

Cycle Networks — Finding the Missing Links

Max Oblein
Supervised by Prof R. E. Wilson

March 30, 2021

Chapter 1

Introduction and background

Cycling is becoming increasingly more popular as a means of transport around large cities like Bristol [1]. This rise is due to a combination of factors, chiefly, increased environmental awareness, personal health, and the huge amount of road traffic meaning cycling is in some cases faster than driving for short journeys.

The National Travel Survey (NTS) [2] reports that cycling currently makes up 2% of all trips, where a trip is defined as a one-way course of travel with a single main purpose, and that the average length of these trips is 23 minutes. This seems a small proportion given that 38% of people surveyed own a bicycle. The third wave of the National Travel Attitude Survey (NTAS) [3] posed questions as to why people don't like to cycle. Their conclusions found that 60% of people agree that "it is too dangerous for me to cycle on the roads". Research suggests that an increased provision of cycling infrastructure is positively correlated with the proportion of people cycling within that city [4]. The key to increasing cycling rates in Bristol is thus to ensure cyclists feel safe and have segregated paths and lanes to use throughout their journeys. It is also worth noting that the covid-19 pandemic has already had a huge impact on the amount of people choosing cycling as a mode of transport. The Bicycle Association's (BA's) study into the growth of the UK cycling market during the pandemic found a 60% increase in bicycle sales since March 2020 [5]. This is largely due to the growing popularity of electric bicycles (E-bikes) the same study by the BA found that E-bike sales increased 92% between April and September 2020, when compared to 2019. Although, it is not just cyclists using segregated paths anymore. In recent years the growth of other micro-mobility methods, such as electric scooters, has increased flow on cycle paths in major cities. These factors all suggest it is worth investigating increased cycling infrastructure in large cities.

Currently, when planning road networks, lots of money is spent with engineering consultants to decide where to build new roads [6]. Cycling is much less common as a mode of transport than driving, so planning the cycle network has a significantly lower budget. This means that selecting where new cycle paths are built or lanes are painted is often an arbitrary decision made by the local authority. Therefore, this project aims to develop a mathematical framework to inform these decisions.

1.1 Previous work

There is some previous work in the area of planning cycle networks, and most studies focus on two areas: the effect infrastructure has on the number of cyclists and the evaluation and proposal of cycle schemes based on geographical factors. The first area motivates the need for cycling infrastructure [4]. The second is useful for the evaluation of routes suggested by this project based on urban factors such as the amount of retail or business land area at both origin and destination of the route [7]. The methods presented by Milakis et al. [8] give a comprehensive methodology for evaluating proposed cycle routes. They use a weighted scoring system to assign good scores to origin-destination (OD) pairs attached to important facilities e.g., universities. In the literature little work has been found in the area of assigning new cycle routes automatically using a mathematical approach.

The main concern when automatically proposing roads for new cycling infrastructure is deciding how best to load a road network with cycling demand. Two papers have been identified for their methods in loading street networks with cycling demand. One approach, formulated by Milakis et al. [8], approaches this problem with the use of large demand centres. The study is based in Athens and the idea is to draw centres around areas of large demand. The centres are chosen based on three main criteria: the land use (retail, offices, education or leisure), the OD matrices from the last two travel surveys in Athens, and the network centres detailed in the master plan of Athens. Milakis et al. identify eight such centres, shown in Figure 1.1 and they are used as the OD pairs from which to generate synthetic cycle journeys. This methodology is a good starting point for loading the network with demand, but a set of only eight nodes does not allow for realistic models of commuting in cities such as Bristol, where workers commute from and to a huge number of different locations. This methodology is also not automatic: it requires a survey of cyclists to decide which land use attributes give the largest demand to a centre, and creating and carrying out a comprehensive survey is costly. The survey used in this case is also only considering existing cyclist's behaviour and therefore cannot predict the impact of new infrastructure on cycling demand.

A second paper by Larsen [9] takes a more granular approach. The methodology begins by separating the study area (Montreal) into 300m grid squares; the demand within each of these squares is then derived using both observed cycling trips (OD data from Montreal travel survey) and potential cycling trips (car trips shorter than 2km). These two data sources give OD pairs and then standard shortest path algorithms are used to model the route taken between them. Flow intensity is then given by the proportion of routes passing through a grid square. This analytical methodology is then combined with a survey of cyclists as to which road they think should be prioritised for new infrastructure to yield a prioritisation index. The map of Montreal in Figure 1.2 with a grid overlay shows the prioritisation for new cycle paths suggested for Montreal. The idea is that any grid square highlighted red show be highly prioritised for new infrastructure.

One part of the Montreal study [9] that could be of particular interest to this project is their study of “Dangling Nodes”. These are defined as grid squares in which the cycle infrastructure ends. The study concluded that the presence of a dangling node in itself is not an indicator of a need

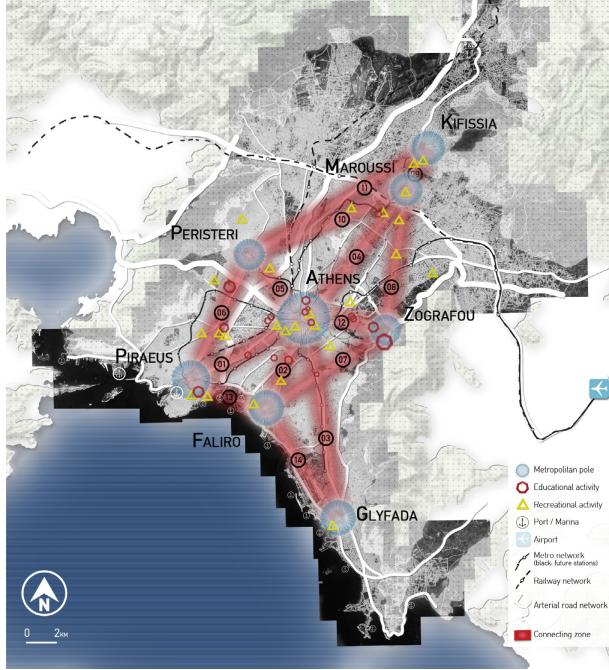


Figure 1.1: This plot shows the centres (highlighted in blue) identified in Athens by Miliakis et al. [8].

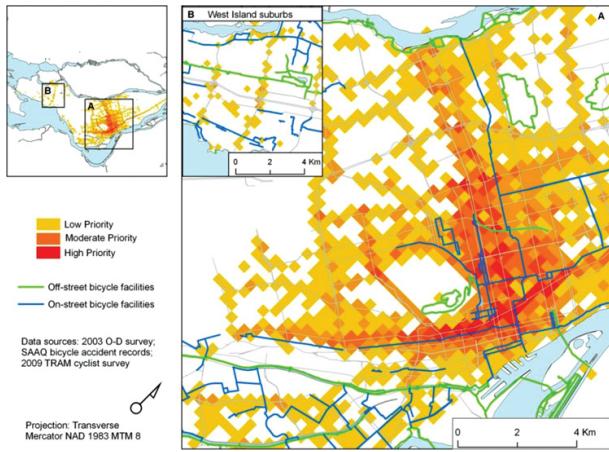


Figure 1.2: Map of Montreal with grid square overlay showing areas highlighted for prioritisation by Larsen et al. [9].

for new infrastructure. Instead they recommend looking at the amount of infrastructure needed to connect the network and performing a case-by-case assessment. The automatic approach proposed in this project may allow for some dangling nodes to be eliminated from consideration, if connecting them to the rest of the network yields no step change in connectivity.

Research by Mauttome et al. [10] applies formal optimisation techniques to the cycle network in some case study cities. The research proposes optimisation of total user and construction costs where both are proportional to distance and user costs are increased on edges without cycling infrastructure. The paper uses a heuristic approach to optimise for user cost on large scale networks, and compares against exact solutions on smaller artificial network examples. However their algorithm does not naturally penalise disconnections in the resulting network, therefore a methodology for identifying links that join up the cycle network is needed.

One tool that has been recommended to us by sustainable transport planners [11] is the the

Cycling Infrastructure Prioritisation Toolkit (CyIPT). This is a combination of smaller tools that aims to provide an interactive map of the UK, detailing areas for proposed new cycling infrastructure. The components of most interest are the propensity to cycle tool (PCT) [12] and the rapid cycleway prioritisation tool (RCPT) [13]. The PCT serves as the CyIPT's demand model, using 2011 census data to provide estimates of OD flows. The PCT also uses geographical factors such as 'hilliness' to reduce the propensity to cycle on any given route. Our key interest in the RCPT is its planned facility to compute cohesive networks, which represent more highly connected cycle networks. However, in the current RCPT documentation there is no formal mathematics presented as to how to find these cohesive networks.

Summary of studies

Study	Methodology
Athens [8]	Select centres throughout the city based on their land use e.g., university. The land uses have a cycling priority index based on an OD survey. Then load demand between centres.
Montreal [9]	Split city into 300m grid squares. Then load underlying road network using OD survey data. Prioritise grid squares based on OD demand and a cyclist survey.
Network Optimisation [10]	OD matrix derived from a 2009 household survey in Montevideo (Uruguay). The algorithm presented then tries to optimise for user cost given the length of edges in the road network.
CyIPT [12], [14]	Demand matrix comes from the PCT lower-level tool which uses 2011 census data for OD pairs along with geographical measures such as incline. These are used to assign a value to each edge as to how likely people are to cycle on it. The CyIPT then takes this demand and proposes schemes with high upside whilst minimising estimated construction cost.

1.2 Research questions

So far a gap has been identified in the area of cycle network design. That is, formal mathematical approaches to network design. All previous work uses primarily geographical and opinion polling to formulate a network upgrade strategy. This project will, through some key research questions, investigate the potential of a mathematical design approach. Three research questions have been identified for study:

- Can a simple model of propensity to cycle give a good approximation of cyclist route choice?
- Can we use a simple heuristic to inform cycle network upgrades in a given city?
- How close does the heuristic approach come to formal optimal network design?



Chapter 2 uses mapping data to create a network representation of Bristol, which is then used to test the impact of the proposed model of cyclists' propensity to cycle. The model is such that a cyclist perceives the length of streets without infrastructure to be longer by a factor related to their personal propensity to cycle. This chapter then tests this model on two case test cities to see whether cyclists with lower propensity to cycle tend to use more cycle paths in their journey.

Chapter 3 investigates the second research question. The key methodology begins with simulating cycling demand using empirical data. Once we can simulate cycling trips we can compute the flow of cyclists over every edge in the network and upgrade those edges with the largest flow counts. Then we investigate whether some of the heuristic's parameters have an impact on the output network for our case study city Bristol.

Finally in chapter 4 we investigate the final research question. We propose some test network cases with edge counts of approximately 20. Then an exhaustive search algorithm is applied to find the optimal cycle network design for our demand model within our upgrade budget. We then test this project's proposed heuristic on the same network with the same demand and budget to see how close to the optimal design it is.



Chapter 2

Data Sources and Initial Analysis

This chapter details the initial analysis of Bristol's street and cycle networks. Initially we discuss the potential data sources to be used and select OpenStreetMaps (OSM) as our primary data source. We then introduce a single parameter model for the perceived length of routes through Bristol's street network for a cyclist with a given propensity to cycle. Finally we show the impact that this propensity to cycle model has on some simple test statistics on the Bristol network. These results are then compared with an objectively better cycle network, the city of Amsterdam, where the network is more expansive and OSM data is more precise.

2.1 Data sources

To create the network needed for the analysis proposed in this project, multiple data sources were investigated to be converted into a network representation for analysis. Three such sources have been identified: Cyclestreets [15], CycleOSM [16], and OSM [17]. Cyclestreets offers a routing engine for cyclist route planning in the UK an example of which is shown in Figure 2.1. Their API allows users to query for cyclist route choices between any two locations in the UK. Cyclestreets give three route options which are: the shortest route, the quietest route, and a balanced route. These options are interesting to cyclists as they have the ability to choose a route based on their propensity to cycle. This project aims to solve the route planning issue with a simple model of a cyclist's propensity to cycle which will be described later in this chapter. This means the extra functionality offered in route planning to avoid roads and other aspects of the Bristol street network should be encapsulated in the cyclists propensity to cycle and so we don't need to use their routing engine. CycleOSM is essentially a condensed version of OSM with only the data required for cyclists present and contains lots of information about cycle parking, repair stations, cycle shops etc. A screenshot of the CycleOSM map of a section of Bristol is shown in Figure 2.2. The map is designed for cyclists to plan their own routes however this project aims to solve the routing problem using shortest route according to a heuristic based on propensity to cycle. This project aims to upgrade streets in a city with new cycling infrastructure, so a map based on purely existing infrastructure is not as useful as the raw OSM data for that city. Both Cyclestreets and CycleOSM are simply built from OSM data, a map from which can be seen in Figure 2.3, and hence it can be used to recreate any functionality they provide for a cyclist. All



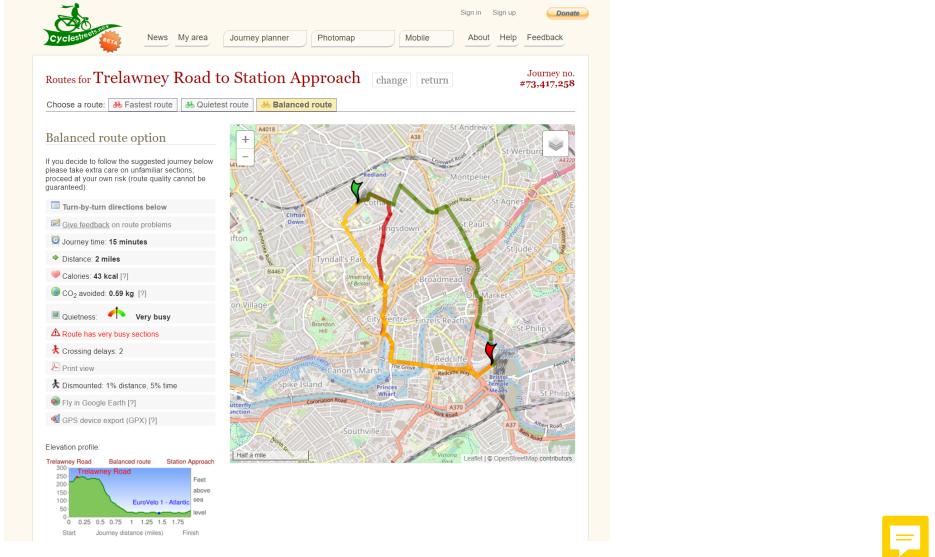


Figure 2.1: An example of the functionality provided by CycleStreets [15]. Shown is a journey planner from the student housing area in Redland to Bristol Temple Meads.

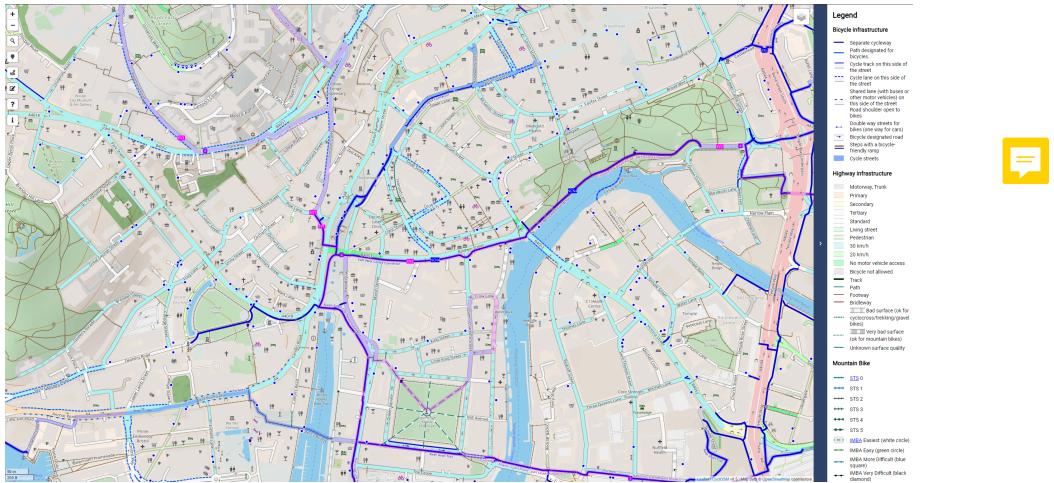


Figure 2.2: An example of the map provided by CycleOSM [16]. Shown is a section of Bristol zoomed in to a point at which all of the extra features of this mapping tool are seen. Blue dots represent bicycle parking and many other bicycle specific locations are highlighted.

the various map sources store data as xml files, a short extract from the xml file returned on an OSM query for Bristol's cycle network is shown in Figure 2.4.

The extract contains the data for only two cycle ways within the cycle network. Clearly these files are difficult work with using only the base python functionality, as in reality there are thousands of ways in the full Bristol road and cycle network. However, there exist an assortment of python toolboxes built specifically for OSM data to streamline the data collection portion of this task, so OSM has been identified as the primary data source for this project. A python package OSMnx [18] has been used for easy conversion of OSM data to a network topology. To obtain the data from OSM we must query the OSM overpass API. OSMnx streamlines the query process, although in the case of this project a custom query is built to obtain all the data

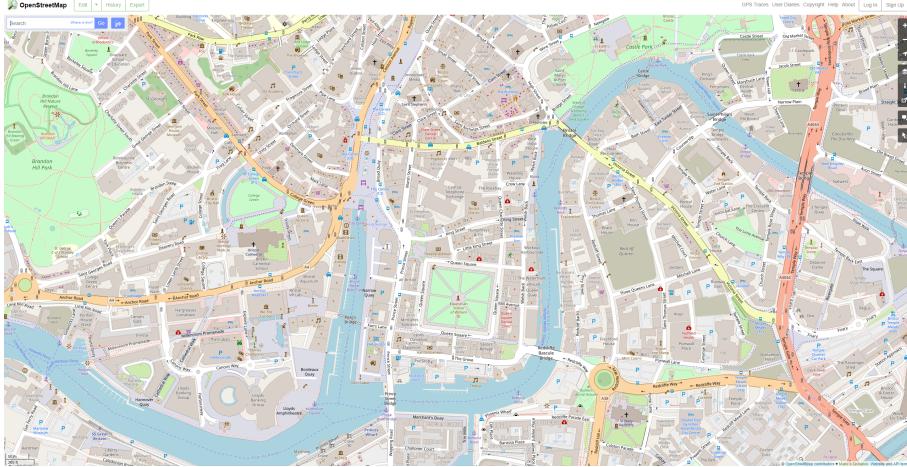


Figure 2.3: An example of the map provided by OSM [17]. Shown is a section of Bristol zoomed in to a point at which all of the extra features of this mapping tool are seen. This tool serves as the basis for the other two. This combined with a set of python packages available for it makes OSM the mapping tool of choice for this project.

Bristol Network	
Raw XML size	> 100 MB
Converted GraphML size	20.5 MB
Number of Nodes	23,856 (OSM nodes representing curvature have been removed)
Number of Edges	54,638
Total street length	1757 km
Total length of cycling infrastructure	160 km (Includes: painted lanes, segregated lanes and, off-road cycleways)



Table 2.1: Key statistics of the Bristol street network.

required to build the cycle network. The first step is to convert the OSM map of Bristol, our selected bounding geography, to a network. This is done by querying OSM’s Nominatim API to get the boundary set for ‘place = Bristol’. After this, the boundary is used, with a small buffer, to form a polygon within which to query the Overpass API for street data, with the option ‘bike’ selected. This yields a set of ‘ways’ (all roads and paths) that can be cycled on within the bounding geography, hence excluding motorways but adding bridleways and segregated cycle lanes to the Bristol street network. This can then be used to create a graph G whose edge set E represents the set of all ways and whose vertex set V represents junctions between ways and points at which ways gain or lose cycling infrastructure. The number of vertices in G will typically be much less than the number in the OSM data as it is not necessary to describe the curvature of each way in our study. The key network statistics for Bristol’s street network are shown in table 2.1.

2.2 Model for propensity to cycle

The first step in the analysis of Bristol’s road network is to identify the edges corresponding to existing cyclepaths. The OSM tags of interest are ‘Highway’, ‘Cycleway’ and ‘Bicycle’, although

Extract from Bristol cycle network xml file

```

<?xml version="1.0" encoding="UTF-8"?><osm version="0.6" generator="Overpass API 0.7.56.9 76e5016d">
<note>The data included in this document is from www.openstreetmap.org.
The data is made available under ODbL.</note>
<meta osm_base="2021-03-21T10:46:58Z"/>

<way id="2955682">
  <nd ref="442725213"/>
  <nd ref="3917460106"/>
  <nd ref="13865038"/>
  <tag k="bicycle" v="designated"/>
  <tag k="designation" v="public_footpath"/>
  <tag k="foot" v="designated"/>
  <tag k="highway" v="cycleway"/>
  <tag k="prow_ref" v="BC66/9"/>
  <tag k="segregated" v="no"/>
  <tag k="surface" v="asphalt"/>
</way>
<way id="2959339">
  <nd ref="13914151"/>
  <nd ref="1045973894"/>
  <nd ref="1045973841"/>
  <nd ref="1045973695"/>
  <nd ref="13914149"/>
  <nd ref="560109337"/>
  <nd ref="613541403"/>
  <nd ref="560109338"/>
  <nd ref="1077037675"/>
  <nd ref="2730704591"/>
  <nd ref="13865044"/>
  <tag k="bicycle" v="designated"/>
  <tag k="designation" v="public_footpath"/>
  <tag k="foot" v="designated"/>
  <tag k="highway" v="cycleway"/>
  <tag k="lit" v="no"/>
  <tag k="prow_ref" v="BC65/3"/>
  <tag k="segregated" v="no"/>
  <tag k="surface" v="compacted"/>
</way>

```



Figure 2.4: An extract from the Bristol OSM .xml file showing the data for two cycle ways within Bristol.

not all of these are present for every way in the network. Our conditions for a way to be identified in the cycle network are: (a) its ‘Highway’ tag takes the value ‘cycleway’; or (b) its tag set includes ‘Cycleway’; or finally (c) its tag set includes ‘bicycle’ which takes the value ‘designated’.

These criteria give rise to a binary indicator variable $\chi_{i,j}$ for the edge connecting nodes i and j which takes values,

$$\chi_{i,j} = \begin{cases} 1, & \text{if } (i,j) \text{ is a cycle path,} \\ 0, & \text{otherwise.} \end{cases}$$

The network with edges taking value $\chi_{i,j} = 1$ highlighted can be seen in Figure 2.5 with a zoomed in section showing the fine detail.

The edges (i, j) of G are weighted to represent the cycled length $l_{i,j}$ of the edge and a parameter $p_{i,j}$ which represents the impedance to cycling. The value of $p_{i,j}$ can in principle take into account lots of real world factors such as: hills, speed limit of roads etc. The analysis presented here will take a very simplified approach by setting

$$p_{i,j} = 1 - \chi_{i,j},$$

Meaning that impedance is zero for a street designated with cycling infrastructure and is maximal otherwise.

Each putative cyclist k should also have a personal parameter ω_k which describes their propensity to cycle on ways without cycling infrastructure.

From this the effective length of the edges in the graph G can be modelled by

$$\hat{l}_{i,j} = l_{i,j}(1 + \omega_k p_{i,j}), \quad (2.1)$$

meaning that an edge without cycling infrastructure has a larger perceived length than its true length according to the cyclist's disposition.

The idea then is that in reality a nervous cyclist would usually go out of their way to use designated cycling infrastructure rather than cycle on the road with other vehicles. This solution models that choice by having a rational agent minimise their perceived route length, meaning where possible using exclusively cycle paths unless the detour required is larger than the length penalty of using a regular road. Obviously in reality a cyclist will not go miles out of their way to use cycling infrastructure but small detours are acceptable.

The literature suggests that the majority of cyclists prefer to cycle on designated cycling infrastructure [4]. This means that ω_k should be strictly positive. The larger the value given to ω_k , the larger the penalisation of routes with no cycling infrastructure. A learner cyclist k' should have a large value for $\omega_{k'}$, as it is highly unlikely that a learner would want to cycle without dedicated infrastructure.

2.3 Uniformly random demand model

Naturally the simplest way of loading the network with cycling demand is to sample OD pairs randomly from the set of all nodes within the network. There are other more sophisticated models for demand such as the gravity model which estimates the number of trips between two points as being proportional to their populations and inversely proportional to the distance between them. In its most simple form the equation for this model is

$$T_{i,j} = A \frac{P_i^\alpha P_j^\beta}{d^\gamma}, \quad (2.2)$$

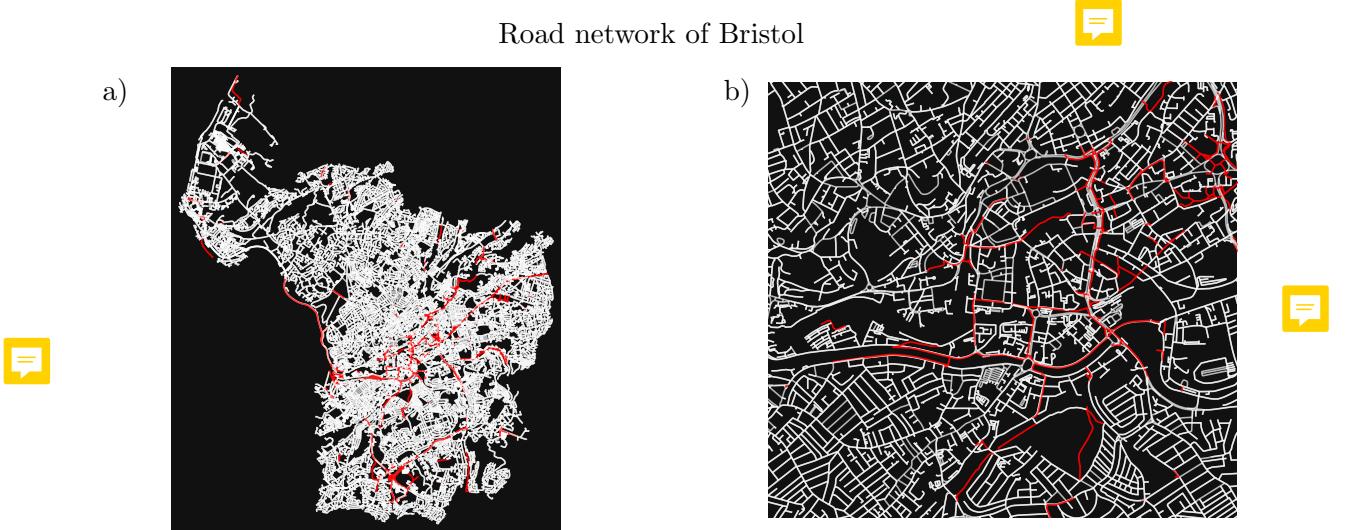


Figure 2.5: Road network of Bristol with edges (i, j) whose $\chi_{i,j} = 1$ highlighted in red. A section in the centre of a) is seen zoomed in b) to show the finer detail of the network.

where P_i and P_j are the populations of two areas, d is the distance between them and

$$A = \sum_i \sum_j \frac{T_{i,j}^{\text{empirical}}}{P_i P_j d^{-2}}. \quad (2.3)$$

The gravity model can be adapted to capture a key aspect of cycling, that cyclists tend to only travel short distances, by increasing the value of γ to further penalise long distance travel. However, this model requires empirical data to tune the parameter A . The uniformly random OD pairs are enough for illustrative purposes in this chapter. In chapter 3 we investigate a better demand model in which we can use empirical data about commuting journeys in Bristol to model demand.

Once we have generated demand on our network we have a set of OD pairs that represent cycling journeys in our city. The next step is to compute the shortest route between each OD pair. OSMnx is built on top of NetworkX, so we use the in build methods for route finding using Dijkstra's algorithm. In larger city cases it may be necessary to use a route finding heuristic like A* to save on computation, but in this project simulation time using Dijkstra's algorithm is sufficiently low for our case study city Bristol.

To test that our model for propensity to cycle is yielding the results we want we must simulate trips for different types of cyclists on a given network to see whether a decreased propensity to cycle leads to higher usage of a cities cycle network, regardless of shortest route. To do this we will simulate trips on the Bristol network. Firstly, with a very high propensity to cycle cyclist ($\omega_k = 0$) and then with a very low propensity to cycle cyclist ($\omega_k = 25$). The idea is that trips for more nervous cyclists use more infrastructure than those of confident cyclists.

Figure 2.6 (a) shows the distribution of percentage of trip length spent on cycling infrastructure when cyclists have penalty factor $\omega_k = 0$. Whereas, Figure 2.6 (b) shows the distribution for cyclists have little propensity to cycle so have a large penalty factor $\omega_k = 25$. In (a) ignoring the huge peak at 0% the proportion of route length spent within the cycle network seems to

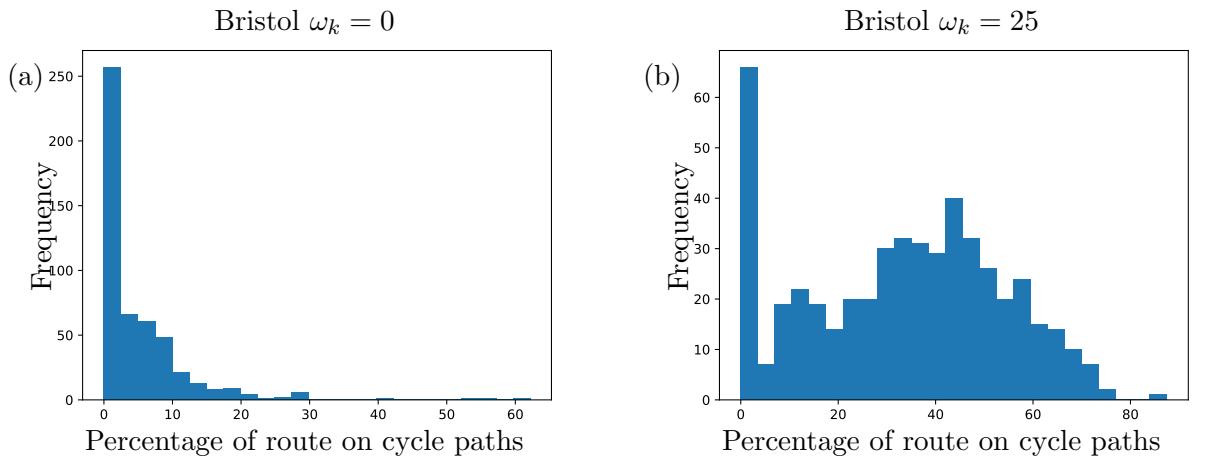


Figure 2.6: Histogram showing the percentage of routes taken on cycle lanes within shortest routes using uniform random OD selection, with parameter $\omega_k = 0$ (a) and $\omega_k = 25$ (b) to show that lower propensity to cycle yields higher cycle lane usage.

decay exponentially, with the vast majority of routes spending less than 15% of their length on cycle infrastructure. This result makes sense given the unconnected nature of Bristol's cycle network, which does not allow for whole routes to use cycling infrastructure. However, when propensity to cycle is decreased in (b) we can see that shortest paths tend to use much more cycling infrastructure. This makes sense as a cyclist who is very nervous is likely to take detours just to use designated infrastructure.

The peak around 0% in both (a) and (b) is due to the random selection of OD pairs giving rise to extremely short routes in areas of Bristol with no cycling infrastructure, so it is impossible for the shortest route to contain cycle lanes. In this computation the mean number of edges used in shortest paths is 120 which, when the network contains around 55,000 edges, is conducive to a large number of short routes. Another explanation for this peak is simply that the cycle network in Bristol is so sparse that most shortest routes cannot use cycling infrastructure. However we have shown that this effect can be reduced by reducing the cyclists propensity to cycle factor.

2.4 Comparison to second bounding geography

A second bounding geography, Amsterdam is also presented in order to compare summary statistics on a more highly connected cycle network. In order to control changes to the model, ω_k is left as 0 in the case of Amsterdam, to see whether connectivity or propensity to cycle on roads has a larger impact on the model result. Amsterdam is selected as a bounding geography not just for its much more comprehensive cycle network but also for its higher quality OSM data when it comes to cycle paths. In Bristol there are cases where new painted cycle paths are not yet documented in OSM, this is because OSM data is populated by users and without a dedicated team it is not possible for OSM to be completely up to date. Amsterdam's cycle network's higher degree of connectivity should yield a higher percentage of time spent on cycle paths within shortest routes.

Figure 2.7 shows that it is the case that in Amsterdam routes use more cycling infrastructure

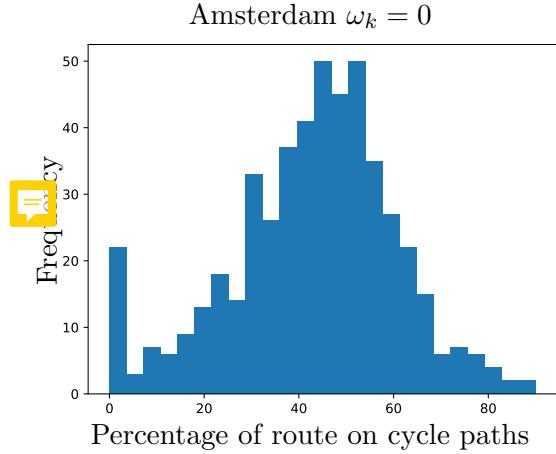


Figure 2.7: Histograms showing that the more connected city of Amsterdam (with $\omega_k = 0$) has a similar percentage of routes on cycle infrastructure to that of Bristol with very low propensity to cycle cyclists.

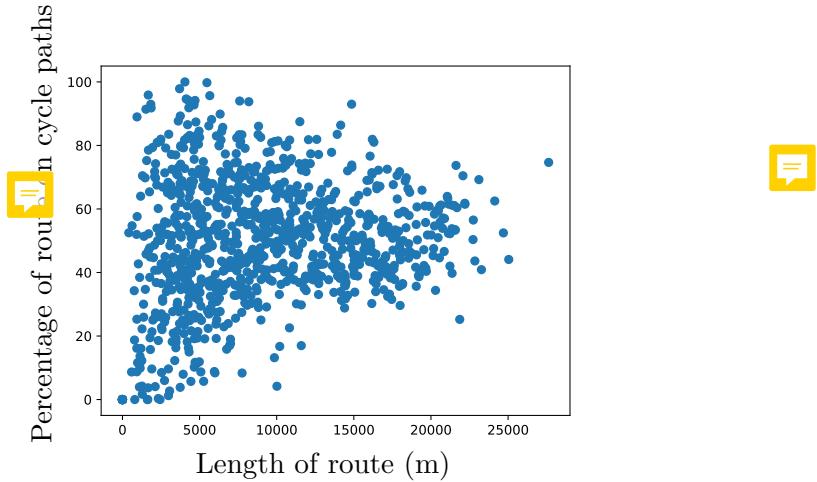


Figure 2.8: This scatter shows the relationship between route length and percentage of route on cycling infrastructure. It serves to motivate the claim that low scoring routes in the Amsterdam network are due to short routes where it is infeasible to detour onto the cycle network.

than Bristol even with $\omega_k = 0$ indicating that the choice of bounding geography and its cycle network have a larger bearing on cyclist experience than simply their propensity to cycle. The issue with the outlier peak at 0% is still somewhat present, but in the case of Amsterdam almost certainly due to a small amount of very short routes that cannot feasibly use cycling infrastructure. Figure 2.8 shows the relationship between route length and the percentage of that route that uses cycling infrastructure. It confirms the theory that low percentage routes in the Amsterdam network are generally very short. It can be seen that the only routes that use no cycling infrastructure have length less than 3km long and therefore a long detour onto the cycle network is infeasible. Further computations (not presented here) have shown that as ω_k is increased for cyclists in Amsterdam, most routes are nearly 100% on cycle paths.

After restricting the model to choose shortest paths with at least 50 edges, the results obtained from the Amsterdam network with an $\omega_k = 0$ yield the same result but without a large peak at 0%. This confirms that this peak is caused by very short routes that simply cannot use cycling

infrastructure.

2.5 Contributions

The main contributions from this section are as follows:

- A model for cyclist route choice by adapting perceived edge length using a cyclist's propensity to cycle.
- Our model for cyclist route choice has been tested on our case study city Bristol. The results show that cyclists with low propensity to cycle do in fact tend to use more cycle paths than those with high propensity to cycle.
- A second bounding geography has been identified as an example of a well designed cycle network.

Chapter 3

Network Upgrade Heuristics

Clearly Bristol’s cycle network has a long way to go to reach the connectivity, and hence user satisfaction, of a city like Amsterdam. The goal of this project is to improve Bristol’s cycle network, by building new cycle paths to ultimately yield a more connected network. In reality councils have a limited budget to build new cycling infrastructure, and wish to design a network that fits within that budget whilst maximising some factor of cyclist satisfaction. The true optimisation problem is computationally infeasible on a road network the size of Bristol’s as it is essentially the network design problem which has been well established as NP-complete in the literature [19]. In its simplest form one must check of all 2^n networks to find which is optimal, where n is the number of nodes. This chapter aims to answer the question as to whether a simple heuristic approach is enough to inform optimal cycle network design in Bristol. We start by detailing the heuristic upgrade process and introducing its parameters. We perform some simple experiments on the Bristol network and introduce a scoring metric to measure how the upgraded network will improve cycling in Bristol. Finally we investigate the impact of varying the heuristic’s parameters on the resulting upgraded network.

3.1 Overview of the heuristic

Figure 3.1 gives a brief overview of the approach this chapter will take to upgrade Bristol’s road network with new cycling infrastructure and table 2.1 briefly summarises the heuristic’s parameters. The heuristic starts with empirical data on commuter flows in Bristol. These are used to compute edge flows, how many cyclists use each edge, for a trip simulation of size N_t . We then upgrade the edges in the network that currently have no infrastructure with cycle paths until a length budget L has been reached. One key aspect of the heuristic that will be examined later in this chapter is that of batching. The question we ask is whether by splitting the upgrade budget into smaller batch upgrades we can achieve a more connected network which will be more desirable for cyclists.

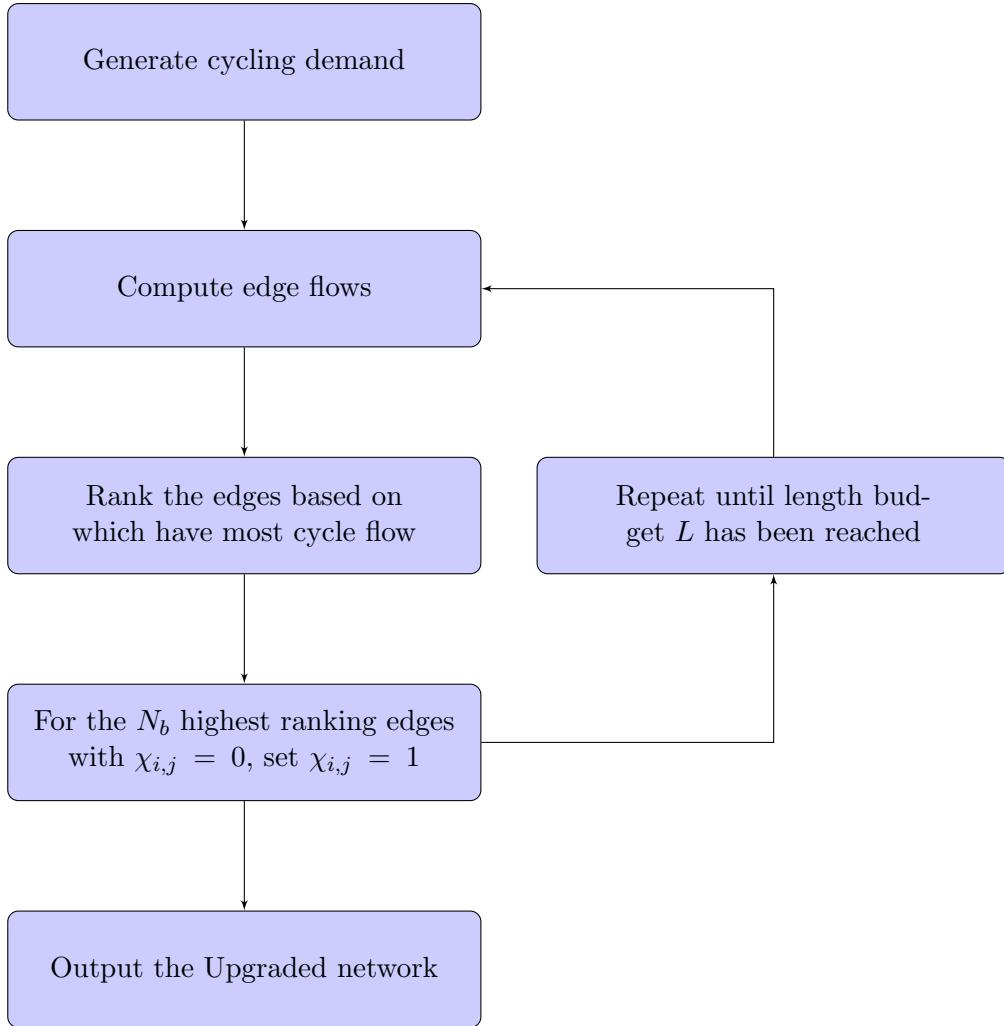


Figure 3.1: Summary of the heuristic upgrade approach proposed in chapter 3.

Key parameters	
Parameter	Definition
L	The total length edges to be upgraded. In case study 160 km to double the amount of infrastructure in Bristol.
N_t	The number of simulated cycling trips in a batch. This value should be large enough to avoid sampling bias but not too large as to focus demand on the centroids.
ω_d	The propensity to cycle factor for a ‘design’ cyclist. This is unique for the upgrade and need not be the same value in test cases.
B	The number of batches required to reach the upgrade budget with the current N_b .
N_b	The number of edges upgraded in a batch. This value is set but will change for the final batch as we go over budget in the final batch.

Table 3.1: Table of the key parameter for the heuristic proposed in Figure 3.1

Census Geography					
Area type	Minimum number of households	Minimum population	Maximum number of households	Maximum population	Number in Bristol
Output area (OA)	40	100	250	625	1368
Lower layer super output area	400	1000	1200	3000	263
Middle layer super output area	2000	5000	6000	15000	55

Table 3.2: The census geography requirements for different output areas and the amount of these areas within Bristol [20].

3.2 Cycling demand generation

Whilst uniformly randomly selecting OD pairs was sufficient to show the impact of a model for cyclist propensity to cycle in Chapter 2, to upgrade a cities cycle network properly a more realistic demand model is required. This section details the use of empirical data for our case study city Bristol, to give us a realistic model for cycling demand on the street network. Bristol, like any large city, owes most of its traffic flow to commuters. If we can simulate commuter journeys and assume they can be made by bicycle we will get a more realistic model for the flows of cyclists through Bristol. This means that census commuting data can be used to predict cycle flows. Unfortunately the most recent census data available in the UK is from 2011 and therefore, is not entirely accurate to today’s commuting trends. Although the methods presented in this report should be repeated on the commuting data from the 2021 census to identify any significant changes. The census flow data is available at three different levels: output areas, middle layer super output areas, and lower layer super output areas whose definitions are shown in table 3.2.

MSOAs are not chosen for this study simply as there are not enough in Bristol, with only 55 being present. This would not give enough variety in journey origins and destinations to effectively model the cycling population of Bristol. OAs are a lot more common in Bristol with around 1400 being recorded within this projects bounding geography. The problem with using OAs is that with only a small population in each one it is difficult to simulate large amounts of demand. With so many OAs we would need a huge amount of simulated trips to avoid sampling bias on OD pair journeys. It is for these reasons that the data used for this study is at LSOA level. There are 263 LSOAs in our bounding geography, giving us good coverage of Bristol but without running into issues of sampling bias.

The methodology detailed for loading demand is inspired by the DataShine commute web tool. It uses the 55 MSOA weighted centroids in Bristol and models flow between these MSOAs from centroid to centroid. Figure 3.2 shows a screenshot from the DataShine web page with flows highlighted to and from one of Bristol’s MSOAs. The colour of the line indicates direction with blue meaning flowing in to the centroid and red meaning flowing out of the centroid. The thickness of the line then shows the amount of journeys flowing between the two centroids. Their tool is concerned with the entire UK, whereas, we are only concerned with Bristol so we take the idea of modeling using centroids and apply it to the ONS’ LSOA commuting data-set for Bristol.

Each LSOA is modelled as its centroid, that is we assume all cycle flow comes from the centre of the LSOA. The data-set contains the longitude and latitude coordinates for each centroid, so the first step is to map these to nodes in the python network structure. The resulting centroid

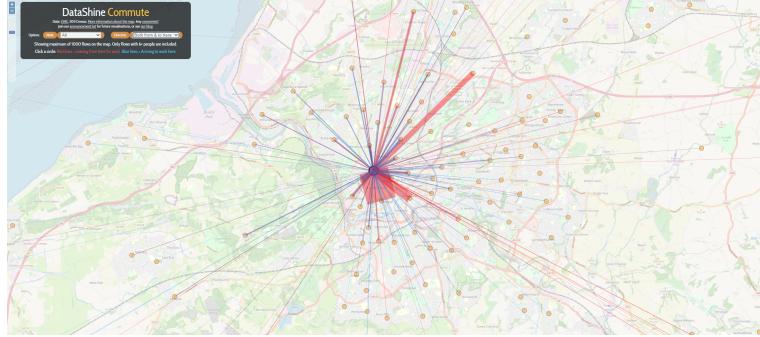


Figure 3.2: A screenshot of the DataShine commute interactive map showing the flows to (blue) and from (red) one of Bristol's 55 MSOAs [21].

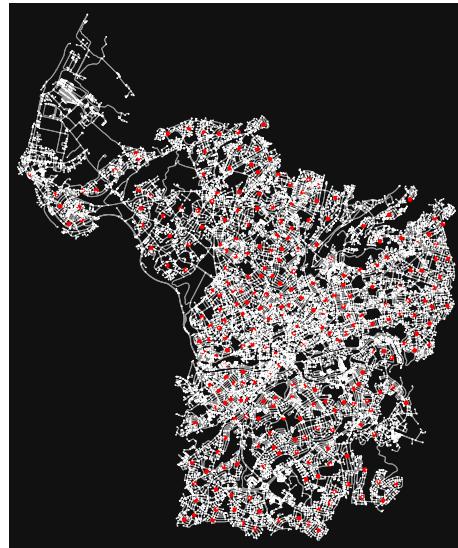


Figure 3.3: The map of Bristol's road network with nodes representing LSOA centroids highlighted in red.

locations for Bristol are plotted in red in Figure 3.3.

The commuting data is represented in a matrix where the element at index i, j gives the amount of people who regularly commute from centroid i to centroid j .

It's worth noting at this point that by taking the assumption of all LSOA flow being generated at its centroid, there is no way to model flow within the LSOA so we set all diagonal elements equal to zero. Since LSOAs are so small this really only eliminates journeys that use at most a couple of edges and therefore should not impact the upgrading of the network significantly.

This matrix allows us to compute a probability vector

$$P(f_i) = \frac{1}{\sum_{i=1}^{N_c}} \begin{pmatrix} f_{i,1} & f_{i,2} & \dots & f_{i,262} & f_{i,N_c} \end{pmatrix}, \quad (3.1)$$

corresponding to each centroid i to represent the probability of a journey that is initialised at i terminates at any other centroid j and where N_c is the total number of LSOA centroids, in this



Figure 3.4: The map of Bristol’s road network with edges coloured based on their $h_{i,j}$ flow value. case 263.

To load the network with cycle flow, journeys are initialised at a random origin centroid and the destination is chosen by sampling from all possible destination centroids using the probability vector $P(f_i)$. The shortest route between the OD pair is then computed using edge lengths adjusted as detailed in section 2.2.

After simulating N_t cycling journeys on the road network we can form an edge flow matrix where the element at index i, j is the number of simulated cycle trips that pass through the edge (i, j) taking a minimum value of 0 and a maximum of N_t . This gives an approximation to the priority given to different edges in the network by the simulated cyclists.

Figure 3.4 shows the Bristol road network with edges coloured corresponding to their flow in a trip simulation of size $N_t = 10,000$. It shows that under this LSOA model with a penalty factor of $\omega_k = 2$, cyclists tend to travel along major spokes with the highest flows being in the city centre which makes sense with a commuting model as most commuters travel to work in the city centre.

This demand loading methodology works well for a British city such as Bristol, which has comprehensive flow data available through the most recent census. However, in other cases empirical data may not be available so loading the network with demand will have to use a more synthetic approach such as the gravity model detailed in section 2.3. For the purposes of this report we will continue with the case study of Bristol and in the next section detail how the network is upgraded and perform some assessment on the Bristol street network.

3.3 Assessment of upgrades using Bristol as a case study

To assess the user experience on an output upgraded network a scoring metric similar to the network assessment in chapter 2 is used although it is formalised here. The methodology is as

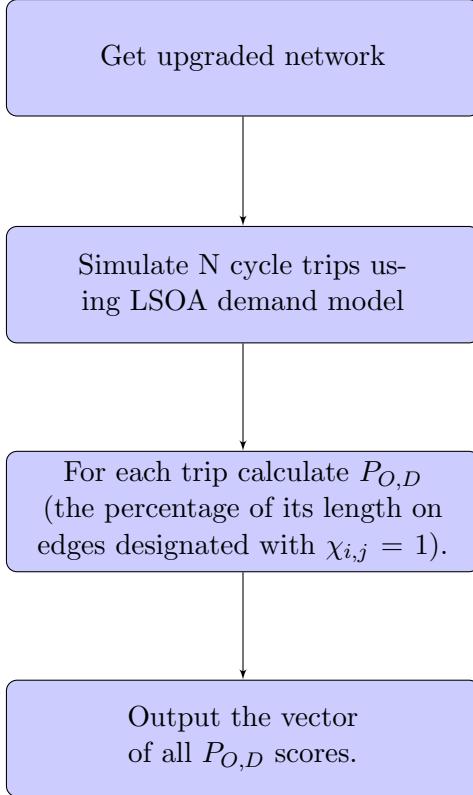


Figure 3.5: Computation of the scoring metric.

shown in Figure 3.5. This metric represents the user experience for cyclists using the upgraded network. This section details experiments performed on our case study city Bristol.

Now that we have a heuristic and a way to assess the impact of implementing it on the network we can upgrade the cycle network. Figure 3.6 shows the upgraded network found by implementing the heuristic on the Bristol with parameters: $L = 160\text{km}$, $N_t = 500$, $N_b = 20$, and $\omega_d = 2$.

In Figure 3.6 it is clear that the main spokes of the Bristol road network that are highlighted with high flow counts in Figure 3.4 are all upgraded with cycling infrastructure if they are not already designated with infrastructure. This makes sense as these edges are the main thoroughfare for cyclists and will be ranked highest at the end of each batch's trip simulation.

Then, once the main spokes have been upgraded the algorithm starts to fill in the lower demand areas that connect Bristol's LSOA centroids to the key network spokes. One example of this effect is seen in the top of the figure. Cycling infrastructure is suggested into this area where demand is not particularly high but, as seen in Figure 3.3 there are LSOA centroids there.

When the upgrade Bristol network is tested as detailed in Figure 3.5, Figure 3.7 (a) is computed. It shows a large increase in the average percentage of their trip that cyclists spend on designated infrastructure. The current Bristol cycle network achieves an average of 34% of journey length on cycling infrastructure. However, the proposed upgraded network achieves an average of 75% of journey length on cycling infrastructure. There are some outlier cyclists in the post-upgrade distribution, these trips are most likely due to test journeys being generated between two LSOA centroids that are very close together with low design demand between them and hence little cycling infrastructure has been generated along the shortest path from origin to destination.

The value of L in this case is 160km which means that after the upgrade has been performed there

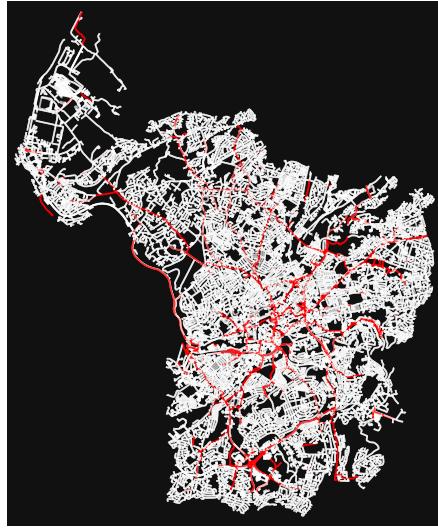


Figure 3.6: The output upgrade network for a batch process with parameters $L = 160\text{km}$, $N_t = 250$, $B = 20$ and $\omega_d = 2$.

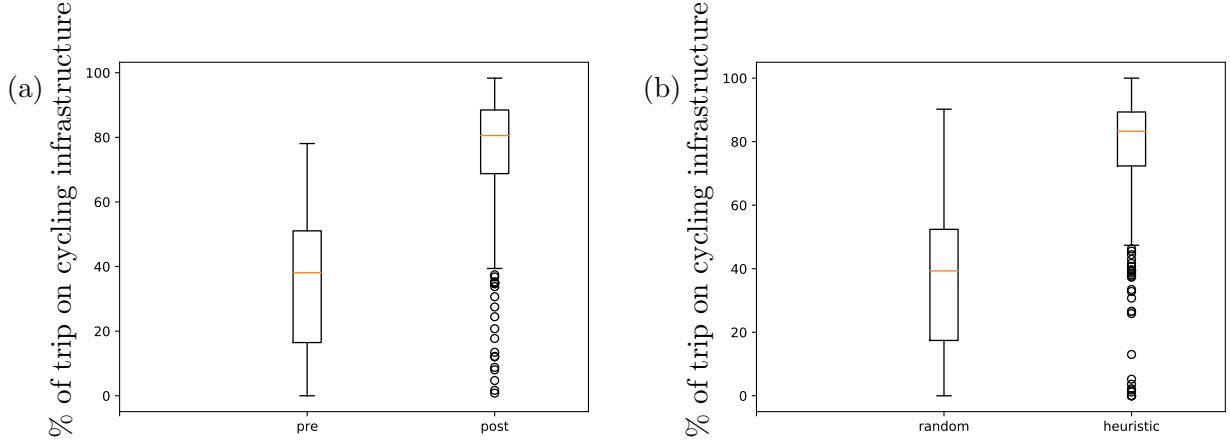


Figure 3.7: The distribution of $P_{O,D}$ scores for (a), the Bristol network before and after applying the upgrade heuristic, and (b), the upgrade heuristic vs randomly assigning the same length of upgraded streets.

is more than twice the cycling infrastructure than in the original network. This will obviously yield an increase in the percentage of the length of cycling trips on designated infrastructure. Therefore, the question is whether the upgrade heuristic performs markedly better than just randomly upgrading the network to contain twice as much infrastructure. To do so we upgrade random edges in the network with cycling infrastructure until the length budget of 160km is met, we then test the $P_{O,D}$ scores of this upgraded network against the network produced by the heuristic. Figure 3.7 (b) clearly shows that the heuristic proposed in this chapter vastly out performs a random process in terms of cyclist satisfaction. The random process only achieves an average of 37% of the length of journeys on infrastructure a 3% gain over the original network with twice as much infrastructure. In contrast the upgrade heuristic proposed here yields a

massive improvement with 78% of the length of journeys using designated infrastructure by upgrading the same number of edges.

3.4 The impact of batching

The upgrade heuristic proposed in this project lends itself to an incremental process, from now on referred to as a batch process, where small portions of the overall length budget are upgraded before cycle flows are recomputed and the next batch of upgrades applied. This process aims to simulate the ‘pull’ of new cycle paths for cyclists travelling between all OD pairs. The idea is that, if new cycling infrastructure is built then existing cyclists would be incentivised to use it over cycling on the roads they usually take, if the new cycling infrastructure does not require them to take a large detour.

In the heuristic this is modelled by new cycle paths having their length shortened by the design cyclist’s propensity to cycle ω_d before the trip simulation for the next batch. This means that in some cases the previously longer route between an OD pair now is a shorter alternative for the cyclist. Over many batches this should mean that flows are increased on the edges at the end of growing contiguous paths of cycling infrastructure and such a connected cycle network should be generated. One issue with large numbers of batches is, to simulate the effect of new infrastructure pulling cyclists to it we must generate a new set of cycling trips for each batch which, for large N_t and large B can become computationally expensive. Therefore, the question we ask is whether varying batch size has a significant enough impact either visually or metrically on the upgraded Bristol network to justify the computational expense.

Figure 3.8 shows the output networks for the two extreme cases of this methodology tested in a smaller upgrade case of length budget $L = 40\text{km}$. The one-shot network (a) is computed by upgrading all 40 km of edges in one batch after simulating 500 trips. Whereas, the single-edge network (b) is computed by upgrading only a single edge for each batch of 100 simulated trips until 40 km of street have been upgraded. The one-shot uses a larger amount of simulated trips simply to account for sampling error, which is not an issue for the single-edge method as trip flows accrue over batches. The zoomed in sections of both networks show the main difference between the two methods. Both (c) and (d) show the same part of the Bristol network shown by the blue boxes on (a) and (b), but (d) using the single-edge method has suggested infrastructure here. This is due to the growing property of a batched approach, the starting edge of the new connected section has been designated by both methodologies but only a batch approach can continue to grow from it.

Figure 3.9 (a) shows the distribution of $P_{O,D}$ scores for both the one-shot and single-edge methods. There is very little variation between the two with single edge only performing slightly better with a marginally higher average $P_{O,D}$ score. Therefore, although there is some visual difference between the suggested networks from both methods there is not necessarily a marked improvement for cyclist experience either way.

Figure 3.9 (b) shows the resulting $P_{O,D}$ distributions from a sweep of values for B and hence N_b , in our test case of upgrading $L = 160\text{km}$ of street. It shows there is actually very little sensitivity

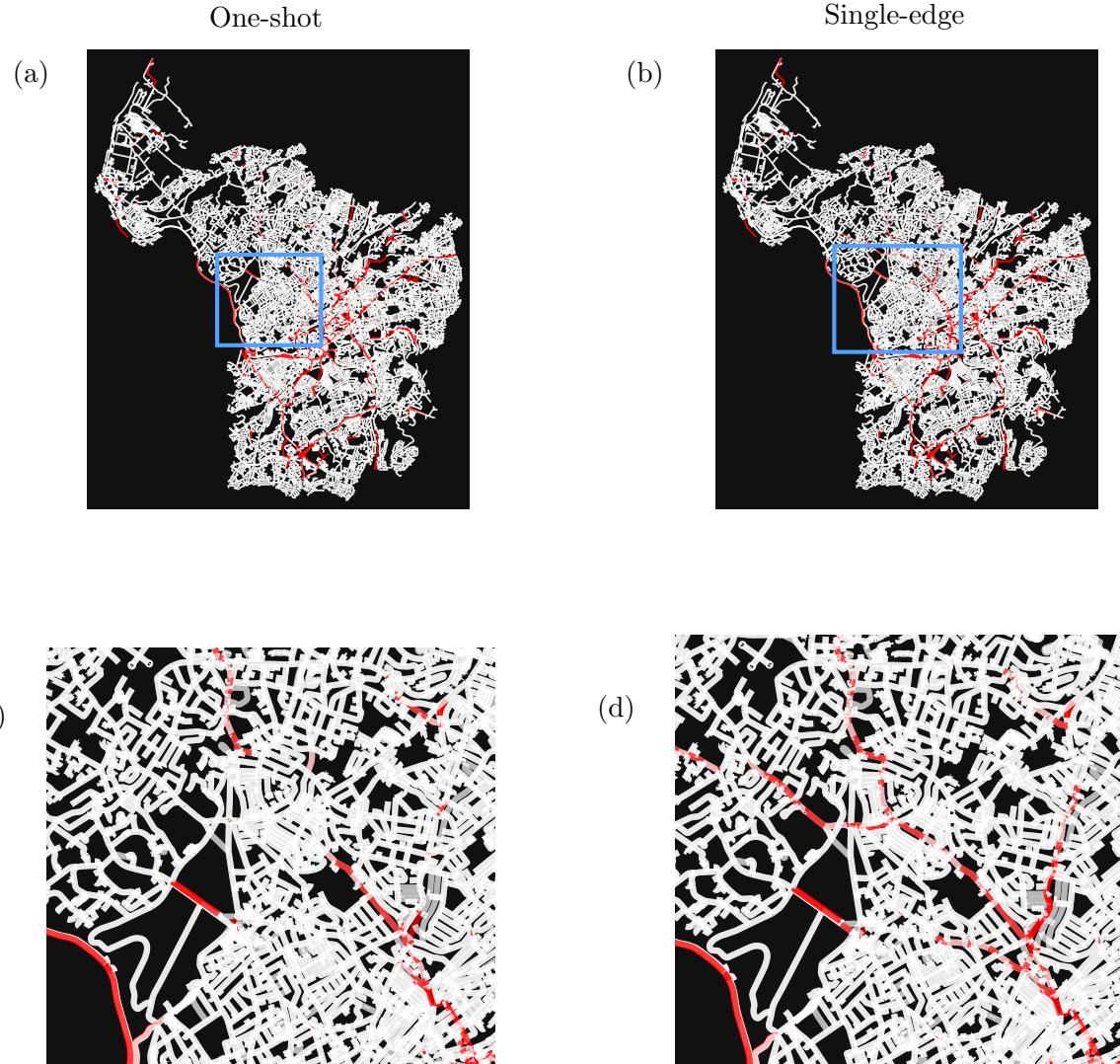


Figure 3.8: Road network of Bristol with edges (i, j) whose $\chi_{i,j} = 1$ highlighted in red. a) shows the network after a one-shot process has been applied to upgrade $L = 40\text{km}$ of street and b) shows the result of a single-edge scheme upgrading $L = 40\text{km}$ also. (c) and (d) show the same zoomed in section of the network in the centre of the map on the left. It can be seen that cycling infrastructure is found here in the single-edge but not the one-shot this is due to the growing property of the single edge.

to the number of batches in the case of the Bristol network, meaning that we can capitalise on the computational saving of reduced batch numbers. This low sensitivity is probably due to the well defined demand spokes seen in Figure 3.4 where these spokes have such heavy flow counts compared to other areas of the network they will always be prioritised for upgrade irrespective of batch sizing. It is possible that other network cases with less well defined demand spokes may have higher sensitivity to the batch size parameters, but this project focuses on the case study of Bristol so none are presented here.

One major correction that must be incorporated into the upgrade heuristic is overshoot. That

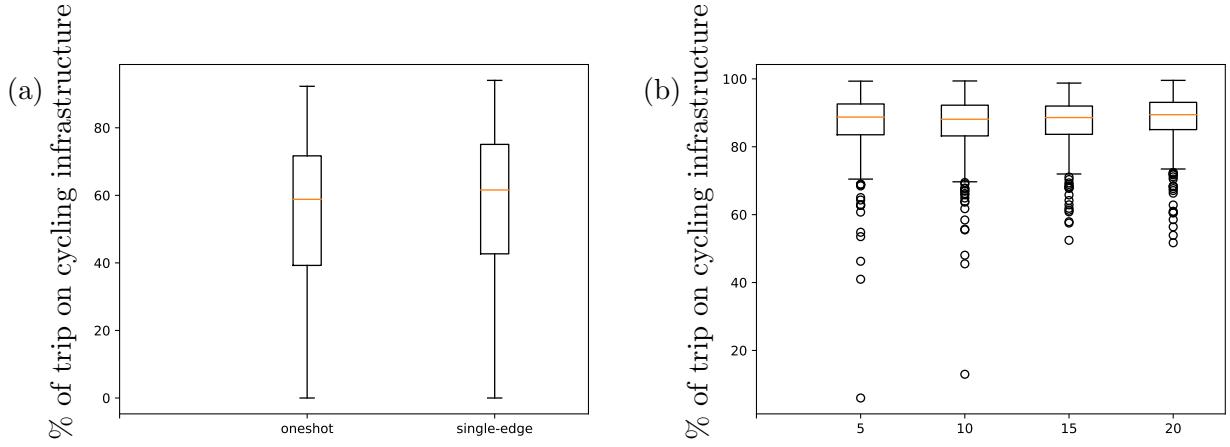


Figure 3.9: The distribution of $P_{O,D}$ scores for (a), the upgraded network for one-shot and single-edge, and (b), when varying the number of batches B .

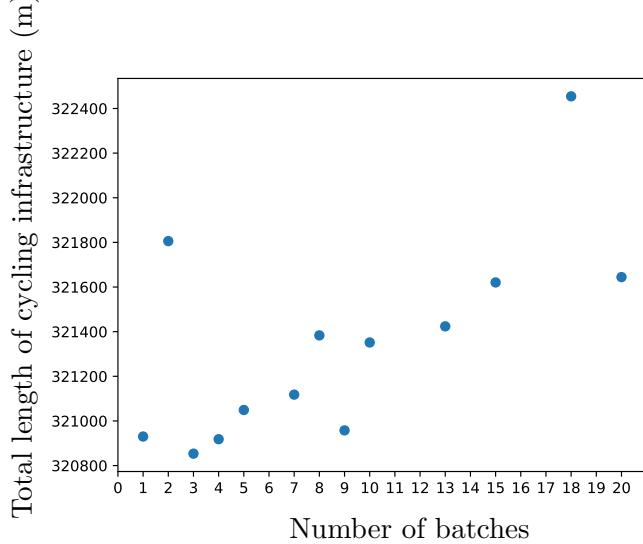


Figure 3.10: Scatter showing the potential for overshoot when designing networks with larger numbers of batches. The target is to double the current infrastructure giving 320km post upgrade.

is, when we divide the overall length budget over B and upgrade the edges for that batch it is appropriate to upgrade edges until we go over budget for that batch. This is required as, in the network representation of the Bristol network, it is not possible to have only a portion of an edge be designated with infrastructure. This methodology will unfortunately cause issues with large numbers of batches with overshooting causing us to upgrade more street than budgeted for. Figure 3.10 shows the impact this overshoot can have with a variety of batch sizes. Obviously there is some potential for random behaviour as the length of the final street in a batch is not considered, so for a given trip distribution it could either be very long or very short. There is clearly a trend to larger overshoot with more batches simply because the overshoot edges have a chance to compound. This result suggests that we should avoid extremely high numbers of batches as it could yield misleading results.

After conducting these experiments with varied batch numbers and sizes it can be seen that in the case of the Bristol network batching does not have a significant impact on upgrade results.

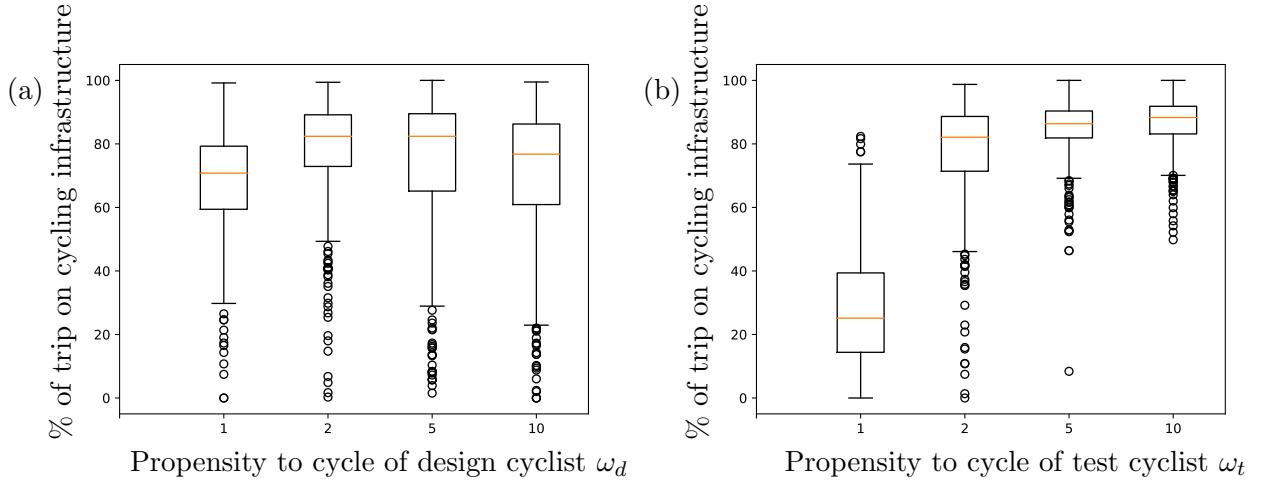


Figure 3.11: The distribution of $P_{O,D}$ scores for (a), the upgraded network for different values of ω_d , and (b), the upgraded network for different values of ω_t .

This is due to the high flow rates on key infrastructure spokes that allows for the existing cycle structure to be connected regardless of batching. However the scope of this project is such that we have not explored the implications of batching on different network setups so there is a chance that in a different city batching may improve the network upgrade suggestion. In chapter 4 we will explore whether batching has an impact on some very small synthetic networks but further work should be conducted on other cities in the future.

3.5 The impact of different design and test cyclists

So far the value of ω has been fixed between upgrading and testing the network, that is to say $\omega_d = \omega_t$. However, it is worth exploring the impact that designing a network for a cyclist with a given propensity ω_d has on the resulting networks fitness for a cyclist with a different propensity ω_t . For example, if we design a network with cyclists who have little preference between cycling on and off designated infrastructure will this have an impact on the experience of a cyclist with a much higher penalty factor in the testing stage. Conversely, if a cyclist has a low penalty factor in the test stage will they prefer routes that are off the cycling infrastructure as they are more direct than the continuous cycle paths desired by more cautious riders.

Figure 3.11 (a) gives the $P_{O,D}$ distributions for four different network designs with four different design cyclists $\omega_d \in [1, 2, 5, 10]$. The networks are then all tested with the same test cyclist with propensity $\omega_t = 2$. It can be seen that whilst all of the upgraded networks present a marked improvement over the original Bristol network, a cyclist who shares the same propensity as the design cyclist has the best experience. In this case the network with $\omega_d = 2$ has both the highest average $P_{O,D}$ score but also the least spread suggesting a better user experience.

Naturally the next step is to fix the design cyclist at $\omega_d = 2$ and see the effect that varying the propensity of the test cyclist in the same range $\omega_t \in [1, 2, 5, 10]$. Figure 3.11 (b) demonstrates that the propensity to cycle of the test cyclist has a huge impact on the distribution of $P_{O,D}$ scores for the upgraded network. In the case $\omega_t = 1$ the test cyclist takes the true shortest route between their origin and destination which may not fall on suggested infrastructure that tends

to be built onto existing infrastructure. We see also that as ω_t is increased the cyclist clearly is happy to take a longer detour to stay on designated infrastructure, with very large ω_t value cyclists spending almost 100% of their journey on cycle paths. This finding suggests that as long as upgraded networks have a connected path between origin destination pairs nervous cyclists will travel a longer distance to use them. This is not true in reality which is why physically a propensity to cycle of 10 is infeasible. The test cyclist with $\omega_t = 2$ seems the most realistic as most journeys do use a lot of designated infrastructure but when trips become too long by trying to stay within the cycle network the cyclist defaults to the shortest path. This correlates with the findings of Krizek et. al. [22] who found that on average cyclists are willing to travel up to 67% further than the shortest path just to use designated infrastructure.

3.6 Process interpretation

The batches in the proposed upgrade algorithm are, until now, a synthetic time-step to encourage connected networks. However, there is a very nice parallel between batches and individual cycle schemes. In reality the council, or anyone else building new infrastructure, want to stagger the implementation of new cycle schemes to gauge the true impact that new infrastructure has on cycling demand. In this model it is appropriate to view each batch as a new scheme perhaps implemented year on year. The characteristic of growing connected cycle paths then follows naturally. If there is new infrastructure, naturally simulated cyclists with high penalty factor ω_d will want to use it to service their journeys. This would in reality then lead to higher flows on roads at the end of this new infrastructure, so long as it doesn't terminate in an LSOA centroid. Therefore the upgrade algorithm should and does prioritise these edges for upgrade in the next batch.

It is in the councils best interest to minimise their cost whilst maximising cyclist user experience. Therefore, it is worth exploring the improvement batch on batch for these incremental cycle schemes. If there is a huge improvement to the $P_{O,D}$ scores in the first couple of batches and only very small improvements after, it may not be worth investing in further improvement. This information is invaluable to someone implementing new cycle schemes as it could reduce costs drastically.

Figure 3.12 shows the score change of improvements made to the Bristol network throughout all 20 batches of the upgrade algorithm. The plot shows that most of the upgrade to the average $P_{O,D}$ scores occurs within the first few batches. It is worth investigating which edges get upgraded first to see if there is some strategy for prioritisation to be found.

Figure 3.13 shows the order in which the upgrades are performed. Edges are coloured base on which batch they are upgraded in with darker edges being upgraded first. This plot indicates that the major infrastructure spokes seen built into the top and bottom of the network in Figure 3.6 are built within the early batches and suggests these are the streets the council should prioritise upgrading first. These upgrades correspond to the significant increase in average $P_{O,D}$ seen in Figure 3.12. The more yellow edges are upgraded in the last couple of batches and tend to be at the end of long spokes, indicating that the algorithm is upgrading in such add to existing infrastructure batch by batch. Finally some of the later batches begin to fill in small gaps near

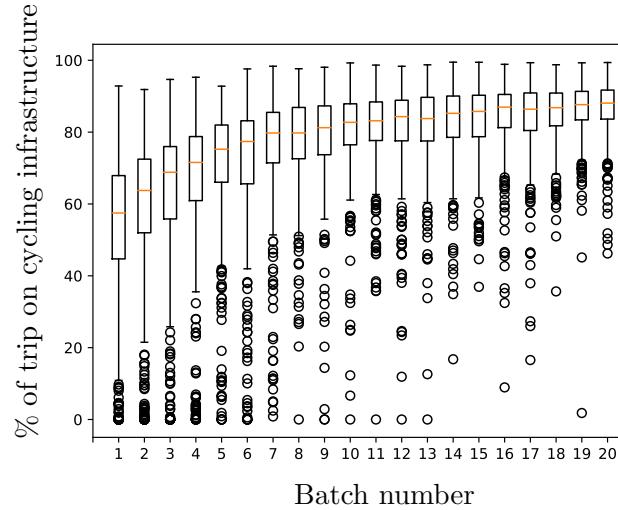


Figure 3.12: Graph showing the improvement in $P_{O,D}$ scores throughout the batches of a 20 batch upgrade of the Bristol network. It shows that most of the upgrades are made in the early batches.



Figure 3.13: The map of Bristol's road network with upgraded edges coloured based on which batch they are upgraded in. The brighter the edge the later it is upgraded.

the locations of LSOA centroids which correspond to the small gains in $P_{O,D}$ spread seen in Figure 3.12.

3.7 Contributions

The major contributions from this section are as follows:

- We have introduced a simple model for loading a cities street network with cycling demand,

provided that empirical data is available.

- We have detailed a heuristic upgrade approach to take a city's cycle street network and upgrade it with new infrastructure up to a specified length budget.
- We have formalised a scoring metric to assess the user experience of a cyclist with given propensity to cycle using the upgraded network.
- We have assessed the performance of this upgrade metric on our case study city Bristol.
- We have demonstrated the effect of a couple of key model parameters on the resulting upgraded Bristol street networks.

This chapter has given a good description of the proposed upgrade heuristic and has shown that it succeeds in suggesting a more connected network conducive to a better cyclist user experience. The question remains as to whether the networks suggested by the heuristic are close to being truly optimal. The next chapter will explore the performance of the heuristic against formal optimal network design on some small network examples.

Chapter 4

Optimal Networks

Bibliography

- [1] C. Allan. Cycling UK's cycling statistics. <https://www.cyclinguk.org/statistics>. Accessed 2020-12-02.
- [2] Conclusions from the National Travel Survey 2019. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/905887/national-travel-attitudes-study-wave-3.pdf. Accessed 2020-11-09.
- [3] Conclusions from the National Travel Attitude Survey wave 3 2019. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/905887/national-travel-attitudes-study-wave-3.pdf. Accessed 2020-11-09.
- [4] J. Dill and T. Carr. Bicycle commuting and facilities in major US cities: if you build them, commuters will use them. *Transp. Res.*, 1828(1):116–123, 2003.
- [5] Key findings from the Bicycle Association's study into covid-19's impact on the cycle industry. <https://www.bicycleassociation.org.uk/news-press/official-industry-stats-reveal-record-extent-of-covid-cycling-sales-growth/>. Accessed 2020-03-28.
- [6] G. Topham. Chancellor announces £27bn for roadbuilding in budget. <https://www.theguardian.com/uk-news/2020/mar/11/chancellor-announces-27bn-for-roadbuilding-in-budget>. Accessed 2020-11-29.
- [7] R. Cervero and M. Duncan. Walking, bicycling, and urban landscapes: evidence from the San Francisco bay area. *Am. J. Public Health*, 93(9):1478–1483, 2003.
- [8] D. Milakis and K. Athanasopoulos. What about people in cycle network planning? Applying participative multicriteria GIS analysis in the case of the Athens metropolitan cycle network. *J. Transp. Geogr.*, 35:120–129, 2014.
- [9] J. Larsen, Z. Patterson, and A. El-Geneidy. Build it. But where? The use of geographic information systems in identifying locations for new cycling infrastructure. *Int. J. Sustain. Transp.*, 7(4):299–317, 2013.
- [10] A. Mauttone, G. Mercadante, M. Rabaza, and F. Toledo. Bicycle network design: model and solution algorithm. *Transp. Res.*, 27:969–976, 2017.
- [11] P. Freestone. Personal communication. October 2020.
- [12] R. Lovelace, A. Goodman, R. Aldred, N. Berkoff, A. Abbas, and J. Woodcock. The propensity to cycle tool: An open source online system for sustainable transport planning. *J. Transp. Land Use*, 10(1):505–528, 2017.
- [13] R. Lovelace and J. Talbot. Rapid cycleway prioritisation tool: Technical specification. <https://www.cyipt.bike/rapid/>. Accessed 2020-12-08.
- [14] DFT. The cycling infrastructure prioritisation toolkit. <https://www.cyipt.bike/report1/>, 2018.
- [15] Cyclestreets main webpage. <https://www.cyclestreets.net/>. Accessed 2020-03-29.
- [16] CycleOSM map of Bristol. <https://www.cyclosm.org/#map=13/51.4515/-2.5657/cyclosm>. Accessed 2020-03-29.
- [17] OSM map of Bristol. <https://www.openstreetmap.org/#map=13/51.4740/-2.5889>. Accessed 2020-03-29.
- [18] G. Boeing. Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Comput. Environ. Urban*, 65:126 – 139, 2017.

- [19] D.S. Johnson, J.K. Lenstra, and A.H.G. Rinnooy Kan. The complexity of the network design problem. *Networks*, 8:279–285, 1978.
- [20] Bristol city council guide to census geography in Bristol.
<https://www.bristol.gov.uk/documents/20182/34008/2011+Geographies+for+2011+Census+data+in+Bristol.pdf/146b77b1-e743-4511-b317-643b6358b571?t=1436544756000>. Accessed 2020-03-29.
- [21] DataShine commute map of Bristol. <https://commute.datashine.org.uk/#mode=allflows&direction=both&msoa=E02003033&zoom=13.0&lon=-2.5790&lat=51.4576>. Accessed 2020-03-21.
- [22] Kevin J Krizek, Ahmed El-Geneidy, and Kristin Thompson. A detailed analysis of how an urban trail system affects cyclists' travel. *Transportation*, 34(5):611–624, 2007.