Prioritizing Road Segments for Investment in Segregated Cycling Infrastructure

A Methodological Framework

Name: Hussein Amr Youssef Ezzeldin Mahfouz

Date: 23 August 2020

Module: CASA0010

Supervisor: Dr. Elsa Arcaute

Word count: 10,873 words

This dissertation is submitted in part requirement for the MSc in the Centre for Advanced Spatial Analysis, Bartlett Faculty of the Built Environment, UCL

Abstract

Understanding the motivators and deterrents to cycling is essential for creating infrastructure that is successful in getting more people to adopt cycling as a mode of transport. Therefore, the aim of this study is to develop a methodology that determines where cycling infrastructure should be added by accounting for cyclist preference and UK government policies to create Low Traffic Neighborhoods. Distance decay, routing, and network analysis techniques are used to examine where future cycling demand in a city will come from, how such demand should be routed on the street network, and how to consider fair distribution of cycling infrastructure that is in line with egalitarian principles.

For routing, it is found that deviating from shortest paths to avoid high-stress road segments can result in acceptable travel time increases for cyclists. Community detection is used to partition the city and evaluate results from utilitarian to egalitarian algorithms on distributing cycling infrastructure. The results show that the egalitarian algorithm provides comparable city-wide results to the utilitarian one.

The aim is for the methodology to serve as a framework that could be tailored to the specific requirements of any urban area. Further research and coordination with local authorities is needed to carry out context-specific routing that accounts for feasibility of reallocating road space to cyclists on certain roads.

Declaration

I hereby declare that this dissertation is all my own original work and that all sources have been acknowledged. It is 10,873 words in length.

Contents

List of Figures

List of Tables

Acronyms

LTN Low Traffic Neighborhood.

MSOA Middle layer Super Output Area.

OD Origin-Destination.

OSM OpenStreetMap.

Acknowledgements

This dissertation benefited from the input of many people. Firstly, I would like to thank Dr. Elsa Arcaute for helping me frame the research topic and providing valuable feedback throughout the duration of the dissertation. I would also like to thank Dr. Maarten Vanhoof and Richard Milton for their very helpful responses to my methodological queries.

The dissertation also benefited from the feedback of Rebecca Flynn, who provided extensive comments on structure and legibility.

This dissertation, and all that I have learned during my year at CASA, would not have been possible without the generous financial assistance of the Chevening Scholarships, the UK government's global scholarship programme, funded by the Foreign and Commonwealth Office (FCO) and partner organisations.

1. Introduction

The past few decades have seen an increased sense of awareness and accountability surrounding the ramifications of our way of life on the future of the planet. The 2015 Paris agreement (UN 2015) was an acknowledgement that fundamental changes to our lives and economies are necessary to reduce the impacts of climate change.

The transport sector is one of many sectors facing transformation, with innovation and policy measures leading to the introduction of 'greener' private vehicles. But as demand for transport rises, resulting in increased congestion, it becomes clear that such vehicles do not scale in cities. The threat of congestion has forced city officials to think more critically about how to move people around. In London, for example, TfL estimates that the annual cost of congestion will reach £9.3 billion by 2030 if there is no shift towards more sustainable modes of transportation (TfL 2018). It therefore comes as no surprise that the importance of public and active modes of transport is gaining more recognition.

The benefits of active transport are not limited to congestion and the environment. It also promises to help alleviate what is referred to as the pandemic of global inactivity. Physical inactivity is on the rise and has become the 4th highest cause of death globally (Kohl 3rd et al. 2012). Various studies have documented the association between active transport and lower risk of disease, including cancer and cardiovascular disease (Celis-Morales et al. 2017; Jarrett et al. 2012; Patterson et al. 2020).

In the wake of the Covid-19 pandemic, and the resulting reduced capacity of public transport, the UK government has pledged to invest billions of pounds to improve walking and cycling infrastructure across the country. This unprecedented sum is an opportunity to reshape cities in a way that improves the wellbeing of citizens, a point that has long being argued by urban researchers and activists. The funding does however come with a warning:

> Inadequate cycling infrastructure discourages cycling and wastes public money. Much cycling infrastructure in this country is inadequate. It reflects a belief, conscious or otherwise, that hardly anyone cycles, that cycling is unimportant and that cycles must take no meaningful space from more important road users, such as motor vehicles and pedestrians - (DfT 2020b)

The funding on its own is therefore no guarantee of a change in commuting across the country; it must be used to design adequate cycling infrastructure that is based on well documented preferences and deterrents to cycling.

1.1 Research Question

The goal of this research is to determine how to prioritize road segments for investments in segregated cycling infrastructure, doing so in a way that makes optimal use of limited budgets. It is understood that optimal solutions are subjective since different stakeholders do not necessarily have the same objective. Nevertheless, the objective of this study is to prioritize cycling infrastructure in a way that helps maximize cycling in the chosen study area. This means providing cycling networks that satisfy current cycling demand and increase the cycling mode share by incentivizing more people to cycle. The question can thus be summarized as follows: *Where should segregated cycling infrastructure be placed so that it benefits the largest amount of current and potential cyclists?*

The work will also look at [1] predicting potential cycling demand[1], [2] routing this demand on the road network in a manner that accounts for stated preference of cyclists, and [3] the effectiveness of different ways of allocating resources. Can we calculate potential cycling demand based on census commuter data[2] and physical geography? What factors need to be considered in order to design a cycling network that encourages non-cyclists to cycle? How do these factors affect decisions on where cycling infrastructure should be added? How do recommendations that focus on maximizing overall utilization of resources differ from those that aim for an equal distribution of resources? These are some of the questions that this research hopes to contribute towards answering.

Manchester is chosen as a case study to evaluate the methodology; According to the 2011 census (ONS 2011), Manchester exhibits modest cycling figures, with only 5% of daily

1. *potential cycling demand* refers to future cycling demand. In this study, it is always higher than existing cycling demand since it is a result of targeted investments in cycling infrastructure made with the aim of increasing cycling mode share.

2. *commuter data* refers to publicly available Origin-Destination (OD) data, normally obtained from a census. This study uses data from the UK census (ONS 2011), which contains aggregate statistics on number of commuters between MSOA pairs, by mode of travel.

commuting trips being cycled. In that regard it is similar to most UK cities.

It should be noted that the focus of the work is not to study a particular city, but to develop a methodology that can be applied to any city, given basic data availability. For this reason, other cities (Nottingham and Birmingham) are used in some parts of the analysis to highlight similarities and differences relating to city size and road network peculiarities.

1.2 Ethical Considerations

The data used in this research is made publicly available by the UK's Office for National Statistics, which ensures aggregation of publicly available data for the purpose of personal data protection. The limitations of the study are well documented and the code for producing the results has also been made publicly available[3] and can be easily reproduced or reviewed.

3. Github repository with code and reproducibility instructions found here

2. Literature Review

As with any research, it is important to put this work in the context of what came before it. The main goal of this research is to suggest cycling networks that encourage more people to commute by bicycle, and so this review looks at research carried out to understand what has been proven to increase the proportion of cyclists on the road. It then looks at how that knowledge has been integrated into methodologies that aim to propose feasible cycling networks. Finally, it considers the main ethical principles underpinning transport analysis and how they are related to the methodological choices of this research.

2.1 What Effects the Decision to Cycle?

There are a number of factors that determine the quality of the cycling network, and various studies have been able to link this network quality to levels of cycling. The quality is affected by factors such as [1] whether or not cycling lanes are separated from motor traffic, [2] how efficient they are at getting cyclists where they need to go, and [3] how well connected the network is. The findings of these studies are outlined below.

2.1.1 Existence of Segregated Cycling Infrastructure

One of the most important determinants for cyclists is the existence of segregated cycling infrastructure on the road network[1]. Winters et al. (2011) evaluated different motivators and deterrents for cycling and found that route conditions and interaction with motor vehicles had a high influence on the likelihood of cycling; cyclists prefer not to share the road with motor vehicles. Different longitudinal studies have reaffirmed this point; A study on the correlation between new cycling infrastructure and cycling uptake in Sydney, Australia found that almost 25% of residents used a new cycleway within 16 months of its construction, with its catchment area being over 3km (Crane et al. 2017). 13% of those who used the cycleway

1. *Segregated cycling infrastructure* refers to road space that is allocated to cyclists only, with physical separation to protect cyclists from other modes of transport.

lived more than 3km away from it, indicating that cyclists are willing to choose less efficient routes in order to commute on safer roads. A longitudinal study on 3 different UK sites that were part of the connect2 program found that proximity to newly constructed pedestrian and cycling infrastructure could predict increases in walking and cycling, but only 2 years after infrastructure investments had been made (Goodman et al. 2014). People living 1km from said infrastructure increased walking and cycling by 45 minutes relative to those living 4km away, indicating that their is latent demand for access to said infrastructure. The routes evaluated were mostly local, and researchers hypothesized that stronger compound effects would manifest as networks became more connected. Aldred, Croft, and Goodman (2019) studied the impacts of the ongoing mini-holland program in London, and found that there was an 18% increase in cycling among people who lived close to the new infrastructure. These areas also saw a positive change in attitudes to cycling compared to the control group. It should be noted that there was no significant reduction in car usage, and the general sentiment of 'too much is being spent on cycling' grew, highlighting the controversial nature of such interventions.

The city of Seville is a successful case study of how to promote utilitarian cycling through accelerated investments in segregated cycling infrastructure. Between 2006 and 2007, the city put in place almost 77km of segregated cycling infrastructure, and by 2011 the network had grown to 164km, increasing the city's almost non-existent cycling mode share to 5% (Marqués et al. 2015). The planners were careful to create a direct network that avoided detours. At the same time, the network was connected, meaning that cyclists could depend on it to get them from residential areas to employment and education hubs. While the network design was based on more than just directness and continuity, these two characteristics have been shown to be fundamental to the success of cycling infrastructure, as explained below.

2.1.2 Directness

Behaviour studies have found that the probability of choosing a route decreases in proportion to its length relative to the shortest route. A study done in Portland, Oregon found that, for commuters, a 1% increase in route length was related to a 9% decrease in probability of utilization (Broach, Gliebe, and Dill 2011). Winters et al. (2010) conducted a similar study and found that while cyclists tend to deviate from shortest routes to those they deem safer, this is only done when the deviation is inside a certain threshold; 75% of trips were found to be less than 10% longer than the shortest route.

2.1.3 Continuity and Density

A study on route choice found that cyclists preferred existing cycling infrastructure to be continuous, especially on arterial roads (Stinson and Bhat 2003). Caulfield, Brick, and McCarthy (2012) conducted a similar survey and found that the majority of respondents stated that connected cycle lanes would encourage them to cycle to work.

While continuity and directness have been shown to positively impact cycling rates, Schoner and Levinson (2014) found that density of the cycling network is also vital [2]. In their study on the bicycle network structure of 74 different US cities, they determine that density is more important for cycling uptake than directness and continuity combined. They therefore warn against expanding a network before densifying it.

2.1.4 Overarching Policies

While segregated, connected, and direct cycling infrastructure is key to achieving high levels of cycling, research has shown that it cannot exist in a vacuum. Wardman, Tight, and Page (2007) developed a mode choice model for the UK and their results showed that improved cycling infrastructure on its own only had modest impacts on mode shift, and even the unlikely scenario of all urban routes being serviced by segregated bike lanes was forecast to increase cycling mode share by only 3%. However, cities that invest in more comprehensive cycling projects show a more significant increase in the number of cyclists as well as the cycling mode share (Pucher, Dill, and Handy 2010). These cities do not just focus on infrastructure, but on general policies as well as restricting car use. Evaluation of policies in Denmark and Germany and the Netherlands has shown that their high cycling mode share is down to a broader set of policies that also include traffic calming, cycling rights of way,

2. making an area's bicycle network denser means adding more cycling routes in the area and thereby giving cyclists more route options

bike parking, integration with the public transport network, and making driving cars both expensive and inconvenient (Pucher and Buehler 2008). While these policies are outside the scope of this research, it is important to recognize their key role in bringing about an increase in levels of cycling.

2.2 Planning Cycling Networks

The studies outlined above lay out the fundamentals for designing cycling networks that generate significant cycling uptake, but they do not propose network-level interventions. The logical next step is to incorporate these fundamentals in our methodology. Buehler and Dill (2016) provide a review of the research carried out to link levels of cycling to the quality of existing cycling infrastructure and note that there has been a recent shift of focus from localized, street or intersection level impact towards studying the whole cycling network. They emphasize the promise of this shift in capturing the network-wide effect of street-level interventions. The section below looks at the different network-level studies that have been carried out, and analyzes their effectiveness in incorporating the fundamentals outlined in Section 2.1. Specific attention is given to optimization and network analysis techniques, such as connected components and community detection, and how they are leveraged to suggest cycling network designs. Inspiration for the methodology of this research is drawn from these techniques.

2.2.1 Connected Components

The ease with which street networks can be abstracted as graphs has made graph theory approaches popular in transport analysis. One such approach is the study of sub-components in a network; are all part of the network connected, or is it made up of multiple disconnected components that are unreachable from one another? Given the importance of well connected cycling networks, this technique has been used for cycling network evaluation and design recommendations.

Betweenness Centrality

Betweenness Centrality (c_b) is one of many measures that are used to evaluate graphs. Put simply, it represents the number of shortest paths passing through a node or link in the network (Bloch, Jackson, and Tebaldi 2019). It can therefore show us the relative importance of links in connecting the network together. Betweeness centrality for link i (c_b^i) can be calculated as follows:

$$c_b^i = \sum_{jk} \frac{g_{jk}^i}{g_{jk}} \tag{2.1}$$

where j and k represent the different nodes in the network, and g_{jk} is the shortest path connecting them. Given that more than one shortest path may exist between two nodes, $\frac{g_{jk}^i}{g_{jk}}$ calculates the fraction of shortest paths between j and k that pass through link i. In transport analysis, where we can represent road segments as links, this measure is very useful; links with high betweenness centrality are normally the road segments with the highest through-traffic.

An example of this would be the work of Natera et al. (2019), who study the cycling network in terms of its disconnected components. They introduce two different algorithms to connect these components by their most critical links[3] and, in doing so, measure the size of the growth of the largest connected component as a function of the kilometers of network added. They observe that small investments at strategic points have a large impact on connectivity in most cases. One limitation of their algorithms is that disconnected components are connected by one link only, which if implemented, would cause this link to have high betweenness centrality. Link redundancy between components is not considered, as the focus is on connectivity and not directness of travel. The concept of connected components is also at the core of the methodology proposed by Olmos et al. (2020). They first calculate potential cycling demand by identifying all non-cycling trips in an OD matrix that could be cycled (distance below a certain threshold), and adding that to existing cycling demand. After

3. *link* refers to a road segment throughout this research

routing this potential cycling demand on the network links, they use percolation theory to filter out the links based on the aggregate flow[4] passing through them. They vary the flow threshold for filtering to identify the minimum flow at which the whole city is connected by a giant component. The results show a cycling network that connects the entire city, and subtracting links intersecting with current cycling infrastructure identifies links proposed for intervention.

The methodologies outlined above have two limitations: [1] they do not account for budgetary constraints, and [2] they focus on connectivity and not directness of travel. Optimization frameworks have been proposed to overcome these limitations.

2.2.2 Optimization

Budget Constraint

Mesbah, Thompson, and Moridpour (2012) propose a bi-level formulation to optimize allocation of cycling lanes to the network without exceeding a set budget. The upper level is the proposed interventions and the lower level is the route choices made by users in reaction to changes in the network. The problem accounts for the effect of cycling lanes on car traffic, and attempts to maximize utilization of said lanes with minimal impact on car travel times. A limitation of this study is that it neglects mode shift, and so car usage is assumed to stay fixed even as improved cycling infrastructure increases incentives for cycling. The emphasis on minimizing impact to car traffic is less relevant in our case, since one of the primary motivations for investing in cycling infrastructure is to reduce car dependency.

Car usage is not considered by Mauttone et al. (2017), who develop an optimization framework that aims to minimize the total user cost of cycling on the network. The aggregate flow on links is obtained by routing cycling demand onto the road network using shortest paths, and the solution is a proposed set of links where cycling infrastructure should be added in order to minimize the overall travel cost of cyclists across the network. The cost of traversing a link is given as a function of its length and whether or not it has cycling

4. *flow* is used throughout this research to refer to the cycling demand when it is routed onto the road network. The flow on any road segment is the cumulative demand on it, resulting from cyclists commuting between various OD pairs

infrastructure. A discontinuity penalty is also added to prioritize connected road segments. By varying the budget constraint (km of cycle infrastructure to be added), they are able to determine an investment threshold at which a high-quality network is created. While this information is undoubtedly useful, it should not be assumed that the necessary budget is available and that the proposed network will be completed in a short period of time; it may be more practical from an implementation perspective to propose links in order of priority. The study also uses existing cycling demand which, as explained by Duthie and Unnikrishnan (2014), reinforces existing cycling patterns and ignores potential cycling demand that could be satisfied by a connected network.

Directness

Duthie and Unnikrishnan (2014) intentionally avoid using commuter data in order to avoid this issue of reinforcing existing cycling patterns. They formulate an optimization problem that aims to minimize the total cost of improving roadway links to meet a desired level of service. This means connecting all OD pairs by links that meet the specified criteria. Unlike Mauttone et al. (2017), they add a constraint to ensure directness, specifying that connecting links between any OD pair cannot exceed a certain multiple of its shortest path length. Adding constraints on deviations from shortest paths ensures that proposed cycling routes are not circuitous and new cycling infrastructure is utilized, as explained in subsection 2.1.2

Palominos and Smith (2020) analyze the road network of London and propose interventions that would benefit active travel and public transportation. They identify important routes by creating shortest paths between all railway and tube stations. The carrying capacity of each route segment is derived from street space analysis (Palominos and Smith 2019), and it enables ranking routes based on available space and shortest paths utilizing them. They find that only 30% of the road network is required to connect all nodes in the study, with utilization of segments exhibiting a left skewed distribution. This again shows that minimal interventions at road segments with high betweenness centrality would result in manyfold improvements in connectivity. The focus on connecting tube and railway stations makes sense from a multi-modal perspective, but it makes the results susceptible to spatial bias, as these nodes are not equally distributed across the city.

2.2.3 Optimization based on Community Detection

The approaches in Sections 2.2.1 and 2.2.2 look at the network as a whole when attempting to improve it. An alternative approach is to identify the different sub-networks that exist within the larger network. Trip patterns in a city are not uniformly distributed geographically, and there are normally localized areas that experience a disproportional number of trips within them. Identifying these areas is referred to as community detection, and it uses the concept of modularity maximization, as defined by Newman (2006). The output is a number of clusters, where the cycling flow on links within the clusters is higher than the flow on links bridging clusters.

Akbarzadeh, Mohri, and Yazdian (2018) use community detection on taxi trips to identify 7 different clusters in the city of Isfahan, Iran. A bi-level optimization problem is then formulated to connect nodes within each community with cycling infrastructure. The objectives of the problem are to [1] minimize travel time between nodes and [2] minimize the total length of the proposed cycling network. To compensate for using taxi trips to detect communities, they only consider connecting nodes in each community that are within reasonable cycling distance of each other.

Bao et al. (2017) adopt a similar methodology, but they use hierarchical clustering to specify the desired number of communities. The main difference in their work is that they set a budget constraint, and so it becomes infeasible to connect all nodes within each cluster. They use a greedy network expansion algorithm where the link with the highest benefit/cost ratio in each cluster is selected, and then the network is grown by adding neighboring links to the solution. The benefit is the flow on the link, and each link is assigned a cost based on current road conditions. Growing the network by adding neighboring links ensures connectivity as it means that the final solution is made up of a connected component in each cluster.

2.3 Underlying Ethical Principles

The methodologies in Section 2.2 are underpinned by different ethical principles, even though these principles are not explicitly acknowledged by the authors. This is important since different ethical principles constitute different problem formulations and targets. Broadly speaking, transport appraisal can be based on either utilitarian or egalitarian principles. The former seeks to maximize the overall benefit, while the latter is concerned with a fair distribution of benefits (Jafino, Kwakkel, and Verbraeck 2020).

Nahmias-Biran, Martens, and Shiftan (2017) criticize the utilitarian approach that has been historically popular in the evaluation of transport investments, explaining how the maximization of overall benefit fails to account for the distribution of that benefit among communities or individuals. Lucas, Van Wee, and Maat (2016) explain how transport studies have traditionally looked at the bigger picture without studying the distribution of investments on the different parts of the study area. They show how policy measures are shaped by their definition of fairness (which has multiple interpretations), and propose an egalitarian approach that ensures the dis-aggregation of transport policy benefits across the study area. Pereira, Schwanen, and Banister (2017) also emphasize the need for a more egalitarian approach to transport planning. They highlight accessibility as a cornerstone of distributive justice, and contend that policies should aim to distribute investments in a way that minimizes spatial variations in accessibility.

A methodological framework that looks at the network as a whole when allocating a limited budget is more in line with utilitarian ethics. Community detection offers a more egalitarian approach if investment is equally distributed between the communities. A similar argument can be made for cycling demand; weighing a network by existing cycling demand risks improving areas with already superior cycling infrastructure and exacerbating spatial inequalities in service provision.

3. Methodology

The main purpose of the research is to propose a methodology for prioritizing road segments for investments in segregated cycling infrastructure. To do so, the work takes inspiration from the different studies mentioned above, and proposes three sub-methodologies that expand on some of their limitations. These limitations include [1] bias inherent when using existing cycling demand, [2] proposing routes that may not correspond to studies on cyclist preference and government policies, and [3] insufficient consideration of the ethical principles underlying the analysis.

In terms of potential cycling demand, a methodology based on analyzing cycling uptake in different areas is proposed. This avoids the issue of reinforcing existing cycling patterns, outlined by Duthie and Unnikrishnan (2014).

To route the calculated potential cycling demand onto the road network, cyclist preferences are considered so as to generate routes that are likely to encourage people to cycle. This expands on the work of Mauttone et al. (2017), by going beyond favoring roads with existing cycling infrastructure to creating a hierarchy of road preference.

Finally, three algorithms are developed to suggest cycling networks based on the aggregated cycling flow on the road network. The first is the most basic, while each of the other two is grounded in one of the two ethical principles outlined above: utilitarianism and egalitarianism.

3.1 Geographical Scale of Analysis

The analysis is heavily dependant on Origin-Destination census data (commuter data). Commuter data is normally based on aggregated zones that do not lend themselves well to the granular nature of active travel. The UK is not an exception, with the most granular commuter data being at the Middle layer Super Output Area (MSOA) level; the average MSOA has a population of 8209 (ONS 2018). Iacono, Krizek, and El-Geneidy (2010) note that such large travel zones are not ideal for understanding route choice behaviour of cyclists and pedestrians. They also give rise to an 'ecological fallacy' whereby average characteristics are assumed to apply to all residents of the aggregated geographical area. This fallacy is evident here, as all trips starting or ending in an MSOA are assigned to a specific point, whereas in reality there are various origins and destinations within each MSOA. Given that more granular data is not publicly available, the study uses MSOA-level commuter data. The methodology is however applicable to more granular commuter data should it become available.

3.2 Calculating Potential Cycling Demand

The literature on cycling is generally in agreement that building cycling infrastructure leads to an increase in the number of cyclists. This phenomenon of *induced demand* has been heavily studied when it comes to road capacity and motorized traffic (Goodwin 1996), and it makes logical sense for it to extend to bicycle networks.

Any study that aims to recommend cycling infrastructure should therefore be based on potential cycling demand, not existing cycling demand. Olmos et al. (2020) calculate potential cycling demand in the city of Bogota as a function of the straight line distance between OD pairs. This study uses a similar approach, but focuses on routed distance and average slope between OD pairs. This is more realistic as straight-line distances are not an accurate representation of distance on the network. Previous research has also confirmed that route hilliness can also be a major barrier to cycling, with an analysis of UK census commuting data showing that a 10% increase with hilliness is correlated with a 9% decrease in cycling mode share (Parkin, Wardman, and Page 2008).

Route distance is based on weighted shortest paths, with the weights reflecting preferences of average cyclists (See section 3.3). The slope between each OD pair is calculated as the average slope of the segments along the routed path between them. This gives us the basis on which to calculate potential cycling demand. This process is done in three steps:

1. **Predict Probability of Cycling Between Each OD Pair Based On Geography**: A generalized linear model is used to predict the probability of cycling for each OD pair based on the two variables mention above: distance $\boldsymbol{d}$ and slope $\boldsymbol{s}$. In reality

the model follows Lovelace et al. (2017) in using both $\boldsymbol{d}$ and $\boldsymbol{\sqrt{d}}$ as predictors. This is done to more accurately represent the bell-shaped nature of cycling frequency relative to distance; very short trips are more likely to be walked than cycled, and so the highest cycling uptake is after a certain distance threshold, as can be seen in Figure 4.2. The probability of cycling $\boldsymbol{P(c_{ij})}$ between i and j is calculated as a function of $\boldsymbol{d}$ and $\boldsymbol{s}$:

$$P(c_{ij}) = \beta_1 d_{ij} + \beta_2 \sqrt{d_{ij}} + \beta_3 s_{ij} \tag{3.1}$$

2. **Accounting for Existing Mode Share** : Assuming that OD pairs with the same geographic characteristics can attract the same number of additional cyclists is naive as it fails to account for existing number of cyclists; OD pairs with a lower cycling mode share have more potential (or latent demand). We therefore use both $\boldsymbol{P(c_{ij}})$ and a scaling factor $\boldsymbol{\gamma_{ij}}$ that accounts for the performance of the OD pair to calculate the fraction of non-cyclists that could be converted to cyclists.

$$U_{ij} = P(c_{ij}) * (Non\,Cyclists_{ij} * \gamma_{ij}) \tag{3.2}$$

This fraction, $\boldsymbol{U_{ij}}$, when added to the current number of cyclists, is referred to as unweighted potential cycling demand $\boldsymbol{PD^U_{ij}}$ [1]. $\boldsymbol{\gamma_{ij}}$ ensures that the fraction of cyclists added $\boldsymbol{U_{ij}}$ is inversely proportional to the existing cycling mode share of the OD pair. To obtain $\boldsymbol{\gamma_{ij}}$, we first determine the performance of each OD pair $\boldsymbol{\alpha_{ij}}$ by getting the ratio of its actual cycling mode share $\boldsymbol{\phi(c_{ij})}$ to its probability of cycling $\boldsymbol{P(c_{ij})}$:

$$\alpha_{ij} = \phi(c_{ij})/P(c_{ij}) \tag{3.3}$$

We then use $\boldsymbol{\alpha_{ij}}$ in a negative exponential function to calculate $\boldsymbol{\gamma_{ij}}$ as follows:

$$\gamma_{ij} = e^{-ln(2)\alpha_{ij}} \tag{3.4}$$

1. The value for $Non\,Cyclists_{ij}$ does not include pedestrians, as we do not aim for a mode shift from walking to cycling.

A negative exponential is used so that OD pairs with better performance ($\boldsymbol{\alpha_{ij}}$) are allocated less additional cyclists than those with low performance. $ln(2)$ is used so that OD pairs with performance = 1 are scaled down by half. OD pairs with performance = 0 are not scaled down. The scaling factor is somewhat arbitrary and further work should be done to determine a more appropriate scaling factor that does not allocate an unreasonable number of potential cyclists to any OD pair.

3. **Scaling Results To Match Mode Share Target**: The above can be sufficient, but it is not aligned with government targets. The UK 2017 target was to double cycling by 2025 (DfT 2017). In response to the covid-19 pandemic, the government has provided emergency active-travel funding to accelerate this modal shift. The 2011 cycling mode share in the UK averaged at 3%, but this varies between cities as shown in Figure 3.1.

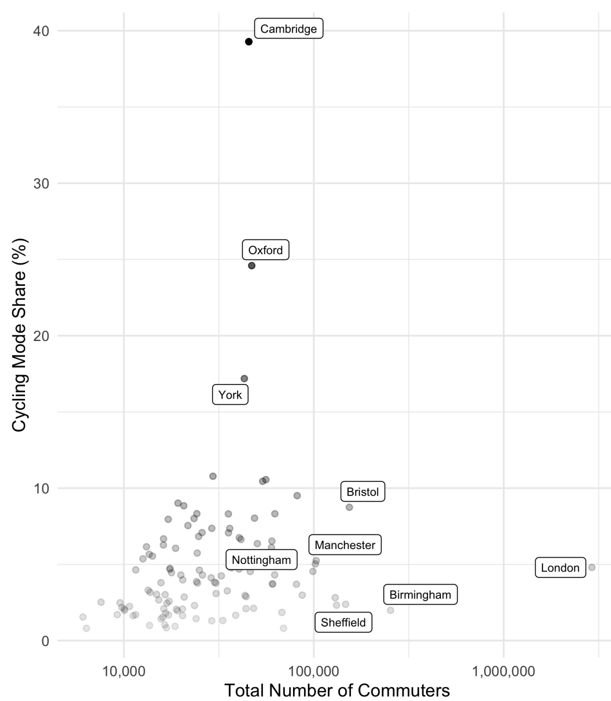


Figure 3.1: Cycling Mode Share In UK Cities

Given the variation in current cycling mode shares among cities, it is clear that they have different potential for an increase in cycling; Cambridge for example has a cycling mode share of 40%, which is comparable to dutch levels, so it's potential for additional cyclists is lower than Manchester, which has a cycling mode share of only 5%. For this study, potential cycling demand is based on a mode shift of 10% towards cycling. A mode shift of 10% is acceptable, since the cities analyzed [2] all have a cycling mode share less than 6%. However, it could be improved by setting specific targets for each city.

A scaling factor $\boldsymbol{\kappa}$ is used to adjust the estimates of the unweighted potential cycling demand ($\boldsymbol{PD^U_{ij}}$) so that they sum up to the mode share target. This factor is:

$$\kappa = PD^W / \sum_{i,j} PD^U_{ij} \tag{3.5}$$

where $\boldsymbol{PD^W}$, the weighted potential cycling demand, is the total number of cyclists in the city if the cycling mode share increased by 10%.

If we multiply the unweighted potential cycling demand $\boldsymbol{PD^U_{ij}}$ for each OD pair by $\boldsymbol{\kappa}$, we obtain the potential cycling demand for that OD pair, $\boldsymbol{PD^W_{ij}}$. This potential demand ensures a 10% cycling mode share increase in the city.

3.3 Routing

The next step is to route the potential cycling demand ($\boldsymbol{PD^W_{ij}}$) between all OD pairs onto the road network. The routing needs to consider cyclist preferences in order to be realistic. Below is an outline of how the routing is done, and what considerations are taken when weighting the road network.

3.3.1 Routing Engine Used

To route potential demand onto the road network, the **dodgr** r package is used (Padgham 2019). The package uses the OpenStreetMaps (OSM) road network and allows the user to assign weights to roads based on their type (see Table 3.1). The routing is done based on

2. Manchester, Birmingham, Nottingham

weighted shortest paths, with the distance along each road segment being divided by a factor to obtain the weighted distance for routing [3].

3.3.2 The Implications of Weighting the Road Network

It is important to stop and consider whether routing based on weighted distances is desired and, if so, what should be considered when assigning weights.

Table 3.1: OSM Road Types

OSM Road Type	**Description**	**UK Equivalent**
Motorway	Road open to high-speed vehicles only	Motorway
Trunk[4]	Roads that don't meet motorway criteria, in descending order of importance and through-traffic	A-Roads with Primary Status
Primary		A-Roads with Non-Primary Status
Secondary		B-roads
Tertiary		Classified unnumbered roads
Unclassified		Unclassified (Intended for local traffic - 60% of UK roads)
Residential	Function is purely residential	
Service	Road that provides access to a facility	
Track	Unpaved - suitable for two track vehicles (mostly rural)	
Cycleway	Designated cycleway - open to cyclists only	
Path	Unpaved - open to non-motorized traffic only	

3. It is more intuitive to multiply when weighting a network, but the dodgr package divides by numbers between 0 and 1, which achieves the same result. For the sake of reproducibility, we stick to the convention used in the package

4. The definition of trunk roads differs between the UK Department for Transport (DfT) and OSM. The DfT defines trunk roads as those that are responsible for distribution of goods and services on a national level. They are roads that connect major towns and cities (DfT 2012). Trunk roads on OSM are equivalent to A-roads in the UK road classification

The argument for weighted routing

Simply obtaining shortest paths between OD pairs is naive and fails to account for the route-choice preference of cyclists (outlined in Section 2.1).

Weighting the segments of the road network to favor roads with existing cycling infrastructure also allows us to determine shortest paths that make use of this infrastructure. Utilizing existing infrastructure makes economic sense, as small investments may lead to large connectivity gains as the disconnected cycling infrastructure gets joined together.

Choosing not to weigh the road segments would give us ideal shortest paths between destinations. The issue here becomes that some routes will certainly go through roads with high levels of motorized traffic, which tend to be the most direct (i.e. trunk roads). This would not be a problem if cycling infrastructure was added to all of the identified shortest paths, but the more likely situation is that budget constraints and restrictions on road space reallocation could limit the investment to only a fraction of the identified road segments. The consequences of that would be a network that goes through main traffic routes with only disjointed cycling infrastructure.

The argument against weighted routing

The purpose of routing potential cycling demand in our case is to suggest road segments for segregated infrastructure; we are routing potential cycling demand that will be created by *future* investments in cycling infrastructure. We are therefore less restricted by existing road types as cycling infrastructure can hypothetically be added to any road segment.

Assuming that the weights would favor routing on existing cycling infrastructure, the quality of the recommendations becomes dependent on how well that infrastructure contributes to direct paths between destinations. It could well be that it was not designed with a network perspective, and would cause overly circuitous routes. Indeed Boisjoly, Lachapelle, and El-Geneidy (2020) note that, in Montreal, connectivity between Boroughs is poor as a result of cycling infrastructure being planned at the Borough level and not at the city level.

Verdict

Routing potential cycling demand on the street network when planning for the future is therefore not a trivial task. It helps to ask: what demographic constitutes this targeted potential demand? Work done by Dill and McNeil (2013) on examining cyclist typologies determined that around 60% of Portland residents fit under the *interested but concerned* category. These were people that enjoyed cycling but avoided it due safety concerns. The key to encouraging this group was to create a low-stress cycling network, not only though segregated infrastructure but also by planning routes that passed through residential streets.

This does not mean avoiding non-residential streets. These streets have main trip attractors and also have more road space that can be re-allocated to cycling infrastructure (Marqués et al. 2015). What this does mean, is that we can adjust the weighting profiles to favour less stressful streets. From the descriptions of the road types in Table 3.1, we can understand their stress levels relative to one another. This information is used to create the weighting profile in Table 3.2 below:

Table 3.2: Weighting Profiles

OSM Road Type	**Weight**
Cycleway	1
Path	0.9
Residential	0.9
Service	0.9
Tertiary	0.9
Track	0.9
Unclassified	0.9
Secondary	0.8
Primary	0.7
Trunk	0.6
Motorway	0

All weights are between 0 and 1. The closer the weight is to 1, the more desirable the road type is for cyclists. Roads with a weight of 0 cannot be traversed by cyclists. The weights are chosen so as to be inversely proportional to the stress level experienced by cyclists on them; given the choice between a primary and secondary road of equal distance, the route will go through the latter.

A weighted distance $\boldsymbol{d_w}$ for each road segment is calculated as following:

$$d_w = d_{unw} \, / \, W \tag{3.6}$$

where $\boldsymbol{d_{unw}}$ is the unweighted distance and $\boldsymbol{W}$ is the weight from Table 3.2. The low weights on trunk and primary roads mean that they will be avoided when routing unless they offer significant distance gains. The marginal difference between the weights of different road types means that the more stressful road will only be selected when there are no routes with comparable distances that avoid it. Leaving space for motorized traffic on trunk and primary roads could also be useful for the creation of Low Traffic Neighborhoods (LTNs), where motorized vehicles are restricted to main roads.

3.4 Prioritizing Road Segments for Segregated Infrastructure

After routing the potential cycling demand onto the road network using weighted shortest paths, we have estimates for the cumulative potential cycling demand passing through all road segments. This cumulative demand (referred to as *flow*) is then used as a basis for determining where best to invest in segregated cycling infrastructure. In doing so, we must account for the motivations and deterrents for cycling identified in Section 2.1, namely direct and well connected routes.

For this purpose, three algorithms have been developed. The first is the most basic, and it shows the logic behind creating a connected and direct cycling network. The next two are more advanced since they utilize existing infrastructure from the beginning and allow us to compare a solution that focuses on utilitarianism to one that focuses on egalitarianism. In all algorithms, links[5] are selected iteratively and the iteration at which each link is added to the solution is recorded. Investments in cycling infrastructure can be limited by budget constraints, so it can be useful to see where best to allocate a defined length of segregated infrastructure.

3.4.1 Algorithm 1: Growth From One Origin

1. Identify link with highest flow and use it as a starting point for the solution
2. Identify all links that neighbor links in the current solution
3. Select neighboring link with highest flow and add it to the solution
4. Repeat steps 2 & 3 until all flow is satisfied or investment threshold is met

This algorithm ensures that the resulting network is connected. It also satisfies the directness criteria, since links on the weighted shortest paths are those that have the highest flow passing through them (this is a result of the routing in Section 2.1). This algorithm is suitable for small urban areas that have no existing cycling infrastructure, and so are not concerned about how their plans align with their existing infrastructure or how this infrastructure should be fairly distributed across the urban area. It can also be used in areas where the network will be overhauled, perhaps because existing cycling infrastructure is judged to be placed in locations that require large detours by cyclists.

It is, however, somewhat restricted by its starting point. There may be links in different parts of the network with very high cycling flow, but only one of them is chosen as a starting point. This can be true for big urban areas where commuter flow is polycentric (i.e. not all directed at one Central Business District). If this were the case, this algorithm would favor one area at the expense of others.

5. *link* refers to a road segment

Identifying Existing Cycling Infrastructure

OSM is used to identify where segregated cycling infrastructure exists. This identification is however not straightforward; the reason why this is the case, and the workaround developed, is explained below.

There are various tags on OSM for tagging cycling infrastructure on roads. The tag used depends on the nature of the infrastructure. Broadly speaking, OSM differentiates between cycle lanes and cycle tracks; the former is on the road, while the latter is separate from the road. We are only concerned with cycle tracks, as they are what fit our definition of *segregated cycling infrastructure*, and are the key to getting people to cycle (see Sections 2.1.1 & 3.3.2). There is more than one way of tagging a cycle track on OSM:

1. highway = cycleway
2. cycleway = track
3. bicycle = designated

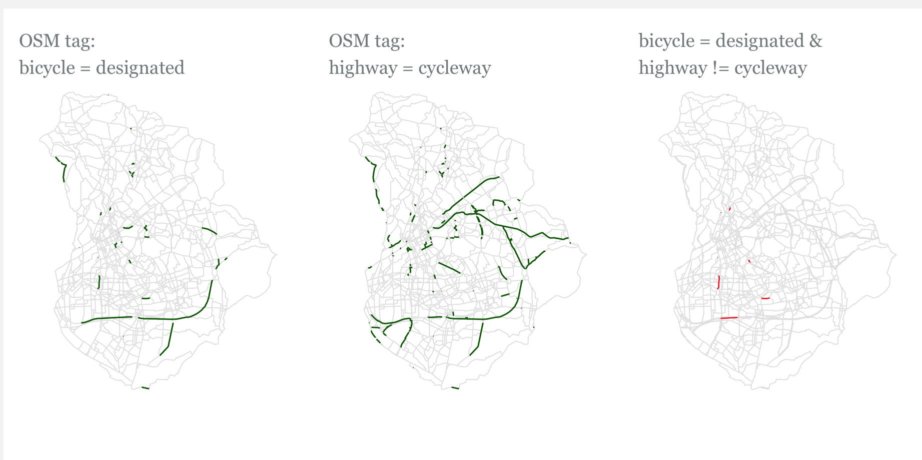


Figure 3.2: Querying OSM for Roads with Segregated Cycling Infrastructure

While the first tag should encompass all segregated cycling infrastructure, inconsistencies in tagging (shown in Figure 3.2) mean that this is not necessarily the case. In the frame on the right we see that some roads have the 'bicycle = designated' tag but are not tagged as 'highway = cycleway'. *To obtain all roads with segregated cycling infrastructure, we identify all roads that meet* ***any*** *of the three tags above.*

3.4.2 Algorithm 2: Utilitarian Growth

1. Identify all links that have segregated cycling infrastructure and add them to the initial solution
2. Identify all links that neighbor links in the current solution
3. Select neighboring link with highest flow and add it to the solution
4. Repeat steps 2 & 3 until all flow is satisfied or investment threshold is met

This algorithm differs from the first one in that it adds all existing cycling infrastructure to the solution from the start. The network growth is therefore less restricted in the beginning, as the pool of neighboring links is bigger. While it may seem intuitive to add all cycling infrastructure from the beginning, there is an important point that should be considered: the existing cycling infrastructure may be very disconnected and inefficient. Adding links that neighbor the existing infrastructure may result in a disconnected network. This is especially true if the budget is small, in which case it may be better to target the investment in one area.

Community Detection

One of the main aims of this research is to incorporate egalitarian principles by fairly distributing investments in cycling infrastructure. One way of quantifying this is to split up the city into smaller geospatial areas and target equal investment in each of those areas. The question remains: How is such a split defined?

In this case, it is defined by looking at cycling commuting patterns; are they randomly distributed across the city, or do they exhibit more localized patterns?

It would make sense for a city the size of Manchester to exhibit the latter when it comes to cycling; cyclists are limited in their commuting distance (see Figure 3.3), and so trip attractors are more likely to have a local catchment area of cyclists.

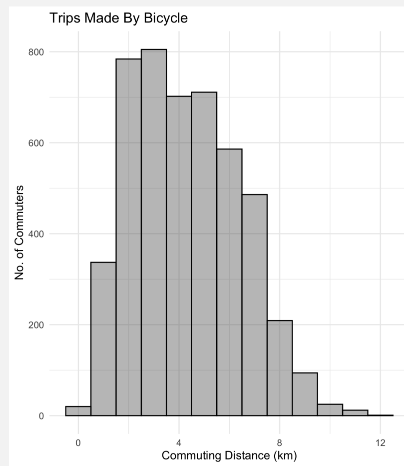


Figure 3.3: Cycling Commuting Distance - Manchester (2011 Census Data)

Community detection, as described in Section 2.2.3, offers us a way to delineate such a split. At its core, it is a way of dividing a network into sub-networks, or communities (Barabási et al. 2016). In our case, the network is the city; the nodes are the population-weighted MSOA centroids and the links connecting each MSOA pair are weighted by the potential cycling demand between them. The Louvian method (Blondel et al. 2008) is used to separate MSOAs into communities. Potential cycling demand is used since we assume that this is what the cycling demand will be once the cycling infrastructure is added. To assign road links to communities, the following steps are carried out:

1. Create links between MSOA centroids and weigh these links by potential cycling demand between them.
2. Use Louvian method to determine optimal number of communities and assign each MSOA centroid to a community.
3. Assign each road link to the same community as the closest MSOA centroid to it.

3.4.3 Algorithm 3: Egalitarian Growth (Focus on Fair Distribution of Resources)

The first two algorithms focus on connectivity and directness, but not on fairly distributing investment. The latter is not a requirement for increasing cycling uptake, but it is fundamental for spatial equity, as explained in Section 2.3. This algorithm incorporates the ideal of fair distribution by using community detection to partition the road network.

The algorithm uses the following logic to ensure fair distribution between communities:

1. Identify all links that have segregated cycling infrastructure and add them to the initial solution
2. Identify all links that neighbor links in the current solution
3. Select *from each community* one neighboring link with highest flow and add it to the solution
4. If there are no more neighboring links in a community, select the link with the highest flow in that community, regardless of connectivity, and add it to the solution
5. Repeat steps 2, 3 & 4 until all flow is satisfied or investment threshold is met

Even though we may end up with a more disconnected network, we will have separate connected networks in each community. Given that communities are defined by having more internal flow than external flow, this is a satisfactory solution.

4. Results

In this section we first look at the cycling uptake across the city and how potential cycling demand is distributed across the different OD pairs in it. We then look at the routing results to see the effect of weighting the road network. Finally we look at the community detection and algorithm results, and compare the performance of the three algorithms.

4.1 OD Pair Performance

To visualize how the cycling mode share varies across the city, we separate the OD pairs into 1 km distance groups. The performance of all OD pairs is obtained by getting the ratio of their cycling mode share to their probability of cycling (See equations 3.1 and 3.3). From Figure 4.1, we can see that the performance is higher in the South West, while OD pairs in the North and East tend to have a lower performance.

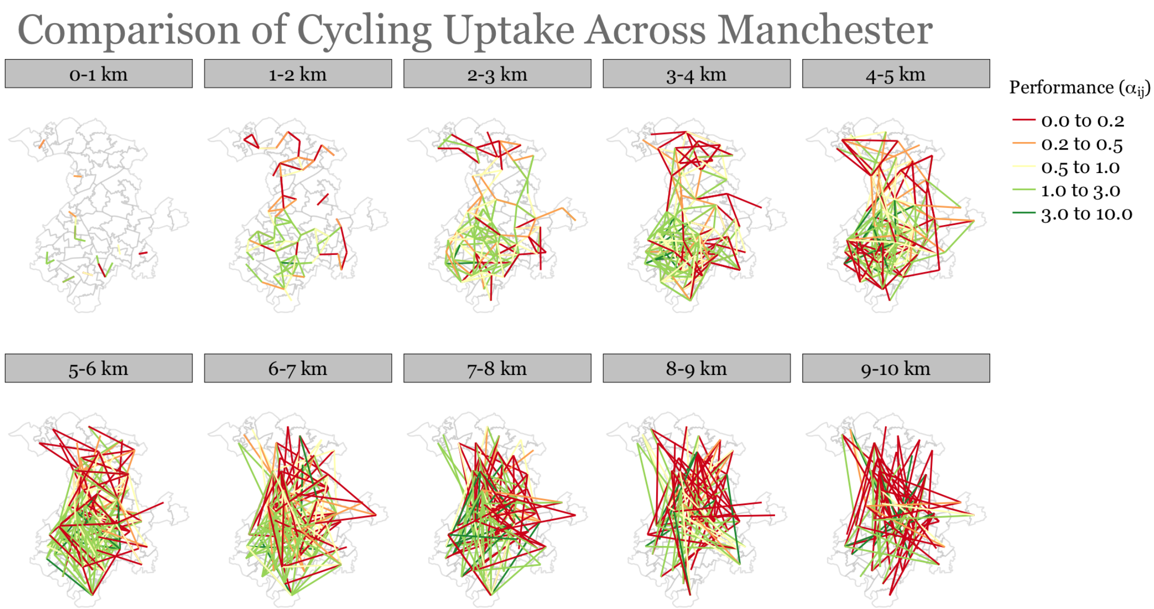


Figure 4.1: Performance $\boldsymbol{\alpha_{ij}}$ = Cycling Mode Share $\boldsymbol{\phi(c_{ij})}$ / Probability of Cycling $\boldsymbol{P(c_{ij})}$

4.2 Potential Cycling Demand

The potential demand calculations show that the current and potential number of cyclists both follow a bell-shaped distribution, with the number of trips peaking around the 3-5km commuting distance and then going back down for longer distances (See Figure 4.2).

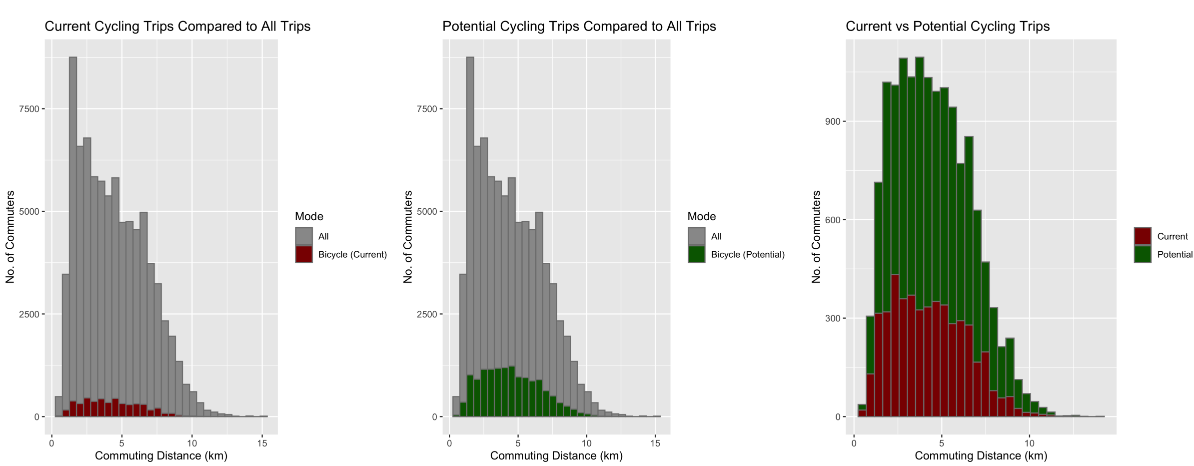


Figure 4.2: Distribution of Potential Cycling Demand

As explained in Section 3.2, the distribution of potential cycling demand is meant to reflect not only the physical geography (i.e. distance and slope), but also the current cycling mode share. OD pairs that have low performance ($\boldsymbol{\alpha_{ij}}$) should be allocated more additional cyclists (See equation 3.2). Figure 4.3 shows that this is the case, with cycling mode share increase being inversely proportional to the OD pair performance.

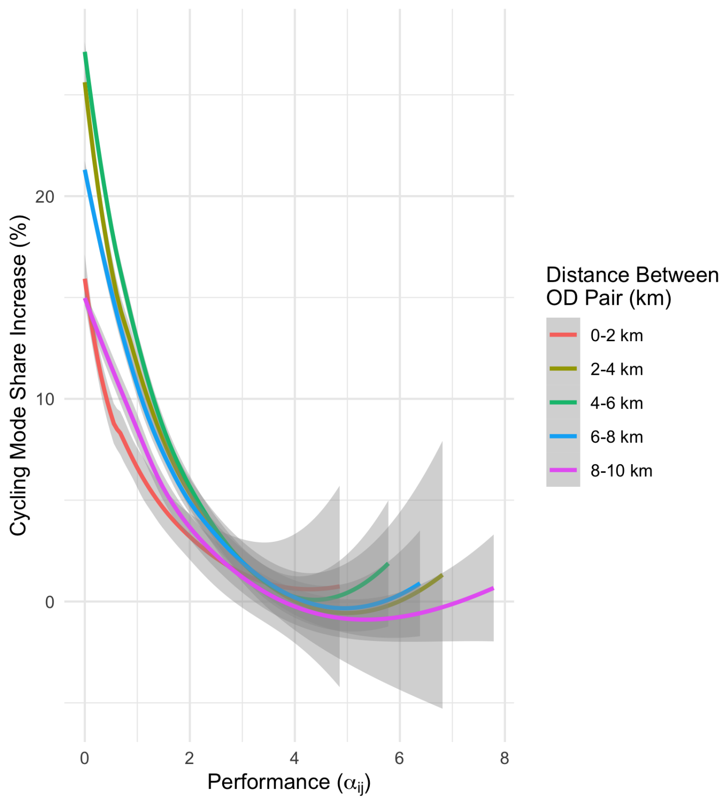


Figure 4.3: Distribution of Potential Cycling Demand Relative to OD Pair Performance

We can also see that the highest cycle mode share increase is in OD pairs between 2-8km apart. This is in line with the bell-shaped distribution of cycling uptake shown in Figure 4.2.

Figure 4.4 shows the geographical distribution of current and potential cycling demand. The increase in the Northern and Eastern regions is consistent with their low existing cycling mode share (See 4.1).

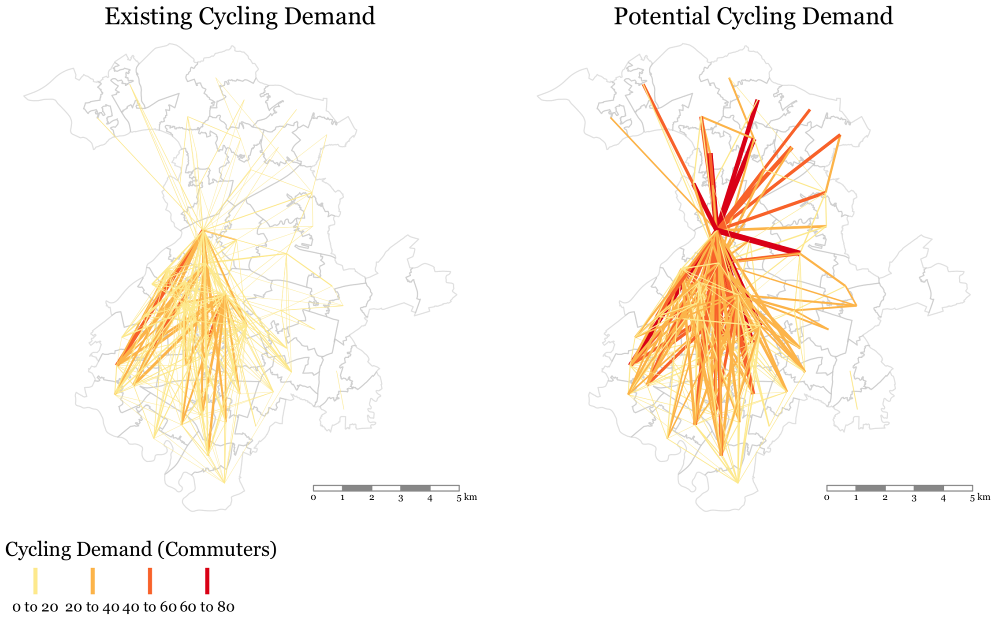


Figure 4.4: Current and Potential Cycling Demand

4.3 Routing

Comparing the cycling demand routed on the weighted and unweighted road network allows us to get a better understanding of the importance of different road types. In the case of Manchester, trunk roads bisect the city and are a major part of unweighted shortest paths (Figure 4.5). On the other hand, cycleways are not part of unweighted shortest paths, and so very little of the cycling demand is routed through them. In the weighted network, cycleways are much better utilized, and the majority of the cycling demand passes through tertiary roads, as expected.

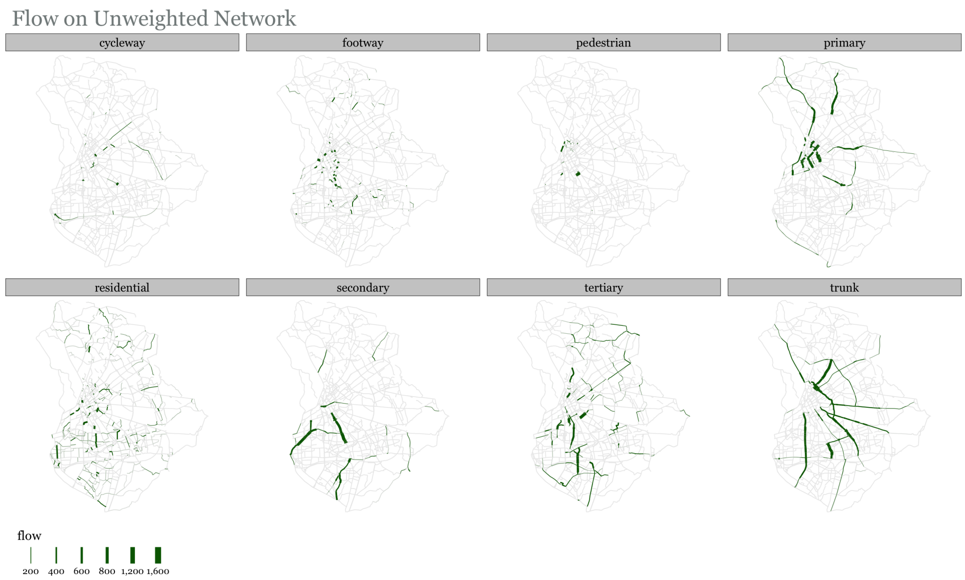


Figure 4.5: Flow Results Based on **Unweighted** Shortest Paths (Manchester)

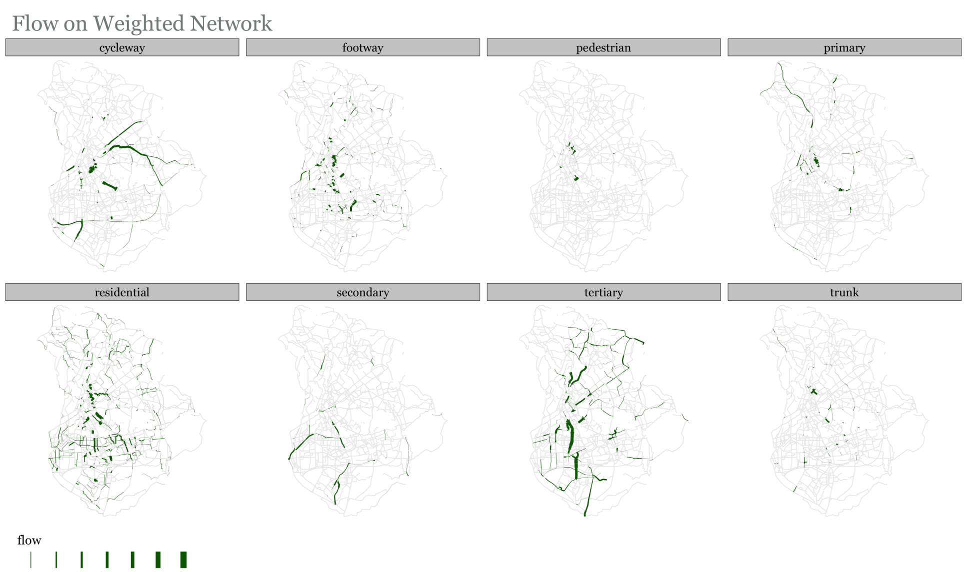
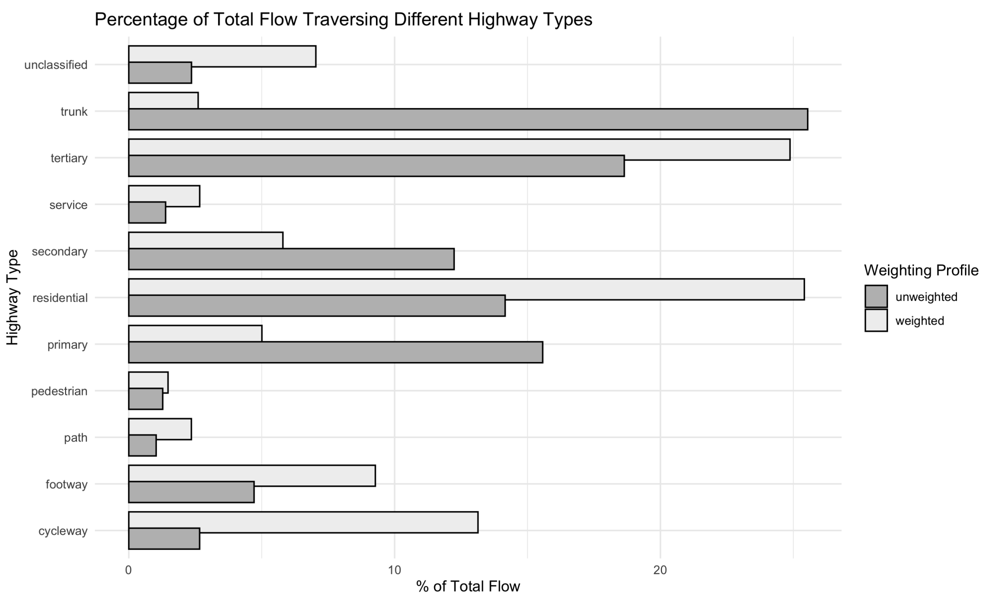
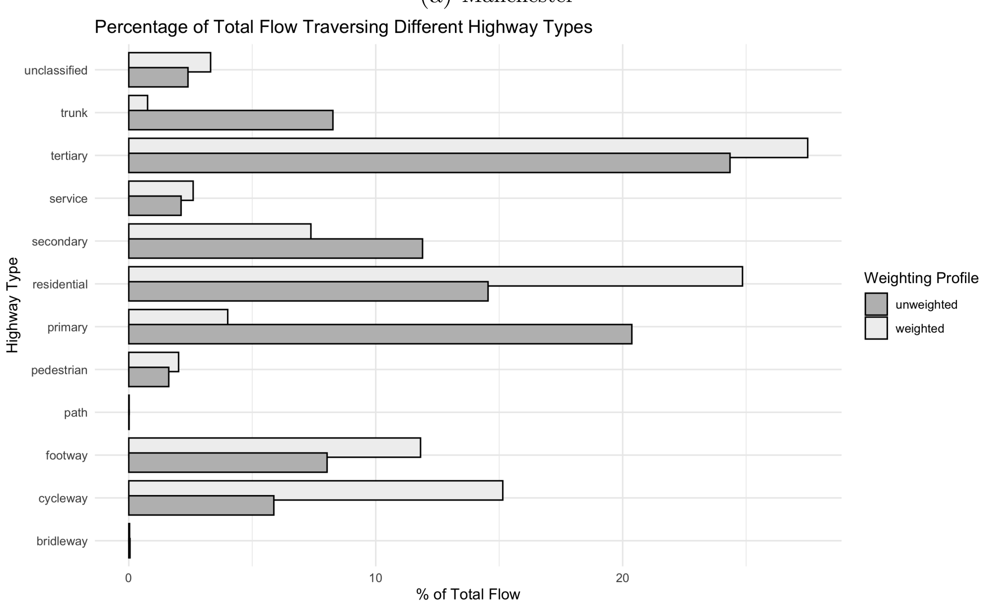


Figure 4.6: Flow Results Based on **Weighted** Shortest Paths (Manchester)

The difference between aggregate flow on weighted and unweighted networks is dependant on the road network of the city. Comparing Manchester to Nottingham, we see that trunk roads are much more important in the former, as over 25% of flow on the unweighted road network passes through them. For Nottingham, less than 10% of the flow on the unweighted network passes through trunk roads, but almost 25% of the flow passes through tertiary roads (Figure 4.7).



(a) Manchester



(b) Nottingham

Figure 4.7: Comparison Of Routing for Road Networks of Different Cities

4.4 Prioritizing Road Segments for Segregated Infrastructure

Community detection carried out based on potential cycling demand reveals that Manchester can be split into three large communities and one small one (Figure 4.8).

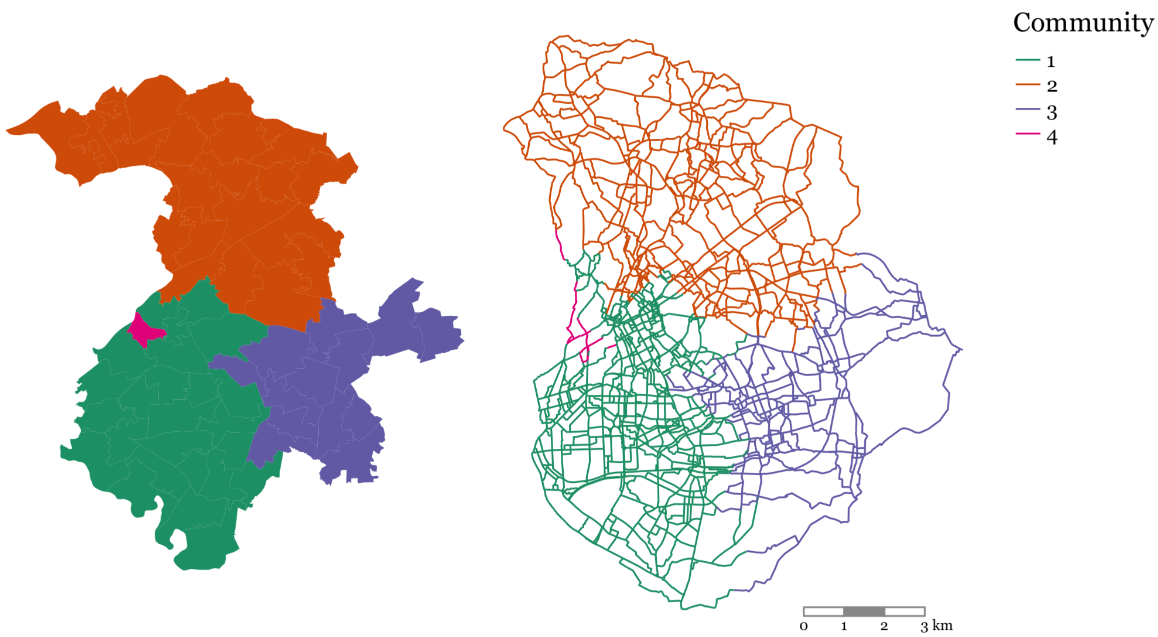
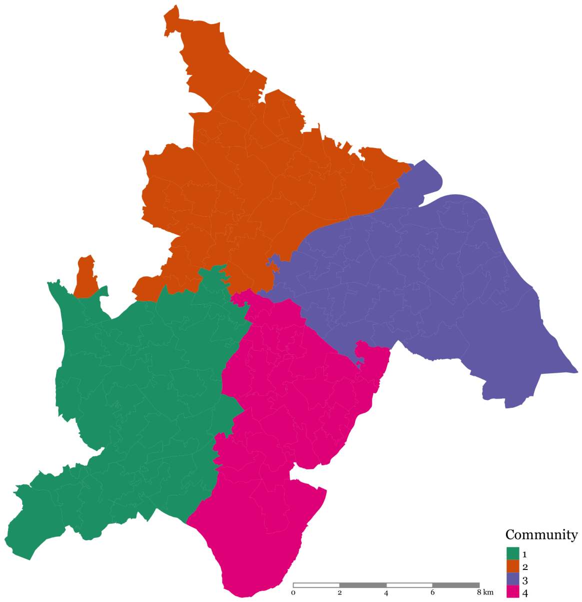


Figure 4.8: Communities Based on Potential Cycling Demand (Manchester)

The number and size of the communities varies between cities, but in general the communities tend to be spatially explicit, as can be seen for Birmingham and Nottingham (Figure 4.9).



(a) Birmingham

(b) Nottingham

Figure 4.9: Community Detection for Different Cities

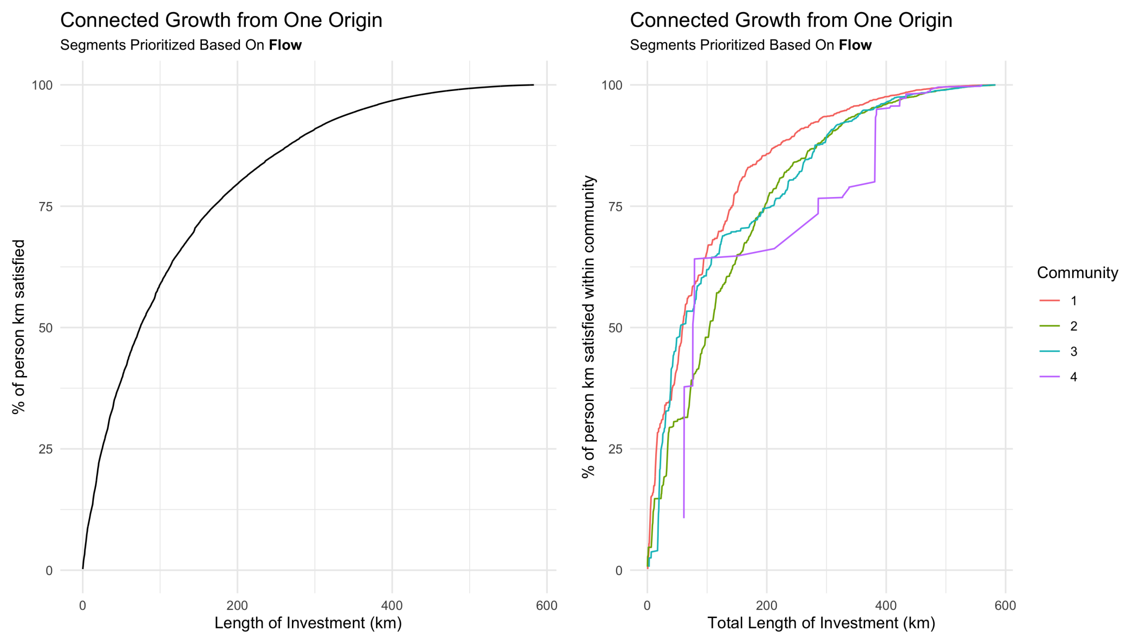
The results of the community detection are used to evaluate the algorithms. For all three algorithms, we look at the *person-km satisfied* as cycling infrastructure is added. Person-km is a measure of the total km cycled on a road segment, so it is the product of the number of potential commuters cycling on that road segment ($flow$) and the length of the segment in km (l). For each road segment, the person-km is equal to $flow * l$. In the case of Manchester, Table 4.1 shows that most of the person-km is in communities 1 & 2, while less than 1% of total person-km on the network is in community 4.

Table 4.1: Total Person-Km in Different Communities (Manchester)

Community	**Person-Km (Total)**	**Person-km (%)**
1	31,676	50.6
2	20,722	33.1
3	9,733	15.5
4	496	0.8

4.4.1 Algorithm 1: Growth From One Origin

Looking at Figure 4.10, we see that 100km of investment in segregated cycling infrastructure is sufficient to satisfy around 60% of the person-km routed on the network. However, when we zoom in on the community level (Figure 4.10b) we see that there is variation in the person-km satisfied. At 100km of investment, over 70% of the person-km in Community 1 are satisfied compared to less than 45% for Community 2

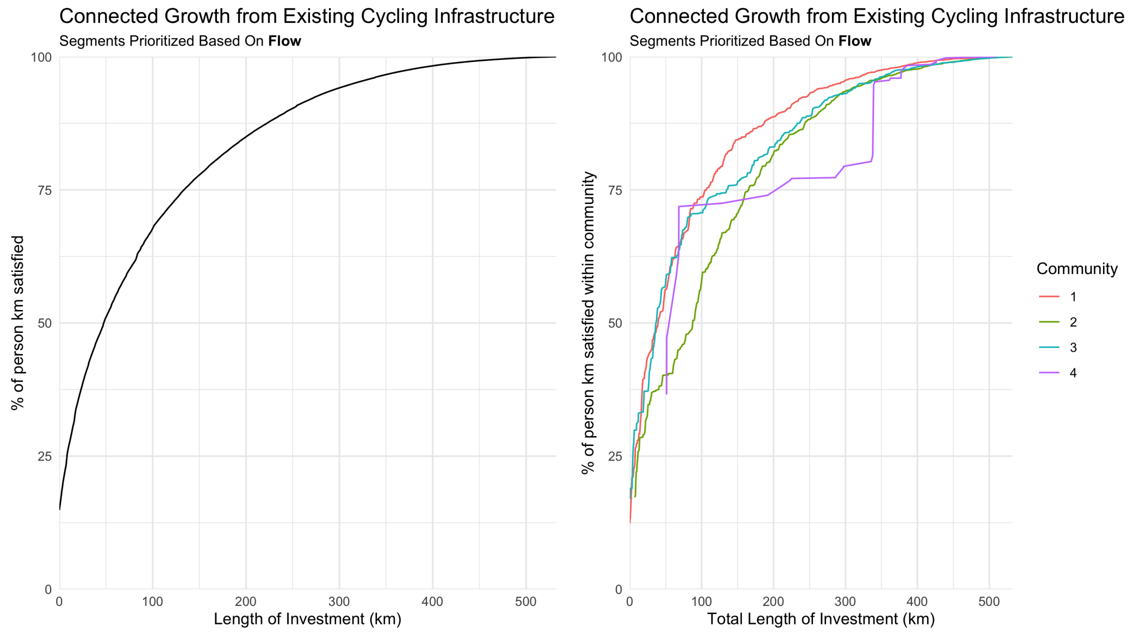


(a) % of Person-Km Satisfied (b) % of Person-Km Satisfied in Each Community

Figure 4.10: Effectiveness of Alg. 1 on Overall and Community-Wide Person-Km Satisfied (Manchester)

4.4.2 Algorithm 2: Utilitarian Growth

Algorithm 2 provides marginally better results than Algorithm 1 when comparing quick gains in person-km satisfied. After 100km of investment, over 70% of person-km on the network is satisfied by a connected network, compared to just 60% for Algorithm 1 (Figures 4.10a & 4.11a). However, these gains are still not distributed equally, with the algorithm largely favouring Community 1, 3 & 4 at the expense of Community 2.



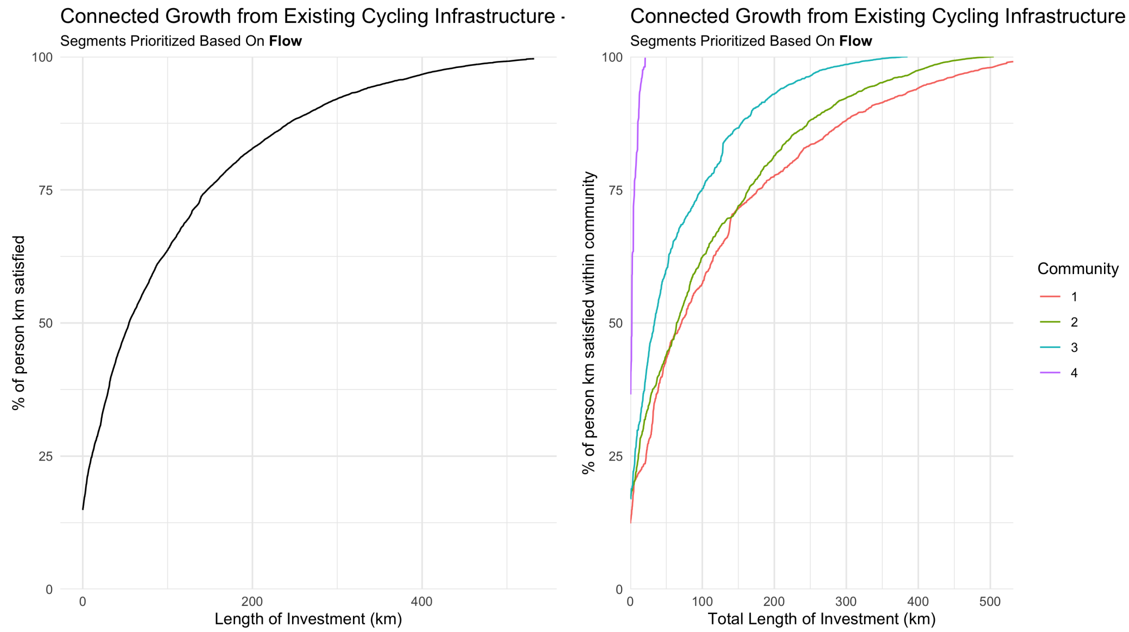
(a) % of Person-Km Satisfied (b) % of Person-Km Satisfied in Each Community

Figure 4.11: Effectiveness of Alg. 2 on Overall and Community-Wide Person-Km Satisfied (Manchester)

4.4.3 Algorithm 3: Egalitarian Growth (Focus on Fair Distribution of Resources)

Looking at the community level, we find that Algorithm 3 provides different results to the preceding algorithms. The allocation of infrastructure is better distributed between communities, as can be seen from the % of person-km satisfied in Figure 4.12b. The exception with community 4 is due to its relatively small size, which means that it requires significantly less investment than the other 3 communities.

Figure 4.12a indicates that the restrictions imposed on the network growth do not seem to have a noticeable effect on the city-wide % of person-km satisfied. Comparing it to Figure 4.11b, we can see that Algorithm 2 provides only marginally quicker city-wide gains than Algorithm 3.



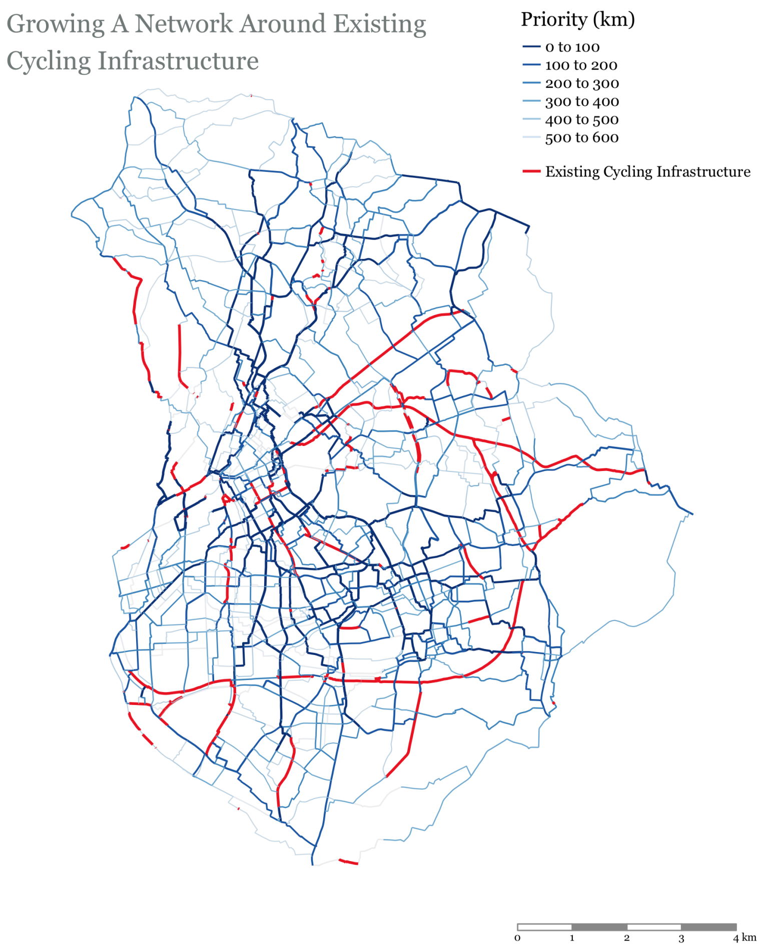
(a) % of Person-Km Satisfied

(b) % of Person-Km Satisfied in Each Community

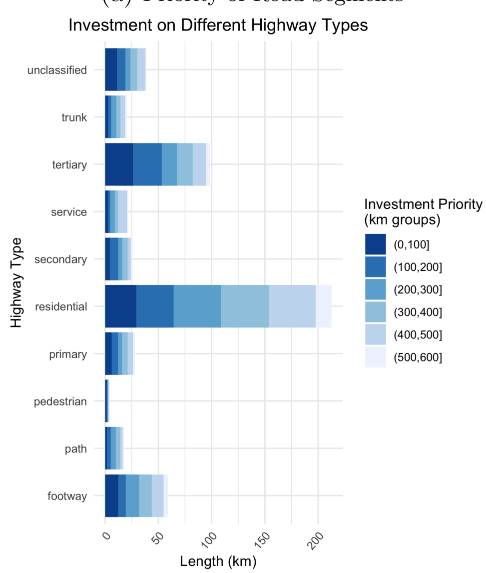
Figure 4.12: Effectiveness of Alg. 3 on Overall and Community-Wide Person-Km Satisfied (Manchester)

Figure 4.13a gives us a visual representation of how the network is grown around existing cycling infrastructure; it shows when each link was added to the solution (first 100km, second 100km, etc). We can see that, generally, road segments around cycling infrastructure are prioritized, except for those neighboring cycling infrastructure on the very periphery. The first 100km is also spatially distributed across the city, with no apparent bias towards a particular area.

It is also important to understand how the different highway types contribute to the proposed network. Figure 4.13b shows that most of the flow will be on residential and tertiary roads, as expected from the weighting profile defined in Table 3.2.



(a) Priority of Road Segments



(b) Contribution of Different Road Types

Figure 4.13: Results of Alg. 3 (Manchester)

4.4.4 Connectivity

Existing cycling infrastructure is made up of many disconnected components. Both Algorithm 2 and 3 start with all existing segregated cycling infrastructure and aim to create an efficient, connected network. Figure 4.14 compares the performance of both Algorithms in terms of connectivity gains. We can see that, for all three cities analyzed, the number of components is gradually reduced as more infrastructure is added (this is true for both Algorithms). However these connectivity gains require considerable investment; for Manchester, both algorithms take almost 150km of investment to reduce the number of components by half. Nottingham and Birmingham show a similar pattern of gradual reduction, requiring around 70 and 250km of infrastructure respectively to reduce the number of components by half. The rate of reduction in number of components appears similar among the three cities, with the differences in results being mainly a function of the initial number of components (which itself appears to be a function of the size of the city). For all cities, there are differences between Algorithm 2 and 3, but these are marginal.

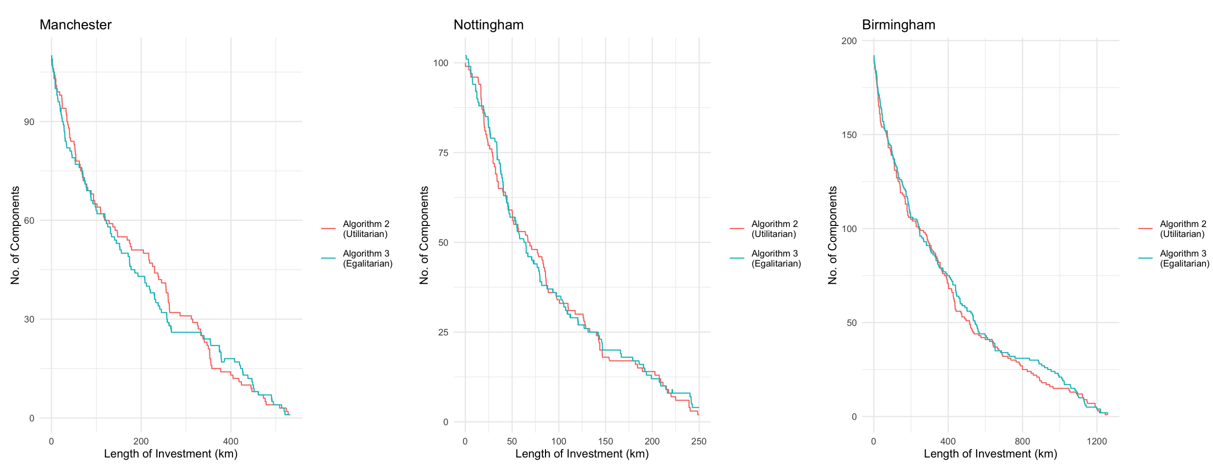


Figure 4.14: Change in the Number of Disconnected Components as Cycling Infrastructure is Added to the Network

Consistent growth can also be seen for the size of the Largest Connected Component in the proposed bicycle network (Figure 4.15). Again, we can see from the results of the three different cities that there is little difference between both Algorithms.

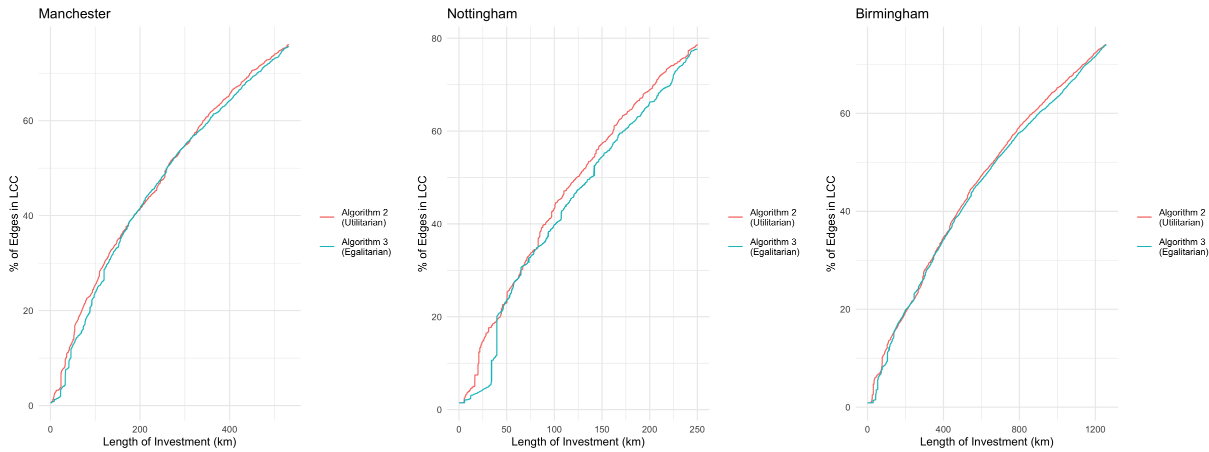


Figure 4.15: Change in the Size of the Largest Connected Component as Cycling Infrastructure is Added to the Network

5. Discussion

5.1 Potential Cycling Demand

The bell shaped distribution of additional trips relative to distance between OD pairs (Figure 4.2) is a good representation of actual cycling patterns. Trips below a certain distance are walkable and so the highest number of cycling trips occurs at a distance beyond that walking threshold.

This is further clarified in Figure 4.3, where we can see that trips in the range of 0-2km have the lowest cycling mode share increase, while trips between 2-6km have the highest. Having a negative correlation between mode share increase and current performance seems intuitive if we consider the problem as one of diminishing returns. The point is made earlier that a city like Cambridge with a cycling mode share of 40% has less potential for additional cyclist than a city that only has a 5% cycling mode share. We extend this logic down to the level of the OD pairs.

It should be noted that the calculations assume a future that is constrained by physical geography; i.e. we consider cycling in the traditional sense. Recently there have been various micro-mobility solutions, including e-bikes, that allow commuters to traverse longer distances and hillier roads with less effort than traditional bicycles. While these modes would probably be associated with less geographical impedance, it is beyond the scope of this work to integrate that into the analysis. Doing so is partially restricted by the lack of data on the proliferation of these modes, which raises the point that perhaps the census data category of 'Bicycle' is too vague, and should be further dis-aggregated to distinguish between traditional bicycles and other forms of micro-mobility.

5.2 Routing

The results of routing potential cycling demand on the weighted and unweighted networks are understandably quite different. From Figure 4.5 we can see that trunk and primary roads are the most efficient means of traversing the road network of Manchester. Both of these road types are classified as Primary A roads according to the UK Department for Transport's road classification (3.1), and are therefore part of the Primary Route Network (PRN) (DfT 2012). The PRN has the widest, most direct roads on the network, and carries most of the through traffic. This includes freight, with all roads in the PRN being required by law to provide unrestricted access to trucks up to 40 tonnes (ibid.).

We choose to avoid routing the potential cycling demand on Primary A Roads for the following 2 reasons:

1. **Logistical Difficulty**: Changes on these roads need to be agreed upon by all affected authorities (ibid.), which may prove to be difficult.

2. **Low Traffic Neighborhoods (LTNs)**: The UK government is aiming to restrict access to motorized vehicles on residential roads to create LTNs (DfT 2020b). This is part of a policy to prevent automobile rat-running and make streets more accessible to cyclists and pedestrians. Under such a policy, Primary A roads would become even more essential for motorized traffic and it would be more difficult to reallocate road space on these roads to cyclists.

Routing potential cycling demand on a weighted network, as done in Section 4.3 is more in line with government policy to create LTNs. Figure 4.6 shows that routing on the weighted network significantly reduces flow on the trunk and primary roads, but does not eliminate it completely. This is intentional, as the impedance on these roads is only slightly higher than remaining road types (See Table 3.2). Potential cycling demand is only routed on these roads if there are no routes through other roads that offer comparable directness.

Banning cycling flow completely on trunk and primary roads may result in excessively circuitous paths, as seen in Figure 5.1. When routing using the weighting profile in Table 3.2, we see that shortest paths increase by less than 5% on average from unweighted shortest paths, with the largest increases still below 30%. When routing on primary and trunk roads is banned, the average increase relative to unweighted shortest paths rises to 10%, with certain locations experiencing more significant negative effects on accessibility.

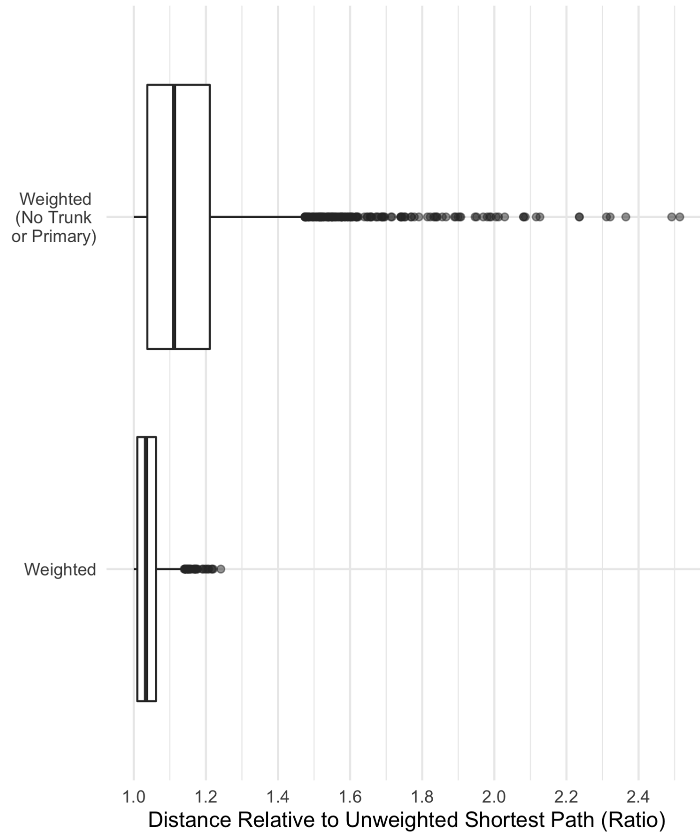


Figure 5.1: Effect of Banning Cyclists from Trunk and Primary Roads for all OD Pairs (Manchester)

Given that cyclists will only deviate from shortest paths by a certain amount to access better cycling infrastructure (as explained in Section 2.1.2), allowing flow on some stretches of trunk and primary roads is necessary to insure cycling uptake and equitable access to cycling infrastructure. In its new vision for walking and cycling, the Department for Transport acknowledges that minimal segregated stretches of bicycle lanes on main roads will be necessary to avoid circuitous cycling networks (DfT 2020b).

Weighting the road network also allows us to better utilize existing cycling infrastructure, as can be seen by the higher flow on cycleways in Figure 4.6. Again, the small differences in impedance between cycleways and other road types mean that cycleways that require significant deviation are not routed on.

It should be reiterated that the weighting profile used for routing has been developed for the purposes of this study. It creates a hierarchy of road preference that is grounded in cyclist preference and government plans to create LTNs. Sensitivity analysis should be done to determine an optimal weighting profile, but given the variation in city road networks (Figure 4.7), these would probably require calibration to the specific city. More accurate routing could be carried out given the availability of road-level data. In such cases we would add additional impedance to specific roads, giving more useful routing results than the current methodology which considers all roads of the same type to be equivalent.

One use-case of such granular data would be to identify roads that serve schools. The Department of Transport notes that the number of school children being driven to school has trebled over the past 40 years (ibid.), and so having cycling infrastructure serving schools is key to achieving the government target of getting more children to cycle. This would not be difficult, as over 75% of children in the UK live within a 15 minute cycle from their school (DfT 2020a). Goodman et al. (2019) show that if dutch levels of cycling were achieved in the UK, the % of children cycling to school could increase from 1.8% to 41%.

In their typology of cyclists, Dill and McNeil (2013) found that a majority of people who say they would never cycle had never cycled to school, whereas confident cyclists were those most likely to have cycled to school. Getting people to cycle from a young age is therefore key to achieving societal change in commuting habits.

5.3 Community Detection

The main purpose of community detection is to evaluate the distribution of investments resulting from the network growth algorithms, so it is necessary for the MSOAs in each community to be spatially connected, and for there to be well-defined boundaries between the communities. Cyclists in the center of a city may commute in any direction, but those on the edges will most likely commute inwards. Given that cyclists on average commute shorter distances than motorized vehicles, it makes sense for well defined communities to exist (as can be seen in Figures 4.8 & 4.9). However, for cities with irregular shapes, this may not necessarily hold up.

In the case of Manchester the communities are indeed spatially explicit. When the communities are mapped onto the road segments, we see that they all have well defined boundaries. The small size of Community 4 is not ideal for our purposes, and so it may make more sense to use the Louvian algorithm for exploratory purposes and then proceed with a graph partitioning algorithm that allows you to explicitly set the number of partitions (as done by Bao et al. (2017))

The results of the community detection are also only as accurate as the potential cycling demand calculations. The partitioning of the city can therefore be seen as further evidence of the potential demand calculations providing reasonable results.

5.4 Algorithms

All three algorithms show that the increase in person-km satisfied is logarithmic as more investment is added. This is an important observation as it shows that small initial investments at strategic locations go a long way in satisfying cycling demand.

Table 5.1: Comparing Algorithm Results

	Investment Length (km)	**Person-km Satisfied (%)**		
		All	**Community**	
			Lowest	*Highest*
Algorithm 1	100	60%	45% (Community 2)	70% (Community 1)
	200	82%	70% (Community 4)	87.5% (Community 1)
Algorithm 2	100	68%	50% (Community 2)	76% (Community 1)
	200	85%	80% (Community 2)	90% (Community 1)
Algorithm 3	100	67.5%	58% (Community 1)	100% (Community 4)
	200	83%	77% (Community 1)	100% (Community 4)

Looking at the person-km satisfied at different stages, we see that the differences between the algorithms are small (Table 5.1). We expect Algorithm 2 to perform better than Algorithm 1, given that it allows segregated infrastructure to be added to any road segment neighboring existing cycling infrastructure. However, it appears that this advantage is only noticeable at the beginning and by the time 200km of investment is added it is almost indistinguishable. The same phenomenon is apparent when comparing Algorithm 2 to Algorithm 3. In theory, the latter is more restricting since one road segment from each community should be selected at each iteration. The results however do not reflect this restriction, with the two algorithms having almost the same % of person-km satisfied at the 100 and 200km mark.

The community-level results show us how effective our algorithms are at distributing the benefits over the study area. Here we see that Algorithm 3 outperforms the other two algorithms. At an investment threshold of 100km, the poorest performing community in Algorithm 3 has 58% of the person-km within it satisfied, compared to 45% and 50% for Algorithms 1 and 2 respectively. This is not surprising since Algorithm 3 grows the network by choosing one road segment from every community at each step (see Section 3.4.3). Given the small size of community 4, a fair distribution of investment would result in 100% of person-km in it to be satisfied after 100km of investment. This is only the case with Algorithm 3, which is further verification of its effectiveness in that regard.

Looking at the length of investment on each road type resulting from Algorithm 3, we see that most of the investment is on residential and tertiary roads (Figure 4.13b). This is true for both the total investment and for the initial 100km. It is important to note that investment in different road types will vary in both nature and cost. The latest report by the Department of Transport states that closure of residential roads to through traffic will be studied as part of the creation of LTNs to facilitate active travel (DfT 2020b). If the closure of such roads aligns with ideal cycling routes, then there would be no need for investing in segregated cycling infrastructure on them. It is therefore evident that convenient cycling routes need to be studied when deciding on the creation of LTNs; if this is done properly then a significant sum of money that would have been spent on segregated cycling infrastructure can be saved.

We also look at how the algorithms create a connected network. Figure 4.14 shows us that in all three of the studied cities, existing cycling infrastructure is made up of many disconnected road segments that do not form part of a whole. Growing the network using Algorithms 2 and 3 results in a gradual reduction in the number of disconnected components, proving that the algorithms fulfill the connectivity requirement of efficient cycling networks. This reduction is exponential, with the number of components being reduced significantly after relatively small investments. The work builds on that of Natera et al. (2019) who showed that small investments can lead to large connectivity gains in a network if placed at strategic locations. In their case, they explicitly solved for connectivity, which meant that there was no redundancy and that the resulting network was not necessarily direct. Here we solve for directness (by first routing potential cycling demand on weighted shortest paths, and then selecting road segments with the highest aggregate flow) but add connectivity constraints, and in doing so we are still able to achieve good connectivity results.

Most importantly, the results show that using community level analysis to build cycling infrastructure using an egalitarian approach does not entail significant trade-offs to city-wide results when evaluated against a utilitarian approach. This is evident in the almost indistinguishable city-wide results of Algorithm 2 and 3 (shown in Figures 4.14, 4.15, and Table 5.1). Coupling that with the fact that the egalitarian approach provides better distribution of infrastructure between communities, it appears to be a promising approach to adopt.

6. Conclusion

The research set out to provide a methodology for designing a cycling network that respects both the needs of the users and the ambitions of the government. The methodology can be used as a starting point for introducing a direct, connected, and low-stress network for any city by leveraging available open data. The only data necessary for reproducing it is the road network (from OSM), its topography (from satellite imagery), and commuter data (from the census). It should be noted that the recommendations are only as good as the granularity of the available data.

A core part of the methodology is determining where there is high potential for cycling, and using this as a basis for recommendations on road space reallocation in order to unlock potential cycling demand. A routing engine (**dodgr**) was used to study the effects of limiting access on different road types, and it was found that reducing cyclist flow on roads with high through-traffic resulted in acceptable increases in commuting distances. As a result, a hierarchy of road preference was used to route potential cycling demand. Algorithms were developed to determine where investments in cycling infrastructure should be prioritized. These algorithms, based on connectivity and egalitarian principles of resource distribution, ensured that whatever the level of investment, the resulting cycling network would improve connectivity of existing cycling infrastructure.

The research hopes to aid in the current shift towards active transport that is being promoted by the UK government. The success of such a shift is contingent on the appeal of the cycling network for prospective cyclists, a fact that was used as the basis for the methodology outlined here.

Bibliography

Akbarzadeh, Meisam, Syed Sina Mohri, and Ehsan Yazdian. 2018. "Designing bike networks using the concept of network clusters." *Applied network science* 3 (1): 12.

Aldred, Rachel, Joseph Croft, and Anna Goodman. 2019. "Impacts of an active travel intervention with a cycling focus in a suburban context: One-year findings from an evaluation of London's in-progress mini-Hollands programme." *Transportation research part A: policy and practice* 123:147–169.

Bao, Jie, Tianfu He, Sijie Ruan, Yanhua Li, and Yu Zheng. 2017. "Planning bike lanes based on sharing-bikes' trajectories." In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining,* 1377–1386.

Barabási, Albert-László, et al. 2016. *Network science.* Cambridge university press.

Bloch, Francis, Matthew O Jackson, and Pietro Tebaldi. 2019. "Centrality measures in networks." *Available at SSRN 2749124.*

Blondel, Vincent D, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. "Fast unfolding of communities in large networks." *Journal of statistical mechanics: theory and experiment* 2008 (10): P10008.

Boisjoly, Geneviève, Ugo Lachapelle, and Ahmed El-Geneidy. 2020. "Bicycle network performance: Assessing the directness of bicycle facilities through connectivity measures, a Montreal, Canada case study." *International Journal of Sustainable Transportation* 14 (8): 620–634.

Broach, Joseph, John Gliebe, and Jennifer Dill. 2011. "Bicycle route choice model developed using revealed preference GPS data." In *90th annual meeting of the transportation research board, Washington, DC.*

Buehler, Ralph, and Jennifer Dill. 2016. "Bikeway networks: A review of effects on cycling." *Transport Reviews* 36 (1): 9–27.

Caulfield, Brian, Elaine Brick, and Orla Thérèse McCarthy. 2012. "Determining bicycle infrastructure preferences–A case study of Dublin." *Transportation research part D: transport and environment* 17 (5): 413–417.

Celis-Morales, Carlos A, Donald M Lyall, Paul Welsh, Jana Anderson, Lewis Steell, Yibing Guo, Reno Maldonado, Daniel F Mackay, Jill P Pell, Naveed Sattar, et al. 2017. "Association between active commuting and incident cardiovascular disease, cancer, and mortality: prospective cohort study." *bmj* 357:j1456.

Crane, Melanie, Chris Rissel, Chris Standen, Adrian Ellison, Richard Ellison, Li Ming Wen, and Stephen Greaves. 2017. "Longitudinal evaluation of travel and health outcomes in relation to new bicycle infrastructure, Sydney, Australia." *Journal of Transport & Health* 6:386–395.

DfT. 2012. *Guidance on road classification and the primary route network.*

———. 2017. *Cycling and walking investment strategy.*

———. 2020a. *Cycling Infrastructure Design.* Technical report.

———. 2020b. *Gear Change: A Bold Vision for Cycling and Walking.* Technical report.

Dill, Jennifer, and Nathan McNeil. 2013. "Four types of cyclists? Examination of typology for better understanding of bicycling behavior and potential." *Transportation Research Record* 2387 (1): 129–138.

Duthie, Jennifer, and Avinash Unnikrishnan. 2014. "Optimization framework for bicycle network design." *Journal of Transportation engineering* 140 (7): 04014028.

Goodman, Anna, Ilan Fridman Rojas, James Woodcock, Rachel Aldred, Nikolai Berkoff, Malcolm Morgan, Ali Abbas, and Robin Lovelace. 2019. "Scenarios of cycling to school in England, and associated health and carbon impacts: Application of the 'Propensity to Cycle Tool'." *Journal of Transport & Health* 12:263–278.

Goodman, Anna, Shannon Sahlqvist, David Ogilvie, and iConnect Consortium. 2014. "New walking and cycling routes and increased physical activity: one-and 2-year findings from the UK iConnect Study." *American journal of public health* 104 (9): e38–e46.

Goodwin, Phil B. 1996. "Empirical evidence on induced traffic." *Transportation* 23 (1): 35–54.

Iacono, Michael, Kevin J Krizek, and Ahmed El-Geneidy. 2010. "Measuring non-motorized accessibility: issues, alternatives, and execution." *Journal of Transport Geography* 18 (1): 133–140.

Jafino, Bramka Arga, Jan Kwakkel, and Alexander Verbraeck. 2020. "Transport network criticality metrics: a comparative analysis and a guideline for selection." *Transport Reviews* 40 (2): 241–264.

Jarrett, James, James Woodcock, Ulla K Griffiths, Zaid Chalabi, Phil Edwards, Ian Roberts, and Andy Haines. 2012. "Effect of increasing active travel in urban England and Wales on costs to the National Health Service." *The Lancet* 379 (9832): 2198–2205.

Kohl 3rd, Harold W, Cora Lynn Craig, Estelle Victoria Lambert, Shigeru Inoue, Jasem Ramadan Alkandari, Grit Leetongin, Sonja Kahlmeier, Lancet Physical Activity Series Working Group, et al. 2012. "The pandemic of physical inactivity: global action for public health." *The lancet* 380 (9838): 294–305.

Lovelace, Robin, Anna Goodman, Rachel Aldred, Nikolai Berkoff, Ali Abbas, and James Woodcock. 2017. "The Propensity to Cycle Tool: An open source online system for sustainable transport planning." *Journal of transport and land use* 10 (1): 505–528.

Lucas, Karen, Bert Van Wee, and Kees Maat. 2016. "A method to evaluate equitable accessibility: combining ethical theories and accessibility-based approaches." *Transportation* 43 (3): 473–490.

Marqués, R, V Hernández-Herrador, M Calvo-Salazar, and JA García-Cebrián. 2015. "How infrastructure can promote cycling in cities: Lessons from Seville." *Research in Transportation Economics* 53:31–44.

Mauttone, Antonio, Gonzalo Mercadante, María Rabaza, and Fernanda Toledo. 2017. "Bicycle network design: model and solution algorithm." *Transportation research procedia* 27:969–976.

Mesbah, Mahmoud, Russell Thompson, and Sara Moridpour. 2012. "Bilevel optimization approach to design of network of bike lanes." *Transportation research record* 2284 (1): 21–28.

Nahmias-Biran, Bat-hen, Karel Martens, and Yoram Shiftan. 2017. "Integrating equity in transportation project assessment: a philosophical exploration and its practical implications." *Transport reviews* 37 (2): 192–210.

Natera, Luis, Federico Battiston, Gerardo Iñiguez, and Michael Szell. 2019. "Data-driven strategies for optimal bicycle network growth." *arXiv preprint arXiv:1907.07080.*

Newman, Mark EJ. 2006. "Modularity and community structure in networks." *Proceedings of the national academy of sciences* 103 (23): 8577–8582.

Olmos, Luis E, Maria Sol Tadeo, Dimitris Vlachogiannis, Fahad Alhasoun, Xavier Espinet Alegre, Catalina Ochoa, Felipe Targa, and Marta C González. 2020. "A data science framework for planning the growth of bicycle infrastructures." *Transportation Research Part C: Emerging Technologies* 115:102640.

ONS. 2011. *2011 Census: Special Workplace Statistics (United Kingdom).* Data retrieved from UK Data Service Census Support, `https://wicid.ukdataservice.ac.uk/`.

———. 2018. "Population estimates for the UK, England and Wales, Scotland and Northern Ireland: mid-2017." *Hampshire: Office for National Statistics.*

Padgham, Mark. 2019. "Dodgr: An r package for network flow aggregation." *Transport Findings. Network Design Lab.*

Palominos, Nicolas, and Duncan A Smith. 2019. *Quantifying and Mapping Streetspace: a Geocomputational Method for the Citywide Analysis of Pedestrian and Vehicular Streetspace.* Technical report. CASA Working Paper 212.

———. 2020. "Identifying and Characterising Active Travel Corridors for London in Response to Covid-19 Using Shortest Path and Streetspace Analysis."

Parkin, John, Mark Wardman, and Matthew Page. 2008. "Estimation of the determinants of bicycle mode share for the journey to work using census data." *Transportation* 35 (1): 93–109.

Patterson, Richard, Jenna Panter, Eszter P Vamos, Steven Cummins, Christopher Millett, and Anthony A Laverty. 2020. "Associations between commute mode and cardiovascular disease, cancer, and all-cause mortality, and cancer incidence, using linked Census data over 25 years in England and Wales: a cohort study." *The Lancet Planetary Health* 4 (5): e186–e194.

Pereira, Rafael HM, Tim Schwanen, and David Banister. 2017. "Distributive justice and equity in transportation." *Transport reviews* 37 (2): 170–191.

Pucher, John, and Ralph Buehler. 2008. "Making cycling irresistible: lessons from the Netherlands, Denmark and Germany." *Transport reviews* 28 (4): 495–528.

Pucher, John, Jennifer Dill, and Susan Handy. 2010. "Infrastructure, programs, and policies to increase bicycling: an international review." *Preventive medicine* 50:S106–S125.

Schoner, Jessica E, and David M Levinson. 2014. "The missing link: Bicycle infrastructure networks and ridership in 74 US cities." *Transportation* 41 (6): 1187–1204.

Stinson, Monique A, and Chandra R Bhat. 2003. "Commuter bicyclist route choice: Analysis using a stated preference survey." *Transportation Research Record* 1828 (1): 107–115.

TfL. 2018. *Cycling action plan.* Technical report.

UN. 2015. "Paris agreement." In *Report of the Conference of the Parties to the United Nations Framework Convention on Climate Change (21st Session, 2015: Paris). Retrived December,* 4:2017. HeinOnline.

Wardman, Mark, Miles Tight, and Matthew Page. 2007. "Factors influencing the propensity to cycle to work." *Transportation Research Part A: Policy and Practice* 41 (4): 339–350.

Winters, Meghan, Gavin Davidson, Diana Kao, and Kay Teschke. 2011. "Motivators and deterrents of bicycling: comparing influences on decisions to ride." *Transportation* 38 (1): 153–168.

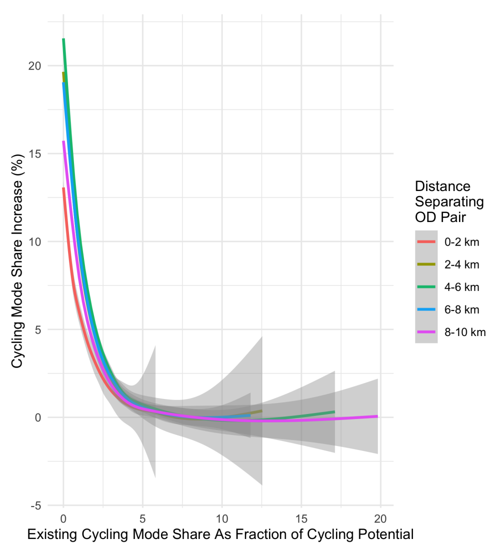
Winters, Meghan, Kay Teschke, Michael Grant, Eleanor M Setton, and Michael Brauer. 2010. "How far out of the way will we travel? Built environment influences on route selection for bicycle and car travel." *Transportation Research Record* 2190 (1): 1–10.

A. Results for Other Cities

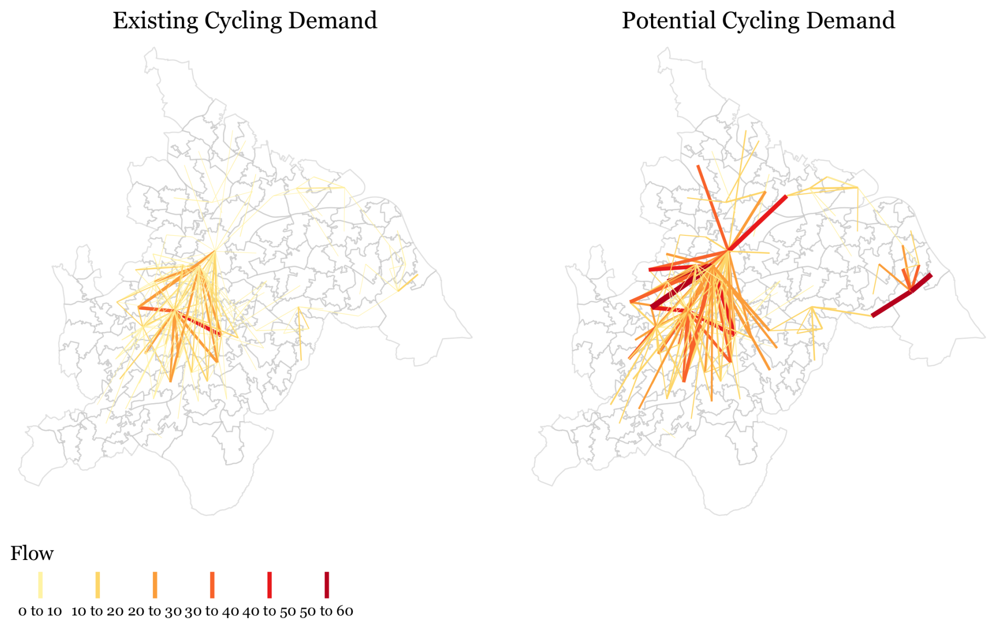
The research was mainly focused on Manchester, but the analysis was carried out for three different UK cities. Below are the results for Birmingham and Nottingham. The results are reproducible for all UK cities available in this ONS dataset. The code for reproducing the results can be found at here.

A.1 Birmingham

Potential Demand

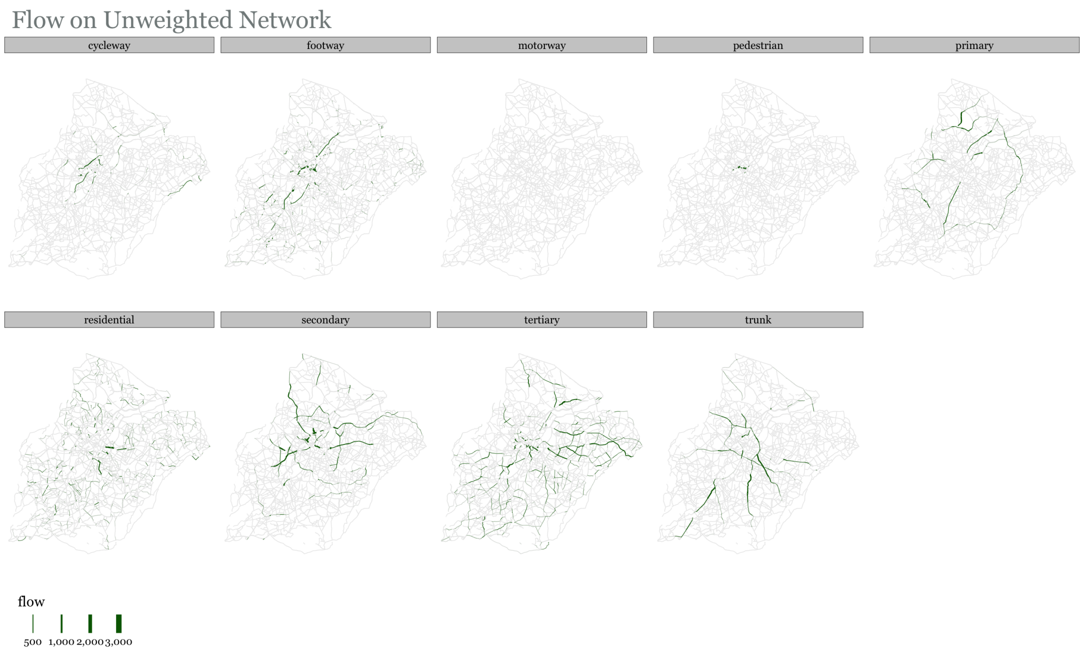


Distribution of Cycling Demand Relative to OD Pair Performance

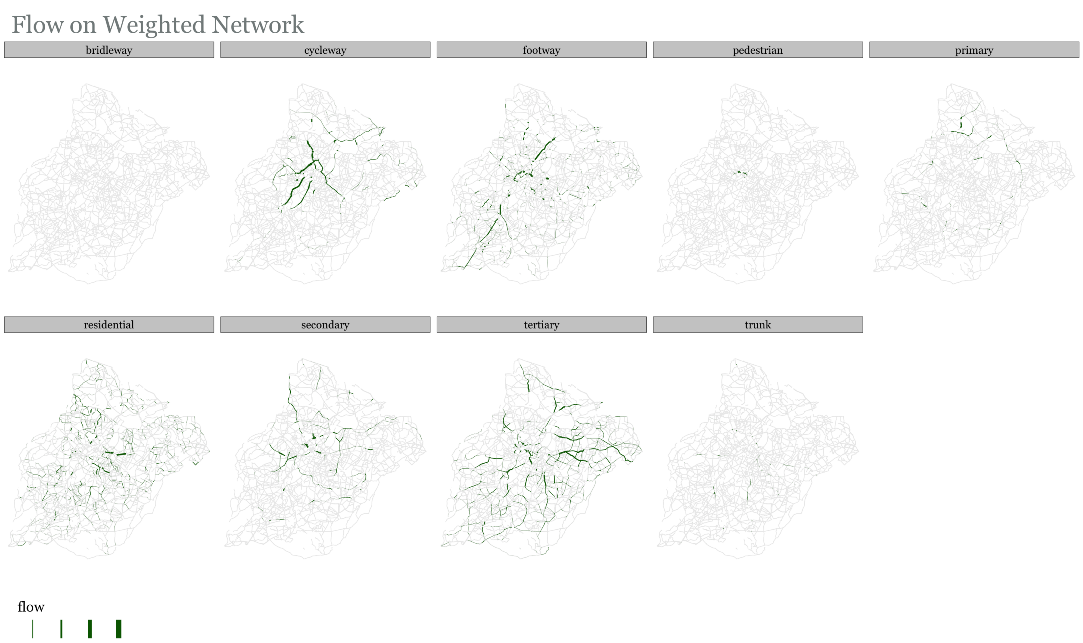


Current and Potential Cycling Flow (Birmingham)

Routing Cycling Flows

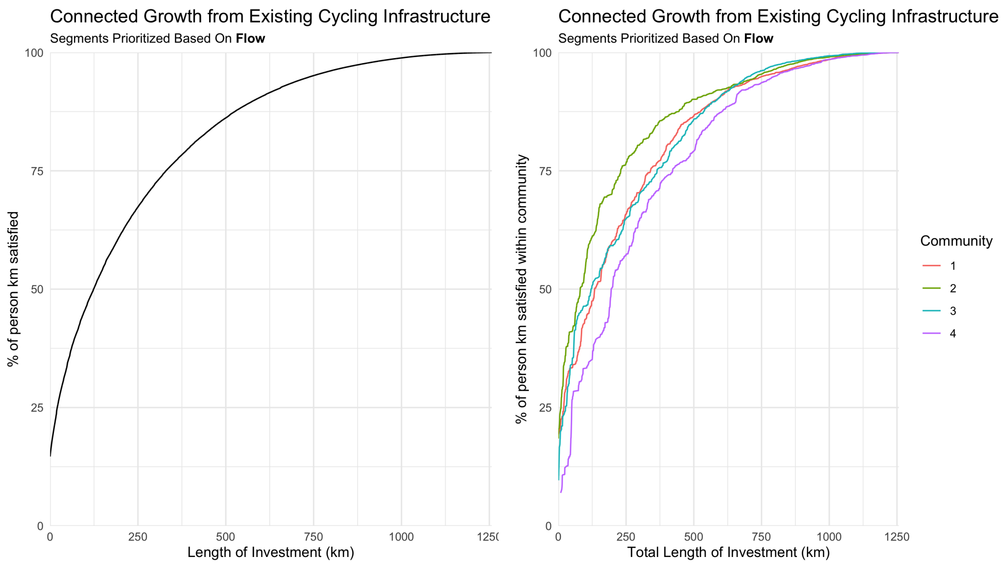


Flow Results Based on **Unweighted** Shortest Paths (Birmingham)



Flow Results Based on **Weighted** Shortest Paths (Birmingham)

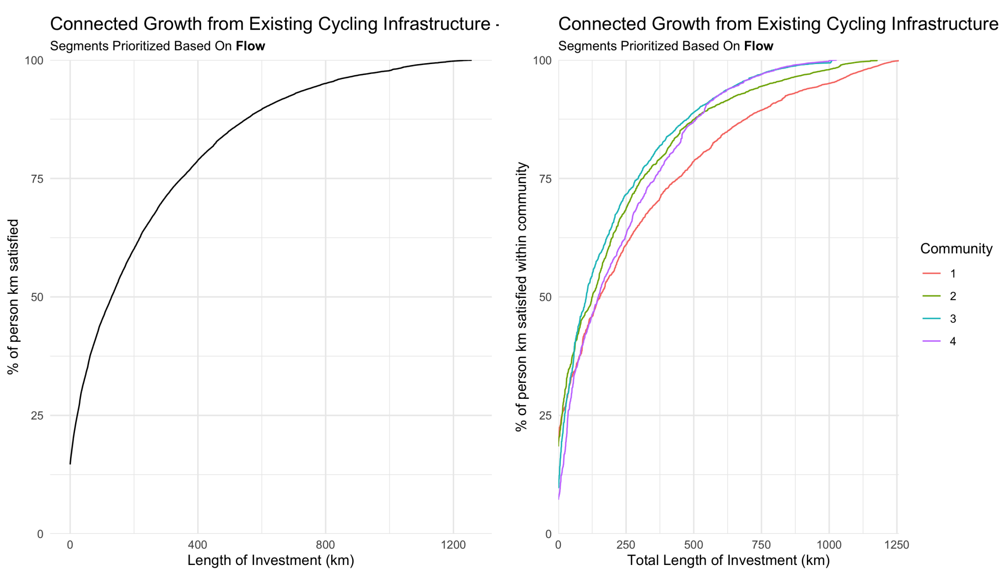
Algorithm 2 (Utilitarian)



(a) % of Person-Km Satisfied (b) % of Person-Km Satisfied in Each Community

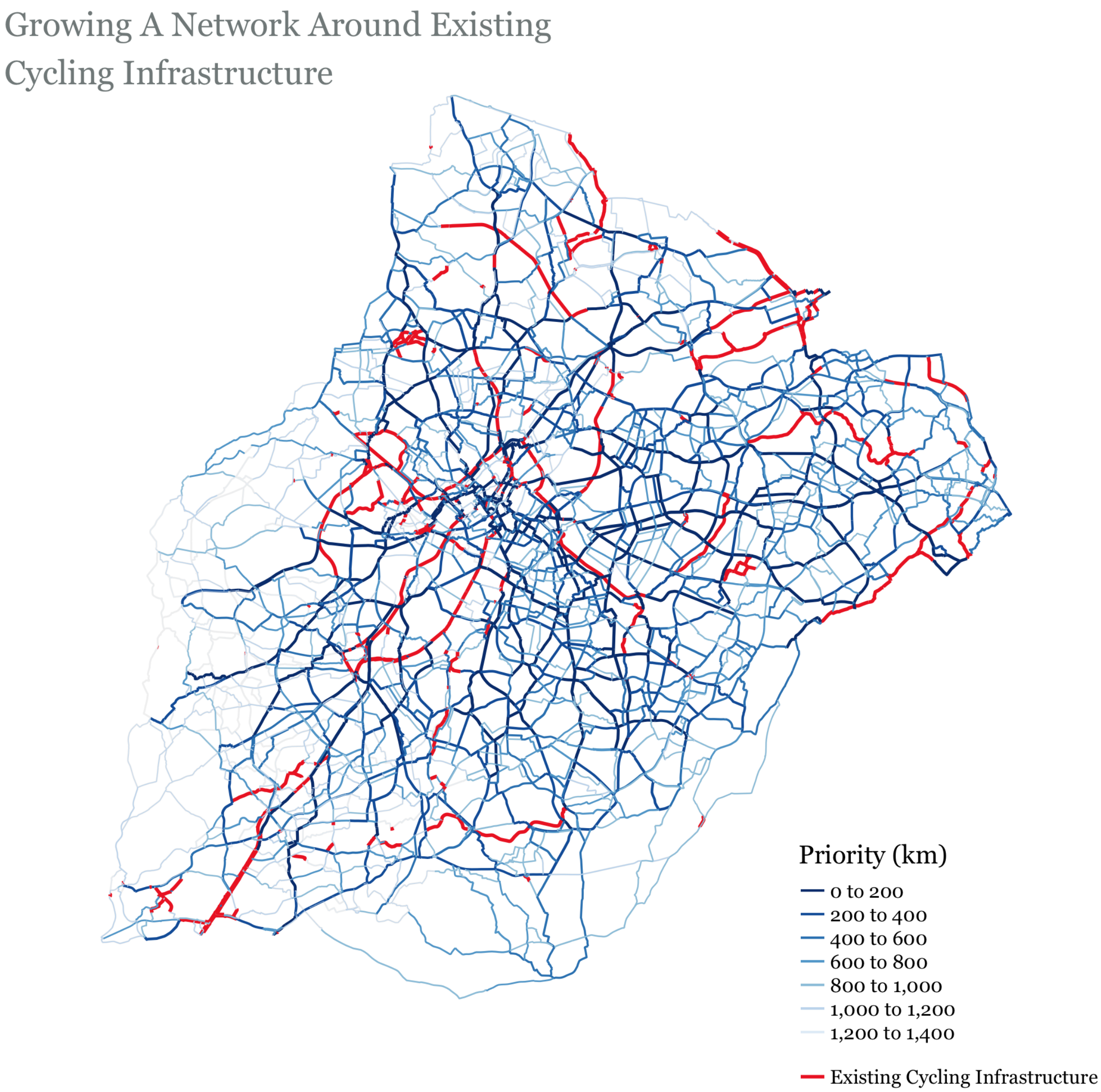
Effectiveness of Alg. 2 on Overall and Community-Wide Person-Km Satisfied (Birmingham)

Algorithm 3 (Egalitarian)



(a) % of Person-Km Satisfied (b) % of Person-Km Satisfied in Each Community

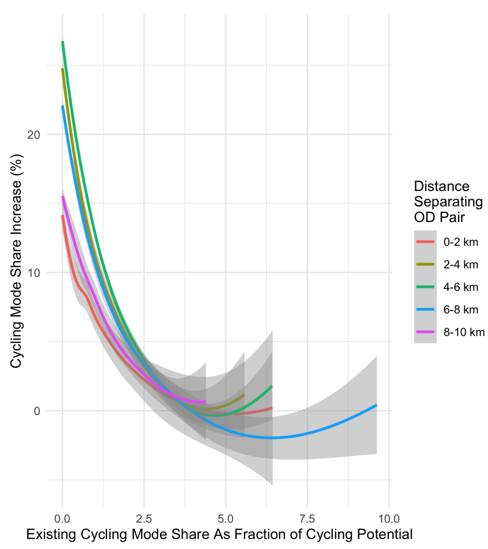
Effectiveness of Alg. 3 on Overall and Community-Wide Person-Km Satisfied (Birmingham)



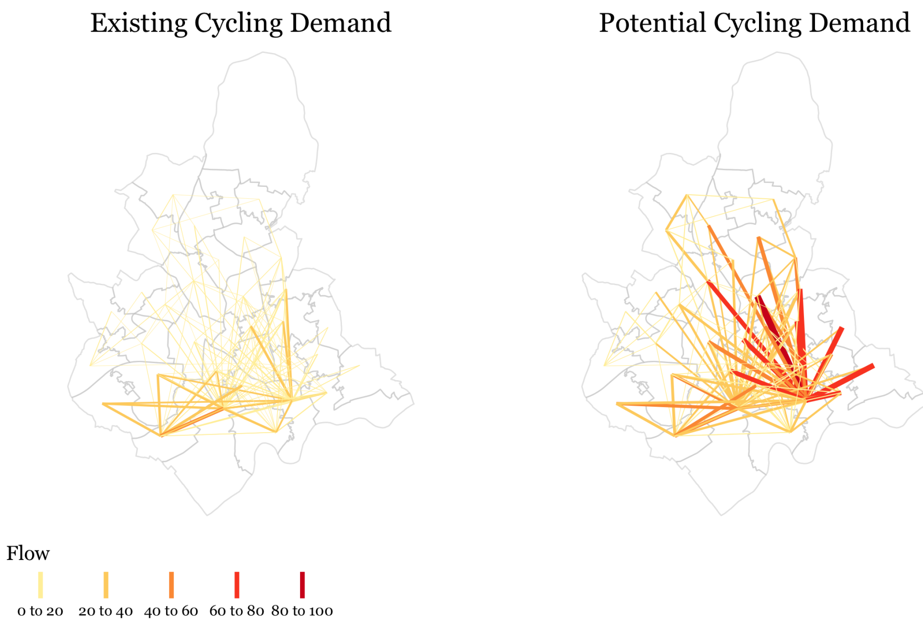
Results of Alg. 3 (Birmingham)

A.2 Nottingham

Potential Demand

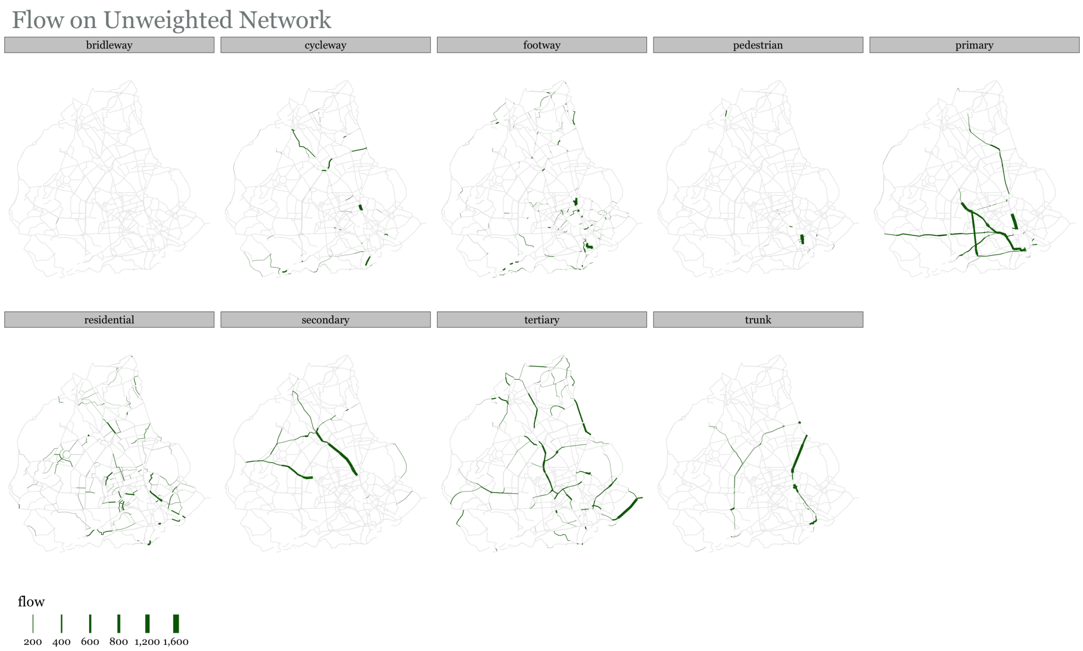


Distribution of Cycling Demand Relative to OD Pair Performance



Current and Potential Cycling Flow (Nottingham)

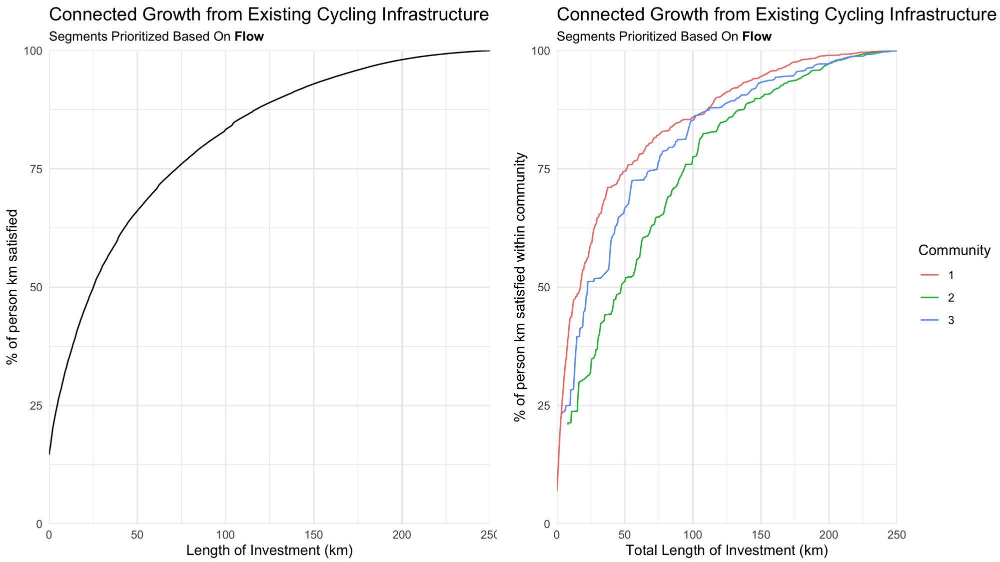
Routing Cycling Flows



Flow Results Based on **Unweighted** Shortest Paths (Nottingham)

Flow Results Based on **Weighted** Shortest Paths (Nottingham)

Algorithm 2 (Utilitarian)

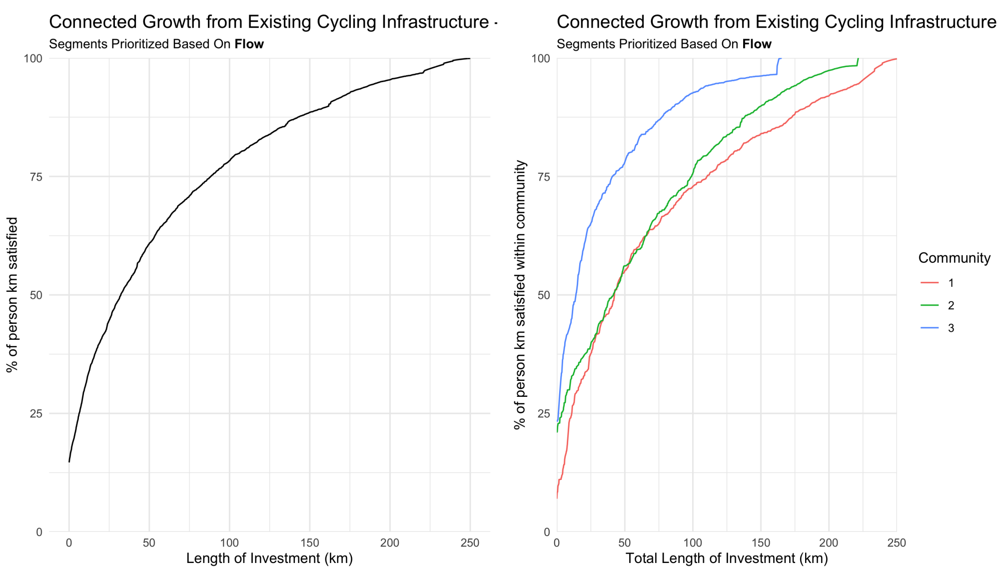


(a) % of Person-Km Satisfied

(b) % of Person-Km Satisfied in Each Community

Effectiveness of Alg. 2 on Overall and Community-Wide Person-Km Satisfied (Nottingham)

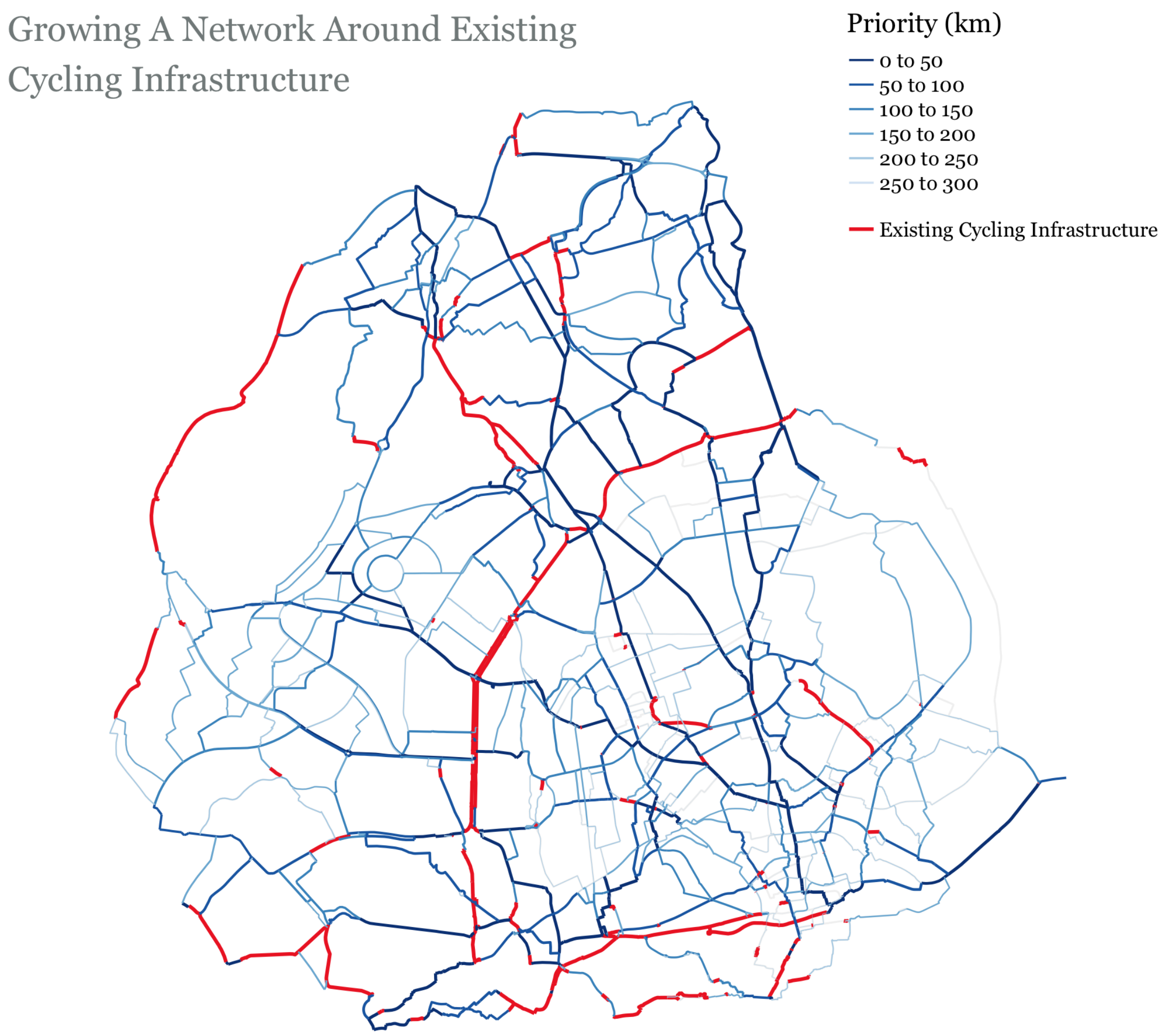
Algorithm 3 (Egalitarian)



(a) % of Person-Km Satisfied

(b) % of Person-Km Satisfied in Each Community

Effectiveness of Alg. 3 on Overall and Community-Wide Person-Km Satisfied (Nottingham)



Results of Alg. 3 (Nottingham)

B. Research Log

Date	Task	Documents Prepared	Challenges / Topics Discussed	Solutions
6- Apr	Meeting with supervisor	Readings on Bus Network Optimization Literature on the Transit Route Network Design Problem	Potential study areas Possibility of focusing on cycling networks instead	
14-Apr	Optimization resources		Can't figure out where to start. Is it linear programming or is it more complex optimization (evolutionary algorithms etc). I have no idea where to start if it is the latter	Read literature (this did not help with implementation)

20-Apr	R vs Python		I am more familiar with python for linear programming (gurobi / cvxpy) but more comfortable with R for spatial analysis	Postponed decision
1-May	Meeting with supervisor	Research Question Datasets on TfL cycling network Literature review of bicycle network Optimization	Look at total flow not just cycling flow (total flow is a proxy for latent demand) Mix connected components with optimization Use Santander cycle data to analyze change in flow since 2011 and compare that to cycling infrastructure investment	I will probably not do linear programming/ optimization

20-May	Finished Draft Literature Review			Better informed about methodologies used. I will stick to network analysis: Route the cycling demand onto the road network and then prioritize road segments for investment based on aggregated cycling demand on them
22-May	Meeting with supervisor	Literature Review Draft Methodology Questions on routing	Literature Review: Mostly good Potential demand: Consider using Santander data to see what mode shift you should model on	
23-May	Routing		Found dodgr package: Could be used to aggregate flow on road segments. This is necessary for next step of analysis	Decided to stick to R.

28-May	Computational Power		Analysis cannot be done for London due to size of road network	Checked possibility of using Myriad.
5-Jun	Potential Cycling Demand	Reproducible github issue	Documented here https://github.com/Hussein-Mahfouz/Bicycle-Network-Optimization/issues	Check github. I also emailed Dr. Maarten Vanhoof and he suggested accounting for existing cycling mode share. OD pairs that have high cycling mode share should not be treated the same as those with low cycling mode share
3-Jun	Network Growth Functions		Functions are very slow. This is due to memory allocation	Rewrite functions with memory allocation in mind (do not add rows to a data frame at every iteration)

12-Jun	Meeting with supervisor	Github issues on my methodological problems	A 'distance decay' function to assign potential cycling demand to MSOA pairs based on distance and slope How to prioritize road segments for cycling infrastructure based on [a] total flow passing over them and [b] current road type/condition Myriad	Dr. Elsa provided a walkthrough of running scripts from the terminal (for Myriad. I did not end up using Myriad in the end as I found it too complicated - it would only have been necessary to carry out the analysis for London)
18-Jun	Routing		I cannot route cycling flow on unweighted shortest paths, as [a] cyclists cannot go on motorways, [b] this does not account for cyclist route choice preference, and [c] this will not route cyclists on existing cycleways	Create dodgr weighting profiles and use them for routing. Weighting profile creates a hierarchy of road preference that avoids stressful streets and routes on cycleways

30-Jun	Finished Draft Analysis	All Figures from analysis sent to supervisor for feedback		
15-Jul	Potential Cycling Demand		Account for existing mode share when assigning potential cycling demand to MSOA pairs	Negative exponential to scale down potential cycling demand of MSOA pair based on existing cycling mode share
31-Jul	Finished Draft Paper			
7-Aug	Meeting with supervisor	Draft Paper	Positive feedback. I need to add topic sentences so that the reader has context. I also need to explain a few topics better	