Cycle Networks — Finding the Missing Links

Max Oblein
Supervised by Prof R. E. Wilson

December 9, 2020

1 Problem statement

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

The National Travel Survey (NTS) [2] reports that cycling currently makes up 2% of all trips, where a trip is defined as a one-way course of travel with a single main purpose, and that the average length of these trips is 23 minutes. This seems a small proportion given that 38% of people surveyed own a bicycle. The third wave of the National Travel Attitude Survey (NTAS) [3] posed questions as to why people don't like to cycle. Their conclusions found that 60% of people agree that "it is too dangerous for me to cycle on the roads". Research suggests that an increased provision of cycling infrastructure is positively correlated with the proportion of people cycling within that city [4]. The key to increasing cycling rates in Bristol is thus to ensure cyclists feel safe and have segregated paths and lanes to use throughout their journeys.

Currently, when planning road networks, lots of money is spent with engineering consultants to decide where to build new roads [5]. Cycling is much less common as a mode of transport than driving, so planning the cycle network has a significantly lower budget. This means that selecting where new cycle paths are built or lanes are painted is often an arbitrary decision made by the local authority. Therefore, this project aims to develop a mathematical framework to inform these decisions.

There is some previous work in the area of planning cycle networks, and most studies focus on two areas: the effect infrastructure has on the number of cyclists and the evaluation and proposal of cycle schemes based on geographical factors. The first area motivates the need for cycling infrastructure [4]. The second is useful for the evaluation of routes suggested by this project based on urban factors such as the amount of retail or business land area at both origin and destination of the route [6]. The methods presented by Milakis et al. [7] give a comprehensive methodology for evaluating proposed cycle routes. They use a weighted scoring system to assign good scores to origin-destination (OD) pairs attached to important facilities e.g., universities. In the literature little work has been found in the area of assigning new cycle routes automatically using a mathematical approach.

The main concern when automatically proposing roads for new cycling infrastructure is deciding how best to load a road network with cycling demand. Two papers have been identified for their methods in loading street networks with cycling demand. One approach, formulated by Milakis et al. [7], approaches this problem with the use of large demand centres. The study is based in Athens and the idea is to draw centres around areas of large demand. The centres are chosen based on three main criteria: the land use (retail, offices, education or leisure), the OD matrices from the last two travel surveys in Athens, and the network centres detailed in the master plan of Athens. Milakis et al. identify eight such centres and they are used as the OD pairs from which to generate synthetic cycle journeys. This methodology is a good starting point for loading the network with demand, but a set of only eight nodes does not allow for realistic models of commuting in cities such as Bristol, where workers commute from and to a huge number of different locations. This methodology is also not automatic: it requires a survey of cyclists to decide which land use attributes give the largest demand to a centre, and creating and carrying out a comprehensive survey is costly. The survey used in this case is also only considering existing cyclist's behaviour and therefore cannot predict the impact of new infrastructure on cycling demand.

A second paper by Larsen [8] takes a more granular approach. The methodology begins by separating the study area (Montreal) into 300m grid squares; the demand within each of these squares is then derived using both observed cycling trips (OD data from Montreal travel survey) and potential cycling trips (car trips shorter than 2km). These two data sources give OD pairs and then standard shortest path algorithms are used to model the route taken between them. Flow intensity is then given by the proportion of routes passing through a grid square. This analytical methodology is then combined with a survey of cyclists as to which road they think should be prioritised for new infrastructure to yield a prioritisation index.

One part of the Montreal study [8] 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. [9] applies formal optimisation techniques to the cycle network in some case study cities. The research proposes optimisation of total user and construction costs where both are proportional to distance and user costs are increased on edges without cycling infrastructure. The paper uses a heuristic approach to optimise for user cost on large scale networks, and compares against exact solutions on smaller artificial network examples. However their algorithm does not naturally penalise disconnections in the resulting network, therefore a methodology for identifying links that join up the cycle network is needed.

One tool that has been recommended to us by sustainable transport planners [10] is the the Cycling Infrastructure Prioritisation Toolkit (CyIPT). This is a combination of smaller tools that aims to provide an interactive map of the UK, detailing areas for proposed new cycling infrastructure. The components of most interest are the propensity to cycle tool (PCT) [11] and the rapid cycleway prioritisation tool (RPCT) [12]. The PCT serves as the CyIPTs demand model, using 2011 census data to provide estimates of OD flows. The PCT also uses geographical factors such as 'hilliness' to reduce the propensity to cycle on any given route. Our key interest in the RCPT is its planned facility to compute cohesive networks, which represent more highly connected cycle networks. However, in the current RCPT documentation there is no formal mathematics presented as to how to find these cohesive networks.

Summary of studies

Study	Methodology
Athens [7]	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 [8]	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 [9]	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 [11], [13]	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.

2 Project plan

Task	Time Frame	Description
Research data sources and formats	Weeks 3 - 5	Research the data sources and formats to be used, and which programming language includes packages that are most applicable. Look for papers of previous work in the subject area.
Build initial network	Weeks 4 -6	After deciding on python as the programming language take the map of Bristol and convert to a network object using OSMnx
Initial modelling	Weeks 6-8	Using uniform random OD pairs generate shortest routes through the network and find the percentage of time spent on cycling infrastructure.
Interim Report Writing and implementing weight adjustment	Weeks 7-9	Write initial problem statement and literature review. When initial modelling is complete begin write up of methods. Implement factor increasing weight of non-cycle edges and add into report after completion.
Proof reading and editing of interim report	Weeks 9-10	Send report to supervisor early week 9 to get feedback. Then implement feedback and proof read before submission on the Friday of week 10. Also implement forcing of longer routes to eliminate peak at 0% cycle usage. Amsterdam is introduced as a more connected case study.
Implement different styles of OD assignment	Weeks 11 - 13	Decide on 'centres' within Bristol to use for OD pairs. Generate routes between centres to simulate a more realistic demand structure. Then use LSOAs from census data to inform where these centres should be placed.
Calculate betweeness centrality edge scores	Week 14	Use this as measure of edge popularity: edges with high scores are recommended for cycle infrastructure development.
Optimisation of edge selection.	Weeks 15-17	After having decided on ranking of edges for upgrade using their edge flow, take top X routes and convert to cycle and recalculate the next set for upgrade. This will take a long time on a large road network so it may be useful to generate some artifical smaller networks for analysis of interesting scenarios.
Draft report and poster	Weeks 17-20	Send draft chapter week 18 but continue to draft rest of report and create poster for hand in to supervisor in week 20.
Take feedback and finish final report and poster	Weeks 20-22	Deadline in week 22 so use supervisor feedback and anyone else willing to read the project to make adjustments before submission.

3 Progress to date

3.1 Data

To create the network needed for the analysis proposed in this project, OpenStreetMaps (OSM) has been identified as the primary data source. A python package OSMnx [14] 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 fist step is to convert the OSM map of Bristol, our selected bounding geography, to a network. This yields a set of 'ways' (all roads and paths) that can be cycled on within the bounding geography, hence excluding motorways etc. 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.

3.2 Prescribing the weights of edges

The first step in the analysis of Bristol's road network is to identify the edges corresponding to existing cyclepaths. The OSM tags of interest are 'Highway', 'Cycleway' and 'Bicycle', although not all of these are present for every way in the network. Our conditions for a way to be identified in the cycle network are: (a) its 'Highway' tag takes the value 'cycleway'; or (b) its tag set includes 'Cycleway'; or finally (c) its tag set includes 'bicycle' which takes the value 'designated'.

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

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

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

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

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

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

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

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

meaning that an edge without cycling infrastructure has a larger perceived length than its true length according to the cyclists disposition. The idea then is that if a given cyclist tries to minimise the effective length of their route, they will prioritise routes that follow cycling infrastructure, provided the detour is not too large.

3.3 Uniformly random OD pairs

Naturally the simplest way of loading the network with cycling demand is to sample OD pairs randomly from the set of all nodes within the network. The next step is to compute the shortest route between each OD pair. At this stage we suppose that $\omega_k = 0$ to model the case where cyclists have equal propensity to cycle on roads with or without cycling infrastructure. We compute 500 such shortest routes and the percentage of time spent on cycle infrastructure is found for each one, see the distribution in Figure 2.

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

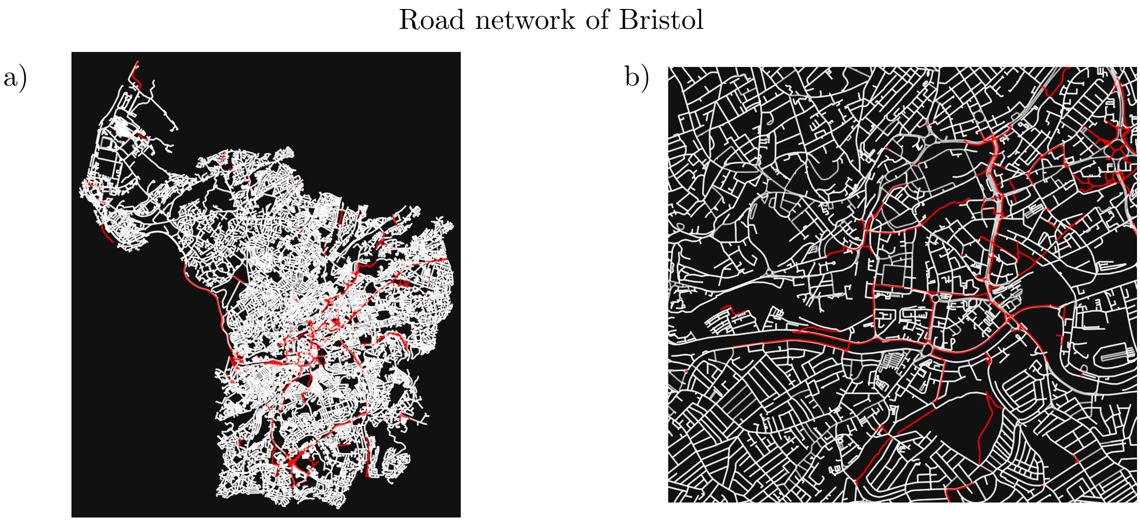

Figure 1: Road network of Bristol with edges (i, j) whose $\chi_{i,j} = 1$ highlighted in red. A section in the centre of a) is seen zoomed in b) to show the finer detail of the network.

length on cycle infrastructure. This result makes sense given the unconnected nature of Bristol's cycle network, which does not allow for whole routes to use cycling infrastructure. The peak around 0% is due to the random selection of OD pairs giving rise to extremely short routes in areas of Bristol with no cycling infrastructure, so it is impossible for the shortest route to contain cycle lanes. In this simulation the mean number of edges used in shortest paths is 120 which, when the network contains $56,000$ edges, is conducive of 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. Some users may be able to increase their percentage of time on cycling infrastructure by allowing for not strictly shortest routes.

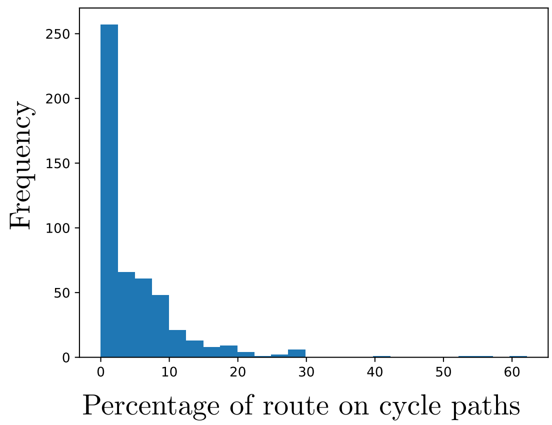

Figure 2: Histogram showing the percentage of routes taken on cycle lanes within shortest routes using uniform random OD selection, with parameter $\omega_k = 0$ so that streets with no cycling infrastructure are not penalised.

3.4 Promote use of cycle paths

The literature suggests that the majority of cyclists prefer to cycle on designated cycling infrastructure [4]. This means that ω_k should be strictly positive. The larger the value given to

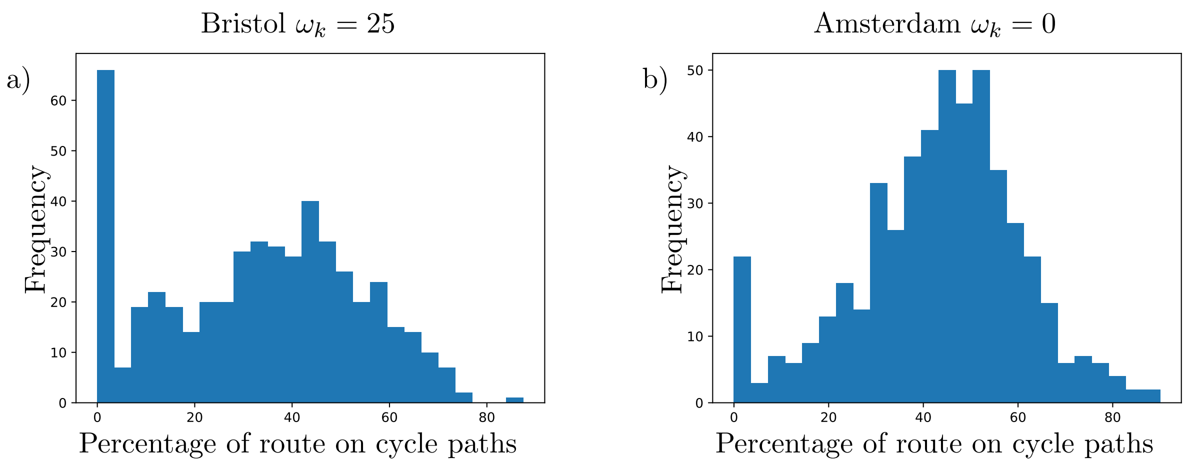

Figure 3: Histograms showing that both increasing ω_k in Bristol a) and the more connected city of Amsterdam (with $\omega_k = 0$) b) have a similar percentage of routes on cycle infrastructure.

ω_k, the larger the penalisation of routes with no cycling infrastructure. If there exists a learner cyclist k', $\omega_{k'}$ should be very large, as it is highly unlikely that a learner would want to cycle without dedicated infrastructure. Furthermore a second bounding geography Amsterdam is also presented in order to compare model results on a more highly connected cycle network. In order to control changes to the model, ω_k is left as 0 in the case of Amsterdam, to see whether connectivity or propensity to cycle on roads has a larger impact on the model result. Amsterdam is selected as a bounding geography not just for its much more comprehensive cycle network but also for its higher quality OSM data. Amsterdam's higher degree of connectivity should yield a higher percentage of time spent on cycle paths within shortest routes.

Figure 3 shows that by decreasing a cyclist's propensity to cycle on roads without infrastructure, we do in fact increase the percentage of length routes spend on cycle paths. However, this effect is small in comparison to using a different bounding geography with a more highly connected cycle network. It is the case that in Amsterdam routes use more cycling infrastructure than Bristol even with $\omega_k = 0$. 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 (not presented here) have shown that as ω_k is increased for cyclists in Amsterdam, most routes are nearly 100% on cycle paths.

After restricting the model to choose shortest paths with at least 50 edges, the results obtained from the Amsterdam network with an $\omega_k = 0$ yield the same result but without a large peak at 0%. This confirms that this peak is caused by very short routes that simply cannot use cycling infrastructure. The problem is much larger in the Bristol network due to its low degree of connectivity, meaning paths would have to use almost the entire network to guarantee the use of cycling infrastructure.

4 Immediate next steps

Work will be conducted to investigate different loading procedures to see how this affects the use of cycle paths in either the Bristol or Amsterdam networks. The flows on these networks will then be used to calculate a metric similar to that of betweenness centrality in the network science literature [15]. The idea is to upgrade those edges (i, j) with high flow rate but whose indicator $\chi_{i,j} = 0$. This should yield an automatic prioritisation method for suggesting roads that should be upgraded with cycling infrastructure. Once these roads have been suggested, the top X candidate edges should have their indicator set as $\chi_{i,j} = 1$ and the analysis should be repeated. This methodology will then be performed on some small artificial examples try to yield an understanding of how this iterative improvement procedure approximates formally optimal network designs.

References

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

[2] Conclusions from the National Travel Survey 2019. `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/905887/national-travel-attitudes-study-wave-3.pdf`. Accessed 2020-11-09.

[3] Conclusions from the National Travel Attitude Survey wave 3 2019. `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/905887/national-travel-attitudes-study-wave-3.pdf`. Accessed 2020-11-09.

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

[5] 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.

[6] 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.

[7] 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.

[8] 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.

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

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

[11] 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.

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

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

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

[15] U. Brandes. On variants of shortest-path betweenness centrality and their generic computation. *Soc. Networks*, 30(2):136–145, 2008.