University College London
Bartlett Centre for Advanced Spatial Analysis

Dissertation

Beyond the Commute: Analysing Urban Cycling Environments for Leisure Cycling, and what this might look like for Hackney, London

By Richy Buttrick
January 2023

This dissertation is submitted in part requirement for the MSc Smart Cities and Urban Analytics at the Centre for Advanced Spatial Analysis, Bartlett Faculty of the Built Environment, University College London.

Submitted: 31 January 2023
Supervisor: Dr. Fulvio D. Lopane
Module ID: CASA0010
Student ID: 21202855
Word Count: 8,972
Pages 8 to 33 inclusive

Abstract

An increasingly urbanised world has led many to look at the quality of life experienced within cities, and how we might like cities to change. Residents' day to day travelling experiences are marked by noise, fumes, danger and a decreased access to our streets, often coinciding with the presence of cars. Contrasting this is a vision of what could be – streets that are safe, spacious, and communal. Urban environments that are rich to experience and pleasurable to travel around. Air that is clean, and a knowledge that fossil fuel consumption is low. A move from passive to active travel is one of the ways that many academics and government bodies think this transition could happen.

One way to increase active travel is to improve cycling infrastructure. Network analysis performed to aid this often focusses on commuting, due to data availability. This study asks what it might be like to use network analysis for those cycling for leisure. It was found in the literature that sensory experience was key to motivating these cyclists.

OSM data filtered for the cycling network in Hackney, London, is used and a range of shortest paths for origin-destination pairs are found. These paths are scored for the sensory experience they involve (referred to as discomfort and derived from the OSM edge type) and the length of the trip. The data is plotted on scatter graphs and a pareto front is found to characterise the network between origin-destination pairs at different geographic scales.

A wide range of lengths at low discomfort (on the pareto front) is found to be desirable between origin-destination pairs. However, using network data less simplified than that used in the literature meant the results needed to be examined on a case by case basis, rather than just checking for flatness of the pareto front. The study could be used to understand differing rates of cycling across London and to analyse the impact of new Low Traffic Neighbourhoods.

Declaration of Authorship

I, Richy Buttrick, hereby declare that this dissertation is all my original work and all sources have been acknowledged. The dissertation is 8,972 words in length.

Signed: R Buttrick ……………………..
Date: 31st January 2021

Table of Contents

List of Figures

List of Terms, Abbreviations, and Acronyms

Biobjective – activity involving two objectives
Passive Transportation – Transport that uses non-motorised means; e.g. walking, cycling
Active Transportation - Transport that uses motorised means; e.g. cars, trains
Walkable - how hospitable the city environment is to walking and cycling
PM 2.5 – air pollutant particulate matter with a width of two and a half microns or less
OSM – Open Street Map, an open data resource of infrastructure maps
MSOA – Middle Layer Super Output Area, a governmentally assigned geographic area
LTN – Low Traffic Neighbourhood, where motor vehicle traffic is intentionally lowered
o-d – origin – destination

Acknowledgements

Firstly, I would like to express the strongest gratitude to my supervisor, Dr. Fulvio Lopane, who has encouraged me many times during this module, and practiced patience when I have encountered problems. He has been generous with his time and always engaged.

Other academics that have given their time generously are Valentina Paz, Ivann Schlosser, Professor Mike Batty, and Professor Elsa Arcaute.

I would also like to thank the Centre For Advanced Spatial Analysis at UCL for their expertise during my Masters.

And lastly, I would like to thank my parents for their limitless support and generosity.

Chapter 1: Introduction

1.1 Context and Motivation

Despite the normalisation of the supremacy of cars within cities, a growing number of people have begun to question their domination. It has begun to seem odd that they occupy so much space whilst being so inefficient and whilst many do not have access to one (Gössling 2020).

Cycling can be seen as an enabler of freedom within the city, and a way to reconnect with the urban environment. This is important as our urban spaces help to shape society and vice versa, in a social-spaitial dialectic (Soja 1980). Transport becomes experiential rather than just functional by enabling connection with the urban environment (Siddall 1987). This is the opposite of the car, which isolates us from the city and from interaction with others.

Beyond this our dependence on passive transport has led to a host of more direct issues. Our cities are becoming increasingly choked with pollution, and citizens are lumbered with health issues such as obesity and asthma. A focus on pollution's health hazards leads onto larger problems that we face – how do cities respond to the climate crisis? A transition is needed towards transport that emits less and reduces our dependence on fossil fuels.

There is a wealth of literature that looks at how to encourage cycling in urban spaces. These cover a wide range of methods, for example, targeting investment so that it is equally distributed across the cycling network (Mahfouz *et al.* 2021), the physical structure of cycleways themselves (Petegem *et al.* 2021), and the provision of publicly accessible bikeshares (Hosford *et al.* 2019). Network analysis can offer excellent insight into the characteristics of infrastructure, however, commuting is often the focus of the analysis because the data is much more readily available. This study responds to the desire for cities to be liveable and explores how we might analyse cities for their capacity for cycling for leisure rather than just commuting.

There is some research that acknowledges cyclists have multiple desires or objectives when cycling. For example, Reggiani *et al.* (2021), which focusses on the length and discomfort of a trip through the lens of a commuter. We look to expand this work into the desires of those cycling for leisure in order to encourage the activity.

1.2 Research Question

This background leads us to ask the questions:

1. How might we analyse urban cycling infrastructure to establish how it serves those cycling for leisure?
2. What does this analysis look like for Hackney, London?

1.3 Scope

This study uses the borough of Hackney in London to perform the suggested quantitative analysis. Hackney is one of 32 boroughs in London and covers 19.06 square kilometres. It is divided into 28 MSOAs, a designated area used in England and Wales. It has a higher rate of commuting by bike compared to many other parts of London (Lovelace *et al.* 2017).

London is an interesting place to look at the prevalence of active travel. Historically the city was the site of a public fight against a proposed sprawling motorway network in the 1960s and 1970s. Three motorway ring-roads were planned to run through the city until public resistance forced a u-turn on the policy (Pharoah and Plowden 1973). Recently the city has been introducing 'Low Traffic Neighbourhoods' (LTNs) in order to decrease motor vehicle traffic and make the city more liveable. There is ongoing political conflict over this (Guardian 2020) and so more research is needed into how this impacts the city. It is our suggestion that this study could be included in this research.

1.4 Ethical Considerations

The study contains very low ethical risk as we did not use any population data so there is no possibility of identifying individuals and we are working with data at MSOA level. All the data is from an open data source and available publicly.

1.5 Report Outline

Chapter 2 will review the existing literature surrounding the topic. This includes a deeper look at the environmental and health impacts of transport culture in cities, and the emphasis towards active travel. Then we will look at the motivations for cycling and how a broader understanding might enable us to look again at our infrastructure. A large part of the literature review is on work done by Reggiani *et al.* (2021) which uses a biobjective analysis to characterise cycling networks, and how this might be developed for leisure cyclists. Lastly, an overview of how London has changed its cycling infrastructure and the aims of local government.

Chapter 3 is a review of the data used for the analysis. Chapter 4 covers the methodology including a provisional selection and cleaning of the data, weighting the edges, and the functions in python used to find paths between origin-destination pairs, and find their discomfort score and length.

Following this in Chapter 5 is the data plotted for different origin destination pairs, that allows for the discussion of the results, limitations and possible further developments laid out in Chapter 6. Lastly, Chapter 7 forms the conclusion.

Chapter 2: Literature Review

2.1 Impacts of current travel culture

2.1.1 Obesity

Over the last thirty years, many countries worldwide have seen rates of obesity double, and in some cases triple or quadruple. An increasing prevalence of childhood obesity will likely mean ballooning obesity rates among adults in the coming years, adding pressure to health care infrastructures. The current state of the USA is seen as a likely template for what many nations will tend towards – its healthcare costs that are directly related to its obesity epidemic are estimated at around US$190 billion per annum, or 21% of healthcare funding. In fact, by 2020, 20% of the world's adult population will be obese (Hruby and Hu 2015).

If we take London for example, 20% of young people do not get enough exercise and between 1997 and 2007 'the percentage of overweight or obese boys aged 2-10 years increased from 24.6% to 29.8 % and for girls, from 22.9% to 28.6%' (Greater London Authority 2011: 13).

This is a complicated issue, with many interlocking causes, but a lack of physical exertion linked to passive transportation is seen as a key environmental factor (Hruby and Hu 2015). In a report by the Greater London Authority, inactivity is cited as a major factor in obesity increase, with active transport e.g. cycling to school, given as an example to reverse this (Greater London Authority 2011). Another study found that there is a statistically significant correlation between how 'walkable' a city is and how overweight its residents were (Congdon 2019).

At the same time, there is a trend for the urbanisation of our planet. The UN has estimated that 4.2 billion people lived in an urban space as of 2019 but that this will increase to 6 billion by 2041. Given this urbanisation, it is important to ask what role cities take in obesity rates and how they might be changed to lower them (Kuddus *et al.* 2020), including passive and active transportation.

2.1.2 Air Pollution

Another global issue affected by urban spaces is air pollution. According to the World Health Organisation, nearly 90% of the world's population currently reside in areas where their air quality limits are surpassed (Liang and Gong 2020). According to Wang et. al (2017) this has a direct impact upon health outcomes, with PM 2.5 particles being linked to shorter life expectancies due to an increased likelihood of respiratory illnesses and supressed immune systems. They also claim that urban spaces are a particular problem, with urbanisation accompanying rapid increases in pollutants (Wang *et al.* 2017). However, Liang and Gong (2020) articulate that the picture is more complicated, with particulate matter levels varying with city size, mixed/uniform land use, density, and poly/monocentricity rather than a more linear correlation between any form of urbanisation and pollution. Yet they do state urban spaces produce around 78% of particulates – a clear statistic to show that the planning and development of our urban spaces need more thought (Liang and Gong 2020).

This is an internationally recognised issue, and a city's passive or active transport is highlighted as a key factor in the literature. The Institute for Public Policy Research's 'London Global Green City' Report (2016b) specifies a modal shift from cars to walking, cycling, and public transport as a way to reduce air pollution levels to a legal level (Institute for Public Policy Research 2016b).

Many impress the urgency of the situation, with a quarter of London school children experiencing pollution levels that break WHO limits (Institute for Public Policy Research 2016b). In 2010 alone, this issue was linked to the loss of 140,743 years from Londoners' lives, and costing £3.7 billion (Institute for Public Policy Research 2016a). The IPPR (2016) cites traffic as the main cause of pollution levels (Institute for Public Policy Research 2016a). In fact, after smoking, it is deemed the most influential factor on health (The Institute for Public Policy Research 2016a).

2.2 How cycling tackles these issues

As previously discussed, the literature supports cycling as an active travel medium to alleviate obesity and air pollution. Following on from this, how might cycling be improved and increased in cities such as London? Interestingly, within the literature cycling is often discussed about in terms of commuting. This is understandable as commuting data is more readily available than data for any other type of trip. But does this focus miss out on wider cycling trends?

A core tradition of transportation modelling is that a cost factor strongly defines models produced. However, Salomon and Mokhrtarian (2001) argue against this, making the case that often travel is desired in itself with no other purpose needed, like, for example, cycling for leisure. They note that there are motivations for traveling other than to reach a destination – 'a sense of speed, motion, control, enjoyment of beauty' (Mokhtarian and Salomon 2001: 695) – which may in fact lead to a traveller covering further distances than the minimum. They do note that it follows that the quality of the travelling experience is a factor in the distance and likeliness to travel. For example congested travel may counteract any desire to spend more time in transit (Mokhtarian and Salomon 2001).

If active travel is a desired outcome for city planning and some travellers make journeys for the journey itself, cycling may be a good transportation mode to support. Guerra and Morris (2015) found that of all typical transport modes, cycling has the largest positive affect upon mood (Morris and Guerra 2015).

Wild and Woodward (2019) agree that cyclists are the most satisfied commuters. Similarly to Salomon and Mokhrtarian (2001), key to their work is the idea that sensory experience by muscular activation or environment are drivers for cycling. If cyclists' experiences are too intense because of excess traffic volume or speed, this will mean sensory overload and decreased motivation. They suggest that focussing on the physical, psychosocial, and social pleasure of cycling, and moving away from the perceived risks, will encourage further take up (Wild and Woodward 2019).

In his book 'Mindful Thoughts for Cyclists: Finding Balance on Two Wheels', Nick Moore (2017), describes how high volume traffic needs a 'minutely focused state' that 'shrink(s)' (Moore 2017: 66) the experience to a 'space just a few inches wide and a million miles long, outside which nothing exists' (Moore 2017: 66). This is the opposite of the sensory experience that we have so far seen may encourage cycling.

Anable and Gatersleben (2005) note that there are diverse reasons for journeys. They claim that what appeals to travellers depends on the purpose of the journey. For commuters, convenience is key and sensory experience less important. For leisure journeys however, a sense of freedom and lack of stress is very important (Anable and Gatersleben 2005).

The 'Cycling Action Plan', put together by the Mayor of London and Transport for London, highlights the fact that there are a range of purposes and types of journeys for cyclists in London. It ascribes at least 10% of all bike journeys to the purpose of meeting friends or family for a social occasion. 13% of trips are for travelling for shopping, and 8% for travelling to school or university. The rest of the trips are made up of commuting, general errands, and cycling for pleasure/exercise (Transport For London 2018). So, a high proportion of trips are being made for reasons other than commuting. This highlights the importance of understanding how London's infrastructure serves these trip-takers, in order to address worsening issues such as obesity and air pollution.

2.3 A Biobjective Network Analysis

Reggiani *et al.* (2021), use network analysis to study the quality of cycling opportunities in urban spaces, in accordance with Anable and Gatersleben's (2005) belief that travellers have different priorities for a journey. They tackle cycling networks with the acknowledgement that there are often more than one type of traveller in a network, each having different desired outcome. A 'biobjective' analysis is established to measure a network for two different objectives that users have – the distance of the journey and the discomfort of the route (with discomfort being defined loosely so that it can be adapted for different situations).

In the paper, a network is split into four different layers, each layer containing the part of the network that has a specific discomfort level (see figure A for an example network split into two layers). Users can travel the network on any of the layers, and switching layers is cost free. Switching layers might entail changing from a cycle lane to a main road (as these have different discomfort levels). A trip taken by a user has a cost score for discomfort and distance, which are the normalised sum of the discomfort and distance of all edges taken during the trip. Total discomfort and distance are normalised by dividing by the Euclidian distance between the origin and destination pair.

The discomfort and length scores of every possible path between two selected nodes in a network is plotted onto a scatter graph, as seen in figure A.

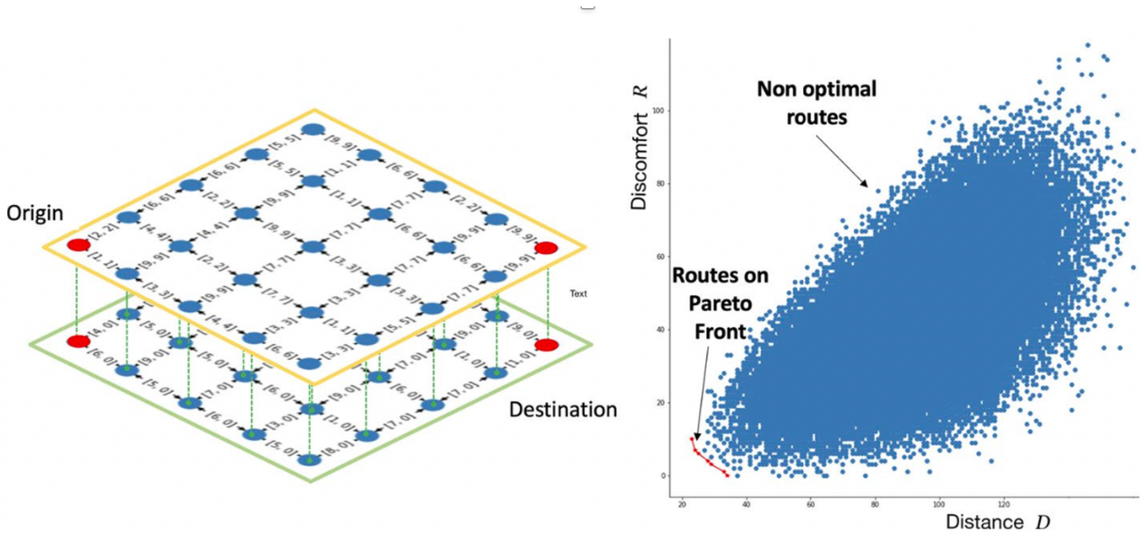


Fig. A. Diagram showing a 5x5 network split into two layers, with an origin and destination, plus the distance and discomfort totals for each route between origin and destination plotted onto a scatter graph. The red line shows the pareto front. Figure from Reggiani *et al.* (2021: 906).

Users are assumed to be rational and so wish to minimise both costs (distance and discomfort) in alignment with traditional network models, meaning that most of the possible trips are not desirable because they have significantly higher levels of discomfort and length than other trips. The most desirable routes are those near the lower ranges of the distance and length scales - Reggiani *et al.* (2021) produces a pareto front here called the bikeability curve, shown in red. This includes the most acceptable routes for different users – those who prefer higher discomfort but shorter routes, those that prefer longer routes with lower discomfort, and all those in between.

A single network wide bikeability curve is produced from the bikeability curves of all origin-destination pairs (see figure B), which 'contains for all possible distance-discomfort trade-off choices, the expected distance, and discomfort of the optimal route for an OD-pair that is sampled from an OD- demand distribution' (Reggiani *et al.* 2021).

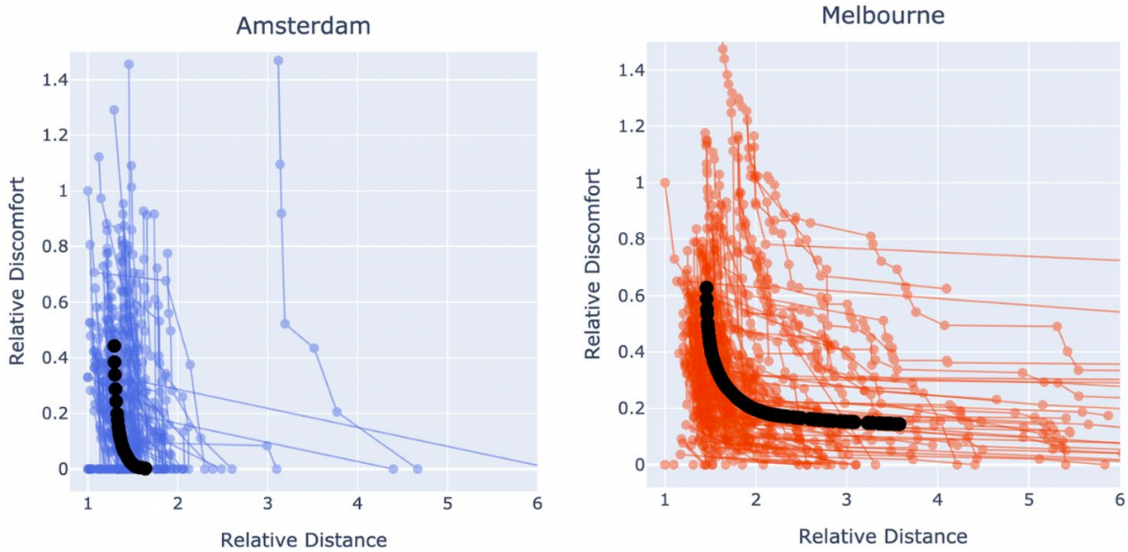


Fig. B. The pareto fronts for Amsterdam and Melbourne combined to one pareto front/bikeability curve to characterise the cities' overall cycling infrastructure. Figure from Reggiani *et al*. (2021).

Reggiani *et al.* (2021) claim that a steeper curve is preferable as it means that the routes contained within it have many different discomfort levels but are all similar distances. This assumes route distance minimisation is desired for all users. However, if, as discussed, there are cyclists that travel for the sake of travel itself and are open to not minimizing the distance travelled, would a flatter curve not be preferable for these users? This would mean a range of paths with different lengths available (those travelling for pleasure then have a range of lengths to choose from), and with low levels of discomfort.

If instead of commuters, we are to focus on those that travel for travel's sake, a pleasurable experience can be taken as an approximation of a low discomfort environment, as discussed earlier using the cited literature by Salomon and Mokhrtarian (2001), Anable and Gatersleben (2005), and Wild and Woodward (2019).

2.4 Discomfort and Network Weighting

Sorton and Walsh (1994) provided an early link between physical environment and the experience of a cyclist. They created ratings between 1 and 5 for different road types based on the stress experienced by a cyclist, taking into account traffic volume, speed and the curb lane width. However, this was based on the reported stress levels of viewers watching video tapes of journeys, and so may be less reliable than reports from cyclists travelling the routes. Landis *et al.* (1997) attempted to build on this by producing a grading system using data from cyclists reporting their stress levels during a journey. Regression analysis was performed to produce a 'Bicycle Level of Service' model of perceived hazard, with a correlation coefficient (r^2) of 0.73, meaning it has a high reliability. The model used a wide range of variables including traffic volume, traffic mix, traffic speed, potential for cross traffic, road surface quality, and width of space allocated for cycling (Landis *et al.* 1997).

People for Bikes' (2018) produced networks with links categorised into high and low stress levels using data from OSM such as traffic speed, number of lanes, parking, cycling facility width, and cycling facility type, e.g. presence of cycle track. However only two levels of categorisation do not accurately reflect the multitude of cyclists' experiences when travelling.

Also using OSM data, Gehrke *et al.* (2020) created weights for a network in order to perform network analysis. The weight was a function of the length of each link, speed of a cyclist on the link (as a function of cyclist type and gradient), cyclist type (either 'Interested but Concerned' or 'Enthused and Confident') and link type, e.g. primary, residential, tertiary. Unfortunately, not all of this data is available to us. The link type was sourced from OSM (Gehrke *et al.* 2020). The paper acknowledges that there are different types of cyclists travelling on the network (an improvement), similarly to Reggiani's paper (Reggiani 2021) which acknowledges cyclists may have different reasons for travelling.

Similarly, Mahfouz *et al.* (2021) drew on the work of Sorton and Walsh (1994) to produce a weighting profile using OSM road types, grading the edges based on whether cycling infrastructure is present and likely stress levels. (Mahfouz *et al.* 2021)

Reggiani *et al.* (2021) suggests four different ways of defining the discomfort weights:

1. Distance travelled on a 'bike friendly' layer. This seems a simplistic approach as roads are not only bike friendly or unfriendly, there is a range of bike 'friendliness' (in Reggiani *et al.*'s (2021) case this is defined as discomfort) for different roads.
2. Number of changes between layers. This signifies the number of times a user changes between, for example, a cycleway and a main road. Changing layers would surely be a deterrent to a cyclist, but this has not been found to be a major source of discomfort in the literature.
3. Risk of accident depending on the time spent on each different layer. This could be linked to stress and so an indicator of the sensory experience of the user, and so a useful analysis.
4. The number of nodes in a path, as intersections are sites of stress. Again, this could be a useful factor for simulating stressful or non-stressful conditions in a journey.

Reggiani *et. al* (2021) settles on assigning four levels of discomfort for the edges:

- bike tracks and lanes – with a weight of 0;
- residential streets, bike tracks and lanes – with a weight of 0.33 x distance on this layer;
- streets where cycling is permitted – with weight of 0.66 x distance on this layer;

- streets used by cars – with weight of 1 x distance on layer.

To find the total discomfort for each path, the distance of each link is multiplied by the weight.

2.5 Limits on trip distances

There is of course a limit on how far cyclists will travel, even when it is for leisure. There has been research into how far travellers are willing to deviate from the shortest path, but this is conflicting. Furth and Mekuria (2016) found that 90% of trips did not exceed the shortest route by 25% (Furth *et al.* 2016). However, Krizek *et al.* (2007) show that combining access to specified cycling paths and cycling during the weekend can result in up to a 93% detour (Krizek *et al.* 2007). This ambiguity within the literature leaves the possibility that changes to a city's cycling infrastructure and the reason the cyclist is travelling can greatly increase the length of journeys.

2.6 How local government has been developing its cycling infrastructure

Interestingly, rates of bike use are not distributed geographically equally over the city. Central and inner London see the most cycling activity with busier bike routes, however, the South West also sees moderately higher rates of cycling (Transport For London 2018). This could be seen as evidence that similar levels of cycling can be extended to all other areas of London.

'The London Plan' (Greater London Authority 2021) is the spatial development strategy for London which is produced and revised roughly every 5 years to coordinate the city's development over the following 20-25 years, and which boroughs must conform to. Within the 2021 plan, cycling is a key tool to transforming the city. It acknowledges that as part of a growing city with twenty first century problems, London needs to move away from cars which it views as space inefficient to more efficient modes of transport such as cycling, specifying that this needs to happen for day to day trips for local amenities as well as commuting to work. 80% of all trips in London should be made by bike, foot, or public transport by 2041. It sets out ten indicators of a healthy street, of which cycling space, controlled noise levels (which motorised traffic is detrimental to), and clean air are three such indicators (Greater London Authority 2021).

Recognition of the need for a shift to cycling has been reflected in local government, with the creation of it's first 'cycling commissioner' in 2016, with a focus on making the activity safer (Mayor of London 2016).

Three boroughs - Waltham Forest, Enfield, and Kingston - have been part of running 'mini holland' programmes in an attempt to emulate the famously successful Dutch cycling culture. This has involved reconstructing junctions so that they are safer for cyclists, reducing the number of motorised vehicles using residential roads, and building segregated cycle lanes that separate cars and bikes (The London Assembly 2020).

Improvements in recent years to cycling infrastructure has occurred at the same time as an increase in cycling and decrease in cyclist deaths – cyclist numbers have tripled between 2000 and 2016 in part due to the 25% of road space that was given over to walking, buses, and cycling by the end of 2016 (Institute for Public Policy Research 2016b).

Chapter 3: Data

Reggiani *et al.* (2021) use a package in Python called OSMnx (Boeing 2017) which allows the extraction of data from open data resource Open Street Map. Their argument for this is that whilst an open data source may contain errors, it is generally more up to date than government records (Reggiani *et al.* 2021). To counter any inaccuracies and improve on computational speeds during analysis, the network for both cities is reduced to a set of nodes that represent local sub-regions, e.g. postal areas. A single edge is established between nodes if they are connected in reality by routes that do not pass through any other sub-regions. The aim of this is the hope that the 'structure and weights of the coarse network represent the structural characteristics of the urban bicycle network graph'. (Reggiani *et al.* 2021: 913).

However, this means the loss of large amounts of information from the system. If the analysis is to be used as a tool for urban planning, it would be useful to point to real routes on the bikeability curve that may need discomfort improvements. And as a planning tool, computational demand is less limited than if it were for a route planning app used by citizens, and so there is scope to perform the analysis for a less simplified system. What if an analysis was performed with data as close to real network as possible, but pre-processed to keep only the most relevant parts?

Work as part of the RUBICON team at UCL's Centre for Advanced Spatial Analysis has been coordinated with other projects to simulate the effects of transport infrastructure developments, new housing projects, and projects aiming to transition to net zero carbon in the UK (Batty *et al.* 2022). To achieve this, new data layers needed to be added to the usual rail, bus, and road layers that are used, including a cycling network layer. It is noted that the cycling network is more difficult to define than the rail network as the routes that can be biked are far more diverse (Batty *et al.* 2022).

Similarly to Reggiani *et al.* (2021), Open Street Map was selected as the data resource, as it was concluded to be more complete than Ordinance Survey data, with a better worldwide dataset, and with around four times the number of edge attribute classifications. This decision is also backed up in a study on OSM data in the United States that shows OSM being more accurate than google maps (Hochmair *et al.* 2015). The python OSMnx (Boeing 2017) package was chosen as the best method for accessing the data based on the characteristics of each possible package, as displayed in figure C.

	osmdata	**osmextract**	**Geofabrik**	**osmnx**
Big data	no	yes	yes	yes
Optimal for building graphs	no	no	no	yes
High details (attribute columns)	yes	yes	no	no

Fig C. Table showing the merits of different methods to access map data (Batty *et al.* 2022). Osmdata (Lovelace and Padgham 2022), and osmextract (Gilardi *et al.* 2022) are packages within R, Geofabrik (Geofabrik 2020) is an online tool, and OSMnx (Boeing 2017) is a package within Python.

Big data was important as it was believed to give more accurate data. An ease of graph building from the data was also deemed important, whereas some packages produced excess attributes, unlike OSMnx (Batty *et al.* 2022).

All edge and node data within London was downloaded using OSMnx (Boeing 2017). A range of networks were produced by filtering out different types of 'highway' combinations (an attribute of the edges e.g. 'motorway', 'cycleway'), as well as a network produced using the built in filter for the bike network. These were benchmarked against a gravity model to see how they would predict cycling flows compared to census data. The network that was selected was the OSMnx (Boeing 2017) network with the 'motorway' highway type removed.

Overall, this produces a more realistic version of a cycling network than using unfiltered OSM data like Reggiani *et al.* (2021), especially when it is not reduced to a simplified network, which is key when thinking of the needs of urban planners.

Chapter 4: Methodology

4.1 Geographic Scope

Though the analysis was aimed at London, cleaning the network for the whole city would have been too large a task for this project and the computational power needed to analyse the network may have outstripped the computational power available. For this reason, it was decided to analyse one borough. Using the Propensity To Cycle Tool (Lovelace *et al.* 2017), it was decided to analyse Hackney borough as that borough corresponds to an area of high rates of cycling in the 2011 census, which can be seen in figures D and E. The settings selected to show the map were 'Trip purpose: commuting' (as opposed to school travel for which there is little variation in cycling rates across London), 'Geography: MSOA', 'Scenario: Census 2011 Cycling', 'cycling flows: none'. For the type of analysis we will perform ideally this decision would be based on data for all types of cycling journeys in London, however this data was not available. Alongside this it was assumed that a higher rate of commutes by bike would mean more of a cycling culture within those MSOAs as residents are more likely to own a bike and have the confidence to cycle in London.

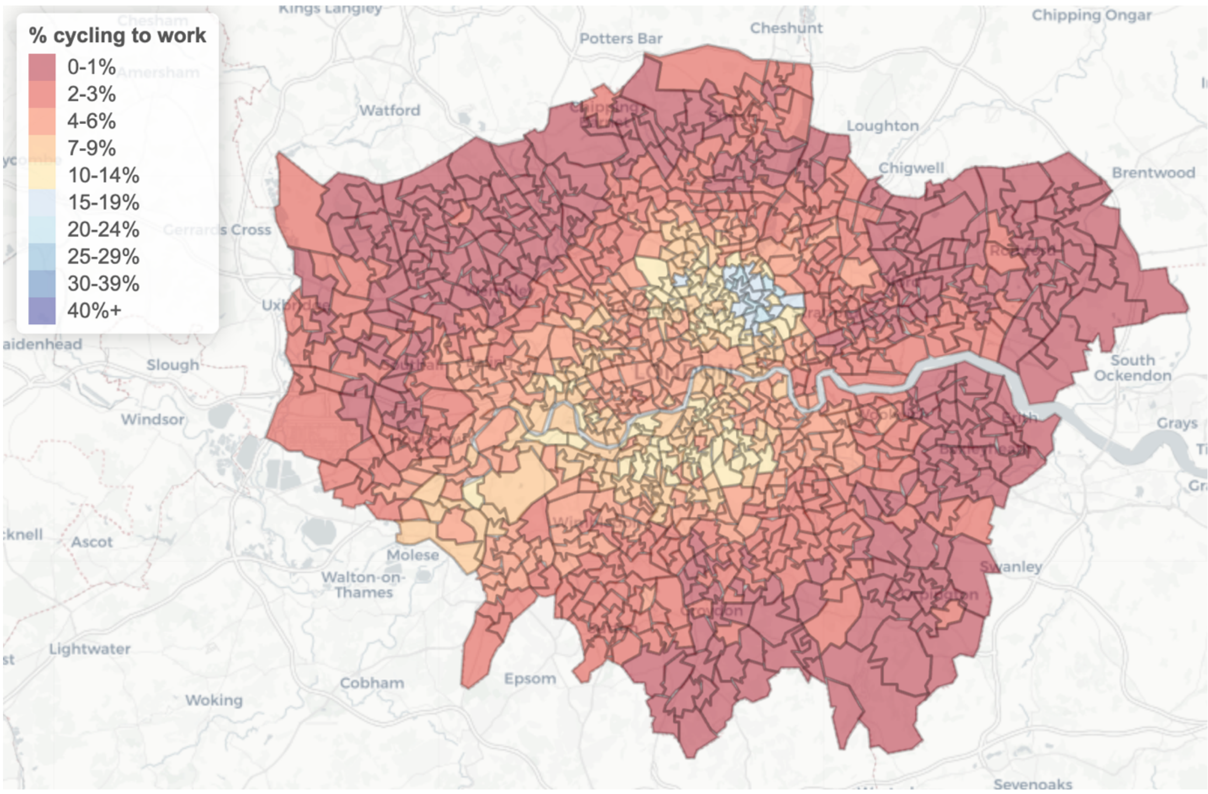


Fig. D. A map of London MSOAs with the % of population that commutes to work by bike from the 2011 census data, assigned as a colour, with a scale. (Lovelace *et al.* 2017)

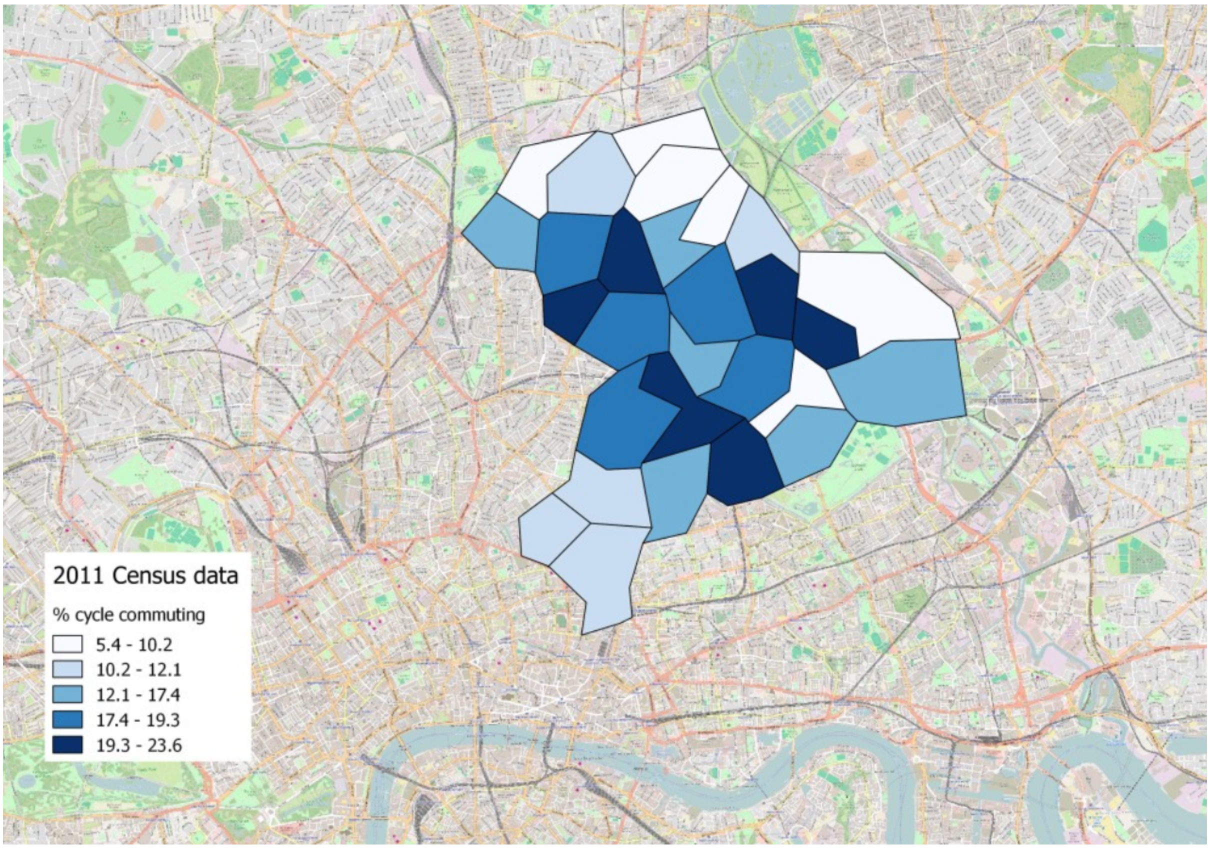


Fig E. 2011 census data for Hackney MSOAs for residents commuting by bike. (Aldred 2016)

4.2 Cleaning and Preparation

The network data for London was provided from the Rubicon team, as discussed in the Data chapter. This was loaded into QGIS software and the Hackney edges were selected by those that were contained within the Hackney borough polygon. This left some edges that had been protruding into Hackney from other boroughs which were not connected to the Hackney network, and which were deleted manually. The remaining network was exported as a shape file.

The MSOA centroids were found by loading the MSOA shapefile, found on the London Datastore website (Mayor of London, 2014) into QGIS software and going to menu, vector, geometry tools, polygon centroid, and 'create a centroid point layer'. This was then exported into a shapefile. MSOA rather than LSOA centroids were used in the analysis for two reasons. The first is that the 28 MSOA centroids provided enough origin destination pair combinations (377) to produce a pareto front. The second is that MSOAs are small enough so that a trip from a centroid to the centroid of a MSOA next door to it is approximately the shortest distance a person would cycle without deciding to walk instead. However, the same analysis could be performed with LSOAs if desired and the data can be found at the same location as the MSOA centroids (Mayor of London, 2014).

4.3 Producing Paths, Lengths, and Discomfort Scores

The code mentioned from here onwards can be found in the Appendix. It was developed using elements from a working paper under construction at the Centre for Advanced Spatial Analytics at UCL (Lopane *et al.* 2022). The cycling network shapefile and MSOA centroids were read into python.

In Reggiani *et al.'s* (2021) paper, the network is split into four layers which are assigned different discomfort levels depending on the link type. By avoiding splitting the Hackney network into layers the computational power needed to find the paths is reduced. This is because the code does not need to consider whether the path switches layers at each node and there are therefore far fewer potential paths between node pairs.

Networkx (Haberg et al. 2008) was selected as the package used to perform the analysis as although it is less computationally efficient compared to other packages, it is very well documented.

In python the nodes in the network that are closest to the locations of the MSOA centroids are found using the 'calc_closest' function, as defined in the code, and compiled into a list. The 'calc_closest' function does this by comparing the distance between each MSOA centroid and each network node. The list of resulting nodes are used as approximations for the MSOA centroids going forward in the analysis. Due to the density of the nodes in the cycling network, it was believed this was a reasonable approximation and would not make the analysis unreliable.

When the network is read in, the label of each node corresponds to its geographic coordinates. To make the code easier to use, the nodes are relabelled using the 'relable_nodes' function as defined in the code, so that the coordinates become an attribute and each node is given a unique integer label.

The number of connected components was found to be 10. Some of the functions we planned to use only work for a connected graph and there were just 27 nodes not connected to the main component. Given the small number of un-connected nodes, it was reasonable to remove them from the data. If there were sub-components of a size significant enough that deleting them would call into question the validity of the analysis, further thought would be required on what action to take. It would be unlikely that this would be the case for any London boroughs, as town planners aim to connect all edges that a cyclist could use. However, if this were the case, it might be worth studying other map data from the area, like Ordnance Survey Maps. If these other data sources suggested there was an edge missing from the OSM cycling network that connected two significant components, it would be reasonable to add it to the OSMnx (Boeing 2017) network.

A requirement of using the 'connected_components' function (used to check if the graph has multiple components) is that the inputted graph is undirected, and so our graph had to first be passed into the 'to_undirected' function. This does lose some information from the network, for example whether a street is one way. However, edges labelled as being one way account for only 0.12% of edges and so could not be expected to have a significant influence on the results.

Each edge in the network has a 'highway' attribute which describes the type of facility of the edge, for example, 'cyleway' or 'living street'. The full list can be found on the OSM Wiki website (OpenStreetMap Wiki 2020). A dictionary named 'highway_comfort' is created where each key is a different highway edge attribute, and each value the discomfort, derived from the stress of traversing the link. The discomfort value was based on a similar method to Mahfouz *et al*. (2021), as outlined in the literature. These are displayed in fig. F.

Highway type	Number of Edges	Percentage of Total Edges	Assigned Discomfort
residential	162949	27.5050048	0.2
footway	135845	22.9299804	0.2
service	111716	18.8571216	0.2
primary	32752	5.52837953	0.4
tertiary	29658	5.00612726	0.2
cyleway	24384	4.11590152	0.1
unclassified	19916	3.36172468	0.2
path	18539	3.12929373	0.2
secondary	11834	1.99752209	0.3
trunk	11448	1.93236715	0.5
track	4467	0.75400804	0.2
pedestrian	2765	0.46671866	0.1
bridleway	1957	0.33033216	0.1
steps	1449	0.24458421	1
trunk_link	1353	0.22837987	0.5
living_street	742	0.12524602	0.1
primary_link	736	0.12423325	0.4
tertiary_link	232	0.03916048	0.2
secondary_link	150	0.02531928	0.3
corridor	77	0.01299723	1
road	27	0.00455747	0.5
elevator	12	0.00202554	1
no	9	0.00151916	1
crossing	2	0.00033759	0.2

Fig. F. All highway attribute types with the number of edges assigned to that type, the percentage of total edges this forms, and the discomfort value.

Approximately 3% of edges had a highway attribute with multiple values e.g. both 'path' and 'bridleway'. This made them difficult to assign a discomfort associated value to. They were not included in the dictionary and so the edges were not assigned discomfort values. Because these consisted of just 3% of edges, they were unlikely to have a significant impact upon the analysis. Not giving these edges a discomfort value was thought more advantageous than removing the edges, as this could have led to the network being split into different components.

It was decided to assign some links a value of 1 as they are not traversable by bike. Removing them could produce multiple graphs which is not desired. They also represent a very small percentage of links and so leaving them in does not have a significant impact upon the analysis. Link type 'road' is defined by OSM Wiki as a link of unknown type (OpenStreetMap Wiki 2022), and so was assigned a discomfort of 0.5 as it the mean of all possible values. Links labelled 'unclassified' are in fact minor roads and so a discomfort of 0.2 was assigned to them.

It is important to note that many link types do not form a significant percentage of links. For example, 'tertiary_link' edges form only 0.026% of edges. A few link types dominate the network. For example, residential, footway, and service links represent 71.6% of all links.

A 'get_route_length' function adds up the lengths of each edge on a path. A 'get_edge_discomfort' function accesses the 'highway' attribute of each edge on a path, extracts the matching discomfort value from the 'highway_comfort' dictionary, and multiplies this by the length attribute. A 'get_route_discomfort' function normalises the discomfort for each path by totalling up the discomfort of the edges for each path and then dividing it by the total length of the path.

The 'k_shortest_paths' function defined in the code uses a pre-existing 'shortest_simple_paths' function (Haberg et al., 2008). 'Shortest_simple_paths' returns a list of paths between a received origin and destination node, and orders them from shortest to longest. Paths with repeated nodes are excluded from the returned list (avoiding edges being repeated in a path and so keeping the paths realistic to those a cyclist might take). 'k_shortest_paths' then selects the first k number of these paths. The weighting is set to length as this means the paths are more reflective of the actual length of paths (as experienced by a cyclist) rather than just the number of edges.

A final 'main' function is defined and employed which uses the previously described functions. This is where the origin node, destination node, and k number of paths is chosen. The function returns the k shortest paths, normalised discomfort, and length for each, and prints this information to a csv file.

The origin and destination nodes should be chosen from the MSOA node list produced previously, as the analysis looks at routes between MSOA centroids rather than any nodes in the network. This is because MSOA nodes make origin/destination node selection relatively easy (it is more challenging to locate other nodes) and it leaves space for further analysis to look at flows between MSOAs, especially if data for this becomes available. The value of k is also set within the function. In our analysis this was set as 500 as it is a high enough number of data points to find a pareto front without taking too long to compute.

From this export, the non-normalised and normalised lengths (produced by dividing path lengths by the length of the shortest path) are plotted against the normalised discomfort values for the node pair. Plotting with normalised lengths allows for comparison of the pareto fronts for origin-destination pairs that exist across different geographic scales. It also makes clear to the reader how much further a cyclist will travel on a path compared to the most direct path. Plotting with non-normalised lengths allows the reader to see how path lengths vary at a real geographic scale. The pareto front is drawn on by joining points on the graph with the lowest length and discomfort values.

4.4 Edge Type Distribution

A large number of the edges (71.6%) are formed from just three highway types – Residential, Footway, and Service. These are fairly evenly distributed across Hackney. However, the Cycleway edges constitute just 4.3% of edges and are more dominant in the East of the borough. This can be seen in the figures below. Red dots represent MSOA centroids, parks are shown in green, and bodies of water in blue.

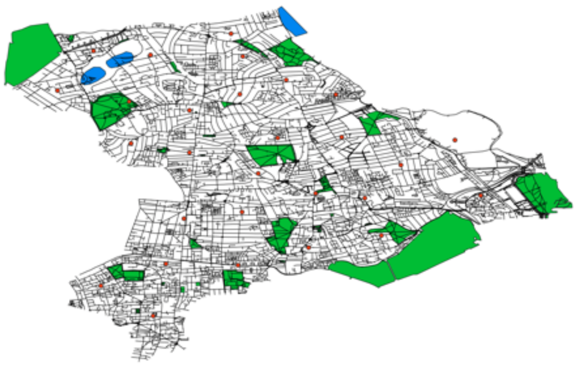


Fig. G. All edges in the network (100%)

Fig. H. Residential edges (28.4%)

Fig. I. footway edges (23.7%)

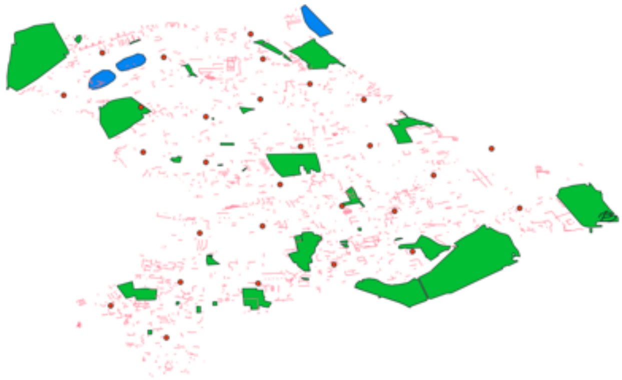


Fig. J. Service edges (19.5%)

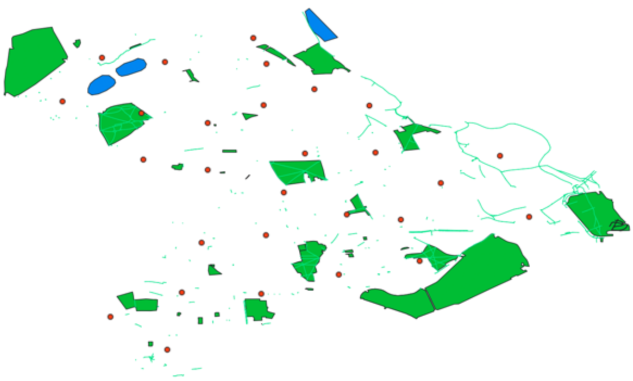


Fig. K. Cycleway edges (4.3%)

Chapter 5: Results

5.1 Origin-destination pairs across a larger scale

Graphs showing the normalised discomfort versus the normalised length of different paths between MSOA centroid combinations. These o-d pairs are cross-borough scale. The pareto front/bikeability curve is visible in red.

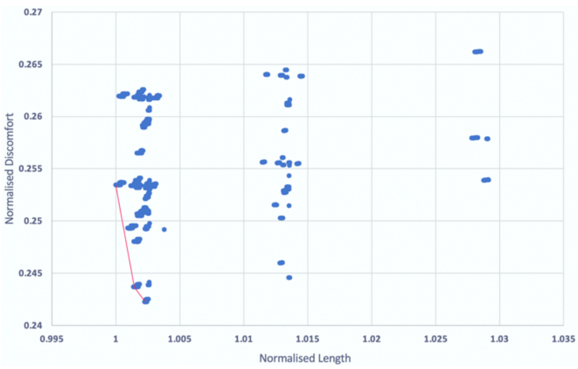


Fig. L

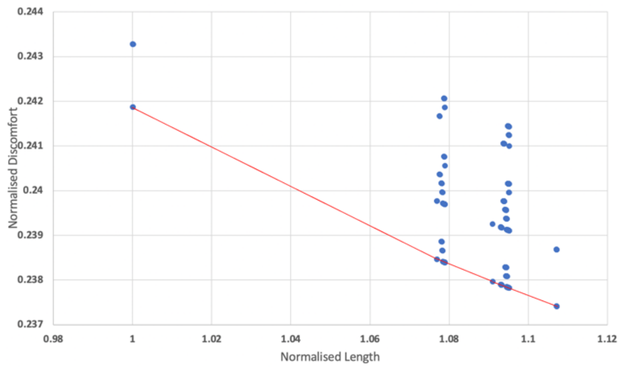


Fig. M

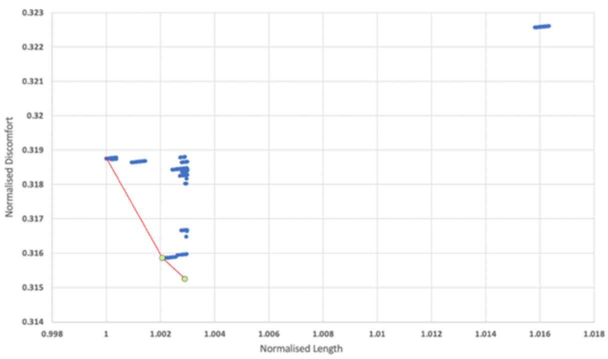


Fig. N

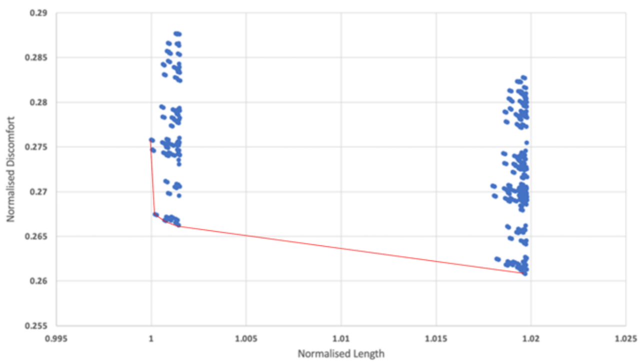


Fig. O

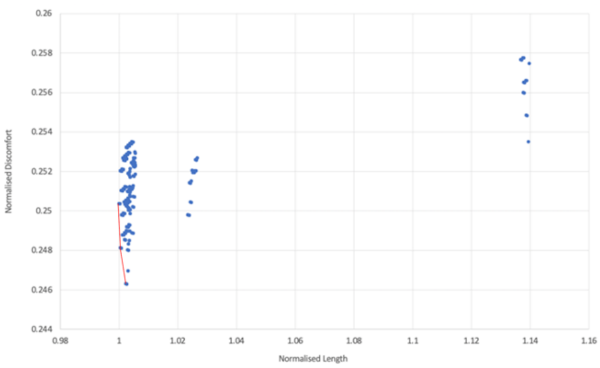


Fig. P

5.2 Origin-destination pairs across a smaller scale

Graphs showing the normalised discomfort versus the normalised length of different paths between MSOA centroid combinations. These o-d pairs are for MSOAs that sit next to each other. The pareto front/bikeability curve is visible in red.

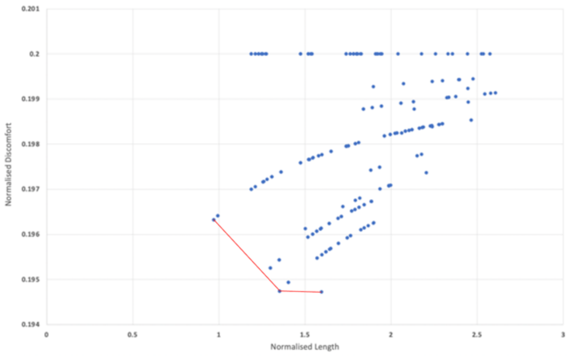


Fig. Q

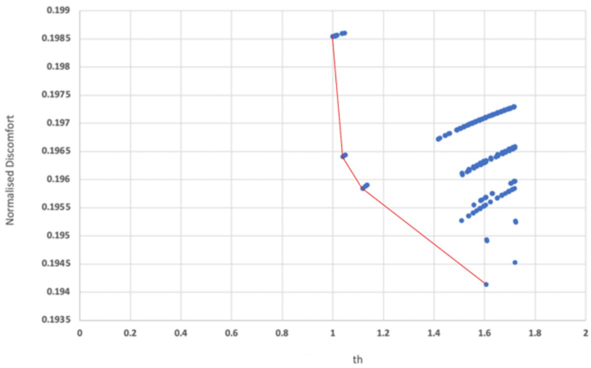


Fig. R

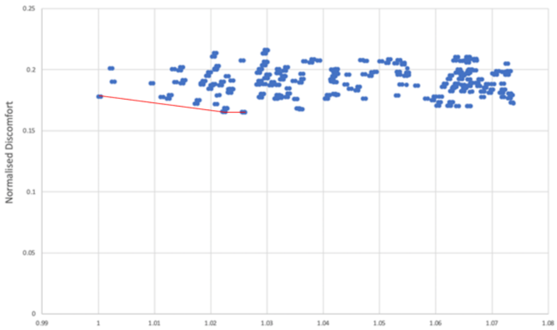


Fig. S

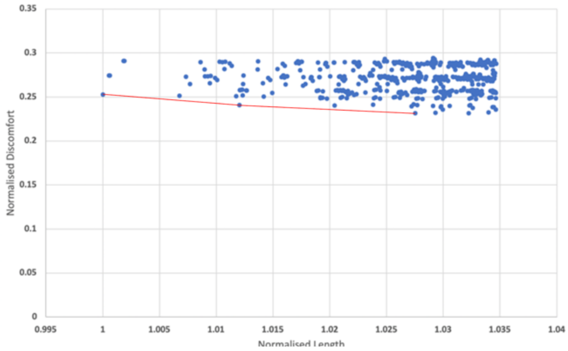


Fig. T

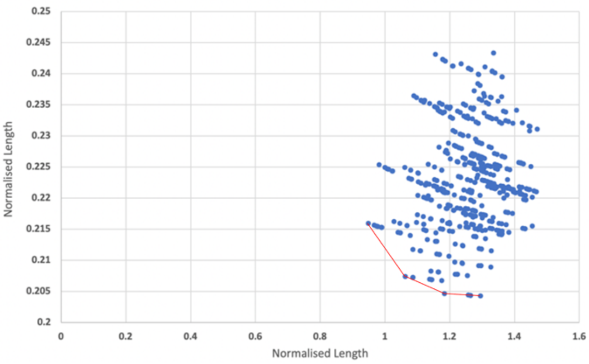


Fig. U

Chapter 6: Discussion

6.1 Interpretation of results

The two graphs with the most preferable pareto fronts are fig. Q and fig. R. This means that it is more preferable to cycle between these two o-d pairs for leisure than the other o-d pairs. This is because they combine both low discomfort values (between 0.1947 and 0.1964 for fig. Q and 0.1943 and 0.1985 for fig. R), whilst also having a greater variation in path lengths, giving the cyclist more choice. Fig. Q has a variation of 1 and 1.65 times the shortest path (65% detour), and fig. R has a variation of 1 and 1.6 times the shortest path (60% detour).

On the other hand, fig. N exhibits poorer characteristics. The range of discomforts on the pareto front is between 0.3153 and 0.3187. These are both greater than the discomfort ranges of fig. Q and R. Fig. N also has a length range of 1.003 (a 0.003% detour), which is far lower than fig. Q and R.

As mentioned previously in the literature review, a lower discomfort across all routes is preferable for a more pleasant sensory experience, so that cycling is encouraged. A larger range of lengths is preferable for those cycling for leisure as there is a greater range of paths lengths to choose from. As referenced in the literature review, there is ambiguity in the amount a cyclist will detour from the shortest path, but Krizek et. al (2007) suggest that this could be up to 93%. Our detour figures fit comfortably within this.

6.2 Reflection on how the results highlight limitations of the analysis

Reggiani *et al.* (2021) suggest that 'The ideal bikeability curve, between an OD pair, should resemble a vertical line that reaches the lowest possible discomfort value and with a relative distance close to one' (Reggiani *et al.* 2021). In our literature review we agree that low discomfort values are preferable for all cyclists, and acknowledge that a steep curve is preferable for those commuting. However, we are interested in those cycling for leisure. We suggest that a flatter curve would be preferable for these cyclists. From our results we can see that in practice it is more complicated. A graph may not appear flat even if it has a wide range of lengths (ideal), as it may also have a wide range of discomforts, even if these are at the lower end of the spectrum (also ideal). For example, fig. M is flatter than fig. R, but it has a far smaller range of lengths and has higher discomfort values. This means that bikeability curve results need to be looked at in a more considered way than just judging them by which is flatter. This would likely also apply to Reggiani *et al.*'s results, with regards to steepness, if they were to use a less simplified network/OSM network as we have done.

Of the five graphs with origin destination pairs at a greater distance, all have a low range of length paths, the greatest being fig. M, with a ratio of 1.11 times the shortest path (an 11% detour). However, three of the graphs with origin destination pairs at a shorter distance have far greater detours – around a ratio of 1.5 times the shortest path (a 50% detour). We would suggest the following reasoning for this. For origin destination pairs that are far apart the shortest path is longer and so has more edges. The likely next shortest path is a small variation on this, perhaps just a few edges forming a detour between two edges on the original chain. Though these two paths are similar they count as two different routes and two points on the scatter graph, and will have similar lengths and discomfort values. If this occurs many times the graph produced has a small range in length and discomfort. On the other hand, the paths between origin destination pairs that are close together have far fewer edges, and so there are fewer possible variations of these paths. As there are few variations of the shortest path, the code moves on to finding paths that deviate more strongly from the direct route, so their length scores deviate more greatly.

This raises a question. Should the number of shortest paths found be proportional to the distance between the origin and destination? This might prevent the previously mentioned phenomenon happening and make it easier to compare bikeability curves for o-d pairs at different scales. On the other hand, it is possible that cyclists would detour less for further apart o-d pairs as they would already be traveling a long way. It may therefore predict real life scenarios well. Further investigation into the maximum detour that cyclists accept across different geographic scales could help address this question.

The previous suggestion that o-d pairs at different smaller scales have less direct routes than those at larger scales is reflected in the graphs in another way. The further apart o-d pairs tend to have clusters of points on the graphs. For example, fig. O has two clusters of points. We would suggest that each group are similar variations of a path. Perhaps each variation has just a few edges that are different for each iteration.

The effect of geographic scale upon the analysis (greater detours at shorter distances and clustering of points at further distances) leads us to question to what extent the graphs from different scales can be compared. This would need to be taken into consideration by any town planners when using the analysis.

These two effects and the suggestion that the curves need more careful consideration than just how flat they are is likely due to the main difference between our analysis and Reggiani *et al.*'s. This is that the network we used was much closer to the real network, whereas Reggiani *et al.*'s was greatly simplified, as discussed in the literature review. Further analysis is needed to establish whether there is a mid-point between a 'closer to real life' OSM network and a simplified network, whereby the bikeability curves produced do not suffer from these effects. This would make the bikeability curves easier to compare, whilst keeping them as accurate as possible by not loosing as much information as when the networks are greatly simplified.

6.3 Limitations of the Methodology

A major limitation of the analysis is the computational power needed to produce the paths due to the number of path possibilities and o-d pair possibilities. We limited the demand by producing the 500 shortest paths and analysing just 10 o-d pairings. However, if similar analysis is used by government organisations, the computing power available could be much higher. If so, an extension to the work could be to perform the analysis for origin-destination pairs across the whole city. This would have several benefits. It would enable analysis at larger scales – at a city wide scale rather than at just borough wide. Pareto fronts from different boroughs could be compared to study how investment inequality has affected cycling, as we know rates differ (Lovelace *et al.* 2017), as well as how obesity rates and air pollution might relate to the network quality. Similarly, pareto fronts could be generated for origin-destination pairs that are suspected to have a high cyclist flow (e.g. through scenic/tourist areas). This would enable policy planners to best identify which parts of the network need further investment. Lastly, it would be possible to see whether the identified effects of geographic scale continue to develop as the distance between o-d pairs is increased further.

Additionally, greater computational power would allow a sensitivity analysis of the number of paths per o-d pair found, in order to discover the optimal number. This would be the smallest number of paths without greatly compromising the pareto front accuracy, and could be found in order to lower computational demand. To do this, the code would be run with different numbers of paths.

In order to do this, the Hackney network in the code would need to be replaced with the London network, which would be very simple. The main problem would be cleaning the network before it was added – deleting detached insignificant sub sections of the network or updating the network with any missing links that cause splits in the network. However, government organisations should have access to higher quality data, and so this may be less of an issue for them. This issue is also common for analysis of real world networks and so not unique to our analysis.

6.4 Further Developments and Uses

There are of course, further factors than the road type that influence the sensory experience of a cyclist, which was established in the literature as key to encouraging leisure cycling (Wild and Woodward 2019), (Anable and Gatersleben 2005). For example green spaces such as parks, and bodies of water, exist in Hackney and across London. Passing by or through one of these is likely to improve the sensory experience of the traveller. In fact, according to Cevero *et. al* (2019), infrastructure that pass by bodies of water may significantly increase levels of bicycle commuting (Cervero *et al.* 2019). It would seem economical to make use of these edges by improving the sensory experience of cyclists using them in other ways (what we have so far referred to as stress or discomfort), to further drive cyclists through them. Our analysis could make it easier to establish which of the edges that pass by water/through green space should be improved. By colour coding any data points on the graphs that have this characteristic, town planners could see which exist on or nearby the pareto front, and so target them with improved cycling infrastructure. The

Rubicon team's data contains a link attribute for both water and green space, and so this would be very achievable.

Sensory experience/stress factors other than this include the presence of signalised intersections. According to Liu and Marker (2020) 'bicycle-car accidents at signalized intersections account for a considerable part of bicycle accidents' (Liu and Marker 2020: 1). If signalised intersection location data is available, then the discomfort value of a path could be increased by a set amount each time such an intersection is encountered, to improve the accuracy of the bikeability curve. If this data does not exist, then a reasonable approach could be to increase the discomfort value an edge if it connects to more than one other edge at each end. This would be reasonable as even non signalised intersections may cause an increase in stress by increasing the likelihood of merging traffic.

Rather than using the shortest path between an o-d pair, it would be preferable to use the Euclidian distance (the distance covered by a straight line between the two points) to normalise the discomfort and length of the paths. This is because the shortest path may not be a particularly direct route and so may not accurately reflect the distance between the two points. This is in fact performed by Reggiani *et al.* (2021) in their analysis.

Our analysis acknowledges that there is more than one reason for using the cycling network in London – leisure cycling as well as commuting – and that these activities have different priorities – discomfort versus length. However, there will be other diversity within the cycling and potential cycling community. Different demographics may have different needs in terms of their sensory experience. If an area has a high population of retirees, might they be more likely to use the cycling infrastructure for leisure than others? Town planners could use census data and bikeability curves to find areas where populations prioritise leisure cycling but the cycling infrastructure has poor leisure bikeability, and target them for improvements.

Similarly, cyclists – especially women – are likely to be caused more stress by poor street lighting during low-light periods due to safety concerns (Uttley *et al.* 2020). Data containing the locations of street lamps could be used to add a discomfort score to any edges that don't have lighting. This could be used to study how bikeability differs across genders and how to improve equality of access.

Hackney council have proposed that 70% of the road network could be development into LTNs (Hackney Council 2023). This has the potential to radically change the experience of cyclists in the borough. Data on which streets are included in LTN zones would enable an analysis of how this might affect the bikeability of the area for those that wish to cycle for leisure. By comparing bikeability curves produced with and without the LTN zones, areas which would be most impacted by the changes could be selected.

Chapter 7: Conclusions

This study explores how bicycle networks might be analysed from an additional perspective, i.e. cycling for leisure instead of commuting, and performs a suggested analysis for Hackney, London. It does this using network data and analysing a range of shortest paths between origin and destination pairs at MSOA level. The normalised length and discomfort of the paths are plotted to produce a pareto front/bikeability curve. The discomfort score was based on the sensory experience of the cyclist, as this was found in the literature to be a large factor in encouraging/discouraging the target group. This was derived from the link type of OSM network data e.g. residential road, cycleway. This deviated from the literature which instead used a much simplified network.

Despite the study's early assertion that a flat bikeabilty curve is preferable for leisure cyclists, using OSM data produced bikeability curves that needed a more nuanced analysis of their shape. However, low discomfort levels across a wide range of lengths was still recognised as preferable. The results also demonstrated that geographic scale had an impact upon the analysis, making it difficult to compare bikeability curves for origin destination pairs at different scales. For larger cross borough scales the data points cluster, and at smaller scales higher rates of detouring become more common. This was not recognised by Reggiani *et al.* (2021) in their research.

Further analysis with fewer limitations on computing power would enable a study of how the model changes at a city wide scale, comparisons of the bikeability of different boroughs with different rates of cycling, and insights into whether air pollution and obesity rates change with network characteristics. An improved understanding of cyclist discomfort using data such as the location of bodies of water and green spaces, street lighting, and signalised intersections would improve the analysis.

Bibliography

Anable, J., Gatersleben, B. (2005) 'All work and no play? The role of instrumental and affective factors in work and leisure journeys by different travel modes', *Transportation Research Part A: Policy and Practice* **39 (2)**: 163–181.

Aldred, R. (2016) *Moving Forward in Hackney*, (http://rachelaldred.org/writing/moving-forward-in-hackney/) - accessed 1 June 2022.

Batty, M., Paz., V., and Schlosser, I., (2022) *Quarterly Report: RUBICON, RUBICON: PI Michael Batty*, Centre for Advanced Spatial Analysis, University College London.

Boeing, G. (2017) 'OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks', *Computers, Environment and Urban Systems* **65**: 126–139.

Cervero, R., Denman, S., and Jin, Y. (2019) 'Network design, built and natural environments, and bicycle commuting: Evidence from British cities and towns', *Transport Policy* **74**: 153–164.

Congdon, P. (2019) 'Obesity and Urban Environments', *International Journal of Environmental Research and Public Health* **16** *(3)*: 464

Furth, P. G., Mekuria, M., and Nixon, H. (2016) 'Network Connectivity for Low-Stress Bicycling', *Transportation Research Record* **2587 (1)**: 41–49.

Geofabrik (2020) *GEOFABRIK // Geofabrik* (https://www.geofabrik.de/geofabrik/) - accessed 25 January 2023.

Gilardi, A., Lovelace, R., Rowlingson, B., Fernandez, S., Potter, N., (2022) *osmextract: Download and Import Open Street Map Data Extracts* (https://CRAN.R-project.org/package=osmextract) - accessed 25 January 2023.

Gössling, S. (2020) 'Why cities need to take road space from cars - and how this could be done', *Journal of Urban Design* **25 (4)**: 443–448.

Greater London Authority (2011) *Childhood Obesity In London.*

Greater London Authority (2021) *The London Plan 2021*.

Hackney Council (2023) *Low traffic neighbourhoods*.

Haberg, A., Schult, D., Swart, P. (2008) *Proceedings of the Python in Science Conference (SciPy): Exploring Network Structure, Dynamics, and Function using NetworkX* (https://conference.scipy.org/proceedings/SciPy2008/paper_2/) - accessed 29 January 2023.

Hochmair, H. H., Zielstra, D., and Neis, P., (2015) 'Assessing the Completeness of Bicycle Trail and Lane Features in OpenStreetMap for the United States', *Transactions in GIS* **19 (1)**: 63–81.

Hosford, K., Winters, M., Gauvin, L., Camden, A., Dubé, A., Friedman, S. M., and Fuller, D. (2019) 'Evaluating the impact of implementing public bicycle share programs on cycling: the International Bikeshare Impacts on Cycling and Collisions Study (IBICCS)', *International Journal of Behavioral Nutrition and Physical Activity* **16 (1)**: 107.

Hruby, A., Hu, F. B. (2015) 'The Epidemiology of Obesity: A Big Picture', *PharmacoEconomics* **33 (7)**, 673–689.

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

Kuddus, M. A., Tynan, E., McBryde, E. (2020) 'Urbanization: a problem for the rich and the poor?', *Public Health Reviews* **41 (1)**: 1.

Landis, B. W., Venkat, V. R. Vattikuti and Brannick, M. T. (1997) 'Real-Time Human Perceptions: Toward a Bicycle Level of Service', *Transportation Research Record* **1578 (1)**: 119–126.

Liang, L., Gong, P. (2020) 'Urban and air pollution: a multi-city study of long-term effects of urban landscape patterns on air quality trends', *Scientific Reports* **10 (1)**: 18618.

Liu, P. & Marker, S. (2020) 'Evaluation of contributory factors’ effects on bicycle-car crash risk at signalized intersections', *Journal of Transportation Safety & Security* **12 (1)**: 82-93.

Lovelace, R., Goodman, A., Aldred, R., Berkoff, N., Abbas, A., and Woodcock, J. (2017) 'The Propensity to Cycle Tool: An open source online system for sustainable transport planning', *Journal of Transport and Land Use* **10 (1)**.

Lovelace, R., and Padgham, M. (2022) *osmdata Guide* (https://cran.r-project.org/web/packages/osmdata/vignettes/osmdata.html#7_References) - accessed 25 January 2023.

Mahfouz, H., Arcaute, E., and Lovelace, R. (2021) '*A Road Segment Prioritization Approach for Cycling Infrastructure*', arXiv:2105.03712

Mayor of London (2014) *Statistical GIS Boundary Files for London - London Datastore*. (https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london) - accessed 24 May 2022.

Mayor of London (2016) *Dr. Will Norman - Walking and Cycling Commissioner,* (https://www.london.gov.uk/who-we-are/what-mayor-does/mayor-and-his-team/dr-will-norman) - accessed 3 June 2022.

McIntyre, N. (2021) Traffic wars: who will win the battle for city streets? *The Guardian*, 25 March.

Mokhtarian, P. L., Salomon, I. (2001) 'How derived is the demand for travel? Some conceptual and measurement considerations', *Transportation Research Part A: Policy and Practice* **35 (8)**: 695–719.

Morris, E. A., Guerra, E. (2015) 'Mood and mode: does how we travel affect how we feel?', *Transportation* **42 (1)**: 25–43.

Moore, N. (2017) *Mindful Thoughts for Cyclists: Finding Balance on Two Wheels*, Brighton: Leaping Hare Press.

Lopane, F., Kalantzi, E., Milton, R., and Batty, M., (2022) A Land-Use Transport-Interaction Framework for Large Scale Strategic Urban Modelling, *CASA Working Paper **232**,* Centre for Advanced Spatial Analysis (CASA), University College London.

OpenStreetMap Wiki (2022) *Key:highway*, (https://wiki.openstreetmap.org/wiki/Key:highway) - accessed 24 May 2022.

Petegem, J. H. v., Schepers, P., and Wijlhuizen, G. J. (2021) 'The safety of physically separated cycle tracks compared to marked cycle lanes and mixed traffic conditions in Amsterdam', *European Journal of Transport and Infrastructure Research* **21 (3)**: 19–37.

Pharoah, T. & Plowden, S. (1973) '*Living Transport - London motorways fight'* (http://www.livingtransport.com/results.php?t=asset&search=245) - accessed 30 January 2023.

Reggiani, G., Oijen, T. V., Hamedmoghadam, H., Daamen, W., Vu, H. L., and Hoogendoorn, S. (2021) 'Understanding bikeability: a methodology to assess urban networks', *Transportation* **49 (3)**: 897–925.

Siddall, W. R. (1987) 'Transportation and the Experience of Travel', *Geographical Review* **77 (3)**: 309–317.

Soja, E. W. (1980) 'The Socio-Spatial Dialectic', *Annals of the Association of American Geographers* **70 (2)**: 207–225.

The Institute for Public Policy Research (2016a) *Lethal and illegal: Solving London's air pollution crisis*.

The London Assembly (2020) *Transforming cycling in outer boroughs: Mini-Hollands programme | London City Hall*, (https://www.london.gov.uk/programmes-strategies/transport/cycling-and-walking/transforming-cycling-outer-boroughs-mini-hollands-programme) - accessed 16 June 2022.

The Institute for Public Policy Research (2016b) *London: Global green city*.

Transport For London (2018) *Cycling action plan*.

Uttley, J., Fotios, S., and Lovelace, R. (2020) 'Road lighting density and brightness linked with increased cycling rates after-dark' *PLOS ONE* **15 (5)**: e0233105.

Wang, S., Zhou, C., Wang, Z., Feng, K., and Hubacek, K. (2017) 'The characteristics and drivers of fine particulate matter (PM2.5) distribution in China', *Journal of Cleaner Production* **142 (4):** 1800–1809.

Wild, K. and Woodward, A. (2019) 'Why are cyclists the happiest commuters? Health, pleasure and the e-bike', *Journal of Transport & Health* **142 (4):** 100569

Appendices

Python

Import Modules

In [164]:

```
import matplotlib.pyplot as plt
import pandas as pd
import geopandas as gpd

import networkx as nx
import os
import osmnx as ox
import csv
import random
from tqdm import tqdm
from math import pow, sqrt
from statistics import mean
from itertools import islice, combinations
from networkx import (
    read_shp,
    has_path,
    connected_components,
    shortest_simple_paths,
)
```

Functions

In [165]:

```
def calc_closest(new_node, node_list):
    """
    Find network nodes that are closest to mosoa centroids
    """
    # Calculate the closest node in a network
    best_diff = 100000
    closest_node = [0, 0]
    for comp_node in node_list.nodes():

        diff = (abs(comp_node[0] - new_node[0]) + abs(comp_node[1] - new_node[1]))
        if abs(diff) < best_diff:
            best_diff = diff
            closest_node = comp_node

    return closest_node
```

In [166]:

```
def relabel_nodes(Network):
    """
    The networks original node lables are cooridnates.
    We want the coordinate to be an attribute, and the label
    to be a unique integer.
    We assign numbers to each node (e.g. 1, 2, 3) so that
    we can refer to specific nodes easily. And we set their original
    label as a coordinate
    """
    # e.g. {(1, 2): (1, 2), (3, 4): (3, 4), ...}
    coordinate_map = {node: node for node in Network.nodes()}
    # assign coordinate attribute to each node
    nx.set_node_attributes(Network, coordinate_map, "coord")

    # e.g. {(1, 2): 0, (3, 4): 1, ...}
    label_map = {}
    for index, node in enumerate(Network.nodes()):
        label_map[node] = index

    # relabel each node with a unique integer.
    # to relabel in place.
    nx.relabel_nodes(Network, label_map, copy=False)
```

In [167]:

```
def get_edges_for_final_calculations(nodes):
    edges = []
    for i in range(len(nodes) - 1):
        edge_data = Network.get_edge_data(nodes[i], nodes[i + 1])
        if not edge_data.get("length"):
            continue
        elif not edge_data.get("highway"):
            continue
        elif edge_data.get("highway").startswith("["):

            continue
        elif edge_data.get("water") == 'true':

            continue
        elif edge_data.get("green") == 'true':
            continue
        else:
            edges.append(edge_data)
    return edges
```

In [168]:

```
def get_route_length(edges):
    """
    Total the length score for the edges in a path
    """
    if len(edges) < 1:
        return "N/A"
    total = 0
    for edge in edges:
        total += edge.get("length", 0)
    return total
```

In [169]:

```
def get_edge_discomfort(edge):
    """
    Access the discomfort value of an edge from the highway dictionary and multiply it by the length of the edge
    """
    highway = edge["highway"]
    comfort = highway_comfort[highway]
    return comfort * edge["length"]
```

In [170]:

```
def get_route_discomfort(edges):
    """
    Total up the discomfort score of a path and normalise this by the length
    """
    if len(edges) < 2:
        return "N/A"
    total = 0
    total_length = get_route_length(edges)
    for edge in edges:
        total += get_edge_discomfort(edge)
    return total / total_length
```

In [171]:

```
def k_shortest_paths(G, source, target, k, weight):
    """
    Instead of returning every single route between 2 nodes, we can just return a
    few (k) of the shortest ones.
    """
    return islice(shortest_simple_paths(G, source, target, weight=weight), k)
```

Input Import Phase

In [172]:

```
Network = nx.read_shp('data/hackney_largest_comp_w_green_water/hackney_largest_comp_w_green_water.shp') # edges of the
Zone_nodes = nx.read_shp('data/hackney_msoa_centroids_WGS84.shp') # MSOA centroids
```

```
/var/folders/5p/6n_rs1y10pz813w594fpqm340000gn/T/ipykernel_49221/2712873902.py:1: DeprecationWarning: rea
d_shp is deprecated and will be removed in 3.0.See https://networkx.org/documentation/latest/auto_example
s/index.html#geospatial. (https://networkx.org/documentation/latest/auto_examples/index.html#geospatial.)
  Network = nx.read_shp('data/hackney_largest_comp_w_green_water/hackney_largest_comp_w_green_water.shp')
# edges of the london cycling network from Rubicon
/var/folders/5p/6n_rs1y10pz813w594fpqm340000gn/T/ipykernel_49221/2712873902.py:2: DeprecationWarning: rea
d_shp is deprecated and will be removed in 3.0.See https://networkx.org/documentation/latest/auto_example
s/index.html#geospatial. (https://networkx.org/documentation/latest/auto_examples/index.html#geospatial.)
  Zone_nodes = nx.read_shp('data/hackney_msoa_centroids_WGS84.shp') # MSOA centroids
```

Establishing the Origin and Destination nodes

In [173]:

```
# count nodes in network before adding centroids
n_nodes_before = nx.number_of_nodes(Network)
print("number of nodes before adding centroids = ", n_nodes_before)
# count nodes in Zone_nodes
n_nodes_centroids = nx.number_of_nodes(Zone_nodes)
print("number of nodes from centroids = ", n_nodes_centroids)
```

```
number of nodes before adding centroids =  9959
number of nodes from centroids =  28
```

In [174]:

```
 # Find the nodes in the cycling network that are closest to the msoa centroids for use as Origins and destinations

msoa_nodes = []

for node in Zone_nodes:
    tuple(node)
    closest_node = calc_closest(node, Network)
    msoa_nodes.append(closest_node)
```

In [175]:

```
print(msoa_nodes)
```

```
[(-0.0687119, 51.5738436), (-0.0913373, 51.5710533), (-0.0824108, 51.5702815), (-0.0662209, 51.5698534),
(-0.0670974, 51.5625788), (-0.0972713, 51.5643323), (-0.0849739, 51.5626654), (-0.0749963, 51.5612488),
(-0.0610042, 51.5566272), (-0.0505092, 51.5565505), (-0.0848222, 51.55541), (-0.0327336, 51.5553349), (-
0.0747861, 51.5543309), (-0.0410454, 51.5525845), (-0.0639857, 51.5504499), (-0.054778, 51.5474863), (-0.
02775, 51.5472625), (-0.0472198, 51.5467082), (-0.0669532, 51.5440138), (-0.0765637, 51.5434902), (-0.044
4388, 51.5407042), (-0.0560914, 51.5387219), (-0.067873, 51.5357922), (-0.0796649, 51.5358918), (-0.08973
75, 51.5323591), (-0.0815383, 51.5273452), (-0.0516762, 51.5637973), (-0.0602466, 51.5661559)]
```

In [176]:

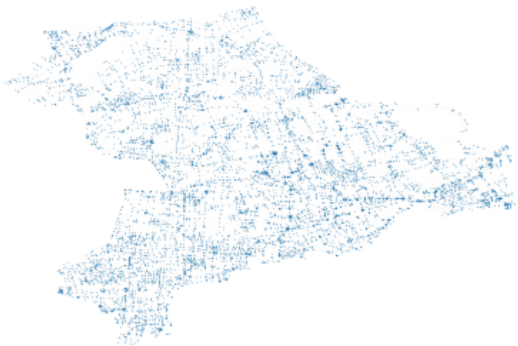
```
# count nodes in network after adding centroids to check consistent
n_nodes_after = nx.number_of_nodes(Network)
print("number of nodes after adding centroids = ", n_nodes_after)
```

```
number of nodes after adding centroids =  9959
```

Creation of the graph and reconfiguration

In [177]:

```
# creation of graph from scratch (and populate it with shapefile information)

pos = {k: v for k,v in enumerate(Network.nodes())} # Map the nodes of the network into a dictionary

X = nx.Graph()  # Empty graph
X.add_nodes_from(pos.keys())  # Add nodes preserving coordinates

l = [set(x) for x in Network.edges()]  # To speed things up in case of large objects
edg = [tuple(k for k, v in pos.items() if v in sl) for sl in l]  # Map the Network edges start and endpoints onto pos

# cleaning out edges with only one value in tuple
final_edges = []
for edge in edg:
    if len(edge) == 2:
        final_edges.append(edge)

X.add_edges_from(final_edges) # add edges to graph

nx.draw(X, pos, node_size=0.01, width=0.01) # draw
```



In [178]:

```
# print out the attributes of example edge
count = 0
for edge in Network.edges():
    if count < 1:
        print(Network[edge[0]][edge[1]])
    count += 1
```


```
{'fid': 22186.0, 'u': 10512502.0, 'v': 2886138738.0, 'key': 0.0, 'osmid': '2514025', 'name': 'Windsor Ter
race', 'highway': 'unclassified', 'maxspeed': '20 mph', 'access': None, 'oneway': 1, 'length': 33.47, 'fr
om': 10512502.0, 'to': 2886138738.0, 'bridge': None, 'lanes': None, 'ref': None, 'junction': None, 'servi
ce': None, 'tunnel': None, 'width': None, 'est_width': None, 'area': None, 'landuse': None, 'green': 'fal
se', 'water': 'false', 'ShpName': 'hackney_largest_comp_w_green_water', 'Wkb': bytearray(b'\x00\x00\x00\x
00\x02\x00\x00\x00\x05\xbf\xb8\x03<\xc8\xc13\xcf@I\xc3\xafc\xec\xa7\xc0\xbf\xb7\xfe\x95\x9c\xb4\xd7\r@I\x
c3\xb3\xc0\x8cL\xaa\xbf\xb7\xfa\x05\xea\xfb\x8b\xfd@I\xc3\xb5CCK\xaf\xbf\xb7\xf4\xa2\xd4\xab\xbe_@I\xc3\x
b6zz\xf0a\xbf\xb7\xf2\x17\xdf\xf0\x96\x00@I\xc3\xb6\xee>6\x9c'), 'Wkt': 'LINESTRING (-0.093799399387719 5
1.5287899880927,-0.0937284 51.5289231,-0.0936588 51.5289692,-0.0935766 51.5290063,-0.0935378 51.529020
1)', 'Json': '{ "type": "LineString", "coordinates": [ [ -0.093799399387719, 51.52878998809274 ], [ -0.09
37284, 51.5289231 ], [ -0.0936588, 51.5289692 ], [ -0.0935766, 51.5290063 ], [ -0.0935378, 51.5290201 ] ]
}'}
```


In [179]:

```
#relable nodes so that coordinates are an attribute, and the label a unique integer.
relabel_nodes(Network)
```

In [180]:

```
# check if Network nodes have key of count, value of position

print(nx.get_node_attributes(Network, "coord"))
```

```
{7512: (-0.0366204, 51.5451374), 5035: (-0.0708876, 51.5482638), 4002: (-0.0244364, 51.5482797), 7121: (-
0.0769705, 51.5516395), 3157: (-0.0500336, 51.548574), 7372: (-0.0846495, 51.5394616), 2739: (-0.0521666,
51.5430779), 5572: (-0.0681042, 51.5568926), 7288: (-0.074496, 51.5396324), 2647: (-0.0429278, 51.556158
3), 3536: (-0.0589857, 51.5410694), 2157: (-0.0219363, 51.5504024), 624: (-0.0701084, 51.534306), 2359:
(-0.0171941, 51.5437541), 7102: (-0.0708505, 51.5524351), 3684: (-0.0584977, 51.5401233), 5413: (-0.02042
44, 51.5454523), 5884: (-0.0665058, 51.5532138), 2962: (-0.0544173, 51.5559223), 3028: (-0.0607548, 51.54
22509), 7719: (-0.0436068, 51.5487915), 9569: (-0.0803804, 51.5620914), 1826: (-0.0641763, 51.5336076), 2
777: (-0.0462877, 51.5411274), 232: (-0.0805492, 51.5230659), 1055: (-0.0910144, 51.5309737), 9750: (-0.0
799555, 51.5548216), 5038: (-0.0709668, 51.5490071), 6718: (-0.07867951321979033, 51.550966522991246), 26
94: (-0.0595026, 51.5579007), 3584: (-0.0572799, 51.5482052), 4853: (-0.0613577, 51.5372209), 5221: (-0.0
626624, 51.5643962), 4397: (-0.0687779, 51.5701777), 4365: (-0.0368213, 51.549384), 6702: (-0.0777431, 5
1.5413398), 2484: (-0.0377991, 51.5542038), 7115: (-0.0712744, 51.5513007), 4964: (-0.0574377, 51.563832
1), 2363: (-0.0180154, 51.5443065), 6466: (-0.0534488, 51.5662394), 9183: (-0.0823948, 51.5681942), 7279:
(-0.0736664, 51.5389363), 4521: (-0.0667235, 51.5375935), 5051: (-0.0701957, 51.5508267), 8729: (-0.08566
22, 51.5733921), 9608: (-0.080201, 51.5547216), 2795: (-0.0462963, 51.5367971), 4614: (-0.063365, 51.5442
413), 9218: (-0.0787592, 51.5748866), 3305: (-0.0438294, 51.5393729), 8302: (-0.0913127, 51.5601276), 612
5: (-0.0547428, 51.5636557), 7516: (-0.0420635, 51.5443795), 6565: (-0.08070223518924935, 51.546417314311
455), 9713: (-0.0729141, 51.5752741), 2817: (-0.0366532, 51.5463005), 4223: (-0.0231397, 51.5444868), 461
0: (-0.0637963, 51.5447483), 4694: (-0.0672302, 51.5572534), 2411: (-0.0178015, 51.5441365), 1668: (-0.07
```

Create MSOA dictionary

In [181]:

```
# save the nodes' labels + attributes in a dictionary - so that can loop through its keys later on

msoa_nodes_dict = {}
nodes_attributes_dict = nx.get_node_attributes(Network, "coord")

for i in msoa_nodes:
    for j in nodes_attributes_dict.keys():
        if i == nodes_attributes_dict[j]:
            msoa_nodes_dict.update({j : nodes_attributes_dict[j]}) # want to save both the label and the value
```

In [182]:

```
# check msoa node dict

print(msoa_nodes_dict)
```

```
{5927: (-0.0687119, 51.5738436), 8976: (-0.0913373, 51.5710533), 9131: (-0.0824108, 51.5702815), 5555: (-
0.0662209, 51.5698534), 5676: (-0.0670974, 51.5625788), 8894: (-0.0972713, 51.5643323), 8782: (-0.084973
9, 51.5626654), 9369: (-0.0749963, 51.5612488), 4795: (-0.0610042, 51.5566272), 2988: (-0.0505092, 51.556
5505), 8624: (-0.0848222, 51.55541), 7886: (-0.0327336, 51.5553349), 9255: (-0.0747861, 51.5543309), 502
1: (-0.0410454, 51.5525845), 4748: (-0.0639857, 51.5504499), 5014: (-0.054778, 51.5474863), 4046: (-0.027
75, 51.5472625), 7604: (-0.0472198, 51.5467082), 4588: (-0.0669532, 51.5440138), 2219: (-0.0765637, 51.54
34902), 3738: (-0.0444388, 51.5407042), 6445: (-0.0560914, 51.5387219), 696: (-0.067873, 51.5357922), 175
3: (-0.0796649, 51.5358918), 1206: (-0.0897375, 51.5323591), 1563: (-0.0815383, 51.5273452), 3322: (-0.05
16762, 51.5637973), 4899: (-0.0602466, 51.5661559)}
```

In [183]:

```
# check number of msoa nodes
print(len(msoa_nodes_dict))
```

```
28
```

In [184]:

```
#view attributes for one edge

count = 0
for edge in Network.edges():
    if count < 1:
        print(Network[edge[0]][edge[1]])
    count += 1
```


```
{'fid': 465642.0, 'u': 6195188336.0, 'v': 6195188338.0, 'key': 0.0, 'osmid': '661733139', 'name': None,
'highway': 'footway', 'maxspeed': None, 'access': None, 'oneway': 0, 'length': 19.404000000000003, 'fro
m': 6195188338.0, 'to': 6195188336.0, 'bridge': None, 'lanes': None, 'ref': None, 'junction': None, 'serv
ice': None, 'tunnel': None, 'width': None, 'est_width': None, 'area': None, 'landuse': None, 'green': 'fa
lse', 'water': 'false', 'ShpName': 'hackney_largest_comp_w_green_water', 'Wkb': bytearray(b'\x00\x00\x00
\x00\x02\x00\x00\x00\x03\xbf\xa2\xbf\xe8\xb8\xbb\x9d\xd0@I\xc5\xc7\x0f\xf4i\xca\xbf\xa2\x9fO\xbb\xff\xf5
\xef@I\xc5\xc7\x1d`c\x17\xbf\xa2\x9e\xd3\x95=\xeaF@I\xc5\xc6x\x1e\xf5\x8f'), 'Wkt': 'LINESTRING (-0.03662
04 51.5451374,-0.0363717 51.545139,-0.036368 51.5451193)', 'Json': '{ "type": "LineString", "coordinate
s": [ [ -0.0366204, 51.5451374 ], [ -0.0363717, 51.545139 ], [ -0.036368, 51.5451193 ] ] }'}
```


Cleaning the graph

In [185]:

```
# check if network is connected and how many components. we need a connected graph
Network = Network.to_undirected()
print(nx.is_connected(Network))
nx.number_connected_components(Network)
```

```
False
```

Out[185]:

```
10
```

In [186]:

```
# print the nodes in each component to identify nodes in sub graphs
components_list = []
[components_list.append(c) for c in sorted(nx.connected_components(Network), key=len, reverse=True)]
print(components_list)
```

```
[{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 2
8, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 5
4, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 8
0, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 1
05, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 1
26, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 1
47, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 1
68, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 1
89, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 2
10, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 2
31, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 2
52, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 2
73, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 2
94, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 3
15, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 3
36, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 3
57, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 3
78, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 3
99, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 4
```

In [187]:

```
# remove nodes that compose sub components of graph

broken_nodes = [1477, 1448, 1449, 7323, 1466, 1467, 1505, 1506, 1498, 1499, 1500, 1501, 8337, 8338, 9021, 1578, 1579,
for i in broken_nodes:
    Network.remove_node(i)
```

Create Highway/Discomfort Dictionary

In [188]:

```
# create a dictionary of the dicomfort values for each highway type

highway_comfort = {
    "bridleway": 0.1,
    "corridor": 1,
    "crossing": 0.2,
    "cycleway": 0.1,
    "elevator": 1,
    "footway": 0.2,
    "living_street": 0.1,
    "path": 0.2,
    "pedestrian": 0.1,
    "primary": 0.4,
    "primary_link": 0.4,
    "residential": 0.2,
    "road": 0.5,
    "secondary": 0.3,
    "secondary_link": 0.3,
    "service": 0.2,
    "steps": 1,
    "tertiary": 0.2,
    "tertiary_link": 0.2,
    "track": 0.2,
    "trunk": 0.5,
    "trunk_link": 0.5,
    "unclassified": 0.2,
}
```

Output the normalised discomfort and lengths

In [189]:

```
def main():
    """
    Uses the other functions to output the normalised discomfort and length for each path, for a set nuumber of
    shortest paths between nodes, entered below. This is then printed to a csv file.
    """

    output_csv_filepath = os.path.join(os.getcwd(), "output.csv")
    start_node = 7604
    end_node = 8782
    number_of_paths_per_pair = 500
    with open(output_csv_filepath, "w") as csvfile:
        writer = csv.DictWriter(csvfile, fieldnames=["start", "end", "count", "length", "discomfort", "environment", "
        writer.writeheader()
        #for i, j in key_node_pairs:
        #    if i == j:
        #        continue
        #    if not has_path(Network, i, j):
        #        continue
        for k, path in enumerate(
            k_shortest_paths(
                Network,
                start_node,
                end_node,
                number_of_paths_per_pair,
                weight="length",
            )
        ):
                #print(path,",")
                edges = get_edges_for_final_calculations(path)
                writer.writerow({
                    "start": start_node,
                    "end": end_node,
                    "count": k,
                    "length": get_route_length(edges),
                    "discomfort": get_route_discomfort(edges),
                    #"environment": get_edge_environment(edges),
                    "steps": len(path) - 2
                })
                csvfile.flush()


if __name__ == "__main__":
    main()
```