted using RStudio's "FuzzyNumbers" package, where α represents the degree of membership, in other words the suitability value, and x represents the input value, in other words the quantitative measure of the environmental factors.

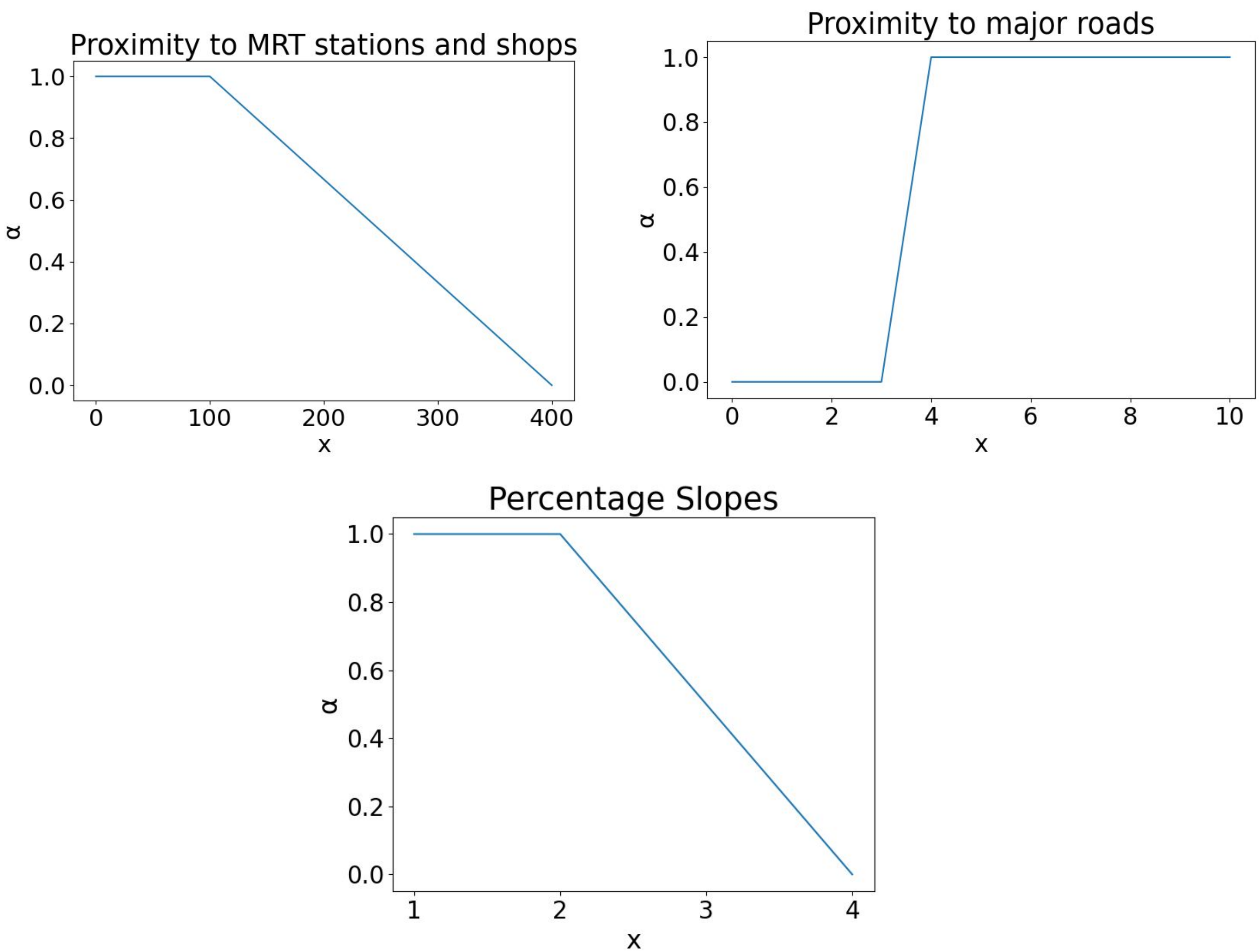


Fig. 9 Fuzzy Membership Functions for the four environmental factors

A dataset of the footpaths in Singapore was first retrieved from Datamall, and serve as the model for which suitability values will be calculated for. To obtain the suitability values for the footpath, data must be obtained for the respective environmental factors. This was done through API calls from Datamall for the Major Roads dataset and Shops Dataset, as well as from data.world for the locations of MRT and LRTs. The elevation values in the slopes dataset was obtained by interpolating data from a SRTM Digital Elevation Model using Python libraries “scipy” and “rasterio”, while the distance value was calculated with EPSG code 3414, a geographical coordinate system with meters as units.

The data that was retrieved was first visualised using Tableau Desktop, as shown in Fig. 10 and Fig. 11 below, to check its level of detail and accuracy, and also to filter out the data of other planning areas in Singapore using the lasso function, such that the data left was for only the Bukit Batok and Bukit Panjang planning areas. The data was then transferred to RStudio for analysis.

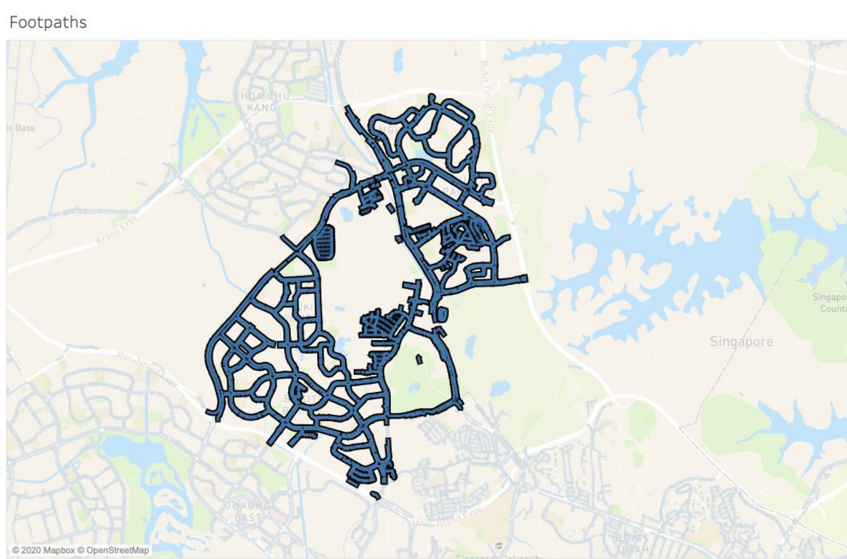
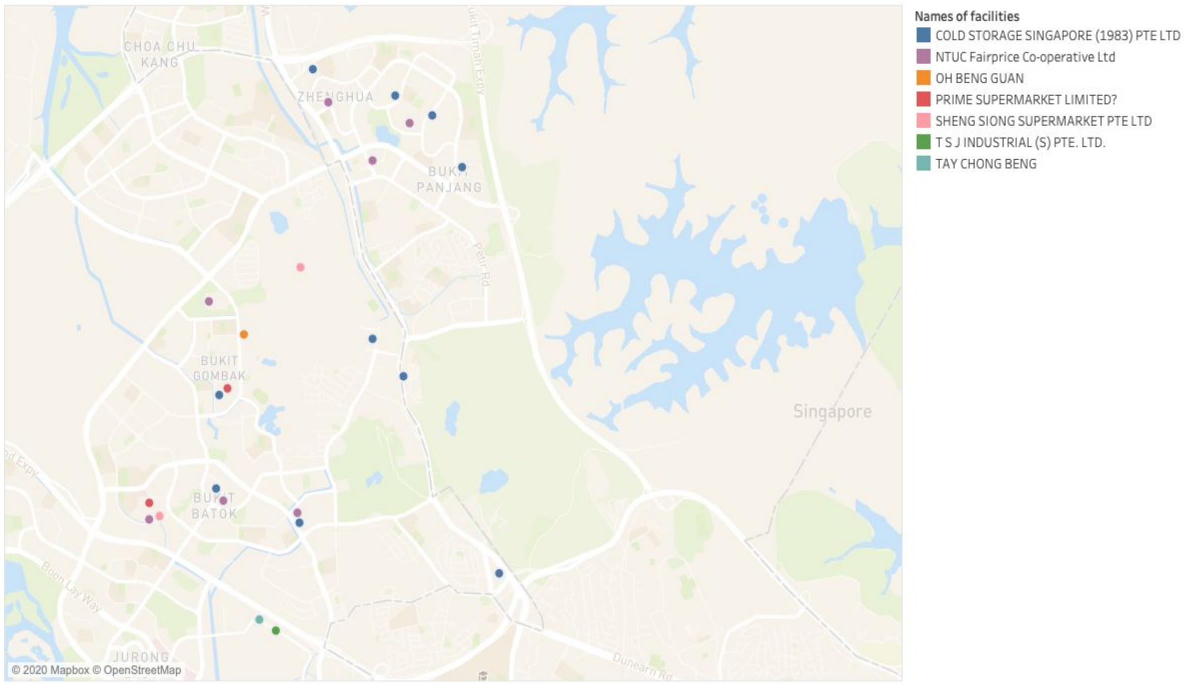
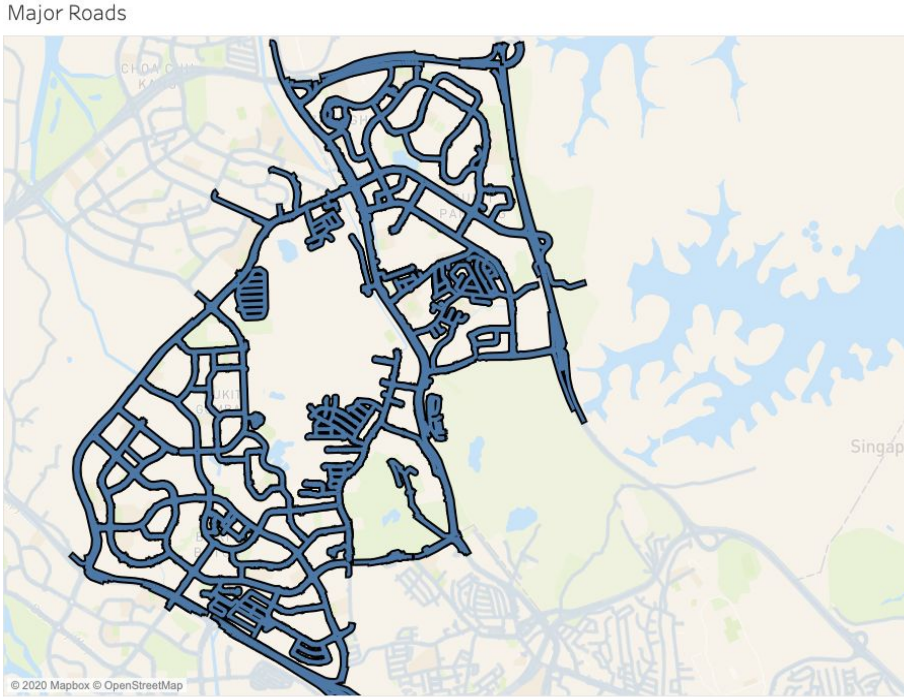


Fig. 10 Visualization of Footpaths data

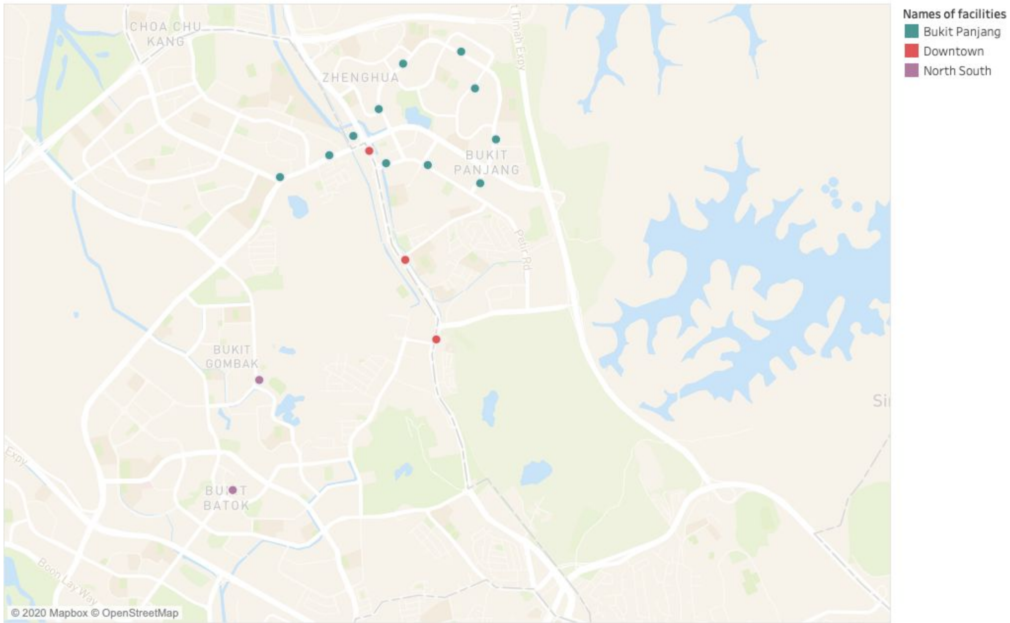
Shops



Major Roads



MRT and LRT Stations



Percentage Slope

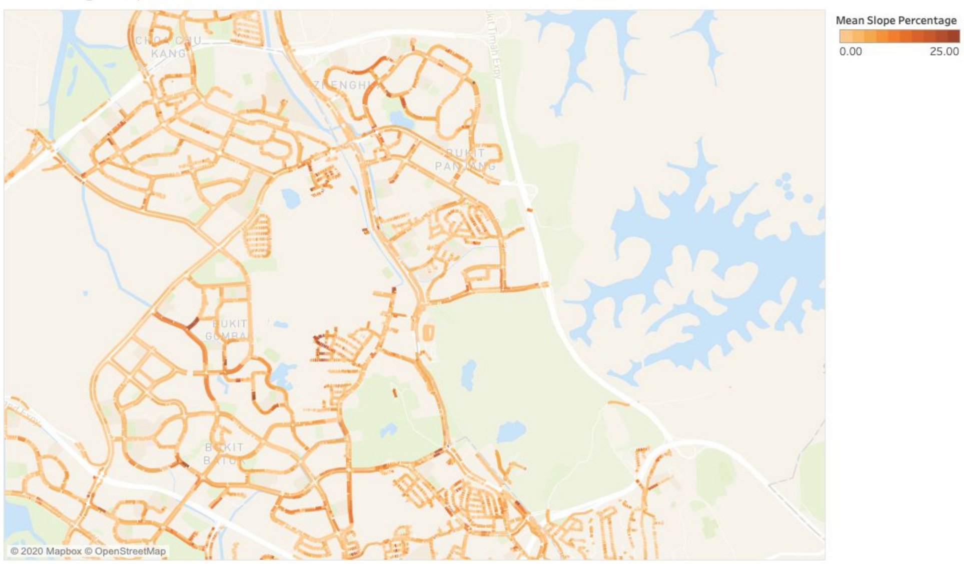
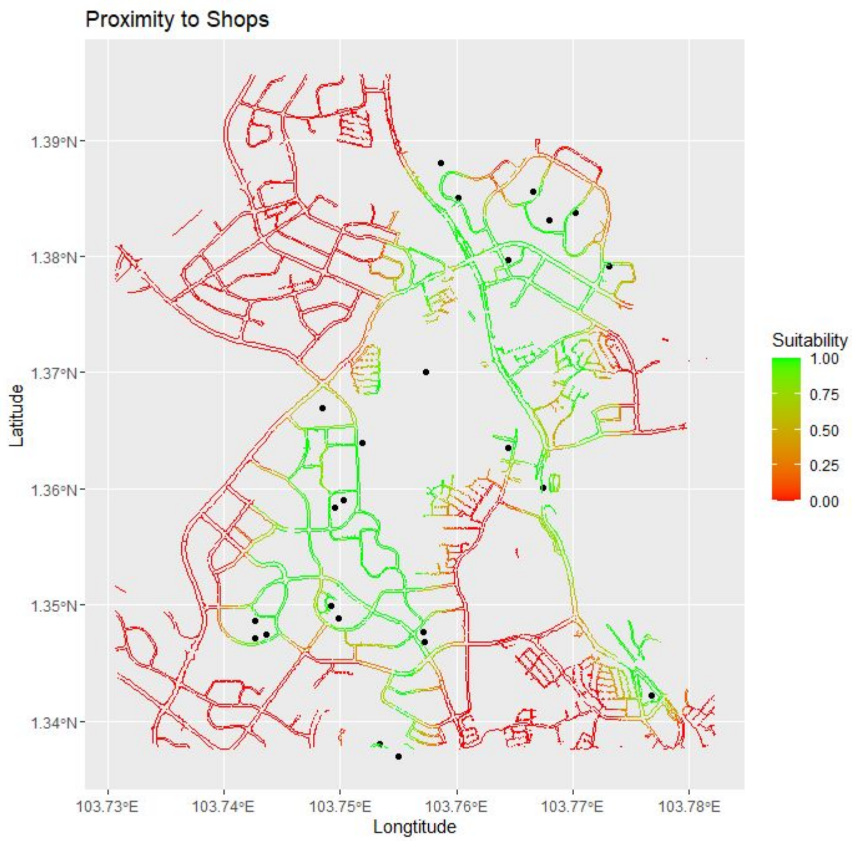


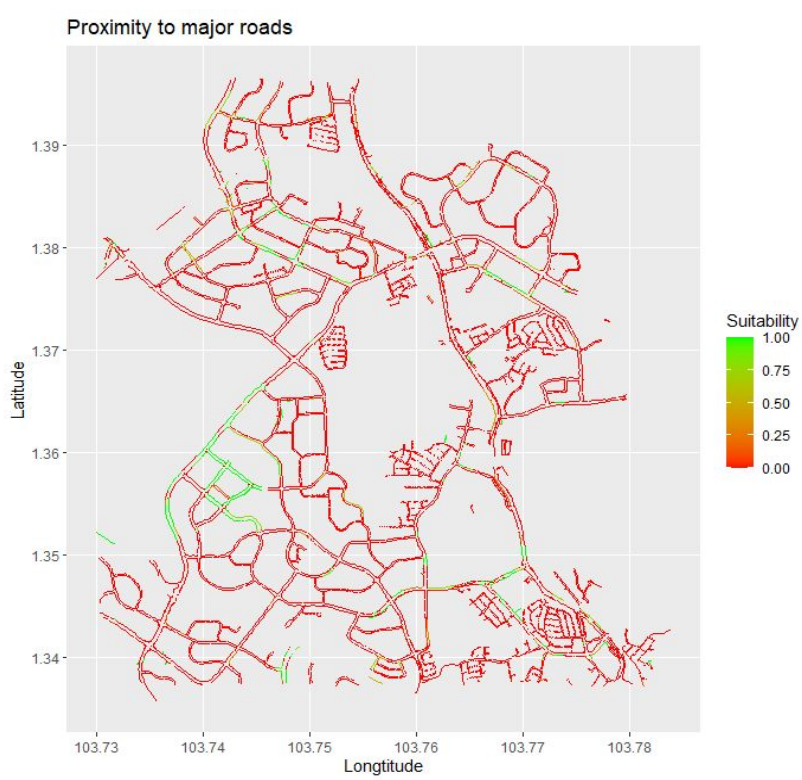
Fig. 11 Visualization of data for four environmental factors

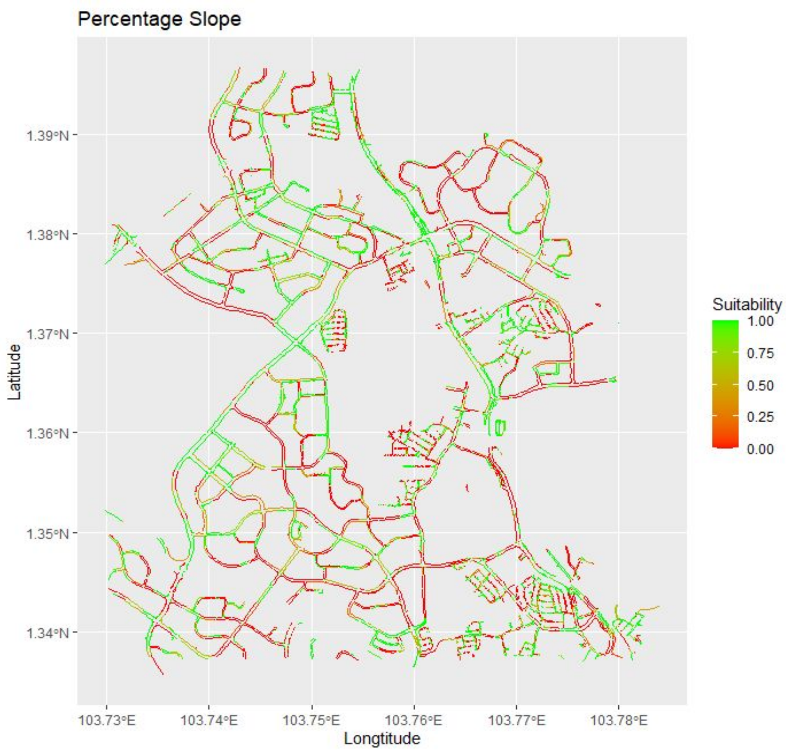
For the following descriptions, some terms are defined as follows. Footpath refers to the entire footpath dataset as a whole, a footpath segment refers to the separate linestrings making up the dataset, and points on the segment refer to the data points that make up each linestring.

Data for two of the environmental factors, locations of shops and locations of MRTs and LRTs, were imported into RStudio. The shortest distance between each point of the footpath and any one of the shops was calculated, before the mean value was taken for all the points on each segment to represent the mean proximity of the footpath segment to a shop. The mean values were inputted into the fuzzy membership function for shops and the respective suitability values were calculated. This was also done for the MRTs and LRTs. With the suitability values for each segment of the footpath on a range of values from 0 to 1, suitability maps were plotted for each of the two factors. A sample of the code for shops is shown in Appendix A.

For the other two factors, Proximity to Major Roads and Percentage Slope, Python was used instead to calculate the suitability values for each segment of the footpath, as it was deemed to be better at handling data calculation of large datasets than R. The shortest distance between the major roads and the footpath was taken, but instead of taking the mean value of all the points on each footpath segment, the point with the shortest proximity to a major road was considered and taken as the proximity of the footpath segment, because a major road is considered a constraint and if one point on the footpath segment does not fulfill the proximity criteria, the entire segment should be affected. The gradient of the slope was converted into percentage slope, and the slope value for each footpath segment is taken to be the mean slope values of the points on the segment. These values were passed through the respective fuzzy membership functions for Major Roads and Slope. The resulting suitability maps are shown in Fig. 12.


Proximity to Shops
Latitude
1.39°N
1.38°N
1.37°N
1.36°N
1.35°N
1.34°N
103.73°E
103.74°E
103.75°E
103.76°E
103.77°E
103.78°E
Longtitude
Suitability
1.00
0.75
0.50
0.25
0.00


Proximity to major roads
Latitude
1.39
1.38
1.37
1.36
1.35
1.34
103.73
103.74
103.75
103.76
103.77
103.78
Longtitude
Suitability
1.00
0.75
0.50
0.25
0.00



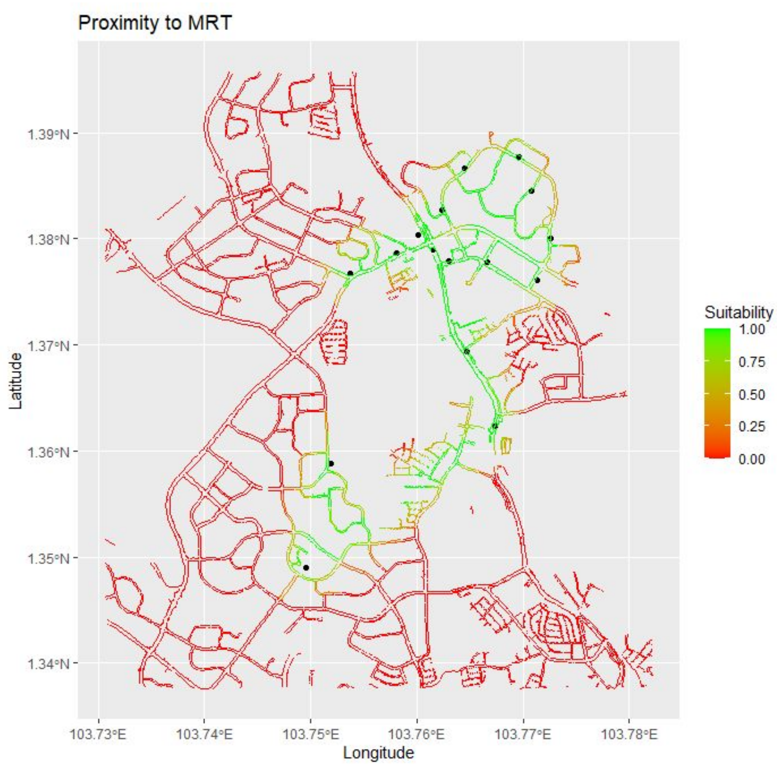


Fig. 12 Suitability maps for four environmental factors

The suitability maps for the four factors were combined into a complete map using the Weighted Linear Combination method with Python, following the formula below for calculation:

$$C = \sum_{i=1}^{n} F_i W_i$$

Where C refers to the final combined value, n refers to the number of factors considered, F_i refers to the suitability values of factors and W_i the weights of the factors respectively.

The Weighted Linear Combination was chosen to be the ideal method to combine suitability values for the four factors as a weighted average, because the factors were ensured to be mutually preference independent of each other as they address different areas of concerns of cyclists, and thus the suitability value of one of the factors would not affect that of another factor. The difference in t statistics from the paired t-tests conducted by Koh and Wong (2013) had also shown that the factors influence a cyclist's choice to use a cycling path to different degrees. Thus, this met the additivity assumption underlying the use of a Weighted Linear Combination as pointed out by Malczewski (2000), allowing the Weighted Linear Combination method to give valid results. The final suitability map is shown in Fig. 13, after the Weighted Linear Combination was applied to the four environmental factors.

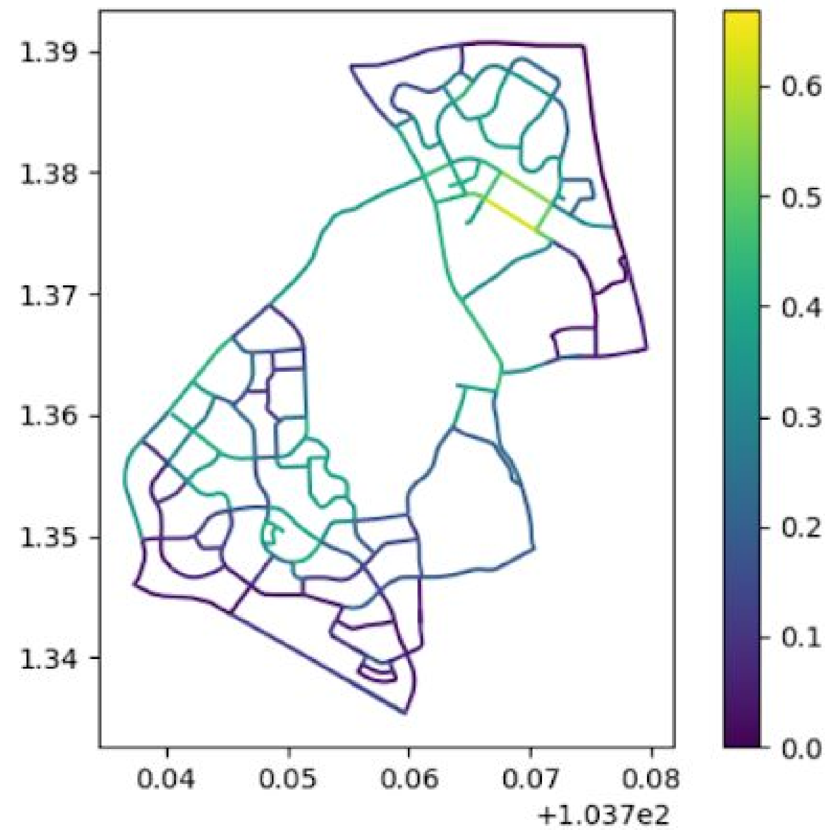


Fig. 13 Combined Suitability Map of four environmental factors

3.3.2 Research Question 2

The second research question of our project is to determine the most efficient placement of cycling paths that meets cyclists' travelling needs while not taking up unnecessary space. For graph theory to be applied to answer this question, Python was first used to plot the edges of the graph, which are the footpath segments, as well as the nodes of the graph, which are the intersection points of the footpaths, as shown in Fig. 14.

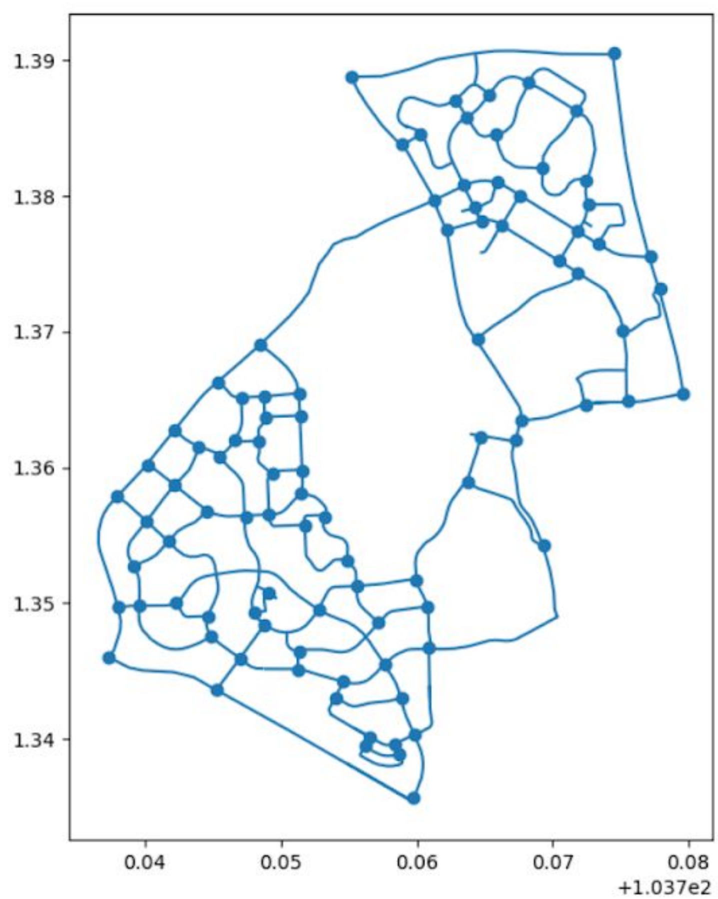


Fig. 14 Footpath network represented as a graph

The graph of the footpath network was overlaid with the combined suitability map produced by the Weighted Linear Combination method. Points were then given to each node on the graph based on the overall suitability value of the footpath segment the node is located at.

Next, the dataset of housing areas was taken from Data.gov.sg where housing areas in the area of focus were filtered out using Python. The dataset of facilities, namely Community Institutions, Healthcare facilities, and Sports and Recreational facilities was obtained and filtered similarly. Both datasets were overlaid onto the graph of the footpath network, where each of the housing areas were highlighted in red and facilities highlighted in orange as shown in Fig. 15.

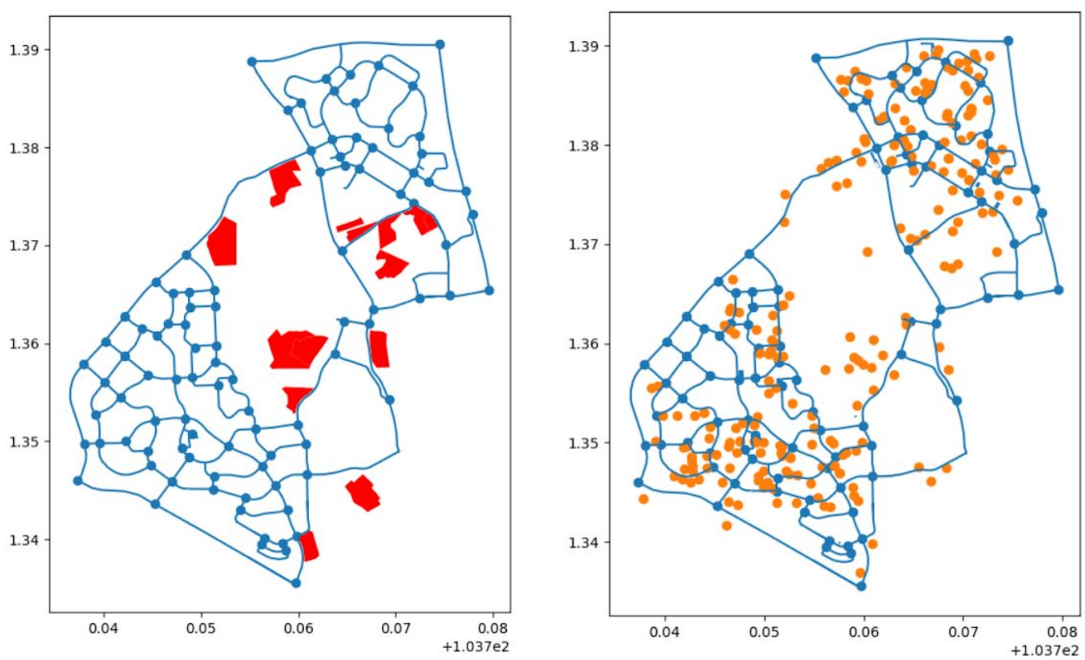


Fig. 15 Visualization of housing area dataset (left)

and facilities dataset (right)

For the housing dataset, a nearest node was chosen for each housing area where a score was given respective to the geometrical area of the housing. Similarly, for the facilities dataset, a nearest node was chosen for each of the facilities.

Then, a shortest path was generated from each of M housing areas to each of N facilities using Dijkstra's algorithm in Python' networkx module. As a result, $M \times N$ paths were generated. They were then layered together to obtain a frequency graph indicating the frequency of each edge being used as the shortest paths as shown in Fig. 16.

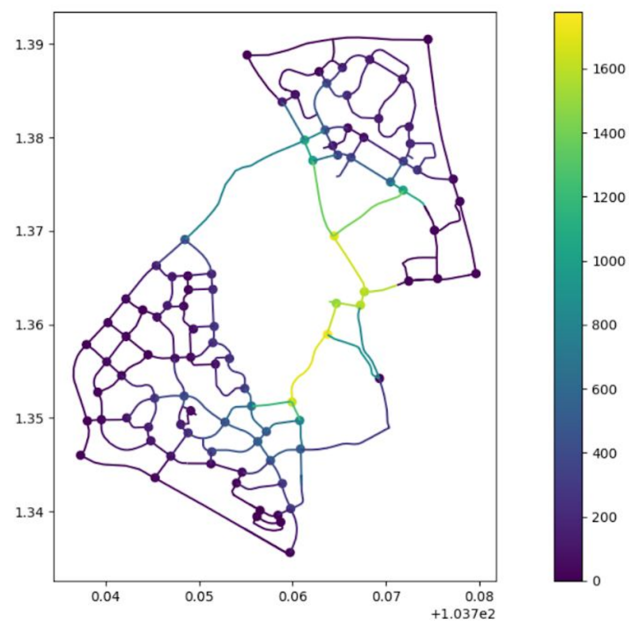


Fig. 16 Frequency of each edge being used as shortest path

The value of each node and edge was then multiplied by their corresponding values in the combined suitability map to form a new graph. Then, a map of the most efficient placement of cycling paths was obtained by filtering the graph with its median value. Hence, a graph was obtained as shown in Fig. 17. The code used for applying the shortest path algorithm, computing the final values of nodes and edges, and filtering the graph with its median value is shown in Appendix B.

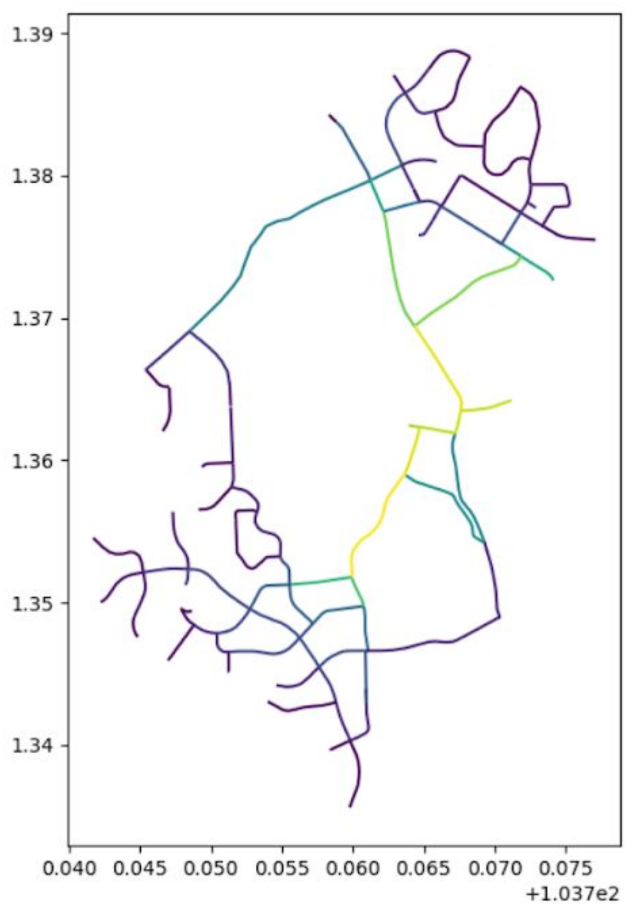


Fig. 17 Graph of most efficient placement of Cycling Paths

The graph was then superimposed onto the actual map of Bukit Panjang and Bukit Batok, and compared with the existing park connectors in the planning areas, as shown in Fig. 18. The proposed cycling paths from our study are outlined in red, while the existing park connectors are outlined in green.

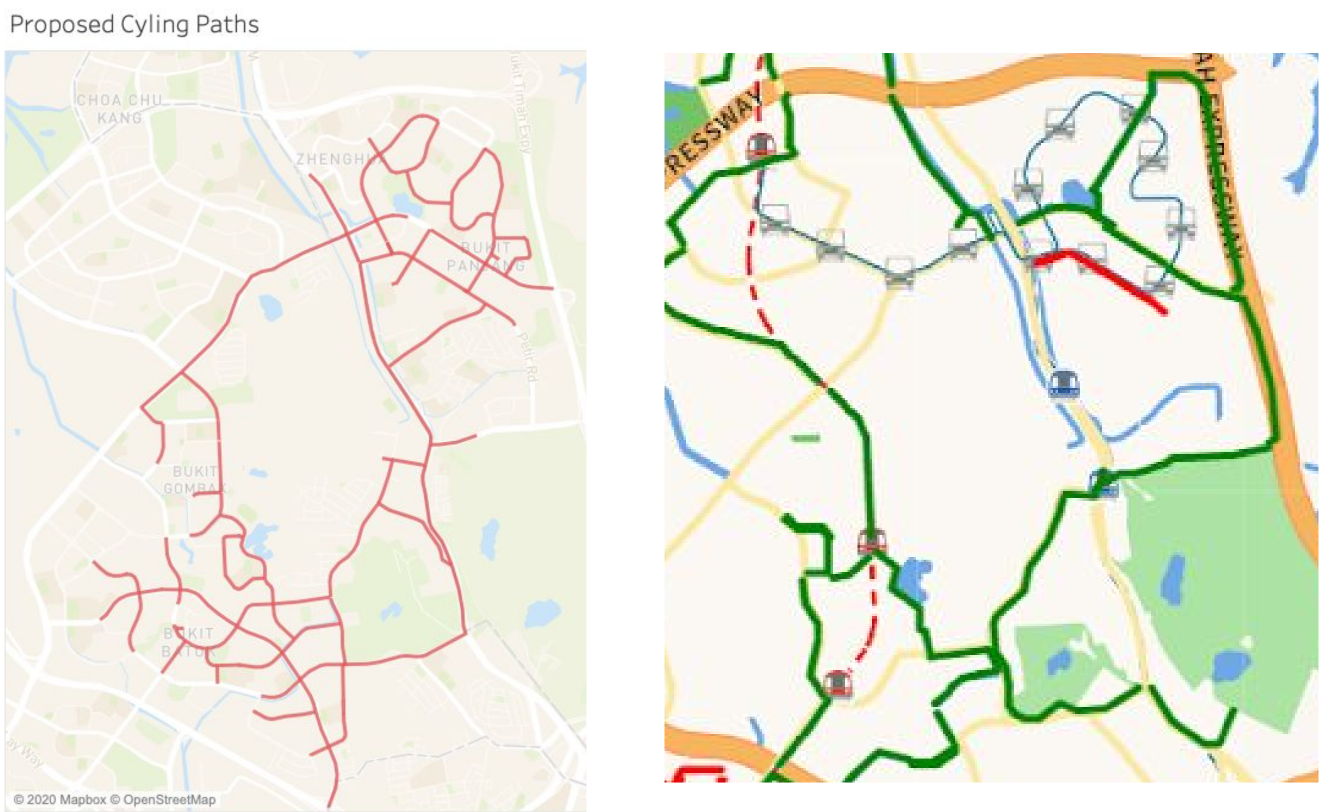


Fig. 18 Proposed model of cycling paths (left)

and existing park connectors in the planning areas (right)

3.3.3 Research Question 3

RStudio was used for determining the most efficient placement of bicycle parking lots in the planning areas. The bicycle parking facilities we focused on were the yellow boxes within the planning area because they are increasingly being introduced to more locations islandwide to address indiscriminate parking caused by bike - sharing schemes (Urban Redevelopment Authority, 2018).

The dataset of Bicycle Parking lots was first taken from Land Transport DataMall. The coordinates of the yellow boxes were then filtered out from the other bicycle parking facilities as a csv file, which was plotted and analysed in RStudio. The map of yellow boxes in the two planning areas plotted in RStudio is shown in Fig. 19.

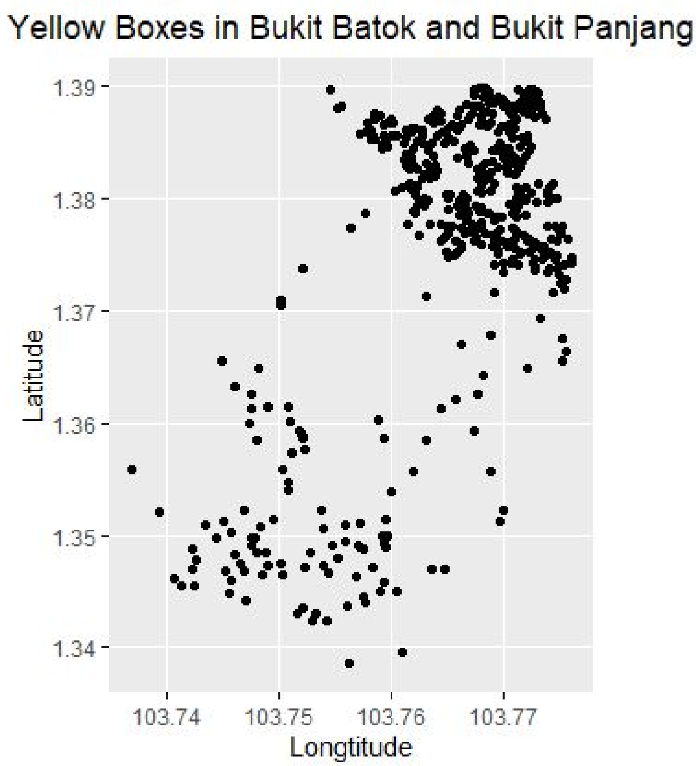
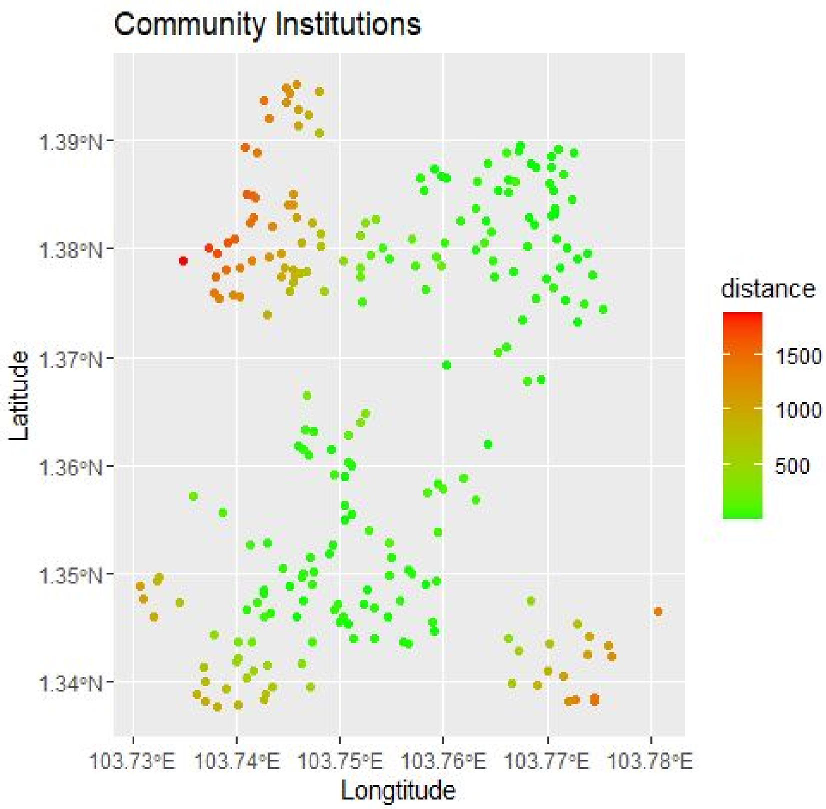


Fig. 19 Visualisation of yellow boxes

Firstly, the suitability of placement of bicycle parking lots in the two planning areas was analysed, in relation to surrounding facilities. The facilities taken into consideration were the same as those used in graph theory analysis for Research Question 2, and the shortest distance between each facility and any bicycle parking lot was calculated, before being represented as a suitability map based on shortest distance to a parking lot, as shown in Fig. 20.



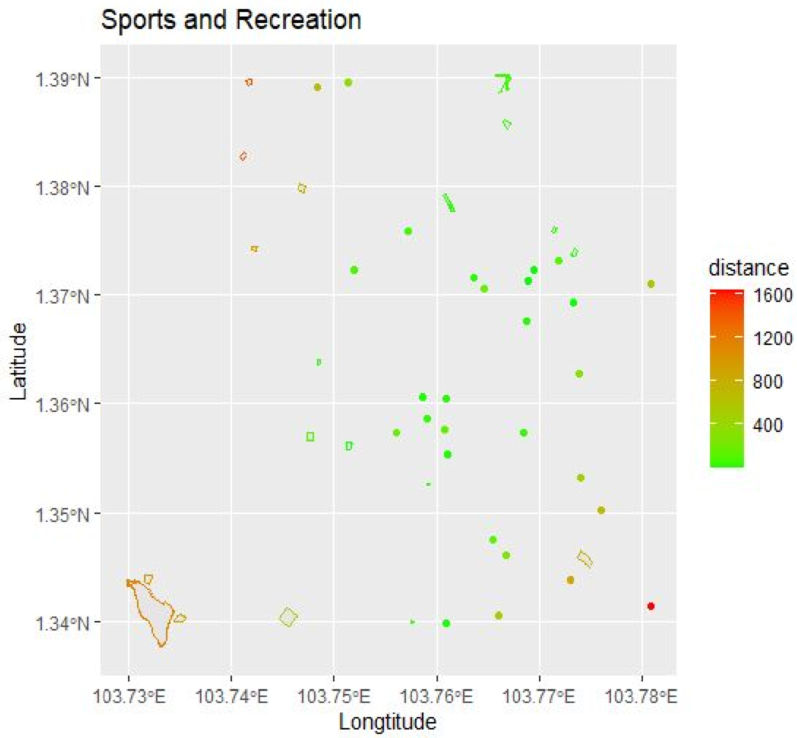


Fig. 20 Suitability maps of shortest distance of facilities to bicycle parking lots

Next, we conducted an analysis of the suitability of placement of each individual bicycle parking lot. The number of times each parking lot was used for comparison to a facility in terms of shortest distance was used to colour - code a map of the bicycle parking lots, as shown in Fig. 21, while the mean distance of each bicycle parking lot to the nearest facilities was also calculated and similarly visualised in a map, as shown in Fig. 22.

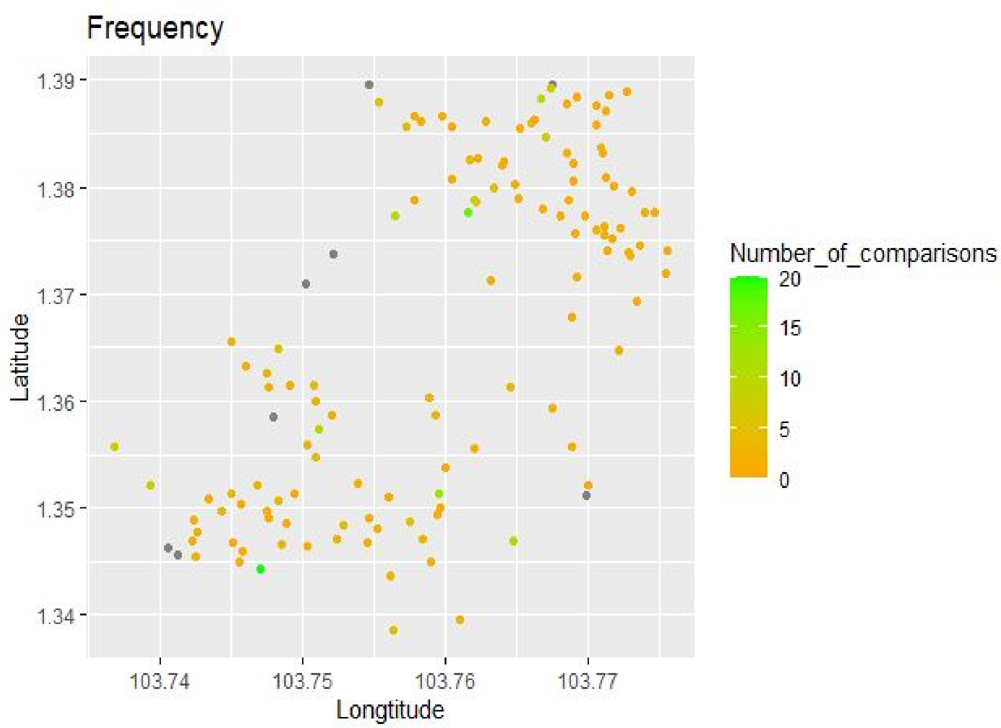


Fig. 21 Suitability map of frequency of comparisons of bicycle parking lots to facilities

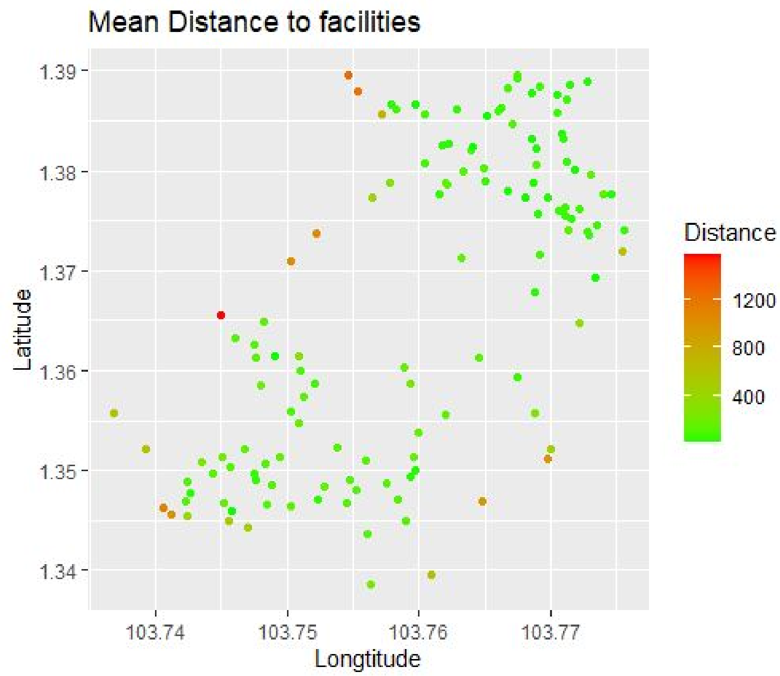


Fig. 22 Suitability map of mean distance of bicycle parking lots to nearest facilities

From the analysis of accessibility of facilities to bicycle parking lots in Fig. 22, almost all facilities in Bukit Batok and Bukit Panjang are within 500 metres of a bicycle parking lot, proving the positioning of bicycle parking lots convenient in proximity to the facilities in the area. As for the analysis of frequency of each bicycle parking lot being compared to a facility for the calculation of shortest distance, almost all of the bicycle parking lots in the Bukit Batok area were compared at least once to a nearby facility, while many of the bicycle parking lots in the Bukit Panjang area have found to be compared zero times with a nearby facility, showing the possibility of over - clustering in the Bukit Panjang area, which causes some of the parking lots to be redundant. From the analysis of mean distance of bicycle parking lots to the nearest facilities in Fig. 22, it can also be concluded that apart from the parking lots in the outskirts of the planning areas, almost all the bicycle parking lots are positioned close to a facility.

To further analyse the connectivity of bicycle parking lots in Bukit Batok, graph theory was applied, by taking the bicycle parking lots as the nodes, and connecting each bicycle parking lot to the four nearest neighbours to form a weighted graph. The actual distance of the shortest route between the bicycle parking lots was measured from geojson.io, and taken to be the weight of the graph. The resultant network, plotted in RStudio, is shown in Fig. 23.

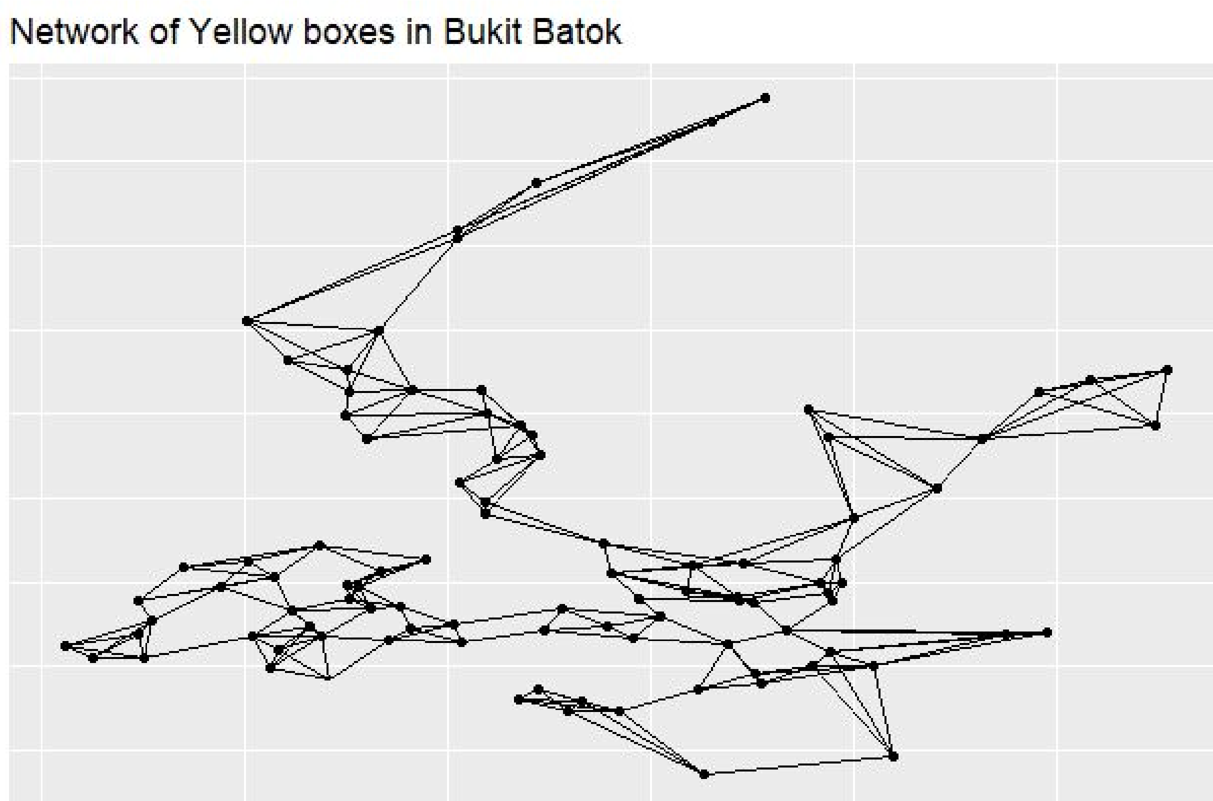


Fig. 23 Network of bicycle parking lots in Bukit Batok

Firstly, the local clustering coefficient of the graph was calculated, which measures local group cohesiveness as the fraction of connected neighbours of each vertex i, out of the total possible number of links that could exist between the neighbours of i. The formula for the local clustering coefficient is as shown.

$$C_i = \sum_{j,h} a_{ij} a_{jh} a_{ih} \frac{1}{k_i(k_i - 1)}$$

(Watts and Strogatz, 1998)

Where i represents the vertex being assessed, j and h represent the neighbouring vertices, $a_{ij} a_{jh} a_{ih}$ represents the elements of the adjacency matrix, and k_i represents the vertex degree. $k_i(k_i - 1)$ acts as a normalizing factor that normalizes the local clustering coefficient to a value between 0 and 1, for comparison with the weighted clustering coefficient. The average local clustering coefficient of the entire network is calculated as shown, by summing up the local clustering coefficient for all nodes in the network and dividing it by the total number of nodes.

$$C = \frac{1}{N} \sum_i c_i$$

$$= \frac{1}{96} (58.15714)$$

$$= 0.606$$

Next, the weighted clustering coefficient, which is a measure of the local cohesiveness that takes into account the importance of the structure on the basis of edge weights, was calculated. The formula for the weighted clustering coefficient is as shown.

$$C_i^w = \frac{1}{s_i(k_i - 1)} \sum_{j,h} \frac{(w_{ij} + w_{ih})}{2} a_{ij} a_{ih} a_{jh}$$

(Barrat, Barthelemy, Pastor-Satorras and Vespignani, 2004)

Where s_i represents the strength of the vertex, k_i represents the vertex degree, w_{ij} and w_{ih} represent the weights of the 2 adjacent edges of vertex i, and $a_{ij} a_{ih} a_{jh}$ represents the elements of the adjacency matrix. Similarly, the average weighted clustering coefficient was found by summing up the weighted clustering coefficient values for all nodes, and dividing it by the total number of nodes.

$$C^w = \frac{1}{N} \sum_i^w c_i^w$$

$$= \frac{1}{96} (56.66841)$$

$$= 0.590$$

The code used to plot the graph, as well as to compute the local and weighted clustering coefficients, is shown in Appendix C. The local and weighted clustering coefficient values for the first 30 nodes in the graph are shown in Appendix D.

Upon comparing the two coefficient values, it has been found that the average weighted clustering coefficient is less than the average local clustering coefficient of the network of bicycle parking lots. According to Barrat, Barthelemy, Pastor-Satorras and Vespignani, 2004, this result would indicate a network in which the topological clustering is generated by edges with lower weight, and in the case of bicycle parking lots, a network in which there is minimal distance between the bicycle parking lots.

Our findings from the generation of suitability maps based on distance from bicycle parking lots to facilities, and from the analysis of the network using graph theory, have shown that the bicycle parking lots in the Bukit Batok planning area are optimally positioned in relation to surrounding facilities, and are well - connected with some level of clustering.

4 Conclusion

4.1 Outcomes and Analysis

For Research Question 1, weightage of factors were proven acceptable and reliable by the use of pairwise comparison. Fuzzy membership functions were used to determine values of suitability based on different environmental factors, and the values were used to generate suitability maps. Weighted Linear Combination was applied to combine all four suitability maps and their suitability values while taking into account the importance of each factor.

For Research Question 2, graph theory was applied to the combined suitability map. Points were generated to each node on the graph based on the overall suitability value of the footpath segment the node is located at. External datasets like housing and facilities were used as indicators of potential demand for cycling and thus taken into consideration in the allocation of points to the nodes and edges. Based on the final values, the best cycling path was determined, that optimises both the environmental factors from the combined suitability map, and efficiency of meeting the demand for cycling.

For Research Question 3, suitability of placement of yellow boxes were analysed in relation to the same dataset used in Research Question 2. Shortest distance between each facility and bicycle parking lots were calculated, after that suitability maps were generated. Suitability of placement of each individual bicycle parking lot was also analysed. It can be concluded that apart from the parking lots in the outskirts of the planning areas, almost all the bicycle parking lots are positioned close to a facility. Graph theory was applied to find the actual distance of the shortest route between the yellow boxes. Local clustering coefficient and weighted clustering coefficient were calculated and compared. It has been found that the average weighted clustering coefficient is less than the average local clustering coefficient of the network of bicycle parking lots. This implies a network in which there is minimal distance between the bicycle parking lots.

4.2 Implications and Recommendations

This project takes on a GIS - Fuzzy set approach to conduct spatial analysis of each segment of the footpath network in the Bukit Batok and Bukit Panjang planning areas, in order to determine the optimal placement of cycling routes to best meet cyclists' needs while minimising the amount of space taken up. The use of the fuzzy logic approach in the GIS method by urban planners has been quite uncommon in spatial analysis (Oh and Jeong, 2011), as such our project proposes a different approach to analysing spatial data and assigning suitability values, that has been proven to give more accurate results.

In addition, our project uses an uncommon approach of Graph Theory in the GIS method to ensure that segments of the footpath with greatest suitability values are utilised in the final model, while also determining the optimal placement of cycling routes to maximise its usage in relation to nearby residences, and minimise the amount of space taken up in the footpath network. Graph theory has also been applied for the analysis of the most efficient placement of bicycle parking spaces in relation to surrounding facilities, after which the connectivity of the network is assessed to determine the level of clustering of bicycle parking lots. This approach can be considered by transport planners as an effective method to maximise placement of bicycle parking spaces based on proximity to surrounding facilities and ensure minimal clustering.

4.3 Possibility of project extension

An assumption that has been made in the course of this project is that people would only cycle for commuting purposes in order to travel from one place to another, and does not take into consideration the preferences of cyclists who cycle for recreational purposes instead. Recreational cyclists have been shown to be more likely to use certain types of bicycle facilities, while regular cyclists are less likely to, with frequent cyclists on average willing to travel greater distances than others (Larsen and El-Geneidy, 2013). As such, future applications of our proposed method can include the characteristics of cycling paths recreational cyclists prefer, and weigh them together with preferences of commuting cyclists, so as to produce a model of a cycling route that caters to a wider range of cyclist needs.

Due to the level of subjectivity involved in the setting of criteria for suitability values and the fuzzy membership functions, different criteria for the allocation of suitability values can be tested, so as to produce a more accurate suitability map with limited level of subjectivity. The Modified S-curve could also have been tried out instead of a linear membership function, as more parameters can be changed to yield different results, out of which the decision maker will be able to choose the most appropriate function for allocation of suitability values. (Peidro and Vasant, 2011).

5 References

American Association of State Highway and Transportation Officials. (2018). *A policy on geometric design of highways and streets 7th edition*. Washington, DC. Available: ftp://www.ahtd.state.ar.us/Outgoing/Roadway/TRC1805/AASHTO%20A%20Policy%20on%20Geometric%20Design%20of%20Highways%20and%20Streets%202018,%207th%20Edition.pdf

Barrat A., Barthelemy M., Pastor-Satorras R., Vespignani A. (2004). The architecture of complex weighted networks. Barcelona, Spain. Available: https://doi.org/10.1073/pnas.0400087101 .

Chua H.X. (2017). *Train Stations in Singapore*. [On-Line]. Available: https://data.world/hxchua/train-stations-in-singapore. Retrieved May 2, 2020.

Data.gov.sg. (2017). *List of Supermarket Licenses*. [On-Line]. Available: https://data.gov.sg/dataset/list-of-supermarket-licences. Retrieved May 16, 2020.

Datamall. (2020). *Geospatial Whole Island*. [On-Line]. Available: https://www.mytransport.sg/content/mytransport/home/dataMall/search_datasets.html?searchText=Geospatial. Retrieved May 16, 2020.

Datamall. (2020). *Kerbline*. [On-Line]. Available: https://www.mytransport.sg/content/mytransport/home/dataMall/search_datasets.html?searchText=kerbline. Retrieved May 16, 2020.

Data.gov.sg. (2020). *LTA Bicycle Rack*. [On-line].Available: https://data.gov.sg/dataset/lta-bicycle-rack. Retrieved 10 July, 2020.

Koh P. P., Wong Y. D. (2013). *Influence of infrastructural compatibility factors on walking and* *cycling route choices*. Singapore. Available: https://doi.org/10.1016/j.jenvp.2013.08.001.

Kapasny. (2016). *Data Science of Visiting Famous Movie Locations in San Francisco.* [On-Line]. Available: https://www.kdnuggets.com/2016/07/visiting-famous-movie-locations-san-francisco.html . Retrieved March 2, 2020.

Land Transport Authority. (2019). *Transport Impact Assessment Guidelines for Developments.* Singapore. Available: https://www.lta.gov.sg/content/dam/ltagov/industry_innovations/industry_matters/development_construction_resources/Street_Work_Proposals/Guidelines/TIA_Guidelines_WCP_Final_07Dec2018.pdf.

Land Transport Authority. (2020). *Cycling Routes*. [On-line]. Available: https://www.lta.gov.sg/content/ltagov/en/map/cycle.html. Retrieved 10 July, 2020.

Larsen, J., El-Geneidy, A. (2013). *A travel behaviour analysis of urban cycling facilities in Montreal Canada*. Canada. Available: http://dx.doi.org/10.1016/j.trd.2010.07.011.

Oh K., Jeong Y. (2001). *The usefulness of the GIS - fuzzy set approach in evaluating the urban residential environment*. Republic of Korea. Available: https://doi.org/10.1068/b2779.

Peidro D., Vasant P. (2010). *Transportation planning with modified S-curve membership functions using an interactive fuzzy multi-objective approach.* Spain. Available: https://doi.org/10.1016/j.asoc.2010.10.014.

Shin H. T., Kai C. (2018). *GIS-MCDA based cycling paths planning: a case study in Singapore.* Singapore. Available: https://doi.org/10.1016/j.apgeog.2018.03.007.

Saaty T. L. (1990). How to make a decision: The Analytic Hierarchy Process. *European journal of operational research*, 48, pg. 9-26. Available: https://doi.org/10.1016/0377-2217(90)90057-I

Urban Redevelopment Authority. (2018). *Walking and Cycling Design Guide*. Singapore. Available: https://www.ura.gov.sg/Corporate/Guidelines/-/media/BD725DB201DB496A93569C8072DD9FD0.ashx

Watts, D., Strogatz, S. (1998). Collective dynamics of 'small-world' networks. *Nature,* 393, pg.440–442. https://doi.org/10.1038/30918

6 Appendices

Appendix A

```
library("tidyverse")
library("ncdf4")
library("ggplot2")
library("sf")
footpaths<-st_read("Footpath.shp")
shops<-read.csv("Supermarket2.csv",TRUE,",")

shopsf<-shopsf %>%st_set_crs(4326)
footpaths<-st_transform(footpaths,4326)
foot2<-st_coordinates(footpaths)
foot2<-foot2 %>% filter(X>103.7307 & X<103.7822 & Y>1.3376 & Y<1.3957)

dist<-geosphere::dist2Line(line=scoords,p=foot2)

dist<-as.data.frame(dist)
dist_final<-cbind(dist,foot3)
dist_final2<-subset(dist_final, select = -c(lon,lat,L1))
colnames(dist_final2)<-c("distance","lon","lat","L1")
dist_final2<-st_as_sf(dist_final2,coords=c("lon","lat"))

dist_ag<-
  dist_final2 %>%
  group_by(L1) %>%
  summarise(distance=mean(distance),do_union = FALSE) %>%
  st_cast("MULTILINESTRING")
dist_ag<-st_as_sf(dist_ag)
diste<-dist_ag[["distance"]]

A3<-FuzzyNumber(0,0,80,400,
        left=function(x) x,
        right=function(x) 1-x
)

e<-data.frame()
iterations=6720
variables=2
for(j in 1:iterations){
  b<-as.numeric(diste[j])
  d<-evaluate(A3,b)
  e<-rbind(e,c(d))
}

dist_ag<-cbind(dist_ag,e)
colnames(dist_ag)<-c("L1","distance","Suitability","geometry")
dist_ag<-st_as_sf(dist_ag,coords=c("lon","lat"))
```

Appendix B

```
import json
import networkx as nx
import geopandas as gpd
import matplotlib.pyplot as plt

d2 = json.loads(open('graphnodes.geojson').read())
d3 = json.loads(open('graphedges.geojson').read())

gdf1 = gpd.read_file('combinedsuitability.geojson') #edges data
gdf2 = gpd.read_file('graphnodes.geojson') #nodes
gdf3 = gpd.read_file('graphedges.geojson') #edges

temp=[]
for x in range(len(gdf2)):
    temp.append(d2['features'][x]['properties']['edges'])

gdf2 = gdf2.assign(edges = temp)

temp=[]
for x in range(len(gdf3)):
    temp.append(d3['features'][x]['properties']['nodes'])

gdf3 = gdf3.assign(nodes = temp)
gdf1 = gdf1.assign(nodes = temp)

G = nx.MultiGraph()

for x in range(len(gdf1.nodes)):
    if len(gdf1.nodes[x]) == 2:
        G.add_edge(gdf1.nodes[x][0],gdf1.nodes[x][1], x, Combined = gdf1.Combined[x], Length =
gdf1.geometry[x].length)


gdfpoints = gpd.read_file('combineddataset.geojson')
gdfarea = gpd.read_file('housingdataset.geojson')
gdfbase = gpd.read_file('base.geojson')
geombase = gdfbase.geometry[0]

bounds = list(gdf2.total_bounds)
gdfpoints = gdfpoints.cx[bounds[0]:bounds[2], bounds[1]: bounds[3]]
gdfarea = gdfarea.cx[bounds[0]:bounds[2], bounds[1]: bounds[3]]
gdfpoints = gdfpoints.reset_index(drop=True)
gdfarea = gdfarea.reset_index(drop=True)
gdfpoints0 =gdfpoints
gdfarea0=gdfarea


l = []
```

```
for i, x in enumerate(gdfpoints.geometry):
    check = True
    if geombase.intersects(x):
        check = False
    if check:
        l.append(i)
gdfpoints = gdfpoints.drop(l)
gdfpoints = gdfpoints.reset_index(drop=True)


l = []
for i, x in enumerate(gdfarea.geometry):
    check = True
    if geombase.intersects(x):
        check = False
    if check:
        l.append(i)
gdfarea = gdfarea.drop(l)
gdfarea = gdfarea.reset_index(drop=True)


la = []
for i, x in enumerate(gdfarea.geometry):
    la.append(-1)
    for j, y in enumerate(gdf2.geometry):
        dist = x.distance(y)
        if la[i] == -1:
            la[i] = j
        else:
            prevdist = x.distance(gdf2.geometry[la[i]])
            if dist < prevdist:
                la[i] = j

lp = []
for i, x in enumerate(gdfpoints.geometry):
    lp.append(-1)
    for j, y in enumerate(gdf2.geometry):
        if j == 91:
            continue
        dist = x.distance(y)
        if lp[i] == -1:
            lp[i] = j
        else:
            prevdist = x.distance(gdf2.geometry[lp[i]])
            if dist < prevdist:
                lp[i] = j


templ = []
for x in la:
    for y in lp:
        if G.has_node(x) and G.has_node(y):
            if nx.has_path(G,x,y):
```

```
            templ += nx.shortest_path(G,x,y,weight='Length')
        else:
            print(x,y)

t=[]

for x in range(len(gdf2)):
    t.append(0)

for x in templ:
    t[x] += 1


n_num = []
for x in t:
    n_num.append(x)
n = len(n_num)
n_num.sort()

if n % 2 == 0:
    median1 = n_num[n//2]
    median2 = n_num[n//2 - 1]
    median = (median1 + median2)/2
else:
    median = n_num[n//2]


gdf2 = gdf2.assign(graph=t)

foo = []
for x in gdf3.nodes:
    cnt = 0
    count = 0
    for y in x:
        cnt = (cnt*count+t[y])/(count+1)
        count += 1
    foo.append(cnt)
gdf3 = gdf3.assign(graph = foo)

templ = gpd.GeoDataFrame()
for x in range(len(gdf3)):
    if gdf3.graph[x] > median:
        templ = templ.append(gdf3[x:x+1])


fig, ax = plt.subplots(sharex = True, sharey = True)

gdf4 = gdf3.append(gdf2)
gdf3.plot(ax=ax, column = 'graph')
templ.plot(column = 'graph')

plt.show()
nx.draw(G)
```

Appendix C

```
library("igraph")
library("tidyverse")
library("ncdf4")
library("ggplot2")
library("sf")

parking<-read.csv("Filtered bicycle parking 2.csv",TRUE,",")
parking2<-parking %>% filter(Longitude>103.7307 & Longitude<103.7822 & Latitude>1.3376 &
Latitude<1.3957)
parking<-as.data.frame(parking)
parking2<-st_as_sf(parking,coords=c("Longitude","Latitude"))
parking_viz<-as.data.frame(parking2)
parking_viz<-merge(parking_viz,links2,by.x=c("Index"),by.y=c("to"))

parking_viz <-parking_viz %>%
  mutate(long = unlist(map(parking_viz$geometry,1)),
         lat = unlist(map(parking_viz$geometry,2)))

ggplot() +
   geom_point(data=parking_viz,aes(x=long,y=lat),colour="black")+
   geom_segment(data=links3,aes(x=fro.x,xend = to.x, y=fro.y,yend = to.y))+
  ggtitle("Network of Yellow boxes in Bukit Batok")+
  theme(
    axis.text.x = element_blank(),
    axis.text.y = element_blank(),
    axis.ticks = element_blank(),
    axis.title.x = element_blank(),
    axis.title.y = element_blank())

links<-read.csv("Graph_Theory_Final.csv",stringsAsFactors=FALSE)

links2<-links[!duplicated(links[,c("to")]),]
nodes<-merge(parking_graph,links2,by.x=c("Index"),by.y=c("to"))
nodes<-nodes%>%subset(select=c(Index))

net<-graph_from_data_frame(d=links,vertices=new_nodes,directed=F)
simple_net<-simplify(net)
plot(net,edge.arrow.size=.4,vertex.label=NA)

trans<-transitivity(simple_net,type="weighted")
trans<-as.data.frame(trans)
sum_trans<-summarise(trans,trans=sum(trans))
average_trans<-summarise(trans,trans=mean(trans))

trans_lo<-transitivity(simple_net,type="local")
trans_lo<-as.data.frame(trans_lo)
sum_translo<-summarise(trans_lo,trans_lo=sum(trans_lo))
avg_translo<-summarise(trans_lo,trans_lo=mean(trans_lo))
```

Appendix D

Vertex Number	Local Clustering Coefficient	Weighted Clustering Coefficient
1	0.5000000	0.4388112
2	0.5333333	0.5291176
3	1.0000000	1.0000000
4	0.8333333	0.8040221
5	1.0000000	1.0000000
6	0.5000000	0.3841438
7	0.8333333	0.8158014
8	0.5333333	0.4796893
9	0.7000000	0.6707341
10	1.0000000	1.0000000
11	0.5333333	0.4900901
12	1.0000000	1.0000000
13	1.0000000	1.0000000
14	0.4285714	0.4248030
15	0.6000000	0.6264629
16	0.4000000	0.3954502
17	0.8333333	0.7922751
18	0.8333333	0.7930274
19	0.2666667	0.2807181
20	0.3000000	0.2127469
21	1.0000000	1.0000000
22	1.0000000	1.0000000
23	0.5000000	0.4146250
24	0.7000000	0.6813305

25	0.5333333	0.4824605
26	0.5333333	0.5172439
27	0.6000000	0.6493802
28	1.0000000	1.0000000
29	1.0000000	1.0000000
30	1.0000000	1.0000000