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 30, 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 inform network upgrade decisions. The model of route choice increases the perceived length of streets without designated cycling infrastructure according to a cyclist's personal propensity to cycle factor. Next, we formulate a simple heuristic that upgrades streets in a city network by equipping them with cycling infrastructure, based on the volume of cyclists in a simulation. We apply this heuristic to the Bristol network and show that it increases cyclist satisfaction according to a metric, which describes the proportion of a journey that is spent on cycling infrastructure. We investigate the application of the heuristic to small synthetic networks where its actions can be more easily understood. In particular, we conclude that in certain examples it is possible for our proposed heuristic to achieve optimal network design. Finally, we generate ensembles of random networks using beta skeleton pruning and after upgrading these networks and scoring them we conclude that on average there is a small gain to be made by switching to a two-batch upgrade scheme rather than upgrading in one shot.

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, 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 approaches to this problem. One approach, also formulated by Milakis et al. [10], uses 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.

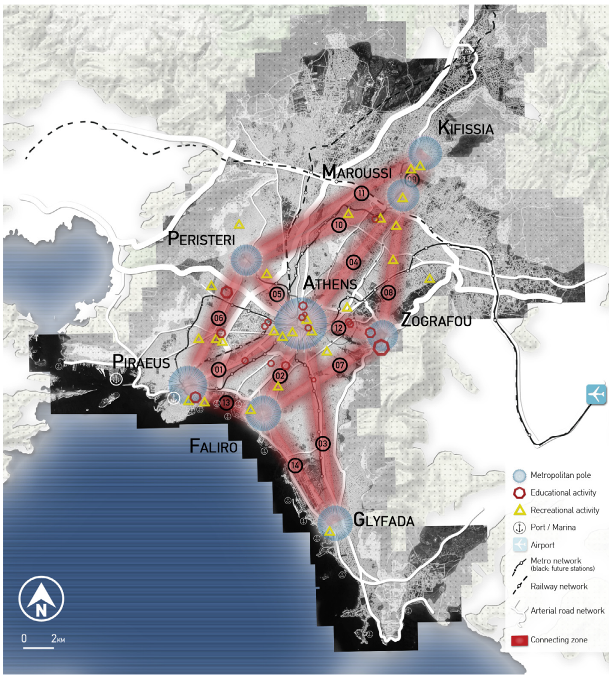


Figure 1.1: Demand centres (blue) in Athens. Reproduced from *What about people in cycle network planning? Applying participative multicriteria GIS analysis in the case of the Athens metropolitan cycle network*, by Milakis et al. (2014) [10].

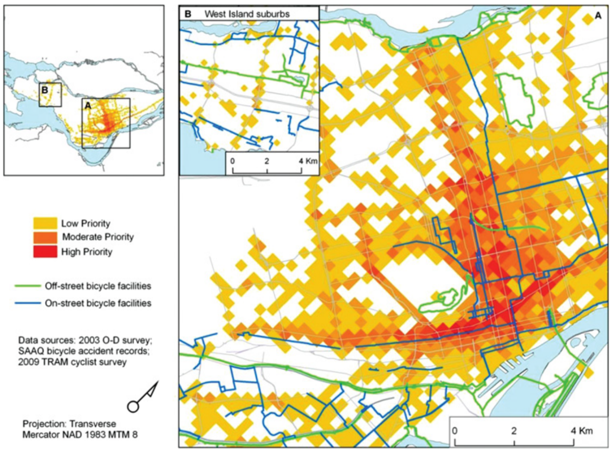


Figure 1.2: Prioritisation grid of Montreal. Reproduced from *Build It. But Where? The Use of Geographic Information Systems in Identifying Locations for New Cycling Infrastructure*, by Larsen et al. (2013) [11].

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 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 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 can achieve the optimal design. Finally we generate ensembles of small synthetic networks to explore whether or not batching the upgrade process gives better upgrade designs.

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 cycle network is more expansive and OSM data is more complete.

2.1 Data sources

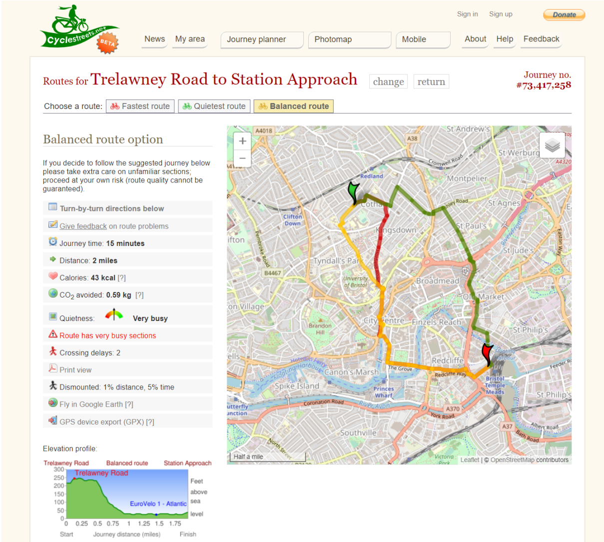
To create the network needed for the analysis proposed in this project, multiple data sources were investigated for conversion into a network representation for analysis. Three such sources have been identified: Cyclestreets [17], CycleOSM [18], and OSM [19].

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 described in very broad terms 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.

Both Cyclestreets and CycleOSM are built from OSM data, a map from which can be seen

(a)



(b)

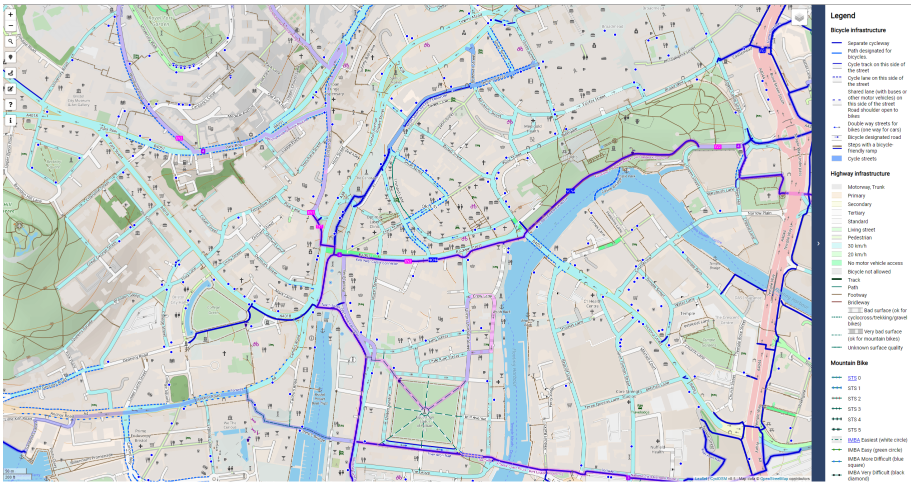


Figure 2.1: Extracts from two alternative data sources. (a) Example of the functionality provided by CycleStreets [17]. Shown is journey planner output from the student housing area in Redland to Bristol Temple Meads. (b) Example of the map provided by CycleOSM [18]. Shown is a section of Bristol zoomed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

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 specialised 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.

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 =

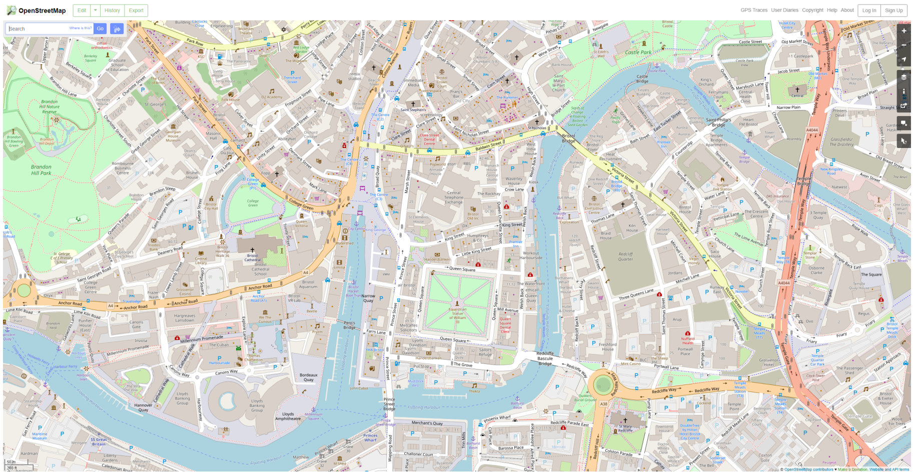


Figure 2.2: Example of the map provided by OSM [19]. Shown is a section of Bristol zoomed in to a point at which all of the extra features of this mapping tool are seen.

Table 2.1: Key statistics of the Bristol street network.

Bristol Network	
Raw XML size	220 MB
Converted GraphML size	20.5 MB
Number of nodes	23,856
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)

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.

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 as having 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'.

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. The Prow_ref field gives the public right of way code [21] used mainly for governance purposes and is not relevant to this report.

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

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

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

Road network of Bristol

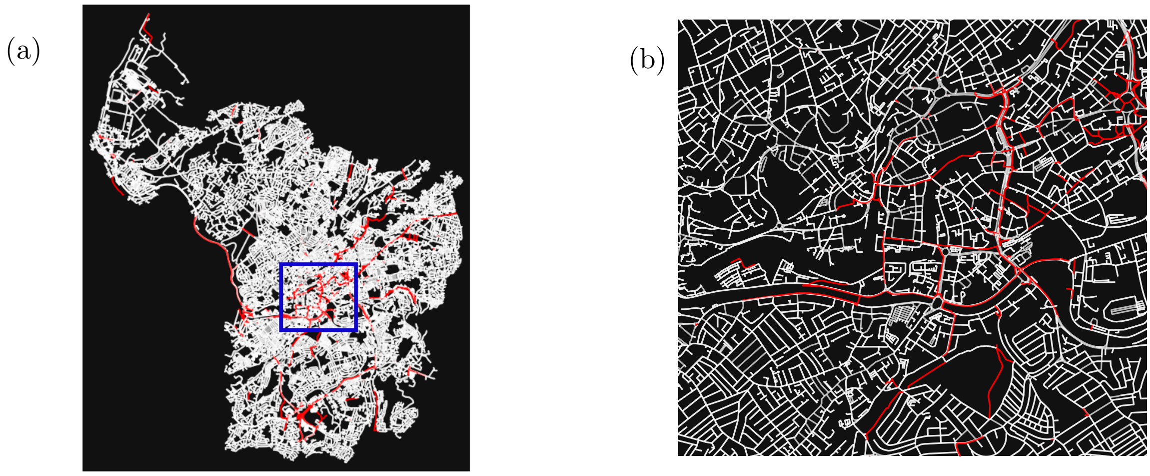


Figure 2.4: Road network of Bristol with edges (i, j) whose $\chi_{i,j} = 1$ highlighted in red. A section in the centre of (a) is seen zoomed in (b) to show the finer detail of the network.

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 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

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

Here $T_{i,j}^{\text{emp}}$ 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 for 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 networks 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 did not prove to be an obstacle.

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 usually allow for complete routes on cycling infrastructure. However, when propensity to cycle is decreased in (b), (c), and (d), we can see that shortest paths (measured by

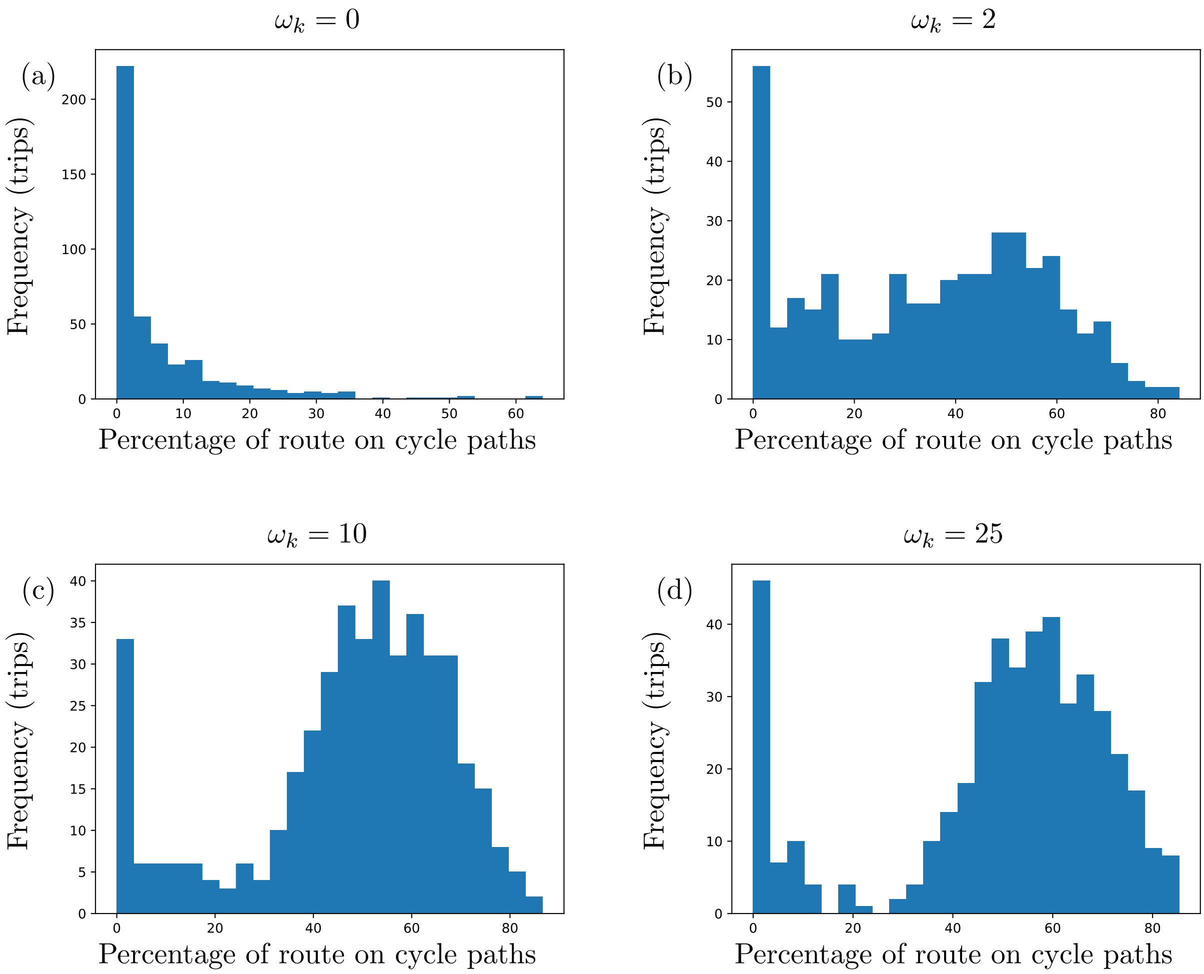


Figure 2.5: Distributions of the percentage of length of routes on designated cycling infrastructure for cycle trips on the Bristol network, using uniformly random OD selection, with $\omega_k = 0, 2, 10, 25$ in (a)-(d) respectively. A larger value of ω_k results in routes with a larger proportion of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

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 shortest routes in the standard sense. In contrast, as $\omega_k \to \infty$, routes are restricted to purely the cycle network where possible.

The default cyclist with $\omega_k = 2$ 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. [25] who found that on average cyclists are willing to travel up to 67% further than the true shortest path to use designated infrastructure.

The peak around 0% in all 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

Table 2.2: Key statistics for the Amsterdam street network.

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)

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

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 i.e., increasing ω_k.

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 it 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 infrastructure 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

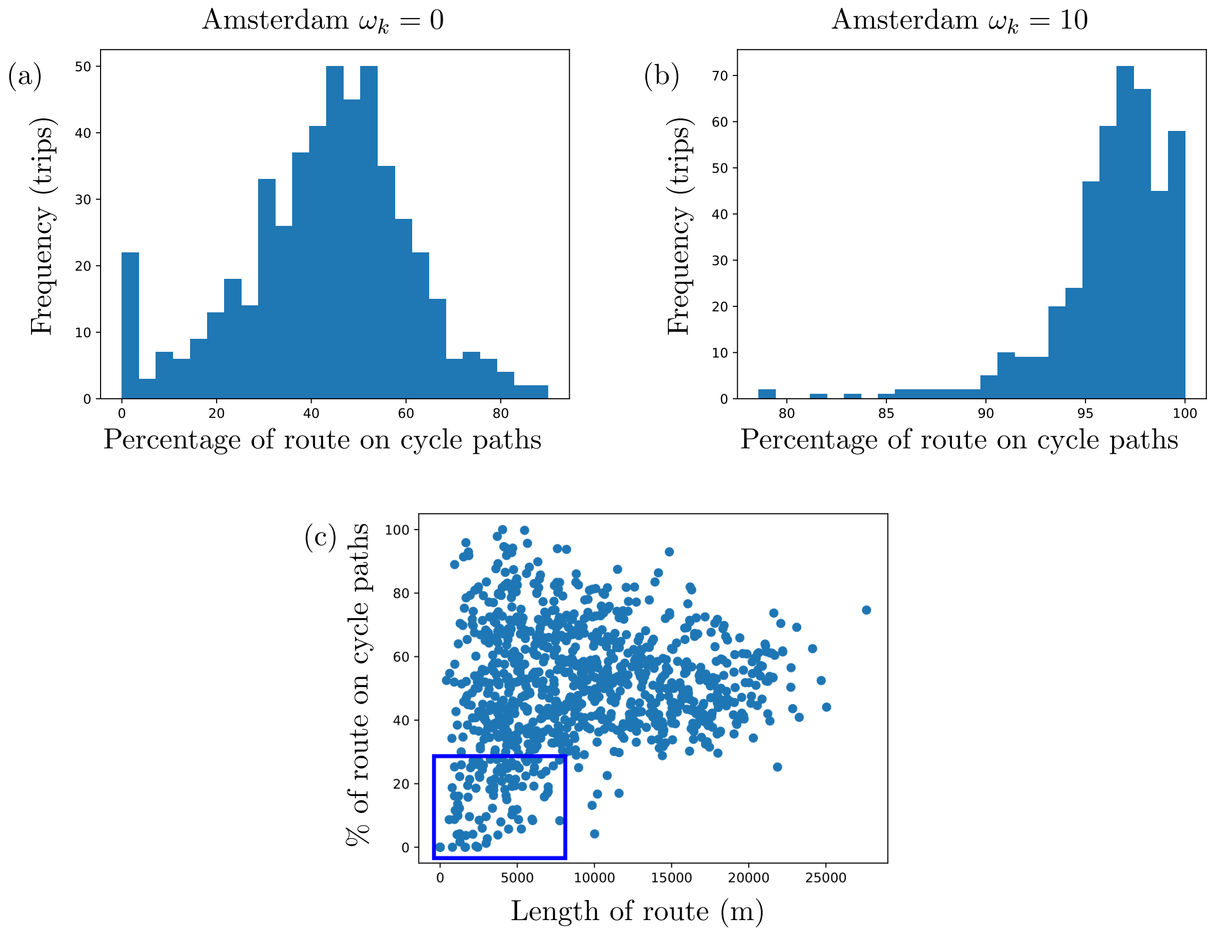


Figure 2.7: (a,b) Distributions of the percentage of length of routes on designated cycling infrastructure for cycle trips on the Bristol network for ω_k = 0 and 10 respectively. Compare Figure 2.5 (Bristol). (c) Low percentage routes tend to be very short (blue box), where it is infeasible to detour onto the cycle network.

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 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 3 km 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.

Chapter 3

Network Upgrade Heuristics

The goal of this project is to help improve a city's cycle network, by suggesting where to build designated cycling infrastructure to ultimately yield a more connected cycle network. Local authorities have a limited budget to build new cycling infrastructure, and might seek to design a network that fits within that budget whilst maximising some factor of cyclist satisfaction. This is a type of network design problem which has been well established to be NP-complete in the literature [26]. In its simplest form one must search through all 2^n possible upgraded networks where n is the number of nodes in the network. In contrast, in this chapter we develop a simple heuristic (and therefore approximate, formally non-optimal) method and test the performance of the heuristic on our case study city (Bristol).

We start (Section 3.1) by describing 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 (Section 3.3) we 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 (Sections 3.4 and 3.5) we investigate the impact of varying the heuristic's parameters, such as the number of upgrade batches or the propensity to cycle of our 'design' cyclist, to measure the impact they have, either metrically or visually on the upgraded network suggestion. Next, (Section 3.6) we discuss the real world interpretations of the heuristic and its parameters. Finally, (Section 3.7) we highlight the key contributions of the chapter towards our goal of answering the question: Can we use a simple heuristic to inform cycle network upgrades in a given city?

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 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 route flows are used to compute edge flows, that is, how many cyclists use each edge, for a trip simulation of size N_t. We then upgrade the most used edges that do not currently have cycling infrastructure, 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' (Section 3.4). The question we ask is whether, by splitting the upgrade budget into smaller batches of upgrades, we can achieve a

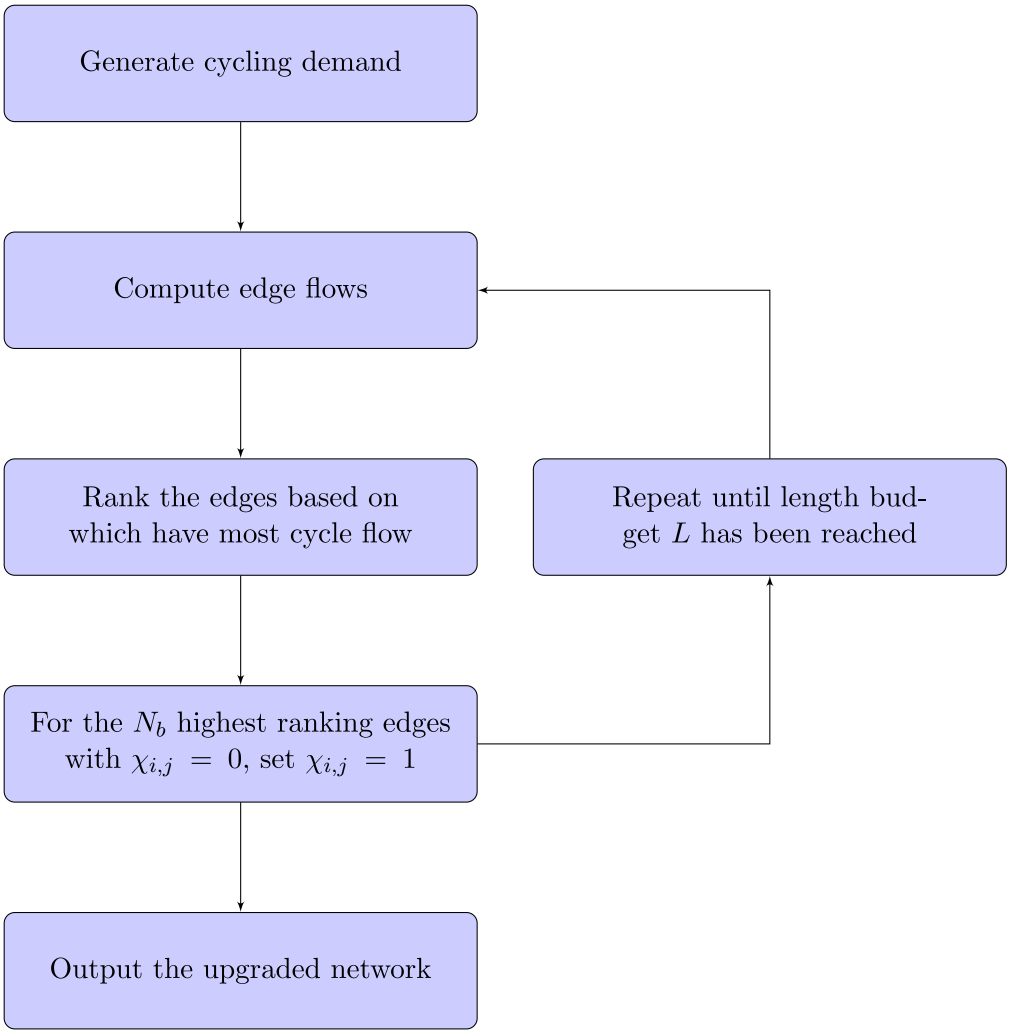


Figure 3.1: Summary of the heuristic upgrade approach.

more connected cycle network which will be more desirable for cyclists.

3.2 Cycling demand generation

Whilst uniformly randomly selecting OD pairs was sufficient to investigate the propensity to cycle model in Chapter 2, to effectively upgrade a city's cycle network a more realistic demand model is required. This section describes the use of more realistic empirical demand data. Bristol, like any large city, owes most of its traffic flow to commuters. If we can simulate commuter journeys and assume that they can be made by bicycle, we will get a more realistic model for the flows of cyclists through Bristol. The data source that we use is derived from the 2011 UK census [27] and provides commuting data at three geographical resolutions: 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 (55) in Bristol. This would not give enough variety in journey origins and destinations to effectively model the cycling population of Bristol. In contrast, there are rather too many (1,368) OAs, and the problem is that the resulting OA-OA OD matrix is sparse, because the census methodology has a minimum reporting threshold to preserve the anonymity of correspondents. It is for these reasons that

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

Key parameters	
Parameter	Definition
L	Total length of edges to be upgraded. We set $L = 160\,\text{km}$ which doubles the amount of infrastructure in Bristol.
N_t	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 LSOA centroids.
ω_d	Propensity to cycle factor for a 'design' cyclist.
B	Number of batches required to reach the upgrade budget with the current N_b.
L_b	Length budget for each batch.

Table 3.2: The census geography requirements for different output areas and the number of these areas within our case study city Bristol [28].

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

the data used for this study is at LSOA level. The resulting OD matrix is 263×263 and can be downloaded from the ONS website [27] by creating a query and selecting place of work and usual residence from a list of Bristol's LSOAs.

A data-set containing the longitude and latitude coordinates for the population weighted centroids for each LSOA in the UK is then employed [29]. These coordinates are used to map the centroids to nodes in the python network structure by matching node IDs in the two data-sets. The resulting centroid locations for Bristol are plotted in red in Figure 3.3.

It is worth noting at this point that the assumption that all LSOA flow is generated at its centroid means we do not consider flow within an LSOA: so we set the diagonal elements of the OD matrix to zero. Since LSOAs are so small this in practice only eliminates journeys that use at most a couple of edges and therefore should not impact the resulting network flows very much.

The OD matrix $\{f_{i,j}\}$ allows us to compute a probability vector for each centroid i, in the form

$$\mathbf{p}^{(i)} = \frac{1}{\sum_{i=1}^{N_c}} \begin{pmatrix} f_{i,1} & f_{i,2} & \dots & f_{i,N_c-1} & f_{i,N_c} \end{pmatrix}, \tag{3.1}$$

which gives the proportion of each flow that originates at centroid i that terminates at centroid j. Here $N_c = 263$ is the number of LSOA centroids.

To load the network with cycle flow, we synthesis N_t journeys that each originate at a randomly chosen origin centroid and the destination is chosen by sampling from all possible destination centroids according to the probability vector $\mathbf{p}^{(i)}$. The shortest route between the OD pair is then computed by Dijkstra's algorithm, using the adjusted edge lengths as described in Section 2.2.

After simulating N_t journeys in this manner we can form an edge flow vector which lists the number of simulated journeys that use each edge, taking a minimum value of 0 and a maximum

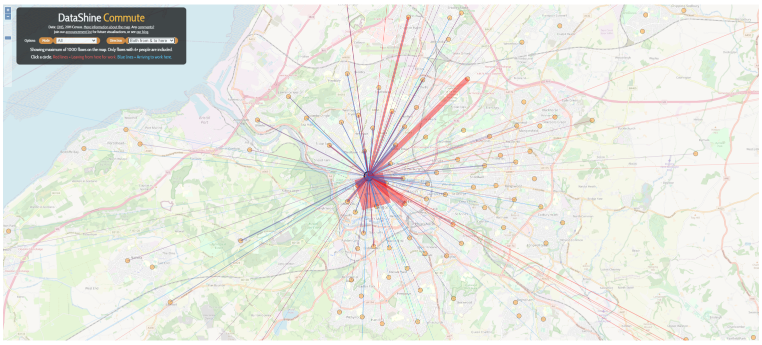


Figure 3.2: A screenshot of the DataShine commute interactive map showing the flows to (blue) and from (red) one of Bristol's 55 MSOAs [30]. The thickness of the line indicates the number of journeys flowing between the two centroids.

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

of N_t. This value is then used to inform the upgrade priority of each edge.

Figure 3.4 shows the Bristol road network with edges coloured corresponding to their flow in a trip simulation of size $N_t = 5{,}000$. 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.

The use of empirical data in this way gives us a more accurate representation of the true flows of commuters than simply using a random OD pair assignment. However, there is one limitation in that we are assuming that all journeys begin and end at a relatively small number of LSOA centroids. This means that we may not recover total coverage of the city street network. Furthermore, we will be artificially increasing the flows on any edge directly connected to a centroid.

In the next section we describe how the network is upgraded and perform some assessment of the heuristic's impact on the Bristol street network.

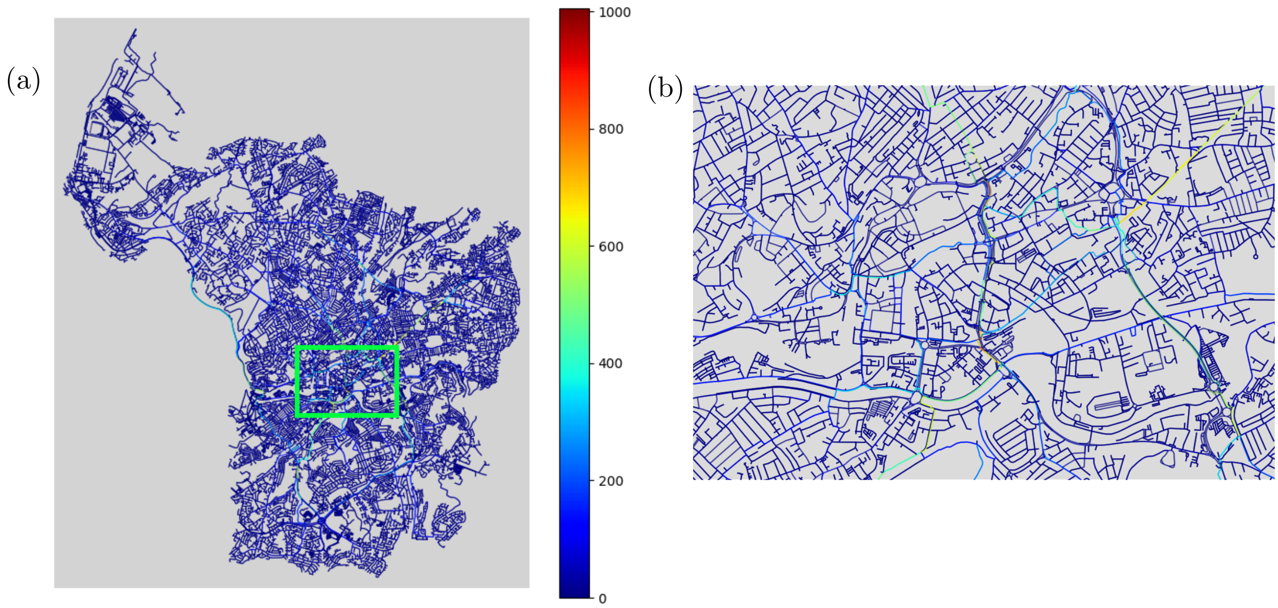


Figure 3.4: (a) map of Bristol's road network with edges coloured based on their cycle flow value. (b) a section in the centre of Bristol (green) to show one of the main stretches of street with high cycle flow. The edges are coloured according to how many journeys use them.

3.3 Assessment of the heuristic

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 3.5 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 3.5 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 with cycle paths if they were not already designated with cycling infrastructure. This is because these edges are the main thoroughfare for cyclist trips between LSOAs, and 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. The green box in Figure 3.5 gives an example of an area where cycling infrastructure is suggested and demand is not particularly high but, as seen in Figure 3.3 there are LSOA centroids there.

To assess the user experience on an upgraded city network the scoring metric from Chapter 2 is used. To recap, we simulate N journeys and for each one calculate a $P_{O,D}$ score by

$$P_{O,D} = \frac{\sum_{i=1}^{N_e} \chi_i L_i}{\sum_{j=1}^{N_e} L_i} \tag{3.2}$$

where i are the edges taken for the journey, N_e is the total number of edges, and L_i is the length of edge i.

This metric represents the user experience for cyclists using the upgraded network. When the

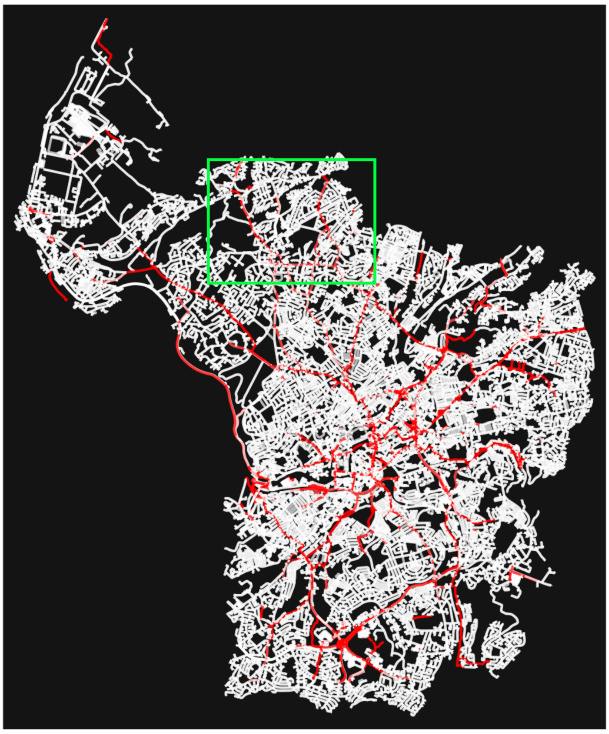


Figure 3.5: The upgraded network for a batch process with parameters $L = 160\,\text{km}$, $N_t = 250$, $B = 20$ and $\omega_d = 2$.

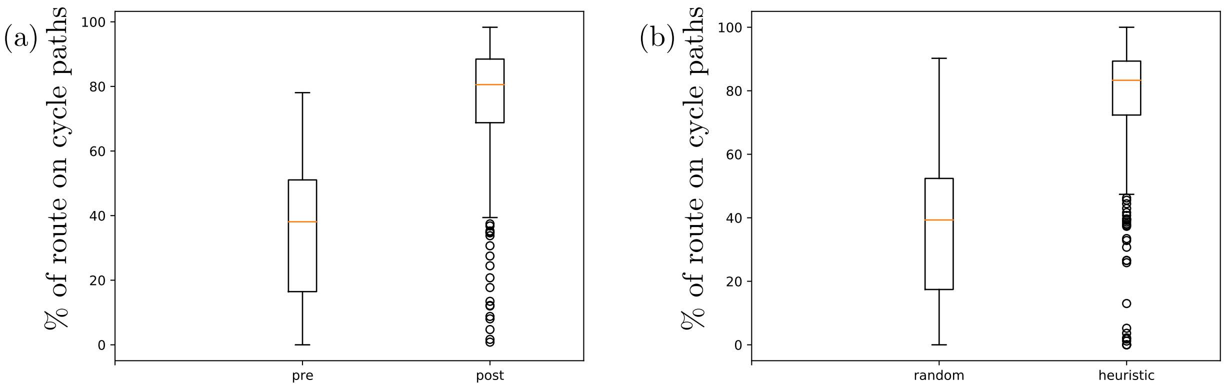


Figure 3.6: The distribution of $P_{O,D}$ scores for (a) the Bristol network before and after applying the upgrade heuristic, and (b) the upgrade heuristic vs randomly assigning the same length of upgraded streets.

upgraded Bristol network is tested using the metric in equation 3.2, Figure 3.6(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%. Note that although the length of the cycling infrastructure has doubled, the proportion has more than doubled, indicating that journeys are being drawn into it by the routing model. There are some outlier cyclists in the post-upgrade distribution: these journeys most likely connect two nearby centroids with low design demand, and hence little cycling infrastructure has been generated along the shortest path from origin to destination. We have looked into this and found a similar pattern to Figure 2.7.

The question we now ask is whether the upgrade heuristic performs markedly better than randomly upgrading the network with the same length budget. Thus we upgrade randomly sampled edges from the network until the length budget of 160km is met, and then test the $P_{O,D}$ scores of this randomly upgraded network against the network produced by the heuristic. Figure 3.6(b) clearly shows that our heuristic vastly outperforms a random process, as we might expect. The random upgrade achieves only an average of 37% of the length of journeys on infrastructure a: 3% gain over the original network but using 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. This analysis confirms, as we would expect, that a connected cycle network yields better cyclist satisfaction.

3.4 Impact of multiple batches

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 infrastructure which via the routing model, draws cyclists towards it. 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. This will then lead to larger flows on streets attached to the new infrastructure and will thus 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 lenght of upgraded edges are shortened by the design cyclist's propensity to cycle ω_d (according to the model in Section 2.2) before the trip simulation for the next batch. Over many batches this should mean that flows are increased on the edges at the end of growing contiguous paths of cycling infrastructure and so a more connected, i.e., "cohesive", 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 due to the need 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 network to justify the computational expense.

Figure 3.7 shows the resulting networks for two extreme cases of this methodology tested in a smaller upgrade with length budget $L = 40$km. The one-shot network (a) is computed by upgrading all 40 km in one batch after simulating $N_t = 500$ trips. Whereas, the single-edge network (b) is computed by upgrading only a single edge in each batch of $N_t = 100$ 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, for which (in effect) trip flows can accrue over batches. The zoomed-in sections of both networks show the main difference between the two methods. (d) shows that in the highlighted section (blue boxes) the single edge upgrade produces a more connected network due to its ability to

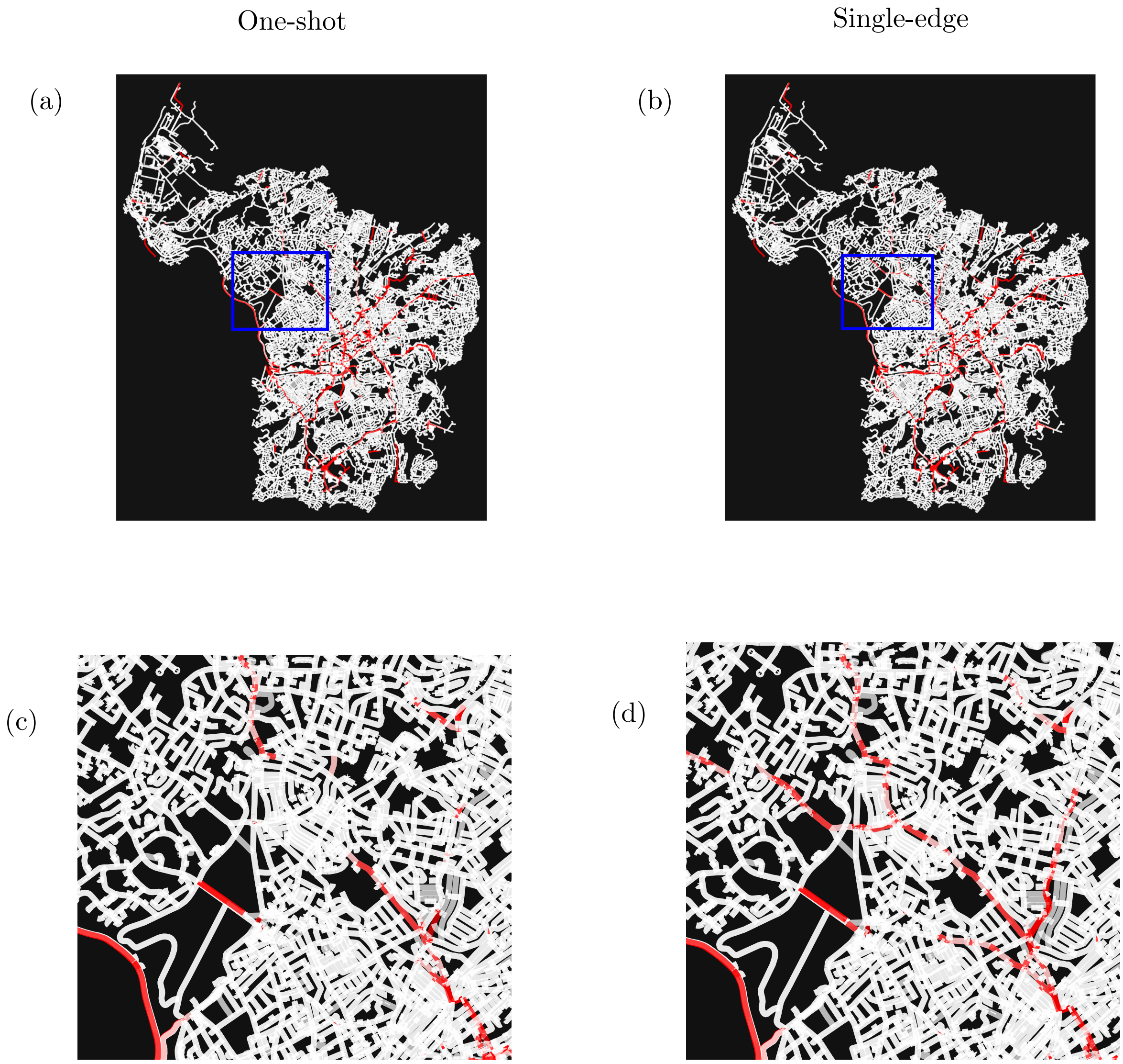


Figure 3.7: Road network of Bristol with edges (i, j) whose $\chi_{i,j} = 1$ highlighted in red. (a) The network after a one-shot process has been applied to upgrade $L = 40$km of street. (b) The result of a single-edge scheme upgrading $L = 40$km also. (c) and (d) show zoom boxes (Highlighted in blue). It can be seen that cycling infrastructure is found here in the single-edge upgrade but not in the one-shot upgrade.

capitalise on the 'pull' of new cycling infrastructure.

Figure 3.8(a) shows the distribution of $P_{O,D}$ scores for both the one-shot and single-edge methods. There is little difference 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 3.7) between the suggested networks from both methods there is not (in this example) a marked metric improvement in choosing a batched process.

Figure 3.8 (b) shows the resulting $P_{O,D}$ distributions from a sweep of values for B. 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

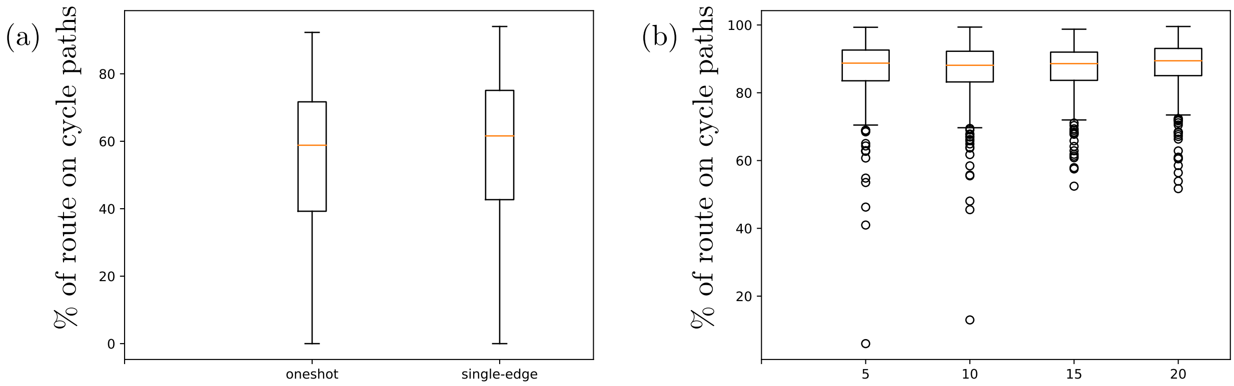


Figure 3.8: The distribution of $P_{O,D}$ scores for (a) the upgraded network for one-shot and single-edge, and (b) when varying the number of batches B.

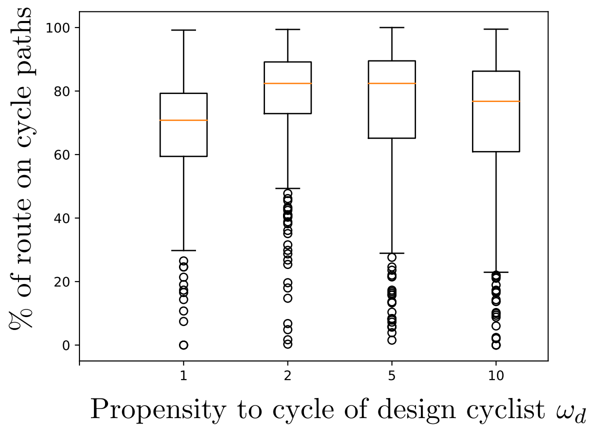


Figure 3.9: Distribution of $P_{O,D}$ scores for a range of design ω_d and test cyclist ω_t set to 2.

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 that they will always be prioritised for upgrade irrespective of batch sizing. Potentially, results for other cities with less well-defined spokes may differ. 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 Impact of different design cyclists

So far the propensity to cycle parameter has been fixed at $\omega = 2$ in both upgrading and testing the network. However, it is worth exploring the impact that designing a network for a cyclist with a given propensity ω_d has on the resulting network's fitness for a cyclist with 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? And vice versa — are networks designed by risk-adverse cyclists useful to those with a high propensity to cycle?

Figure 3.9 displays the $P_{O,D}$ distributions for four different network designs with generated

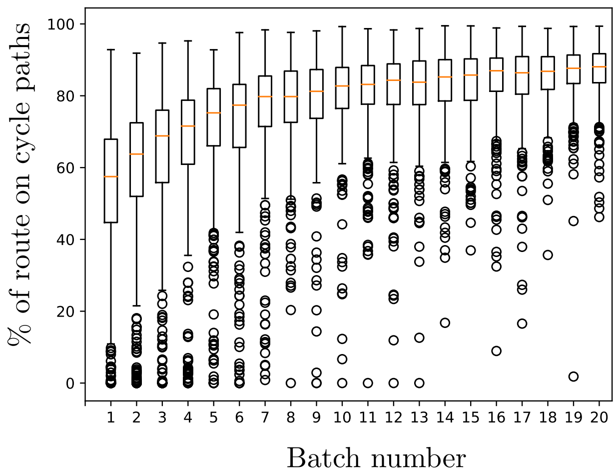


Figure 3.10: Improvement in $P_{O,D}$ scores throughout the batches of a 20 batch upgrade of the Bristol network. It shows that most of the benefit is delivered in the early batches.

by design cyclists with $\omega_d = 1, 2, 5, 10$. 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 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 overall experience.

3.6 Interpretation as street improvement over time

The batches in the proposed upgrade algorithm have been, until now, iterations of a synthetic loop whose aim is to encourage more connected "cohesive" networks. However, there is a nice parallel between batches and the natural incremental implementation of real-world cycling schemes over time. In reality the responsible local authority will want to stagger the implementation of new cycle schemes in order to spread the costs. Thus the batches of the upgrade heuristic might be viewed as the incremental improvements in a real-world network that would be carried out year on year. This observation suggests that if the local authorities plans are adapted year on year in response to incremental improvements in the cycling infrastructure, mirroring the structure of our heuristic algorithm, then naturally more cohesive networks will result. Conversely, if there is high confidence in the predictive power of our routing model, we might view the upgrades suggested in each batch as the optimal ones to perform year-on-year in a precomputed multi-year plan.

Let us investigate this multi-year plan idea. Figure 3.10 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 3.11 shows the order in which the upgrades are performed. This plot indicates that the major infrastructure spokes seen built into the top and bottom of the network in Figure 3.5 are built within the early batches and suggests these are the streets that the local authority should

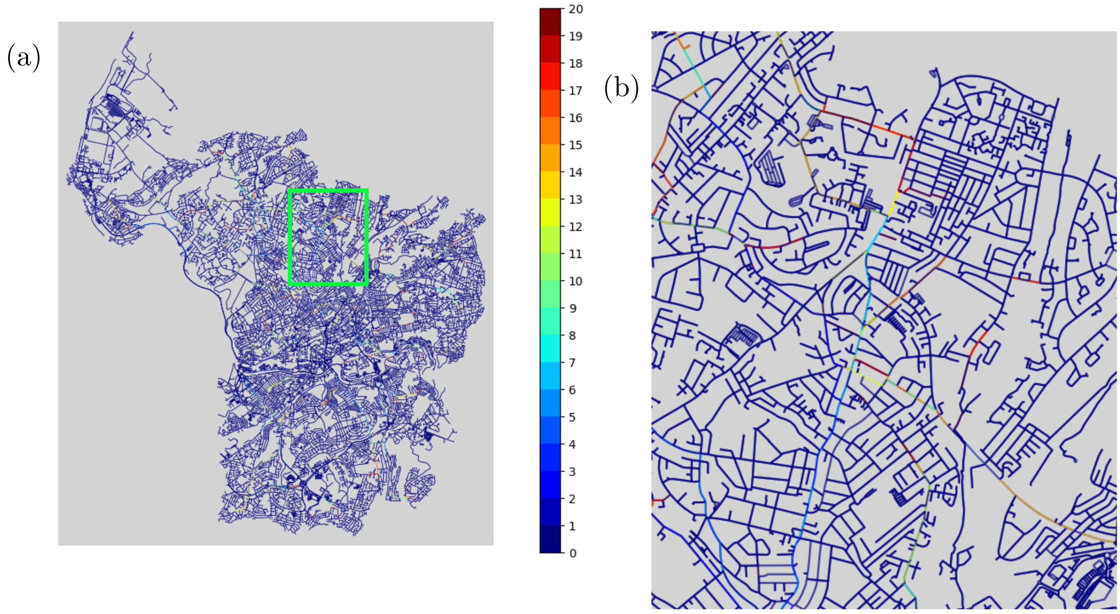


Figure 3.11: (a) Map where colour indicates the batch number of each edge's upgrade. (b) Zoom box shows how contiguous path grows northwards out of city centre from batch to batch.

upgrade first. These upgrades deliver the largest increase in average $P_{O,D}$ seen in Figure 3.10. In contrast, the edges that are upgraded in the last couple of batches and tend to be at the end of long spokes. 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 3.10. Figure 3.11 (b) shows the ability of the batched heuristic to build contiguous paths. We can see that the next batch's upgrades tend to be attached to those of the previous batch, yielding a contiguous cycle path.

3.7 Contributions

The main contributions from this section are as follows:

- We have introduced a simple model for loading a city's street network with synthetic cycling demand, provided that empirical data is available.
- We have developed a heuristic approach to upgrade a city's cycle network up to a specified length budget.
- We have developed a 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 that in the case of the Bristol network batching has little impact on the upgraded network.

- We have demonstrated that the best results are achieve for a test cyclist that shares the same propensity to cycle as the cyclist used to design the network.

Chapter 4

Synthetic Network Experiments

This chapter explores the heuristic proposed in Chapter 3 on small synthetic networks. We start by adapting both the heuristic and scoring metrics to a new synthetic network framework. Next (Section 4.1), we introduce a very simple 'cross' network and perform some experiments on it to explain how the heuristic works and to highlight how a batched heuristic might deliver better networks. Then (Section 4.2), we introduce a process for generating ensembles of random networks using beta-skeleton pruning to reduce the number of paths within the network and investigate the difference between one-shot and batched upgrades across these ensembles. Finally (Section 4.3) we discuss the main contributions from this chapter.

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 each of a small number (four) of OD pairs, rather than using an agent based formulation that samples from a very large number of OD pairs. The upgrade approach used here is thus structurally 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 will use the $P_{O,D}$ metric introduced in Chapter 3, averaged over all routes and normalised by demand such that it outputs a value $0 \leq S \leq 1$ with zero meaning that the shortest route for all OD pairs uses no cycling infrastructure and one meaning that all shortest routes between OD pairs use exclusively cycling infrastructure. We also introduce a new metric that measures the demand weighted average effective length of the routes between all (four) OD pairs according to the model proposed in Section 2.2.

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 behind the algorithms output.

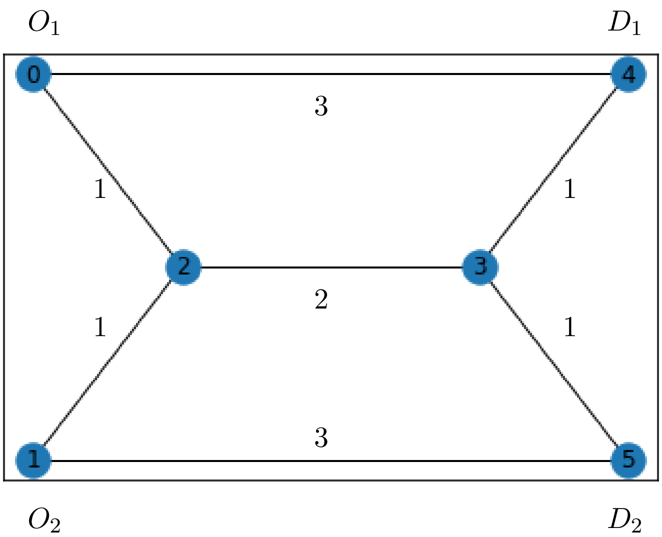


Figure 4.1: 'Cross' network setup.

4.1 Cross network experiments

To illustrate the impact of our upgrade heuristic a symmetric cross network is used, see 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

$$d = [d_{04} \quad d_{05} \quad d_{14} \quad d_{15}] = [1.5 \quad 1 \quad 1 \quad 1.5] \tag{4.1}$$

so that there is slightly more demand from top-to-top and bottom-to-bottom than diagonally across the network. During the graph setup each edge is given a value $\chi_{i,j}$ as described in Section 2.2.

Again for simplicity, we initialise the graph with no pre-existing cycling infrastructure. The goal is then to upgrade the network with a budget L=6 of half of the overall network length. As established in Section 2.3 and confirmed in Section 3.5, a choice of $\omega_k = 2$ is considered appropriate to model 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.

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 only one of these can be upgraded within our budget so (1,5) is chosen by numerical ordering in the ranking list and we move down to the next highest flow edges as these may fit within our budget; 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; 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

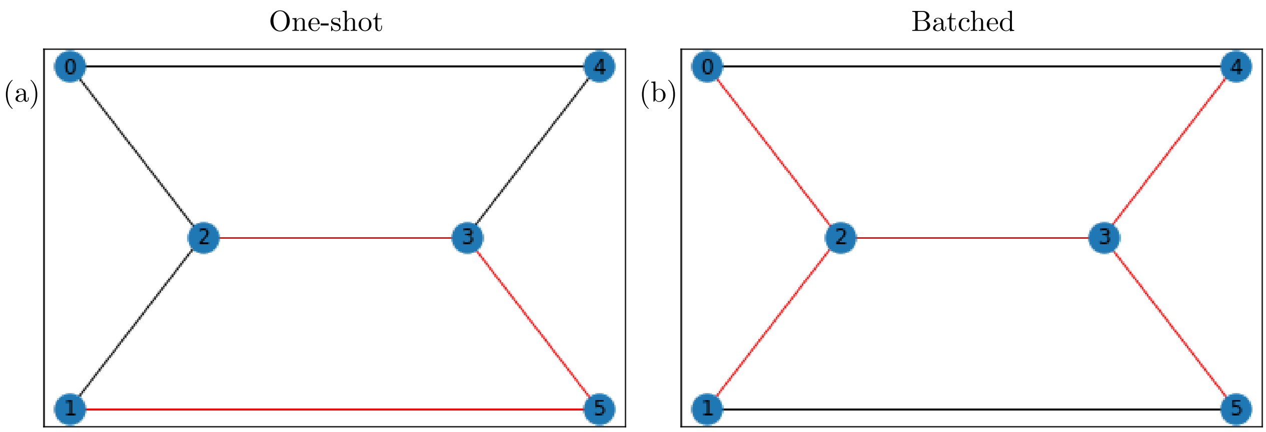


Figure 4.2: (a) the output network resulting from upgrading the budget $L = 6$ in one-shot it is worth noting any combination of the middle edge a long edge and a short edge can result from random selection to break edge flow ties. The one-shot approach does not achieve a symmetric design as the upgrade budget allows for only one long edge to be upgraded after the middle edge. (b) the output network resulting from upgrading the budget $L = 6$ in two batches. The batched method does result in a symmetric design which overall satisfies more cyclist's needs.

[(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 naturally we move down the list to the first edge that fits within our budget, meaning (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; 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)].

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.

The two networks in Figure 4.2 are then assessed with the metrics outlined earlier 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. It is also clear that the batched process yields an average shorter effective route length according to our propensity to cycle model.

Table 4.1: The metric scores for both the one-shot and batched upgrades on the 'cross' network.

Upgrade Method	Average $P_{O,D}$ score normalized by demand	Average effective route length
One-shot	0.55	4.9
Two batch	1.0	4.0

As we discussed in Chapter 3, the heuristic proposed here is not usually optimal. It is worth exploring the optimal network design for the synthetic network to establish whether, in this small test case, our heuristic can achieve the 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.

We have already 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.

4.2 Experiments on ensembles of random networks

In some joint work with the project supervisor we aim to generate ensembles of small random networks to then upgrade and score those random networks to see whether batching is consistently the best approach. The networks are generated via Delaunay triangulation and then pruned as a beta skeleton according to a formulation proposed in [31]. An ensemble of these networks is generated and for each we compute the upgraded network for a budget of half the overall length of the network using both a one-shot and batched approach.

The generation of the random networks has two key parameters:

- N: The number of nodes in the generated network.
- β: The pruning factor (larger β gives you a more sparse network).

Figure 4.3 shows the generated networks for two different values of β for a network with $N = 20$ nodes. Clearly (a), with a lower value of β, produces a network more similar to a street network with multiple routes connecting OD pairs. In this case for ease of computation we want to reduce this number of routes slightly so we use a value of β of 1.35 to maintain multiple routes but not yield an over-connected network.

To establish the impact of batching on these networks we then apply two upgrade schemes across all networks in the ensemble. Firstly we upgrade 50% of the network's length in one batch and secondly we upgrade 50% of the network's length over two 25% batches. In this case we apply a correction such that the final edge of the batch is upgraded if the remaining batch

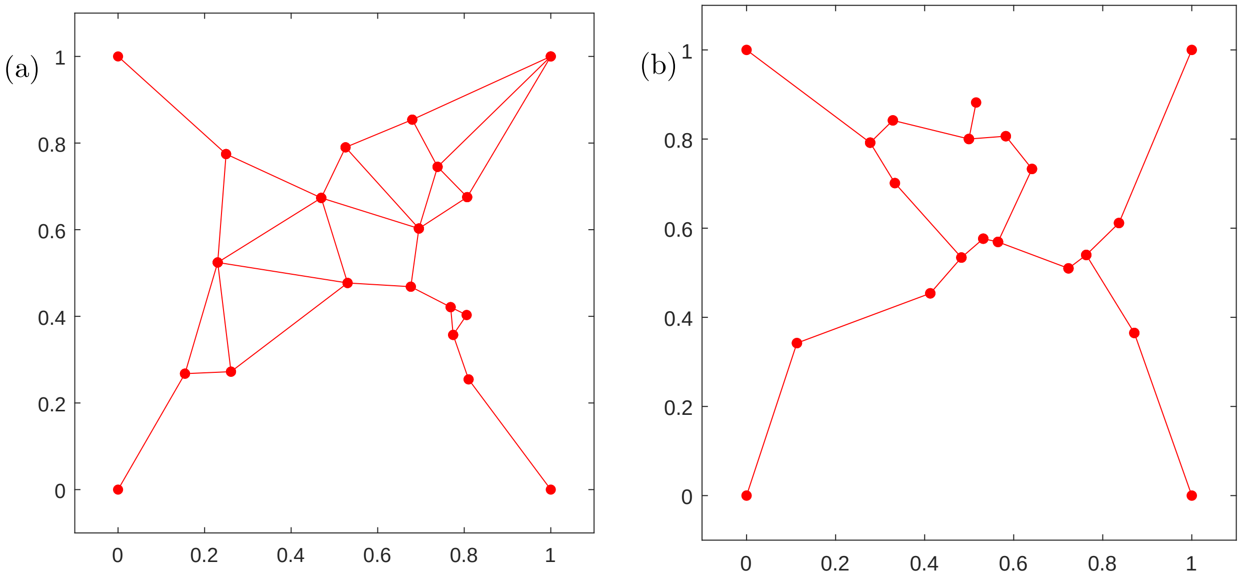


Figure 4.3: Two randomly generated networks generated with $N = 20$ nodes. (a) is generated with $\beta = 1$ such that the graph is very well connected and (b) is generated with $\beta = 2$ such that the graph is very sparsely connected. In reality a value of β in between will produce a more city like network structure.

Table 4.2: Average metric scores for one-shot and two-batch upgrades schemes on an ensemble of 1000 randomly generated networks. Results show that on average a two-batch upgrade scheme performs slightly better than a one-shot scheme.

Upgrade Method	Flow weight proportion of length on cycling infrastructure	Flow weighted effective length
One-shot	0.7731	1.8006
Two batch	0.7742	1.8000

budget can upgrade over half of the edge's length, else this budget is rolled over to the next batch. Once again tie-breaking is solved simply by index ordering of the edges. To assess the difference between one-shot and batching for these networks we again use the scoring metrics established earlier in the chapter. Table 4.2 shows the average scores for these metrics across an ensemble of 1000 randomly generated networks. The scores show that on average there is a small improvement gained by switching to a two batch scheme and capitalising on the pull of cycle paths for a cyclist with propensity to cycle $\omega_k > 1$.

This result leads us to question whether the batched upgrade always gives the same result or slightly better than a one-shot upgrade. To investigate we take a larger batch of $N = 500$ beta skeleton networks and plot the difference between the $P_{O,D}$ scores for both methods. Figure 4.4(a) shows the resulting scatter plot. It shows that the batched approach is sometimes worse than a one-shot upgrade. However, there are more occasions where batching performs better leading to the positive average delta. Unfortunately, no clear causality is observed in the networks that yield either the largest positive or negative delta. It is clear that for the vast majority of networks the two methods perform identically, this is probably because the networks are not complex enough, with only 20 nodes and four OD pairs, to allow the differences in the two methodologies to be observed. It is possible with some more complex (with either more nodes or more OD pairs or both) synthetic networks we would see more comprehensive results.

It is also worth investigating the impact of using different values for design ω when computing

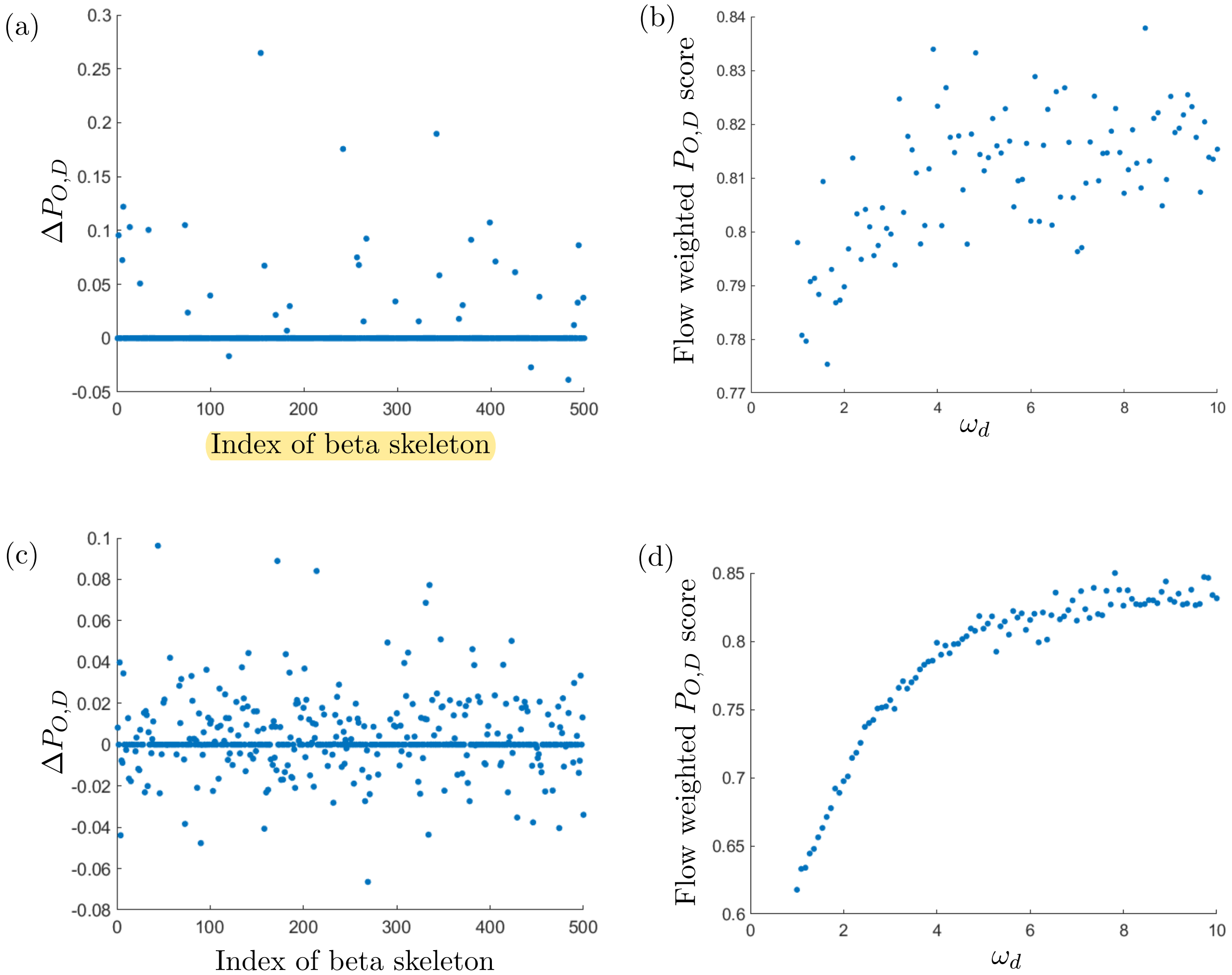


Figure 4.4: (a) A scatter of the difference between batched and one-shot upgrade $P_{O,D}$ scores across and ensemble of 500 beta skeletons. (b) A scatter showing the trend in batched $P_{O,D}$ scores as design ω is increased. (c) A scatter of the difference between batched and one-shot upgrade $P_{O,D}$ scores across and ensemble of 500 beta skeletons when using SUE demand assignment. (b) A scatter showing the trend in batched $P_{O,D}$ scores as design ω is increased when using SUE demand assignment.

metrics across the ensembles of random networks. We fix $\omega_t = 2$ and vary our design cyclist in the range $1 \leq \omega_d \leq 10$. We have established that the two-batch method on average performs better than a one-shot for both of our metrics. Therefore, we focus on the variation in flow weighted $P_{O,D}$ and flow weighted effective length scores for a two-batch upgrade across our parameter space. Figure 4.4(b) shows that as we increase ω_d the batched $P_{O,D}$ scores tend to increase as well. This is due to the increased pull effect of upgraded streets for higher values of ω_k. In Section 2.3 we established $\omega_k = 10$ is an extreme value that means cyclists will use exclusively designated infrastructure where possible this, of course, will lead to higher scores in our metric.

One other area of interest is using stochastic user equilibrium (SUE) assignment to distribute demand over a set of shortest routes rather than putting all demand on the strictly shortest route. When we use this the scores for one-shot and two-batch are as shown in Table 4.3. The scores once again show that there is slight improvement made by switching to a batch upgrade process. We again check whether batched upgrades always perform the same or better than one shot. Figure 4.4(c) shows that when using SUE assignment there is a greater chance that

Table 4.3: Average metric scores for one-shot and two-batch upgrades schemes on an ensemble of 1000 randomly generated networks when using SUE assignment for demand assignment. Results show that on average a two-batch upgrade scheme performs slightly better than a one-shot scheme but on average the effective length of routes are longer than using strictly shortest route assignment.

Upgrade Method	Flow weight proportion of length on cycling infrastructure	Flow weighted effective length
One-shot	0.7061	2.2052
Two batch	0.7078	2.2045

batching performs worse than one-shot upgrades. This is is because when we spread demand over the network we do not capitalise on the pull of upgraded streets quite as much since we are not focused solely on shortest routes. After this result is observed, we investigate the effect of ω_d on a batched upgrade scheme using SUE assignement. Figure 4.4(d) shows that there is a clear metric increase by increasing our ω_d. However, as observed in Section 2.3 there are diminishing returns at higher ω_d values as routes simply cannot fall on cycling infrastructure due to our constrained budget.

4.3 Contributions

The main contributions from this section are:

- We have introduced a simple cross network example to demonstrate the steps in which our proposed heuristic upgrades a network.
- We have investigated the impact of our heuristic on ensembles of randomly generated networks and shown that on average a two-batch upgrade method performs better than a one-shot approach.
- We have introduced the idea of using SUE assignment to spread demand out over multiple short routes rather than strictly the shortest route to give a more realistic demand assignment for these smaller synthetic networks and shown that metric trends seen previously still hold.

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 4.1) 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.

We investigated the impact of batching for ensembles of randomly generated networks. We averaged two key scoring metrics across an ensemble of randomly generated networks for both one-shot and two-batch upgrade schemes. The results showed that on average there is a minor improvement gained by switching to a batched approach, this is due to the ability of a batch approach to capitalise on the pull effect of new infrastructure highlighted in Section 3.6. It was found that this result is not universal and that there are some random networks for which one-shot performs better than batching. We demonstrated that the results found in Sections 2.3 and 3.5 hold for these random networks and that as we increase our design omega we do achieve higher scores in our metrics.

Finally, we introduced SUE demand assignment to spread demand throughout the synthetic networks and we found thatthe result that batching is slightly better still holds, although we note that it is more likely that batching could yield worse performance than one-shot if we use SUE assignment. We also find that when varying our design ω in a scheme using SUE assignment, improvement is made by increasing ω_d up to a point of diminishing returns which corresponds with our findings in Chapter 2.

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 real city network.

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] G. K. 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] K. J. Krizek, A. El-Geneidy, and K. Thompson. A detailed analysis of how an urban trail system affects cyclists' travel. *Transportation*, 34(5):611–624, 2007.

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

[27] 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.

[28] 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.

[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] 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.

[31] A. Espinosa Mireles De Villafranca. *On emergent traffic patterns in synthetic road network ensembles: an experimental mathematics approach*. PhD thesis, University of Bristol, 2020.