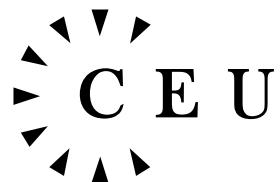


# **Multimodality and sustainability in urban networks**



**Luis Guillermo Natera Orozco**

Department of Network and Data Science  
Central European University

Supervisor: Federico Battiston  
External Supervisor: Michael Szell

A Dissertation Submitted in Partial Fulfillment of the Requirements  
for the Degree of Doctor of Philosophy in Network Science

2021

Luis Guillermo Natera Orozco: *Multimodality and sustainability in urban networks*, ©  
2021  
All rights reserved.

## RESEARCHER DECLARATION

---

I Luis Guillermo Natera Orozco certify that I am the author of the work Multimodality and sustainability in urban networks. I certify that this is solely my own original work, other than where I have clearly indicated, in this declaration and in the thesis, the contributions of others. The thesis contains no materials accepted for any other degrees in any other institutions. The copyright of this work rests with its author. Quotation from it is permitted, provided that full acknowledgement is made. This work may not be reproduced without my prior written consent.

### **Statement of inclusion of joint work**

I confirm that Chapter 2 is based on a paper which was written in collaboration with Laura Alessandretti, Michael Szell and Federico Battiston. Dr. Battiston and I concived the idea of doing a review about the application of multiplex networks to the study of urban mobility systems. All authors contributed to the writing of the paper. Dr. Battiston endorses this statement with his signature below.

I confirm that Chapter 3 is based on a paper which was written in collaboration with Federico Battiston, Gerardo Iñiguez, and Michael Szell. I, with Dr. Battiston and Dr. Szell concived the idea of the overlap census. I collected the data and carried out the analyses. All authors developed the methods used, contributed to the writing of the paper on which the chapter is based and gave final approval for publication. Dr. Szell endorses this statement with his signature below.

I confirm that Chapter 4 is based on a paper which was written in collaboration with Federico Battiston, Gerardo Iñiguez, and Michael Szell. I, with Dr. Battiston and Dr. Szell concived the research. I collected, processed and cleaned data, and carried out the computational analysis. All authors designed the algorithms, analysed the data and results, and helped draft the manuscript. Dr. Szell endorses this statement with his signature below.

I confirm that Chapter 5 is based on a paper which was written in collaboration with Dávid Deritei, Anna Vancsó, and Orsolya Vásárhelyi. I with Dr. Deritei and Dr. Vásárhelyi concived the research and methodology, all authors collected, processed and cleaned data, and carried out the computational analysis. Together with Dr. Deritei and Dr. Vásárhelyi we analyzed the results. All author helped draft the manuscript and gave final approvement for its publication. Dr. Vásárhelyi endorses this statement with her signature below.

Signature of PhD Candidate:

February, 2021

Signature of Dr. Federico Battiston, endorsing statement of joint work:

February, 2021

Signature of Dr. Michael Szell, endorsing statement of joint work:

February, 2021

Signature of Dr. Orsyola Vásárhelyi , endorsing statement of joint work:

February, 2021



## ABSTRACT

---

In this thesis, I apply network science methods to the study of cities and their mobility infrastructure. The general aim of this work is to provide new insights about the underlying urban mobility infrastructure, to better understand it and propose methodologies that can improve the sustainable mobility options in cities. To achieve that, I work with network science and computational methods to unravel the city multilayer mobility network, first to study it as a whole, and then I switch the focus to the active mobility layers.

Each chapter tackles different questions related to the multimodal mobility options. First, I cover the existing literature in the domain of multimodal urban network, spanning from the infrastructure and dynamics to the available measures, tools and data to work with such phenomena. Second, I propose a new method to extract the multimodal profile from a city's multiplex transport network, and show how it can be applied to identify similarities between cities. Third, I zoom-in into the bicycle layer of the multimodal network, to show how fragmented it is, and propose an algorithmic approach to find the missing links. After applying the algorithms to different cities I show that a focussed investment, can have a significant impact improving the bicycle infrastructure connectivity. Moreover, I present a data-driven, network-based method to quantify the liveability of a city based on pedestrian accessibility to amenities and services, safety and environmental variables. Finally, I discuss the main contributions of this thesis and outline possible applications and open questions for the future.



## ACKNOWLEDGEMENTS

---

To be written



# CONTENTS

---

<b>Contents</b>	<b>i</b>
<b>List of Tables</b>	<b>iii</b>
<b>List of Figures</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Related Work</b>	<b>3</b>
2.1 Introduction . . . . .	3
2.2 Multilayer networks: a framework for multimodality . . . . .	4
2.3 Multimodal infrastructures . . . . .	6
2.3.1 Modeling urban infrastructure . . . . .	7
2.3.2 Characterizing multimodal infrastructure . . . . .	10
2.4 Multimodal mobility . . . . .	15
2.4.1 Modeling urban mobility . . . . .	16
2.4.2 Characterizing multi-modal mobility. . . . .	18
2.5 Open data & Tools . . . . .	23
2.5.1 Data . . . . .	23
2.5.2 Tools . . . . .	24
<b>3 Extracting the multimodal fingerprint for transportation networks</b>	<b>27</b>
3.1 Introduction . . . . .	28
3.2 Data . . . . .	28
3.3 A local measure of multimodal integration . . . . .	30
3.4 Identifying clusters of cities with the Overlap Census . . . . .	31
<b>4 Data-driven strategies for optimal bicycle network growth</b>	<b>33</b>
4.1 Introduction . . . . .	34
4.2 Data . . . . .	35
4.3 Defining bicycle network growth strategies and quality metrics .	36

4.3.1	Algorithmic approach . . . . .	37
4.3.2	Evaluating the algorithms impact . . . . .	39
4.4	Growing bicycle networks shows stark improvements with small investments . . . . .	42
4.5	Different cities have different optimal investment strategies . . . . .	44
4.6	Discussion . . . . .	48
<b>5</b>	<b>Life quality as walkability</b>	<b>51</b>
5.1	Introduction . . . . .	52
5.2	Data . . . . .	53
5.3	Quantifying Life Quality . . . . .	54
5.3.1	The services index: $Q^{\text{services}}$ . . . . .	56
5.3.2	Safety index: $Q^{\text{safety}}$ . . . . .	58
5.3.3	Environmental index $Q^{\text{environment}}$ . . . . .	58
5.4	Results . . . . .	59
5.4.1	Evaluation . . . . .	60
5.5	Discussion . . . . .	62
<b>6</b>	<b>Conclusion</b>	<b>63</b>
<b>7</b>	<b>Appendices</b>	<b>85</b>
7.1	Life quality as walkability . . . . .	85
7.1.1	Secondary data sources . . . . .	85
7.1.2	Categorization . . . . .	86
7.2	Data-driven strategies for optimal bicycle network growth . . . . .	86
7.2.1	Data . . . . .	86
7.2.2	Bicycle network improvement . . . . .	87
7.2.3	Bicycle network and 30 km/hr streets . . . . .	88

## LIST OF TABLES

---

3.1	Layers measures for analyzed cities . . . . .	29
4.1	Measures for analyzed cities . . . . .	36



## LIST OF FIGURES

---

3.1	Manhattan multiplex network . . . . .	29
3.2	Schematic overlap census . . . . .	30
3.3	Overlap Census clusters . . . . .	32
4.1	Multimodal configuration . . . . .	35
4.2	Algorithms schematic representation . . . . .	41
4.3	Algorithmic improvement in Budapest . . . . .	43
4.4	Normalized increase in kilometers inside the largest connected component ( $\ell_{LCC}$ ) . . . . .	44
4.5	Cities bicycle connectivity improvement 5 new kilometers . . . . .	45
4.6	Cities bicycle connectivity improvement 35 new kilometers . . . . .	47
4.7	Bicycle infrastructure service area . . . . .	48
5.1	Budapest pedestrian network . . . . .	55
5.2	Budapest neighborhoods life quality index . . . . .	60
5.3	Budapest life quality index correlation with real estate prices . . . . .	61
7.1	Connected component size distribution for analyzed cities . . . . .	86
7.2	Normalized increase in nodes inside the largest connected component ( $n_{LCC}$ ) . . . . .	87
7.3	Normalized increase in kilometers inside the largest connected component ( $\ell_{LCC}$ ) . . . . .	88
7.4	Normalized increase in kilometers inside the largest connected component ( $\ell_{LCC}$ ) versus the fraction of extant kilometers to be added. . . . .	89
7.5	Bike-car directness $\Delta$ per invested kilometers. . . . .	90
7.6	Kilometers gain in the largest connected component. . . . .	91
7.7	Normalized increase in nodes inside the largest connected component ( $n_{LCC}$ ) . . . . .	92
7.8	Bike-car directness $\Delta$ per invested kilometers. . . . .	93
7.9	Kilometers gain in the largest connected component. . . . .	94



# CHAPTER 1

---

## INTRODUCTION

---

With more than 50% of the human population living in cities, solving their challenges have become a pri

solving the challenges of urban mobility has become one of the key requisites of ensuring urban sustainability. The evolution of urban agglomeration has been accompanied by new mobility options that have enabled to cover larger distances evolving from horsecars, to bicycles, trams, cars and nowadays new micromobility options.

As urban agglomerations have grown, in the 20th century

The car-centric design of cities has fundamental implications on human health and on the efficient organization of society, such as air pollution and road fatalities, urban sprawl, and on the spread of epidemics. The complex system of urban transport infrastructure is shaping the dynamics of human mobility, a network embedded in space with special properties. The availability of data sets on urban infrastructure networks has enabled first insights on the topological constraints on human movements. Large-scale mobile phone data sets, on the other hand, have been used as proxies to understand mobility patterns in cities, finding predictable human mobility motifs related to socio-economic activities in cities. The use of such massive data sets has also provided insights into the possibilities of shareability, and into the efficient allocation of space for different types of transportation.

Despite these first efforts, the interplay between infrastructure networks and mobility flows is still unknown. From the perspective of network science, both are multilayer networks. 1) Multilayer infrastructure networks are the set of networks of physical infrastructures dedicated to multiple mobility options and the connections between those networks. 2) Multimodal mobility is the use of different transportation options in a city, from walking, personalized or collec-

tive mobility to motorized mobility. Here we propose to explore multimodal mobility and the underlying infrastructure networks following a network science approach to understand the effects of the network on the mobility choices. The research goal is to quantify topological properties such as the redundancy of specific infrastructure networks, their completeness and their impact on urban mobility.

### **Aim and structure of the thesis**

The primary aim of this thesis is to contribute to the better understand of the structural properties of multimodal transportation networks in urban areas. For that, I build on the tools and methods of network science to analyze the underlying complexity of urban mobility infrastructure. After analyzing the multimodal infrastructure, I propose new network science based methods to improve the connectivity of bicycle infrastructure, and develop a methodology to measure the life quality in terms of pedestrian accessibility.

This thesis is structured as follows:

- Chapter 2: We review previous work on multimodal transportation and mobility research from the complex system perspective.
- Chapter 3: We propose a method to extract the multimodal profile from a city multiplex transport network, and apply it to fifteen cities, finding clusters of cities with similar multimodal infrastructure.
- Chapter 4: In this chapter, we focus our attention in the bicycle layer of the multimodal network, and propose algorithmic approaches to improve its connectivity.
- Chapter 5: We present a methodology to measure the quality of life in a city based on the pedestrian accessibility to amenities and services. We applied the methodology to Budapest and show how it can be used to capture inequalities in neighborhoods.

# CHAPTER 2

---

## RELATED WORK

---

In this chapter we provide a comprehensive overview of complex systems approaches to multimodal transportation and mobility in urban networks. We cover multiple approaches that had been used to model cities using the multiplex framework, measures to quantify centralities and changes in the given networks and layers. After reviewing the mobility infrastructure and measures, we move our focus to the mobility dynamics on top of those layers, covering the study of urban mobility dynamics in public transport systems and at individual level. Finally we conclude the chapter with an overview of the available data and computational tools for the study of urban mobility networks<sup>1</sup>.

### 2.1 Introduction

In this first chapter we offer a comprehensive review of multimodal transportation and mobility research focusing on recent complex systems approaches. In such approaches, the city is studied as a complex system [1, 2], in which especially urban transport infrastructure, such as streets, sidewalks, bicycle lanes and public transportation systems can be well modelled and understood using methods from network science. From this perspective, single-layer spatial networks, especially transportation networks, have been widely studied [3, 4, 5], finding different topological properties [6, 7, 8, 9, 4, 10, 11, 12], distribution of centrality metrics [13, 14, 15], and network growth processes [16, 17]. Further topics studied include impacts of the street networks on pedestrian volume [18], accessibility and vitality of cities [19, 20, 21], and resilience and growth

---

<sup>1</sup>A stand alone of this chapter is currently being prepared as a paper, to be submitted to Transportation Reviews. No preprint is yet available

of different transportation networks [22, 23, 24].

The most recent of these approaches can be seen as the beginning of the emerging field of Urban Data Science, which exploits large-scale new urban data sets with tools combining geoinformatics, data and network science [25, 26].

In this section we focus on the combined use of single-layer networks as multilayer networks to characterize the multimodal transportation infrastructure of cities and the human mobility taking place on them. We follow the primal approach to networks [27], where streets and mobility infrastructure constitute the network links, and intersections (bus stops, subway stations, etc.) constitute the nodes of the network.

The remainder of this section is arranged as follows. Section 2.2 introduces the mathematical concept of multilayer networks and related theoretical research underlying network science approaches to the topic. In Section 2.3 we cover urban infrastructures, measures to quantify their multimodality and models to reproduce the interconnected structures of real-world transportation systems. In Section 2.4 we focus on mobility, flows and navigation across these multimodal systems, and implications for transportation choices. In Section 2.5 we cover the relevant open datasets and the main software tools which can be used to analyse multimodal transportation systems.

## 2.2 Multilayer networks: a framework for multimodality

Over the last decades, networks have emerged as a versatile tool to understand, map and visualise the interconnected architecture of a wide range of complex systems [28, 29, 30, 31], in particular spatially-embedded ones [4]. Formally, a network – or graph –  $\mathcal{G} = (\mathcal{N}, \mathcal{L})$  consists of a set of nodes  $\mathcal{N}$ , and a second set  $\mathcal{L}$  of edges, describing connections among unordered pairs of elements of the first set. This information can be conveniently stored into an adjacency matrix  $A = a_{ij}$ , where  $i = 1, \dots, N$  are the nodes, and  $a_{ij} = 1$  if there is a link between nodes  $i$  and  $j$ ,  $a_{ij} = 0$  if there is no link between  $i$  and  $j$ . In transportation systems [3], nodes can represent the stations of a network, and links direct connections between them. The adjacency matrix can also include weights  $W = w_{ij}$ , where  $w_{ij}$  are positive real numbers, for instance describing how strongly connected two nodes are. For spatial systems, weights are often taken as the reciprocal of the distance between two nodes, or the time it takes to travel from one to another, i.e.  $w_{ij} = 1/d_{ij}$  or  $w_{ij} = 1/t_{ij}$ .

More recently, network scientists have put a lot of effort in characterising the

structure of systems which are formed by different interconnected networks. Also widespread in social and biological networks, these structures are natural for transportation systems. Think for instance of the largest transportation hubs in worldwide cities, where stations are routinely served by bus, underground and railway infrastructures.

Indeed, most urban transportation systems systemically rely on the interplay between different mean of transportation. These systems can be conveniently described by *multiplex* or *multilayer* networks. Here we introduce the so-called *vectorial* formalism for multilayer networks [32, 33], widely used in most papers on multimodal transportation. We note that an alternative description can be provided by a more mathematically involved *tensorial* framework [34, 35].

In multilayer networks, links of different types, describing for instance a different mean of transportation, are embedded into different *layers*. Each layer  $\alpha$ ,  $\alpha = 1, \dots, M$ , is described by an adjacency matrix  $W^{[\alpha]} = w_{ij}^{[\alpha]}$ . In a multimodal urban transportation networks with three layers,  $\alpha = 1$  can represent the bus network,  $\alpha = 2$  the underground network, and  $\alpha = 3$  the urban railway network. The full transportation system  $\mathcal{M}$  can be described as  $\mathcal{W} = \{W^{[1]}, \dots, W^{[M]}\}$ . Nodes  $i = 1, \dots, N$  are labeled in the same order in all networks.

In the case of transportation networks, identifying nodes of different networks (urban location) as the same station might not provide the most complete description of the multimodal network. Think for instance of the largest stations in mega-cities, like King's Cross - St. Pancras in London, Grand Central Station in New York or Hongqiao transportation hub in Shanghai. All of those are identified by a unique location (node index)  $i$  across the different transportation layers. However, sometimes switching from one mean of transportation to another within the same station might require a non negligible fraction of time and effort, given the complexity and size of the overall infrastructure.

For this reason, it is often relevant to complement the description of the *intra-layer* connections present in the system, with *inter-layer* links associated to the cost, physical distance or time required to switch layers. Inter-layer links between layers  $\alpha$  and  $\beta$  at a node  $i$  can be encoded through the inter-layer matrix  $C_i = c_i^{[\alpha\beta]}$ , and all such inter-layer connections can be stored in the vector  $C = \{C_1, \dots, C_N\}$ . In this case, the full multiplex structure of the system is described by taking into account both intra-layer and inter-layer connectivity, hence  $\mathcal{M} = (W, C)$ . Inter-links may be neglected for many measures focusing on diversity [33], as well as correlations [36] across the layers of the systems, relevant to assess the different roles and geographical spanning of the different mean of transportations of a multilayer network.

Multilayer networks are a natural framework for multimodal networks. Indeed, one of the pioneering works introducing the framework and concept of ‘layered complex networks’ [37] explicitly focused on the case of transportation systems, where a first layer encoded the physical infrastructure of the system, and the second one described the flows on such infrastructure. Other early works on the topic also dealt with interconnected systems at the worldwide level, focusing on different modes of transport such as the multiplex airline networks [38].

Noticeably, multimodal infrastructures seem to possess exclusive characteristics, different from other multilayer networks. For instance, when tools to assess the redundancy of the different layers are considered, transportation networks are often found to be irreducible [39]. Differently from many biological systems, where layers often duplicate information to guarantee the interconnected system a high level of robustness, the layers of a multiplex transportation systems are purposefully engineered to be different, in order to maximise efficiency [40]. As a byproduct of this feature, multimodal systems are also often highly fragile [41], and sensitive to disruptions or failures of a single infrastructure [42]. For the reader interested in further material on the topic, we refer to the early reviews [32, 35] and textbook [43] covering the field. Ref. [44] provides a more recent eye-bird view of the field. A thorough review of the measures and models used to analyse such systems can be found in Ref. [45], whereas Ref. [46] gives a theoretical overview of spreading and diffusive processes on such systems. In the following sections of this review, we focus on findings of more direct relevance to the research community working with multimodal transportation and urban mobility.

## 2.3 Multimodal infrastructures

Multiple approaches have been followed to study the structure of cities, and since the 1950s fields such as Architecture, Urbanism and Transport Planning have grown a large body of literature studying the form and structure of cities and their transport systems. As cities grow and add different transportation modes, understanding the transportation infrastructure and its interconnected nature is crucial to fully capture real-world patterns of urban mobility. With the growth of the fields of Complex Systems and Network Science new tools and models have been developed to study the complexity behind urban systems, and specifically mobility infrastructure.

In particular, multimodal urban infrastructures can be represented as multilayer networks, in which each layer  $\alpha$  represents a mobility infrastructure (e.g. subway, light railway, bus service, pedestrian or bicycle infrastructure), the set

of nodes  $\mathcal{N}$  are locations (e.g. bus stops, intersections, subway stations), and the set of edges  $\mathcal{L}$  in layer  $\alpha$  are the infrastructure links between nodes in the same layer (e.g. subway lines, bus routes, bike lanes, see also Section II). Modeling infrastructures is of great importance for the understanding of how urban systems work, and the design of new sustainable mobility options.

In the following we will review the scientific literature on the modelling (Section 2.3.1) and the characterization of multimodal urban transportation systems from a complex systems perspective (Section 2.3.2).

### 2.3.1 Modeling urban infrastructure

How do transportation layers are coupled and grow? In this section we review the main models used to reproduce and understand multimodality in the city, and to simulate and test different scenarios with realistic data.

One of the first contributions to the topic was provided by de Cea et al. [47]. In which the authors tackle the problem of modelling a city and their transportation options, an approach used by different cities and transportation agencies to make rational decisions with respect of future development of their transportation infrastructure. Although, models have been used to plan and simulate the effects of new transportation options, most of them failed to consider congestion associated with transit modes, this was a shortcoming when the models are used to predict equilibrium for future years.

The model proposed by de Cea et al. [47], takes the road network as the base, in which the links has an average operating cost that takes into account different mobility options (e.g. cars, taxi, etc.). For every pure public transportation layer a new network is defined, with their unique nodes and links. For these layers, the cost function of the public transportation links depend on the sum of travel, waiting, and transfer time, as well as on vehicle flow over the road network and the passenger flow in the existing services. The model considers the existence of combined trips (e.g. car/metro, bus/metro, etc). As for the equilibrium conditions the model's basic assumption is that every user chooses her route to minimize their average operation cost (Wardrop's first principle). This means that at equilibrium, routes with flow will have an equal (minimum) cost, while those without flow will have an equal or greater cost than the minimum.

By taking into account the different layers in the system, flows, and travel cost, the authors are able to solve first a toy model, showing that their model is able to find the equilibrium in the trips between origin and destination. Finally, the authors show how the model can be applied into real world scenarios, it was used in Santiago (Chile) to evaluate the metro line 5. For the evaluation the model evaluated 13 user classes, 11 transport modes, and 450 zones (that means 439 trip matrices). The modelling results overestimate the demand in

the morning peak period by about 15 – 20%, an acceptable range for these type of predictions.

A similar philosophy was deployed by Li et al. [48]. Differently from the previous work in which the model takes in consideration the availability of routes, here the authors also introduce the parking behaviour and time spent while looking for parking of auto commuters. Taking into account these considerations, the authors focused on the interplay and impact of park-and-ride (P&R) schemes to encourage users to switch from car travel to subway and public transportation options when traveling to the cities' central area.

The model by Li et al. [48] considers the effects of traffic conditions on travel demand, and incorporates elastic demand into the model to capture commuters' responses to traffic congestion and availability of parking supply. The responses of a user include the decision to switch to another transportation option, or to not make the trip at all. The model, in the public transportation layer, also takes in consideration a level of discomfort resulting from the degree of crowding in transit vehicles. The final assumption of the model is that the mobility options for the commuters are: car, walk-metro, and park-and-ride.

When the authors analyzed the numerical simulations of the model, they found that the available parking spaces and cost in the P&R locations heavily influence the choices of the agents. In the simulations, the authors found that it is possible to have a P&R scheme that can simultaneously balance the parking demand and supply in the central area and at the P&R site such that the parking resource is fully used. Furthermore, when considering the frequency and fare of metro lines, the findings suggest that with a fixed frequency the balance of parking demand and supply in the P&R appears earlier (i.e., at a lower metro fare level) than in the cities' center. By contrast, for a fixed metro fare, the balance of parking demand and supply occurs earlier (i.e., lower frequency of metro lines) in the cities' center than in the P&R site.

The use of real world transportation infrastructure as a modeling framework for multilayer networks were first explicitly considered by Gil [49]. There, the authors proposed to use open data from Open Street Map to model the urban mobility network of a given city. The proposed model is composed of three layers, first the street network (private transportation) in which the nodes are intersections and links are streets, this layer is the main interface with the other layers in the system, meaning that all other layers have to be connected to it. Second, the public transport layer that represents the stations as nodes and link them whenever there is a public transport service between two stations. This layer is then linked to the street layer by the stations and their closest street intersection. Finally, the third layer the authors incorporated into their model is the land use, even if it is not an integral part of multimodal transportation, the

rational behind is that it is an integral part to measure accessibility and urban form.

After building their model, the authors applied it to analyze the Randstad city-region in the Netherlands. They tested the model under different parameters and layer combinations, measuring their closeness and betweenness centralities, and looking for significant statistical correlations of the analysis results with ground truth data, such as land us location, traffic, and passenger counts. The authors found that the betweenness centrality in the public transportation layer is a good indicator to passenger flows, the results improve when new layers are added into the model, except when adding the private transport layer. When the same measure were applied to the private transport layer the analysis did not capture the main roads and urban centers of the city, according to the study this effect has to do with the fact that the private layer contribute to a different phenomena in the region than the public transport.

As shown in the previous study, the private and public transport infrastructure contribute to different phenomena in the city, and thus each one can be investigated separately. For instance in Aleta et al. [50] the authors analyzed public transportation as a multilayer network following two approaches, the first one in which they modeled each bus/tram/subway line as a layer of the system. This approach is useful to take into consideration transfer times and synchronization between transport modes, however with this approach it is not possible to evaluate the importance of a transportation mode. The second approach is the most common one, in which all lines of the same mode are combined in a *superlayer*. The authors approach was to use both modelling options, they showed that the *superlayers* are fundamental to study the interdependency and resilience of the system while to have a realistic model of human mobility the single line per layer perspective should be adopted.

Using the two proposed models, the authors investigated the public transport systems for nine European cities. With the first model, the authors looked at the overlapping degree (sum of the node's degree in all layers). What the authors found was that even if all cities are different, the public transport infrastructure has some universal properties, the maximum overlapping degree is quite similar in all the networks, even though their number of layers -thus their theoretical maximum overlapping degree- differs a lot. This kind of generalized behaviour can be explained when we consider the city space in which the networks are embed. Thus, the real theoretical maxima (for both overlapping degree and edge overlap) are constrained by physical space and/or citizens' interest.

With the second proposed model, Aleta et al [?] proceeded to analyze the superlayers, first by comparing the overlapping degree with the superlayer ac-

tivity (fraction of nodes that are active in  $1, 2, \dots, M$  layers). They found that, surprisingly, the nodes with the highest overlapping degree are not the ones with the highest superlayer activity. In fact, the nodes (transport stops) belonging to just one superlayer are the ones which tend to have maximum overlapping degree. Those nodes that are present in only one superlayer, surely have a major role for the mobility of the system (think bus stops with multiple bus lines overlapping). As the authors pointed out, if we think in disruption of the system, it is easier to move a bus stop to a street nearby, even if it has multiple lines, than solve a disruption in a subway station. The authors later investigate the interdependence of the superlayers to find the importance of the superlayers based on the number of shortest paths that make use of the superlayer. As shown, the two proposed models are useful and complementary to investigate public transport systems, their functioning and to assess their vulnerabilities.

The different models presented in this section, shown how the multilayer networks framework has been used to investigate the structure, function and vulnerabilities of the system. In the next section we cover some measures to quantify the effects of these structures and their interconnections.

### 2.3.2 Characterizing multimodal infrastructure

In the previous section we saw how to model the different mobility options as multilayer networks. The empirical study of multimodal transportation infrastructure has revealed different structural properties for this specific kind of networks. But how to quantify the effectiveness of their interconnectdeness? Here we discuss a number of measures used to capture the multiplexity of multimodal infrastructures, such as the importance of different nodes, the system's resilience, and the similarity between layers.

**Paths** At a more global scale, multimodality is often associated to the ability of an agent to navigate the system. The navigation of an agent (or group of agents) in a transportation network can be defined by different types of paths.

A first possibility is to consider the quickest path, which neglects waiting times between transportation modes and is computed using the largest speed in the edges, and assuming a perfect synchronization between the different transportation systems. This would be the equivalent to find the shortest path between  $i$  and  $j$  in a weighted-single layer network. However, the transportation networks also have a temporal dimension. In order to find a path that allows for a change between two -or more- different transportation options we must find a time-respecting path, this is the shortest path between nodes  $i$  and  $j$  that takes into account the departures and arrivals constraints usually given by timetables

and more recently by GTFs Feeds. Furthermore, for multimodal transportation networks the walking transfer time between modes has to be taken into account when computing time-respecting paths.

Besides the time constraints to find a viable path between  $i$  and  $j$  in a multilayer transportation system, we also have to take into account the validity of the path [51, 52], meaning that the path is viable if the proposed sequence of modes is within a set of constraints. These constraints can be defined in terms of cost, not only the cost of traversal of one link, but also the interchange between modes, and also in terms of sequence, as the shortest path can be a combination of subway-bus-subway-private car-subway that might solve the shortest path, but it will be an ‘illogical path’ and an unlikely choice for a user. Hence, when looking for viable paths we look for the shortest path with no more than a predefined number of modal transfers.

The task to find viable paths is one of the most important problems in urban transportation, as it has the potential to help user to find the most efficient paths in the city. In their work Lozano et al. [52] proposed an algorithm to find such paths when the agent establishes her limitations on the number of modal transfers.

When looking for shortest and viable paths the topology of the network has an effect, and certain layers of the multimodal network are of utmost importance. As shown by Aleta et al. [?], just by computing shortest paths, and disregarding the travel speed, the rail systems (trams, subway) contribute to most of the shortest paths in a city. As the authors points out, the reason behind this seems clear, the rail network connects distant points in straighter routes than bus or other transport modes.

**Spatial outreach** The availability of different transportation modes, such as subways or tramways, affects the accessibility of certain locations in the city. To measure this effect the spatial outreach is defined following Strano et al. [53] as the average Euclidean distance from node  $i$  to all other nodes in the same layer  $\alpha$  that are reachable within a given travel cost,  $\tau$ :

$$L_\tau(i) = \frac{1}{N(\tau)} \sum_{j|\tau_m(i,j) < \tau} d_e(i, j), \quad (2.1)$$

where  $d_e(i, j)$  is the Euclidean distance between nodes  $i$  and  $j$ , and  $N(\tau)$  is the number of nodes reachable on the multilayer network within a travel cost  $\tau$ .

When investigating the travel outreach, Strano et al. [53] modified the average speed (traversal time of a link) in the layers of the system to measure their effects. They found that when the metro speed increases compare to the street

layer speed, a continues area of high-outreach nodes emerges in the city center and around the nodes that have connections to the high-speed layer. In other words, as the velocity in layer  $\beta$  increases the nodes that are closer to the interchange nodes in layer  $\alpha$  improves their accessibility, implying that a person can travel from this areas to faraway places.

The outreach and accessibility has also been studied in multimodal transport networks using isochrones (accessible area from a point within a certain time threshold). The work by Biazzo et al. [20] used this approach to measure accessibility in different urban areas. The authors measure the accessibility (in an aggregated hexagon containing multiple intersections, no by single nodes), using the isochrones computed by the combination of public transit and pedestrian infrastructure. The isochrones where then used to obtained scores that capture how well a city is served by the public transit and how accessible is a specific area of the city to the rest of the city.

**Interdependency** As shown in the previous measures the structure of the multimodal transportation network has different effects on how people can move in a city. One of this effects is the fraction of shortest paths that go from one node to all the others that makes use of two or more different layers. This topological interaction between different layers in a multilayer network can be described as a *coupling*. For transportation networks a better approach is to measure the interdependency [54, 55, 53]

$$\lambda_i = \frac{1}{N-1} \sum_{j \neq i} \frac{\psi_{ij}}{\sigma_{ij}}, \quad (2.2)$$

where  $\psi_{ij}$  is the number of shortest paths between  $i$  and  $j$  that use edges in two or more different layers, and  $\sigma_{ij}$  is the amount of shortest paths between  $i$  and  $j$ . The quantity of interdependency of node  $i$  lies in the interval  $[0, 1]$ , in which the larger values means that most of the paths go through two or more layers and closer to 0 means that most of the paths go through just one layer. By taking the average over all nodes  $\lambda = 1/N \sum_i \lambda_i$  we obtain the interdependency of the system.

To take into account the nature of transportation networks this metric can be modified to calculate the interdependency of a layer in the system [50], denoting  $\psi_{ij}^\alpha$  as the number of shortest paths between nodes  $i$  and  $j$  that uses two or more layers and at least one of them is layer  $\alpha$ . The layer interdependency for layer  $\alpha$  is defined as:

$$\lambda^\alpha = \frac{\sum_i \sum_{i \neq j} \psi_{ij}^\alpha}{\sum_i \sum_{i \neq j} \psi_{ij}}, \quad (2.3)$$

this measure tells us how many paths go across two or more layers and at least one part of the path belongs to layer  $\alpha$ . Note that the normalization is over the multilayer shortest paths to find the layers that are used to change to another one.

When applying this measure to multimodal transport networks [50] found that the metro and trams layers play an important role in concentrating shortest paths. For Madrid, the authors found that more than 40% of the trips have at least one link in the metro layer, even if the metro layer has only 241 nodes while the bus layer has 4590 nodes. The reason behind the importance of the metro and tram layers seems clear, as they connect distant points with straighter routes than the other transportation options in the system.

**Centrality** When analyzing a city and their mobility options one of the first task is to locate the transportation hubs, nodes in the system that have connections between different layers, and enable an agent to move and transfer between mobility options. At the same time, being central in one layer can mean that a node has multiple connections to other nodes in the system, for example a subway station that has multiple lines connecting on it.

In the street layer the number of possible connections from one node to the rest of the nodes in the system is constrained by the physical space, one intersection has a limited number of streets/sidwalks that can intersect there. However, when analyzing a single layer there are other centralities measures that can provide insights about the importance of a node or a link. One of these measures is the betweenness centrality [56], that in the absence of mobility data, is a good proxy to find bottlenecks in the system. The betweenness centrality measure the amount of shortest paths that go through one node/edge in the network. When calculating it for the node  $i$  in layer  $\alpha$  the betweenness centrality can be computed as follows:

$$\text{bc}_i^\alpha = \frac{1}{(N^\alpha - 1)(N^\alpha - 2)} \sum_{j,k \in N^\alpha} \frac{\sigma_{j,k}^\alpha(i)}{\sigma_{j,k}^\alpha}, \quad (2.4)$$

where  $\sigma_{j,k}^\alpha$  is the total number of shortest paths between  $j$  and  $k$  in layer  $\alpha$ , and  $\sigma_{j,k}^\alpha(i)$  is the number of shortest paths between  $j$  and  $k$  that goes through node  $i$ . Similarly, we can define the betweenness centrality of node  $i$  in layer  $\alpha$  for the multilayer network  $\mathcal{M}$  as:

$$\text{bc}_i^{\text{multi}} = \frac{1}{(N^\alpha - 1)(N^\alpha - 2)} \sum_{j,k \in N^\alpha} \frac{\sigma_{j,k}^{\text{multi}}(i)}{\sigma_{j,k}^{\text{multi}}}, \quad (2.5)$$

where  $\sigma_{j,k}^{\text{multi}}$  is the total number of shortest paths between  $i$  and  $j$  in the multiplex network, and  $\sigma_{j,k}^{\text{multi}}(i)$  is the total number of shortest paths between nodes  $j$  and  $k$  that goes through node  $i$  in any layer.

As cities grow and new lines and options are added into the mobility system, new interconnections between layers have to be added, thus the betweenness centrality also evolves. Ding et al. [57] studied how the centralities evolved when the rail network of Kuala Lumpur grew from a tree-like structure to a more complex one. Their findings suggest that as the network grows the average shortest path in the multilayer network decreases dramatically, also affecting the betweenness centrality, as new nodes are able to serve as interchange between layers, thus enabling new shortest paths along the system.

The results from Ding et al. [57] are in line with the previous findings by Strano et al. [53], where they analyzed the coupling effect between the street and subway networks in London and New York. The authors show that the introduction of new layers and their interconnections plays an important factor on decentralizing congestion in the street layer. The analysis of betweenness centrality shows that the subway systems have an effect as decentralizing force, moving the congestion (measured as betweenness) from internal street routes and bridges, to inter-modal places located at the terminal points of the subway system which might be used as interchange points for suburban flows into the city center.

**Resilience** The structure of the multimodal system also plays an important role to understand the resilience of the transportation network. In a single layer network when a node fails it can not be traversed by any single path, as shown by de Domenico et al. [58] for the single layer representation of the public transportation network in Great Britain. However, if the node is part of a multimodal transportation network it can still be accessed by any of the other layers. To measure this resilience de Domenico et al. [42] used random walks and coverage methods to study London's transportation system, showing that the multiplexity of the system enhances its resilience to failure compared with a single layer. In a similar approach Baggag et al. [22] used random walks to measure the coverage and robustness in multimodal transportation networks, in this case the authors introduced a limitation in the form of a given time budget or steps, showing that the layers interconnection helps to find paths from apparently isolated areas of the system. And more recently Ferretti et al. [23] used the multiplex framework to model Singapore's public transportation infrastructure and test its resilience against floods in the city in different scenarios, either removing nodes or edges from the multimodal network.

As with the resilience, the study of the multimodal structure has been a use-

ful proxy to understand congestion in transportation systems when dynamical data is not available. As shown by Stano et al. [53] and later by Ding et al. [57] the spatial distribution of betweenness centrality changes when a new layer is introduced into the system, specifically when adding a subway the congestion shifts from internal street routes and bridges to intermodal places in terminal points of the subway network. In a similar sense, the work of Solé-Ribalta et al. [59] shows that the main driver in traffic dynamics and congestion in multimodal transport networks, is the change of paths from the least efficient layers to the most efficient one.

**Overlap census** One interesting problem when dealing with multimodal transportation networks is the characterization, and comparison between different cities based on their multimodal mobility options. For multiplex networks in general Nicosia et al. [36] proposed measures to capture nontrivial correlations in multiplex networks. And more recently, Brodka et al. [60], presented an overview of different metrics to compute similarities between layers in multiplex networks.

However a method to compare the mobility options and its infrastructure of different cities had yet to be proposed, a possible approach to this is the overlap census [61], that we cover in detail in Chapter 3.

## 2.4 Multimodal mobility

Understanding urban travel is paramount for a range of real-world applications, including planning transportation [62] and designing urban spaces. Starting from the 1950s, a large body of literature in the fields of Geography and Transportation has studied how people move and use transportation technology.

As the transport infrastructure becomes increasingly multi-modal, modelling how individuals make travel decisions in complex interconnected networks is critical. In recent years, the scientific understanding of Human Mobility has dramatically improved, also due to the widespread diffusion of mobile-phone devices and other positioning technologies, which allowed to gather large-scale geo-localized datasets of human movements and develop increasingly realistic behavioural models. Concurrently, these recent developments were made possible by the dramatic growth of the fields of Complex Systems and Network Science, which brought together ideal tools to study interconnected systems (for a comprehensive review of the recent literature stream of Human Mobility, the reader can refer to [63]). Despite these advancements, our understanding of multimodal mobility in urban systems remains limited, also

due to the difficulties related to collecting comprehensive data across multiple transportation modalities.

In the following, we overview the scientific literature on multi-modal mobility. While our focus will be on multi-modality, we will inevitably touch upon some of the concepts related more broadly to modelling of urban travel. In Section 2.4.1, we briefly overview existing models, focusing on latest advances driven by the Complex Systems literature. In Section 2.4.2, we review measures and empirical findings, with a focus on recent studies based on passively collected data sources.

### 2.4.1 Modeling urban mobility

Modelling travel demand in a multimodal system involves understanding how individuals make decisions in a constantly changing complex environment. The most diffused family of models for travel demand in the Geography and Transportation literature are the so-called *four-step models* proposing that each trip results from four decisions [64]: whether to make a trip or not, where to go, which mode to use, and which path to take. For simplicity, these have been largely considered as independent, sequential choices. We refer the reader to the excellent review by McNally [64], for a comprehensive overview about this modeling approach.

In recent years, the literature in Complex Systems has modeled travel behavior on multiplex networks using different approaches, that we briefly review in this section. It is important to remark that, also due to the lack of empirical data on the mechanisms driving human navigation, most models are based on the simplistic assumption that individuals are rational, homogeneous and have unlimited knowledge. Research based on novel data sources will be key to develop mobility models on multilayer networks that include realistic elements such as limited knowledge and cognitive limitations.

**Random walks.** A random walk process is a prototypical model of mobility on a network, and it is defined by a walker that, located on a given node  $i$  at time  $t$ , hops to a random nearest neighbor node  $j$  at time  $t + 1$ . In the case of multilayer networks, the walk between nodes and layers can be described with four transition rules accounting for all possibilities [42]: (i)  $P_{ii}^{\alpha\alpha}$ , the probability for staying in the same node  $i$  and layer  $\alpha$ ; (ii)  $P_{ij}^{\alpha\alpha}$  the probability of moving from node  $i$  to  $j$  in the same layer  $\alpha$ ; (iii)  $P_{ii}^{\alpha\beta}$  the probability of staying in the same node  $i$  while changing to layer  $\beta$ ; (iv)  $P_{ij}^{\alpha\beta}$  the probability of moving from node  $i$  to  $j$  and from layer  $\alpha$  to  $\beta$ , in the same time step. These probabilities depend

on the strength of the links between nodes and layers (e.g. the frequency of vehicles and the cost associated to switching layers).

Notwithstanding their simple formulation, random walks provide fundamental hints to understand all types of diffusion processes on networks and measure the dynamical functionality of a network. For example, random walks processes were studied to measure the navigability of multiplex networks [42]. To this end, one can measure the coverage of the multiplex network  $\rho(t)$ , defined as the average fraction of distinct nodes visited by a random walker in a time shorter than  $t$  (assuming that walks started from any other node in the network), and describing the efficiency of a random walk in the network exploration:

$$\rho(t) = 1 - \frac{1}{N^2} \sum_{i,j=1}^N \delta_{i,j}(0) \exp[-\mathbf{P}_j(0) \mathbb{P} \mathbf{E}_i^\dagger], \quad (2.6)$$

where  $\mathbf{P}_j(0)$  is the supravector of probabilities at time  $t = 0$ , the matrix  $\mathbb{P}$  accounts for the probability to reach each node through any path of length  $1; 2; \dots$  or  $t + 1$ , and  $\mathbf{E}_i^\dagger$  is a supra-canonical vector  $i$  allowing to compact the notation. In the work by De Domenico et al. [42], the authors provided an alternative representation of equation 2.6 building upon the eigendecomposition of the supra-Laplacian. They showed that the ability to explore a multiplex is influenced by different factors, including the topological structure of each layer and the strength of interlayer connections and the exploration strategy. Further, they showed that the multilayer system is more resilient to random failures than its individual layers separately because interconnected networks introduce additional paths from apparently isolated parts of single layers, and thus enhance the resilience to random failures.

Random walks can be used to assign a measure of importance to each node in each layer, by measuring the asymptotic probability of finding a random walker at a particular node-layer as time goes to infinity, the so-called ‘occupation centrality’[65]. In the work by Solé-Ribalta et al. [65], the authors provided analytical expressions for the occupation centrality in the case of multilayer networks.

**Agent-based models accounting for congestion.** Several works have focused on modelling agents aiming at minimizing total travel time in a congested network [66, 67, 68, 59].

In Bassolas et al. [67], the authors developed an agent-based models describing individuals mobility through a multilayer transportation system with limited capacity. The routing protocol used by individuals for planning is adaptive with local information. In the absence of congestion, individuals follow the

temporal optimal path of the static multilayer network calculated by the Dijkstra algorithm. If there are line changes, they estimate, besides the change walking penalty, an additional waiting time of half the new line period (the real waiting time will be given by the vehicles location in the line when the individual arrives at the stop). Individuals' route is only recalculated when a congested node, whose queue is larger than the vehicle's capacity, is reached. They investigate analytically (for simple networks) and via numeric simulations the robustness of the network to exceptional events which give rise to congestion, such as demonstration concerts or sport events. The study revealed that the delay suffered by travellers as a function of the number of individuals participating in a large-scale event obeys scaling relations. The exponents describing these relations can be directly connected the number and line types crossing close to the event location. The study suggested a viable way to identify the weakest and strongest locations in cities for organizing massive events.

In a similar way, in the work by Manfreid et al. [68] the authors introduce a limit to the nodes capacity of storing and processing the agents. This limitation triggers temporary faults in the system affecting the routing of agents that looks for uncongested paths.

### 2.4.2 Characterizing multi-modal mobility.

Traditionally, multi-modal mobility models are calibrated using data from travel surveys: *Revealed Preference* surveys retrieve actual travel information from the respondents, while *Stated Preference* surveys expose the travelers to various hypothetical scenarios and record their choices [69]. Studies based on survey data have provided insights into how multi-modal travelers value aspects such as the different travel time components (in-vehicle time, walk time, access time, wait time...) [70], service quality [71], and travel costs [69], and heterogeneities across socio-demographic groups [72]. Due to the high costs associated with data collection and inherent biases in self-reported data, these studies suffer of serious limitations, including small sample sizes, data inaccuracy and incompleteness [73, 74]. Covering empirical results from travel surveys is outside the scope of this review, and we refer the reader to [69] for a comprehensive introduction to the topic.

In recent years, the empirical research on Human Mobility has taken new directions. A growing body of literature has focused on quantitative descriptions of human movements from large, automatically collected data sources, such as mobile phone records, travel cards and GPS traces [63]. In this section, we overview recent empirical findings on multi-modal mobility in the field of Complex Systems, which focused on two important aspects: the dynamics of *public transport systems*, whose study was driven by the availability of public

transport data such as schedules and positions of stop and stations; and *individual multi-modal behaviour*, driven by the availability of data collected using ‘smart travel cards’ and GPS data.

The existing empirical research on multi-modal travel based on passively collected data-sources is far from being comprehensive. Most studies have focused on public transit, such that the interplay between public and forms of private transportation such as walking, driving and cycling has been poorly characterized. Further, several studies are based on public transport schedules instead of real-time data, thus neglect important effects deriving from congestion. The increasing availability of high-resolution GPS trajectories collected by individual mobile phones and sensors installed on private and public transport vehicles [63], will be key to fill these gaps in the literature.

### Public transport systems dynamics.

Over the last decade, the availability of detailed public transport schedules shared by public transport companies has allowed to better estimate travel times and characterize transport systems.

**Efficiency** To satisfy the demand of large number of individuals while reducing energy and costs, multilayer transport systems must achieve high efficiency. One aspect concerns the *synchronization* between the network layers, because the more layers are synchronized, the less users have to wait for vehicles. The synchronization inefficiency  $\delta(i, j)$  [75, 76] for nodes  $i$  and  $j$  can be measured as the ratio of the time-respecting travel time  $\tau_t(i, j)$ , which accounts for walking and waiting times and the fact that the speed of vehicles varies during the day, and the minimal travel time  $\tau_m(i, j)$ , where one assumes that vehicles travel at their maximum speed and transfer are instantaneous:

$$\delta(i, j) = \frac{\tau_t(i, j)}{\tau_m(i, j)} - 1 \quad (2.7)$$

Using the synchronization inefficiency, it was shown [75] that, on average in the UK, 23% of travel time is lost in connections for trips with more than one mode. Interestingly, across several urban transport system in the UK, the synchronization efficiency  $\delta(i, j)$  obeys the same scaling relation with the path length  $\ell(i, j)$ :

$$\delta(i, j) \approx \delta_{min} + \frac{\delta_{max} - \delta_{min}}{\ell(i, j)^v}, \quad (2.8)$$

with  $v \approx 0.5$ , and where  $\delta_{max}$  and  $\delta_{min}$  are the maximum and minimum values of  $\delta(i, j)$  for a given urban transport system.

Further, it has been shown that the average synchronization inefficiency  $\bar{\delta}$  for a given urban system follows:

$$\delta \sim \Omega^{-\mu}, \quad (2.9)$$

where  $\Omega$  is the total number of stop-events per hour (e.g. the number of times a vehicle stops), and  $\mu = 0.3 \pm 0.1$ .

Other studies focused on the efficiency in terms of ability to satisfy users demand. In Alessandretti et al. [77], the authors introduced a method based on non-negative matrix factorization to compare the network of commuting flows and the public transport network. The methodology, applied to various public transport systems in France, showed that, while in Paris the transportation system meets the overall demands, it does not in smaller cities, where even people with access to fast public transportation prefer to use car. In Sui et al. [78], the authors proposed three topological metrics to quantify the interaction between public transport network and passengers' flow and applied it to study differences between the cities of Chengdu and Qingdao in China. In Hollecsek et al. [79] the authors used data mining approaches to compare the use of public and private transportation and identify the existence of weak transportation connections.

**Congestion.** Congestion can dramatically alter travel time estimates for routes that use popular network links in a urban system, but adding network layers in a multilayer transport system can help reduce global congestion [80]. Given a network with edges  $e$  and flows  $j_e$ , one can quantify the total time lost in congestion as

$$T_c(\mathbf{j}) = \sum_{e \in \mathcal{L}} j_e(t_e^* - t_e(j_e)), \quad (2.10)$$

where  $\mathbf{j}$  is the vector of flows whose  $e$ th element is the flow along edge  $e$ ,  $t_e^*$  is the free flow time on edge  $e$  (in the absence of congestion), and  $t_e(j_e)$  is the congested travel time in the presence of flow  $j_e$ . This measures can be used to analyze the impact of changes in flow along a route. The quantity

$$\Delta_p = -\nabla T_c(\mathbf{j}) \cdot \mathbf{e}_p \quad (2.11)$$

where  $p$  is a path and  $\mathbf{e}_p$  is the vector whose  $e$ th component is 1 iff  $e \in p$ , quantifies the impact of removing a single unit of flow from  $p$  on the global congestion function  $T_c$ . Chodrow et al. [80], quantified how the creation of a planned metro network in Riyadh would affect congestion, by quantifying the change  $\Delta_p$  as a function of the speed ratio between the street and metro systems:

$$\beta = v_c/v_m \quad (2.12)$$

The authors showed that, as the subway speed increases the global congestion is reduced, but increases locally close to key metro station.

**Navigability** As cities and their transportation systems become increasingly complex and multimodal, it is important to quantify our difficulty navigating in them. It has been shown that multilayer transport system are characterized by limited navigability, implying that finding one's way is cognitively challenging [81]. To quantify the difficulty of navigating between two nodes  $s$  and  $t$  in a network, one can compute the total information value of knowing any of the shortest paths to reach  $t$  from  $s$ :

$$S(s \rightarrow t) = -\log_2 \sum_{\{p(s,t)\}} P[p(s,t)] \quad (2.13)$$

where  $p(s,t)$  is the set of shortest paths between  $s$  and  $t$  (note that there can be more than one with the same length) and  $P[p(s,t)]$  is the probability to follow path  $p(s,t)$ , making the right choice at each intersection along the path [82]:

$$P[p(s,t)] = \frac{1}{k_s} \prod_{j \in p(s,t)} \frac{1}{k_j - 1}, \quad (2.14)$$

In Gallotti et al. [81] the authors quantified the amount of information an individual needs in order to travel along the shortest path between any given pair of metro stations, in the single-layer metro networks for 15 large cities. They found that this information has an upper bound of the order of 8 bits, corresponding to approximately 250 connections between different routes. Further, studying several among the largest multi-layer transport networks (metro/buses/light rail), they showed that the amount of information one needs to know to travel between any two points exceeds the identified cognitive limit of 8 bits in 80% the cases, suggesting multi-layer are too complex for individuals to navigate easily.

### Individual multi-modal behaviour.

In recent years, data collected via smart travel cards has dramatically improved our ability to characterise multimodal behaviour in urban transport systems, overcoming some of the limitations related to collecting and analyzing survey data [73, 74]. Smart-card automated fare collection systems allow passengers to make journeys involving different transport modes using magnetic cards and automatic gate machines. As these systems identify and store the location and

time where individuals board and, in some cases, alight public transport, they collect accurate descriptions of individual travel [83]. Concurrently, advancements were made possible by the development of methodologies allowing to identify typical travel patterns [84].

**Route choices.** Smart-card data has allowed to quantify how individuals navigate multilayer networks. One of the key findings is that individuals do not choose optimal paths (those with shortest travel time), especially when the system is congested. Focusing on the bus and subway trips of 2.4 million passengers in Shenzhen (China), Zheng et al. [85] studied the coupling (see eq.2.2) between the bus and subway layers. In contrast to previous studies [53], the authors characterized the coupling  $\lambda$  engendered by passengers' behaviour rather than structural properties of the multilayer network. Under their definition, the *coupling* between layers is the fraction of multimodal trips actually undertaken by passengers, rather than the fraction of multimodal shortest paths (see eq.2.2). The authors find that this 'behavioural' coupling correlates weakly with the empirical speed ratio measured between the two layers over time, implying that passengers choose unimodal trips even when multimodal trips may be preferred because one of the two layers is congested. This finding highlights that the speed ratio of different network layers, which was regarded as a key factor in determining coupling strength [53, 80], may have a negligible effect on travelers' route selections, possibly because passengers do not have a full view of the status of traffic. Instead, the authors showed that the coupling between layers is generated by long-distance trips originating from nodes served by a single transport layer.

**Power-law distribution of displacements** The availability of large-scale data sources has allowed to reveal that individual mobility patterns display universal properties. One key finding is that the distribution  $P(\Delta r)$  describing the probability of travelling a given distance  $\Delta r$  is characterized by a power-law tail  $P(\Delta r) \sim \Delta r^{-\beta}$ , with  $1 \leq \beta \leq 2$  [63]. Interestingly, it was shown that this observation could be the effect of the widespread use of multimodal transportation infrastructure [86, 87]: In Zhao et al. [87], the authors used GPS data to show that mobility using a single mode can be approximated by a lognormal distribution, but the mixture of the distributions associated with each modality generates a power-law. In Gallotti et al. [86], the authors found that a simple model where individual trajectories are subject to changes in velocity generates a distribution of displacements with a power-law tail. In fact, individuals using multimodal infrastructure are subject to drastic changes in velocity. In Verga et al. [88], the authors showed that the travel speed  $v$  increases with travel distance

according to the power-law functional form  $v \sim r^\alpha$ , where  $\alpha \approx 0.5$ . This dependence is due to the hierarchical structure of transportation systems and the fact that waiting-times (parking, take-off, landing, etc) decrease as a function of trip distance.

## 2.5 Open data & Tools

During the last years there has been an increase in data availability and tools to study multimodal transport networks, in this section we highlight some of the datasets that are available, and computational tools to analyze the data as a complex networks and more specifically as multiplex transport networks.

### 2.5.1 Data

There are two main types of data available, the one that captures the infrastructure, and the one about dynamics. For the first one the largest open data repository that has been used to analyze transportation networks is OpenStreetMap, a high quality [89, 90, 91, 63], collaborative effort to map the world. For the later, different data sets have been published making use of General Transit Feeds Specification (GTFS), smart cards to access public transportation, and some origin destination studies that took into account the mobility options.

Using the data from OpenStreetMap it is possible to generate transportation networks, specifically for the single layer case multiple datasets have been publish (for an overview see [12, 92]). With the same data it is possible to build multimodal transport networks, [93] proposed a methodology to extract the different layers from OpenStreetMap data and build a multiplex network linking the layers by the intersections between transport modes. Following this approach we [94] made public the multiplex transportation data for fifteen cities in the world including cities from developed countries, like the United States of America, to cities in developing countries like Mexico or Colombia.

In 2015 Gallotti et al. [95] published the multilayer temporal network of public transport in Great Britan, a large data set that contains not only the public transport layers in cities, but also the national level covered by coach, planes and ferries. This dataset has the property to include the temporality of the trips encoded in the links, a link is active during a trip and encodes the travel time as the arriving time to the destination node minus the departure time from the origin node. This dataset also includes a static non-temporal network, in which the weight of each edge is the minimal travel-time.

The use of timetables and transit feeds has been useful to bridge and complement the structure with dynamical data. More recently with the digitalization

of transport schedules and timetables by public transit operators new data has been made publicly available, an extensive collection of GTFS datasets is available at <https://transitfeeds.com/>. This datasets includes the stops, routes and timetables of public transport in multiple cities and providers around the world. Although this data can be analyzed as single layer of public transportation, it can also be used to create multiplex networks in which each layer is a line or transport provider that are interconnected based on their shared stations, a similar approach that the one followed by Alete et al.[50]. Kujala et al. [96] published a collection of 25 cities' public transport networks, containing cities from North America, Europe, and Oceania. The dataset differentiates between public transport modes (tram, subway, rail, bus, ferry, cablecar, gondola, funicular), and also includes the walking layer, enabling the construction of multilayer networks when taking into account different transport modes.

A recent dataset that contains travel time and distance information for the Helsinki Region has been made available by Tenkanen et al. [97]. In this dataset the authors included the travel time and distance information between all the 250 meters statistical grid centroids for the Helsinki Capital Region. The data includes the travel time matrices for 2013, 2015 and 2018. The dataset is multimodal, the included transport modes are walking, cycling, public transport and private car. The travel times were calculated using a door-to-door principle, making the information between different travel modes and year comparable.

### 2.5.2 Tools

Multiple computational tools have been developed to work with complex networks, such as Networkx [98], graph-tool [99], and igraph [100], enabling research in multiple topics.

For the transportation and multimodal analysis different tools have been developed, such as OSMnx [101], a Python package that downloads and constructs street networks from OpenStreetMap, it can download other infrastructure networks too, allowing to download different layers from a city and build their multimodal transport network. In a similar sense, Pyrosm [102] is a python library for reading data from OpenStreetMap's Protocol Buffer Format files (\*.osm.pbf) into GeoDataFrames, making it easy to extract multiple datasets from a single file. The General Transit Feed Spec has played an important role to make available more data than can be used to build the public transport layer in a multimodal transport network. Google [103] developed a Python library to parse, validate and build GTFS files.

Specifically to work with multiplex and multilayer networks different tools have been developed. De Domenico et al. [104] published muxViz a tool to visualize and analyze multilayer networks, this tool provides a Graphical User

Interface that lets the user interact and analyze with the network.

MAMMUL, is a collection of algorithms developed by Nicosia et al. [105] for the analyzes and modelling of multilayer networks. The functions included in the collection cover from structural properties, such as node, edge, and layer basic properties, to the analysis of dynamics on multilayer network, such as random walks. Although, the tool was developed to work with general multilayer networks, it can be adapted to work with multimodal transport networks.

multiNetX [106] is a python package that builds on top of Networkx and extends its capabilities to work with multilayer networks, this allows to create undirected weighted and unweighted networks, analyse the spectral properties of the adjacency or Laplacian matrix, and visualize dynamical processes by coloring the nodes and links in the visualizations.

Pymnet by Kivela et al. [107] is one of the most recent Python libraries to work with multilayer networks. The library handles general multilayer networks, including multiplex networks with temporal variables, and has multiple network analysis methods, transformations, and models for its analysis and visualization.

The multinet library for the analysis of multilayer networks by Magnani et al. [108], is a Python and R library that provides tools and algorithms to work with multilayer networks including community detection and visualization methods.

Rivero et al. [109] developed *multiplex*, an R library to work with multilayer networks, apart from manipulating and visualizing multigraphs, this package offers multiple functions to work with algebraic systems - such as the partially ordered semigroup, and balance or cluster semirings - their decomposition, and the enumeration of bundle patterns occurring at different levels of the network.



## CHAPTER 3

---

# EXTRACTING THE MULTIMODAL FINGERPRINT FOR TRANSPORTATION NETWORKS

---

Urban mobility increasingly relies on multimodality, combining the use of bicycle paths, streets, and rail networks. These different modes of transportation are well described by multiplex networks. Here we propose the overlap census method which extracts a multimodal profile from a city's multiplex transportation network. We apply this method to 15 cities, identify clusters of cities with similar profiles, and link this feature to the level of sustainable mobility of each cluster. Our work highlights the importance of evaluating all the transportation systems of a city together to adequately identify and compare its potential for sustainable, multimodal mobility<sup>1</sup>.

---

<sup>1</sup>A stand alone of this chapter has been published in Transport findings [61]

### 3.1 Introduction

In the previous chapter, we survey the literature on the topic of multimodal transportation networks. In this chapter, we provide an original contribution to the topic by suggesting a method to extract the multimodal profile from a given city's multiplex transport network.

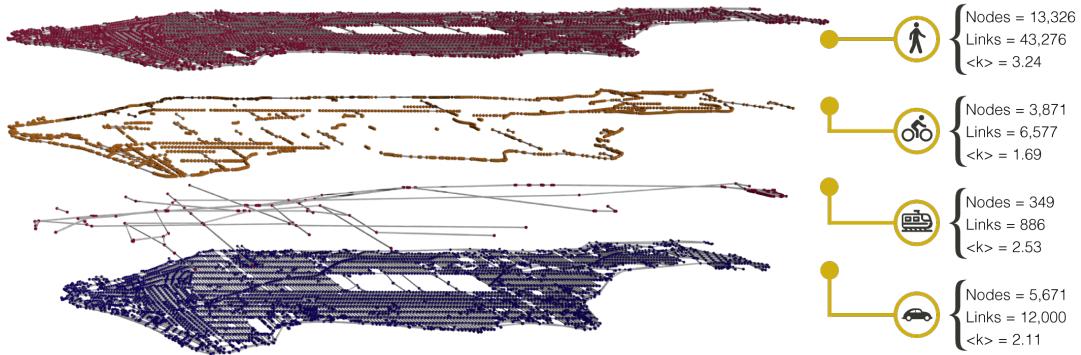
The infrastructure of different modes of transportation can be described as a mathematical object, the multiplex transport network [54, 17, 110, 33, 75, 42, 53, 50, 111]. A city's multiplex transport network contains the layer of streets and other coevolving network layers, such as the bicycle or the rail networks, which together constitute the multimodal transportation backbone of a city. Due to the car-centric development of most cities [112], streets form the most developed layers [113, 114] and define or strongly limit other layers: For example, sidewalks are by definition footpaths along the side of a street and make up a substantial part of a city's pedestrian space [113]. Similarly, most bicycle paths are part of a street or are built along the side. Yet, the different layers of a multimodal network typically serve as diverse channels to permeate a city. Here we consider the transport networks of 15 world cities and develop an urban fingerprinting technique based on multiplex network theory to characterize the various ways in which transport layers can be interconnected, identifying the potential for multimodal transport. Using clustering algorithms on the resulting urban fingerprints, we find distinct classes of cities, reflecting their transport priorities.

### 3.2 Data

We acquired urban transportation networks from multiple cities around the world, defined by their administrative boundaries, using OSMnx [101] a Python library to download and construct networks from OpenStreetMap (OSM). OSMnx simplifies the OpenStreetMap's raw data to retain only nodes at the intersections and dead ends of streets, and the spatial geometry of the edges, generating a length-weighted nonplanar directed graph [14]. These data sets are of high quality [89, 90] in terms of correspondence with municipal open data [115] and completeness [63]. The various analyzed urban areas and their properties are reported in Table 3.1. Figure 3.1 shows the different network layers for Manhattan, one of our analyzed cities.

Code to replicate our results is available as Jupyter Notebooks (<https://github.com/naturaluis/Multimodal-Fingerprint>) and data can be downloaded from Harvard Dataverse [94].

To characterize the city as a multiplex network [32, 35, 45] we follow the



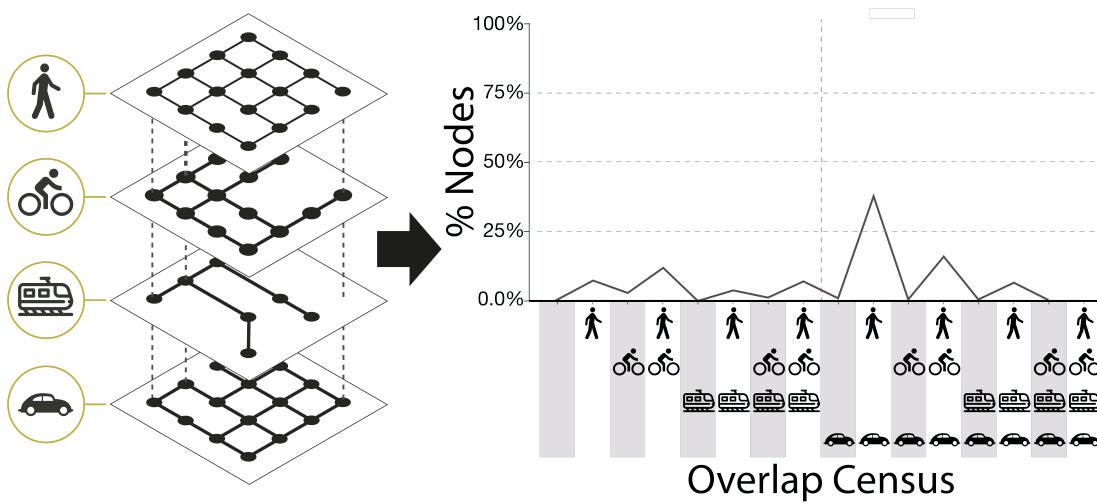
**Figure 3.1.** (*Map plot left*) Multiplex network representation of Manhattan with the four analyzed layers of transport infrastructure (pedestrian paths, bicycle paths, rail lines, and streets), with data from OpenStreetMap. (*Right*) Network information for each layer, number of nodes, links and average degree  $\langle k \rangle$ .

	Pedestrian			Bicycle			Rail			Street			Population
	Nodes	Links	$\langle k \rangle$	Nodes	Links	$\langle k \rangle$	Nodes	Links	$\langle k \rangle$	Nodes	Links	$\langle k \rangle$	
Amsterdam	23,321	33,665	2.89	34,529	35,619	2.06	1,096	1,655	3.02	15,125	21,722	2.87	872,680
Barcelona	20,203	30,267	3.00	7,553	7,647	2.02	249	249	2.00	10,393	15,809	3.04	1,600,000
Beihai	2,026	2,978	2.94	0	0	0.00	59	62	2.10	2,192	3,209	2.93	1,539,300
Bogota	81,814	121,038	2.96	9,760	9,651	1.98	166	165	1.99	62,017	91,197	2.94	7,412,566
Budapest	73,172	106,167	2.90	10,494	10,318	1.97	1,588	1,964	2.47	37,012	52,361	2.83	1,752,286
Copenhagen	30,746	41,916	2.73	13,980	13,988	2.00	276	369	2.67	15,822	20,451	2.59	2,557,737
Detroit	47,828	78,391	3.28	3,663	3,626	1.98	20	21	2.10	28,462	45,979	3.23	672,662
Jakarta	140,042	191,268	2.73	248	231	1.86	58	54	1.86	138,388	188,637	2.73	10,075,310
LA	89,543	128,757	2.88	14,577	14,428	1.98	173	221	2.55	71,091	101,692	2.86	3,792,621
London	270,659	351,824	2.60	62,398	60,043	1.92	2,988	3,535	2.37	179,782	219,917	2.45	8,908,081
Manhattan	13,326	21,447	3.22	3,871	3,777	1.95	349	436	2.50	5,671	9,379	3.31	1,628,701
Mexico	108,033	158,425	2.93	5,218	5,278	2.02	370	364	1.97	95,375	140,684	2.95	8,918,653
Phoenix	111,363	157,075	2.82	35,631	35,979	2.02	105	138	2.63	73,688	102,139	2.77	1,445,632
Portland	50,878	72,958	2.87	24,252	24,325	2.01	230	340	2.96	35,025	49,062	2.80	583,776
Singapore	82,808	110,612	2.67	12,981	12,947	1.99	683	740	2.17	50,403	66,779	2.65	5,638,700

**Table 3.1.** Measures for the administrative area of analyzed cities. The number of nodes, links and average degree ( $\langle k \rangle$ ) for each layer in all cities of our dataset are highly diverse due to the varying developmental levels and focus of transport. The range of population in the analyzed cities goes from half million people to ten million people living in Jakarta, this allows to have a range of different sizes and cover different developmental stages.

approach presented in section 2, in which the system has  $M$  layers and  $N$  nodes that can be active in one or more layers in the system. Layers follow a primal approach [27] where nodes represent intersections (that may be present in one or more layers), and links represent streets (denoted by  $s$ ), bicycle paths and designated bicycle infrastructure ( $b$ ), subways, trams and rail infrastructure ( $r$ ), or pedestrian infrastructure ( $p$ ). Construction of these intersection nodes fol-

lows the topological simplification rules of OSMnx [101]. This recent approach has been useful to demonstrate how cities grow [17, 110], how efficient [75] and dense they are, and to capture the tendency of travel routes to gravitate towards city centers [111]. Each layer  $\alpha = 1, \dots, M$  is described by an adjacency matrix  $A^{[\alpha]} = a_{ij}^{[\alpha]}$  where  $a_{ij}^{[\alpha]} = 1$  if there is a link between nodes  $i$  and  $j$  in layer  $\alpha$  and 0 otherwise. The multiplex urban system is then specified as a vector of adjacency matrices  $A = (A^{[1]}, \dots, A^{[M]})$ .



**Figure 3.2.** Schematic of multiplex layers in a city (left) and its transformation to the overlap census (right). In the overlap census, the vertical red line gives a visual separation of the left from the right half where nodes become active in the street layer. High spikes in the right half indicate car-centricity.

### 3.3 A local measure of multimodal integration

Whereas one of the simplest features of single layer networks is the degree distribution, in multiplex networks a node can have different degrees in each layer, which inform us about the multimodal potential of a city through the different roles that its intersections play. If a city has nodes that are mainly active in one layer but not in others, there is no potential for multimodality. On the contrary, in a multimodal city we expect to find many transport hubs that connect different layers, such as train stations with bicycle and street access, i.e. nodes that are active in different multiplex configurations. Note that even in a multimodally “optimal” city there will be a high heterogeneity of node activities due to the different speeds and nature of transport modes, implying, for example,

a much lower density of nodes necessary for a train network than for a bicycle network. Still, if we had a way to see and compare all combinations of node activities in the system, we could learn how much focus a city puts on connecting different modes. We can define such a fingerprint using the multiplex network formalism

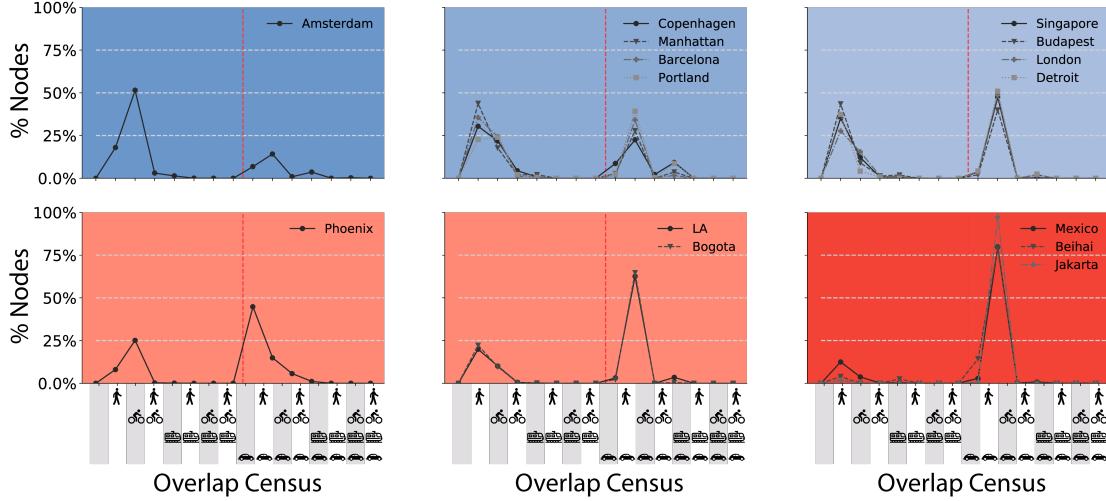
In a multimodal city, we expect to find many transport hubs that connect different layers, such as train stations with bicycle and street access, i.e. nodes that are active in different multiplex configurations. Here we propose a method to assess all such combinations of node activities in the system, helping us to learn how well connected different modes are. For each city, we build a profile based on the combinations of node activities, and refer to it as *overlap census* (Figure 3.2). The overlap census captures the percentage of nodes that are active in different multiplex configurations and provides an “urban fingerprint” of its multimodality [50]. To define the overlap census formally, given a multiplex transport network with  $M$  layers the overlap census is a vector of  $(2^M) - 1$  components, which accounts for the fractions of nodes that can be reached through at least one layer.

In Fig. 3.2 we show a schematic of how the overlap census is built: taking the multiplex network, and calculating the percentage of nodes that overlap in different configurations. The multiplex approach addresses the multimodality of a city: it not only counts how many nodes or links there are in each layer, but it shows how they are combined, revealing the possible multimodal mobility combinations in the city. Understanding the possibilities for interchange between mobility layers provides us with a better understanding of urban systems, showing us the complexity and interplay between layers.

## 3.4 Identifying clusters of cities with the Overlap Census

Due to the expected heterogeneity of node activities in different layers, the overlap census of a specific city is also expected to be heterogeneous and hard to assess on its own. Therefore, a good way to assess a city’s overlap census is by comparing it with the overlap census of other cities. We find similarities between cities via a k-means algorithm fed with fifteen vectors (one per city), where each vector contains the percentages of nodes active in each possible configuration. The algorithm separates the 15 analyzed cities into six different clusters [Fig. 3.3].

On the left half of the overlap census, we show the configurations in which nodes are not active in the street layer, while the right half contains car-related



**Figure 3.3.** Clusters of cities based on similarity of their overlap census. We find six different clusters using a k-means algorithm (coloured areas), which explain more than 90% of the variance.

configurations [Fig. 3.3]. These clusters of cities are useful to explain similarities in infrastructure planning in different transport development paths [116, 11], with clusters of car-centric urbanization (like Mexico, Beihai, and Jakarta) opposed to clusters that show a more multimodal focus in their mobility infrastructure (like Copenhagen, Manhattan, Barcelona, and Portland). In the extreme cluster that contains only Amsterdam, close to 50% of nodes are active in the bicycle layer, whereas in the Mexico-Beihai-Jakarta cluster more than 50% of nodes are active in the street-pedestrian configuration. The concentration of nodes in just one configuration informs not only about the mobility character of the city, i.e. Amsterdam being a bicycle-friendly city, but unveils the importance of explicitly considering overlooked layers and their interconnections. For example, Singapore, Budapest, London, and Detroit have two main peaks indicating that most of their nodes are either active in the street-pedestrian or only in the pedestrian configuration. This is not the case in Los Angeles and Bogota, where the majority of nodes are active in the car-pedestrian combination, i.e. the pedestrians have to share most of the city with cars. Our multimodal fingerprint unravels how different transport modes are interlaced, helping identifying which layer (or set of layers) could be improved to promote multimodal, sustainable mobility.

To summarize, we propose the new “overlap census” method based on multiplex network theory allowing to rigorously identify and compare the multimodal potential of cities.

## CHAPTER 4

---

# DATA-DRIVEN STRATEGIES FOR OPTIMAL BICYCLE NETWORK GROWTH

---

Urban transportation networks, from sidewalks and bicycle paths to streets and rails, provide the backbone for movement and socioeconomic life in cities. To make urban transport sustainable, cities are increasingly investing to develop their bicycle networks. However, it is yet unclear how to extend them comprehensively and effectively given a limited budget. Here we investigate the structure of bicycle networks in cities around the world, and find that they consist of hundreds of disconnected patches, even in cycling-friendly cities like Copenhagen. To connect these patches, we develop and apply data-driven, algorithmic network growth strategies, showing that small but focused investments allow to significantly increase the connectedness and directness of urban bicycle networks. We introduce two greedy algorithms to add the most critical missing links in the bicycle network focusing on connectedness, and show that they outmatch both a random approach and a baseline minimum investment strategy. Our computational approach outlines novel pathways from car-centric towards sustainable cities by taking advantage of urban data available on a city-wide scale. It is a first step towards a quantitative consolidation of bicycle infrastructure development that can become valuable for urban planners and stakeholders.<sup>1</sup>

---

<sup>1</sup>A stand alone version of this chapter has been published in Royal Society Open Science [24]

## 4.1 Introduction

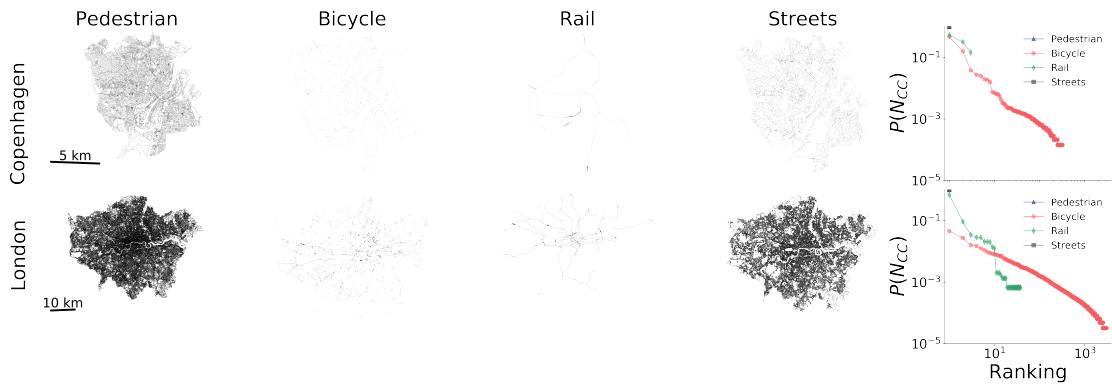
In the first two chapters we investigated the interconnected network of urban transportation systems. Here to make urban transport sustainable, we shift our attention to the bicycle layer. In particular, to the bicycle infrastructure, for which it is still not clear how to extend them comprehensively.

Most modern cities have followed a car-centric development in the 20th century [112] and are today allocating a privileged amount of urban space to automobile traffic [113, 114]. From a network perspective, this space can be described as the street layer of a larger mathematical object, the multiplex transport network [54, 53, 50]. A city's multiplex transport network, as previously shown in Chapter 3, contains other network layers that have co-evolved with the street layer, such as the bicycle layer or the rail network layer, Fig. 4.1. Due to the car-centric development of most cities, street layers are the most developed layers and define or strongly limit other layers: For example, sidewalks are by definition footpaths along the side of a street and make up a substantial part of a city's pedestrian space [113]; similarly most bicycle paths are part of a street or are built along the side.

From an urban sustainability perspective, this situation is suboptimal because the unsustainable mode of automobile transportation dominates sustainable modes like cycling. Consequently, urban planning movements in a number of pioneering cities are increasingly experimenting with drastic policies, such as applying congestion charges (London) [117] and repurposing or removing car parking (Amsterdam, Oslo) [118, 119, 120]. These efforts agree in one common goal, together with the literature on cycling safety [121, 122, 123, 124] and with cost-benefit analysis [125]: Protected bicycle lanes need to be extended considerably to create complete bicycle networks that provide a safe infrastructure for cycling citizens. Although scattered efforts in this direction have shown preliminary success, a quantitative framework for developing and assessing systematic strategies is missing [126].

In this chapter we analyze the bicycle infrastructure network from the 14 world cities we previously analyzed in Chapter 3, the cities in the study are from leading bicycle-culture countries like Netherlands, to car-centric countries like Great Britain, the USA, or Colombia. We first uncover network fragmentation within the bicycle dedicated infrastructure. Then, to improve a city's vital dedicated bicycle infrastructure [127, 128, 129, 130], we develop algorithms for connecting disconnected graphs based on concrete quality metrics from bicycle network planning [131] and apply them to the empirical bicycle networks via network growth simulations. We find that localized investment into targeted missing links can rapidly consolidate fragmented bicycle networks, allowing to

significantly increase their connectedness and directness, with potentially crucial implications for sustainable transport policy planning.



**Figure 4.1.** (*Map plots, left*) Networks representing various layers of transport infrastructure (pedestrian paths, bicycle paths, rail lines, and streets) for Copenhagen and London, with data from OpenStreetMap. (*Right*) Connected component size distribution  $P(N_{cc})$  as a function of the ranking of the component for all considered network layers and cities. All layers are well connected except the bicycle layer: Copenhagen has 321 bicycle network components despite being known as a bicycle-friendly city, while London's bicycle layer is much more fragmented, featuring over 3000 disconnected components. Copenhagen's largest connected bicycle component (leftmost data point) spans 50% of the network, but London's only less than 5%.

## 4.2 Data

We acquired street and bicycle infrastructure networks from multiple cities around the world using OSMnx [101], the same library we described in details in the previous chapter (see Chapter 3.2). OSMnx simplifies the OpenStreetMap's raw data to retain only nodes at the intersections and dead ends of streets, or any other mobility infrastructure. In this case we downloaded the pedestrian, bicycle, rail and street infrastructure. The resulting networks retain the spatial geometry of the edges, generating a length-weighted nonplanar directed graph [14]. As we have shown, the data from OpenStreetMap are of high quality [89, 90, 115] and completeness: More than 80% of the world is covered by OSM [63]. In particular, OSM's bicycle layer has better coverage than proprietary alternatives like Google Maps [132]. We collect data from a diverse set of cities to capture different development states of bicycle infrastructure networks; from consolidated networks like Amsterdam and Copenhagen, less developed ones like Manhattan and Mexico City, to rapidly developing cities like Jakarta

and Singapore. The various analyzed urban areas and their properties (number of nodes  $N$ , number of connected components  $CC$ , and population) are reported in Table 4.1. Code to replicate our results is available as Jupyter Notebooks (<https://github.com/naturaluis/bicycle-network-growth>) and data can be downloaded from Harvard Dataverse [94].

	walk			bike			rail			drive			Population
	$N$	$CC$	$\ell(km)$	$N$	$CC$	$\ell(km)$	$N$	$CC$	$\ell(km)$	$N$	$CC$	$\ell(km)$	
Amsterdam	23,321.0	1.0	2,075.67	34,529.0	355.0	972.08	1,096.0	8.0	288.72	15,125.0	1.0	2,010.49	872,680
Barcelona	20,203.0	1.0	2,122.6	7,553.0	122.0	229.19	263.0	29.0	105.98	10,393.0	1.0	1,551.44	1,600,000
Bogota	81,814.0	1.0	8,686.51	9,760.0	171.0	367.33	166.0	12.0	20.2	62,017.0	1.0	7,383.69	7,412,566
Budapest	73,172.0	1.0	7,746.12	10,494.0	257.0	336.13	1,588.0	20.0	522.06	37,012.0	1.0	5,332.97	1,752,286
Copenhagen	30,746.0	1.0	2,286.66	13,980.0	321.0	417.01	276.0	3.0	123.56	15,822.0	1.0	1,547.3	2,557,737
Detroit	47,828.0	1.0	6,769.46	3,663.0	53.0	141.06	20.0	3.0	11.54	28,462.0	1.0	5,624.49	672,662
Jakarta	140,042.0	1.0	13,947.96	248.0	19.0	8.44	60.0	8.0	81.24	138,388.0	1.0	14,194.2	10,075,310
LA	89,543.0	1.0	14,329.92	14,577.0	230.0	653.16	173.0	9.0	90.82	71,091.0	1.0	13,324.46	3,792,621
London	270,659.0	1.0	23,846.62	62,398.0	3,023.0	1,281.71	2,988.0	38.0	1,045.39	179,782.0	1.0	18,154.52	8,908,081
Manhattan	13,326.0	1.0	1,320.78	3,871.0	105.0	111.42	349.0	5.0	197.51	5,671.0	1.0	1,022.13	1,628,701
Mexico	108,033.0	1.0	14,547.18	5,218.0	52.0	332.37	371.0	18.0	253.48	95,375.0	1.0	13,732.39	8,918,653
Phoenix	111,363.0	1.0	14,314.0	35,631.0	141.0	1,221.18	105.0	4.0	71.64	73,688.0	1.0	11,841.49	1,445,632
Portland	50,878.0	1.0	5,324.78	24,252.0	198.0	596.36	230.0	2.0	132.36	35,025.0	1.0	4,583.47	583,776
Singapore	82,808.0	1.0	8,633.13	12,981.0	104.0	339.39	683.0	14.0	428.66	50,403.0	1.0	6,635.37	5,638,700

**Table 4.1.** Measures for the administrative area of analyzed cities. The number of connected components ( $CC$ ) and nodes ( $N$ ) for each layer in all cities of our dataset are highly diverse due to the varying developmental levels and focus of transport.

We characterize each city street and bicycle infrastructure as a primal network, [27] in which nodes are intersections, while links represent bicycle paths, and designated bicycle infrastructure. This recent approach has been useful to demonstrate how cities grow [17, 110], how efficient [75] and dense they are, and to capture the tendency of travel routes to gravitate towards city centers [111]. This network is described by an adjacency matrix  $A = \{a_{ij}^{[\alpha]}\}$  where  $a_{ij} = 1$  if there is a link between nodes  $i$  and  $j$  and 0 otherwise.

### 4.3 Defining bicycle network growth strategies and quality metrics

Across all cities considered, we find that almost all network layers are made up of one giant component, except for the bicycle layer which is always fragmented into many disconnected components (see Table 4.1). This discovery is remarkable given that the fragmentation occurs also in bicycle-friendly cities like Copenhagen (Fig. 4.1), showing that cycling infrastructure can be suboptimal even in the leading cycling cities on the planet. To quantify such an underdevelopment in the sustainable mobility infrastructure of cycling, we focus on the single layer of bicycle networks and on two well-established metrics in

bicycle infrastructure quality assessment [133, 134, 135, 131, 136]: *connectedness* and *directness*. Connectedness indicates “the ease with which people can travel across the transportation system” [131], and it is related to answering the question “can I go where I want to, safely?”. Directness addresses the question “how far out of their way do users have to travel to find a facility they can or want to use?”, and can be measured by how easy it is to go from one point to another in a city using bicycle infrastructure versus other mobility options, like car travel.

As our main approach, we choose to measure connectedness and directness over the designated bicycle infrastructure only, without considering travel on streets. Although it is possible to cycle on streets, growing evidence from bicycle infrastructure and safety research is unveiling serious safety issues for cycling when mixed with vehicular traffic [121, 122, 123]. However, we also tested our algorithms on a combination of bicycle infrastructure plus streets for which the maximum speed is 30 km/h, following common best-practice reasoning that low speed limits can make streets safe for cycling [137]. The results of these additional simulations are available in the Supplementary Information; they do not differ significantly from the case of designated bicycle infrastructure presented below, as the developed algorithms follow the same rules to connect the multiple components in both cases of segregated bicycle infrastructure only and of included bikeable streets.

To quantify connectedness, we first measure the number of disconnected components of each city’s bicycle network. It is no surprise that car-centric cities have a highly fragmented bicycle infrastructure: for example, London has more than 3,000 disconnected bicycle infrastructure segments. However, even bicycle-friendly cities like Copenhagen have over 300 disconnected bicycle path components – see the connected component size distribution  $P(N_{cc})$  in Fig. 4.1. This infrastructure fragmentation in the bicycle layer poses a challenge for a city’s multimodal mobility options [61] and for the safety of its cycling citizens [138, 139].

### 4.3.1 Algorithmic approach

There are various approaches in developing automated strategies for bicycle infrastructure planning. Hyodo et al. [140] have proposed a bicycle route choice model to plan bicycle lanes taking into account facility characteristics. Other studies have used input data from bicycle share systems [141] or origin destination matrices [142] to plan bicycle lanes. More recently, taxi trips have been used to identify susceptible clusters for bicycle infrastructure [143]. Here we attempt an alternative approach: Since hundreds of bicycle network components already exist in most cities, we aim at consolidating the existing infrastructure by making strategic connections between components rather than starting from

scratch.

---

**Algorithm 1** Largest-to-Second. The algorithm takes the bicycle network  $G$  and a list of its weakly connected components  $wcc$ , then it iterates over the weakly connected components, sorts them by their size (number of nodes inside each component), locates the closest pairs of nodes between the first and the second components. The process is repeated until all the components have been connected.

---

```

1: procedure L2S
2:    $G \leftarrow$  bicycle network graph
3:    $wcc \leftarrow$  components of network  $G$ 
4:   for  $i$  in  $\text{length}(wcc)-1$  do
5:     sort  $wcc$  by components size
6:      $cc \leftarrow$  two biggest components from  $wcc$ 
7:      $i\_j \leftarrow$  closest nodes between  $cc_0$  and  $cc_1$ 
8:     connect  $cc_0$  and  $cc_1$  in  $i\_j$ 
```

---



---

**Algorithm 2** Largest-to-Closest. The algorithm takes the bicycle network  $G$  and a list of its weakly connected components  $wcc$ , then it iterates over the weakly connected components, sorts them by their size (number of nodes inside each component), locates the largest connected component and the closest of the remaining components, the components are connected. The process is repeated until all the components have been connected.

---

```

1: procedure L2C
2:    $G \leftarrow$  bicycle network graph
3:    $wcc \leftarrow$  components of network  $G$ 
4:   for  $i$  in  $\text{length}(wcc)-1$  do
5:     sort  $wcc$  by components size
6:      $cc_0 \leftarrow$  biggest component from  $wcc$ 
7:      $cc_n \leftarrow$  closest component to  $cc_0$ 
8:      $i\_j \leftarrow$  closest nodes between  $cc_0$  and  $cc_n$ 
9:     connect  $cc_0$  and  $cc_n$  in  $i\_j$ 
```

---

Our approach takes into account the currently available bicycle infrastructure and uses an algorithmic process to improve the network by finding the most important missing links step by step. This way we focus on optimizing the connectedness metric, growing the bicycle infrastructure by making it more connected, merging parts into fewer and fewer components. We develop two iterative greedy algorithms that we check against a random and a minimum

investment approach. The first algorithm, *Largest-to-Second* (L2S), identifies in each step the largest connected component in the bicycle infrastructure network and connects it to the second largest (see algorithm 1 for details). The second algorithm, *Largest-to-Closest* (L2C), also identifies the largest connected component, but connects it to the closest of the remaining bicycle infrastructure components (see algorithm 2 for details). In both algorithms, components are connected through a direct link between their two closest nodes. We use this technique as an approximation to the underlying street-shortest path – since the most relevant shortest 100 connections typically range from 14 to 500 meters, roughly the length of two blocks, this approximation is reasonable. The algorithms repeat this process until there are no more disconnected components in the network.

To have a random baseline, we compare our algorithms with a *Random-to-Closest* (R2C) component approach. In each step of this baseline approach, one component is picked at random and connected with the closest remaining one (see algorithm 3 for details). This baseline allows us to model a scenario where infrastructure is developed following a systematic but random linking approach – in urban development this corresponds to uncoordinated local planning that randomly connects close pieces of bicycle infrastructure. We also implement a second baseline, the extreme case of *Closest-Components* (CC), which prioritizes connecting the closest two components disregarding their size (see algorithm 4 for details). This CC approach is equivalent to an “invest as little as possible” development strategy – it builds up a minimum-spanning-tree-like structure following a modified Kruskal’s algorithm [144]. All four algorithms connect components optimizing a well-defined criterion, finding the critical missing links in the network, and adding one new link per iteration. See Fig. 4.2 for a schematic of the four algorithms.

We apply the algorithms to the bicycle infrastructure inside the political demarcation of the cities, however it is possible to extend the methods and include bicycle highways and cross-city trails, since they use as input a set of spatial network components to connect. We opt to not include cross-city links, since they are a special case only available in a few regions and where adequate intra-urban bicycle infrastructure has already been established [145, 146].

### 4.3.2 Evaluating the algorithms impact

To test how much cities improve their bicycle layers using these four algorithms, we define two metrics on the bicycle layer that operationalize the notion of connectedness: i)  $n_{LCC} = \frac{N_{LCC}}{N}$ , the fraction of nodes from the bicycle infrastructure inside the largest connected component ( $N_{LCC}$ ) compared to the total number of

---

**Algorithm 3** Random-to-Closest. The algorithm takes the bicycle network  $G$  and a list of its weakly connected components  $wcc$ , then it iterates over the weakly connected components, randomly picks a component and connects it to the closest of the remaining components. The process is repeated until all the components have been connected.

---

```

1: procedure R2C
2:    $G \leftarrow$  bicycle network graph
3:    $wcc \leftarrow$  components of network  $G$ 
4:   for  $i$  in  $\text{length}(wcc)-1$  do
5:      $cc_{ran} \leftarrow$  random component from  $wcc$ 
6:      $cc_n \leftarrow$  closest component to  $cc_{ran}$ 
7:      $i\_j \leftarrow$  closest nodes between  $cc_{ran}$  and  $cc_n$ 
8:     connect  $cc_{ran}$  and  $cc_n$  in  $i\_j$ 
```

---

**Algorithm 4** Closest-Components. The algorithm takes the bicycle network  $G$  and a list of its weakly connected components  $wcc$ , then it iterates over the weakly connected components, calculate the distance between available components and connect the two closest ones. The process is repeated until all the components have been connected.

---

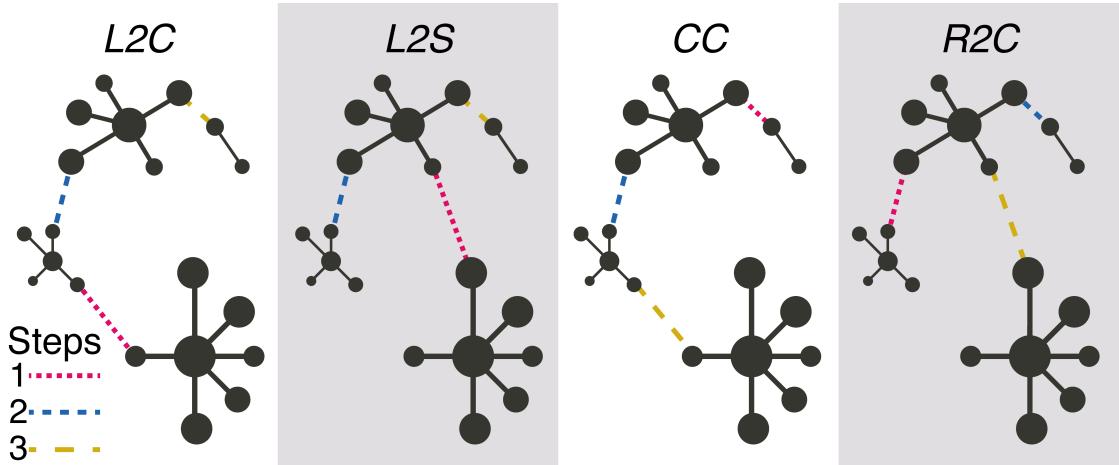
```

1: procedure CC
2:    $G \leftarrow$  bicycle network graph
3:    $wcc \leftarrow$  components of network  $G$ 
4:   for  $i$  in  $\text{length}(wcc)-1$  do
5:      $\Delta_{min} \leftarrow$  closest components in  $wcc$ 
6:      $cc_0 \leftarrow$  first component for  $\Delta_{min}$ 
7:      $cc_1 \leftarrow$  second component for  $\Delta_{min}$ 
8:      $i\_j \leftarrow$  closest nodes between  $cc_0$  and  $cc_1$ 
9:     connect  $cc_0$  and  $cc_1$  in  $i\_j$ 
```

---

nodes from the same type of infrastructure ( $N$ ), and ii)  $\ell_{LCC} = \frac{L_{LCC}}{L}$ , the fraction of link kilometers inside the bicycle infrastructure largest connected component ( $L_{LCC}$ ) compared to the total number of link kilometers in the bicycle network ( $L$ ). Both metrics take values between 0 and 1, where 1 means that there is only one connected component. An intermediate value, for example 0.2, means that the largest connected component contains 20% of all bicycle intersections or path kilometers. Executing our algorithms step by step these metrics can only grow, approaching 1 when the process is complete and they terminate. What distinguishes the algorithms is *how fast* these values grow.

We quantify directness through the metric: iii) bicycle-car directness  $\Delta$ ,



**Figure 4.2.** Schematic representation of algorithms to improve bicycle network infrastructure: Largest-to-Closest (L2C) finds the largest component and connects it with the closest one; Largest-to-Second (L2S) connects the largest component with the second largest; Closest-Connected (CC) connects the two closest components; and Random-to-Closest (R2C) picks a random component and connects it to the closest.

which answers the question “how direct are the average routes of bicycles compared to cars?” via the ratio between average distance by car and average distance by bicycle. For example, if the shortest car-route from west to east Manhattan is 4 km and the shortest route on the bicycle network between these two points is 5 km, the bicycle-car directness is  $4/5 = 0.8$ . Note that if the bicycle network is a subset of the street network, then  $\Delta$  cannot be larger than 1. Formally we write  $\Delta = \frac{\langle \delta_{ij}^b \rangle_{ij}}{\langle \delta_{ij}^s \rangle_{ij}}$ , where  $\langle \delta_{ij}^s \rangle_{ij}$  is the average car-route distance, and  $\langle \delta_{ij}^b \rangle_{ij}$  is the average length of the shortest bike-route between  $i$  and  $j$ . In each iteration of any of our algorithms, we implement this measure by randomly selecting one thousand pairs of origin-destinations nodes and then averaging the corresponding street/bicycle distance. To avoid undefined values due to disconnected components in the bicycle layer, we add the following condition: If a node from the pair  $i$  and  $j$  is in a different component, we assign the value  $\delta_{ij}^b = 0$ . This condition also ensures consistency of growing directness values while the algorithm merges more and more nodes into the same component.

Finally, in order to measure the cumulative efficiency of our algorithms, we define the metric: iv)  $G_{LCC}$  as the relative gain of bicycle path kilometers in the largest connected component. For example,  $G_{LCC} = 1.5$  means that the algorithm has increased the largest connected component’s original size by 150%.

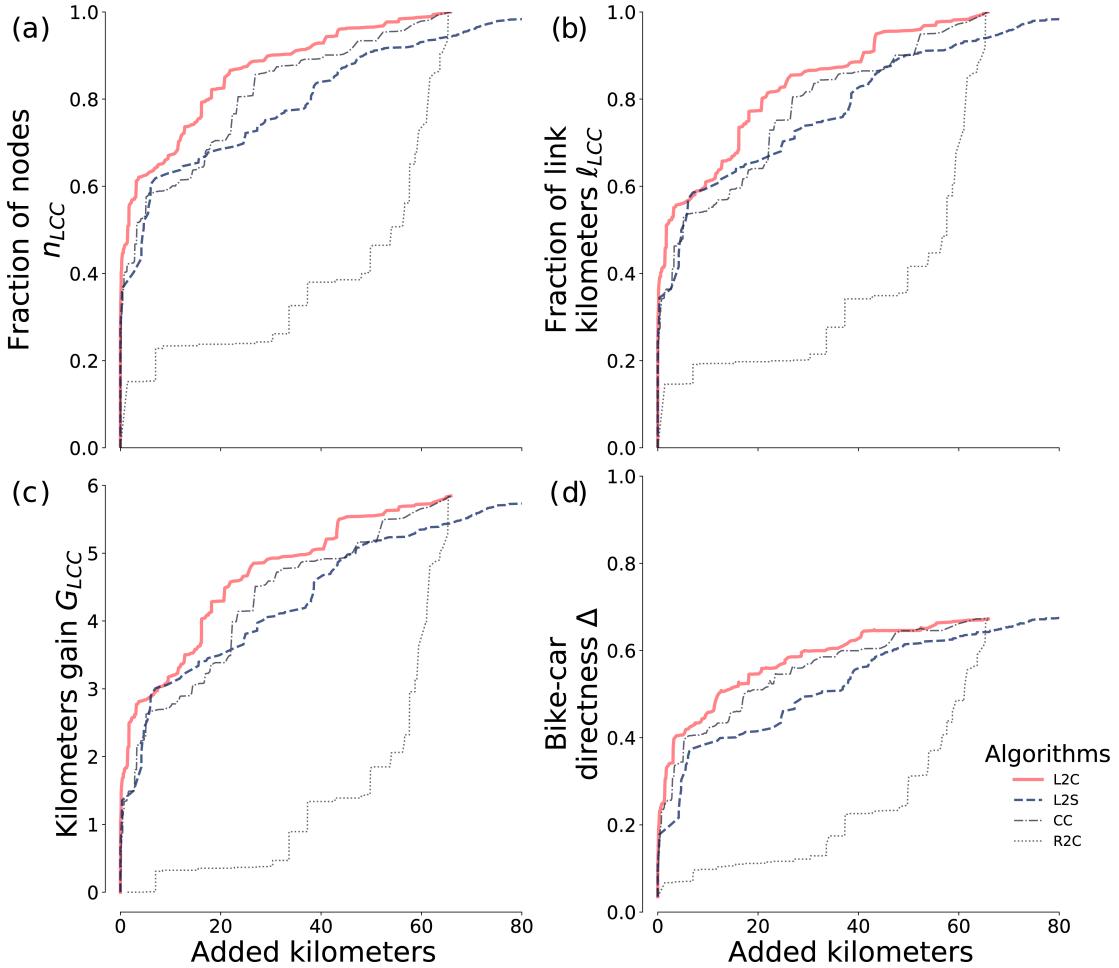
Formally,  $G_{LCC} = \frac{L_{LCC} - L_{LCC_0}}{L_{LCC_0}}$ , where  $L_{LCC_0}$  is the sum of kilometers in the largest connected component before the algorithm runs. As with all other metrics,  $G_{LCC}$  is monotonically increasing with the growth algorithm, and reaches  $\frac{1 - \ell_{LCC_0}}{\ell_{LCC_0}}$  at the end of the dynamics.

## 4.4 Growing bicycle networks shows stark improvements with small investments

We demonstrate in Fig. 4.3 the power of the various growth strategies by showing the initial state of the bicycle layer for the case of Budapest and its state after 85 iterations of the *Largest-to-Closest* algorithm: At this point the network has almost quadrupled the size of its largest connected component (from 82 km to 313 km), with a negligible investment of just less than 5 km (corresponding to 1.4% of the previously existing bicycle infrastructure) in new connecting bicycle paths. In terms of connectedness, it goes from 15% to 56% connected. This rapid increase shows that the city can easily improve its bicycle infrastructure with small investments. For some extreme cases, like Bogota, with the same 5 km investment (an increase of 1.3% to the previously existing infrastructure) the bicycle-car directness increases from 6% to almost 48% and connectedness from 34% to 89%. [Similar encouraging results hold for other cities \(See Section 7.2.2\)](#).

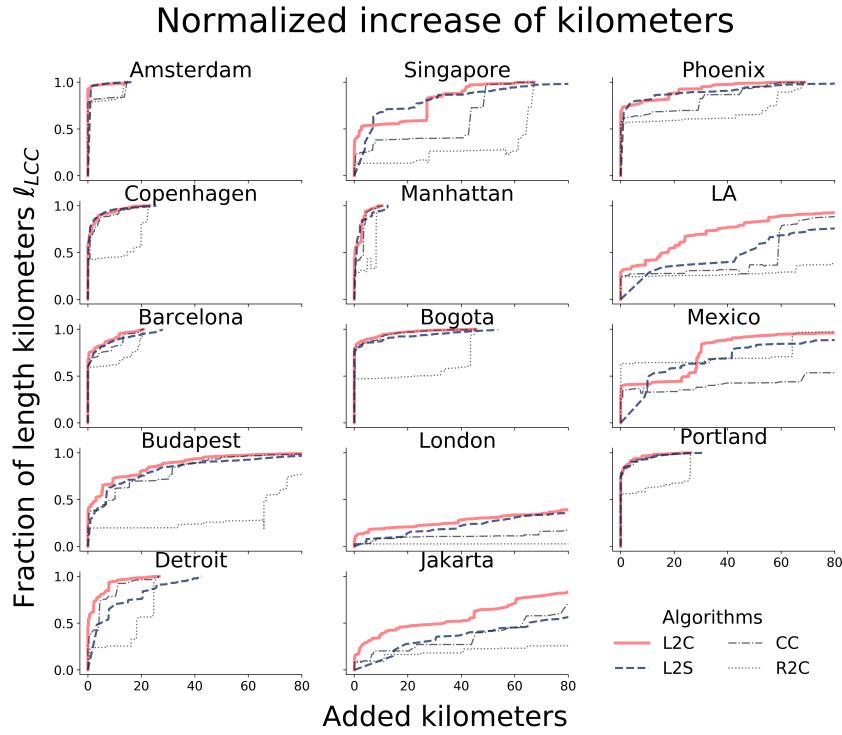
The fraction of nodes inside the largest connected component increases rapidly with newly added links for all considered algorithms except *Random-to-Closest*, Fig. 4.3(a). The *Largest-to-Closest* algorithm performs better than the others, even more than *Closest-Components* which prioritizes minimum investments in the network. Since we are considering bicycle infrastructure, a better practical measure than the number of intersections is the number of kilometers that can be cycled using only designated paths. Figure 4.3(b) shows how this measure improves in a similarly explosive way: with an investment of only 20 km (5.9% of the existing infrastructure), the largest connected component will contain 80% of the original bicycle infrastructure. Results for the kilometer gain  $G_{LCC}$  are shown in Fig. 4.3(c). Three of the four algorithms rapidly gain new kilometers, but as the invested new kilometers grow, each algorithm follows a different gain rate. Also for this metric, *Largest-to-Closest* is the algorithm with the best performance.

We also measure the bicycle-car directness ratio, Fig. 4.3(d). The bicycle-car directness  $\Delta$  improves as the algorithms consolidate the network. These improvements are, however, indirectly driven by the improvement of connect-



**Figure 4.3.** (a) Normalized increase in nodes inside the largest connected component ( $n_{LCC}$ ). (b) Normalized increase in kilometers inside the largest connected component ( $\ell_{LCC}$ ). (c) Kilometers gain ( $G_{LCC}$ ). (d) Bicycle-car directness ( $\Delta$ ). Measures in (b-e) are plotted as a function of the sum of added links in kilometers, for the case of Budapest (for all cities see Fig. 7.1)

edness, which boosts the accessibility of bicycles to different areas of the city. The flattening of the curves at a value considerably smaller than 1 (around 0.65) shows that cars will always outperform bicycles in terms of directness, having on average at least 33% shorter paths in the city. This suboptimal flattening is a natural consequence of the algorithms optimizing for connectedness only, not adding “redundant” connections. Nevertheless, the measure shows that, similar to connectedness, with a relatively negligible investment of bicycle path kilometers into the system, the bicycle network’s directness improves drastically, even in the greediest case where the shortest possible missing link is added in



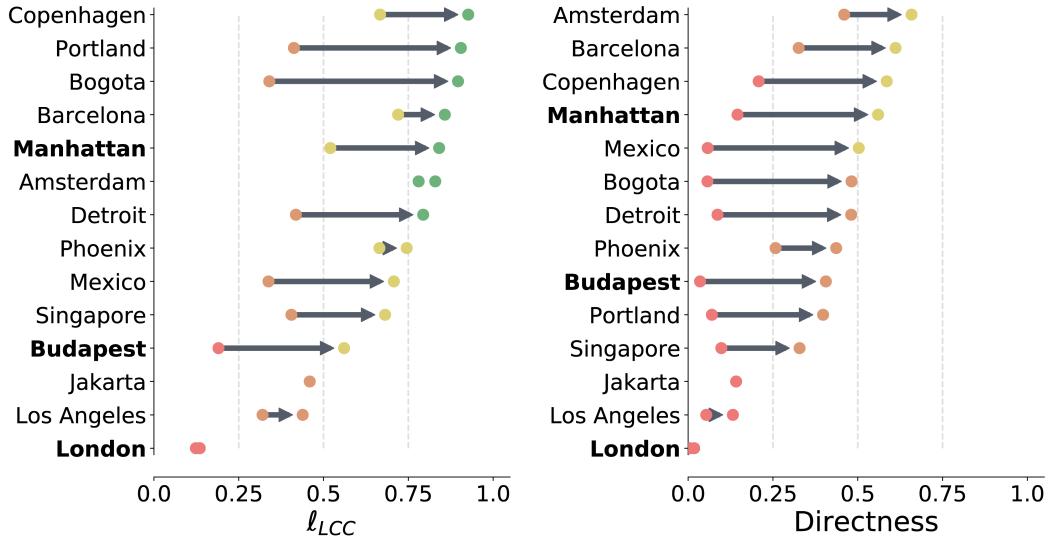
**Figure 4.4.** Normalized increase in kilometers inside the largest connected component ( $\ell_{LCC}$ ).

every iteration. This result holds for all analyzed cities (see SI). The large differences between the baseline *Random-to-Closest* and our two algorithms (*Largest-to-Second* and *Largest-to-Closest*) show the importance of following an approach that consolidates and grows the largest connected component. Similar results are obtained for all the cities when taking into account the combination of bicycle infrastructure and safely bikeable streets ( $\leq 30\text{ km/h}$ ), in Figure 4.4 we show the increase in bikable kilometers when taking into consideration bikable streets and designated bicycle infrastructure combination for all the analyzed cities (see Section 7.2.3 for the rest of the measures).

## 4.5 Different cities have different optimal investment strategies

Differences arise in the state of the bicycle layer and its improvement after applying a growth algorithm. To see this effect, we rank how cities improve using

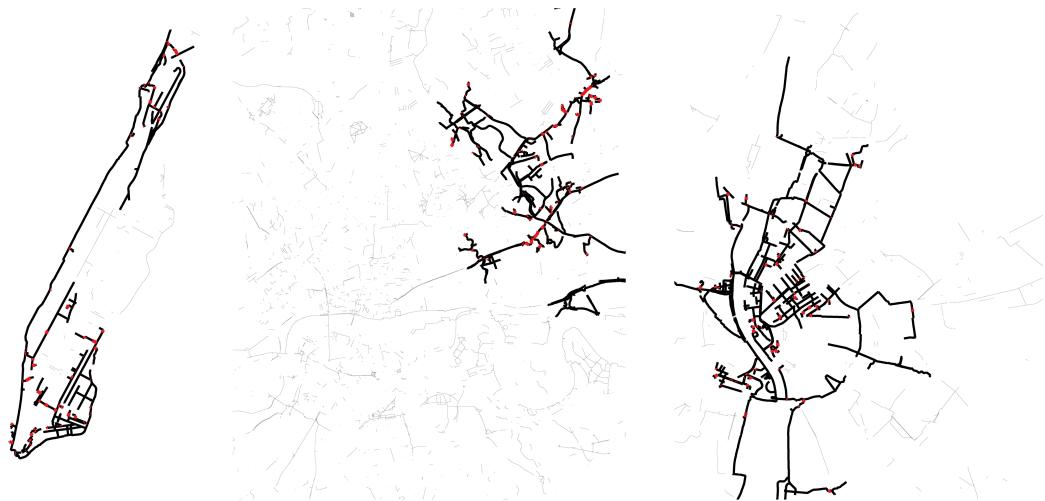
### 5 km investment



Manhattan

London

Budapest



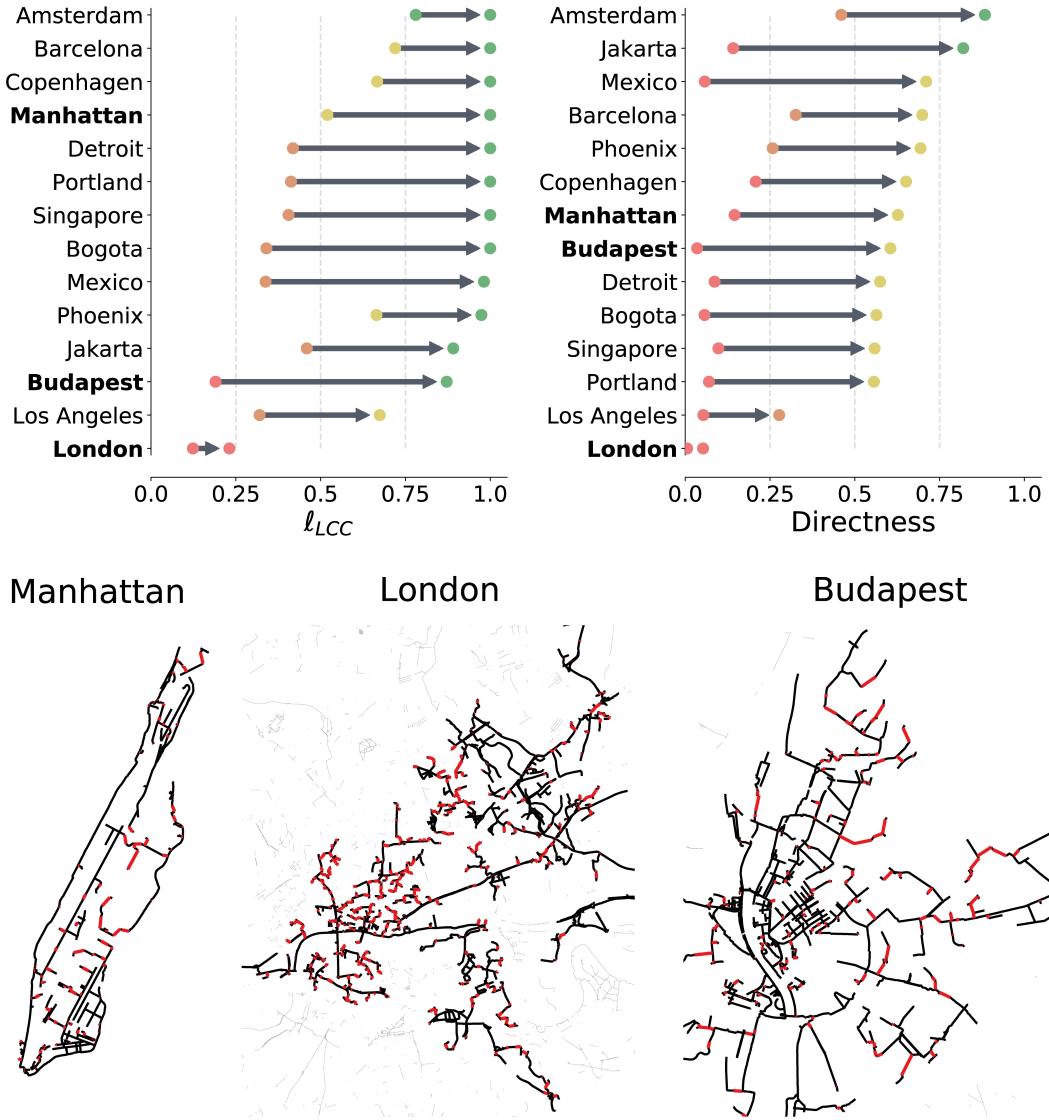
**Figure 4.5.** Cities improvement and ranking using the Largest-to-Closest algorithm. We report the improvement and ranking on the fraction of total kilometers of bicycle infrastructure in the largest connected component ( $\ell_{LCC}$ ) and in the bicycle-car directness ( $\Delta$ ). Dotted lines show thresholds of 25%, 50%, and 75%. Plots (a-b) show investment strategies of 5 km and 35 km, respectively, the Manhattan, London, and Budapest plots show the suggested new links (red) after adding 5 km and 35 km, the newly created largest connected component (black), and the remaining separated components (grey).

the *Largest-to-Closest* algorithm in two different investment scenarios: investing either i) 5 km, or ii) 35 km. Figure 4.5 shows how cities improve when investing 5 km of bicycle infrastructure. We see that some cities get above 75% of their existing infrastructure connected, meaning that their bicycle layer only needs a small extension. On the other hand, cities like London, Los Angeles, and Jakarta need a larger investment to improve. Concerning bicycle-car directness, cities reach lower values due to the focus of the algorithms on completeness. In the worst performing cities like Los Angeles, a covered length close to 50% can be reached easily, while the bicycle-car directness ratio stays below 20%, showing that it is much harder to gain an acceptable bicycle infrastructure in cities where cars are overprioritized. The 35 km investment strategy shows that most cities can get at least 75% of their bicycle infrastructure connected, Fig. 4.6. The worst performing outlier is London, due to its bicycle layer containing more than 3000 connected components scattered around 1600 km<sup>2</sup> (see Table 4.1). In terms of bicycle-car directness London also performs badly, while Amsterdam is the best performing one.

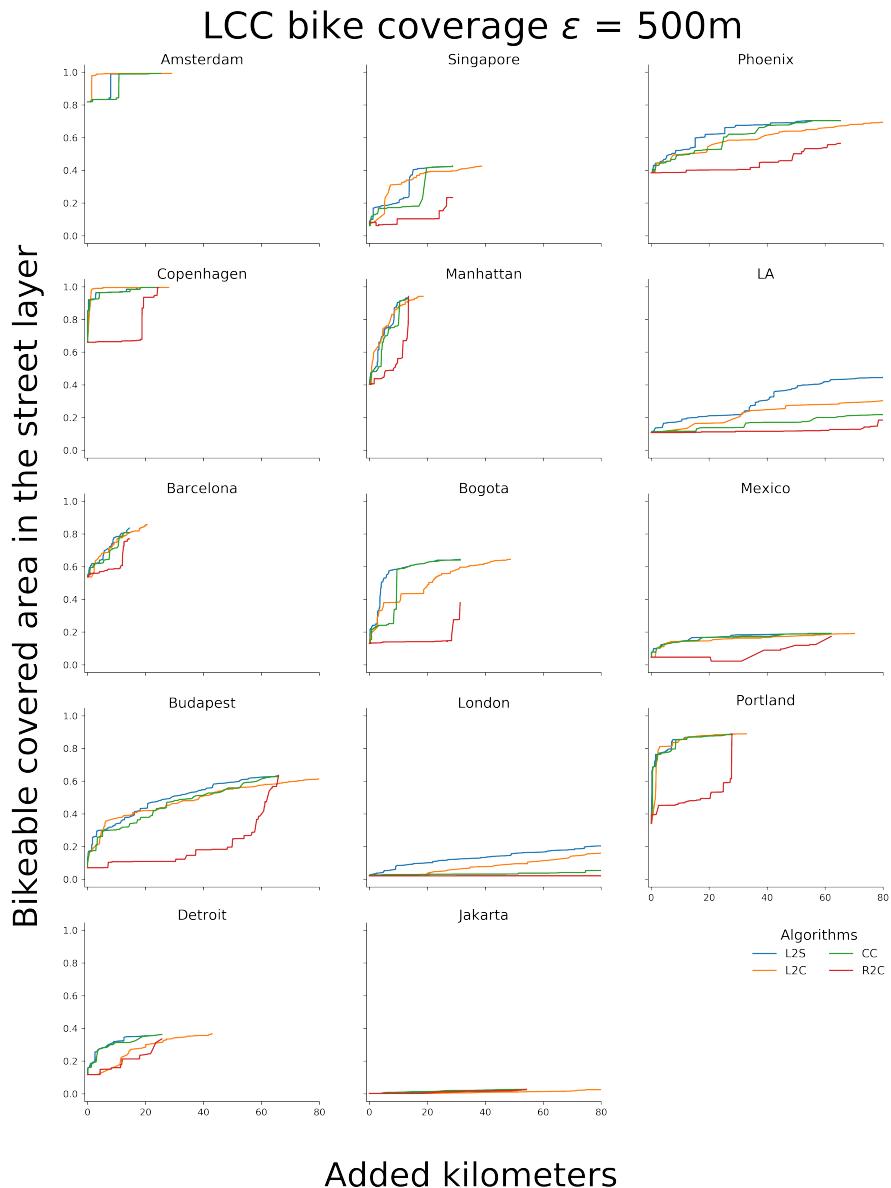
The four proposed metrics capture the impact of newly created connections on the various components of the bicycle network. By linking previously disconnected neighbourhoods with a sustainable mode of transport, our approach focuses on consolidating bicycle infrastructure networks, thus making cities more cohesive and green. It does not, however, focus on growing the bicycle network into large areas of the city not currently served. To test to which extent such connectedness-based algorithms bring an indirect benefit for coverage, we measured the proportion of the city that is covered and reachable by bicycle with an epsilon of 500 meters around the bicycle infrastructure of the largest connected component, and calculated the percentage of nodes in the street layer that are covered by the bikeable area. The results of this measurement show a wide range of effects: Cities with an already high coverage above 80% (Amsterdam, Copenhagen) reach near instantly 100%, cities with an intermediate coverage (Manhattan, Bogota, Budapest) follow a more linear progression per added kilometer, while underdeveloped or sprawling cities (LA, London, Jakarta) show negligible growth (Fig. 4.7).

While our present goal (consolidation of the bike network) is intended to show the potential of our approach, an extension towards the exploration of new city areas will increase further the real-world applicability of our results. We consider this extension an interesting line of future research in the challenge of developing optimal data-driven strategies of transport network growth, potentially informed by theoretical frameworks such as optimal percolation [147, 148]. Besides, the use of other network metrics, such as network efficiency [40], might unveil new dimensions characterising the impact of the

### 35 km investment



**Figure 4.6.** Cities improvement and ranking using the Largest-to-Closest algorithm. We report the improvement and ranking on the fraction of total kilometers of bicycle infrastructure in the largest connected component ( $\ell_{LCC}$ ) and in the bicycle-car directness ( $\Delta$ ). Dotted lines show thresholds of 25%, 50%, and 75%. Plots (a-b) show investment strategies of 5 km and 35 km, respectively, the Manhattan, London, and Budapest plots show the suggested new links (red) after adding 5 km and 35 km, the newly created largest connected component (black), and the remaining separated components (grey).



**Figure 4.7.** Area of influence (coverage) for the bicycle infrastructure in the street layer.

proposed algorithms on the development of the bicycle infrastructure.

## 4.6 Discussion

Our starting point showed that a common characteristic of cities is the fragmentation of their bicycle networks. We have proposed the use of data-driven algo-

rithms to consolidate bicycle network components into connected networks to improve efficiently sustainable transport. We have shown that connecting the bicycle infrastructure in an algorithmic way rapidly improves the connectedness and directness of the bicycle layer. These algorithms, when compared with two baselines, highlight the usefulness of growing the bicycle network on a city-wide scale (considering all areas of the city) rather than randomly adding local bicycle infrastructure. Improving the connectivity of bicycle lanes and paths improves not only the network itself, but also promotes the use of bicycles as means of transportation in a city, improving the health of its inhabitants [149].

Improving bicycle infrastructure one link at the time (by identifying suitable components to connect) is only the first step towards a systematic framework for realistic bicycle network growth strategies. Our current approach is not the last word in this development, since it does not yet explicitly optimize for directness and does not account for transport flow. Further, our proposed approach helps implement a more connected transport network which can improve the possibilities for multimodal transport. This could be starting point for implementing truly multimodal strategies, such as integration with public transportation, or bicycle parkings in transportation hubs [131].

In our algorithms, each new link works as a bridge between components, potentially having large betweenness centrality. Such high-betweenness segments could become overused and create bottlenecks in practice. To improve this situation, it would be necessary to create links in the network that act as redundant paths. In doing so, directness and coverage would also be improved, along with the network's robustness to interruptions. This is an interesting and possibly demanding task that we leave for future research, as the new links would have to be created in a coherent manner balancing trade-offs between network structure and mobility dynamics. We anticipate that complementing OpenStreetMap data with additional information on the use of traffic flow and movement data, like trips from bike share systems or origin-destination matrices, possibly from alternative sources such as municipalities and transportation agencies, might further improve the algorithms by better detecting underserved and optimal areas in the city where new links should be created. Despite these various possibilities for qualitative updates to the studied growth strategies, our first models have demonstrated the capability to generate substantial improvements with minimal effort.

The use of data-driven algorithms to identify crucially missing links in bicycle infrastructure has the potential to improve the mobility infrastructure of cities efficiently and economically. This approach is not only useful for planning city structure, but could also be used together with simulating mobility flows and to provide insights on how the system will behave after new measures are

implemented. Ultimately, planning cultures and processes will also have to be accounted for [150]. We anticipate that a future stream of work should include longitudinal studies [151] in multiple cities, along with algorithmic simulations to first model and simulate possible changes to the transport network, and then to test those models with ground truth data, to compare the evolution of infrastructure and mobility dynamics between cities with different transport priorities.

## CHAPTER 5

---

### LIFE QUALITY AS WALKABILITY

---

Life quality in cities is deeply related to the mobility options, and how easily one can access different services and attractions. The pedestrian infrastructure network provides the backbone for social life in cities. While there are many approaches to quantify life quality, most do not take specifically into account the walkability of the city, and rather offer a city-wide measure. Here we develop a data-driven, network-based method to quantify the liveability of a city. We introduce a life quality index (LQI) based on pedestrian accessibility to amenities and services, safety and environmental variables. Our computational approach outlines novel ways to measure life quality in a more granular scale, that can become valuable for urban planners, city officials and stakeholders. We apply data-driven methods to Budapest, but as having an emphasis on the online and easily available quantitative data, the methods can be generalized and applied to any city.<sup>1</sup>

---

<sup>1</sup>A stand alone version of this chapter has been published in Complex Networks and Their Applications VIII [21]

## 5.1 Introduction

In this chapter, we shift our attention from the bicycle layer, to the pedestrian one. We focus on measuring the liveability of a city using its pedestrian infrastructure to compute how easily it is to access different types of amenities that serve multiple purposes, such as cultural, educational and access to day-to-day goods.

As we have mentioned before, the evolution that most cities followed during the 20th century promoted a car-centric vision [112, 113, 114], neglecting the presence of other mobility options. From a liveability perspective, this situation is suboptimal because the automobile infrastructure dominates and defines the walkable area, increasing car traffic, air pollution and deteriorating walkable conditions.

The concept of walkability is an important factor to consider in connection with liveability. Liveability refers to an environment from an individual perspective [152] which includes "a vibrant, attractive and secure environment for people to live, work and play and encompasses good governance, a competitive economy, high quality of living and environment sustainability" [153]. Thus in a liveable city, there must be an emphasis not only on sustainable transportation and built environment to reduce the harm on nature [154, 155] but also encouraging citizens to walk for supporting their physical and mental well-being [156]. However, improving walkability is more complex than we would think. Walking should be an available, safe and well-connected mode of transportation, but as Speck put it well, it should be interesting and comfortable as well, to have a feeling of the streets as 'outdoor living rooms' [157].

The pedestrian infrastructure that sustains walkability in a city can be described as a network [27]. This approach has been useful to identify street patterns [8, 11] and its evolution [17, 110], measure the morphology of cities [158], and how the streets connectivity impacts on pedestrian volume [159].

The various approaches to create a walkability index or so-called walk score consider mainly the following components: safety and security [160, 161]; convenience, attractiveness and public policy [162, 157], connectedness [163], but also reckon with the land use mix and residential density of the certain area[164]. Another approximation rather accents the importance of its effect on air pollution, health problems, travel costs and even on the sense of community[165]. Thus measuring walkability not only captures the propensity to walk in a city but also includes the components a liveable city must have and support, under the umbrella of sustainability.

There are good examples of how sustainable city development initiatives tackle growing inequalities with data-driven approaches. Long Island used city

data to analyze which amenities are needed to increase the quality of life in a newly built environment [166], other cities are investing in smart technologies to develop public transport, connecting spatially discriminated areas [167, 168].

Since the number of components which should be taken into consideration in creating a walkability index is high, the types of data are also mixed and thus difficult to integrate. While the information on connectedness, security, residential density, etc. is quantitative and in general easily available, gaining opinion about attractiveness, convenience, or even about the feeling of security is more complicated. Here we propose to use a data-driven approach as a proxy to quantify life quality, making it reproducible and easily expanded to include different data sources. We apply our methods to Budapest, but as having an emphasis on the online and easily available quantitative data, the methods can be generalized and applied to any city.

## 5.2 Data

We work with three different data sources: networks, points of interest and city attributes. To acquire the pedestrian network and points of interest, we used OSMnx [101], the same tool we used in the previous chapters (see Chapter 2.5). As we have shown, the data from open street maps is of high quality [89, 90, 115, 63].

The majority of points of interest were downloaded from OpenStreetMap, from different classification keys (amenity, tourism, shop, office, leisure) using OSMnx [101]. We filtered the points of interest using the districts' demarcation [169], to get only the data within Budapest boundaries, having, as a result, more than 39,000 data points. We complement the data sets with secondary data sources as specialized directories of doctors and childcare facilities (see Section 7.1.1).

We categorize the points of interest in six main categories: I) Family friendliness (Access to education and daycare, and family support services), II) Access to health care and sport facilities, III) Art and culture (e.g. museums, exhibitions), IV) Nightlife (e.g. bars, restaurants), V) Environment (air quality and access to green areas), and VI) Public Safety.

The district-level data (population and crimes) were obtained from the Hungarian Police's public database. The crime statistics are reported as the number of crimes committed in public places in 2018 [170]. Population data is coming from the 2016 micro-census conducted by the Hungarian Statistical Bureau [171]. We took into account the air pollution, this data set coming from National Air Pollution Measurement Network [172], containing the geolocation of the air quality stations and different measures (annual median concentration

of carbon monoxide, nitrogen dioxide, and PM10 dust).

Accuracy of the Life Quality Index (LQI) model highly depends on how comprehensive the distribution of listed services. We use OSM as our key data source, but to achieve a more comprehensive and country-specific database we collect publicly available data from various Hungarian websites for each category.

The network contains all the sidewalks and pedestrian designated infrastructure, it is conceptualized as undirected, nonplanar and primal network [27]. The pedestrian network is described as a weighted graph, with its adjacency matrix  $W = w_{ij}$  where the weight  $w_{ij}$  contains the length between  $i$  and  $j$  if connected, and 0 otherwise.

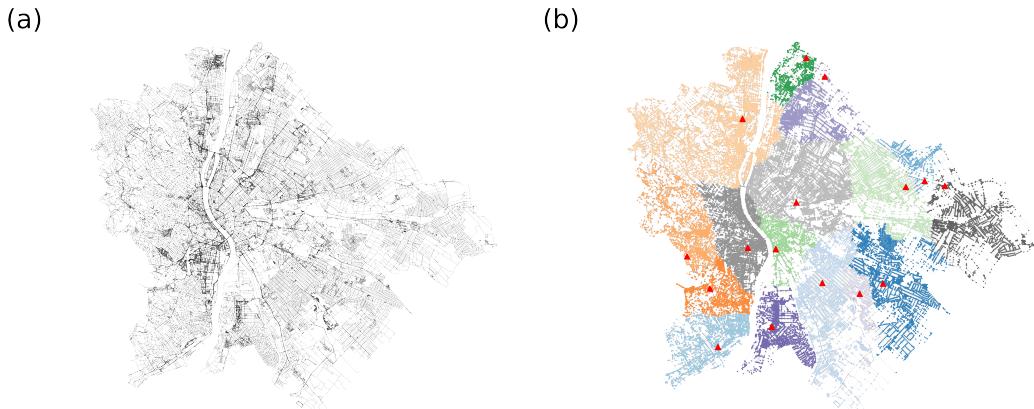
We assigned properties to the nodes of the network, matching the nodes with their corresponding districts, then assign nodes as attributes based on the district level data (population and crimes, see section 5.3.2). For the pollution data, we calculated the corresponding Voronoi cells for the air quality stations, and matched the nodes with them. Finally, we divided the pollution by the number of nodes in each corresponding cell and assigned the value to the nodes (See section 5.3.3). For the edges, we encoded their length  $\ell_{ij}$  along with the traversal time  $Tt_{ij}$  between nodes  $i$  and  $j$  calculated as  $Tt_{ij} = \frac{\ell_{ij}}{ps}$  where  $ps$  is the pedestrian speed as a constant rate of 5km/h.

## 5.3 Quantifying Life Quality

The life quality of a person is largely subjective and hard to quantify. However, it is both intuitive and has been scientifically shown that the environment and personal well-being strongly correlate [173]. Thus using environmental factors as proxies, life quality and livability becomes quantifiable [174].

The main environmental factors we consider in our model are: the availability of services and amenities, the quality of the infrastructure, environmental factors and safety. The goal of our model is to quantitatively characterize the immediate environment of residents in the space of factors that affect life quality.

The fundamental framework of our model and our calculations is the network representation of Budapest's pedestrian infrastructure. The nodes of the network represent intersections, while links are sidewalks and pedestrian infrastructure. The output of our model is an index, that characterizes every node of the Budapest network, giving a high-resolution quality-landscape of the city. The index is ultimately a number aggregated from multiple sub-categories, and its main value is highlighting inequalities and relative deficiencies within the



**Figure 5.1.** (a) Network representing Budapest pedestrian structure. The network was built following a primal approach, where the edges are sidewalks and pedestrian infrastructure, and nodes are intersections. (b) The graph-Voronoi tessellation of the Budapest network, generated using a subset of 15 parks as seeds. The color of the nodes represents the cell they belong to and the highlighted red dots are the seeds of each cell. The distance measure between two points is defined as the weighted shortest path on the graph, the weights being the average time required to cross a given edge.

city.

The final value of the index is a weighted sum, characterizing every node (intersection) in the network:

$$Q_i = w^{\text{services}} \tilde{Q}_i^{\text{services}} + w^{\text{safety}} \tilde{Q}_i^{\text{safety}} + w^{\text{environment}} \tilde{Q}_i^{\text{environment}} \quad (5.1)$$

In the equation  $i$  represents an individual node in the network. The “tilde” above the  $Q$  terms means that the values of the different category indices are normalized within the category. The weights  $w$  assigned to every term are arbitrary and are highly context-dependent. All terms of the equation are discussed in the following sections.

The weights of the different  $Q$  indices in the final aggregation highly depends on the context and the nature of the problem. Here we present the values we used to generate the results of this study, that were agreed upon consulting with experts, and policy makers.

The weights of the sub-indices from equation (5.1) are of the following values:

- $w^{\text{services}} = 0.7$ ;

- $w^{\text{safety}} = 0.1$ ;
- $w^{\text{environment}} = 0.2$

### 5.3.1 The services index: $Q^{\text{services}}$

The number quantifying each node in terms of how well it is connected with amenities and services is a weighted sum of sub-categories as well.

$$Q_i^{\text{services}} = \sum_c w^c Q_i^c \quad (5.2)$$

where,  $c$  denotes categories (family, culture, health, sport, and nightlife), and  $w^c$  the importance of category  $c$ .

The category weights used in equation (5.2), aggregating  $Q^{\text{services}}$  are:

- $w^{\text{family}} = 0.3$ ;
- $w^{\text{health}} = 0.3$ ;
- $w^{\text{culture}} = 0.15$ ;
- $w^{\text{sport}} = 0.15$ ;
- $w^{\text{night life}} = 0.1$

Some categories, like family, have further subcategories. Even though we have also had data and made a separate analysis on tourism, its effects on life quality of the residents are ambiguous, so we decided to omit it from the index.

What sets categories apart is that they incorporate different sets of amenities, with a few overlaps. The details of the categorization of amenities are included in the Appendix.

For every service/amenity class we have a given set of points of interest (POI) along with where the amenities of that class are available, with exact geolocation. We assign every POI of a given amenity class (e.g supermarket, pharmacy, school, etc.) to the nearest node on the infrastructure network. Each set of POIs organically generates a spatial partitioning of the city with one partition per POI. The partition of a POI is the set of all the nodes from which that particular POI can be reached faster than any other POI of the same class.

Mathematically these partitions are called graph-Voronoi cells [175, 176], where every node of a cell is assigned to its closest seed (POI). Distance, in this case, is not euclidean or geometric distance, but the distance on the network, where we use the weighted shortest path between two nodes as the distance

measure. The weight of links is a temporal parameter encoding the average time required to cross the represented street from one end to the other, thus the weight is a simple product of average speed and length of the street. This is in principle very similar to the way navigation systems find routes between points. For an example of a graph-Voronoi partitioning see Figure 1 (b).

To assess how well connected a node is to amenities we consider the following factors:

- How important is an amenity - weight ( $w_a$ )
- How long does it take to reach the amenity - time to reach ( $t_{ia}$ )
- Relatively how many nodes (or people) does the amenity share with - exclusivity ( $P_a$ )

From the three factors, the latter two are calculated using the city infrastructure network. The index for an amenity class, from the perspective of node  $i$ , is proportional with its importance (weight) and it is inversely proportional with the time to reach the closest POI from  $i$  and with the degree of exclusivity.

$$q_i^a = \frac{w_a}{(P_a + 1)(t_{ia} + 1)} \quad (5.3)$$

There can be certain singular cases when a Voronoi cell is empty ( $P_a = 0$ , i.e. no residents in the area) or the node  $i$  in question is right at the POI ( $t_{ia} = 0$ ). To avoid anomalies in the index we added 1 to both parameters.

The index of one category is proportional to the sum of its amenity-indices (calculated in (5.3)). To treat this number on the right scale (in practice we can get very large and very small numbers) we take the natural logarithm of the sum across amenities.

$$Q_i^c = \log\left(\sum_a q_i^a\right) : \quad (5.4)$$

As we have mentioned earlier the final services index is the weighted sum of the indices of the sub-categories.

$$Q_i^{services} = \sum_c w^c Q_i^c$$

Finally we normalize the values of  $Q_i^{services}$  so its values are comparable to the other values of the final  $Q$  equation (5.1):

$$\tilde{Q}_i^{services} = \frac{Q_i^{services} + |\min(Q^{services})|}{\max(Q^{services}) + |\min(Q^{services})|} \quad (5.5)$$

### 5.3.2 Safety index: $Q^{\text{safety}}$

The safety index is calculated across districts based on the number of crimes committed per one hundred thousand residents. Since the highest resolution data available to us was on the district level, every node  $i$  in the same district will have the same safety index value. The crime index:

$$Q_i^{\text{crime}} = \frac{N_{\text{crime}}^{\text{district}}}{n_i^{\text{district}}}$$

Where  $N_{\text{crime}}^{\text{district}}$  is the number of crimes committed in a district in a year, and  $n_i^{\text{district}}$  is the number of nodes in the district. The safety index is one minus the normalized crime index.

$$\tilde{Q}_i^{\text{safety}} = 1 - \frac{Q_i^{\text{crime}}}{\max(Q^{\text{crime}})} \quad (5.6)$$

### 5.3.3 Environmental index $Q^{\text{environment}}$

The environmental index is made up of two components: air pollution ratio and ratio of natural areas.

#### Air pollution ratio

We use the data provided by Budapest's air pollution measuring stations for the year 2018. For this study, we used the yearly median value of three polluters: carbon monoxide, nitrogen dioxide, and PM10 dust-pollution. As an approximation, we project the geometric Voronoi cells of the measuring stations onto the city map and each node will receive the pollution metrics of the geometrically closest station. We divide these values with the yearly upper health limit for the given polluter to assess to what degree do these values approximate the health limit. Thus the air pollution index of one node is formalized as follows:

$$C_i = \frac{c_i^{\text{CO}}}{c_{\text{limit}}^{\text{CO}}} + \frac{c_i^{\text{NO}_2}}{c_{\text{limit}}^{\text{NO}_2}} + \frac{c_i^{\text{Pm10}}}{c_{\text{limit}}^{\text{Pm10}}}$$

Where  $c_{\text{limit}}^{\text{CO}} = 3000 \text{ g/m}^3$ ,  $c_{\text{limit}}^{\text{NO}_2} = 40 \text{ g/m}^3$ ,  $c_{\text{limit}}^{\text{Pm10}} = 40 \text{ g/m}^3$  are the yearly upper limits based on data from the Hungarian Air Quality Network [172].

#### Ratio of natural areas

For this index, we have data on the neighborhood level, which is a more granular level of administrative partitioning the city than the districts are. In this

case, we project the same index onto every node in the same neighborhood. We consider as natural areas forests, parks and water surfaces (ponds, rivers, etc). The index:

$$T_i = \frac{R_{water}^{nh(i)} + R_{forest}^{nh(i)} + R_{park}^{nh(i)}}{\max(T)}$$

Where  $R_x^{nh(i)}$  is the relative surface area of natural area  $x$  within the neighborhood that  $i$  belongs to ( $nh(i)$ ). In other words, the surface area of a natural area is divided by the number of nodes in the neighborhood and the surface area of the neighborhood. Thus  $R_x^{nh(i)} = \frac{T(x)}{T(nh(i))n_{nh}}$ , where  $T(x)$  is the surface area of  $x$  natural area,  $T(nh)$  is the surface area of  $nh$  neighborhood and  $n_{nh}$  is the number of nodes in neighborhood  $nh$ . The final environmental index:

$$Q_i^{environment} = \frac{1 + T_i}{1 + C_i},$$

That after a normalization is:

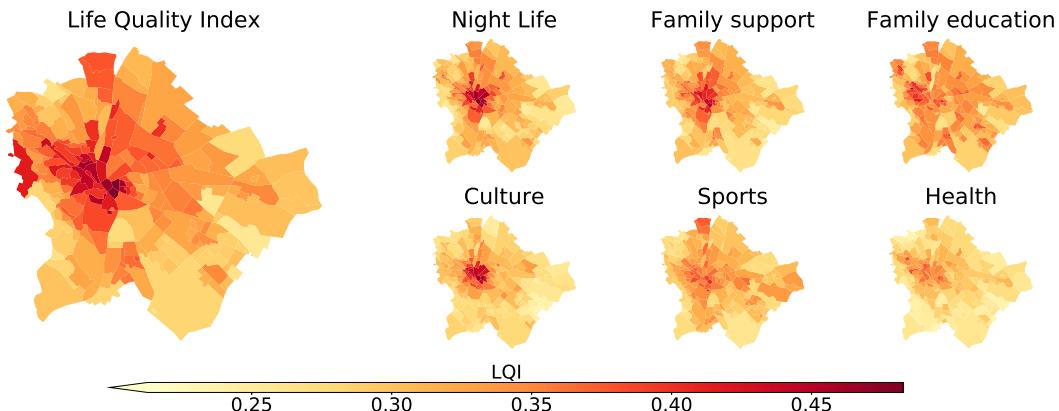
$$\tilde{Q}_i^{environment} = \frac{Q_i^{environment}}{\max(Q^{environment})}$$

## 5.4 Results

We quantify life quality in terms of each category (family support, education healthcare, sport, culture, nightlife, environment), and an overall measurement which contains all 6 categories and crime rate normalized by the population for the city of Budapest. Our method allows us to measure life quality for each intersection of the city, which helps to capture within neighborhood inequalities too. Analysis on the category level is beneficial for targeted policy interventions for better service allocation.

Figure 5.2 shows our overall life quality index (LQI) and by categories. Heatmaps reveal important features of Budapest. Similarly to most European cities life quality is much better in the inner districts [177, 178], especially in the case of Night Life and Culture.

Budapest is divided by the Danube river into two main parts: Buda and Pest. The river does not only serve as a geographical border but due to historical reasons, it also divides the citizens by social status. Hilly Buda, on the West side of the river, used to be the capital of the country, with the residence of the former Hungarian king. On the other side, the mainly flat Pest used to be the agricultural supporter of the aristocrats in Buda [179]. Even though the city has changed dramatically since the Monarchy, the division of Buda and Pest



**Figure 5.2.** Budapest neighborhoods, average life quality by categories and aggregated life quality index.

persists, and our life quality index captures it well. However certain services are legally guaranteed to be evenly distributed in the city, such as education and healthcare, for precise modeling one should take into account private care too, which highlights inequalities. So, the traditional division of Buda and Pest is even visible in categories where there should not be that much of a difference (Education, Family Support, Healthcare).

Results also highlight that category LQI-s are highly correlated, less liveable neighborhoods are constant regardless of the amenity category, and well-performing neighborhoods do not change either. It is caused by two main factors: the lack of amenities and the relatively high walking distances in the suburbs.

The compact city concept focuses on building more sustainable and livable cities while designing practical neighborhoods where citizens can maintain everyday life without a car [180]. Since, the walkability of a neighborhood highly correlates with its liveability [181] and the suburbs in Budapest do not show any compact city design features, both long distances and the lack of amenities effects suburban habitats lives negatively.

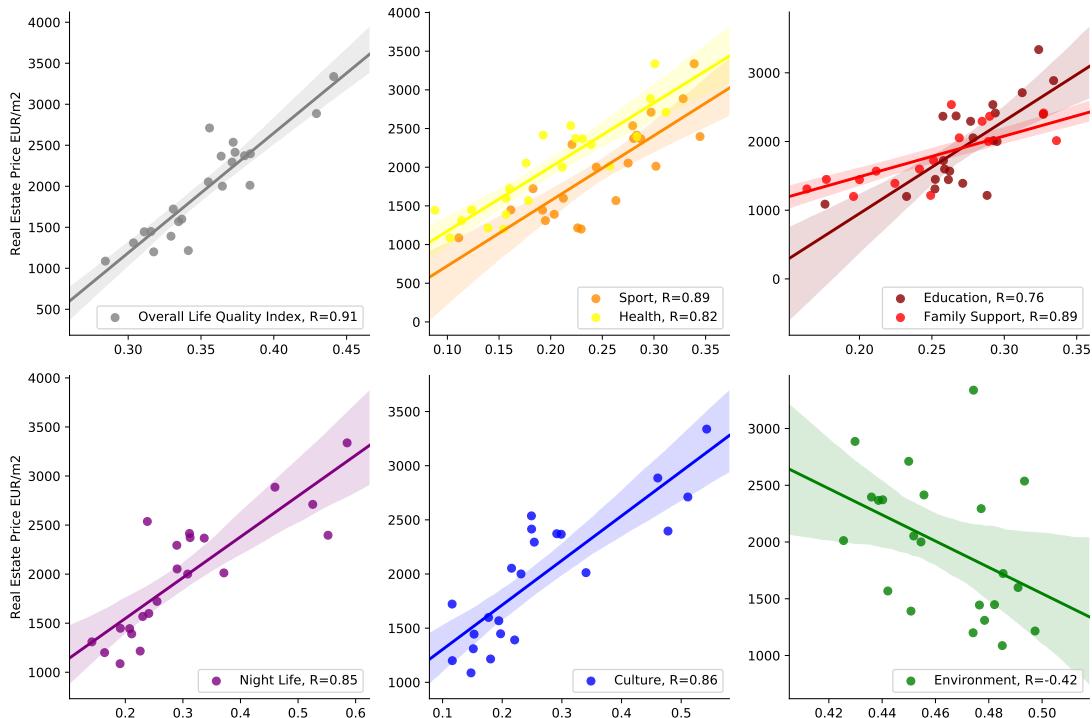
#### 5.4.1 Evaluation

Multiple methods have been developed to evaluate the accuracy of quality of life metrics: Scholars used expert validation with geographic visualization [182, 183], correlations with socioeconomic characteristics [184] and surveying citizens' perceptions of the conditions of life [185].

Our evaluation is based on the micro-economical *hedonic approach* of esti-

mating the values of public goods. In a capitalist market, real estate prices reflect the recognition of a neighborhood's characteristics: Prices are formed based on demand, more desirable places are more expensive, due to the underlying assumption of providing a higher quality of life. [178] Estimating neighborhoods life-quality with real estate prices has a long tradition in urban literature [186, 187, 188], therefore we adopt this method to evaluate our model.

We collected the average  $m^2/EUR$  price for all 23 districts of Budapest in January 2019 [189] and correlated each LQI category averaged by district with it. Figure 5.3 shows that our overall LQI correlates the most ( $R=0.91$ ) with the real-estate prices. Most of its components have a positive correlation with real-estate prices, except the environment which is calculated based on air pollution and green surface proximity. The life quality (LQI) in Budapest is much higher in densely populated downtown districts, which are lack of green surface and suffers from high air pollution due to heavy traffic.



**Figure 5.3.** Districts of Budapest Life Quality Index (LQI) and its components correlated with real estate prices ( $m^2/EUR$ )

**Summary** Locals of Budapest, like in most European cities, traditionally values downtown areas. The relative closeness to CBD, good access to public trans-

port, and vital city life kept it as a desirable area for living [190]. However, in recent years, the city is facing new challenges: due to gentrification [191] and over-tourism (eg.:Airbnb) real estate prices are sky-rocketing in downtown areas. In contrast with the early 2000-s when (upper) middle-class moved to the suburbs, nowadays, lower-income families and young professionals are leaving the downtown behind in hope for more affordable living.

As our findings show, Budapest is quite centralized and the quality of life highly correlates with real estate prices, which possibly lead to even more inequalities in the future. This spatial discrimination with longer traveling time, less fulfilling environment, and potential segregation reduces the chances of upward mobility and the quality of life of individuals [192].

## 5.5 Discussion

We have proposed a methodology to quantify life quality as a function of walkability on urban networks. We have used open data to capture inequalities between neighborhoods and districts in the city. We have shown that the real estate market reflects the life quality that our methods found.

A data-driven approach for quantifying life quality at such a granular level like our proposed method can help decision-makers to tackle social and environmental challenges better. Designing compact, liveable neighborhoods, considering also the upcoming environmental crisis is the number one priority of many cities worldwide.

The use of open data sources and algorithmic approaches adds up towards a systematic framework for understanding urban liveability. Our current approach is not the last word in this development since it does not yet account for multiple other variables, such as the quality of services and infrastructure, and other qualitative variables. To capture the more specific indicator of liveability in different cities it would be necessary to work with more granular and city-dependant data.

We anticipate a future stream of research focused on the use of worldwide open data sets to quantify urban liveability, including longitudinal studies in multiple cities, along with algorithmic modeling, simulations, and machine learning approaches, to first quantify the liveability, propose changes and test them with the ground truth data.

## CHAPTER 6

---

# CONCLUSION

---

Conclusion for chapters and thesis

### **Future Work**

What is next



---

## BIBLIOGRAPHY

---

- [1] Batty, M.: *The New Science of Cities*. MIT Press, ??? (2013)
- [2] Lobo, J., Alberti, M., Allen-Dumas, M., Arcaute, E., Barthelemy, M., Bojorquez Tapia, L.A., Brail, S., Bettencourt, L., Beukes, A., Chen, W.-Q., et al.: *Urban science: Integrated theory from the first cities to sustainable metropolises* (2020)
- [3] Lin, J., Ban, Y.: Complex Network Topology of Transportation Systems. *Transport Reviews* **33**(6), 658–685 (2013). doi:10.1080/01441647.2013.848955
- [4] Barthelemy, M.: Spatial networks. *Physics Reports* **499**(1-3), 1–101 (2011). doi:10.1016/j.physrep.2010.11.002. 1010.0302
- [5] Ding, R., Ujang, N., Hamid, H.B., Manan, M.S.A., Li, R., Albadareen, S.S.M., Nochian, A., Wu, J.: Application of Complex Networks Theory in Urban Traffic Network Researches. *Networks and Spatial Economics* **19**(4), 1281–1317 (2019). doi:10.1007/s11067-019-09466-5
- [6] Jiang, B., Claramunt, C.: Topological analysis of urban street networks. *Environment and Planning B: Planning and Design* **31**(1), 151–162 (2004). doi:10.1068/b306
- [7] Cardillo, A., Scellato, S., Latora, V., Porta, S.: Structural properties of planar graphs of urban street patterns. *Physical Review E* **73**(6), 1–7 (2006). doi:10.1103/PhysRevE.73.066107. 0510162
- [8] Barthélémy, M., Flammini, A.: Modeling Urban street patterns. *Physical Review Letters* **100**(13), 138702 (2008). doi:10.1103/PhysRevLett.100.138702
- [9] Batty, M.: The size, scale, and shape of cities. *Science* **319**(5864), 769–771 (2008). doi:10.1126/science.1151419

- [10] Strano, E., Viana, M., Costa, L.d.F., Cardillo, A., Porta, S., Latora, V.: Urban street networks, a comparative analysis of ten European cities. *Environment and Planning B: Planning and Design* **40**(6), 1071–1086 (2013). doi:10.1068/b38216
- [11] Louf, R., Barthelemy, M.: A typology of street patterns. *Journal of the Royal Society, Interface* **11**(101), 20140924 (2014). doi:10.1098/rsif.2014.0924
- [12] Boeing, G.: A multi-scale analysis of 27,000 urban street networks: Every us city, town, urbanized area, and zillow neighborhood. *Environment and Planning B: Urban Analytics and City Science* **47**(4), 590–608 (2020). doi:10.1177/2399808318784595
- [13] Crucitti, P., Latora, V., Porta, S.: Centrality Measures in Spatial Networks of Urban Streets. *Physical Review E* **73**(3), 1–4 (2008). doi:10.1103/PhysRevE.73.036125. 0504163
- [14] Boeing, G.: Planarity and street network representation in urban form analysis. *Environment and Planning B: Urban Analytics and City Science* **47**(5), 855–869 (2020). doi:10.1177/2399808318802941
- [15] Kirkley, A., Barbosa, H., Barthelemy, M., Ghoshal, G.: From the betweenness centrality in street networks to structural invariants in random planar graphs. *Nature Communications* **9**(1), 2501 (2018). doi:10.1038/s41467-018-04978-z
- [16] Makse, H.A., Havlin, S., Stanley, H.E.: Modelling urban growth patterns. *Nature* **377**(6550), 608–612 (1995). doi:10.1038/377608a0
- [17] Strano, E., Nicosia, V., Latora, V., Porta, S., Barthélemy, M.: Elementary processes governing the evolution of road networks. *Scientific Reports* **2**(1), 296 (2012). doi:10.1038/srep00296
- [18] Hajrasouliha, A., Yin, L.: The impact of street network connectivity on pedestrian volume. *Urban Studies* **52**(13), 2483–2497 (2015). doi:10.1177/0042098014544763
- [19] De Nadai, M., Staiano, J., Larcher, R., Sebe, N., Quercia, D., Lepri, B.: The death and life of great italian cities: A mobile phone data perspective. In: Proceedings of the 25th International Conference on World Wide Web. WWW '16, pp. 413–423. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE (2016). doi:10.1145/2872427.2883084

- [20] Biazzo, I., Monechi, B., Loreto, V.: General scores for accessibility and inequality measures in urban areas. *Royal Society Open Science* **6**(8) (2019). doi:10.1098/rsos.190979
- [21] Natera Orozco, L.G., Deritei, D., Vancsó, A., Vásárhelyi, O.: Quantifying life quality as walkability on urban networks: The case of Budapest. In: Cherifi, H., Gaito, S., Mendes, J.F., Moro, E., Rocha, L.M. (eds.) *Complex Networks and Their Applications VIII*, pp. 905–918. Springer, Cham (2020)
- [22] Baggag, A., Abbar, S., Zanouda, T., Srivastava, J.: Resilience analytics: coverage and robustness in multi-modal transportation networks. *EPJ Data Science* **7**(1), 14 (2018). doi:10.1140/epjds/s13688-018-0139-7
- [23] Ferretti, M., Barlacchi, G., Pappalardo, L., Lucchini, L., Lepri, B.: Weak nodes detection in urban transport systems: Planning for resilience in Singapore. In: *Proceedings - 2018 IEEE 5th International Conference on Data Science and Advanced Analytics, DSAA 2018*, pp. 472–480 (2019). doi:10.1109/DSAA.2018.00061
- [24] Natera Orozco, L.G., Battiston, F., Iñiguez, G., Szell, M.: Data-driven strategies for optimal bicycle network growth. *Royal Society Open Science* **7**(12), 201130 (2020). doi:10.1098/rsos.201130
- [25] Organizers, Kang, W., Oshan, T., Wolf, L.J., Discussants, Boeing, G., Frias-Martinez, V., Gao, S., Poorthuis, A., Xu, W.: A roundtable discussion: Defining urban data science. *Environment and Planning B: Urban Analytics and City Science* **46**(9), 1756–1768 (2019)
- [26] Resch, B., Szell, M.: Human-centric data science for urban studies. *ISPRS International Journal of Geo-Information* **8**(584) (2019). doi:10.3390/ijgi8120584
- [27] Porta, S., Crucitti, P., Latora, V.: The network analysis of urban streets : a primal approach. *Environment and Planning B: Planning and Design* **33**(5), 705–726 (2006). doi:10.1068/b32045
- [28] Albert, R., Barabási, A.-L.: Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**(1), 47 (2002)
- [29] Dorogovtsev, S.N., Mendes, J.F.F.: Evolution of networks. *Adv. Phys.* **51**(4), 1079–1187 (2002)

- [30] Newman, M.E.J.: The structure and function of complex networks. *SIAM Rev.* **45**, 167–256 (2003)
- [31] Boccaletti, V., Latora, Y., Moreno, M., Chavez, M., Hwang, D.-U.: Complex networks: Structure and dynamics. *Physics Reports* **424**(4), 175–308 (2006). doi:10.1016/j.physrep.2005.10.009
- [32] Boccaletti, S., Bianconi, G., Criado, R., del Genio, C.I., Gómez-Gardeñes, J., Romance, M., Sendiña-Nadal, I., Wang, Z., Zanin, M.: The structure and dynamics of multilayer networks (2014). doi:10.1016/j.physrep.2014.07.001. <https://arxiv.org/pdf/1407.0742.pdf>
- [33] Battiston, F., Nicosia, V., Latora, V.: Structural measures for multiplex networks. *Phys. Rev. E* **89**, 032804 (2014). doi:10.1103/PhysRevE.89.032804
- [34] De Domenico, M., Solé-Ribalta, A., Cozzo, E., Kivelä, M., Moreno, Y., Porter, M.A., Gómez, S., Arenas, A.: Mathematical formulation of multilayer networks. *Phys. Rev. X* **3**, 041022 (2013). doi:10.1103/PhysRevX.3.041022
- [35] Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J.P., Moreno, Y., Porter, M.A.: Multilayer networks. *Journal of Complex Networks* **2**(3), 203–271 (2014). doi:10.1093/comnet/cnu016
- [36] Nicosia, V., Latora, V.: Measuring and modeling correlations in multiplex networks. *Physical Review E* **92**(3), 032805 (2015)
- [37] Kurant, M., Thiran, P.: Layered complex networks. *Physical review letters* **96**(13), 138701 (2006)
- [38] Cardillo, A., Gómez-Gardenes, J., Zanin, M., Romance, M., Papo, D., Del Pozo, F., Boccaletti, S.: Emergence of network features from multiplexity. *Scientific reports* **3**(1), 1–6 (2013)
- [39] De Domenico, M., Nicosia, V., Arenas, A., Latora, V.: Structural reducibility of multilayer networks. *Nature communications* **6**(1), 1–9 (2015)
- [40] Latora, V., Marchiori, M.: Efficient behavior of small-world networks. *Physical review letters* **87**(19), 198701 (2001)
- [41] Buldyrev, S.V., Parshani, R., Paul, G., Stanley, H.E., Havlin, S.: Catastrophic cascade of failures in interdependent networks. *Nature* **464**(7291), 1025–1028 (2010)

- [42] De Domenico, M., Solé-Ribalta, A., Gómez, S., Arenas, A.: Navigability of interconnected networks under random failures. *Proceedings of the National Academy of Sciences* **111**(23), 8351–8356 (2014). doi:10.1073/pnas.1318469111
- [43] Bianconi, G.: *Multilayer Networks: Structure and Function*. Oxford university press, Oxford (2018)
- [44] Aleta, A., Moreno, Y.: Multilayer Networks in a Nutshell. *Annual Review of Condensed Matter Physics* **10**(1), 45–62 (2019). doi:10.1146/annurev-conmatphys-031218-013259
- [45] Battiston, F., Nicosia, V., Latora, V.: The new challenges of multiplex networks: Measures and models. *European Physical Journal: Special Topics* **226**(3), 401–416 (2017). doi:10.1140/epjst/e2016-60274-8
- [46] De Domenico, M., Granell, C., Porter, M.A., Arenas, A.: The physics of spreading processes in multilayer networks. *Nature Physics* **12**(10), 901–906 (2016)
- [47] de Cea, J., Fernández, J.E., Dekock, V., Soto, A.: Solving network equilibrium problems on multimodal urban transportation networks with multiple user classes. *Transport Reviews* **25**(3), 293–317 (2005). doi:10.1080/0144164042000335805
- [48] Li, Z.C., Lam, W.H.K., Wong, S.C., Zhu, D.L., Huang, H.J.: Modeling park-and-ride services in a multimodal transport network with elastic demand. *Transportation Research Record* **1994**(1994), 101–109 (2007). doi:10.3141/1994-14
- [49] Gil, J.: Analyzing the configuration of multimodal urban networks. *Geographical Analysis* **46**(4), 368–391 (2014). doi:10.1111/gean.12062
- [50] Aleta, A., Meloni, S., Moreno, Y.: A Multilayer perspective for the analysis of urban transportation systems. *Scientific Reports* **7** (2017). doi:10.1038/srep44359
- [51] Battista, M.G., Lucertini, M., Simeone, B.: Path composition and multiple choice in a bimodal transportation network. volume 2: Modelling transport systems. In: *World Transport Research. Proceedings of the 7th World Conference on Transport Research*. World Conference on Transport Research Society (1996)

- [52] Lozano, A., Storchi, G.: Shortest viable path algorithm in multimodal networks. *Transportation Research Part A: Policy and Practice* **35**(3), 225–241 (2001). doi:10.1016/S0965-8564(99)00056-7
- [53] Strano, E., Shai, S., Dobson, S., Barthelemy, M.: Multiplex networks in metropolitan areas: generic features and local effects. *Journal of the Royal Society, Interface* **12**(111) (2015). doi:10.1098/rsif.2015.0651
- [54] Morris, R.G., Barthelemy, M.: Transport on coupled spatial networks. *Physical Review Letters* **109**(12), 128703 (2012). doi:10.1103/PhysRevLett.109.128703
- [55] Nicosia, V., Bianconi, G., Latora, V., Barthelemy, M.: Growing multiplex networks. *Physical Review Letters* **111**(5), 058701 (2013). doi:10.1103/PhysRevLett.111.058701
- [56] Freeman, L.C.: A set of measures of centrality based on betweenness. *Sociometry* **40**(1), 35–41 (1977)
- [57] Ding, R., Ujang, N., bin Hamid, H., Manan, M.S.A., He, Y., Li, R., Wu, J.: Detecting the urban traffic network structure dynamics through the growth and analysis of multi-layer networks. *Physica A: Statistical Mechanics and its Applications* **503**, 800–817 (2018). doi:10.1016/j.physa.2018.02.059
- [58] de Regt, R., von Ferber, C., Holovatch, Y., Lebovka, M.: Public transportation in great britain viewed as a complex network. *Transportmetrica A: Transport Science* **15**(2), 722–748 (2019)
- [59] Solé-Ribalta, A., Gómez, S., Arenas, A.: Congestion Induced by the Structure of Multiplex Networks. *Physical Review Letters* **116**(10), 1–5 (2016). doi:10.1103/PhysRevLett.116.108701
- [60] Bródka, P., Chmiel, A., Magnani, M., Ragozini, G.: Quantifying layer similarity in multiplex networks: a systematic study. *Royal Society Open Science* **5**(8), 171747 (2018). doi:10.1098/rsos.171747
- [61] Natera Orozco, L.G., Battiston, F., Iñiguez, G., Szell, M.: Extracting the multimodal fingerprint of urban transportation networks. *Transport Findings* (2020). doi:10.32866/001c.13171
- [62] Patriksson, M.: The Traffic Assignment Problem: Models and Methods. Courier Dover Publications, ??? (2015)

- [63] Barbosa, H., Barthelemy, M., Ghoshal, G., James, C.R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J.J., Simini, F., Tomasini, M.: Human mobility: Models and applications. *Physics Reports* **734**, 1–74 (2018)
- [64] McNally, M.G.: The four step model. *Handbook of transport modelling* **1**, 35–41 (2000)
- [65] Solé-Ribalta, A., De Domenico, M., Gómez, S., Arenas, A.: Random walk centrality in interconnected multilayer networks. *Physica D: Nonlinear Phenomena* **323**, 73–79 (2016)
- [66] Tan, F., Wu, J., Xia, Y., Tse, C.K.: Traffic congestion in interconnected complex networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* **89**(6), 062813 (2014). doi:10.1103/PhysRevE.89.062813
- [67] Bassolas, A., Gallotti, R., Lamanna, F., Lenormand, M., Ramasco, J.J.: Scaling in the recovery of urban transportation systems from massive events. *Scientific Reports* **10**(1), 1–13 (2020). doi:10.1038/s41598-020-59576-1
- [68] Manfredi, S., Di Tucci, E., Latora, V.: Mobility and Congestion in Dynamical Multilayer Networks with Finite Storage Capacity. *Physical Review Letters* **120**(6), 068301 (2018). doi:10.1103/PhysRevLett.120.068301
- [69] Arentze, T.A., Molin, E.J.: Travelers' preferences in multimodal networks: design and results of a comprehensive series of choice experiments. *Transportation Research Part A: Policy and Practice* **58**, 15–28 (2013)
- [70] Abrantes, P.A., Wardman, M.R.: Meta-analysis of uk values of travel time: An update. *Transportation Research Part A: Policy and Practice* **45**(1), 1–17 (2011)
- [71] Wardman, M.: A review of british evidence on time and service quality valuations. *Transportation Research Part E: Logistics and Transportation Review* **37**(2-3), 107–128 (2001)
- [72] Nobis, C.: Multimodality: facets and causes of sustainable mobility behavior. *Transportation Research Record* **2010**(1), 35–44 (2007)
- [73] Chen, C., Ma, J., Susilo, Y., Liu, Y., Wang, M.: The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation research part C: emerging technologies* **68**, 285–299 (2016)

- [74] Zannat, K.E., Choudhury, C.F.: Emerging big data sources for public transport planning: A systematic review on current state of art and future research directions. *Journal of the Indian Institute of Science*, 1–19 (2019)
- [75] Gallotti, R., Barthelemy, M.: Anatomy and efficiency of urban multimodal mobility. *Scientific Reports* **4**(1), 6911 (2014). doi:10.1038/srep06911
- [76] Barthelemy, M.: *The Structure and Dynamics of Cities*. Cambridge University Press, ??? (2016)
- [77] Alessandretti, L., Karsai, M., Gauvin, L.: User-based representation of time-resolved multimodal public transportation networks. *Royal Society Open Science* **3**(7), 160156 (2016). doi:10.1098/rsos.160156
- [78] Sui, Y., Shao, F., Yu, X., Sun, R., Li, S.: Public transport network model based on layer operations. *Physica A: Statistical Mechanics and its Applications* **523**, 984–995 (2019). doi:10.1016/j.physa.2019.04.269
- [79] Holleczeck, T., Yu, L., Lee, J.K., Senn, O., Ratti, C., Jaillet, P.: Detecting weak public transport connections from cellphone and public transport data. In: *Proceedings of the 2014 International Conference on Big Data Science and Computing*, pp. 1–8 (2014)
- [80] Chodrow, P.S., Al-Awwad, Z., Jiang, S., González, M.C.: Demand and congestion in multiplex transportation networks. *PLoS ONE* **11**(9) (2016). doi:10.1371/journal.pone.0161738
- [81] Gallotti, R., Porter, M.A., Barthelemy, M.: Lost in transportation: Information measures and cognitive limits in multilayer navigation. *Science Advances* **2**(2), 1500445–1500445 (2016). doi:10.1126/sciadv.1500445
- [82] Rosvall, M., Trusina, A., Minnhagen, P., Sneppen, K.: Networks and cities: An information perspective. *Physical Review Letters* **94**(2), 028701 (2005)
- [83] Pelletier, M.-P., Trépanier, M., Morency, C.: Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies* **19**(4), 557–568 (2011)
- [84] Ma, X., Wu, Y.-J., Wang, Y., Chen, F., Liu, J.: Mining smart card data for transit riders' travel patterns. *Transportation Research Part C: Emerging Technologies* **36**, 1–12 (2013)

- [85] Zheng, Z., Huang, Z., Zhang, F., Wang, P.: Understanding coupling dynamics of public transportation networks. *EPJ Data Science* **7**(1), 23 (2018). doi:10.1140/epjds/s13688-018-0148-6
- [86] Gallotti, R., Bazzani, A., Rambaldi, S., Barthelemy, M.: A stochastic model of randomly accelerated walkers for human mobility. *Nature communications* **7**(1), 1–7 (2016)
- [87] Zhao, K., Musolesi, M., Hui, P., Rao, W., Tarkoma, S.: Explaining the power-law distribution of human mobility through transportationmodality decomposition. *Scientific reports* **5**(1), 1–7 (2015)
- [88] Varga, L., Kovács, A., Toth, G., Papp, I., Néda, Z.: Further we travel the faster we go. *PloS one* **11**(2), 0148913 (2016)
- [89] Haklay, M.: How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets. *Environment and Planning B: Planning and Design* **37**(4), 682–703 (2010). doi:10.1068/b35097
- [90] Girres, J.F., Touya, G.: Quality Assessment of the French OpenStreetMap Dataset. *Transactions in GIS* **14**(4), 435–459 (2010). doi:10.1111/j.1467-9671.2010.01203.x
- [91] Ferster, C., Fischer, J., Manaugh, K., Nelson, T., Winters, M.: Using OpenStreetMap to inventory bicycle infrastructure: A comparison with open data from cities. *International Journal of Sustainable Transportation*, 1–10 (2019). doi:10.1080/15568318.2018.1519746
- [92] Boeing, G.: Street Network Models and Indicators for Every Urban Area in the World (2020). 2009.09106
- [93] Gil, J.: Building a multimodal urban network model using openstreetmap data for the analysis of sustainable accessibility. In: *OpenStreetMap in GIScience*, pp. 229–251. Springer, Cham (2015)
- [94] Natera Orozco, L.G.: Multiplex urban networks. Harvard Dataverse, V1, <https://doi.org/10.7910/DVN/GSOPCK> (2019). doi:10.7910/DVN/GSOPCK
- [95] Gallotti, R., Barthelemy, M.: The multilayer temporal network of public transport in Great Britain. *Scientific Data* **2** (2015). doi:10.1038/sdata.2014.56

- [96] Kujala, R., Weckstrom, C., Darst, R.K., Mladenovic, M.N., Saramaki, J.: A collection of public transport network data sets for 25 cities. *Scientific Data* **5**(1), 1–14 (2018). doi:10.1038/sdata.2018.89
- [97] Tenkanen, H., Toivonen, T.: Longitudinal spatial dataset on travel times and distances by different travel modes in Helsinki Region. *Scientific Data* **7**(1), 1–15 (2020). doi:10.1038/s41597-020-0413-y
- [98] Hagberg, A., Schult, D., Swart, P.: Exploring network structure, dynamics, and function using networkx. In: Proceedings of the 7th Python in Science Conference (SciPy2008), pp. 11–15 (2008)
- [99] Peixoto, T.d.P.: graph-tool. graph-tool, V2.35, <https://graph-tool.skewed.de/> (2020)
- [100] Csádi, G., Nepusz, T.: Igraph. V0.8.3 <https://igraph.org/> (2006). <https://igrpah.org>
- [101] Boeing, G.: OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems* **65**, 126–139 (2017). doi:10.1016/j.comenvurbssys.2017.05.004
- [102] Tenkanen, H.: pyrosm. V.0.5.0 <https://github.com/HTenkanen/pyrosm> (2020). doi:10.5281/ZENODO.3818244. <https://github.com/HTenkanen/pyrosm>
- [103] Google: transitfeed. V.1.2.16 <https://github.com/google/transitfeed> (2018). <https://github.com/google/transitfeed>
- [104] De Domenico, M., Porter, M.A., Arenas, A.: MuxViz: A tool for multilayer analysis and visualization of networks. *Journal of Complex Networks* **3**(2), 159–176 (2015). doi:10.1093/comnet/cnu038. 1405.0843
- [105] Nicosia, V., Battiston, F.: mammult: Metrics And Models for MUL-Tilayer networks. GitHub repository. <https://github.com/KatolaZ/mammult> (2015). <https://github.com/KatolaZ/mammult>
- [106] Kouvaris, N.E., Hata, S., Díaz-Guilera, A.: Pattern formation in multiplex networks. *Scientific Reports* **5**(1), 1–9 (2015). doi:10.1038/srep10840
- [107] Kivelä, M.: Pymnet. GitHub repository. <https://github.com/bolozna/multilayer-networks-library> (2018). <http://mkivelä.com/pymnet/>

- [108] Magnani, M., Rossi, L., Vega, D.: Multiplex network analysis with r. *Journal of Statistical Software* (2020). Accepted for publication
- [109] Rivero Ostoic, J.A.: Algebraic analysis of multiple social networks with multiplex. *Journal of Statistical Software* **92**(11), 1–41 (2020). doi:10.18637/jss.v092.i11
- [110] Barthelemy, M., Bordin, P., Berestycki, H., Gribaudi, M.: Self-organization versus top-down planning in the evolution of a city. *Scientific Reports* **3**(1), 2153 (2013). doi:10.1038/srep02153
- [111] Lee, M., Barbosa-Filho, H., Youn, H., Holme, P., Ghoshal, G.: Morphology of travel routes and the organization of cities. *Nature Communications* **8**(1), 2229 (2017). doi:10.1038/s41467-017-02374-7
- [112] Jacobs, J.: *The Death and Life of Great American Cities*, p. 458. Random House, New York, New York, USA (1961)
- [113] Gössling, S., Schröder, M., Späth, P., Freytag, T.: Urban Space Distribution and Sustainable Transport. *Transport Reviews* **36**(5), 659–679 (2016). doi:10.1080/01441647.2016.1147101
- [114] Szell, M.: Crowdsourced Quantification and Visualization of Urban Mobility Space Inequality. *Urban Planning* **3**(1), 1 (2018). doi:10.17645/up.v3i1.1209
- [115] Ferster, C., Fischer, J., Manaugh, K., Nelson, T., Winters, M.: Using Open-StreetMap to inventory bicycle infrastructure: A comparison with open data from cities. *International Journal of Sustainable Transportation*, 1–10 (2019). doi:10.1080/15568318.2018.1519746
- [116] Rodrigue, J.-P., Comtois, C., Slack, B.: *The Geography of Transport Systems* (Third Edition), p. 432. Routledge, ??? (2013). doi:10.1080/10630732.2011.603579. <https://www.routledge.com/The-Geography-of-Transport-Systems-4th-Edition/Rodrigue-Comtois-Slack/p/book/9781138669574>
- [117] Eliasson, J.: Lessons from the Stockholm congestion charging trial. *Transport Policy* **15**(6), 395–404 (2008). doi:10.1016/j.tranpol.2008.12.004
- [118] Littke, H.: Revisiting the San Francisco parklets problematizing publicness, parks, and transferability. *Urban Forestry and Urban Greening* **15**, 165–173 (2016). doi:10.1016/j.ufug.2015.12.010

- [119] Bliss, L.: How to Cut 10,000 Parking Spaces Without Anyone Complaining. Citylab (2019)
- [120] Nieuwenhuijsen, M.J., Khreis, H.: Car free cities: Pathway to healthy urban living. Pergamon (2016). doi:10.1016/j.envint.2016.05.032
- [121] Reynolds, C.C.O., Harris, M.A., Teschke, K., Cripton, P.A., Winters, M.: The impact of transportation infrastructure on bicycling injuries and crashes: A review of the literature. Environmental Health: A Global Access Science Source **8**(1), 47 (2009). doi:10.1186/1476-069X-8-47
- [122] Teschke, K., Harris, M.A., Reynolds, C.C.O., Winters, M., Babul, S., Chipman, M., Cusimano, M.D., Brubacher, J.R., Hunte, G., Friedman, S.M., Monro, M., Shen, H., Vernich, L., Cripton, P.A.: Route infrastructure and the risk of injuries to bicyclists: A case-crossover study. American Journal of Public Health **102**(12), 2336–2343 (2012). doi:10.2105/AJPH.2012.300762
- [123] Pucher, J., Buehler, R.: Safer cycling through improved infrastructure. American Journal of Public Health **106**(12), 2089–2091 (2016). doi:10.2105/ajph.2016.303507
- [124] Aldred, R., Goodman, A., Gulliver, J., Woodcock, J.: Cycling injury risk in london: A case-control study exploring the impact of cycle volumes, motor vehicle volumes, and road characteristics including speed limits. Accident Analysis & Prevention **117**, 75–84 (2018)
- [125] Gössling, S., Choi, A., Dekker, K., Metzler, D.: The social cost of automobility, cycling and walking in the european union. Ecological Economics **158**, 65–74 (2019)
- [126] Gössling, S.: Why cities need to take road space from cars-and how this could be done. Journal of Urban Design, 1–6 (2020)
- [127] Dill, J., Carr, T.: Bicycle commuting and facilities in major u.s. cities: If you build them, commuters will use them. Transportation Research Record: Journal of the Transportation Research Board **1828**(1), 116–123 (2013). doi:10.3141/1828-14
- [128] Schoner, J.E., Levinson, D.M.: The missing link : bicycle infrastructure networks and ridership in 74 US cities. Transportation **41**(6), 1187–1204 (2014). doi:10.1007/s11116-014-9538-1
- [129] Hull, A., O'Holleran, C.: Bicycle infrastructure: can good design encourage cycling? Urban, Planning and Transport Research **2**(1), 369–406 (2014). doi:10.1080/21650020.2014.955210

- [130] Buehler, R., Dill, J.: Bikeway networks: A review of effects on cycling. *Transport Reviews* **36**(1), 9–27 (2016). doi:10.1080/01441647.2015.1069908
- [131] Hannah Twaddell (ICF), Eliot Rose (ICF), Joseph Broach (PSU), Jennifer Dill (PSU), Kelly Clifton (PSU), Claire Lust (PSU), Kimberly Voros (Alta), Hugh Louch (Alta), Erin David (Alta): Measuring Multimodal Network Connectivity. U.S. Department of Transportation, Federal Highway Administration (2018)
- [132] Hochmair, H.H., Zielstra, D., Neis, P.: Assessing the completeness of bicycle trail and designated lane features in openstreetmap for the united states and europe revised, pp. 111–137 (2012)
- [133] Krizek, K.J., Roland, R.W.: What is at the end of the road? understanding discontinuities of on-street bicycle lanes in urban settings. *Transportation Research Part D: Transport and Environment* **10**(1), 55–68 (2005). doi:10.1016/j.trd.2004.09.005
- [134] Movement, U.: *Cycling in the city*. Urban Movement, London (2013)
- [135] Systems, U.: *Pedestrian and Bicycle Facility Design Guidance*. Region of Peel, Toronto (2014)
- [136] Beck, B., Chong, D., Olivier, J., Perkins, M., Tsay, A., Rushford, A., Li, L., Cameron, P., Fry, R., Johnson, M.: How much space do drivers provide when passing cyclists? understanding the impact of motor vehicle and infrastructure characteristics on passing distance. *Accident Analysis & Prevention* (2019). doi:10.1016/j.aap.2019.03.007
- [137] Initiative, G.D.C., of City Transportation Officials, N.A.: *Global Street Design Guide*. Island Press, ??? (2016)
- [138] Dill, J.: Bicycling for transportation and health: The role of infrastructure. *Journal of Public Health Policy* **30**(SUPPL. 1), 95–110 (2009). doi:10.1057/jphp.2008.56
- [139] Chataway, E.S., Kaplan, S., Nielsen, T.A.S., Prato, C.G.: Safety perceptions and reported behavior related to cycling in mixed traffic: A comparison between brisbane and copenhagen. *Transportation Research Part F: Traffic Psychology and Behaviour* **23**, 32–43 (2014). doi:10.1016/j.trf.2013.12.021
- [140] Hyodo, T., Suzuki, N., Takahashi, K.: Modeling of bicycle route and destination choice behavior for bicycle road network plan. *Transportation*

- Research Record: Journal of the Transportation Research Board **1705**(1), 70–76 (2000). doi:10.3141/1705-11
- [141] Bao, J., Ruan, S., He, T., Zheng, Y., Li, Y., Ruan, S., Li, Y., Zheng, Y.: Planning Bike Lanes based on Sharing-Bikes' Trajectories. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '17, pp. 1377–1386. ACM Press, New York, New York, USA (2017). doi:10.1145/3097983.3098056
- [142] Mauttone, A., Mercadante, G., Rabaza, M., Toledo, F.: Bicycle network design: Model and solution algorithm. In: Transportation Research Procedia, vol. 27, pp. 969–976. Elsevier, ??? (2017). doi:10.1016/j.trpro.2017.12.119
- [143] Akbarzadeh, M., Mohri, S.S., Yazdian, E.: Designing bike networks using the concept of network clusters. Applied Network Science **3**(1), 12 (2018). doi:10.1007/s41109-018-0069-0
- [144] Kruskal, J.B.: On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. Proceedings of the American Mathematical Society **7**(1), 48 (1956). doi:10.2307/2033241
- [145] Hildebrandt, E., Wittkowski, A.: Expert study on the state of the art: Bicycle-highways and Pedelecs. Technical Report November, Planungsgemeinschaft Verkehr, Hannover (2013)
- [146] Taciuk, A., Davidson, G.: Practitioner's Guide to Planning, Designing, and Implementing Bicycle Highways in North America. In: TAC 2018: Innovation and Technology: Evolving Transportation - 2018 Conference and Exhibition of the Transportation Association of Canada, pp. 1–22 (2018)
- [147] Achlioptas, D., D'Souza, R.M., Spencer, J.: Explosive percolation in random networks. Science **323**(5920), 1453–1455 (2009). doi:10.1126/science.1167782
- [148] Morone, F., Makse, H.A.: Influence maximization in complex networks through optimal percolation. Nature **524**(7563), 65–68 (2015). doi:10.1038/nature14604
- [149] Mueller, N., Rojas-Rueda, D., Salmon, M., Martinez, D., Ambros, A., Brand, C., de Nazelle, A., Dons, E., Gaupp-Berghausen, M., Gerike, R., Götschi, T., Iacobassi, F., Int Panis, L., Kahlmeier, S., Raser, E.,

- Nieuwenhuijsen, M.: Health impact assessment of cycling network expansions in European cities. *Preventive Medicine* **109**, 62–70 (2018). doi:10.1016/j.ypmed.2017.12.011
- [150] Zhao, C., Carstensen, T.A., Nielsen, T.A.S., Olafsson, A.S.: Bicycle-friendly infrastructure planning in beijing and copenhagen-between adapting design solutions and learning local planning cultures. *Journal of transport geography* **68**, 149–159 (2018)
- [151] Carstensen, T.A., Olafsson, A.S., Bech, N.M., Poulsen, T.S., Zhao, C.: The spatio-temporal development of copenhagen's bicycle infrastructure 1912–2013. *Geografisk Tidsskrift-Danish Journal of Geography* **115**(2), 142–156 (2015)
- [152] Heylen, K.: Liveability in social housing: three case studies in flanders. In: ENHR Conference 'Housing in an Expanding Europe: Theory, Policy, Implementation and Participation', Ljubljana, Slovenia (2006)
- [153] Shamsuddin, S., Hassan, N.R.A., Bilyamin, S.F.I.: Walkable environment in increasing the liveability of a city. *Procedia-Social and Behavioral Sciences* **50**, 167–178 (2012)
- [154] Campbell, S.: Green cities, growing cities, just cities?: Urban planning and the contradictions of sustainable development. *Journal of the American Planning Association* **62**(3), 296–312 (1996)
- [155] Jabareen, Y.: Planning the resilient city: Concepts and strategies for coping with climate change and environmental risk. *Cities* **31**, 220–229 (2013)
- [156] Frank, L.D., Sallis, J.F., Conway, T.L., Chapman, J.E., Saelens, B.E., Bachman, W.: Many pathways from land use to health: associations between neighborhood walkability and active transportation, body mass index, and air quality. *Journal of the American planning Association* **72**(1), 75–87 (2006)
- [157] Speck, J.: *Walkable City: How Downtown Can Save America, One Step at a Time*. Farrar, Straus and Giroux, New York, New York, USA (2012)
- [158] Boeing, G.: The morphology and circuitry of walkable and drivable street networks. In: D'Acci, L. (ed.) *The Mathematics of Urban Morphology*, pp. 271–287. Springer, Cham (2019)
- [159] Hajrasouliha, A., Yin, L.: The impact of street network connectivity on pedestrian volume. *Urban Studies* **52**(13), 2483–2497 (2015)

- [160] Quercia, D., Aiello, L.M., Schifanella, R., Davies, A.: The digital life of walkable streets. In: Proceedings of the 24th International Conference on World Wide Web, pp. 875–884 (2015)
- [161] Silva, J.P., Akleh, A.Z.: Investigating the relationships between the built environment, the climate, walkability and physical activity in the arabian peninsula: The case of bahrain. *Cogent Social Sciences* **4**(1), 1502907 (2018)
- [162] Krambeck, H.V.: The global walkability index. PhD thesis, Massachusetts Institute of Technology (2006)
- [163] Southworth, M.: Designing the walkable city. *Journal of urban planning and development* **131**(4), 246–257 (2005)
- [164] Carr, L.J., Dunsiger, S.I., Marcus, B.H.: Walk score<sup>TM</sup> as a global estimate of neighborhood walkability. *American journal of preventive medicine* **39**(5), 460–463 (2010)
- [165] Wheeler, S.M., Beatley, T.: Sustainable Urban Development Reader. Routledge, UK (2014)
- [166] Childs, S.: A Case for Data-Driven City Planning. (2018 (accessed June 3, 2019)). <https://medium.com/citiesense/a-case-for-data-driven-city-planning-8cd7d9332a>
- [167] Kaushik, V.: How Technology Empowers Data-Driven Urban Planning. (2017 (accesed June 3, 2019)). <https://www.getrevue.co/profile/TGIC/issues/how-technology-empowers-data-driven-urban-planning-59814>
- [168] Fitzgerald, M.: Data-Driven City Management. (2016 (acsesed June 3, 2019)). <https://sloanreview.mit.edu/case-study/data-driven-city-management/>
- [169] Budapest Districts. <https://data2.openstreetmap.hu/hatarok/index.php?admin=9>
- [170] Police, H.: Bunugyi Statisztikak. *Data available at request.* <http://www.police.hu/a-rendorsegrol/statisztikak/bunugyi-statisztikak>
- [171] Hungarian Population Data. (2016). <https://www.ksh.hu/mikrocenzus2016/>

- [172] Network, A.M.: National Air Pollution Measurement Network,. *Data available at request.* <http://levegominoseg.hu/manualis-merohalozat>
- [173] Rosow, I.: The social effects of the physical environment. *Journal of the American Institute of Planners* **27**(2), 127–133 (1961)
- [174] Kahneman, D., Krueger, A.B.: Developments in the measurement of subjective well-being. *Journal of Economic perspectives* **20**(1), 3–24 (2006)
- [175] Erwig, M.: The graph voronoi diagram with applications. *Networks: An International Journal* **36**(3), 156–163 (2000)
- [176] Deritei, D., Lazar, Z.I., Papp, I., Jarai-Szabo, F., Sumi, R., Varga, L., Regan, E.R., Ercsey-Ravasz, M.: Community detection by graph voronoi diagrams. *New Journal of Physics* **16**(6), 063007 (2014)
- [177] Hohenberg, P.M., Lees, L.H.: *The Making of Urban Europe: 1000-1994.* Harvard University Press, ??? (1996)
- [178] Brueckner, J.K., Thisse, J.-F., Zenou, Y.: Why is central paris rich and downtown detroit poor?: An amenity-based theory. *European economic review* **43**(1), 91–107 (1999)
- [179] Kövér, G., Gyáni, G.: Magyarország társadalomtörténete a reformkortól a második világháborúig. *Osiris* (1998)
- [180] Dittmar, H., Ohland, G.: *The New Transit Town: Best Practices in Transit-oriented Development.* Island Press, ??? (2012)
- [181] Rogers, S.H., Halstead, J.M., Gardner, K.H., Carlson, C.H.: Examining walkability and social capital as indicators of quality of life at the municipal and neighborhood scales. *Applied Research in Quality of Life* **6**(2), 201–213 (2011)
- [182] Rinner, C.: A geographic visualization approach to multi-criteria evaluation of urban quality of life. *International Journal of Geographical Information Science* **21**(8), 907–919 (2007)
- [183] Gavrilidis, A.A., Ciocănea, C.M., Niță, M.R., Onose, D.A., Năstase, I.I.: Urban landscape quality index–planning tool for evaluating urban landscapes and improving the quality of life. *Procedia Environmental Sciences* **32**, 155–167 (2016)
- [184] Talen, E.: Pedestrian access as a measure of urban quality. *Planning Practice and Research* **17**(3), 257–278 (2002)

- [185] Santos, L.D., Martins, I.: Monitoring urban quality of life: The porto experience. *Social Indicators Research* **80**(2), 411–425 (2007)
- [186] Roback, J.: Wages, rents, and the quality of life. *Journal of political Economy* **90**(6), 1257–1278 (1982)
- [187] Blomquist, G.C., Berger, M.C., Hoehn, J.P.: New estimates of quality of life in urban areas. *The American Economic Review*, 89–107 (1988)
- [188] Lora, E.A., Powell, A.: A new way of monitoring the quality of urban life (2011)
- [189] ingatlannet: Budapest Real Estate Statistic. <https://www.ingatlannet.hu/statisztika/Budapest>
- [190] Cassiers, T., Kesteloot, C.: Socio-spatial inequalities and social cohesion in european cities. *Urban Studies* **49**(9), 1909–1924 (2012)
- [191] Garcia, B.: Cultural policy and urban regeneration in western european cities: lessons from experience, prospects for the future. *Local economy* **19**(4), 312–326 (2004)
- [192] Gobillon, L., Selod, H., Zenou, Y.: The mechanisms of spatial mismatch. *Urban studies* **44**(12), 2401–2427 (2007)
- [193] Sport Associations in Budapest. (2019 (accesed June 3, 2019). [https://sportmegoldasok.hu/klub\\_budapest.ht](https://sportmegoldasok.hu/klub_budapest.ht)
- [194] Kindergartens & Daycare, Primary Education, Secondary Education. (2019 (accesed June 3, 2019). <http://budapest.imami.hu/iskolak-ovodak-bolcsodek/>
- [195] Art and Music Schools. (2019 (accesed June 3, 2019). <http://budapest.imami.hu/iskolak-ovodak-bolcsodek/muveszeti-es-sportiskolak>
- [196] Pediatricians, Gynecologists. (2019 (accesed June 3, 2019). <http://budapest.imami.hu/egeszseg>
- [197] Social Welfare. (2019 (accesed June 3, 2019). <http://szocialisportal.hu/intezmenykereso>
- [198] Culture Centers. (2019 (accesed June 3, 2019). <https://holmivan.valami.info/kultura-lista-220/kultura>, Budapest
- [199] Indoor Playgrounds. (2019 (accesed June 3, 2019). <http://budapest.imami.hu/szolgaltatok/szulinapi-party>

- [200] Healthcare (hospitals, Private and Public Clinics, Specialists). (2019 (accesed June 3, 2019)). <https://www.budapestinfo.eu/gyogyitas>
- [201] Fitness and Training Facilities. (2019 (accesed June 3, 2019)). <https://www.budapestinfo.eu/sportolas/fitnesstermek>
- [202] Outdoor Fitness Training Facilities. (2019 (accesed June 3, 2019)). <https://www.google.com/maps/d/u/0/viewer?msa=0&hl=en&ie=UTF8&t=h&ll=47.481378772238784%2C19.130302999999913&spn=1.329955%2C3.757785&source=embed&mid=1qIGN-mnKuGrvNxo0ENmACm78RP8&z=11>
- [203] Thermal and Spa. (2019 (accesed June 3, 2019)). <https://www.budapestinfo.eu/furdozes>
- [204] Playgrounds and Parks. (2019 (accesed June 3, 2019)). <https://zoldkalauz.hu>



# CHAPTER 7

---

## APPENDICES

---

### 7.1 Life quality as walkability

#### 7.1.1 Secondary data sources

- Sport associations in Budapest [193]
- Kindergartens, daycares, primary and secondary education [194]
- Art and music schools [195]
- Child health services [196]
- Social welfare system (eg.: elderly care) [197]
- Culture centers [198]
- Indoor playgrounds [199]
- Healthcare (hospitals, private and public clinics, specialists) [200]
- Fitness and training facilities [201]
- Outdoor fitness facilities [202]
- Thermal baths and spa [203]
- Playgrounds and parks [204]

### 7.1.2 Categorization

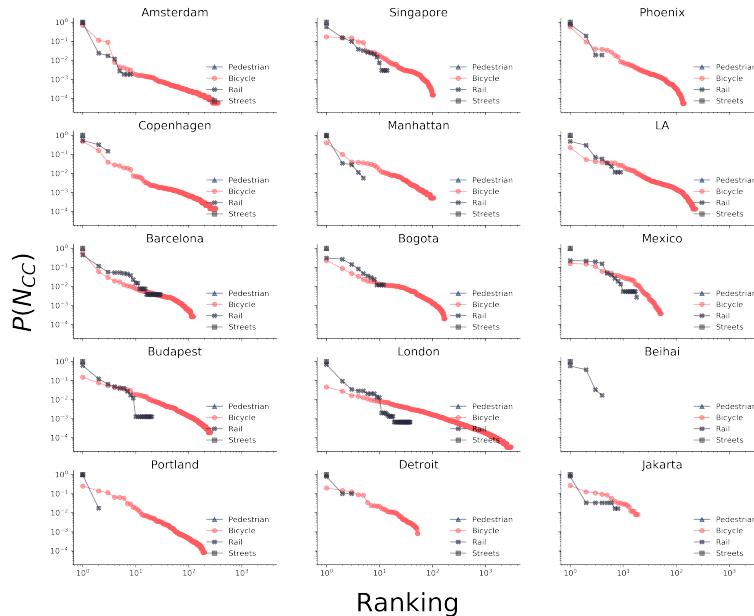
The amenities used for each category are as follows:

- Culture

## 7.2 Data-driven strategies for optimal bicycle network growth

### 7.2.1 Data

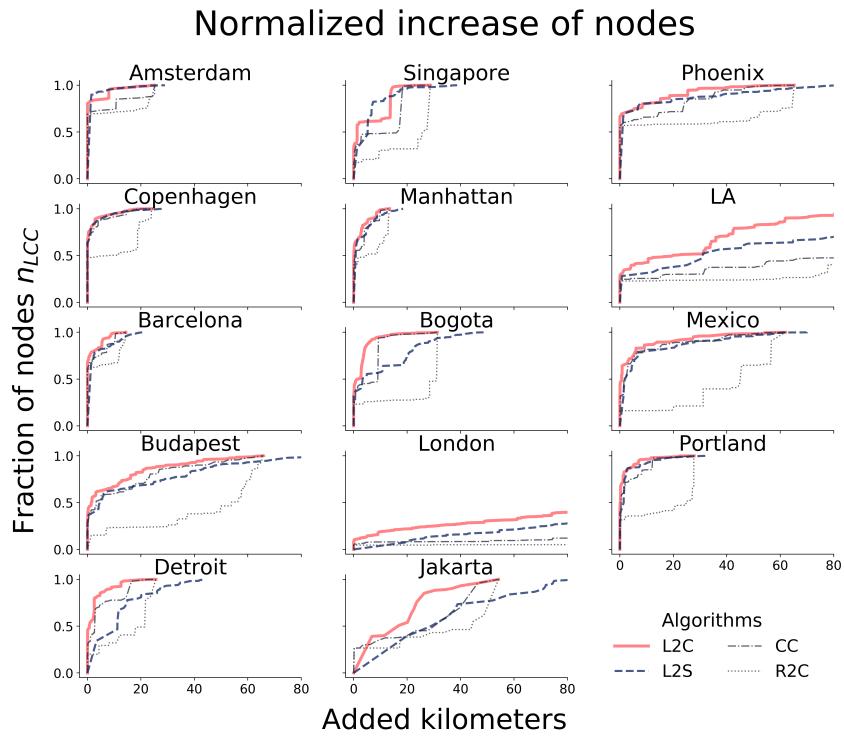
Figure 7.1 shows the connected component size distribution  $P(N_{cc})$  for all considered layers and cities.



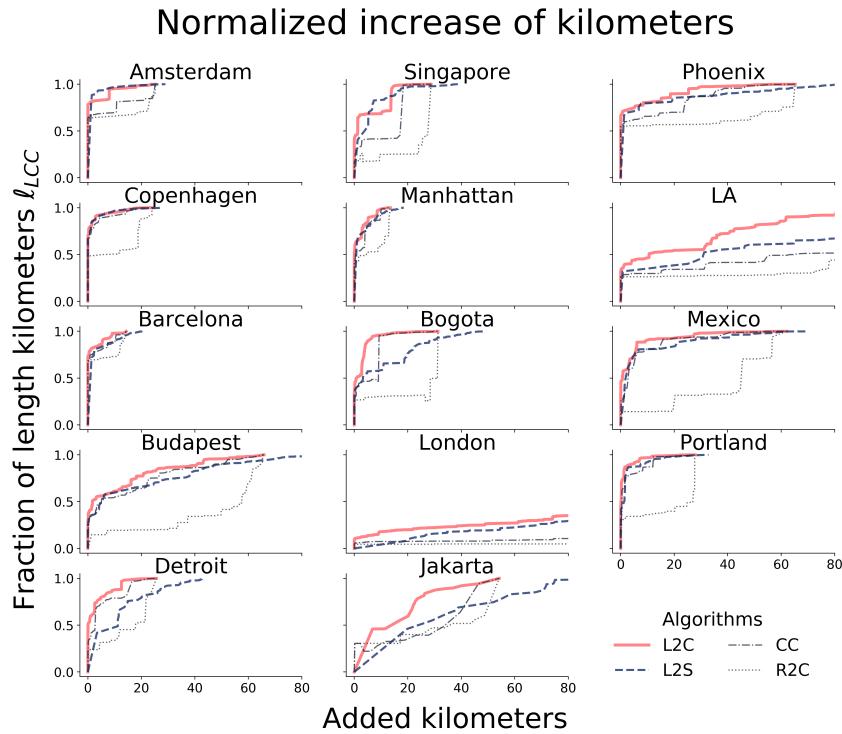
**Figure 7.1.** The connected component size distribution [ $P(N_{cc})$ ] for all cities and layers is well connected except in the bicycle layer. London has the most fragmented bicycle infrastructure layer, with more than 3000 components

### 7.2.2 Bicycle network improvement

Here we show the improvement of the bicycle network after the implementation of the algorithms. We measure the improvement with four different metrics. Two of them implement the notion of connectedness: i) Fraction of nodes inside the largest connected component compared to the total number of nodes in the bicycle layer, and ii) the fraction of link kilometers inside the largest connected component. In Figure 7.2 and ?? we show these two measures for fourteen different cities. We also quantify iii) bicycle-to-car directness to answer the question “how direct are the average routes of bicycles compared to cars?”. Finally, in order to measure the cumulative efficiency of our algorithms, we define the metric: iv)  $G_{LCC}$  as the relative gain of bicycle path kilometers in the largest connected component. In Figures 7.5 and 7.6 we report these two measures for all algorithms and cities considered.



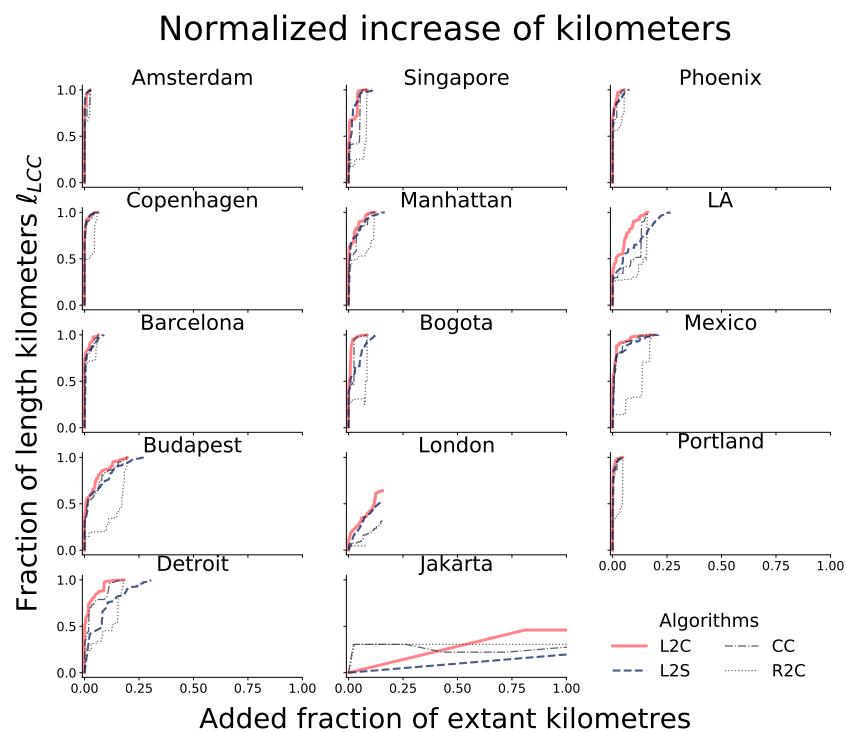
**Figure 7.2.** Normalized increase in nodes inside the largest connected component ( $n_{LCC}$ ).



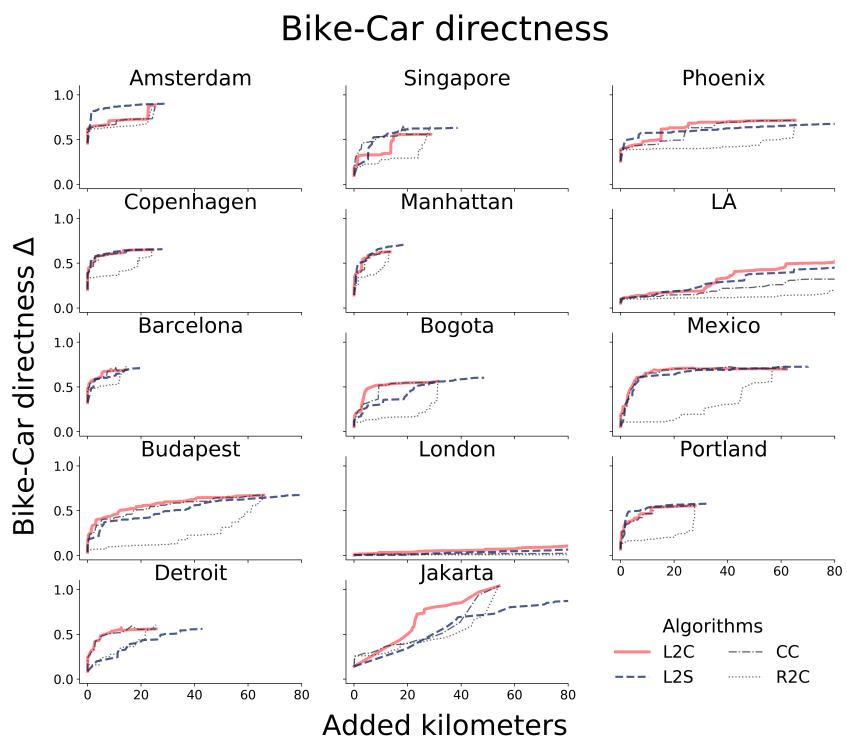
**Figure 7.3.** Normalized increase in kilometers inside the largest connected component ( $\ell_{LCC}$ ).

### 7.2.3 Bicycle network and 30 km/hr streets

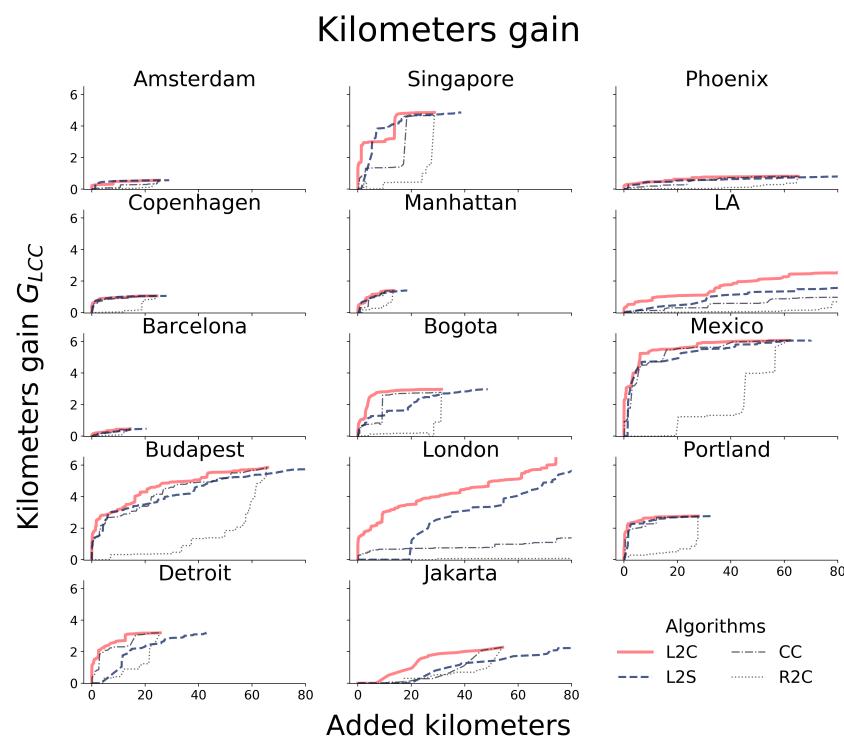
We applied the algorithms to the bicycle infrastructure and all the bikeable streets, those with a speed limit of 30 km/hr or less (see Figures 7.7, 4.4, 7.8, and 7.9).



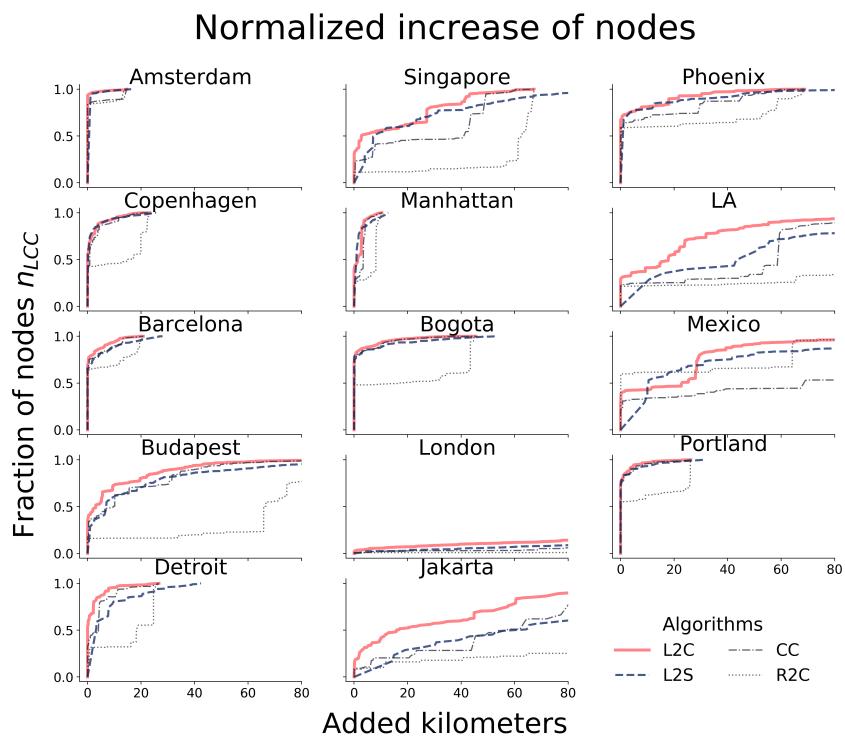
**Figure 7.4.** Normalized increase in kilometers inside the largest connected component ( $\ell_{LCC}$ ) versus the fraction of extant kilometers to be added.



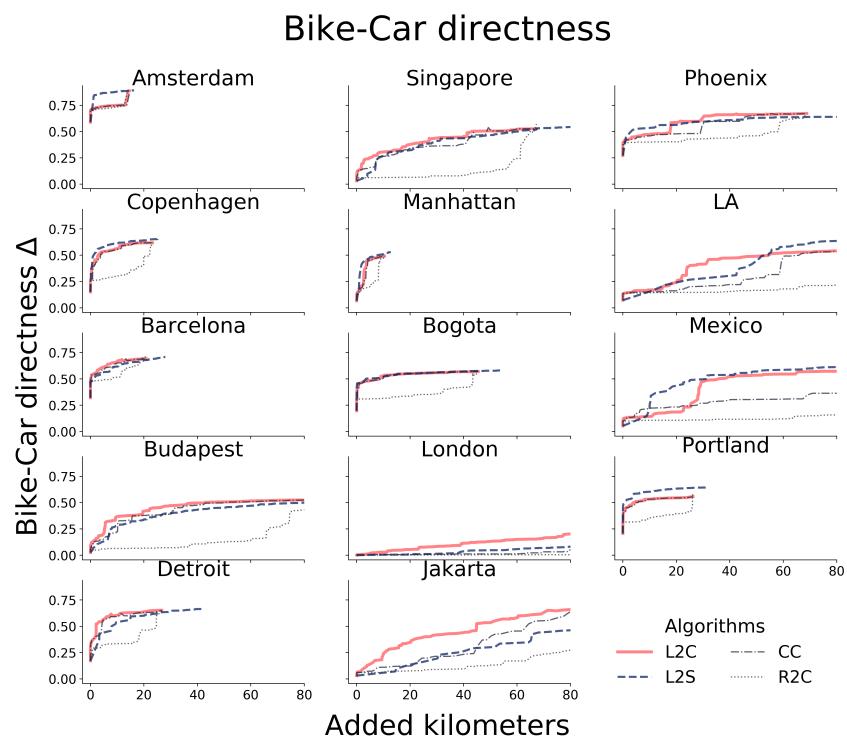
**Figure 7.5.** Bike-car directness  $\Delta$  per invested kilometers.



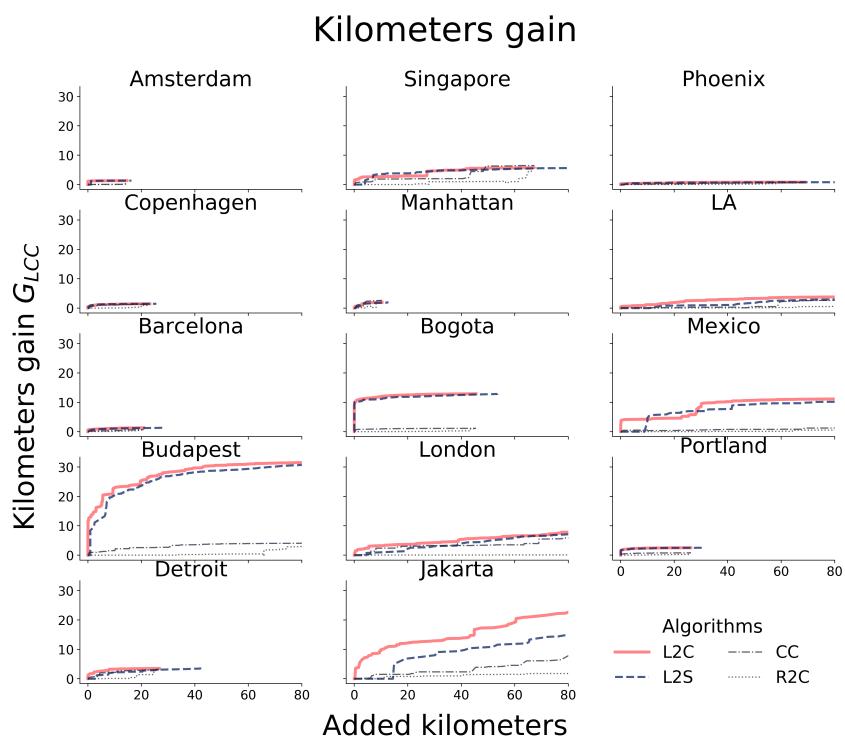
**Figure 7.6.** Kilometers gain in the largest connected component.



**Figure 7.7.** Normalized increase in nodes inside the largest connected component ( $n_{LCC}$ ).



**Figure 7.8.** Bike-car directness  $\Delta$  per invested kilometers.



**Figure 7.9.** Kilometers gain in the largest connected component.

