UNIVERSITY OF BRISTOL

DEPARTMENT OF ENGINEERING MATHEMATICS

Cycle Networks — Finding the Missing Links

Max Oblein (Engineering Mathematics)
mo17165@bristol.ac.uk

Project thesis submitted in support of the degree of Master of Engineering

Supervised by Prof R. E. Wilson

April 28, 2021

Abstract

Cycling is becoming an increasingly more popular mode of transport around large cities across the UK [1]. However, research suggests that by increasing the amount of designated cycling infrastructure even more people will choose to cycle [2]. This project aims to create a simple model of cyclist route choice and use this to simulate cyclists on a city street network to inform upgrade decisions. We propose a simple model of route choice that increases the perceived length of streets without designated infrastructure according to a cyclist's personal propensity to cycle factor. Next, we formulate a simple upgrade heuristic that upgrades streets in a city network with cycling infrastructure based on the amount of cycle flow along the street in a simulation of cyclists drawn from an empirical demand model using perceived length to find shortest routes between their origin and destination. We apply this heuristic to the Bristol network and show that it increases cyclist satisfaction according to our scoring metric, which measures the amount of a cyclist's journey that is spent on cycling infrastructure. Finally, we investigate applying the heuristic to small synthetic networks where the impacts are more visible. This also allows us to compare to formal optimal design and we conclude that in certain cases it is possible for our proposed heuristic to achieve optimal network design.

Contents

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) [3] 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) [4] 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". Indeed, research suggests that an increased provision of cycling infrastructure is positively correlated with the proportion of people cycling within that city [2]. 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 number of people choosing cycling as their transport mode for essential trips. 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 also found that E-bike sales increased 92% between April and September 2020, when compared to 2019. However, it is not just cyclists using segregated cycling paths anymore. In recent years the growth of other micro-mobility methods, such as electric scooters, has also increased [6]. However, it is worth noting that under current UK legislation [7], electric scooters cannot be ridden on cycle paths or pavement, although these rules are presumably subject to change in the near future. These factors all suggest it is worth investing in 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 [8]. Cycling is much less common as a mode of transport than driving, so planning the cycle network receives 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 academic work in planning cycle networks, and most studies focus on two areas: 1. the effect infrastructure has on the number of cyclists and 2. the evaluation and proposal of cycle schemes based on geographical factors. The first area motivates the need for cycling infrastructure [2]. The second is useful for the evaluation of routes that might be suggested by this project, based on urban factors such as the amount of retail or business land area at both the origin and destination of the route [9]. The methods presented by Milakis et al. [10] 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 contrast, in the literature little work has been found in 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, also formulated by Milakis et al. [10], 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. They identify eight such centres, see 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. The methodology of [10] 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 manual work. The survey used in this case is also only considering existing cyclist's current behaviour and therefore cannot predict the impact of new infrastructure on changes in cyclist behaviour, or indeed behaviour of travellers who take up cycling because of the new infrastructure.

A second paper by Larsen et al. [11] 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 each pair. 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, see 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 be highly prioritised for new infrastructure.

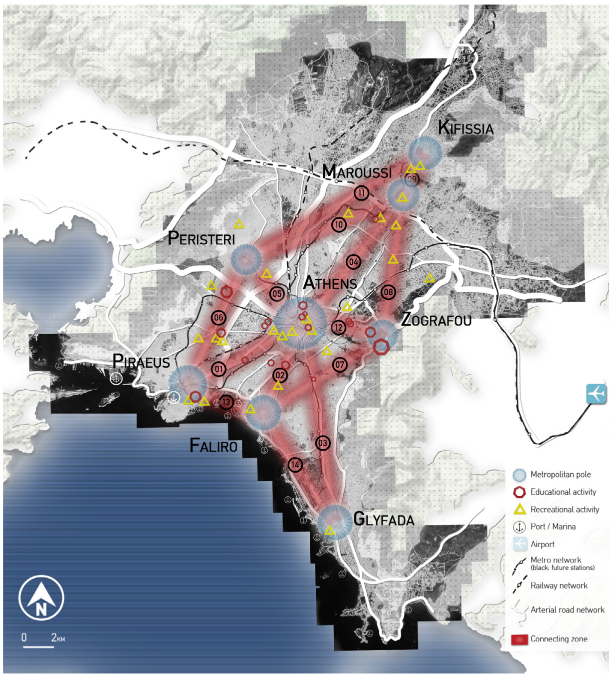

Figure 1.1: Demand centres (blue) in Athens. Reproduced from *What about people in cycle network planning? Applying participative multicriteria GIS analysis in the case of the Athens metropolitan cycle network*, by Milakis et al. (2014) [10].

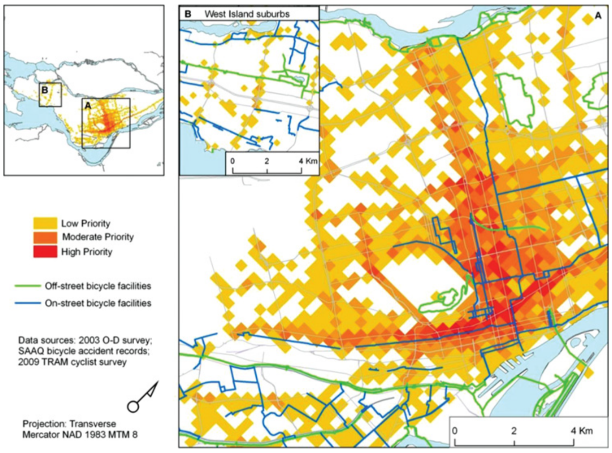

Figure 1.2: Prioritisation grid of Montreal. Reproduced from *Build It. But Where? The Use of Geographic Information Systems in Identifying Locations for New Cycling Infrastructure*, by Larsen et al. (2013) [11]

One part of the Montreal study [11] 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 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 Mauttone et al. [12] applies formal optimisation techniques to the cycle network in some case study cities. They propose minimisation 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 optimisation method on large scale networks, and compares against exact solutions on smaller artificial network examples. However the 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 [13] 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 to us are the propensity to cycle tool (PCT) [14] and the rapid cycleway prioritisation tool (RCPT) [15]. 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 (but hitherto, not implemented) facility to compute "cohesive" networks, which, in more standard network science terminology, we interpret to mean highly connected. Such networks are thought to be highly desirable, as they enable journeys to take place almost entirely on dedicated cycling infrastructure, without the dangerous short gaps where cyclists are required to mix with cars, buses, lorries etc. Unfortunately, in the current RCPT documentation there is no formal mathematics presented as to how to find such "cohesive" networks.

Summary of studies

Study	Methodology
Athens [10]	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 [11]	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 [12]	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 [14], [16]	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 whist 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 data and opinion polling to formulate a network upgrade strategy. This project will, through some key research questions, investigate the potential for a more mathematical, automatic design approach. Three key research questions have been identified:

- 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 a new model we propose for 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 a personal propensity to cycle. We show how cyclists with a lower propensity to cycle tend then (by shortest route principles) to select routes with a higher proportion of cycling infrastructure. Finally, the various metrics for Bristol are compared and contrasted with a second case study city (Amsterdam), in which the cycling provision is far superior.

Chapter 3 investigates the second research question. We develop a methodology that begins with simulating cycling demand using empirical census data. Simulated cycling trips are then used to compute the flow of cyclists on each edge in the network and we then upgrade those edges with the largest flows. The methodology has several parameters and we investigate how their tuning effects proposed network improvements in our case study city (Bristol).

In Chapter 4 we investigate the final research question. We start by modifying the heuristic and scoring functions for the purpose of testing synthetic networks. We then introduce a simple cross network example to demonstrate the steps of the upgrade heuristic. The optimal design is found for this network and we are then able to gain some insight as to whether our heuristic approximates this formal optimal. Finally we generate ensembles of small synthetic networks to explore whether or not batching the upgrade process gives a more optimal output network.

Finally in Chapter 5, we present conclusions and opportunities for further work.

Chapter 2

Data Sources and Initial Analysis

This chapter details the initial analysis of Bristol's street and cycle networks. Initially (Section 2.1) we discuss the potential data sources to be used and select OpenStreetMaps (OSM) as our primary data source. We then (Section 2.2) introduce a new single parameter model for the perceived length of routes for a cyclist with a given propensity to cycle. We then show (Section 2.3) the impact that this propensity to cycle model has on some simple test statistics on the Bristol network. These results are then compared (Section 2.4) with an objectively better cycle network, the city of Amsterdam, where the network is more expansive and OSM data is more complete.

2.1 Data sources

To create the network needed for the analysis proposed in this project, multiple data sources were investigated for conversion into a network representation for analysis. Three such sources have been identified: Cyclestreets [17], CycleOSM [18], and OSM [19].

Cyclestreets offers a routing engine for cycle route planning in the UK, an example of which is shown in Figure 2.1 (a). Their API allows users to query for cycle 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 thus are able to choose a route based on their propensity to cycle. So far as we have been able to determine, the algorithm which Cyclestreets uses to compute these route options is only available on top level in the public domain. However, note that as we will introduce our own propensity to cycle model (see Section 2.2), we would expect the routes that it will generate to encapsulate all of the possibilities offered by Cyclestreets, but with extra flexibility in terms of tunable parameters, modular re-use in other code, etc.

CycleOSM is essentially a condensed version of OSM with only the data required for cyclists present and it contains lots of information about cycle parking, repair stations, cycle shops etc. A screenshot of the CycleOSM map for part of Bristol is shown in Figure 2.1 (b). However, we aim to upgrade streets in a city with new cycling infrastructure, so a map based on purely existing cycling infrastructure, and not containing other streets, is not suitable for this purpose.

(a)

(b)

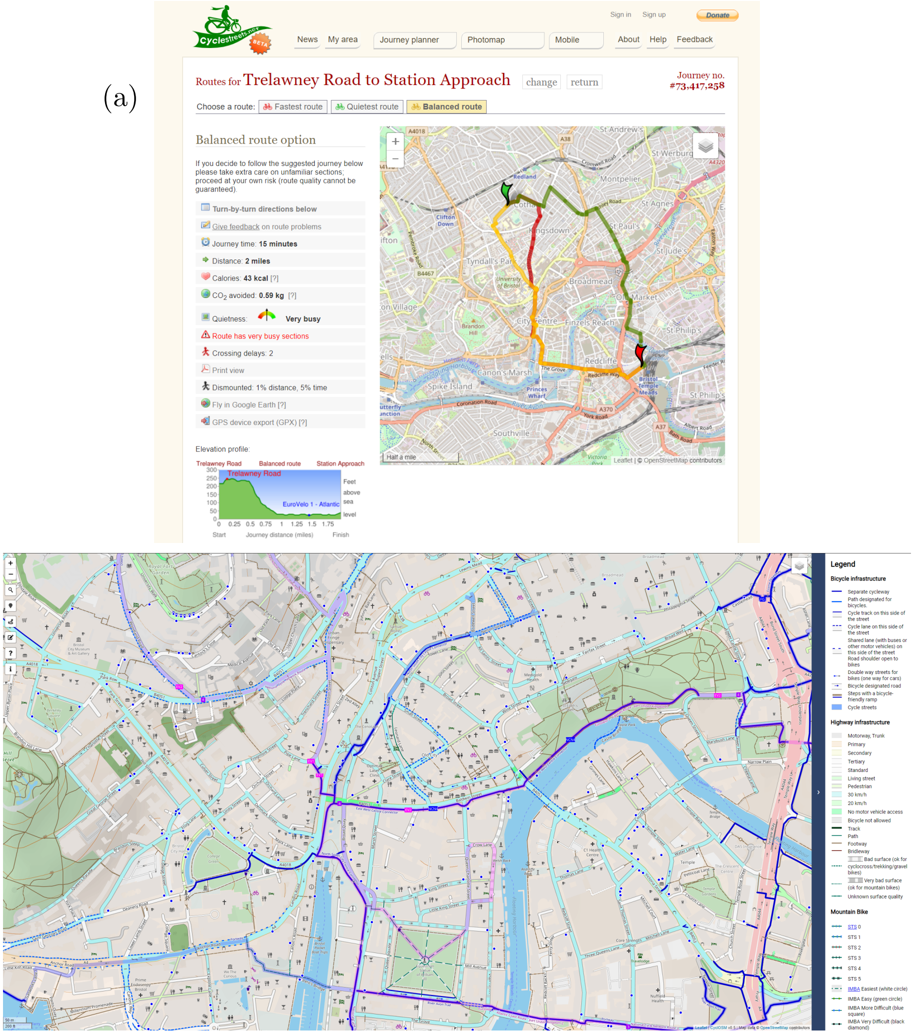

Figure 2.1: Extracts from two alternative data sources. (a) gives an example of the functionality provided by CycleStreets [17]. Shown is journey planner output from the student housing area in Redland to Bristol Temple Meads. (b) gives an example of the map provided by CycleOSM [18]. Shown is a section 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.

Both Cyclestreets and CycleOSM are simply built from OSM data, a map from which can be seen in Figure 2.2. This fact, combined with the assortment of python packages available for OSM analysis are the reasons why OSM is selected as the primary data source for this project.

All 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 2.3. The extract contains the data for just two cycle ways. Clearly these files are difficult to work with using only the base python functionality, as in reality there are thousands of ways in the full Bristol road and cycle network. Therefore, a python package OSMnx [20] 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 required to build the cycle network.

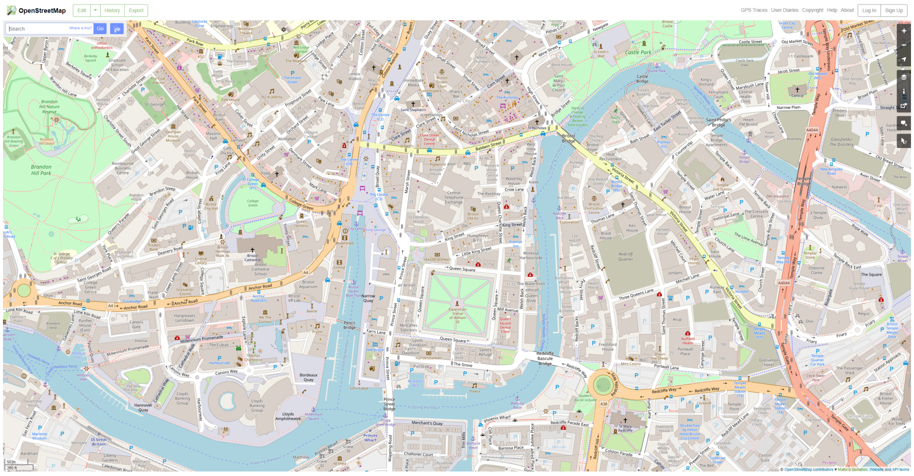

Figure 2.2: An example of the map provided by OSM [19]. Shown is a section of Bristol zoomed in to a point at which all of the extra features of this mapping tool are seen. This tool serves as the basis for the other two. This combined with a set of python packages available for it makes OSM the mapping tool of choice for this project.

Bristol Network	
Raw XML size	220 MB
Converted GraphML size	20.5 MB
Number of nodes	23,856 (OSM nodes representing curvature have been removed for this project)
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.

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 Nomanatim 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. However, we maintain the correct arc length of each street despite the node removal process. The key statistics for Bristol's street network are shown in Table 2.1.

Extract from Bristol cycle network xml file


```xml
<?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 2.3: An extract from the Bristol OSM .xml file showing the data for two cycle ways within Bristol. One field that requires explanation is the prow_ref field. This field gives the public right of way code for the way [21] but is used mainly for governance purposes and is not relevant to the work presented in this report.

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 cycle paths. The OSM tags of interest are 'Highway', 'Cycleway' and 'Bicycle', although not all of these are present for every way in the network. Our conditions for a way to be identified with designated cycling infrastructure 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

Road network of Bristol

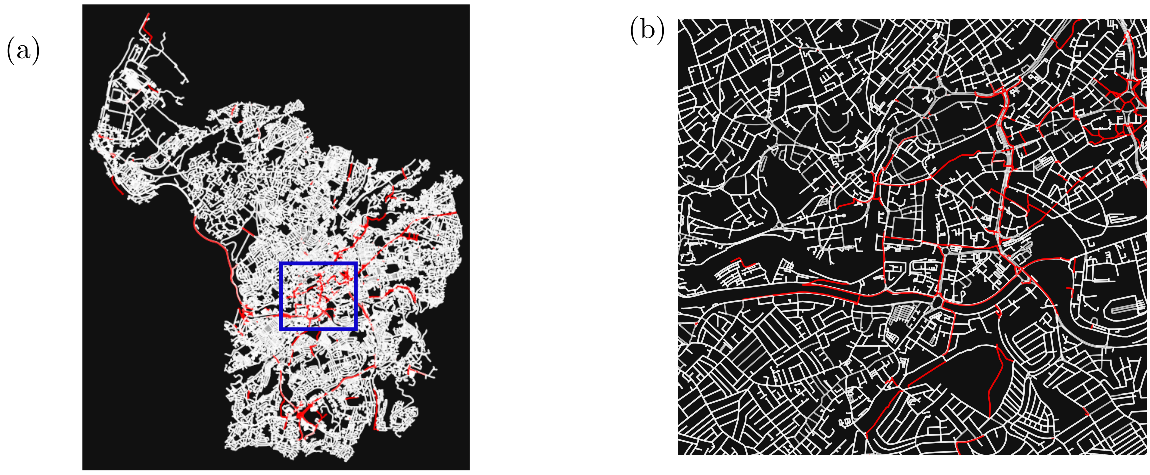

Figure 2.4: Road network of Bristol with edges (i, j) whose $\chi_{i,j} = 1$ highlighted in red. A section in the centre of (a) is seen zoomed in (b) to show the finer detail of the network.

j which takes values

$$\chi_{i,j} = \begin{cases} 1, & \text{if edge } (i,j) \text{ is designated with cycling infrastructure,} \\ 0, & \text{otherwise.} \end{cases}$$

The complete Bristol street network with edges taking value $\chi_{i,j} = 1$ highlighted can be seen in Figure 2.4 with a zoomed-in section showing fine detail.

In our approach 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

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

which means that impedance is zero for a street with cycling infrastructure and is maximal (one) otherwise.

Each putative cyclist k will 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

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

which means that an edge without cycling infrastructure has a larger perceived length than its true length, to a degree that relates to the cyclist's disposition.

The idea 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 cars etc. To achieve this, we model each cyclist as a rational agent who minimises their perceived route length, meaning where possible using exclusively cycle paths unless the detour required is larger than the length penalty of using a a road that is not designated with cycling infrastructure. 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 [2]. 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 thus have a large value for $\omega_{k'}$, as it is highly unlikely that a learner would want to cycle without dedicated infrastructure, unless it is completely unavoidable.

2.3 Uniformly random demand model

Naturally the simplest way of loading the network with cycling demand is to sample origin-destination (OD) pairs randomly from the set of all nodes within the network. In the literature, there are other more sophisticated models for demand, such as the gravity model [22], which models the number of trips between two points as proportional to their populations and inversely proportional to the distance between them. In its most simple form the gravity model may be written as

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

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

$$A = \sum_i \sum_j \frac{T_{i,j}^{\mathrm{emp}}}{P_i P_j d^{-2}}. \tag{2.3}$$

Here $T_{i,j}^{\mathrm{emp}}$ is the number of recorded trips between point i and point j in the empirical flow data used to train the model parameters.

The gravity model can be adapted to capture a key aspect of cycling, namely that cyclists tend to only travel relatively 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. This is a complicated process, and therefore we take the view that uniformly random OD pairs are enough for illustrative purposes in this chapter. In Chapter 3 we investigate a better demand model in which we use empirical census data on commuting journeys.

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. The network package OSMnx is built on top of NetworkX, and we use its built-in implementation of Dijkstra's shortest-path algorithm [23]. In larger network cases it may be necessary to use a route finding heuristic such as A* [24] 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 desire, we simulate trips for different types of cyclists on the Bristol network, to see whether a decreased propensity to cycle leads to higher usage of a city's network of designated cycling infrastructure, even though such ways will not constitute shortest routes according to geographical distance. We consider a variety of cyclists with various propensity values $\omega_k \in [0, 2, 10, 25]$ with larger values representing more nervous cyclists. To repeat: the idea is that trips for more nervous cyclists use more designated cycling infrastructure than those of confident cyclists.

Figure 2.5 shows a comparison of the percentage of trip length using designated cycling infrastructure for cyclists with a variety of propensity to cycle parameters. In (a) (ignoring the huge peak at 0%) the proportion of route length spent within the cycle network seems to decay exponentially, with the vast majority of routes spending less than 15% of their length on cycle infrastructure. This result makes sense given the disconnected 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), (c), and (d), we can see that shortest paths (measured by perceived length) tend to use much more cycling infrastructure. This makes sense and confirms the validity of our modelling approach: a cyclist who is very nervous is likely to take detours just to use designated infrastructure. It is clear that we are approaching a limit to the effectiveness of the propensity to cycle parameter as both (c) and (d) are very similar distributions with very different ω_k values. This analysis serves to demonstrate the effect of ω_k on route choice. For instance, an ω_k value of zero means that cyclists choose purely the shortest route between OD pairs. Whereas, an ω_k value of ∞ means that routes are restricted to purely the cycle network where possible. Therefore, the analysis performed later in this project is done with an ω_k value of 2 to simulate true cyclist behaviour, choosing cycle paths where possible given that the required detour is not too large.

The peak around 0% in both all of the figures is due to the random selection of OD pairs giving rise to extremely short routes in areas 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 parameter.

2.4 Comparison with second bounding geography (Amsterdam)

A second bounding geography, Amsterdam (see Figure 2.6) is also presented in order to compare summary statistics on a more highly connected cycle network. The key network statistics for Amsterdam are given in table 2.2. Amsterdam is selected as a bounding geography due to its much more comprehensive cycle network and 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, because OSM data is populated by users and without a dedicated team it is not possible for OSM to be completely up to date. The idea is that Amsterdam's cycle network's higher degree of connectivity should yield a higher percentage of time spent on cycle paths within shortest routes for a cyclist of any given propensity to cycle. 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.

Figure 2.7 (a) shows that it is the case that in Amsterdam routes use more cycling infrastruc-

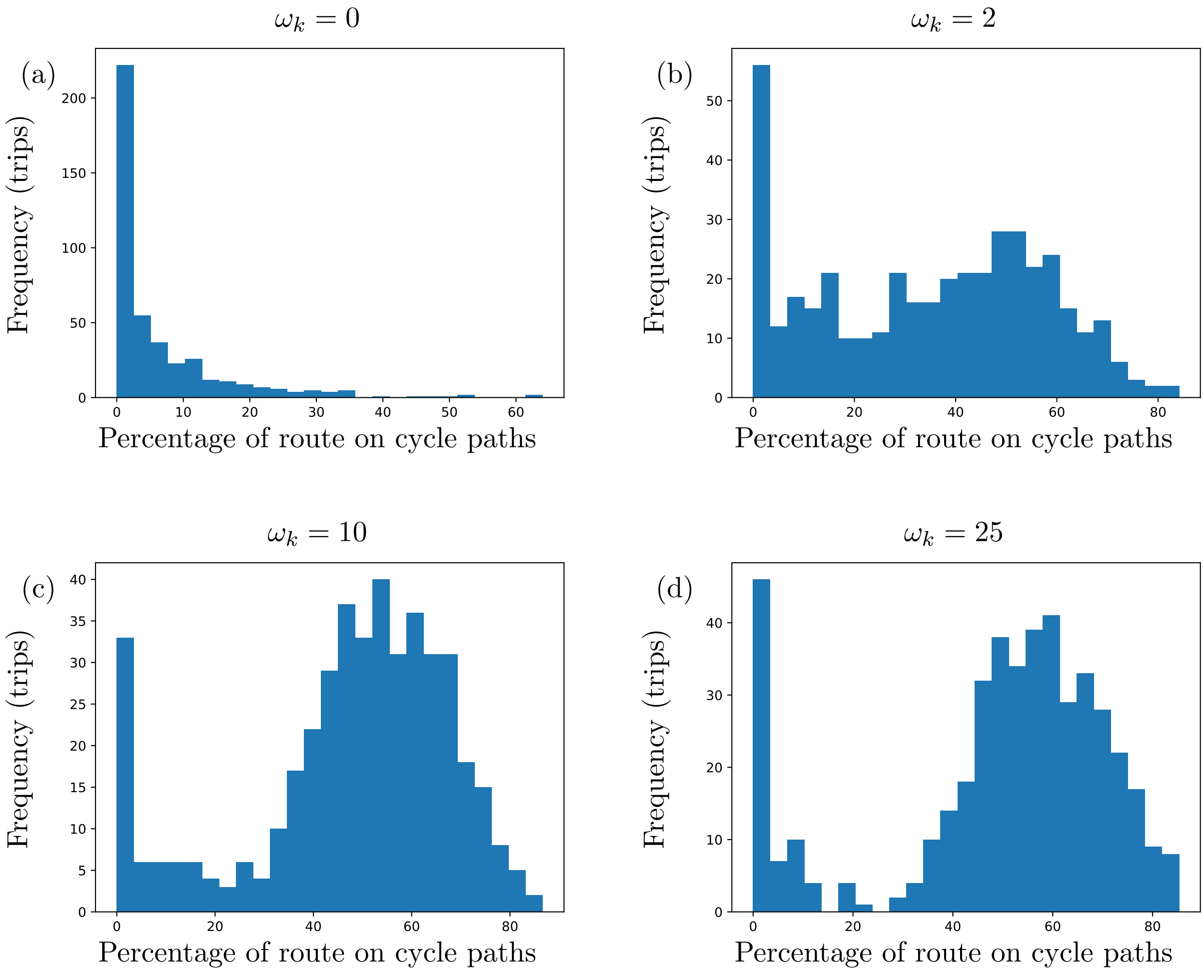

Figure 2.5: Distributions of the percentage of length of routes on designated cycling infrastructure for cycle trips on the Bristol network, using uniform random OD selection, with $\omega_k = 0, 2, 10, 25$ in (a)-(d) respectively. A larger value of ω_k results in routes with a larger proportion of their length on cycling infrastructure.

Amsterdam Network	
Raw XML size	825 MB
Converted GraphML size	24 MB
Number of Nodes	27,101 (OSM nodes representing curvature have been removed)
Number of Edges	65,751
Total street length	2689 km
Total length of cycling infrastructure	1035 km (Includes: painted lanes, segregated lanes and, off-road cycleways)

Table 2.2: Key statistics of the Amsterdam street network.

ture 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. Further computations, see in Figure 2.7 (b), have shown that as ω_k is increased for cyclists in Amsterdam, most routes are nearly 100% on cycle paths. Figure 2.7 (c) shows the

Figure 2.6: The road network of Amsterdam with edges that are designated with cycling infrastructure highlighted in red.

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.

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 chapter are as follows:

- A model for cyclist route choice by adapting perceived edge length using a cyclist's propensity to cycle.
- Using our case study city (Bristol), we verify that cyclists with low propensity to cycle do indeed tend to use more cycle paths than those with a high propensity to cycle.
- By comparison of two case study cities (Bristol and Amsterdam), we show that proportion of trips spent on cycle paths is a good measure of the quality of cycling infrastructure.

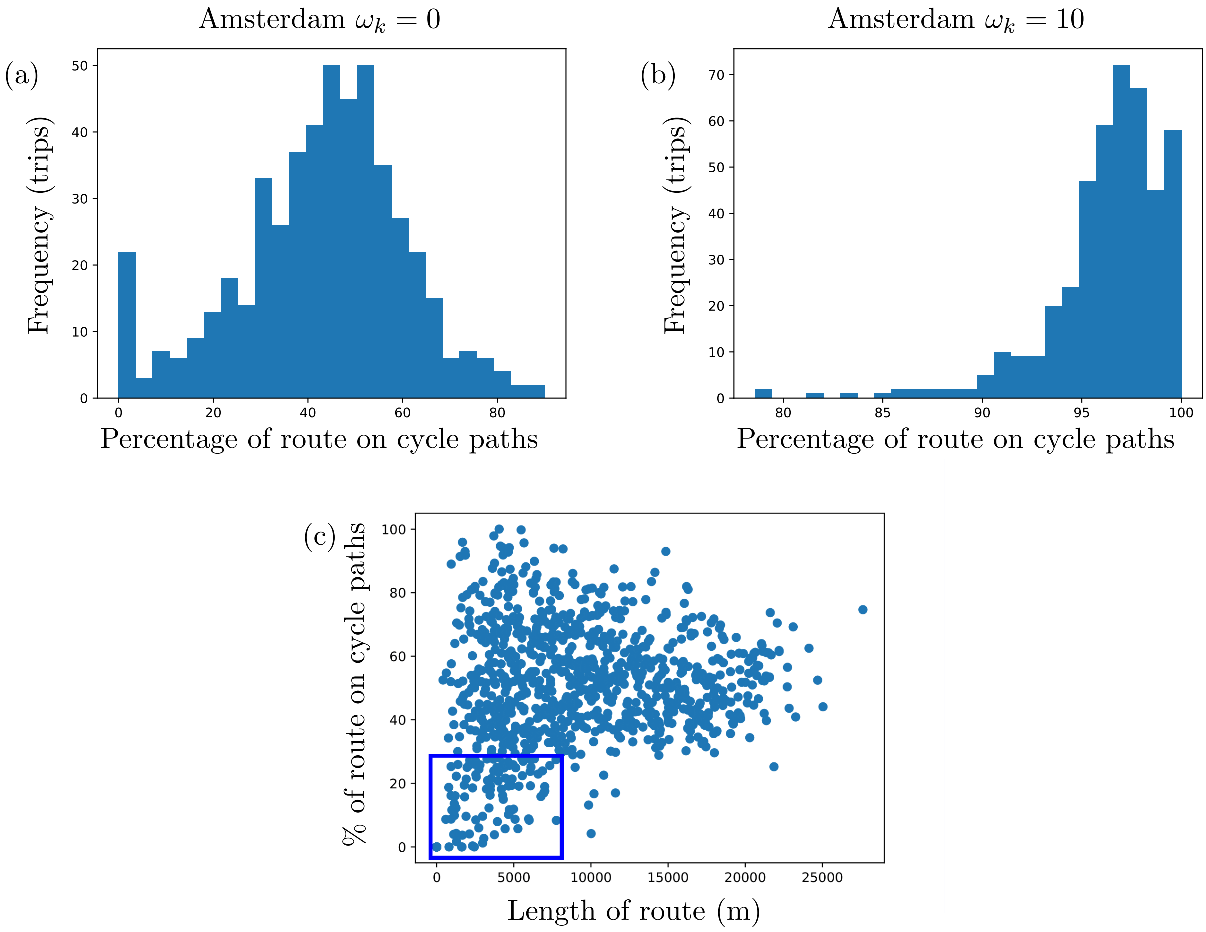

Figure 2.7: (a) distribution showing that the more connected city of Amsterdam (with $\omega_k = 0$) has a similar percentage of routes on cycle infrastructure to that of Bristol with very low propensity to cycle cyclists (figure 2.5 (a)). (b) shows that as we increase ω_k to an extreme value (e.g., 10) almost all cycling trips in Amsterdam use designated cycling infrastructure for 100% of their length. (c) shows a scatter demonstrating the relationship between route length and percentage of route on designated 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. The blue box indicates the very short routes that yield very low percentage of route length on designated cycling infrastructure.

Chapter 3

Network Upgrade Heuristics

The goal of this project is to improve a city's cycle network, by building designated cycling infrastructure to ultimately yield a more connected cycle 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 city road network as it is essentially the network design problem which has been well established as NP-complete in the literature [25]. In it's simplest form one must check all 2^n configurations of designating edges with cycling infrastructure to find which is optimal within budget, 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 cities by applying the heuristic to our case study city (Bristol). We start (Section 3.1) by detailing the heuristic upgrade process and introducing its parameters. We then introduce (Section 3.2) a methodology for loading city street networks with cycling demand using empirical census data. Next we (Section 3.3) perform some experiments on our case study city (Bristol) and introduce a scoring metric to measure how the upgraded network will improve cycling in a given city. Then we (Sections 3.4 and 3.5) investigate the impact of varying the heuristic's parameters on the resulting upgraded network of our case study city (Bristol). Finally (Section 3.6) we discuss the real world interpretations of the heuristic and its parameters.

3.1 Overview of the heuristic

Figure 3.1 gives a brief overview of the approach this chapter will take to upgrade city street networks with new cycling infrastructure and Table 2.1 briefly summarises the heuristic's parameters. The heuristic starts by estimating cycling demand using empirical census data on commuter flows in the city. 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 cycling infrastructure, according to their edge flows, 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 cycle network which will be more

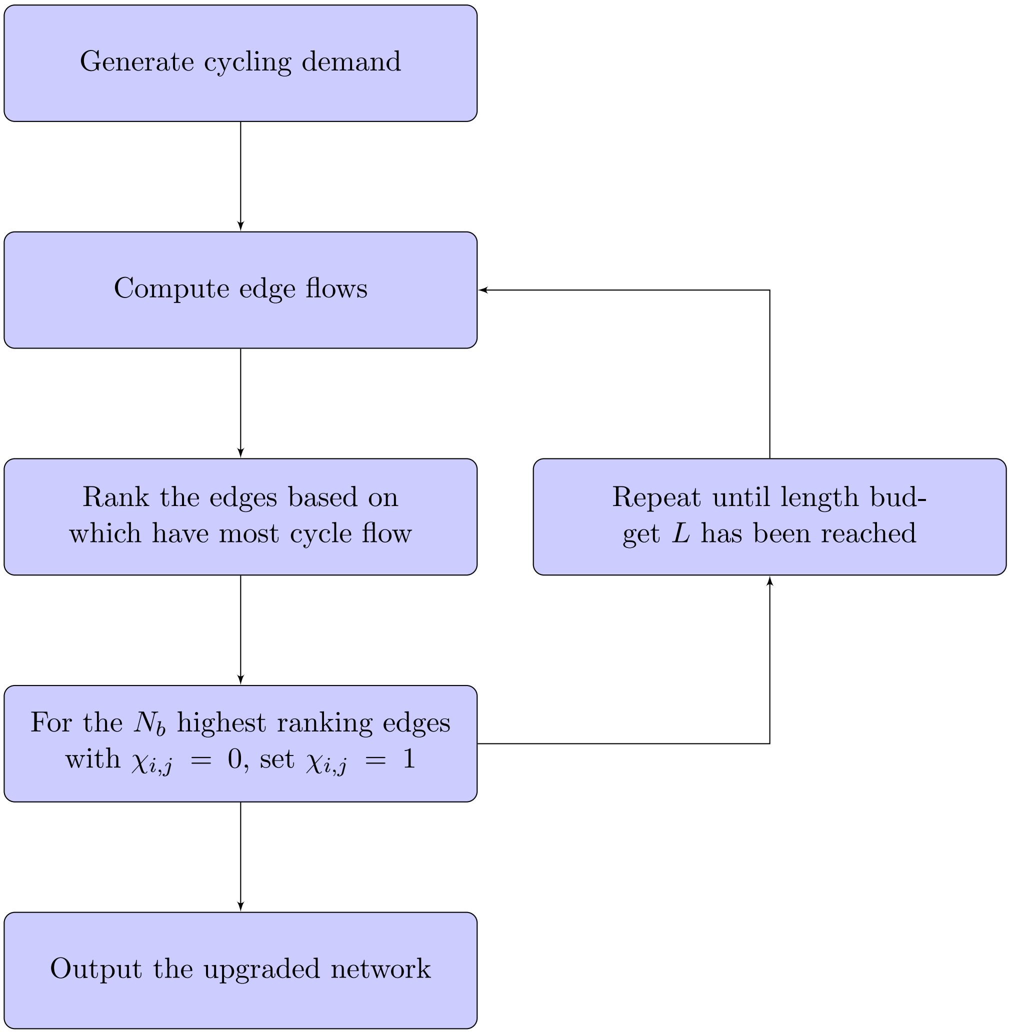

Figure 3.1: Summary of the heuristic upgrade approach proposed in chapter 3.

desirable for cyclists.

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 effectively upgrade a city's cycle network 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. 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

Key parameters	
Parameter	Definition
L	The total length edges to be upgraded. In our 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 our case study city Bristol [26].

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 the bounding geography used for Bristol in this report. 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 case study city, giving us good coverage of Bristol but without running into issues with sampling bias.

The methodology detailed for loading demand is inspired by the DataShine commute web tool. They use MSOA weighted centroids and model flow between these MSOAs from centroid to centroid across the UK. Figure 3.2 shows a screenshot from the DataShine web page with flows highlighted to and from one of the MSOAs within our case study city. The colour of the line indicates direction with blue flowing in to the centroid and red flowing out of the centroid. The thickness of the line then illustrates the amount of journeys flowing between the two centroids. In our use case we are only concerned with the demand structure of a given city so we use the same centroid methodology to simulate flows between all LSOAs in the city.

The data-set is downloaded from the ONS website [28] by creating a query and selecting place of work and usual residence from a list of LSOAs. The analysis of this project requires all LSOAs in a given city which can be selected from a drop down list. We then need the locations of the centroids of the LSOAs in our city. A data-set containing the longitude and latitude coordinates

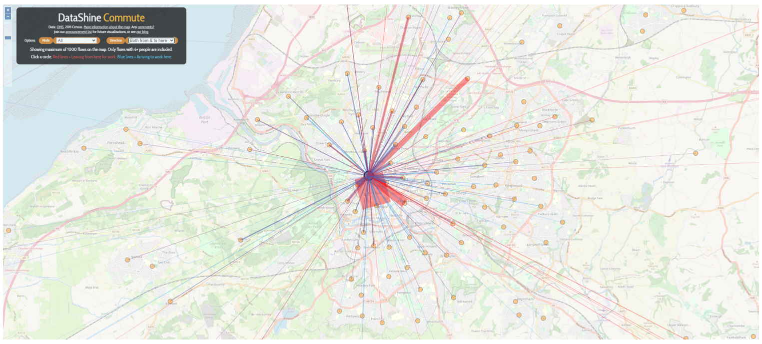

Figure 3.2: A screenshot of the DataShine commute interactive map showing the flows to (blue) and from (red) one of Bristol's 55 MSOAs [27].

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

for the population weighted centroids for each LSOA in the UK is used [29]. These coordinates are then used to map the centroids to nodes in the python network structure of our city by matching node IDs in both data-sets. The resulting centroid 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 the assumption of all LSOA flow being generated at its centroid means we do not consider 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.

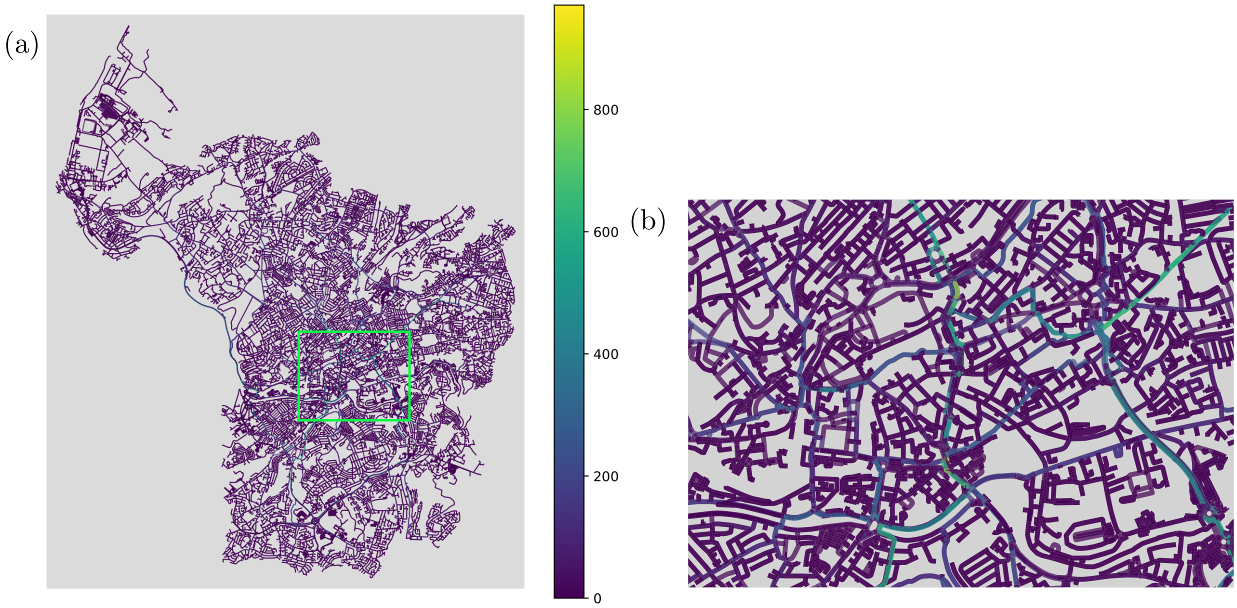

Figure 3.4: (a) map of Bristol's road network with edges coloured based on their cycle flow value. (b) a section in the centre of Bristol (green) to show one of the main stretches of street with high cycle flow. The edges are coloured according to how many journeys use them.

This matrix allows us to compute a probability vector

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

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 our case study this is 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 according to the probability vector $P(f_i)$. The shortest route between the OD pair is then computed by Dijkstra's algorithm, 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 value is then used to inform edge upgrade priority.

Figure 3.4 shows the Bristol road network with edges coloured corresponding to their flow in a trip simulation of size $N_t = 5000$. 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. This fits with a commuting model as we would expect most commuters to travel to work in the city centre.

Using empirical data gives us a more accurate representation of the true flows of commuters within a large city than simply using a random OD pair assignment. However, there is one concession made in that we are restricting the OD matrix to only contain nodes representing

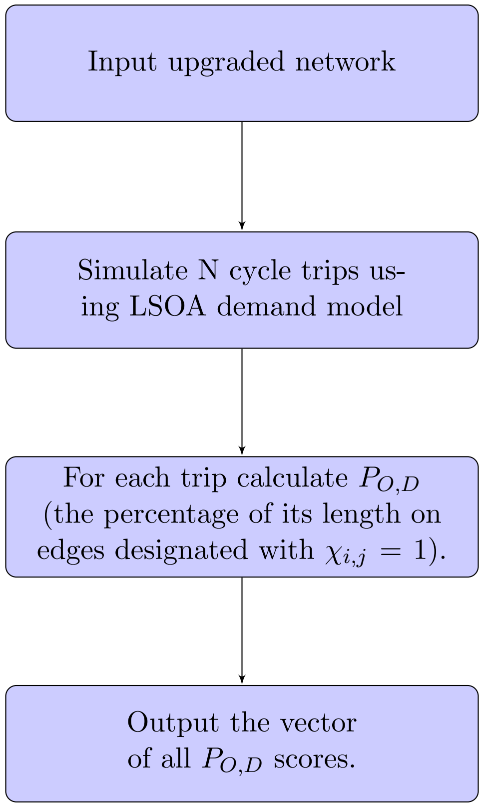

Figure 3.5: Overview of the scoring metric.

the centroids of LSOAs. This means we may not get total coverage of the city street network. It is also worth noting that by assuming all flows originate at a centroid we will be artificially increasing the cyclist flows on any edge directly connected to a centroid.

This demand loading methodology works well for a British city such as Bristol, which has comprehensive flow data available from 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 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 of the heuristic's impact on the Bristol street network.

3.3 Assessment of heuristic

To assess the user experience on an output upgraded city network a scoring metric similar to the network assessment in chapter 2 is used, although it is formalised here. The methodology is as 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) to assess the ability of our heuristic to propose networks conducive to a better cyclist experience.

Figure 3.6 shows the upgraded network found by implementing the heuristic on the Bristol street network with parameters: $L = 160$km, $N_t = 500$, $N_b = 20$, and $\omega_d = 2$.

The heuristic is designed in such a way that it prioritises building infrastructure on areas of the network with large flows of cyclists. Figure 3.6 clearly shows that the main spokes of the Bristol road network that were highlighted with high flow counts in Figure 3.4 are all upgraded

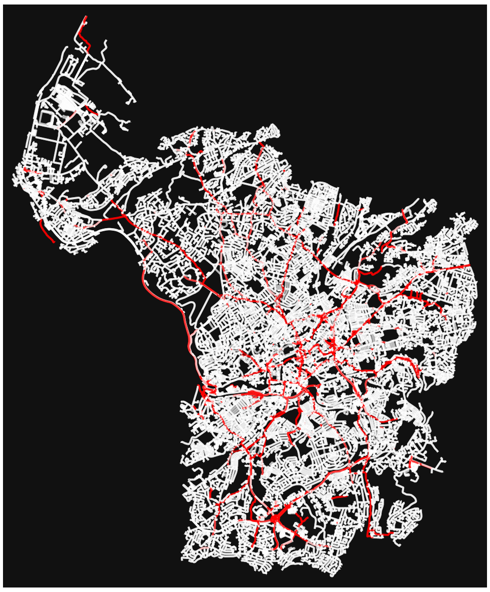

Figure 3.6: The output upgrade network for a batch process with parameters $L = 160$km, $N_t = 250$, $B = 20$ and $\omega_d = 2$.

with cycle paths if they are not already designated with cycling infrastructure. This is because these edges are the main thoroughfare for cyclist trips between LSOAs so have high flow rates and will, therefore, 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 upgraded 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 length that cyclists spend on designated cycling 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. Figure 3.7 (c) supports this claim with the blue box indicating that routes with lower $P_{O,D}$ scores tend to have shorter route length.

The value of L in this case study experiment is 160km which means that after the upgrade has been performed there is more than twice the cycling infrastructure than in the original Bristol street network. This will obviously yield an increase in the percentage of the length of cycling trips on designated cycling infrastructure. Therefore, the question we ask is whether the upgrade heuristic performs markedly better than randomly upgrading the network to contain twice as

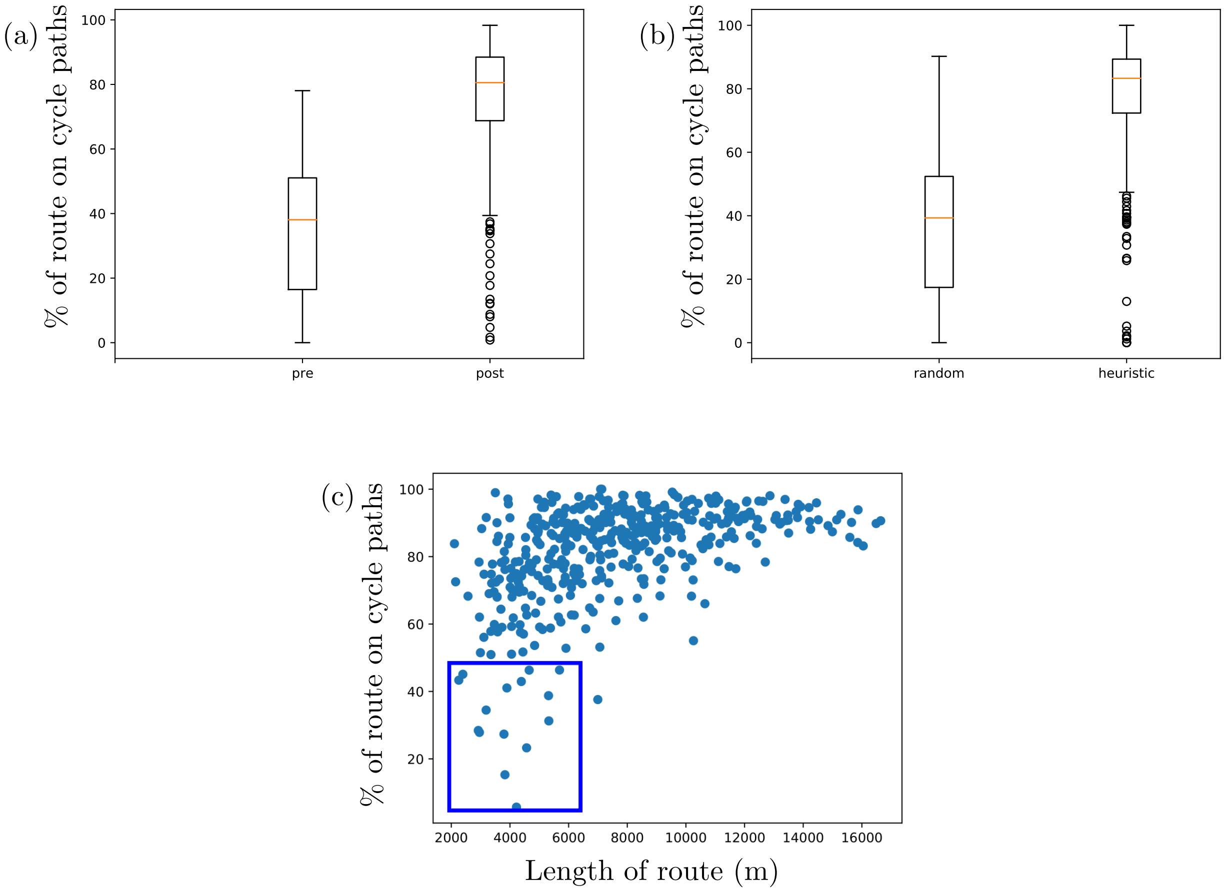

Figure 3.7: The distribution of $P_{O,D}$ scores for (a), the Bristol network before and after applying the upgrade heuristic, and (b), the upgrade heuristic vs randomly assigning the same length of upgraded streets. (c) gives the relationship between route length and $P_{O,D}$ scores on the upgraded Bristol network. The blue box indicates the section of the plot that supports the argument that routes with low scores tend to be shorter without the option to use cycling infrastructure.

much cycling infrastructure. To do so we upgrade randomly sampled edges from 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 cycling infrastructure by upgrading the same number of edges. This finding suggests that the more connected cycle network proposed by the upgrade heuristic does, in fact, yield better cyclist satisfaction than a disconnected cycle network with the same amount of infrastructure.

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. This will then lead to larger flow counts on streets attached to the new infrastructure and will lead to those edges being prioritised for upgrade in the next batch, ultimately yielding upgrades along contiguous paths of edges.

To model the pull of new infrastructure the length of edges upgraded with cycling infrastructure are 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, is now a shorter alternative for the cyclist if it uses edges with designated infrastructure. 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 more connected cycle network should be generated.

One issue with large numbers of batches is that, to simulate the effect of new infrastructure pulling cyclists to it, we must generated a new set of cycling trips for each batch. If we use large N_t and large B the heuristic can become computationally expensive to generate these trips for every new batch. 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 $L = 40$km. The one-shot network (a) is computed by upgrading all 40 km of street 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 $N_t = 2500$ simply to account for sampling error, which is not an issue for the single-edge method, which uses $N_t = 100$ as trip flows can 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) shows that by using the single-edge method infrastructure is suggested in this area of the Bristol network 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, see Figure 3.8, between the suggested networks from both methods there is not necessarily a marked metric improvement for cyclist experience by choosing a batch process.

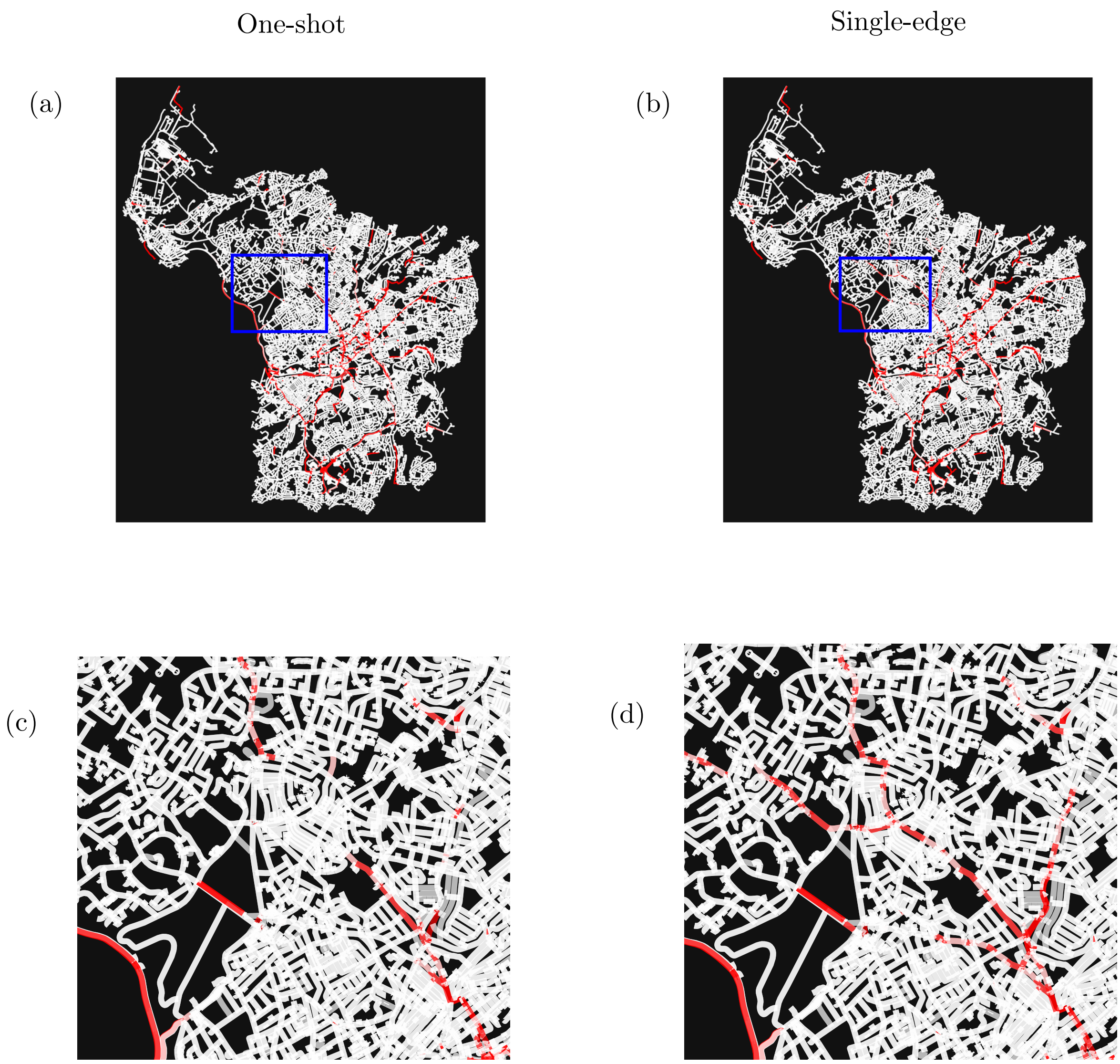

Figure 3.8: Road network of Bristol with edges (i, j) whose $\chi_{i,j} = 1$ highlighted in red. a) shows the network after a one-shot process has been applied to upgrade $L = 40$km of street and b) shows the result of a single-edge scheme upgrading $L = 40$km also. (c) and (d) show the same zoomed in 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.

Figure 3.9 (b) shows the resulting $P_{O,D}$ distributions from a sweep of values for B, in our case study upgrading $L = 160$km of street in Bristol. It shows there is actually very little sensitivity 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 in Bristol, see Figure 3.4. 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 city networks with less well defined demand spokes may have higher sensitivity to the batch size parameters, but the computation presented in this

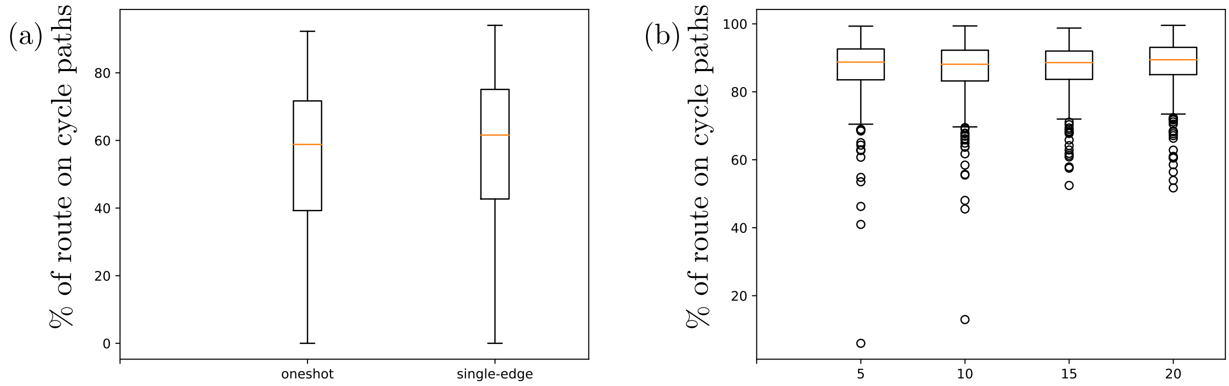

Figure 3.9: The distribution of $P_{O,D}$ scores for (a), the upgraded network for one-shot and single-edge, and (b), when varying the number of batches B.

project is only for our case study city (Bristol).

One property of the batch process that should be investigated is that of overshoot. That 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 a city street network, it is not possible to have only a portion of an edge designated with infrastructure and without overshoot very small batch budgets may yield no upgrade at all. 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. 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 city 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 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 as $\omega_k = 2$ 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

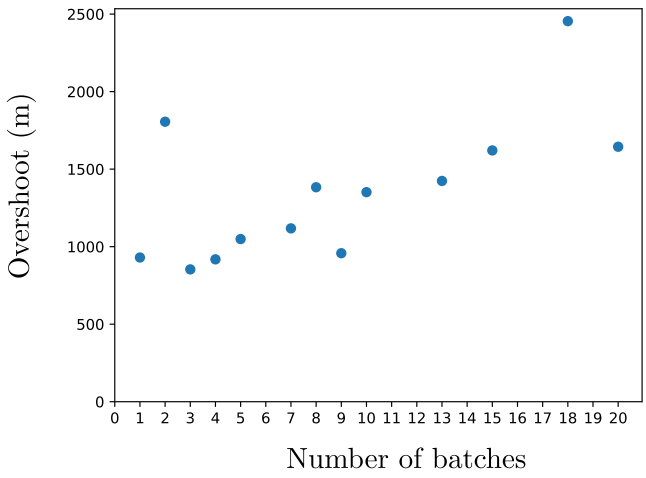

Figure 3.10: Scatter showing the potential for overshoot when designing networks with larger numbers of batches. Points show the overshoot on a post upgrade target of 320km of street designated with cycling infrastructure to double the amount in our case study city (Bristol).

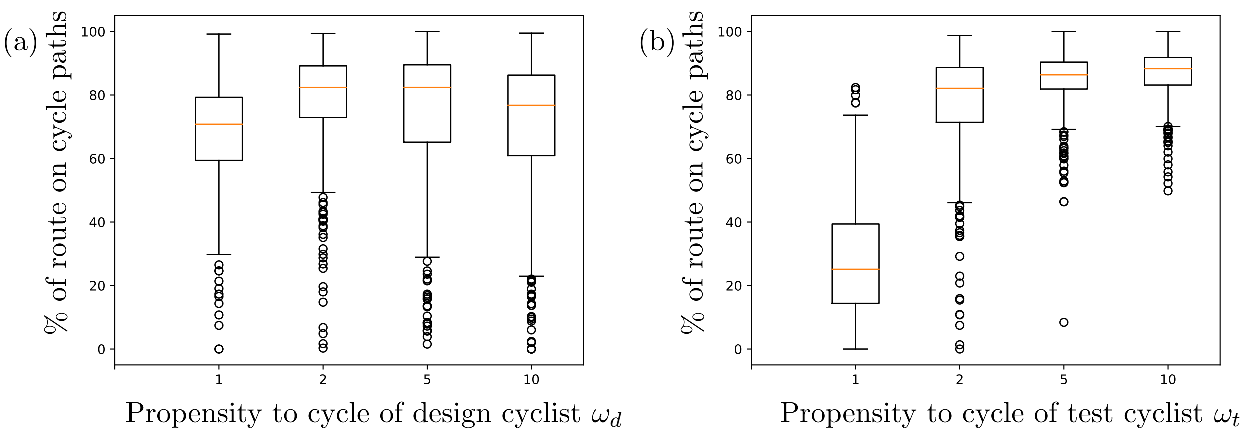

Figure 3.11: The distribution of $P_{O,D}$ scores for (a), the upgraded Bristol network for different values of ω_d with $\omega_t = 2$, and (b), the upgraded Bristol network for different values of ω_t with $\omega_d = 2$.

a different propensity ω_t. For instance, 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 observe the effect that varying the propensity of the test cyclist in the same range $w_t \in [1, 2, 5, 10]$ has on our $P_{O,D}$ metric. 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. This route may not fall on suggested cycling infrastructure which tends to be built onto the end of existing infrastructure which was not on the shortest route for cyclists in the original network. We also find that, as expected, when ω_t is increased the cyclist 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, as cyclists will not take very large detours to use designated infrastructure. This 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 choice is supported by the findings of Krizek et. al. [30] who found that on average cyclists are willing to travel up to 67% further than the true shortest path to use designated infrastructure.

3.6 Interpretation as street improvement over time

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 the heuristic 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 nervous 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.

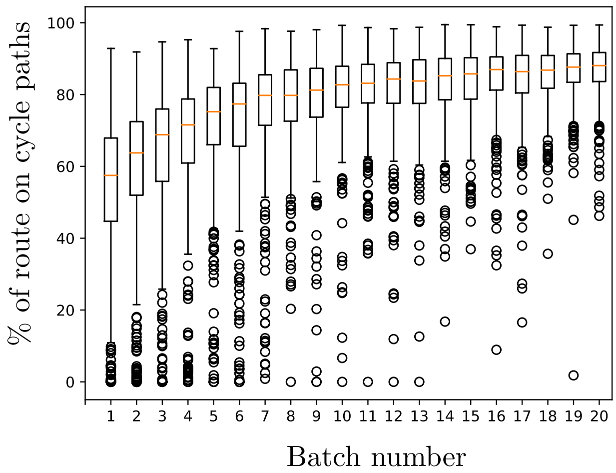

Figure 3.12: Graph showing the improvement in $P_{O,D}$ scores throughout the batches of a 20 batch upgrade of the Bristol network. It shows that most of the upgrades are made in the early batches.

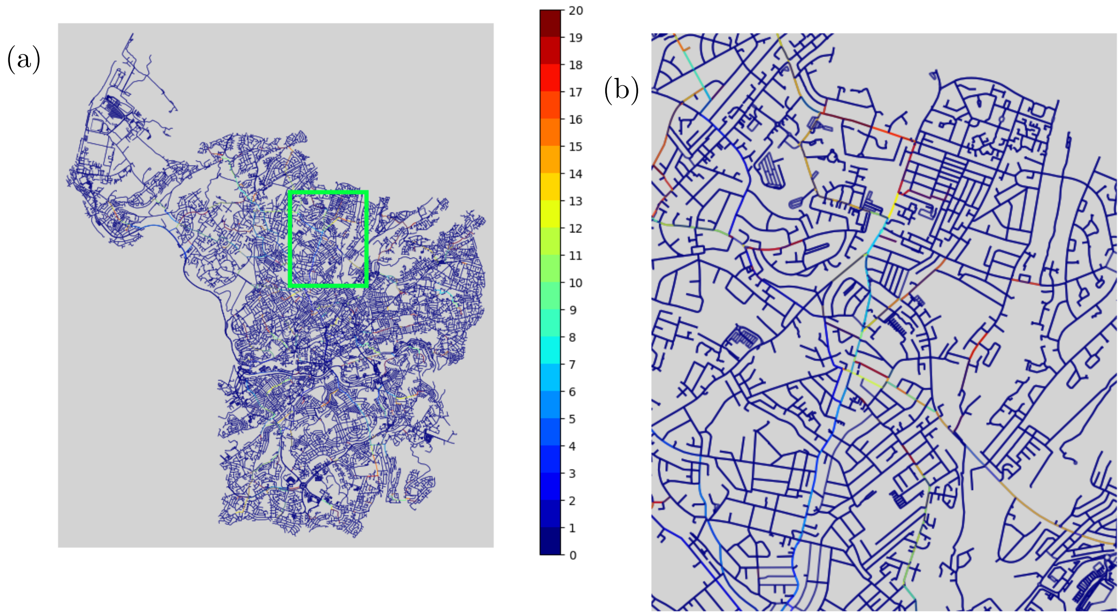

Figure 3.13: In (a) we see the map of Bristol's road network with upgraded edges coloured based on which batch they are upgraded in. The brighter the edge the later it is upgraded. (b) shows a zoomed in segment in the north east of Bristol to show that cycle paths tend to grow out from the city centre (the edges further north are upgraded in later batches).

Figure 3.13 shows the order in which the upgrades are performed. Edges are coloured based 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 the locations of LSOA centroids which correspond to the small gains in $P_{O,D}$ spread seen in Figure 3.12. Figure 3.13 (b) is a zoomed-in section of the top of the Bristol network to show up close the ability of the heuristic to build contiguous paths. In the plot we can see that the next batch's upgrades tend to be attached to the previous batch's yielding a continuous cycle path. This, when seen over the whole network, means that the upgrade heuristic tends to suggest connected cycle networks which is the goal of this project.

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.
- Using our case study city (Bristol) we have verified that the upgrade heuristic does, in fact, provide a visual and metric improvement to a city's cycle network.
- We have demonstrated the effect of a couple of key model parameters on the resulting upgraded Bristol street networks. For example, we have shown that the best results are found when matching ω_d and ω_t.

Chapter 4

Synthetic Network Experiments

This chapter explores the effects of the heuristic proposed in Chapter 3 on small synthetic networks. We start (Section 4.1) by adapting both the heuristic and scoring metrics to a new synthetic framework. Next (Section 4.2), we introduce a very simple cross network and perform some experiments on it to highlight the impact of key parameters of the heuristic in an example for which those impacts are easily visualised. Then (Section 4.3), we introduce a process for generating ensembles of random networks using beta-skeleton pruning to reduce the number of paths within the network. Finally (Section 4.4), we sweep over a few key parameters to show the impact they have on the average metric scores of upgraded networks within the random ensemble.

4.1 Heuristic adaptations

The key difference between the OSM street networks used in Chapters 2 and 3 and the synthetic networks proposed here, is that, in this case, we formulate demand as a continuous value assigned to an OD pair rather than using an agent based formulation. The upgrade approach used here is the same as detailed in Figure 3.1 but with flows allocated as continuous variables for all edges in shortest routes using the specified OD demands.

To assess upgraded networks we again need a scoring metric. Again it is worth noting that, in these smaller examples, demand is framed as a continuous value for each OD pair rather than being agent based as in Chapter 3. The networks are scored with

$$S = \frac{\sum_{(i,j)\in OD} d[i,j] \times \frac{\text{Length of cycle paths}(i,j)}{\text{Length}(i,j)}}{\sum d}. \tag{4.1}$$

This function outputs a value in the range $S \in [0, 1]$ with 0 meaning that the shortest route for all OD pairs uses no cycling infrastructure and 1 meaning that all shortest routes between OD pairs use exclusively cycling infrastructure. With these adaptations made in the next section we introduce an example network and illustrate the impact of the heuristic upgrade process on it.

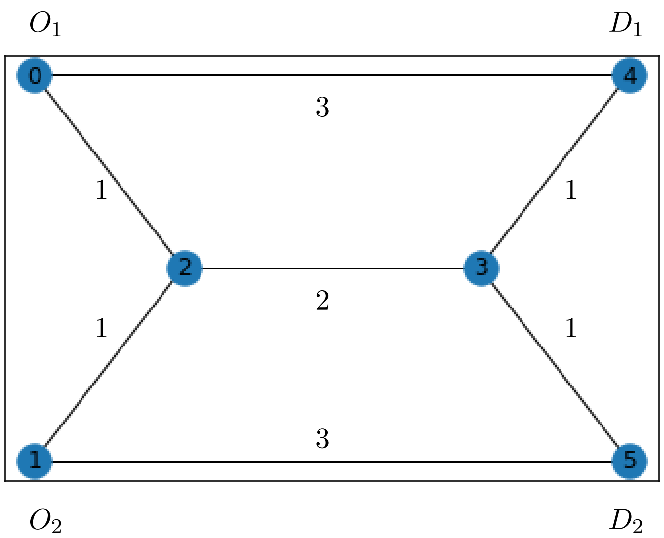

Figure 4.1: Cross network setup.

4.2 Cross network experiments

To illustrate the impact of this project's proposed network upgrade heuristic the cross network is used. To keep things as simple as possible a symmetric network design is chosen which can be seen in Figure 4.1.

The network contains 4 OD pairs ($OD = [(0,4) \quad (0,5) \quad (1,4) \quad (1,5)]$) whose demands are also symmetric and are described in the demand vector

$$d = [d_{04} \quad d_{05} \quad d_{14} \quad d_{15}] = [1.5 \quad 1 \quad 1 \quad 1.5] \tag{4.2}$$

so that there is slightly more demand from top-top and bottom-bottom than in the cross network cases. During the graph setup each edge is given a tag called 'cycle' whose usage is described as

$$G[i \quad j][\text{'cycle'}] = \begin{cases} \text{True}, & \text{if edge } (i,j) \text{ is designated with cycling infrastructure,} \\ \text{False}, & \text{otherwise.} \end{cases}$$

Again for simplicity, we initialise the graph with no pre-existing cycling infrastructure. The goal is then to upgrade the network such that half of the overall length of the network is designated with cycling infrastructure, in this case the upgrade budget is $L = 6$. As established in Section 2.3 and confirmed in Section 3.5, a choice of $\omega_k = 2$ is appropriate to encapsulate the route choice of a general cyclist. Now that we have established the setup of the problem to be analysed, we explore the impact of batching on the resulting upgraded cross networks.

In Section 3.4 we introduced the idea of splitting the upgrade budget into batches to see if it had any impact either visually or metrically on the output network for our case study city Bristol. In that case, there was very little difference between the metric scores of the two networks but there was a small visual difference, with batching providing a more connected output cycle network. Therefore, it is worth exploring the impact of batching on our cross network as we can easily see the difference in the output networks and even interpret the steps

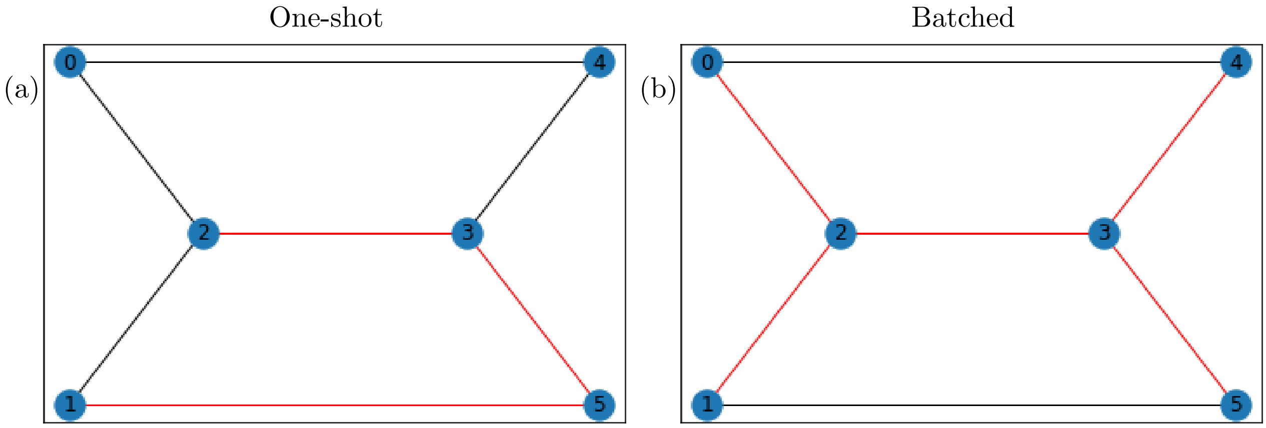

Figure 4.2: (a) the output network resulting from upgrading the budget $L = 6$ in one-shot it is worth noting any combination of the middle edge a long edge and a short edge can result from random selection to break edge flow ties. (a) the output network resulting from upgrading the budget $L = 6$ in two batches.

behind the algorithms output.

Figure 4.2 shows the output networks for two cases: (a) upgrading the cross network in one batch, and (b) upgrading the cross network in two batches. Two batches is the only feasible batching case for the cross network as anything more will mean that the top and bottom 'long' edges could not be upgraded as they would not fit within the batch budget. Visually the key take away is that the one-shot network does not achieve a symmetric design leading to low satisfaction for many cyclists. In this setup it is worth noting that any combination of the middle edge, one short edge, and one long edge is a potential output from a one-shot approach, but this combination is selected simply due to index order in the edge flow sorted list.

Lets take the one-shot network and detail the heuristic steps that give this output: 1. Shortest routes are computed for each OD pair [(0,4), (0,2,3,5), (1,2,3,4), (1,5)], 2. Edges in those shortest paths are assigned with the respective OD demand, 3. Since we upgrade in one batch the batch budget is 6, 4. Edge (2,3) has the highest flow of 2 so is the first edge assigned with cycling infrastructure, 5. Edges (0,4) and (1,5) now share the next highest flows with 1.5 and (1,5) is chosen by numerical ordering in the ranking list, 6. Edges (0,2), (1,2), (3,4), and (3,5) now all share the final and lowest flow but are the only edges short enough to fit in budget so edge (3,5) is the final edge selected for upgrade, and finally 7. After upgrade the shortest paths according to propensity weighted length are [(0,2,3,4), (0,2,3,5), (1,2,3,4), (1,5)].

Now for the two batch upgrade case: 1. Shortest routes are computed for each OD pair [(0,4), (0,2,3,5), (1,2,3,4), (1,5)], 2. Edges in those shortest paths are assigned with the respective OD demand, 3. Edge (2,3) has the highest flow of 2 so is the first edge assigned with cycling infrastructure, 4. Since the batch budget is 3 the next highest ranking edges (0,4) and (1,5) are too long to be upgraded so edge (3,5) is chosen for upgrade, 5. The new shortest routes according to propensity weighted length are now [(0,2,3,4), (0,2,3,5), (1,2,3,4), (1,2,3,5)], 6. Edges (0,2), (1,2), and (3,4) are now the highest ranking edges according to flow and their length satisfies the batch budget and hence they are all upgraded, and finally 7. This upgrade step does not change the shortest paths in the output network which are still [(0,2,3,4), (0,2,3,5), (1,2,3,4), (1,2,3,5)].

Upgrade Method	Score
One-shot	0.55
Two batch	1.0

Table 4.1: The results from the scoring metric on both the one-shot and two batch upgraded networks.

The two networks in Figure 4.2 are then assessed with the scoring metric outlined in equation 4.1 and the results are displayed in Table 4.1. These scores clearly show that, in the case of the cross network, upgrading in batches gives a significant improvement over upgrading in one-shot. This is due to the ability of a batch process to simulate the pull of new cycle paths for cyclists. This means that the shortest paths and therefore, the edge flows have the opportunity to change between the batches yielding a connected cycle network. The output network for two batches allows for all shortest routes to be entirely encapsulated within the cycle network that has been suggested. In the next section we compare this result to a brute force attempt to solve the optimal network design problem on the cross network.

As we discussed in Chapter 3, the heuristic proposed here is not strictly optimal. It is worth exploring the optimal network design for the synthetic network to establish whether, in this small test case, our heuristic can approximate optimal design. To find the optimal network design according to our scoring metric we should consider all possible combinations of designating cycling infrastructure to the edges within the synthetic network i.e, its power set. The power set of a network's edge set grows as 2^n for an edge set of size n. Clearly this makes solving for the optimal network impossible on larger networks but fortunately it is possible for the cross network whose power set contains $2^7 = 128$ edge combinations. We then search over all of these combinations for the highest scoring network that fits within our upgrade budget.

In Section 4.2 we established that a two batch upgrade process produces an output network that achieves the maximum score of 1 for our metric. It is no surprise then, that formal optimal design also yields the network shown in Figure 4.2 (b). This means that for this specific example it is possible for our proposed heuristic to yield formal optimal network design. In the next section we will investigate the performance of the upgrade heuristic on randomly generated networks according to two metrics. We will also compare smaller random networks to their formal optimal design to see if the results reported in this section hold more generally.

4.3 Generating random ensemble networks

- talk about network generation using delauny triangulation and beta-skeleton pruning.
- introduce the parameters of the networks and set some initial values
- maybe discuss the SUE stuff here but possibly later

4.4 Experiments on the ensembles

- Compare one vs two batch for these ensembles and plot the differences

- Experiment with different design and test omega values
- explore the impact of lambda which introduces non-shortest routes

Chapter 5

Conclusions and Further Work

In Chapter 1 we identified three key research questions to be investigated throughout this report: 1. Can a simple model of propensity to cycle give a good approximation of cyclist route choice? 2. Can we use a simple heuristic to inform cycle network upgrades in a given city? and 3. How close does the heuristic approach come to formal optimal network design?

In Chapter 2 we investigated the first of these questions. We developed a single parameter model for cyclist route choice using their individual propensity to cycle. Through simulation on the Bristol street network (Section 2.3) we showed that as you increase a cyclist population's propensity to cycle (i.e, cyclists are more nervous) the proportion of routes that use designated cycling infrastructure increases. We identified the extreme value of this parameter as for any value of $\omega_k \geq 10$ cyclists use exclusively cycling infrastructure if at all possible. Analysis concluded that a choice of $\omega_k = 2$ is a good model for the average cyclist as it allows for short detours from their shortest route to use designated infrastructure but will not lead to unnecessarily long detours this choice was later confirmed (Section 3.5). We then assessed this model on a more connected cycle network (Amsterdam) to confirm that $\omega_k = 10$ is an extreme value and in that case nearly all cyclist routes have 95% or more of their length on designated infrastructure. Therefore we conclude that for both: 1. a sparsely connected and 2. a well connected cycle network, a model of route choice using propensity to cycle is a good model for cyclist behaviour.

In Chapter 3 we investigated the potential of a simple upgrade heuristic to inform network design on large scale city networks using our case study city Bristol. We showed (Section 3.3) that by applying our upgrade heuristic to the Bristol network cyclists spend on average 80% of their route on cycling infrastructure compared with only 40% for the current Bristol network giving a huge improvement to cyclist satisfaction. We also demonstrated that the heuristic provides a significant improvement over upgrading the same length budget using a random assignment showing that the connected networks generated by our heuristic provide a better experience for cyclists. We have also shown that, in our case study city Bristol, varying parameters such as the amount of batches or the design ω have very little metric impact, although we cannot confirm that this is the case universally. We conclude that the heuristic proposed in this report does inform cycle network design in our case study city and we suggest that it could be applied to other cities.

In Chapter 4 we investigated the ability of our heuristic to approximate formal optimal network design, and the impact of batching our upgrade process using some very small synthetic network setups. Using a carefully designed cross network (Section 4.2) we demonstrated that a better network design is achieved using a batched upgrade and in fact this batched process achieves the same output network as formal optimal network design. This means that, at least in the case of small networks, we can say that our heuristic does not only approximate optimal design but can in fact achieve it.

Finally we investigated into the impact of batching for ensembles of randomly generated networks ...

There are some areas of this project for which we identify the potential for further work. Firstly, in modelling cyclist route choice it may be worth investigating adding extra terms to the formula for perceived length to model other key factors in cyclist route choice e.g., incline or speed limit. Secondly, we suggest applying the upgrade heuristic to other city street networks where the existing infrastructure is more well connected or there does not exist significant demand spokes to investigate the effect of these factors on the output network design. Finally, it would be pertinent to explore the idea of using SUE assignment to allow demand to be spread out over multiple short routes rather than the pure shortest route and to see if this has a significant impact on the suggested network design of a city.

Bibliography

[1] C. Allan. Cycling UK's cycling statistics. https://www.cyclinguk.org/statistics. Accessed 2020-12-02.

[2] J. Dill and T. Carr. Bicycle commuting and facilities in major US cities: if you build them, commuters will use them. *Transp. Res. Rec.*, 1828(1):116–123, 2003.

[3] 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.

[4] 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.

[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 2021-03-28.

[6] The rise in popularity of e-scooters - what is the legal position in the UK? https://www.lexology.com/library/detail.aspx?g=32fce955-08a1-4e83-8b91-412e48c74894. Accessed 2021-04-01.

[7] Government guidance on powered transporters. https://www.gov.uk/government/publications/powered-transporters/information-sheet-guidance-on-powered-transporters. Accessed 2021-04-01.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] A. Mauttone, G. Mercadante, M. Rabaza, and F. Toledo. Bicycle network design: model and solution algorithm. *Transp. Res. Rec.*, 27:969–976, 2017.

[13] P. Freestone. Personal communication. October 2020.

[14] 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.

[15] R. Lovelace and J. Talbot. Rapid cycleway prioritisation tool: Technical specification. https://www.cyipt.bike/rapid/. Accessed 2020-12-08.

[16] DFT. The cycling infrastructure prioritisation toolkit. https://www.cyipt.bike/report1/, 2018.

[17] Cyclestreets main webpage. https://www.cyclestreets.net/. Accessed 2021-03-29.

[18] CycleOSM map of Bristol. https://www.cyclosm.org/#map=13/51.4515/-2.5657/cyclosm. Accessed 2021-03-29.

[19] OSM map of Bristol. https://www.openstreetmap.org/#map=13/51.4740/-2.5889. Accessed 2021-03-29.

[20] G. Boeing. Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Comput. Environ. Urban*, 65:126 – 139, 2017.

[21] Osm wiki page decribing the prow_ref tag. https://wiki.openstreetmap.org/wiki/Key:prow_ref. Accessed 2021-04-19.

[22] George Kingsley Zipf. The p 1 p 2/d hypothesis: on the intercity movement of persons. *American sociological review*, 11(6):677–686, 1946.

[23] Documentation for Networkx's shortest path function. https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.shortest_paths.generic.shortest_path.html#networkx.algorithms.shortest_paths.generic.shortest_path. Accessed 2021-04-01.

[24] Description of the A* algorithm. https://en.wikipedia.org/wiki/A*_search_algorithm. Accessed 2021-04-01.

[25] D.S. Johnson, J.K. Lenstra, and A.H.G. Rinnooy Kan. The complexity of the network design problem. *Networks*, 8:279–285, 1978.

[26] 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 2021-03-29.

[27] 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 2021-03-21.

[28] Data table for census commuting data. (Query from here). https://www.nomisweb.co.uk/query/construct/summary.asp?reset=yes&mode=construct&dataset=1228&version=0&anal=1&initsel=. Accessed 2021-04-06.

[29] Download the population weighted centroids of lsoas across the UK. https://geoportal.statistics.gov.uk/datasets/b7c49538f0464f748dd7137247bbc41c_0?geometry=-30.065%2C50.531%2C25.812%2C55.170. Accessed 2021-04-06.

[30] 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.