Doctoral Thesis

Traffic operations on urban grid networks

Author(s):
Ortigosa Marin, Javier

Publication Date:
2015

Permanent Link:
https://doi.org/10.3929/ethz-a-010616357

Rights / License:
In Copyright - Non-Commercial Use Permitted
TRAFFIC OPERATIONS ON URBAN GRID NETWORKS

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by
JAVIER ORTIGOSA MARIN

Enginyer de Camins, Canals i Ports, UPC (BarcelonaTech)
Master of Science, Chalmers University of Technology

born on 26.10.1982

citizen of
Spain

Accepted on the recommendation of
Dr. Monica Menendez
Prof. Kay W. Axhausen
Prof. David M. Levinson
Prof. Nikolas Geroliminis

2015
Foreword

Discussions on the relation between form and function of urban structures are as relevant today as they were one hundred years ago, especially when considering that the urban layout is the backbone to urban transportation systems. Hence, it is very important to link traffic engineering and urban planning disciplines for a better design and redesign of cities. The work of Javier Ortigosa uses abstract grid networks to derive insights on the relation between urban structure and traffic performance. He looks first into different street configurations (e.g., one-way vs. two-way streets), and how they affect the overall traffic operations (including travel times and distances, as well as the overall network capacity). He then investigates different strategies for space removal, and their effects on traffic performance both at the local and at the network level. Last, he explores how macroscopic models can be employed to assess and evaluate urban networks; bringing some of the concepts presented before into a more pragmatic realm.

Overall, the thesis is rather interesting from both a practical and a scientific perspective. The work provides innovative methodologies and thought-provoking results. The topic also seems very timely, as many cities re-think the organization of their downtown networks (with conversions between one-way and two-way streets), as well as their space allocation policies in order to create more sustainable urban spaces. As a matter of fact, some of the presented findings could be useful beyond the scope of this thesis to understand traffic phenomena in relation to urban networks, as well as to implement different traffic management and control schemes. On behalf of the Traffic Engineering research group at the ETH Zurich, I thank Javier Ortigosa for using his endless creativity to build a bridge between urban planning and transport engineering. I hope such bridge help to inform better urban policies and traffic management strategies.

Dr. Monica Menendez
Director of research group Traffic Engineering
Institute for Transport Planning and Systems (IVT)
ETH Zurich
Abstract

The main purpose of this PhD thesis is to reach a better understanding of traffic operations behavior in urban networks. We want to furnish urban planners and traffic engineers with a general perspective of how street patterns affect traffic, which later, they can take into account when planning. This research is motivated by studying how planning of urban road space can be done in a more sustainable way, and by how innovative urban traffic management techniques can be more efficiently implemented.

This first part of the dissertation is devoted to an in-depth study of street configurations on urban grid networks: two-way streets (TW), one-way streets (OW), and an intermediate solution, two-way streets with prohibited left turns (TWL). To date, no consensus has been reached on which street network configuration provides the optimal trade-off between these characteristics. On the one hand, strictly focusing on the movement of cars, two-way street networks provide a higher accessibility but also offer less capacity at intersections. On the other hand, one-way street networks provide intersections with higher capacities and travel speeds, but force drivers to travel longer routes on average. The analysis was carried out employing three different methods: analytical formulations, Static Traffic Assignment (STA), and Dynamic Traffic Assignment (DTA) based on microsimulation. The three methods present very different traffic assignment levels of complexity, computational times, and evidently research outcomes. Employing all of them makes this research robust and results more complete.

TW networks provide the shortest distance traveled between two points and the highest route redundancy and flexibility. However, they present the lowest vehicle capacity at intersections which penalizes networks heavily. TWL networks offer the best trade-off between distance travelled and capacity at intersections. The main disadvantage is that they have a very limited route redundancy which translates into more heterogeneous spread of congestion. OW networks supply the highest capacity per movement at intersections although they also trigger the largest distance traveled. OW networks spread congestion more homogeneously than TWL networks because they have a higher route redundancy.

The second part of this thesis deals with the removal of space in urban settings. Our aim is to understand and quantify how this removal affects drivers and the overall system. We also study abstract grid networks representing urban environments. First, we look at different strategies of link removal in the grid networks; later, we study the
effects by only removing lanes. Same as in the first part of the dissertation, we employ two methods to model traffic: the STA and the microsimulation with a DTA module.

Results clearly show that it is possible to remove certain amount of links from an urban grid without losing connectivity and worsening traffic conditions excessively. Evidently, traffic impacts depend on the removal strategy. Removing streets from the center creates the highest impacts because it affects the redundancy of routes, and hinders traffic flows from spreading more homogeneously. In that case, intersections with really high traffic loads will easily trigger congestion in the system. Removing streets from the perimeter, instead, allows the system to keep the high connectivity in the center that allows spreading traffic loads more evenly. Lane removal does not change the network connectivity but it makes traffic distribution more heterogeneous. Results show that central removal impacts the system more than peripheral removal in terms of capacity drop, congestion appearance and gridlock time. Congestion propagation depends on two factors: the number of one-lane streets in networks, and the origin of congestion. If congestion is originated in the perimeter, it spread over the network faster than if congestion started in the center. Although networks do not present clear congestion propagation patterns, central removal has a tendency to propagate congestion more from the center, whereas peripheral removal has a slight disposition to propagate congestion more from the perimeter.

The third part of the dissertation dealt with the amount and location of fixed monitoring resources that a city should have to measure a reliable Macroscopic Fundamental Diagram (MFD) control scheme. This research is especially important for practitioners as we study a real case: the city of Zurich. Most cities do not have much traffic information a priori, so we propose different blind strategies to select only a certain amount of the links to be monitored to create an MFD. The results (obtained with simulation data) show that independently of the strategy used, a minimum of 25% of the links ensures a fairly accurate MFD. Selecting links randomly but with a certain weight towards central activity places (e.g. for the case of Zurich, the central train station) not only provides good results but also maintains a low variability, especially if at least 15% of the links are selected. If cities already have traffic counts or a simulation model where values of flows and densities can be estimated, the selection of links can be done in a very efficient way. Finally, like in the case of Zurich, some cities might already have some traffic counts for macroscopic control even if the coverage is not really high. Employing those same links or streets to create MFDs proved, for the case of Zurich, to be a very efficient strategy.
Zusammenfassung


Netzwerke mit Gegenverkehrsstrassen weisen die kürzesten gefahrenen Distanzen zwischen zwei Punkten, sowie die höchste Routenredundanz und Flexibilität auf. Sie haben jedoch die tiefste Knotenkapazität, was sich nachteilig auf solche Netzwerke auswirkt. TWL-Netzwerke zeigen die beste Ausgewogenheit zwischen gefahrener Distanz und Kapazität an Knotenpunkten. Der grösste Nachteil ist, dass die Routenredundanz ziemlich limitiert ist und sich somit Stauaufkommen heterogener verteilen. OW-Netzwerke haben die grösste Kapazität an Knoten. Sie erzeugen allerdings die längsten gefahrenen Distanzen. In OW-Netzwerken breiten sich Staus
homogener als in TWL-Netzwerken aus, da eine höhere Routenredundanz vorhanden ist.


Der dritte Teil der Arbeit befasst sich mit der Anzahl und der Position von fix installierten Überwachungsmassnahmen, um mithilfe deren ein zuverlässiges Kontrollschema, ein sogenanntes makroskopisches Fundamentaldiagramm (MFD), messen zu können. Dieser Forschungsansatz ist speziell für Leute aus Praxis von hoher Relevanz, da mit der Stadt Zürich ein reales Fallbeispiel untersucht wird. Ein Grossteil der Städte besitzt von jeher nicht ein Übermass an Verkehrsdaten, weshalb vereinfachte
Acknowledgements

During this dissertation, many times I have felt like I was “swimming against the stream”. Thanks to all of you who support me, help me, guide me, and comfort me, this dissertation has come to a successful conclusion.

My advisor, Monica Menendez, would say: “When life gives you lemons, make lemonade”. In this story, I, the lemon, have needed a lot of squeezing to produce decent lemonade—see document attached. I am very grateful for her many efforts, time, and patience which she devoted to me, and made me a better researcher. Our Judging–Perceiving\(^1\) interactions have been overall very fruitful although sometimes exhausting and consuming. I am sorry for that, Boss.

And\(^2\) I am also very grateful to Monica for hiring me in the first place and trusting me with many other responsibilities and activities. Despite all the efforts, when I look back, I am very happy and satisfied to have contributed a little bit to create this research group, the traffic engineering group (SVT). Thank you, SVTers, for your help and company during these years. Also I want to thank all my colleagues at the Institute of Transport Planning and Systems (IVT), and those at EPF Lausanne and at other Swiss Universities who I have met at least once a year, thanks to STRC.

I did not do this journey alone. Along the way, I found a strong research ally, Vikash V. Gayah, who helped me a lot, especially in the first part of this dissertation, gave me confidence and further connected me to the research world. I also counted on my office comrade, Qiao Ge, who was always there when I needed help and he never complained; 谢谢.

I would like to thank Prof. Kay W. Axhausen for his valuable inputs and advice during my dissertation as well as my other thesis examiners, Prof. David M. Levinson and Prof. Nikolas Geroliminis. Thanks to you, my research is now stronger.

This research was supported by the ETH Research Grant ETH-38 13-1, “TERRAIN: Traffic Operations on Urban Grid Networks”. Taking advantage of this, I want to thank all the ETH personnel that were always very helpful with my enquiries. Additionally, I


\(^2\) This conjunction at the beginning of the sentence has been included intentionally 😊.
wish to thank Christian Heimgartner for providing us with the Zurich simulation model that was key in the third part of this research.

I would probably not be here today if I had not been at the Center for Innovation in Transport (CENIT, BarcelonaTech) before. I want to thank the people I met during my time there for introducing me into this research world.

Last but not least, I want to thank my family and my friends. I mean it. It is very difficult to mention here all of you since you are many and in many different places. Both to family and friends, I feel very lucky to have all of you, here and there, to give me the strength and support that I need. That goes especially to my parents, Angel and Laura, to whom I owe everything I am.
Table of contents

Foreword............................................................................................................................i
Abstract...........................................................................................................................iii
Zusammenfassung.............................................................................................................v
Acknowledgements ........................................................................................................ix
Table of contents ..........................................................................................................xi
List of figures ..................................................................................................................xv
List of tables ..................................................................................................................xvii
Chapter 1. Introduction ..............................................................................................1
  1.1. Background and motivation ..............................................................................1
  1.2. Research objectives .........................................................................................4
  1.3. Thesis outline ..................................................................................................6
  1.4. Outcomes of this research ...............................................................................8
    1.4.1. Research tools developed .......................................................................9
    1.4.2. Main findings ..........................................................................................9
PART 1. Street configuration .....................................................................................13
Chapter 2. One-way vs. two-way street grid networks comparison .....................15
  2.1. Introduction ....................................................................................................15
  2.2. Analytical formulations for low traffic volumes ...........................................16
    2.2.1. Two-way streets ....................................................................................17
    2.2.2. Two-way streets with prohibited left turns .........................................20
    2.2.3. One-way streets ...................................................................................21
    2.2.4. Average trip length and average turns per trip ...................................22
    2.2.5. Travel time on networks ......................................................................23
  2.3. Static Traffic Assignment for high traffic volumes ........................................25
2.3.1. Average trip length and average turns per trip ........................................ 27
2.3.2. Travel time on networks .......................................................................... 29
2.4. Different demand patterns .......................................................................... 33
2.5. Causes of general trends in network behavior ............................................ 35
   2.5.1. Comparison between analytical approach and simulation ....................... 35
   2.5.2. Heterogeneity of congestion and route redundancy .............................. 37
2.6. Conclusions.................................................................................................. 39

Chapter 3. Traffic dynamics on one-way and two-way street grid networks ........................................... 43
3.1. Introduction.................................................................................................... 43
3.2. Methodology.................................................................................................. 45
   3.2.1. Network characteristics ......................................................................... 45
   3.2.2. Dynamic Traffic Assignment module in VISSIM 6 ................................. 46
   3.2.3. Network Exit Function and other macroscopic indicators ...................... 48
3.3. Results ........................................................................................................... 49
   3.3.1. Comparison of full NEFs ....................................................................... 49
   3.3.2. Influence of the route choice ................................................................. 51
   3.3.3. Congestion speed and demand served .................................................. 53
   3.3.4. Influence of the left-turn pocket in TW networks ................................. 55
   3.3.5. Macroscopic convergence of the NEF .................................................... 56
3.4. Conclusions.................................................................................................... 59

PART 2. Space removal ....................................................................................... 63

Chapter 4. Street removal on grid networks ....................................................... 65
4.1. Introduction.................................................................................................... 65
4.2. Methodology.................................................................................................. 66
   4.2.1. Network design ..................................................................................... 66
   4.2.2. Link removal strategies ......................................................................... 68
4.2.3. Traffic indicators.............................................................................................. 69
4.3. Results ...................................................................................................................... 69
  4.3.1 Network performance indicators ................................................................. 70
  4.3.2. Network intersections ...................................................................................... 73
  4.3.5. Demand level variation .................................................................................... 76
4.4. Conclusions ........................................................................................................... 76

Chapter 5. Traffic dynamics of lane removal on grid networks .....................79
5.1. Introduction ............................................................................................................ 79
5.2. Methodology ......................................................................................................... 81
  5.2.1. The one-way grid network .............................................................................. 81
  5.2.2. Lane removal strategies ................................................................................... 82
  5.2.3. Macroscopic traffic indicators ....................................................................... 84
  5.2.4 Congestion propagation model ....................................................................... 84
5.3. Results .................................................................................................................... 86
  5.3.1. Network capacity ............................................................................................ 87
  5.3.2. Time to congestion and gridlock ..................................................................... 88
  5.3.3. Traffic inhomogeneity ..................................................................................... 91
  5.3.4. Origin and spread of congestion ..................................................................... 94
  5.3.5. Level of service and traffic demand ............................................................... 97
5.4. Conclusions ......................................................................................................... 99

PART 3. Urban traffic monitoring resources ......................................................103

Chapter 6. Study on the number and location of measurement
  points for an MFD perimeter control scheme .....................................................105
  6.1. Introduction ............................................................................................................ 105
  6.2. Methodology ........................................................................................................ 108
    6.2.1. Creating an MFD with p links ..................................................................... 108
    6.2.2. Evaluating the accuracy of a pMFD ............................................................ 108
6.2.3. Strategies for creating incomplete MFDs ............................................. 110

6.3. Results. Case study: the inner city of Zurich ........................................ 112

6.3.1. The complete MFD of Zurich .............................................................. 113

6.3.2. Quasi-optimal selection strategy ......................................................... 114

6.3.3. Blind selection strategies ................................................................. 115

6.3.4. Use of ZuriTraffic links to create a pMFD ............................................ 117

6.3.5. Different demand patterns .............................................................. 121

6.3.6. Applicability and limitations ............................................................. 122

6.3. Conclusions ......................................................................................... 123

Chapter 7. Conclusions ................................................................................ 125

7.1. Research findings ................................................................................. 126

7.2. Recommendations .............................................................................. 128

7.3. Future research .................................................................................. 131

Bibliography .............................................................................................. 133
List of Figures

Figure 1.1. Streets and patterns ................................................................................... 1
Figure 1.2. Eixample plan in Barcelona. ..................................................................... 5
Figure 1.3. Thesis outline ............................................................................................ 8
Figure 2.1. The three network configurations ........................................................... 17
Figure 2.2. Distance and turns (analytical)................................................................. 18
Figure 2.3. Average trip length and turns per trip (analytical). ................................. 23
Figure 2.4. Average travel time (analytical).............................................................. 25
Figure 2.5. Average trip lengths and turns (STA) ..................................................... 28
Figure 2.6. Average travel times (STA)................................................................... 30
Figure 2.7. Average V/C ratios at intersections (STA) ............................................. 32
Figure 2.8. Average trip length and average travel times and V/C ratios................. 34
Figure 2.9. Examples of path redundancy and flow spread (STA) ......................... 38
Figure 2.10. CDF of travel time (STA)..................................................................... 39
Figure 3.1. Detail of an intersection (DTA) .............................................................. 45
Figure 3.2. Flow chart of the DTA process ............................................................... 47
Figure 3.3. NEFs for TW, TWL, and OW................................................................. 50
Figure 3.4. NEFs for different krc values................................................................. 52
Figure 3.5. Congestion and gridlock times for TW, TWL, and OW ......................... 54
Figure 3.6. Congestion and gridlock times for different left turn pockets.............. 56
Figure 3.7. Absolute errors on trip ending rates and aggregated NEFs............... 58
Figure 4.1. The network model ................................................................................. 67
Figure 4.2. Flow distribution in different network configurations ......................... 70
Figure 4.3. Total travel time indexed for removal strategies.................................... 71
Figure 4.4. Network performance indicators (1/2).................................................. 72
Figure 4.5. Network performance indicators (2/2) ................................................... 73
Figure 4.6. Number and type of maneuvers at intersections ..................................... 75
Figure 4.7. Total travel times to evaluate the gain in capacity ................................. 75
Figure 4.8. Total travel time for different demand loads .......................................... 76
Figure 5.1. The 10 network scenarios analyzed.......................................................... 83
Figure 5.2. Congestion propagation model................................................................ 86
Figure 5.3. NEFs and network capacity for removal strategies .............................. 88
Figure 5.4. Evolution of congestion for removal strategies ..................................... 90
Figure 5.5. Evolution of congestion during the DTA iterations ............................... 91
Figure 5.6. Average density, std. dev. of lane removal scenarios ............................. 92
Figure 5.7. Traffic distribution for central and peripheral removal .......................... 93
Figure 5.8. Evolution of the Maximum Connected Component ............................... 95
Figure 5.9. First links to become congested ............................................................. 96
Figure 5.10. Demand to maintain the same LOS vs. Space gain .............................. 98
Figure 6.1. Tabu Search structure........................................................................... 111
Figure 6.2. Complete MFD for the inner city of Zurich ......................................... 114
Figure 6.3. Accuracy vs. network coverage of the quasi-optimal case .................... 115
Figure 6.4. Accuracy vs. network coverage of selection strategies ....................... 116
Figure 6.5. Accuracy vs. network coverage for random selection ......................... 117
Figure 6.6. Individual fundamental diagrams ZuriTraffic links ............................. 118
Figure 6.7. MFD with ZuriTraffic links and complete MFD................................. 120
List of Tables

Table 2.1. Different networks analyzed .................................................................27
Table 2.2. Errors between analytical formulations and simulations.........................36
Table 3.1. Comparison of STA and DTA methods ...................................................60
Table 6.1. Average errors between individual ZuriTraffic links and a complete MFD. .................................................................................................................119
Chapter 1

Introduction

1.1. Background and motivation

How do urban patterns affect traffic? Are there networks that can handle traffic better than others? How do changes in street infrastructure affect mobility?

These are some of the questions that inspired the present research. Urban planning and design often see transportation from a static perspective, not taking into consideration all interactions and relations that occur in urban networks. Traffic operations are a characteristic of an urban place, hence, urban planners should have a deeper understanding of how transportation flows distribute in urban networks and how network changes affect the overall urban system. As Marshall (2005) represents in Figure 1.1, streets and patterns are the intersection between transportation engineers and urban planners. Typically, the former see street patterns as something given and focus on studying operations. The latter, instead, often consider street patterns design as a land use distribution where a certain amount of urban space should be devoted to that purpose (Marshall, 2005); without thinking about the implications of transportation flow distributions.

![Figure 1.1. Streets and patterns (Marshall, 2005).](image-url)
First, our motivation is to reach a better understanding of traffic operations behavior in urban networks. We seek to generalize these learnings in a simple and easy way so they can serve as liaison between traffic engineers and urban planners. In other words, we want to furnish urban planners with a general perspective of how street patterns affect traffic, which later, they can take into account when planning. At the same time, we want to encourage traffic engineers to carry out research aiming at influencing urban planning decisions. We believe urban problems have to be solved in a transdisciplinary way.

There are numerous studies on how the urban form (location, land use, size and shape of settlements, population density, road network, and urban layout (Stead and Marshall, 2001)) affects travel behavior. Many look at real cities and how the urban form factors are related to travel characteristics such as mode choice, vehicle miles traveled, or travel time (e.g. Frank and Pivo, 1994; Cervero and Gorham, 1995; Levinson and Kumar, 1997; Crane and Crepeau, 1998). However, the connection to mobility is weak, as mostly, the interaction between urban pattern and demand is covered but not traffic operations.

The spatial characteristics of urban patterns have been extensively studied. Researchers have developed indicators to describe network properties such as connectivity (e.g. Garrison and Marble, 1961; Kansky, 1963) or centrality measures (e.g. Freeman, 1979; Sheikh Mohammad Zadeh and Rajabi, 2012) supported by graph theory concepts. With the progress of technology and computational tools, urban morphology studies have evolved considerably. Urban scientists have established theories of how network elements interact (e.g. Alexander et al., 1977; Hillier and Hansen, 1984); and have identified topological properties in large scale networks, making metropolis more comparable (e.g. Jiang and Claramunt, 2004; Jiang, 2007; Lämmer et al., 2006; Chan et al., 2011). Even so, among the spatial analysis studies, only some works take into account traffic modeling, either implementing some traffic assignment (e.g. Scott et al., 2006; Xie and Levinson, 2009; Sheikh Mohammad Zadeh and Rajabi, 2012; Vitins et al., 2012) or collecting real traffic data (e.g. Hackney et al., 2007; Parthasarathi and Levinson, 2011).

Second, this research is motivated by studying how planning of urban road space can be done in a more sustainable way. Nowadays, we are aware that space in cities is scarce. Yet so much is still devoted for private mobility and so little, in comparison, to other transportation modes and activities. Up to the end of the 19th century and beginning of the 20th century, most cities were developed as compact, mixed-use communities. However, with the development of modern architecture and the widespread use of private vehicles, urban planners started separating land uses and
INTRODUCTION

creating residential suburbs. This led to what is known today as urban sprawl (Steuteville and Langdon, 2009; Sieverts, 1997) which in turn, increased car dependency.

Several decades ago, some researchers already alerted that an increase of road capacity could actually lead to an increase of traffic and eventually, travel times and congestion (Downs, 1962; Smeed, 1968; Thomson, 1972). Transportation scientists defined that phenomenon as induced demand, i.e. the change on demand because of the user’s perception that an increase of capacity can deliver better travel times (Hills, 1996; Goodwin, 1996; Lee, 1999). Many researches have modeled and quantified this effect looking at the relation between the amount of increased travel, and increased road supply (Hansen and Huang, 1997; Barr, 2000; Noland, 2001; Cervero and Hansen, 2002; Cervero, 2003; Levinson and Karamalaputi, 2003; Weiss and Axhausen, 2009). At the same time, many have studied the opposite effect. In some cases, when reductions of network capacity have taken place (e.g. construction work, natural disasters, urban measures) there has been a reduction of the overall amount of traffic (Goodwin et al., 1998; Cairns et al., 2002; Hunt, 2002; Clegg, 2007; Zhu et al., 2010; Xie and Levinson, 2011); as Cairns et al. (2002) call it: disappearing traffic.

Planners are shifting towards a shared use of space in cities between cars and citizens, reclaiming the streets for high quality urban spaces (European Commission, 2004). Road space needs to be better allocated and more efficiently organized; in other words, cities need to do more with less. On the one hand, we want to study how streets can be more space-efficient designed. To date, there is no consensus on basic street configurations and which one of them provides the best mobility degree (Walker et al., 2000; Meng and Thu, 2004; Gayah and Daganzo, 2012; Boyles et al., 2014). We want to understand how street patterns hinder traffic performance and which street configurations are able to handle traffic better. On the other hand, we want to provide an objective and quantitative answer to the effects of shifting road space to other uses. Actions in that direction bring a lot of controversy which is often based on subjective opinions, e.g. traffic will become extremely chaotic (Cairns, 2002). Common sense tells us that traffic will worsen in the short term, as car demand remains fairly constant and supply is reduced. The question is: to what degree? And: is it possible to minimize its impact?

Third, we want to bring closer traffic engineering tools and models to urban planners and practitioners. We look especially at city-wide macroscopic traffic models because they can feed innovative traffic control mechanisms to improve the quality of cities. The first macroscopic models linked network properties and traffic performance (Smeed, 1966; Thomson, 1967; Wardrop, 1968; Zahavi, 1972). Later, researchers
developed models that related traffic performance in a city and the accumulation of vehicles in that city (Herman and Prigogione, 1979; Herman and Ardekani, 1984; Mahmassani et al., 1987) following the Fundamental Diagram on a link. However, they did not describe congested states. It was not until recently that Daganzo (2007) and Geroliminis and Daganzo (2008) proposed the Macroscopic Fundamental Diagram (MFD) that connects the number of trips in a city and the number of vehicles circulating in it. The MFD covers all traffic states and can feed traffic control schemes in cities. However, despite its potential, the use of the MFD outside the research arena, in real cities, has been very limited. We want to study how these access control mechanisms can be implemented easily and with fewer resources, and help practitioners to understand them. In addition, we believe that these macroscopic models can also be applied to assess and evaluate city networks, i.e. we aim at using the MFD as a network performance indicator.

1.2. Research objectives

The general objective of this research is to better understand urban structures and their interactions with traffic, in order to find more efficient ways of organizing traffic and recovering road space for other transportation modes or urban uses. We believe that this research can help traffic engineers and urban planners to identify strategies for maintaining an adequate level of traffic performance in cities, while supporting decisions for improving other transportation modes and activities given the limited amount of available space. Contrary to many urban design/planning studies, here we emphasize an in-depth analysis of traffic. This is necessary in order to explore the consequences of different urban policies from the traffic operations perspective.

Our research focuses mainly in one particular street pattern: the grid. In addition to the author’s fascination with his home city grid plan in Barcelona (Figure 1.2); the grid is very suitable for modeling purposes and at the same time it is extremely flexible and generalizable:

“The grid is by far the commonest pattern for planned cities in history. It is universal both geographically and chronologically. No better urban solution recommends itself as a standard scheme for disparate sites, or as a means for equal distribution of land or the easy parceling and selling of real estate.” (Kostof, 1991).

We are aware that our findings cannot be directly exported into other cities. However, we believe that the general trend and the general behavior can be translated into dense urban cities with a grid-like structure.
In this research work, we do not consider the induced demand. Instead, we focus on traffic operations, on the response of networks to the same demand load, on the supply side. This way, networks are more comparable and not dependent on further modeling, which might overestimate induced demand (Hymel et al., 2010) and is dependent on location-specific behavioral factors. Moreover, when certain modifications in networks are carried out, our idea is to study the very short-term impacts, when demand has not been able to rearrange. With that more conservative perspective or upper bound we can also reduce the alarmist thought that traffic will become extremely chaotic (Cairns et al., 2012).

In this research work, we do not consider the induced demand. Instead, we focus on the response of networks to the same demand load—i.e. the supply side. This way, networks are more comparable and not dependent on further modeling. Notice that demand estimation can be rather inaccurate (Hymel et al., 2010), and is dependent on location-specific behavioral factors. Moreover, as we modify networks, we want to study the very short-term impacts of our space removal policies, when demand has not been able to rearrange itself. This, evidently, is a conservative approach, representing approximately an upper bound for the negative effects that could be expected from those policies. Such perspective should help us reduce the alarmist thought that traffic will become extremely chaotic (Cairns et al., 2012) if road space is removed. Any other scenario, considering e.g. a demand reduction, would probably yield better traffic performance and less negative outcomes.
Summarizing, the research objectives of this thesis derive from the main motivations exposed in the previous section and can be summarized with the following questions:

1) Are there more efficient practices to organize traffic in grid networks?

We ambition to study different traffic organization means in the urban grid patterns considered. Put differently, how directions of streets and turns at intersections can be arranged to achieve a higher degree of mobility.

2) Which are the impacts on traffic of removing road space in cities?

We want to evaluate the impacts of removing space in grid networks. Is it possible to minimize traffic impacts while removing space? We also want to mimic these urban planning strategies that target space removal in cities.

We expect to obtain several recommendations and tools for scientists and practitioners that can also be summarized in two main focus areas:

3) What can urban planners learn from this?

We seek to furnish urban planners with recommendations and know-how on traffic operations to support decision making.

4) How can traffic tools be more accessible to urban planners and practitioners?

We aim at developing and adapting traffic tools and models to evaluate the different street patterns, and urban planning policies. We want to encourage further research to bridge the gap between traffic engineering and urban planning. In addition, we also study the real implementation of macroscopic traffic schemes in cities to help practitioners carry out innovative practices.

1.3. Thesis outline

This PhD thesis has 7 chapters that are organized in three conceptual parts. The first part is devoted to the study of different street configurations on grid networks. The second part deals with the effects of space removal on grid networks (see Figure 1.3). Finally the third part studies the implementation of macroscopic traffic schemes in real cities. As Figure 1.3 illustrates, in each chapter, the traffic tools and models employed are different. In Chapters 2 and 4 we are looking at traffic from a static perspective, whereas in Chapters 3, 5, and 6 we are studying the traffic dynamics. Each one of these approaches has advantages and disadvantages, as we explain throughout the dissertation. We believe that the combination of both techniques can make our findings stronger.
The title and content of the chapters is the following:

Part 1. Street configuration:

- **Chapter 2.** One-way vs. two-way street grid networks comparison. In this chapter, we analyze one-way, two-way, and two-way streets with prohibited left turns grid networks. We employ analytical formulations and a static traffic assignment method to study different network sizes and demand patterns.

- **Chapter 3.** Traffic dynamics of one-way vs. two-way street grid networks. Following the analysis of the previous chapter, we compare the different network configurations simulating traffic dynamics. We study the routing behavior and we employ macroscopic network indicators.

Part 2. Space removal:

- **Chapter 4.** Street removal on grid networks. In this chapter, we quantify the effects of removing streets on a grid network following different strategies: central, peripheral, and random. Traffic indicators are obtained with a static traffic assignment method.

- **Chapter 5.** Traffic dynamics of lane removal on grid networks. In this chapter, the space removal is focused on lanes and not streets as previously. We also consider different scenarios and traffic is analyzed dynamically from simulation data. Our focus is on how traffic is distributed and how congestion propagates.

Part 3. Urban traffic monitoring resources:

- **Chapter 6.** Study on the number and location of measurement points for an MFD perimeter control scheme. In this chapter, we study how an MFD can be obtained in a real case study: the city of Zurich. Particularly, we look at incomplete MFDs, i.e. obtained with incomplete information due to fewer monitoring resources. We evaluate their accuracy and we investigate on some strategies to pick the streets to be monitored.
1.4. Outcomes of this research

In this last section of the introduction chapter, we summarize the main research outcomes that this research has produced and that are detailed along the dissertation. We divide these outcomes in tools we developed and adapted, and the main research findings we produced.
1.4.1. Research tools developed

- **Analytical formulations.** We created analytical formulations for finite regular grids with uniform demand that describe the average trip length, the number of turns, and also approximate travel time in uncongested conditions.

- **Static Traffic Assignment (STA).** We implemented a static traffic assignment model based on the Leblanc et al. (1975) algorithm and with special detail on intersections to describe the routing effects due to the increase of travel time. This model is applied to the different network configurations and to the quasi-grid structures originated from link removal.

- **Dynamic Traffic Assignment (DTA).** We simulated the different networks with a microsimulation software (VISSIM 6) and we have adapted the DTA module to reach traffic equilibrium and obtain the effects of traffic dynamics, e.g. spillbacks. We used this model with the different street configuration networks and with the cases of lane removal.

- **Network performance indicators.** We employed different network performance indicators based on travel time, distance traveled, and maneuvers made to compare the different networks and scenarios.

- **Macroscopic indicators.** We employed macroscopic indicators such as the Network Exit Function (NEF) to evaluate the capacity of the networks. In addition, we have created our own indicators to determine when networks become congested and gridlocked; and to assess how many vehicles are able to reach their destination.

- **Accuracy methodology for incomplete MFDs.** We developed a methodology to compare the accuracy of MFDs created with information from only some of the links or streets in the city. This methodology allowed us to discriminate with choice of links could actually generate a better MFD for traffic control purposes.

- **Search algorithm for combinations of links.** We created a search algorithm to find the best combination of links to provide an MFD. To do so, we adapted a Tabu Search metaheuristic algorithm that enumerates the solution domain very efficiently.

1.4.2. Main findings

The first part of this research is devoted to the study of urban grid configurations, particularly two-way street (TW) networks, two-way street with prohibited left turns (TWL) networks, and one-way (OW) networks. The main findings are the following:

- TW street networks provide the lowest distance traveled but are heavily penalized because vehicle capacity at intersections is the lowest. TWL provide the better compromise between distance traveled and capacity at intersections, however they have less route redundancy that causes a more heterogeneous spread of traffic. OW
networks have higher flows because average distance traveled is longer, but they are able to distribute congestion more homogeneously.

- When looking at traffic dynamics, how drivers are routed through the different paths that connect network nodes has a determinant effect. If vehicles are distributed more evenly among different routes, OW have a higher advantage, if vehicles are routed through the shortest path route, TWL have a higher advantage.

- We have elaborated some recommendations for urban planners when choosing one network configuration over another one. Summarizing, TW networks are suitable for networks with low traffic volumes, networks that do not get congested, and also networks with long links. OW networks are adequate for dense cities that present congestion frequently and that might also present heterogeneous demand patterns. Links in OW networks should be shorter to gain advantage over the other ones. Finally, TWL networks are suitable for a various range of scenarios between low demand and congested cities.

The second part of this dissertation focuses on the study of space removal in urban grid networks. This is analyzed in two phases: full link removal, and lane removal. The main findings are:

- The full grid is a robust traffic system that allows a certain degree of link and lane removal without affecting traffic conditions considerably. Evidently, the space removal strategy influences the amount of space that can be removed. When we look at link removal, although central removal is the most restrictive strategy, the system allows certain degree of removal. On the other hand, peripheral removal is the least restrictive strategy. The random removal lies in between the two extreme strategies and could emulate a urban planning policy as combination of central and peripheral strategies.

- The center of the grid plays a key role distributing traffic. The availability of many different routes is the key to achieve more homogeneous flow distributions and hence, not extreme traffic loads in particular intersections. The high connectivity in the center of the grid is what guarantees the redundancy of routes making the network more efficient and resilient.

- Lane removal does not network structure but the traffic distribution of the system. Removal scenarios present always more heterogeneous distributions not only than the initial case but, for some case, also than the final scenario, the full one-lane network. The existence of higher capacity links encourages drivers to take certain routing that ends up being disadvantageous.

- Congestion propagation speed depends on both: the number of one-lane streets that networks have, and where congestion is originated. If congestion is originated in the perimeter, the spread over the network happens faster than if congestion started in the center. Although networks do not present clear congestion propagation patterns, central removal has a tendency to propagate congestion more from the center
whereas peripheral removal has a slight disposition to propagate congestion more from the perimeter.

- Some removal actions provide a better road-space return than others. As a matter of fact, from an urban planning perspective, once some removal steps have been taken (either for perimeter or central), it is better to move to the final one-lane grid because the last removal scenarios worsen traffic and do not provide such space gain as the full one-lane network.

The third part of this dissertation looked at the amount and location of monitoring resources to create accurate MFDs to support traffic control mechanisms in urban areas. The main findings are:

- Most cities do not have that much traffic information a priori, so ideally the selection strategies should be supported only by affordable information. The results show that independently of the strategy used for link selection, a minimum of 25% of the links ensures a fairly accurate MFD.

- We evaluated different link selection strategies (blind strategies) to create an MFD based on common criteria to place fixed monitoring devices: total random selection, distance to the center selection, hierarchy of the street selection, and downstream traffic signal presence selection. The distance selection strategy not only provides the best results but also maintains a low variability. This indicates that cities where monitored links are closer to the center might be better equipped to create accurate MFDs.

- If traffic engineers have information about flows and densities in the city, selecting the links quasi-optimally (i.e. with metaheuristic algorithm to search for link combinations close to the optimal) brings substantial benefits and savings in detection points. We see how with a 5% coverage of links selected quasi-optimally, we can achieve the same accuracy as the best cases with random selection and 40% coverage.

- For the particular case of the city of Zurich, results show that by aggregating the data from the ZuriTraffic links we are able to observe better when the network goes from uncongested to congested states. Using the 23 ZuriTraffic links (5% of the total) to create a pMFD yields an average error in density ratios of 8.7 ppts, significantly below the error from any of the four blind strategies we evaluated.
PART 1

Street configuration

This first part of the dissertation is devoted to an in-depth study of street configurations on urban grid networks. Planners and policy makers are often tasked with deciding between one-way and two-way streets in urban transportation networks, particularly when making the choice to change a street configuration. There are many factors that go into this decision including: safety, urban design, livability, and traffic operations (Hocherman et al., 1990; Stemley, 1998; Lyles, 2000; Walker et al., 2000; Tindale and Hsu, 2005). To date, no consensus has been reached on which street network configuration provides the optimal between these characteristics.

Strictly focusing on the movement of cars, two-way street networks provide a higher accessibility but also offer less capacity at intersections. Instead, one-way street networks provide higher travel speeds due to higher capacity intersections, but force drivers to travel longer routes on average. This trade-off has received little quantitative attention in the literature and is the focus of these two chapters. Walker et al. (2000) qualitatively discuss the extra distance and movements required in a one-way street network, but quantitative metrics and details are not provided. Meng and Thu (2004) analyze operations after the conversion of a two-way grid into a one-way grid with a CORSIM simulation. This study concludes that one-way streets perform better than two-way streets mainly because of the higher average travel speeds observed. However, higher average speeds do not necessarily mean that people get to their destinations faster if they have to travel longer distances. To overcome this gap, Gayah and Daganzo (2012) examine the maximum rate at which trips could be completed in various grid networks, including one-way, two-way, and two-way with prohibited left turns. Using this new metric, this study is the first to quantify the potential trade-offs between additional distance traveled and capacity available at intersections by combining them into a single metric: trip-serving capacity. The study shows that one-way street networks outperform two-way street networks when trips are long, and two-way street networks with prohibited left turns always outperform one-way street networks. However, the work uses abstract networks of infinite size and optimistically assumes that congestion is evenly distributed across the network. Additionally, the metric
proposed only considers capacity flows and cannot be used to provide insights into operations during either free-flow or congested conditions. Boyles et al. (2014) employ simulation techniques to study abstract grid networks under different levels of congestion. However, the authors focus primarily on developing a STA model that is able to represent the low-conflict intersections maneuvers of Eichler et al. (2013) innovative networks, and do not emphasize the differences between the different network configurations.

Chapter 2 looks at that problem employing, first, analytical formulations, and later, a STA method. Analytical formulations consider a fixed traffic assignment but they provide simple and useful insights. The STA instead, although static, considers the effects of higher traffic flows and represents how drivers reroute to have lower travel times. The traffic performance indicators employed are basically, average trip length and average travel time. Chapter 3 completes this analysis looking at dynamics of congestion with a microsimulation software integrated in a DTA framework. In that case, we are able to capture traffic queues and spillbacks. There, traffic performance is measured with macroscopic indicators that estimate the network capacity and when the system becomes congested.

The outcomes of this research will furnish urban planners and policy makers with a better understanding of the level of mobility provided by each of these types of urban street network layouts.
Chapter 2

One-way vs. two-way street grid networks comparison

This chapter is based on the following research papers:


• Ortigosa, J., V.V. Gayah, and M. Menendez. Street configuration evaluation in urban grid patterns. Working paper.

2.1. Introduction

Debates between two-way and one-way street network configurations started when cars were first introduced into urban environments and are still ongoing. In this paper, we provide additional insights on this issue by analyzing aggregate traffic behavior on three different street configurations: one-way streets, two-way streets, and a compromise solution between them: two-way streets with prohibited left turns. This study is focused on abstract finite grids operating with both, low and high levels of traffic demand.

We employ analytical formulations to compare the three network types under low demand scenarios. These formulations rely on a fixed traffic assignment scheme in which drivers are assumed to minimize their overall travel distance and number of turning maneuvers, and are used to reveal general insights. Previous authors have developed analytical formulations for the analysis of urban grid networks (e.g. Gayah and Daganzo, 2012; Doig et al., 2013). Our analytical formulations are created for finite urban grids and provide exact expressions for distance traveled and turns made, as well as they approach the average travel time. For high demand volumes, vehicles do not necessarily follow the shortest path distance as they seek to minimize their travel time. We develop a STA model for this purpose. The STA is an iterative algorithm that
approaches the system to the user equilibrium (Wardrop, 1952). It is used to examine the behavior of networks under more realistic situations, especially under higher demand levels when the analytical formulations may no longer be valid.

In general, the study reveals that TW networks provide the shortest distance traveled but they are severely penalized by restrictions in capacity at the intersections due to more complicated phasing schemes, which leads to higher average travel times. OW networks offer the longest travel distances but have intersections with very high capacity. They perform better in congested scenarios. Finally, TWL provide the best compromise between distance traveled and intersection capacity, which makes them suitable for many different scenarios with low and medium congestion. Their main weakness is the lack of route redundancy that could lead to very large travel times when the network is very congested.

The rest of the chapter is organized as follows. First, we introduce analytical formulations to describe operations on uncongested networks where trip decisions do not depend on traffic volumes. Second, we present the STA model used to analyze grids in more congested scenarios where decisions depend on traffic volumes. Finally, we analyze the results and present the main conclusions of this research.

2.2. Analytical formulations for low traffic volumes

We consider here an abstract, square grid network that has $n$ intersection$^3$ nodes per side ($n^2$ nodes in total). The origins and destinations of trips (i.e. demand nodes) are assumed to be located in the midpoint of every link to emulate the generation and attraction of trips along the street. The demand is initially assumed to be uniform, which means that every demand node exchanges the same amount of trips, $T$, with all the other demand nodes. This assumption is later relaxed in Section 2.3.3 with the aid of the STA algorithm. Figure 2.1 shows the three kinds of network configurations we analyze: two-way streets with all the turns allowed, two-way streets with prohibited left turns, and one-way streets. For simplicity, we assume TW networks have a left-turn pocket to facilitate left-turn vehicle storage.

When congestion is not present in a network, i.e. traffic volumes on links and at intersections are low, vehicle routing on the network is likely to be independent of the number of vehicles traveling. We assume that drivers route themselves to minimize personal travel time, minimizing the total distance traveled and, according to previous

---

$^3$ Note that also nodes on the perimeter, including the corners, are considered as intersection nodes.
studies (this type of routing decision has been verified using empirical data—Mannering et al., 1994; Abdel-Aty and Abdalla, 2004—and used to study the aggregate behavior of under-saturated networks—Doig et al., 2013), also minimizing the number of turning maneuvers performed.

Our analytical formulations seek to relate the total distance traveled and total number of turning maneuvers made when vehicles operate in this way. The analytical formulations are obtained by aggregating the generalized expressions of the different metrics for each network type considered. For example, a network with \( n \) intersection nodes per side has a total of \( 2n(n - 1) \) demand nodes, and hence, \( 2n(n - 1)(2n(n - 1) - 1) \) OD pairs.

### 2.2.1. Two-way streets

The total distance traveled (in units of block length), \( TD_{TW} \), and total number of turns (both left and right turns) required to complete all trips, \( TR_{TW} \), were calculated for networks of various sizes considering the geometries that exist in the finite grid with bidirectional streets. The result is a polynomial expression that relates these metrics to the size of the network, \( n \). Figure 2.2a shows the traveled distances from a given node \( i \) to the rest of the nodes in a TW network. These distances are divided into their two components, X and Y, i.e. the part of the trip done along the X dimension, and the part done along the Y dimension.
Figure 2.2. a) Distance in blocks from node $i$ to the rest of the nodes divided into the two components, X and Y. b) Turns from node $i$ to the rest of the nodes.

The total distance traveled is decomposed into different terms depending on where the nodes are located: between two nodes in vertical streets, $TD_{TW}^{VS}$; between two nodes in horizontal streets, $TD_{TW}^{HS}$; between a node in a horizontal and a node in a vertical street, $TD_{TW}^{VHS}$; and between a node in a vertical and a node in a horizontal street, $TD_{TW}^{VHS}$. Taking advantage of the symmetry of the network, this expression can be simplified into two summands:

$$TD_{TW} = TD_{TW}^{VS} + TD_{TW}^{HS} + TD_{TW}^{VHS} + TD_{TW}^{VHS} = 2TD_{TW}^{VS} + 2TD_{TW}^{VHS}. \quad (2.1)$$

Let us consider the distance between nodes in vertical streets. In (2.2), the first summand corresponds to the total distance in the X-axis between nodes in vertical streets, whereas the second summand corresponds to the total distance in the Y-axis between nodes in vertical streets. Note that these are only valid for the nodes with origins and destinations located in different streets. Hence, the third summand accounts for the trips between nodes in the same street.

$$TD_{TW}^{VS} = 2T(n-1)^2 \sum_{r}^{n-1} \sum_{i}^{r} i + Tn(n-1) \left( n-1 + 2 \sum_{r}^{n-2} \sum_{i}^{r} i \right) + 2Tn \sum_{r}^{n-2} \sum_{i}^{r} i$$

$$= \frac{1}{3} Tn(n-1)^3(n+1) + \frac{1}{3} Tn(n-1)^2(n^2 - 2n + 3) + \frac{1}{3} Tn^2(n-1)(n-2)$$

$$= \frac{1}{3} Tn(n-1)(2n^3 - 3n^2 + 2n - 2). \quad (2.2)$$

Let us now develop the expression for the total distance between nodes located in vertical streets and nodes located in horizontal streets. We also divide it into the
distance traveled along the X-axis, and the distance traveled along the Y-axis. In this case, because of the symmetry, both summands are equal.

\[
TD_{TW}^{VHS} = Tn(n - 1) \left( 2 \sum_r \sum_l i - \frac{n(n - 1)}{2} \right) + Tn(n - 1) \left( 2 \sum_r \sum_l i - \frac{n(n - 1)}{2} \right)
\]

\[
= \frac{1}{3} Tn^2(n - 1)^2(2n - 1).
\]

(2.3)

Finally, substituting (2.2) and (2.3) into (2.1), we obtain the expression of the total distance traveled.

\[
TD_{TW} = 2TD_{TW}^{VS} + 2TD_{TW}^{VHS}
\]

\[
= \frac{2}{3} Tn(n - 1)(2n^3 - 3n^2 + 2n - 2) + \frac{2}{3} Tn^2(n - 1)^2(2n - 1)
\]

\[
= \frac{2}{3} Tn(n - 1)(4n^3 - 6n^2 + 3n - 2).
\]

(2.4)

We obtain the total turns made in a similar way. This case is much simpler because turns do not depend on the distance to the origin node. As Figure 2.2b shows, between nodes in vertical streets in TW networks there will be always two turns needed, whereas if one node is in a horizontal and one is in a vertical street there will be only one turn needed. We consider that the block face of the destination is not relevant otherwise the routing might change. Following the same logic as before:

\[
TR_{TW} = TR_{TW}^{VS} + TR_{TW}^{HS} + TR_{TW}^{VHS} + TR_{TW}^{HVS} = 2TR_{TW}^{VS} + 2TR_{TW}^{VHS}.
\]

(2.5)

Developing each summand:

\[
TR_{TW}^{VS} = Tn(n - 1)(n - 1)(n - 1) = n(n - 1)^3.
\]

(2.6)

and:

\[
TR_{TW}^{VHS} = \frac{1}{2} Tn(n - 1)(n - 1)n = \frac{1}{2} n^2(n - 1)^2.
\]

(2.7)
Thus, the total number of turns is:

\[ TR_{TW} = 2Tn(n - 1)^3 + Tn^2(n - 1)^2 = 2Tn(n - 1)^2(3n - 2). \] (2.8)

The total number of trips made in the network, \( N \), was also determined by calculating the number of trips exchanged between each pair of demand nodes. That is, by taking the product of the total number of OD pairs calculated before, and the number of trips exchanged between each node pair, \( T \).

\[ N = 2Tn(n - 1)(2n^2 - 2n - 1). \] (2.9)

The ratio of (2.4) to (2.9) yields the average distance traveled (i.e. average trip length), \( AD_{TW} \), as a function of network size, \( n \):

\[ AD_{TW} = \frac{1}{3} \left( \frac{4n^3 - 6n^2 + 3n - 2}{2n^2 - 2n - 1} \right). \] (2.10)

Similarly, the ratio of (2.8) to (2.9) yields the average number of turns per trip, \( AT_{TW} \), as a function of network size, \( n \):

\[ AT_{TW} = \frac{(n - 1)(3n - 2)}{(2n^2 - 2n - 1)}. \] (2.11)

### 2.2.2. Two-way streets with prohibited left turns

The same procedure as in the previous section was followed for TWL networks. The total distance traveled, \( TD_{TWL} \), and the total number of turns made, \( TR_{TWL} \), in these two-way networks with prohibited left turns are:

\[ TD_{TWL} = \frac{8}{3} Tn \left( n^4 - n^3 - \frac{9}{4} n^2 + \frac{13}{4} n - 1 \right), \text{ and} \] (2.12)

\[ TR_{TWL} = 4Tn(n - 1)^2(2n - 1). \] (2.13)

Since the total number of trips made, \( N \), remains the same, the ratio of (2.12) to (2.9) yields the average trip length, \( AD_{TWL} \):

\[ AD_{TWL} = \frac{4}{3} \left( \frac{n^4 - n^3 - \frac{9}{2} n^2 + 3n - 1}{(n - 1)(2n^2 - 2n - 1)} \right). \] (2.14)
Similarly, the ratio of (2.13) to (2.9) yields the average number of turns per trip, \( AT_{TWL} \):

\[
AT_{TWL} = \frac{2(n - 1)(2n - 1)}{(2n^2 - 2n - 1)}. \tag{2.15}
\]

### 2.2.3. One-way streets

There are different possible configurations of OW grids depending on the size of the network and street direction pattern. However, we focus only on those in which trips between any pair of demand nodes can be feasibly made. This leaves only square grids with an even number of intersection nodes per side, and those in which perimeter streets form a traversable loop. In this case, developing the expressions as in the previous case was too complex. We employed a simple Matlab program with a shortest path assignment to calculate the total distance traveled or the total turns made. With the aggregated values for several \( n \) cases we obtain the exact polynomial expression that follows that curve—the sum of square errors between both expressions equals zero. We also employed this method to validate the expressions developed manually. Under these conditions, the total distance traveled, \( TD_{OW} \), and the total number of turns made, \( TR_{OW} \), in these one-way networks are:

\[
TD_{OW} = \frac{2}{3} T \left( 4n^5 + 2n^4 - \frac{9}{2} n^3 - 29n^2 + 2n + 24 \right). \tag{2.16}
\]

\[
TR_{OW} = T(n - 1)(10n^3 - 6n^2 - 12n - 8). \tag{2.17}
\]

Likewise, since the total number of trips made, \( N \), remains the same; the ratio of (2.16) to (2.9), and the ratio of (2.17) to (2.9), provide the average trip length, \( AD_{OW} \), and the average number of turns per trip, \( AT_{OW} \), respectively:

\[
AD_{OW} = \frac{1}{3} \left( 4n^5 + 2n^4 - \frac{9}{2} n^3 - 29n^2 + 2n + 24 \right), \text{and} \tag{2.18}
\]

\[
AT_{OW} = \frac{(5n^3 - 3n^2 - 6n - 4)}{n(2n^2 - 2n - 1)}. \tag{2.19}
\]
2.2.4. Average trip length and average turns per trip

The analytical formulations reveal some interesting observations about these three network types. For example, expressions (2.10), (2.14), and (2.18) are all asymptotic to three unique parallel lines which serve as a bound on the average trip length (in blocks):

\[ AD_{TW} > \frac{2n - 1}{3}, \]  \hspace{1cm} (2.20)

\[ AD_{TWL} < \frac{2n + 2}{3}, \text{ and} \]  \hspace{1cm} (2.21)

\[ AD_{OW} > \frac{2n + 5}{3}. \]  \hspace{1cm} (2.22)

Notice that these lines ((2.20), (2.21), (2.22)) provide lower bounds on average travel distance for TW and OW networks, and an upper bound for TWL networks. As Figure 2.3 shows, these bounds are fairly tight and can be used as a good approximation for the average trip length, particularly when the network is large. These equations also reveal that trips in a TWL network are at most one block longer than in a TW network, and trips in an OW network are approximately two blocks longer than trips in a TW network. The equations also reveal that the relative differences in average travel distance become smaller as \( n \) grows. Thus, for larger networks, intersection capacities can be expected to play a larger role in overall traffic operations, while travel distance might play more of a role when the network is smaller. These values also correspond well to those derived for infinitely large networks in Gayah and Daganzo (2012).
Figure 2.3. a) Trends for average trip length and, b) Trends for average number of turns per trip for network sizes \(n=4-16\).\(^4\)

Similarly, as Figure 2.3b shows, the average number of turns per trip tends to 1.5 in the case of TW networks, 2 for the case of TWL networks, and 2.5 for the case of OW networks. The latter is a lower bound while the first two are upper bounds. Thus, on a one-way network, at least 1 extra turn per trip is required in comparison to a two-way network, and at least 0.5 extra turns are required in comparison to a two-way network with prohibited left turns. These values all confirm expectations that OW networks are more restrictive than TWL and TW networks, and that TWL networks are more restrictive than TW networks.

2.2.5. Travel time on networks

The analytical formulas derived in this section do not provide direct information on travel times. However, they can be used to obtain an approximate expression of travel times for scenarios with low traffic volumes. To do so, we assume that: 1) cars travel at free flow speed on all links; and, 2) intersections are under-saturated such that delays during a single cycle can be calculated using a deterministic queuing diagram.

The travel time on any one link is then \(l/v_f\), where \(l\) is the link length and \(v_f\) is the free flow speed. Assuming a uniform rate of vehicle arrivals to all intersections, \(\lambda\), red time, \(R\), cycle time, \(C\), and saturation flow, \(\mu\), the average delay at any single intersection is:

\[\frac{\lambda R}{\mu}\]

\(^4\) As stated earlier, \(n\) is the number of intersection nodes per grid side.
When the demand is very low, the flow ratio, $\frac{\lambda}{\mu}$, is very small and can be neglected. Since the average trip length expressions are given in units of block lengths traveled, they also correspond to the number of intersections a car traverses on its trip. The delay experienced at the intersection depends on the movement being performed. Through and right turn movements (also left turn movements in OW networks depending on the intersection) share the same capacity so they experience the same delay. However, left-turning vehicles on TW networks experience longer delays because there is generally less green time available for this movement. Since half of all turns on TW networks can be expected to be left turns, the average travel time for TW networks, $ATT_{TW}$, is:

$$ATT_{TW} = (AD_{TW} - \frac{1}{2}AT_{TRW}) R_{tr}^2 + \frac{1}{2}AT_{TW} R_{le}^2 + AD_{TW} \frac{l}{v_f},$$

(2.24)

where $R_{tr}$ is the red phase for through and right turn movements and $R_{le}$ the red phase for left turn movements. The first term corresponds to the average intersection delay experienced by through and right turn movements; the second term corresponds to the average intersection delay experienced by left turn movements; and, the third term corresponds to the travel time on the links. For TWL and OW, the expressions are simpler because there is only one red phase, $R$, that is equal in both cases:

$$ATT_{TWL} = AD_{TWL} \left( \frac{R^2}{2C} + \frac{l}{v_f} \right), \text{ and}$$

(2.25)

$$ATT_{OW} = AD_{OW} \left( \frac{R^2}{2C} + \frac{l}{v_f} \right).$$

(2.26)

The street length (i.e. the block size) is also an important factor to consider. Among cities with urban grid structures, street lengths can vary widely. Some measures of U.S. grids (Nairn, 2010) show this variability: Carson City, 230 ft (~70m); Portland, 260 ft (~79 m); Houston, 330 ft (~101 m); Austin, 350 ft (~107 m); Sacramento, 410 ft (~125 m); Columbus, 480 (~146 m); or Salt Lake City, 780 ft (~238 m). In this work, two lengths are examined (100 m and 200 m), which are considered representative of real block sizes. Figure 2.4 depicts this average travel time for these two link lengths, 100 m and 200 m, and different network sizes.
TW and TWL networks perform similarly for small size networks. However, as soon as the networks become larger (i.e. networks have more nodes per side), the number of trips increases and TW networks lose competitiveness in comparison to TWL networks. OW networks slightly improve their travel times in relation to the other network types when network size increases; however, they still always appear to perform the worst, especially when links are longer. Figure 2.4 also shows how TW networks gain an advantage when links are longer because the time spent overcoming distance becomes a larger portion of the total travel time compared to the time spent delayed at intersections.

![Figure 2.4: Average travel time according to (2.24), (2.25), and (2.26) for: a) 100 m link length, and b) 200 m link length, and network sizes n=4-16. For TW, C=60 s, Gtr=20 s, and Gle=5 s; and for TWL and OW, C=60 s, and R=27.5 s.](image)

2.3. Static Traffic Assignment for high traffic volumes

In the previous section, we assumed vehicles use a fixed assignment method that minimizes total travel distance and number of turning maneuvers, and consequently, delay formulations do not depend on the traffic volume. However, these assumptions lose their applicability as trip demands grow and congestion occurs in the network. In these scenarios, drivers will still seek the route that provides them with the shortest travel time, but this might no longer be the one that minimizes the number of turning maneuvers performed and total travel distance. When no driver can improve his/her situation, the system has reached the user equilibrium (Wardrop, 1952). In this section we employ an STA method to compare again how different grid networks (TW, TWL, and OW) behave under different demand loads that might create inhomogeneous congestion patterns (Doig et al., 2013). STA has two main advantages over the analytical formulations: a) it describes the behavior of trips when the network is more
congested; and, b) it captures the travel time as a function of the traffic load in the different links and intersections.

In the Traffic Assignment Problem (Beckmann et al., 1956) a travel time function is defined for every link in the network. If this function only depends on the volume of that link and it is continuous, differentiable and monotonic, then the problem has a unique equilibrium solution (the problem is a convex optimization). This was first solved by Nguyen et al. (1973) and LeBlanc et al. (1975) employing the Frank and Wolfe algorithm (Frank and Wolfe, 1956). This algorithm presents many advantages (e.g. small storage requirements), but also some drawbacks, e.g. slow convergence, tailing. Nowadays, newer and more efficient algorithms have been developed (e.g. Bar-Gera, 2002; Dial, 2006). Nevertheless, given the macroscopic nature of our analysis, and the number of networks that we are analyzing, we have developed a STA based on LeBlanc et al. (1975).

The most common delay formulation is the one from the Bureau of Public Roads (1964) function (BPR) which we employ to represent link delays. Modeling intersections, however, is more complex. To overcome the asymmetry of the problem, intersections are modeled based on (Koutsopoulos and Habbal, 1994) employing 12 dummy links. For each one of them, representing an intersection movement (e.g. left turn), a 3 piece-wise differentiable delay function is created and adjusted to the Highway Capacity Manual (2010) formulation (HCM-2010). Obviously, the STA employed here is less accurate describing traffic operations than DTA programs based on microscopic simulations, which are able to represent spillbacks and other effects in congested scenarios. However, the latter method is also computationally more complex, more unstable reaching convergence, and has more factors influencing the results (e.g. random seed, temporal distributions, intersection design, software employed). Hence, we believe that the next logical step after the analytical formulations is an STA algorithm that represents in a simple way how cars reroute themselves due to the volume of traffic in streets and intersections. This simplicity is the key to understand the properties of the different networks when accommodating traffic.

All the streets are modeled with the same characteristics: two lanes (one in each direction in the case of TW and TWL; the two in the same direction in the case of OW) with a saturation flow of 1,800 veh/h-lane. Every intersection node is assumed to be signalized. A 60 s cycle length is assumed with 2.5 s of lost time per phase change, a total lost time of 5 s per cycle in TWL and OW networks, and 10 s per cycle in TW networks. At intersections, on TW networks, four-phase signals are used: two dedicated left-turn phases (5 s each) and two through/right-turn phases (20 s each). A dedicated left-turn pocket is also assumed to exist at these locations for TW networks. At
intersections on OW and TWL networks, two-phase signals are used: one for each perpendicular direction of movement (27.5 s each).

We have detected few cases, in TW and TWL networks for low to medium demands, where the STA was unstable (in certain iterations, the algorithm did not find a feasible step length). This instability occurred especially for TWL networks, when differences in cost matrixes between iterations were very small. In cases where these instabilities appeared, we established the all-or-nothing solution as current flow matrix (step length=1). That allowed continuing the iterative process. Notice, however, that these instabilities only happen in some particular cases and we believe that their effects are negligible given the macroscopic nature of our results.

In our experiments we analyze networks of different sizes (number of intersection nodes per side), and we apply different levels of demand, $T=0.01$, $T=0.10$, and $T=0.25$ trips between each two demand nodes (see Table 2.1). The total number of trips is not only function of the demand level but also of the size of the network, e.g. a 16x16 network will have 418 times more trips than a 4x4 network. Finally, we also analyze different block sizes considering 100 m and 200 m links.

<table>
<thead>
<tr>
<th>Nodes per side</th>
<th>Total number of nodes</th>
<th>Total number of links</th>
<th>Area (km$^2$)</th>
<th>Range of total number of trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>16</td>
<td>24</td>
<td>0.09–0.36</td>
<td>5.5–55.2–138.0</td>
</tr>
<tr>
<td>6</td>
<td>36</td>
<td>60</td>
<td>0.25–1.00</td>
<td>35.4–354–885.0</td>
</tr>
<tr>
<td>8</td>
<td>64</td>
<td>112</td>
<td>0.49–1.96</td>
<td>124.3–1243.2–3108.0</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>180</td>
<td>0.81–3.24</td>
<td>322.2–3222–8055.0</td>
</tr>
<tr>
<td>12</td>
<td>144</td>
<td>264</td>
<td>1.21–4.84</td>
<td>694.3–6943.2–17358.0</td>
</tr>
<tr>
<td>14</td>
<td>196</td>
<td>364</td>
<td>1.69–6.76</td>
<td>1321.3–13213.2–33033.0</td>
</tr>
<tr>
<td>16</td>
<td>256</td>
<td>480</td>
<td>2.25–9.00</td>
<td>2299.2–22992–57480.0</td>
</tr>
</tbody>
</table>

2.3.1. Average trip length and average turns per trip

Following a similar structure as in Section 2.2, the average trip lengths are now calculated for three different demand levels (See Figures 2.5a–c). These results are highly consistent with the analytical formulations (Figure 2.3a), and suggest that the presence of congestion does not considerably impact average distance traveled. Only for the demand level of $T=0.25$ and a network of size $n=16$ we do see how the average trip length of TWL increases, however we do not include those results in Figure 2.5. According to the simulation results, in this case more than 40% of the intersection approaches have Volume to Capacity ratios (V/C) larger than 1. This explains the
significant increase in traveled distances and times, and makes these results somewhat unreliable.

Figure 2.5. Average trip lengths and turns with the simulation model for demand levels of $T=0.01$, 0.1, and 0.25; network sizes $n=4-16$, and links of 100 m.

The average number of turning maneuvers per trip appears to be fairly consistent with the analytically derived counterparts for TW and TWL networks. However, OW networks have much higher turning rates than predicted by the analytical formulations. One reason for this is that there is no difference in delay between turning or through maneuvers since these movements share the same green phase. Therefore, if there are
several routes with the same total distance and different number of turns, the vehicles will choose the one that is less congested, even if it needs to perform more turning maneuvers. This should serve to spread congestion more uniformly across the network in the simulation (which will be discussed in Section 2.4.2). Note, however, that in reality vehicles might choose straighter routes because there might be some additional benefits to doing so, like signal coordination; however, our model does not account for that.

2.3.2. Travel time on networks

In this section, we examine travel times using the simulation model, which incorporates the impacts of congestion and demand-related delays at intersections. Figure 2.6 illustrates the travel times for different sizes, different levels of demand, and link lengths. When the demand level is really low and the links are 100 m long (Figure 2.6a), we observe high consistency with the results of the analytical formulations shown in Figure 2.4a. In this scenario, TWL networks provide the shortest travel times when networks are small (i.e. \( n \) is low); this is followed very closely by TW networks and finally OW networks. As the size of the networks increases, the difference between TWL and TW increases because the latter type is penalized by the additional signal phases required at the growing number of intersections. For every intersection, TWL and OW provide additional total green time due to less lost time (e.g. in this simulation, 5 extra s of green time), which results in lower average delays. Overall, this diminishes the performance of TW in comparison to the other network types when demand and network size increase.

When traffic demand increases and networks become more congested, we see that travel times increase in an exponential manner with network size. Also, we see how TWL and OW get closer until OW becomes more competitive than TWL (e.g. when \( n=14 \) and demands exceed 0.1 trips per node pair). Notice that these behaviors are not represented by the analytical formulations since the latter are due to the effects of congestion, specifically: the path redundancy of the networks, the heterogeneity of congestion, and the delay formulation employed. We examine in detail all these effects in Sections 2.4.2 and 2.4.3. As for the case of Figure 2.5, results for \( n=16 \) are not presented in Figures 2.6c and 2.6f because the system is too congested and results might be unreliable.
Figure 2.6. Average travel times with the simulation model for demand levels of $T=0.01, 0.1$, and $0.25$, network sizes $n=4-16$ and links of 100 m and 200 m.

TW networks offer the shortest routes with the least number of turns, but are penalized by the reduced-capacity at the intersections. However, as links become longer the ratio between travel time on the links and travel time at the intersections increases, and this allows the TW networks to become more competitive. This case is depicted on Figures 2.6d–f, which provide simulation results for links that are 200 m long. In this new scenario, the general trends repeat but we observe that the curve for TW travel times becomes lower (and the curve for the OW travel times becomes higher) than the
case where links are 100 m long. For example, when networks are $n=4$, TW networks are better for all the demand levels. Networks with higher average trip lengths like OW are more penalized because travel time on links is higher.

Increasing the green time for left turns makes TW networks more inefficient. The green time considered (5 s) is a realistic assumption. Nevertheless, for the sake of completeness, we have also simulated the networks with a shorter green left time, 2.5 s, that although unrealistic might improve the overall performance of the system. However, even with this shorter value that could be considered as the upper bound, TW networks are still worse than TWL and OW networks.

Let us now look at how congested the different networks are. Obviously, the bottlenecks of the system are intersection approaches as they have the lowest capacity due to the traffic signals. Figure 2.7 shows the V/C ratios at intersection approaches for the demand levels analyzed. The V/C ratio at intersections is a great indicator of congestion because provides a fairly good idea of how much the networks are loaded. On average, OW networks yield higher V/C ratios, which means that these networks become more congested. That is explained because vehicles need to travel longer and hence, flows are also higher. Notice, however, that average V/C ratios, alone, fail to show the variability across intersections. To address that, Figure 6 also shows the standard deviation of the V/C ratios. We can see then that even though the average V/C ratio is lower for TWL compared to OW, the variance is much higher, i.e. congestion levels are more heterogeneous. Since delay formulations are not linear, a higher heterogeneity in V/C ratios will cause a higher delay.
Another reason because OW become better in congested scenarios is the type of delay functions we employ at intersections. The HCM-2010 formulations are dependent on the total saturation flow of each approach (in addition to the V/C ratio, green time and cycle length). Because both lanes are used to serve a single direction in the OW network, the saturation flow at the intersection is twice that of the TWL network (3,600 veh/h as opposed to 1,800 veh/h). Hence, for any given V/C ratio, the HCM-2010
methodology provides a much higher delay for the approach with a lower saturation flow (TW and TWL) than for the approach with a higher saturation flow (OW).

### 2.4. Different demand patterns

Thus far, we have focused on uniform demand patterns because it is the simplest and more generalized case, and because one could argue that demand in very dense city centers is more or less uniform. However, to understand if the previous results are generalizable to other scenarios, here we consider two additional demand patterns:

a) Higher trip generation and attraction rates in the perimeter nodes. The nodes of the perimeter generate and attract 3 times more trips than the other nodes.

b) A single quadrant (out of four) in the grid that generates and attracts 2 times more trips than the other.

Interestingly, the average trip distance and number of turns calculated analytically are very close to those presented in Section 2.2 under these different demand patterns. For example, as Figure 2.8b shows, the difference in average trip lengths across networks is very close to the one presented in Figure 2.3.
a) Alternative demand patterns

*Pattern a*

*Pattern b*

b) Trip length (analytical)

c) Travel time (STA)

d) V/C ratios (STA)

Figure 2.8. Average trip length (analytical), and average travel times and V/C ratios (STA) for the two non-uniform demand patterns considered for $l=100m$, network sizes $n=4-14$, and demand level $T=0.25$. 
The different behavior from different demand patterns arises when the static simulation is employed. There, trips are allocated to different routes as certain links and intersections increase their volume to capacity ratios. Figure 2.8c and 2.8d show the simulated average travel times and V/C ratios for the two non-uniform demand patterns considered, respectively. These plots show a similar behavior to the previous plots with uniform demand (Figures 2.6 and 2.7). However, since these demand patterns concentrate more trips in certain regions, congestion arrives earlier. In other words, these demands create more inhomogeneous congestion patterns so average travel times start increasing fast even for lower demands and smaller networks. Hence, TWL networks lose their advantage faster. When \( n = 12 \), OW networks appear to handle congestion better. Based on these results, one would expect that the more concentrated the demand becomes, the better OW networks perform. This is reasonable as OW networks can handle better directional traffic flows as compared to TW and TWL networks.

2.5. Causes of general trends in network behavior

The previous results suggest that TWL networks provide the lowest average travel times on a network for most of the cases analyzed. TW networks only provide a better option when links are long and networks are small. OW networks are the best choice when networks are large and heavily loaded or when demand is concentrated. Here, we look more closely at the differences in the analytical and simulation results, and some of the reasons for these general findings—specifically, the nature of congestion and route redundancy under congestion.

2.5.1. Comparison between analytical approach and simulation

As we discussed, analytical formulations are suitable for uncongested scenarios whereas simulations describe better congested situations. Note, however, that the analytical formulations are very accurate for some metrics. Figure 2.3a (analytical) and Figures 2.5a-c (simulation) show that the average trip length is certainly the metric that is less influenced by congestion; hence, analytical formulations are suitable to accurately describe distance traveled for many different demand and network size scenarios. Moreover, average turns per trip are quite consistent for TW and TWL networks, except for very congested cases. OW simulations, however, show more turns than analytical formulations because in the case of equal length routes, turning or going through both cause the same delay.

Figures 2.4 and 2.6 present travel times for analytical formulations and simulations, respectively. This is the most sensitive indicator of the congestion level. When demands
are very low, the simulation results (Figure 2.6a) are very similar to the results obtained from the analytical formulations (Figure 2.4a). However, as trip demand increases, travel time magnitudes increase as well, and the fit between simulations and the analytical formulations deteriorates. Figure 2.4 was calculated assuming that the flow ratios on (2.24), (2.25), and (2.26) were 0 considering the system very uncongested. It is then reasonable that results differ from the simulations when congestion arises as flow ratios and V/C ratios differ significantly from 0. We have calculated analytically the average flow per intersection approach, \( \lambda \), for each network type, and applied them to expressions (2.24), (2.25) and (2.26). If we do so, analytical formulations provide rather reliable results on average. Table 2.2 shows the errors between the analytical approach and the simulation for TW networks. Notice that only in cases with high demands and large network sizes we observe errors above 10% (bold values in Table 2.2). Obviously, these results cannot be used for detailed traffic assignment purposes or to compare two network configurations because the errors are computed in absolute values. Nevertheless, they show once again, that analytical formulations might be rather accurate for an aggregate-level analysis, especially when including a reasonable flow ratio for high traffic demands. Otherwise, analytical formulations are still very useful, when assuming uncongested intersections as they can save a lot of computational efforts for more general and macroscopic analyzes. Last, but not least, the link length also plays a role in the accuracy. The longer the links the better the accuracy as the relative weight of the distance in the total travel time is higher. Thus, errors in estimating intersection delays become less important.

Table 2.2. Absolute values of relative (total) error between analytical formulations and simulations for average travel time on TW networks in units of percent difference (in s).

<table>
<thead>
<tr>
<th>Nodes per side</th>
<th>Demand level</th>
<th>0.01</th>
<th>0.05</th>
<th>0.1</th>
<th>0.15</th>
<th>0.2</th>
<th>0.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>9.0% (5 s)</td>
<td>8.7% (5 s)</td>
<td>8.4% (5 s)</td>
<td>8.2% (5 s)</td>
<td>7.9% (4 s)</td>
<td>7.6% (4 s)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7.1% (6 s)</td>
<td>6.5% (5 s)</td>
<td>5.7% (5 s)</td>
<td>4.9% (4 s)</td>
<td>4.1% (3 s)</td>
<td>3.4% (3 s)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>6.4% (7 s)</td>
<td>5.0% (5 s)</td>
<td>3.4% (4 s)</td>
<td>1.8% (2 s)</td>
<td>0.2% (&lt;1 s)</td>
<td>1.3% (1 s)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>5.9% (8 s)</td>
<td>3.4% (4 s)</td>
<td>0.4% (1 s)</td>
<td>2.5% (3 s)</td>
<td>5.3% (7 s)</td>
<td>8.0% (11 s)</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>5.3% (8 s)</td>
<td>1.2% (2 s)</td>
<td>3.7% (6 s)</td>
<td>8.4% (14 s)</td>
<td>12.8% (21 s)</td>
<td>17.0% (29 s)</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>4.6% (9 s)</td>
<td>1.7% (3 s)</td>
<td>9.1% (17 s)</td>
<td>15.8% (31 s)</td>
<td>22.0% (45 s)</td>
<td>37.2% (78 s)</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>3.8% (8 s)</td>
<td>5.4% (12 s)</td>
<td>15.6% (35 s)</td>
<td>24.8% (58 s)</td>
<td>147.4% (364 s)</td>
<td>794.8% (2081 s)</td>
<td></td>
</tr>
</tbody>
</table>
2.5.2. Heterogeneity of congestion and route redundancy

In transportation networks, path redundancy is desired because drivers have reasonable alternatives when some routes are blocked. Moreover, the presence of redundant routes may imply that the network is reliable and resilient to localized disruptions, both desirable qualities for transportation networks. Here, we study the redundancy of routing in these networks.

Given any two points on a TW, TWL, or an OW network, there is only one route that simultaneously minimizes both distance and the number of turns; thus, no flexibility in route choice exists (Figure 2.9a). If the condition of minimizing the number of turns is relaxed and only the distance minimization is considered, the redundancy of each network type changes. In a TW network, the number of shortest paths between two intersections nodes is \( \binom{t_l}{d_s} \), where \( t_l \) is the total length of the trip in blocks, and \( d_s \) is the minimum orthogonal component in complete blocks. As an example, in Figure 2.9a, there will be 6 different minimum distance routes between demand nodes \( i \) and \( j \) (or between intersections nodes \( n_i \) and \( n_j \)) as \( t_l = 4 \) and \( d_s = 2 \). TWL networks represent the other extreme by offering no flexibility at all. For example, in Figure 2.9a, all alternative routes require a longer travel distance. OW networks lie somewhere in between these two extremes: some OD pairs can be served by only a single route while others have multiple routes with the same (minimum) travel distance (see Figure 2.9a). Notice also that extra turns in OW networks do not impose an extra delay.

Let us now illustrate the path redundancy of the three networks with an example employing the simulation scheme (Figure 2.9b). In an empty network, we load different demand levels from point A to point B, and we see how the traffic assignment routes them through the network. Notice that \( T \) here is much larger than before because we are only considering one OD pair with travel demand in an otherwise empty network. In this figure, the thickness of each line represents the level of congestion on each link. Notice that TW networks appear to be able to spread flows among many different links especially when the level of demand is the highest. TWL networks are especially rigid: only when the system is very congested flows spread and even then, they do so only slightly. The alternative routes have considerably longer distance and/or more turns. OW networks keep a medium spread that is fairly constant with the different levels of demand. This happens because the differences in travel time between the shortest path route and the alternative ones are not as high as in the TWL networks. It is worth pointing out that this particular A to B combination allows more shortest path routes between the two points than other combinations in the OW network.
Figure 2.9. a) Examples of path redundancy in TW, TWL, and OW networks; b) Flow spread between an origin and destination with different levels of demand.

Figure 2.10 depicts a Cumulative Distribution Function (CDF) of travel times on the three different networks for a network size of 14 nodes and two demand levels, $T=0.10$ and $T=0.25$. The CDFs provide more insights into how the three networks accommodate trips differently. The CDFs for OW networks are steeper in the central portion and flatter in the borders of the domain, which shows that there is lesser variability across travel times. Examination of various percentiles confirms this. This further supports the above findings on the resilience of OW networks compared to TWL networks. OW networks are able to offer more consistent travel times in congested conditions. Trips are then routed more equally among routes, making them more resilient in congestion. However, when demands are fairly low ($T=0.10$), we see that TWL networks appear to have shorter travel times for almost all percentiles on the CDF. For higher demands, the shortest trips appear to be completed more quickly on
TWL networks than OW networks, while the reverse is true for longer trips. In all cases, these results also verify that TW networks are always the least favorable ones with the highest travel time variability.

![CDF of travel time for TW, TWL and OW](image)

Figure 2.10. CDF of travel time for TW, TWL and OW \(n=14\) networks and demand levels of \(T=0.1\) and \(T=0.25\).

### 2.6. Conclusions

In this chapter, we studied different configurations of streets in finite urban grids: two-way streets, two-way streets with prohibited left turns, and one-way streets. We considered a perfect finite grid network that emulates a typical urban environment and studied the behavior of traffic under two main scenarios: low traffic volumes and high traffic volumes. The former scenario was studied using analytical formulations, while the latter was studied using simulations that implement Static Traffic Assignment. The results presented here provide a general overview of how the different street layouts behave and what factors urban planners and traffic engineers should consider.

When traffic volumes are low, i.e. networks are not congested, drivers are likely to select the route that provides the shortest travel distance, which is independent of the behavior of other vehicles. In this context, analytical formulations can be applied providing very useful and compact expressions to calculate total distance traveled, average trip length, average turns per trip, and average travel time (Section 2.2 and Section 2.4.1). Analytical formulations are not only useful because of their simplicity, but because they allow us to identify behavioral trends. We find that TW networks achieve the lowest average trip length, whereas trips are on average one block longer for TWL networks and two blocks longer for OW networks. Regarding average turns per trip, TW networks require approximately 1.5 turns per trip, whereas TWL and OW networks require 2 and 2.5, respectively. Nonetheless, despite providing the least distance traveled and lowest number of turns, TW networks are penalized when it
comes to travel time because vehicle capacity at intersections is the smallest out of the three networks. As Figure 2.3 shows, TWL networks provide the better compromise between distance traveled and delay at intersections in uncongested scenarios.

When traffic volumes are high, car flows have a high impact on travel times. Hence, we developed a STA program that, despite its simplicity, approximates the system to the user equilibrium. TWL networks are still the best configuration for the majority of scenarios because of the better balance between distance traveled and delay at intersections. However, when demand is high and networks are congested, TWL networks lose that advantage because they are not able to spread congestion as well as OW networks (see Section 2.3.2), and because the HCM-2010 delay formulation employed penalizes more streets with lower saturation flows (as is the case of TWL). In addition to travel times, we also analyzed V/C ratios at intersections (Figure 2.6). This analysis shows how traffic is spatially distributed, and it reveals that OW networks ensure a more homogeneous spread of vehicles and lower travel times because the V/C variability is lower, although the average V/C ratios are higher. We have studied this phenomenon in depth (Section 2.5) by examining the redundancy of routes and the trip distributions on each network. We found that OW networks are able to spread traffic more homogeneously because alternative routes to the shortest path do not differ significantly from it. By contrast, TWL networks have less route redundancy and alternative routes involve an additional travel distance of 4 blocks. In other words, if the shortest path routes are very congested, if the STA distributes cars through the alternative paths, TWL networks worsen their performance. When streets are longer, the distance traveled has a greater effect on the travel time; therefore, in congested cases TW networks provide the best results for smaller networks.

Our initial analysis used a uniform demand pattern with different demand levels. We think that this configuration is the most transferable, simple, and the most representative of many cities. Moreover, analysis of non-uniform demand patterns (Section 2.4) reveals that the previous trends and comparisons among networks are still valid. The major difference is that OW networks gain a slight advantage compared to TWL networks as non-uniform demand patterns create more heterogeneous flows on the network, and OWs can normally handle better directional flows.

Overall, there is not a unique network that provides superior traffic operations for all conditions studied; each network presents particular advantages and disadvantages. After carrying out extensive evaluation under various scenarios, we summarize the main characteristics of each network type as follows:

- **Two-way networks** provide the shortest distance traveled and are able to spread congestion better due to the redundancy of shortest path routes. They are, however,
penalized by the reduced capacity offered at the more complicated intersections. This street configuration is suitable for very small, uncongested networks, and for networks with long streets to avoid queue spillback concerns. Moreover, the TW configuration can be easier applied to incomplete grid networks (as we see in Chapter 4) because the availability of routes maintains connectivity better.

- **Two-way networks with prohibited left turns** present the best compromise between distance traveled and capacity at intersections. That makes them suitable for the widest range of scenarios including both small and large networks, with short and long streets and low to moderate congestion. The main weakness of this network type is the rigidity of shortest path routes, i.e. the extra penalty imposed by alternative paths. This causes an inhomogeneous distribution of congestion and reduced resilience in cases with high levels of congestion. This is especially critical for networks with non-uniform demand patterns. However, for networks with demand patterns that result in uniform congestion distributions across all links (e.g. see Doig et al (2013)), this might not be as large of a concern.

- **One-way networks** are only the best option in a few scenarios. Their main disadvantage is that they force vehicles to travel the longest distances on average. These additional travel distances create a large penalty to overcome in networks with long streets. However, in networks with short streets and high levels of congestion, OW networks offer a better spread of the congestion than TWL networks as alternative routes to the shortest path do not impose a significant penalty. We see that especially when demand patterns are very heterogeneous, as OW networks are able to direct flows better to certain areas of the network. In addition, OW networks can still provide full connectivity with only one lane devoted to car movement (as we see in Chapter 5).

In the next chapter we will focus on the dynamics of traffic for the different networks and see how, spillbacks, can affect network performance, and we will see how this performance is affected by routing behavior.
Chapter 3

Traffic dynamics on one-way and two-way street grid networks

This chapter is based on the following research papers:


- Ortigosa, J., M. Menendez, and V.V. Gayah. Analysis of network exit functions for different urban grid network configurations. Transportation Research Record: Journal of the Transportation Research Board, No. 2491, 2015, pp. 12–21.

- Ortigosa, J., V.V. Gayah, and M. Menendez. Street configuration evaluation in urban grid patterns. Working paper.

3.1. Introduction

In the previous chapter, we studied different street configurations for urban grid networks: two-way street networks, one-way street networks, and two-way street networks with banned left-turns. We analyzed vehicle operations on these networks for a variety of congested and non-congested scenarios, and concluded that TWL networks offer the best trade-off between travel distance and vehicle-moving capacity at intersections under most cases. However, we identified that TWL networks offer little redundancy in shortest path routing options—only one shortest distance routing option exists between any origin-destination pair, whereas TW networks, and also OW networks to a lesser extent, offer multiple shortest path options for all ODs. This might make TWL networks more susceptible to congestion. However, we employed a STA method that did not account for the impacts of queue spillbacks and traffic dynamics, and used a deterministic (all-or-nothing) assignment that assumed full knowledge of travel times on all available routes.
The main objective of this chapter is precisely to further research on that network comparison employing tools based on network traffic dynamics, particularly macroscopic network-wide traffic models. These models have become promising tools in traffic network operations research. Of particular interest are relationships between average network flow and density (the MFD), and analogously trip-completion rate and the number of vehicles circulating (known as the Network Exit Function or NEF) (Daganzo, 2007; Geroliminis and Daganzo, 2008; Daganzo and Geroliminis, 2008). MFDs and NEFs have been widely used to study aggregate traffic behavior on small urban networks. These relationships are generally expected to arise when demands are uniform, congestion patterns are reproducible and drivers adaptively route to avoid congestion (Mazloumian et al., 2010; Daganzo et al., 2011; Gayah and Daganzo, 2011), although they have also been used to study the aggregate behavior of large-scale networks with non-uniform demand patterns (Geroliminis and Sun, 2011; Mahmassani et al., 2013; Saberi et al., 2014). This parsimonious tool is particularly useful for the design of network-wide traffic control strategies (Haddad and Geroliminis, 2012; Keyvan-Ekbatani et al., 2013; Yildirimoglu and Geroliminis, 2014; Ramezani et al., 2015). Consequently, as we study in the third part of this dissertation, several research articles have examined how MFDs/NEFs can be appropriately estimated and traffic data collected to help inform these strategies (Gayah and Dixit, 2014; Leclercq et al., 2014). Less attention has been paid to using MFD/NEFs to study street network design. To the authors’ knowledge, only Knoop et al. (2014) and Mühlich et al. (2015) employ the MFD to examine different hierarchical network structures. However, there is tremendous potential for using the MFD and NEF to study how various network topologies accommodate traffic. Summarizing, the contribution of our research is a comparison of these three urban grid street network configurations using the macroscopic NEF model that accounts for the impacts of queue spillbacks and traffic dynamics.

We employ a microsimulation software (VISSIM) that uses a DTA to capture dynamic effects that arise in heavily congested scenarios. DTAs are typically more computationally demanding (so fewer iterations and simulations can be performed) and more unstable as convergence is not guaranteed. However, they more realistically emulate driver behavior by routing according to previous travel time experience without assuming perfect travel time information. Most of MFD research papers use microsimulations that employ fixed routing schemes. In this thesis we employ an iterative DTA method (Chiu et al., 2011) that seeks to achieve user equilibrium conditions.

The NEF-based results show that network performance is generally worse on TW networks than the other two types because TW networks are penalized by the reduced
capacity offered at intersections. The comparison of TWL and OW networks depends on how drivers route themselves between origins and destinations. OW networks generally offer higher trip completion rates if drivers spread out across multiple routing options, while TWL networks offer higher trip completion rates if drivers tend to use the route that provides minimum travel time. This occurs because TWL networks offer a better trade-off between distance traveled and capacity at intersections, while OW networks offer more routing choices between origin-destination pairs.

The rest of the chapter is divided in three more sections: Methodology, Results, and Conclusions.

### 3.2. Methodology

In this section, we describe the three network configurations studied and the characteristics of the microsimulation model and DTA. We also describe how the NEF plots were generated to compare the performance of the three network configurations.

#### 3.2.1. Network characteristics

We consider a 10x10 grid network consisting of 180 links with each link 120 m long, and with a saturation flow of 1800 veh/h. Three street configurations exist: two-way streets, two-way streets with prohibited left turns, and one-way streets. Two travel lanes are provided on each link: the two lanes are used to travel in the same direction for OW networks, while one travel lane is provided for each direction in TWL and TW networks. TW networks also include an additional 25 m turning-lane at intersections to allow cars to make left turns; see Figure 3.1.

![Figure 3.1. Detail of an intersection in: a) TW network, and b) OW network.](image-url)
Every intersection is signalized with a 60-s cycle length, and no offset is applied between adjacent traffic signals. As previously discussed, TW networks are disadvantaged because they must accommodate conflicting left-turns at intersection. Thus, we assume that TW networks have a four-phase operation: 19 s of green time for through and right movements, 5 s of green for left movements (vehicles also use part of the lost time in amber for the left turns), and 12 s of total lost time (3 s per each change of phase). For TWL and OW networks, a simpler two-phase operation is used: 27 s of green time for through and turn movements, and 6 s of lost time (3 per each change of phase).

Trip origins and destinations occur at mid-block locations called demand nodes, which exist on each of the 180 links in the network. We consider a spatially uniform demand pattern in which each origin produces the same number of trips (on average), and each destination is equally likely to be chosen by a vehicle. While not realistic on large urban networks, such uniform demand patterns might arise on small urban networks with homogeneous design, of the type studied here. Two 3-hour simulation scenarios are considered. The first scenario has an average of 9,666 total trips during the first hour, 16,110 during the second and 22,554 during the third (demand levels of 0.30–0.50–0.70). The second scenario has 6,444, 12,888 and 19,332 trips during the three hours, respectively (demand levels of 0.20–0.40–0.60). These demand levels are chosen to provide an appropriate balance between uncongested and congested traffic states and a more continuous NEF.

Multiple random seeds are tested to consider stochastic variation in the trip generation and distribution processes to verify that the resulting macroscopic measures are independent of small demand fluctuations. Nonetheless, the demand patterns are still fairly uniform as the mean values of the trips generated at each origin varied by just 6.6% across simulation runs, and the coefficient of variation of trip starts across all origins was always less than 0.3.

3.2.2. Dynamic Traffic Assignment module in VISSIM 6

The network is simulated with the VISSIM software using the DTA module to consider how vehicles might route themselves in each of the three street network configurations. Researchers have fairly recently devoted many efforts to tackle the DTA problem (Peeta and Ziliaskopoulos, 2001; Chiu et al., 2011) and most relevant traffic simulators now incorporate DTA modules in their framework (Barceló, 2010). DTA methods are iterative processes that attempt to provide user equilibrium traffic states (Wardrop, 1952). The main advantage of DTA over STA is that it incorporates traffic dynamics by using Dynamic Network Loadings (DNL)—the traffic representation of a network
given the vehicle routes—as opposed to volume-delay functions. DNLs can be analytical functions or based on simulations.

The DTA module of VISSIM is an iterative process, details of which can be found in Fellendorf and Vortisch (2010) and PTV (2014); see Figure 3.2 for an overview of the DTA logic. For each OD pair from $i$ to $j$, VISSIM identifies $R$ potential shortest path routes available. The travel time of any route $r$ is calculated by averaging the travel times over the past $h$ iterations using the Method of Successive Averages (MSA). The value $h$ is selected to balance the trade-offs between stable travel time estimation (high $h$) and convergence efficiency (low $h$). For each of the $R$ routes available between OD pair $i$-$j$, the probability of a vehicle taking any particular route $r$ is provided by:

$$p_{i,j}^r = \frac{U_{i,j}^r k_{rc}}{\sum_{s=1}^{R} U_{i,j}^s k_{rc}},$$

(3.1)

where $U_{i,j}^r$ is the utility provided by route $r$ and equal to the inverse of its travel time; and $k_{rc}$ is known as the Kirchoff parameter that controls the relative weight of alternative routes. The value of $k_{rc}$ controls the sensitivity of driver routing to travel times and significantly impacts the relative comparison of the network types, as will be shown in the following section.

Figure 3.2. Flow chart of the DTA process in VISSIM (PTV, 2014).
In the DTA module, VISSIM assigns vehicles to routes according to (3.1) at regular intervals. We select a 20-min interval for the routing updates. Routing decisions are repeated based on the travel times obtained from previous iterations. The complete DTA simulation needs multiple iterations until the simulation converges to a stable solution. The convergence indicator chosen for this study is that the travel times along individual routes do not change by more than 5% between consecutive iterations. Other convergence indicators were tested (including stable link flows and link travel times); however, these indicators produced similar network results while taking longer to converge.

3.2.3. Network Exit Function and other macroscopic indicators

Many MFD research papers use microsimulations that employ fixed routing schemes obtained with STA methods, as opposed to a true DTA framework, e.g.: San Francisco, CORSIM (Geroliminis and Daganzo, 2008); Orlando, VISSIM (Gayah and Dixit, 2014); or the case of Zurich in VISSIM presented in Chapter 6. Other MFD studies employ static methods but allow drivers to change routes at every intersection according to the travel times measured in real-time; i.e. one-shot assignment-simulation (Chiu et al., 2011). Saberi et al. (2014) uses this methodology in DYNASMART-P to analyze the MFDs of real city networks (e.g. Chicago, Salt Lake City, Long Island) with different percentages of drivers that are allowed to adapt their routes. This tree structure routing more evenly spreads congestion and provides a large improvement when compared to fixed-route simulations. However, there is no learning process and rerouting options are limited once a driver is en-route; hence, traffic behavior might be far from reality. In this thesis we employ an iterative DTA method (Chiu et al., 2011) that seeks to achieve user equilibrium conditions.

A 5-min analysis period is used to report aggregate traffic measures and the NEF. Each period is further disaggregated into 30-s intervals, during which the number of vehicles on the network and trip completion rate are calculated to obtain the NEF. For each 5-min period, we averaged the accumulation values (i.e. number of active vehicles in the network) that the microsimulator provides every 30 s. The trip ending rates are aggregated for all the 30-s intervals of each 5-min period.

We also explore the time the network becomes both congested and gridlocked, and the total demand served during the simulation runs. The start of the congested period is defined as the first time a pre-specified number of vehicles (in this case, 20 vehicles during any 30-s interval) are not able to enter the network from demand nodes due to queue spillbacks (i.e. when the latent demand reaches 20 taking into account all the nodes). Gridlock is defined as when fewer than 10 trips are able to reach their
destination during a 30-s interval (for comparison, in a non-congested state more than 100 trips would end during a 30-s interval). Lastly, the numbers of vehicles that enter the network and reach their destination are tracked at 30-s intervals to help identify the number of vehicles to complete their trip and the number of vehicles that cannot leave their origin due to queue spillbacks (i.e. the latent demand).

### 3.3. Results

Multiple DTA simulations are performed for each of the three network configurations being studied. Each simulated 3-hour period requires between 20 and 40 iterations before convergence is achieved (i.e. until travel times on all routes do not change more than 5% between consecutive iterations). For the simulation results presented here, we use $h=5$ for the MSA calculations; however, tests show that the comparison of the three networks is the same if a value of $h=1$ is used.

#### 3.3.1. Comparison of full NEFs

The NEFs obtained from the multiple simulation runs are presented in Figure 3.3. Note that data from the first 5 min of each hour are discarded, as this is considered a warm-up period (i.e. time needed for the network to get loaded).

Figure 3.3 reveals that the shapes of the three NEFs are generally similar. However, the TW network NEF is significantly smaller than the TWL and OW network NEFs. The OW network NEF also appears to be slightly larger than the TWL network NEF. This suggests that for the range of network accumulations examined, OW networks are able to serve trips at a slightly higher rate than TWL networks and at a much higher rate than TW networks. Using the 10 highest trip completion rates observed as a measure of trip-serving capacity, TW networks have a capacity of 1,469 trips per 5 min, while TWL and OW networks have 1,729 and 1,821 trips per 5 min, respectively. Notice that the trip-serving capacities of OW and TWL networks are greater than the capacity of TW networks by 24% and 17%, respectively, even though the green time provided at intersections is just 12.5% greater for OW and TWL networks than for TW networks. Additionally, OW networks have 5% higher trip-serving capacities than TWL networks, even though both provide the same green time. Thus, the differences are caused not just by the green phase time provided, but also by the characteristics of each network, including routing options. The NEFs also unveil that the relative ranking of the three network configurations hold under uncongested and congested conditions (the increasing and decreasing portions of the NEFs, respectively).
These results differ slightly from results presented at the previous chapter employing analytical formulations and STA. There, we found that trip-serving capacities are generally higher on TWL networks than OW networks, while TW were the least efficient. The previous rankings were primarily due to the fact that OW networks require vehicles to travel the longest distances on average, and TW networks offered the lowest vehicle-moving capacities. As discussed in (Daganzo, 2007), both of these features are associated with lower trip completion rates. Some of these factors still arise in the DTA-based simulation performed here that accounts for the impact of queue spillbacks and congestion dynamics. For example, our simulations showed that average trips lengths are about 15–18% longer in TWL networks than TW networks, even though average delays are 13–18% higher in TW networks than TWL networks. However, unlike Chapter 2, we found that average trips lengths were approximately 5–8% longer in TWL networks than OW networks, which leads to 7–9% higher delays in TWL networks than OW networks. As will be shown in the following section, the relative rankings of TWL and OW networks change depending on the value of the
Kirchoff parameter selected—i.e. network performance is sensitive to driver route choice logic.

Note that our results assume that signal timings have zero offsets and ignore any impact of local coordination. Since signal coordination might provide an advantage to one-way streets, this might suggest that OW networks would further benefit from the consideration of signal coordination. However, we believe that the benefits obtained from coordination (if any) would tend to be insignificant at a network level. In general, coordination can only be provided for one of the two directions (North-South or East-West) at once; thus, any travel time benefits in the coordinated direction are likely to be offset by the additional travel time for vehicles traveling in the conflicting direction. Since we focus on a uniform demand pattern in this study, there are no clear directional flows that might be prioritized. According to Girault et al. (2015), in regular and symmetric networks, signal coordination strategies do not seem to consistently provide an advantage regarding macroscopic traffic magnitudes. Furthermore, these benefits would only be observed in the uncongested (increasing) section of the NEF and would not affect the observed capacity. This is because capacity is traditionally dictated by the total green time available at intersections and is not influenced by offsets unless blocks are extremely short; see Daganzo and Geroliminis (2008).

3.3.2. Influence of the route choice

Vehicle routing in the DTA module is based on (3.1), which uses travel times in previous simulation iterations to calculate the probability that a vehicle selects a specific route. As an example of this route choice logic, let us consider an OD pair with three possible paths and the following historical travel times based on the MSA algorithm: 10 min, 15 min, and 20 min. Using the default Kirchoff coefficient value $k_{rc}=3.5$, these paths will be chosen with probabilities of 0.75, 0.18, and 0.07, respectively. However, if a value of $k_{rc}=11.75$ was selected instead, the probability of choosing the path with the lowest travel time will increase to 0.99 and the other paths will only be chosen with a combined probability of 0.01. In this second case, the simulation routes flows through the shortest (travel time) path.

To examine the influence of this route choice behavior on the network performance, the NEFs of TWL and OW networks are provided for various values of $k_{rc}$ in Figure 3.4. This figure clearly shows that the performance of the TWL networks improves more than that of the OW networks as $k_{rc}$ increases. In fact, TWL networks offer higher trip completion rates for all uncongested and capacity accumulations with $k_{rc}>11.75$. These results are consistent with (Gayah and Daganzo, 2012), which assumed that vehicles are always routed through the shortest path, and with the results of the previous
chapter which employed a STA. In addition, TWL networks become congested and gridlocked later. OW networks also improve capacity slightly with \( krc > 11.75\) (to 1,834 trips every 5 min) although more variability exists during the convergence process as \( krc \) increases. Looking at the shape of the NEFs we see how TWL networks gain in trip making efficiency because of the high number of traffic states concentrated around the capacity point (Figure 3.4). TWL networks also perform better than OW networks in uncongested situations as \( krc \) increases, although OW networks tend to outperform TWL networks in congestion. Although it is not illustrated in Figure 3.4, it should be noted that TW networks also experience a substantial increase in capacity with this new route choice behavior (from 1,469 to 1,631 trips per 5 min), although their performance is still much worse than that of the other two networks.

Figure 3.4. NEFs of TWL and OW networks for different \( krc \) values: a) 3.5, b) 6, c) 11.75, and d) 20.

We also observe a change in the other macroscopic traffic performance indicators under different routing conditions. For \( krc = 20 \), average trip length is between 10 and 12% higher and average delay is between 2–3% higher for OW networks than TWL networks. Average speed is also between 1.5 and 2% higher for TWL networks. When
$k_{rc}=20$ the route choice model is more similar to the static results in which traffic flows are not spread between routes with small differences in travel times.

Summarizing, the performance of the different networks partially depends on the routing strategy adopted by drivers. If drivers are very willing to use alternate routes with small increases in travel times, OW networks will offer better performance than TWL networks. However, if drivers are very sensitive to travel time changes across competing routes, then TWL networks will outperform OW networks. These results suggest that OW networks might be more redundant than TWL networks—i.e. they provide alternative shortest-path routes that have little to no increases in travel time, whereas alternative shortest-path routes in TWL networks have much larger travel times. This reiterates the findings in the previous chapter where we unveiled the lack of redundancy of TWL networks.

### 3.3.3. Congestion speed and demand served

Here, we explore the amount of time required for the networks to become congested and the amount of demand they are able to serve. Figure 4a plots the time required for the networks to reach congestion and gridlock, while Figure 4b shows the number of vehicles that reached their destinations, are left circulating in the network, and have not entered the network at the end of each DTA iteration due to blockage at entrance location. Both sets of graphs are based on the results from a single random seed; however, similar results were obtained for other random seeds tested.
Figure 3.5. a) Congestion and gridlock times, and b) distribution of vehicles reaching destination, in the network, and remaining at origin at the end of each iteration; for the three network configurations and for a single random seed.

TW networks not only get congested and gridlocked faster, but have a lower portion of vehicles reaching their destination—in the case presented, fewer than 50% of the vehicles reach their destination in the last iteration of the DTA. TWL and OW networks
are consistently better than TW networks. During the last iteration of the simulation, congestion times are similar, although OW networks gridlock later. By the end of the DTA, about 70% of the vehicles are able to reach their destinations in the OW networks, approximately 10% more than in the TWL networks. In addition, we see how OW behavior is much more consistent across the DTA iterations, whereas TWL networks present a higher variability. These results were obtained with $krc=3.5$. As we mentioned in the previous section, when $krc>11.75$ the trends invert and OW networks present faster congestion and gridlock times instead.

### 3.3.4. Influence of the left-turn pocket in TW networks

According to Haddad and Geroliminis (2013), the length of the left turn pocket influences congestion propagation. A longer left-turn pocket would help to separate left-turning vehicles from through and right-turning vehicles, reducing the chance that left-turn queues spill over and block adjacent traffic lanes. Figure 3.6 depicts the comparison of TW networks using the default length, $le=25$ m, and a shorter left-turn pocket, $le=5$ m. Both congestion and gridlock start sooner in the case of the short left-turn pocket, and not even 50% of the cars are able to reach their destinations by the end of the DTA. In addition, convergence takes longer when shorter left-turn pockets are used.
3.3.5. Macroscopic convergence of the NEF

In this section, we investigate how the NEF shape changes across the various DTA iterations. To do so, we calculate for every pair of consecutive iterations the absolute errors (between them) in the trip ending rates for each 5-min interval. The results are presented in Figure 3.7a. These changes highlight the variability of TW and TWL networks across simulation iterations when vehicles are routed across multiple routes (when \(k_{rc}=3.5\)). Note that OW networks experience low and consistent errors, which shows, once again, their robustness in different routing options. In Figure 3.7b, we show the NEF of the last 15 iterations of the convergence process aggregated in groups of 5 iterations each. In this way, we can see the change in the NEFs until convergence is achieved. One common characteristic for the three cases is that the uncongested branch is fairly constant throughout the iterations. Most of the variability occurs in the congested branch of the NEF. As a matter of fact, when we exclude more congested...
states for the comparison, the error values are significantly lower. This is consistent with Chapter 6 where the MFD estimation provides lower errors in the uncongested branch. Nonetheless, the general shape of the NEF does not appear to change significantly; what changes is the spread of observed values along the congested branch. In other words, the macroscopic convergence of the NEF is achieved sooner than the DTA convergence with the criterion we employed. This suggests that even NEFs obtained before full convergence of the DTA might be useful to describe macroscopic behavior, and doing so would help to substantially save computational time in future studies.
Figure 3.7. a) Average absolute error of the trip ending rates between consecutive iterations, for the three network configurations; b) NEFs aggregated across 5 iterations for the three network configurations, and a single random seed.
3.4. Conclusions

This chapter continued the study between TW, TWL, and OW street configurations from Chapter 2. Here we look at the traffic dynamics employing a microsimulator integrated on a DTA framework. In addition, we employ network macroscopic tools like the NEF to describe the performance and the properties of the different networks. We describe below our findings.

Most studies that use the NEF (or MFDs) to study traffic network performance rely on microsimulation programs that employ STA routing schemes. STA is very robust and fast, but it is more unrealistic as it does not fully consider traffic dynamics. In this study, we seek to achieve user equilibrium conditions through the DTA module in VISSIM. This methodology is much more computationally demanding and the convergence slower and more unstable. However, the use of DTA improves upon previous studies by: a) considering the dynamics of traffic (e.g. spillbacks) that have a clear impact on travel time in these networks; and b) better reflecting driver behavior as it allocates demand through different routes according to some travel time criteria while allowing for randomness in individuals’ behavior. Table 3.1 summarizes the characteristics of both methods. Notice that the results of the two methods cannot be directly comparable because traffic indicators are different as well as the network scenarios analyzed.

Our study shows that TW networks are penalized the most by congestion due to the reduced capacity offered at intersections, and that they cannot achieve the levels of trip completion rates offered by TWL and OW networks. In other words, the NEFs of TW networks are lower than the ones for TWL and OW networks. The lower performance of TW networks is also observed through additional metrics, such as the time until the network becomes congested and gridlocked, and the proportion of demand that is able to be served during the simulated period. The TW network performance is also affected by the length of left-turn pockets at intersections. According to our simulations, shorter left-turn pockets trigger congestion faster in the system as left-turning vehicle queues start to spillover and block through-moving and right-turning vehicles in adjacent lanes.

The comparison of TWL and OW networks depends on how vehicles are routed using the DTA algorithm. Specifically, the Kirchoff parameter, which regulates the percentage of vehicles routed through alternative paths, plays a key role in the performance of all networks. The default value spreads drivers more evenly among alternative paths, even when they have slightly higher travel times than the shortest path available. In that context, OW networks provide higher trip-ending rates than TWL networks—i.e. OW networks produce a higher NEF. They also take longer to get
congested and gridlocked, and the number of vehicles reaching their destination by the end of the simulated period is higher. Another characteristic of OW networks is the performance stability across all the DTA iterations (as demonstrated in Figure 3.7). However, the results change significantly when larger values of the Kirchoff parameter are selected. In those cases, vehicles have a higher probability of being routed through the fastest path (according to the MSA average from previous iterations), and alternative options are more or less ignored. In that case, the performance of TWL networks improves considerably: they exhibit the highest trip completion rate, the latest congestion and gridlock times, and the highest percentage of vehicles reaching their destinations by the end of the DTA. The performance of OW and TW networks also improves slightly, although the performance variability across DTA iterations also increases.

<table>
<thead>
<tr>
<th>Table 3.1. Comparison STA and DTA methods</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Static Traffic Assignment</strong> (LeBlanc et al., 1975)</td>
</tr>
<tr>
<td><strong>Initial routing</strong></td>
</tr>
<tr>
<td><strong>Trip distribution among paths, driver’s behavior</strong></td>
</tr>
<tr>
<td><strong>Travel time</strong></td>
</tr>
<tr>
<td><strong>Traffic description characteristics</strong></td>
</tr>
<tr>
<td><strong>Convergence</strong></td>
</tr>
</tbody>
</table>

In summary, the performance of the different networks is sensitive to how drivers are routed within the network. That is consistent with the findings on Chapter 2, where we first revealed the lack of route redundancy of TWL networks in comparison to TW and OW networks. Even though TWL networks perform better than OW networks for most of the scenarios, when drivers are routed more evenly across the different paths, OW
networks outperform TWL ones. In very congested scenarios, or in cases where demand is not uniform, OW might present an advantage over TWL networks as they are more flexible.

Finally, the shape of the NEFs provides a good indicator of the overall performance of the networks. In general, the capacity of the network (provided by the height of the NEF) is the primary indicator of how much demand the network is able to serve and when congestion is produced. Additionally, we found that the shape of the NEF does not change significantly from iteration to iteration after a certain point, especially for TW and OW networks, even when full DTA convergence is not reached. These findings indicate that future works could employ only a certain fixed number of iterations, e.g. 15 or 20, as a reliable proxy for the NEF and save a considerable amount of computational time.
PART 2

Space removal

Since the 1950s, many cities have experienced urban changes to devote more space and infrastructure to cars. This, in turn, has modified the travel behavior and the activities carried out in cities. This cycle has then become unsustainable: increasing capacity for cars, increases induced demand, and negative externalities (Section 1.1). Fortunately, in the last decades, cities are trying to revert this cycle and shift towards more sustainable urban environments. One direction to reduce these externalities is aiming at recovering space from cars to find a more balanced mode share and higher living standards. The European Commission (2004), for example, presents the positive experience of many European cities that have restricted car usage in their city centers. The result is that cities not only gain in sustainability and livability, but traffic overall is reduced. Also in the U.S., these urban planning policies are gaining popularity after several cities have started removing freeway links (Billings et al., 2013). This phenomenon, called traffic evaporation (Cairns et al., 2002) and based on empirical evidence, describes the reduction of car demand when infrastructure capacity is restricted (Goodwin et al., 1998; Hunt, 2002; Clegg, 2007; Zhu et al., 2010; Xie and Levinson, 2011).

The urban space is rather scarce, and different transportation modes often compete for it (e.g. cars, buses, trams, bicycles, pedestrians). Taking space away from cars can provide the required space for improving the performance of other transportation modes, making them more appealing, and increasing their demand over time. The positive effects of space removal in cities (e.g. traffic reduction) are typically observed in the medium-long term, because of the rearrangement of traffic demand. In the short term, the traffic tends to worsen since demand has not yet changed and the supply is reduced (European Commission, 2004). Most of the decisions regarding the removal of road space or the spatial reorganization of cities are highly controversial, partly because their effects are not well understood yet. This research sheds some light on the topic by analyzing how different modifications to urban networks can affect their ability to cope with traffic.
Networks are composed of many elements, and often it is hard to evaluate how they individually affect the performance of the whole system. Much research has been devoted to analyze the impacts of the failure of network elements. Reliability studies, for example, started with computer and biological networks (Von Neumann, 1956). Concepts were later extended to transportation networks to evaluate the impact of different types of disruption (e.g. construction works, congestion, accidents), and to better plan evacuations and the response to other critical situations, e.g. natural disasters (e.g. Bell, 2000; Sakakibara et al., 2004; Jenelius et al., 2006; Tampère et al., 2007; Knoop et al., 2008; Qiang and Nagurney, 2008; Jenelius, 2009; Erath et al., 2009; Mattson and Jenelius, 2015). There are also some studies that particularly focus on how the demand rearranges in the network once an element fails (Parthasarathi et al., 2003; Scott et al., 2006; Zhu et al., 2010; Xie and Levinson, 2011). Notice that some of these papers, although highly related to our topic, often overlook the dynamic properties of traffic (such as the spread of congestion). Moreover, in our research, we consider the removal of space as a consequence of urban policies (space recovery) whereas most of the research to date focuses on finding a critical link to measure the reliability of the network with or without it.

In Chapter 4, we explore the behavior and the properties of traffic based on a two-way grid network and different sub-patterns created by removing links from this initial structure. The purpose of this removal is to see how vulnerable the grid pattern is to these changes, and to understand where the removal is more feasible (i.e. produces less impact to the system). Traffic analysis is done with the STA method presented in Chapter 2. In Chapter 5, we look instead at lane removal in a one-way urban grid network and also evaluate different strategies. A lane removal is a smaller step but it does affect street capacity and network flows. It is also more likely to happen as road space is partially devoted to e.g. dedicated public transportation lanes. If Chapter 4 is more focused on overall indicators such travel time, in Chapter 5 we focus more on the dynamics of traffic. Hence, we employ the microsimulation model and the DTA, employed in Chapter 3, to study traffic dynamics. In addition, we analyze the trade-off between demand and road space estimating the amount of trips which would shift to other modes when space is reduced.
Chapter 4

Street removal on grid networks

This chapter is based on the following research papers:


4.1. Introduction

The main purpose of this chapter is to better understand the removal of space in urban settings, and to provide some results showing that it is possible to remove streets from a city without worsening traffic excessively. The idea is to evaluate, from a general perspective, the impact of space removal in cities, and to provide some general recommendations to support the decision making process. We want to understand how different demand magnitudes and patterns (spatial, temporal, and topology-based) can influence the performance of traffic.

According to Cairns et al. (2002), after evaluating different cases of space removal, the traffic predictions are often too alarmist (i.e. the traffic situation predicted is much worse than it really is). Cairns et al. (2002) present some short term evaluations (less than one year) of the traffic impacts of link/capacity removal. For many of them, the traffic volumes in the alternatives routes increased after the removal, but the total number of cars decreased. Unfortunately, this evaluation lacks details on traffic impacts on drivers (e.g. how much did the travel time worsen?) and on the city (e.g. was there really a reduction of the veh-km traveled in the overall area?). Clegg (2007) studies more in depth the effects of traffic disruptions and proposes some models to represent routing behavior. There are also several research papers dealing with the study of individual cases like Zhu et al. (2010) that provide a detailed analysis on how traffic rearranges after the collapse of the I-35W Mississippi river bridge. Cairns et al. (2002)
expect when space is removed, some car demand shifts as the utility of driving the own vehicle decreases. This effect might occur in the medium long term. Our research, however, does not consider changes in demand and aims at studying how traffic behaves when these road space changes happen. In that sense, Scott et al. (2006) carry out a similar methodology when calculating the robustness index of the links they remove. Nevertheless, they only remove one link at a time which does not represent the result of an urban planning action.

We seek to address that type of question, by understanding and quantifying how the removal of links affects traffic on an urban area (both on drivers and on the city) in a short term (before demand rearranges). For that purpose, we create an abstract grid network composed of 100 nodes and two-way streets. This TW configuration is chosen to be able to keep connectivity as much as possible despite link removal. The link removal strategies seek to represent city planning policies aiming at recovering space for other activities. Links, or streets, are removed following three different strategies: in a total random manner, focusing on the center of the grid, and focusing on the perimeter of the grid. Up to 30% of the total links of the full grid are removed. In this chapter we employ the same STA method as in chapter 2 to represent traffic and we also employ a uniform demand pattern. When we remove space, demand does not change and we are able to depict the worst case scenario.

The rest of the chapter is organized in the following sections: Methodology, Results, and Conclusions.

4.2. Methodology

4.2.1. Network design

Our analysis starts with a regular grid (square blocks) where every node has four street connections and represents a four-way signalized intersection (except in the perimeter). Every link represents a two-way street with saturation flow of 1,800 veh/h for each direction, and they are all 111 meters long. The total area of analysis considered is 1 km² (100 nodes, Figure 4.1a). The idea is to emulate reality (in this case a dense urban network) as accurately as possible but without the loss of generalization.

The intersections play a determinant role in the traffic assignment model and specific links had to be created to represent their delay (Section 2.3.1). In addition, in this case, origins and destinations are on the intersection nodes and hence, we needed to add 8 more dummy links (Figure 4.1b). We have chosen this configuration, different from other chapters where origins and destinations are in the middle of the link, as otherwise
when links are removed, demand would also decrease—the origins and destinations would disappear. Although streets only have one lane per direction, in intersections, for every incoming direction an extra lane is added to separate cars that want to turn left (4-leg and 3-leg intersections, Figures 4.1c and 4.1d) or in some particular cases, cars that want to turn right (3-leg intersections, Figure 4.1d). In any case, if the links have a width of two lanes, when approaching the intersection the width will be of 3 lanes to allow the different turning maneuvers.

![Image](image.png)

Figure 4.1. The network model.

Same as in the other chapters, for the sake of simplicity we consider a uniform demand model where every node exchanges the same amount of trips with all the other nodes. Different levels of the number of trips exchanged between nodes provide different levels of demand in the network. In previous chapters we considered a wide range of demands because the objective was to see how networks deal with different loads, from uncongested to very congested conditions. Instead, here we try more realistic conditions to see the impacts of link removal. A rough estimation with macroscopic data from (Ajuntament de Barcelona, 2010), and population and traffic distribution assumptions, give us that in the central and grid-like neighborhood of
Barcelona, *The Eixample*, around 10,000 trips/h-km² are performed on average. That magnitude would account for a value of $\tau=1.12$ trips (11,088 trips in the 1 km² grid). As we will see later, although this value does not overload the system when the grid is complete, as soon as links are removed, some intersections start to show really high volume to capacity ratios. Notice that the streets of *The Eixample* have higher capacity than those in our grid. This level of demand is higher than the ones considered in previous chapters because also there are less origins and destinations as trips are generated on intersection nodes.

For each 4-leg intersection (without link removal) we have considered 17 s of effective green for the straight and right turn movements, and 8 s for the left turn movement, out of a total cycle of 60 s including 10 s of lost time. The left green time is higher than in previous chapters to provide slightly more flexibility because once links are removed, less routes are available and vehicles might be forced to make more left turns. Considering a saturation flow of 1,800 veh/h, the total capacity for the straight and right turn movements combined is 510 veh/h, and 240 veh/h for the left turn movement. Since the straight and right turn movements share the capacity, the delay is allocated jointly. This is the function of the dummy link situated before the straight and right movements. When a link is removed, the intersection becomes a 3-leg node. In that setting, we consider three phases of 15 s each and a total of 15 s lost time. All the movements have equal capacity of 450 veh/h. Notice that the lost time is larger than for 4-leg nodes. Although not very realistic, this is a consequence of averaging capacities given the large number of possible combinations in 3-leg nodes. As we mentioned in Chapter 2, the analysis time for calculating the delay with the HCM-2010 expression is one hour, and the incremental delay factor ($k$) (see HCM-2010 formulation) is 0.5.

### 4.2.2. Link removal strategies

The idea is to remove links from the grid following an iterative process. For a valid removal, all nodes must remain connected, i.e. there is at least one possible path between all OD pairs. Three strategies are used. With the first strategy, every combination of links removed is selected randomly. With the second and third strategy, we implement a random process influenced by the distance to the geometric center of the square that the grid occupies. For every link, the Euclidian distance from its middle point to the center is calculated. We use the Euclidian distance instead of the shortest path in the graph because the earlier one does not depend on the link removal process or other properties of the grid (e.g. scale of the blocks), and yet provides a good measure of geometrical centrality. With the central removal strategy, links are more likely to be removed in the center, whereas with the peripheral removal strategy, links are more likely to be removed in the perimeter of the grid.
As soon as we remove links, the intersections might change. Many 4-leg intersections (12 possible movements) become 3-leg intersections (6 possible movements); and in the case of only two links meeting on a node, since there are not any conflicting movements, full capacity is assigned to the two possible directions.

4.2.3. Traffic indicators

Not all of the traffic indicators can be extracted from the STA. Since the model is static, the analysis possibilities are limited, as the time dimension and the spatial spread of queues are not considered. The indicators employed to represent the traffic and travel behavior in this chapter are:

- **Total travel time**: the main indicator of the cost for the system.
- **Total distance traveled**: an indicator on the sustainability of the network (i.e. how efficiently the system is able to connect the demand nodes).
- **Average speed**: an indicator of how congested the network is.
- **Volume to capacity ratio**: same as the speed, a proxy for congestion. In addition, an indicator of how demanded and important within the network a specific link or intersection is. The average and maximum V/C ratios across the network will be calculated.
- **Delay at intersections**: a quantification, in travel time terms, of the impact of intersections.
- **Maneuvers at intersections**: an indicator of how the intersections are used. The percentage of right turn, left turn and straight movements are studied.

4.3. Results

This section presents the results obtained after running our traffic assignment model. Starting from the full grid network (Figure 4.2a), we follow an iterative process removing 3 links on every iteration until 30% of the links are removed. Since the process is random, different network structures are obtained every time. For that reason, we repeat this process 60 times (obtaining 60 network cases at every iteration). The same procedure is repeated for all the removal strategies: random (Figure 4.2b), central (Figure 4.2c), and peripheral (Figure 4.2d). In total we carry out 3,240 STA simulations. The demand employed ($\tau=1.12$) corresponds to a magnitude of 11,088 trips/h-km² (see Section 4.2).
4.3.1 Network performance indicators

Figure 4.3 depicts the evolution of the total travel time (in indexed values) when we remove links following the different strategies. Figures 4.3a, 4.3b and 4.3c show the results for the individual strategies when up to 20% of the links are removed. Figure 4.3d presents the whole trend (only with the average values) for all 3 strategies. We observe that the total travel time increases drastically when 30% of the links are removed.
There are many measures of centrality in a network, e.g. betweenness (Freeman, 1979; Sheikh Mohammad Zadeh and Rajabi, 2013). For simplicity, let us look at the flow that a link carries. We could say that the higher the flow is, the more routes that will pass through that link and hence the more relevance that this link has within the system. Given the characteristics of our network, central links concentrate more flow because the network is homogeneous and demand uniform (Figure 4.2a). Notice that in more heterogeneous networks (e.g. with different street hierarchies) these conditions might not hold. As depicted in Figure 4.2, the effect of link removal is significantly worse for central removal (Figure 4.3b) than for peripheral removal (Figure 4.3c). Logically, random removal values are between these two strategies (Figure 3a) and the results present more variability (i.e. more differences in values for a given removal percentage). Evidently, especially for the random removal, taking as a reference the average value does not guarantee that certain removal configurations do not make the system worse.

Figure 4.3. Total travel time indexed for: a) random, b) central, c) peripheral removals; and d) all the strategies (average values).
A complete grid structure (every node has 4 connections) allows to take many different routes with the same distance. When certain routes are more congested, cars can take faster alternatives without being penalized with additional travel distance. However, the link removal not only makes certain routes more congested, but it also reduces connectivity and then the availability of alternative routes. Figure 4.4a depicts a similar relationship as in Figure 4.3d for the different removal strategies, but as a function of total distance traveled. The less harmful strategy is the peripheral removal whereas the central removal increases distance considerably. Since with this strategy the center loses connectivity but not the perimeter, trips must take the surrounding roads to reach the central nodes. This triggers the rapid increase in total distance traveled. However, it can be observed how the distance reaches a point (around 1.8) where it does not increase anymore. As mentioned earlier, the total distance traveled cannot increase indefinitely because the space is limited. Moreover, when links from the outskirts start to be removed, vehicles cannot take longer (but faster) routes anymore. Remember that certain level of connectivity must be preserved as all nodes remain always connected.

The average speed calculated here is the relationship between the total travel time and the total distance traveled. The starting average speed is approximately 16 km/h. As figure 4.4b shows, during the first stages of link removal, this speed is maintained (except for the central removal where the value drops quickly from the beginning). After 15% of links are removed, the system worsens rapidly and the averages drop to values close to 4 km/h. Although that can seem unrealistic, some empirical works (Menendez and Ge, 2012) have shown that speeds on cities can be very low in peak hours (e.g. 4.5 km/h for certain routes in Zurich, Switzerland).

![Figure 4.4. Average values of: a) indexed distance traveled, b) average speed, for the three removal strategies.](image)
4.3.2. Network intersections

Intersections are the bottleneck of the system triggering most of the increases in travel time. In the full grid, time spent on an intersection accounts for approximately 70% of the total travel time. However, when 30% of the links are removed, intersections account for approximately 84% of the total travel time for peripheral removal, and 97% for central removal (Figure 4.5a). Although with the random removal strategy some configurations are very loaded (e.g. a network divided into two parts only connected by one link), on average the central removal still performs the worst. The delay at intersections is related to the load on the intersections (Figure 4.5b). The maximum values of the V/C ratio at intersections follow the same evolution as the other indicators: central removal presents the worst values and peripheral removal the best ones. Moreover, although not shown here for the sake of space, the difference between average and maximum V/C ratios increases as a function of the removal rate. This is not surprising as the system is used less homogeneously when we remove links.

![Figure 4.5. Average values of: a) total travel time spent on intersections, and b) maximum V/C ratios on intersections, for the three removal strategies.](image)

We have looked as well at the number of maneuvers at the intersections (Figure 4.6). Not surprisingly, for central removal not only the delay on intersections is bigger, but the number of movements on intersections is almost twice as that from the peripheral removal. This makes sense as routes are longer now, and cars must navigate through more intersections on every trip. Figure 4.6d shows how many movements (on average) were performed on each of intersection. The peripheral removal, which has the best results for each indicator, presents a significantly higher share of flows in 4-leg intersections in comparison to the other strategies. Retaining the largest number of 4-leg intersections (keeping the central area fairly well connected) allows the system to better absorb the link removal and to reroute trips homogeneously on the grid (Figure 4.2). In
other words, the high connectivity in the center offers shorter routes (less overall flows), but most importantly, offers more alternatives to allocate flows. This property of the full grid, that the peripheral removal is able to keep better than the other strategies, is the key to not having excessive loads at intersections, and limiting the increase in total travel time.

When links are removed, the number of alternatives for any given route decreases. This creates very loaded paths that impact greatly some intersections. On the other hand, when links are removed, a higher capacity is given to the intersections involved in the removal because there are less conflicting movements on them. Figure 7 shows the difference in total travel time between the link removal process carried out (Figure 4.3), and a link removal process without giving more capacity to the intersections (i.e. each movement on each intersection has the same capacity as in the initial grid independently of the links removed). It can be observed how for low removal rates, giving more capacity to intersections creates some cases where the total travel time is inferior to the full grid case (Figure 4.7a). In those cases the removal of links makes the system better in terms of travel times due solely to the gain in capacity at the intersections. Peripheral removal presents more cases of lower total travel time for low removal rates, as it can be observed in Figure 4.3c. We also see how for low link removal rates, the difference between the two curves is small in comparison to high removal rates, where the extra intersection capacity represents significant savings in total travel time (Figure 4.7b). This makes sense, as the main cause of increase in travel times is the extreme load that some intersections must cope with. If we restrict the extra capacity gain at intersections due to link removal, the V/C ratio becomes even bigger and generates even more delay. The gains in travel time are seen better in congested cases (e.g. high removal cases). However, this gain is somehow fictitious because removing links with the purpose of gaining capacity at intersections, only improves the system in the first stages and only for some cases (Figure 4.7a). After that, with more link removal the system will always get worse.
Figure 4.6. Number and type of maneuvers at intersections (average values) for: a) random removal, b) central removal, and c) peripheral removal; and d) percentage of movements at 4-leg intersections.

Figure 4.7. Total travel times for random removal to evaluate the effects of the gain in capacity at intersections: a) 0–10% link removal, and b) 0–30% link removal.
4.3.5. Demand level variation

The results presented above are restricted to this particular model and to the demand pattern and values considered. Our demand model is uniform. The number of trips interchanged between nodes (in our case \( \tau = 1.12 \) trips) determines the demand load (11,088 trips in the 1 km\(^2\) grid). We considered that this number of trips loads the system effectively and represents a dense urban setting. However, for the sake of completeness, we ran our model considering higher demand loads (\( \tau = 1.62 \) trips producing 15.878 trips in total; and \( \tau = 2.12 \) trips producing 20.778 trips in total). The overall patterns remain the same, and the rest of the findings still hold, although evidently, the more loaded the system is, the less capacity it has to absorb the effects of the link removal (see Figure 4.8). This is especially relevant for high demands (e.g. \( \tau = 2.12 \) trips) where small amounts of removal could create significant increases in travel time.

![Graph showing total travel time for different demand loads (\( \tau = 1.12; 1.62; \) and 2.12).](image)

4.4. Conclusions

Cities across the world are starting to recover space, previously devoted to cars, for other uses. This initiative, however, often brings a lot of controversy because, in the short term, without rearrangement of demand, the removal of space most likely will make the traffic situation worse. That fear often hinders urban policies that aim at space recovery. The goals of this research are precisely to better understand the removal of space in urban settings, and, to quantify how this removal affects drivers and the overall system. For that purpose we have studied an abstract grid network that represents an urban environment where we have removed links in three different ways: randomly, focusing on the center of the grid, and focusing on the perimeter. We have analyzed 60
removal sequences for each strategy, from the full grid to 30% of the links removed. In each iteration we remove 3 links and we assign traffic following a static macroscopic scheme based on the FW algorithm so the system tends towards the user equilibrium.

The results of this research help us to understand how grid urban systems behave. In our case, where all streets have the same characteristics and demand is uniform, central streets concentrate more flow. In more heterogeneous networks, these findings might not hold. Nevertheless, despite the fact that our analysis focuses on a very particular case, we believe that the behavior can be extrapolated to other networks. Summarizing, in the light of the results obtained we can say that:

• The full grid is a robust traffic system that allows a certain degree of link removal without affecting traffic conditions considerably. It is possible for a high connected city with an urban structure similar to a grid pattern, to recover space from cars without penalizing drivers significantly.

• The link removal strategy clearly influences the number of links that can be removed. In our case, the peripheral removal provides the best results and the central removal the worst ones.

• The intersections are the critical points in our network. Oversaturated intersections create delays that affect significantly the whole system. In addition, since the signal control is fixed, there is more capacity that is wasted because although some movements might be very loaded, others might not.

• The availability of many different routes is the key to achieve more homogeneous flow distributions and hence, not extreme traffic loads in particular intersections. The high connectivity in the center of the grid is what guarantees the redundancy of routes making the network more efficient and resilient.

• The gain in capacity at intersections by limiting movements when links are removed, reduces the overall travel time only for certain cases at low removal rates (this is more likely to happen for peripheral removal). However, as soon as links are removed, having less connected intersections limits the spread of traffic and the availability of routes, worsening the system rapidly. This is the case of the central removal.

• The high connectivity in the center also offers shorter routes to drivers. This leads to less overall distance traveled, which in turn, is directly related to the total flows in the network.

These conclusions do not necessarily advocate for prohibiting removal in the center which is also the most important place to actually recover space for other activities. In fact, the same argument of the central connectivity could be employed to increase the presence of public transportation lines. Although central removal is the most restrictive strategy, the system allows certain degree of removal. On the other hand, peripheral
removal is the least restrictive strategy. The random removal lies in between the two extreme strategies and can emulate a urban planning policy as combination of central and peripheral strategies. One interesting exercise would be to reproduce different removal patterns representing real world experiences and see in which type of category they fall into. Peripheral removal brings lower traffic impacts, but it does increase the traffic flows in the city center. That can be counterproductive because we actually worsen the city center that is typically the most valuable area.

The next chapter investigates a less drastic removal strategy, i.e. lane removal.
Chapter 5

Traffic dynamics of lane removal on grid networks

This chapter is based on the following research papers:

• Ortigosa, J., and M. Menendez, M. Traffic impacts of removing lanes on one-way grid networks. Presented at the Transportation Research Board 95th Annual meeting.


5.1. Introduction

A high quality urban space is a determinant factor to have a more compact, livable, and sustainable urbanization. More space efficient networks (Eichler et al., 2013) and urban planning measures are needed to better allocate road space. In the last chapter we studied how the removal of links (i.e. streets) from an abstract grid network affected traffic. These results show that depending on how road space is removed, the negative effects on traffic conditions can be reduced. As mentioned in previous chapter, our research studies the effects of infrastructure changes in the short and medium term. Although demand remains, drivers might seek faster routes and avoid these space restrictions by travelling longer distances. That phenomenon changes the traffic distribution of the system and hence the overall performance. This chapter expands on the analysis from Chapter 4 by considering a DTA that takes into account spillback effects and better traffic dynamics. Moreover, now we consider lane removal and not full street removal, as we believe these cases are more common in cities, e.g. dedicated bus lanes, introduction of bicycle or parking lanes, or widening of street curbs. Lane removal is a smaller step, but also is a more feasible action to be implemented in cities.

The removal of lanes affects network’s capacity. We know that even in a homogeneous network, traffic is inhomogeneously distributed (Doig et al., 2013). What
are then the effects on traffic when the network is heterogeneous because of lane removal? How do drivers react to capacity reductions? Can space gains be obtained with minimum traffic performance deterioration? We believe that answering these questions is helpful to support urban planning decisions. In this chapter, we study two removal strategies: central removal (where space in the city center is recovered), and peripheral removal (the opposite strategy). With these two strategies we want to emulate what urban planners would do. On the one hand, the most valuable space to recover would be in the city center. On the other hand, and based on the results from Chapter 4, it seems reasonable starting the recovery from the perimeter as, intuitively, streets carry less traffic and the impacts and controversy of these actions could be minimized.

In Chapter 3 we saw that tools such as the MFD proposed by Daganzo (2007) and Geroliminis and Daganzo (2008) are able to provide a general understanding of traffic in cities. It also allows us to study how the urban infrastructure accommodates traffic (Knoop et al., 2014; Mühlich et al., 2015) as we did in Chapter 3. We again employ the Network Exit Function as the main tool to analyze the performance of the different lane removal scenarios because it does not depend on the network type. We are also using the indicators presented in Chapter 3 that describe when the system becomes congested and how many vehicles are able to reach their destinations. This study pays especial attention to how congestion spreads and grows in networks. Ji and Geroliminis (2012) and Ji et al. (2014) studied the dynamics of congestion by looking at congested streets that are connected and how they spread on the urban network. Some of the indicators they provided are very useful to our present research. Additionally, we compare our simulation results with a theoretical and simplified congestion spread model to identify any common propagation patterns.

The space trade-off analysis in urban traffic is getting more and more attention by traffic researchers, especially when looking at the interactions with other transportation modes, particularly between cars and buses or trams. A higher presence of buses and trams in a city will decrease the space available for cars and hence the capacity of the network (Gonzales et al., 2011, Zheng and Geroliminis, 2013; Zheng et al., 2013; Chiabaut et al., 2014; Geroliminis et al., 2014; Ortigosa et al., 2015). Geroliminis et al. (2014) presented the 3D-vMFD which is a 3-dimensional representation of many MFDs for difference public transportation frequencies. This concept is especially interesting as it provides another dimension to the MFD control algorithm that is dependent on the city planning strategy.

Cairns et al. (2002) introduced the concept of disappearing traffic as the car trips that are no longer done by car because of a reduction of road infrastructure. This shift in
demand is caused by the increase of travel times triggered by the capacity reduction. As mentioned in Section 1.2, we are not modeling this induced demand and we assume that demand remains the same so we can assess the very short-term situation. However, we have included a section studying the amount of demand that needs to be reduced or lost, in each scenario to sustain free flowing traffic conditions. We compare these values with the percentage of space gained by approximating some expressions for the demand elasticity.

Following the analysis from previous chapters, we study an abstract grid network. We consider a one-way street network since this configuration is the only one able to keep one lane for vehicles without losing connectivity. We are using a 10x10 abstract grid with two-lane one-way streets. Parallel to Chapter 3, the routing behavior is represented with a microsimulation software, VISSIM 6, that has a DTA. The iterative behavior ensures that drivers adapt to infrastructural changes.

The rest of the chapter is divided in three more sections: Methodology, Results, and Conclusions.

5.2. Methodology

The following section presents the tools we are using in our research, and the characteristics of the networks and the lane removal scenarios we study.

5.2.1. The one-way grid network

This paper employs a 10x10 one-way grid network, with 180 links with two lanes each and 120 m long. There is a fixed signalization system (no offset) that has 60 s of cycle time, 26 s of green time and 3 s of yellow time per direction, in addition to 2 s of all red time. Trips are generated and ended in the middle of each link. Demand is uniform, i.e. each node sends the same amount of trips to the rest, although we do consider different levels of demand. Our study network represents a dense urban environment with high living density and mixed activities, as well as the presence of other transportation modes and urban space uses.

The simulation model employed is a microsimulation model with DTA, VISSIM 6. For further details, see Chapter 3 where the DTA algorithm is explained. Drivers are routed in each iteration until the system reaches convergence: travel time in each individual path does not change more than 5% between consecutive iterations. Finally, to account for the stochasticity of the simulation we consider several random seeds (4 in total).
5.2.2. Lane removal strategies

Employing the analytical formulations (based on shortest path trip assignment) presented in Chapter 2; we see how the flow distribution on a homogeneous one-way grid network with uniform demand tends to concentrate the flows in the center. The maximum flow values, however, are obtained in the perimeter as networks are finite and trips need higher detours to get to the nodes on the perimeter.

We remove lanes from the full two-lane network to a full one-lane network (Figure 5.1). Notice that all streets remain one-way throughout the removal process. To do so, we employ two strategies: central and peripheral removal (Figure 5.1). Cities might prefer to remove space from cars in the city center where there is a higher density of demand for public transportation and many other activities are concentrated. This corresponds to the first strategy, the central strategy. Cities might also plan the opposite strategy as they might be concerned about the traffic impacts that a central removal can create. In that case, we believe that removing lanes from the perimeter where (except for the border effect) flows are lower could be a reasonable strategy for a city aiming at minimizing traffic impacts. The removal logic follows the same reasoning as in Chapter 4.
Figure 5.1. The 10 network scenarios analyzed: a) initial and final scenarios; b) the two strategies: central and peripheral removal.
5.2.3. Macroscopic traffic indicators

The macroscopic traffic analysis is based on the Network Exit Function (NEF) which is equivalent to the MFD, and relates the number of trips ended in the network with the number of cars that are circulating in it. We compile network traffic measures every 30 s into 5-min intervals. The advantage of the NEF is that it does not depend on the network topology: changes in the number of lanes would change the MFD as they would translate in flow and density changes. Hence, networks would not be comparable. In addition to the NEF, we also analyze when the networks become congested and gridlocked by looking at the trip completion rate and at the number of vehicles that are unable to exit their origins employing the indicators from Chapter 3. These indicators are very useful to compare networks that might exhibit similar capacities but not during the same amount of time.

In this chapter, and in order to consider the effects of lane removal on the spatial homogeneity, we are using several indicators which describe how traffic is distributed. We are first looking at link vehicle densities, employing the average or mean density, and the standard deviation of the sample. Our indicator to represent the distribution homogeneity is the coefficient of variation (standard deviation/mean). These values are obtained for each link of the network every 2.5 min using the link evaluation function in VISSIM. In addition, we want to actually see how congestion is spread on the network. Ji and Geroliminis (2012) and Ji et al. (2014) employed the graph theory concept of the Maximum Connected Component (MCC) to describe congestion. We are using that concept to illustrate the physical growth of gridlock. We define the MCC as the subgraph with connected links that are jammed. Obviously, a jammed link is considered as the one with jam density. We compare the total amount of jammed links and the amount of connected jammed links (MCC). The search algorithm of the MCC is simple: starting from one jammed link, we evaluate the links connected to it (i.e. that a vehicle can travel from one link to the other) and that are also saturated.

5.2.4 Congestion propagation model

One of the purposes of this PhD thesis is to provide simple formulations that help us understand the behavior of traffic operations in urban grid networks. We propose a congestion propagation model in one-way urban grid networks by making some simplifying assumptions about traffic behavior. Later, we compare this idealized propagation with the results obtained from simulations.

We consider a regular grid network with a homogeneous traffic distribution where flows and densities are close to capacity values. At a certain point, congestion starts in
some of the links. To be consistent with the symmetry of the network, congestion is triggered in 4 links at the same instant. $\Delta t$ is the time needed for these congested links to spillback onto the upstream links. Consequently, these new links need also $\Delta t$ time units to spillback onto the links upstream of them. This time interval remains constant independently of the number of links that are already congested. Evidently, in our model, vehicles do not reroute and there are only two possible traffic states for links: uncongested or jammed. Discretizing time in $\Delta t$ intervals, we illustrate in Figure 5.2 how congestion propagates.

Figure 5.2a depicts congestion propagation when it starts in the center, whereas Figure 5.2b illustrates the process when propagation starts from the perimeter. Propagation curves are shown in Figures 5.2c and 5.2d. We chose to depict a high network size ($n=20$) to be able to see clearly the pattern these curves describe. Figure 5.2c describes how many new links become congested at each interval. Central propagation has a clear shape that can be approximated to an isosceles triangle. It presents a peak value of $8(n - 1)$ congested links that occurs: in the $\frac{n}{2} + 2$ interval (if $\frac{n}{2}$ is even); or in the $\frac{n}{2} + 1$ interval (if $\frac{n}{2}$ is odd). $n$ represents network size. The triangular expression translates into a cumulative curve composed of two equal parabolas of opposite sign (Figure 5.2d). Peripheral propagation curves do not present a clear and generalizable expression for small networks ($n<14$). However, they do present also a triangular shape (Figure 5.2c), in this case asymmetric and with a higher peak, almost twice as much as for the central removal one. In other words, congestion speed propagation is higher and networks need less time intervals to gridlock. Notice that in peripheral propagation, congestion starts from 4 separate links whereas these 4 links are together in central propagation. This division on the propagation front allows a higher congestion growth speed.
5.3. Results

In this section, as Figure 5.1 presents, results from a total of 10 network cases with different demand levels (from 0.3 to 0.65 trips between nodes) are presented. Two of these scenarios correspond to the full grid either with two lanes or one lane. There are 4 scenarios representing central removal, and 4 scenarios representing peripheral removal. We employ several random seeds to average results and give consistency to our findings. Data obtained from the simulation consists of: flow and density measurements (for each link every 2.5 min), and network indicators (e.g. total trips ended, number of vehicles in the network) every 30 s. The DTA is an iterative process and each hour of simulation needs around 20 iterations (i.e. 20 one-hour simulations) until convergence (same criterion as in Chapter 3, when travel times in links do not differ more than 5% in consecutive iterations).
5.3.1. Network capacity

As discussed earlier, the NEFs are created to calculate the capacity of the different networks. Each NEF is created with data from many one-hour simulations with different demand levels and random seeds. In Figure 5.3a we plot the 6 NEFs corresponding to central removal scenarios together to observe the range of capacities. In Figure 5.3b, the same operation is repeated for peripheral removal. Evidently, for every scenario there exists a NEF, and hence, there is a network capacity value. We average the 10 highest values of capacity (including all random seeds) to estimate the capacity for each scenario. These values are plotted in Figure 5.3c.

Figures 5.3a and 5.3b look fairly similar. Bear in mind that the NEFs corresponding to the full two-lane and one-lane networks are plotted in both figures. Nonetheless, we observe, in Figure 5.3a, more spread in the capacity region. Instead, in Figure 5.3b, we see a more defined free flow curve and more concentrated points in that capacity region. Moreover, the central removal NEF (Figure 5.3a) is wider than the peripheral removal NEF (Figure 5.3b). Additionally, peripheral removal NEF presents more extreme traffic states—either uncongested or jammed—while central removal presents a more continuous range of traffic states.

Looking at the average capacity (Figure 5.3c), results indicate how removing lanes in the center has a higher impact than removing them in the perimeter. In fact, by removing only the 4 central lanes, the capacity decrease is approximately of 9%. Alternately, removing 62 lanes in the perimeter (the outer crown) translates in only 6% of capacity reduction. The difference between central and peripheral removal reaches a maximum of approximately 20% of the full two-lane network capacity. These results are consistent with previous chapter’s findings and indicate how in grid networks with uniform demands, the central links play a key role on the connectivity of the system. Capacity reductions are not proportional to the amount of space removed. They are not only dependent on the strategy but they are also case dependent. The last scenario, the full one-way network, has a capacity approximately 33% lower than the full two-lane network. Put differently: removing half of the road space reduces one third the traffic capacity. Finally, we observe that the last peripheral removal case (point 5P in Figure 5.3c) presents lower capacity than the full one-lane scenario (point 6). We expand on that issue in the following sections.
Figure 5.3. a) NEFs overlapped for central removal for all levels of demand and random seeds; b) NEFs overlapped for peripheral removal for all levels of demand and random seeds; c) Network capacity for the different scenarios and removal strategies.

5.3.2. Time to congestion and gridlock

Capacity measures from the previous section explain how many trips can be served by networks at their peak performance. However, this capacity state might be only temporal; as soon as queues spillback the system might become congested and eventually gridlocked. In this section we study when the different scenarios become congested under several demand levels. To do so, we employ the indicator introduced in Chapter 3 that detects when networks start getting congested—i.e. latent demand increases because vehicles cannot leave their origins- and when they become gridlocked. Figure 5.4 depicts the average times to congestion and gridlock for both removal strategies. We choose three demand levels (0.40–0.45–0.50) that provide a wide range of suitable results for comparison, and are neither too congested nor too uncongested. These results are averaged across random seeds.
In general, looking at Figures 5.4a–f, results are consistent with the previous section (Figure 5.2): peripheral removal scenarios are able to cope with higher demand loads than central removal without becoming congested. Specifically, we observe a clear difference in the evolution of congestion and gridlock curves between both strategies. While central removal gradually deteriorates traffic performance (Figures 5.4a–c), peripheral removal does so very abruptly (Figures 5.4d–f). The largest difference occurs between the 3P and 4P scenarios (Figures 5.4d–f) which also present the highest capacity drop between consecutive peripheral removal scenarios (Figure 5.3c). Overall, peripheral removal is more resilient, however, once congestion starts, in cases with a high lane removal rate, the system deteriorates fast and the marginal cost of any additional km of lane removal is high.

Along the same lines, another apparent distinction between central and peripheral removal is how fast different networks become gridlocked once they reach congestion. We see how peripheral removal cases present particularly short gridlock times. There are several factors that could explain this phenomenon. On the one hand, as illustrated in Figure 5.4f, in the 2P scenario the time difference between congestion and gridlock is shorter than 3P or 4P cases because not all random seeds get congested. Since these curves are obtained averaging all random seeds, the cases where congestion does not appear are counted as zero, resulting in a lower average difference. On the other hand, there is a difference on congestion propagation speed. We believe that this propagation speed is mainly influenced by: i) the amount of one-lane links that a network has, and ii) how congestion is originated and propagated. In networks with one-lane streets, and for equal demand level, congestion propagation is higher than networks with two-lane streets because links spillback earlier. That could be the case of 4P scenario in Figure 5.4f. In Section 5.3.4 we look into more detail how congestion propagates in networks to further understand these differences.

Last but not least, Figure 5.4d shows what we first saw in Figure 5.3c: in peripheral removal, the 5P case performs worse than the full one-lane network. Although the 5P network has 4 extra lanes, capacity levels as well as congestion and gridlock times are lower than in scenario 6. This effect is more relevant when demand levels are low and systems only become congested at the end of the simulation. Figure 5.5 depicts the congestion evolution along DTA iterations for the scenarios 5P and 6 (similar to Chapter 3) and a lower demand level, 0.35. Notice that this is for a particular random seed because every case needs different number of iterations to converge. However, the same pattern can be observed for other random seeds as well. The full one-lane network (Figure 5.5b) is able to improve more along the iterations of the DTA algorithm with more efficient rerouting—i.e. by distributing congestion more homogeneously. The improvements for the 5P scenario occur more gradually, and are bounded to a
performance level lower than that of scenario 6. Even when more DTA iterations are performed, case 5P remains at that performance level. This is due to the actual layout of the network. The two-lane roads in the center for the scenario 5P (Figure 5.5a) create a more inhomogeneous traffic distribution than the full one-lane case. More details on traffic spatial inhomogeneities are discussed below.

Figure 5.4. Evolution of congestion for central (a, b, c) and peripheral (d, e, f) removal for different demand levels.
5.3.3. Traffic inhomogeneity

In the previous sections we saw how different removal scenarios deteriorate the system. Now, we study the evolution of the spatial traffic distribution, looking at vehicle densities on links. If all links had the same density, i.e. all links with the same traffic state, traffic would be completely homogeneous. Notice, however, that we should not expect to see this, as most of our scenarios consist of a mix of one-lane and two-lane streets, leading automatically to different densities per km of road. Particularly, we look at the main statistical indicators that describe the density distribution: the mean and the standard deviation. Figures 5.6a–c show the average density and standard deviation for all time intervals and random seeds for the demand levels we previously employed. Figures 5.6d–f illustrate the variation coefficient which explains how disperse the distribution is.
In scenarios with more space removal and higher demand levels, the average density increases as networks become, in general, more congested. These figures are totally consistent with previous sections. Central removal presents a rapid increase of densities that translates in a decrease of flows. They deteriorate more than peripheral removal scenarios. Similarly, we see how for low demand levels, the last peripheral removal scenario (5P) presents worse indicators than all the other scenarios, including the full one-lane network. For low space removal levels, peripheral removal leads to a lower spatial inhomogeneity (i.e. lower coefficient of variation) than central removal.
However, as the removal level and/or the demand level increase, the level of inhomogeneities caused by the two removal strategies becomes the same: the one for central removal decreases, and the one for peripheral removal increases. As expected, the level of inhomogeneity is the same for the full one-lane and the full two-lane networks; both layouts lead to the same routing process.

Traffic distribution is further illustrated in Figure 5.7 showing the density spread for each network case for a given demand level (0.45) and a time interval (30–32.5 min). We observe how networks spread congestion differently. The full two-lane and one-lane network scenarios spread congestion more homogeneously, yet with a higher vehicle concentration in the center of the grid. Figure 5.6 shows how as soon as space is removed from the center, vehicles look for faster routes in the perimeter (where two-lane links are) creating more inhomogeneous distribution and concentrating many vehicles in the perimeter. That effect ends up being counterproductive to the system as we see it in scenario 5C. The benefits of having two-lane roads in that case are minimal in comparison to the congestion created at the perimeter by these two-lane roads which, in addition spreads rapidly to the one-lane area. In Figure 5.7 we observe that when space is removed from the perimeter, vehicles concentrate even more in the center. That distribution is really inefficient for cases 4P and 5P, where such a high density in the center triggers the appearance of congestion. In both cases, once congestion reaches the one-lane portion of the grid, it spreads fast. The next section explains the origin and spread of congestion.

Figure 5.7. Vehicle density distribution for central removal (a) and peripheral removal (b) for a demand level of 0.45 and time interval 30–32.5 min.
5.3.4. Origin and spread of congestion

In Figure 5.8, we analyze the evolution of the totally congested or jammed links and the Maximum Connected Component. As presented in the methodology section, the MCC is the subgraph of links that are completely jammed (Ji et al., 2014). The spread of these jammed links causes the network to gridlock. Figures 5.8a and 5.8b depict how fast the number of congested links and the MCC grow for central and peripheral removal cases. Figure 5.8c does the same for the regular cases, full two-lane network and full one-lane network, scenarios 1 and 6 respectively.

Figure 5.8 shows that the growth of the congested links and the MCC start separately, i.e. not all jammed links belong to the MCC. When the system is more congested, these two values converge, i.e. practically all congested links belong to the MCC. Additionally, we see how the growth of congested links is faster in cases 6 (Figure 5.8c), 5P, and 4P (Figure 5.8b) than case 5C (Figure 5.8a). The former ones have a higher proportion of one-lane streets than the latter ones. Congestion propagation in those cases is really fast, e.g. in scenario 6, after 30 min of simulation the MCC has 68% of the network links (Figure 5.8c). Although cases between removal strategies cannot be directly compared as they do not have the same network length, we notice a relation between one-lane streets and propagation speed. In Figure 5.3 we already noticed that peripheral removal cases presented more extreme traffic states, going fast from uncongested to very congested conditions. Another possible reason for that different behavior is analyzed below and is related to where congestion is originated.
As we saw in the last section, there are clear and distinct traffic distribution patterns, i.e. peripheral removal concentrates more flows in the center than central removal and vice-versa (Figure 5.7). However, the spread pattern of jammed links does not show a clear difference between removal strategies. We collect for each case and interval the first links to be jammed. We aggregate this data in Figure 5.9 showing which links have a higher frequency of triggering congestion. We also represent in Figure 5.9 the different regions of the grid network, emphasizing in red the 4th crown as it is the region where the highest frequency of congested links for both removal cases occurs.

As we consider more intervals (Figures 5.9b, 5.9c, 5.9e, and 5.9f) there is a slight difference between both removal strategies. For the case of central removal, there are more links congested in the center of the grid whereas in the perimeter removal there are more congested links in the outskirts of the grid (we consider the 3x3 central network composed of 60 links as center, and the other 120 links as perimeter). Case-individual plots show the same tendency, although for the sake of simplicity, only the aggregated results are presented in Figure 5.9. This means that there is a slight tendency for central removal to start congestion more towards the center; and also, a tendency for peripheral removal to start congestion more in the periphery.
In the light of these results (Figure 5.9) we see that there might be slight similarities with theoretical propagation curves presented in Section 5.2.4. Although congestion is triggered in the 4th crown of the grid (Figure 5.9), there is a certain tendency depending on the removal strategy. From this, we can infer that central removal presents a curve more in the lines of central propagation (Figure 5.2a) and peripheral removal more similar to peripheral propagation (Figure 5.2b). The difference in congestion propagation combined with the higher presence of one-lane streets for perimeter
removal cases, explain why peripheral removal cases collapse so rapidly once congestion starts.

5.3.5. Level of service and traffic demand

Employing the structural engineering simile, the previous sections have dealt with the “ultimate limit state”: we studied under which conditions and levels of demand, networks collapse. This last section, instead, studies networks and demand levels from the Level Of Service (LOS) perspective. The question is not how many trips networks can handle before collapsing, but how many trips networks can handle while maintaining similar LOS conditions for drivers.

We gather data from all simulations (all random seeds, demand levels, and network scenarios) and we select all those demand levels that maintain an uncongested Level of Service for each scenario. In other words, for a given scenario, and averaging all random seeds, we look for the maximum demand level that the network can cope with in uncongested, close to capacity, conditions. To find these values the three following conditions must apply: i) average travel time should remain constant and not deteriorate excessively; ii) we accept that congestion can appear at the end, but gridlock cannot be reached; and iii) all vehicles should be able to at least, leave their origins. We are aware that these conditions might seem not strict enough as we do not differentiate between uncongested levels. Nevertheless, the differences between uncongested levels, in terms of travel time, are negligible in comparison to the increase due to appearance of congestion. We need to consider that the average trip length, employing analytical formulations presented in Chapter 2, in uncongested conditions is approximately 1 km and a difference in uncongested speeds is negligible in comparison to queue spillbacks, for example. We analyze all the feasible demand levels from 0.30 to 0.65 trips every 0.025 trips increments. In the cases where even this interval was too large, as it presented too much variability, we have chosen the middle demand level point between two consecutive values.

Figure 5.10a illustrates the evolution of total trips that networks can handle to maintain uncongested conditions. Figure 5.10a resembles Figure 5.3c which shows the evolution of capacity values although evidently, demand values are slightly lower in comparison to Figure 5.3c as the LOS indicator that we employ is more restrictive than the capacity. In any case, this figure is consistent with all the previous findings. In Figure 5.10b we look at the shift of demand—i.e. the percentage of demand we would need to shift to keep the same LOS- in comparison to the percentage of space gained. Peripheral removal strategies are always located under the 1:1 proportion meaning that for a given reduction of demand (in percentage) we would always get a higher
percentage of possible space removal. Central removal, instead, does not bring a good return on space, especially in the first stages. In fact, as it can be seen in this figure, the difference on demand between the third scenario and the full one-lane network is really low in comparison to the possible space gain. Central and peripheral removal are relatively extreme strategies. As we saw in the previous chapter, other strategies of space removal fall in between these two.

![Figure 5.10](image-url) a) Maximum number of trips that networks can have to maintain the same LOS; b) Comparison between the percentage of demand shift to maintain the same LOS and the percentage of space gain.

Figure 10b represents the demand and network length magnitudes in a relative scale. We can actually fit these two magnitudes with the least squares method to the following curves, $f_c$ and $f_p$, that approximate the percentage of trips that need to be shifted to maintain equivalent LOS:

\[
f_c(l) = 4.544\ln(139.25l),\]  \tag{5.1}
\[
f_p(l) = 0.806l,\]  \tag{5.2}

where $l$ is the network length gain (recovered for other uses) in percentage values. Differentiating (5.1) and (5.2) we can also find the elasticity of demand vs. network length or urban space:

\[
\frac{df_c(x)}{dl} = \frac{4.544}{l}, \]  \tag{5.3}
\[
\frac{df_p(x)}{dl} = 0.806. \]  \tag{5.4}
These expressions might not be extremely accurate but they can provide planners a rough estimation of how much demand should be reduced for a certain removal strategy to maintain the same LOS. Compared to the existing literature, these elasticity values are high because we are assuming that the LOS is maintained, and because our demand pattern does not change either. As said earlier, this is a conservative approach that does not take into account the disappearing demand, hence, these values could be considered as an upper bound.

5.4. Conclusions

Urban planning and design require a more detailed analysis on how infrastructure usage affects mobility in cities. In this context, urban areas need to rethink traffic behavior in “slimmer” scenarios where less road space is devoted to cars. In Chapter 4 we dealt with full link removal and in this one we have studied the lane removal process. It is a smaller step, but it is a common and feasible action that cities implement when they provide more space to pedestrians, bicycles, public transportation, or other activities. Particularly, we have studied two removal strategies that aim to mimic urban planners’ way of thinking: removing lanes from the center as that is the most valuable road space to be recovered; or removing lanes from the perimeter to minimize traffic impacts. Same as in the other chapters, we study abstract grids with a simulation model, in this case the one employed in Chapter 3 that captures dynamics of traffic and has a Dynamic Traffic Assignment.

As expected, results show that lane peripheral removal impacts are lower than central removal. We have generated the NEFs of the different removal scenarios (Figure 5.1) and estimated their capacity(Figure 5.3). If the full one-lane network has approximately two thirds of the initial grid capacity, both strategies tend differently to that value. When only 4 of the central links are removed, the capacity drops by 9% whereas if we started the lane removal from the perimeter, not using 62 lanes would only translate in only 6% reduction of capacity. The difference between central and peripheral removal reaches a maximum of approximately 20% of the full two-lane network capacity. The last scenario, the full one-way network, has a capacity approximately 33% lower than the full two-lane network. Put differently: removing half of the road space reduces one third the traffic capacity.

Employing the indicators developed in Chapter 3, we measured the evolution of congestion and gridlock for the different networks. In general, results are consistent with the capacity findings: peripheral removal scenarios are able to cope with higher demand loads than central removal without becoming congested. Specifically, we observe a clear difference in the evolution of congestion and gridlock curves between
both strategies. While central removal gradually deteriorates traffic performance, peripheral removal does so very abruptly. Overall, peripheral removal is more resilient, however, once congestion starts, in cases with a high lane removal rate, the system deteriorates quite fast and the marginal cost of any additional km of lane removal is quite high.

At the beginning of the chapter, we brought up one research question: is it possible that heterogeneous networks create more homogeneous traffic distributions? Given our networks and lane removal characteristics, that does not happen but rather the opposite. Removal scenarios present always more heterogeneous distributions than the full scenario. When central removal is carried out, cars are routed more on the perimeter creating higher flows and densities there. Instead, when we remove lanes from the perimeter, the behavior is the opposite and the grid center is more loaded. This heterogeneity creates some unusual situations. That is the case, mainly, of the last scenarios of peripheral removal. The 5P scenario in particular, which has all one-lane streets except the 4 central streets, performs consistently worse than the full one-lane network which has slightly less road space. The reason is that vehicles are still attracted to the center because of the two-lane roads, creating a worse traffic distribution. Put differently, the pattern characteristics encourage drivers to have a routing behavior that is more prejudicial to the system.

We have looked at how congestion levels spread on the networks, particularly looking at the Maximum Connected Component of jammed links. The slope of the MCC curves presented in Figure 5.9 show that in the scenarios with more one-lane streets, peripheral removal and the full one-lane case, congestion grows significantly fast, reaching gridlock in about 10 min once congestion arises. Although networks spread traffic very differently, there are not such clear patterns when analyzing the origin of congestion. The 4th crown of the grid, almost in the perimeter, presents the highest amount of links triggering congestion. After few minutes, however, central removal presents more congested links in the center and peripheral removal more in the perimeter.

Section 5.3.5 presented an idealized model of congestion propagation for one-way urban grid networks. With that model, we have studied two cases of propagation: i) when congestion starts in the center; and ii) when congestion starts in the perimeter. This model can be used by practitioners to approximate the propagation of congestion. When comparing results from the idealized model we see a similar pattern: central removal congestion growth resembles the theoretical curve representing central propagation. In addition, peripheral removal congestion growth presents similar characteristics to peripheral propagation. These similitudes, together with the fact that
there is a higher presence of one-lane streets in peripheral removal scenarios, explains why congestion propagates faster in peripheral removal scenarios than in central removal scenarios.

Finally, we studied the trade-off between demand and space removal. We obtained, for each scenario and removal strategy, the maximum number of trips that could be handled in uncongested conditions. Consequently, we compared the percentage of demand that needs to be shifted (from full two-lane case) vs. the space gains obtained. Results show how peripheral removal provides a better return on space gains except for the two last cases which are not worth to implement as they do not improve the one-lane scenario and use more road space. The same occurs with central removal. In addition, we approximated to a logarithmic and a linear curve these relationships. Differentiating these curves, we obtained some approximate values of demand elasticity on space removed. These magnitudes could be also useful to urban planners. Central and peripheral removal are relatively extreme strategies. As we saw in the previous chapter, other strategies of space removal fall in between these two, and are expected to present intermediate elasticity values.
PART 3

Urban traffic monitoring resources

The macroscopic modeling of traffic in cities has driven many research efforts in the last decades because it presents important advantages over other models that typically need too many inputs, and yet might be inaccurate describing the unpredictable behavior of drivers and congestion (Daganzo, 2007). The Macroscopic Fundamental Diagram has been widely established in the scientific community (Geroliminis and Daganzo, 2008; Daganzo and Geroliminis, 2008). In Chapters 3 and 5, we employed the Network Exit Function, equivalent to the MFD, in certain traffic distribution conditions, and we saw that it was very useful to assess and evaluate traffic properties in networks. Its construction, however, is associated with a requirement for a large amount of information, as some other previous macroscopic models (e.g. Thomson, 1967; Wardrop, 1968; Godfrey, 1969; Herman and Prigogine, 1979). Building an NEF in real life is really costly because one needs to keep track of all trips made and the number of vehicles circulating a network for every time slice considered. The MFD is more feasible as it requires flow and density measurements on network streets. Nevertheless, monitoring all the streets in a city network can be quite costly. In this last part of the dissertation we are looking at how to build MFDs with fewer monitoring resources so traffic engineers and practitioners can actually obtain these macroscopic tools without really high investments, and in some cases with existing monitoring infrastructure in the city.

Our research focuses on the monitoring resources required by cities as they try to implement the MFD for macroscopic traffic control, its main application. The MFD allows traffic engineers to assess traffic conditions (i.e. traffic state) for the whole city at any given time, by measuring flow and density at certain locations. Access control mechanisms can be implemented to prevent the system from exceeding the critical density and becoming congested. However, despite its potential, the use of the MFD outside research arena, in real cities, has been very limited. The goal of this chapter is to evaluate the data requirements for a possible implementation of an MFD control scheme in an urban area. Particularly, we study the accuracy of MFDs created using
only a percentage of the links (i.e. streets). This is especially useful because monitoring resources are often scarce, and most cities do not have access to the large amount of information that is typically associated with the construction of an MFD.

We evaluate several strategies that cities typically use to place fixed monitoring devices (e.g. loop detectors), and compared them with solutions employing all city links. The results show that independently of the strategy used for link selection, a minimum of 25% of network coverage, according to our accuracy methodology, ensures an average error in density ratios below 15 percentage points (ppts). Based on the particular case of the city of Zurich, we also analyze the feasibility of implementing an MFD control scheme with the links that are currently monitored. Results are very encouraging, showing an average error below 9 ppts. Although all results were obtained with a VISSIM microsimulation model of the inner city of Zurich, we believe the knowledge and methodology presented here can be transferred to other urban areas. In fact, we are hopeful that this research can contribute to making the implementation of an MFD control scheme feasible for many cities.
Chapter 6

Study on the number and location of measurement points for an MFD perimeter control scheme

This chapter is based on the following research paper:


6.1. Introduction

It is known that the trips carried out in a city, under certain conditions of homogeneity, can be related to the accumulation of vehicles in that urban area (Geroliminis and Daganzo, 2008). In addition, it has been demonstrated that the number of trips is proportional to the average flow on the links (i.e. streets) of the network. Therefore, a well-defined curve, the MFD, can be obtained by aggregating measures of flows and densities of, e.g. loop detectors. That relationship allows traffic engineers to assess traffic conditions (i.e. traffic state) for the whole city at any given time, by measuring flow and density at certain locations. Access control mechanisms can be implemented to prevent the system from exceeding the critical density and becoming congested. The MFD proposed in Daganzo and Geroliminis (2008) is a theoretical curve, mainly, because it requires a homogeneous spread of congestion across the urban network. Many scientific works (e.g. Mazloumian et al., 2010; Geroliminis and Sun, 2011; Daganzo et al., 2011; Mahmassani et al., 2013; Saberi et al., 2014) have described the different phenomena that cause the real aggregate flow-density curves to be under the upper theoretical bound, the MFD. For the sake of simplicity, however, we will refer to the obtained curves as MFDs although, in reality, there is only one curve that should be
called MFD, and the ones we will use here are probably just approximations to that real one.

The general goal of this chapter is to analyze the data (monitoring resources) required for a possible implementation of an MFD control scheme in an urban area. Given that many cities worldwide do not have enough access to floating car data yet, this study is limited to data obtained with fixed monitoring devices (e.g. loop detectors, video cameras). We study, particularly, the accuracy of MFDs created using only a percentage of the links (i.e. streets) selected according to different strategies. We create a complete MFD (obtained with information from all the links), and incomplete MFDs (obtained with information from only $p$ links). We then develop a methodology to assess the accuracy of the incomplete MFDs in relation to the complete MFD. Other research works have also looked at the accuracy of incomplete MFDs with data from links (Keyvan-Ekbatani et al., 2013) or from probe vehicles (Gayah and Dixit, 2013). However, the current work presents a systematic and generalized choice of links and a specific methodology to assess the MFD accuracy.

Herein, when we talk about location, we refer to the streets to be measured but not the exact location of the detector or the number of them within the street. That topic has been studied by other researchers (e.g. Buisson and Ladier, 2009; Courbon and Leclercq, 2011). The optimal placement of the detector within the street might be affected by queues from traffic signals, and other causes of local inhomogeneities. Here, we assume that we can properly measure flow and density in certain streets using the link evaluation function from VISSIM. This could be equivalent in reality to placing several detectors in the same link, or using other fixed measurement technologies (e.g. video cameras that track long sections of the link). The question is then: given that a certain number of streets are monitored, what accuracy can we expect from the resulting MFD?

A similar question has been already studied for determining the optimal location of detectors to achieve accurate origin-destination (OD) matrixes (e.g. Yang and Zhou, 1998; Ehler et al., 2006; Fei et al., 2007). In that case, the methodologies involve analyzing many different combinations of links by simulating traffic with an OD matrix obtained with the data from the detectors. Computationally, the problem is very costly. Luckily, our traffic inputs and hence the traffic simulation do not depend on the link selection, and the simulation needs to be done only one time. Nevertheless, the costs of building an accurate model can be very high and that reason justifies our research.

For creating the incomplete MFDs, we use different approaches. First, we formulate a minimization problem to find a quasi-optimal solution. Using a Tabu Search algorithm, we find the best combination of $p$ links to maximize the accuracy of the resulting
incomplete MFD. Although this is not a feasible alternative for a real implementation, it is used here to find an upper bound for the accuracy given $p$. Second, in order to account for more realistic situations, we introduce four strategies that represent common practices for placing loop detectors in urban areas. These strategies typically do not need any traffic information a priori. The goal is to emulate the type of data that cities may have based on different common schemes for placing loop detectors (e.g. closer to the center, in the biggest streets, upstream of signalized intersections).

This research is applied to the case of Zurich. We employ a microsimulation model of the inner city of Zurich (Heimgartner, 2012) where we evaluate the different selection strategies. The city of Zurich currently employs an adaptive control system, ZuriTraffic, as a macroscopic operational scheme for reducing congestion in the central area. The system constantly monitors several links in the inner city. It also estimates, through a demand model elaborated in 2007 (Kanton-Zurich, 2007), what proportion of the traffic on each of the monitored links comes from the different roads accessing the city. When there is a drop in the level of service (LOS) on the monitored inner city links, the traffic lights on the roads accessing the city (and supplying the traffic to those specific links) are adjusted automatically.

This system represents a clear step towards more efficient urban operations. However, it also presents some drawbacks: a) it is based on a static demand model that needs to be frequently updated, b) it defines only two traffic scenarios, congested and uncongested, and c) it only measures the individual traffic state of some streets. For these reasons, we believe that a control scheme based on the MFD can bring further benefits to the city. The theoretical MFD relationship is independent of the demand (Geroliminis and Daganzo, 2008), it continuously assesses the traffic states within the city and can adapt easily to the capacity and traffic requirements, and it represents the aggregate traffic conditions in a city, and not some links individually. In order to evaluate (and possibly improve) the current ZuriTraffic scheme used in the city of Zurich, we create individual MFDs corresponding to each of the monitored links. We also create an MFD aggregating the data from the 23 ZuriTraffic links. The objective is to demonstrate that an MFD control scheme could be feasible in Zurich, and it would not require (in terms of data) any additional monitoring resources, besides those already used by ZuriTraffic.

The rest of the chapter is organized as follows. First, we introduce the methodology for creating both the complete and the incomplete MFDs, and the different strategies to select links to create incomplete MFDs. Second, we present the results of applying those methodologies to our case study: the inner city of Zurich. Finally, we summarize the findings and describe some possible extensions of this research.
6.2. Methodology

6.2.1. Creating an MFD with $p$ links

To calculate the points of the MFD, we find the weighted flow and density averages across all the streets of the urban area chosen. A point in the MFD is defined by an average weighted flow, and an average weighted density (Geroliminis and Daganzo, 2008). These average flows per lane, $q_t$, and average densities per lane, $k_t$, are obtained for every time slice $t$, e.g. every 5 min. $q^i_t$ and $k^i_t$ correspond to the total flow and density of link $i$ for time slice $t$, $m^i$ corresponds to the number of lanes on link $i$, and $l^i$ corresponds to the length of link $i$.

$$q_t = \frac{\sum_i q^i_t l^i}{\sum_i m^i l^i}, \quad (6.1)$$

$$k_t = \frac{\sum_i k^i_t l^i}{\sum_i m^i l^i}. \quad (6.2)$$

Let us consider the case of two aggregate flow density relationships. One of them is what we call complete MFD since it is the diagram found using data from all the links. The other one is what we call $p$MFD (i.e. incomplete MFD) because it has been created in the same way but using data only from $p$ links (with $p<Z$, $Z$ being the total number of links). For every time slice, the flows and densities in the streets are averaged according to (6.1) and (6.2). The difference is that whereas in the complete MFD all the links are used to compute that average, only $p$ links are used for the $p$MFD.

6.2.2. Evaluating the accuracy of a $p$MFD

Given a $p$MFD created with $p$ links, here we evaluate how accurate it is compared to the complete one (the MFD). To do so, we propose to use density ratios, as we believe that absolute values are not as relevant for the implementation of realistic macroscopic traffic control strategies (e.g. access control strategies). Two systems might have similar capacities or critical densities but, due to inhomogeneities in traffic, different behaviour (e.g. different times when they reach congestion). This might render a monitoring scheme for perimeter control either useless or highly inaccurate if it only considers absolute values. The ratio between the density at a certain time slice and the critical density indicates how close or far the system is from congestion. A 100% accurate $p$MFD according to this methodology is one that presents exactly the same density ratios as the complete MFD for each time slice. Employing the reservoirs’ simile used by Daganzo (2007) would correspond to comparing how full (in % terms) each...
reservoir is at every time slice. One disadvantage of using this accuracy method, however, is that it does not provide an absolute value for the number of cars that can be let in or out of the system when using perimeter control. In other words, extrapolating the number of cars to enter or exit an area using a 100% accurate $pMFD$ (according to our methodology) might not be realistic. To overcome this, the control strategy needs to be calibrated to see how it affects the $pMFD$. This calibration should relate the flows entering and exiting the city with the different traffic states in the $pMFD$.

Considering a point, $a$, with density $k_a$, in an MFD with critical density $k_{cr}$ and jam density $k_j$, we define the uncongested density ratio, $R_{un}$ (if $k_a < k_{cr}$), and the congested density ratio, $R_{con}$ (if $k_a > k_{cr}$), as follows:

$$R_{un} = \frac{k_a - k_{cr}}{k_{cr}}.$$  \hfill (6.3)

$$R_{con} = \frac{k_a - k_{cr}}{k_j - k_{cr}}.$$ \hfill (6.4)

For a given sample of traffic states, $k_{cr}$ is calculated by averaging the densities associated with the maximum flow values (top 1%). $k_j$ is calculated by averaging the maximum density values (top 1%). Let us now compare two MFDs. For the same time slice, $t$, one density point is generated in each of them: $k_t$ in the MFD and $k'_{cr}$ in the $pMFD$. Consider that $k_{cr}$ and $k'_{cr}$ are the critical densities, and $k_j$ and $k'_{j}$ the jam densities, of the MFD and the $pMFD$, respectively. Then, the error in density ratios for the time slice $t$ is:

$$\Delta R_t(MFD,pMFD) = \begin{cases} 
\frac{k_t - k_{cr}}{k_{cr}} - \frac{k'_{t} - k'_{cr}}{k'_{cr}} & \text{if } k_t < k_{cr} \\
\frac{k_t - k_{cr}}{k_{cr}} - \frac{k'_{t} - k'_{cr}}{k'_{cr}} & \text{if } k_t > k_{cr}.
\end{cases}$$ \hfill (6.5)

Since this is a difference of ratios (i.e. percentages), the resulting units of (6.5) are percentage points (ppts). Note that the value that determines which density ratio to use (uncongested or congested) is the density of the complete MFD and not the incomplete one. Finally, we can average the absolute errors for all the time slices to obtain what we define as the average density error between the two MFDs:

$$\overline{\Delta R}(MFD,pMFD) = \frac{1}{u_t} \sum_t |\Delta R_t(MFD,pMFD)|,$$ \hfill (6.6)
where $u_t$ is the total number of time slices considered ($u_t$ must be equal for the MFD and the $p$MFD). The same concept of ratios could be applied to the flows. However, in all the scenarios we have tested, the error in flows is always less restrictive than the one in densities. This is reasonable as a flow value (and ratio) can represent two traffic states (congested and uncongested), whereas a density value represents a unique traffic state. Therefore, through the remainder of this chapter, we refer only to the error between density ratios.

### 6.2.3. Strategies for creating incomplete MFDs

#### The optimal $p$MFD

Given all possible links to build a $p$MFD, it is possible to select the combination of $p$ links that minimizes the average density error, hence that creates the best possible $p$MFD (according to our methodology). This minimization problem has the following objective function:

$$\min \Delta R(MFD, pMFD),$$

subject to:

1. the $p$ links are included in $Z$, the total number of links;
2. the $p$ links are not repeated; and
3. all time slices, $u_t$, are used in the comparison.

#### Quasi-optimal selection of $p$ links

Since the enumeration of the $\left(\begin{array}{c} Z \\ p \end{array}\right)$ combinations might be computationally infeasible (e.g. for the Zurich case with $Z=457$ and $p=23$ links, it would be necessary to evaluate $3.3e38$ combinations), a metaheuristic method based on the Tabu Search (Glover, 1989) is developed employing the software MATLAB. This method is not exact, but it approaches the optimal solution, and it is computationally very efficient. To find the exact solution, we would need to develop an exact method (e.g. based on branch and bound techniques), but in this case the potential gains in precision do not justify the additional computational efforts.

The Tabu Search method aims at finding the optimal solution by creating permutations in the neighborhood of an initial solution. In contrast with greedy or heuristics methods, Tabu Search has some mechanisms to ensure a better and more exhaustive search in the domain.
presented in Figure 6.1a. Our movement to create neighborhood permutations is a simple swap of links. On each iteration, \( p(Z - p) \) neighborhood permutations are evaluated (Figure 6.1b). To ensure a proper search in the whole domain, the method establishes several strategies to avoid cyclic movements. It assigns a Tabu tenure to links that have been recently swapped to avoid being chosen again within a certain number of iterations. The Tabu Search allows hill climbing, so the solution might get worse (i.e. not improving the local minimum) during certain number of iterations. If the local minimum is not improved in a predefined number of iterations, the system is restarted, taking a random solution to diversify the search. For more information on Tabu Search algorithms, see Glover (1989).

![Figure 6.1. a) Tabu Search structure employed in the search of the quasi-optimal solution. b) Example of the neighborhood permutations.](image)

**Blind strategies to select the \( p \) links**

The search for a quasi-optimal \( p \) combination of links to generate a \( p \)MFD is only possible because we have full traffic information a priori (i.e. densities and flows for all links in the urban area). Evidently, this is not a feasible alternative in reality, as most cities only count with a limited amount of traffic information. Hence, here we examine some blind strategies (i.e. selections of \( p \) links based on affordable information). We present four simple blind strategies that follow the typical rationale for placing loop
detectors in urban areas. The first blind strategy is presented just for comparison purposes; the other three are often used by city traffic engineers to determine the location of loop detectors independently of MFD control schemes. The four blind strategies will later be evaluated based on a microsimulation of the city of Zurich.

- **Totally random selection:** the selection is performed randomly. In our model, every link is given a random number between 0 and 1, and the $p$ highest values are selected. This strategy is used as the basis for comparison.

- **Distance to the center selection:** the selection is carried out giving more importance to more centric streets (a common practice in many cities). The concept does not refer purely to the geometric center but to what planners understand as the most important area in terms of traffic. The Euclidian distance to this central area is calculated and standardized. The selection of the links is still, to a certain degree, random. The standardized values are multiplied by a random 0–1 number. The $p$ highest resulting values are selected.

- **Hierarchy of the street selection:** the selection is carried out giving more importance to bigger streets (also a common practice that leads to the placement of detectors in the biggest and most important streets). We use as an indicator of hierarchy the number of lanes that the streets have. The number of lanes might change along the street, so we consider the highest number of lanes a link has throughout its length. We select only the streets with more than one lane. Among these streets, the selection is random. As before, we multiply them by a random 0–1 number. The $p$ highest resulting values are selected.

- **Signal selection:** the selection considers only the streets that have a traffic signal downstream. Typically, cities locate loop detectors upstream of signalized intersections to control them. Hence, there is a high probability of finding already existing loop detectors in this type of settings. Among the streets with a signalized intersection downstream, the selection is random. As before, we multiply them by a random 0–1 number. The $p$ highest resulting values are selected.

6.3. Results. Case study: the inner city of Zurich

In this section, we evaluate the quasi-optimal solution, the four blind strategies mentioned above, and the ZuriTraffic control scheme, with a case study of the inner city of Zurich. The main input to our research is the measurement of flows and densities in the streets of Zurich. These data inputs are based on a traffic microsimulation model implemented in VISSIM 5.4, and representing the central area of Zurich (around 2.6 km$^2$). In this model, several transportation modes are included. For our purposes, we consider only cars, vans, trucks and buses. Trams are excluded as they circulate on specific tracks and for the most part on separate lanes. The demand corresponds to the 5–6 pm afternoon peak on a working day. The VISSIM model emulates the signal
prioritization scheme for public transportation that the city has implemented. This model has been calibrated using flows from vehicle counts (Heimgartner, 2012; Menendez and Ge, 2012).

The network is filtered so only the links with flows larger than zero are considered. Connectors are excluded. In those cases where a street is composed by several links, all of them are grouped as one link as long as there is not a source or a sink (i.e. demand origin or destination) in the middle of the street (in this case two different links are counted). Each link only represents one movement direction, hence bidirectional streets have two links. In total, we have 457 filtered links.

Our simulation study only considers the afternoon peak. But to obtain an MFD, we need to have a representation of many different traffic states. That is why, keeping the same OD pattern, we scale the level of trips considering 25 factors from 0.1 to 2.5. We are aware of the limitations of this scaling since the route choice remains the same. In other words, drivers are not adaptive and do not change their routes in response to congestion. The implications of this assumption (often used in this kind of simulation) can be found in Daganzo et al. (2011) and are illustrated below. For every demand loading, we run the simulation with five different random seeds to cover different cases. In total, we carry out 125 one-hour simulations.

6.3.1. The complete MFD of Zurich

We create the MFD of the inner city of Zurich employing (6.1) and (6.2). Figure 6.2a shows the resulting MFD. Although overall it seems well-defined, it is noticeable that the congested branch presents more scatter than the uncongested branch. In addition, it shows a steep drop to the jam state. There are two reasons for such scatter and drop.

The first reason is the inhomogeneity of traffic distribution across the links. Although the urban network considered is fairly small and streets have similar characteristics, the original data of the simulation (demand from 5 to 6 pm) show that congestion is distributed inhomogeneously (i.e. for every time slice, there is a large variability in densities). According to Mazloumian et al. (2010), this might be due to inhomogeneity in OD patterns, road infrastructure, and traffic control. The demand modeled only accounts for a one-hour period (5–6 pm) in the inner city of Zurich. In addition, there are a large number of interactions with other modes (buses, trams, pedestrians) that can significantly affect traffic on different areas of the city. As some research works indicate (Mazloumian et al., 2010; Geroliminis and Sun, 2011; Knoop et al., 2012), the inhomogeneity implies that the network is not as efficiently used as it could be. As a result, the traffic states we obtain are lower states (in terms of production rate) than the ones in the theoretical MFD (when congestion is distributed homogeneously).
The second reason is related to the traffic simulation itself. As analyzed by Daganzo et al. (2011), most of the simulators do not consider drivers’ adaptation to traffic conditions (they have a fixed route choice for all demand levels). If the loading speed (i.e. the demand or trip levels) is increased, traffic states with higher flows are obtained at the beginning, but since there is no adaptation, some streets gridlock whereas some others are still empty. When that happens, there is a sudden drop in flows (Figure 6.2b). If drivers could adapt (using alternative routes and hence spreading more homogeneously) the system would support a higher density before collapsing. The experiences with real data (e.g. Yokohama in Geroliminis and Daganzo (2008)) show that real drivers most likely adapt, so the jam density takes higher values.

Summarizing, the theoretical MFD is an upper bound of all the curves obtained with real or simulated data because the conditions of homogeneity are not always fulfilled. Therefore, it is reasonable to assume that the real MFD for the city of Zurich could be wider (i.e. it would have a higher jam density). Nevertheless, for the remainder of this chapter, the complete MFD shown in Figure 6.2a will be considered as the ’true’ complete MFD for comparison purposes.

### 6.3.2. Quasi-optimal selection strategy

Based on the Tabu Search algorithm explained in the Section 6.2.3 here we find the quasi-optimal solutions for different monitoring/coverage levels (covering between 5 and 40% of the total links). Each execution was terminated after 250 iterations. We
allowed 20 hill-climbing iterations (i.e. iterations that do not improve the local minimum) before restarting the initial solution. In our case, the restarting process changed only 20% of the links randomly. In the swapping process of finding neighborhood permutations (Figure 6.1b), a tabu tenure of five iterations was imposed to the links that were deselected, and a tabu tenure of six iterations was imposed to the links that were just selected. The comparative results to the complete MFD are presented in Figure 6.3. With only 5% coverage, if links are selected quasi-optimally, it would be possible to achieve an average difference in density ratios of 3 ppts. This difference is reduced to less than 1 ppt if 40% of the links are selected. Although the quasi-optimal selection is generally infeasible in real life, the calculations here give us a lower bound for the possible error (or upper bound for accuracy) that can be obtained with different levels of coverage.

Figure 6.3. Accuracy vs. network coverage of the quasi-optimal solution in comparison to the complete MFD.

6.3.3. Blind selection strategies

Here, we evaluate the accuracy of the four blind strategies described in the Section 6.2.3. To do so, we look at different coverage levels (between 5 and 40%). Each strategy has a random component, therefore, 40 different cases have been considered for each level of coverage and strategy. For the distance selection strategy, we considered two locations for delimiting the central area (in terms of traffic movements): the main train station (Zurich HB), and the tram and bus Station Bellevue; so we calculated the distance to these two centers, and picked the smallest of the two. Figure 6.4 shows the average, minimum, and maximum error in density ratios for the random, distance, hierarchy, and signal selections.
The average line (which is an average of the 40 different selections for each coverage level) is almost always below 15 ppts. As expected, this value decreases when coverage increases. In fact, when 35–40% of the links are covered, this value is approximately 5 ppts for all strategies. All the strategies present very similar results, at least concerning the average line. Therefore, it seems that there is no benefit in discriminating certain links concerning distance, hierarchy or if they are located upstream of a traffic signal, since a random selection can provide similar or even better results (for some cases). Notice that none of the tested scenarios with the four blind strategies approached the quasi-optimal solution. The difference between the blind strategies and the quasi-optimal selection strategy is obviously reduced when the coverage increases. However, with only 5% coverage, if links are selected quasi-optimally, it is possible to achieve similar accuracy as the best case for a 40% coverage when placing links according to any of the four blind strategies.

Here, we also look at the variability of results across different solutions (i.e. the spread between maximum and minimum error). Notice that a distance selection provides consistently less variability. Previous works have demonstrated that
bottlenecks in the center might limit the capacity of more peripheral streets (Zheng et al., 2012). That could be one reason to explain how links that are closer to the center are more representative and consistent with the overall network traffic states defined in the complete MFD. Nevertheless, none of these strategies should be expected to provide extremely accurate results, as by nature, they do not use representative links. When using any of these blind strategies for implementing real MFD control schemes, due to the limited amount of information that we typically have in real life, we recommend looking at the maximum error as an upper bound. Again, it seems that the distance selection performs the best. Although as a general rule we could say that a 25% coverage should guarantee an error below 15 ppts with any of the four blind strategies.

![Figure 6.5. Accuracy vs. network coverage for random selection considering: a) network coverage in percentage of links, and b) network coverage in percentage of length.](image)

Last, but not least, we have investigated the possible variability in the network coverage definition. So far, for the sake of simplicity, we have considered percentage of links. However, monitoring some links or others might involve different percentages of network coverage in terms of length. Figure 5 shows the errors for a random selection strategy with both definitions. The behavior of both cases is very similar. We attribute that to the fact that the street lengths in the inner city of Zurich are not that diverse, and percentages keep the same proportions.

### 6.3.4. Use of ZuriTraffic links to create a \( p \)MFD

As explained in Section 6.1, the city of Zurich currently employs an adaptive control system as a macroscopic operational scheme for reducing congestion (ZuriTraffic). The system uses the individual measurements of 23 links from the inner city (see Figure 6.6b) to regulate traffic entering and exiting the city. We create here a \( p \)MFD with those links to evaluate the system. Figure 6.6a shows some of the individual fundamental
diagrams for the ZuriTraffic links obtained with the data from the microsimulation (these 4 links were randomly selected among the 23 ZuriTraffic links).

Figure 6.6. a) Individual fundamental diagrams for four different ZuriTraffic links. b) Map of the inner city of Zurich and the location of the monitored streets.
Table 6.1. Average errors between individual ZuriTraffic links and a complete MFD.

<table>
<thead>
<tr>
<th>Link number</th>
<th>Total average error in ppts</th>
<th>Uncongested average error in ppts</th>
</tr>
</thead>
<tbody>
<tr>
<td>104</td>
<td>34.4</td>
<td>15.8</td>
</tr>
<tr>
<td>105</td>
<td>30.8</td>
<td>7.9</td>
</tr>
<tr>
<td>535</td>
<td>26.3</td>
<td>12.4</td>
</tr>
<tr>
<td>393</td>
<td>19.2</td>
<td>13.7</td>
</tr>
<tr>
<td>856</td>
<td>46</td>
<td>15</td>
</tr>
<tr>
<td>658</td>
<td>22.6</td>
<td>13.6</td>
</tr>
<tr>
<td>361</td>
<td>29.6</td>
<td>14.8</td>
</tr>
<tr>
<td>385</td>
<td>70.4</td>
<td>16.6</td>
</tr>
<tr>
<td>384</td>
<td>23</td>
<td>15.6</td>
</tr>
<tr>
<td>896</td>
<td>18.3</td>
<td>19.2</td>
</tr>
<tr>
<td>897</td>
<td>1043.9</td>
<td>17.6</td>
</tr>
<tr>
<td>567</td>
<td>36</td>
<td>14</td>
</tr>
<tr>
<td>874</td>
<td>37.1</td>
<td>22.5</td>
</tr>
<tr>
<td>922</td>
<td>5985.8</td>
<td>13.5</td>
</tr>
<tr>
<td>923</td>
<td>25</td>
<td>28.7</td>
</tr>
<tr>
<td>773</td>
<td>24.4</td>
<td>23.2</td>
</tr>
<tr>
<td>774</td>
<td>70.4</td>
<td>15.9</td>
</tr>
<tr>
<td>828</td>
<td>18.9</td>
<td>16.7</td>
</tr>
<tr>
<td>829</td>
<td>105.3</td>
<td>12</td>
</tr>
<tr>
<td>421</td>
<td>74.8</td>
<td>30.5</td>
</tr>
<tr>
<td>865</td>
<td>82.9</td>
<td>13.8</td>
</tr>
<tr>
<td>864</td>
<td>7233.2</td>
<td>20.6</td>
</tr>
<tr>
<td>262</td>
<td>31.1</td>
<td>29.2</td>
</tr>
</tbody>
</table>

Applying the methodology presented in Section 6.2, we calculate the accuracy of the 23 fundamental diagrams (MFDs with \( p \) links when \( p=1 \)) corresponding to all the individual ZuriTraffic links. The goal is to evaluate how well they represent the traffic states of the city (i.e. how close they are to the complete MFD). Table 6.1 shows the average errors for two scenarios: i) considering all the traffic states from the complete MFD and the individual link, and ii) considering only the uncongested traffic states in the complete MFD, and the corresponding states in the individual link. We evaluate this second scenario because often we are just interested in predicting if we are approaching congestion. In addition, it seems that the fundamental diagrams for the individual links present a lot of scatter in the congested branch, while they are fairly consistent in the uncongested one. Notice that when considering all the traffic states, the average errors are substantially high. Errors for the uncongested part are lower, but still reach values up to 30 ppts. The links with the largest total average error (see Table 6.1, links: 897, 922, 864) correspond to individual links that are really far from reaching capacity (or
getting congested) at times when the overall system (the complete MFD) is congested. Therefore, their fundamental diagrams only have a small uncongested part (Figure 6.6a shows the fundamental diagram for link 922) which gets compared with very congested states in the complete MFD, creating a very large error according to the accuracy methodology employed here. As a result, it is possible that control strategies (i.e. changes in the traffic lights accessing the city) are being implemented either too early or too late if based only on individual links. Early access control could starve the city when it is not really necessary. Late access control would not help to prevent congestion.

![Figure 6.7. MFD with ZuriTraffic links and complete MFD.](image)

We now evaluate how representative are the measurements when aggregating the 23 ZuriTraffic links (i.e. incomplete MFD created only with ZuriTraffic links corresponding to 5% of the links of our network). Figure 6.7 shows both the incomplete MFD created with ZuriTraffic links and the complete MFD.

Using the same accuracy methodology, this combination of links provides an average difference in density ratios of 8.7 ppts. This error is substantially lower than those obtained with the individual approach. Furthermore, when carefully inspected, we see that the largest errors are located in the congested part rather than in the uncongested one. That is due to the shape of both fundamental diagrams. The absolute values of the MFD created with ZuriTraffic are higher, because ZuriTraffic monitors on average the most important streets with higher capacities and critical densities. In addition, the ZuriTraffic MFD is narrower in relation to its critical and jam densities (i.e. the congested branch has a steeper slope). As a result, the largest errors are concentrated on
the most congested traffic states. Averaging only the errors in the uncongested part provides an average difference of 3.3 ppts. This value is again substantially lower than those obtained with the individual approach, and close to the quasi-optimal solution. We believe then that aggregating data from the links that are currently being used provides a fairly accurate MFD and can lead to better control strategies. The largest errors, located at the end of the congested branch, might have no effect on control strategies, since these quasi gridlock states are unlikely to happen in a city, as demand would most likely adapt before then.

6.3.5. Different demand patterns

As previously explained, all the experiments conducted here come from a simulation model that employs the afternoon peak demand (from 5 to 6 pm on a working day). Although different demand factors were employed, the demand pattern was kept the same. Theoretically, the MFD, understood as the upper bound of the aggregated flow density curves, is independent of the demand since it requires homogeneity across the city. However, as explained earlier, our complete MFD curve is not the theoretical upper bound since we observe variability in the spread of the congestion (drivers do not adapt, and hence, the city could in theory accommodate more demand). Thus, a valid question is: are the results shown above still valid for a different demand pattern?

To evaluate the sensitivity of our research findings to the demand pattern employed, and with the limitation of not having another calibrated demand pattern, we generated five random demand scenarios (changing randomly the number of trips exchanged in the different OD pairs, but keeping the same overall number of trips). The results show large discrepancies between complete MFDs (from 13 up to 35 ppts error compared to our original MFD). This would indicate that demand patterns do affect the shape of the complete MFD (i.e. the aggregate flow density relationship found with all the links of the city), and hence the link selection combination that works for one specific demand pattern might not be valid for other ones. However, this is not a conclusive proof, as a possible explanation for this error is related to how the demand is changed in the simulation. In the simulation, the route choice parameters are maintained (the ones from the original demand), so drivers do not adapt, not only in the congested branch but in most of the traffic states even when the system should theoretically be uncongested. As a result, the complete (with modified demand pattern) MFDs show also some scatter in the uncongested branch. In other words, the MFD with the original demand is, in the uncongested branch, an upper bound of the other MFDs. We should stress that the MFDs created with the other demands present systematically lower flow traffic states although the total number of trips for all the cases is the same.
For the sake of completeness, and despite the drawbacks explained above, we also evaluated the accuracy of the ZuriTraffic MFD for every new demand pattern. Errors ranged between 8.6 and 33.9 ppts. However, due to the same reasons stated above, it is hard to draw any conclusions. More realistic demand patterns should be used to further study the impact of the demand on the link selection strategies.

In the case where different demand patterns create multiple MFDs, or if traffic becomes very inhomogeneous within the city; one possible solution is to further divide the study area into different spatial (more) homogeneous clusters. For that, Ji et al. (2012) proposed a partitioning algorithm to group streets with similar density characteristics.

### 6.3.6. Applicability and limitations

In this chapter, we have studied the possibility of creating a $p$MFD for control purposes using only $p$ links (i.e. a portion of all the links). Our results show that this is actually possible, which could lead to substantial savings in monitoring costs. That is encouraging especially for practitioners, who could now in theory achieve a macroscopic control in cities at a lower cost. The strategies studied here for link selection are based on the premise that the monitoring resources are already in place (i.e. monitored links have been selected without having traffic information a priori). This research work, however, still has some limitations:

- The inhomogeneity of the congestion in our simulation model (based only in the afternoon peak, from 5 to 6 pm) does not guarantee that the results are independent of the demand patterns.
- The transferability of these findings to other cities remains unclear, as only data from the city of Zurich have been analyzed.

Moreover, the differences between the blind strategies and the optimal strategy raise the possibility of the existence of other more advanced strategies that could yield bigger gains.

As previously seen, the blind strategies provide reasonable accuracy but still with a high variability and, results do not seem to change much from one strategy to another. This, as previously explained, is not unexpected, as by nature these strategies do not use representative links. Instead, the ZuriTraffic links selection, with only 5% of the links, performs much better. These findings motivate the search for more advanced blind strategies.
6.3. Conclusions

In recent years, the existence of the MFD has generated a lot of research activity, mostly on the theoretical side and not in the real implementation side. One of the reasons is the scarcity of monitoring resources that make it impossible for most cities to have access to the large amount of information that is typically associated with the construction of an MFD. The goal of this research was to evaluate the basic data requirements for a possible implementation of an MFD control scheme. Particularly, we studied the accuracy of MFDs created using only a percentage of the links selected according to strategies typically used by practitioners to place loop detectors or other fixed monitoring devices.

We developed an accuracy methodology to compare two aggregate flow-density relationships: on the one hand, the complete MFD with data from all the links, and on the other hand the incomplete MFD ($p$MFD) with data from only $p$ links. This methodology, based on density ratios, is simple and focuses on the validity of a $p$MFD as a control mechanism, i.e. to evaluate if the incomplete MFD is able to explain the evolution of congestion as accurately as the complete MFD. With that methodology, we created a quasi-optimal search algorithm that finds the best selection of $p$ links. The results from this algorithm are employed as a lower bound for evaluating the different selection strategies. All these comparisons are possible because we employ a VISSIM microsimulation model of the inner city of Zurich. We assume (using the link evaluation function of VISSIM) that we are able to properly measure the flow and density in links, although in reality achieving that might suppose using the average from several loop detectors or data from other monitoring sources.

Most cities do not have that much traffic information a priori, so ideally the selection strategies should be supported only by affordable information. We evaluated different $p$ link selection strategies to create an MFD based on common criteria used by cities to place fixed monitoring devices. Four typical blind selection strategies were considered: total random selection (for comparison purposes), distance to the center selection, hierarchy of the street selection, and downstream traffic signal presence selection. The results show that independently of the strategy used for link selection, a minimum of 25% of the links ensures an average error in density ratios below 15 ppts. The distance selection not only provides good results but also maintains a low variability, especially if at least 15% of the links are selected. Selecting quasi-optimally the links brings substantial benefits and savings in detection points. We see how with a 5% coverage of links selected quasi-optimally, we can achieve the same accuracy as the best cases with random selection and 40% coverage.
In this chapter, we also analyzed the ZuriTraffic control scheme, used nowadays by the city of Zurich. The system, although very innovative, presents three main drawbacks that we believe could be addressed by converting it to an MFD control scheme. Results show that by aggregating the data from the ZuriTraffic links (instead of analyzing them individually), we are able to observe better when the network goes from uncongested to congested states. As manifested earlier, the use of these links individually does not necessarily reflect the traffic state of the entire system, the MFD does. In addition, the MFD control scheme would not be dependent on demand models, and would provide more flexibility in the control scenarios. As a matter of fact, the city would not even need to collect more data, just to use the already collected data in a different way. Decisions should be made on the aggregate data rather than on the individual links. Using the 23 ZuriTraffic links (5% of the total) to create a pMFD yields an average error in density ratios of 8.7 ppts, significantly below the error from any of the four blind strategies we evaluated.

These results emphasize the need for researching into more accurate advanced blind strategies. Better strategies could help further reduce the monitoring requirements for the implementation of MFD control schemes. In addition to this, more experiments with other cities are needed to verify the transferability of the results, and to evaluate the effects of different demand patterns on the selection strategies. Finally, we think that further research on combining these data from fixed monitoring devices with floating car data could create MFDs more efficiently.
Chapter 7

Conclusions

Overall, we believe that this PhD dissertation brings traffic engineering and urban planning perspectives together for a better design of streets and patterns; and a deeper understanding of traffic operations in urban areas. We claim that urban planners need to incorporate traffic engineering concepts in the urban design process. At the same time, we encourage traffic engineers to tackle urban planning issues employing traffic analysis perspective and modeling tools.

Research-wise, the main contribution of this dissertation is an in-depth analysis of different urban planning problems from the traffic operations perspective, providing further understanding and recommendations in a simple and pragmatic manner. The first part is devoted to the study of different street configurations on grid networks. The second part deals with the effects of space removal on grid networks. The third part studies the implementation in real cities of macroscopic traffic schemes. Through the 6 main chapters of this dissertation we have employed a wide range of analysis tools to examine problems from the traffic engineering perspective. To the author’s knowledge, this is the first research work to combine 3 completely different traffic assignment methods to analyze, for example, the street configuration problem. Because of that, we have a more complete and detailed view of networks’ features. Another distinctive characteristic of our research is the use of traffic engineering models and techniques as indicators to evaluate network properties. These indicators are very helpful to compare networks’ performance. In addition, we believe that the work presented in the second part of this thesis is novel on evaluating road space removal (as a result of an urban policy) in a systematic and abstract way. Finally, we have studied the implementation of macroscopic control schemes in cities based on the MFD, and how practitioners could actually monitor traffic employing more efficiently the city’s resources. This is especially useful for cities as it bridges the theoretical research concepts to the real applications.

Our research mainly focuses on the analysis of urban grid patterns which are the most common patterns in planned cities. We believe that the abstract character of our
experiments can contribute to more transferable and understandable findings for researchers and practitioners. However, the third part of this dissertation does include a real case: the city of Zurich.

In this research work, we do not considered the induced demand. Instead, we focus on the response of networks to the same demand load—i.e. the supply side. This, evidently, is a conservative approach, representing approximately an upper bound for the negative effects that could be expected from those policies. Such perspective should help us reduce the alarmist thought that traffic will become extremely chaotic (Cairns et al., 2012) if road space is removed. Any other scenario, considering e.g. a demand reduction, would probably yield better traffic performance and less negative outcomes.

The dissemination to society of our findings is a priority for us. Notice that throughout the dissertation we put special relevance on the implications that certain findings might have in practice. In this last section we summarize all research findings as well as the recommendations oriented to urban planners and traffic engineers. Finally, we propose possible research extensions of this dissertation.

7.1. Research findings

The first part of this PhD thesis focuses on the study of different street configurations in finite urban grids. This study covered the debate between two-way streets and one-way streets. In addition, it analyzed an intermediate solution, two-way streets with prohibited left turns.

The analysis was carried out employing three different methods: analytical formulations, Static Traffic Assignment, and Dynamic Traffic Assignment based on microsimulation. The three methods present very different traffic assignment levels of complexity, computational times, and evidently research outcomes. Employing all of them makes this research robust and results more complete. Analytical formulations are simple and basic, but not that accurate. On the other hand, they reveal behavioral trends in a clearer manner. The opposite case is the microsimulation tool which captures many traffic interactions but also requires more efforts in computing and interpreting results.

First, TW networks provide the shortest distance traveled between two points and the highest route redundancy and flexibility. However, they present the lowest vehicle capacity at intersections which penalizes networks heavily. Second, TWL networks offer the best trade-off between distance traveled and capacity at intersections. They are the best option for many different scenarios. The main disadvantage is that they have a very limited route redundancy which translates into more heterogeneous spread of congestion. That becomes a problem in highly congested scenarios, when drivers are
routed through different paths, or when demand is heterogeneous. Last, OW networks supply the highest capacity per movement at intersections although they also trigger the largest distance traveled. The route redundancy of OW networks lies between TW and TWL networks, making them very suitable to highly congested cases, heterogeneous demand patterns, or with diversifying routing options.

The second part of this PhD thesis dealt with the removal of space in urban settings. Our aim is to understand and quantify how this removal affects drivers and the overall system. We also studied abstract grid networks representing urban environments. First, we looked at different strategies of link removal in the grid networks; later, we studied the effects by only removing lanes. Same as in Part 1, we employed two methods to model traffic: the STA and the microsimulation with a DTA module. This diversification allows us to take advantage of their potentialities and compensate their weaknesses.

Results clearly show that it is possible to remove certain amount of links from an urban grid without losing connectivity and without worsening traffic conditions excessively. Evidently, traffic impacts depend on the removal strategy. Removing streets from the center creates the highest impacts because it affects the redundancy of routes, and hinders traffic flows from spreading more homogeneously. In that case, intersections with really high traffic loads will easily trigger congestion in the system. Removing streets from the perimeter, instead, allows the system to keep the high connectivity in the center that allows spreading traffic loads more evenly.

Lane removal does not change the network structure (i.e. connectivity) as link removal does, but it affects the traffic distribution in the system. Removal scenarios always present more heterogeneous distributions, not only than the initial case, but, for some cases, also than the final scenario: the full one-lane network. Results show that central removal impacts the system more than peripheral removal in terms of capacity drop, congestion appearance and gridlock time. Congestion propagation depends on two factors: the number of one-lane streets in networks, and the origin of congestion. If congestion is originated in the perimeter, it spread over the network faster than if congestion started in the center. Although networks do not present clear congestion propagation patterns, central removal has a tendency to propagate congestion more from the center, whereas peripheral removal has a slight disposition to propagate congestion more from the perimeter.

The third part of this dissertation looked at the amount and location of monitoring resources (i.e. streets to be monitored) to create accurate MFDs to support traffic control mechanisms in urban areas. In recent years, the existence of the MFD has generated a lot of research activity but it has not yet been translated to real
implementations. We believe that one possible reason is the significant amount of data that these models require.

We developed an accuracy methodology to compare two aggregate flow-density relationships: on the one hand, the complete MFD with data from all the links, and on the other hand the incomplete MFD (pMFD) with data from only $p$ links. This methodology, based on density ratios, is simple and focuses on the validity of a $p$MFD as a control mechanism, i.e. to evaluate if the incomplete MFD is able to explain the evolution of congestion as accurately as the complete MFD. We proposed some blind strategies for monitoring specific links without a priori traffic information, and we compared them with a quasi-optimal search algorithm that finds the best selection of $p$ links (the upper bound in terms of accuracy). Results show how these blind strategies can actually produce accurate MFDs using less monitoring resources.

7.2. Recommendations

In this section, we describe the practical implications that our research findings may have. We summarize a set of recommendations targeted to urban planners, traffic engineers, and other practitioners involved in one of the two topics covered by this dissertation.

There is not a perfect urban street configuration. Every network configuration studied presents particular advantages and disadvantages. Urban planners should then consider which configuration fits their purposes better. The main advantage of TW networks is that they minimize distance traveled and they offer a lot of routing flexibility. However, they handle congestion poorly. These networks could be employed in small cities but also in city centers as long as there are mechanisms to control the number of vehicles entering the area. The flexibility in routing options makes TW networks appropriate for urban areas where many disruptions may occur, and for incomplete networks where connectivity is a relevant issue. In addition, this configuration is suitable for long street networks where distance traveled is determinant and spillbacks are restrained due to the length of the links.

TWL networks can be applied to many different urban scenarios from low to moderate congestion levels. They are useful for urban settings where cities want to keep the two-way configuration and maintain at the same time a good level of capacity at intersections. There are already many cities that implement local prohibition of left turns to improve capacity at intersections. In North America, for example, there are cases of left turn prohibition during peak hours (e.g. Toronto, Vancouver, Minneapolis), but there are not, to the author’s knowledge, cases of extensive TWL networks. By doing so, cities could also take advantage of the TWL network effects studied in this
dissertation. In addition, TWL could be convenient for cities with high presence of two-
way public transportation corridors, especially if capacity for vehicles is further reduced
due to public transportation prioritization schemes. For example, the city of Barcelona
has a one-way grid with a high presence of public transportation, especially buses.
However, the fact that all streets are one-way segregates public transportation stops and
makes transfers more difficult. TWL networks need certain uniformity on routing
and demand, hence they are appropriate for homogeneous urban systems that are not
prone to have many traffic disruptions.

OW networks are the best option for heavily congested cities where capacity at
intersections is significantly more important than distance traveled. This configuration
is optimal for networks with short streets and high traffic loads. They are also robust to
demand changes, which makes them suitable for dynamic and changeable cities. They
are the only configuration capable of providing one-lane streets without losing
connectivity. Therefore, they are suitable for “slimmer” scenarios where urban planners
can recover space for other activities, e.g. networks with high presence of public
transportation, bike lanes, pedestrian areas, or on-street parking. That is the case of one-
way grid networks like Mannheim (Germany) or Glasgow (UK) where there are many
streets with only one lane devoted to cars and the rest of the road space is devoted to
other activities.

In cities, actions towards reclaiming road space for other transportation modes and
activities are more common. Cities like San Francisco (Billings et al., 2013),
Copenhagen, or London (European Commission, 2004) have actually recovered
successfully space from cars to other activities and transportation modes. We have seen
how link and lane removal actions can be carried out in urban areas without worsening
traffic conditions excessively. If minimizing traffic impacts is the priority for planners,
peripheral removal is the most suitable strategy. It is the most conservative approach as
the center, which handles and spreads traffic flows more homogeneously, remains
untouched. However, this minimization of traffic impacts on the system comes with an
additional price: the increase of traffic flows in the center. The paradox is that although
traffic impacts are minimized, an urban planning measure focused on improving urban
space quality might end up loading more the city center.

Removing links and lanes in the center creates, undoubtedly, a higher traffic impact
in the short term. Nevertheless, a central removal can also be used as: a political
statement, to control the number of cars in a system employing congestion as
disincentive, or as a policy measure to achieve a shift to more sustainable transportation
modes. In Barcelona, for example, in recent years, several lanes where converted to
bicycle lanes discouraging road transportation and promoting the bicycle use. Our
results show that central and peripheral removal to be the lower and upper bound strategies respectively. The random link removal strategy, studied in Chapter 4, lies between those too. Obviously, there are many possibilities that fall into that random strategy. However, if links were chosen correctly, traffic impacts might be minimized. To choose these links, we have seen that the key is to maintain flexibility in routes without loading certain intersections excessively.

Although this dissertation does not consider demand modeling, we study which percentage of demand should be reduced to maintain initial LOS traffic conditions while removing lanes. Clearly central removal imposes much harder conditions to fulfill, especially in the first removal stages. For example, a 30% of demand reduction it is needed to only remove 4 lanes in the center of the grid. These values and elasticities are presented in Chapter 5. Despite demand shift values are high, real measurements have demonstrated that a considerable amount of traffic can be reduced when capacity restrictions apply. For example, in Cairns et al. (2002), the median traffic reduction of 70 road capacity restriction cases from different countries was approximately 11%. Additionally, we also see that after having removed a certain amount of lanes, it is not worth continuing these gradual removal steps as the final scenarios may perform equally or even worse than the full one-lane network scenario. In that situation, planners should directly convert networks to the full one-lane case.

The third part of the dissertation dealt with the amount and location of fixed monitoring resources that a city should have to measure reliable MFDs that can feed traffic control mechanisms. This research work is especially important for practitioners. Most cities do not have that much traffic information a priori, so we consider different blind strategies that do not require these data. The results (obtained with simulation data from the city of Zurich) show that independently of the strategy used, a minimum of 25% of the links ensures an average error in density ratios below 15 ppts. The distance selection (i.e. selecting links randomly but with a certain weight towards central activity places, e.g. for the case of Zurich, the central train station) not only provides good results but also maintains a low variability, especially if at least 15% of the links are selected. Evidently, if cities already have traffic information a priori, the selection of links can be done in a very efficient way. For comparison purposes, if data from all links were available, an optimal link selection with only 5% of the links would achieve the same accuracy as a random selection with 40% of the links. Finally, like in the case of Zurich, some cities might already have some macroscopic traffic control mechanisms even if the traffic count coverage is not really high. Employing those same links or streets to create MFDs proved a very good and efficient strategy for the case of Zurich.
7.3. Future research

Finally, in this last section, we detail some of the possible extensions that our research could have.

The abstract grid networks studied in this dissertation are homogeneous, regular, and have a fixed traffic signal configuration. We believe that some of the most interesting extensions to our research could be to consider less ideal and homogeneous conditions. In that direction, researchers could look into quasi-like grid structures, or hybrid configurations between OW and TW networks. In addition, following the example of Eichler et al. (2013), further research could be focused on optimizing grid networks configuration to fulfil certain predefined criteria.

Concerning the space removal in urban areas, there exist other strategies that could be studied which fit in the logic of urban planners and traffic engineers. One possibility would be to remove corridors as a result of creating dedicated bus lanes. This removal strategy not only is feasible and common in many cities, but it can help to further understand car and bus interactions. We believe that removing corridors is a strategy that falls between peripheral removal (lower bound) and central removal (upper bound) concerning traffic impacts. Additionally, researchers could combine both link and lane removal; and apply them to different network configurations. The study of space rearrangement instead of removal, i.e. creating larger grid blocks, is also a promising research extension. On-going research efforts are focusing in that direction.

This thesis has looked uniquely to how different networks deal with the same amount of demand, i.e. we looked at the infrastructure supply properties. We considered that demand remains in the short-term and that is a conservative approach. In the future, demand modeling should be included to provide more realistic medium and long term scenarios. In addition, we believe that the possibility of shifting transportation modes should be included. Representing the mode shift could help to better evaluate the road space trade-offs in urban environments. In that direction, research works like (Zheng and Geroliminis, 2013) that combine mode choice and macroscopic traffic evaluation could be employed to influence urban planning decisions.

The analytical propagation model is a promising tool to study traffic spread behavior. However, the connection between the simulation networks and the model is not that clear. Future work could look into different modifications of the analytical model to fit more realistic propagation behaviors. In addition, researchers could also look at bigger networks and see if congestion propagation patterns are more identifiable there. At the same time, further research can work on describing the growth of the MCC in grid networks depending on factors such as lane removal or demand levels.
Regarding the traffic analysis tools, many new simulation models (e.g. mesoscopic models) are appearing to cover the limitations of classic tools. On the one hand, the STA does not consider traffic dynamics; on the other hand, the microsimulation software has many parameters to handle and has very high computational cost. Bear in mind, that the present research is one of the few research works, to the author’s knowledge, that generates MFDs with a full DTA microsimulation. Future research can be focused in finding more efficient methods for conducting traffic analysis in urban grid networks.

Finally, concerning the third part of this dissertation, our results emphasize the need for researching into more accurate strategies that can bring even more monitoring savings. Monitoring methods are changing, as vehicles and persons carry devices with Global Positioning System (GPS) transmitters. We believe that floating car and passenger data will be the main input for traffic control algorithms. Further research should focus on the combination of these data sources (fixed and floating devices) to have more accurate MFDs.
Bibliography


Barr, L. Testing for the Significance of Induced Highway Travel Demand in Metropolitan Areas. Transportation Research Record: Journal of the Transportation Research Board, No. 1706, 2000, pp. 1–8.


Cerdà, I. *Teoría General de la Urbanización y aplicación de sus principios y doctrinas a la reforma del ensanche de Barcelona.* Imprenta general Española, Madrid, 1867.


Doig, J., V.V. Gayah, and M. Cassidy. Inhomogeneous flow patterns in under-saturated road networks and implications for the MFD. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2390, 2013, pp. 68–75.


Last accessed December, 2015.


Ortigosa, J., M. Menendez, and H. Tapia. Study on the number and location of measurement points for an MFD perimeter control scheme: a case study of Zurich. EURO Journal on Transportation and Logistics, Vol. 3, No. 3-4, 2014, pp. 245–266.

Ortigosa, J., M. Menendez, and V.V. Gayah. Analysis of network exit functions for different urban grid network configurations. Transportation Research Record: Journal of the Transportation Research Board, No. 2491, 2015, pp. 12–21.


