Author:
Espinosa Mireles De Villafranca, Alonso

Title:
On Emergent Traffic Patterns in Synthetic Road Network Ensembles

An Experimental Mathematics Approach

On Emergent Traffic Patterns in Synthetic Road Network Ensembles: An Experimental Mathematics Approach

By

ALONSO ESPINOSA MIRELES DE VILLAFRANCA

Department of Engineering Mathematics
UNIVERSITY OF BRISTOL

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of DOCTOR OF PHILOSOPHY in the Faculty of Engineering.

AUGUST 2020

Word count: thirty three thousand and fifty six

Abstract

This thesis presents a computational investigation of traffic equilibrium models on synthetic networks that represent stylised street maps. Borrowing from statistical physics and complexity science, a network ensemble approach is used to examine the relationship between network structure and traffic equilibria. The network family constructed ($\alpha\beta$-networks) capture node distributions that range from square grids to networks with uniformly randomly distributed nodes. Cost-function parameters for roads are defined according to an endogenous supply provision heuristic that incorporates local network structure, and the static traffic assignment problem (STAP) is solved for ensembles of $\alpha\beta$-networks with a range of morphologies. How the networks' griddedness and road density affect the performance of traffic equilibria is explored. A key finding is that traditional network theory statistics do not correlate well with transportation efficiency. Also, less grid-like networks are more sensitive to the choice of demand structure. For mixtures of selfish and altruistic vehicles, the road density and network size are found to be key features determining the pathway to optimal performance. Finally, the $\alpha\beta$-networks are used to investigate how to recover a network fundamental diagram (NFD). It is found that combining the STAP with an ensemble approach is enough to determine how the uncongested branch of the NFD depends on network morphology. Finally, a method is proposed for recovering the congested branch of the NFD by using projected dynamical systems and Filippov systems, in order to capture dynamics of congestion that that lie beyond the STAP's scope.

Dedication and Acknowledgements

"There is no idea, however ancient and absurd, that is not capable of improving our knowledge."
– Paul K. Feyerabend

Nothing happens in a vacuum (except perhaps the Casimir effect), so I would like to thank those who have supported me throughout my PhD.

Firstly I would like to thank the institutions whose scholarships made this research possible: the Mexican Council of Science and Technology (CONACYT) and the University of Bristol.

I would especially like to thank my supervisor, Prof. Eddie Wilson, for giving me the freedom to explore my research interests, for his input, high research standards and the uncountable espressos (and pints) that undoubtedly made this thesis better than it otherwise would have been. Similarly, I am grateful for the conversations with Dr Richard Connors on traffic assignment and network structure, they were invaluable and helped shape the flavour of this thesis.

Thanks to my wife, Marika, for her unwavering support, for keeping me sane during the writing-up, and for rubbing off some of her Finnish sisu on me.

Thanks to my mother, Leonor, for always being an inspiration and an example of perseverance in the face of adversity. To my grandparents the academics, Nonó and Tati. To my Abuelo Jorge, for his unconditional support to my education. To my father, Jorge, for transmitting his love of computers to me. Also thanks to my brothers, Pablo, Daniel, and Pedro for always giving me their honest opinions, whether I wanted it or not.

Lastly, I want to thank my fellow students at Bristol, as well as the friends I made during my time in the UK, with a special mention to Julia, Rich, Louis K., Frank, and Oli.

For Ricardo.

AUTHOR'S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ALONSO ESPINOSA MIRELES DE VILLAFRANCA DATE: .....

TABLE OF CONTENTS

List of Figures

CHAPTER 1

INTRODUCTION

We live in an urbanised world that depends heavily on transportation. Already in 2018, 55% of the world's population were urban dwellers [11], and the UK is the most urbanised of OECD countries [12]. In the UK 80% of person-miles travelled in a year are road miles, similarly over 80% of domestic freight is transported on roads and motorways [13]. In Mexico City, the world's fourth largest megacity, it is estimated that a commuter can spend, in aggregate, 45 days of each year stuck in traffic [14]. Transportation issues filter through to all aspects of daily life, with congestion being an increasingly prominent concern. The European Environment Agency estimates that time losses in urban areas due to congestion have a cost equivalent to 2% of GDP [15]. Arguably, of even greater concern are the health and environmental effects caused by vehicular exhaust emissions, which are getting worse due to congestion [16]. This has led to policy interventions, such as the London and Antwerp low emission zones.

Understanding transport, and the ways in which we study and research mobility and transportation, has societal importance. There is ample evidence that altering the physical road network (or 'improving' infrastructure) fails to address congestion issues (see section 1.1 below). This thesis suggests that what is needed instead is a complete re-think of how the modelling of transport systems is approached: we focus on traffic equilibria from the point of view of complexity science. As the title implies, we will adopt an experimental approach in which well-understood models are used to explore features of emergent network effects on urban road systems.

Complexity science is the study of systems that are composed of many interacting components, and in which interactions between the components are the main determinant of system wide behaviour. As such, complexity science has much to offer to transport research. Overall, complexity science is defined more by the objects it studies than its mode of investigation. Indeed, there is no clear consensus on what makes a system complex [17], however there are systems that are considered poster examples of complexity. One such system is the human brain, for which

consciousness can be thought of as an emerging trait. Another is the collective motion of living creatures [18, 19]. Waldrop [20] introduces complexity science through different kinds of emergent phenomena, ranging from the collapse of the Soviet Union to the evolution of the eye. An in-depth discussion about complexity and what constitutes a complex system, as well as the much-used term 'emergent behaviour' is presented by Ladyman *et al.* in [17], where they attempt to pin down what are the traits shared by systems that are accepted to be complex. These systems tend to have been long studied in other fields, such as trophic webs and reaction diffusion equations, making the methods of study of complexity an eclectic set of techniques and tools borrowed from their originating areas; from dynamical systems to experimental biology. For example, the brain has long been studied from a medical or biological perspective [21], however the use of network theory is a more recent development which is now standard in computational neuroscience [22].

Complex phenomena tend to be identified as adaptive and emergent, which means that the structure and evolution of complex systems are driven by the components of the system itself. For example, in a transportation context, all road network users interact by affecting each other's travel times indirectly through the state of the traffic network. Yet the traffic patterns that determine these costs are, in turn, formed by the users' actions. Feedback loops such as these are an extremely useful concept in electronics and control theory, and they have now made their way even to the study of gene-regulation networks, to which they are central.

Complexity science, then, can be described as the science of learning systems [23], where learning is understood in terms of the adaptive behaviours of phenomena that arise in the interactions of multiple agents. This shifts the focus from the components themselves to the main features of their interactions, that give rise to patterns in the system's behaviour that would seem unexpected when considering the components in an isolated way. In other words, the core assumption of complexity science is that complexity emerges from simple rules, which however do not predict the outcome of every situation, but enable agents to adapt to their environment through feedback and learning over time.

1.1 Networks, Cities, Traffic, and Complexity

Cities fit all the characteristics of complex systems. They are made up of different components: physical infrastructures, different sectors of society, industry, government and so on. They are also primarily a collection of living creatures, namely humans. In terms of dynamics, transport and mobility clearly show emergent patterns, such as traffic jams that arise from the interactions of agents. In the last couple of decades, with the increasing recognition for the need for multidisciplinary research to solve the challenges of modernity [24], there have been many important developments in the study of cities. Bettencourt *et al.* [25] exploit the non-linear scaling relations that cities follow (for example see [26, 27]) to define metrics that capture phenomena local to particular cities. They also find that negative traits scale in similar ways. The inspiration for

this type of analysis is mainly the scaling in 'metabolic rates' exhibited by [28, 29], and the acceptance that even though there is central planning in the evolution of a city, it is ultimately self-organisation, an aggregate phenomenon, that drives the evolution of a system as complex as a metropolitan area [30]. As a consequence (and made explicit in the title of [29]: *Cities as organisms: Allometric scaling of urban road networks*), cities are now widely thought of more as living organisms than as containers for human life.

These scaling ideas have also influenced studies on the road networks of cities, for example [31], where the natural abstraction of a road network to a mathematical graph has led to new ways of approaching the study of urban networks. The use of network theory has led to new insights into the structure of cities and their evolution (a few interesting studies [32–34]).

Increasingly, it has been found that the evolution of networks in nature, of which road networks are no exception, follow similar underlying principles [35]. It is also important to note that while there are a large number of complex network models in use for a large array of applications, spatial networks that are planar are ubiquitous in nature, and while they might appear overly simplistic as models, they are far richer than might be expected.

What has emerged is an understanding of the myriad of ways in which complex systems are 'active' and have agency to modify themselves. This also offers new ways of conceptualising what cities are and how they behave [36], which in turn has led to a much better understanding of the problems urban areas now face, and offers ways to find solutions to these problems. One of the more important problems that cities face is increasing congestion. As a temporary measure (and in the USA the main policy [37], as cited by [1]), the strategy used to solve congestion problems has been to increase supply, however building more roads simply provides more road miles to suffer from congestion. Figure 1.1 (reproduced from [1]) shows the vicious cycle of congestion that is tied to the dependence on automobiles as the main method of transportation.

For the examination of the interactions amongst agents, borrowing ideas from game theory has proven to be essential for transport research. For example, Wardrop user equilibria [38], where users choose their paths through a network in order to minimise their travel time, are actually Nash equilibria for congestion games on networks (first identified by Charnes and Cooper in [39]). Therefore techniques used for analysing the inefficiency of equilibria, namely the *price of anarchy*, have made their way to the study of flows on networks, where they have been extremely useful in the context of internet routing [40, 41] and in the context of transport networks [2]. The price of anarchy is the ratio between the system costs at equilibrium to the costs of the system at optimal performance. It therefore measures the factor by which a system underperforms.

Roughgarden [42] shows that the upper bound for the price of anarchy in transportation networks, and commodity flow networks in general, depends on the choice of cost (or disutility) functions for the individual roads rather than on the structure of the networks themselves. His provocatively titled paper *The price of anarchy is independent of the network topology* [43] has inspired much research into the applications of the price of anarchy. However, the title is

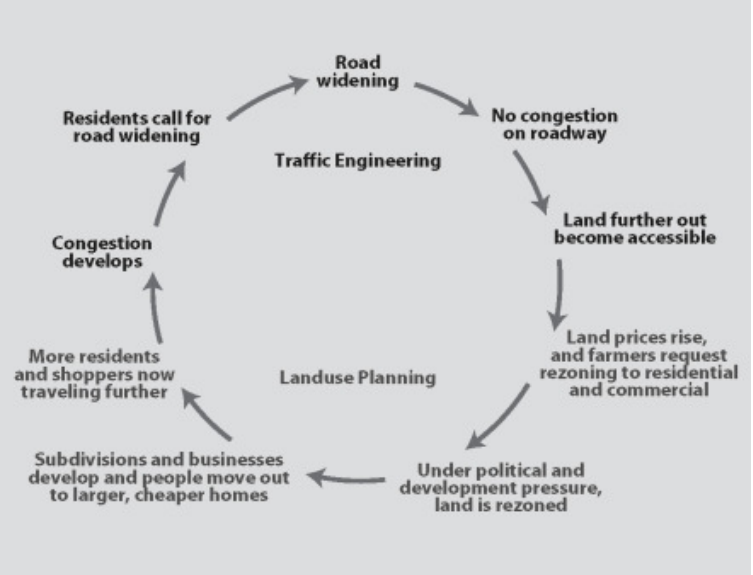


Figure 1.1 Vicious cycle of congestion. Reproduced from Toth, 2007 [1] (where it is figure 8-2)

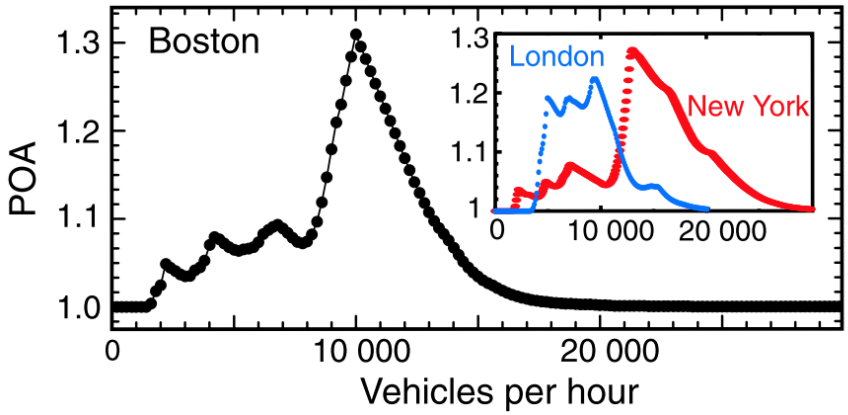
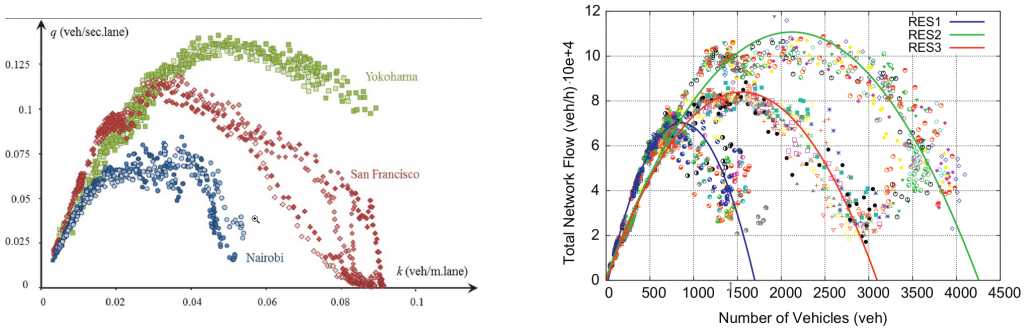


Figure 1.2 Price of anarchy as a function of demand shown for Boston, New York and London. Reproduced from Youn *et al.*, 2008 [2]

deceptive. The actual price of anarchy observable in networks depends heavily on the network and on the distribution of journeys that are made. As an example, in figure 1.2 we reproduce a figure from [2] showing the price of anarchy for different road networks. The price of anarchy will play a prominent role in this thesis, and in fact we will address the extent to which it actually depends on network structure in chapter 3. One of our key findings is that, even with significant variance when averaged over networks of similar morphology, mean values of the price of anarchy across network ensembles of similar structure (yet with high intra-ensemble variability) are consistently sensitive to even slight variations in network structure.

There is a need for more research on the behaviour of the price of anarchy and network structure, and this thesis answers to that need. The literature is mostly concerned with bounds



(a) (b)

Figure 1.3 Examples of macroscopic fundamental diagrams. (a) Data for Yokohama, San Francisco and Nairobi (reproduced from [3], where data from [4–6] is used). (b) Macroscopic fundamental diagrams of three (homogeneous) partitions for San Francisco, a heterogeneous network (reproduced from Abdoulas *et al.*, 2013 [7]).

that are not tight and can be deceptive when compared to actual achieved values of the price of anarchy on real-world, as well as synthetic networks. O'Hare [44] observes that low price-of-anarchy values might seem to indicate that the equilibrium reached by selfish users is quite efficient, however if total system costs are high, it can under-account for inefficiencies due to delays on important roads.

There are also few systematic studies on how complex network structure affects the price of anarchy. For example a paper by Rose *et al.* [45], in similar spirit to [46] and the present work, is interested in how network structure affects the price of anarchy. They, however, approach the construction of a heterogeneous network by manipulating the travel cost functions of some links, as the networks they use are lattices. They do this, in part, because they aim to find exact values for changes in the price of anarchy, which, however, quickly becomes problematic when networks increase in complexity. The authors remark that even with simple networks, the consideration of all possible paths becomes analytically intractable.

Another work that studies the relationship between price of anarchy and network structure is [47] by Smith *et al.*, where they investigate the influence of equilibrium efficiency as affected by the properties of the nodes in the network. Here, the difficulty of dealing with causal effects of network properties on the price of anarchy is also highlighted.

With regards to the role of local network structure in emergent large-scale traffic patterns, the observation of the so-called *macroscopic fundamental diagram* by Geroliminis and Daganzo [5], has been very influential. Figure 1.3 (reproduced from [3] and [7]) shows some macroscopic fundamental diagrams. Apart from a good theoretical understanding in how it emerges, and how to approximate it using standard traffic flow theory (for example [48]), this work has also facilitated the observation that under favourable circumstances, the partitioning of road networks can lead to sub-networks exhibiting different, yet well-defined, macroscopic fundamental diagrams

(see [49], for example). From a complexity science point of view, the most striking observation is that the resulting partitions tend to have uniform structure. That is, the macroscopic relationships fundamentally depend on the underlying network structure. In this thesis, we argue that the structure, in conjunction with the induced routing options, are more important than actual traffic dynamics for the emergence of the macroscopic patterns of the aggregated traffic variables; at least up to the critical point (where flow and occupancy become anti-correlated). The realisation that network structure is more important than what is implied by the theoretical bounds of the price of anarchy has led to studies that aim to be applicable in future traffic management. For example, in [3], the authors use both approximations of the macroscopic fundamental diagram, together with mathematical expressions for Wardropian equilibria, to propose a dynamic-tolling scheme strategy. More generally, the effect of partitioning a large network into individual components with independent behaviour has given rise to the possibility of cordoning strategies inspired by the success of congestion charging. Some examples of these studies are [7, 50, 51].

Bringing back network theory into the discussion, recently Kirkley *et al.* [52] identify betweenness centrality as an 'invariant' of planar graphs. Betweenness is a measure from network theory that measures how many shortest paths a particular node or edge in a network belongs to. It has been used to approximate the volume of traffic flow on roads of street networks (for example in [31]) under the assumption that human drivers choose shortest paths, thus edges that belong to many of the shortest paths will suffer more congestion.

While betweenness may serve as a proxy for when assignment calculations are deemed too cumbersome to carry out, the result from Kirkley *et al.*, that planar networks show similar distributions of betweenness, actually suggests that it is not a good proxy for traffic volumes on streets. Street networks have a mostly planar structure [53, 54], yet even networks with very similar structures have widely varying behaviour in terms of price of anarchy. This has been observed by O'Hare [46], and we obtain similar results in chapter 3.

Another link between network performance and network structure is captured by the Braess paradox [55], where the addition of a link, and a new route through a network, can cause a deterioration in the performance of the user equilibrium traffic pattern. While the Braess paradox is well understood [56], and its prevalence in complex (realistic) networks is estimated to be high (Steinberg and Zangwill [57] estimate its occurrence in large networks at a rate of 50%), it is not currently known to what extent it is responsible for unwanted performance issues of actual real-world systems.

There has been much research into classes of networks for which the Braess paradox can be eliminated altogether. Milchtaich [58] shows that in a class of networks (series-parallel networks), composed by adding links in series with networks made up of parallel bundles of edges, the paradox does not occur. This result, while not necessarily helpful in the design of road networks, highlights the claim that there is much to be discovered with regards to network structure and its influence on network performance.

In view of the discussion in this section, we identify that there is a gap in the literature in understanding the subtleties of the effects of network structure on emerging traffic patterns. Furthermore, it is accepted that in many situations an analytical treatment of paths is not always possible. Therefore there is a need to develop a deeper experimental intuition of the connection between network structure and network performance, in order to drive further theoretical progress. In the following sections we narrow down to the technical subject matter of this thesis and lay out our contribution to the field of transport networks.

1.2 Research Questions

Our main interest lies with equilibrium traffic patterns, their efficiency, and how these change with increasing demand on the one hand, and with changes in user behaviour on the other. At the heart of this thesis sits the development of a network model for which morphological characteristics can serve as independent experimental variables. Therefore, the model seeks to be useful in a variety of transport research applications.

The first application is in studying to what extent the price of anarchy is sensitive to network structure and to the origins and destinations of journeys in the network. The second is to explore how network morphology, and therefore the underlying route structure of a network, affects the gains in efficiency that could be leveraged by using autonomous vehicles as altruistic users of the network. Lastly, network structure and the emergence of a macroscopic fundamental diagram are shown to be intimately linked [5]. Therefore, we aim that our network model is of use in understanding what is the most basic connection between these two things.

In order to address these overarching concerns this thesis addresses the following research questions:

RQ1 How simple can a network model be, and still remain useful in understanding real-world transport properties?

RQ2 What insights can a highly-stylised routing model, such as the static traffic assignment problem, provide into the role played by network morphology in the traffic patterns (and their efficiency) that arise on these networks?

RQ3 Given that autonomous vehicles can provide a mechanism for achieving optimal network performance, is the transition to optimality provided by them, across networks with different morphologies, a good one?

RQ4 In the context of understanding global (network-wide) behaviour of networks, what 'minimal' modifications can be made to the *static traffic assignment problem* to build a network fundamental diagram from the bottom up?

To answer these research questions, we construct a family of random networks that capture different types of network structure and morphology. These networks are then equipped with travel-time cost functions, that allows us to find equilibrium traffic patterns on them and examine how network structure affects the efficiency of these networks. We not only study traffic patterns that solve the classic static traffic assignment problem (RQ2), but also traffic equilibria of mixed user classes (RQ3), where one class behaves altruistically, as could be plausibly implemented with autonomous vehicles.

Research question RQ4 brings us back to the central motivation of complexity science, the emergence of structure from underlying interaction. The structure we aim to recover is the relationship between network-wide aggregated traffic variables. The interactions among users of the networks are captured through each other's influence on the cost functions of the edges.

We base our method of inquiry on the doctoral work of O'Hare [44], in which he studies the efficiency of synthetic road networks, how it is affected by different levels of travel demand, and finally how is affected by the structure of the networks. He shows how fluctuations in the price of anarchy come about from the differences in active links and routes between the optimal traffic assignment and the so-called user equilibrium assignment [53].

O'Hare remarks that the price of anarchy is much lower than the theoretical bounds and further, that it is very sensitive to variation in individual network structure. His results indeed show that the spread of performance metrics in general vary significantly across networks of similar types (that is, drawn from the same experimental ensemble).

We depart from O'Hare's methodology in two important ways. The first is the way we develop our network model. We focus on constructing networks that are plausible proxies for real world networks. We take care to avoid introducing arbitrary choices of variables, for example, by implicitly defining the travel-time functions of the streets according to the local network topology. We also normalise the parameters of the networks to ensure a fair comparison of networks with different number of edges. The second difference, is that we use potential applications (described in section 1.3 below), beyond the study of price of anarchy, to guide the network design. The networks that we construct aim to model homogeneous 'patches' of network, and by giving them periodic boundary conditions, single origin-destination pairs give rise to traffic patterns that would otherwise require more complex journey structures to generate them.

The style of analysis in this thesis is based on ideas from complexity science. We will touch upon equilibrium, network structure, and collective behaviour. In terms of model types, we will traverse different scales in different parts of the thesis. At the microscopic end of the scale spectrum, we have the geometrical construction of a family of synthetic road networks. At the macroscopic end, we explore the emergence of collective patterns such as the network fundamental diagram.

Inspired by computational statistical physics, we use ensembles of networks of different morphologies, in combination with classic models from transport theory, to probe new ways in

which we can qualitatively recover traffic behaviour that is more usually the concern of much more sophisticated (or complicated) models. We build mainly on Wardropian equilibria [38], and make experimental use of the *static traffic assignment problem* [53, 59]. We view the simplicity of the model as an advantage, and we use it to show how the structure and morphology of road networks plays a fundamental role in determining their transport properties.

1.3 Structure of this Thesis

To address the research questions stated above, this thesis has been structured as follows. Chapters 2 and 3 are concerned with answering the first two research questions RQ1 and RQ2, while chapters 4 and 5 are devoted to the third and fourth research questions (RQ3 and RQ4), respectively.

In chapter 2, a family of random planar graphs, which we will call '$\alpha\beta$-networks' is developed. These networks are based on β-skeletons [60], and we have designed them to capture a smooth transition from a perfect grid structure to a network whose nodes are uniformly randomly distributed. We examine the spatial and structural properties of these networks, and show that they are a suitable model for experimentally controlling network morphology.

The aim of chapter 3 is to explore the extent to which particularities in network structure determine network performance in a routing context. We use the static traffic assignment problem with affine cost functions as a transport equilibrium model. We develop a heuristic for allocating cost-function parameters in a way that makes networks with different morphologies and numbers of edges directly comparable in terms of their provision of supply. To measure efficiency of the traffic patterns across networks of different morphologies and with different choices of origin and destination patterns, we use the price of anarchy, to show that the spatial distributions of intersections and their clustering is more important for network efficiency than topological features. We also find that the price of anarchy as a function of demand is sensitive to variations in network morphology across ensembles of networks.

In chapter 4, we use the $\alpha\beta$-networks to explore the transition to optimality that might be realised with the introduction of autonomous vehicles as a routing-control mechanism. We use a simple mixed equilibrium model such as the ones used to study the uptake of route guidance [61, 62]. We apply it to ensembles of road networks to see how efficiency increases with the penetration rate of altruistic autonomous vehicles. Significant space is devoted to understanding the mechanism in the reduction of costs on small networks before applying the model to our $\alpha\beta$-networks. We find that, as the networks become more complex, the curves that trace the decrease in costs towards optimal smoothen out as the discrete effects observed in small networks overlap. We also discuss what types of network structures are better suited to this type of amelioration of congestion.

In chapter 5, we focus on a topical area in the fields of transport and mobility: the network

fundamental diagram. Once again, the starting point is static traffic assignment, from which we derive network-wide relationships between aggregated traffic variables. By necessity, static assignment is not enough to capture the dynamic nature of a true macroscopic fundamental diagram. Thus, we show that slight modifications to it are enough to recover, at least qualitatively, behaviour observed empirically and in more sophisticated traffic flow simulations reported in the literature. By examining a specific application in depth, namely the network fundamental diagram, we return to the broader subject of complexity: how simple rules lead to emergent global patterns. In this case, by considering aspects of dynamical systems that have found applications in transport theory [63, 64], a novel way of approaching the time-dependent behaviour of congested traffics networks is proposed: network traffic dynamics interpreted as a Filippov system [65]. Finally, chapter 6 presents our conclusions and describes possibilities for future work.

CHAPTER 2

SYNTHETIC NETWORKS

The methodological backbone of this thesis is the use of computational experiments as a way of probing what role network morphology plays in the emergence of traffic patterns. We begin by constructing a network model which is used to generate the experimental ensembles of networks used throughout this thesis. The purpose of this chapter is to identify appropriate variables which will be used to describe important network structures, and to set up the experiments that will be discussed in this work.

Since graphs and networks are applied in diverse transportation problems, the context of the modelling problem heavily influences what is considered relevant to the network structure. We aim to model homogeneous 'patches' of road networks that can be thought of as embedded in a larger network. In this chapter we lay out the key features of the road networks that are captured by the model which we construct and which we will use in our discussion throughout the the rest of the thesis.

In section 2.1, the context in which the model sits is introduced, and the motivations behind the design choices are highlighted. The network construction method is described in detail in section 2.2, which involves sampling the locations of the nodes, the method by which the edges of the network are wired up, and the boundary conditions of the networks. Properties of the generated networks are explored in section 2.3. Finally in section 2.4, an alternative view and variation upon the model is presented, where the geometric correspondence of nodes with grid points is relaxed; the nodes can then be obtained from a non-homogeneous point process. The purpose is to show that the network model can be generalised and cast as a random point process, which is common practice in modelling network infrastructure. Finally section 2.5 summarises our contributions.

A condensed overview of the network model constructed in this chapter can be found in the proceedings of the Traffic and Granular Flow conference of 2017 [66]; this chapter is far more

comprehensive. It is worth noting that this model has already been found useful in the wider transportation research community, for example, in the context of intersection signalling and controllability of traffic networks [67].

2.1 Background

In the transport modelling literature, equilibration algorithms and optimisation routines tend to be tested on benchmark networks; the canonical example being Sioux Falls. Alternatively, bespoke test-networks appropriate for the authors' needs are used.

The need for testing algorithms and models on networks with different structure is a consequence of the importance of the connectivity of the networks. Networks with different structures can have drastically different behaviour. A key example of this is the Watts-Strogatz small-world network model [68]. While it is not a spatial model, it shows that rewiring a small proportion of edges in a network can drastically change global properties. The way spatial networks are wired is intimately related to their embedding in space. Therefore, development of appropriate network models with specific domain applications, which consider this embedding as well as additional restrictions like planarity are necessary. The abstraction from the physical street map of a city to a network is initially straightforward; streets are edges, and intersections are nodes. This is called the primal representation [69] (for an example in an application where different representations are discussed, see [70]).

The structure of road networks is not only important in terms of their statistical and topological properties, but also in the efficiency of traffic routing itself through the networks. For example, in [2], Youn *et al.* use the *price of anarchy* (PoA) as an efficiency measure. Using the PoA, they show that network structure (both of individual networks and of samples from random families of graphs) perform differently in terms of transport efficiency. Overall, these kinds of systematic studies of the dependency of traffic performance on network morphology are, however, rare (both [32] and [44] touch upon this point). The principal reason for this is the difficulty involved in comparing different networks (with possibly vastly different characteristics) to each other. In a way, networks having different numbers of nodes and edges to each other already means they are fundamentally different.

Additionally, the number of existing cities is small compared to the numerosity of possible networks that can be constructed. On the one hand, real cities do not present the variability needed for a systematic study of morphology; there are too few cities of similar types. On the other hand, cities are too irregular; their shapes are constrained by geographical features such as rivers and topography. The solution to this, apparently contradictory problem, is to use synthetic networks. In the field of network science, random graph models are used to approximate or simulate real-world systems. Some models are particularly useful for their mathematical properties, and have allowed for the understanding of underlying mechanisms

for the emergence of statistical properties of networks [68, 71]. An example of such models are preferential attachment models, yielding degree distributions that follow power laws (so-called, scale-free networks). While it is possible to find traffic equilibria on these networks, their usefulness for understanding the routing behaviour in cities is limited. Many of these networks have no spatial description whatsoever.

Another approach for network construction is to generate or *grow* them according to some prescribed rules, for example (resource) minimisation principles informed by the system to be modelled. Network growth models can be used to study the dynamics of network evolution. Additionally, they can be used to validate models that attempt to describe how the road network configuration evolves, such as in [32]. Many types of physical networks lend themselves to be studied in similar ways, and not necessarily in a time dependent way. For example, infrastructure networks [72], leaf venation patterns [73], and even slime moulds [74, 75].

Investigations that take the network science approach, including some of the ones discussed above, tend to take a purely empirical approach in measuring transport variables of known networks, or are concerned mostly with the structure of the infrastructure itself. In other cases, due to how recent this way of studying cities and collective human behaviour is, studies are more of the 'proof-of-concept' type. This means that the focus often is on the insights offered by network-theoretic measures into the network structure; for example, betweenness centrality has been identified as important in [31] and [70] and even used as a proxy for traffic flow.

Initial interest in how network structure can lead to counter-intuitive effects (exemplified by the Braess paradox [55] and discussed in detail in [56]), helped solidify the concept that in large complex networks, effects like the Braess paradox are actually commonplace [57]. In this last paper by Steinberg and Zangwill [57], the arguments are quite involved, and as has been pointed out in the literature as well [76], the equations are quite opaque, relying on applying Cramer's rule to a very large system of equations. Perhaps due to the difficulty in obtaining analytical results, as well as the difficulty in empirical observation, counter-intuitive effects due to routing and availability of paths remain under-studied.

The purpose of constructing our network model is to contribute in clarifying the connections between network measures and the transport properties of road networks. In this way, we also wish to build upon ways of theorising about transportation problems, and traffic pattern formation, from a complex systems perspective. In practice, this means we will use instances of road networks generated with identical morphological parameters to create experimental ensembles of networks. In the rest of the thesis we will talk about the response of these networks when subjected to different (computational) experimental treatments.

We adopt a statistical ensemble approach, since is appropriate to the task at hand, in fact in the course of this research, other authors have also found it suitable for tackling similar questions to ours [45, 47]. In this light, honing models that are useful in identifying important network features is central to the wider field of transportation.

O'Hare [44] constructs his network ensembles by pruning triangulations of random nodes. The pruning of the edges is done randomly with the (uniform) probability of edge removal serving as a model parameter. In order to guarantee that the resulting networks remain connected when pruned in this way, a minimum spanning tree for the network is found, and edges on this tree are protected from the pruning. This means that O'Hare's network model captures a transition where the networks change from being mesh-like to tree-like.

A common starting point in network analysis is the examination of how tree-like networks are; trees can be conceptualised as an 'extreme' or boundary-case type of network structure. Watanabe, in [77], identifies the tree structure as important to road networks, and tries to quantify networks generated by proximity graphs depending on how tree-like or grid-like they are, primarily by looking at a metric he introduces as the *grid-tree* proportion.

In general, the models used for studying physical networks can be roughly split into two types. The first type we will call *sampling* models, which include the traditional random network models used in network science such as Erdős-Renyi [78], Barabási-Albert [71], and small-world networks [68]. These tend to be described and characterised statistically, while the construction method for these types of networks is usually probabilistic in nature (often trying to satisfy some constraints). As an example, the configuration model (a standard model for complex networks, see for example [69]) and its variants can be used to generate random networks with a given degree sequence. The second type of models are *generative*, in which the networks are grown or generated according to some heuristic, which is usually abstracted from physical or physical-like processes. Examples of these are cellular automata models of urban growth [79,80], and growth of leaf patterns and infrastructure networks following some optimality principle [32]. These models tend to emulate or capture an underlying process of how the system under study behaves, as well as to test whether the models proposed to construct such networks agree with experimental or real-world observations.

The model developed in this chapter has aspects of both approaches and draws from the method developed by O'Hare in [44] (and used as an analysis method in [46]) for studying the performance of networks with different structures. Key requirements for a model suitable for investigating these kinds of networks include:

- That the networks be (strongly) connected, so that for randomly chosen origins and destinations, the journeys can actually be made.
- That the networks' morphology parameters capture aspects of real-world road networks, or at least of networks widely used to benchmark algorithms and models.

2.1.1 Some definitions

So far we have discussed approaches to studying networks in general, and road networks, more specifically. Before going into more technical detail it is appropriate to define the basic

mathematical concepts that we will be using. Since we will be abstracting road networks as graphs, we give use the following (standard) definition [81, 82].

Definition 2.1.1 (Graph/Network)**.** *A* graph *or* network*, G, consists of the pair*

$$G = (\mathcal{V}, \mathcal{E}), \tag{2.1}$$

where $\mathcal{V}$ is a (non-empty) vertex set, and $\mathcal{E}$ is a (possibly empty) set of edges. An edge, $e \in \mathcal{E}$, is composed of a pair vertices ($e = (\eta, \nu)$ for η and $\nu \in \mathcal{E}$).

This is a very general definition that allows for any vertex set $\mathcal{V}$, which is what makes graph-based models so widely applicable. The edges represent relations between nodes. As described in the, edges are given as pairs of nodes. A directed network has the additional requirement that the vertices (or nodes) composing the edges are ordered; the standard way to depict directed edges is as arrows pointing from the source node to the target node.

Definition 2.1.2 (Directed network)**.** *A network, $G = (\mathcal{V}, \mathcal{E})$, is directed if the vertex pairs that make up the edges are ordered. That is, edges with the same vertices but inverted order are distinct,*

$$(\eta, \nu) \neq (\nu, \eta), \tag{2.2}$$

for $\nu, \eta \in \mathcal{V}$.

An edge $e \in \mathcal{E}$ is said to be *incident from* ν_i and *incident to* ν_j. Edges will also be referred to simply by their indices, $i \sim j$ where the convention will be that i is the source or tail of the edge and j the target or head.

2.2 Assembling Random Networks

We construct the networks that we use in this thesis in two stages. The first stage is the arrangement of the nodes on the plane (section 2.2.1); the second stage is the wiring of the network by forming the β-skeleton of the set of nodes. This is covered in section 2.2.2.

The construction method aims to provide parameters that can be changed to study the properties of the networks' structure. For example, in chapter 3 it will be used to find correlations between the traffic-routing properties of the networks and their morphology. Since the possible variations in morphology are quite numerous due to the combinatorial nature of graphs, the construction method presented here aims to capture two characteristics of networks that have been shown to have an impact in terms of transport properties: on one hand, how ordered the nodes are and, on the other, the redundancy of edges. Thus, a two-parameter family of random planar graphs is obtained, with the geometric parameters providing a quantification of these properties.

The first parameter of the model, α, measures the griddedness of the networks or how regular the positioning of the nodes is. This way, α gives a measure of how far along the spectrum

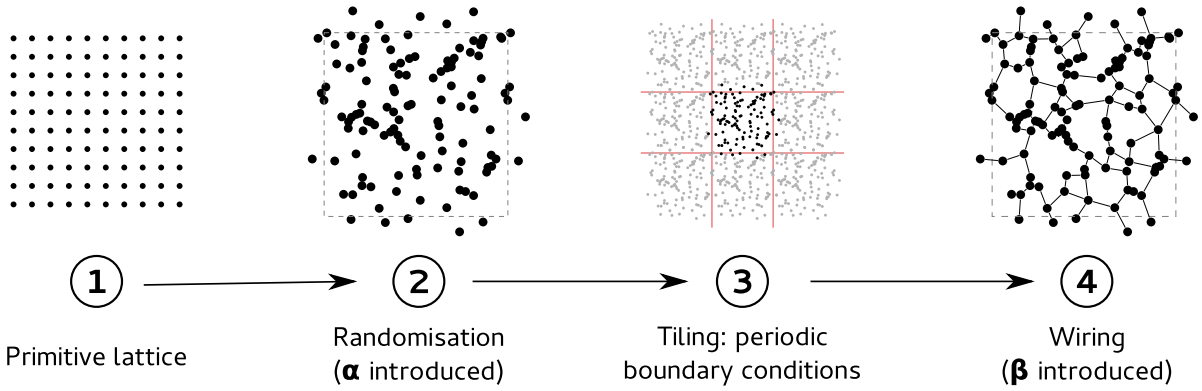


Figure 2.1 The four main steps of the network construction method are shown. The randomisation of the lattice and the edge wiring are parametrised by α and β, respectively. This yields a two-parameter family of random graphs.

between a perfect grid and perfectly random the nodes are positioned. The second parameter is the construction parameter of the β-skeleton [60], a graph model that can be used to reconstruct 'shapes' of sets of points. There is evidence for β-skeletons being useful for re-constructing networks from their node-sets, as is shown in [8]. The generation process has two steps:

1. Generating the node-set $\mathcal{V}$.

2. Wiring the network (constructing the edge set $\mathcal{E}$).

The construction of the node-set is covered in section 2.2.1 and the wiring of the nodes in section 2.2.2. A visual summary of the network construction method is shown in figure 2.1: the stages where the morphology parameters of the $\alpha\beta$-networks are introduced are explicitly identified.

2.2.1 Scrambling the grid

The position of the network nodes on the plane, or more specifically in the unit square, are determined by starting with a primitive square lattice, which is then perturbed according to the procedure described as follows.

The parameter that captures how lattice-like the nodes are arranged, α, will be referred to occasionally as the *griddedness* of the network. The griddedness is defined in the unit interval ($\alpha \in [0,1]$) such that the distribution of nodes changes from a regular square grid (for $\alpha = 0$) to a uniformly random distribution of points, when $\alpha = 1$. In order to determine the positions of the nodes, we start with a primitive regular square lattice with N points arranged in n rows and n columns (*i.e.*, a $n \times n$ grid). The lattice is centred inside the unit square $[0,1]^2$ (as shown in figure 2.2). The gap, or shortest distance, between the exterior points of the grid and the boundary of $[0,1]^2$ is $1/2n$ and the spacing between the points, or lattice constant, is $1/n$. We associate a node with each lattice point to obtain the randomised position of the actual nodes. For each

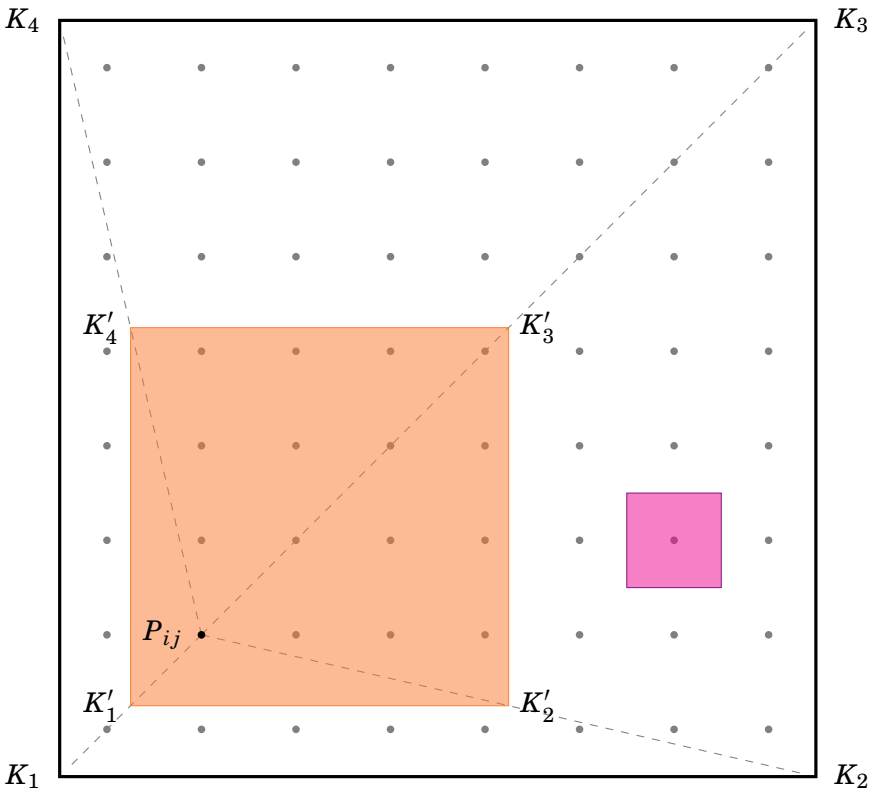


Figure 2.2 Primitive lattice for node-set construction (for this example $n = 8$). The lattice constant is determined by the n as well as the gap between the exterior nodes and the unit square's boundary. The construction of the re-sampling boxes is shown, as is the separation between nodes and the gap between the exterior nodes and the boundaries. The purple square shows the cell associated to a particular node. Its size is determined by the node density.

point P in the lattice we define a re-sampling box from which the actual position of the node will be drawn. The size of the re-sampling box is determined by α which we now use to define its corners K'_1, K'_2, K'_3, K'_4. Each K'_i for $i = 1, \ldots, 4$, is defined as the point on the line segment PK_i that divides it into proportions with ratio α in the form

$$\alpha = \frac{PK'_i}{PK_i}. \tag{2.3}$$

The distance of corner K'_i from the lattice point P is a proportion α of the distance to K_i, the corresponding corner of $[0,1]^2$. So as α increases, K'_i slides along the line segment PK_i between the point P itself ($\alpha = 0$) to the corner of the unit square K_i. Figure 2.2 illustrates both the lattice itself and the construction of the re-sampling box.

The re-sampling boxes are square: the two corners of a side of the re-sampling box (say, the top) are displaced by the same amount (upwards) towards the unit square boundary. This can be verified by calculating the length of one of the edges. Due to the construction technique, the re-sampling boxes are not centred around their corresponding lattice points except for $\alpha = 0$.

Rather, with increasing α, the re-sampling boxes are displaced towards the centre of $[0,1]^2$. Using the sub-index notation p_{ij} for the point on the primitive lattice in the i-th row and j-th column,

$$p_{ij} = \begin{pmatrix} \frac{1}{2n} + \frac{j-1}{n} \\ \frac{1}{2n} + \frac{i-1}{n} \end{pmatrix}; \tag{2.4}$$

more compactly,

$$\mathbf{p}_{ij} \cdot \hat{\mathbf{e}}_1 = \frac{2j+1}{2n},$$
$$\mathbf{p}_{ij} \cdot \hat{\mathbf{e}}_2 = \frac{2i+1}{2n}.$$

Each corner K_r of the re-sampling boxes is given by

$$\mathbf{p}_{ij} - K_r = \mathbf{p}_{ij} + \alpha(K_r - \mathbf{p}_{ij}).$$

The position of the left boundary of the j-th box in a row is given by

$$l_j = \frac{1}{n}(1-\alpha)\left(j - \frac{1}{2}\right).$$

Similarly, the position of the right boundary is

$$r_j = \alpha + \frac{1}{n}(1-\alpha)\left(j + \frac{1}{2}\right).$$

As α increases the trajectory followed by the centre of the box traces the line segment between p_{ij} and (1/2, 1/2), the centre of $[0,1]^2$. For the particular case when n is odd, the sampling box corresponding to the lattice point at (1/2, 1/2) remains centred. This 'bunching up' of the boxes towards the centre as α grows results in higher densities of points where the re-sampling boxes overlap.

Similarly, as α increases and the re-sampling boxes cover a larger area of $[0,1]^2$, the nodes diffuse out to the edges of the unit square. The distribution achieves uniformity when all the re-sampling boxes coincide with the unit square at $\alpha = 1$; the nodes also become identically and independently distributed. The separation between adjacent sampling boxes is the same for all consecutive interior pairs of boxes, either row-wise or column-wise. Take the horizontal gap between consecutive boxes in a row, S_j and S_{j+1}, the distance between the right boundary of box S_j and the left boundary of box S_{j+1}, is independent of the index:

$$l_{j+1} - r_j = \frac{1}{n}(1-\alpha) - \alpha. \tag{2.5}$$

In particular, the shrinking gap between the right boundary ∂S_{ij} of box S_{ij} and incoming left boundary $\partial S_{i(j+1)}$ of box $S_{i(j+1)}$ vanishes at

$$\alpha = \alpha_{\text{crit}} = \frac{1}{n+1}. \tag{2.6}$$

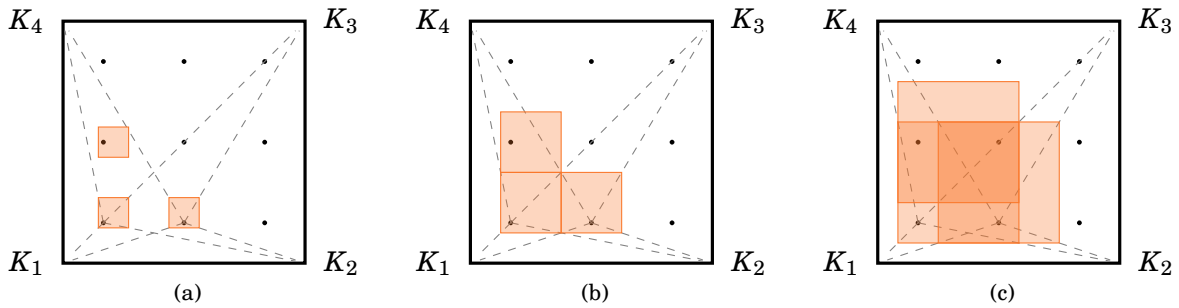


Figure 2.3 As α increases, $\alpha \to 1$, the re-sampling boxes expand and slide towards the centre of the unit square, $[0, 1]^2$. The panels show three of the re-sampling boxes: the bottom left-most box and its neighbouring (column- and row-*wise*) boxes for three different values of α. (a) $\alpha = \frac{1}{2}\alpha_{\text{crit}}$, for $\alpha \leq \alpha_{\text{crit}}$ the boxes are disjoint. (b) At $\alpha = \alpha_{\text{crit}}$ (for $n = 3$, we have$\alpha_{\text{crit}} = \frac{1}{4}$). (c) For $\alpha \geq \alpha_{\text{crit}}$ (here $\alpha = 2_{\text{crit}}$) there are multiple overlaps, including with boxes corresponding to further neighbours.

The right hand side of equation 2.5 does not depend on the labelling of the boxes, therefore all the boxes begin overlapping at α_{crit}. Figure 2.3 shows how the re-sampling boxes grow as α increases. Initially for small α the boxes are disjoint; they keep growing until at α_{crit} all boxes overlap with their neighbours. We can also see the way the boxes are pushed towards the centre as α increases. The centre of each box is also displaced from its corresponding lattice point towards the centre of the square.

When $\alpha = \alpha_{\text{crit}}$, the re-sampling boxes tile the central region of the square. There is still a gap between the outermost boundary of $\bigcup_i S_i$, and the boundary of the unit square; there is always a gap for $\alpha \leq 1$.

However, despite the detailed discussion of $\alpha \to 1$, the main empirical regime of interest for α is before the re-sampling boxes begin to overlap. When this happens, the lattice ordering of the nodes begins to completely break down and nodes can land arbitrarily close together. To make this more explicit, and for notational convenience we define the normalised griddedness, $\hat{\alpha}$, for sets of nodes (with the same number of nodes N) by

$$\hat{\alpha} = \frac{\alpha}{\alpha_{\text{crit}}}. \tag{2.7}$$

2.2.2 Skeletonising the points

In order to wire together the nodes from the scrambled lattice, the β-skeleton of the node-set is constructed. Introduced by Kirkpatrick and Radke [60], β-skeletons provide a way to determine the 'shape' of a set of points. They are a type of proximity graph model that, due to its generality and straight-forward geometric construction, has found widespread application, from stochastic and computational geometry, to analysing the distribution of mass in the universe [83]. Jaromczyk and Toussaint review different types of proximity graphs, and their properties and algorithms for constructing them in [84]. The β-skeleton is a type of proximity graph called an empty region

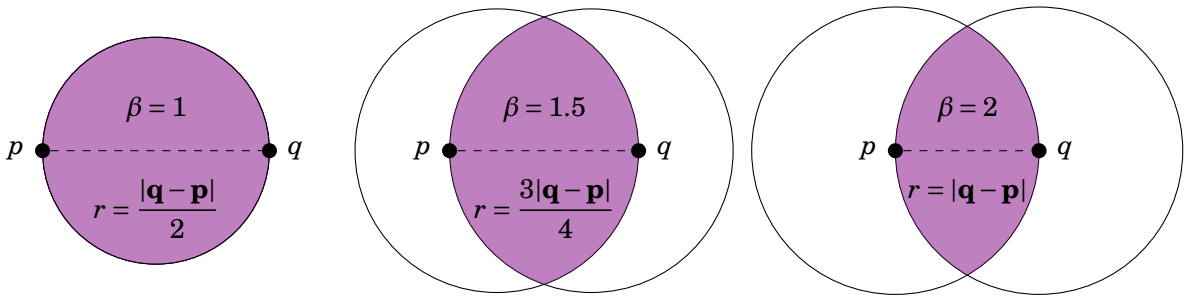


Figure 2.4 The construction of the exclusion (or template) region for the lune-based β-skeleton as the intersection of two disks with radius $r = \frac{\beta|\mathbf{q}-\mathbf{p}|}{2}$. When (a) $\beta = 1$, the skeleton recovers the *Gabriel graph*. In (b) the lune for an intermediate value $\beta = 1.5$ is shown. For $\beta = 2.0$ (b), the skeleton coincides with the *relative neighbourhood graph*. Note that the scale is different for each of the diagrams, which highlights how much more elongated the lune is at (c), that even when scaled down it is still taller than in (b).

graph [85]. In empty region graphs an exclusion region is defined for every pair of nodes $p, q \in \mathcal{V}$. Two points are deemed to be neighbours if the exclusion region is empty of any other nodes, that is if $\mathcal{V} \setminus \{p,q\} \bigcap R = \emptyset$. Neighbouring points are then connected with an edge.

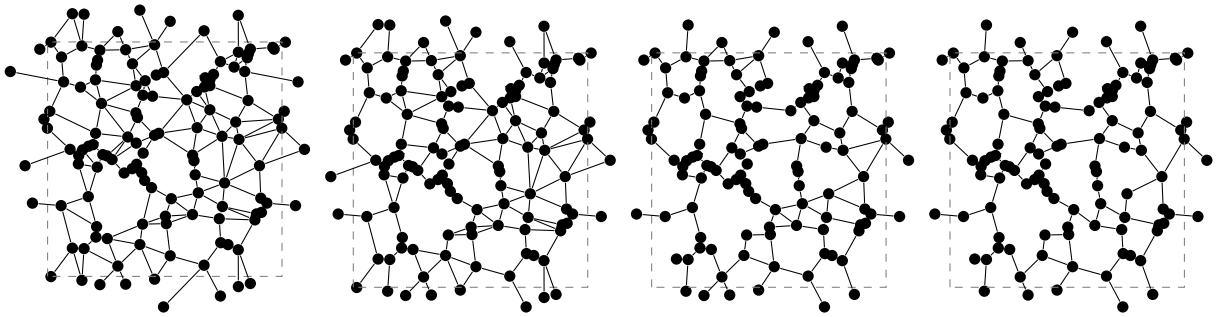
The type of β-skeleton we use is the the *lune*-based β-skeleton, named after the lens-shaped region, the lune, that plays the role of the exclusion region. For every pair of points $\mathbf{p},\mathbf{q} \in V$, the lune is the intersection of two disks of radius $r = \beta|\mathbf{q}-\mathbf{p}|/2$. Each disk is associated to one of the points $\mathbf{p}$ or $\mathbf{q}$ which it contains in its interior, with the other point lying on its boundary. The centres of the disks lie on the line defined by p and q. Figure 2.4(a) illustrates the construction of a lune corresponding to a pair of points for $\beta = 1$, 1.5, and 2.

Intuitively, the β-skeleton is a generalisation of a Delaunay triangulation, or more accurately the Delaunay triangulation can be seen as the limit of the β-skeleton as $\beta \to 0$. In the limit, the exclusion region gets 'pushed out' from between pairs of points to triplets of points.

Figure 2.5 shows several β-skeletons constructed from the same set of nodes but with different values of β. From the figure, it can be seen that this definition of the exclusion region is only valid for $\beta \geq 1$. When $\beta = 1$, the β-skeleton and the Gabriel graph of $\mathcal{V}$ coincide. The nodes straddle opposite sides of a diameter of each of the disks, meaning further displacement along the axis defined by them would make the nodes not contained in their corresponding disks.

Note that there are alternative definitions for the β-skeleton [84, 86], including for $\beta < 1$, as well as the circle-based skeleton. We will, however, restrict our discussions to the lune-based skeletons. For $\beta > 2$, the resulting skeleton can be disconnected, therefore we restrict the range by considering only $\beta < 2$. Note that for $\beta < 1$, the lune-based skeletons have to be defined differently (in which case they coincide with circle-based β-skeletons) so we shall set $\beta = 1$ as the lower bound for our constructions. Examples of the networks produced, with the model are shown in figure 2.6.

In term of road networks, β-skeletons can be useful for reconstructing street networks of



(a) $\beta = 1.0$ (b) $\beta = 1.25$ (c) $\beta = 1.75$ (d) $\beta = 2.0$

Figure 2.5 The β-skeleton for the same uniformly distributed node-set of 100 nodes ($\alpha = 1$): (a) $\beta = 1$; (b) $\beta = 1.25$; (c) $\beta = 1.75$; and (d) $\beta = 2$. The β-skeleton coincides with the Gabriel graph at $\beta = 1$ and the Relative Neighbourhood graph (RNG) at $\beta = 2$. Observe the monotonicity (equation 2.8) of the β-skeleton: for $\beta_1 < \beta_2$ then $G_{\beta_2} \subseteq G_{\beta_1}$, so that edges are successively removed (and never added) as we move to the right across the figure panels.

real cities. Osaragi and Hiraga in [8] show that in some cases, the topology of a road network can be reconstructed with high accuracy by constructing the β-skeleton, taking the actual street intersections as the node-set. The authors show that agreement between the reconstructed skeleton and the original network can be as high as $\sim 80\%$ for sections of the Tokyo road network.

Disagreement between the original network and the reconstructed one tends to happen where there are prominent geographical features, such as rivers or abrupt changes in topography. In the case of such obstructions, the β-skeleton over-connects the node-set, instead there should be gaps due to the limited connectivity.

In fact, Osaragi and Hiraga [8] show that a range of cities have good agreement with β-skeletons for $1 \leq \beta \leq 2$. They show that for $1.15 \leq \beta \leq 1.45$, the agreement rate between the original network and the reconstruction is maximised. In a related study, Maniadakis and Varoutas [87] examine the similarity of a much richer set of real-world networks and compare them to β-skeletons. In particular, they look at the same networks as [88–91], as well as 100 samples chosen from Greek cities. Among the features they compute are network length, average node degree, node density, and the number of edges. In concord with Osaragi and Hiraga, they find that for $1 \leq \beta \leq 2$, the error between the properties of the β-skeletons and the city networks can be as low as 10%, they find the best agreement for $\beta \in [1.2, 1.4]$.

Based on the reasoning above, we shall only consider β in the interval $[1,2]$. This range encompasses the region in which β-skeletons are known to represent some road networks morphology well. In addition, the end-points of the interval correspond to values of β for which the β-skeletons coincide with well-known graph models: the *Gabriel graph* (GG) (when $\beta = 1$), and the *relative neighbourhood graph* (RNG) ($\beta = 2$). For a given node-set $\mathcal{V}$, the edges of it's corresponding β-skeletons are 'monotonic' with respect to β, that is, given $\beta_1 > \beta_2$, G_{β_2}, is a

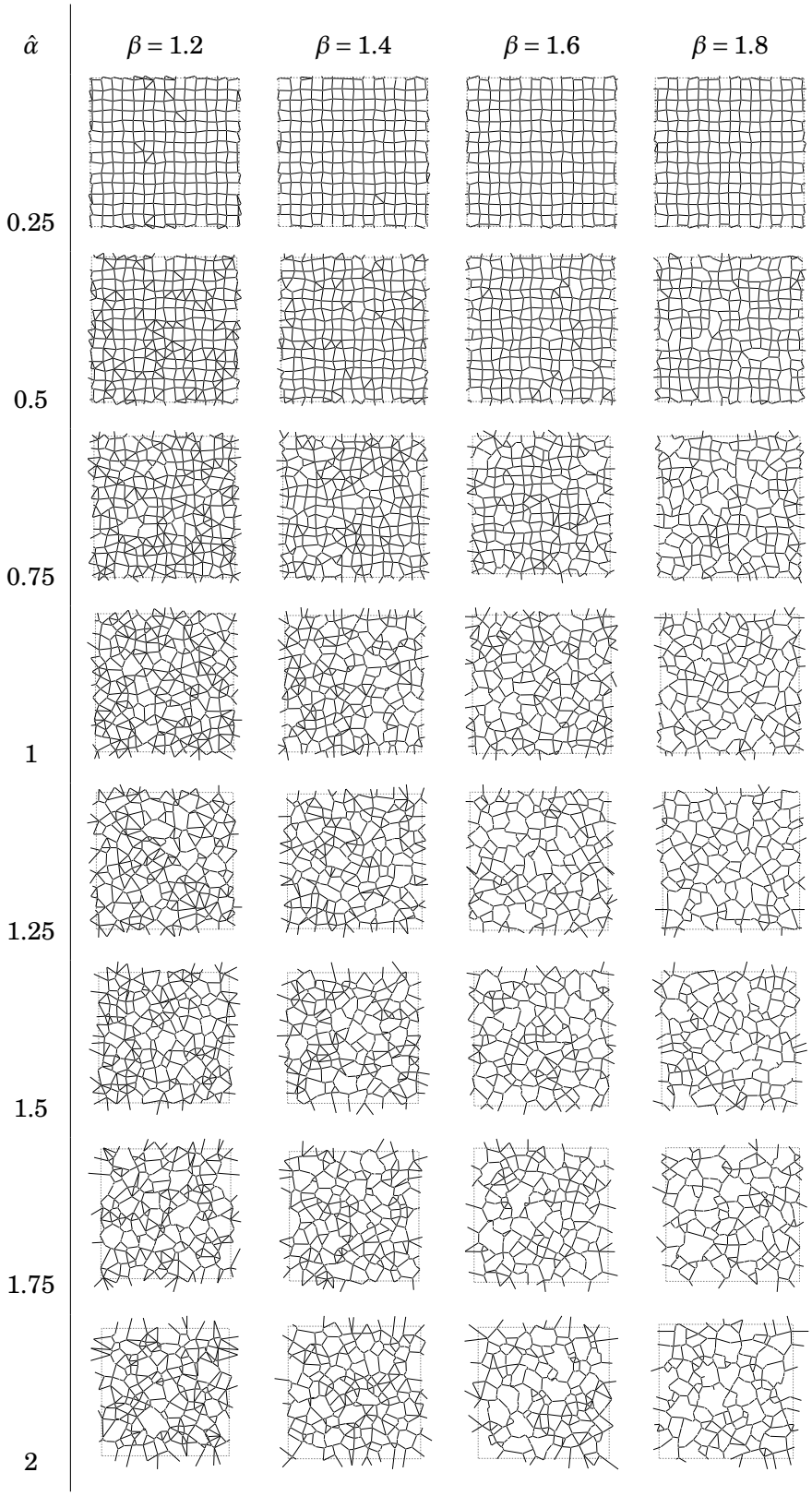


Figure 2.6 Examples of network structures for different values of $\hat{\alpha}$ and β. Networks are sampled from ensembles based on the 15×15 grid ($N = 225$).

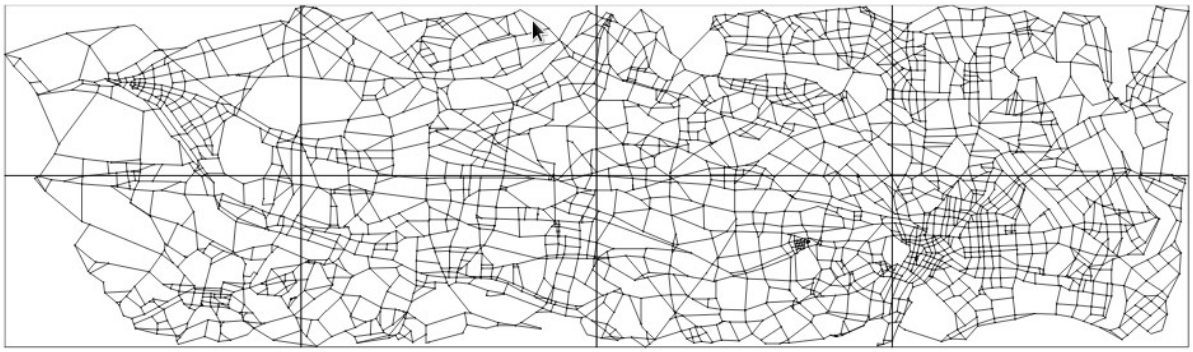


Figure 2.7 The Tokyo (a) road network sections used in [8] by Osaragi and Hiraga for reconstruction by β-skeletons. Reproduced from Osaragi and Hiraga, 2014 [8].

subgraph of G_{β_1}. Thus for $\beta \in (1,2)$, the following nested structure of the skeletons holds,

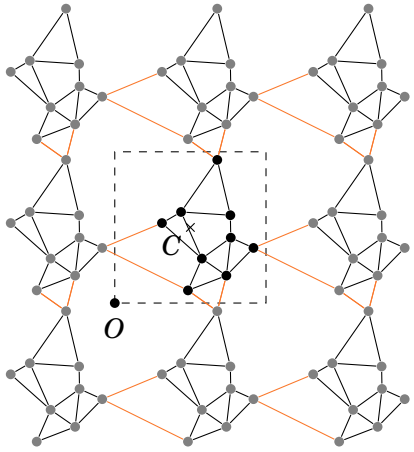
$$\mathrm{GG}(\mathcal{V}) \subseteq G_\beta(\mathcal{V}) \subseteq \mathrm{RNG}(\mathcal{V}). \tag{2.8}$$

In order to represent road networks, the β-skeletons of the node-sets are transformed into directed networks by replacing each of the skeleton's edges with two directed edges with opposite orientations. Cities, however, often have one-way streets, which would mean that not all edges would appear with their corresponding partner in the opposite direction. To simplify matters, we will model all streets as two-way.

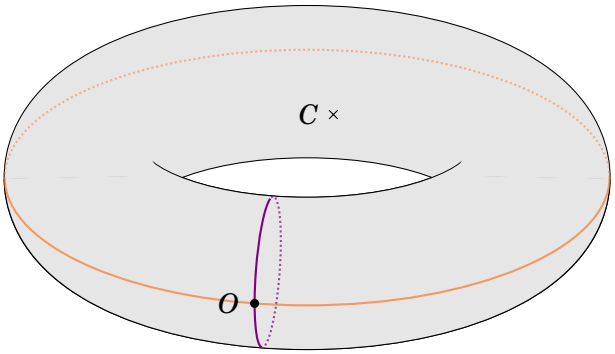
2.2.3 Boundary conditions and scaling

The synthetic networks that we construct will be used to model patches of a homogeneous road network of a specific morphological type. In order to represent a patch embedded in the network, periodic boundary conditions are imposed on the unit square where the nodes are defined by associating parallel sides with each other, essentially removing any 'hard' boundary to the region (*i.e.*, $[0,1]^2$) in which the network is defined in. The motivation for using periodic boundaries is two-fold: firstly it reduces hard-boundary effects (for example, macroscopically, flow cannot cross a hard boundary but rather must be tangential to it), as will be examined in more detail in section 2.3. Secondly, when we use the networks for traffic experiments, we may load them quite evenly with very simple origin-destination demand structures. From the point of view of a single node, the network is more isotropic, as travel demand from any other node can reach it without a drastic change of travel direction.

To achieve periodic boundary conditions, copies of the node-set are tiled around the unit square, and the β-skeleton for the super-set of nodes is constructed. Image nodes from across the boundaries are associated with their counterpart node in $\mathcal{E}$, the original node-set. The length of the cross-boundary edges is then modelled as the Euclidean distance between the image node and the real node. That is, for $\mathbf{p} \in [0,1]^2$ points shifted an integer number of times laterally or



(a) Tiling for periodicity



(b) Embedding in torus T^2

Figure 2.8 Periodic boundary conditions induce an embedding in a torus T^2. (a) Images of the original node-set are tiles around a 9 node network (*i.e.*, a *Moor* neighbourhood). Connecting nodes across boundaries imposes toroidal periodic boundary conditions. Edges that cross boundaries are coloured orange. (b) The torus showing the boundaries of the unit square as well as where the origin and the centre of the square are located in the embedding.

vertically

$$\mathbf{p} = \mathbf{p} + \mathbf{z},$$

are equivalent for any $\mathbf{z} \in \mathbb{Z}^2$. The tiling and edge association is depicted in figure 2.8.

The construction of the β-skeleton requires that all triplets of nodes are compared; each node is checked for containment in the exclusion region corresponding to all other node pairs. This means that the complexity of the (brute-force) construction of the skeleton is $\sim O(n^3)$. However since not all the nodes of the image tiles are effectively reachable, the number of operations can be reduced by considering only the nodes a few lattice-lengths from the boundaries.

Introducing periodic boundary conditions changes the topology of the space in which the network is embedded. Effectively the network has been mapped onto the unit torus T^2. The main consequence of this is to effectively reduce the scale of the networks, since nodes that are initially far away (on opposite sides of the square) become close by the corresponding unit square boundaries being associated. Therefore embedding the graphs in the torus (T^2) reduces the diameter of the graph. The farthest that two points can get from each other is by being at a corner and in the centre: this means that the diameter scales with the diagonal of the square, in the form $D(G) \sim 1/\sqrt{2}$.

In the following section we look at how increasing the order of the networks reduces the boundary effects, by examining the impact of network size on some centrality measures.

2.3 Properties of the $\alpha\beta$-networks

In this section we explore how the network structure of the $\alpha\beta$-network ensembles change as the morphological parameters α and β change. We vary $\hat{\alpha} = \alpha/\alpha_{\text{crit}}$ over the range $[0,2]$ and β between 1.2 and 1.8, and generate 100 random instances. We then consider how various ensemble statistics depend on $\hat{\alpha}$ and β. The $\hat{\alpha}$ range is chosen to capture the transition from an ordered lattice to a more uniform distribution, centred around α_{crit}. The range of β was chosen based on the observations of Osaragi and Hiraga [8], that real-world networks can be fit well with $\beta = 1.4$, as well as from visual inspection of the structure of the $\alpha\beta$-networks.

As β increases, triangles become less abundant. This is a consequence of the shape of the lunes: lunes of different β are not similar (in the geometric sense). The apices of lunes with larger β extend farther laterally in relation to the length of the axis joining the node pair. Therefore, a larger β means that nodes that are farther away (in the lateral directions) have a higher probability of being clipped by the lune and thus breaking potential triangles. This effect, of scaling the exclusion region via β, has comparable effects on the $\alpha\beta$-networks to existing constructive methods in the literature for generating complex planar networks that use resource minimisation procedures: in particular, models which limit growth to minimise resource consumption [32].

However, in our case this clipping is achieved in a much simpler manner: by the elongated shape of the lunes that extend transversely to the axis joining a given pair of nodes. The presence of additional nodes (which for larger lunes can be farther away in a perpendicular direction) in the lunes, prevents nodes that might otherwise be close from connecting: thus, preventing the formation of triangles (which in the resource-optimising models are a redundant use of resources).

To visualise the effect of morphological parameters on the number of edges, we plot histograms of the distribution of number of edges for 20 ensembles (sampled from 100 networks with $N = 225$) in figure 2.9. The most notable effect is the global reduction in edges as the griddedness of the networks is lost. The variable β simply spreads the distributions of the ensembles. From figure 2.9, we can see how selecting $\beta = 1.2$, 1.4, 1.6 and 1.8, in turn ensures that the distributions of the number of edges have minimal overlaps (so that they form a good range for experiments).

The relationship between the number of nodes and the total network length is shown in figure 2.10. We can see that the total network length scales L_{G} according to $L_{\text{G}} \sim N^{1/2}$. This is coincides with Barthélemy's [73] findings for real-world road networks.

We now turn to the morphological characteristics of our network model. First we focus on the geometrical properties of the network and then we examine various centrality measures and network statistics.

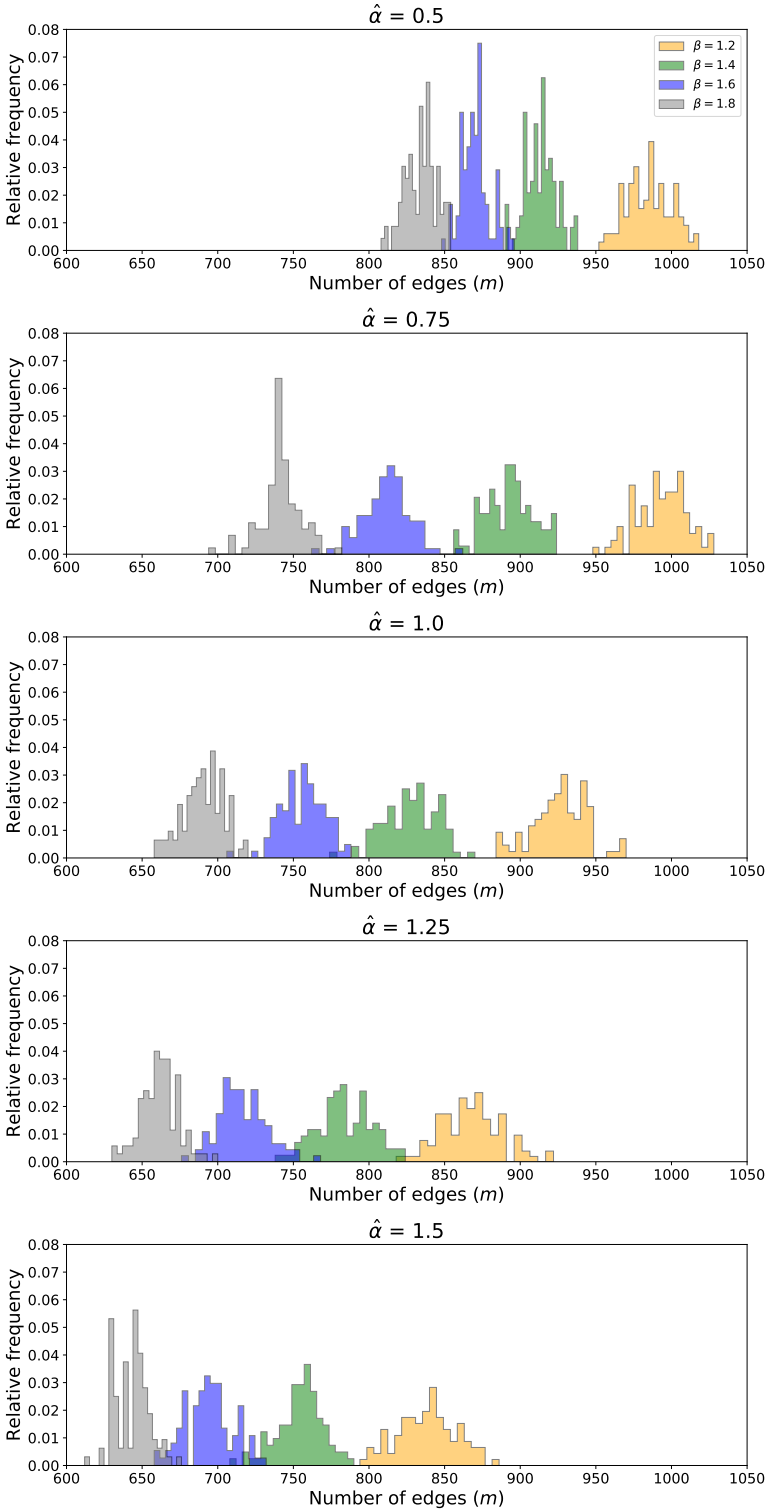


Figure 2.9 Histograms of the number of edges for 20 ensembles of networks for 225 nodes. The morphological parameters are $\hat{\alpha} = 0.5, 0.75, 1, 1.25$, and 1.75 (with $\alpha_{\text{crit}} = 0.0625$) and $\beta = 1.2, 1.4, 1.6$, and 1.8.

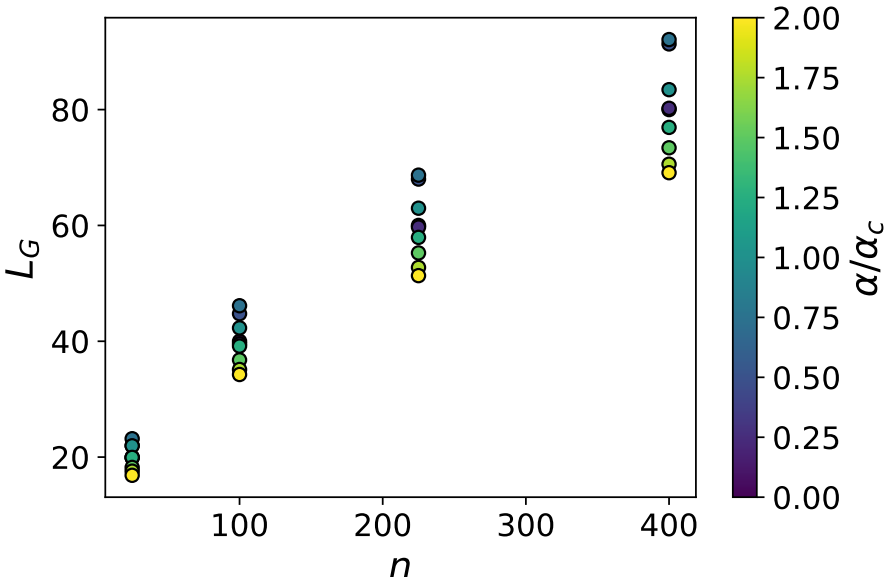


Figure 2.10 Mean total network length, L_{G}, for eight $\alpha\beta$-network ensembles as a function of the number of nodes. Here $\beta = 1.2$. Note L_{G} scales like $N^{1/2}$.

Geometrical properties

The distribution of edge lengths is governed by two factors: the distribution of the nodes, and the skeleton parameter β. The distribution of nodes determines the inter-nodal distances. It also influences how likely it is for there to be nodes in the exclusion regions of the β-skeleton. For example, in regions with high node density, long edges will be less likely to connect due to a higher probability of finding intervening nodes in the exclusion regions (the lunes). The parameter β determines the scale of the exclusion lunes relative to the distance between node pairs, with the area scaling as $\sim \beta^2$. In figure 2.11, the distributions of edge-lengths are shown for 20 ensembles (based on a 15×15 grid, *i.e.*, $N = 225$). These distributions show the effects of both increasing $\hat{\alpha}$ as well as changing β. The most notable effect is the broadening (as well as flattening) of the distributions as $\hat{\alpha}$ increases. The effects of the overlapping boxes are seen in the vanishing of the short-edge tail as α approaches α_{crit} (*i.e.*, when $\hat{\alpha} = 1$). This is a consequence of how, as the re-sampling boxes expand, the total area of multiple-box overlaps grows as well. As regions with increasing numbers of overlapping boxes appear, the number of nodes per cluster can increase as well. Thus, there are more nodes that can land in the small lunes, which reduces the relative frequency of the short edges.

The key observation drawn from the edge-length distributions is that griddedness has a significantly larger effect on the distribution of edge lengths of $\alpha\beta$-network ensembles than the β. Higher β values do slightly lower the mean edge-length of an ensemble, as is expected when the β-skeleton lunes are larger, and can thus capture more distant nodes, to prevent long edges from

wiring. However this effect is overshadowed by the increasing size of the overlapping boxes, as can be seen by the leftward shift in the distributions of figure 2.11.

Figure 2.12 shows the ensemble mean of edge lengths for 36 ensembles with different values of $\hat{\alpha}$ and β ($N = 225$). It can be seen that for small $\hat{\alpha}$, β first has a lengthening effect on the roads. This can be understood by examining figure 2.6, where increasing β yields a larger proportion of roads diagonal to the original lattice structure. From the contour plot for the standard deviation of the mean road length (figure 2.12(b)), we once again observe the effect captured in figure 2.11: $\hat{\alpha}$ is more important in modifying the distribution of edge lengths, clearly shown by the increase in the standard deviation as $\hat{\alpha}$ increases.

Degree distribution

The degree distribution of bounded networks is heavily restricted, mainly because the networks obtained from beta skeletons are planar [84]. This means that the mean degree of the nodes is bounded, since for planar graphs we have that the mean degree satisfies

$$\langle k \rangle \leq 6. \tag{2.9}$$

For toroidal graphs — when periodicity at the boundaries of $[0,1]^2$ — there is an equivalent bound [92],

$$\frac{E}{V} \leq 3, \tag{2.10}$$

which also gives $\langle k \rangle \leq 6$ since

$$\langle k \rangle = \frac{2E}{V}. \tag{2.11}$$

In contrast, for commonly used types of network models, such as Erdős-Renyi networks, Barabasi-Albert, and other non-spatial networks [68, 71, 78], the degree distribution is not as useful or as informative of network structure. Therefore, to characterise the effect of varying the construction parameters α and β, other network statistics are required: for example centrality measures or clusterng coefficients.

Centrality measures and the clustering coefficients

The definition of a vast battery of network statistics and centrality measures, that capture different aspects of network structure, has been among the important contributions of network science. For example, centrality measures try to capture the importance of the edges and nodes in a graph according to their relationship with the other nodes and edges. Depending on what is of interest, some centrality measures might be more useful than others. Masucci *et al.* [54], propose that when families of networks are studied, taking into account several centrality measures simultaneously can yield insights into differences in structure in what the authors call *multiple*

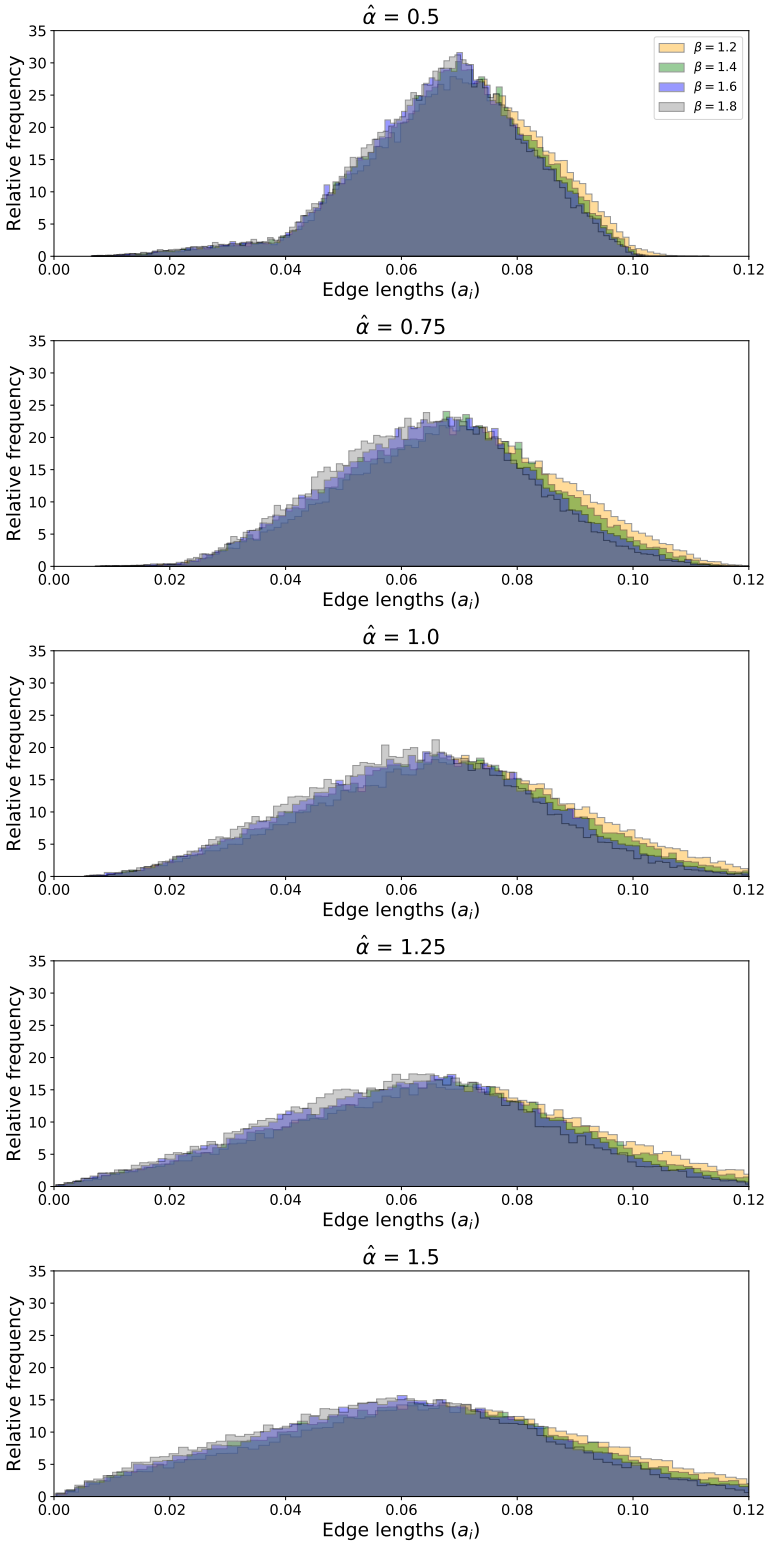
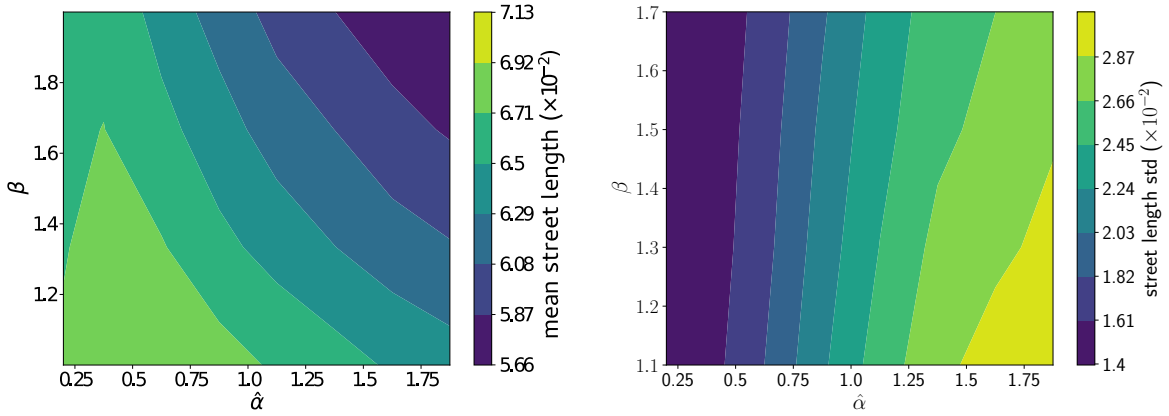


Figure 2.11 Histograms of the edge lengths for 20 ensembles of networks for 225 nodes. The morphological parameters of these ensembles are $\hat{\alpha} = 0.5, 0.75, 1, 1.25$, and 1.75 (with $\alpha_{\text{crit}} = 0.0625$) and $\beta = 1.2, 1.4, 1.6$, and 1.8. From the overlap of the distributions, it can be seen that the normalised griddedness $\hat{\alpha}$ is much more important than β in determining edge length statistics.



(a) Mean edge lengths

(b) Edge length standard deviation

Figure 2.12 (a) Ensemble mean edge lengths and (b) the standard deviation of edge lengths (b) is shown for ensembles with different values of α and β. The networks used for these figures have $N = 225$.

centrality analysis or MCA. Therefore, we examine several measures related to the networks properties that are most affected by α and β to select the most appropriate ones for describing the changing structure of the $\alpha\beta$-networks.

Crucitti *et al.* [70] identify *betweenness centrality*, *closeness centrality* and *information centrality* amongst useful measures in studying city networks. For example, they find that the information centrality distribution of city networks can be used to distinguish between planned and self-organised cities. Masucci *et al.* [54] use the closeness centrality in combination with the *clustering coefficients* to quantify higher order correlations, that is, correlations between nodes that are more than one hop away, for the London road network.

The *betweenness centrality* of a node is a measure that captures the proportion of shortest paths in a network that pass through that node, written in the form

$$C_b(v) = \sum_{i \neq j} \frac{\sigma(i,j|v)}{\sigma(i,j)}, \tag{2.12}$$

where $\sigma(i,j|v)$ is the number of shortest paths between nodes i and j that contain node v. Here, $\sigma(i,j)$ is the total number of shortest paths between i and j. In a similar way, the *edge betweenness centrality* of edge e is given by

$$C_b(e) = \sum_{i \neq j} \frac{\sigma(i,j|e)}{\sigma(i,j)}, \tag{2.13}$$

where $\sigma(i,j|e)$ is the number of paths that contain edge e.

Closeness centrality is a measure of how close a node is to the rest of the network,

$$C_c(i) = \frac{N-1}{\sum_{j \neq i} d(j,i)},$$

where $d(j,i)$ is the shortest-path distance between nodes j and i. Closeness centrality is of special interest since it is a (transformed) version of the *index of integration* [93] concept from the *space syntax* discipline. This implies that closeness centrality can potentially serve as a methodological tool to bridge the conceptual gaps between the transport, and urban and architectural fields, although we will not pursue this line of enquiry further.

Information centrality ([94] includes a discussion in relation to other centrality measures) captures the decrease in network efficiency if the edges incident at node v are removed. It can be defined as follows,

$$C_i(v) = \frac{\Delta E}{E} = \frac{E(G) - E(G')}{E(G)}. \tag{2.14}$$

Here $E(G)$ is the network efficiency of graph G, and G' is the subgraph obtained by removing node v along with the edges incident to it. The network efficiency itself is given by

$$E(G) = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{d(i,j)}, \tag{2.15}$$

where $d(i,j)$ is the length of the shortest path between nodes i and j. The efficiency is the average of the reciprocal shortest path, with high efficiencies obtained for short path lengths.

The information centrality of a node captures the increase of the length of shortest paths by having to avoid using that node. Nodes with high betweenness centrality have high information centrality as well. However, the converse is not necessarily true, since the efficiency also depends on the lengths of the paths. Thus if the removal of a node causes the next best paths to be much larger for G' than for G, it can also have a large information centrality.

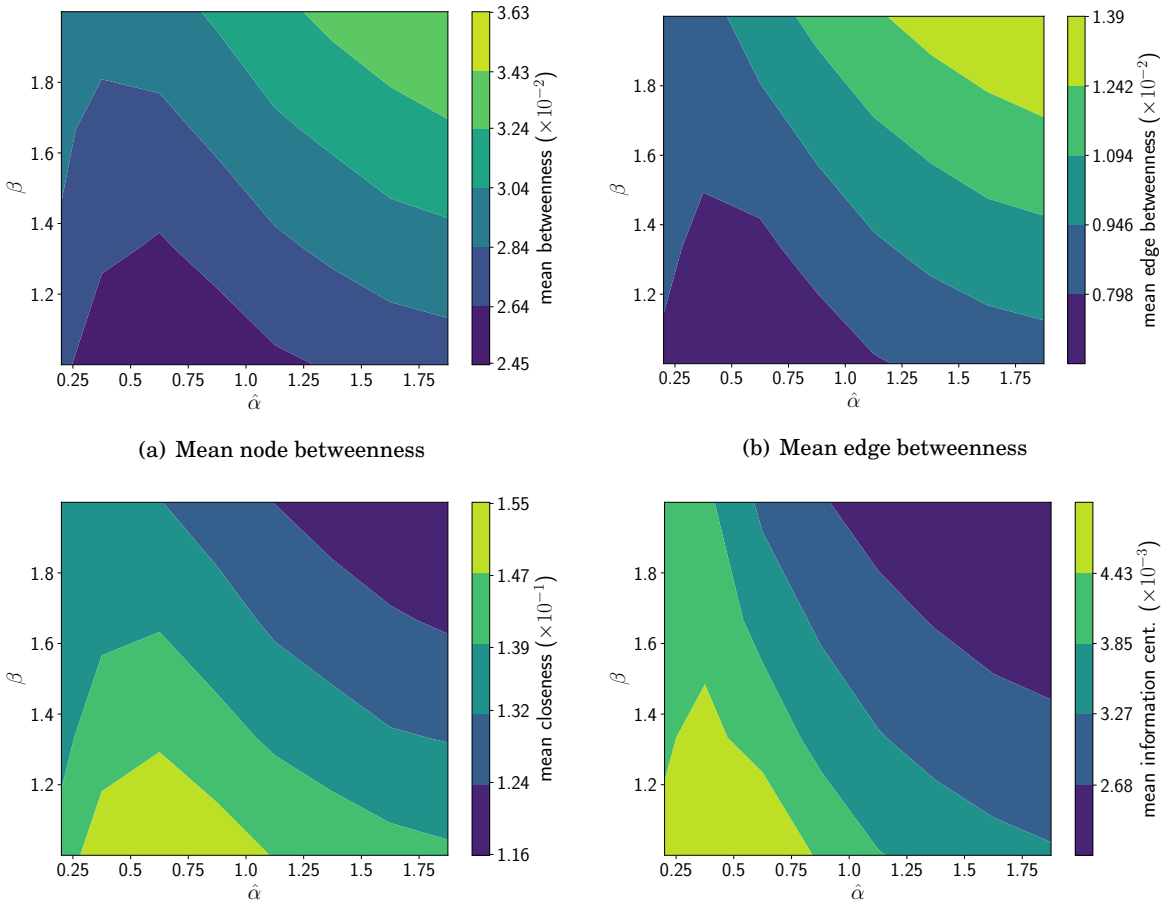
As mentioned above, a limiting case of the β-skeleton is a Delaunay triangulation (as $\beta \to 0$), and increasing β has the effect of removing edges in general from a given network. This means that β also influences the amount of triangles in a network. The clustering coefficient of a node measures how many triangles the node belongs to, in proportion to the maximum number of triangles it could form with its neighbours, in the form [69]

$$CC(v) = \frac{2\tau_v}{k_v(k_v - 1)}. \tag{2.16}$$

Here, τ_v is the number of subgraphs of G with three edges and three nodes (*i.e.*, triangles), where one of the nodes of the subgraph is v and the other two are its neighbours. The denominator counts the maximum number of triangles that these nodes can form. To characterise the prevalence of triangles across a whole network, we will use the network average clustering coefficient,

$$\langle CC \rangle_{\mathcal{N}} = \frac{1}{n} \sum_{v \in \mathcal{V}} CC(v). \tag{2.17}$$

In figure 2.13, heat maps of the ensemble means of edge and node betweenness, and closeness as well as information centralities are shown for ensembles of $\alpha\beta$-networks with $N = 100$.



(a) Mean node betweenness

(b) Mean edge betweenness

(c) Mean closeness

(d) Mean information betweenness

Figure 2.13 Mean topological statistics are shown for ensembles with $N = 225$ and different values of $\hat{\alpha}$ (from 0 to 2 in steps of size 0.25) and β (1.2, 1.4, 1.6, and 1.8). That is the contour interpolation is done for 36 ensemble means, where 100 networks per ensemble were sampled. (a) betweenness, (b) edge betweenness, (c) closeness, and (d) the information betweenness.

In terms of the dependency on $\hat{\alpha}$, it is clear that the sharp decrease in the node distance around $\hat{\alpha} = 0.5$ (see figure 2.12(a) above) also affects these centrality measures across all values of β. The edge and node betweenness increases with $\hat{\alpha}$ for $\hat{\alpha} > 0.5$, as can be seen in the figure by the transition to lighter colours. The closeness and the information betweenness experience the opposite effect as the networks achieve more complex structures, as can be seen in the exemplar networks of figure 2.6 above. For example, a feature that directly affects the closeness and the information betweenness is the higher abundance of nodes of degree one for larger $\hat{\alpha}$ and β.

For networks with periodic boundary conditions, the distribution of node degree is fairly homogeneous. Since there are no peripheral nodes, there is more uniformity in the adjacency properties of the nodes. In contrast, for the bounded case, the nodes at the boundaries of the unit square have less neighbours; they are farther away from the bulk of the network.

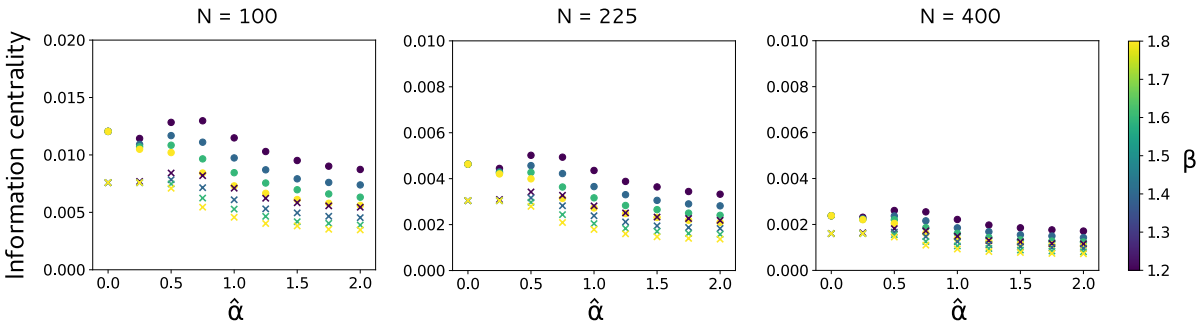


Figure 2.14 The information centrality for ensembles of three different networks sizes are shown. The different markers represent the type of boundary condition: circles for periodic (toroidal) boundaries and crosses for bounded networks. As the ensemble size grows it is clear how the difference between the centrality diminishes for both different periodic conditions as well as amongst themselves. As the ensembles grow from $N = 100$ to $N = 400$, there is a drop of about an order of magnitude in the spread of information centrality.

Information centrality is sensitive to boundary effects. Thus, we can use the information centrality to determine appropriate experimental sizes for networks, so that boundary effects are small. In figure 2.14 we see that as the ensemble size increases, the difference in information centrality — both in magnitude as well as in relative terms — becomes small quickly. From the decrease in the order of magnitude in the spread of the information centrality, we can conclude that ensembles of $N = 225$ are large enough for boundary effects to have decayed to a large extent.

In figure 2.15, we display the clustering coefficient (CC) as a function of $\hat{\alpha}$ and β. For $\hat{\alpha} = 0$ the network is a square grid and therefore there are no triangles. This means that the clustering coefficient is also zero. As $\hat{\alpha}$ increases to 0.75 there is a nonlinear increase in the CC which peaks around $0.75 \leq \hat{\alpha} \leq 1.0$. Ensembles with low β achieve higher CC values, and then as $\hat{\alpha}$ keeps increasing, the CC appears to decay towards a constant value. For lower β there are more triangles as the β-skeleton $G_\beta(\mathcal{V})$ becomes a larger subgraph of the Delaunay triangulation since,

$$\mathrm{DT}(\mathcal{V}) \subseteq \mathrm{GG}(\mathcal{V}) \subseteq G_\beta(\mathcal{V}). \tag{2.18}$$

We can see from figure 2.15(b), that the CC variation with respect to β, although always quadratic, depends heavily on $\hat{\alpha}$. The (decreasing) quadratic dependence on β can be understood with simple scaling arguments: β parametrises a linear variable of the lune of the β-skeleton (which grows with β), therefore the probability of it being empty (approximating the density of nodes as uniform) varies like $-\beta^2$.

In figure 2.16 we visualise the landscape of the clustering coefficient (CC) above the $\hat{\alpha}\beta$-plane. The CC goes through a transition as the boxes overlap. Compared to the other network metrics above, the CC clearly captures the overlapping of the re-sampling boxes in terms of network properties: there is a clear transition point (captured by the peak) at $\hat{\alpha} = 1$. The CC, therefore,

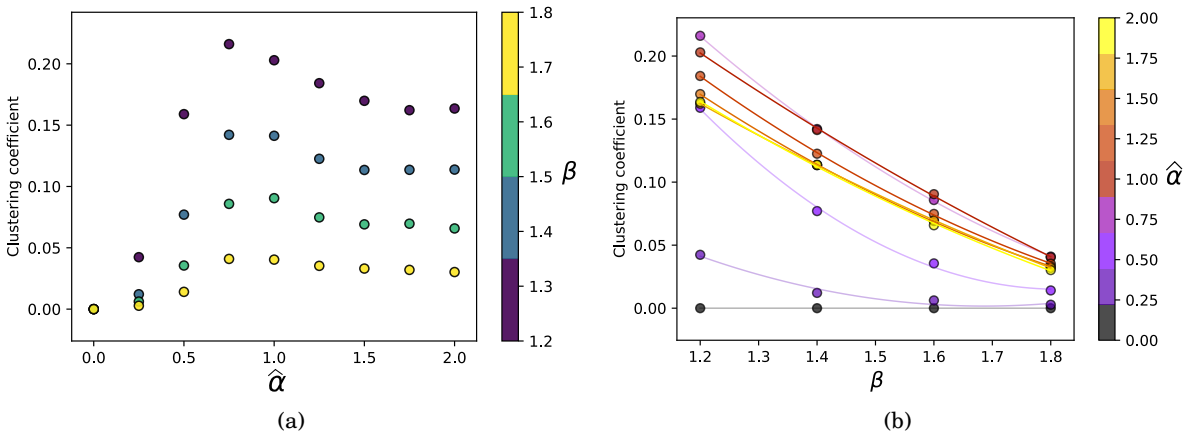


Figure 2.15 The clustering coefficient as a function of (a) $\hat{\alpha}$ and (b) β. Network size is $N = 400$ with periodic boundary conditions, 100 networks were sampled per ensemble. Observe quadratic curves of fit for the ensembles (grouped and coloured according $\hat{\alpha}$).

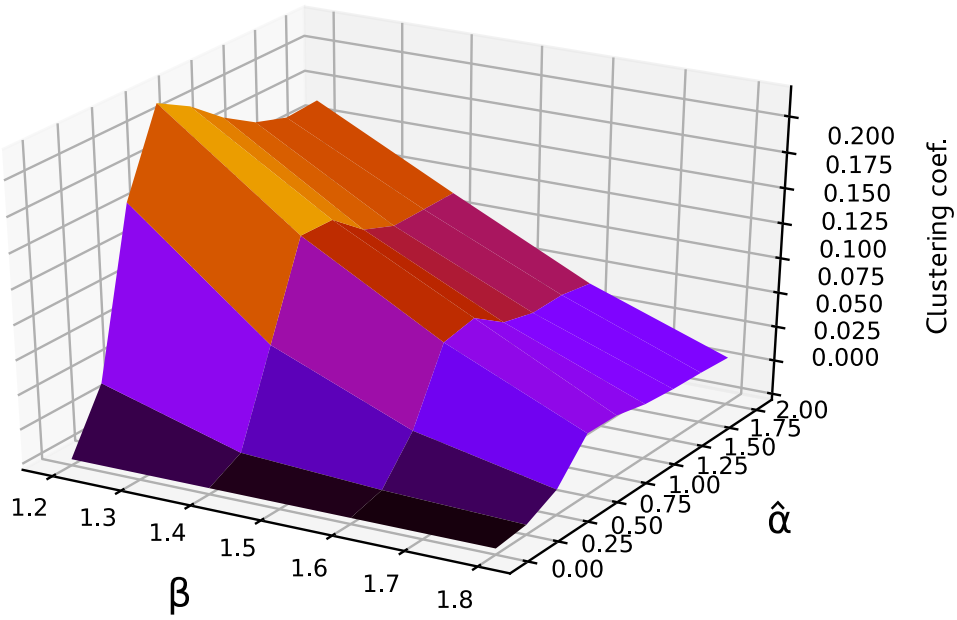


Figure 2.16 Approximation of the mean clustering coefficient surface for network size $N = 400$ ensemble, calculated by taking ensemble means for 100 network samples per ensemble. Shading of surface is according to clustering coefficient value.

tracks the break-down of the lattice order structure. Furthermore, it also captures the subsequent change in node distribution as it approaches uniformity.

2.4 Point Processes, Intensities, and Entropy

So far we have approached the construction of the node-set as a perturbed lattice, where the noise is modulated by α. We now re-examine the construction process, comparing it with a *spatial point processes*. The comparison will highlight the resulting distributions of the nodes inside the unit square.

There are two reasons why the comparison with point processes is important in its own right. Firstly, the use of stochastic processes (such as point processes and line processes [95]) is widespread in the field of wireless communications, so it is useful to have comparisons with standard models from similar application fields. Secondly, stochastic point processes are used as flexible ways of generating point distributions, which has made them important in theoretical studies of spatially embedded graphs (see for example, Penrose's work on random geometric graphs [96]).

We begin by approximating our re-sampling model with a *Poisson point process*. The construction of the equivalent point-process model will not be stationary (except perhaps in the large n limit, which also implies an area scaling).

The expected number of points in a region $A \subseteq [0,1]^2$ can be written as the sum of the indicator function for each of the re-sampling boxes, in the form

$$\mathbb{E}(|V \cap A|) = \frac{1}{\alpha^2} \int_A \sum_{k=1}^{n} \mathbb{1}_{S_k} \mathrm{d}x\mathrm{d}y. \tag{2.19}$$

Here $\mathbb{E}(|V \cap A|)$ is the expected number of nodes found in region A, for an instance of an $\alpha\beta$-network. Essentially, this is the expected number of boxes that each point in A belongs to (given by the sum of indicator function $\mathbb{1}_{S_k}$), multiplied by the probability density of each box (which is a function of α, since it is uniform over each S_k).

The indicator function for each S_k is a square step function in 2D, which can be expressed in terms of of the Heaviside function Θ, in the form

$$\mathbb{1}_{S_k}(\mathbf{x}) = [\Theta(x - l_k) - \Theta(x - r_k)][\Theta(y - b_k) - \Theta(y - t_k)]. \tag{2.20}$$

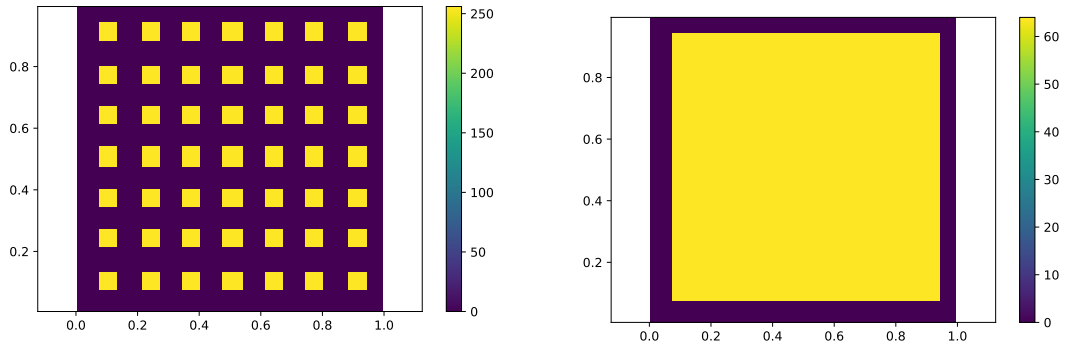
here the Heaviside function is given by

$$\Theta(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1, & \text{otherwise.} \end{cases}$$

Figure 2.17 shows how this changes as α grows. Substituting expression 2.20 in the sum of equation 2.19 yields

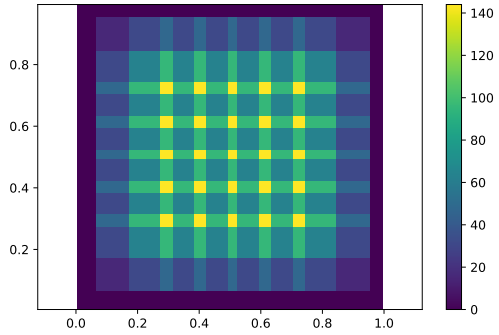
$$\mathbb{E}(|V \cap A|) = \frac{1}{\alpha^2} \int_A \sum_{k=1}^{n} [\Theta(x - l_k) - \Theta(x - r_k)][\Theta(y - b_k) - \Theta(y - t_k)] \mathrm{d}x\, \mathrm{d}y. \tag{2.21}$$

The expected value of the number of points can be used as an intensity measure of the *Poisson point process*, which means that equation 2.21 can be used as an intensity to approximate the

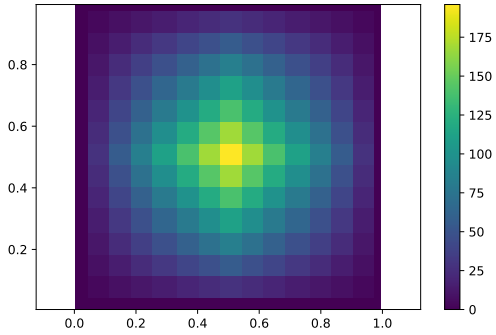


(a) $\alpha = \frac{1}{2}\alpha_{\text{crit}}$ ($\hat{\alpha} = 1/2$)

(b) $\alpha = \alpha_{\text{crit}}$ ($\hat{\alpha} = 1$)



(c) $\alpha = 2\alpha_{\text{crit}}$ ($\hat{\alpha} = 2$)



(d) $\alpha = 4\alpha_{\text{crit}}$ ($\hat{\alpha} = 4$)

Figure 2.17 The sum of densities $\sum_k \mathbb{1}_{S_k}(\mathbf{x})$ function of points for different values of α. In (b), when $\alpha = \alpha_{\text{crit}}$ the re-sampling boxes begin to overlap and while the distribution looks it, it is not a uniform distribution over the central square since each node has to be contained inside its α-box. The probability density functions are reminiscent of the diffraction pattern due to a square grating.

point sets of our model with an inhomogeneous Poisson process. While this can be useful, it means that only the average number of nodes, $\mu(A) = \mathbb{E}(|A \cap V|)$,on a network can be specified,

In effect, equation 2.21 integrates to $\mathbb{E}(|A \cap V|) = N$, which is the number of nodes specified by the original lattice. In summary, a set of points obtained from an inhomogeneous Poisson process will have the same average number of nodes as the $\alpha\beta$-networks. The difference between the two approaches is that there are no constraints on how many points can belong to each box, allowing for greater clustering of the points, which in turn is important for the construction of the β-skeletons.

In order to recover an equivalent formulation to our method for constructing the $\alpha\beta$-networks, we have to condition the process on there being a single point for each re-sampling box, which would yield a *binomial point process*.

We can take equation 2.21 as a probability density function,

$$\rho(x,y) = \frac{1}{N\alpha^2} \sum_{k=1}^{N} [\Theta(x-l_k) - \Theta(x-r_k)][\Theta(y-b_k) - \Theta(y-t_k)], \tag{2.22}$$

that approximates the distribution of nodes. For a given node-set $\mathcal{V}$, we can calculate the entropy of the spatial distribution of the nodes as

$$H(\mathcal{V}) = -\sum_{v \in \mathcal{V}} p_v \log p_v, \tag{2.23}$$

where the probabilities p_v are given by $p_v = \rho(v_x, v_y)$. This way, for $\alpha = 0$, when there are no overlaps of the re-sampling boxes, all the p_v's are the same, that is, $p_v = 1/N$. This leads to an entropy of

$$H\big|_{\alpha=0} = \log N. \tag{2.24}$$

At the other extreme of griddedness $\alpha = 1$, we also have the $p_v = 1/N$; we recover the uniform density of nodes. The entropy of the set of nodes also yields $H\big|_{\alpha=1} = \log N$.

The entropy of the node distributions measures how 'surprised' we should be of a particular distribution. In the case of the perfect grid, the positions of the nodes are prescribed and therefore there is only one configuration, leading to a baseline value of entropy for when none of the re-sampling boxes overlap. For the uniform distribution of nodes, any position is as likely as any other, so no distribution is any more surprising than any other. For uniform probabilities then, the entropy takes the minimum value of $\log N$.

For intermediate values of α, because of the overlaps of the re-sampling boxes (the Heaviside functions), the entropy of the node-sets can take different values than $H = \log n$. For example, if we do not get many points in the overlaps, it means that there are more terms with low p_v, which increases the entropy of the distribution. Occurrences of events with low p_v, which are 'surprising', increase the entropy. These behaviours can be seen in figure 2.18.

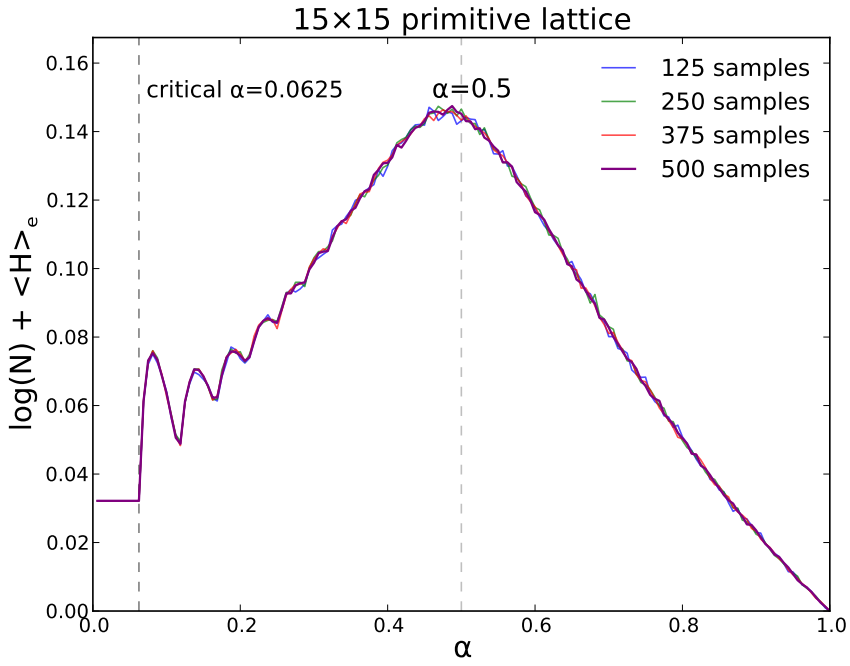


Figure 2.18 Mean entropy estimate $\bar{H} = \log N + \langle H(\mathcal{N})\rangle_e$ for increasing α. The different curves are approximations of the ensemble entropy for different numbers of sampled node-sets; the purple (smoothest) curve is for 500 samples. The order transitions as the re-sampling boxes overlap with each other can be clearly seen starting with the first at α_{crit}, the subsequent humps are due to overlaps at higher order (*e.g.*, when boxes overlap with second neighbours).

2.5 Discussion

In this chapter we have discussed the construction of a random planar network model that is used for generating experimental ensembles of networks throughout the rest of this thesis. We have discussed where the model is situated in terms of existing models in the literature and what has motivated its design choices.

The network measure, that will be used as a descriptive variable for the structure of the networks, is the clustering coefficient, as we have shown it to capture the break-down of the lattice order and the subsequent transition as the node distribution approaches uniformity. We have also identified, that in order to compare networks of different sizes, the *normalised griddedness* $\hat{\alpha} = \alpha/\alpha_{\text{crit}}$ is most appropriate.

In the literature there is a clear, and sometimes explicit, call for more network models that can serve as "toy models of road networks" (Aldous and Shun, [97]). The evidence is also clear that the effects of network structure on transport performance need to be better understood. Rather than single toy network models, what is needed are a variety of models that each account for specific important properties of networks. The model presented here is designed to investigate the role of spatial order and edge redundancy in transportation networks, as well as to answer

the call to develop detailed, purposeful 'toy' models. In the following chapters layers, will be added to the model to allow for several computational experiments revolving around traffic assignment.

In summary, the key contributions of this chapter can be captured in the following statements:

C2.1 The development of a network model, in the way set out by O'Hare [44], for the experimental treatment of traffic assignment in complex networks, with a focus on the effects of morphology on transportation.

C2.2 The network model recovers broad statistical similarities with mid-size networks used commonly in the literature, as well as picking up on qualitative similarities (for example the relative proportions of quadrangles to triangles).

C2.3 We have identified the clustering coefficient (CC) as one of the main attributes of morphology for spatially embedded 'planar' networks, with the model capturing a transition in the CC due to a smooth transition to stochasticity.

CHAPTER 3

TRAFFIC ASSIGNMENT ON NETWORK ENSEMBLES

This chapter consists of two main parts. The first part — section 3.1 up to section 3.4 — consists of preparing the $\alpha\beta$-networks that we developed in chapter 2 for solving the static traffic assignment problem (STAP) on them. We discuss parameters that define the cost functions of the network edges, the choice of *origin-destination* (OD) pairs, and the demand range for numerical experiments. We also explain how to associate units of length (*i.e.*, the scale) and traffic volumes to the $\alpha\beta$-networks, to enable quantitative comparisons with other (even real-world) network models. The overall aim is to prepare the $\alpha\beta$-networks for use in as broad a range of traffic applications as possible (exemplified in chapters 4 and 5). The second part of this chapter (section 3.5) presents a numerical investigation of the Wardropian equilibria that emerge on the $\alpha\beta$-networks through the solution of the STAP on them. This investigation focuses on the *price of anarchy* (PoA) as a function of the travel demand on the networks. The aim is to determine the sensitivity of the PoA to *classes* of network structure by carrying out numerical experiments on ensembles of a selected range of $\hat{\alpha} = \alpha/\alpha_{\text{crit}}$ and β.

We begin by introducing the context of the STAP in section 3.1, touching on concepts of Wardropian equilibrium, commodity flows, congestion games, and optimisation. Section 3.2 gives the STAP's mathematical formulation, following Patriksson [59].

Section 3.3 details the link-cost functions for the $\alpha\beta$-networks and discusses units, dimensions and scale of the networks. The allocation of network supply that we develop takes into account the local network structure at intersections. By considering a finite amount of road resource, it guarantees that each road's sensitivity to congestion depends on the capacity of the intersection it feeds. This 'resource allocation' heuristic also guarantees that $\alpha\beta$-networks within an ensemble (which may have different numbers of edges), as well as across different ensembles, are comparable in terms of their infrastructure supply.

Section 3.4 discusses the appropriate choice of origin-destination (OD) pairs. We show how the OD pairs are selected so that the flows 'explore' the network fully. Since in the experiments that follow we will make heavy use of the PoA as a signal for studying morphological effects, it is important for the traffic flows to load the networks as evenly as possible.

Section 3.5 presents the results of the computational experiments prepared in the previous sections. We first compare the effects of the resource allocation heuristic of section 3.3 with the existing methodology from the literature for assigning cost-function parameters. We also touch upon the distributions of vehicles on the roads and the delays due to congestion as travel demand on the networks increases. Finally, in section 3.6 the contributions stemming from this chapter are discussed.

3.1 Background

The STAP provides a simple way of approaching traffic routing on networks. Due to its general formulation, it applies also to commodity flow settings [59]. It is a well-understood model that is easy to implement and has high explicative power. In the context of road traffic, it is ultimately based on intuitive behavioural and optimisation principles (see Wardrop's criteria p42). In this regard, it embodies the key ideas of complexity science: simple interactions (of the road users with each other via the roads they share) give rise to pattern formation, which in this case is the equilibrium traffic state of the network (to which each user contributes, due to the roads being sensitive to congestion). In contrast to stochastic user equilibrium approaches (SUE) [98] (another well-known traffic equilibrium paradigm), the STAP does not need to compute routes through the network explicitly, but rather can be computed in terms of link flows alone — this is a significant simplification that makes STAP most suitable for our parsimonious approach.

The STAP models 'users' that need to get from their node of origin to their destination, and that try do so by choosing a minimal-cost route through the road network. Each user's choice of route depends on how congested the network is, as this has a direct impact on their travel time. In its simplest formulations, like the one we use here, the STAP assumes that users have perfect knowledge of the network's current traffic state and that the travel costs that users experience are only affected by the traffic volumes on links that lie on their chosen route. The word *static* makes explicit that the STAP takes a steady-state view of journeys being made per-unit time on the network. This, in turn, makes the traffic flows akin to stationary currents: traffic flows and costs that we consider are costs and flows per-unit time in a time-invariant world (see [99], for a detailed discussion).

Perhaps the best way of thinking of equilibrium in transport networks is to appeal to intuition and common sense. After all, agents that compose real transport networks are human. This is what Wardrop's two principles of traffic equilibrium [38] try to capture:

1. All used routes between an origin and a destination cost the same, while all unused routes

have equal or greater cost.

2. The average costs for users in the network are minimised.

Each principle (or criterion) yields a different traffic assignment. The first describes *user equilibrium* (UE) where users are selfishly driven to minimise their personal travel times. As such, with perfect information or after learning the network, all users originating and terminating their journeys at the same nodes experience the same travel times. The rationale is that if they were aware of a lower cost route, users would switch to it, thus, equilibrating the costs of the high-cost and low-cost routes. The second principle describes a *system optimal* (SO) equilibrium, where minimal average costs per user also implies minimal total system cost. The difference between UE and SO assignments is at the core of the concept of PoA as we discuss below, and it will also be central to our work in chapter 4.

We now define the stylised concept of road network which we use in the remainder of this thesis. We take road networks to be composed of a graph (chosen from the $\alpha\beta$-network family) together with a vector of cost functions, having as elements the cost function on each edge of the network.

Definition 3.1.1 (Road network)**.** *A* road network*, $\mathcal{N}$, is the pair*

$$\mathcal{N} = (G, \mathbf{c}). \tag{3.1}$$

The graph G (defined as in 2.1.1 of chapter 2, $G(\mathcal{V}, \mathcal{E})$) represents the structure. The components of the cost-function vector $\mathbf{c}$ *are the travel-cost functions associated to each link.*

In general the cost function $c_i : \mathbb{R}^{|\mathcal{E}|} \to \mathbb{R}$ of each link can depend on the flows on every other link and is therefore a function of the traffic pattern, or flow vector $\mathbf{x} \in \mathbb{R}^{|\mathcal{E}|}$. We can now give a more explicit statement of the STAP.

Definition 3.1.2 (Static traffic assignment problem: STAP)**.** *We are given a road network $\mathcal{N}$, a set of origin-destination pairs Ω, and specified travel demands d_ω for each $\omega \in \Omega$. The STAP consists of finding a traffic pattern $\mathbf{x}^\star$ that satisfies one of Wardrop's principles, and meets the trip demands.*

Among the first to interpret the traffic assignment problem in game theoretic terms were Charnes and Cooper [39], where players are OD pairs (we will revisit similar ideas in chapter 4). In contrast, Rosenthal [100] considers a finite number of discrete players that are the users of the network themselves, their strategies being the routes they take. In the limit of an infinite number of players, congestion games are known to coincide with Wardropian equilibria. Patriksson gives a comprehensive review of the development of congestion games and traffic assignment in [59], highlighting parallels between Kirchhoff's circuit laws and commodity flows. The correspondence between congestion and potential games, where a potential function to be minimised captures the players' interactions, is dealt with in detail by Rosenthal and by Monderer and Shapley in [101].

When thinking of users' transport route choice as a congestion game, the natural interpretation we adopt is to consider the network users as the players. Each user's chosen strategy induces a choice of resources: the streets of the network along their chosen route. In turn, the cost of a particular strategy depends on the congestion (which is simply the flow volume) on the network, which in turn is determined by the aggregate strategies of all the players on the network. The costs incurred by players can be obtained from a single function, the Beckmann functional, which also makes this formulation of the STAP a potential game.

In summary, the result of the collective interactions of the players have to be specified: this is done via flow-dependent cost functions associated to travelling down a street (edge). Each (infinitesimal) player's cost is affected by the strategy (or route choice) of the others. The cost (or *disutility*) to each player is their individual travel time. The way in which the strategies of the players interact is reflected in the edge costs, due to their sensitivity to congestion. Before turning to to the choice of cost functions and OD pairs for the $\alpha\beta$-networks, we briefly go over the basics of the STAP formulation.

3.2 Static Traffic Assignment

The STAP can be expressed as an optimisation problem with linear constraints [53]. A comprehensive treatment of models and methods related to static assignment is presented by Patriksson [59], of which, we are interested in the most simple way of setting the problem for the $\alpha\beta$-networks. We now express the solution of the STAP as an optimisation problem.

The SO assignment, which corresponds to Wardrop's second criterion, is achieved when the function minimised is the *total cost*

$$\Phi_{\mathrm{SO}}(\mathbf{x}) = \sum_i c_i(x_i)x_i. \tag{3.2}$$

For user equilibrium (UE), in which the users are assumed to minimise their personal travel costs, the objective function is the Beckmann functional [102]

$$\Phi_{\mathrm{UE}}(\mathbf{x}) = \sum_i \int_0^{x_i} c_i(\xi_i)\,\mathrm{d}\xi_i. \tag{3.3}$$

In the Beckmann functional, integration of the cost functions, up to the level of flow they carry, accounts for cost incurred by a user that arrives at the link that already carries x_i units of flow. The area under the curve does not take into account the cost imposed on other users already on the link. This contrasts with the objective function for SO which considers the full cost to all users.

For affine cost functions c_i of the form

$$c_i(x_i) = a_i + b_i x_i, \tag{3.4}$$

which we use exclusively in this thesis, the total cost function can be expressed compactly in terms of the cost-function parameters, a_i and b_i, by first assembling parameter vectors $\mathbf{a}$ and $\mathbf{b}$

that have for the i-th element the corresponding parameter of the cost function for edge i. We can then write the total cost function of the system in the form

$$\Phi_{\mathrm{SO}}(\mathbf{x}) = \sum_{i=1}^{m} x_i c_i(x_i) = \mathbf{x}\cdot\mathbf{a} + \mathbf{x}\cdot D_{\mathbf{b}}\,\mathbf{x}, \tag{3.5}$$

where $D_{\mathbf{b}}$ is the square matrix with diagonal elements matching those of $\mathbf{b}$ and all off-diagonal entries set to zero, in the form

$$D_{\mathbf{b}} = \begin{pmatrix} b_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & b_m \end{pmatrix}. \tag{3.6}$$

In the case of UE (when the cost functions are affine), then the Beckmann functional can be expressed (from equation 3.3) as

$$\Phi_{\mathrm{UE}}(\mathbf{x}) = \mathbf{x}\cdot\mathbf{a} + \frac{1}{2}\mathbf{x}\cdot D_{\mathbf{b}}\,\mathbf{x}, \tag{3.7}$$

which does not incorporate the marginal cost caused to others by choosing a particular path. The coefficient of 1/2 in the quadratic term reflects the smaller contributions (compared to SO) that users contribute to the objective function. As introduced above, the cost inflicted by any user on the others 'already' on the same link is not considered (see Roughgarden [43] for a full treatment of the relationship between UE and SO assignments).

Now that we have defined the objective functions, we consider the constraints. We start by considering the case for a single OD pair. For the flow variables to be physically meaningful we require that $x_i \geq 0$ for all edges ($i = 1, \dots, m$). The other requirement is that flow is conserved at intersections. That is, the incoming and outgoing flows at the intersections have to cancel out, once trips that originate or terminate at the intersection nodes have been accounted for. The conservation constraints can be written compactly in terms of the incidence matrix S of the network and a demand vector, $\mathbf{d}$, defined as

$$d_i = \begin{cases} -d, & \text{if node } i \text{ is the origin,} \\ d, & \text{if node } i \text{ is the destination, ,} \\ 0, & \text{otherwise,} \end{cases} \tag{3.8}$$

where d is the magnitude of demand associated to the OD pair. The directed incidence matrix (also known as the node-link incidence matrix), S, encodes the structure of the road network and can be defined as

$$S_{ij} = \begin{cases} -1, & \text{if edge } j \text{ is outgoing at node } i, \\ 1, & \text{if edge } j \text{ is incoming at node } i, . \\ 0, & \text{otherwise.} \end{cases}$$

Ensuring that the incoming flow at each node equilibrates the outgoing flow can be expressed as the matrix equation

$$S\,\mathbf{x} = \mathbf{d}. \tag{3.9}$$

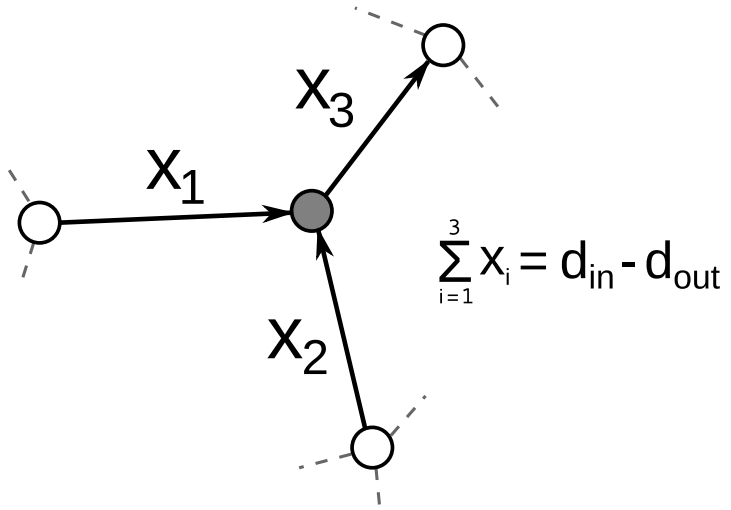


Figure 3.1 The sum of incoming and outgoing flows to each intersection node equal the total demand that originates and terminates there. The conservation equation shown for the central node is assembled with that for other nodes to give equation 3.9.

The left hand side of each row is the difference of incoming and outgoing flows, where $\mathscr{I}_k$ is the set of links incoming to node k and $\mathscr{O}_k$ is the set of outgoing links for node k. The right hand side of equation 3.9, is zero except at the origin and destination, where the sign of the right hand side ensures that d units of flow either originate or terminate at node k accordingly. Thus, more explicitly, equation 3.9 is equivalent to the N equations,

$$\sum_{j \in \mathscr{I}_k} x_j - \sum_{j \in \mathscr{O}_k} x_j = \begin{cases} -d, & \text{if } k \text{ is the origin,} \\ d, & \text{if } k \text{ is the destination,} \\ 0, & \text{otherwise,} \end{cases} \tag{3.10}$$

for the index $k = 1, \ldots, N$, running over all nodes.

Figure 3.1 depicts the constraints at an exemplar node. We use the convention of demand originating at the node (d_{out} in the figure) as negative and the terminating demand (d_{in}) as positive to match the definition of S.

We can now specify the STAP (for a single OD) as a quadratic optimisation programme by replacing the objective function $\Phi(\mathbf{x})$ with $\Phi_{\text{UE}}(\mathbf{x})$ or $\Phi_{\text{SO}}(\mathbf{x})$, depending on whether we solve for UE or SO,

$$\begin{aligned} \underset{\mathbf{x}}{\text{Minimise}} \quad & \Phi(\mathbf{x}) \\ \text{Subject to} \quad & S\mathbf{x} = \mathbf{d} \\ & \mathbf{x} \geq 0. \end{aligned} \tag{3.11}$$

The use of affine cost functions that simplifies the STAP into a quadratic optimisation problem. Added to that, the constraints are captured by a linear matrix equation. Thus, the STAP, as

expressed by the minimisation problem 3.11, is a quadratic optimisation program with affine constraints.

For multiple OD pair, more care must be taken. In this case, for each OD pair ω in a given set Ω of OD pairs, we can consider the total flow on each link to be composed of the flows corresponding to all OD pairs, that is

$$x_i = \sum_{\omega \in \Omega} x_{i\omega}, \quad \forall \omega \in \Omega, \tag{3.12}$$

where $x_{i\omega}$ corresponds to the flow on link i due to OD pair ω. Thus, ensuring flow conservation at the nodes requires an equation equivalent to equation 3.9 for each OD pair, which results in $|\Omega|$ matrix equations

$$S\,\mathbf{x}_\omega = \mathbf{d}_\omega, \quad \forall \omega \in \Omega. \tag{3.13}$$

Here, $\mathbf{x}_\omega$ is the vector of flows on the network associated to OD pair ω, and the demand vectors $\mathbf{d}_\omega$ are defined as in equation 3.8, but now corresponding to each ω. Therefore, for multiple OD pairs the STAP can be expressed as,

$$\begin{aligned} \underset{\mathbf{x}}{\text{Minimise}} \quad & \Phi(\mathbf{x}) \\ \text{Subject to} \quad & S\,\mathbf{x} = \mathbf{d}_\omega, \quad \forall \omega \in \Omega \\ & \mathbf{x}_\omega \geq 0, \quad \forall \omega \in \Omega \\ & \mathbf{x} = \sum_{\omega \in \Omega} \mathbf{x}_\omega, . \end{aligned} \tag{3.14}$$

The relationship between the link-flow formulation of the STAP and and a formulation based on route flows is presented in detail by Patriksson in [59], as well as as an extensive discussion on algorithms and general methods. As mentioned previously, by using affine cost functions the STAP becomes a quadratic optimisation problem, for which there are a number of general-purpose industry-standard numerical solvers (we use Gurobi [103], which for quadratic optimisation problems uses a barrier algorithm, see for example [104]). In the implementation of multiple ODs (see chapter 5), we restrict ourselves to having the nodes being at most an origin and destination to a single flow.

Price of Anarchy

The price of anarchy (discussed extensively in [43]) is a measure for quantifying how far away a user equilibrium traffic assignment is from system optimal.

Definition 3.2.1 (Price of anarchy)**.** *The price of anarchy (PoA) for a given transport network, $\mathcal{N}$, under a particular demand* $\mathbf{d}$ *and OD set Ω is the ratio of the total system costs under UE flows*

compared to the optimal cost at SO, and can be expressed in the following ways

$$\begin{aligned} \text{PoA} &= \frac{\sum_{i=1}^{m} c_i(x_i^{UE}) x_i^{UE}}{\sum_{i=1}^{m} c_i(x_i^{SO}) x_i^{SO}}, \\ &= \frac{\mathbf{x}^{UE} \cdot \nabla \Phi_{SO}(\mathbf{x}^{UE})}{\mathbf{x}^{SO} \cdot \nabla \Phi_{SO}(\mathbf{x}^{SO})}, \\ &= \frac{\mathbf{x}^{UE} \cdot \mathbf{c}(\mathbf{x}^{UE})}{\mathbf{x}^{SO} \cdot \mathbf{c}(\mathbf{x}^{SO})}. \end{aligned} \tag{3.15}$$

As we discuss below, the PoA is useful to set a scale for the global demand multiplier on a network for a particular demand pattern. Since PoA is only greater than one when the SO and the UE (the *Nash* equilibrium) of the system do not coincide, it can be used to identify the demand interval in which a network strays from optimal performance.

For a traffic model with no hard capacities on the edges — for example, when affine cost functions are used — the PoA, when viewed as a function of demand (for a fixed OD pattern) can be broadly characterised by curves that fit within a uni-modal envelope. For example, in [2], this is shown by using averaged PoA profiles calculated for networks with different types of structures.

Roughgarden [43] has shown that for a wide class of cost functions, the PoA achieved an arbitrary networks is bounded and that the upper bound of the PoA depends on the type of cost functions used. Initially, Roughgarden and Tardos [42] had proven that for affine cost functions PoA $\leq 4/3$. This might initially suggest that the inefficiency of UE is not particularly large, however for less idealised functions the situation gets worse. In [43] proofs for more general functions are provided, for example, for general polynomial functions of degree less than or equal to p, the corresponding bound is

$$\text{PoA} \leq \frac{1}{1 - p(p+q)^{-(p+1)/p}}. \tag{3.16}$$

For link cost functions that have a hard capacity (*i.e.*, the cost diverges for finite flow level s_i), for example $c_i(x_i) = 1/(s_i - x_i)$ the situation gets worse, since the PoA becomes unbounded, with an upper bound existing only in the case where the minimum capacity of any edge exceeds the total demand on the network.

The PoA is a useful probe, not only in measuring the inefficiency of a congestion game played on a network, but also into the interaction between network structure, routes, costs and OD structure. It is most revealing when used as a function of the demand on a network.

For small demand values when the contribution to travel times is mostly due to the free-flow travel time, the PoA is close to unity: the UE and SO costs are similar due to drivers mostly taking the shortest free-flowing routes in both cases.

For very high demand values, the situation is slightly more complicated. In numerical experiments [2, 44] and in this thesis, it is eventually seen to decay to one as demand on the network keeps increasing. The argument offered by Youn *et al.* [2] (for affine cost functions) is that the ratio of objective functions for SO and UE tends to a constant since the quadratic terms in the sums dominate, thus $\Phi_{\mathrm{SO}}/\Phi_{\mathrm{UE}} \to 2$. Since the ratio of the objective functions tends to a constant, both objectives are minimised by the same asymptotic flow pattern.

O'Hare [44] offers more insight as to what happens as demand increases. He shows that the behaviour of the PoA as a function of demand depends on the set routes that are used at the corresponding demand level for the UE and SO assignment. This intuitively makes sense since the differences in cost will stem from the differences in the link flows. Furthermore, O'Hare proves that for cost functions of the form $c_i(x_i) = a_i + b_i x_i^{\beta}$ the UE and SO link flows (expressed as a function of demand d) are related to each other by the relation,

$$x_i^{\mathrm{SO}}(d/\sqrt[\beta]{\beta+1}) = \frac{1}{\sqrt[\beta]{\beta+1}} x_i^{\mathrm{UE}}(d). \tag{3.17}$$

This relation between UE and SO allows relates demands at which changes in the used-link sets between UE and SO assignments, namely, $\eta_{\mathrm{SO}} = \dfrac{1}{\sqrt[\beta]{\beta+1}} \eta_{\mathrm{UE}}$.

Essentially this means that the same links are added or removed to the used-link set at different values, with the changes for SO occurring earlier. When the number of used routes for SO increase the PoA increases, when the number of used routes increases for UE, the PoA decreases. Considering SO first, for large demand levels, eventually all routes are used with no possibility of further changes, eventually the UE assignment 'catches' up and has the same route-set, yet there is a discrepancy between the link flows. O'Hare conjectures (based on analytical calculations for simple networks, and on numerical evidence for larger examples) that once this happens the decay to one follows a power law relation $\mathrm{PoA} \sim d^{-2\beta}$. In short, at high demands, the PoA is low because routing options are bad even for SO routing.

In contrast, the PoA is high when there is a large gap between the costs of the UE and SO assignments, which happens when the contribution to costs due to congestion are comparable to the free-flow cost [2]. In terms of network utilisation, this is where for SO, some users take uncongested routes with large free-flow travel times, which helps to decongest the routes with a large delay.

In figure 3.2, we reproduce the graph from [2] that shows average PoA curves for ensembles of networks as functions of demand. The main finding of Youn *et al.* [2] is that different network structures have similar $\langle \mathrm{PoA} \rangle_{\mathrm{ens}}$ curves: they are roughly unimodal in their form. Note further that for lattice networks, the form of PoA is quite robust to the number of OD pairs that are chosen: a theme we return to in section 3.4.

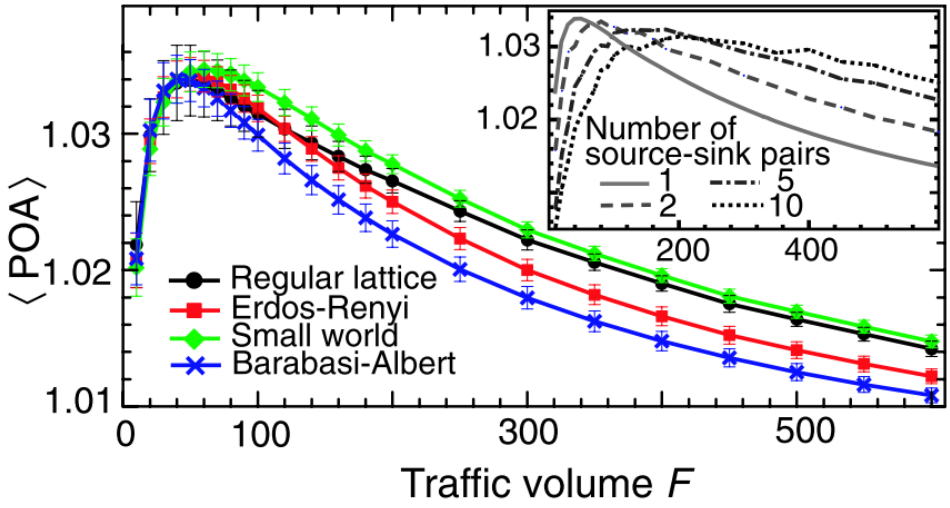


Figure 3.2 Reproduced from [2]: the averaged PoA as a function of demand for ensembles of different network types. The three types of non-planar graph models used by Youn *et al.* are canonical models from complex network theory; Barabási-Albert, Erdös-Rényi, and Small World. Note how the decay in PoA profile for the regular lattice (a planar graph) is different than for the other networks: it is slower and in the large demand limit approximates the behaviour of small-world networks. The inlay shows the ⟨PoA⟩ for a regular square lattice, where the curves are averages over ensembles with different numbers of OD pairs.

3.3 Equipping Skeletons for Traffic Assignment

In order to use the $\alpha\beta$-networks of chapter 2 as transport networks, we need to equip them with cost functions $\mathbf{c}$ (see definition 3.1.1) that capture the time it takes to travel down each link of the network. Depending on the application, different cost functions can be used, for example, polynomial functions like the widely used BPR functions [53] for roads, or functions for mean waiting-times for queues (such as M/M/1) when solving Wardropian equilibrium for assigning jobs to networks of machines [99].

Since we are using affine cost functions (equation 3.4), we need to specify two parameters for each edge: a_i, which is the free-flow travel time, and b_i, which captures the sensitivity to congestion. In order to construct useful ensembles of transportation networks from the $\alpha\beta$-network family of graphs, we need to systematically allocate the parameters for each edge in a way that is not arbitrary. An important consideration is that we need to be able to compare networks that have different numbers of edges. Recall, from chapter 2, that the distributions of the number edges across the same network ensemble are broad. Moreover, ensembles need to be comparable amongst each other and networks of different ensembles differ even more in terms of the number of edges that they have.

The natural approach is to set the a_i parameters, which model the free-flow travel time, as proportional to the length of the corresponding street. In fact, since we are working in

dimensionless units, we simply prescribe the a_i equal to the Euclidean length of each edge. In contrast, the b_i, which are the coefficients that capture the delay due to congestion, require more care.

The goal is that the cost functions ensure that the STAP incorporates features from the network structure in a 'natural' way (from the complexity science point of view), from the bottom up. Below we develop a method, endogenous to the network model, that can be interpreted intuitively as fixing the amount of 'road infrastructure supply'. This method of allocating cost functions also has a satisfying interpretation as equivalent to fixing an 'intersection capacity' as we will discuss further.

Infrastructure supply

As stated above, the free-flow travel time of each edge, parameter a_i, is chosen as the length of the edge:

$$a_i = \ell_i. \tag{3.18}$$

To specify the cost functions fully, we still need to define one more parameter for each edge. Recalling the affine cost functions, equation 3.4, the gradient of the congestibilty term, b_i, seems to be a kind of inverse capacity or sensitivity to congestion of the link. A naïve interpretation is that it is inversely related to the width of the road (or the number of lanes).

Our guiding principle is to strive for parsimony. In this spirit, we seek a simple heuristic that assigns the b_i in an endogenous way for each network, and that is consistent across all networks and ensembles. We also want to avoid modifying the elegant optimisation formulation of the STAP: we want the local network structure in the vicinity of an edge to define the cost functions.

To start, we will take the naïve interpretation to heart. If we associate the length and the width of street i to a_i and $1/b_i$ respectively, then we can set the following sum equal to constant,

$$\sum_{i=1}^{m} \frac{a_i}{b_i} = s, \tag{3.19}$$

where s is a constant across network ensembles, capturing that all networks are supplied with the same amount of total 'road area'. Furthermore, we may take $s = 1$ without loss of generality, by jointly re-scaling the demands, consequent flows, and b_i parameters — further details follow below.

There is still an excess of degrees of freedom; to fully define each b_i we need another $m - 1$ equations. In order to bring in the local topological structure of the network, we jump to a node-centric view of the network, and focus on the intersections. We assume that each intersection has a constant 'intersection capacity', λ, that is split evenly amongst all incoming edges. Thus the inverse of b_i is the same for all edges that are incident to the same node,

$$\frac{1}{b_i} = \frac{\lambda}{k_v}, \tag{3.20}$$

where k_v is the in-degree of node v to which edge e_i is incident to. Or more succinctly,

$$b_i = \frac{k_v}{\lambda}. \tag{3.21}$$

From a transportation perspective this is also a satisfying view; we suppose that incident streets at the same node share the intersection capacity (green time) — we do not consider the re-allocation of green time as part of the STAP — this would be a significant complication, as explained by Smith [105].

By making λ constant across all the intersections, we can write it out explicitly (by re-arranging equation 3.19, and substituting into equation 3.21) in terms of a sum over the nodes, in the form

$$\lambda = \frac{s}{\sum_{v \in \mathcal{V}} \frac{1}{k_v} \sum_{i \in \mathcal{I}_v} a_i}. \tag{3.22}$$

Here $\mathcal{I}_v$ is the set of edges that are incident into node v.

The denominator of equation 3.22 is the average length of incoming edges of each node summed over all nodes, which is the total network length L_G (introduced in section 2.3), so that

$$\lambda = \frac{s}{L_\mathrm{G}}. \tag{3.23}$$

We now make the following two observations. Firstly, it is easy to verify that λ depends on the network morphology; as we saw in chapter 2, L_G depends heavily on the griddedness α. Secondly, we have arrived at this relation by making simple assumptions that are natural due to the model's structure.

In summary, the parameters a_i and b_i for the affine cost functions (equation 3.4) are defined as follows:

- a_i is set as the non-dimensional Euclidean length of the corresponding edge.
- b_i is set by dividing the intersection capacity, λ (from equation 3.22), equally between roads incoming at each node.

The limitation of this method is that all roads in-bound at an intersection have the same congestibility parameter, failing to capture some of the natural heterogeneity in real-world road networks. However, choosing a more detailed way of distributing the intersection capacity implies introducing additional parameters or heuristics, inconsistent with our parsimonious approach. However, note that assigning the b_i based on the node at which the edge terminates means that the corresponding edge in the opposite direction has a different delay term in its cost function.

Scaling and units

So far, the variables and parameters of the model such as traffic flows, costs, the a_i and b_i, and even demand, have been treated as purely mathematical, dimensionless, quantities. We now discuss their units in order to see how the networks scale, and to bring our analysis closer to the network models used by practitioners.

Under the conditions laid out by the STAP, there is a fixed travel demand that has to be satisfied per unit time. The arising equilibrium traffic patterns on a network are therefore stationary flows. The way travel demand is specified, as trips that have to be satisfied per unit time, means that the total number of trips aggregated across all OD pairs have units of vehicles/time, or we might say

$$[d] = \frac{\mathrm{V}}{\mathrm{T}}, \tag{3.24}$$

where V are vehicle units (perhaps PCUs) and T represents units of time. The flow on a given link, x_i, can be thought of as the number of vehicles that pass trough a cross-section of the link. The flows are also given in vehicles per unit time,

$$[x_i] = \frac{\mathrm{V}}{\mathrm{T}}. \tag{3.25}$$

By construction we have assigned the numerical value of a_i as the Euclidean length of the link, yet we have referred to it as the free-flow travel time. Given that our networks are always embedded in the unit square, regardless of the number of nodes they have, choosing units for a_i implicitly determines both the scale of the network as well as the density of the nodes: because in the STAP set-up we are using, a_i is the free-flow travel time of the links

$$[a_i] = \mathrm{T}. \tag{3.26}$$

It is perhaps more intuitive to fix the scale of the network by scaling the unit square, since it allows the network to be shrunk or enlarged to match a desired or reasonable value of a_i: for example, such that a characteristic speed (such as the speed limit) is involved in the form

$$a_i = \frac{\ell_i}{v_{\max}}, \tag{3.27}$$

where the Euclidean length of the road is ℓ_i as above, and $v_{\max}$ is the characteristic highest speed (note that under STAP all vehicles on a link travel at the same speed in order to satisfy the stationary flow conditions).

To understand the scaling of b_i, recall that the cost functions take the affine form of equation 3.4, and as is standard, we take the cost (or dis-utility) to be the travel time on link i. The b_i represent an impedance to flow, so their reciprocal is a type of conductance, with units

$$\left[\frac{1}{b_i}\right] = \frac{\mathrm{T}^2}{\mathrm{V}}. \tag{3.28}$$

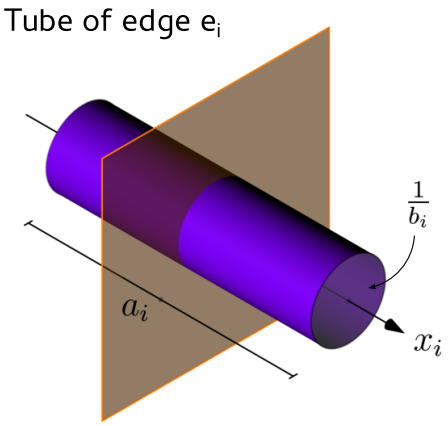


Figure 3.3 Diagram to aid visualisation of the units in the model. A link (road) can be represented in three-dimensional space by a cylinder with length a_i and areal section $1/b_i$. Due to the static nature of the assignment, Edie flows and cross-sectional flows are the same. This 3D representation makes the units of the length of link (a_i) and the area $1/b_i$ congruent.

To visualise the relationships between the units of the free-flow term and the delay (congestion) term, in figure 3.3 a link is depicted as a tube in three-dimensional space. In this space, lengths are time (if we consider all vehicles to be identical and discrete, then V is dimensionless), and $1/b_i$ is the cross-sectional area of the tube.

3.4 Choice of ODs

We now focus on the design of the origin-destination (OD) demand structure and the implications it has for the numerical experiments that we develop in the remainder of this thesis. The general setting is the $\alpha\beta$-networks with periodic boundary conditions as developed in chapter 2 — as a model of a homogeneous 'patch' of a city. Traditional STAP analyses load their networks with a very large number of OD flows. In contrast, we show how the periodic boundary conditions enable an even loading when there are very few OD pairs. The introduction of periodic boundaries ensures that the otherwise peripheral nodes, close to the edges of the unit square, potentially can connect to nodes across the boundary: thus significantly reducing the magnitude of boundary effects, making the 'peripheral' nodes more like the rest of the network's nodes in terms of adjacency properties.

The first OD pair we will consider defines its origin to be the node closest to the bottom left corner of the unit square and its destination to be the node closest to the centre of the square. Figure 3.4 shows UE traffic patterns for this OD pair and for both a periodic and a non-periodic network. For low demand values, vehicles can take the direct routes from the origin to the destination, in as straight a manner as allowed by the links. As congestion increases, and with it, the cost of these 'direct' routes, it is clear how the periodic conditions ensure a more even loading

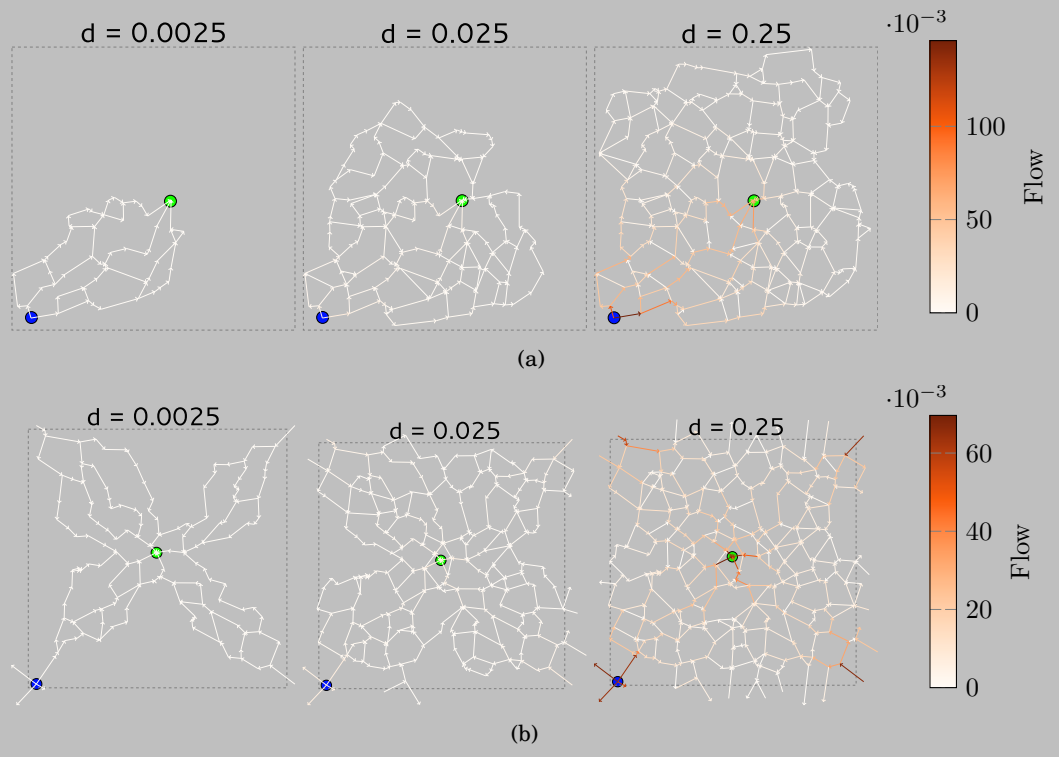


Figure 3.4 The filling of a network (a) without periodic boundary conditions, and (b) with periodic boundary conditions, as demand is increased. There is a single OD pair (origin:blue, destination: green). The demand increases from left to right across the figure. Note flow scales differ between (a) and (b): the busiest links are significantly less congested under periodic boundary conditions, due to a more even spreading of demand. Here $N = 225$ nodes, $\hat{\alpha} = 3/4$, and $\beta = 1.4$. Links with flow computed $< 10^{-6}$ are not shown.

of the network. Thus, with just a single OD pair and periodic conditions, the network already behaves a bit like a network 'patch' embedded in a larger network. At a more fundamental level, the set of used paths to which each edge belongs are more uniform for the periodic networks.

In order to test how the periodic conditions allow us to approximate a true multi-OD structure, in section 3.5.2 we will use two OD pairs set up so that they both make ample use of the whole network, and so that their flows 'cross' each other. The first OD pair will be the same as described above. The second pair is generated by shifting the first laterally by 1/2 on the unit square. That is, the origin is the node closest to the midpoint of the lower boundary of the unit square $(1/2, 0)$; for the destination, we choose the node closest to the midpoint of the left boundary $(0, 1/2)$. Figure 3.5 shows the two OD pairs, and how the image origins and destinations across the periodic boundaries yield flows in multiple directions on the network. Henceforth we refer to the OD pairs as ω_1 and ω_2 respectively.

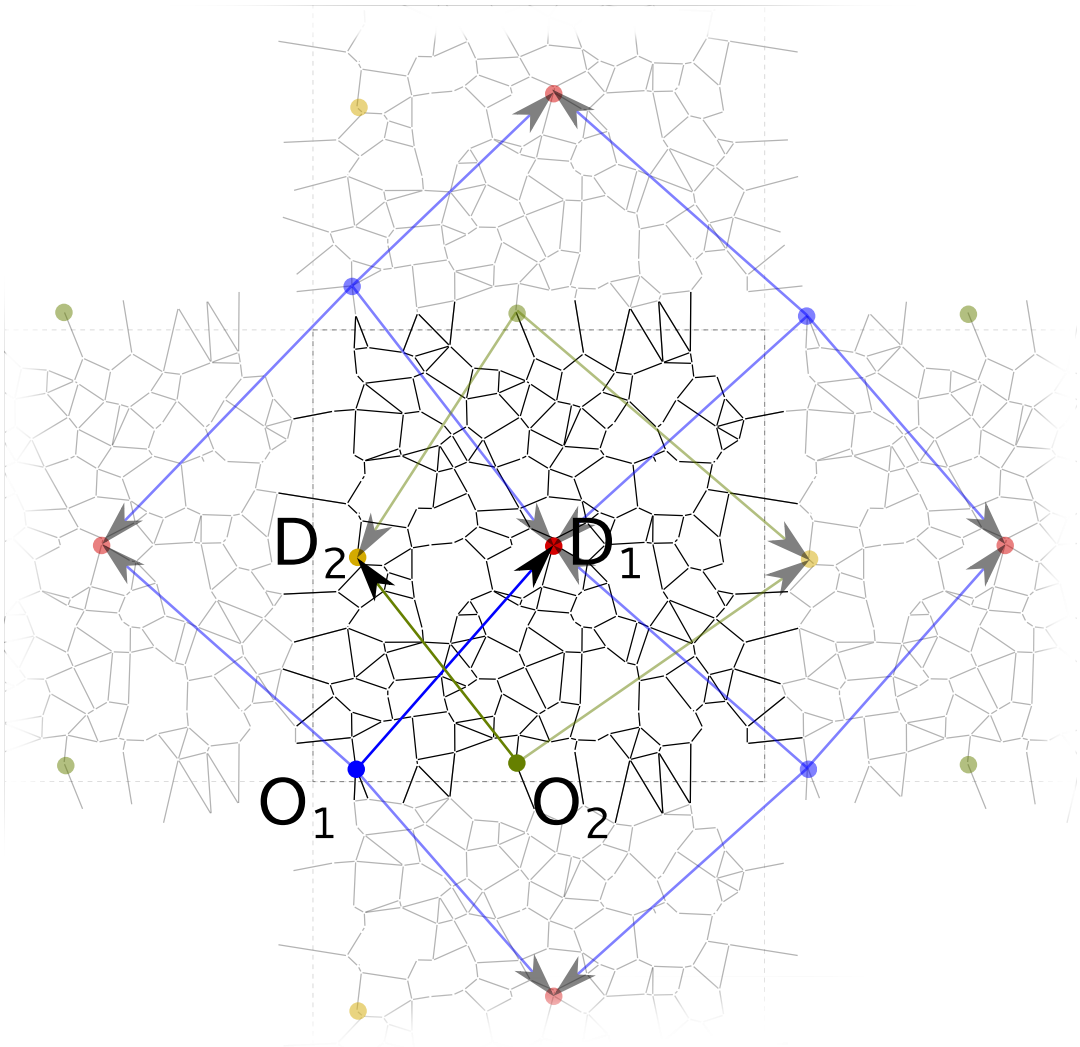


Figure 3.5 Two OD pairs, ω_1 (with origin O_1 and destination D_1) and ω_2, with origin and destination labelled O_2 and D_2 respectively. Here $\hat{\alpha} = 1.25$, $\beta = 1.5$, and $N = 225$. Introducing periodic boundary conditions makes each OD pair load the network more like four ODs with a common destination node. For simplicity, only arrows within the unit square are drawn for ω_2.

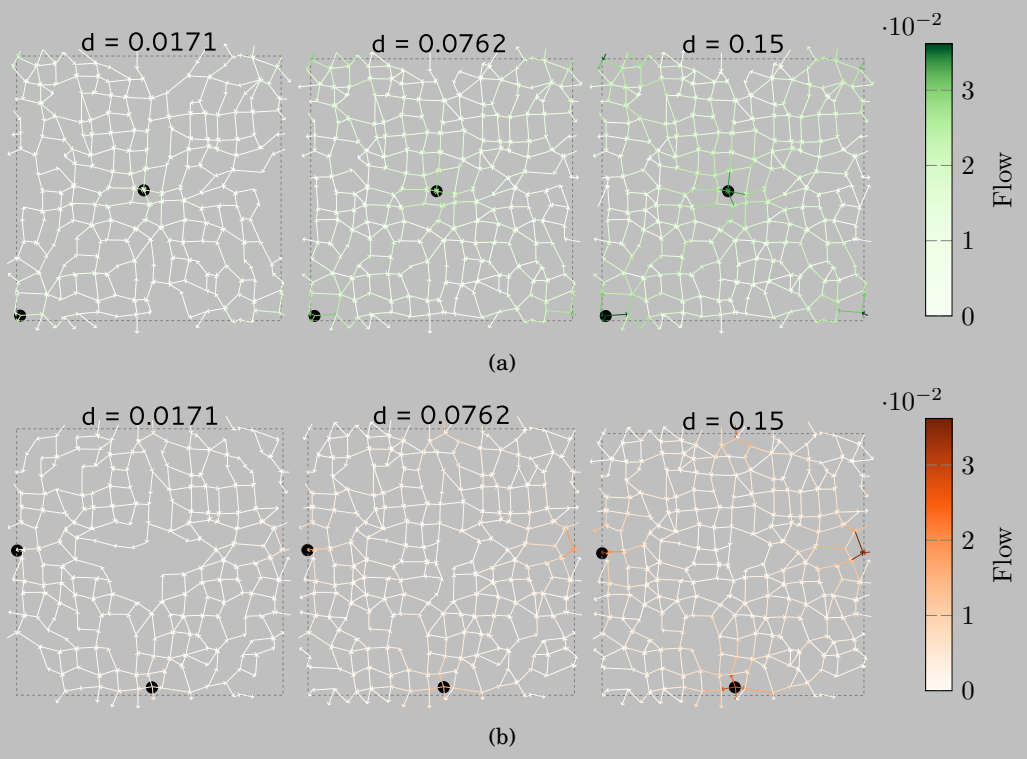


Figure 3.6 UE traffic patterns for a network ($\hat{\alpha} = 1.25$, $\beta = 1.4$, and $N = 225$) associated to the flow of (a) OD pair ω_1, and (b) ω_2. Note that in terms of coverage of the network, the flows from both OD pairs (marked in black) complement each other, giving rise to broadly isotropic loading.

Figure 3.6 shows the filling of a network with the UE flow patterns for the two ODs ω_1 and ω_2. Note that the traffic patterns reflect the way the OD pairs are related by a lateral translation of 1/2; the gap in the used links of the pattern occurs at the edges of the unit square for OD pair ω_1, and at the centre of the square for ω_2.

Individually, each OD pair ω_1 and ω_2 makes ample use of the network. The periodic conditions simulate demand that originates beyond the boundaries of the unit square that have a common destination inside it, making the arising flows more 'isotropic'. Both ω_1 and ω_2 have their origin and destination separated by approximately the diameter of the network, which means that the users have to cross as much of the network as possible. Note that the diagonal crossing of the flows also ensures that many of the same links are used by the flows of each OD pair. All this guarantees that the network is more akin to a sub-network embedded in a larger network, yet avoids prescribing incoming flows from external nodes or centroids.

In section 3.5.2 we will compare the average PoA curves of $\alpha\beta$-network ensembles loaded under a single OD pair with those arising from solving the STAP for a two-OD regime where the global demand is split equally between the OD pairs. The way that the networks are loaded

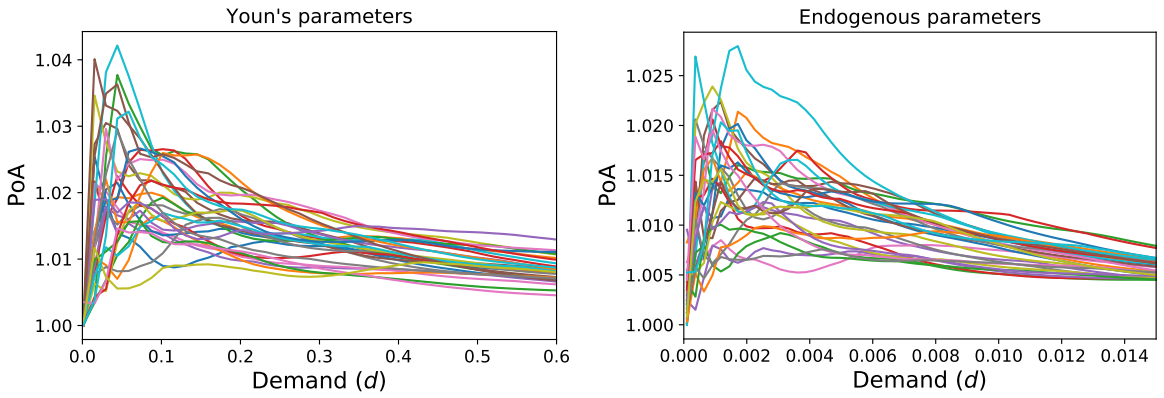
guarantees that flows from both OD pairs share as many links as possible, compared to, for example, would happen by taking ω_1 together with the OD pair where ω_1's origin and destination are interchanged.

By simulating a larger number of OD pairs by introducing the periodic conditions on the networks, the achieved effect is a more even loading of the networks. The effects an even loading of the network has in terms of the PoA, is that as more links of the networks are used, there are less routes that remain unused through more of the demand range. As O'Hare [44] explains, the changes in PoA with demand, the peaks and troughs occur when the used routes under SO and UE assignments change. The rule of thumb being that when the set of used routes for SO increases PoA increases, and when the set of used routes for UE increases, PoA decreases. This effect can be understood by considering that when the SO used-route set increases and the UE does not (UE route changes tend to lag behind SO ones) the difference in flows (which is optimal under SO) is greater, creating a larger difference between the costs of both assignments. In terms of the evenness of loading the networks, having a more even loading that permeates the network, ensures that the active-route set is larger at lower demands. This reduces the demand range needed to be covered in order to capture the peak and decay of PoA. As far as the ensemble view is concerned, a more even loading reduces the intra-ensemble variability of the PoA of the different networks. This happens since the whole network structure is sampled, thus reducing the instances in which only specific local structures are sampled in the initial region of the demand range which leads to more particular PoA profiles.

3.5 Numerical Results

In this section we present results from numerical experiments of static assignment on $\alpha\beta$-networks. We begin by comparing our endogenous method of defining parameters with the random selection of Youn *et al.* in [2] in section 3.5.1. Then we proceed with comparing the behaviour of the averaged PoA curves for two different OD demand structures: a single OD, and two ODs (ω_1 and ω_2 from section 3.4), set up so that the flows connecting each OD cross each other.

The way in which we select the demand range for solving the STAP on ensembles of $\alpha\beta$-networks is designed to capture high PoA values and enough of the decay, so that the different ensembles become distinguishable. For example, figure 3.7, below, shows how different parameter choices scale the demand range, and with it, the peak region for the PoA curves for the same networks.



(a) (b)

Figure 3.7 PoA traces for 30 networks sampled from the ensemble with $N = 225$, $\hat{\alpha} = 1.5$ ($\alpha = 0.09375$) and $\beta = 1.6$. (a) Cost-function parameters are assigned as in [2]. (b) Parameters are assigned according to the method from section 3.3. In each case the same networks are used with the same OD pair (ω_1) for each network.

3.5.1 Comparison of Youn *et al.*'s parameter allocation method with the endogenous method

In order to compare the implicit (endogenous) method of allocating cost-function parameters (as described in section 3.3) with the random choice of Youn *et al.*, we carried out the following experiment.

We sampled 30 graphs from the $\alpha\beta$-network ensemble with parameters $N = 225$, $\hat{\alpha} = 1.5$, and $\beta = 1.6$. Two sets of road networks were created from the same graphs. The first had cost function parameters assigned according to Youn *et al.*'s [2] method: edge lengths were assigned uniform random integers between 1 and 3, while the b_i were assigned uniform random integers between 1 and 100; corresponding edges with opposite orientations are given the same parameters. The second set of road networks were assigned cost functions in the manner laid out in section 3.3.

The STAP was solved for both sets of networks for 60 demand values. To calculate the PoA both UE and SO assignments were calculated. The networks with Youn *et al.*'s parameter choice required a demand range starting at $d = 0.001$ and ending at $d = 0.85$, with a demand step of $\Delta d \simeq 0.0144$. For the set of networks with cost-function parameters chosen according to our methodology, the demand range starts at $d = 0.0001$, finishes at $d = 0.016$, and has a demand step of $\Delta d \simeq 0.000279$. Figure 3.7 displays the PoA curves for all of the networks.

For the random choice of parameters (Youn *et al.*'s method), the peak PoA values appear to be higher (note the difference in scales), sharper, and the PoA curves seem to decay with a more pronounced concavity. That is, after a sharp rise, due to the initial large discrepancy between UE and SO assignments, further increase in demand causes the UE assignment to approximate SO at a faster rate. These qualitative observations are confirmed in figure 3.8, which shows the

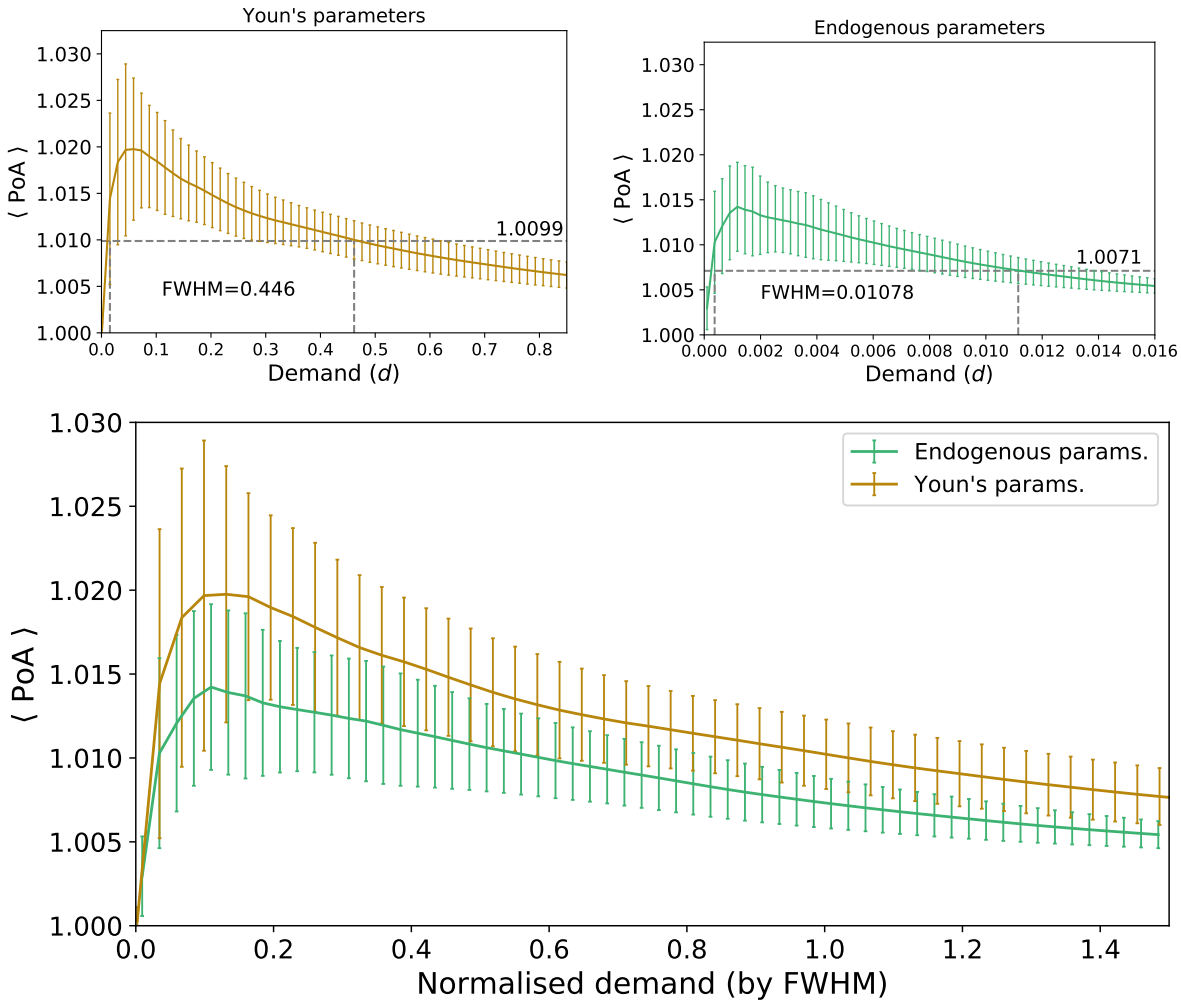


Figure 3.8 Average PoA for the same networks as figure 3.7. In total, 60 values of demand were calculated for an approximately equivalent range for both parameter choices. The full width at half-maximum (FWHM) is shown: note that the baseline is taken as one rather than zero. The bottom figure shows the ⟨PoA⟩ for both cost-parameter choices, with the demand range normalised by the FWHM.

average PoA curves from the individual traces from figure 3.7. The bottom plot of figure 3.8 shows ⟨PoA⟩ *vs* demand normalised by the FWHM.

As the PoA decays, the (sample) standard deviation behaves differently in each case, although this is hard to see due to the large error bar size. So, to better appreciate the dispersion around the mean, figure 3.9 displays the coefficient of variation as a function of demand.

Comparing the coefficients of variation highlights the importance of taking a transport-centric approach to network design. The uniformly random assignment of integers yields a broader distribution of cost-function parameters. In turn, this causes a persistently larger standard deviation of the PoA throughout the demand range. Even though the decrease in the coefficient

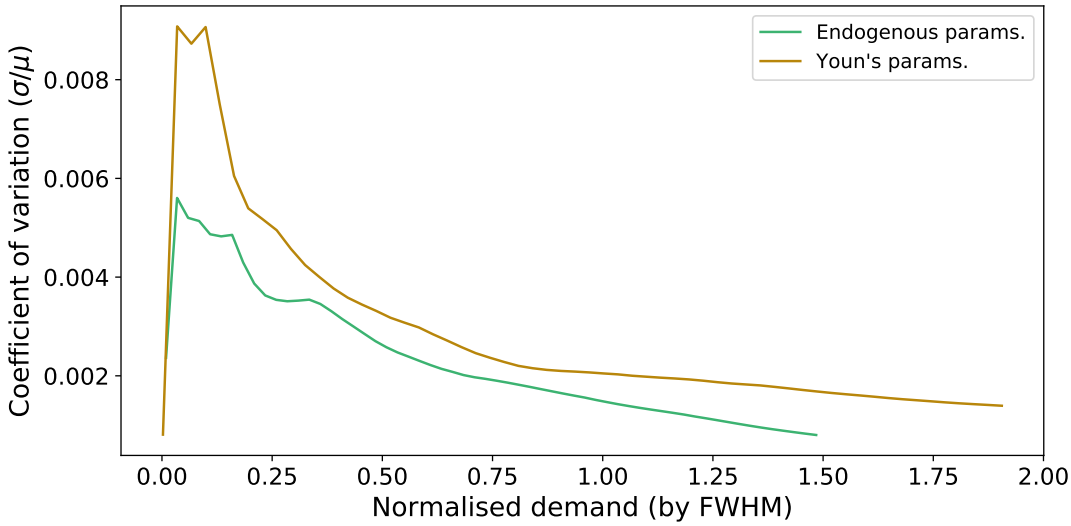


Figure 3.9 The coefficient of variation $(\sigma(\text{PoA}_i)/\langle\text{PoA}\rangle_i)$ for both cost-function parameter choices. The coefficient of variation is consistently lower for the networks with endogenously defined parameters.

of variation is faster after the PoA peak, it eventually settles into a more gradual decrease as demand increases further. In contrast, for the endogenous choice of parameters, the fall in variation accelerates in the decaying tails of the ⟨PoA⟩ curve. This is a crucial observation for the usefulness of PoA as an experimental tool, since it is in the tails where clear distinctions can be made between the different ensembles.

Figure 3.10 compares both mean PoA curves on the same scale by normalising the ⟨PoA⟩ of both ensembles by each of their maximum values. Once again, as in figure 3.8, the demand is scaled by the FWHM. Our observations made about the PoA traces of individual networks are now more concrete. The shape of the peak is more pronounced for Youn's parameters; the magnitude of the error bars also reflects this. The curve for the endogenous parameters rises faster but then tapers off slower, the error bars tightening much quickly around the mean value. Compare this to the behaviour of Youn *et al.*'s regular lattice (the black curve) from figure 3.2.

We conclude that allocating cost-function parameters with our implicit method is a successful way in reducing the intra-ensemble variability than what is obtained when assigning them in a uniformly random way. From the results of Youn *et al.*, the clearest behavioural difference between the ⟨PoA⟩ curves for the different ensembles is between the lattice and the non-spatial network models. The curve peaks sooner and decays slower than the other ensembles, cutting across the curves of the other ensembles.

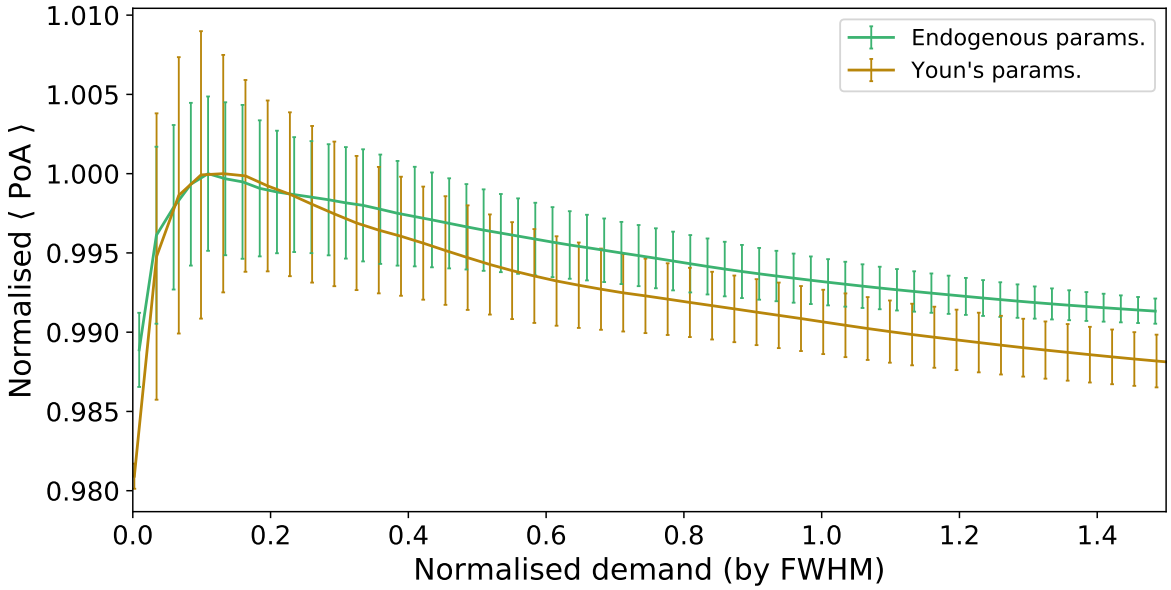


Figure 3.10 The normalised ⟨PoA⟩, with respect to peak PoA value, for the same data as in figures 3.7 and 3.8. The error bars (one standard deviation on either side) have also been scaled accordingly.

3.5.2 Comparison of assignments on ensembles: one OD *vs* two ODs

We now compare the averaged PoA curves between network ensembles that are subjected to different OD demand structures: two maximally separated OD pairs (ω_1 and ω_2) where the total demand is equally split between them (see section 3.4 above), and the averaged PoA curves for a single OD pair: ω_1. The morphology parameters for the ensembles are the combinations of $\beta = 1.2$ and 1.4, and $\hat{\alpha} = 0.75$, 1, and 1.25. The STAP is solved for the demand range $0.00001 \leq d \leq 0.01$, in steps of size 0.0003.

Figures 3.11 and 3.12 show the averaged PoA curves for the six ensembles, for the two-OD case, compared with one OD. The single OD used was ω_1. The first feature of note (figure 3.11) is that there is little variation between the PoA curves, although some fine details can be picked out.

For $\hat{\alpha} = 0.75$ (relatively grid-like networks), both OD set-ups yield rather similar PoA curves. In contrast to Youn *et al.*'s results for the lattice, we observe that grid-like $\alpha\beta$-networks with periodic boundary conditions are robust to changes in OD structure. Apparently, the periodic $\alpha\beta$-network ensures that small numbers of OD pairs make ample use of the network, more characteristic of multiple OD pairs. This result is further supported by the comparatively low standard deviation.

Still focusing on the differences between the PoA curves, we note that the largest differences are seen for critical griddedness $\hat{\alpha} = 1$. For the more 'realistic' values of β (*e.g.*, of 1.4, as found by Osaragi and Hiraga [8]), the demands at which the PoA curves peak, as well as the maximum

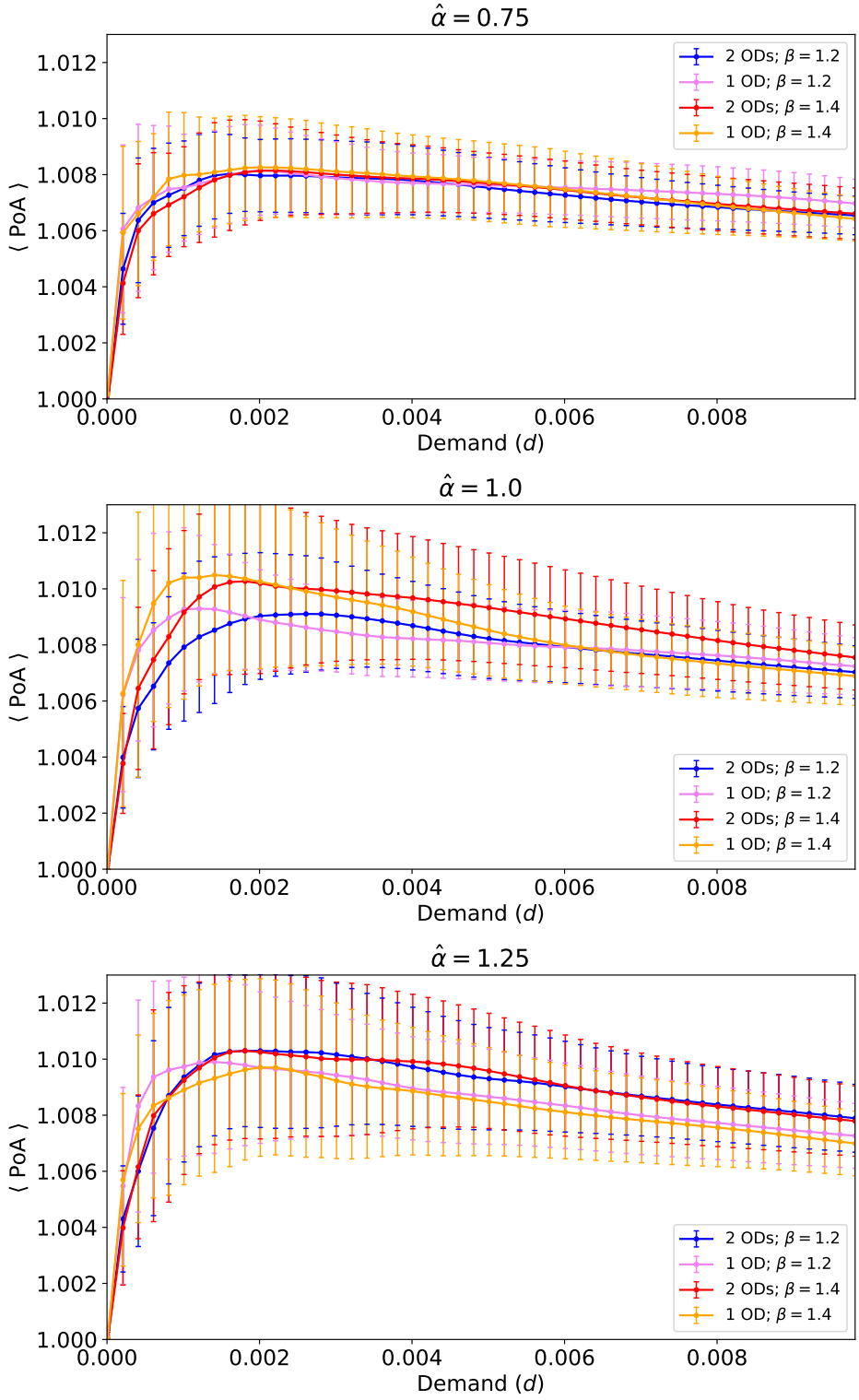


Figure 3.11 Average PoA profiles are shown for six ensembles (45 networks each) with $\hat{\alpha} = 0.75, 1, 1.25$; $\beta = 1.2$ and 1.4; and $N = 225$. The STAP was solved for these ensembles under two different OD regimes: a single OD pair (ω_1), and for two OD pairs (ω_1 and ω_2), assigned half the total travel demand each. Error bars are one standard deviation to either side of the mean.

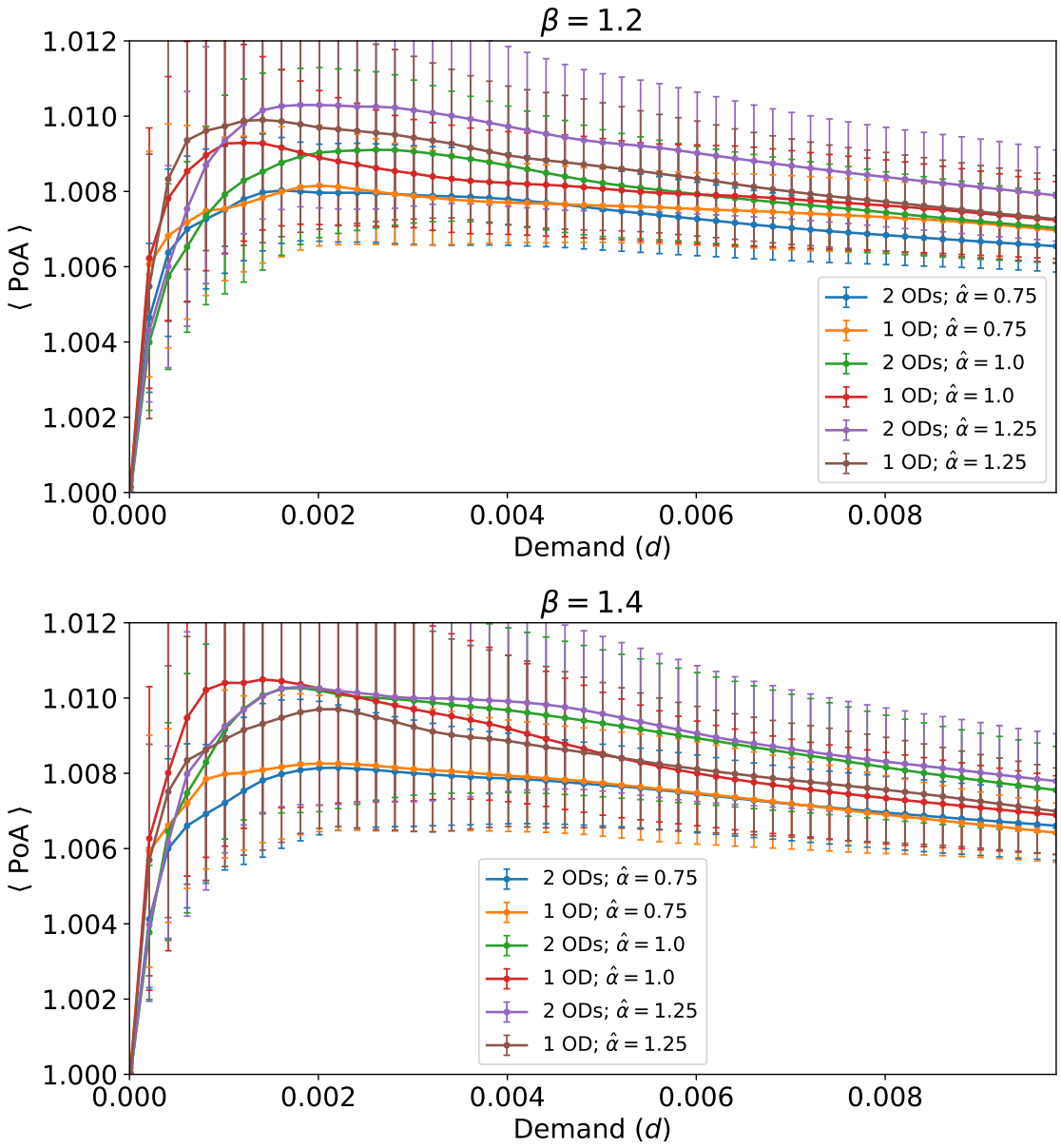


Figure 3.12 Average PoA profiles are shown for six ensembles (45 networks each, with $\hat{\alpha} = 0.75, 1, 1.25$; $\beta = 1.2$ and 1.4; and $N = 225$) under single and two OD pair regimes (same $\langle \text{PoA} \rangle$ curves as in figure 3.11). Curves are grouped according to β. Error bars are one standard deviation to either side of the mean.

PoA reached are closer than for $\beta = 1.2$.

As the griddedness increases further to $\hat{\alpha} = 1.25$, the PoA curves once again come closer together. However, the dependence on OD structure begins to show in the way the curves for the same number of OD pairs follow each other very closely.

When examining the PoA curves grouped according to β, the robustness to OD structure for

low $\hat{\alpha}$ is starker. The single-OD and double-OD case are the most similar amongst the curves. Interestingly, for $\beta = 1.5$, the PoA ensembles with $\hat{\alpha} > 0.75$ show that the OD structure is more important than the griddedness of the networks. This suggests that for more realistic networks, the interactions between vehicles belonging to different OD pairs is more important than the underlying network structure, once the grid structure breaks down.

Overall, the more regular networks — with lower $\hat{\alpha}$ and β — show the least sensitivity to OD structure, at least when the loading is even, and the interactions between OD pairs is important.

3.5.3 Discussion of results

When using the PoA as a measure of network performance, we have seen that the main limitation is the large variance observed within the ensembles. The large variation is a testament to the sensitivity of the PoA to network structure. In fact, for larger $\hat{\alpha}$ (lower griddedness) and higher β (lower road density), the spread around the mean is larger, and therefore the $\langle$PoA$\rangle$ less smooth.

From the numerical experiments of section 3.5.2, we can conclude that implementing periodic boundary conditions is a robust way of simulating multiple OD flows while solving the STAP for a single OD pair, when the griddedness of the networks is high. Another finding is that past the critical griddedness ($\hat{\alpha} \geq 1$), as the roads become less dense ($\beta = 1$), the OD structure has a greater effect on the $\langle$PoA$\rangle$ than the griddedness. This can be seen by the overlap, especially within the decaying tails of the average PoA curves for the two-OD scenario, irrespective of whether $\hat{\alpha}$ is 1.0 or 1.25.

3.6 Discussion

In this chapter we have discussed how the $\alpha\beta$-network family is assigned cost functions to solve the STAP on ensembles of these networks. Special care was taken to avoid introducing artefacts by way of our boundary conditions, and in ensuring that our sampled networks remain comparable for different parameter values. With the attention to detail regarding the boundary conditions introduced in chapter 2, different types of OD pairs were studied, allowing us to explore the sensitivity of averaged PoA curves to morphological variation and to changes in OD demand structure.

From our numerical experiments in section 3.5, we have found that the griddedness parameter α can modulate how sensitive the traffic performance is to changes in the density of edges (determined by β) and to the OD structure.

To bring the discussion back to the relationship between network measures and traffic performance on networks, we reproduce a figure from [66] (where the $\alpha\beta$-networks were originally presented). The figure compares the integrated average PoA curves (IPoA, for *integrated price of anarchy*), the algebraic connectivity, the mean street length and the mean degree of twenty $\alpha\beta$-network ensembles. The ensembles take all combinations of $\beta = 1, 1.2, 1.4, 1.6, 1.8$, and 2 with

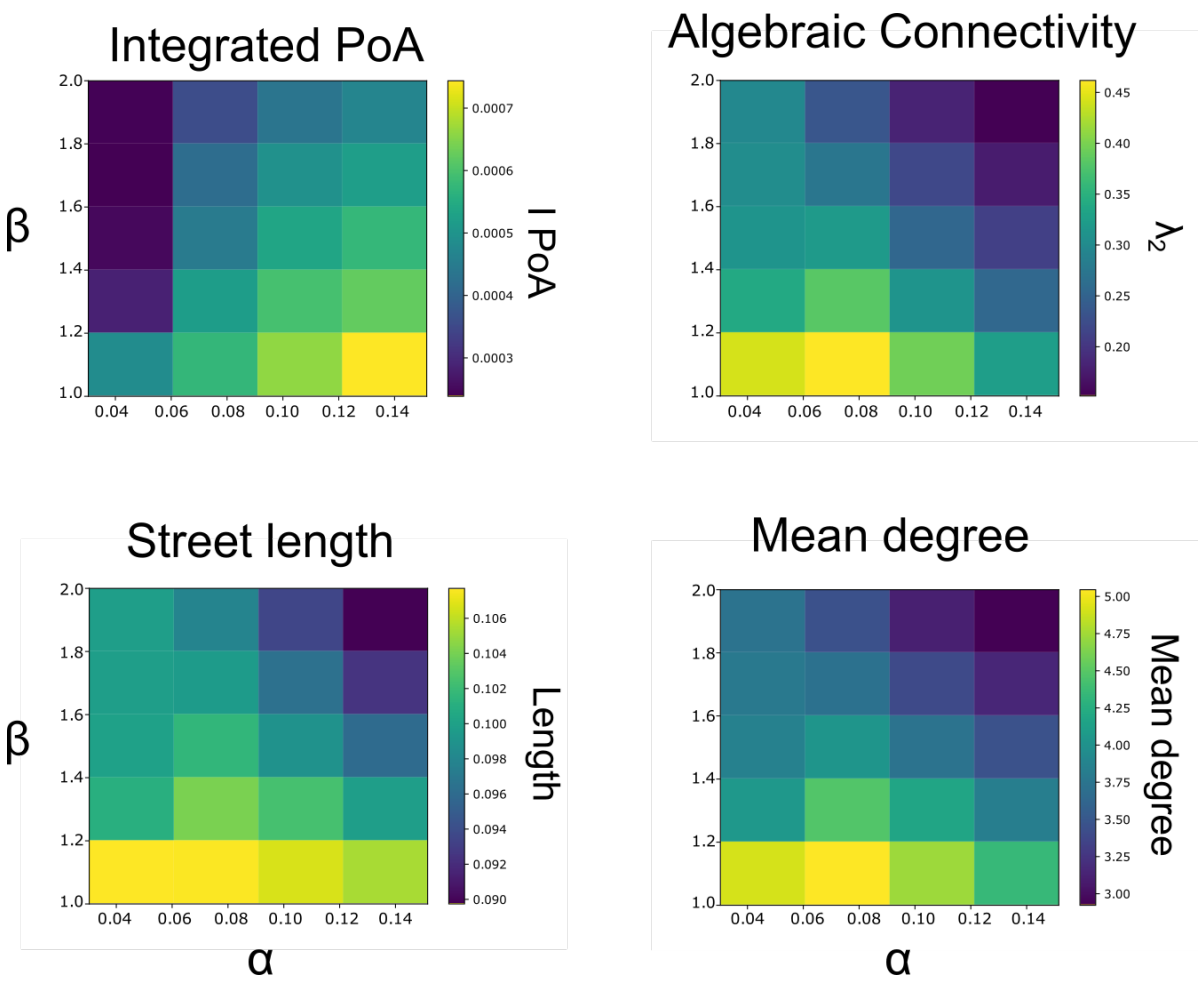


Figure 3.13 The averaged *integrated price of anarchy*, ⟨IPoA⟩, is shown as a way of capturing the total inefficiency of the ensembles across the demand range studied. It is compared to the mean degree and the algebraic connectivity (two of the standard network-theoretic measures discussed in chapter 2). We also compare it to the mean *street length*, which as discussed is slightly better correlated with the IPoA.

$\hat{\alpha} = 0.88, 1.1, 0.32$, and 1.54. In this case the networks are based on a 10×10 grid ($N = 100$) to reduce computation time, due to the larger number of demand steps and the larger number of networks sampled from the ensembles. The numerical integration of the IPoA was done as a Riemann sum with the value of the PoA as the height of the columns, over the demand range with 100 demand values, starting at $d = 0.001$, ending at $d = 0.3$, and with a demand step-size of $d = 0.003$.

Due to the fact that each mean PoA point is calculated for an average of 100 networks, over 100 demand values, the ⟨IPoA⟩ is quite a coarse measure, as well as computationally expensive relative to the statistical certainty it provides. However, as we have seen (like with figures 3.11 and 3.12), even with the large variation, this type of aggregated measure is informative of the bulk behaviour.

It is clear from inspection of figure 3.13 that there is little correlation between network theory statistics and the performance of road networks. The better correlation is with the physical property of mean road length (and therefore L_{G}), however it is also not great. We propose that if theoretical explanations for the behaviour of the PoA that stem from network properties are to be found, they need to be found in more than just network theory statistics.

On a more general level, the maximum value that the PoA reaches is far from the theoretical upper bound of 4/3 for affine cost functions. More realistic cost functions such as BPR functions or other polynomials can yield larger PoA values, yet they are also simplified functions. A possible extension that could be more revealing, would be to include capacity constraints on the links in the networks. This can be done in two types of ways, the first is by explicitly including hard capacities as constraints into the optimisation problem. The second, is to include it via the cost functions themselves. For example, as mentioned in section 3.2, cost functions of the form $c_i(x_i) = 1/(s_i - x_i)$, that arise as average delay functions from queuing systems, have an asymptote at $x_i = s_i$, naturally representing a flow capacity. Depending on the demand levels in relation to the link capacities s_i, if $d_{\max > \min i s_i}$ the PoA can be unbounded [43]. These cost functions, could represent delays due to queues at intersections due to traffic signal control. In this case, the PoA might more realistically capture the inefficiency of a network's state due to selfish routing. Studies such as the ones presented in this chapter, but include capacities, or diverging cost functions, might yield much richer and detailed results that the ones presented here that more accurately reflect the effects of network and demand structure on the efficiency of traffic equilibria.

From the results in this chapter and from existing studies in the literature (for example [2,44]) the PoA has strengths as well as weaknesses in comparing network ensembles. Among it's strengths, is that due to the unimodal behaviour, ensembles of vastly different morphologies can be compared. The way the PoA decays one it has reached its peak value, gives some information as to the morphological structure of the networks. For example, in figure 3.2 from [2] the PoA curve for the regular lattice (black curve) has its peak close to the Erdös-Rényi and Barabási-Albert ensembles, yet the tail resembles the small-world networks more. In our numerical experiments, the early peaking of PoA and its slower decay when using morphology-aware parameters compared more random networks is also observed (see figure 3.8). Unfortunately, the PoA only captures the inefficiency ratio of networks in the aggregate. The distribution of flows and costs amongst different links in the networks is ignored by the PoA. A possible measure to capture how effective different network structures are evenly distributing flows and costs could be the Gini coefficient of the flow and cost distributions. This metric would capture how efficient the networks are at spreading the traffic flows and more effectively utilising the infrastructure supply. Alternatively, to reflect the large intra-ensemble variability clustering of the PoA profiles (using functional distance measures, for example) could yield further insights as to different types of network features that play a role in effecting specific types of PoA profiles, thus enabling a more detailed study into local network structure and routing inefficiency.

In a very recent paper, Alonso *et al.* [106] numerically examine the eigenvalue spectrum of β-skeletons to show that these graphs have interesting spectral properties, and they suggest different ways of analysing β-skeletons. This suggests possible avenues forward for future studies, like the one in this chapter, can benefit by taking a spectral approach. We propose, for example, to decompose patterns and link costs in terms of network eigenvectors. This would help re-examine whether abstracting road networks directly to graphs, although the most direct approach, is the best approach for future transport research.

One of the main goals of this thesis is to bring transportation and network modelling closer together. By building on O'Hare's methodology [44], a direct application has been provided for extending the use of the PoA from a quantity of theoretical interest, to an experimental tool for understanding the relationship between network structure and traffic equilibria performance. The key contributions of this chapter are the following:

C3.1 The endogenous method for allocating cost functions for investigating the effects of infrastructure supply on networks makes the $\alpha\beta$-network family a general-purpose synthetic network model that avoids arbitrary parameter choices.

C3.2 The average PoA curves for single and multiple ODs show that implementing periodic boundary conditions is a robust way of achieving traffic patterns that are similar to those arising from more complex demand structures.

C3.3 More random networks exhibit stronger sensitivity to OD demand structure, across large demand ranges, than the more robust grid-like networks.

C3.4 Traditional network theory statistics fail to correlate with transportation efficiency as measured through the PoA. Actual links between network theory and transportation efficiency are more subtle and deeper than what network-theoretic narratives would suggest.

CHAPTER 4

MIXED WARDROPIAN EQUILIBRIA

In this chapter, we use the STAP on the $\alpha\beta$-networks family as an experimental method to study transitions in traffic equilibria. The discussion is embedded within the context of multi-class equilibrium, where the user classes differ with respect to their routing behaviour or *strategy*. We split the network users into two classes: selfish or altruistic. The mixed traffic assignment is then studied for different proportions of the *penetration rate* of the altruistic class. We also discuss the behaviour of the mixed equilibria in general terms by building up from examples of simple networks.

The interest in mixed equilibria as a network control mechanism stems from the apparent advent of the autonomous vehicle (AV). One can picture a setting in which a benevolent dictator assigns routes to the altruistic user class in an attempt to drive the system towards optimal — thus using the flow of the altruistic user class as an indirect control mechanism on the costs experienced by the selfish users. We approach this problem from the context of *route guidance systems*, mainly developed as a response to the then-increasing use of satnavs [61], where multi-class equilibria can be modelled as variants of *Stackelberg* games. One of the aims of this chapter is to redirect the discussion onto AVs, and the possible gains in efficiency, from a focus on the removal of the human driver as a potential source of increased efficiency (by forming platoons and negotiating intersections, for example), to the behaviour of the system as a whole by focusing on routing.

We leverage well-understood and simple models as experimental tools. This allows the focus to be on the effects of the shifts in traffic patterns themselves and on the role of network morphology. Some of the material in this chapter was presented at the MFTS symposium held in Ispra, Italy, in 2018. The chapter is structured as follows.

Section 4.1 describes the context of the mixed equilibrium (ME) models that we base our numerical investigations on. These are models that were developed in the 1990s [62, 107] for the

purpose of considering the effects that the uptake of route guidance could have on the equilibrium efficiency of traffic networks. This is still a relevant discussion that now revolves around autonomous vehicles, since automation could provide a way of enforcing altruistic behaviour in a proportion of the vehicles on the roads.

In section 4.2 we provide the details of the implementation of the algorithm that we use in this chapter for the numerical solution of the ME STAP. It is a variant of the models described in section 4.1, which means it inherits the stability properties of the equilibrium.

The transition to optimality of small networks, and how the equilibrium traffic patterns change by introducing an altruistic class of vehicles, are discussed in section 4.3. The aim is to explain the mechanism by which the system costs are reduced when a proportion of the legacy self-optimising vehicles are replaced by altruistically-routed AVs that seek to optimise the total-system cost.

The change in costs due to the increasing penetration rate, γ, of the AVs in the user fleet is then explored in section 4.4. Finally section 4.5 discusses our contributions in the round.

4.1 Background

We begin this section by re-visiting the *mixed equilibrium* (ME) model of van Vuren and Watling [61]. The literature on route guidance (for example see [62]) gives insight to the changes in traffic patterns that *autonomous vehicles* (AVs) and widespread *mobility as a service* (MaaS) might cause.

Route guidance and routing games

In [61], van Vuren and Watling prove the existence of a static equilibrium assignment for a two-class model, where one class behaves according to Wardrop's UE criterion, and the other class seeks to minimise the total system cost. The resulting traffic pattern is a mixed Wardropian equilibrium where the human-driven vehicles, which we call selfish vehicles (SVs), *self-optimise*, whilst the AVs (also a suitable abbreviation for *altruistic vehicles*) collectively try to optimise for the whole system. For cost functions of the form

$$c_i(x_i) = a_i + b_i x_i^p, \tag{4.1}$$

where $a_i \geq 0$, $b_i > 0$ and $p > 1$, van Vuren *et al.* [107] show that an equilibrium exists that is unique in terms of aggregate link-flows. Our affine cost functions are a simple case of this functional form, with $p = 1$. Even for a simple network, however, [107] finds that a large proportion of guided vehicles might be needed before the ME and SO coincide. For BPR-type functions

$$\begin{aligned} c_1(x_1) &= 10.8 + 197x_1^4 \\ c_2(x_2) &= 3149x_2^4 \end{aligned} \tag{4.2}$$

on a Pigou network and demand $d = 2800$, SO costs are not achieved until the system optimising vehicles are at $\sim 64\%$ of the total demand.

In the mixed equilibrium (ME), the altruistic class of vehicles manages to reduce the total network costs by incurring extra costs themselves, which highlights that even though UE is inefficient (in terms of the cost incurred by each user travelling between the same origin-destination pair) it is still a *fair* assignment in the sense that for a single OD pair, the paths taken by all users have equal costs. The converse to UE being fair is that not all the altruistic vehicles' costs are equal; some take longer routes to minimise congestion along links used by the selfish users. Therefore the different classes can share at most one route (when edges have distinct cost functions). This concept of fairness only works for a single OD case, since if more OD pairs were to be considered, the way vehicles one OD pair utilise the networks can have an important impact on the flows associated to the other ODs. For example, a long-range OD pair with high demand, where vehicles traverse large sections of the network, can significantly affect the costs of other OD pairs that have smaller flows associated with them by congesting the links of their best routes. In a case like this, fairness would have to be evaluated across different OD pairs and with respect to the magnitudes of demand associated with each one.

The idea of ME can be further generalised to more complex types depending on the vehicle classes and their objective function, that is, the *potential function* of the congestion game. However, from a more applied point of view, changing the type of game might also be of interest. For example, in an economically inspired set-up [108], Yang *et al.* study an extension of van Vuren and Watling's mixed equilibria for more complex routing classes that correspond to strategies borrowed from economic competition scenarios (namely, Cournot-Nash competition). They prove the existence of the mixed equilibrium for a mixture of three classes: a system optimising class following SO routing, one self-optimising class following UE routing, and a finite number of separate players belonging to a Cournot-Nash class where each player tries to optimise their own class's average cost. In the application of their model to exemplar networks, they use (like us) affine cost functions. A noteworthy result is that for a mix of only Cournot-Nash and UE players, SO is not achieved unless the Cournot-Nash class controls all of the flow. In the autonomous vehicle context, this type of ME could arise with a mixture of selfish users and providers of AVs that optimise for aggregated costs of their own class.

Yang *et al.* then compare their multi-class equilibrium with a different model obtained by treating the classes as players that take turns in a leader-follower Stackelberg congestion game (as in [109]). The models and algorithm seem to differ in small values of flow volumes, but otherwise their results are consistent with Van Vuren and Watling; system costs can indeed be driven to optimal, but it can require a large proportion of the flow to be controlled by the SO class.

In a more recent paper by Sharon *et al.* [110], the ME is solved on six exemplar networks taken from [111]. The observed behaviour confirms observations in the examples of [61, 107, 108] on the proportion of vehicles needed to reach SO: between 13% to 53% of altruistic vehicles are

needed to achieve SO costs. The trend in Sharon *et al.*'s results shows that, generally, networks that are larger and with more complex structure require higher proportions of SO vehicles in order to reach optimal performance. Therefore, extending the basic ME model of van Vuren *et al.*, or even changing the types of interactions (for example, setting up the ME as a Stackelberg game or multi-objective optimisation problem), yields qualitatively the same results. Furthermore, as we shall see in this chapter, our use of affine cost functions does not qualitatively change the behaviour of the ME systems (a contribution to the field in its own right).

As discussed in chapter 3, the concept of PoA has been studied from both the transport perspective and as a scheduling optimisation problem for assigning jobs to machines. In [99], Roughgarden studies multi-class strategies in a setting equivalent to a network with a single OD pair and m parallel links. In Roughgarden's proposed strategies, however, the simple structure of the implicit network is exploited to prove rigorous theory, which means that this type of approach is not directly applicable to our $\alpha\beta$-networks. Other approaches include control theory: Nilsson et al. [112] cast the mixed equilibrium STAP as a control problem with the class-split proportion as a control variable.

4.2 Equilibration Algorithm

The mixed assignment model we use is equivalent to the one in [107], where the SVs' objective is to minimise the total system cost. The AVs are thus modelled by a modified Beckmann functional (see section 3.2) that takes into account the flow split. The contribution to the selfish vehicles' (SV) cost due to the AV flow can be included in the objective function by being incorporated into the free-flow cost term that is seen by the selfish class. The functional for the SV cost is therefore

$$\begin{aligned}\Phi^{\mathrm{SV}}\left(\mathbf{x}^{\mathrm{AV}},\mathbf{x}^{\mathrm{SV}}\right) &= \mathbf{x}^{\mathrm{SV}}\cdot\left(\mathbf{a}+D_{\mathbf{b}}\,\mathbf{x}^{\mathrm{AV}}\right)+\frac{1}{2}\mathbf{x}^{\mathrm{SV}}\cdot D_{\mathbf{b}}\,\mathbf{x}^{\mathrm{SV}}\\ &= \mathbf{x}^{\mathrm{SV}}\cdot\mathbf{a}+\mathbf{x}^{\mathrm{SV}}\cdot D_{\mathbf{b}}\,\mathbf{x}^{\mathrm{AV}}+\frac{1}{2}\mathbf{x}^{\mathrm{SV}}\cdot D_{\mathbf{b}}\,\mathbf{x}^{\mathrm{SV}}.\end{aligned} \tag{4.3}$$

Here the flow vectors $\mathbf{x}^{\mathrm{AV}}$ and $\mathbf{x}^{\mathrm{SV}}$ are the link flows of the altruistic and selfish vehicles respectively. In contrast, for the AVs, their objective function is the total system cost. However, they only represent a fraction of the flow so the aggregated flow of SVs and AVs is used in the form

$$\begin{aligned}\Phi^{\mathrm{AV}}(\mathbf{x}^{\mathrm{AV}},\mathbf{x}^{\mathrm{SV}}) &= \Phi^{\mathrm{SO}}(\mathbf{x}^{\mathrm{AV}}+\mathbf{x}^{\mathrm{SV}})\\ &= (\mathbf{x}^{\mathrm{AV}}+\mathbf{x}^{\mathrm{SV}})\cdot\mathbf{a}+(\mathbf{x}^{\mathrm{AV}}+\mathbf{x}^{\mathrm{SV}})\cdot D_{\mathbf{b}}\,(\mathbf{x}^{\mathrm{AV}}+\mathbf{x}^{\mathrm{SV}}).\end{aligned} \tag{4.4}$$

Each class can be seen as a player controlling a proportion of the flow. If we continue with the idea of using centrally-routed autonomous vehicles to reduce total network costs, the SO player can be thought of as a benevolent dictator that controls the AVs' routes. The penetration rate of the AVs in the fleet, γ (with $0 \le \gamma \le 1$), is the proportion of vehicles routed by the altruistic player.

Compare equations 4.3 and 4.4. The objective function of the SVs consists of adding to the standard Beckmann functional (see equation 3.7) a bi-linear form that captures the contribution

of the AVs to the SVs costs: $\mathbf{x}^{\mathrm{SV}} \cdot D_{\mathbf{b}} \mathbf{x}^{\mathrm{AV}}$. That is,

$$\Phi^{\mathrm{SV}}\left(\mathbf{x}^{\mathrm{AV}}, \mathbf{x}^{\mathrm{SV}}\right) = \Phi^{UE}(\mathbf{x}^{\mathrm{SV}}) + \mathbf{x}^{\mathrm{AV}} \cdot D_{\mathbf{b}} \mathbf{x}^{\mathrm{SV}}. \tag{4.5}$$

The AV objective involves more terms, since the congestion term is a quadratic form of the aggregate flows for clarity, equation 4.4 can be re-written as,

$$\Phi^{\mathrm{AV}}(\mathbf{x}^{\mathrm{AV}}, \mathbf{x}^{\mathrm{SV}}) = \Phi^{\mathrm{SO}}(\mathbf{x}^{\mathrm{AV}}) + \Phi^{\mathrm{SO}}(\mathbf{x}^{\mathrm{SV}}) + 2\mathbf{x}^{\mathrm{AV}} \cdot D_{\mathbf{b}} \mathbf{x}^{\mathrm{SV}} \tag{4.6}$$

The last term captures the costs to each class caused by the presence of the other, hence the factor of two.

Inspired by Yang *et al.*, and the practical approach of market economic models, we take a Stackelberg approach in the implementation of our numerical schemes. We can choose, say the UE ($\mathbf{x}^{\mathrm{SV}}$) class, as the leader, and for a particular loading of the network we let them re-equilibrate; meanwhile keeping the SO flows ($\mathbf{x}^{\mathrm{AV}}$) fixed. The shift in flows as the SVs reach equilibrium changes the cost that the AVs experience. The additional cost to the AVs, due to the SV' presence can be thought of as modifying the free flow costs by the addition of $D_{\mathbf{b}}\mathbf{x}^{\mathrm{SV}}$.

The solution is therefore arrived at by iterating over the user classes, for each, minimising its objective function (Φ^{SV} or Φ^{AV}, accordingly) while keeping the flow of the other one fixed. This approach is also known as diagonalisation.

Figure 4.1 shows a flow chart of the equilibration procedure. The sub-problems of solving for UE or SO are described in chapter 3. The alternating iteration, between shifts in SV and AV flows, is also guaranteed to converge for convex cost functions as established by van Vuren et al. [107] and discussed above.

Separating optimisation routines into sub-problems is a standard way of approaching transport equilibria due to their complexity (again we refer to Patriksson, [59], for a comprehensive treatment of traffic assignment). In our case, the sub-problems correspond to the alternating iteration between solving for UE for the SVs and for SO in the AV case. Further, in our work, we use the UE assignment as a simple approach to generate the initial condition, $\mathbf{x}_0$, when calculating ME on $\alpha\beta$-networks.

In terms of convergence, the alternating routine in practice, is fast. Typically, it converges to the ME within a few iterations to the desired tolerance (10^{-7}), the magnitude of flow changes as the algorithm converges is shown in figure 4.2, which shows how flows are re-arranged after each class has rearranged itself on the network. It should not be surprising that this part of the routine is quick, since Roughgarden's bound of PoA $\leq$ 4/3 (for affine cost functions) also bounds the magnitude of the flow-shift.

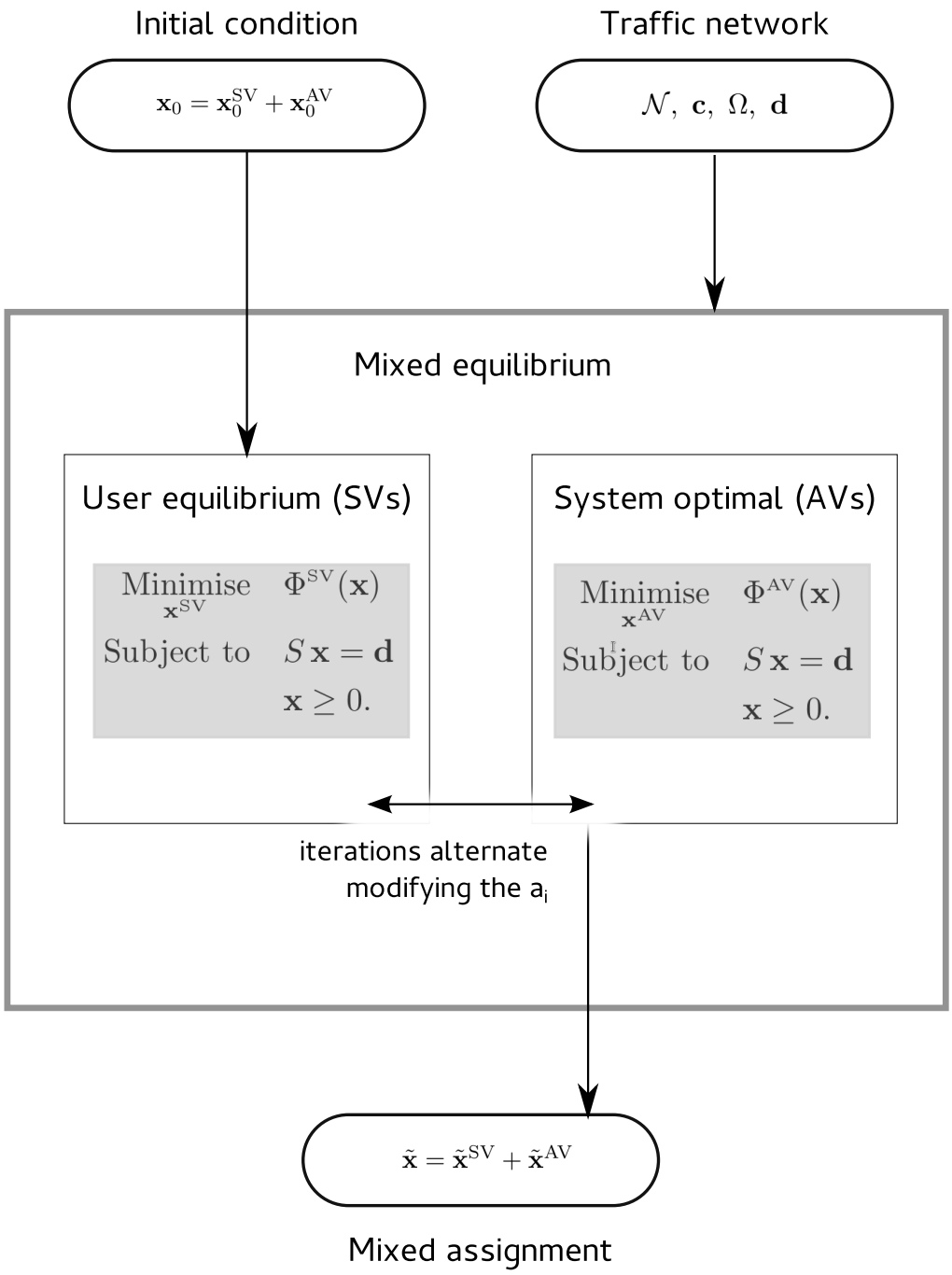


Figure 4.1 Flow chart of the equilibration procedure for the mixed equilibrium (from section 4.2). The inputs are an initial guess for link flows (for example the UE assignment can be used) as well as the traffic network and travel demand. The resulting assignment is the ME assignment.

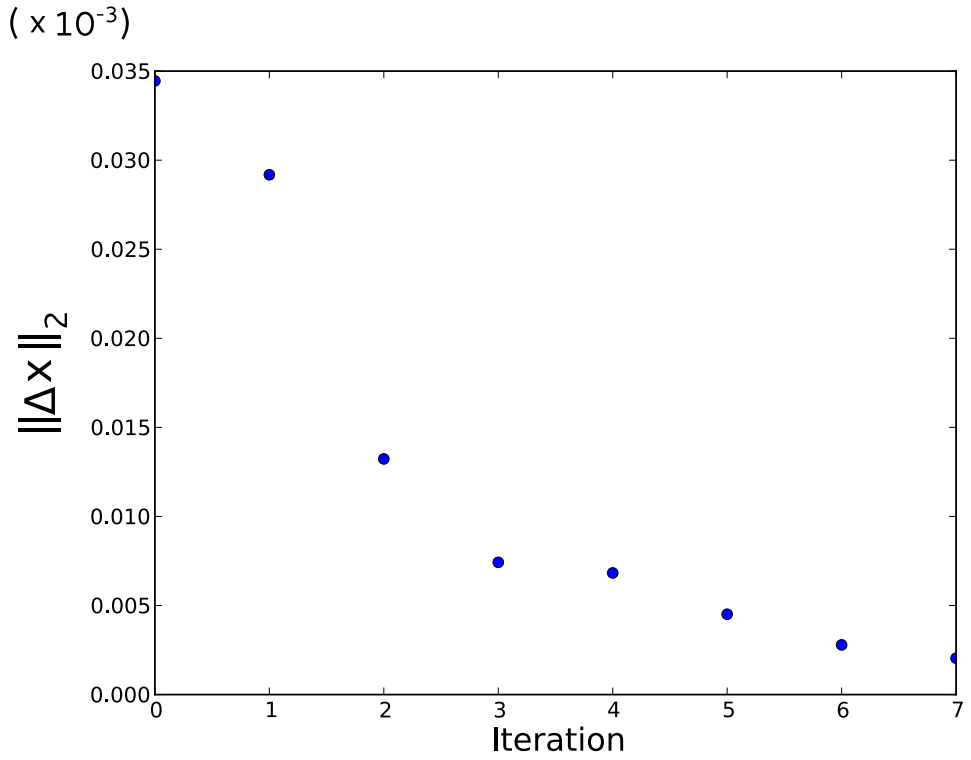


Figure 4.2 Changes in the assignment with consecutive iterations of the alternating routine of a 400 node $\alpha\beta$-network with $\hat{\alpha} = 1.5$ (corresponding to $\alpha \simeq 0.0714$). The total demand is $d = 0.001$ with penetration rate $\gamma = 0.5$. Within a few iterations, the changes to the ME assignment are small. The tolerance for convergence of the cost function was 10^{-7}.

4.3 Transition to Optimality (Small Networks)

In this section we look at some simple networks (where the small number of links is comparable to the number of routes) in order to understand the effects of substituting some of the travel demand by AVs; that is, by increasing the penetration rate, γ, of (altruistically routed) AVs in the vehicle fleet. As has been mentioned above, when the penetration rate γ approaches one, the system approaches optimal costs. However, as we shall see, this is not necessarily a smooth transition due to the discrete structure of networks and the constraints they impose on the possible route sets. The networks that we examine are the Pigou parallel link network and the symmetric Braess network, where the symmetry allows for a treatment equivalent to a two-route network.

We begin with the simplest example, that is a Pigou network, where one of the links is completely insensitive to congestion (*i.e.*, $b_i = 0$). We choose two demand levels, $d = 1$ and $d = 1.4$ (see figure 4.3), that lie on either side of the higher-cost link's activation demand value under UE (which we call the switching demand). This way we can show the effects of crossing the switching demand of the higher-cost link (for the UE assignment) with the SVs, as their volume on the network decreases with increasing γ. The ME transitions to optimality are shown in figure 4.4.

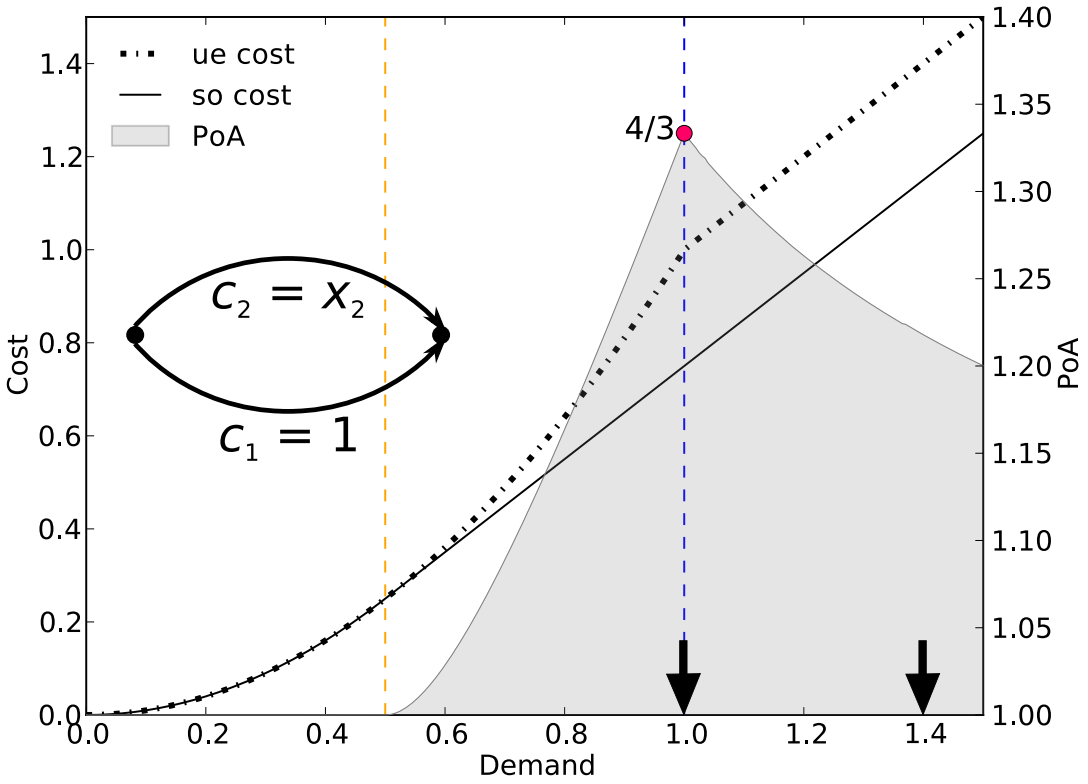


Figure 4.3 The total cost and the PoA for the Pigou network are shown. As shown in the inlay, the cost parameters for this example are $\mathbf{a} = (1, 0)^\top$ and $\mathbf{b} = (0, 1)^\top$. The activation demands of link 1 are shown by the dashed lines; orange for UE and blue for SO. The black arrows show the demand values that we use in figure 4.4: $d = 1$ and $d = 1.4$.

The Pigou network once again demonstrates its explicative power. It exhibits the underlying mechanism through which vehicles are replaced as the proportion of vehicles attempting to solve for SO (the AVs).

For total demand chosen above the switching value, there is no decrease at all in system cost for small γ. This results in a plateau in the cost curve until $\gamma = 0.4/1.4 \simeq 0.29$, which is when the AVs take over the vacancy left by the SVs. The cost does not start to decrease until the proportion of AVs fully covers the original flow of SVs. Then, additional AVs still take the (now more) expensive route, which lets the SVs use the low cost link (link 2) by themselves. What this means is that as γ keeps increasing, this route becomes less utilised until the SO traffic assignment is reached.

For the Braess network, the cost and the PoA as a function of demand are shown in figure 4.5. The demand values at which there are expansions or contractions in the active link set are marked with vertical dashed lines. Thus, for $d = 1$ (which is above the deactivation demand), as SVs are replaced with AVs, the switching threshold for SO (orange dashed line) will be crossed, since at $\gamma = 1$, the SO flow is recovered. This is shown in figure 4.6, along with the flow volumes for both classes along the inner and outer routes. As there is slack between $d = 1$ and the switching

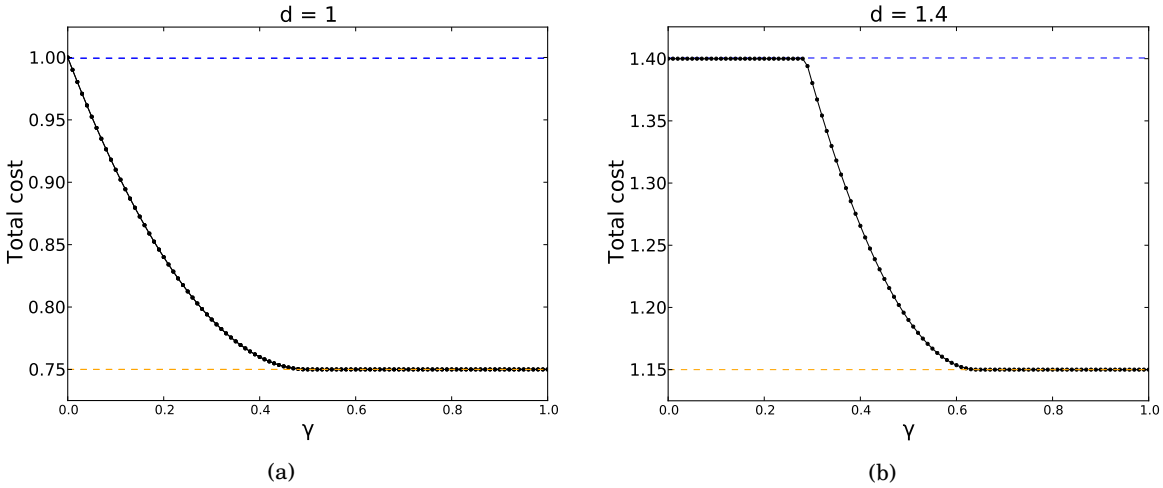


Figure 4.4 ME transitions to optimality for the Pigou network as penetration rate γ increases for two values of demand: (a) $d = 1$ and (b) $d = 1.4$. We can see from the cost plateau on (a) that for demands where both classes share links (above the UE activation threshold), introducing AVs does not immediately result in cost reductions. As the AVs take the congestion-sensitive link, the SVs swap to the other road. In aggregate there is no change in assignment until the proportion of AVs is high enough (in this case $\gamma \simeq 0.29$) so that the travel demand for SVs actually falls below the UE activation threshold of $d = 1$.

demand, similar to the Pigou network example of figure 4.4, there is an initial plateau because the flow corresponding to the SVs is initially replaced by the AVs with no redistribution across the network. In figure 4.6(b), this effect is seen in the cross-over region of the green and blue curves corresponding to the flows on the outer routes. In contrast with the Pigou example, the Braess network does not reach optimal costs until $\gamma = 1$.

If we fix the penetration rate and increase the demand on the networks, we can see how the traffic pattern shifts from SO to UE at different demands. This is because the switching demands for SO are smaller than for UE (in fact, for affine cost functions, the switching demands are twice those of SO [99]). Therefore, the potential efficiency gain that might result from the introduction of AVs is contingent on the congestion level of the network; if the network is so congested that the UE and SO assignments coincide, there is no gain to be made. Clearly, how this might play out in larger networks might be very complicated indeed — see section 4.4.

In figure 4.7, we show the flows on the Pigou and Braess network under ME as a function of total demand. For the Pigou network, this is a simple picture since the flow space has only two coordinates. For the Braess network, since we have chosen **a** and **b** so that the outer-route costs are the same, we sum their flows and show them on the y-axis. The blue curves are the SO costs and the red are the UE ones. The ME (dashed lines) follow the SO at low demands, before crossing to the UE assignment. For higher γ, there is more slack since it takes a higher demand to reach the switching threshold when the volume of the SVs is lower.

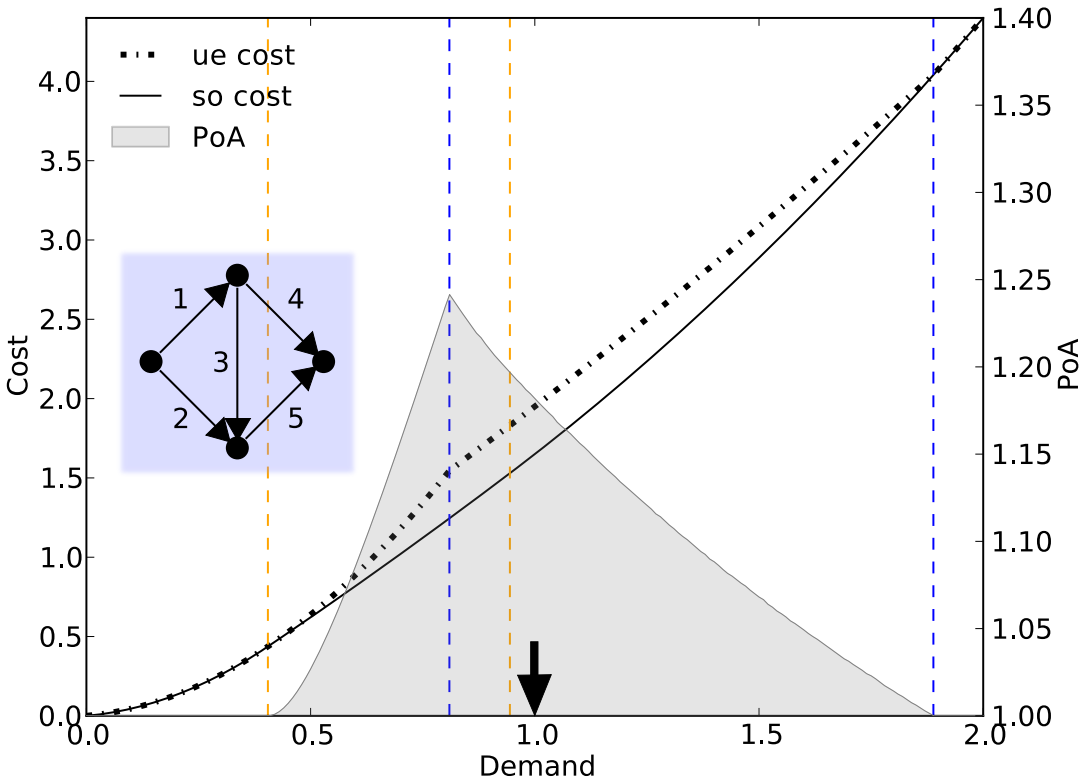
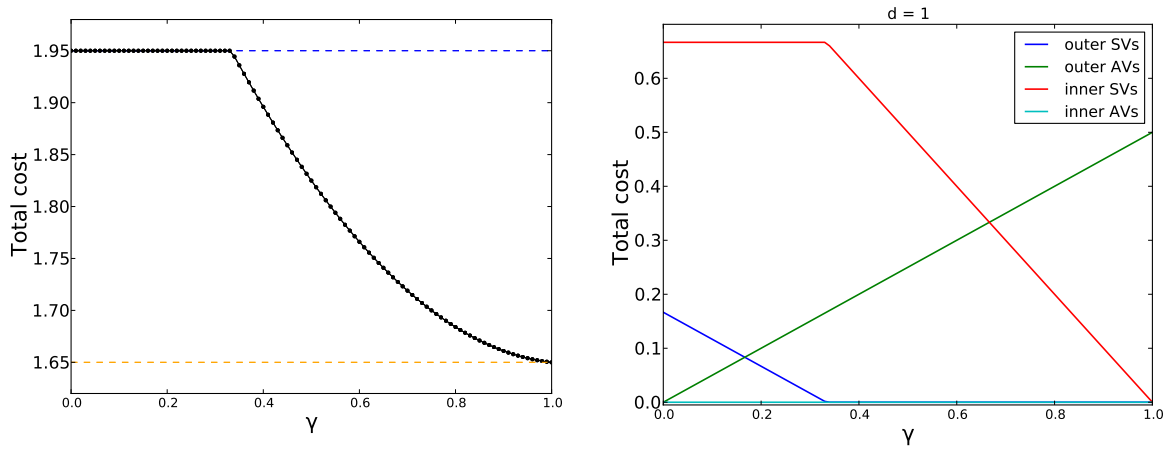
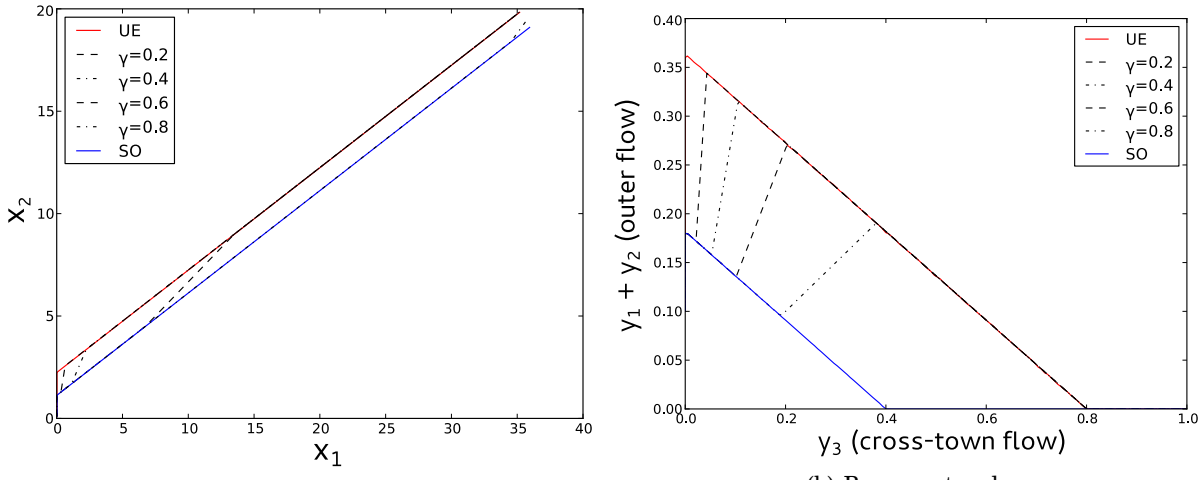


Figure 4.5 Total cost and PoA for the Braess network with parameters: **a** = $(0.1, 1.0, 0.05, 1.0, 0.1)^\top$ and **b** = $(1.0\ 0.1\ 0.05\ 0.1\ 1.0)^\top$. The vertical dashed lines show the activation demand of the outer links (2 and 4) as well as the deactivation demand for the link 3. The blue lines correspond to SO and the orange lines to UE. For each pair of lines, the first corresponds to the activation of the outer routes and the second to deactivating the inner route. The arrow on the axis marks the demand to be used in the ME example, see figure 4.7.



(a) (b)

Figure 4.6 Transition to optimality under ME with increasing penetration rate for the Braess network. (a) Total cost relative to the UE and SO costs. (b) Flows for the inner route and one of the (symmetric) outer routes for both AV and SV classes. Compare with the transition for the Pigou network in figure 4.4. For the Braess network, the system does not reach SO cost until $\gamma = 1$. The reason can be seen in (b): there is non-zero SV flow on the inner route that congests the outer routes which the AVs take exclusively.



(a) Parallel 2-link network (b) Braess network

Figure 4.7 The variation in equilibria for the Pigou network (a) and the Braess network (b), as demand increases. Mixed equilibria for different values of γ are denoted by the dashed lines. The blue and red curves correspond to SO and UE respectively. These curves in the flow space are parametrised by demand. The UE assignment follows the SO assignment (vertical line along the ordinate axis) until the first activation is reached.

4.4 Transition to Optimality (Large Network Ensembles)

The transitions to optimality for larger and more complex networks are messier than for the simple examples developed above. This is because of the great variety of potential flow shifts that happen as more vehicles chose altruistic paths, due to the explosion in the number of paths causing many more cross-overs between cost functions. This in turn causes many plateaus in the costs due to the reservoir effect discussed in section 4.3 above.

In section 4.4.1 we examine the transition to optimal for a single $\alpha\beta$-network. We show that the typical behaviour of class-costs of a single $\alpha\beta$-network is representative of the ensemble behaviour, with AVs incurring higher costs than the original average UE costs until the penetration rate γ is high ($\gamma \simeq 0.8$). The networks studied are based on a 10×10 grid (*i.e*, $N = 100$). In section 4.4.2 we increase N to $N = 225$ to seek size dependent effects.

4.4.1 Preliminary observations for $\alpha\beta$-networks

Figures 4.8 and 4.9 show the transition from UE to SO through the increase in γ of a ME for an $\alpha\beta$-network with $N = 100$ nodes, $\hat{\alpha} \simeq 0.09$ ($\alpha = 0.61$), and $\beta = 1.4$.The ME was calculated for one OD pair, corresponding to ω_1 of chapter 3. Figure 4.8(a) shows the PoA profile and the cost difference between the UE and SO assignments. The value for the aggregate demand ($d = 0.01$) was selected so that the travel demand for the SV class traverses most of the PoA peak as γ increases. The cost transition is shown in figure 4.8(b). The total cost of the assignment ($\Phi_{\mathrm{SO}}(\mathbf{x})$) is normalised so that the UE cost baseline is zero and the UE cost is one, that is

$$\hat{\Phi}(\mathbf{x}) = \frac{\Phi_{\mathrm{SO}}(\mathbf{x}) - \Phi_{\mathrm{SO}}(\mathbf{x}^{\mathrm{SO}})}{\Phi_{\mathrm{SO}}(\mathbf{x}^{\mathrm{UE}}) - \Phi_{\mathrm{SO}}(\mathbf{x}^{\mathrm{SO}})}. \tag{4.7}$$

Compare figure 4.8 and figure 4.4. The form of the cost transition clearly takes a quite different qualitative form the small scale examples of section 4.3. In figure 4.9 we show the difference in costs 'per-vehicle' of the SVs and AVs. Since we have a continuous demand model, this cost-per-vehicle becomes the cost-density per unit of traffic flow. In general, the $\alpha\beta$-networks show a transition in accordance with the literature (namely, [61] and [113]); the total cost for the SV class (as well as the aggregated cost) begins to decrease as soon as AVs are introduced, yet it is not until around 30-40 % of the vehicles are altruistic, that there is a sharp decrease in the normalised total cost.

Observe (figure 4.9(b)) for a single $\alpha\beta$-network, that the sharp reduction in total cost at $\gamma \simeq 0.3$ is a result of the steady decrease of the SV cost in conjunction with the start of the decrease in the AV cost after its initial rise. On closer inspection, we can observe rather the complex structure in the cost reduction. Note that the peak in AV costs mirrors a dip in the SV cost. Note that the exemplar network reaches SO costs before $\gamma = 1$. This finding is consistent with Yang *et al.*'s [108] results for their multiclass equilibria, where SO costs are reached before the whole user fleet are

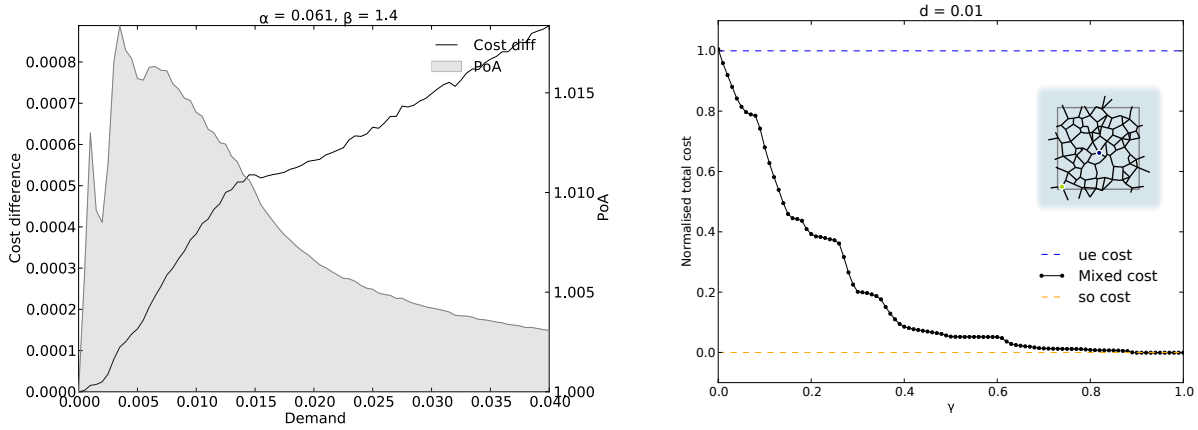


Figure 4.8 Demand range and transition to optimality of network of an $\alpha\beta$-network drawn from an ensemble with parameters $N = 100$, $\alpha = 0.61$, and $\beta = 1.4$. (a) The PoA and difference between UE and SO costs is shown. (b) Transition from UE to SO for the ME as a function of the penetration rate γ, as captured by the total cost of the system. The total cost shown is normalised according to equation 4.7.

AVs, however, the exemplar network we have used is a much more complex network than either of their examples.

The averaged per-vehicle costs of the ME for four different ensembles are shown in figure 4.10. The ensemble parameters are $\hat{\alpha} = 1/3$ and $1/2$ and $\beta = 1.4$ and 1.6. The PoA curves are averages for 10 networks in each ensemble for 20 values of γ which are uniformly spaced between $\gamma = 0$ and $\gamma = 1$ inclusive. A single OD demand was prescribed according to section 3.4. The range $\hat{\alpha} < 1$ (where re-sampling boxes do not overlap) highlights the effect of numerosity of paths, as the number of edges, and hence the number of paths, is higher for lower $\hat{\alpha}$.

In a similar way to the per-vehicle class-costs of our exemplar network, see figure 4.10, the cost density of the AVs does not go below the initial cost of the UE assignment until around $\gamma = 0.8$ or higher. This is an important finding because at that stage the majority of users incur higher costs than at the initially 'inefficient' UE assignment. However, in the bulk of the system, the benefits to the small SV class are so large that at first glance, it seems that the whole system operates 'better'.

4.4.2 UE to SO transitions as the network size is increased

In terms of the total costs, it is useful for us to have a simple model to use as a yardstick for the transitions. For our model to be meaningful, it has to capture a 'reasonable' way of displacing the equilibrium in link flow space: the space whose coordinates are given by $\mathbf{x}$ ($\subseteq \mathbb{R}_+^m$). Due to the model's simplicity, we start with a UE to SO transition that follows a straight line in the

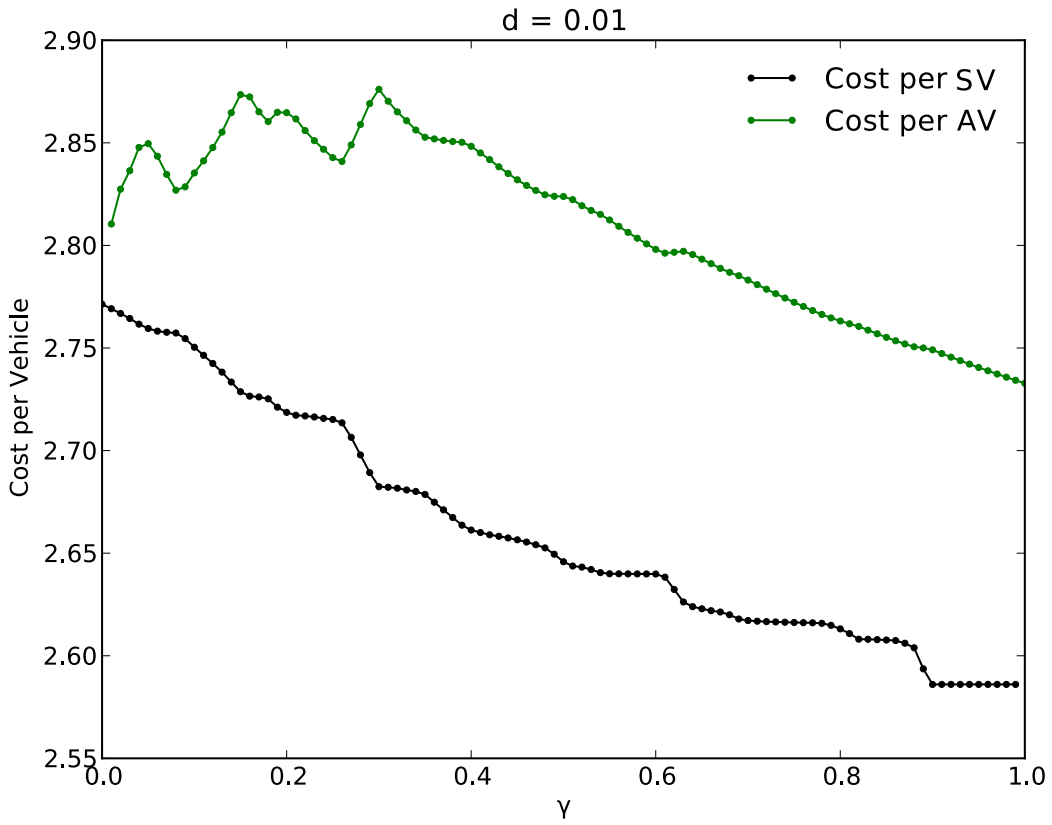


Figure 4.9 The cost-per-vehicle as a function of γ for both classes is shown for the same network as used for figure 4.8). The reduction in costs for the SVs starts as soon as the AV class appears. We can see how the AVs carry more of the system costs, and at first their costs can even go up. Eventually the reservoir is exhausted, and as the AVs become more influential in the network, the cost for AVs decreases as well (at around $\gamma = 0.3$).

flow space, and we refer to is as the *linear transition*. We parametrise this transition by $\gamma \in [0,1]$. Therefore, we define the linear transition, $\mathbf{x}^{\mathrm{LT}}(\gamma)$, by

$$\mathbf{x}^{\mathrm{LT}}(\gamma) = (1-\gamma)\mathbf{x}^{\mathrm{UE}} + \gamma \mathbf{x}^{\mathrm{SO}}. \tag{4.8}$$

Since networks from ensembles with different parameters have different network lengths, $L_{\mathcal{N}}$, in order to compare assignment costs between them, the curves will be normalised with respect to the UE cost as was done in figure 4.8(b) by using equation 4.7.

In figure 4.11, we show transitions of individual networks drawn from a single ensemble ($N = 100$, $\hat{\alpha} = 2$ and $\beta = 1.6$) together with the mean ensemble transitions costs for six different ensembles (all with $N = 100$). Figure 4.11(a) illustrates the intra-ensemble variation. It is clear that the behaviour of individual networks is very diverse, which is strong evidence that local structure has global consequences. The main difference between the networks of the same

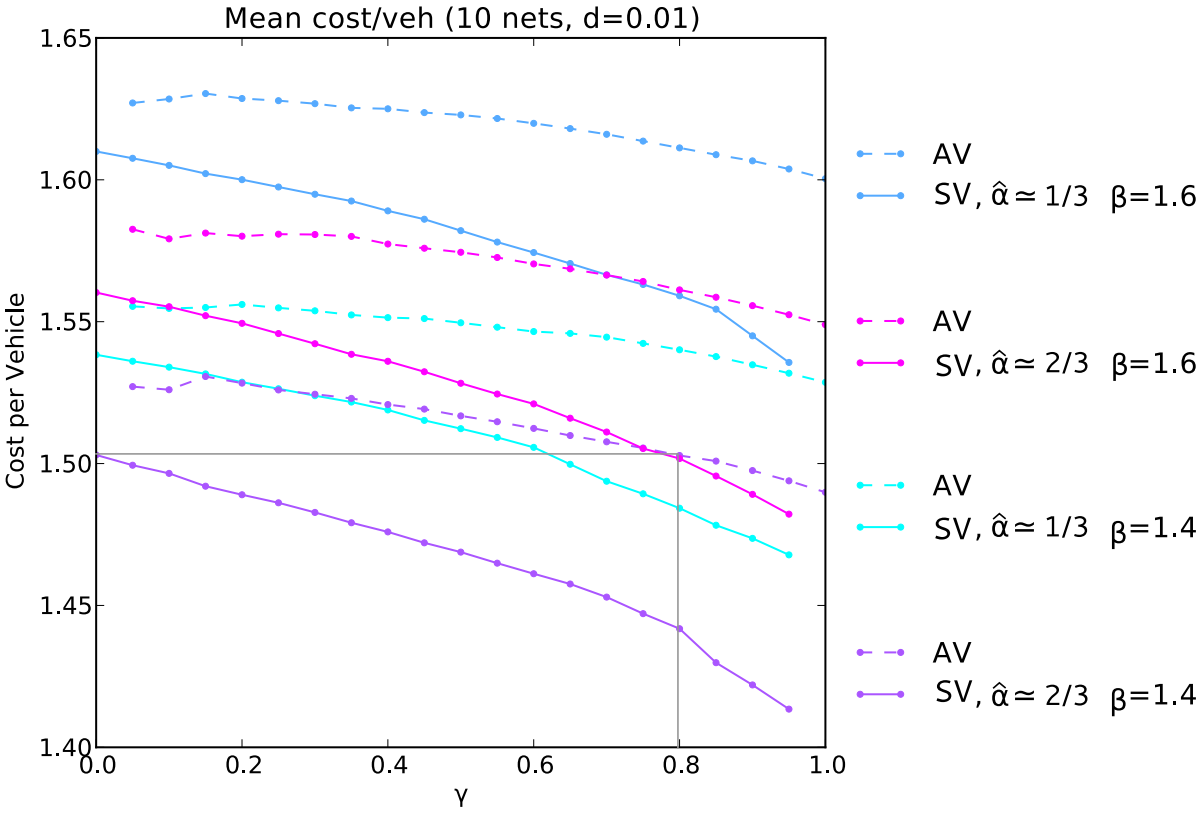


Figure 4.10 Class breakdown of the costs-per-vehicle for four $\alpha\beta$-ensembles ($N = 100$). Costs are averaged over 10 networks. Here $\hat{\alpha} \simeq 1/3$ corresponds to $\alpha = 0.03$, and $\hat{\alpha} \simeq 2/3$ to $\alpha = 0.06$. The box line shows the γ value ($\gamma \simeq 0.8$) at which the costs for the AVs drop to the original UE costs for the ensemble with $\hat{\alpha} = 2/3$ and $\beta = 1.4$. Of the four ensembles shown here, $\gamma \simeq 0.8$ is the lowest value for which AVs experience costs lower than the original UE costs.

ensemble is the shape and location of empty islands determined by the way the intersections cluster (refer back to figure 2.6, page 22).

Compared to the cost curve for the linear transition, the variation is high. However, as we increase the sample size (the number of networks per ensemble) the mean transition curves for all ensembles display a collapse, apparently onto the linear transition.

In terms of the linear transition itself, it is indeed different for all the networks. However, in terms of normalised costs the difference is smaller than the thickness of the line itself. In fact the averaged linear transition converges to the curve $(\gamma - 1)^2$. It is also interesting, yet not surprising, that for 30 networks, the mean costs fall slightly below the transition. This is easily understood by recalling that due to the set-up of the problem, shifts in equilibria due to ME can only lower the total cost, whereas the linear transition only guarantees that the flows are feasible.

As the networks increase in size, the effects of the multiplicity of routes are expected to dominate the behaviour of the networks. Therefore, in figure 4.12, we show the mean transition costs for four ensembles. The first important observation is that the cost transitions fall consistently

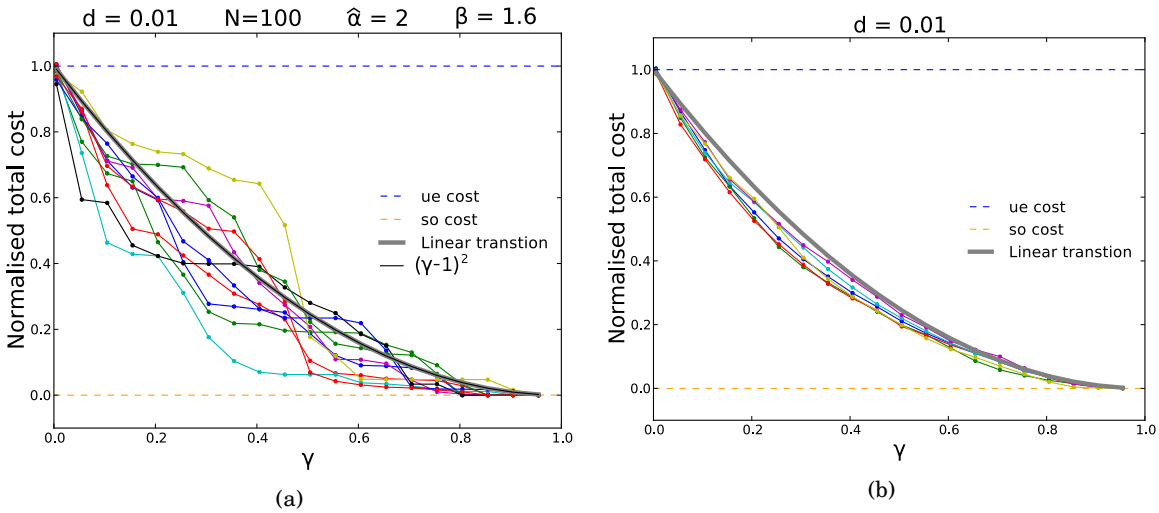


Figure 4.11 Comparison of transitions of ensembles compared to the linear transition. (a) Cost transitions for 10 networks for the ensemble with $N = 100$, $d = 0.01$, $\hat{\alpha} = 2$ (corresponding to $\alpha \simeq 0.18$), and $\beta = 1.6$; the linear transition for the ensemble is the grey line. The thin black line is the curve $(\gamma - 1)^2$. (b) Mean (of 30 networks) transitions for six different ensembles. For each α, two ensembles are used with different β: for $\hat{\alpha} = 1/4$ ($\alpha \simeq 0.023$), $\beta = 1.2$ and 1.8; for $\hat{\alpha} = 2/3$ ($\alpha \simeq 0.06$) $\beta = 1.4$ and 1.6; and for $\hat{\alpha} = 1.75$ ($\alpha \simeq 0.16$) $\beta = 1.2$ and 1.8. The values of β were chosen to have a larger range of road densities for some of the ensembles.

below the linear-transition curve, except towards the $\gamma = 1$ end of the ME. We can also see, that for higher β, the effects of α are much more significant in discriminating between the ensembles. Changing the triangulation of the networks reveals their underlying randomness or griddedness. Surprisingly, the larger networks show more variation, meaning that global properties arise from the underlying clustering of the nodes (recall section 2.3, page 11).

As the networks are scaled up in size, statistical patterns start to emerge. In terms of the best type of structure that the $\alpha\beta$-networks can have, the lower the β is, the faster the ME reaches SO. The number of triangles in the networks reflect the number of possible short-cuts (dependent on cost-function parameters, of course), therefore the results show that increasing the number of routes in the networks increases the efficiency of ME, across almost all values of γ. Contrary to the widespread belief that grids are inherently efficient, in this application it is not the griddedness of the network that matters the most.

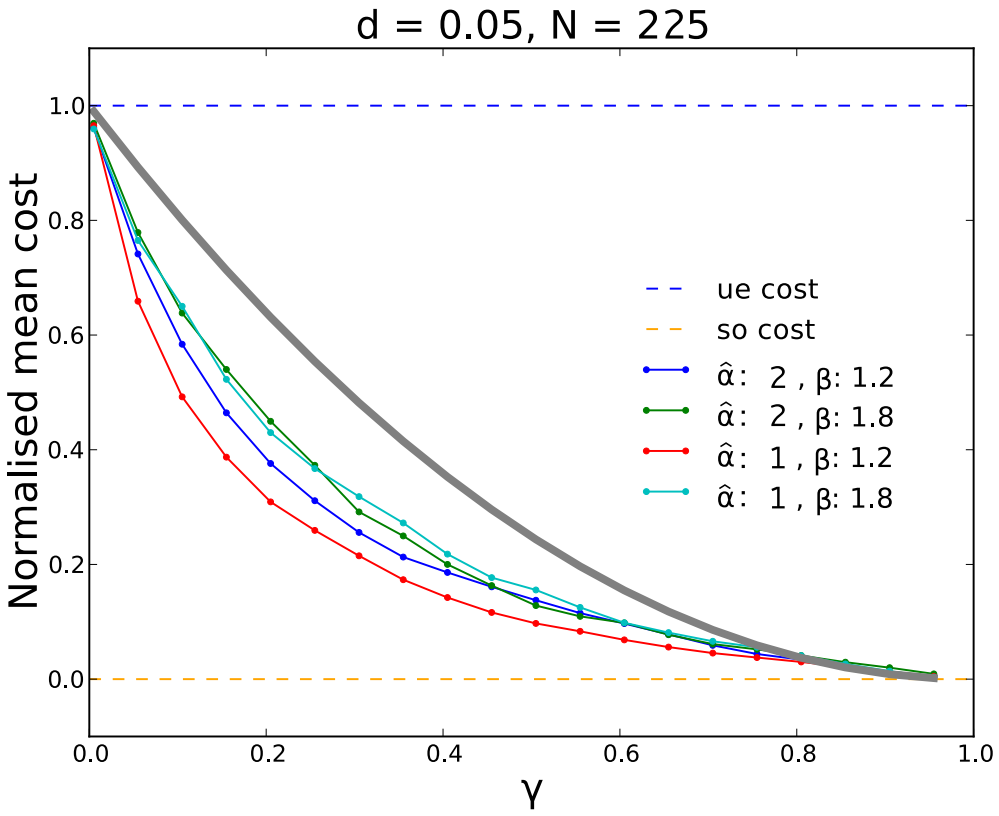


Figure 4.12 Cost transitions of four ensembles based on 15×15 primitive lattices are displayed ($N = 225$). The grey curve indicates the costs of the linear transition $\mathbf{x}^{\mathrm{LT}}$. The samples from each ensemble consist of 10 networks. Compared to figure 4.11(b): the effect of β is more pronounced. Note the spacing between the red and turquoise curves of both $\alpha = \alpha_{\mathrm{crit}} = 0.0625$ ($\hat{\alpha} = 1$), in contrast with overlap of the green and turquoise curves. The overlapping curves correspond to ensembles with $\beta = 1.8$, regardless of the griddedness of the networks.

4.5 Discussion

For small $\alpha\beta$-networks (100 nodes), as well as for demand levels where the cost gap between UE and SO total costs are small, the averaged cost transitions follow the linear transition closely. As route diversity increases by increasing the size of the networks and thus the node density, better efficiencies than the linear transition can be achieved.

On average, as can be seen from the concavity of the transition curves in figures 4.11 and 4.12, most of the cost reduction occurs in the range of $0 < \gamma < 0.4$. The pathways followed by individual networks show that, the intra-ensemble variation is high and individual networks can vary from the mean considerably.

For small networks ($N = 100$) Ensembles of different morphology (as well as total network length), behave like the linear transition between a UE and an SO assignment. As the networks increase in size, and thus so do the available routes, the ME transition to optimality improves

relative to the costs of the linear transition. As the routes increase, the structural features of the networks become more significant, as is shown in figure 4.12. The ensembles with higher β ($\beta = 1.8$) respond better to the introduction of AVs, with consistently lower costs across all values of the penetration rate γ than either of the ensembles with $\beta = 1.2$. As discussed in chapter 2, the effect of low β is that the density of edges is higher, which reinforces the effect of the number of available routes: with a larger number of edges there are more routes available to the users. This means that the AVs can spread themselves out more, allowing them to not interfere with each other's costs as much as for the less densely connected networks.

We remark that an issue that deserves further study is the 'fairness' in the ME assignment. Even when AVs are in the majority ($0.5 \leq \gamma \lesssim 0.8$), they can still experience costs higher than the original costs of the UE assignment. The model we have used, is perhaps to stringent in the requirement of having the AVs achieve SO costs. For one, at low penetration rates, the effect they can potentially have on the network is expected to be small regardless of the network topology; requiring the AVs to minimise some other function (for example the total cost to the AV user class) might cause the cost of the AV class under ME to not be as high.

This chapter serves as a proof-of-concept of how simple, perhaps even overly-artificial models, can be used for the understanding of the efficiency (and inefficiency) of traffic equilibria in a mixed user-class situation.

By using the canonical exemplar networks, Braess and Pigou, we have shown that the demands at which links (and routes) switch on and off determine for which intervals of the penetration rate, γ, there is a cost reduction under ME. We have also shown, by means of numerical experiments, that for more complex networks, the ME presents the same reduction mechanism as for smaller networks with plateaus and quadratic decreases in costs with γ.

We find that under the right demand levels, AVs can indeed drive the system to optimal. There are already significant improvements for γ between 0.2 and 0.4: typically the equilibrium reaches the mid-point between the UE and the SO cost in this interval. This reduction, however, can be deceiving, since depending on the demand and the network, the UE and SO assignments can be very close together. The contributions from the present chapter can be condensed into the following statements.

C4.1 The reduction in costs for small networks happens in a piecewise way that is governed by the switching demands of the active link set. Complex networks, with more overlaps in the switching demands, thus exhibit smoother cost transition curves. Cost reduction occurs mainly due to AVs being displaced to higher cost routes, rather than by encouraging SVs to take better ones.

C4.2 Larger scale networks — that better sample local structure in a single graph — show that route diversity is more important than local network-structure patterns. Larger networks experience larger improvement to aggregate costs than smaller ones, under ME assignment.

C4.3 In this route assignment application, the density of roads (and the number of triangles) matters more than the griddedness of the networks, contrasting with the view that grid networks are inherently efficient.

C4.4 The use of AVs is ultimately a good way of improving system costs. However, it is not until very high penetration rates that the AVs (which at that point are in the majority) experience lower costs than the original UE assignment. The pathway to widespread adoption of AVs thus seems extremely challenging, at least from the point of view of route assignment.

CHAPTER 5

AN EMERGENT FUNDAMENTAL DIAGRAM

This chapter is concerned with the emergence of macroscopic relationships between traffic variables. We use the $\alpha\beta$-networks as an experimental tool to explore how routing and structure determine aggregate behaviours. The ultimate aim of this line of research is to explain the emergence of a *network fundamental diagram* (NFD [114], or *macroscopic fundamental diagram* MFD [5, 48]) from morphological structures of a road network.

The technical challenge faced in this chapter is to maintain the simplicity of Wardropian routing — and the advantages of its static nature — while being able to experimentally 'sample' a phenomenon that is dynamical. For the 'free-flowing' branch (beginning in section 5.2), we exploit the ensemble approach we have been using in this thesis so far. The randomness of the $\alpha\beta$-networks and intra-ensemble differences in road length can be used in our favour, because a large diversity of routing situations are thus realised for any given ensemble. For a morphological network type (measured by $\hat{\alpha}$ and β) the different networks themselves serve as microscopic experimental observations from which we extract macroscopic flows.

The congested branch [115] is harder to obtain, because the analogy to congestion from previous chapters is inadequate in this new context. In an MFD setting, congestion is due to dynamic effects and queuing, not to distributions of the uniformly moving traffic that is considered in the STAP. To address this challenge (in section 5.3), we make use of the blurry boundary between the steps of an optimisation algorithm and their interpretation as a dynamical system. We do this by considering the STAP as a *projected dynamical system* [64]. We lay out the foundations for making the jump to a quasi-static assignment. The novelty here, is in the application of projected dynamical systems [64] (primarily developed in the context of economic equilibria) on small networks, as an experimental method for sampling congested (which we will now understand as sub-optimal) assignments.

5.1 Background

The *macroscopic fundamental diagram* (MFD) attempts to describe, in a similar manner to the *fundamental diagram* of traffic flow theory, how relationships between observed traffic variables are related to each other. The traditional fundamental diagram for a road, or motorway, captures the relationship between traffic flow, speed, and density. In a similar manner, the MFD attempts to capture relationships between aggregated traffic variables in more complex systems, *e.g.*, city road networks.

Determining how network structure gives rise to traffic patterns has been of interest for many decades. For example, Mahmassani *et al.* [116] show that clear relationships between macroscopic traffic variables can be observed in simulations on gridded networks. The widespread availability of sensors to make large scale quality measurements of traffic has more recently led to empirical evidence for the existence of an MFD ([5] by Geroliminis *et. al.*) in Yokohama. This success has spurred the study of conditions for the emergence of an MFD [48, 117] and into ways of deriving MFDs (like in [118]) and approximating them, for example, [119] by Leclercq and Geroliminis.

These studies have shown that a given network may be partitioned into 'homogeneous-enough' sub-regions, each of which has its own MFD. In order to identify MFDs, networks are partitioned using procedures that deliberately minimise data scatter. The resulting subdivided networks tend to be structurally uniform, yet present qualitatively different morphologies for the different sub-networks. Thus, from a traffic management point of view, it shows that network traffic may be modelled by a set of reservoirs that exchange flows [120], which in turn has increased the interest in research on cordoning strategies.

The main result of interest, from the morphological point of view, that we have been building in this thesis, is that 'homogeneous enough' networks can give rise to well defined, unique MFDs. Thus, the $\alpha\beta$-network family is an appropriate tool to to examine how sensitive is the MFD to slight statistical variations in network structure. Using the STAP (see chapter 3) as a foundation, we also, a priori, evade the common oversimplification (as discussed in [121]) of considering trip lengths of all vehicles to be the same. Wardropian equilibria (whether UE or SO) naturally yield distributions of lengths of trips, even for flows belonging to the same OD pair. This avoids introducing additional considerations about user classes based on trip length (as is done in [122] by Batista *et al.*). Even when the partitioning is good, the partition boundaries may fluctuate depending on the demand and traffic state of the network (see, for example [49] by Ji and Gerolimins). To fit the reservoir description of a network, the sub-networks need to be manipulated (sometimes heavily), and in a way that depends on demand and the traffic state.

In a recent study [9] by Mariotte *et al.*, the authors calibrate and validate a multi-reservoir MFD model for the city of Lyon (we reproduce two different partitions that they obtain in figure 5.1). They note the importance of the path length diversity and the usefulness of considering Wardropian equilibria in estimating path flows from induction loop data. They conclude that while methodologically useful in some cases, Wardropian equilibrium routing is a naïve and can

Figure 5.1 The Lyon road network partitioned into five and ten reservoirs respectively. Reproduced from [9], by Mariotte *et al.* (corresponding figure 4 in their paper).

fail to describe real-world flows. The simplicity inherent in the STAP, however, will be useful to us as a conceptual starting point.

Our focus is on the emergence of equilibrium patterns from collective behaviour as a consequence of morphological properties of the networks themselves. Conceptually, this differs from the normative approach [49, 117]. We adopt the lesser-used term *network fundamental diagram* (NFD), to highlight that the aggregation we consider is over the network. In fact, the unaggregated traffic variables are already macroscopic. The coarse graining done to obtain the MFD reflects a scaling-up of the system in terms of the complexity (and variation) of a transport system's routing options. As a starting point we ask the question: does starting with homogeneous network regions, or patches, lead to the emergence of a macroscopic fundamental diagram for the network?

The NFD consists of two regimes formed by a congested and an uncongested branch. The uncongested branch is the most understood, partly because it is the easier branch to observe empirically, it exhibits the behaviour of the network under steady flows. The congested branch can be understood in terms of queuing and spill-back effects, however, the transient nature of traffic jams reduces the frequency of empirical observations with the scatter-minimising partitioning approach. Therefore, in section 5.2 we show how, even with high intra-ensemble variation, uniform networks yield distinct uncongested NFDs. In the subsequent sections we change gears and focus on how to take advantage of Wardropian parsimony and incorporate a type of time-dynamic that will yield feasible, yet sub-optimal assignments. From the point of

view of uncapacitated static assignment, congested network states and sub-optimal ones are synonymous.

5.2 Uncongested Branch of the NFD

The static nature of the STAP allows us to easily draw the correspondence between free-flowing traffic and the time invariant flows of equilibrium traffic patterns. The subtlety is the instant propagation of traffic conditions through the network due to the simplicity of the STAP model. We also note that, as in the previous chapters, the choice of OD pattern is crucial. To study the emergence of an NFD in a morphology-centric way, simplifications of OD patterns, and simulations of different traffic states of the network, must be suitably made. Due to the dynamic nature of the MFD, when it exists, it persists even when the data points are not causally correlated amongst themselves. As a proxy for multiple observations at uncorrelated time intervals, we make use of the ensemble formulation. The (macroscopic) traffic state of a single network plays the role of an observation point of the traffic states accessible to an 'ideal network', abstracted in terms of its morphological parameters α and β.

In order to avoid building a trivial model, we need to ensure that even across networks of the same morphology type and equal demand levels, there is diversity in the realised traffic states. The way we achieve this is by setting up an open system, where vehicles in transit through the sub-network do not necessarily have their origins or destinations within it. The sub-networks we consider are composed of the nodes and edges completely contained within the central quadrant of each $\alpha\beta$-network. We refer to these sub-networks as the region of study, A, for which the NFD will be studied. Figure 5.2 illustrates the way we partition the networks, and how the central quadrant satisfies the desired 'openness' of the region in terms of having incoming and outgoing traffic as well as having journeys originate or terminate within A.

The general idea of our method is as follows. For a given ensemble (determined by α, β and N), networks are sampled and the STAP is solved for multiple OD pairs and for different demands. To avoid overloading OD pairs that are close to each other, for a given global demand d_Ω, the demand corresponding to each individual OD pair, d_{ω_i}, is proportional to the Euclidean distance between the origin and destination nodes. That is,

$$d_{\omega_i} = d_\Omega \frac{|\omega_i|}{\sum_{\omega\in\Omega}|\omega|}, \tag{5.1}$$

where $|\omega_i|$ is the euclidean distance (modulo the periodic metric introduced by the toroidal boundary conditions) between the origin and destination nodes of OD pair ω_i.

Macroscopic traffic quantities are then aggregated over the central quadrant of each of the networks. The traffic quantities for each quadrant, and for each demand value are taken as a traffic state 'observed' on the particular ensemble: thus, allowing for many different traffic states that are representative of the ensemble, rather than of a particular individual network.

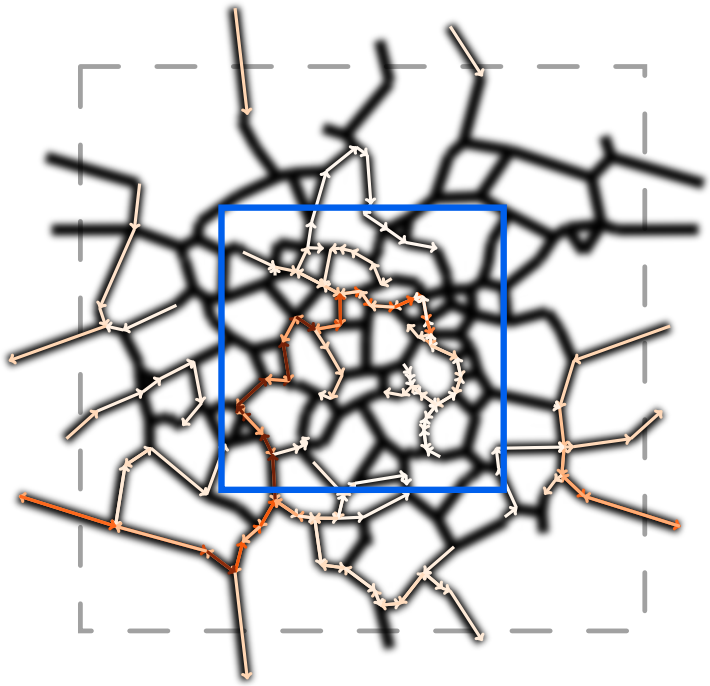


Figure 5.2 An $\alpha\beta$-network (with $N = 100$ nodes) with the central quadrant, region A, shown with the blue square. The traffic pattern is the UE pattern corresponding to ten randomly chosen origins and destinations. In our scheme, a single node can belong to at most two OD pairs, but cannot be the origin (or destination) of both.

5.2.1 Experimental set-up

In this section we describe the set-up for the numerical experiments. The key feature is the way demand is increased on the networks. To approximate a realistic loading, the demand is increased in fixed steps. To simulate that time is elapsing and that vehicles are reaching their destinations (which has to be externally imposed on the stationary flows given by the STAP), for each demand value some of the cost of each edge is incorporated in to the free-flow travel-time. For demand at time-step t, d^t, we modify the free-flow travel time of each edge as follows,

$$a_{i,t} = a_i + \frac{1}{2} b_i x^{\mathrm{UE}}_{i,t-1}. \tag{5.2}$$

This is similar to modelling many time-slices, where at slice t, users are influenced by the traffic state at $t-1$. Only half of the delay term is incorporated to the a_i to represent some users from time slice $t-1$ having reached their destinations by time-slice t. Even though we talk about time slices, the stationarity of the traffic patterns mean that from $\mathbf{x}^{\mathrm{UE}}_t$ to $\mathbf{x}^{\mathrm{UE}}_{t+1}$, the difference is not that a fixed Δt has elapsed, but that the network has reached an equilibrium under the new conditions. Keeping the changes in demand small enough (therefore maintaining $\Delta\mathbf{x}^{\mathrm{UE}}$ small as well) is what gives the illusion of time passing, without the need to consider more sophisticated interventions to make the STAP into a dynamic (or quasi-dynamic) assignment model (see for example [123],

although we are deliberately avoiding the type of sophistication needed to make the routing match observational data). A comprehensive discussion of how different traffic assignment models are related to each other is presented by Bliemer *et al.* in [124].

The macroscopic traffic variables we calculate correspond to the well-known microscopic ones involved in the relationships referred to as the (classical, *i.e.*, link-based) fundamental diagram. The network level definitions of variables that we use are *trip production* P_A and region density ρ_A. The production is calculated by

$$P_A = \frac{\sum_{i \in A} v_i x_i}{L_A}, \tag{5.3}$$

where L_A is the total road length of all roads fully contained in A and v_i is the speed of vehicles on edge i. Trip production is the total veh$\cdot$km travelled in region A per unit time. The region density is given by

$$\rho_A = \frac{\sum_{i \in A} x_i c_i(x_i)}{L_A} = \frac{k_A}{L_A}, \tag{5.4}$$

where k_A is the number of vehicles in the region, or more specifically, the aggregated vehicle occupation of all the links completely contained in A.

Figure 5.3(a) schematically shows how a hypothetical loading and unloading of demand is done. As a sense check of our experimental approach, in figure 5.3 we show the production NFD observed during loading and unloading of a single network with our method. We note the different path taken during the loading phase and the unloading phase. In our model, this is a consequence of the perceived costs to users at step t, since the observed free-flow costs incorporate some of the demand from the previous step. Thus, in the unloading phase there is a significant reduction in production. Note, however, that qualitatively this behaviour is the same as the hysteresis effects that are observed in fully dynamical simulations [125] as well as observationally [10]. Encouraged by this finding, that with the quasi-dynamic modification to the STAP we can qualitatively recover the same type of network behaviour that is observed under fully dynamic conditions, we continue with the ensemble approach.

The networks we use have $N = 225$ nodes each (generated from a 15×15 grid). The ranges chosen for the morphological parameters are $\alpha \in \{0.2, 0.4, 0.6, 0.8, 1\}$ and $\beta \in \{1.2, 1.4, 1.6, 1.8\}$. For each (α, β) pair an experimental ensemble of 100 networks is used. A random OD pattern of 20 origins and destinations is used for each network ($\sim 10\%$ of the nodes). The total network demand is varied in a range that explores the demand region with the largest cost gap between the UE assignment and optimal costs. The demand for each OD pair is scaled to be proportional to the Euclidean distance between each OD pair (equation 5.1). This way, we avoid overloading links incident to origins and destinations. The total demand of the networks was increased from 0.2 to 4 in steps of 0.1. From each demand value to the next, the free-flow costs are modified according to equation 5.2.

Figure 5.4 shows the speed-density relationship for four different ensembles. The data points corresponding to each ensemble form striations that are clearly separated from each other,

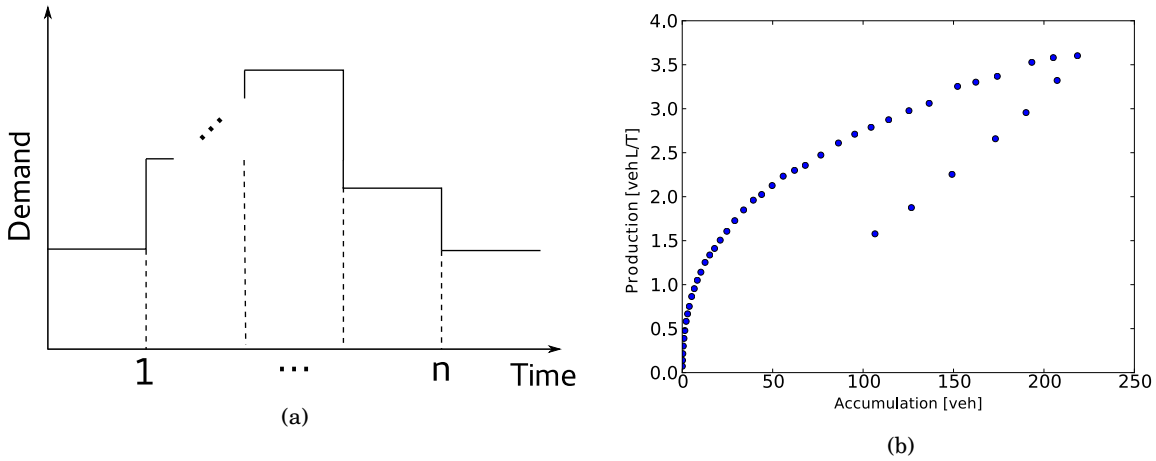


Figure 5.3 (a) Increase and decrease of demand used with the scheme expressed in equation 5.2 and (b) the response of a network of $N = 225$ ($\alpha = 0.06$ and $\beta = 1.4$) nodes for demand increasing from $d_\Omega = 0.2$ to $d_\Omega = 4$ in steps of $\Delta d_\Omega = 0.1$ and subsequently decreasing to $d_\Omega = 1.6$ in steps of size $\Delta d_\Omega = -0.4$. Note the difference in loading and unloading, which is similar to the hysteresis shown in [10].

showing the formation of well-behaved relationships between the network traffic variables. In figure 5.5 the trip production (P_A) *vs* density (ρ_A) is shown for 20 ensembles. Each data point is an average of the density and production over a whole ensemble subjected to the same global demand. The curves of fit are of the form $P_A = C\sqrt{\rho_A}$, with fit coefficient C. They are coloured according to the mean observation region length L_A to show the trend of the family of P_A *vs* ρ_A curves with network length. Each data point in figure 5.5 corresponds to the ensemble average of P_A over 100 networks.

There is a near-monotonic relationship between region length and performance, which is a sense-check on the results: as infrastructure supply increases so does trip production. However, the effects of network topology can be seen more clearly in figure 5.6. For mid-α values the spread of the NFD curves due to β means ensembles with similar α can be hard to distinguish by performance alone. Near the turning point $\alpha \simeq 0.5$, the coefficient lines are parallel and close. As α increases the production curves for ensembles with very different griddedness (see lines for $\alpha = 0.2$ and $\alpha = 0.8$ in figure 5.6) yield similar production curves but can be distinguished due to their values of β.

For $\alpha = 1$, the linear trend of the fit coefficient is not as good. In these ensembles, due to the interplay between clustering of nodes and the edge density, the networks show more diversity, thus exhibiting more complex behaviour.

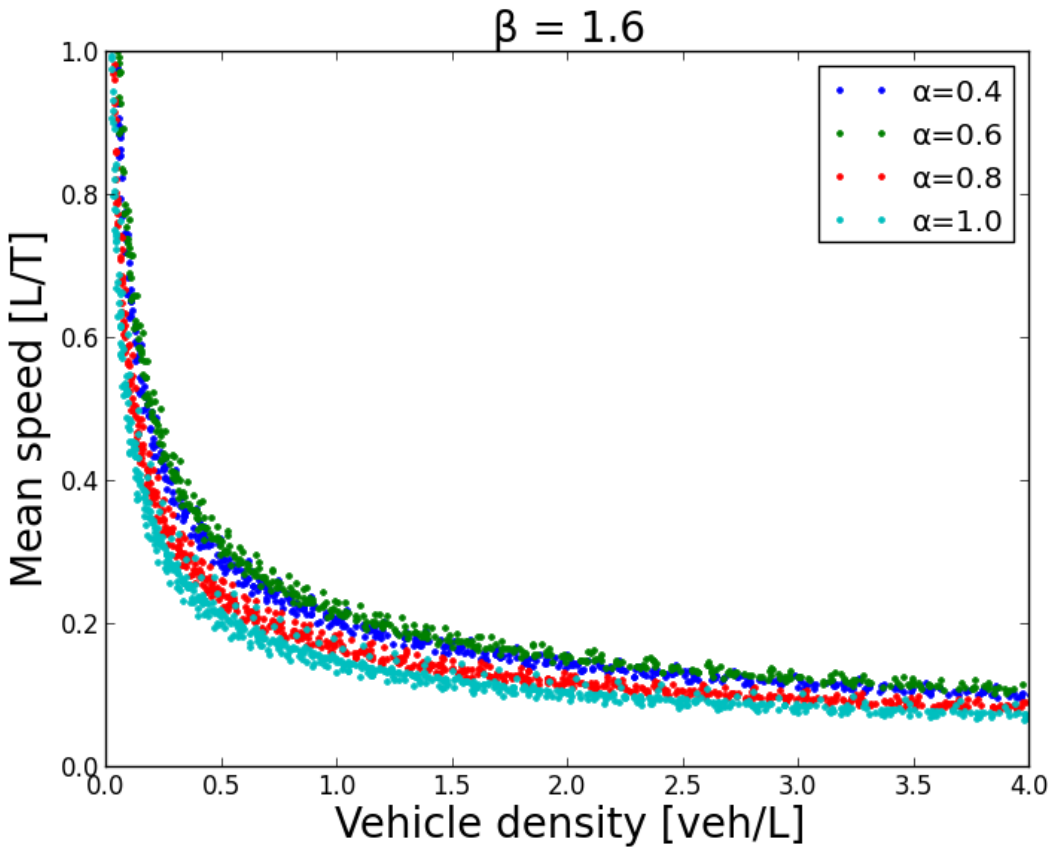


Figure 5.4 The (vehicle) mean speed against density is shown for four ensembles ($N = 225$) that cover a large range of α for $\beta = 1.6$. Each data point corresponds to a particular demand value for one of the networks sampled for each ensemble.

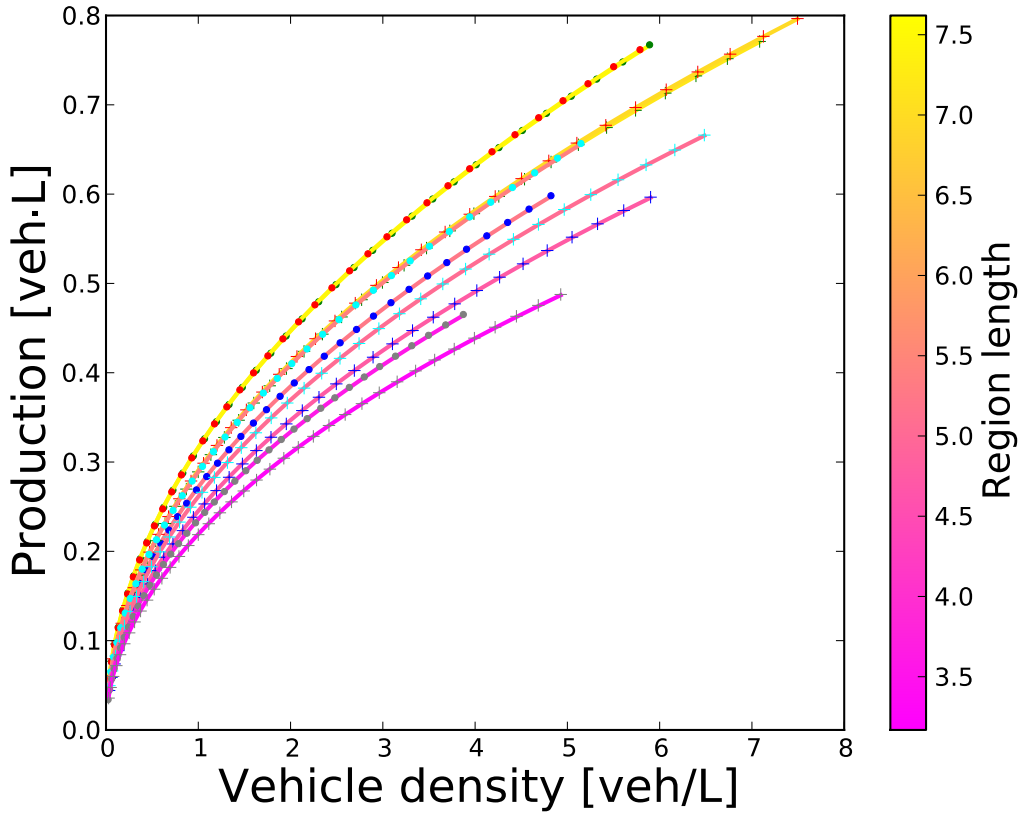


Figure 5.5 The increase of trip production P_A with respect to vehicle density ρ_A. The marker colours blue, green, red, cyan, grey correspond to $\alpha = 0.2, 0.4, 0.6, 0.8$, and 1, respectively. The different shapes correspond to different values of β. The fitted curves are coloured according to the region length L_A and are of the form $P_A = C\sqrt{\rho_A}$. In general L_A is larger for networks that perform better, however the trend is not smooth.

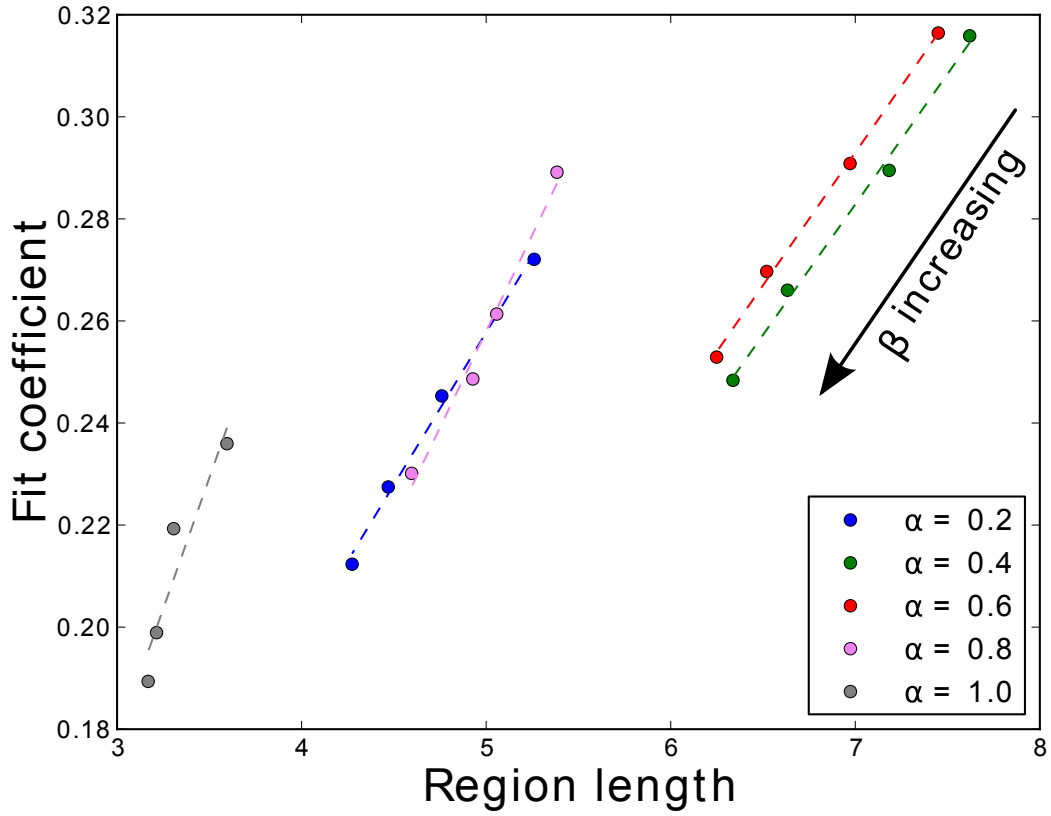


Figure 5.6 The fit coefficient C is plotted against the L_A. The coefficients depend linearly on the region length which is strongly modulated by α. Note the cross-over of the $\alpha = 0.2$ (blue) and $\alpha = 0.8$ (violet) lines reflects the staggering of the production curves in figure 5.5.

5.3 Towards the Congested Branch

Building up the congested branch from the STAP is considerably more complicated than the uncongested section of the NFD. This is mainly because the type of congestion is dynamic in nature. In more realistic (or detailed) traffic settings, delays and congestion are mostly due to queuing at intersections. For examplethe two-fluid theory of Herman and Prigogine [126] is built precisely on this idea.

As mentioned at the beginning of the present chapter, we are concerned with how to equip the STAP so as to obtain different congested states in a dynamic way. Central to the existence of the NFD is its time invariance, that is, the persistence of the structure over long periods of observation. Also, in the congested branch of the NFD, many traffic jams are expected to share a common structure regardless of their microscopic details. This is reminiscent of orbits converging to strange attractors in chaotic dynamical systems. In the limit where an orbit converges onto the attractor, a large part of the phase space is explored as an effect of topological transitivity (a standard result from dynamical systems theory, for example refer to [127] or [128]). Modern dynamical systems theory has shown the usefulness of piecewise-smooth dynamical systems, in both applications ([129]) and in new methodological techniques to deal with hard classical problems dealing with friction and compliance (*e.g.*, [130, 131]).

In what follows, instead of extending the STAP model to move closer towards the dynamical regime, we revisit models showing convergence to Wardropian equilibria and relate them with so-called Projected Dynamical Systems (PDS). A detailed monograph of these systems is [64] by Nagurney and Zhang. These dynamical systems are a type of non-smooth system inspired by ideas from variational inequalites from optimisation theory and projected gradient descent algorithms. The main idea is that for some variational inequality problems, a corresponding constrained dynamical system can be constructed so that its stationary states coincide with solution sets to the variational inequality formulation.

5.3.1 A dynamical twist: projections and flow-swaps

In contrast to the optimisation formulation, in which equilibrium states are found as the (global) optimal of an objective function within a constraint set, one can consider a dynamical system in 'flow-space' that has the Wardropian *user equilibrium* flow pattern as a stable steady state. The equations of motion would then describe how users of a transport network react (in subsequent journeys) to the costs of routes (of their current journey). The way the flows move around in this flow-space represent how users shift between routes to reach equilibrium from an initial state. The time evolution of the system tracks the traffic pattern throughout the equilibration procedure until no better costs can be achieved by any user.

The structure of the flow space is governed by the network structure. In order to understand the effects of route and link conservation constraints on the flow space, after defining the

dynamical systems we examine, we will give three examples of different networks that capture different situations: the Pigou network, where the links coincide with the routes; the Braess network to generalise to higher dimensions; and lastly, the lollipop network with three links and two routes, to explicitly show the structure of the flow space when the systems are cast in terms of link flows as opposed to route flows.

The constraints of the STAP define a convex feasibility region. In fact, they define a polyhedron in flow-space. If we imagine a flow swapping mechanism that pushes the route flows in the direction where the costs decrease, then when the flow vector is in the interior of the constraint set, the change is in the direction opposite to the cost vector $\mathbf{C}(\mathbf{y})$. If the flow vector lies on the boundary, then the trajectory of the flow vector evolves parallel to the projection of the vector $-\mathbf{C}$. If we call the convex constraint set K, then the projected dynamical system

$$\dot{\mathbf{x}} = \Pi_K(\mathbf{x}, -\mathbf{F}(\mathbf{x})), \tag{5.5}$$

captures the behaviour of the equilibrating network when the vector field $\mathbf{F} = \mathbf{C}$. The operator $\Pi_K(\mathbf{p}, \mathbf{v})$ is the projection of vector $\mathbf{v}$ at $\mathbf{p}$ onto the convex set K. That is,

$$\Pi_K(\mathbf{p}, \mathbf{v}) = \lim_{\delta \to 0} \frac{P_K(\mathbf{p} + \delta \mathbf{v}) - \mathbf{x}}{\delta}. \tag{5.6}$$

Since K is a convex polyhedron defined by the demand conditions for each OD pair ($\sum_{p \in P_\omega} y_p = d_\omega$), the projection P_k is well defined. Figure 5.7 shows the general behaviour of the projection onto K. Points in the exterior are projected onto the bounding faces, possibly more than one, in which case the resulting $P_K(\mathbf{v})$ will lie on an edge or vertex of the constraint polyhedron rather than just on a face.

For the case where the closest point of K to the vector $\mathbf{v}$ being projected is a vertex (like $\mathbf{p}_2$ in figure 5.7) the normal vector to K is not well defined. For each face of K that the vertex belongs to there is a normal vector defined. We show the direction of the normals at the top vertex with the dashed red lines. At the vertices, the possible vectors that can be thought of as normal to K are contained in the normal cone at the vertex. The normal cone is comprised of the conic combinations of the normals of all the faces to which the corner (or edge) in question belongs.

To show that the right hand side of equation 5.5 is not a smooth function, we observe what happens to the projection of a vector as the point where it is being applied to approaches a corner of K. In figure 5.8 we show a vector field $\mathbf{F}$ and the projections of $\mathbf{F}$ at two points $\mathbf{p}$ and $\mathbf{q}$ along a face of K. If we consider displacing $\mathbf{p}$ along the face towards $\mathbf{q}$ in the direction of the projection $\Pi(\mathbf{p}, \mathbf{F}(\mathbf{p}))$ (shown as red arrows), then when the top vertex $\mathbf{q}$ is reached, the projection of the field, now $\Pi(\mathbf{q}, \mathbf{F}(\mathbf{q}))$, points tangential to the left face; the projected vector suffers a jump discontinuity as the top vertex is reached.

For vectors in the interior of the constraint set, $\mathbf{x} \in K^\circ$, the projection map is,

$$\Pi_K(\mathbf{x}, \mathbf{v}) = \mathbf{v}. \tag{5.7}$$

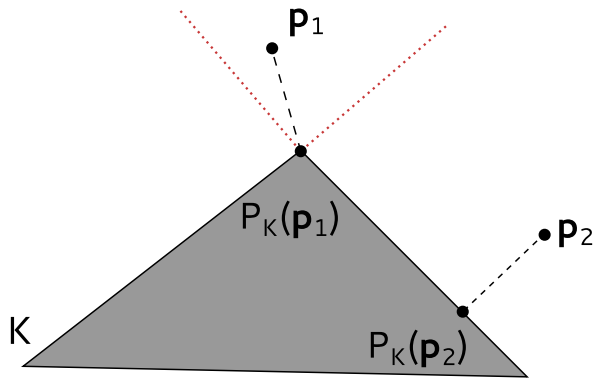


Figure 5.7 For the triangle K projections of two points $\mathbf{p}_1$, $\mathbf{p}_2$ are shown. $P_K(\mathbf{p}_1)$ lies on the corner since it is contained in the normal cone to K at the top corner. For projections that lie on one face only are like that shown for $\mathbf{p}_2$, with $P_K(\mathbf{p}_2)$ being the point that minimises the distance to the closest plane that defines the faces of K.

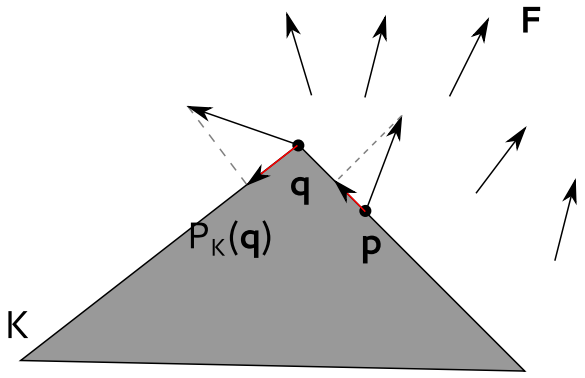


Figure 5.8 The projection of a field $\mathbf{F}$ is shown for two points on the boundary of a convex polyhedral set K. The projection of the field at point $\mathbf{p}$, $\Pi(\mathbf{p},\mathbf{F}(\mathbf{p}))$, has a jump discontinuity at the vertex: visualised as following the displacement of point $\mathbf{p}$ as it slides towards $\mathbf{q}$ along the boundary of K.

In terms of a projected dynamical system following equation 5.5, if $\mathbf{x}$ is in the interior of K then its velocity is just given by the unconstrained vector field (in this case $\mathbf{v}$).

For vectors in the boundary of K, $\mathbf{x} \in \partial K$, the projection can be obtained by subtracting from $\mathbf{v}$, its component normal to K. If the projection lies on a face of K then the normal is the normal vector to the plane that defines the face. If, like at point $\mathbf{q}$, the normal has to be chosen with more care, its projection can be written as

$$\Pi_K(\mathbf{x},\mathbf{v}) = \mathbf{v} + \gamma(\mathbf{x})\mathbf{n}^\star(\mathbf{x}), \tag{5.8}$$

where the appropriate (inward) normal vector $\mathbf{n}^\star$ is the one that aligns the most with $\mathbf{v}$,

$$\mathbf{n}^\star(\mathbf{x}) = \arg\max_{\mathbf{n}\in\mathbf{n}(\mathbf{x})} (-\mathbf{n}\cdot\mathbf{v}). \tag{5.9}$$

The set of possible normal vectors to K at $\mathbf{x}$ is the set of all vectors $\mathbf{n}(\mathbf{x})$ which define planes of support of K at $\mathbf{x}$.

Given the linear constraints we will be dealing with, care must be taken at corners of K. Going back to figure 5.7, any vector contained within the dotted red lines at the top vertex (for instance $\mathbf{p}_1$) belongs to this set of normals. The scalar γ in equation 5.8 is the distance from K to $\mathbf{v}$. And we only need to project it if points outwards to the set, so it can be written as,

$$\gamma(\mathbf{x}) = \max\{0, -\mathbf{n}\cdot\mathbf{v}\}. \tag{5.10}$$

This way of expressing Π_K is given in [64]. Here we give it to make it clear that while the projected dynamical view might seem cumbersome, the actual projection mechanism can be understood geometrically in a simple way: if the vector $\mathbf{v}$ points outwards of K at some $\mathbf{x}$ on the boundary, then project it back onto ∂K along the 'best' tangent. If the point $\mathbf{x}$ is either an interior point or if it is a boundary point but $\mathbf{v}$ points inwards, then $\Pi_K(\mathbf{x},\mathbf{v}) = \mathbf{v}$.

The main reason for giving the form for Π_K as in equations 5.8-5.10, is to show the similarity with the flow swapping mechanism of Smith [63]. Equation 5.10 essentially generalises the $(\cdot)_+$ operation used by Smith to keep flows positive. The difference is that the constraint set K in general can arise due to different constraints.

In general, the projected dynamical system described by,

$$\dot{\mathbf{x}} = \Pi_K(\mathbf{x}, -\mathbf{F}(\mathbf{x})), \tag{5.11}$$

when compared to the corresponding unconstrained dynamical system that evolves according to

$$\dot{\mathbf{x}} = -\mathbf{F}(\mathbf{x}) \tag{5.12}$$

will have different fixed points (to start off, they have to be contained in K). And even when the projected system has fixed points arising from the projected dynamics, the type of stability does not have to correspond to that of the unconstrained system.

To give an example (inspired by example 2.1 in [64]), the phase portrait of the following projected dynamical system is shown in figure 5.9,

$$\dot{\mathbf{x}} = \Pi_{K_1}(\mathbf{x}, -\mathbf{H}(\mathbf{x})), \tag{5.13}$$

where $\mathbf{x} \in \mathbb{R}^2$, and the constraint set K_1 is the triangle defined by

$$\begin{aligned} x_1 &\geq 0, \\ x_2 &\geq 0, \\ x_1 + x_2 &\leq 1, \end{aligned} \tag{5.14}$$

and the vector field $\mathbf{H}$ is given by

$$\mathbf{H}(\mathbf{x}) = \mathbf{H}_1(\mathbf{x} - \mathbf{c}), \tag{5.15}$$

with

$$\mathbf{H}_1(x_1, x_2) = \begin{pmatrix} -x_2 \\ -4x_1 \end{pmatrix}, \tag{5.16}$$

and

$$\mathbf{c} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \tag{5.17}$$

The vector field $\mathbf{H}$ rotates around $\mathbf{c}$, and for the unconstrained system, the trajectories would be ellipses around $\mathbf{c}$. That is, $\mathbf{c}$ is a centre. However all trajectories starting in K_1 converge to the top-left vertex at (0, 1).

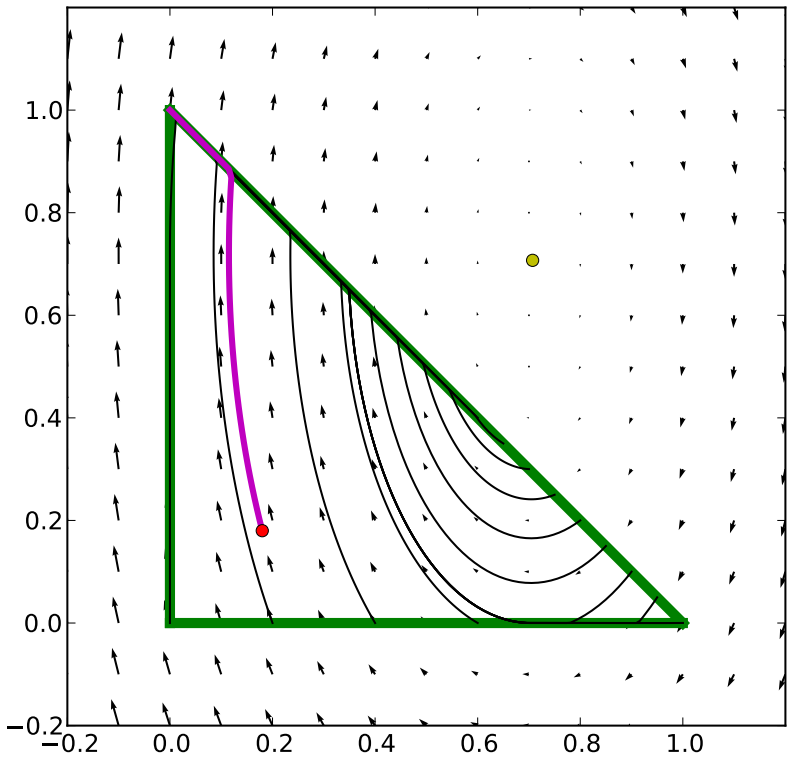


Figure 5.9 The phase portrait for the system defined by equation 5.15 is shown. The boundary of the feasibility region K is shown in green, several trajectories are shown and one trajectory with initial condition in the interior of K that converges to a fixed point of the projected system is shown in magenta, with initial condition shown in red. The fixed point for the unconstrained system (the yellow dot) is not asymptotically stable; it is a centre, whereas the fixed point for the projected dynamical system is asymptotically stable.

The main point of this example is to show that the equilibrium properties of the projected system in general can be very different to the unconstrained system.

In [64], conditions for the existence and uniqueness of an asymptotic steady state are given for projected dynamical systems which are equivalent to equilibrium states of variations of the STAP. The dynamics reflect the flow equilibration dynamics that can be thought of as a dynamical process through which the users swap to 'better' routes over time and while doing so converge to a *user equilibrium* pattern.

For a network with a fixed demand, the route equilibration process defined by

$$\dot{\mathbf{y}} = \Pi_{K_{\mathcal{N}}}(\mathbf{y}, -\mathbf{C}_{\mathcal{N}}(\mathbf{y})), \tag{5.18}$$

where Π_K is defined as above and $K_{\mathcal{N}}$ is the polytope defined by the demand constraints of the transport network $\mathcal{N}$. These constraints consist of the standard non-negativity of flows,

$$y_p \geq 0, \quad \forall p \in P_\omega, \tag{5.19}$$

for all routes in the network (*i.e.*, for all ω OD pairs) along with the conservation of flows for each OD pair,

$$\sum_{p \in P_\omega} y_p = d_\omega, \quad \forall \omega \in \Omega_{\mathcal{N}}. \tag{5.20}$$

The cost vector has been labelled $\mathbf{C}_{\mathcal{N}}$, to highlight that the route costs depend fundamentally on the network structure,

$$\mathbf{C}_{\mathcal{N}}(\mathbf{y}) = \mathbf{C}(R\,\mathbf{x}), \tag{5.21}$$

with R being the the route-link incidence matrix of $\mathcal{N}$.

Example 1: two parallel links (Pigou's example)

To place the projected dynamical systems back into the transport context we return to the famous example of Pigou [132], also widely used illustratively by Roughgarden [43] for its simplicity. Consider a network consisting of only two parallel links, as is shown in figure 5.10.

The number of routes is the same as the number of links. Mathematically, this means that the route-link incidence matrix reduces to the identity matrix in two dimensions (R, in equation 5.21, is the 2×2 identity matrix), so we have

$$y_i = x_i, \quad i = 1,2. \tag{5.22}$$

We will use cost functions

$$\begin{aligned} C_1(\mathbf{y}) &= 1, \\ C_2(\mathbf{y}) &= y_2, \end{aligned} \tag{5.23}$$

so that the first route has constant costs and the second route's cost is proportional to the route flow. For simplicity we are assuming that the free-flow cost vanishes for zero flow.

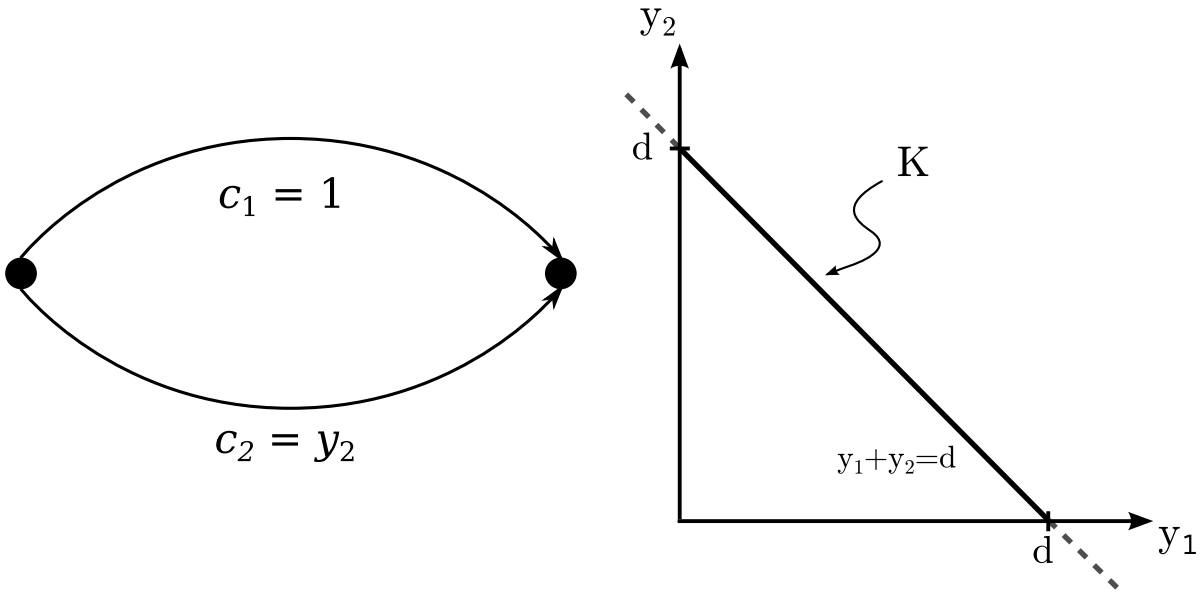


Figure 5.10 Pigou's example: (a) Parallel 2-link network. (b) The 2-simplex that makes up the constraint set K for this network is the the intersection of the line $y_1 + y_2 = d$ with the positive quadrant $\mathbb{R}^2_+$.

The corresponding projected equation for the Pigou network for a demand level $d = 1$ is given by

$$\dot{\mathbf{y}} = \Pi_K(\mathbf{y}, -\mathbf{C}(\mathbf{y})), \tag{5.24}$$

for $\mathbf{C}$ defined according to equations 5.23, and K is the line segment defined by the line $y_1 + y_2 = 1$ for $y_1, y_2 \geq 0$.

The phase portrait of the system along with the trajectory for an initial condition $\mathbf{y}_0$ converging to the user equilibrium flow $\mathbf{y}^{\text{UE}}$ is shown in figure 5.11 (a). At the equilibrium flow,

$$\mathbf{y} = \mathbf{y}^{\text{UE}} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{5.25}$$

the field

$$-\mathbf{C}\left(\mathbf{y}^{\text{UE}}\right) = -\begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{5.26}$$

is normal to the constraint set K and we have the asymptotic steady state of the system.

Given the expression of the projection from equation 5.8, the parameter γ is the magnitude of $\mathbf{C}$ itself and $\mathbf{n}^\star$ is the normal to $y_1 + y_2 = 1$,

$$\mathbf{n}^\star = -\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}. \tag{5.27}$$

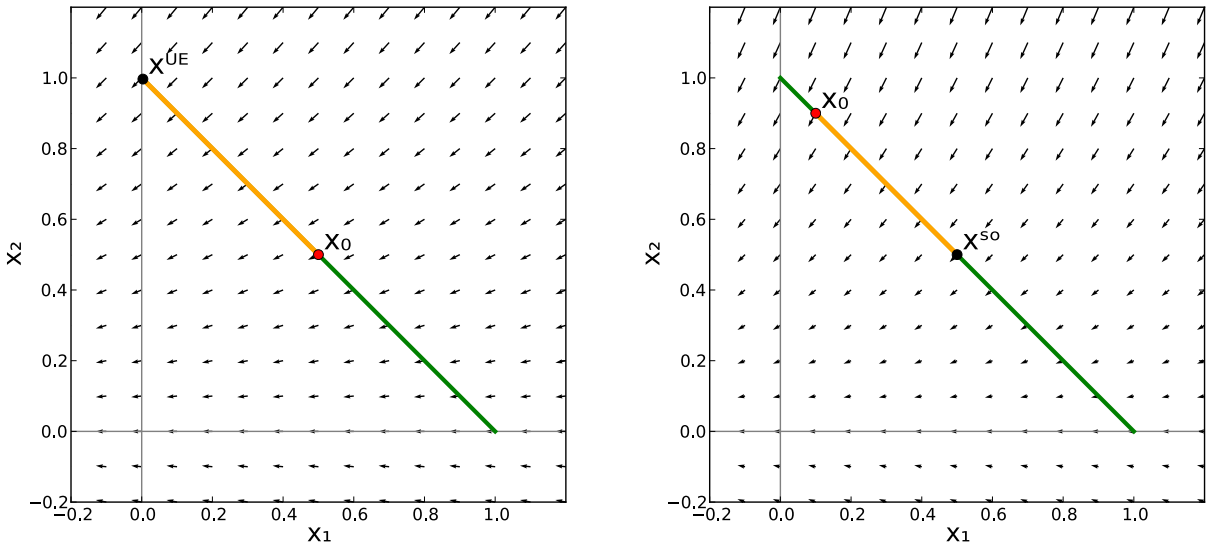


Figure 5.11 (a) Phase portrait for the two-link parallel (Pigou) network defined by equation 5.18 with cost functions 5.23. This dynamical system is equivalent to solving for UE. The initial condition $\mathbf{y}_0 = (1/2, 1/2)$ is shown in red, the trajectory towards $\mathbf{y}^{\mathrm{UE}}$ in orange. (b) Phase portrait for the two-link parallel network. The lines $x_1 = 0$ and $x_2 = 0$ are shown in grey. The convex set $K_{\mathcal{N}}$ to which the dynamical system is restricted is $x_1 + x_2 = 1$, given by route-flow conservation shown in green. The orange line shows the trajectory for initial condition $\mathbf{x}_0 = (0.1, 0.9)$ (shown in red) as it converges to the user equilibrium pattern $\mathbf{x}^{\mathrm{SO}} = (1/2, 1/2)$ (shown in black). At $\mathbf{x}^{\mathrm{SO}}$, the field is orthogonal to K and its projection vanishes.

An equivalent system can be defined that has the SO flow pattern as its asymptotic steady state. Thanks to the equivalence noted by Dafermos and Sparrow [133], the system with transformed cost functions has for its user equilibrium the system optimal of the original set-up. For the example that we are using, the transformed cost functions are,

$$\tilde{C}_1(\mathbf{y}) = 1 \tag{5.28}$$

$$\tilde{C}_2(\mathbf{y}) = 2x_2. \tag{5.29}$$

A phase space portrait for UE is shown in 5.11(a), with a trajectory for an initial condition showing the convergence towards

$$\mathbf{y}^{\mathrm{SO}} = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}. \tag{5.30}$$

The phase portrait of the system, along with the trajectory for an initial condition $\mathbf{y}_0$ converging to the system optimal flow $\mathbf{y}^{\mathrm{SO}}$, is shown in figure 5.11(b). At $\mathbf{y}^{\mathrm{SO}} = (1/2, 1/2)$ the projected field onto K vanishes which makes it a stationary point, since $\tilde{\mathbf{C}}(\mathbf{x}^{\mathrm{SO}}) = (1, 1)$, which is parallel to $\hat{\mathbf{n}}_{K_1} = \frac{1}{\sqrt{2}}(1, 1)$.

The dimensionality of the projected dynamical system corresponding to a network grows with the number of routes between its OD pairs. This makes the visualisation of the dynamical

system challenging, but it is possible to get some intuition from looking at a slightly more complex network than the parallel two-link network.

In general, the constraint set for the number of routes is a simplex of dimension $|P_\omega| - 1$ defined by the flow conservation equation

$$\sum_{p \in P_\omega} y_p = d_\omega, \tag{5.31}$$

along with the non-negativity requirement that $y_i \geq 0$, for $i = 1, \ldots, |P_\omega|$. To visualise this we turn to the Braess network.

Example 2: symmetric Braess network

The Braess network allows us to see how networks behave when link and flow variables are not the same. The virtue of the Braess network is that it has a small number of routes. It has two disjoint routes that interact only via a third route that shares a link with each of the them. This is a type of second order interaction since the route-flows of the disjoint routes do not appear in each other's cost functions. However, their interaction can result in the famous Braess paradox [55] The Braess network and the corresponding flow-space, the (two-dimensional) simplex as defined by equation 5.31, are displayed in figure 5.15.

The Braess network has five links and three routes, so it is an exception in terms of the numerosity of routes. For the edge and node numbering as is shown in figure 5.15, the Braess network has the adjacency matrix

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{5.32}$$

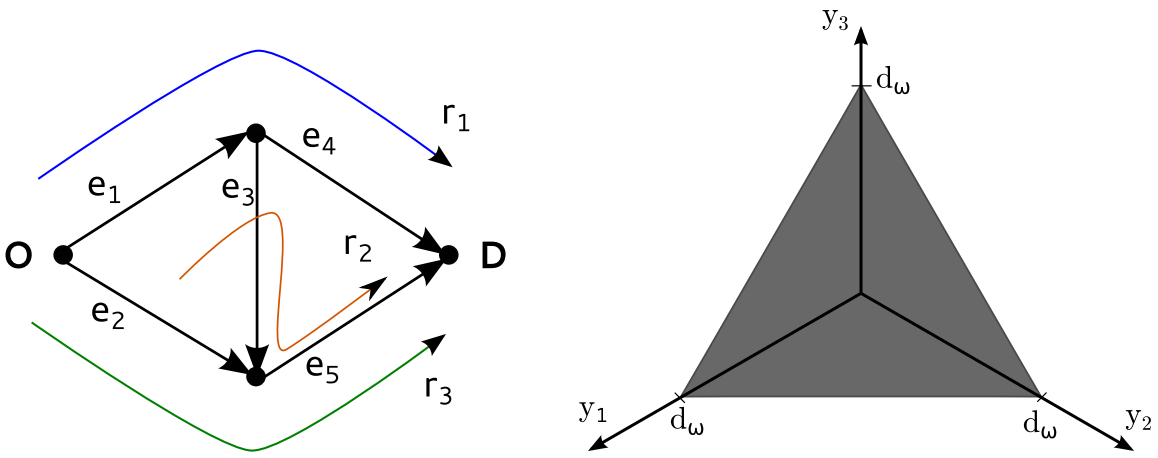


Figure 5.12 (a) The Braess network along with the (b) 2-simplex that is the constraint set K of the route flows, $K \subset \mathbb{R}^3$.

We label the routes in the following way,

$$r_1 = \{e_1, e_4\}, \tag{5.33}$$

$$r_2 = \{e_1, e_2, e_5\}, \tag{5.34}$$

$$r_3 = \{e_2, e_5\}, \tag{5.35}$$

which means that r_1 and r_3 are the outer routes and r_3 uses the cross-town link e_3.

We can express the composition of the link flows in terms of the route flows through the route-link incidence matrix R, which for our current example we have (from equations 5.33-5.35)

$$R = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix}. \tag{5.36}$$

The relationship between the route and the link variables can be expressed as

$$\mathbf{x} = R^\top \mathbf{y}, \tag{5.37}$$

which yields the following compositions of the link flows, which can be verified by inspection of the routes and figure 5.15 (a),

$$\begin{aligned} x_1 &= y_1 + y_2, \\ x_2 &= y_3, \\ x_3 &= y_2, \\ x_4 &= y_1, \\ x_5 &= y_2 + y_3. \end{aligned}$$

The corresponding flow conservation equations can be written in terms of the route variables by using expression 5.36,

$$S R^\top \mathbf{y} = \mathbf{d}.$$

This yields the constraints

$$\begin{aligned} -y_1 - y_2 - y_3 &= -d, \\ (y_1 - y_1) + (y_2 - y_2) &= 0, \\ (y_3 - y_3) + (y_2 - y_2) &= 0, \\ y_1 + y_2 + y_3 &= d, \end{aligned} \tag{5.38}$$

where the first and last expression reduce to the definition of the simplex and the middle two expressions are trivial.

We now write out the ODE of the projected dynamical system in the form

$$\dot{\mathbf{y}} = \Pi_{K_{\mathcal{N}}}(\mathbf{y}, -\mathbf{C}_{\mathcal{N}}(\mathbf{y})), \tag{5.39}$$

where $K_{\mathcal{N}}$ is the simplex defined by equation 5.38 and the cost vector is given by

$$\begin{aligned} \mathbf{C}_{\mathcal{N}}(\mathbf{y}) &= R\,\mathbf{c}(R^{\top}\mathbf{y}), \\ &= R\,\mathbf{a} + R\,D_{\mathbf{b}}R^{\top}\mathbf{y}, \end{aligned} \tag{5.40}$$

where R is the same as in equation 5.36. For the current example we use affine cost functions of the form $c_i(x_i) = a_i + b_i x_i$, with parameters

$$\begin{aligned} a_1 &= 1/2, \quad b_1 = 1, \\ a_2 &= 1, \quad b_2 = 1/2, \\ a_3 &= 1/10, \quad b_3 = 1/10, \\ a_4 &= 1/2, \quad b_4 = 1, \\ a_5 &= 1, \quad b_5 = 1/2. \end{aligned} \tag{5.41}$$

Figure 5.13 shows the phase portrait of the projected Braess system for a demand of $d = 1$. These parameters and demand value were selected to ensure that the Braess network is in a regime, where all three routes are used and where interesting behaviour from the assignment point of view, namely the Braess paradox exists (for a thorough discussion of the Braess networks refer to [56]).

Since there are three variables that are restricted to a plane triangle, we can transform the trajectories into barycentric coordinates. This effectively is like looking at the triangle resting in the corner of the positive orthant from an orthogonal perspective. The orbits of different initial conditions are shown converging to the UE equilibrium flow of

$$\mathbf{y}^{\mathrm{UE}} = \begin{pmatrix} 0.412 \\ 0.176 \\ 0.412 \end{pmatrix}. \tag{5.42}$$

The orange orbit corresponds to the initial condition $\mathbf{y}_0 = (1, 0, 0)$.

An example of the use of PDS to find equilibria at different demands is shown in figure 5.14, where the simplex has been scaled according to demand to trace the trajectory of the UE equilibrium across it. The PDS for the Braess network (as in figure 5.13) was integrated for demand values in the interval $(0, 2]$ and the stationary equilibria (UE) found.

Nagurney and Zhang [64] prove the equivalence of the steady state of this kind of dynamical system with the user equilibrium that is found by the convex optimisation problem of equation 3.11 (p46). After all, they both satisfy Wardrop's criterion. The point is that from a purely optimisation point of view, it is inconsequential how one gets to the equilibrium, unless it exploitable in the implementation of an algorithm. From a dynamical systems point of view, how one approaches equilibrium is perhaps the central focus. An aspect of this can be seen in the usefulness of Lyapunov functions for proving the existence of a Wardropian equilibria.

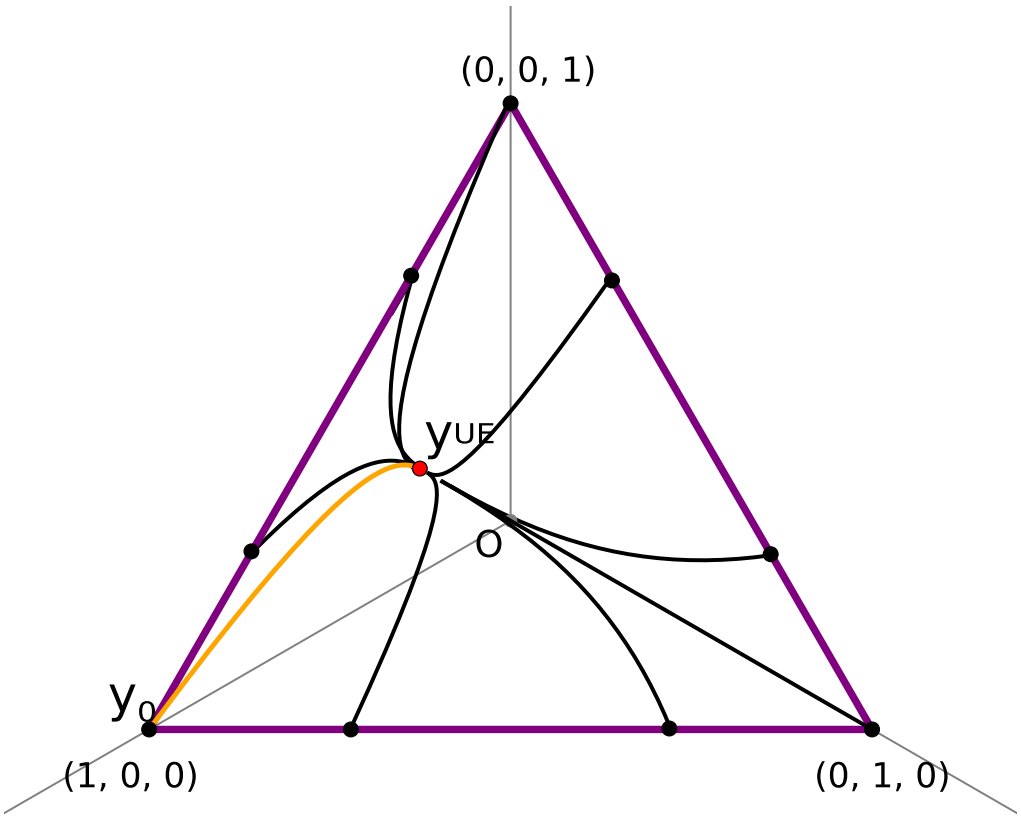


Figure 5.13 Trajectory converging to UE for the Braess network is shown. The y_i axes are shown in light grey. Cost function parameters are from equation 5.41 and demand, $d = 1$. Trajectories converge to the UE traffic pattern $\mathbf{y}^{\mathrm{UE}} = (0.412, 0.176, 0.412)$, additional orbits are drawn with their initial conditions marked with black dots. All initial conditions converge to $\mathbf{y}^{\mathrm{UE}}$, albeit slowly.

Example 3: lollipop network

A useful example for visualising the type of feasibility sets that arise when considering the STAP in link-flow variables, is the so-called lollipop network. The geometry of the flow space changes in the case of link variables. As opposed to route variables, the feasible region is no longer composed of the simplex obtained from slicing the positive orthant with the total demand plane. The lollipop network (shown on the left pane of figure 5.15) is a converse example to the Braess network. It has one more link than the number of routes, and its flow-space can be visualised in three dimensions. It consists of two parallel links in series with another edge.

The link flow-space defined by the flow conservation equation at each of the nodes is

$$\begin{aligned} x_1 &= d, \\ x_1 - (x_2 + x_3) &= 0, \\ x_2 + x_3 &= d. \end{aligned} \tag{5.43}$$

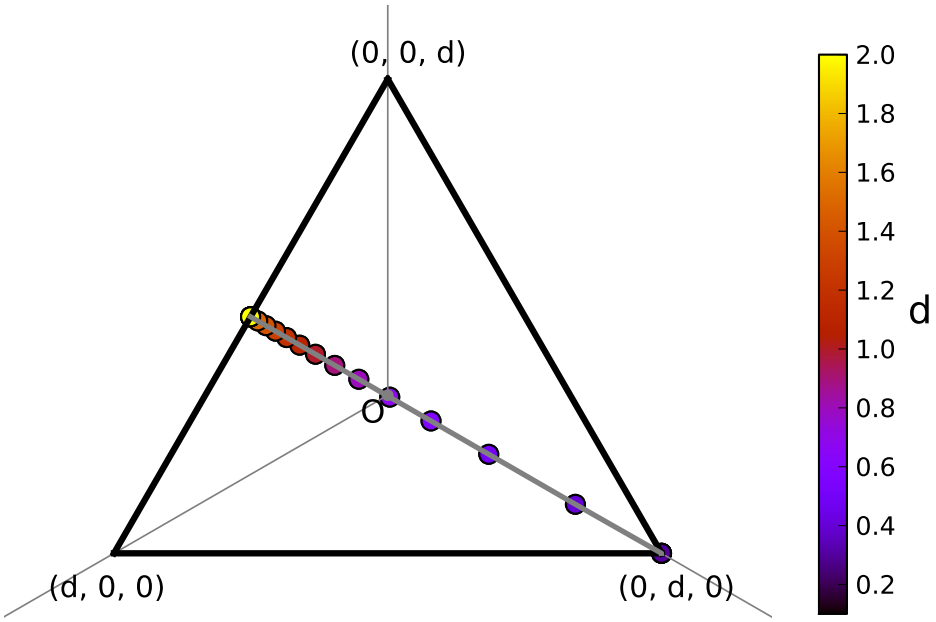


Figure 5.14 The user equilibrium for the Braess network is shown with the same set-up as in figure 5.13, but with increasing demand, $d \in (0,2]$. The path that the equilibrium point traces across the simplex as demand is increased is shown by the coloured dots, starting with only the cross-town route r_2 being used at $(0, d, 0)$, for low demands, and ending with the flow split evenly between the outer routes r_1 and r_3. Flows are normalised according to demand to make the simplices at all demand levels congruent. One can imagine the simplex as growing with demand from the origin, being displaced away from the vertex of the orthant, with the points marking the relative displacement of the equilibrium.

By virtue of being the only edge connected to the origin, e_1 is forced to carry all the demand, imposing the first flow conservation constraint from equations 5.43.

The geometry of these flow-spaces (whether in route or link variables), whilst important from the perspective of efficient algorithms, is not something widely visualised in the literature. By taking a constructivist tack, and showing the increasing complexity of the flow spaces for these small networks, we highlight that the feasibility sets for multiple ODs are simplicial complexes. Thus Nagurney's PDS take a simple form, since the convex sets in which the systems have an equilibrium have a lower dimension than the flow space they are embedded into.

An encompassed case

In [63], Smith takes a dynamical systems approach to prove that a simple system, following a dynamical route-flow equilibration mechanism, converges to a Wardropian equilibrium. We now discuss Smith's model in detail, so as to relate it to the PDS treatment described above. Smith takes Wardrop's criteria as a starting point, yet follows through with the behavioural explanation

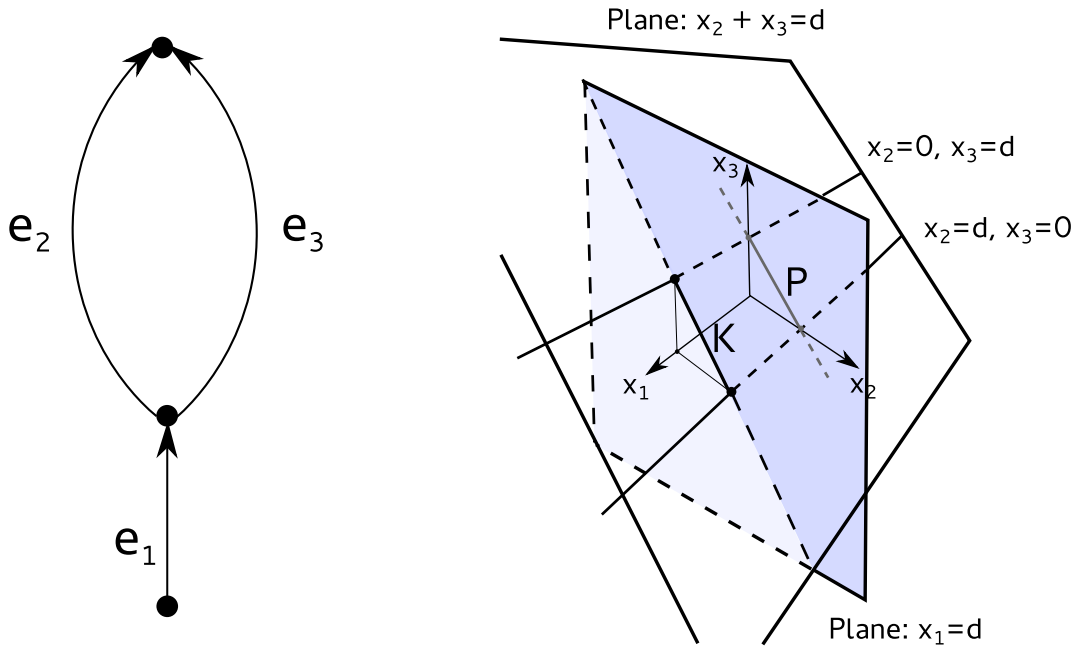


Figure 5.15 (a) The lollipop network consisting of two routes that share edge e_1. (b) The link-flow space of the lollipop network for a travel demand of d. The feasible set K (a straight-line segment) is defined by the flow conservation constraints equations 5.43, corresponding to each node; in this case, the intersection of two planes. The constraint planes are: $x_1 = d$ (the blue shaded plane) and $x_2 + x_3 = d$. This structure is the equivalent to the Pigou simplex in route variables. In fact, for affine constraints, the lollipop network is equivalent to Pigou's network if the travel-time of e_1 is added to the free-flow costs of e_2 and e_3.

of users learning the system by changing to better routes until the system equilibrates. In Smith's model, for any given two routes that connect the same OD pair, for example r and s, users will swap from route r to route s at a rate proportional to the difference in route costs, and will only swap to routes with lower costs. For notational convenience, Smith defines the operator $(\cdot)_+$ as

$$x_+ = \max\{0, x\}, \tag{5.44}$$

to keep track of positive and negative cost differences. This is used to determine the direction of swapping between routes with the help of vectors Δ_{rs}, which represent the flow swap directions from r to s in the form

$$(\mathbf{\Delta}_{rs})_i = \begin{cases} -1, & \text{if} \quad i = r, \\ 1, & \text{if} \quad i = s, \\ 0, & \text{otherwise.} \end{cases} \tag{5.45}$$

These $\mathbf{\Delta}_{rs}$ vectors point in the direction of changes to $\mathbf{y}$ that satisfy the demand conservation constraints. The flow equilibration mechanism is expressed as

$$\dot{\mathbf{y}} = \sum_{r,s \in P_\omega} y_r [C_r(\mathbf{y}) - C_s(\mathbf{y})]_+ \mathbf{\Delta}_{rs}; \tag{5.46}$$

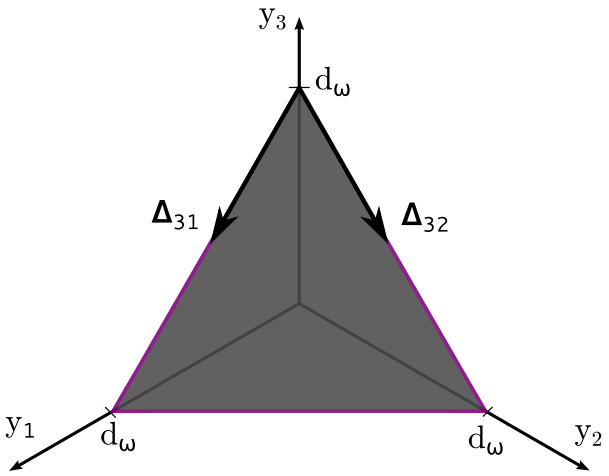


Figure 5.16 Visualisation of the route-flow simplex for a hypothetical network for OD pair ω, along with some of the swap vectors $\mathbf{\Delta}_{rs}$ which span the (extended) simplex.

Explicitly writing out the double sum yields

$$\dot{\mathbf{y}} = \sum_{r \in P_\omega} y_r \sum_{s \in P_\omega} [C_r(\mathbf{y}) - C_s(\mathbf{y})]_+ \mathbf{\Delta}_{rs}. \tag{5.47}$$

It is easier to see that the terms of the second sum, the $[C_r(\mathbf{y}) - C_s(\mathbf{y})]_+ \mathbf{\Delta}_{rs}$, are the projections of the truncated cost difference, $[C_r(\mathbf{y}) - C_s(\mathbf{y})]_+$, onto the vectors $\mathbf{\Delta}_{rs}$. These vectors span the affine space in which the simplex of route flows is contained. Each $\mathbf{\Delta}_{rs}$ lies on a hyperplane and is parallel to the line $y_s - y_r = \text{constant}$. The following anti-symmetric relation $\mathbf{\Delta}_{rs} = -\mathbf{\Delta}_{rs}$ holds, however in equations (5.46) and (5.47), only one of these terms will survive due to the truncation operation $(\cdot)_+$; $C_t(\mathbf{y}) > C_k(\mathbf{y})$ means that the term containing $\mathbf{\Delta}_{tk}$ vanishes.

Given the preceding observations, the inner sum in equation 5.47 is the projection onto the simplex; the product with the appropriate $\mathbf{\Delta}_{rs}$ is taken care of by the truncation of the cost difference $[C_r(\mathbf{y}) - C_s(\mathbf{y})]_+$, which simply projects the difference inside the brackets onto the positive octant $\mathbb{R}^3_+$.

Each row y_r in equation 5.47 is composed of the incoming flows due to swapping onto route r minus the ones swapping to other cost routes. The rate of swapping is also proportional to the flow on the higher cost route, which is where the y_r factor appears. The possible directions of flows shift are the $\mathbf{\Delta}_{rs}$ vectors, which lie on (and span) the simplex of route flows defined by the flow conservation constraints $\sum_{p \in P_\omega} y_p = d_\omega$, see figure 5.16. The clipping operation $(\cdot)_+$ acts as a valve only allowing flow swaps that improve (reduce) travel time.

The term $C_r(\mathbf{y}) - C_s(\mathbf{y})$ can be constructed from a graph Laplacian matrix where the graph represents the routes as nodes and routes that join the same OD pair are connected. In this way, routes that connect the same ω form cliques. For a single OD, the route graph is the $|P_\omega|$-complete graph and we can re-write equation 5.46 as

$$\dot{\mathbf{y}} = -L_R \, \mathbf{C}(\mathbf{y})_+ \sum_{r,s \in P_\omega} \mathbf{\Delta}_{rs} y_r. \tag{5.48}$$

where L_R is the Laplacian of the route graph.

This approach is coherent with Wardrop's principle of user equilibrium. It also brings us back conceptually to the territory of dynamical systems. Static traffic assignment assumes steady state traffic. The existence proof by Smith [63] conceptualises the equilibrium traffic pattern $\mathbf{x}$ as an asymptotic steady state of a dynamical process in which users swap to routes that are less costly than the ones they are on. Visualising an equilibrium traffic pattern as an attractive set of a dynamical system that evolves in flow-space helps with the intuitive understanding of different algorithms

The jump from a static view to a dynamical one can also be seen in [64], where a class of discontinuous dynamical systems are studied (inspired in part by the flow-swapping dynamics of Smith) whose stable stationary points coincide with the solution set of the corresponding optimisation problem with equilibrium constraints. In the following section, we take our cue from Smith and use the truncation operator $(\cdot)_+$ to connect the PDS of Nagurney *et al.* to *piecewise-smooth* dynamical systems.

5.4 A Filippov System

In this short section we bring together the observations on PDS and traffic networks above and propose that they can be cast as Filippov dynamical systems [134]. Filippov systems are a class of non-smooth dynamical systems for which the state space can be partitioned into sub-regions for which the equations of motion can be defined in a piecewise way. That is, for the sub-regions S_i into which the phase space of the system is split, we can write

$$\dot{\mathbf{x}} = \mathbf{F}_i(\mathbf{x}), \quad \text{for } \mathbf{x} \in S_i. \tag{5.49}$$

If we consider the Braess network in terms of route variables as above (figure 5.16), we can reduce the phase space to the plane that contains the feasible simplex. This plane can be split into four regions: the simplex, and the regions of the plane that lie outside the positive orthant. What we are trying to achieve is to recover the trajectories of the PDS without having to use the projection operator $K_{\mathcal{N}}$, and recover the sliding along the boundary from the combined effect of the $\mathbf{F}_i$ piecewise fields. That is, we want sliding modes (see chapter 2 of [135] by di Bernardo *et al.*) that are equivalent to the sliding portions of the PDS trajectories.

Since we do not want the trajectories to escape the simplex, the $\mathbf{F}_i$ that lie outside of the it simply need to cancel any component that is parallel to the outwards normal $\mathbf{n}^\star$ at the corresponding boundary of the simplex. At each point of the boundary, we can then define the field as

$$\mathbf{F}(\mathbf{x})|_{\partial K} = \mathbf{F}(\mathbf{x}) + \mathbf{n}^\star(\mathbf{n}^\star \cdot \mathbf{F}(\mathbf{x}))_+ . \tag{5.50}$$

Note that the terms are added due to the definition of the PDS with a negative sign (see equation 5.12).

The main reason why we propose taking this approach is that the method, of dividing the phase space into regions where the vector fields are individually defined, is more general and simpler than characterising the projection operator once the feasibility regions become more exotic: which happens when multiple OD pairs are involved as well as for larger networks with many more constraints than the lollipop network described above. Furthermore, piecewise-smooth dynamical systems are now a well-developed area [136] for which efficient numerical algorithms have been specially developed, for example [137].

5.5 Discussion

In this chapter we have looked at how to use the STAP and related methods to construct the NFD from the bottom up. Via computational experiments on network ensembles, we confirm that the key factor that determines the NFD relationship (for the uncongested branch) between trip production and density is the infrastructure supply. This is captured in our model, mainly through the total road-length of the network region under observation.

We have thus found that the bulk physical properties of the network play a more important role than structural features of the networks in the emergence of the NFD. As a first approximation, the morphological properties are important in the way they indirectly affect the total road length. This is seen in how the fit coefficients of the NFD depend on the clustering coefficient (captured by our β). We note that our method begins with networks that are structurally homogeneous, since that is the way the $\alpha\beta$-network ensembles have been constructed. Even with affine cost functions, the different ensembles show distinct behaviour. This is slightly surprising considering the nuanced qualitative change in morphologies presented by the different ensembles.

A detailed understanding of the effects of morphology on the NFD is crucial to understand the routing properties of individual cities. This is especially true in a context where the NFD is to be leveraged as tool for traffic control strategies. However, unpicking these macroscopic relationships is challenging due to the complexity of defining appropriate experimental and theoretical frameworks for investigation. We have seen that the $\alpha\beta$-network family lends itself well to modelling networks with homogeneous structure.

In terms of the traffic model, the STAP is conceptually rich and can be used directly for capturing uncongested traffic states of networks. For the uncongested states, the simplified STAP, as used in this thesis, proves to be more problematic, and in fact inappropriate. A modification that would be appropriate to the STAP that could lead to a naturally emerging MFD would be to modify the model to incorporate exit-flow capacity constraints for the network links and queue storage constraints at nodes. Static and quasi-static models such as these have proven useful in obtaining more realistic flow estimates in road networks [138] than the standard STAP, and have been shown to be easily extended to incorporate sophisticated modelling techniques as well [139]. For our purposes, including queuing (even without spillback) and capacities, along

with our ensemble approach would provide access to congested network states. The queues would contribute to the link occupation, and reduce trip production rates, allowing us to 'observe' a congested state of a network. Obtaining equilibrium solutions with queues for different networks in an ensemble, would provide a data point for each network (as in figure 5.4) might allow for an ensemble NFD to emerge.

Additionally, we have begun to show how the Wardropian PDS might be used to build the congested branch of the NFD from numerical simulations. For different initial conditions, the transient trajectories, as the orbits converge to equilibrium, would provide the congested (in STAP parlance: sub-optimal) traffic states of the network. Since such dynamical systems are deterministic, this approach would also allow additional investigation into the hysteresis effects observed in simulations as well as from observational data.

In view of the above, the key contributions from this chapter are summarised as follows.

C5.1 The dominating factor in determining the uncongested branch of the NFD is the infrastructure supply (region length).

C5.2 Quantitatively, coefficients of fit of the square-root NFDs (for a fixed griddedness) depend linearly on β, that is, triangles are of key importance.

C5.3 The (quasi-dynamic) PDS that converges to Wardropian equilibria is equivalent to a Filippov system. By treating the STAP in such a way, sub-optimal assignments and their orbits in flow-space can potentially be used to 'observe' different network traffic states.

In terms of next steps in this work, viewing traffic assignment as a Filippov system (a type of non-smooth dynamical system), opens up several new lines of research. A particularly interesting idea is the potential understanding of types of chaos (such as chaotic attractors) that can present themselves in non-smooth systems even for a small number of variables. The identification of traffic models and dynamical systems carried out in the second part of this chapter (from section 5.3 onwards) naturally leads to further associations. For example, the sliding along the boundaries, due to the projection, can be replaced with collision maps (such as is done in dynamical billiards, [140]). This could potentially lead to insights into unintended effects of traffic control. For example, at the boundary of a feasible region, where the flow on some links is zero, replacing the sliding mechanism with an instant collision (also a non-smooth map) could be interpreted as traffic being forcefully redirected, a common situation when traffic accidents occur. The theoretical motivation behind this would be the chaotic behaviour exhibited by inelastic billiards, possibly leading to a better understanding of the nature of chaos in traffic systems.

The emergence of NFDs for the synthetic $\alpha\beta$-networks highlights an important question regarding what can be learned from idealised experiments, and how applicable to real-world networks findings from them actually are. For the most part, the importance of the results, in our case, are qualitative in nature. One of the main findings is that for the congested branch of the NFD (see fig 5.5) the fit parameters of the squared-root NFDs, even though strongly dominated

by the region length, they still depend on the morphological parameters; the griddedness and the road density are important. Quantitatively, however, our model is still fairly removed from real-world networks. Thus, the specific functional dependence cannot be directly applied to real urban networks. Our abstracted model, though, allows us to explore the features that might be important to keep in mind when understanding real-world NFDs and partitioned networks. Future studies can be guided by observations from bottom-up reconstructions such as the one presented in this chapter, and thus hopefully allow the conceptual simplification of a fully dynamic theory of MFD emergence.

Balancing abstraction and applicability in mathematical modelling is a delicate matter, and in the search for generality, realism can be lost. In our aim to keep the equilibrium models as simple as possible, fine details, such as the functional forms obtained for the NFDs are not directly applicable or comparable to existing models or networks. However, some of the relationships between morphology and routing performance are able to be discovered.

CHAPTER 6

CONCLUSIONS

In this thesis we have set out to investigate traffic equilibria through a complexity science lens. From this viewpoint, we have designed experiments that investigate how simple a network traffic model can be, yet preserve explicative power. We have revisited the classic static traffic assignment (STAP) and used it in conjunction with a random-network model designed for parsimony, that captures stylised features of road networks. Our goal has been to draw general conclusions about how the structure of road networks effects performance metrics (such as price of anarchy, and total system cost) and network-aggregated traffic variables (like the network fundamental diagram).

We have developed the $\alpha\beta$-network model (chapter 2), and an implicit heuristic for defining cost-function parameters, for use with the STAP (chapter 3). Considerable attention was devoted to setting up the networks for numerical experiments. For example, periodic boundary conditions were exploited to enable small numbers of origin-destination (OD) pairs to yield traffic patterns similar to those due to much more complex set-ups. A key finding is that network statistics, such as betweenness centrality, even though often used as proxies for traffic flow, do not correlate with network performance in a meaningful way. Then, in chapter 4, we applied the $\alpha\beta$-networks to investigate how networks with different structures might benefit from replacing a proportion of their vehicle fleet with altruistically-routed (autonomous) vehicles. We have shown that the transition to optimal costs depends on network size and complexity. Surprisingly, larger and more complex networks enhance the effectiveness of the intervention, which goes against the commonly-held belief that grid networks are inherently efficient. In chapter 5 we applied the $\alpha\beta$-network model to study the emergence of a network fundamental diagram (NFD). We used the STAP to understand the relationships between network-wide aggregated traffic statistics and their dependence on network morphology. By analysing projected dynamical systems (PDS) [64],

the structure of the flow-space induced by the STAP, and a simple dynamical system from Smith [63] that converges to a Wardropian equilibrium, we proposed that a method to recover the congested branch of the NFD is to cast a flow equilibration system as a piecewise-smooth Filippov dynamical system. Filippov systems are naturally equipped to handle hard boundaries in their phase space, and also exhibit rich bifurcation structures that do not occur in classical dynamical systems.

We now re-state the research questions that were framed at the beginning of the thesis and examine how and to what extent we have addressed them. The questions are:

RQ1 How simple can a network model be, and still remain useful in understanding real-world transport properties?

RQ2 What insights can a highly-stylised routing model, such as the static traffic assignment problem, provide into the role played by network morphology in the traffic patterns (and their efficiency) that arise on these networks?

RQ3 Given that autonomous vehicles can provide a mechanism for achieving optimal network performance, is the transition to optimality provided by them, across networks with different morphologies, a good one?

RQ4 In the context of understanding global (network-wide) behaviour of networks, what 'minimal' modifications can be made to the *static traffic assignment problem* to build a network fundamental diagram from the bottom up?

RQ1. Drawing from chapter 3 and the existing transportation research literature, it is clear that the STAP is a highly idealised model that does not capture the full richness of real-world traffic flow patterns. However, we have shown that simple network models allow us to explore the factors — such as network structure, and OD choice — that affect the efficiency of traffic equilibria. Indeed, a very simple choice of OD pairs on the $\alpha\beta$-network model has allowed us to distinguish performance features, such as the PoA, a measure that is known to behave in very similar ways for networks with vastly different structures. As such, while we do not explore features of real-world networks, we have shown that our $\alpha\beta$-networks are indeed a useful tool for deepening the understanding of Wardropian UE performance.

There are two points that benefit from further discussion, the first is to what extent does this model allow insights to be transferred to more realistic settings, and the second is how would this potentially be evaluated. To address the first point, the ideal solution would be to compare the behaviour of different networks models, and then these models with real city networks. One potential example would be networks generated according to some growth model that resemble urban growth, for example the model from [35] where networks are grown according to a resource optimisation principle and result in networks with features shared by urban networks. Other random network models, such as the one in [141] are fitted to resemble specific road networks

and could be a step in bridging the gap towards more realism. However, in order to verify how our findings can translate directly to real-world networks, while still using morphologically homogeneous networks, an avenue forward could be to use sub-networks as they are partitioned in MFD studies in the literature. One possible example is the Lyon sub-networks from [9] discussed in chapter 5. Where the different reservoirs have their particular structure. The challenge, however, becomes the parameters choice for measuring morphological characteristics, and the diversity of networks. In terms of parameter choice, based on the findings of this thesis, road density, node distribution statistics and the clustering coefficient of the networks are proposed. In terms of diversifying test samples to use real-world networks within an ensemble framework, a possibility is to make ensembles according to OD structure in order to sample network structure in different combinations.

RQ2. Consider our achievements in chapter 3. We have demonstrated the importance of ensuring that the cost-function parametrisation has a measurable impact on the ensemble-wide performance of transport networks. We have shown, as captured by contributions C2.2 and C3.3 (page 68), that the griddedness of networks plays an important role in determining their sensitivity to OD structure, and that more complex networks shift the onset of peak PoA to higher demand values.

RQ3 can be answered directly. From chapter 4, we have found that altruistically-routed AVs can indeed provide system-wide benefits in terms of reductions in total system cost. However the benefits might only be realised when a high proportion of vehicles are altruistically routed. Also, our analysis has shown that the costs incurred by the AVs, in ameliorating the costs for the other users, would make this mode of transport highly undesirable at first, suggesting that the pathway to the adoption of AVs is very challenging.

RQ4. In chapter 5 we have shown that with an ensemble approach, the STAP is suitable as-is for explaining the uncongested branch of the NFD. Furthermore, we conjectured that a dynamical approach like [63], cast as a Filippov system, is suitable for accessing sub-optimal, and thus more congested network states. Viewed from the perspective of Nagurney's PDS, the static optimisation and dynamical approaches are closer than they might seem. The orbits of such dynamical systems, before they converge to equilibrium, can be thought of as congested states of the networks, which aggregated with our ensembling approach might yield the congested branch of the NFD: however, the details remain to be addressed. Filippov systems also exhibit different types of bifurcations, and thus, this approach could potentially lead to further understanding of the complex behaviour, and perhaps routes to chaos, of transportation systems.

In conclusion, this thesis shows that simple models, both for networks and for traffic, are still relevant in deepening the understanding of transportation issues. We have validated the world-view of complexity science, that simple interactions between 'agents' are fundamental in pattern formation. Furthermore, planar networks, often ignored because of the restrictions on their degree distributions, can give rise to complex pattern formation just as much as more

common models from network science. As we go forward, we suggest that the transportation research field would benefit from drawing more explicitly from complexity science than even we have done.

References

[1] G. Toth.
Reducing growth in vehicle miles traveled: Can we really pull it off?
In D. Sperling and J. S. Cannon, editors, *Driving Climate Change*, pages 129–142. Academic Press, Burlington, 2007.

[2] H. Youn, M. T. Gastner, and H. Jeong.
Price of Anarchy in Transportation Networks: Efficiency and Optimality Control.
Physical Review Letters, 101(12):128701, 2008.

[3] Mahyar Amirgholy and H. O. Gao.
Modeling the dynamics of congestion in large urban networks using the macroscopic fundamental diagram: User equilibrium, system optimum, and pricing strategies.
Transportation Research Part B: Methodological, 104:215 – 237, 2017.

[4] N. Geroliminis and C. F. Daganzo.
Macroscopic modeling of traffic in cities.
In *Transportation Research Board 86th Annual Meeting*, number 07-0413. No. 07-0413, 2007.

[5] N. Geroliminis and C. F. Daganzo.
Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings.
Transportation Research Part B: Methodological, 42(9):759–770, 2008.

[6] E. Gonzales, C. Chavis, Y. Li, and C. F. Daganzo.
Multimodal transport in Nairobi, Kenya: Insights and recommendations with a macroscopic evidence-based model.
In *TRB 90th Annual Meeting Compendium of Papers DVD*, 2011.

[7] K. Aboudolas and N. Geroliminis.
Perimeter and boundary flow control in multi-reservoir heterogeneous networks.
Transportation Research Part B: Methodological, 55:265 – 281, 2013.

[8] T. Osaragi and Y. Hiraga.
Street network created by proximity graphs: its topological structure and travel efficiency.
In *Proceedings of the AGILE'2014 International Conference on Geographic Information Science, Castellón, June, 3-6, 2014.* AGILE Digital Editions, 2014.

[9] G. Mariotte, L. Leclercq, S. F. A. Batista, J. Krug, and M. Paipuri.
Calibration and validation of multi-reservoir MFD models: a case study in Lyon.
Transportation Research Part B: Methodological, 136:62–86, 2020.

[10] N. Geroliminis and J. Sun.
Hysteresis phenomena of a macroscopic fundamental diagram in freeway networks.
Transportation Research Part A: Policy and Practice, 45(9):966–979, 2011.

[11] United Nations Publications.
World Urbanization Prospects: The 2018 Revision.
United Nations, 2019.

[12] Secretary of State for Business, Energy and Industrial Strategy.
Industrial Strategy. Building a Britain fit for the future.
Her Majesty's Stationery Office, 2017.

[13] Government Office for Science.
A time of unprecedented change in the transport system.
Foresight, 2019.

[14] Instituto Mexicano para la Competitividad.
El costo de la congestión: vida y recursos perdidos [The cost of congestion: life and resources lost].
`https://imco.org.mx/costo-la-congestion-vida-recursos-perdidos/`.
Accessed: 08-07-2020.

[15] The European Commission.
Developments and Forecasts on Continuing Urbanisation.
`https://ec.europa.eu/knowledge4policy/foresight/topic/continuing-urbanisation/developments-and-forecasts-on-continuing-urbanisation_en`.
Accessed: 08-07-2020.

[16] European Environment Agency.
Stronger measures needed to tackle harm from air pollution.
`https://www.eea.europa.eu/highlights/stronger-measures-needed`.
Accessed: 08-07-2020.

[17] J. Ladyman, J. Lambert, and K. Wiesner.
What is a complex system?
European Journal for Philosophy of Science, 3(1):33–67, 2013.

[18] R. Olfati-Saber.
Flocking for multi-agent dynamic systems: algorithms and theory.
IEEE Transactions on automatic control, 51(3):401–420, 2006.

[19] I. D. Couzin, C. C. Ioannou, G. Demirel, T. Gross, C. J. Torney, A. Hartnett, L. Conradt, S. A. Levin, and N. E. Leonard.
Uninformed individuals promote democratic consensus in animal groups.
Science, 334(6062):1578–1580, 2011.

[20] M. M. Waldrop.
Complexity: The Emerging Science at the Edge of Order and Chaos.
Simon and Schuster, 1993.

[21] R. R. Llinás.
The contribution of Santiago Ramon y Cajal to functional neuroscience.
Nature Reviews Neuroscience, 4(1):77–80, 2003.

[22] T. Trappenberg.
Fundamentals of Computational Neuroscience.
OUP Oxford, 2009.

[23] B. Davis and E. Simmt.
Understanding learning systems: Mathematics education and complexity science.
Journal for Research in Mathematics Education, pages 137–167, 2003.

[24] H. Ledford.
How to solve the world's biggest problems.
Nature News, 525(7569):308, 2015.

[25] L. M. A. Bettencourt, J. Lobo, D. Strumsky, and G. B. West.
Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities.
PloS One, 5(11):e13541, 2010.

[26] M. Batty.
Cities as Complex Systems: Scaling, Interaction, Networks, Dynamics and Urban Morphologies.
Springer, 2009.

[27] E. Arcaute, E. Hatna, P. Ferguson, H. Youn, A. Johansson, and M. Batty.
Constructing cities, deconstructing scaling laws.
Journal of the Royal Society Interface, 12(102):20140745, 2015.

[28] G. B. West and J. H. Brown.
Life's universal scaling laws.
Physics Today, 57(9):36–43, 2004.

[29] H. Samaniego and M. E. Moses.
Cities as organisms: Allometric scaling of urban road networks.
Journal of Transport and Land Use, 1(1):21–39, 2008.

[30] C. Kühnert, D. Helbing, and G. B. West.
Scaling laws in urban supply networks.
Physica A: Statistical Mechanics and its Applications, 363(1):96 – 103, 2006.

[31] S. Lämmer, B. Gehlsen, and D. Helbing.
Scaling laws in the spatial structure of urban road networks.
Physica A: Statistical Mechanics and its Applications, 363(1):89–95, 2006.

[32] M. Barthélemy and A. Flammini.
Optimal traffic networks.
Journal of Statistical Mechanics: Theory and Experiment, 2006(07):L07002, 2006.

[33] G. Boeing.
Urban spatial order: Street network orientation, configuration, and entropy.
Applied Network Science, 4(1):67, 2019.

[34] E. Strano, V. Nicosia, V. Latora, S. Porta, and M. Barthélemy.
Elementary processes governing the evolution of road networks.
Scientific Reports, 2:296, 2012.

[35] M. Barthélemy and A. Flammini.
Modeling urban street patterns.
Physical Review Letters, 100(13):138702, 2008.

[36] M. Batty.
The New Science of Cities.
MIT press, 2013.

[37] T. J. Lomax and D. L. Schrank.
The 2005 Urban Mobility Report.
Texas Transportation Institute, Arlington, TX, 2005.

[38] J. G. Wardrop.
Road paper. some theoretical aspects of road traffic research.
Proceedings of the Institution of Civil Engineers, 1(3):325–362, 1952.

[39] A. Charnes and W. W. Cooper.
Extremal principles for simulating traffic flow in a network.
Proceedings of the National Academy of Sciences of the United States of America, 44(2):201, 1958.

[40] E. Koutsoupias and C. Papadimitriou.
Worst-case equilibria.
In *Annual Symposium on Theoretical Aspects of Computer Science*, pages 404–413. Springer, 1999.

[41] C. Papadimitriou.

Algorithms, games, and the internet.
In *Proceedings of the Thirty-Third Annual ACM Symposium on Theory of Computing*, pages 749–753, 2001.

[42] T. Roughgarden and É. Tardos.
How bad is selfish routing?
Journal of the ACM, 49(2):236–259, 2002.

[43] T. Roughgarden.
The price of anarchy is independent of the network topology.
Journal of Computer and System Sciences, 67(2):341–364, 2003.

[44] PhD thesis.

[45] A. Rose, R. D. O'Dea, and K. I. Hopcraft.
Price of anarchy on heterogeneous traffic-flow networks.
Physical Review E, 94(3):032315, 2016.

[46] S. J. O'Hare, R. D. Connors, and D. P. Watling.
Mechanisms that govern how the price of anarchy varies with travel demand.
Transportation Research Part B: Methodological, 84:55–80, 2016.

[47] O. Smith, J. Crowe, R. D. O'Dea, and K. I. Hopcraft.
The price of anarchy in flow networks as a function of node properties.
Europhysics Letters, 127(1):18001, 2019.

[48] C. F. Daganzo and N. Geroliminis.
An analytical approximation for the macroscopic fundamental diagram of urban traffic.
Transportation Research Part B: Methodological, 42(9):771 – 781, 2008.

[49] Y. Ji and N. Geroliminis.
On the spatial partitioning of urban transportation networks.
Transportation Research Part B: Methodological, 46(10):1639–1656, 2012.

[50] J. Ortigosa, 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, 3(3-4):245–266, 2014.

[51] M. Ramezani, J. Haddad, and N. Geroliminis.
Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control.
Transportation Research Part B: Methodological, 74:1–19, 2015.

[52] A. Kirkley, H. Barbosa, M. Barthélemy, and G. Ghoshal.
From the betweenness centrality in street networks to structural invariants in random planar graphs.

Nature Communications, 9(1):1–12, 2018.

[53] Y. Sheffi.
Urban Transportation Networks.
Prentice-Hall, Englewood Cliffs, NJ, 1985.

[54] A. P. Masucci, D Smith, A Crooks, and M. Batty.
Random planar graphs and the London street network.
The European Physical Journal B, 71(2):259–271, 2009.

[55] D. Braess.
Über ein paradoxon aus der verkehrsplanung.
Unternehmensforschung, 12(1):258–268, 1968.

[56] E. I. Pas and S. L. Principio.
Braess' paradox: Some new insights.
Transportation Research Part B: Methodological, 31(3):265–276, 1997.

[57] R. Steinberg and W. I. Zangwill.
The prevalence of Braess' paradox.
Transportation Science, 17(3):301–318, 1983.

[58] I. Milchtaich.
Network topology and the efficiency of equilibrium.
Games and Economic Behavior, 57(2):321–346, 2006.

[59] M. Patriksson.
The Traffic Assignment Problem: Models and Methods.
Courier Dover Publications, 2015.

[60] D. G. Kirkpatrick and J. D. Radke.
A framework for computational morphology.
In *Machine Intelligence and Pattern Recognition*, pages 217–248. Elsevier, 1985.

[61] T. Van Vuren and D. P. Watling.
A multiple user class assignment model for route guidance.
Transportation Research Record, pages 22–22, 1991.

[62] D. P. Watling and T. Van Vuren.
The modelling of dynamic route guidance systems.
Transportation Research Part C: Emerging Technologies, 1(2):159–182, 1993.

[63] M. J. Smith.
The stability of a dynamic model of traffic assignment — an application of a method of Lyapunov.
Transportation Science, 18(3):245–252, 1984.

[64] A. Nagurney and D. Zhang.
Projected Dynamical Systems and Variational Inequalities with Applications.
Springer, New York, 1996.

[65] A. F. Filippov.
Differential Equations with Discontinuous Righthand Sides: Control Systems.
Springer Science & Business Media, 2013.

[66] A. Espinosa Mireles de Villafranca, R. D. Connors, and R. E. Wilson.
Static traffic assignment on ensembles of synthetic road networks.
In *International Conference on Traffic and Granular Flow*, pages 29–36. Springer, 2017.

[67] M. Rinaldi and F. Viti.
Investigating the relationship between controller locations and dynamic traffic control in generic transportation networks.
Transportation Research Record, 2674(5):172–182, 2020.

[68] D. J. Watts and S. H. Strogatz.
Collective dynamics of 'small-world' networks.
Nature, 393(6684):440–442, 1998.

[69] M. Newman.
Networks: An Introduction.
Oxford University Press, 2010.

[70] P. Crucitti, V. Latora, and S. Porta.
Centrality measures in spatial networks of urban streets.
Physical Review E, 73(3):036125, 2006.

[71] R. Albert and A.-L. Barabási.
Statistical mechanics of complex networks.
Reviews of Modern Physics, 74(1):47, 2002.

[72] G. Fu, S. Wilkinson, and R. J. Dawson.
A spatial network model for civil infrastructure system development.
Computer-Aided Civil and Infrastructure Engineering, 31(9):661–680, 2016.

[73] M. Barthélemy.
Spatial networks.
499(1-3):1–101, 2011.

[74] A. Tero, S. Takagi, T. Saigusa, K. Ito, D. P. Bebber, M. D. Fricker, K. Yumiki, R. Kobayashi, and T. Nakagaki.
Rules for biologically inspired adaptive network design.
Science, 327(5964):439–442, 2010.

[75] M.-A. I Tsompanas, G. C. Sirakoulis, and A. I. Adamatzky.

Evolving transport networks with cellular automata models inspired by slime mould.
IEEE transactions on cybernetics, 45(9):1887–1899, 2014.

[76] A. Rapoport, V. Mak, and R. Zwick.
Navigating congested networks with variable demand: experimental evidence.
Journal of Economic Psychology, 27(5):648–666, 2006.

[77] D. Watanabe.
Evaluating the configuration and the travel efficiency on proximity graphs as transportation networks.
Forma, 23(2):81–87, 2008.

[78] P. Erdös and A. Rényi.
On random graphs.
Publicationes Mahematicae, 6(26):290–297, 1959.

[79] C. Andersson, K. Frenken, and A. Hellervik.
A complex network approach to urban growth.
Environment and Planning A, 38(10):1941–1964, 2006.

[80] D. P. Ward, A. T. Murray, and S. R. Phinn.
A stochastically constrained cellular model of urban growth.
Computers, Environment and Urban Systems, 24(6):539–558, 2000.

[81] R. J. Trudeau.
Introduction to Graph Theory.
Dover, 1993.

[82] F. Harary.
Graph Theory.
Addison-Wesley series in mathematics. Addison-Wesley Pub. Co., 1969.

[83] F. Fang, J. Forero-Romero, G. Rossi, X.-D. Li, and L.-L. Feng.
β-Skeleton analysis of the cosmic web.
Monthly Notices of the Royal Astronomical Society, 485(4):5276–5284, 03 2019.

[84] J. W. Jaromczyk and G. T. Toussaint.
Relative neighborhood graphs and their relatives.
Proceedings of the IEEE, 80(9):1502–1517, 1992.

[85] J. Cardinal, S. Collette, and S. Langerman.
Empty region graphs.
Computational Geometry, 42(3):183–195, 2009.

[86] M. Kowaluk and G. Majewska.
β-skeletons for a set of line egments in r^2.

In *International Symposium on Fundamentals of Computation Theory*, pages 65–78. Springer, 2015.

[87] D. Maniadakis and D. Varoutas.
Fitting planar proximity graphs on real street networks.
In *Proceedings of ECCS 2014*, pages 11–20. Springer, 2016.

[88] A. Cardillo, S. Scellato, V. Latora, and S. Porta.
Structural properties of planar graphs of urban street patterns.
Physical Review E, 73(6):066107, jun 2006.

[89] E. Strano, M. Viana, L. da Fontoura Costa, A. Cardillo, S. Porta, and V. Latora.
Urban street networks, a comparative analysis of ten European cities.
Environment and Planning B: Planning and Design, 40(6):1071–1086, 2013.

[90] J. Peponis, D. Allen, D. Haynie, M. Scoppa, and Z. Zhang.
Measuring the configuration of street networks: the spatial profiles of 118 urban areas in the 12 most populated metropolitan regions in the us.
In *6th International Space Syntax Symposium Proceedings*, 2007.

[91] S. H. Y. Chan, R. V. Donner, and S. Lämmer.
Urban road networks — spatial networks with universal geometric features?
The European Physical Journal B, 84(4):563–577, 2011.

[92] D. B. West.
Introduction to Graph Theory, volume 2.
Prentice hall, 1996.

[93] A. Turner.
From axial to road-centre lines: a new representation for space syntax and a new model of route choice for transport network analysis.
Environment and Planning B: Planning and Design, 34(3):539–555, 2007.

[94] K. Stephenson and M. Zelen.
Rethinking centrality: Methods and examples.
Social Networks, 11(1):1–37, 1989.

[95] J. G. Andrews, R. K. Ganti, M. Haenggi, N. Jindal, and S. Weber.
A primer on spatial modeling and analysis in wireless networks.
IEEE Communications Magazine, 48(11):156–163, 2010.

[96] M. Penrose.
Random Geometric Graphs.
Oxford University Press, 2003.

[97] D. J. Aldous and J. Shun.
Connected spatial networks over random points and a route-length statistic.

Statistical Science, 25(3):275–288, 2010.

[98] C. F. Daganzo and Y. Sheffi.
On stochastic models of traffic assignment.
Transportation Science, 11(3):253–274, 1977.

[99] T. Roughgarden and É. Tardos.
Bounding the inefficiency of equilibria in nonatomic congestion games.
Games and Economic Behavior, 47(2):389 – 403, 2004.

[100] R. W. Rosenthal.
A class of games possessing pure-strategy Nash equilibria.
International Journal of Game Theory, 2(1):65–67, 1973.

[101] D. Monderer and L. S. Shapley.
Potential games.
Games and Economic Behavior, 14(1):124 – 143, 1996.

[102] M. J. Beckmann, C. B. McGuire, and C. B. Winsten.
Studies in the Economics of Transportation.
RAND Corporation, 1955.

[103] LLC Gurobi Optimization.
Gurobi Optimizer Reference Manual.
`http://www.gurobi.com`, 2020.

[104] S. Boyd and L. Vandenberghe.
Convex Optimization.
Cambridge University Press, 2004.

[105] M. J. Smith.
The existence of an equilibrium solution to the traffic assignment problem when there are junction interactions.
Transportation Research Part B: Methodological, 15(6):443–451, 1981.

[106] L. Alonso, J. A. Méndez-Bermúdez, and E. Estrada.
Geometrical and spectral study of β-skeleton graphs.
Physical Review E, 100(6):062309, 2019.

[107] T. van Vuren, D. van Vliet, and M. J. Smith.
Combined equilibrium in a network with partial route guidance.
In *Traffic control methods. Proceedings of the Fifth Engineering Foundation Conference*, 1990.

[108] H. Yang, X. Zhang, and Q. Meng.
Stackelberg games and multiple equilibrium behaviors on networks.
Transportation Research Part B: Methodological, 41(8):841–861, 2007.

[109] Y. A. Korilis, A. A. Lazar, and A. Orda.
Achieving network optima using Stackelberg routing strategies.
IEEE/ACM Transactions on Networking, 5(1):161–173, 1997.

[110] G. Sharon, M. Albert, T. Rambha, S. Boyles, and P. Stone.
Traffic optimization for a mixture of self-interested and compliant agents.
In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[111] Transportation Networks for Research Core Team.
Transportation networks for research.
`https://github.com/bstabler/TransportationNetworks`.
Accessed: 11-2019.

[112] G. Nilsson, P. Grover, and U. Kalabić.
Assignment and control of two-tiered vehicle traffic.
In *2018 IEEE Conference on Decision and Control (CDC)*, pages 1023–1028. IEEE, 2018.

[113] L. Zhang.
Behavioral foundation of route choice and traffic assignment: Comparison of principles of user equilibrium traffic assignment under different behavioral assumptions.
Transportation Research Record, 2254(1):1–10, 2011.

[114] V. L. Knoop, D. de Jong, and S. P. Hoogendoorn.
Influence of road layout on network fundamental diagram.
Transportation Research Record, 2421(1):22–30, 2014.

[115] V. L. Knoop and S. P. Van Lint, H.and Hoogendoorn.
Traffic dynamics: Its impact on the macroscopic fundamental diagram.
Physica A: Statistical Mechanics and its Applications, 438:236–250, 2015.

[116] H. Mahmassani, J. C. Williams, and R. Herman.
Performance of urban traffic networks.
In *Proceedings of the 10th International Symposium on Transportation and Traffic Theory*, volume 14, pages 1–20. Elsevier Amsterdam, The Netherlands, 1987.

[117] N. Geroliminis and J. Sun.
Properties of a well-defined macroscopic fundamental diagram for urban traffic.
Transportation Research Part B: Methodological, 45(3):605–617, 2011.

[118] D. Helbing.
Derivation of a fundamental diagram for urban traffic flow.
The European Physical Journal B, 70(2):229–241, 2009.

[119] L. Leclercq and N. Geroliminis.
Estimating MFDs in simple networks with route choice.
Transportation Research Part B: Methodological, 57:468–484, 2013.

[120] J. Haddad and N. Geroliminis.
On the stability of traffic perimeter control in two-region urban cities.
Transportation Research Part B: Methodological, 46(9):1159–1176, 2012.

[121] G. Mariotte, L. Leclercq, and J. A. Laval.
Macroscopic urban dynamics: Analytical and numerical comparisons of existing models.
Transportation Research Part B: Methodological, 101:245 – 267, 2017.

[122] S. F. A. Batista, L. Leclercq, and N. Geroliminis.
Estimation of regional trip length distributions for the calibration of the aggregated network traffic models.
Transportation Research Part B: Methodological, 122:192–217, 2019.

[123] S. Nakayama and R. D. Connors.
A quasi-dynamic assignment model that guarantees unique network equilibrium.
Transportmetrica A: Transport Science, 10(7):669–692, 2014.

[124] M. C. J. Bliemer, M. P. H. Raadsen, L. J. N. Brederode, M. G. H. Bell, L. J. J. Wismans, and M. J. Smith.
Genetics of traffic assignment models for strategic transport planning.
Transport Reviews, 37(1):56–78, 2017.

[125] N. Mühlich, V. V. Gayah, and M. Menendez.
An examination of MFD hysteresis patterns for hierarchical urban street networks using micro-simulation.
In *94th Annual Meeting of the Transportation Research Board*, 2014.

[126] R. Herman and I. Prigogine.
A two-fluid approach to town traffic.
Science, 204(4389):148–151, 1979.

[127] S. Wiggins.
Introduction to Applied Nonlinear Dynamical Systems and Chaos.
Springer Science & Business Media, 2003.

[128] Edward Ott.
Chaos in Dynamical Systems.
Cambridge university press, 2002.

[129] S. Banerjee and G. C. Verghese.
Nonlinear Phenomena in Power Electronics.
IEEE, 1999.

[130] S. Webber and M. R. Jeffrey.
Micro-slip as a loss of determinacy in dry-friction oscillators.
International Journal of Bifurcation and Chaos, 29(06):1930015, 2019.

[131] S. J. Hogan and K. U. Kristiansen.
On the regularization of impact without collision: the Painlevé paradox and compliance.
Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 473(2202):20160773, 2017.

[132] A. C. Pigou.
The Economics of Welfare.
Macmillan, 1920.

[133] S. C. Dafermos and F. T. Sparrow.
The traffic assignment problem for a general network.
Journal of Research of the National Bureau of Standards B, 73(2):91–118, 1969.

[134] A. F. Filippov.
Equations with the right-hand side continuous in x and discontinuous in t.
In *Differential equations with discontinuous righthand sides*, pages 3–47. Springer, 1988.

[135] M. di Bernardo, C. Budd, A. R. Champneys, and P. Kowalczyk.
Piecewise-smooth Dynamical Systems: Theory and Applications, volume 163.
Springer Science & Business Media, 2008.

[136] M. R. Jeffrey.
Modeling with nonsmooth dynamics, volume 7 of *Frontiers in applied dynamical systems*.
Springer, March 2020.

[137] P. T. Piiroinen and Y. A. Kuznetsov.
An event-driven method to simulate Filippov systems with accurate computing of sliding motions.
ACM Transactions on Mathematical Software, 34(3), May 2008.

[138] M.C.J. Bliemer, M.P.H. Raadsen, E.-S. Smits, B. Zhou, and M.G.H. Bell.
Quasi-dynamic traffic assignment with residual point queues incorporating a first order node model.
Transportation Research Part B: Methodological, 68:363–384, 2014.

[139] Michiel CJ Bliemer and Mark PH Raadsen.
Static traffic assignment with residual queues and spillback.
Transportation Research Procedia, 38:647–667, 2019.

[140] N. Chernov and R. Markarian.
Chaotic Billiards.
Number 127. American Mathematical Soc., 2006.

[141] W. Peng, G. Dong, K. Yang, and J. Su.
A random road network model and its effects on topological characteristics of mobile delay-tolerant networks.

IEEE Transactions on Mobile Computing, 13(12):2706–2718, 2014.

List of Abbreviations

STAP	Static traffic assignment problem
OD	Origin-destination
UE	User equilibrium
SO	System optimal
PoA	Price of anarchy
MFD	Macroscopic fundamental diagram
NFD	Network fundamental diagram
PDS	Projected dynamical systems