You Can't Get There From Here:

Defining and Assessing Urban Cycle Networks

Name: Hugh Kelley

Course: CASA0010, Master of Science, Smart Cities and Urban Analytics

Centre for Advanced Spatial Analysis

Bartlett Faculty of the Built Environment

University College London

Date: 30 August 2019

Word Count: 9,634

Supervisor: Dr. Elsa Arcaute

This dissertation is submitted in partial fulfillment of the requirements for the degree of Master of Science in Smart Cities & Urban Analytics, Bartlett Faculty of the Built Environment, University College London.

I, Hugh Kelley, confirm that the work presented herein is my own. Where information has been derived from other sources, I confirm that this has been indicated. It is 9,634 words in length.

All maps displayed herein use data from:

All maps are projected in `UTM-30 EPSG:322300`.

Code used to produce results and analysis and latex files for this document available on Github at `https://github.com/HughKelley/Dissertation`.

Database available upon request via Github.

1 Acknowledgments

I received valuable assistance from a number of people while working on this dissertation. First and foremost was Dr. Elsa Arcaute who supervised and guided this work. Her availability and the thoughtfulness with which she considered my questions and guided the reading and background research for this work is much appreciated.

A number of other people at the UCL Bartlett Centre for Advanced Spatial Analysis assisted as well. These include Mateo Neira and Nicolas Palominos for their advice regarding networks and OpenStreetMap data. Thank you to Matthew Klovski for providing feedback and suggestions on a draft.

This dissertation would not have been possible without the support of my partner, Carolyn, my parents, Robert and Ellen and my brothers, Brendan and William.

2 Abstract

Accurate representations of street networks from the point of view of an urban cyclist should account for the level of stress that the cyclist experiences when traveling on the network. When streets with a level of stress that exceed a cyclists tolerance are removed from the network, some journeys become impossible or unacceptably long. To increase cycling's share of trips in a city, building a network of infrastructure that connects destinations without exceeding a certain stress tolerance is a key priority. This dissertation implements this methodology using data from OpenStreetMap (OSM), a community generated spatial database. It finds that while difficult, it may be possible to rely on the OSM classification of streets and tagging of cycle infrastructure where professionally collected survey data is not available. Streets tagged "Primary" and "Trunk" are nearly irreplaceable in the London street network. Once these streets are removed over 40% of journeys become impossible and the length of the remaining routes increase by approximately 30%.

Contents

List of Figures

List of Tables

3 Introduction

> "When you get these incomplete networks you have this situation where great, you have this fresh new bike lane, you're excited you know?" Maerowitz said. "This infrastructure is working for you and then suddenly it's gone and you have to go back out into traffic again." - Juhasz 2019.

3.1 The Challenge of Cycling in Cities

What makes a city comfortable for cycling? The conditions are obvious when one sees them, but defining the characteristics quantitatively is challenging. More than half of the world's population lives in cities(*World Urbanization Prospects* 2018). Cycling is of obvious importance to the well-being of the world's urban population in terms of the macro-climate crisis, local pollution problems, public health, wealth disparities, and urban traffic congestion. A commonly researched question is "What factors have the most significant effect on cycling rates?" Only recently, however, have researchers begun to study cycling from a network perspective. A comprehensive understanding of cycling networks is hindered by multiple challenges.

First is the scope of the analysis, considering individual changes to roads instead of the status of the entire transportation network from the cyclist's perspective. The majority of work on this question approaches the matter from a perspective of discrete policy intervention and the effect of marginal improvements on cycling. This approach fails to build a comprehensive theory of cycling rates that a network analysis approach can offer. Only a network approach can offer a complete understanding of the level of service for cyclists in a community.

The second consideration is the difficulty of obtaining data used by the studies that do take a comprehensive network approach. The most successful study used high quality data regarding street characteristics that the authors obtained from the local government. In many situations for financial or political reasons this data may not be available. This dissertation therefore uses open data to replicate the methods of previous work.

In this context, this dissertation contributes to the understanding of cycling infrastructure and cycling behavior by using London as a case study. London is an attractive case because it has a considerable cycling population but does not yet have the level of infrastructure that exists in world leading cities like Copenhagen (Mayor 2017). Thus London's position at a mid-point of cycling infrastructure development allows for the identification of strengths and weaknesses to a live program of improvement.

3.2 Research Question

The goal of this dissertation is to build and interpret quantitative measures for the quality of London's cycling network. First, OpenStreetMap data will be assessed as the primary data source for building representations of London's cycling infrastructure. Second, metrics will be calculated and compared for the different network representations. Third the comparisons will be used as the basis for drawing conclusions about the London cycling infrastructure network. The most important objective of this dissertation is to specify how data and methods can be combined and improved to provide for a fuller understanding of the current quality of the networks and to indicate the best ways to improve the network in the future.

The dissertation will address a few secondary questions. First, to what extent is transportation space in London a zero sum game? Does improving the experience of cyclists require taking space from motorists? Second, are there any key differences between the grid-like networks studied in past research in North American Cities and the more tree-like network of London streets?

3.3 Ethical Risks

This project relies entirely on publicly accessible data. For this reason, ethical risks are not present in the research methodology.

3.4 Research Structure

First, existing work on this research area will be reviewed and the techniques to be built upon will be identified in section 4. Then section 5 will describe a general methodology for defining the strength of cycling infrastructure in a city. Section 6 will describe the data available regarding London's cycling infrastructure in the context of past work, and what is available for other cities and through other channels that were not available to this research. The steps taken for data cleaning, transformation and joining will be specified. Section 7 will describe the implementation of this methodology for London, identifying the scope of the case study, defining the exact data collected and transformed and the specific tools used, and reporting the results of the implementation. Finally section 8 will set forth conclusions, identify opportunities to improve the methodology and the quality of the data, and make key recommendations for further improving the London cycling infrastructure network.

After critically reviewing this existing research, a methodology for using open source tools and data to estimate the relative strength of a cycle network and a method for prioritizing improvements will be specified.

The review that follows takes an initial look at typical literature considering how to promote cycling, recent attempts to use network analysis to accomplish this same goal, and closes with a look at work analyzing data sources relevant to this investigation.

4 Literature

4.1 Cycling Behavior

The most important contribution from literature that tries to predict cycling behavior is the focus on multiple types of cyclists, and the factors that influence each type's decision to cycle. Most often, researchers identify four types of cyclist, "strong and fearless", "enthused and confident", "interested but concerned", and "no way no how" (Dill and McNeil 2013). This categorization is sometimes changed so that the final category is dedicated to children with their set of safety requirements for a suitable cycling environment (Mekuria, Furth, and Nixon 2012). The literature separates the decision to cycle into how often someone who is willing to cycle generally chooses to do so (cycling frequency), and the decision to cycle at all, with different factors influencing each decision (Stinson, Bhat, et al. 2005).

Despite behavioral differences between these categories of cyclists, research has shown that all cyclists are willing to sacrifice time and energy for increased safety on their route (Winters et al. 2011). Indeed, psychological research showed that fear is a significant factor during an urban cycling trip (Ellett, Kingston, and Chadwick 2018). This is important given the sensitivity cyclists show to efficiency (Wuerzer and Mason 2015). The impact of a route change can be very significant; for instance a higher frequency of stop signs on a route can double the energy required for a journey (Fajans and Curry 2001). Given the common trade-off between efficiency and safety, it was found that the effect of infrastructure improvements is very dependent on context. Effect is a function of the change in safety, and the importance of the location to trips (Kondo et al. 2018). Improvements that meaningfully increase safety at important locations have the largest effect.

Several studies looked at the importance of perception in behavior change, assuming that a real change in safety is irrelevant if it is not perceived by potential cyclists as a change (Li et al. 2012) and (Parkin, Wardman, and Page 2007). This gives rise to literature that focuses on a behavior and attitude change approach from psychology that prioritizes change in habits and perception over infrastructure, with changes to the built environment only used where required to change perception (Savan, Cohlmeyer, and Ledsham 2017). This should be considered in the context of research showing that cyclist perceptions of danger are generally accurate (Vandenbulcke, Thomas, and Panis 2014). Thus it seems reasonable to conclude that although the decision to cycle is a fairly complex mix of factors, when an urban commuter is deciding whether to cycle, the most important factors are safety, and efficiency. An efficient network of safe street-edges, connecting important place-nodes, would be expected to have a meaningful effect on the rate of cycling in an urban area.

4.2 Cycling Networks

Buehler and Dill (2016) is a very useful introduction to the literature on cycling networks. Unfortunately it concludes that very little true network analysis has been developed for cycle networks. They found that the majority of papers could be categorized into those that focus exclusively on nodes, and those that focus on edges of the dual graph, where intersections are nodes and streets are edges. At the time of writing there were 115 papers citing Buehler and Dill's review; however all but five of them fail to take the central recommendation that *If individual characteristics of a network's links and nodes contribute to cycling levels, it logically follows that a network of such features would as well...*. The "Toward Studying the Whole Bicycling Network" section of Buehler and Dill's review is a good overview of attempts up to 2016. The key findings were that continuity and connectivity of infrastructure are valued by cyclists. Of particular interest is the Schoner and Levinson (2014) study of the relationship between network characteristics and cycling mode share in 74 US cities, finding that the density of the network had the highest elasticity of effect on cycling rate.

Several of the works reviewed contribute new ways of measuring "quality" of the infrastructure. These quality measures include a Bicycle Compatibility Index (BCI) (Klobucar and Fricker 2007), Bicycle Level of Service (BLOS) (Lowry et al. 2012), and Level of Traffic Stress (LTS) (Mekuria, Furth, and Nixon 2012). Each of these can reasonably be viewed as an attempt to measure the "safety" of a network link. These studies generally lacked a rigorous method for prioritizing nodes by importance or defining a sample set of trips between nodes. Improvements in this area will be addressed in the section reviewing network analysis literature.

Buehler and Dill note that a key challenge to using the network methods reviewed is data availability, this dissertation hopes that network analysis can be a method for reducing rather than extending the amount of data necessary to understand a cycle network, as network statistics could be used to replace some empirical measurements as discussed below. They further criticize the approaches as lacking empirical validation. Gathering accurate cycle traffic data and safety data is an immense challenge as demonstrated by the flow estimation techniques of Gosse and Clarens (2014) and the safety estimate techniques of Puchades et al. (2018). The latter focuses on near misses as a proxy for predicting actual safety incidents, but acknowledges the difficulty of collecting near miss data without human observation.

Since the publication of Buehler's review, the papers extending the full network analysis method have had moderate success. Akbarzadeh, Mohri, and Yazdian (2018) use taxi trip data to weight the links between destinations in order to build communities of nodes that tend to be origin and destination pairs. While this is a novel approach to prioritizing edges, it seems likely that taxis trips tend to be used for one time trips which could be very different from daily journey to work trips (National Statistics 2011).

Doorley et al. (2019) focus on building cycle infrastructure to maximize a function of travel costs, infrastructure costs, health, traffic accidents, and pollution. While this is an interesting approach, it addresses a more political question in the sense that the key result of the algorithm is to recommend a specific amount of investment in cycle infrastructure to maximize the costs and benefits to all road users, The authors fail to recognize the prioritizing the goals of public policy is a normative and subjective exercise and that a model designed to give an "objective" answer to this question inherently reflects the author's preferences and when calibrated to "the real world" reflects the biases and preferences of the status quo, rather than the true ideal outcomes preferred by a population. Instead, modeling, especially for urban planning purposes, should accept an exogenous goal and implement it as efficiently as possible. For instance, cycle infrastructure, is explicitly intended to reduce motor vehicle use, it would make no sense to then use a model that determines the efficient level of motor vehicle use, the political process has already determined the answer and merely asks for implementation recommendations from the modeler.

Mauttone et al. (2017) similarly focuses on an optimization framework for cycling networks, choosing a subset of streets that are "suitable to building cycle infrastructure". This is odd in the sense that the goal of building cycling infrastructure is to *create* streets that are suitable for cycling, not merely identify them. Similar to Doorley et al. (2019), they identify a cost to building cycle paths which they seek to balance against the benefits.

Overall, it is not clear that a model for building cycle paths should be particularly cost sensitive. Gu, Mohit, and Muennig (2017) found a very high return on investment to the budget for cycling infrastructure in New York City. The very idea of using network analysis for the development of cycle networks emphasizes the potential non linearity of the effect of building more infrastructure, with usage accelerating as the network approaches "completeness" in some form. In addition, cities tend to combine cycling infrastructure improvement with other required improvement and maintenance activities, mitigating the costs by being opportunistic in implementation.

Lastly, Osama, Sayed, and Bigazzi (2017) use a number of predictors including network statistics to predict bike travel within zones of Vancouver similar to Schoner and Levinson (2014). They found a positive coefficient for the density of the bike network in a zone. fThus while network analysis has been applied to cycling infrastructure there is not a consensus on the methodology to be used. Additionally a definition of "quality" or "safety" has not been clearly established.

The most direct inspiration for this work comes from Furth, Mekuria, and Nixon (2016). That analysis built representations of the San Jose cycle network for different types of users employing data collected and maintained by the local government. Data used included the locations of cycle tracks and shared paths, the width of street lanes, the width of bike lanes on those streets, the volume of traffic by lane for each street, right of way in intersections and the structure of each intersection, and the frequency of bike lane blockages for each street. The analysis then calculated from the data on journeys to work a "connectivity ratio" that is the percentage of commuter trips that are possible for a given representation of the network.

Furth notes that "These connectivity methods do not necessarily require use of the LTS classification scheme. They can be applied with any classification scheme that distinguishes high- and low-stress segments." This is a vital component of this dissertation as the majority of data that Furth uses is not available for London. In the methodology section, a process for building similar networks using data from OpenStreetMap will be specified. Furth also notes the computational burden of the analysis as a limiting factor.

The second important inspiration for this dissertation is Boisjoly, Lachapelle, and El-Geneidy (2019). This work used a survey of cyclist route choices between sets of destinations in Montreal to estimate the probable path for a larger set of destinations. With these paths, they estimated the average directness of a journey in the city to identify neighborhoods with relatively low directness of cycling routes. Their analysis considered for a given route the percentage of the routes distance that occurred on cycling infrastructure and the directness of the route relative to the directness of the shortest possible route between the given origin and destination. It looked particularly at the idea that "there is often a trade-off between route directness and quality of route" (Boisjoly, Lachapelle, and El-Geneidy 2019). The routes were predicted using data from a survey of 1,525 cyclists, which collected data on their cycle trips regarding usage of bicycling infrastructure. The study specified a cost function that allowed for the expression of stress and distance in common units by penalizing high stress edges, increasing their distance cost, and reducing the distance cost of low stress edges. The cost function was estimated from the survey of cyclists. It measures how far from the shortest route the cyclist diverted in order to use a piece of cycling infrastructure.

The strength of the Boisjoly study is that it does not require high detail data on the street network, only the simple street connectivity network and data on the location of cycling infrastructure. The weakness of the study is twofold. It requires substantial survey data on route choice from network users and it uses minimal discrimination across the quality of cycle infrastructure, and across the level of stress for a given street.

4.3 OpenStreetMap Data Quality

A central research question of this work is "how sufficient is Open Street Map (OSM) data for replicating the studies considered above." This is because both the Furth and Boisjoly studies rely on data unavailable for London, cycling route choice survey data and high detail street characteristic data.

There is a strong body of existing research on the quality of OSM data. This research can be divided into two focus areas and two types of methodology. The two focus areas are locational accuracy of features and the completeness of attribute tagging and description for features. The two methods for assessing the quality of the data are extrinsic and intrinsic. The extrinsic method uses an external professionally collected dataset. The intrinsic analysis attempts to solely use an analysis of OSM data to assess itself.

Studies on locational accuracy are consistently extrinsic. Haklay (2010) compared OSM in the UK to the government produced Ordnance Survey data for roads in the UK finding that there was about 24% coverage of the UK and that features tended to be very close to their location in the Ordnance survey data. This is of minimal use to assessing cycle networks however, because the characteristics of the features, streets, are of much more importance than the precision of their locations.

Assessing the accuracy and completeness of feature attribute tagging in OSM is a more recent endeavor and more difficult due to the lower availability of data, the more frequently changing nature of the data as road works are undertaken and the very open structure of tagging of attributes in OSM.

The use of OSM data for specialized routing applications has been considered by Mobasheri et al. (2017). In this study, they considered the quality of sidewalk/pavement tagging in OSM from multiple cities for the purpose of routing wheelchair users. This study was looking for information about the surface type, incline, and width of pedestrian ways. They found about 17% coverage of sidewalks in Hamburg Germany and that coverage was best where density of features and tags was highest. This work combined extrinsic and intrinsic analysis of the question but did not take the opportunity to validate the intrinsic analysis with the extrinsic analysis. They concluded that large parts of the cities considered had OSM data that could support specialized routing if the data quality was confirmed by additional checks.

Hochmair et al. (2013) researched the completeness of bicycle features in OSM. They used Google Maps to extrinsically validate the OSM data for bike trails in the US and Europe. They found that coverage was fairly high, but they only considered fully separated bicycle infrastructure like segregated lanes and off-road trails, which is of limited use to the analysis of urban cycling networks where many cycling infrastructure features exist on shared roads.

Finally, Zielstra, Hochmair, and Neis (2013) considered the impact of bulk uploads of geospatial data to OSM. They found that in the United States, while government collected data had a higher level of completeness for motor vehicle related street network data, data for pedestrian related features was higher in OSM. This raises the possibility that for a well mapped area, OSM could be the best possible source of data for cycling, depending on the priorities of local governments in their data collection efforts. This is especially important in the context of the new Transport for London Cycling Infrastructure Database, which will be imported to OSM over the next few years (London 2019). Despite the possibility that there are meaningful problems with current cycle related data in OSM today, OSM is likely to be the highest quality source of this data in the near future.

Ultimately, the quality of OSM data for cycling related data is uncertain and the methodology and analysis will address a qualitative attempt to understand accuracy in London, with the expectation that it is at least as high as anywhere else in the world.

5 Methods

5.1 Scope

The scope of the study will be defined to maximize the study's representation of London's cycling commutes, capturing as many trips by origin and destination as possible as well as considering the location of road casualties in London. This will be done in the context of the computing resources available, where travel time calculations for the origin destination matrix must be reasonable, less than one million pairs.

5.2 Defining Networks

The first step in this investigation is to build a data set that represents the London cycling network as accurately as possible. This representation needs to reflect the fact that different cyclists are willing to use different streets as a function of the perceived safety of the street and the level of confidence of that cyclist. Thus, the data set will be multiple representations of the city cycling network that each represent a level of confidence, only including streets with a certain level of safety.

A key question then is, how to quantify "safety". In a perfect world, this would be done empirically. This would involve a combination of cycle traffic volume collection, cycle traffic behavior observation, and interviews with a representative set of cyclists and non-cyclists about their decision making. All of this data could be compared to the cycling environment in different locations to find cyclist sensitivity to different factors. In the absence of the required data, this dissertation will rely on the implications of OSM definitions for various types of streets.

Networks will be defined as subsets of "ways" in the OSM street network. OSM uses tags to associate street characteristics with the geometries that make up the map. Table 2 contains the definitions from the OSM wiki page for each of the tags used.

Using the network filters, the connectivity data is downloaded from OSM via the Overpass API as a JSON file. A graph is built with the nodes and ways from the Overpass data with edges coming from the "way" elements, and nodes coming from the intersection of the edges as defined in OSM as well as the end of an edge. This requires simplification because the OSM data includes many more nodes than just the endpoints and intersections. The network will be simplified by removing nodes with degree 2, where the node simply connects one street to one other, unless there is a difference in directionality between them. That is, when a street changes from two way to one way, a node will be placed at the intersection to denote that change. Each of these steps can be accomplished using a combination of Python packages `OSMnx`, (Boeing 2017) and `Networkx` (Hagberg, Swart, and S Chult 2008).

5.3 Defining Origins and Destinations

London was divided into 4765 Lower Super Output Areas (LSOA) for the 2011 census ((GLA) 2014). One Origin/Destination (O/D) point will be selected for each LSOA in the scope. QUANT uses the node of highest degree in a given LSOA. For calculating a sample of routes on the networks, this analysis will use the node closest to the centroid of each LSOA.

5.4 Assessing Networks

A key point of interest is the network structures that result from different filters. For different filters, the count of nodes and edges, and therefore the overall density of the network will be considered. The size of the largest component will be considered as well as the number of O/D pairs connected by the network.

The distance of the shortest path between each origin and destination will be calculated for each network with more than 50% of pairs connected. These data will be converted to measures of directness, dividing by the straight-line distance between the two points, and into travel times, dividing by the speed of a cyclist as estimated by Google Maps.

In addition to calculating the multi-directed graphs for different OSM filters, distances will be calculated for undirected versions of the graph. This will be used to investigate how building infrastructure for cyclists to travel safely against traffic could further raise accessibility or replace the need to build infrastructure that takes space from other users on main roads.

Finally, the mean and shape of the cycling distribution will be compared to the shape of the QUANT transport distribution.

6 Data

6.1 OpenStreetMap

OSM is a mapping project started in 2004 to collect volunteered geographic information. It consists of geometries drawn by users, either in person, as they travel through a city or remotely, looking at donated satellite images of cities. This data is accessible via the overpass API from several hosts. Boeing (2017) describes using the OSM Overpass API query language as "notoriously difficult".

Second to the actual geometry of a "way" (streets and paths), a node (single point on the map), or relation (collection of ways nodes and other relations) are tags. Tags specify what a particular geometry is, what its characteristics are, and rules for use or other characteristics of the geometry. This allows for differentiation on the map between public and private areas, specification of what exactly a node is referencing, an intersection, mailbox, or a business location, or the type of traffic allowed or commonly seen on a street way.

OSM allows users to tag features with any key value pair they see fit. This allows for maximum flexibility in handling real world conditions but also allows for errors and inconsistencies that will be addressed in the Analysis Section 7.

6.2 Transport for London Cycling Infrastructure Database

The Transport for London Cycling Infrastructure Database (CID) data was in the process of being publishing during the period of research for this work (London 2019). While it was not available at the time of publication, it is notable because it promises to raise the quality and volume of OSM data for London substantially. It includes 2,000km of cycle lanes as well as hundreds of thousands of parking spaces, cycling related signs and other relevant features. (OpenStreetMap 2019).

Literature review section 4.3 addressed the possibility for bulk data uploads to dramatically enhance OSM data quality, and this may be one example. The data was professionally surveyed.

6.3 QUANT

The QUANT dataset of public transport travel times was calculated as part of an effort to build models of interactions between areas in the UK (*QUANT* 2019). Of the 22 million origin-destination pairs, 5.8 million pairs had no public transport link. To clean this data, the set will be cut down to match the scope of the investigation. Where there is no link between an origin destination pair, a link will be constructed by combining the walking time from the origin to the node of highest degree in another LSOA where public transport is available to the destination LSOA.

The walking speed used will be taken from Google and the distance is the straight-line distance between two points. This approach is less accurate than actually finding the walking route between the two points but this level of detail was not computationally feasible.

The QUANT data provides a point of comparison for cycling travel times to be calculated.

6.4 2011 census journey to work data

The 2011 census asked each household where they lived, where they worked and how they traveled to work the preceding week (National Statistics 2011). Thus data for origin and destination by mode of transport is available. This data will be used to help define the optimal scope for the analysis.

Unfortunately, the data is publicly available only at the borough level, because of privacy concerns around high resolution home and work location data becoming personally identifiable. In order to calculate a meaningful connectivity ratio, following Furth, Mekuria, and Nixon 2016 this resolution would need to match the LSOA level resolution of the origin and destination pairs of distance calculations.

This data is shared through the Nomis Labor Force website as multi-sheet excel pivot tables (National Statistics 2019). Making the data usable requires stripping the meta data headings from each sheet, importing the book to a `pandas` dataframe by sheet, melting from a pivot table to long data with origin, destination, and count columns, adding the sheet name that identified the mode of travel as a column, appending each individual sheet together into a single dataframe, and pushing the dataframe to the `PostgreSQL` database.

6.5 LSOA boundaries and data

LSOA boundaries were obtained from National Statistics (2016). The LSOA boundaries were determined as a part of the 2011 census containing approximately 5000 people each. These are used, first to specify a boundary for the scope of the study and then to specify origin and destination nodes on the network. The node closest to the centroid of each LSOA is selected as the origin and destination for that LSOA. Where LSOA's were comprised of multiple polygons, the centroid of the largest polygon is used. These polygons are used for the production of maps seen in the Analysis Section 7.

6.6 Road KSI data

Data on those killed or seriously injured on London streets is available through the London Datastore. (London 2014). This was used to assist in determining the scope of the investigation. While it was hoped that the data could be used to build an estimate of danger to cyclists on London's streets several obstacles prevent this. The first is the change in infrastructure over time and lack of data about the exact infrastructure present at the time and location of each incident. Second was the lack of high resolution data about cyclist volumes. An area that has a particularly high KSI rate may be especially dangerous, but it may be relatively safe after adjusting for cyclist miles traveled, which is unknown. London has begun collecting some data on cyclist volumes although this remains fairly sparse.

6.7 Data import, storage, cleaning, and joining

Data import was done in Python (Rossum 1995) using the `csv`, `pandas` (McKinney 2011), `geopandas` (Jordahl 2014), `json`, and `OSMnx` (Boeing 2017) packages.

Data from OSM was converted from `JSON` to a `pandas` dataframe, tag strings parsed, edges truncated and geometry simplified. Geometry was converted to `Well-Known-Text` (WKE) format and multi-lines were broken into single line geometries for compatibility with the PostgreSQL database (Stonebraker and Rowe 1986).

Once the data was cleaned, it was passed to a PostgreSQL database using the `SQLAlchemy` package (Bayer 2010). PostGIS was used for calculating distances, associating nodes with centroids (Ramsey et al. 2005). QGIS was used to melt polygons into outer boundary and for the construction of visualizations (QGIS Development Team 2009). The excellent DBeaver database client was used for interaction with PostGreSQL (*dBeaver* 2019).

7 Analysis and Results

7.1 Scope

Scope was defined to capture the largest computationally feasible network with a simple set of rules.

The first rule was to restrict the network to "Inner London". This has the advantage of a formal designation by the GLA for each borough (London Councils 2019). The second rule was to restrict the scope to the area north of the River Thames. This has two advantages. First it further refines the study area to a higher density segment of the city with a higher portion of journeys to work done by bicycle. Second, it removes the need for a trip to cross the river via a bridge. Including bridges in the analysis would have made drawing conclusions more difficult as the ability to cross the river would have been a deciding factor in the possibility of a journey and a key determinant of the journey's distance, requiring a cyclist to go far out of their way to use one of only a dozen bridges potentially available.

Table 1 contains some data about journeys to work in the scope of the study area. The area defined captures 18% of the population, 16.8% of the working adults, and 25% of the jobs in Greater London. 8.2% of the journey's to work are contained in this area. Additionally, rates of cycling are higher in Inner London than in the periphery as seen in the 5% rate of cycling between locations within the area compared to the 2.4% London average. Finally, the area contains 25% of the traffic incidents in which a cyclist was killed or seriously injured between 1989 and 2004 (London 2014).

7.2 Defining Networks

There are three possible ways to construct a set of ways and nodes from OSM: a positive filter; a negative filter; and selecting by relation. A positive filter specifies tags that a

Mode Share Within Scope	All Modes	Bicycle	% by bicycle
Origin in scope	981,354	46,832	4.8%
Destination in scope	1,454,606	48,461	3.3%
Both in scope	479,882	24,843	5.2%
All journeys	5,852,298	140,180	2.4%

Table 1: Journeys to work by location and type

way or node must have to be included. A negative filter includes all ways and nodes without the tags specified. A relation is the OSM term for a collection of ways and nodes that belong to a set identified by an OSM contributor.

The key challenge is described by Furth, Mekuria, and Nixon (2016) as: "to measure miles of designated bike facilities can be misleading. Some designated bicycling facilities involve LTS values that most people will not tolerate."

OSM relations in London identified as being cycle routes are mapped in Figure 4. The query for selecting this set can be found in the Appendix table 2. It is clear that this is an extensive network, but the density is fairly low and for almost any trip a user would need to venture beyond the relation geometries. Further, many of the ways included in the cycling relations are in fact no different from other streets: as seen in Figure 1 a picture of Brandon Road. This has no safety improvements for cyclists, and it has been observed that speeds can far exceed the 20 mph limit. Thus, the cycle relations on their own are insufficient for building representative networks. It is not certain that Brandon Road is substantially safer than Euston Road which is not part of a cycling relation and is tagged `trunk`, depicted in Figure 2.

A good example of the difficulty of building a good network representation is Castle Baynard Street depicted in Figure 3. It connects the Central London part of Cycle Super Highway 3 with the East London section that continues out to Limehouse. This is a tunnel that serves as a bike path and as a driveway to an underground car park. A network built from the `relation[route=bicycle]` set of ways and nodes would include

tag	count	definition
bridleway	46	For horse riders
crossing	2	A crosswalk or zebra crossing
cycleway	4726	Designated Cycleway
living street	153	Pedestrians have legal priority over cars. Low speeds.
no	2	Not an official tag
path	2432	Generic, including footpaths, cycle paths, bridleways, tracks.
pedestrian	1650	Roads mainly/exclusively for pedestrians.
permissive	2	Not an official tag
primary	6668	Important roads linking larger towns.
primary link	97	link roads associated with primary roads.
residential	30021	Roads that serve housing, without connecting settlements.
road	4	A highway of unknown type.
secondary	3272	Less important than primary.
secondary link	46	link roads associated with secondary roads.
service	16474	Access roads to industrial or business parks, etc.
tertiary	5602	Less important than secondary
tertiary link	26	Link roads associated with tertiary roads.
track	492	Roads for mostly agricultural or forestry uses.
trunk	2683	Important roads that aren't motorways.
trunk link	134	Link roads associated with trunk roads
unclassified	7637	less important than tertiary. Artefact of UK road system.

Table 2: Highway tags used in OSM London

this but the street has no dedicated cycling infrastructure and is therefore tagged as `highway = unclassified`. Part of the problem is the lack of consistent tagging, it only takes one line segment missing a tag to render a significant part of cycle infrastructure useless. This also reflects the fact that getting somewhere within London nearly always requires leaving cycle infrastructure at some point and using main roads.

In the case of the relation, the Camden Hackney Quietway was examined in person. Figure 1 shows an image of Brandon Road, a part of the Quietway. This way is tagged `highway=unclassified` and connects to Agar Grove, tagged `highway = tertiary`. The unclassified marker indicates that the street is less important than a tertiary street but is neither residential or a service road. As can been seen in the image, Brandon Road has no actual cycling infrastructure. There is a tag noting the max speed is 20 mph. While speed limits could be a part of an effective estimate of safe streets for cycling the completeness of speed limit tagging in OSM is poor and the level of enforcement or observation of speed limits may be very low. It has been suggested that as many as 80% of drivers exceeded these limits (Parker 2019).

In other cases, OSM underestimates the quality of cycling infrastructure. For instance the intersection of Mile End Road and Cambridge Heath Road in the borough of Tower Hamlets is a high traffic intersection both for automobiles and for cyclists. It is an integral part of the Stratford to Aldgate cycle super highway. This intersection has been reworked to be safer for cyclists. In OSM however, it is labeled `highway=trunk`, due to its high traffic nature. It is way `7058092014`. There is also a tag `cycleway:left=lane` indicating that there is a cycle lane on the left side of the street. (OpenStreetMap contributors 2017)

A positive filter was explored to include exactly the edges and nodes that met a given level of stress for cycling. This method was challenged by the fact that a single road segment missing a tag or tagged inaccurately could disconnect a network part. Further,

index	id	compatible with OSMnx	filter text
1	u_1	yes	undirected version of d_1
2	d_1	yes	no additional filters to row 10
3	u_2	yes	undirected version of d_2
4	d_2	yes	["highway"!~"primary|primary_link|trunk|trunk_link"]
5	d_3	yes	["highway"!~"secondary|secondary_link|primary| primary_link|trunk|trunk_link"]
6	d_4	yes	["highway"!~"tertiary|tertiary_link|secondary|secondary_link|primary|primary_link|trunk|trunk_link"]
7	d_5	yes	["highway"!~"living_street—residential|tertiary|tertiary_link|secondary|secondary_link|vprimary|primary_link|trunk|trunk_link"]
8	bicycle relations	no	relation[route=bicycle]
9	primary and trunk	no	["highway"="primary—primary_link—trunk—trunk_link"]
10	filters d_1 through d_5 also include		way["highway"]["highway"!~"footway|steps|corridor|elevator|escalator|motor|proposed|construction|abandoned|platform|raceway"]["area"!~"yes"]["bicycle"!~"no"]["service"!~"private"]["access"!~"private"]
11	all cycle infrastructure	no	relation[route=bicycle]({{bbox}}); way[highway=cycleway]({{bbox}}); way[highway=path][bicycle=designated]({{bbox}});

Table 3: Filters used to define networks.

Figure 1: Camden Hackney Quietway

Figure 2: Euston Road

Figure 3: Castle Baynard Street

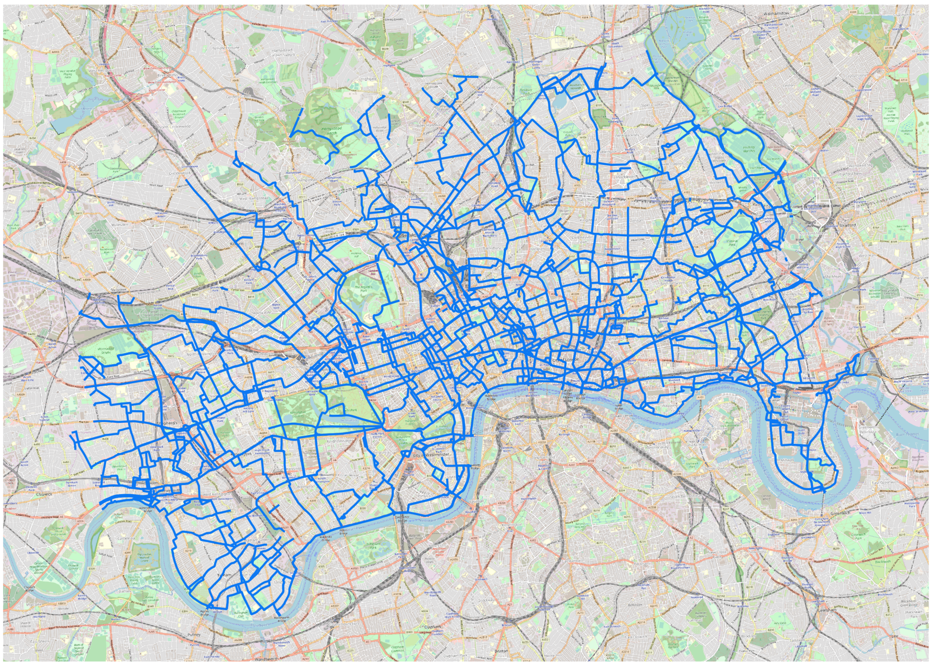

Figure 4: Relations identified as bicycle related.

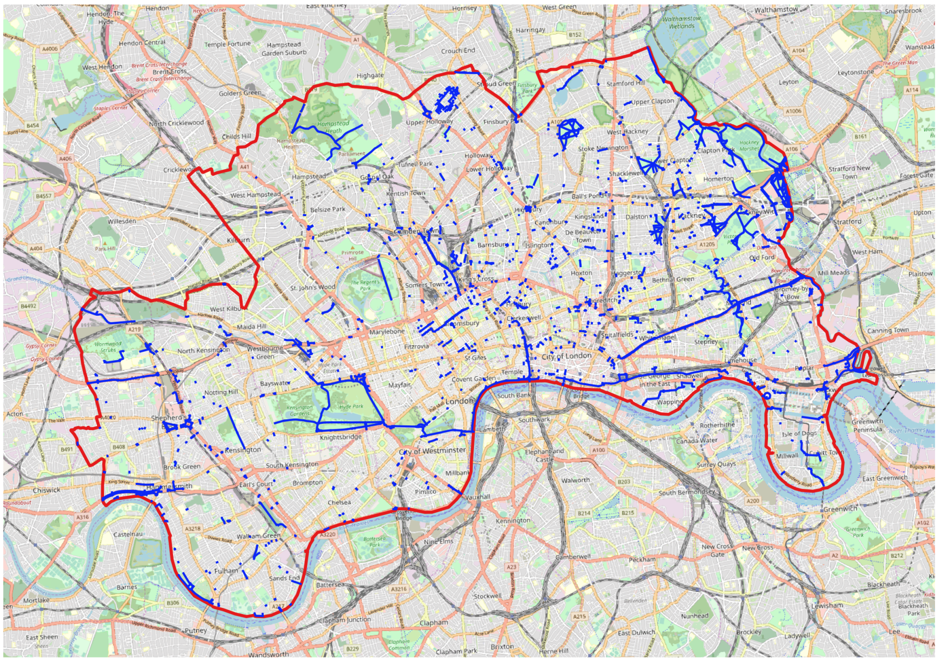

Figure 5: Highways identified as cycleways.

there was not a method found to work correctly for selecting roads with any one of many tag values. Table 2 details all the possible tags related to cycling that were found in the London OSM data. In the Overpass Turbo application, this could be accomplished by using multiple subqueries as detailed in the Overpass Turbo Cycling network query listed in row 11 of Table 3. In `OSMnx` this multiple query statement structure was not an option. Thus a method for building full and accurate networks of OSM geometries using a positive filter was not found.

The negative filter, implemented as a modification of the filters built into `OSMnx version 0.11dev`, was the most successful. The first filter included all ways and nodes not explicitly tagged with values indicating that cycling was not allowed. Row 10 of Appendix 3 contains the Overpass API filter used for this. The second filter was the same as the first but excluded streets tagged as `primary` and `trunk`, row 4 of Table 3. These are the tags used for the highest priority street types as seen in Table 2 describing standard tags

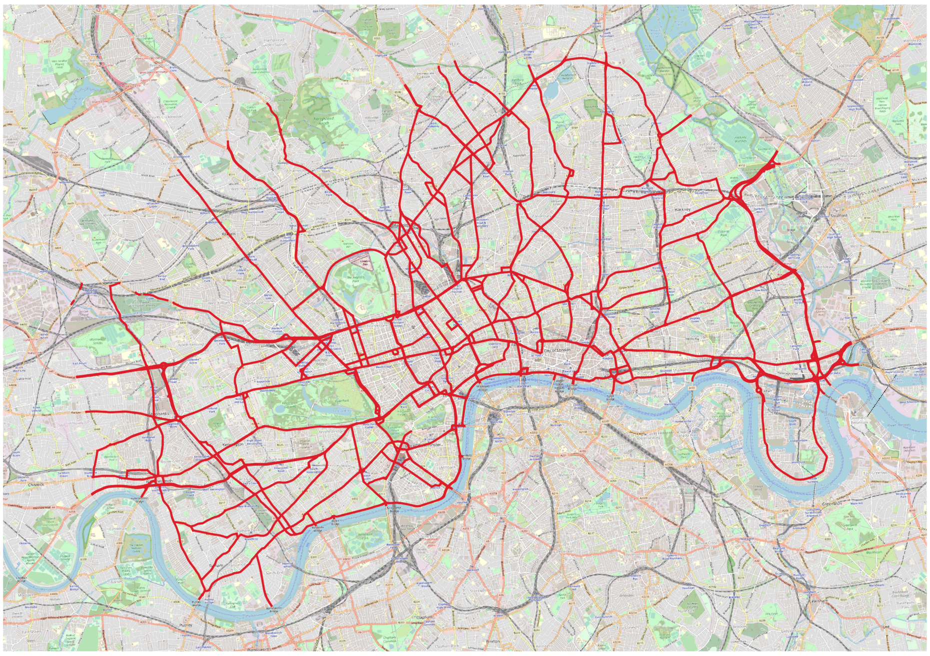

Figure 6: Highways tagged primary or trunk

for the `highway` key in OSM. These streets are visualized in Figure 6 The third filter restricted secondary streets in addition to the restrictions of the second filter. The fourth filter restricted tertiary streets leaving only living and residential streets in addition to non-motor vehicle ways like canal paths and segregated cycle lanes. The final filter, row 7, restricted all edges where a cyclist might interact with motor vehicles. Samples of these filters are displayed in Figures 7 through 10

Lastly, two more networks were specified. These were networks `u_1` and `u_2`, all streets and all streets but `primary` and `trunk`, with the directionality of the streets removed. These networks were used to test the effect of direction restrictions on travel times and whether removing direction restrictions for cyclists could successfully replace the need to use some busier and more dangerous streets.

The weakness of the negative filter was the inability to include streets tagged as multiple types. Negative filter 2 excluded any highway tagged primary regardless of whether it

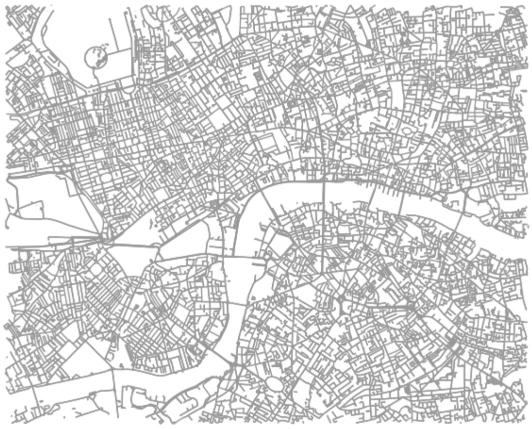

Figure 7: D1: most confident cyclists

Figure 8: D5: no interaction with cars

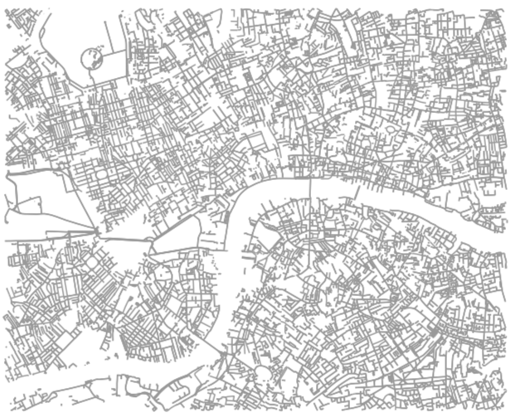

Figure 9: D2: no primary or trunk streets

Figure 10: D4: residential & living streets

was also tagged with `cycleway` or `cycleway:left=lane`. Thus the negative filters also do not fully reflect the reality of the London street network for a cyclist. However, it was the best method for testing the importance of street types and the effect of directionality on travel times.

7.2.1 Quant Network

The QUANT dataset has 23,377,225 pairs of origins and destinations. The subset matching the scope of this investigation has 799,236 pairs. That subset has 200,041 pairs missing distances where there was no connection between the origin and destination transit hubs.

To complete the missing distances, the QUANT travel times were built into a `Networkx Multidigraph`. There was an edge for each node pair that represented the walking time between the nodes, calculated as the straight-line distance between the nodes divided by the Google walking speed (3 mph converted to 4.83 kph and converted to 0.0805 kilometers per minute) giving a number of minutes of walking time between the nodes that was in the same units as the number of minutes of public transit time between the nodes.

Then for each pair of nodes missing a transit time, the shortest path on the graph was calculated. Each travel time may therefore be a combination of walking and riding public transit between nodes. The QUANT travel times augmented with the ability to walk directly between nodes as well as use public transit, are referred to below as "QUANT+". Considerations around comparing the QUANT travel times with the cycling travel times are discussed below.

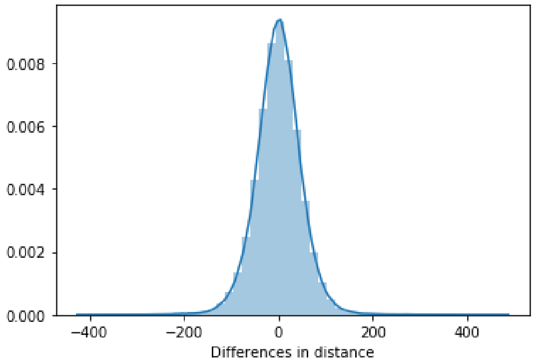

Figure 11: Node v. Centroid distances

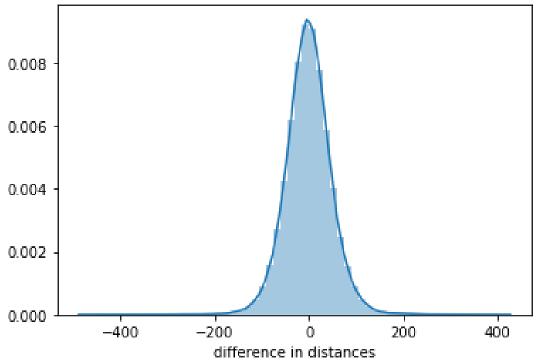

Figure 12: Node v. QUANT distances

7.3 Origins and Destinations

One origin/destination node was selected for each LSOA. Each such node was selected from the set of nodes across all network definitions that was closest to the centroid of the LSOA polygon. This is essentially a sampling technique. While individual nodes may give strange results due to the specifics of their locations, it is expected that the average results for 894 nodes that yield 798,342 O/D pairs will be a sufficiently large sample that individual idiosyncrasies balance out.

As seen in Figure 11 the distribution of the differences between the distance between centroids and the distance between actual nodes is well balanced.

Node `5816785884`, closest to the centroid of LSOA `E01004225`, is found at the entrance to a garage at the end of a one way street, making every other node in the network inaccessible on the directed versions of the networks. The edge leading to this node is tagged `service` so perhaps service streets should have been excluded as they are frequently dead ends.

Because urban density increases as one approaches the center of a city, there could be a slight bias towards node distances being lower than centroid distances. This is because there is a higher probability that the closest node will be on the central side of the

network	u_1	d_1	u_2	d_2	quant	quant+	d_3	d_4	d_5
% pairs connected	100	100	63.1	58.2	74.9	100	21.9	1.4	0.03
average directness *	1.20	1.23	1.53	1.80	-	-	-	-	-
min travel time*	0.4	0.4	0.4	0.4	0.1	-	-	-	-
mean travel time*	33.0	33.8	41.1	47.4	22.6	-	-	-	-
max travel time*	88.8	91.2	114.0	124.0	51.9	-	-	-	-
travel time std. dev*	16.8	17.1	20.4	23.6	8.6	-	-	-	-

Table 4: Network routing statistics
$*$ includes only pairs connected by all four networks.

centroid than the outside of the centroid. This was investigated and as seen in Figure 11, the distribution of differences in straight-line distances was well balanced around 0. The same consideration was checked for the differences between distances for cycling nodes and the distances between the nodes used in the QUANT calculations. As seen in Figure 12, that distribution was also well balanced around zero.

7.4 Travel Times

Travel times were converted from distances using a walking speed of 3 mph and a cycling speed of 8 mph based on data taken from Google maps estimates for journeys in London. These were converted to kilometers per minute, walking: 0.0805 and biking: 0.215.

Table 4 contains statistics about the distances and travel times calculated. To make a direct comparison, the table shows travel times for only O/D pairs that were connected for each of the four networks for which routes were calculated.

7.4.1 Street Types

Removing `primary` and `trunk` highways raises travel times by about 33%. Nearly 40% of origin destination pairs are disconnected in this scenario.

The network does not really fracture into sub-components as was found with regard to the San Jose street network in Furth, Mekuria, and Nixon (2016). Where connectivity

networks	u_1	d_1	u_2	d_2
			Δ time (mins)	
u_1		0.79	8.16	14.46
d_1	0		7.38	13.68
u_2	36.9	36.9		6.30
d_2	47.8	47.8	4.7	
	Δ % connected			

Table 5: Changes between networks, % connected and directness

goes down, it is usually because an individual node became disconnected from all others, not because a sub group was separated off. Most nodes not connected to another node, were not connected to any other nodes.

7.4.2 Directedness

The difference between directed and undirected distances is relatively small. This indicates that building multidirectional cycle infrastructure on side streets is probably not a good way to improve cyclist access.

7.4.3 Public Transit

QUANT public transit times are substantially faster than cycling times. There was no origin destination pair where cycling was found to be faster than public transport.

However, transit times are from hub to hub, so in most cases there is additional walking time required to get from the transit hub to the actual destination that would be less of a factor for a journey by bicycle. It is very possible that the methodology used to calculate the QUANT travel times allows for effective comparison of travel times between O/D pairs within the QUANT data set but is not appropriate for comparison with another separately calculated dataset like the cycling travel times calculated in this dissertation.

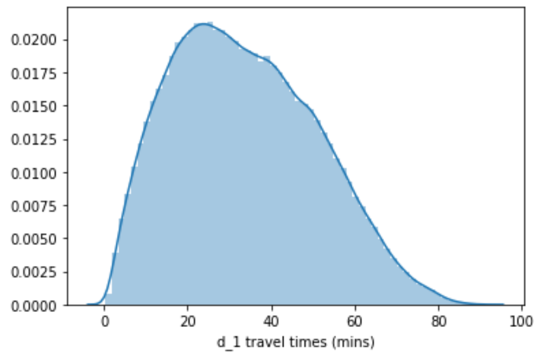

Figure 13: D1 Travel Time Distribution

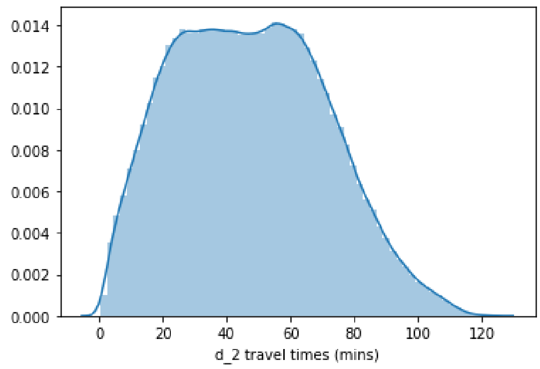

Figure 14: D2 Travel Time Distribution

7.4.4 Distributions of Travel Times

Not only are the two modes of travel different in their means, the shape of the distribution is also noticeably different. This is seen in Figures 13 through 15. The distribution of times for QUANT is noticeably more balanced than the cycling travel times. This is consistent with the option to use higher speed transit modes for longer journeys. Short journeys may be fastest by bus but for a long journey, a less direct path that uses more underground services will be faster due to the substantially higher speeds.

7.4.5 Changes in Routing

The data exhibits three types of changes in routing that increase the distance a cyclist would have to travel.

The first is a large detour added to an otherwise fairly direct path. Figure 16 shows the shortest route between Finsbury Park and the north side of Hyde Park. Removing the `primary` and `trunk` highways to produce the shortest route in Figure 17, a substantial detour through North Kensington is added, nearly doubling the total distance traveled. Furth, Mekuria, and Nixon (2016) assume that such a trip would not take place, either the cyclist would use a higher stress route or the trip would not be accomplished by bicycle. The limit they set was a 25% increase in the trip distance to avoid a high stress

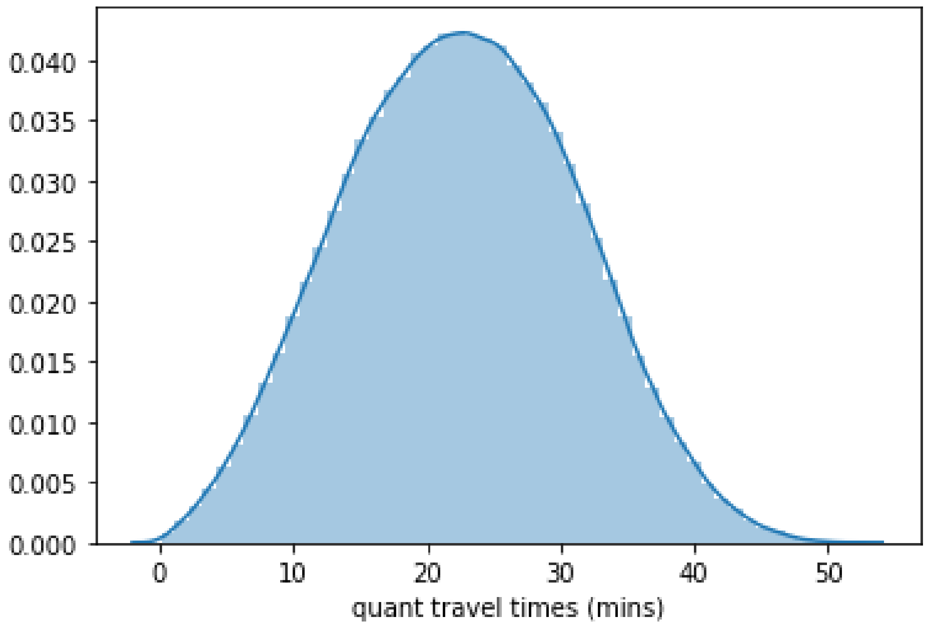

Figure 15: Distribution of QUANT public transit travel times

road.

The second type of routing change is the addition of a large number of small turns and detours throughout the trip. Consider the difference between the routing on directed network 1 in Figure 16 compared to the routing of the shortest path between the same origin and destination on directed network 2 in Figure 17, where `primary` and `trunk` streets are excluded. The network 2 route never strays very far from the network 1 route but the distance increases by 10%. Comparing by distance is informative but it is useful to keep in mind that traveling through intersections probably lowers the overall speed of a trip so that actual travel time and power expended probably increase by more than distance increases.

A third illustrative example has to do with the compounding of detours. Figure 20 shows the shortest route between Dalston and North Kensington. Avoiding `primary` and `trunk` streets increases distance by 30% as seen in Figure 21. This is because, while much of

Figure 16: Example 1, Network 1

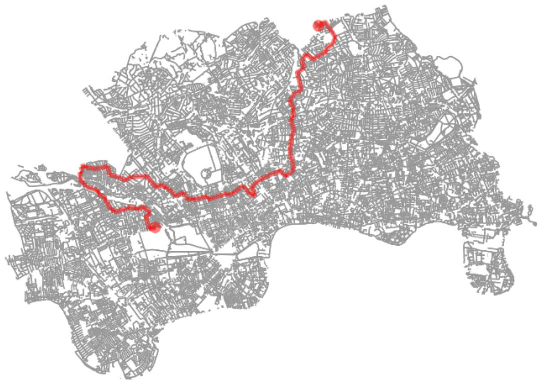

Figure 17: Example 1, Network 2

Figure 18: Example 2, Network 1

Figure 19: Example 2, Network 2

Figure 20: Example 3, Network 1

Figure 21: Example 3, Network 2

the Figure 20 route is along the Regents Canal Path, a required stretch of Euston Road between the end of the path in Angel and Regents part is a `trunk` highway. Avoiding this stretch renders the canal path useless to the route in Figure 21. Thus in true network fashion, isolated changes can have large effects on the importance of other segments of the network.

Finally, in exploring individual changes in routing, it was noticed generally that trips through the northern-most and western-most sections of the area investigated tended to be the trips with dramatic increases in distance.

7.4.6 Maps

To explore whether this trend in northern and western London was real maps were built showing the average values for each LSOA. In each map, areas within the boundary that are not colored are LSOA's that are not connected to any other LSOA. For directed network 2, this is about 100 of 900 areas.

It can be seen in Figure 22 that directness decreases substantially in the northern and western parts of the area as represented by the darker red areas there. This is consistent with the anecdotal observation that many of the larger changes in routing distance involved passing through or around these areas.

A different pattern is seen in comparing cycling time with public transport time, in Figure 23. Here Canary Wharf is relatively difficult to access by bicycle. This may be influenced by the fact that it is on the periphery of the area of investigation. It could also be influenced by public transit that travels across the river and through South London. This routing option was not available to the cycling routes as it was outside the scope of the network defined. It is expected that the quickest option for cycling to Canary Wharf would involve Mile End Road or Cycle Super Highway Three on the northern side of the river, within the scope of this analysis, and not crossing to the south side of the river

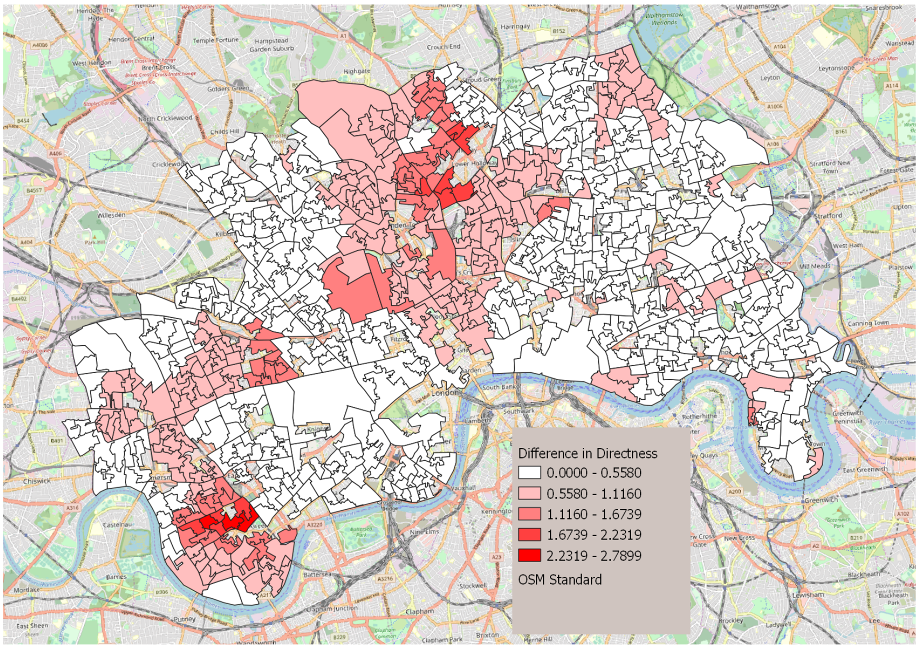

Figure 22: Difference in directness D1 and D2

(outside the scope) due to the difficulty of crossing back over at the eastern end of the river near canary wharf.

7.5 Notes About Computation

Several related works mention computational limitations to their analysis as a factor determining the scope. These sources did not provide data about their computations, which could have informed this analysis. That data is included here in the hope of assisting future research planning and the search for possible improvements in efficiency.

Runtimes for the distances between nodes were long. Computations were done on an Intel i74700HQ processor with the database contained on the internal solid state drive.

Seen in Table 6, runtimes increased with the number of connected origin destination pairs, since unconnected pairs did not require the calculation of a shortest path. Thus d_1 travel times took longer than d_2. Additionally, the undirected network calculations

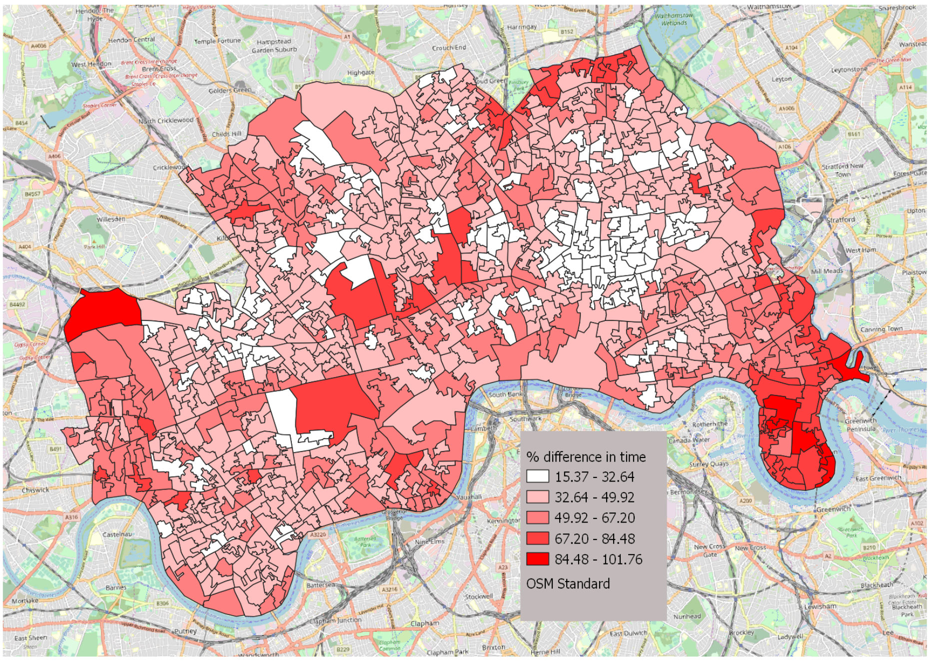

Figure 23: % Difference in travel time, QUANT+ and D1

network	1 directed	1 undirected	2 directed	2 undirected
time(hours)	24	72	9	36

Table 6: Calculation times for routes

took substantially longer than the directed networks because there were significantly more route possibilities with more edges available at each node.

Lastly, runtimes for the 200,000 missing pairs in the QUANT data were very long. This was because the walking distance between a node and all other nodes was added as an edge. This meant that each node had a degree of almost 1000. A better approach would be to only add edges where the walking distance was some reasonable value.

8 Conclusions

8.1 Results and Recommendations

The primary conclusion, consistent with personal observation around London is that cycling on unprotected parts of high volume, main streets is a required part of moving around the city by bicycle. A high percentage of trips require the use of these types of highways.

Further, a reasonable representation of the London cycling network could be constructed from Open Street Map data using `OSMnx` and `Networkx`. With some specific improvements to the software and methodology a good job could be done of estimating accessibility in London by bicycle. Access to computing resources is a key consideration. The structure of the OSM data and the structure of the Overpass API query language presented some difficulty but this was essentially solvable with a more manual process for collecting cleaning and converting the data into a network. The largest possible problem with the construction of networks using OSM data in this analysis was the potential for a complete lack of data on some feature of the street network. Given the density of the data in London this was probably not a significant concern. With the ongoing addition of the TfL Cycling Infrastructure Database, OSM is likely the highest quality source for this data available going forward.

There were significant regional differences in the effect of removing primary and trunk streets and those looking to improve the London cycling network may want to focus their efforts on western and northern London within the scope of the analysis. Removing direction restrictions for cyclists is not an effective way to improve cycling safety and the directness of travel by bicycle. This did not mediate much of the change in travel times that resulted from removing primary and trunk streets.

In contrast to the work of Furth, Mekuria, and Nixon (2016), the London network did not fracture into sub-components as the San Jose street grid did in that analysis. In the tree-like street network of London decreases in connectivity came through the complete separation of single nodes from the network. A node that lost connectivity with any other node nearly always lost connectivity with all other nodes. Thus while Furth et al's study focused on connecting disconnected components, improvements to London's network should perhaps be targeted toward reducing trip times by improving the directness of safe routes.

8.2 Limitations

Selection of origin and destination locations was an important determinant of the results of this research. QUANT uses the transit station of highest degree as the origin and destination points, meaning that there is no walking time included in the calculation. Additionally, QUANT travel times are probabilistic in the sense that the precise start time affects the travel time because there may be a wait time for a bus or train. This is a key advantage that cycling has over public transport, a more linear relationship between distance and travel time. In one instance, the selection of a node at the end of a one way service street meant that it was connected to another node only on the undirected networks. Some LSOA's that were disconnected from the network after the removal of edges continued to contain nodes that were connected to the network. Thus the selection criteria could have a significant impact on the outcomes and could possibly be improved.

The network representations could also be improved. `OSMnx` makes working with the Overpass API easier by abstracting away a lot of the process. Unfortunately this ease comes with a cost, a diminished ability to build an accurate network using all of the possible attributes of OSM geometries. A key feature for `OSMnx` that would improve the ability to define networks accurately would be the option to "stack" multiple types of requests into a full network. This would enable the collection of several sets of edges and nodes with several filters, as can be accomplished with the Overpass Turbo OSM tool (*Overpass Turbo* n.d.).

It was subsequently found that Google speed estimates for cycling can vary dramatically and the 13 kph speed used may be fairly conservative. Further tests of Google Maps speed estimates being closer to 15 kph. This does not change the overall indication of the data that public transit is faster than cycling, but the magnitude may not be as great as indicated by the estimates. An empirical study of observed cycling speeds would be the optimal way to settle this question.

Lastly, this dissertation was unable to adjust for the demand for travel between O/D pairs because of the lack of access to high resolution journey to work data. While this is unfortunate, the differences in connected pairs and in average travel times between different network definitions are so large that it is highly unlikely that adjusting for relative travel demand would significantly alter the conclusions.

8.3 Opportunities for Improvement and Extension

Beyond improving the implementation of the methodology used in this dissertation, the key improvement to the methodology generally would be to move from using OSM tags for highway importance, which is a discrete set of values, to a continuous estimation of the stress for a given street based on all of the possible data. As found in the Section 4 review of literature, there are not clear methods for addressing this estimation problem because of the nature of the data and the difficulty of measuring cyclist volumes at a high resolution. To date, no methodology has implemented a continuous estimation although Boisjoly, Lachapelle, and El-Geneidy (2019) approach a basic version of this by measuring the portion of a trip completed on designated cycle infrastructure.

While removing edges from the network did have a significant effect on connectivity, reducing the connected origin/destination pairs by 40%, the study showed that less confident cyclists have route options available to them if they are willing to go out of their way along less direct paths. The phenomenon referred to as "rat-running", where traffic filters off main routes onto side roads is a key reason that OSM highway types may not actually be good indications of traffic stress or danger. A "quietway" may get relatively more traffic than it is built to handle relative to primary routes if the quietway is a good route between destinations. This is consistent with anecdotal observations of the author that on some quiet ways traffic behavior is more dangerous and less predictable than on primary routes. It is not clear that dense slow moving traffic is more dangerous than sparse high speed traffic on back roads. If this was found to be a significant consideration in quietways it would suggest that adding "gates" that allow cyclists but not motor vehicles to pass through may be preferable to other cycling infrastructure for improving the safety of that route.

Several data sets could potentially augment this type of analysis further. London has several decades of traffic incident data with spatial data for each incident. This could possibly provide for validation of the high stress nature of OSM highways tagged primary and trunk. With the beginning of collection on cyclist volumes in some parts of the city, this could become possible in the future. Additionally, high resolution journey to work data for London exists but is not publicly available due to privacy concerns. Access to this data would allow for a connectivity ratio like that of Furth, Mekuria, and Nixon (2016), which adjusts the percentage of possible trips of a reasonable length by the number of commuter trips that are associated with that origin/destination pair.

8.4 Conclusions

Based upon this analysis one can conclude that there is still plenty of room for improvement for the London Cycle Network. Particularly in the northern and western sections of inner London, many journeys require large detours or the use of high volume main streets. Looking forward, future analyses can consider to what extent this conclusions changes as the result of more accurate data and methodologies as well as to what extent this changes as the result of the Mayor of London's program to improve and extend the network of cycling infrastructure across Greater London. Overall it is gratifying to see data methodology and real world conditions improving the ability of urban commuters to rely on cycling as a primary mode of transit.

References

(GLA), Greater London Authority (2014). *LSOA Atlas.* retrieved from data.london.gov.uk, `https://data.london.gov.uk/dataset/lsoa-atlas`.

Akbarzadeh, Meisam, Syed Sina Mohri, and Ehsan Yazdian (2018). "Designing bike networks using the concept of network clusters". In: *Applied network science* 3.1, p. 12.

Bayer, Mike (2010). "SQLAlchemy Documentation". In:

Boeing, Geoff (2017). "OSMnx: New Methods for Acquiring, Constructing, Analyzing, and Visualizing Complex Street Networks". In: *Computers, Environment and Urban Systems* 65, pp. 126–139.

Boisjoly, Geneviève, Ugo Lachapelle, and Ahmed El-Geneidy (2019). "Bicycle network performance: Assessing the directness of bicycle facilities through connectivity measures, a Montreal, Canada case study". In: *International Journal of Sustainable Transportation*, pp. 1–15.

Buehler, Ralph and Jennifer Dill (2016). "Bikeway networks: A review of effects on cycling". In: *Transport Reviews* 36.1, pp. 9–27.

dBeaver (2019). Version 6.1.4. retrieved from `https://github.com/dbeaver/dbeaver`.

Dill, Jennifer and Nathan McNeil (2013). "Four types of cyclists? Examination of typology for better understanding of bicycling behavior and potential". In: *Transportation Research Record* 2387.1, pp. 129–138.

Doorley, Ronan et al. (2019). "Designing cycle networks to maximize health, environmental, and travel time impacts: An optimization-based approach". In: *International Journal of Sustainable Transportation*, pp. 1–14.

Ellett, Lyn, Jess Kingston, and Paul Chadwick (2018). "State paranoia and urban cycling". In: *Psychiatry research* 266, pp. 341–344.

Fajans, Joel and Melanie Curry (2001). "Why bicyclists hate stop signs". In:

Furth, Peter G, Maaza C Mekuria, and Hilary Nixon (2016). "Network connectivity for low-stress bicycling". In: *Transportation research record* 2587.1, pp. 41–49.

Gosse, C Alec and Andres Clarens (2014). "Estimating spatially and temporally continuous bicycle volumes by using sparse data". In: *Transportation Research Record* 2443.1, pp. 115–122.

Gu, Jing, Babak Mohit, and Peter Alexander Muennig (2017). "The cost-effectiveness of bike lanes in New York City". In: *Injury prevention* 23.4, pp. 239–243.

Hagberg, Aric, Pieter Swart, and Daniel S Chult (2008). *Exploring network structure, dynamics, and function using NetworkX.* Tech. rep. Los Alamos National Lab.(LANL), Los Alamos, NM (United States).

Haklay, Mordechai (2010). "How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets". In: *Environment and planning B: Planning and design* 37.4, pp. 682–703.

Hochmair, Hartwig H et al. (2013). "Assessing the completeness of bicycle trails and designated lane features in OpenStreetMap for the United States and Europe". In: *Transportation Research Board Annual Meeting.*

Jordahl, K (2014). "GeoPandas: Python tools for geographic data". In: *URL: https://github.com/geopandas/geopandas.*

Juhasz, Aubri (Aug. 2019). "In NYC, Cycling Deaths Increase But Gears Turn Slowly On Safety Measures". In: *All Things Considered.* URL: `https://www.npr.org/2019/08/14/751218425/in-nyc-cycling-deaths-increase-but-gears-turn-slowly-on-safety-measures`.

Klobucar, Michael S and Jon D Fricker (2007). "Network evaluation tool to improve real and perceived bicycle safety". In: *Transportation Research Record* 2031.1, pp. 25–33.

Kondo, Michelle C et al. (2018). "Where do bike lanes work best? A Bayesian spatial model of bicycle lanes and bicycle crashes". In: *Safety science* 103, pp. 225–233.

Li, Zhibin et al. (2012). “Physical environments influencing bicyclists’ perception of comfort on separated and on-street bicycle facilities”. In: *Transportation Research Part D: Transport and Environment* 17.3, pp. 256–261.

London Councils (2019). *List of inner/outer London Boroughs*. retrieved from londoncouncils.gov.uk, `https://www.londoncouncils.gov.uk/node/1938`.

London, Transport for (2014). *Pedal Cyclist Casualties, Killed and Seriously Injured.* data retrieved from data.london.gov.uk, `https://data.london.gov.uk/dataset/pedal-cyclist-casualties-killed-and-seriously-injured`.

— (2019). *Cycling Infrastructure Database*. data retrieved from data.london.gov.uk, `https://data.london.gov.uk/dataset/cycling-infrastructure-database`.

Lowry, Michael B et al. (2012). “Assessment of communitywide bikeability with bicycle level of service”. In: *Transportation research record* 2314.1, pp. 41–48.

Mauttone, Antonio et al. (2017). “Bicycle network design: model and solution algorithm”. In: *Transportation research procedia* 27, pp. 969–976.

Mayor (2017). *Mayor and Commissioner set out vision for getting Londoners active.* `https://www.london.gov.uk/press-releases/mayoral/setting-out-a-vision-for-getting-londoners-active`. Accessed: 2019-08-24.

McKinney, Wes (2011). “pandas: a foundational Python library for data analysis and statistics”. In: *Python for High Performance and Scientific Computing* 14.

Mekuria, Maaza C, Peter G Furth, and Hilary Nixon (2012). “Low-stress bicycling and network connectivity”. In:

Mobasheri, Amin et al. (2017). “Are crowdsourced datasets suitable for specialized routing services? Case study of OpenStreetMap for routing of people with limited mobility”. In: *Sustainability* 9.6, p. 997.

National Statistics, Office for (2011). *Location of usual residence and place of work by method of travel to work, UK Census 2011.* data retrieved from nomisweb.co.uk, `https://www.nomisweb.co.uk/census/2011/wu03uk`.

National Statistics, Office for (2016). *Lower Layer Super Output Area (LSOA) boundaries*. data retrieved from data.gov.uk, https://data.gov.uk/dataset/fa883558-22fb-4a1a-8529-cffdee47d500/lower-layer-super-output-area-lsoa-boundaries.

— (2019). *Nomis, Official Labour Market Statistics*. retrieved from https://www.nomisweb.co.uk.

OpenStreetMap (2019). *OSM Wiki page for TfL Cycling Infrastructure Database*. retrieved from wiki.openstreetmap.org, https://wiki.openstreetmap.org/wiki/TfL_Cycling_Infrastructure_Database.

OpenStreetMap contributors (2017). *Planet dump retrieved from https://planet.osm.org*. https://www.openstreetmap.org.

Osama, Ahmed, Tarek Sayed, and Alexander Y Bigazzi (2017). "Models for estimating zone-level bike kilometers traveled using bike network, land use, and road facility variables". In: *Transportation Research Part A: Policy and Practice* 96, pp. 14–28.

Overpass Turbo (n.d.). https://overpass-turbo.eu/. Accessed: 2019-08-24.

Parker, Nick (2019). "20mph Zones No-Go". In:

Parkin, John, Mark Wardman, and Matthew Page (2007). "Models of perceived cycling risk and route acceptability". In: *Accident Analysis & Prevention* 39.2, pp. 364–371.

Puchades, Víctor Marín et al. (2018). "The role of perceived competence and risk perception in cycling near misses". In: *Safety science* 105, pp. 167–177.

QGIS Development Team (2009). *QGIS Geographic Information System*. Open Source Geospatial Foundation. URL: http://qgis.org.

QUANT (2019). retrieved from http://quant.casa.ucl.ac.uk/.

Ramsey, Paul et al. (2005). "Postgis manual". In: *Refractions Research Inc* 17.

Rossum, Guido (1995). *Python Reference Manual*. Tech. rep. Amsterdam, The Netherlands, The Netherlands.

Savan, Beth, Emma Cohlmeyer, and Trudy Ledsham (2017). “Integrated strategies to accelerate the adoption of cycling for transportation”. In: *Transportation research part F: traffic psychology and behaviour* 46, pp. 236–249.

Schoner, Jessica E and David M Levinson (2014). “The missing link: Bicycle infrastructure networks and ridership in 74 US cities”. In: *Transportation* 41.6, pp. 1187–1204.

Stinson, Monique A, Chandra R Bhat, et al. (2005). “A comparison of the route preferences of experienced and inexperienced bicycle commuters”. In: *TRB 84th annual meeting compendium of papers*. 05-1434.

Stonebraker, Michael and Lawrence A Rowe (1986). *The design of Postgres*. Vol. 15. 2. ACM.

Vandenbulcke, Grégory, Isabelle Thomas, and Luc Int Panis (2014). “Predicting cycling accident risk in Brussels: A spatial case–control approach”. In: *Accident Analysis & Prevention* 62, pp. 341–357.

Winters, Meghan et al. (2011). “Motivators and deterrents of bicycling: comparing influences on decisions to ride”. In: *Transportation* 38.1, pp. 153–168.

World Urbanization Prospects (2018). `https://population.un.org/wup/Publications/Files/WUP2018-KeyFacts.pdf`. retrieved:26 August 2019.

Wuerzer, Thomas and Susan G Mason (2015). “Cycling willingness: Investigating distance as a dependent variable in cycling behavior among college students”. In: *Applied geography* 60, pp. 95–106.

Zielstra, Dennis, Hartwig H Hochmair, and Pascal Neis (2013). “Assessing the Effect of Data Imports on the Completeness of OpenStreetMap A United States Case Study”. In: *Transactions in GIS* 17.3, pp. 315–334.