Cycle Networks - Finding the Missing Links

Max Oblein
Supervised by Prof R. E. Wilson

December 6, 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 in these large cities meaning cycling is in some cases faster than driving for short journeys.

The National Travel Survey (NTS) 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 [2]. However this number is still only a very small amount of the population, even though 38% of people surveyed own a bicycle. The third wave of the National Travel Attitude Survey (NTAS) 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" [3]. Research suggests that an increased amount of cycling infrastructure is positively correlated with the amount of people cycling within that city [4]. The key to increasing cycling rates in Bristol is to ensure cyclists feel safe and have paths and lanes to use throughout their journeys.

Currently when planning road networks lots of money is spent with engineering consultants to decide where the optimum place to build the new road is [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 rather than a mathematically informed decision.

There is some previous work in the area of planning cycle networks but most studies focus on two areas; the effect infrastructure has on the amount of cyclists and the evaluation and proposal of cycle schemes based on geographical factors. The first of these areas serves to motivate the need for cycling infrastructure [4]. The second is useful for evaluation of routes suggested by this project based on urban factors in Bristol. The main factor used is 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 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 to best load a road network with cycling demand. This demand allows for high volume roads to be highlighted for new development. Two papers are identified for their methods in loading street networks with cycling demand. The first 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 Origin destination (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 8 such centres and they are used as the origin destination pairs for cycle routes to be generated between. This methodology is a good starting point for loading the network with demand but a set of only 8 nodes does not allow for realistic models of commuting in cities such as Bristol where workers commute to a huge number of different locations. This methodology is also not automatic it requires 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 cyclists and therefore cannot predict the impact of new infrastructure on cycling demand.

The second paper by Larsen [8] takes a more granular approach. The methodology presented in this paper 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 cases give OD pairs and then standard shortest path algorithms are used to model the route taken between them and demand is given by the proportion of routes passing through a grid square. This is only one part of their model for prioritisation of new cycle infrastructure, it is combined with a survey of cyclists as to which road they think should be prioritised for new infrastructure. This paper, like the Athens study, takes a huge amount of input from cycle surveys and opinion polls these methods are costly and do not allow for an automatic method to be applied to an arbitrary city.

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 the recommend looking at the amount of infrastructure needed to connect the network and assessing case by case. The automatic approach proposed in this project may allow for some dangling nodes to be eliminated from consideration.

Research by Mauttone et al. [9] uses optimisation techniques on a network structure of the cycle network in some case study cities. The research proposes optimization of user and construction costs where both are proportional to distance and user costs are increased on edges without cycling infrastructure. This creates a rather computationally difficult problem with analysis on larger networks taking up to 18 hours to find an optimal solution. This report however details a method for simply identifying candidate edges to be considered for cycle infrastructure using only network metrics to try and speed up the process. This does mean that options presented may not be strictly optimal but further analysis may be performed on the suggested networks.

One tool that is highlighted by sustainable transport planners is the The Cycling Infrastructure Prioritisation Toolkit (CyIPT). This tool is a combination of smaller tools that aims to provide an interactive map of the UK detailing areas for cycling infrastructure. The part of this tool that is of most interest is the propensity to cycle tool (PCT) which aims to map the potential of areas in the UK to cycle [10]. The tool uses 2011 census data to give the OD pairs to estimate demand. The tool goes further using 'hilliness' and length of route to add a decay term to the propensity to cycle. This tool is used to demonstrate the effects of different scenarios based on baseline propensity to cycle. The ideas presented in this tool are useful for developing demand on the road network to be analysed in this project.

Another newer tool created as part of the CyIPT is the Rapid cycleway prioritisation tool (RCPT) which is developed to assign new temporary cycle routes during the covid-19 pandemic. The demand model used in this tool is the PCT so this tool comes with the drawbacks of that tool. The tool includes a section on cohesive networks which aims to explore connecting existing cycling infrastructure. However, there does not seem to be a formally mathematical or network science approach behind developing these cohesive network structures. This is the problem that this project aims to solve.

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 $300\mathrm{m}$ grid squares load underlying road network with 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 in Uruguay. The algorithm presented then tries to optimise for user cost given the length of edges in the road network.
CyIPT [10], [11]	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	Weeks 3 - 5	Research the data 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 Origin Destination (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	Initially write 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 introduce 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. The use LOSAs from census data to inform where these centres should be placed.
Calculate betweeness centrality edge scores	Week 14	Use this as measure of how used edges are and if road segments have high scores they are suggested 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 need for the analysis proposed in this project OpenStreetMaps (OSM) is identified as the primary data source. A python package OSMnx [12] has been used for easy conversion of OSM data to a network topology. The fist step is to convert the OSM map of Bristol, our selected bounding geography, to a network. This yields a set of 'ways' of all roads that can be cycled on within the bounding geography hence excluding motorways etc. This can then be used to create a graph G whose edges 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 cardinality of the vertex set of G which be less than that of the set of nodes in the OSM data as it is not necessary to describe the curvature of each way.

3.2 Prescribing the weights of edges

One problem with OSM is the cluttered nature of the tags given to edges. Therefore the first step in the analysis of Bristol's road network is to identify the edges corresponding to existing cyclepaths. The tags of interest are 'Highway', 'Cycleway' and 'Bicycle' although not all of these are present for every way in the network. The conditions for a way to be considered in the cycle network are,

- Its 'Highway' tag takes the value 'cycleway' or,
- Its tag set includes 'Cycleway'
- Its tag set includes 'bicycle' which must take 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,

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

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

Te 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 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 formulated as

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

meaning that an edge with cycling infrastructure has a large perceived length than its true length according to the cyclists disposition.

Initial analysis is conducted assuming cyclists have equal propensity to cycle on both roads with or without infrastructure i.e. $\omega_k = 0$

Road network of Bristol

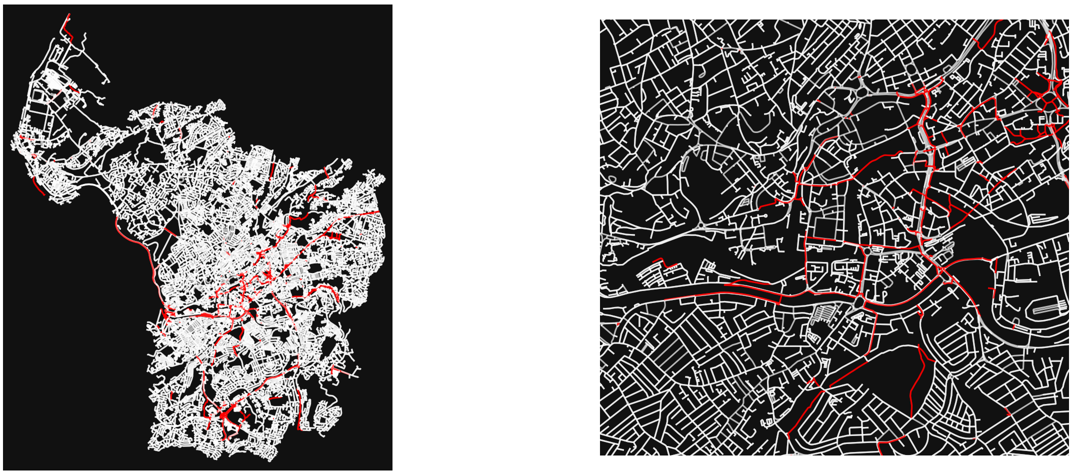

Figure 1: Road network of Bristol with edges (i, j) whose $\chi_{i,j} = 1$ highlihhted in red.

3.3 Uniform random Origin destination 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 OD pairs. To try and find a pattern 500 such shortest routes are calculated and the percentage of time spent on cycle infrastructure is found for each one. The histogram in Figure 2 shows the distribution of this percentage measure.

Ignoring the huge peak at 0% the amount of time spent within the cycle network seems to decay exponentially with anything higher then 15% being zero excluding noise. This makes sense due to the huge discontinuities in the cycle network which do not allow for whole routes to be contained within the network. Some routes may be able to make use of more cycle lanes by allowing for the route between origin and destination to not strictly be the shortest. 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 edges which, when the network contains 56000 edges, is conducive of a large number of short routes. This leads to lots of short OD paths in sections of the network with no cycling infrastructure which gives rise to the large peak at 0% in Figure 2. Another explanation for this peak is simply that the cycle network in Bristol is so sparse most shortest routes cannot use cycling infrastructure.

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 not be zero. The larger the value given to ω_k the

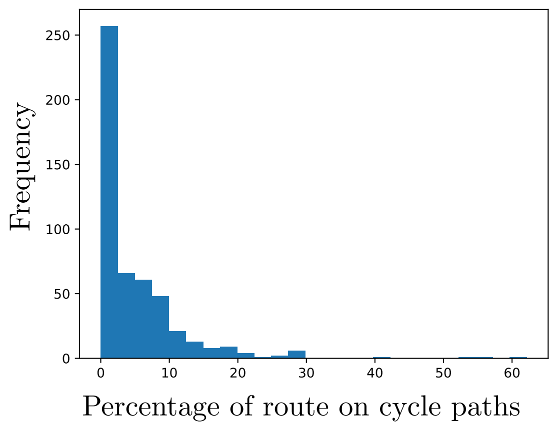

Figure 2: Histogram showing the percentage of route taken on cycle lanes within shortest routes using uniform random OD selection. with parameter $\omega_k = 0$ so that streets with no cycling infrastructure are not penalised.

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. This should improve the demand model on Bristol's cycle network but a second bounding geography Amsterdam is also presented in order to show model results on a 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. Amsterdam is selected as a bounding geography not just for its much more comprehensive cycle network but also for the increased amount data for it in OSM. The higher degree of connectivity should yield a higher percentage of time spent on cycle paths within shortest routes.

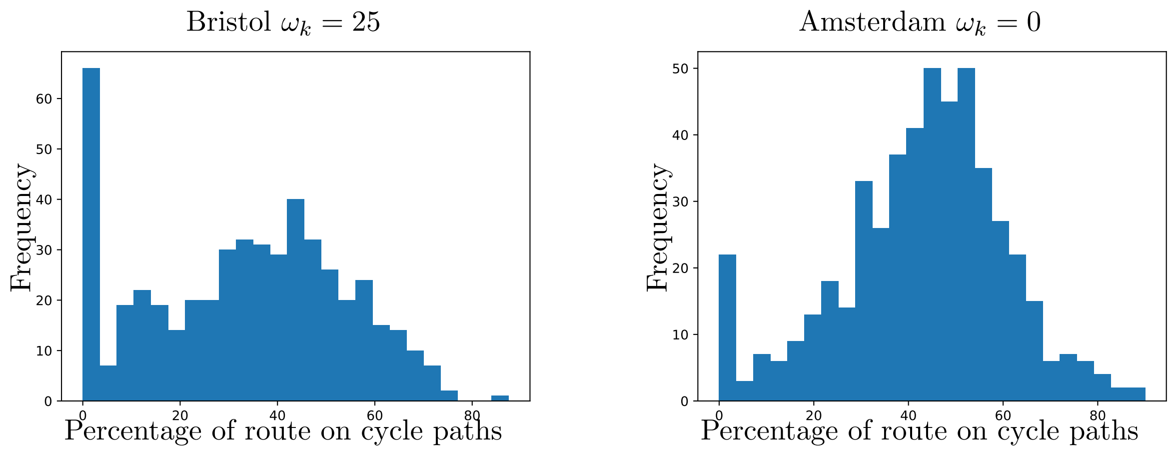

Figure 3: Histograms showing that both increasing ω_k in Bristol and switching to a more connected Amsterdam have similar effects on the percentage of routes on cycle infrastructure.

Figure 3 shows that by decreasing a cyclists propensity to cycle on roads without infrastructure we do in fact increase the percentage of time 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 more routes use cycle paths even with $\omega_k = 0$.

The issue with the outlier peak at 0% is still somewhat present in the case of Amsterdam almost certainly due to a small amount of very short routes that cannot feasibly use cycling infrastructure. As ω_k is increased for cyclists in Amsterdam the distribution shows that most routes are nearly 100% on cycle paths. This means that the question of filling in missing links is more pertinent in a city with good existing structure whereas in cities like Bristol it may be more beneficial to install new schemes rather than connecting existing 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 expected result without a large peak at 0%. This confirms that this peak is caused by very short routes that simply cannot use cycling infrastructure. It is extenuated in the Bristol network due to its low degree of connectivity and paths would have to be extremely long to guarantee use of the cycling infrastructure.

4 Further work

Further work should 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 should then be used to calculate a metric similar to that of betweenness centrality in the network science literature [13]. 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 for upgrade 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 should then be performed on some small artificial examples to yield some interesting results in a much smaller time frame than simply optimising the entire street network of a city for cycling.

References

[1] C. Allan, "Cycling uk's cycling statistics," 2019.

[2] "Conclusions from the national travel survey 2019." `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/906698/walking-and-cycling-statistics-england-2019.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," *Transportation research record*, vol. 1828, no. 1, pp. 116–123, 2003.

[5] G. Topham, "Chancellor announces £27bn for roadbuilding in budget," Mar 2020.

[6] R. Cervero and M. Duncan, "Walking, bicycling, and urban landscapes: evidence from the san francisco bay area," *American journal of public health*, vol. 93, no. 9, pp. 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," *Journal of Transport Geography*, vol. 35, pp. 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," *International Journal of Sustainable Transportation*, vol. 7, no. 4, pp. 299–317, 2013.

[9] A. Mauttone, G. Mercadante, M. Rabaza, and F. Toledo, "Bicycle network design: model and solution algorithm," *Transportation research procedia*, vol. 27, pp. 969–976, 2017.

[10] 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," *Journal of transport and land use*, vol. 10, no. 1, pp. 505–528, 2017.

[11] l. b. U. o. L. The CyIPT team, "The cycling infrastructure prioritisation toolkit," 2018.

[12] G. Boeing, "Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks," *Computers, Environment and Urban Systems*, vol. 65, pp. 126 – 139, 2017.

[13] U. Brandes, "On variants of shortest-path betweenness centrality and their generic computation," *Social Networks*, vol. 30, no. 2, pp. 136–145, 2008.