The Influence of Structure in Supply and Demand on the Performance Characteristics of Road Traffic Networks

An exploration of how methodological approaches from network science can be implemented for a transportation research problem

Steven James O'Hare

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

Institute for Transport Studies

March, 2015

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

The right of Steven James O'Hare to be identified as Author of this work has been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

Acknowledgements

I would like to acknowledge the following individuals.

First and foremost, my thanks are due to my supervision team of Richard Connors and David Watling for their support, tuition, critical input and encouragement over the course of my research. Their guidance has made the last three and a half years an often thoughtful, always interesting and ultimately rewarding journey.

I would also like to thank my fellow PhD students at the university for their friendship and for providing numerous opportunities to discuss research related ideas and problems.

My thanks are also due to the University of Leeds for funding my research through the award of a University Research Scholarship, which provided me with peace of mind and stability during the course of my studies.

Finally, I would like to convey special thanks to my partner, Joanne, my parents, and the rest of my friends and family for their support and encouragement.

Abstract

Recently, researchers in the field of Network Science have begun to study how the structural properties of road traffic networks affect their performance characteristics. An understanding of how different structures of network infrastructure and travel demand combine to yield different performance characteristics would be useful because it could help identify how existing road traffic networks could be used more effectively or how structural features, which yield desirable performance characteristics, could be built into the construction of new road traffic networks.

Thus far, however, these studies have been restricted to numerical experiments with synthetic networks that do not provide plausible representations of real road traffic networks. Furthermore, these studies have used a disparate range of parameter settings for supply and demand structure, making it difficult to generalise their findings, and have provided no explanations for their conclusions.

To address these deficiencies, this thesis proposes an investigative framework for studying the effects of structure on the performance characteristics of road traffic networks. This framework comprises an experimental part, which describes how to design and conduct numerical experiments so as to provide useful insights into how performance varies with respect to specific aspects of network structure; and an analytical part, which focuses on developing explanations for patterns uncovered numerically.

This thesis then demonstrates the application of this framework to an investigation of how two performance indicators; the average link Volume-to-Capacity ratio and the Price of Anarchy, vary with respect to four aspects of road traffic network structure. As part of this investigation, a simple model of road network generation is presented that produces spectrums of plausible, synthetic road traffic network ensembles, which vary with respect to specific aspects of structure. Focussing on the variation of the Price of Anarchy with travel demand, this thesis then establishes theory that explains several features of the variation shown numerically.

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Research Outline

Over the last twenty years, advances in instrumentation and computing have significantly increased the amount of data that exists on a wide range of natural and manmade systems. Examples of such systems include biological systems, such as food webs and human metabolic networks; technological systems, such as the Internet and World Wide Web; infrastructure systems, such as freight distribution networks and national power grids; and social systems, such as friendship groups (Newman, 2003). The increasing availability of data on such a vast array of different systems has stimulated the emergence of a new field of research, called *Network Science*, in which researchers, from a variety of disciplines, have begun to study the characteristics of such systems. The aims of network science include: 1) to characterise the structural properties of networks that underlie real-world systems, 2) to develop models that explain the formation of such structures, and 3) to investigate how the structural properties of networks affect the emergent characteristics of the systems they support (Newman, 2010).

Within this trend, road traffic networks have been one of the many subjects of study. There have been empirical studies, which have sought to describe the structural characteristics of network infrastructure and patterns of travel demand in urban areas from across the world (Barthelemy, 2011). Theoretical studies have proposed generative models of network growth, which describe the formation and evolution of road traffic networks over time (Barthelemy and Flammini, 2009, Courtat et al., 2011). There have also been numerical studies that have drawn comparisons between the performance characteristics of different types of networks from the network science literature, using ensembles of synthetically generated networks and traffic equilibrium modelling techniques (Sun et al., 2012, Wu et al., 2008b, Wu et al., 2008a, Wu et al., 2006, Youn et al., 2008, Zhao and Gao, 2007, Zhu et al., 2014).

The network science approach to the study of networks has been recognised as a new perspective from which urban areas, in particular, can be studied and understood. In particular, it has been argued that network science *"has the potential to enrich current approaches to city planning"* (Batty, 2008). However, as an emerging research field, network science is not without criticism, and there are still opportunities for researchers from other fields to make significant contributions towards its development (Alderson, 2008, Havlin et al., 2012). This is certainly the case for road traffic networks (Lin and Ban, 2013). Indeed, one of the main criticisms made of empirical studies of network structure and theoretical models of network growth proposed thus far is that they have focussed almost exclusively on network

connectivity, and have therefore neglected other domain relevant features, such as link capacities and junction types, which also play a defining role in the systemic characteristics of road traffic networks. In addition, numerical studies of the influence of structure on the performance characteristics of road traffic networks have focussed on network types that are not plausible for real road traffic networks and also lack a systematic investigative approach, which makes it difficult to generalise their findings to other network types. There is therefore significant scope for additional contributions to this field within each of the identified research themes.

The goal of this thesis is to explore how contributions and methodological approaches from network science can be more appropriately and systematically applied within the specific context of the third theme of network science research, which addresses the research question of how the performance characteristics of road traffic networks vary with respect to the structural properties of supply and demand. This thesis therefore aims to directly address the criticisms that have been made of existing approaches, which were identified towards the end of the preceding paragraph.

1.2 Justification for the Research

1.2.1 Why investigate the influence of structure in supply and demand on the performance characteristics of road traffic networks?

Transport has a significant impact on economic prosperity (Eddington, 2006), and also plays a critical role in the functioning of modern society. Road traffic networks, at both the urban and interurban level, play an integral role in this impact, and account for 90% of all passenger traffic kilometres travelled and approximately 68% of all freight tonnes moved in the United Kingdom (UK)[1], and approximately 86% of all commuting trips in the United States of America (USA)[2]. Road traffic networks are also critically important in low-income countries as they act as a catalyst for economic development (Rodrigue et al., 2006). Road traffic networks provide people with a means to commute between homes and workplaces, to visit commercial and leisure facilities, and to access public services such as hospitals and schools.

---

[1] Tables TSGB0101 and TSGB0401 of Transport Statistics Great Britain (TSGB) 2014 (Department for Transport). Available at www.gov.uk/government/statistics/transport-statistics-great-britain-2014

[2] Table 1-41 of National Transportation Statistics from the Bureau of Transportation Statistics and available at: www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national transportation_statistics/html/table_01_41.html

Given their importance, it is perhaps unsurprising that national and regional governments invest substantial amounts of money in road traffic networks. For example, in 2012/13, public authorities in the UK spent approximately £6.3 billion on road infrastructure[3]. Furthermore, in December 2014, the UK government announced a road investment strategy[4] in which it proposed to invest £15.7 billion over six years, between 2015 and 2021, in one-hundred and twenty-seven road schemes across the UK, with projects including new roads, road widening and junction improvement schemes, and the use of new technologies in the provision of smart motorways. These examples illustrate how road traffic networks attract a significant amount of investment.

Decisions on how to spend such funding in the UK are made under an appraisal framework, which guides responsible bodies in how to identify appropriate schemes for investment. This is a sequential process, which begins with the identification of problems in the transport system that need to be addressed. Such problems are typically identified through reference to transport objectives, for example, with respect to the efficiency of networks, and also to standards, which provide a quantifiable yardstick against which problems can be identified. A common problem that is often identified in road traffic networks (particularly in urban areas) is congestion, which Ortúzar and Willumsen (2001) define as arising when *"demand levels approach the capacity of a facility and the time required to (travel through it) increases well above the average under low demand conditions."* Congestion has been estimated to cost the UK economy at least £20 billion per year (Goodwin, 2004).

Once a problem has been identified in a specific location, solutions are then developed in order to mitigate its impact. Such solutions can include supply-side policy interventions, such as improvements to traffic signals at junctions, road widening or the construction of a bypass; or demand-side policy interventions, such as demand management measures in the form of parking controls or congestion charging schemes. The impacts of selected interventions are then typically tested within a transport model in order to evaluate their effectiveness, and solutions that are deemed to offer value for money and be suitably desirable with respect to other social, economic and environmental objectives are then implemented. This process is then repeated as other problems are identified.

Whilst this approach has been adopted as a standard method and has been in use for many years, several authors have argued that it also has flaws. Chief amongst these flaws is that this

---

[3] Table RDE0101 of TSGB 2014

[4] Available at www.gov.uk/government/publications/road-investment-strategy-overview

approach to planning focuses narrowly on transport problems at a specific location, at specific point in time, and that it therefore lacks a systematic consideration of the characteristics of road traffic networks at a global scale or of how the effects of many individual decisions compound over time. Ortúzar and Willumsen (2001) argue that *"by narrowing a transport problem we may gain the illusion of being able to solve it"* but that *"transport problems have the habit of 'biting back', of reappearing in different places and under different guises."* Xie and Levinson (2009), citing Curry (1964), also highlight that, whilst each individual decision to intervene in a specific problem may be rational in isolation, when all interventions are viewed as a whole, the process can appear to be completely random. Without a systematic consideration of the influence of patterns in supply and demand structure at a global scale, it is also unclear whether policy interventions can be successfully transferred between networks with different structures and whether they will be as effective.

An understanding of how different configurations of supply and demand structure combine to yield different performance characteristics would be useful for both researchers and policy makers, as it would allow them to design more appropriate and/or effective transport solutions for existing road traffic networks based on their individual structural characteristics. Such understanding also has potential implications for the construction of new road traffic networks as it could lead to an understanding of desirable combinations of supply and demand structure that could be built-in and so hard-coded into the make-up of new networks.

1.2.2 Why use methodological approaches from network science as the starting point?

Accepting the premise, argued above, that the study of how the structural properties of supply and demand affect the performance characteristics of road traffic networks is worthwhile, the most obvious follow-up question to the stated goal of this thesis is 'why use network science as the starting point?'. Indeed, looking to the transportation literature, it is evident that research communities in geography, urban studies and transportation have all made prominent contributions within each of the three identified research themes of network science. For example, geographers have developed theories that describe the development of urban form as a complex combination of social, environmental and economic factors (Pacione, 2005), and have also proposed classification systems for street patterns within road traffic networks (Marshall, 2005). In addition, several studies in transportation research have focussed on the effects of road network structure on network performance; for example, see Tsekeris and Geroliminis (2013) and Ortigosa and Menendez (2014). Given these contributions, it is reasonable to question what network science has to offer that is different in comparison with these other approaches.

It is argued here that there are several key differences between these research disciplines and network science. Firstly, geographical approaches tend to be based upon qualitative descriptions of network patterns, which make them difficult to apply within numerical experiments. This contrasts with the network science approach, which is based upon numerical datasets and can therefore be more naturally applied within numerical experiments. Secondly, transportation research studies are typically location specific and focus on individual case studies; for example, Tsekeris and Geroliminis (2013) focussed on an idealised concentric city network pattern and Ortigosa and Menendez (2014) focussed on a grid network. The narrow focus of transportation studies makes it difficult to evaluate the generalizability of their findings with respect to the wide array of structures that are known to exist in real road traffic networks. This contrasts with approaches in network science, which look for broad-scale commonalities in network phenomena across a wide range of network structures.

For these reasons, it is argued that network science is a sensible and interesting starting point for the research described in section 1.1.

1.3 Research Objectives

Within the context of the goal of exploring how the performance characteristics of road traffic networks vary with respect to the structural properties of supply and demand, the research described within this thesis had two objectives:

1. To develop a systematic methodological approach, incorporating methods from network science, for investigations of how network performance varies with respect to the structural properties of supply and demand in road traffic networks. This approach should be generally applicable to a wide range of performance phenomena and should also provide an intelligible foundation for further research.
2. To apply this methodology to identify and characterise relationships between one or more aspects of supply and demand structure in road traffic networks, and one or more measures of network performance.

1.4 Research Scope

By *road traffic networks*, this thesis refers specifically to transport systems in which travellers move about in private, motorised vehicles through an arrangements of roads and junctions across a geographical area such as a city, a region or a country. This thesis does not consider other modes of transport such as urban rail systems, nor does it consider bus systems, bicycle or pedestrian flows, which also make use of road networks. The definitions of other important terms are defined as follows:

- *Supply* refers to the physical infrastructure composed of roads and junctions, and their capacity to provide for the movements of travellers.
- *Demand* refers to the magnitude and distribution of the movements of travellers between different locations when aggregated across the geographical area served by a network.
- *Performance characteristics* refers to the many measures and methods that have been developed to quantify how well a road traffic network achieves its function of providing for the movements of travellers.

In the context of this final bullet point, it is acknowledged that there are many performance measures and methods that are commonly used and which each highlight different aspects of road traffic network performance. In working towards the objectives described in section 1.3, this thesis focuses on two specific aspects of performance (to be defined shortly). The application of this methodology with other performance measures is left as a goal for future research.

As a consequence of the stated goal of exploring how contributions and methodological approaches from network science can be applied to investigate how network structure affects performance in road traffic networks, this thesis devotes considerable space to a comprehensive review of contributions from this field. As a consequence, it is highlighted that whilst other fields, most notably geography, spatial science and urban studies, have made significant contributions to the study of urban form and the structural characteristics of transportation networks, a comprehensive review of all past work in these areas was beyond the scope of this research and is not included. The interested reader is instead referred to Marshall (2005), Pacione (2005) and Ducruet and Beauguitte (2014) and references therein for further reading on contributions from these areas. This thesis does, however, attempt to explicitly acknowledge the important contributions from these fields and to provide a selection, if not an exhaustive compendium, of references to literature in relevant locations.

1.5 Description of Original Contributions

This thesis makes four original contributions to existing literature.

The first contribution of this thesis is an investigative framework that can be used by researchers to study the effects of structure in supply and demand on the performance characteristics of road traffic networks. This framework comprises an experimental component and an analytical component. The experimental part of the framework proposes a way of designing, conducting and recording the results of numerical experiments. This approach includes use of the network science method of drawing comparisons between the performance of ensembles of synthetically generated networks but then adds to it by: 1)

applying empirical data on the structural characteristics of real road traffic networks to make the networks under study more plausible than those used in existing literature, and 2) focussing experiments on comparing performance across a spectrum of network ensembles, which vary with respect to specific aspects of network structure rather than between isolated ensembles as have been used in existing network science studies. The analytical component of the framework then focuses on developing explanations to explain patterns uncovered by the numerical experiments. This investigative framework sets out a systematic approach to tackling what is shown to be high dimensional problem. It is shown how it addresses the deficiencies of existing approaches in existing literature and also how it can be applied to a wide range of research questions.

As a second contribution, this thesis then demonstrates the application of the experimental part of the above framework to an investigation of how two performance indicators; average link Volume-to-Capacity ratio and the Price of Anarchy (which measures the inefficiency of selfish routing in road traffic networks) vary with respect to four aspects of supply and demand structure in the specific context of urban road traffic networks. The four aspects of network structure that are the focus of these experiments are travel demand density, and the size, density and connectivity of network supply structure; selected because empirical studies in network science have shown there to be considerable variation in these features across different urban areas. As part of this investigation, a simple model of road network generation is presented that is able to produce spectrums of synthetic network ensembles, which provide plausible representations of urban road traffic networks and which also vary with respect to each of the aforementioned structural dimensions. Several challenges and opportunities for further research are identified as a result of this investigation; in particular, with respect to the computational burden of numerical experiments and the lack of empirical data on several key aspects of the structure of supply and demand in urban road traffic networks.

Focussing on the variation of the Price of Anarchy with respect to travel demand, the third original contribution made by this thesis is the establishment of theory that characterises four mechanisms that govern the variation observed in the numerical experiments described above. This contribution thereby demonstrates the application of the analytical part of the proposed investigative framework. Through a series of theorems, propositions and conjectures, this section of the thesis characterises the different effects of the mechanisms that govern the variation of the Price of Anarchy and also provides a series of numerical examples to illustrate these results. This theory is shown to be applicable in the general setting of road traffic networks with multiple Origin to Destination (OD) movements and continuous, differentiable, separable and strictly increasing link cost functions under the User Equilibrium

(UE) and System Optimal (SO) modelling paradigms. This section also includes several additional results for a commonly studied special case of the UE and SO models in which network links have BPR-like cost functions[5]. In particular, it is shown that there is a systematic relationship between link flows under UE and SO, and also that the Price of Anarchy has power law decay for high demand.

Following observations made regarding the small size of the Price of Anarchy in the numerical experiments, the fourth (and final) contribution of this thesis is an explanation for these observations and the proposal of an alternative measure of the inefficiency of selfish routing that complements the existing measure. It is shown that this new measure; called Price of Anarchy Delays, is subject to the same upper bounds as the Price of Anarchy but that it also achieves larger values that are closer to this upper bound. The usefulness of this new measure in practical applications is also discussed.

1.6 Thesis Structure

The remainder of this thesis is divided into seven chapters. The content of these chapters is described in the paragraphs that follow. Figure 1.1 then illustrates how the six chapters that make up the main body of the thesis are interlinked.

Chapter 2 summarises and provides a critical review of relevant literature. This review focuses on the key contributions and methodological approaches of network science, the structural properties that have been shown to exist in supply and demand in real road traffic networks, and the effects of such structural properties on the performance characteristics of road traffic networks. This review also identifies gaps and deficiencies in approaches used in existing literature to which this these seeks to contribute new results and understanding.

Chapter 3 proposes an investigative framework for studying the effects of network structure on performance in road traffic networks. This chapter begins with a discussion of the main challenge that is faced by numerical experiments; namely, of how to select networks from the huge, multi-dimensional space of all possible configurations of supply and demand structure. This chapter then describes the drawbacks of how approaches in existing literature have addressed this challenge thus far, which then feeds into the proposal of the investigative framework.

Chapters 4 and 5 demonstrate the application of the numerical part of the investigative framework and focus on how two performance indicators; the average link Volume-to-

[5] Cost function composed of a constant term plus a monomial of a single variable with a positive power

Capacity ratio and the Price of Anarchy, vary with respect to the four dimensions of network structure; the density of travel demand and the size, density and connectivity of network supply structure. Chapter 4 describes the model of network generation that was used in these experiments and demonstrates how it is capable of producing spectrums of plausible, synthetic road traffic network ensembles. Chapter 5 then describes the numerical experiments and parameter settings used with the aforementioned model, presents the results of the numerical experiments and then provides a discussion of what the results show.

Motivated by the results of these numerical experiments, chapter 6 explores how the Price of Anarchy varies with respect to travel demand; thereby demonstrating the application of the analytical part of the investigative framework proposed in chapter 3. This chapter begins by characterising the existence of four mechanisms that govern the variation of the Price of Anarchy with travel demand and then characterises the effect of these mechanisms on the Price of Anarchy itself through a series of theoretical results. The final part of this chapter is then devoted to a series of numerical examples, which illustrate these theoretical results.

Chapter 7 focuses on why values of the Price of Anarchy observed in chapter 5 are small and goes on to propose a new measure; called Price of Anarchy Delays, as an alternative measure of the inefficiency of selfish routing. This chapter includes a discussion of the complementary perspective that this new measure provides and also proves that it is subject to the same upper bounds as the Price of Anarchy.

Chapter 8 evaluates the extent to which the aims and objectives of this thesis have been met and acknowledges limitations of the research presented. This chapter also highlights opportunities for further research.

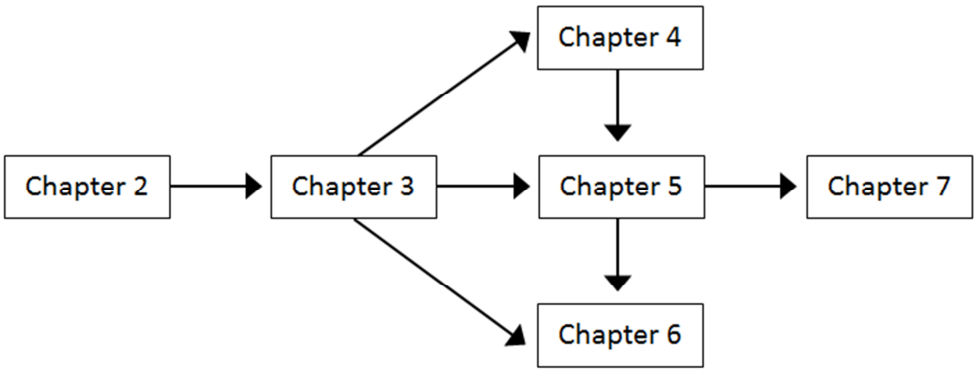

Figure 1.1 - A representation of how the chapters within this thesis are interlinked

2 Literature Review

2.1 Introduction

This chapter summarises and provides critical review of literature relevant to the research described in chapter 1 and so demonstrates the current state of knowledge. Importantly, this chapter identifies gaps in the literature to which this research seeks to contribute new methods, results and understanding.

With the research objectives described in section 1.3 in mind, the following Literature Review Questions were posed in order to define the scope of this literature review.

1. What are the key contributions and methodological approaches used in network science?
2. What structural properties have been shown to exist in supply and demand in road traffic networks?
3. How have the effects of supply and demand structure on the performance characteristics of road traffic networks been studied thus far and what have such studies found?

These questions are addressed in the three sections that follow. Section 2.2 focuses on introducing the main contributions that network science has made to the study of networked systems; specifically with respect to characterising their structural properties and to studying how structure affects their performance characteristics. Sections 2.3 and 2.4 then focus specifically on road traffic networks. Section 2.3 describes literature on the structural properties of supply and demand in road traffic networks. Section 2.4 then describes literature on how the performance characteristics of road traffic networks are affected by network structure. Whilst focussing primarily on contributions from network science, the material presented in these sections also makes reference to important contributions and literature from geography and transportation research. This chapter then concludes, in section 2.5, with a summary of main findings, which are categorised according to the above questions.

2.2 Network Science: Origins and Contributions to the Study of Networked Systems

The US National Research Council describes research under the umbrella of network science as *"the study of network representations of physical, biological, and social phenomena leading to predictive models of these phenomena."* As stated in the introduction, the broad aims of network science include: 1) the characterisation of the structural properties of real-world networked systems; 2) the proposal of models to explain the formation of such structures; and 3) the investigation of the effects of the structural properties of networks on the emergent

characteristics of the systems they support (Newman, 2010). This section describes the most significant empirical and theoretical contributions from network science to the study of networked systems under each of these aims and begins with a brief historical overview.

2.2.1 A Brief Historical Overview

The origins of network science are commonly credited to Leonhard Euler and his 1735 solution to the seven bridges of Konigsberg problem (Boccaletti et al., 2006, Costa et al., 2011, Newman, 2003). This problem sought a proof for whether it was possible for a person to cross all seven bridges in the city, shown on the left-hand side of Figure 2.1, in one uninterrupted walk without crossing any bridge twice. By transforming the problem into the simplified representation shown on the right-hand side of Figure 2.1, Euler realised that for a walk to exist, all of the intermediate islands (the blue circles) visited on such a walk must have an even number of connections with other islands. Otherwise, no matter which path was chosen, the traveller would become stuck on an island with an odd number of connections. Euler showed that such a walk is impossible in the setup shown in Figure 2.1 because every island has an odd number of bridges.

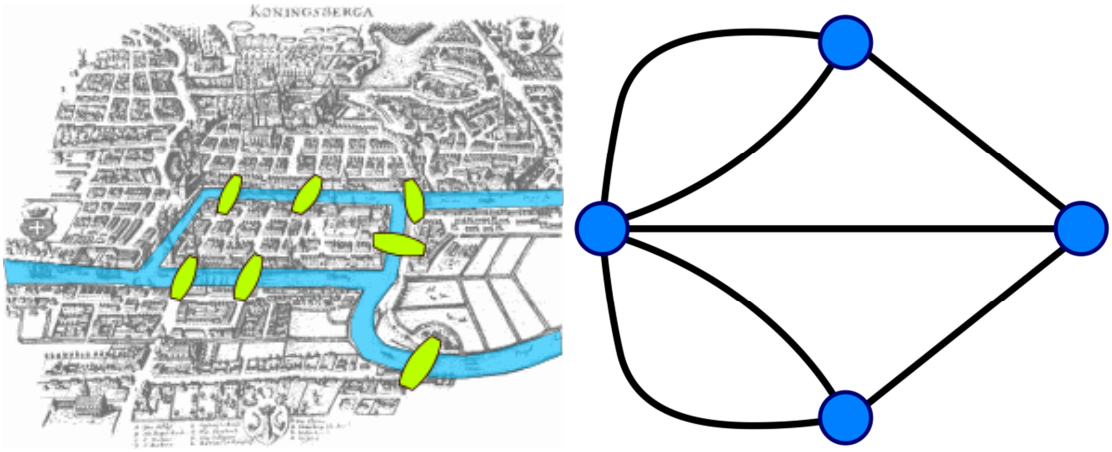

Figure 2.1 - Seven Bridges of Konigsberg Problem (left) and Graph Representation (right)[6]

The important advance that Euler made in solving this problem was the realisation that it was the connectivity properties of each island that mattered rather than the spatial aspects of the problem; e.g. the lengths of the bridges or the size of the islands. The simplified representation shown on the right-hand side of Figure 2.1 is one of the first examples of what came to be known as a *graph*, with the blue circles referred to as *vertices* and the black lines referred to as *edges*. Over time, *graph theory*, as it became known, has developed into a substantial body of material, which has produced a range of definitions and methods for characterising the

[6] Images downloaded from http://en.wikipedia.org/wiki/Seven_Bridges_of_Konigsberg

structural characteristics of graphs, which eventually led on to the formation of network science. Some of these key definitions and methods are described in section 2.2.2.

In the age of modern science, the earliest commonly referenced examples of the application of tools from graph theory to the study of real-world networked systems came in the field of social networks analysis in the first half of the twentieth century (Newman, 2003). Due to limitations on the availability of data and computing power at this time, these early works were predominantly limited to small networks that could be drawn by hand and to the study of the properties of individual vertices. An example of such research is that undertaken by Jacob Moreno, who studied the number of friendships of individuals within friendship groups in order to identify those people who were well connected and those who were isolated (Moreno, 1934). Moreno studied this idea by developing a graphical representation of friendship groups, called a sociogram, much like Euler did for the Konigsberg bridge problem. The idea that different network components have different levels of *importance* was later formalised in the concept of *centrality* introduced by Bavelas (1948) and in the centrality measures of Nieminen (1974), Sabidussi (1966) and Freeman (1977) (Freeman, 1979).

There was also an early interest in networks for their ability to transmit material between different nodes; for example, in the experiments by Milgram (1967) on path lengths in social networks. In these experiments, random individuals in Nebraska and Kansas, in the Midwest of the USA, were asked to try to get a letter to a specific individual in Boston, on the east coast. Participants who did not know this person were instructed to send the letter to someone they knew who they thought was more likely to know the target individual. Although the experiments suffered from high refusal rates, the letters that did arrive did so in six steps on average. This is the origin of the *small-world* phenomenon; the idea of a large network that can be traversed in a surprisingly small number of steps.

Given the data limitations that existed for these early studies, an important theoretical development in the study of networks was the development of generative models that could produce synthetic representations of networked systems. The first prominent example of such a model was the random graph model of Erdös and Rényi (1959) (Albert and Barabasi, 2002, Boccaletti et al., 2006, Newman, 2003), which was studied in a series of papers in the 1960s. This model, and other versions that followed, provided alternative sources of data for the study of networked systems and, at the time of their publication (and given a lack of evidence to the contrary) were thought to provide realistic representations of such systems (Newman, 2003).

As the availability of datasets and computing power increased in the latter half of the twentieth century, it became clear, however, that the random graph model did not reproduce

properties of networks that were increasingly being observed empirically (Newman, 2003). For example, in many systems it was found that the distribution of the numbers of edges k attached to each vertex obeyed a *power law* of the form $p(k) \propto k^{-\gamma}$, for some $\gamma > 0$, in contrast to the poisson distribution that was known to exist in random graphs. This eventually culminated, at the turn of the century, with the publication of two seminal papers that proposed two new models of real-world networked systems: the small-world model of Watts and Strogatz (1998), which modelled the small-world phenomenon previously highlighted by Milgram (1967); and the preferential attachment model of Barabasi and Albert (1999), which provided an explanation for the power law phenomena described above and defined a new class of so-called *scale-free* networks. These models are described in more detail in section 2.2.3.

Following publication of these new models, there was an explosion of new empirical research, which primarily focussed on uncovering whether the signatures of small-worlds and power laws could be detected in real-world systems. Over time, such studies found that a wide range of networked systems of a variety of different types - social, technological, informational and biological - did indeed have such signatures and therefore similar structural characteristics. Examples of such results are described in more detail in section 2.2.4. Many of these studies also proposed new structural measures that have been used to characterise various different features of network structure.

In more recent years, several studies have also gone on to study the effects of structure on traffic flow phenomena and the performance characteristics of networked systems; for example, with respect to jamming phenomena in network routers on the Internet. These studies are described in more detail in section 2.2.5.

This overview illustrates how network science and its antecedents in graph theory have made a number of key contributions to networks research; in particular, in the form of methods to characterise network structure, theoretical models to produce representations of networks, empirical studies of the structural properties of real networked systems, and studies of the effects of structure on the emergent characteristics of networked systems. Further detail on each of these areas is provided in sections 2.2.2 to 2.2.5.

2.2.2 Characterising Networked Systems and their Structural Properties

As described at the beginning of section 2.2.1, the characterisation of networked systems and their structural properties began in graph theory. This section therefore begins, in section 2.2.2.1, with a summary of definitions and methods from this field. Section 2.2.2.2 then presents a short survey of measures and methods that have since been proposed for

characterising the structural properties of networks, which have been used by empirical studies in network science.

2.2.2.1 Graph Theory: Definitions and Notation

Stated formally, a *graph* G is an ordered pair (V, E) consisting of a set of *vertices* V that are connected to each other by a set of *edges* E, together with an incidence function φ_G, which associates an unordered pair of *vertices* from V with each *edge* of G (Bondy and Murty, 2008). Networks can also be described using the same terminology, although vertices and edges are often instead referred to as nodes and links. This thesis will adopt this practice.

Two examples of graphs are shown in Figure 2.2. In this figure, *nodes* are represented by the circles and the *links* are represented by the lines. If two nodes are connected to each other then they are called *adjacent*, and the number of links that are incident with an individual node is called the *degree* of that node, which is denoted by k. Visual representations of graphs can also be described as having *faces*; the white spaces enclosed by the nodes and links.

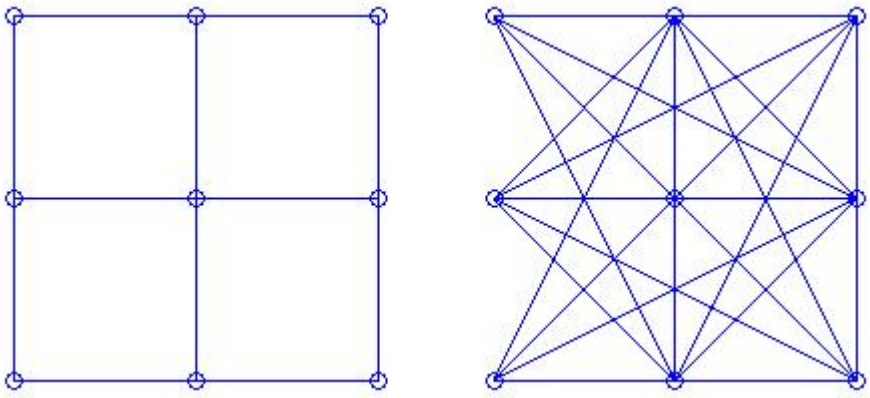

Figure 2.2 - Examples of Graphs, Vertices/Nodes (the circles) and Edges/Links (the lines)

All of the links in the graphs shown in Figure 2.2 are *unweighted* and *undirected*. However, graphs can also have links that are *weighted* or *directed*. In the former case, links are *weighted* by a numerical value, which is typically used to represent length or connection cost. Links that are *directed* can be traversed in only one direction. Graphs that contain only directed links are called *directed graphs* or *digraphs*. A link that has the same start and end node is called a *loop*, and two or more links that share the same start and end node are referred to as *parallel*. A graph that has no loops or parallel links is called *simple* and a *complete graph* is a simple graph in which every pair of nodes are adjacent.

In addition to visual representations, such as those displayed in Figure 2.2, graphs can also be represented by an *adjacency matrix*, which is particularly useful for storing graphs within a computer. An adjacency matrix is a square matrix whose dimension is equal to the number of

nodes and whose entries a_{ij} represent the number of links between node i and node j, for all node pairs ij. A simple graph has an adjacency matrix whose entries $a_{ij} = 0$ or 1 $\forall i, j$, and a complete graph has an adjacency matrix that has entries $a_{ij} = 1$ $\forall i, j$ for which $i \neq j$ and $a_{ij} = 0$ $\forall i, j$ for which $i = j$, i.e. on the principal diagonal.

With respect to travel within graphs, Graph Theory defines a *walk* or *path* as an ordered sequence $v_0 e_1 v_1 \dots v_{l-1} e_l v_l$, whose terms alternate between nodes and links (Bondy and Murty, 2008). Two links are said to be *connected* if a path exists between them and a *connected graph* is a graph in which a path exists between every pair of nodes. A *cycle* is a path that starts and ends at the same node. The *shortest path* between two nodes is the path using the minimum number of links (or of shortest length if the links are weighted), and the *diameter* of a graph is the length of the longest shortest path over all node pairs.

A special type of graph that will be particularly relevant in this thesis are *planar* graphs, which refer to those graphs that can be drawn in the plane in such a way that their links meet only at nodes and nowhere else (Bondy and Murty, 2008). As examples, the left-hand graph in Figure 2.2 is planar but the right-hand graph is not planar. The most well-known property of planar graphs is Euler's formula, which relates the number of nodes n, the number of links m and the number of faces f to each other as follows: $n - m + f = 2$. Using this formula, an upper bound for the average node degree, commonly denoted $\langle k \rangle$, of a planar graph can be derived; this being $\langle k \rangle \leq 6$, which leads to an upper bound on the maximum number of links in a planar graph of $m = 3n - 6$ (Barthelemy, 2011). Planar graphs with this many edges are known as *maximal planar graphs*. At the opposite extreme to maximal planar graphs are *trees*, which are connected graphs that have no cycles and for which the number of links is $m = n - 1$.

2.2.2.2 Commonly Used Measures of Network Structure

Over time, a large number of *measures* have been proposed to quantify and so characterise the structural properties of networked systems. This section describes a selection of measures that have been commonly used by empirical studies in network science and which are also referred to in later sections of this thesis. It should be noted that the list of measures presented here is by no means exhaustive and that many more measures have been proposed and used; the interested reader is referred to the reviews of Newman (2003), Boccaletti et al. (2006) and Barthelemy (2011) for further examples.

Before describing such measures, however, it is worth noting that many of these measures can be applied to both *graphs* and *networks*. The same is also true for the terms defined in section 2.2.2.1. For this reason, the terms *graph* and *network* are often used interchangeably in the literature. This is also true of the text presented in this thesis thus far. Strictly speaking

however, these terms do refer to different things; in broad terms, a graph is a *"mathematical concept"* whilst a network refers to *"an interconnected system of things"*[7]. However, in practical terms, in networks research, this distinction (or abuse of terminology as some may call it) is often not important.

As stated in section 2.2.1, amongst the earliest proposed measures of network structure were the centrality measures of Nieminen (1974), Sabidussi (1966) and Freeman (1977). These were, respectively, *degree centrality*, which measures the importance of a node based on its degree; *closeness centrality*, an extension of degree centrality that measures importance based on how near, in terms of the number of links in the shortest path, a node is, on average, to all other nodes; and *betweenness centrality*, which measures the importance of a node based on how many shortest paths between other nodes travel through it. Using the notation introduced in section 2.2.2.1, the degree, closeness and betweenness centralities are defined respectively for a node i as follows:

$$Degree\ Centrality\ = k_i = \sum_{j \in V} a_{ij}$$

$$Closeness\ Centrality\ (i) = \frac{1}{\sum_{j \in V} d_{ij}}$$

$$Betweenness\ Centrality\ (i) = b_i = \sum_{j,k \in V; j \neq k; k \neq i} \frac{n_{jk}(i)}{n_{jk}}$$

where a_{ij} are adjacency matrix entries, d_{ij} is the number of links in the shortest path between i and j, n_{jk} is the number of shortest paths between node j and k, and $n_{jk}(i)$ is the number of shortest paths between node j and k that pass through node i. The fraction in the betweenness centrality measure accounts for the fact that there may be more than one shortest path between two nodes. The statistical distribution of the degree centrality measure over all nodes in a network is one of the most commonly used methods that is used by empirical studies in network science to characterise network structure, particularly with respect to the identification of power laws and *scale-free* networks, which were mentioned in section 2.2.1. These networks are defined more formally in section 2.2.3.3.

The identification of a *small-world* signature in the structure of a network uses two different network measures: *average shortest path length* and *clustering coefficient* (Watts and Strogatz, 1998). Average shortest path length, denoted $\langle l \rangle$, for a network can be calculated as follows:

---

[7] http://efoundations.typepad.com/efoundations/2008/01/graphs-networks.html

$$\langle l \rangle = \frac{1}{n(n-1)} \sum_{i \neq j} d_{ij}$$

where d_{ij} can either be the number of links in the shortest path between i and j or based on the total cumulative length of those links. It is known that for a d-dimensional regular lattice network, $\langle l \rangle \sim n^{1/d}$ (Barthelemy, 2011), where n equals the number of nodes.

Whilst average shortest path length focuses on paths, the *clustering coefficient* focuses on the connectivity properties of individual nodes and measures the extent to which the neighbours of a node i are also adjacent to each other. The clustering coefficient C_i for a node i can be calculated as follows (Watts and Strogatz, 1998):

$$C_i = \frac{2h_i}{k_i(k_i - 1)}$$

where h_i is the number of links between the nodes that are adjacent to node i. In recent years, a somewhat related aspect of structure that has attracted increasing attention is the existence of *community structures* within networks, which refer to groups of nodes that are highly interconnected with each other but have very few connections with other nodes outside the group. Measures that are currently used to characterise communities are very complex and there is currently no commonly agreed method (Barthelemy, 2011); for example, see Fortunato (2010) for an extended discussion of this point.

The final measure that is mentioned is with respect to the concepts of *assortativity* and *disassortativity* in networks. This refers to, respectively, the prevalence for whether individual nodes in networks tend to be adjacent to nodes that have similar degree or whether individual nodes tend to be adjacent to nodes that have different degrees. This is an interesting feature because it begins to highlight the existence (or lack of) hierarchical features within the connectivity structure of networks. Assortativity can be measured via the *assortativity coefficient* (Newman, 2002), which can be calculated as follows:

$$\Gamma = \frac{c \sum_i j_i k_i - \left[c \sum_i \frac{1}{2}(j_i + k_i)\right]^2}{c \sum_i \frac{1}{2}(j_i^2 + k_i^2) - \left[c \sum_i \frac{1}{2}(j_i + k_i)\right]^2}$$

where j_i and k_i are the degrees of the nodes at the ends of the ith link, m equal to the number of links and $c = 1/m$. $\Gamma > 0$ indicates an assortative network whereas $\Gamma < 0$ indicates a disassortative network. $\Gamma = 0$ indicates that there is no correlation between link connections and node degree.

The above measures are commonly used by empirical studies in network science to characterise the structural properties of networked systems. As should be evident from their definitions, many of these measures apply to individual network components, such as

individual nodes, links or faces. The typical approach taken by empirical studies that have applied these measures has therefore been to examine the average value or statistical distribution of each measure across all such components within a network.

Despite the success of empirical studies, as will be shown in section 2.2.4, in highlighting a range of interesting features and similarities between networks from a range of different disciplines, the approach to the selection of measures and their application has been subject to the criticism that studies often lack a clear purpose. In other words, that, in some cases, measures are created and applied to data simply because they can be rather than with an intended purpose in mind. Newman (2003) best articulates this criticism as:

"…while we are beginning to understand some of the patterns and statistical regularities in the structure of real-world networks, our techniques for analysing networks are at present no more than a grab-bag of miscellaneous and largely unrelated tools. We do not yet, as we do in some other fields, have a systematic program for characterizing network structure…"

This is a theme that is also apparent in empirical studies of supply structure in road traffic networks, which are the focus of section 2.3.1.

2.2.3 Key Theoretical Models of Network Structure

Theoretical models for networked systems are useful because they can be used to identify the key mechanisms that underpin the formation and evolution of networks observed in the real world. The random graph, small-world and preferential attachment network models are the most cited in the network science literature and have each inspired a range of associated models. This section defines these models and describes some of their properties.

2.2.3.1 Random Graph Model

In the random graph model of Erdös and Rényi (1959), a given number of nodes n is used to create a graph where each possible (undirected) link between each pair of nodes is present with a given and uniform probability p. Examples of random graphs generated on twenty-five nodes for three values of p are shown in Figure 2.3.

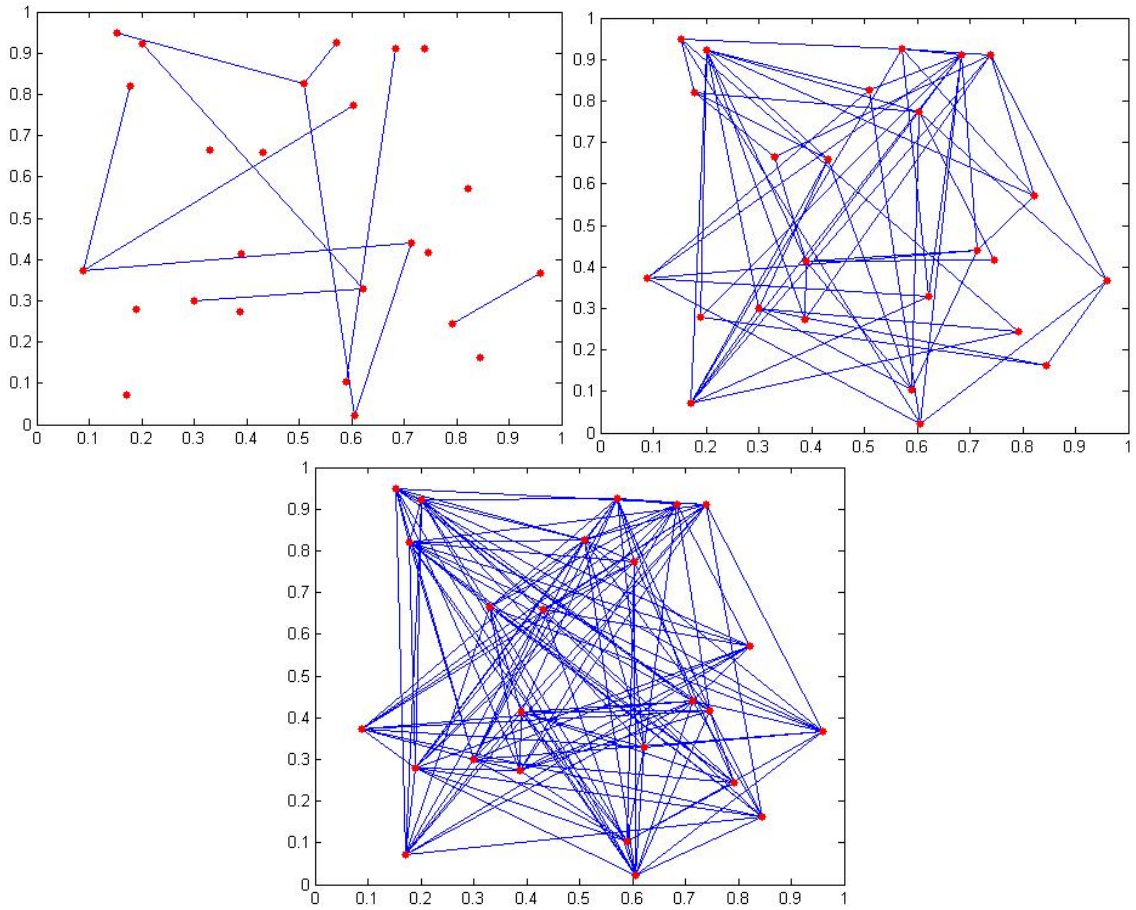

Figure 2.3 - Three examples of random graphs generated by the model of Erdös and Rényi (1959) for $p = 0.03$, $p = 0.2$ and $p = 0.5$ respectively

At $p = 0$, no links are added to the network, whilst at $p = 1$, every possible link is present and the model produces a complete graph. In the spectrum between these extremes, graphs have a variable number of edges and one of the key findings of Erdös and Rényi (1959) was a critical probability p_c above which a giant connected component exists, which connects the majority of nodes in the graph.

2.2.3.2 Small-World Model

Starting from a two-dimensional regular lattice with n nodes, the small-world model of Watts and Strogatz (1998) randomly rewires links with a given and uniform probability p. Examples of graphs generated using this model for three different values of p are shown in Figure 2.4. $p = 0$ produces a regular lattice network whereas $p = 1$ generates a random graph of the type described in section 2.2.3.1.

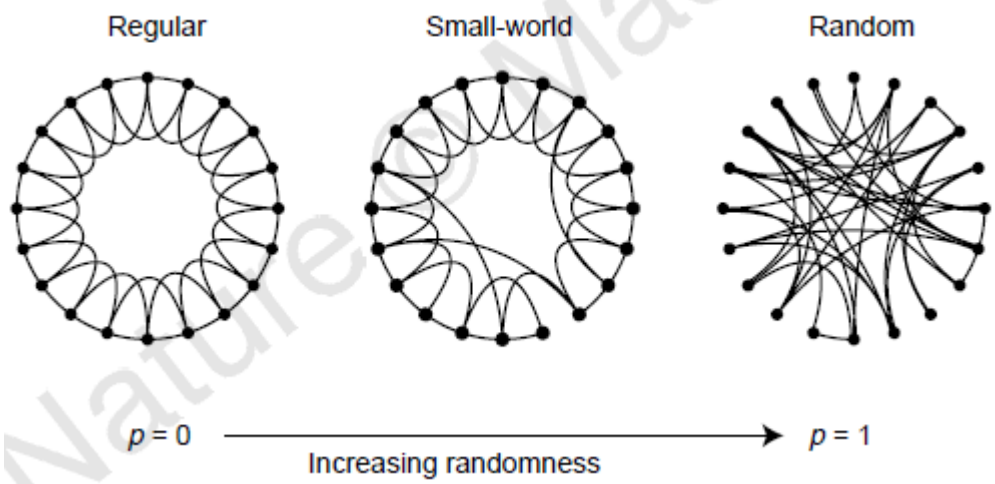

Figure 2.4 - Examples of graphs generated by the small-world model (Figure 1, Watts and Strogatz (1998))

It is known that two dimensional lattice networks have an average shortest path length (here measured as the number of links) $\langle l \rangle \sim \sqrt{n}$, and that such networks are also highly clustered, as is shown on the left-hand side of Figure 2.4. Through the process of random rewiring, Watts and Strogatz (1998) recognised that it was possible to generate graphs that retained this high level of clustering and that also had much smaller average shortest path lengths of the order $\langle l \rangle \sim \log(n)$, which were more akin to random graphs. It was shown that a very small value of p could generate networks that remained highly clustered like lattice networks but also had small average shortest path lengths like random graphs. This effect is illustrated in Figure 2.5, taken from Watts and Strogatz (1998), which plots the values of clustering coefficient C and average shortest path length (denoted by L in their paper as opposed to $\langle l \rangle$) as functions of p between $p = 0$ and $p = 1$. The values of $C(p)$ and $L(p)$ in this graph have been normalised by their respective values at $p = 0$ in order to emphasise the differing ways in which these two measures change as p increases. Also note that the p values are plotted on a log scale; this illustrates that the critical value of p at which the small world effect appears is very small. This figure demonstrates how the addition of relatively few links creates shortcuts between nodes that were previously connected by much longer paths; thereby leading to a lower average shortest path length.

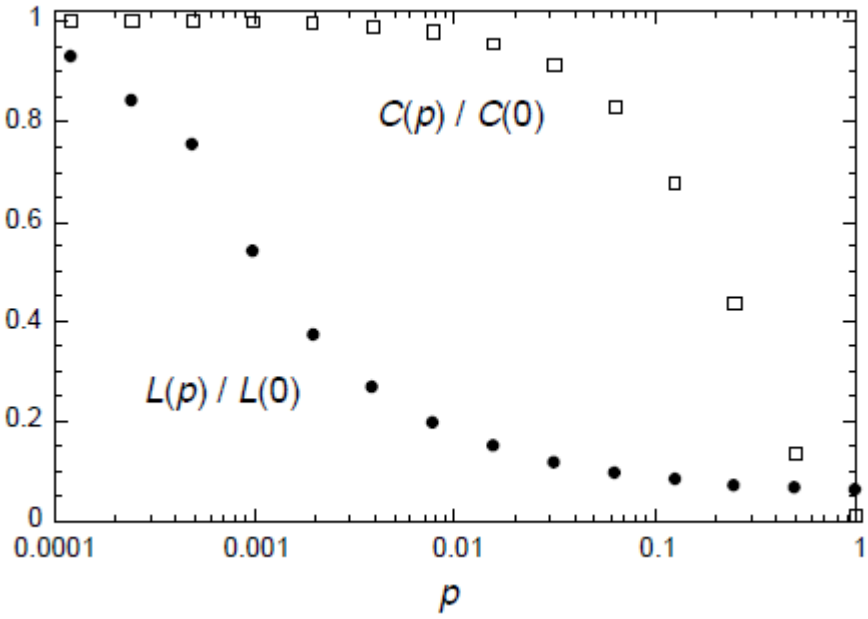

Figure 2.5 - Variation of average shortest path length L and clustering coefficient C with respect to p in the small-world model. (Figure 2, Watts and Strogatz (1998))

2.2.3.3 The Preferential Attachment Model

The final model introduced in this section is the preferential attachment model of Barabasi and Albert (1999), which is sometimes referred to as the BA model. The generative process within this model is different to the random graph and small-world network models because instead of starting with a predetermined set of nodes, which are then connected to each other according to some rules, this model gradually adds new nodes to a domain, which are then connected to existing nodes through the addition of links. The BA model has two defining components: (a) the continuous expansion of the network through the addition of nodes, such that each new node is immediately connected to the network guaranteeing a connected network, and (b) that new nodes attach preferentially to nodes that are already well connected in the network, i.e. to those nodes that are already of high degree.

In notation, the network starts with a given small number of nodes m_0 at time $t = 0$. The first component of the model is then incorporated by adding one new node at each time step, which is then connected to $m \leq m_0$ other nodes already present in the network. The second component is then incorporated by setting the probability that a new node connects to a node i as $\Pi(k_i) = k_i / \sum_j k_j$. After being run for a sufficient length of time, this process results in a network with a power law degree distribution of the form $p(k) \sim k^{-\gamma}$ with exponent $\gamma = 2.9 \pm 0.1$.

These networks are characterised by the existence of hub nodes; that is nodes with degree significantly higher than most other nodes. This feature sets these networks apart from

networks generated using the small-world and random graph models, which have exponential scaling in the tails of their node degree distributions. Exponential scaling decays much faster than power law scaling and so high degree nodes are much more rare in small-world and random graphs than they are in networks generated by the preferential attachment model. The difference between random and scale-free networks is illustrated in Figure 2.6, which highlights hub nodes in grey.

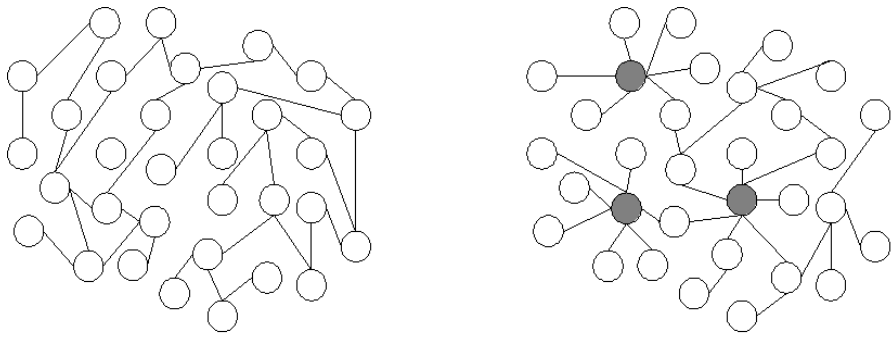

Figure 2.6 - Examples of Random (left) and Scale-Free (right) networks[8]

Networks generated by the preferential attachment model are referred to as *scale-free networks* precisely because there is no characteristic range that can be used to describe the values sampled from a power law distribution, as is possible with a Poisson distribution.

2.2.4 Empirical Studies of the Structural Properties of Networked Systems

As described in sections 2.2.1, empirical studies in network science have studied the structural properties of real-world networked systems using measures such as those identified in section 2.2.2. In many of these studies, the main focus has been on whether a network displays small-world or scale-free properties against the null hypothesis that the network is random. These studies are the subject of section 2.2.4.1. It is also possible to identify a class of empirical studies that focus specifically on spatially constrained networks. Findings for these studies are the subject of section 2.2.4.2.

2.2.4.1 Detection of Small-World and Scale-Free Signatures

The method used to identify a small-world signature in a network compares calculated values of average shortest path length (calculated as the number of links) $\langle l \rangle$ and average clustering coefficient C, with values for the same measures, denoted $\langle l \rangle_{rand}$ and C_{rand}, calculated on a random counterpart graph; that is, a random network that has the same number of nodes and links (Watts and Strogatz, 1998). A small-world signature is said to have been detected if

[8] Image from http://commons.wikimedia.org/wiki/File:Scale-free_network_sample.png

$\langle l \rangle \approx \langle l \rangle_{rand}$ and $C \gg C_{rand}$, i.e. if the average shortest path length of the network is approximately similar to a random graph but it exhibits much greater clustering between nodes. After defining the small-world model, Watts and Strogatz (1998) went on to demonstrate evidence for the existence of small-world signatures in the neural network of a nematode worm, the power grid in the USA and a network of collaborations between actors in films. This evidence is shown in Figure 2.7 ($L = \langle l \rangle$ in this figure).

	L_{actual}	L_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

Figure 2.7 - Evidence for small-worlds in the film actors network, US power grid and neural network of a nematode worm (C. elegans). (Table 1, Watts and Strogatz (1998))

The typical method used to identify whether a network has a scale-free signature is the production of a plot of node degrees k against probability $p(k)$ on a doubly logarithmic scale and to observe whether a straight line pattern emerges in the data points. This approach has been used, for example, by Faloutsos et al. (1999) and Albert et al. (1999) for the Internet and by Redner (1998) for two networks of citations between scientific papers. The plots produced by Redner (1998) are shown in Figure 2.8 as examples.

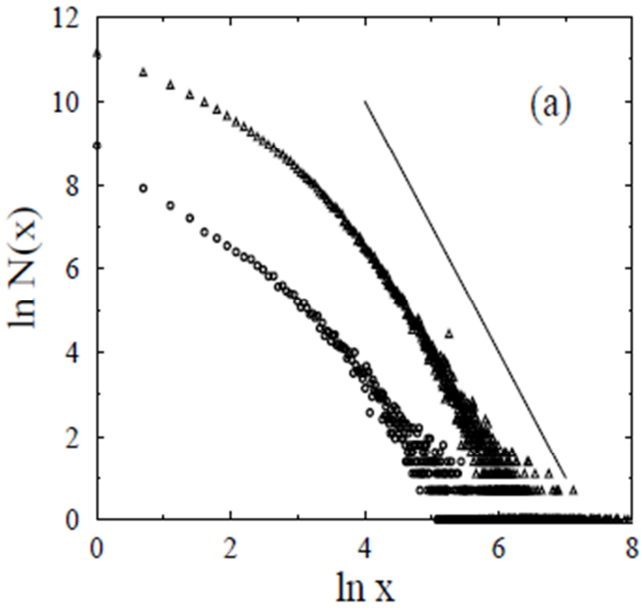

Figure 2.8 - Evidence of power law scaling in the degree distribution of two networks of citations between scientific papers. (Figure 1a, Redner (1998))

The above cited examples provide a small illustration of how empirical studies in network science have uncovered that many real-world network systems share similar structural characteristics. The survey papers of Albert and Barabasi (2002) and Costa et al. (2011)

summarise many more examples of the application of this approach and include results on the structural properties of a diverse range of systems including the Internet and World Wide Web, Biological networks, Social networks and Infrastructure networks. However, in recent years, there has also been criticism of these findings from other research disciplines. For example, statisticians have criticised the method of producing a log-log plot to detect a power law as overly simplistic at best or misleading and erroneous at worst, with some research finding that power laws do not exist where they were previously claimed to exist by others (Clauset et al., 2009). The most prominent example of this criticism is a claim that the Internet does not actually have a power law node degree distribution; for example, by Doyle et al. (2005). There has also been criticism that, in focusing primarily on the structural properties of the graphs that underlie networked systems, studies in network science neglect other important domain relevant information, which significantly limits the usefulness of their findings (Alderson, 2008, Havlin et al., 2012). For example, all of the studies cited in this section thus far focussed solely on the connectivity properties of networks, and omitted other characteristics of nodes and links. Whilst more recent empirical studies, which are described in the next section, have begun to include some of these other characteristics, primarily in the form of link lengths, there is still some way to go on this issue. As will be shown in sections 2.3 and 2.4, this criticism can also be made of studies of road traffic networks.

2.2.4.2 The Structural Properties of Spatially-Constrained Networks

In some of the networks discussed in the previous section; such as the citation network and network of actor collaborations, the difference in cost between connections of differing lengths is relatively small. In contrast, in networks such as the Internet and transportation networks, this is not the case because their geographical embedding imposes constraints on their formation and operational characteristics. This typically manifests in increased costs of long distance connections, which therefore require a strong economic reason for their construction (Barthelemy, 2011). These differing constraints make such networks an interesting class to study within the broader family of networks.

Barthelemy (2011) provides a review of empirical studies of these networks, which include studies of transportation networks such as:

Airline networks – This includes work by Barrat et al. (2004) on the worldwide air-transportation network and Barrat et al. (2005) on the link between airport size and the magnitude of travel. These studies found that the *"airport connection graph is … a clear example of a spatial (non-planar) small-world network displaying a heavy-tailed degree distribution and heterogeneous topological properties"*

Bus and Subway (Underground) networks – This includes work by Sienkiewicz and Holyst (2005) and von Ferber et al. (2009) on the public transport networks of 22 polish and 15 worldwide cities respectively. In comparison with airline networks, these networks appear to have smaller average node degrees and longer average shortest path lengths. Barthelemy (2011) hypothesised that this is a consequence of the more restrictive spatial constraints that exist in these networks.

Cargo ship networks – This includes work by Hu and Zhu (2009) and Kaluza et al. (2010) on the worldwide cargo ship network. In contrast to bus networks, Barthelemy (2011) hypothesised that these works appear to show that such networks are less constrained by their spatial embedding and that long distance links are actually less costly than short distance links in such networks.

In his conclusions, Barthelemy (2011) highlighted that spatial networks can be broadly split into categories: planar networks, such as bus and subway networks; and spatial, non-planar networks, such as airline and cargo ship networks, where nodes have a geographical embedding but where links can intersect. Networks in the latter category appear to have more similarities in structure to the networks studied in the empircal studies described in section 2.2.4.1. As will be shown in section 2.3, road traffic networks fall into the former category. Barthelemy (2011) also identified several important influences of spatial constraints on the structural properties of networks. Firstly, that they restrict the occurence of high degree nodes and usually produce a degree distribution that is peaked around the average degree rather than a power law like many other real-world networked systems; secondly, that spatial constraints limit the length of links and, for planar networks, the link length distribution is usually peaked; and thirdly, that restrictions on node degree in planar networks constrain the formation of hub nodes in favour of short links, which tends to lead to highly clustered networks.

Like the studies of the previous section, the studies highlighted here have been criticised for stripping out domain relevant information. For example, Derrible and Kennedy (2011) criticised the studies of bus networks because they did not include the fact that bus networks are composed of transit lines and that the ability to transfer between transit lines is not explicitly recognised.

2.2.5 Studies of the Effects of Structure on the Performance Characteristics of Networks

In addition to contributions highlighted in the preceding sections, the network science literature also contains many studies that have investigated the effects of network structure on traffic flow dynamics and performance phenomena in networked systems. As described in

chapter 1, the goal of this thesis is to study how methods from network science can be used to investigate this question in the context of road traffic networks, so this literature is particularly important.

Studies in network science have been principally inspired by the existence of jamming phenomena in the Internet. Jamming occurs when routers malfunction or suffer a drop in performance, which results in a redistribution of flow and therefore congestion to other routers in the network (Boccaletti et al., 2006). The primary aim of such research has typically been to find the critical level of demand, denoted λ_c, at which the system moves from a free-flowing to a congested (or jammed) state and to test the effect of different network structures and different routing strategies on this critical level of demand.

The methodological approach used within such studies is based on models and has typically used either real network data from the Internet or canonical models from the network science literature, such as those described in section 2.2.3, to generate large ensembles of network topologies to be tested. The performance characteristics of these different ensembles of networks are then compared to each other under an assignment of demand using a *packet transmission model* with various different routing strategies. This comparison is based on the average performance across all networks within each ensemble. Examples of the application of this approach are described in Table 2.1.

A *packet transmission model* is designed to provide a basic representation of how the Internet routes data packets between different nodes in a network. In this model, nodes are either routers, which accept data packets as input and then forward them on through the network, or hosts, which are the origins and eventual destinations of data packets. The links, meanwhile, represent cables connecting routers and hosts to each other so forming a representation of the network of the Internet. The router nodes in these models determine the pattern of traffic flow because they dictate the rate, denoted β_i, at which packets can be sent on through the network. If the number of packets arriving at a router is greater than this rate then the excess is stored in a queue to be transmitted at a later point in time and all subsequently received packets are stored at the back of this queue. Once a packet reaches its destination it is removed from the network. As the above description suggests, this is a micro-simulation model within which the current positions of all individual packets are tracked and recorded over increasing time steps t.

Examples of eight studies that have used this approach are summarised in Table 2.1. The main aspects that vary between papers are, on the supply side, the size and connectivity structures of the underlying network topologies, and the rates β_i at which routers can forward packets per unit t. The main things that vary on the demand side are the rates λ at which packets are

introduced into the network and the extent to which congestion is accounted for within the routing strategies tested. Various measures are used to measure network performance. One common measure is the critical demand generating rate λ_c at which the model moves from a free-flow to congested state, which is determined by the *order parameter* (Arenas et al., 2001):

$$\eta = \lim_{t \to \infty} \frac{\langle \Delta\Theta \rangle}{\lambda \Delta t}$$

where $\Delta\Theta = \Theta(t + \Delta t) - \Theta(t)$, $\Theta(t)$ is the total number of packets in the network at time t, and $\langle \cdot \rangle$ is the average over time windows of Δt. This measure records the difference between the rate at which packets are introduced to the network, λ, and the rate at which packets reach their destinations. If the former is greater than the latter then the number of packets in the model is increasing over time and the system is said to have transited into a congested state.

Reference	Supply Structure				Demand Structure		Main Performance Indicator
	Network Topologies	Network Size	Network Realisations	Router Capacities	Demand Structure	Route Choice	
Ohira and Sawatari (1998)	Lattice	$n = 25, 50, 100, 150, 250$	Not reported	One packet per unit t, FIFO queuing	Random with a generation rate: λ packets per unit t	Based on Shortest Path (SP)	Average Travel Time
Echenique et al. (2004)	Subset of the Internet	$n = 10^4$	200	One packet per unit t, FIFO queuing	Random p packets generated. at $t = 0$	Shortest Path and Congestion aware routing	Maximum Packet Travel Time
Arrowsmith et al. (2005)	- Random - Scale-Free with exponent $\gamma = 3$ - Scale-Free with exponent $\gamma = 2$	$n = 500$, $\langle k \rangle = 3$	Not reported	One packet per unit t, FIFO queuing	Random with a generation rate: λ packets per unit t	Shortest Path Routing	Number of delivered packets, Average Travel Time
Zhao et al. (2005)	- Cayley Tree	$n = 1093$, $\langle k \rangle = 2$	50	Each node i has a transmission rate β_i units per unit t. Two scenarios: - $\beta_i \propto k_i$ - $\beta_i \propto b_i$	Random with a generation rate: λ packets per unit t	Shortest Path Routing	Order Parameter η
	- Lattice - Random - Scale-Free	$n = 1000$, $\langle k \rangle = 4$					

Echenique, Gomez-Gardenes and Moreno (2005)	Subset of the Internet	$n = 11174$	Not reported	One packet per unit t, FIFO queuing	Random with a generation rate: λ packets per unit t	Shortest Path and Congestion aware routing	Number of packets in the system, Order Parameter η
Tadic et al. (2007)	- Random - Scale-Free	$n = 1000$	Not reported	One packet per unit t, LIFO queuing	Random: a constant number p packets generated in network at each unit t	Random routing of packets until within two steps of destination where routing becomes deterministic	Travel Time Distribution
Tang and Zhou (2011)	- Scale-Free - Subset of the Internet	$n = 2000$	100	Each node i can send at most k_i packets per unit t. Two scenarios: - FIFO observed - FIFO ignored	Random with a generation rate: λ packets per unit t	- Shortest Path based on topological distance - Shortest path with links weighted by betweenness	Order Parameter η
Gavalda et al. (2012)	- Scale-Free - Random	$n = 5000$, $\langle k \rangle = 8$	50	One packet per unit t, FIFO queuing	Random with a generation rate: λ packets per unit t	Shortest Path and Congestion aware routing	Order Parameter η

Table 2.1 - Studies of Network Performance Phenomena in the Network Science Literature

The eight papers summarised in Table 2.1 used a range of supply and demand structures:

- On the supply side, there are two types of topology that have been used; synthetic and real, with synthetic topologies being much more common. In part, this is probably due to the ease with which they can be generated; for example, by using the generative models described in section 2.2.3. A wide range of network sizes are also evident. With the exception of Zhao *et al.* (2005), router capacity is typically one packet transmission per unit t.
- On the demand side, packets are randomly generated at a rate λ between a random origin and random destination in every paper. The appropriateness of such a distribution of demand is not commented upon in any of the papers but it is worth considering whether such a demand profile is realistic and also how other demand profiles would affect the results. These are all open questions. As stated earlier, routing strategies vary from shortest path routing to those that include some awareness of congestion.

In the context of the wide range of supply and demand structures identified by Table 2.1, all of these papers vary some aspect of structure in order to observe the effects on performance. For example, there are comparisons between different network topological structures; e.g. random versus scale-free, and also between different routing principles. There are also comparisons between different generating rates λ and network sizes n. These comparisons can be categoried into two types: *'named comparisons'*; like those between different topological structures, and *'parameterised comparisons'*; like those between different demand generating rates. A named comparison is essentially a comparison of two discrete entities. In contrast, a parameterised comparison allows comparison between many points over a spectrum of values. It is argued that comparisons of the latter type are more desirable because they provide a better indication of how performance varies between two extremes. The former comparison type is less useful because it is unclear how similar or different the networks being compared are to each other. It is important to note that parameterised comparisons exist only in those experiments for which the demand level is varied.

With respect to the performance measures, in addition to the order parameter, metrics based around travel time are also common throughout the referenced papers. The key point to note about the results is that they are averages over a number of realisations of each network type. This is necessary because the synthetic networks are generated using models that have a stochastic component, which, when rerun, therefore generate different networks, albeit with the same general properties. There is also randomness in the demand profile so even when using subsets of the real Internet network, the results presented are averages over many

simulations. All of the results presented in these papers are *average* comparisons *between* the various different network structures used; for example between the average performance of networks with random and scale-free topological structures. Such a comparison ignores the possibility that there may be variation *within* each group of networks, between networks that have similar supply and demand structures.

The papers summarised in Table 2.1 provide the following two main findings:

1. Routing Strategies that include some amount of congestion-awareness lead to higher (i.e. better) critical generation rates λ_c than routing strategies based purely on shortest paths.

In these papers the standard reference point for the performance of a routing strategy is routing by shortest paths. The above finding means that strategies that include some knowledge of the amount of traffic in other parts of the network; either locally or globally, improve route selection and mean that a network can accommodate more demand before becoming congested.

2. The critical generation rate λ_c is lower (i.e. worse) for tree and lattice network topologies than it is for scale-free network topologies but all of these have a lower critical generation rate than random networks.

In other words, random network topologies can accommodate more traffic than scale-free networks which, in turn, can accommodate more traffic than tree and lattice networks for a similar level of congestion. This finding indicates an interest in how the underlying network topology affects system performance. Arrowsmith et al. (2005) argued that the reason for this is that the underlying structure of scale-free networks means that many shortest paths travel through a small subset of hub nodes, which can forward only one packet per unit time and so become congested very quickly. This is in contrast to random networks where shortest paths are much more evenly distributed, a fact which is highlighted by the distribution of the betweenness centrality.

2.3 The Structural Properties of Supply and Demand in Road Traffic Networks

This section describes empirical studies of the structural properties of supply and demand in road traffic networks. Section 2.3.1 focuses on the structural properties of network supply (the physical infrastructure of junctions and roads) and section 2.3.2 focuses on describing generative models that have been proposed for road traffic networks. Section 2.3.3 then focuses on the structural properties of travel demand (characteristics of the movements of individual travellers between locations in a network).

Most of the content that is presented in these sections is derived from the network science literature and uses similar approaches to those used in empirical studies described in section 2.2. Perspectives are also provided from other research disciplines that have made contributions in these areas; specifically geography and the transportation literature.

2.3.1 The Structure of Supply in Road Traffic Networks

In the network science literature, empirical studies of the structural properties of supply in road traffic networks have, thus far, focussed exclusively on networks in urban areas. Such studies typically focus on a sample of urban areas and are based upon network data from geographic databases. Examples of data sources include: the NXI GESTATIO laboratory database, which was used by Buhl et al. (2006) for forty-one urban areas in Africa, Asia, Europe and South America; the Tele Atlas MultiNet^{TM} geographic database, which was used by Lammer et al. (2006) and Chan et al. (2011) for twenty urban areas in Germany; the TIGER database, which was used by Jiang (2007) and Zhang et al. (2011) for forty-one urban areas in the USA; and the Ordnance Survey and Integrated Transport Network (ITN) datasets, which were used by Masucci et al. (2009) and Gudmundsson and Mohajeri (2013) for forty-one urban areas in the United Kingdom.

Once extracted, the network data for each urban area are subjected to a processing stage in which the level of detail is simplified and any network that is outside the area of interest is removed. The remaining network for each urban area are then converted into an undirected, simple graph of nodes and links. Two approaches to this conversion process are evident in the literature: the *primal* approach and the *dual* approach. The primal approach to the study of network structure uses an intuitive graphical representation of a network in which nodes represent junctions and edges represent road segments between junctions. The dual approach, in contrast, uses an alternative representation in which nodes represent 'streets' and links represent intersections between streets. These approaches are described in more detail in later sections. As with studies of general networked systems, the structural properties of the primal/dual graphs for each urban area are then characterised by measures that quantify the characteristics of individual network components; the nodes, links and faces of the graph. The statistical distributions of such measures over all components within each network are also examined.

Examples of measures that have been used in empirical studies of the structural properties of road traffic networks are described in section 2.3.1.1. Sections 2.3.1.2 and 2.3.1.3 describe the main findings of empirical studies under the primal and dual approaches. A critical review of network science is then provided in section 2.3.1.4, which includes perspectives from geography.

2.3.1.1 Measures of the Structural Properties of Road Traffic Networks

In addition to the measures described in section 2.2.2, which are generally applicable to all network types, a separate class of measures have also been specifically developed in the context of road traffic networks. Derrible and Kennedy (2011) credited Garrison and Marble (1964, 1962, 1965) and Kansky (1963) for the development of some of the first measures of this type, which actually emanated from transport geography.

The three measures proposed by Garrison and Marble (1964, 1962, 1965) were called the alpha, beta and gamma indices. The α-index is a measure of network density with the following formula:

$$\alpha = \frac{m - n + 1}{2n + 5}$$

This measure is known to range between $\alpha = 0$ for a tree (as in this case $m = n - 1$) and $\alpha = 1$ for a maximal planar graph. This measure was later referred to as *meshedness* by Buhl et al. (2006) and denoted by M. The β-index relates the total number of links to the total number of nodes in the following formula:

$$\beta = \frac{m}{n}$$

Although the beta index is exactly the same as degree centrality averaged over all nodes, Derrible and Kennedy (2011) described it as an *"indicator of network complexity"*; the argument presumably being that the higher the average degree, the more complicated the network. Finally, the γ-index is also a measure of network density, which relates the total number of links in a network to the maximal possible number of links. For planar networks, Eulers Formula (see section 2.2.2.1) shows that the maximum number of edges in a planar graph is $3n - 6$ and so the formula for the γ-index is:

$$\gamma = \frac{m}{3n - 6}$$

This measure also ranges between values close to $\gamma = 0$ for trees and up to $\gamma = 1$ for a maximal planar graph.

Kansky (1963) also proposed three measures of network structure, which were average link length, average traffic flow per vertex and an indicator of system spread, which divided total network length by the diameter of the network.

An alternative measure for network efficiency that has been proposed is the *route factor*, denoted $Q(i,j)$, which relates the shortest path distance through a network between two nodes to the Euclidean distance between two nodes (Barthelemy, 2011). This measure has the following formula:

$$Q(i,j) = \frac{d_R(i,j)}{d_E(i,j)}$$

where $d_R(i,j)$ is the distance through the network and $d_E(i,j)$ is the Euclidean or 'crow flies' distance. Route factor is also referred to as *circuity* by Levinson and El-Geneidy (2009).

Lammer et al. (2006) also presented a series of measures to quantify the shape and structure of the faces of planar networks; i.e. the internal blank spaces enclosed by nodes and links. Lammer et al. (2006) referred to faces as cells and proposed measures including the *cells neighbourhood degree* k_c, which was defined as the number of adjacent cells to a cell c; the area of cells A_c; and the form factor, which has the following form:

$$\phi_c = \frac{4}{\pi(A_c/D_c^2)}$$

where D_c denotes the maximum distance between two points in the cell. Values of the form factor range between $\phi_c = 0$ for an infinitely narrow cell and $\phi_c = 1$ for a perfect circle (Barthelemy, 2011).

2.3.1.2 Findings under the Primal Approach

An example of the primal representation of an urban road traffic network is shown in Figure 2.9 below. As can be seen, the physical junctions in the network on the left-hand side are represented by nodes in the graphical representation on the right-hand side of Figure 2.9. Nodes are then connected to each other by straight lines representing the road segments. The city blocks (in grey) on the left-hand side are represented by the faces of the graph on the right-hand side of Figure 2.9. As previously stated, faces are typically referred to as cells in the network science literature.

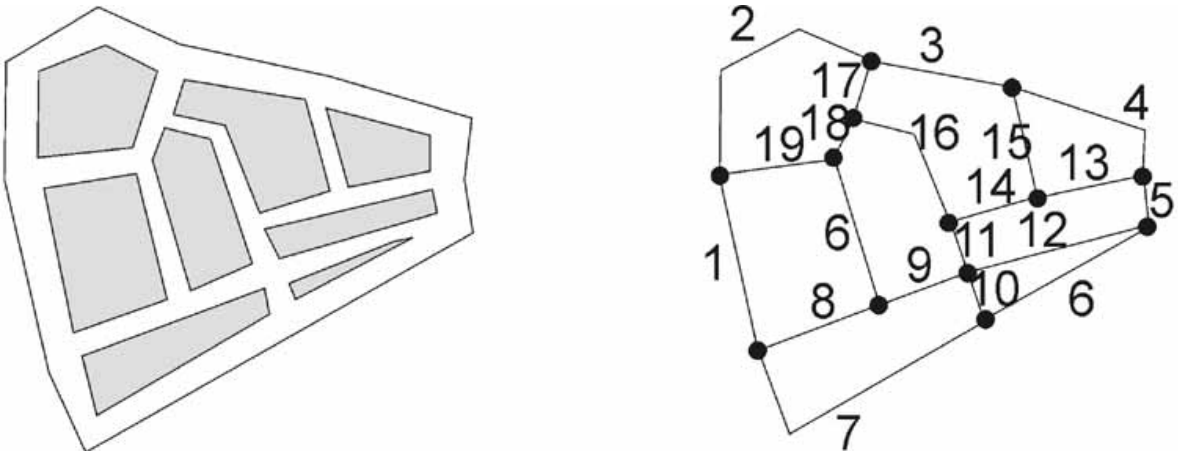

Figure 2.9 - Primal Representation of a Road Traffic Network (Figure 2, Porta et al. (2006b))

The findings of empirical studies under the primal approach can be categorised as focussing on the *microscopic* properties of nodes, links and cells and the overall *macroscopic* properties of networks. These aspects are addressed in sections 2.3.1.2.1 and 2.3.1.2.2.

2.3.1.2.1 Microscopic Properties of Urban Road Traffic Networks

Empirical studies have found that, at the *microscopic* level of nodes, links and cells, supply networks in a large number of urban areas share several common structural features.

Perhaps unsurprisingly, urban road traffic networks are predominantly planar (Buhl et al., 2006). As stated in section 2.2.4.2, it is known that planarity can significantly restrict other structural features in road traffic networks; such as the average node degree, $\langle k \rangle$, to values $\langle k \rangle \leq 6$ (Barthelemy, 2011). Indeed, in a study of the twenty largest cities in Germany, Chan et al. (2011) reported the very narrow range of $\langle k \rangle \in [3.17,3.31]$. Masucci et al. (2009) reported that London has $\langle k \rangle \approx 2.44$, which is close to a tree. This can be attributed to a massive presence of dead-ends in their graphical representation of the London network; approximately 30% of all nodes had degree 1. For the networks analysed by Buhl et al. (2006), $\langle k \rangle \in [2.02,2.87]$. However, Chan et al. (2011) attribute the difference in their findings to be because of their deletion of nodes of degree $k \leq 2$.

The probability distribution $P(k)$ of node degree is heavily constrained by the planarity constraint with nodes of degree $k \geq 5$ found to be very rare across several studies (Buhl et al., 2006, Chan et al., 2011, Masucci et al., 2009). In a study of this distribution on the networks of Germany, Chan et al. (2011) found that the tail ($k \geq 5$) of the distribution decayed as an exponential. However, they also identified that nodes of degree $k = 4$ occurred much more often than would be expected where the distribution to be exponential for all values of k. Masucci et al. (2009) described it as misleading to claim a distribution type but also found a sharp drop off in high degree nodes. Based on this, Barthelemy (2011) concluded that urban road traffic networks have a topological structure that is significantly different to other networks analysed in network science, which, it has been suggested, have power law scaling in their node degree distributions.

Turning to link based measures, Masucci *et al.* (2009) found that the probability distribution $P(l)$ of link lengths in London was well fitted by a power law with an exponential cut-off of the form $P(l) \propto \exp\left[-\frac{145}{l} - \frac{l}{2000}\right] l^{-3.36}$. Chan et al. (2011) did not identify a distribution type for their networks but did highlight a plateau in the distribution up to 100m, which then tailed off as longer links became less frequent. The plot of $P(l)$ in Masucci *et al.* (2009) displays the same feature. Barthelemy (2011) proposed that the decay in the probability of long distance

links is another consequence of the planarity constraint under which long distance links are conjectured to be particularly expensive to build and therefore rare.

Chan et al. (2011) also studied the angles between adjacent and opposite arms at junctions and found that the frequency distributions of these angles in a sample of German cities displayed two pronounced peaks at $\theta = 90^{\circ}$ and $\theta = 180^{\circ}$. Angles of $\theta = 90^{\circ}$ were more abundant and showed a larger dispersion than those about $\theta = 180^{\circ}$. For nodes of degree $k = 4$ the distributions were found to be peaked at $\theta = 90^{\circ}$, whereas for nodes of degree $k = 3$ the distribution had peaks at $\theta = 90^{\circ}$ and $\theta = 180^{\circ}$, with 90° angles roughly twice as common as 180° angles. Similar findings were also reported by Strano et al. (2013). Chan et al. (2011) provided three explanations for these findings: 1) the shortest Euclidean connection between a node and a nearby link is provided by a perpendicular connection, 2) rectangular cells are preferable for buildings and 3) angles significantly different to $\theta = 90^{\circ}$ make for complex turning movements for traffic. Using double angle distributions (the angle of deviation between straight-ahead movements) to measure the *"straightness"* of crossing roads, Chan et al. (2011) found a peak at $\theta = 180^{\circ}$ for nodes of degree $k = 4$ and peaks at $\theta = 180^{\circ}$ and $\theta = 270^{\circ}$for nodes of degree $k = 3$. These results suggest that in urban road traffic networks, links meet at nodes as would be expected in a grid-like structure.

Turning to the structural properties of cells; i.e. the city blocks, Lammer et al. (2006) found that the distribution of cell areas $p(A_c)$ in the city of Dresden followed a power law of the form $p(A_c) \propto A_c^{-1.9}$. Masucci et al. (2009) reported a similar finding for the city of London, which Barthelemy and Flammini (2008) have argued is in sharp contrast to a grid structure where cell areas are typically of a similar size. Focussing on the shape of cells, Chan et al. (2011), in light of their findings on link angles, proposed two measures of the *"rectangularity"* of cells to assess the deviation of the structure of an urban road traffic network from a perfect rectangular grid. For one of these measures; the squared cosine $\langle \cos^2 2\theta_l \rangle$, where θ_l range over all angles between adjacent links in a network, all networks were found to lie within a small range around approximately 0.7. Chan et al. (2011) found that the form factor measure, defined in section 2.2.2.2, fell between 0.3 and 0.6 for most of the cells in all of the cities they studied, with very few examples of values above 0.6. This was linked to the abundance of nodes of degree $k = 3$ and $k = 4$.

With respect to the radial distribution of nodes, links and cell areas, Masucci et al. (2009) and Chan et al. (2011) both showed that London and cities in Germany have a high density of nodes, short links and small cell areas in their city centres, but that their networks become more dispersed as distance from the centre increases.

Overall, these results suggest that there are several common rules that govern the structure of supply in road traffic networks across a broad range of urban areas.

2.3.1.2.2 Macroscopic Properties of Urban Road Traffic Networks

In contrast to results at the microscopic level, at the macroscopic level the network science literature points to a high degree of structural variation between road traffic networks in different urban areas, particularly with respect to the size, density and connectivity of network infrastructure. For example, Chan et al. (2011) showed that urban areas in Germany vary considerably with respect to the geographical extent and the number of nodes in their road traffic networks, with areas ranging between 141 km^2 and 891 km^2; see Table 2.2. With respect to the density of network infrastructure, Chan et al. (2011) also illustrated that the number of nodes per square kilometre, denoted ϱ_n, varied considerably, from 9.8 nodes per km^2 in Bielefeld to 35.6 nodes per km^2 in Munich. Finally, with respect to network connectivity, Courtat et al. (2011) found a range of values for the meshedness measure (defined in section 2.2.2.2) across ten French cities, with values $M \in [0.2, 0.47]$; see Table 2.3. Cardillo et al. (2006) and Buhl et al. (2006) also found considerable variability in meshedness.

City	Population	Area (km^2)	Nodes (n)	Population Density (population/km^2)	Node Density (ϱ_n)
Berlin	3,392,425	891	19,931	3,807	22.4
Hamburg	1,728,806	753	9,044	2,296	12
Munich	1,234,692	311	11,058	3,970	35.6
Cologne	968,639	405	5,395	2,392	13.3
Frankfurt	643,726	249	3,911	2,585	15.7
Dortmund	590,831	281	3,281	2,103	11.7
Stuttgart	588,477	208	3,612	2,829	17.4
Essen	585,481	210	4,093	2,788	19.5
Dusseldorf	571,886	218	3,124	2,623	14.3
Bremen	542,987	318	3,827	1,708	12
Duisburg	508,664	233	2,837	2,183	12.2
Leipzig	494,795	293	3,753	1,689	12.8
Nuremberg	493,397	187	3,543	2,638	18.9
Dresden	480,228	328	3,346	1,464	10.2
Bochum	388,869	146	2,233	2,663	15.3
Wuppertal	363,522	168	1,750	2,164	10.4
Bielefeld	324,815	259	2,546	1,254	9.8
Bonn	308,921	141	2,094	2,191	14.9
Mannheim	308,759	145	2,674	2,129	18.4
Karlsruhe	281,334	173	2,204	1,626	12.7
Minimum	***281,334***	***141***	***1,750***	***1,254***	***9.8***
Average	***740,063***	***296***	***4,713***	***2,355***	***15.5***
Maximum	***3,392,425***	***891***	***19,931***	***3,970***	***35.6***

Table 2.2 - Size and Density of Supply Networks in twenty German Cities (Chan et al., 2011)

City	Connectivity (M)	City	Connectivity (M)
Angoulme	0.28	Grenoble	0.32
Avignon	0.23	Lyon	0.47
Caen	0.29	Rennes	0.26
Carcassonne	0.2	Rouen	0.38
Dijon	0.33	Troyes	0.28

Table 2.3 - Connectivity of Supply Networks in ten French Cities (Courtat et al., 2011)

2.3.1.3 Findings under the Dual Approach

One of the main criticisms of the primal approach from a network science structural analysis point of view is the restriction that its geographical embedding imposes on values of structural measures, for example on node degree. The dual approach removes these constraints. To explain the rationale for this alternative approach, Wagner (2008) used the example of giving someone directions to demonstrate that, in the mind of a traveller, a road traffic network is not viewed as a collection of distinct road segments but rather as a collection of contiguous streets and that when giving directions it is common to identify only those junctions where a turn is required and not every intermediate junction that is crossed. The starting point for this visualisation was the seminal work on the social logic of space by Hillier and Hanson (1984) and it is otherwise known as Space Syntax, as described for example by Batty (2004).

The dual representation of a network is generated by first identifying the contiguous streets according to some rule(s). The identified streets are then converted into nodes in the dual graph and edges are drawn between them if those streets intersect. There are several different rules by which contiguous streets have been identified in the literature. Three of these approaches are shown in Figure 2.10.

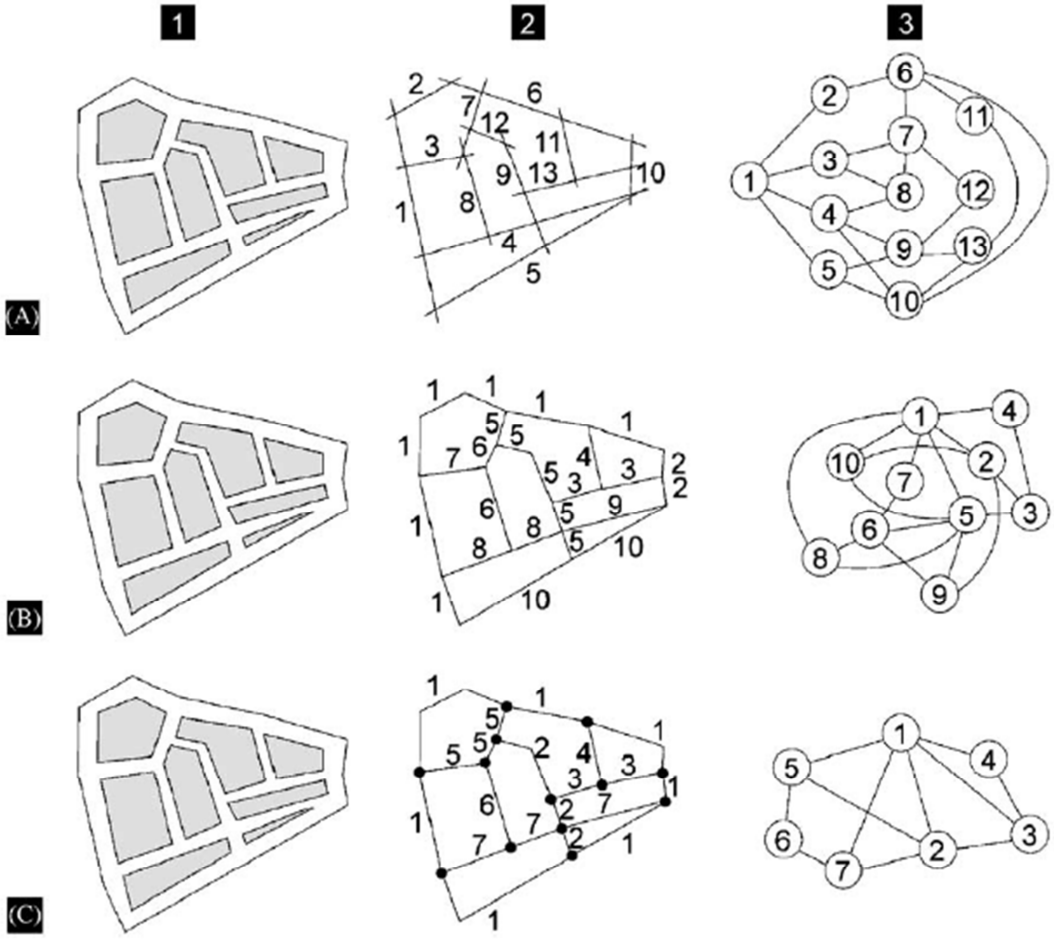

Figure 2.10 - Dual Representations of Road Traffic Networks (Figure 2, Porta et al. (2006a))

The approach used by Hillier and Hanson (1984) was to use axial lines; this approach is shown in the second image of row A in Figure 2.10. The lines represent contiguous sections of road that offer direct lines of sight. The second approach, shown in row B, used by Jiang and Claramunt (2004) identifies contiguous streets using street names. The third approach, shown in row C, proposed by Thomson (2004) is termed an Intersection Continuity Negotiation (ICN) model and is based on examining each node in turn and joining the two edges at each node that form the largest convex angle into one contiguous road segment and so on until all edges at each node have been examined and joined together. In this way it uses a 'principle of good continuity' where pairs of road segments at each node forming the most straight movement are joined to each other. This relaxes the strict line of sight rule that is part of the formation of axial maps, which reduces the number of nodes of degree two and therefore maintains continuity for curved roads.

A fourth approach, not shown in the above figure, used by Jiang and Liu (2009) for example, interpolates between the axial mapping and ICN model approaches and is called the natural streets approach. In this approach the same principle of continuity, as described for the ICN

model, is used but only up to a certain upper threshold ϕ of deviation from a straight line. A setting of $\phi = 0^{\circ}$ allows no deviation from the straight line and returns something similar to the axial map, whilst a large ϕ will allow a larger deviation between contiguous segments and the ICN model map.

The final images in each row of Figure 2.10 illustrate that, depending on the chosen method of conversion, it is possible to generate many different dual graphs from the same road network. This is one of the main criticisms of the dual approach.

The main advantage of the dual approach is that it releases networks from their geographical constraints and so, for example, the degree distribution of nodes of the dual graph can vary much more widely than in the primal graph. In analyses of the pattern of this degree distribution, Porta et al. (2006a) found that in real road traffic networks of a sufficient size, the distribution obeys a power law, making these networks similar to those analysed in other disciplines within network science. In analysing clustering coefficient and path lengths of dual graphs, Porta et al. (2006a) also found evidence for a small-world structure, although these findings were somewhat limited because they came from one square mile samples of only six cities. Further evidence for this phenomenon was found in a much larger study carried out by Jiang (2007). In this work, the node degree distributions of the dual representations of the full road traffic networks of forty US cities, generated via the natural streets approach with a threshold degree of $\phi = 70°$, were found to obey power laws with exponents $\gamma \approx 2$. Jiang (2007) also concluded that *"about 80% of streets within a street network have degrees or lengths less than the average of the network, while 20% of streets have length or degrees greater than the average. Out of the 20%, there are less than 1% of streets which can form a backbone of the street network"*. Using the named streets approach Kalapala et al. (2006) analysed the degree distributions of the national highway networks of Denmark, England and the United States, and concluded that they too obeyed power laws with exponents $\gamma \in [2.2,2.4]$.

In releasing road traffic networks from their geographical constraints the above examples show how the dual representation has been able to uncover recurrent patterns, which point towards the existence of highly connected and central streets. However, the many methods that exist to generate dual representations raise questions of how sensitive the above results are to the conversion model chosen; something that Porta et al. (2006a) highlighted as being a question of whether results truly reflect properties of the underlying network or are a facet of the generation process chosen. The primal representation of network structure has been the model of choice for traffic modelling because it retains the link to geographical space and the influence this has travel patterns. This is lost when the dual approach is used because a street

is compressed to a single point. Some authors have developed and applied models of traffic flow on the dual graph, for example Hu et al. (2008) and Zeng et al. (2009), but it is still unclear what the benefits are for doing this.

2.3.1.4 Summary and Critical Review

The results of empirical studies presented in sections 2.3.1.2 and 2.3.1.3 highlight that there are both similarities and differences in the structural properties of road traffic networks across different urban areas. However, these results should be viewed with the following caveats.

Firstly, there is significant variation between studies with respect to how raw network data is processed. For example, with respect to identifying the boundary for each urban area, a wide range of methods have been used in existing studies. For example, Cardillo et al. (2006) and Buhl et al. (2006) used an artificial square boundary; whereas, Masucci et al. (2009) defined the boundary using a circle centred on the 'centre' of the city. In contrast to these geometric approaches, Jiang (2007) used boundaries for administrative regions, which are likely to be more related to political geography than any particular physical aspect of the network. Further differences are evident with respect to the simplification of network data and, in particular, the deletion (or non-deletion) of nodes of degrees one and two. For example, Chan et al. (2011) chose to remove all nodes of degree one (and their associated links) that represented dead-ends, whereas Masucci et al. (2009) chose to retain them. These inconsistencies make it difficult to compare results between different studies as such decisions can have a significant impact on the values of structural measures. This can be seen in the results presented in section 2.3.1.2.1.

Secondly, all empirical studies, published to date, have neglected to study other aspects of supply structure; for example, the distribution of capacities, road widths, numbers of lanes and speed limits across network links, structural patterns in the road hierarchy for an urban area and the distribution of junction types. These structural features influence traffic flow but their structural characteristics remain unknown.

Thirdly, and finally, each study used a different selection of network measures to characterise structure in a unique sample of cities. This means that there are very few cases where the *same* network measure was applied to different datasets. With such limited crossover between studies, it cannot be concluded that all urban road traffic networks share all of the properties described in sections 2.3.1.2 and 2.3.1.3. Thus far, empirical studies of network structure appear to represent the application of lots of different measures to lots of different datasets without a clear goal. This was a criticism of other empirical studies in network science that was highlighted in section 2.2.2.2.

There is therefore a need for further empirical studies of the structural properties of urban road traffic networks, which should aim to address the above methodological inconsistencies, using an approach that focuses on the analysis of structure with traffic flow applications in mind, and which uses a broad selection of network measures to study a large sample of urban areas. This is an important goal for future work.

2.3.1.4.1 Perspectives from Transport Geography

Sections 2.3.1.2 and 2.3.1.3 have shown that network science has taken a quantitative approach to the characterisation of structure in urban road traffic networks. However, this is not the only approach. Researchers in geography, spatial science and urban planning have also made contributions to the categorisation of road pattern types. Typically these approaches place much greater emphasis on the development of urban form as being a result of a complex combination of social, environmental, technological and economic factors of which transport influences are only one part of the story. However, transport is still recognised as an important influence; indeed, Pacione (2005) highlighted the relationship between transport and urban structure, showing how the earliest cities had high densities because walking and horses provided the main modes of transport, that this was followed by the streetcar/tram era in which urban growth occurred along transit corridors, and that this was then followed by the car era with lower density development and urban sprawl.

It would appear that approaches to the categorisation of road network structures in geography are based more upon qualitative descriptions of patterns rather than numerical measures. For example, Marshall (2005) lists thirty-two different categorisation systems that have been proposed by geographers in the last century to classify road pattern types in urban areas. One of the most prominent messages that was conveyed by Marshall (2005) is of how difficult it is to classify structure within urban road traffic networks and that networks are often a mishmash of many different patterns that have developed over a long timeframe. Indeed, Marshall (2005) stated that several geographers consider the endeavour to be *"futile"* or *"impossible"*, and quoted Hanson (1989) as stating that:

"Time and time again, authors suggest that all towns are made up of a limited vocabulary of urban forms, yet when called upon to specify the elements of that vocabulary, the temptation to multiply categories seems to be irresistible."

Marshall (2005) highlighted that the establishment of a typology *"depends crucially on the purpose of that typology"*, i.e. that the reason for wanting to characterise the structural properties of networks should help to define the methods used in doing so. Recalling the

quote of Newman (2003) presented at the end of section 2.2.2.2, this is something that is conspicuously lacking from existing studies in network science.

In the context of the research question and objectives presented in the introductory chapter, this thesis argues that one such purpose could be to focus on the characterisation of structure within the context of studying the performance characteristics of traffic flows on road networks. Such a purpose has implications for the approach used. It could mean that empirical studies focus on the most heavily trafficked roads in their datasets instead of including every single road in an urban area. This approach acknowledges that roads in networks are of differing to traffic. For example, roads in residential areas could be omitted from such studies because they are lowly trafficked with the focus instead being placed on the structural characteristics of major roads and distributer roads. Demand data could also be used to determine the boundary of the network considered for each urban area. It could be that some urban areas draw traffic from a wider geographical area than other urban areas. Such approaches would require the simultaneous consideration of traffic flow data alongside supply data. An additional suggestion is that empirical studies could look to separate urban road traffic networks into different layers and study their properties independently. For example, there could be layers representing primary roads, such as A roads in the UK or freeways in the USA, and other local roads.

2.3.2 Generative Models of the Supply Structure of Road Traffic Networks

As with research for general networked systems, empirical analyses of the structural properties of road traffic networks have been followed by the development of models to represent their formation and growth; with the aim being to uncover the key mechanisms that have led to the creation of networks observed in the real world. In this section two recent examples of models from network science that have been proposed for the formation of road traffic networks are presented and briefly critiqued.

The two models that are highlighted here are similar in construction to the scale-free model of Barabasi and Albert (1999) in that nodes are added in a sequential process and are then connected to the network over a series of time steps rather than starting out with a predefined distribution of nodes in space. These models accord with the way in which the development of transport networks is viewed in geography as the result of an evolutionary process over a long time frame (Levinson, 2005, Xie and Levinson, 2009).

In the first model, proposed by Barthelemy and Flammini (2009), network growth is modelled using a local optimality principle where the addition of new centres at each time step stimulates the growth of new roads from the closest available points that are visible to the

new node. This latter condition ensures that the planarity of the network is retained. To encourage the formation of cycles these new centres are allowed to stimulate the growth of more than one new road at a time. Two variants of the model are presented.

In the first variant of this model, new nodes appear in way that creates a uniform distribution across a plane. The model is able to reproduce empirically observed structural properties of real road traffic networks such as constrained average node degrees. An example of the sequential generation of a road traffic network using this model is shown in Figure 2.11.

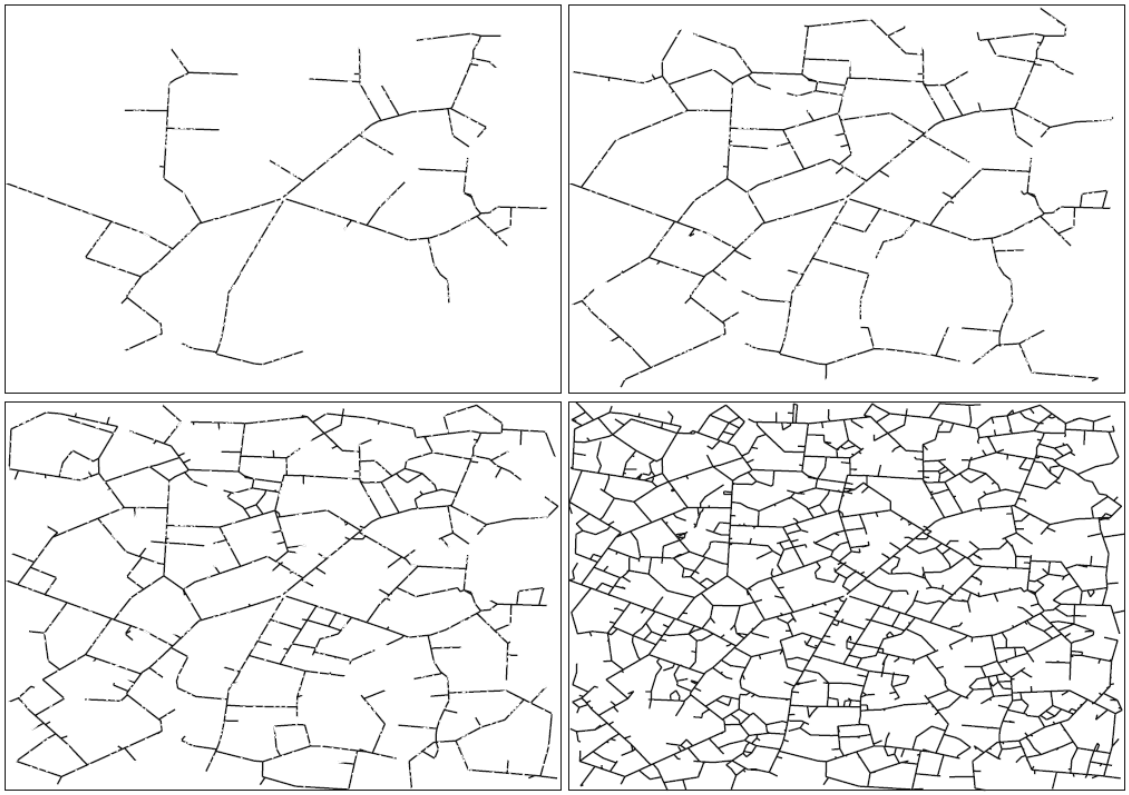

Figure 2.11 - Evolution of the first variant of the model of Barthelemy and Flammini (2009) over four time steps (Figure 3, Barthelemy and Flammini (2009))

Changing how nodes appear in the plane such that the concentration decreases as distance from the centre increases induces networks to be created with cell size distributions that are similar to results that have been observed empirically by Lammer et al. (2006). They also model the effect of a river on where nodes are positioned, finding that the resulting network creates bridges equally spaced along the river. Examples of networks produced using these different node placement rules are shown in Figure 2.12. The visual similarity of these networks to what one would expect in a real city is clearly evident.

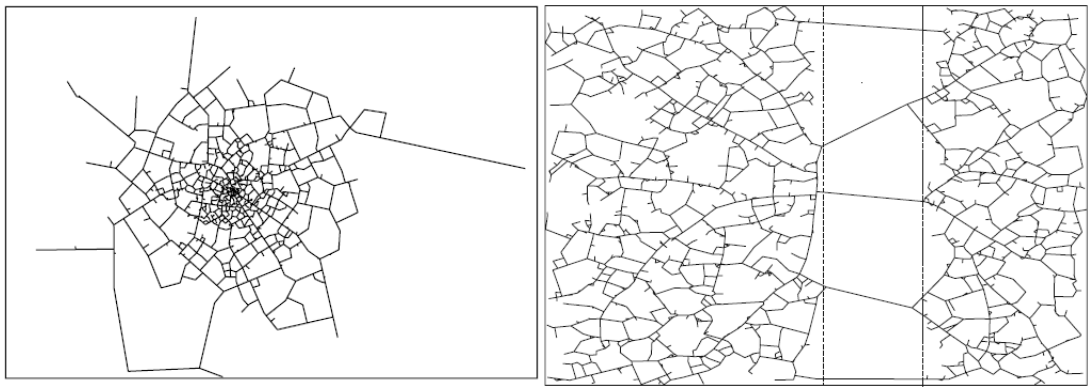

Figure 2.12 - Networks produced by the first variant of the model of Barthelemy and Flammini (2009) for the case of a non-uniform node distribution (left) and the existence of a river (right) (Figures 5 and 6, Barthelemy and Flammini (2009))

In the second variant of the model, Barthelemy and Flammini (2009) refined the node placement mechanism so that new nodes are placed as a function of population density and transport accessibility. To achieve this, they divide the plane into several square sectors and then calculate the probability of the addition of a new centre in each one of these square sectors based on accessibility and rent price. Rent price is modelled as a strictly increasing function of the number of nodes already placed in each sector; the higher the number of nodes, the higher the rent. Accessibility is measured using the average betweenness centrality of the nodes within each sector; the higher the betweenness centrality, the lower transportation costs are in that sector. The probability for a new node to appear within a given sector is then calculated using a logit model with utilities based on the total sum of rental and transport costs. A parameter λ is used to adjust the weights of rental and transport costs in the calculation of utilities; such that when λ is small, rental costs are of greater influence than transport costs, whereas when λ is large, the opposite occurs.

The effects of this on the networks that are produced by this model are shown in Figure 2.13. When λ is small, a uniform distribution of nodes is recovered, which produces results like the first variant of the model. When λ is large however, the network becomes concentrated around few centres.

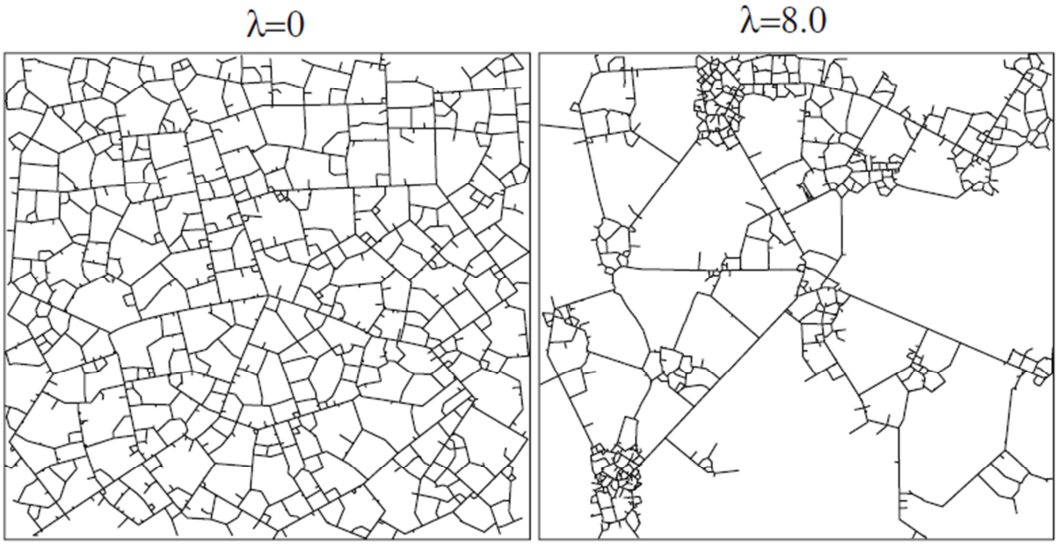

Figure 2.13 - Networks produced by the second variant of the model of Barthelemy and Flammini (2009) (Figure 11, Barthelemy and Flammini (2009))

The second model that is referenced in this section is that of Courtat et al. (2011). This model uses similar principles to the model of Barthelemy and Flammini (2009) but is significantly more complex and encompasses significantly more variables. Like the model above, network generation is modelled as an evoluationary process over time with the addition of new nodes to a domain, which are then connected to existing nodes in the network through the construction of links. The difference between the two models lies in the additional complexity of the rules by which these two steps occur with both including several parameters that can be adjusted to affect the shape and pattern of the resulting network.

The appearance of a new node is governed by a *"potential field"*, which, given the existing network, controls the probablity of where the new node will appear. The settings of the potential field can be adjusted to control how close new nodes are positioned relative to the existing network. Once a new node has been generated and positioned, the set of *"visible"* existing nodes is identified; that is, the set of nodes for which a straight line segment connecting to the new node could be added without violating the planarity of the network. Controlled by yet more parameters, the new node is then connected to between one (the closest) and all of the nodes in this set. The parameters provide a great amount of flexibility in the model and lead to many different visual structures of network. The varying effects of two of these parameters are shown in Figure 2.14 below.

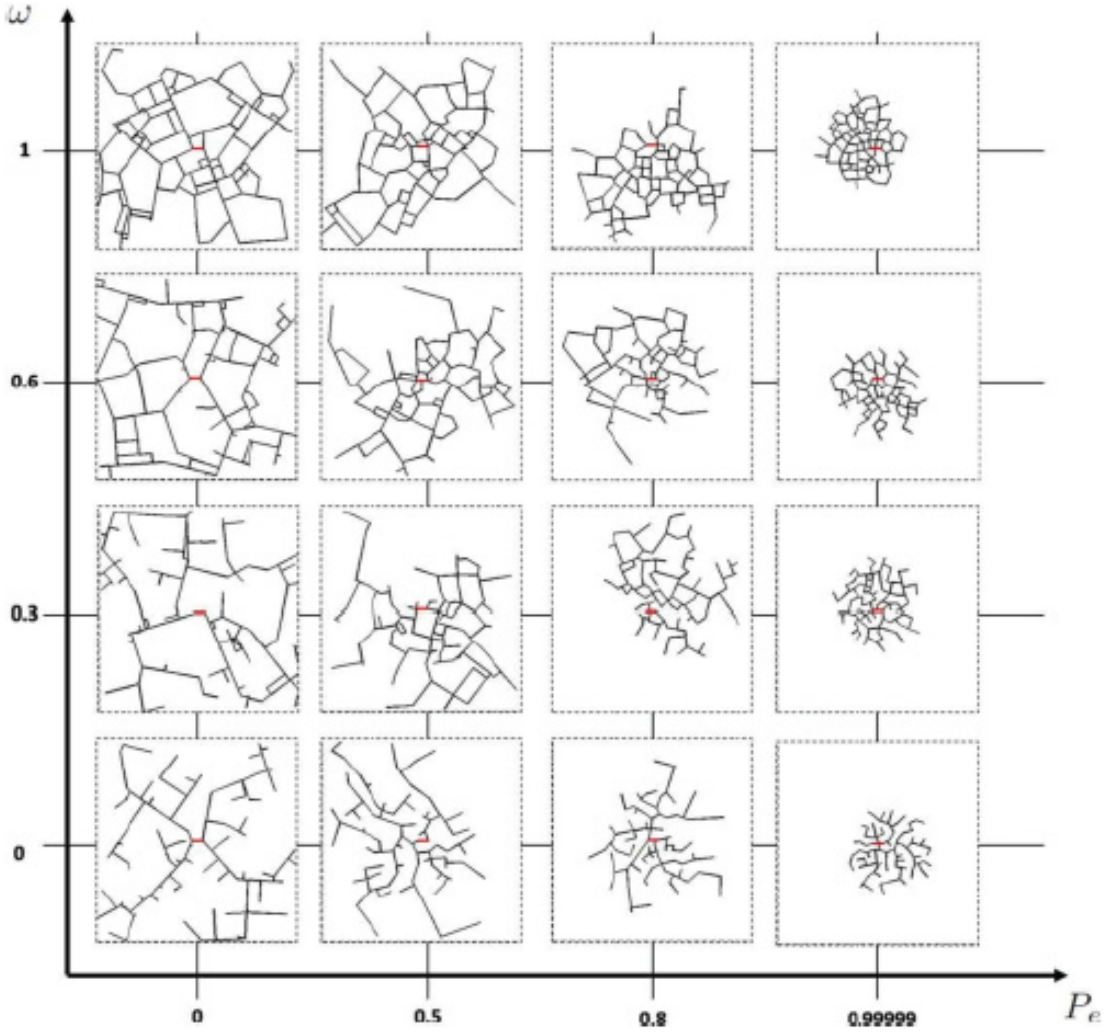

Figure 2.14 - Example networks produced by the model of Courtat et al. (2011) (Figure 9, Courtat et al. (2011))

The P_e term on the x-axis controls how strictly new nodes appear according to the potential field. High values indicate a high level of organisation and therefore uniform spread of new centres whilst low values lead to the random positioning of new centres. The ω term on the y-axis controls how many connections are made between new centres and the existing network. As can be seen low values of ω result in treelike structures whilst high values produce networks with many loops. An example of the evolution of a road network generated by this model over five steps ((a) to (e)) is shown in Figure 2.15.

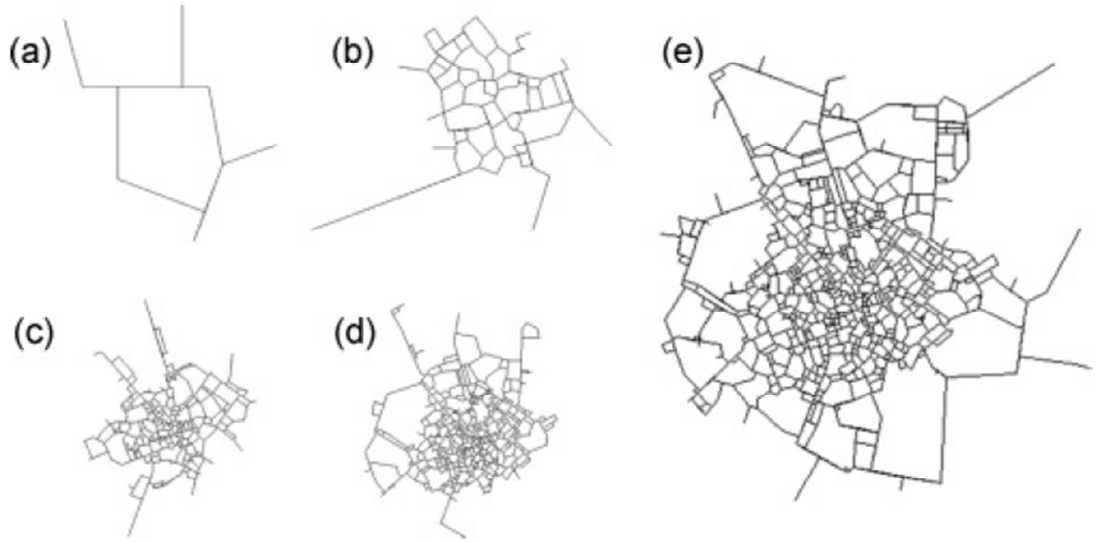

Figure 2.15 - Example evolution of transport networks produced by the model of Courtat et al. (2011) (Figure 11, Courtat et al. (2011))

The networks produced by both of the above models, as shown across Figure 2.11 to Figure 2.15, are certainly impressive and do appear to produce networks that resemble real urban road networks. However, there are several areas for development:

Firstly, whilst these models are clearly able to produce networks that are visually similar to real urban road networks, there has been only a limited amount of empirical analysis to determine whether the networks produced are really representative of such networks and whether they share characteristics such as those identified in section 2.3.1.

Secondly, with respect to the evolutionary behaviour of the networks shown across the snapshots in Figure 2.12 and Figure 2.15, the rate of network growth appears to be constant over time. It is known in transport geography that urban road networks do not have constant growth rates and that real networks go through three stages: *"birth"*, where the growth rate is low as only a few important links are built; a *"growth"* stage, in which the construction of the network accelerates, and a *"maturity"* stage, in which the network reaches saturation and slows down (Levinson, 2005). Neither of the above models includes such a mechanism for varying the growth rate; and, in particular, the idea of a point of saturation where the model reaches a natural conclusion.

On a related issue, there is also no evidence for either model to show whether the order in which links are constructed as the network grows reflects the pattern of growth that has occurred in the development of real networks. This is hard to verify because datasets of network evolution over long timescales are hard to find but the recent work of Strano et al. (2012) on the evolution of the road network of the Groane area, to the north of Milan, over

the last two centuries is illuminating. Using seven snapshots of the road network between 1833 and 2007, they found that network growth was governed by two elementary processes: *"exploration"*; in which new roads branch out into new areas, and *"densification"*; in which roads between these branches are filled in and there is an increase in the density of roads. The models highlighted above appear to include the first but not the second of these behaviours.

The models also develop road networks independently of the world around them, i.e. they generally exclude potential influences of topography, although Barthelemy and Flammini (2009) do simulate the effect of a river with visually representative results, but also the effects of other cities and towns in the wider geographical area, which will affect the appearance of main routes through a city. Courtat et al. (2011) touch on this in a final example but it would be interesting to see further work on the impact of such influences and whether these models continue to produce realistic structures compared to the real-world. Widening the scope of this point, there are also other external influences on road networks. Xie and Levinson (2009) highlight that *"transport development represents a complex and dynamic process that involves a magnitude of dimensions, which may be topological, morphological, technical, economic, managerial, social or political."* Although many of these may be difficult to represent within a modelling framework.

A final criticism of these models is that they include only the growth and formation of the topological structure of the road network, and do not include other aspects of supply structure, such as a road hierarchy, or a representation of how population and travel demand has evolved with the network over time. With respect to the development of a road hierarchy, the work of Levinson and Yerra (2006) is particularly interesting as they were able to show, albeit only for a grid network, that a hierarchy of major and minor roads can appear as the result of a *"decentralised process"* without the direction of human design. With respect to the inclusion of a representation of population and travel demand, many geographical approaches focus on the dynamics of population within urban settlements; for example by Makse et al. (1995). An interesting future research area would be to see whether these approaches can be modelled together.

2.3.3 The Structure of Demand in Road Traffic Networks

The notion that travel demand in a road traffic network has structural properties refers to the existence of patterns in how travellers use road infrastructure. The form and characteristics of travel patterns are of interest to many different fields; for example, geographers and transport planners are interested in the spatial distribution of activities, epidemiologists are interested in modelling the spread of infectious diseases (e.g. Yashima and Sasaki (2014)) and, in a

commercial context, businesses are interested in where people travel so that they can most effectively place their advertisements (Barthelemy, 2011).

In comparison with studies of the structure of network supply, empirical studies of the structural characteristics of travel demand are more limited in their level of detail. This is primarily a reflection of the greater difficulty that exists in collecting data on travel demand. There are several contributory factors to this difficulty. Unlike road infrastructure, travel demand is constantly changing from hour-to-hour, day-to-day and over weeks, months and years. Demand also has many different dimensions; its properties can be studied for individual travel modes (e.g. private car, public transport or active modes), different traveller types/purposes (e.g. commuter, business or leisure) and, as will be shown, at different spatial scales. The challenges of collecting data for each of these dimensions vary in scope and difficulty.

Traditionally, data on the characteristics of travel demand have been collected using a combination of roadside interview surveys and household based travel diaries in which a sample of travellers are asked to provide details of their journeys. Such approaches are widely known to be prone to bias, human error and typically achieve only low sample rates. These disadvantages are compounded by the significant costs of such surveys. However, despite their known deficiencies, these methods are still commonly used by public bodies and transport consultancies today.

These traditional approaches contrast with new data collection methods that have been proposed and used in parts of the academic literature, which typically make much greater use of technology and so reduce the burden on members of the public to accurately report their travel habits. Examples include use of link traffic counts and routing information derived from GPS (Parry and Hazelton, 2012) or number plate surveys (Castillo et al., 2008) to derive travel demand patterns across a geographical area. In more recent years, there has also been significant growth in the number of papers that study the characteristics of travel demand using mobile phone datasets, for example, the variation in aggregate call volume data in mast coverage areas across a city (Sevtsuk and Ratti, 2010) or the records of calls made by individual phones across a day (Caceres et al., 2013, Iqbal et al., 2014, Kang et al., 2012). The latter datasets have been used to develop methods to derive data on movement patterns across urban areas.

The purpose of the subsections that follow is to summarise and critically review the key findings of existing studies of the structural properties of travel demand. These studies are categorised into three groups, which each focus on the characteristics of travel demand at different spatial scales. Section 2.3.3.1 describes studies of the broad characteristics of travel

demand at the aggregate level of overall populations. Section 2.3.3.2 then summarises research on the structure of travel movements within more limited geographical areas, such as between distracts or in a city, at the scale of *origin-destination matrices* (defined at the beginning of that section). Finally, section 2.3.3.3 focuses on studies of the distribution of travel demand across network links within an urban area and attempts that have been made to use structural measures, such as those identified in section 2.2.2, to characterise travel demand. Section 2.3.3.4 then provides a summary of main findings and critical review.

2.3.3.1 Broad Characteristics of Travel Demand

In the UK, data on the broad characteristics of travel demand across the country are collected via the National Travel Survey[9], which uses a series of face-to-face interviews and paper based travel diaries in which participants are asked to log their movements over a one week period. These data are used to present an overall statistical summary of travel habits with information presented, for example, to characterise the total volume of demand, distance travelled and trip durations segmented by mode and travel purpose. One-off studies at the scale of urban areas have also been undertaken along similar lines; for example, see Mokhtarian and Chen (2004) for a list of over fifty such surveys.

Analysis of such datasets led Zahavi (1977) to hypothesise that individual travellers within urban areas have a personal travel time budget; the idea being that an individual would typically travel only for so long for different trip purposes. Zahavi (1977) also proposed that this budget does not vary significantly over time or across different geographical regions (Barthelemy, 2011). In support of this hypothesis, Kolbl and Helbing (2003) found that between the years 1972 and 1998, the average daily travel time in the UK for a range of transport modes has been approximately constant. Levinson and Wu (2005) also found empirical evidence to suggest that travel budgets within cities are constant over time but also found variation in budgets between cities. In their analysis of over fifty surveys, Mokhtarian and Chen (2004) concluded that whilst such patterns do exist at the most aggregate scales, there is considerable variability in travel time budgets when travellers are disaggregated by socio-economic factors. Barthelemy (2011) concluded that these studies *"point to the possible existence of universal features of human movement"* but that further empirical and theoretical studies are required before a *"'unified theory' of human travel behaviour"* can be determined.

[9] https://www.gov.uk/government/collections/national-travel-survey-statistics

2.3.3.2 Urban and Interurban Mobility Patterns

Studies of the structural properties of travel movements within urban areas, such as towns and cities, or between different urban areas, such as between districts within a country, are typically undertaken through study of a mathematical construct called an *Origin-Destination (OD) matrix*. Within this representation of demand, the geographical region under study is subdivided into several disparate areas, called zones, and the volume of travel between all pairs of zones is then recorded. Zones typically represent areas that share some congruent internal properties such as a residential area or business park within an urban area or different administrative regions within a country. One relatively recent development is the recognition that an OD matrix can also be converted into a network representation in which nodes represent origin and destination zones, links represent the existence of traffic flow between two zones and link weights represent traffic flow volumes.

Empirical studies of the structural characteristics of OD matrices - many of which are from the network science literature - have been undertaken for both urban and interurban travel, for different modes of transport and across a wide range of different countries.

For example, with respect to the properties of interurban travel, De Montis et al. (2007) studied the structure of the demand network created by commuting movements between the 375 municipalities on the island of Sardinia and concluded that there is a *"rich-club phenomenon"* in which there are a small number of regions with high total traffic flows, which have busy connections between them, and a significant number of smaller regions that act as *"satellites"* of larger cities. Their conclusion was that these structural features created an *"overall network structure* [that is] *widely punctuated with star-like subsystems pivoting around important urban poles."* In their analysis of commuting patterns between districts in Germany, both Patuelli et al. (2007) and Reggiani et al. (2011) also uncovered heterogeneities in demand structure, with the latter highlighting the existence of twelve hubs that dominate in terms of traffic volume.

Focussing on urban travel patterns, Chowell et al. (2003) used an agent-based simulation model to simulate the movements of 1.6 million people in the city of Portland in the USA. The model was calibrated using census data, vehicle ownership records, public transport timetables and information about travel movements from a travel survey undertaken in the city. Using the network created by these movements, the authors undertook a structural analysis using node degree and the clustering coefficient measures and were able to uncover power laws in the distributions of traffic. The uncovering of these laws again highlights strong heterogeneities in the movements of travellers within OD matrices. Similar heterogeneity in the distribution of travel movements across a city were also uncovered by Gao et al. (2013),

who used a dataset of seventy-four million mobile phone call records for the city of Harbin in northeast China.

Turning to a study undertaken within the transportation research community, Gutierrez and Garcia-Palomares (2007) analysed differences in travel patterns in Madrid between 1988 and 1996, using data from mobility surveys carried out in those years, and found that there had been a significant shift in the structure of Madrid from a monocentric organisation, where most trips are to a dominant city centre, to a polycentric organisation, with increased suburb to suburb travel. Their assertion was that this is a result of a process of decentralisation in employment. Using oyster card data, Roth et al. (2011) uncovered a similar polycentric structure of demand in London and identified multiple centres that both attract and generate large amounts of flow at different times of day, on different days and across different weeks.

All of these studies highlight a trend towards an increasingly complex, heterogeneous distribution of demand in OD matrices at the level of both urban and interurban travel. Given the existence of such patterns, it is unsurprising that several models have been proposed for characterising the distribution of travel demand. The most well-known model is that based on the gravity law and which posits that the volume of travel between two zones is proportional to the travel populations within those two zones and inversely proportional to the cost of travel or distance between them. Several forms of gravity based models are used in practice, which each use different functional forms to represent the deterrence of travel costs. Both Patuelli et al. (2007) and Reggiani et al. (2011) attempted to fit gravity models to explain the patterns uncovered in their empirical analyses but were unable to successfully fit either an exponential or power form deterrence function. A more successful attempt at fitting a gravity law model was presented by Jung et al. (2008), who analysed the network created by the movements of traffic on the interurban highway network of South Korea. Using a dataset comprising total movements between the top thirty cities (by population) as recorded by the toll plazas sited at all entry and exit points to the network, they successfully fitted a gravity law with a power law form deterrence function.

It is worth noting that several other models for travel demand have also been proposed; for example, the intervening opportunities model, which proposes that *"the number of persons going a given distance is directly proportional to the number of opportunities at that distance and inversely proportional to the number of intervening opportunities"* (Stouffer, 1940); and, more recently, the radiation model, which focuses on commuting flows by way of modelling how individuals accept job offers with respect to benefits and distance (Simini et al., 2012). The latter model, which has the advantage of being parameter free, has been shown to produce better estimates of travel demand patterns than the gravity model (Simini et al.,

2012), although not for large cities like London (Masucci et al., 2013). Masucci et al. (2013) argue that *"commuting at the city scale still lacks a valid model and that further research is required to understand the mechanism behind urban mobility"*.

2.3.3.3 The Distribution of Travel Demand on Network Links

At the finest spatial scale, several studies have focussed on the structural characteristics of demand volumes and how they are distributed across individual links within an urban area.

For example, Jiang (2009) investigated the assertion, quoted in section 2.3.1.3, by Jiang (2007) that *"a minority of streets account for a majority of traffic flow"* by studying traffic flows in the city of Gavle, Sweden. This investigation used GPS data from the movements of taxi cabs during one week in October 2007 as a proxy for the distribution of travel demand on network links, and then used the dual representation, described in section 2.3.1.3, to create a representation of streets in the city. By calculating the lengths of these streets, the authors found that the top 20% of streets by length carry 80% of the traffic flow. Whilst there are question marks over the methodological approach taken in this study - for example, of whether taxi drivers are a good representation of the broad pattern of travel demand given the greater network knowledge they are likely to have and their different travel habits, and of use of the dual representation - it seems plausible that, in general, there would be a large amount of variation between the volumes of traffic using different roads within an urban area. (This adds further weight to the argument, put forward in section 2.3.1.4.1, that studies of supply structure should consider that different roads are of different importance in the context of traffic flows).

Related to such findings, several papers within the network science literature have attempted to draw correlations between structural measures and traffic flows on network links, with several pointing towards the betweenness centrality measure, defined in section 2.2.2.2, as a natural candidate for such tests. Indeed, Lammer et al. (2006) studied the distribution of betweenness centrality on the road network of Dresden, Germany, and found that it follows a power law. This led to a conclusion that *"high values of* [betweenness] *can be interpreted as a high concentration of traffic on the most important intersections."* Barthelemy (2011) went further by stating that *"the betweenness centrality is in itself interesting since it points out the important zones which potentially are congested."*

There have been several attempts to find empirical evidence to support the use of structural measures, such as betweenness centrality, as representations of travel demand. For example, Kurant and Thiran (2006a) and Kurant and Thiran (2006b) studied the explanatory power of several topological indicators for demand flows across three network examples: the public

transport network of Warsaw, Poland; the railway network of Switzerland and the railway network formed by major trains and stations in Europe. In each representation, the number of services per day at each network node were used as a proxy for demand flows. The first measures for which correlations were sought were node degree and betweenness centrality but both provided very low statistical correlations. Two adjusted versions of the betweenness measure were then used in attempts to increase the correlation by adding additional network data: a *"restricted betweenness"* measure, in which the summation in the betweenness measure was restricted to movements between nodes that act as the termination point of services; and a measure called *"simple load"*, in which the summation was further restricted to movements between node pairs for which there existed a service between them. It was only this final measure that was able to provide a *"rough approximation"* for traffic flows. Moving to evidence from road traffic networks, Bono et al. (2010) used traffic flow data extracted from the TeleAtlas MultiNet dataset and the UK Department for Transport (DfT) website, plus freely available data on the internet, to analyse correlations between two structural measures and traffic flows on the road networks of London, Manchester and Birmingham in the UK, and Forsyth County in the USA. In testing three versions of each measure, it was found that structural measures provided a good representation of demand flows only when additional network information, such as travel times, was included as a weight.

It is unsurprising that pure topological measures, such as the betweenness centrality, provide poor representations of traffic flows. By definition, the measure relies upon finding shortest paths between all pairs of nodes in a network and so implicitly assumes a uniform distribution with one unit of demand between all node pairs, which contrasts with the heterogeneous structure of demand identified by the studies in section 2.3.3.2. These measures, and the studies that use them, also ignore that travellers often use multiple routes between OD pairs because of congestion effects, and that the distribution of travel demand varies over time. These features make the distribution of traffic flow in a road network significantly different to, for example, the distribution of water flow in a river system.

2.3.3.4 Summary and Critical Review

The literature cited in sections 2.3.3.1 to 2.3.3.3 demonstrates that whilst there is some evidence for general patterns and aggregate features of human mobility, there is also a considerable amount of complexity in demand patterns. The empirical evidence highlighted in section 2.3.3.2 in particular demonstrates that the distribution of travel demand is highly heterogeneous between different urban areas, within urban areas and with respect to traffic volumes on network links. However, it is notable that no attention has been given to the pattern of travel demand for road traffic networks in urban areas. The focus of studies has

instead been on broader spatial scales or on the distribution of demand for other travel modes; perhaps for which data are more easily accessible. Whilst there is no evidence to suggest that the distribution of demand within urban road traffic networks should be dramatically different to the heterogeneous distribution that has been uncovered, further empirical studies are required to test this assertion.

It is worth noting here that transportation planning theory recognises that travel demand is endogenous and that it does depend on the distribution of land uses within an urban area and on the availability of network supply. The structural characteristics of demand evolve over time, with network supply, as a result of decisions taken by individuals, companies and network operators (Ortúzar and Willumsen, 2001). Travellers can respond in the short-term by varying their departure time, route choice or trip frequency, and in the long-term in car purchase decisions, or by changing their residential location or place of work. Network operators responses include decisions, for example, to build new roads or introduce traffic management initiatives. The structure of travel demand is therefore intricately interwoven with the structure of network supply. However, to the knowledge of the author, there is currently no empirical research of whether there is a systematic relationship between the distribution of travel demand and network structure, or the nature of that relationship.

2.4 The Effects of Structure on the Performance Characteristics of Road Traffic Networks

This section describes studies of the effects of structural properties of supply and demand - such as those described in the previous section - on the performance characteristics of road traffic networks. Section 2.4.1 describes how the performance characteristics of road traffic networks can be measured and section 2.4.2 describes how traffic flows are typically modelled. Sections 2.4.3 and 2.4.4 then describe contributions from the transportation and network science communities, which have used such measures to study network performance and how it is affected by network structure.

2.4.1 Measures of the Performance Characteristics of Road Traffic Networks

As stated in the introductory chapter, in the context of road traffic networks, performance refers to *how well* a network fulfils its function of providing for the movements of travellers. The notion that networks can have performance characteristics highlights how the concept of performance has many facets - of which efficiency, reliability and vulnerability are some examples - which can each be quantified by a wide range of measures and methods. This section introduces a selection of these measures, focussing on those that are referred to in later sections, and also highlights how they are commonly used.

The majority of the simplest and most commonly used performance measures are essentially variations in the use of data on traffic flows on network links and at nodes, travel costs/times between nodes and distances travelled. This data can be collected for individual nodes and links, or can be aggregated to various levels of detail, for example, to denote travel times between OD pairs on individual routes, average link speeds across all network links, or the total distance travelled by all travellers in a network. These measures are similar to those used by the references cited in section 2.2.5, which study the performance of packet transmission models. In order to refer to the above data items, this thesis will use the following notation. The volumes of traffic flow on a link i and on a path k between the rth OD pair in a network will be denoted by x_i and f_k^r respectively. The costs of travel on a link i and path k between the rth OD pair will then be denoted by c_i and C_k^r respectively.

In the specific context of road traffic networks, performance measures can also be further aggregated by time period (across one or more hours, a day, a week or a year) or disaggregated by user class (e.g. split up by cars, heavy goods vehicles, etc.) or trip purpose (e.g. commuters, business travellers, etc.). Measures can also be mixed with supply characteristics; for example, a commonly used indicator of congestion is the link volume-to-capacity ratio, denoted V/C ratio, which measures the amount of flow on a link as a proportion of the capacity of a link.

Many of these measures are regularly reported for groups of roads or individual urban areas by public authorities in many countries. For example, the Department for Transport in the UK regularly reports on the average speeds on 'A' roads in weekday morning peak periods[10] and on the delays in travel time experienced by travellers on the strategic road network[11]. In the USA, the Texas Transportation Institute publishes an annual urban mobility report that reports on total yearly delays experienced by travellers in one-hundred and one urban areas across the country (Schrank et al., 2012).

As will be described in the next section, many of the above measures are regularly used by public authorities and by commercial organisations.

Building on the simple measures outlined above, the academic research community has published many papers that have proposed more complex measures and methods to quantify

---

10 https://www.gov.uk/government/statistics/congestion-on-local-a-roads-england-jan-to-mar-2014

11 https://www.gov.uk/government/publications/reliability-of-journeys-on-the-highways-agency-s-motorway-and-a-road-network

the performance characteristics of road traffic networks. For example, focussing specifically on network efficiency, Nagurney and Qiang (2007) proposed the following measure:

$$\varepsilon = \frac{\sum_r \frac{q_r}{C_r}}{n_r}$$

in which q_r represents the total demand and C_r denotes the cost of travel for the rth OD pair and n_r denotes the number of OD movements in the network. This quantity represents the average throughput of the network per unit of cost. Another measure of network efficiency, which instead focuses on the efficiency of route selections by individual travellers, is the Price of Anarchy. This measure was proposed by Koutsoupias and Papadimitriou (1999) and Papadimitriou (2001) and is calculated as the ratio of the Total Travel Cost, denoted TTC, when individual travellers selfishly choose routes to minimise their individual travel costs, to the TTC when individual travellers instead choose routes such that the TTC across all travellers is minimised. These route choice concepts are more commonly referred to as the User Equilibrium (UE) and System Optimal (SO) routing principles respectively. A formula for the Price of Anarchy is shown in equation (1).

$$\rho = \frac{TTC^{UE}}{TTC^{SO}} = \frac{\sum_i x_i^{UE} c_i\left(x_i^{UE}\right)}{\sum_i x_i^{SO} c_i\left(x_i^{SO}\right)} \quad (1)$$

The application of network measures, such as those described above, requires data on traffic flows and travel costs. This data can either be obtained through observations of real road traffic networks or extracted from computer models of road traffic networks. One of the most commonly used approaches to modelling road traffic - especially in commercial contexts - are traffic equilibrium modelling techniques. As this modelling approach is used in papers described in later sections of this literature review, a brief outline of these techniques and how they are applied is provided in section 2.4.2, which is based on Sheffi (1985) and Ortúzar and Willumsen (2001). A more thorough description of notation and the mathematical conditions underpinning these techniques is provided in section 4.4.

2.4.2 Modelling Road Traffic Networks

Most transport models provide a snapshot of a limited number of aspects of a traffic network; for example the transportation characteristics of a transport system in a specific city, on an average weekday AM peak period in 2014, for road passenger traffic. In modelling transport, it is normal to focus on those dimensions that are most important; for example, peak hour models are often used because the main interest is in the performance of the transport system when it is under the most pressure.

The classic transport planning model has four stages: trip generation, which quantifies productions and attractions for zones within the model area based on, for example, population and economic activity; trip distribution, in which departures from each zone are connected to arrivals in other zones using a gravity model, for example; mode choice, in which each trip is allocated to a particular mode and traffic assignment; in which trips in an OD matrix are loaded onto network links.

The assignment of travel demand to network links is most often based on the principle of equilibrium. This principle of equilibrium states that if this balance is disturbed then market forces will naturally force them back into equilibrium. So, for example, if the demand for a product is higher than the supply then producers will increase prices and/or production to take advantage of this fact, whereas if the demand for a product is lower than the supply then producers will drop prices or reduce production.

In road traffic networks, supply and demand are characterised by the layout and the level of service of the roads and junctions in a network and the magnitude of travel demand between different points in that network. As in the general case above, these factors interact with each other. For example if, when aggregated, a significant number of travellers use a particular stretch of road then the level of service on that road will drop, which may consequently encourage some travellers to seek alternative routes. The interaction between these congestion effects and traveller decisions forms continuous feedback loops. A point of equilibrium occurs where a balance, by some definition, has been achieved between these congestion effects and traveller decisions.

Wardrop (1952) defines two principles of route choice that have a mathematical formulation that yields such a point of equilibrium:

1. Individual travellers choose routes such that they each, selfishly, minimise their individual travel time (or cost). This is called the User Equilibrium (UE) principle.
2. Individual travellers each, unselfishly, choose routes such that the total travel time (cost) in a network when aggregated across all travellers is minimised. This is called the System Optimal (SO) principle.

When applying these techniques, travel demand is aggregated in an OD matrix and the level of service provided by each link is represented by a *link performance function*. For a link i these link performance functions describe how travel time t_i (in general, travel cost c_i) varies in relation to the volume of flow x_i on the link, and sometimes also traffic flows on adjacent links. A typical visualisation of such a function is provided in Figure 2.16.

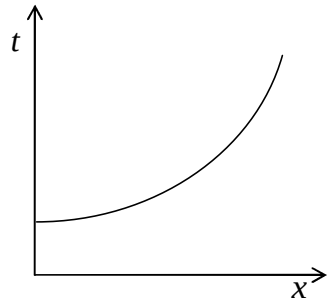

Figure 2.16 - Example of a link cost function used in transport models

As illustrated by Figure 2.16, in order to maintain physical realism these link performance functions are assumed to be continuous, positive and strictly increasing functions of the flow on the individual link that they represent, i.e. $t_a = t_a(x_a)$. The dependence of travel time on link flow allows the effects of congestion to be included. The most commonly used link performance functions typically take the form of polynomials with positive coefficients; for example, the BPR cost function (Bureau of Public Roads, 1964). However, other functions are possible. The BPR function has been criticised because it tends to *"underestimate delays at junctions"* and also *"when demand is close or above the capacity of the link"* (Ortúzar and Willumsen, 2001). This makes such cost functions less appropriate in urban areas where junctions are the major determinant of travel times.

The equilibrium modelling approach also has other features that make it unrealistic. For example, it assumes that when traffic is assigned, it simultaneously appears on every link on its route through the network. This is known as the steady state assumption. In reality however, traffic is dynamic and changes over time. There are other traffic modelling approaches, such as dynamic traffic assignment and microscopic simulation models, which include more realistic features of traffic flow but also typically include many more parameters and therefore require much more input from data. However, having highlighted these criticisms, equilibrium models continue to be well used outside academia and are generally considered to be effective and reasonable representations of traffic flows.

2.4.3 Studies of the Effects of Network Structure on Performance in Transportation

This section describes the contributions of the transportation community to studying the effects of network structure on performance in road traffic networks. Section 2.4.3.1 focuses on empirical studies of network performance whereas section 2.4.3.2 describes theoretical studies.

2.4.3.1 Empirical Studies of the Performance Characteristics of Road Traffic Networks

In general, most empirical studies of the performance characteristics of road traffic networks are more commonly undertaken by transport consultancies than by the transportation research community. These studies are usually individual case studies that focus on the effects of proposed supply side or demand side interventions, such as the construction of a bypass or imposition of a congestion charging scheme. By focussing on isolated individual case studies, these studies reveal very little about how global structural properties of supply or demand in road traffic networks affect network performance.

Although fewer in number, there are also examples of individual case studies in the transportation research literature. For example, Tsekeris and Geroliminis (2013) studied a concentric city model - which provides an idealised representation of an urban area - and found that an increase in compactness, as travel demand increases, maximised efficiency with respect to congestion. A further example is provided by Ortigosa and Menendez (2014) who studied the effects of the removal of links from a grid network and found that a strategy of link removals from the geometric centre of a network is the most detrimental to performance. Again, these studies reveal little about how network structure affects network performance in general.

Two notable exceptions to this general picture of individual case studies are provided by Levinson (2012) and Parthasarathi et al. (2012). These papers used network and travel data for fifty cities in the USA to study correlations, via regression models, between several measures of network structure, including the α, β and γ indices and the route factor measure defined in section 2.2.2.2, and urban mobility indicators. Both papers derived measures of network structure using network data from the TIGER database, referred to in section 2.3.1. Levinson (2012) also used performance indicators from the Texas Transportation Institutes annual urban mobility report, referenced in section 2.4.1. Levinson (2012) found that *"larger cities have more delay, longer commutes and less travel per person"* and also have *"more connected road networks, … are more accessible, and are less hierarchical."* Parthasarathi et al. (2012) found that aspects of network structure that lead to increased network travel distances, such as the route factor, lead to reduced actual travel distances by travellers. They also found that street density is negatively correlated with actual trip lengths but also that the proportion of highways in a road traffic network is positively correlated with actual trip lengths. Whilst both of these papers present interesting links between network structure and performance characteristics, it should be noted that the R-squared values in their regression models are particularly small; this is especially true for the results of Parthasarathi et al. (2012).

2.4.3.2 Theoretical Studies of the Performance Characteristics of Road Traffic Networks

Whilst the transportation research community has made very few empirical contributions to the study of how performance characteristics vary with network structure, several transportation researchers have made theoretical contributions with respect to producing upper bounds for one particular measure of network performance; the Price of Anarchy.

This focus on *worst-case* values of the Price of Anarchy appears to be particularly interesting to transportation researchers. For example, it has been used to establish upper bounds for the maximum efficiency gains of road pricing (Han and Yang, 2008, Yang et al., 2010) and car number plate based traffic rationing schemes (Han et al., 2010). The Price of Anarchy has also been used to establish upper bounds on the maximum efficiency loss in competitions between providers of private road infrastructure (Liu et al., 2011, Xiao et al., 2007), in traffic networks where only a minority of travellers have access to advanced traveller information (Liu et al., 2007), and in traffic networks where some travellers choose to follow shortest paths, oblivious to the effects of congestion (Karakostas et al., 2011).

The first upper bound for the Price of Anarchy was produced by Roughgarden and Tardos (2002), who demonstrated that the Price of Anarchy has a maximum value of $4/3$ in traffic networks with affine link cost functions. Generalisations and extensions of this result have since followed to families of traffic networks with separable, polynomial link costs (Roughgarden, 2003, Dumrauf and Gairing, 2006); non-separable, symmetric costs (Chau and Sim, 2003); and non-separable, asymmetric costs (Perakis, 2007). Upper bounds have also been produced in the context of elastic demand assignment for traffic networks with non-separable, symmetric cost maps (Chau and Sim, 2003); and non-separable, asymmetric and non-linear costs (Han et al., 2008).

In each of the above instances, the upper bounds that have been presented depend only on characteristics of the cost functions, such as the value of the highest power across all network links or the degree of link cost asymmetry. For example, for traffic networks with separable, polynomial link costs, Roughgarden (2003) showed that the Price of Anarchy is bounded above by equation (2), where p is the value of the highest power across all network links.

$$\left[1 - p(p+1)^{-(p+1)/p}\right]^{-1} \tag{2}$$

Upper bounds for the Price of Anarchy that have been produced more recently include characteristics of demand. For example, for traffic networks with separable, polynomial link costs, Correa et al. (2008) showed that tighter upper bounds than those presented by Roughgarden (2003) could be derived provided the free-flow travel cost on each network link is at least a non-zero, fixed proportion of its travel cost under a UE assignment of travel

demand. Englert et al. (2010) also showed that the maximum increase in the Price of Anarchy, due to an increase in demand, could be bounded for traffic networks with separable, polynomial link costs and a single OD pair.

These studies show that the Price of Anarchy has a maximum value across broad families of networks. However, in focussing on the worst-case value of the Price of Anarchy *across* broad families of road traffic networks, the above studies neglect the variation that occurs *within* families of traffic networks, between road traffic networks that may have very different demand and supply structures. Evidence of this variation is revealed by numerical studies in network science, which are the subject of the section that follows.

2.4.4 Studies of the Effects of Network Structure on Performance in Network Science

Studies in network science of the effects of network structure on performance in road traffic networks have, thus far, focussed exclusively on the urban context, and appear to be inspired by and have used a similar methodology to studies of the performance characteristics of the Internet, which were described in section 2.2.5.

The approach taken by most of these studies has been to use canonical models from the network science literature to generate large ensembles of synthetic networks, each of which have different structural characteristics. Frequently used models include the random graph model of Erdös and Rényi (1959), the scale-free network model of Barabasi and Albert (1999) and the small-world network model of Watts and Strogatz (1998); for which example networks were shown in section 2.2.3. The performance characteristics of these different ensembles of networks are then compared, under an assignment of travel demand using the modelling approach described in 2.4.2, in order to determine which ensemble has the best performance on average, when taken across all networks from within each ensemble.

Table 2.4 describes seven numerical studies that used this approach, alongside traffic equilibrium modelling techniques, to study the variation of several different performance indicators as total demand was increased in a variety of different network ensembles.

Study	Network Topologies	Network Size (n nodes, m links, $\langle k \rangle$: Average Node Degree)	Number of Network Realisations	Link Travel Time/Cost Functions t_i (t_{0i}: free-flow travel time, x_i: link flow, cap_i: link capacity)	Demand Structure	Main Performance Indicator
Wu et al. (2006)	Random; Scale-Free; Small World	$n = 400$; $m = 1400$; $\langle k \rangle = 7$	25	$t_i = t_{0i}\left[1 + 0.15\left(\frac{x_i}{cap_i}\right)^4\right]$ - $t_{0i} \in (0, 0.1]$ randomly selected for each link - $cap_i \in [20, 60]$ randomly selected for each link	Random	Proportion of links over Capacity
Zhao and Gao (2007)	Regular Ring; Random; Scale-Free; Small World	$n = 500$; $m = 1000$; $\langle k \rangle = 4$	50	$t_i = t_{0i}\left[1 + 0.15\left(\frac{x_i}{cap_i}\right)^4\right]$ - $t_{0i} \in (0, 1]$ randomly selected for each link - $cap_i = 10000$ for each link	Uniform	Total Travel Time
Youn et al. (2008)	Sub-networks of: - Boston - London - New York	 - $n = 88, m = 246$ - $n = 82, m = 217$ - $n = 125, m = 319$	One of each	$t_i = t_{0i}\left[1 + 0.2\left(\frac{x_i}{2000k_i}\right)^{10}\right]$ - $t_{0i} = d_i/35$ where d_i is the length of each link i - k_i = number of lanes on each link i	Single OD pair	Price of Anarchy
	1D Regular Lattice; Random; Scale-Free; Small World	$n = 100$; $m = 300$; $\langle k \rangle = 6$	50	$t_i = a_i + b_i x_i$ - $a_i \in \{1, 2, 3\}$ randomly allocated to each link - $b_i \in \{1, 2, \ldots, 100\}$ randomly allocated to each link	Single OD pair	Price of Anarchy
Wu et al. (2008a)	Random; Scale-Free	$n = 100$; $m = 1350$; $\langle k \rangle = 2.7$	100	$t_i = t_{0i}\left[1 + 0.15\left(\frac{x_i}{cap_i}\right)^4\right]$ - $t_{0i} \in (0, 0.1]$ randomly selected for each link - $cap_i = C\ \forall i$ but the value of C is not defined	Not reported	Proportion of links over capacity
Wu et al. (2008b)	Regular Lattice; Random; Scale-Free; Small World	$n = 100, \ldots, 1000$; $m = 100, \ldots, 1000$; $\langle k \rangle = 2$	50	$t_i = t_{0i}\left[1 + 0.15\left(\frac{x_i}{cap_i}\right)^4\right]$ - $t_{0i} \in (0, 1]$ randomly selected for each link - cap_i is not defined	Random	Price of Anarchy

Sun et al. (2012)	Scale-Free with variable community structure	$\boldsymbol{n = 100, 160, 220}$; $\boldsymbol{m = 400, 640, 880}$; 4 communities	20	$\boldsymbol{t_i = t_{0i}\left[1 + 0.15\left(\frac{x_i}{cap_i}\right)^4\right]}$ - $\boldsymbol{t_{0i}}$ randomly selected for each link - $\boldsymbol{cap_i = 60\ \forall i}$	Random	Proportion of links over capacity
Zhu et al. (2014)	Scale-Free; Small World	$\boldsymbol{n = 1000}$; $\boldsymbol{m = 3000}$; $\boldsymbol{\langle k \rangle = 6}$	Not reported	$\boldsymbol{t_i = t_{0i}\left[1 + 0.15\left(\frac{x_i}{Ce_i}\right)^4\right]}$ - $\boldsymbol{t_{0i} = 1}$ for every link $\boldsymbol{i}$ between nodes $\boldsymbol{i1}$ and $\boldsymbol{i2}$ - $\boldsymbol{Ce_i = \min(Cn_{i1}/k_{i1}, Cn_{i2}/k_{i2})}$, for which i) $\boldsymbol{Cn_j}$ is fixed and ii) $\boldsymbol{Cn_j = f(k_j)}$	Uniform; Gravity Model	Volume to Capacity ratio (V/C)

Table 2.4 - Summary of Network Science Studies of the effects of Network Structure on Performance

The columns in Table 2.4 illustrate that these studies used a broad range of supply and demand structures. On the supply-side, significant variation is evident with respect to the numbers of nodes and links used, and also the parameter settings for link cost coefficients, which, in most cases, were chosen either randomly, from within a given range, or were fixed at one value, which was then applied to each link in each network. Significant variation is also evident on the demand-side. For example, Wu et al. (2006), Wu et al. (2008b) and Sun et al. (2012) used a *random* structure of demand wherein, as travel demand increased, each increment of total demand was wholly allocated to a randomly selected origin-destination node pair. Whereas, in contrast, Zhao and Gao (2007) and Zhu et al. (2014) used a uniform structure of demand in which each increment of total demand was spread evenly across all origin-destination node pairs.

The studies described in Table 2.4 all found that network performance does indeed vary with respect to supply and demand structure. For example, Figure 2.17 depicts results from two papers in Table 2.4 that show how congestion increases in three different network topologies as travel demand is increased.

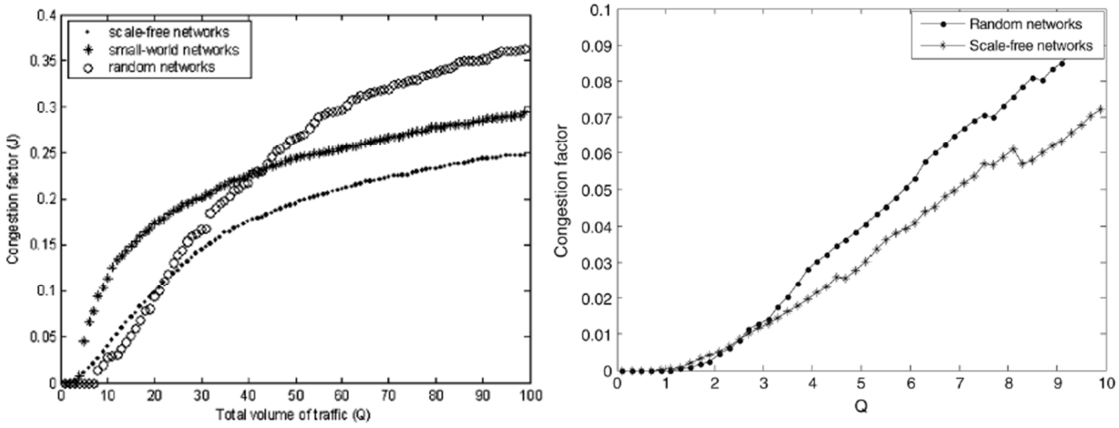

Figure 2.17 - Proportion of links over capacity against Demand (*Q*) for random, small-world and scale-free networks (Left: Figure 1, Wu et al. (2006). Right: Figure 1, Wu et al. (2008b))

However, there are also inconsistencies in their findings. For example, for the Price of Anarchy, whilst Youn et al. (2008) found that scale-free networks performed the best, followed by random and lattice networks, and that small-world networks performed the worst, Wu et al. (2008b) found a different ordering in which scale-free networks performed the best, followed by small-world and random networks, and that lattice networks performed the worst. These differences are clearly the result of differences between the two studies in the selected configurations of supply and demand structure, but it is difficult to understand the exact reasons for these differences because such a large number of aspects of structure are different between the two studies.

From a broader perspective, the studies presented in Table 2.4 do not provide a clear connection with the empirical studies described in section 2.3. Indeed, all of the studies based on synthetic networks used structures of supply that are not plausible for urban road traffic networks because they do not replicate the structural features that have been observed in real networks. With respect to topological structure, random, small-world and scale-free graph models typically produce non-planar networks. Moreover, the values chosen for free-flow travel costs t_{0i} in each study also indicate that the networks used did not have a geographical embedding.

Furthermore, from a methodological perspective, it is not clear that a comparison of the average performance of each ensemble of networks is appropriate because none of these studies provide justification for whether this is a suitable summary statistic. Such justification would require discussion of the distribution of performance across network realisations within each ensemble. However, this analysis is undertaken only by Youn et al. (2008), for which results are shown in Figure 2.18, and even then it is only through provision of error bars that represent one standard deviation. This assumes that performance is symmetrically distributed across networks within each ensemble but no evidence is provided to support this.

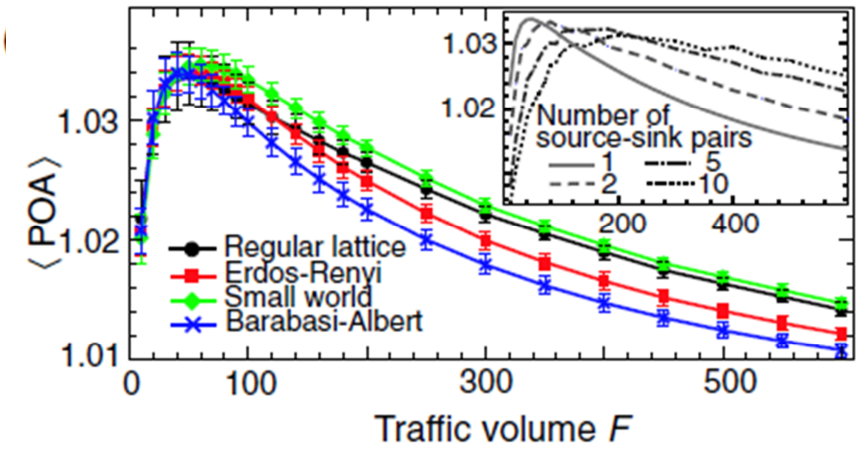

Figure 2.18 – Price of Anarchy against Demand for four synthetic network topologies (Figure 3b, Youn et al. (2008))

It is also highlighted that three of the seven studies in Table 2.4 did not provide complete descriptions of all of the parameter settings used; specifically the studies of Wu et al. (2008a), Wu et al. (2008b) and Zhu et al. (2014). These studies cannot, therefore, be reproduced and independently verified by other researchers.

The one study in Table 2.4 that used real network data is the first experiment of Youn et al. (2008), which studied how the Price of Anarchy varies as travel demand is increased in three,

single OD sub-networks of the Boston, London and New York road networks. This variation is shown in Figure 2.19.

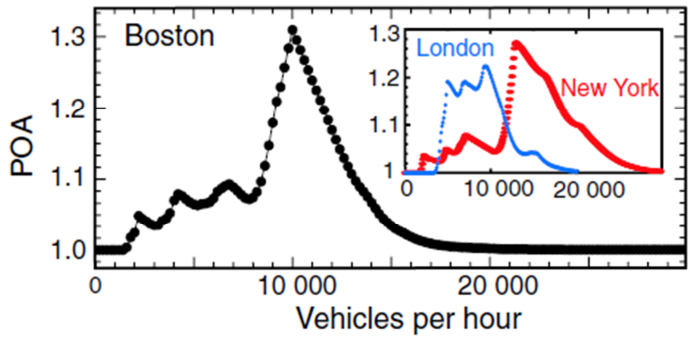

Figure 2.19 – Price of Anarchy against Demand for three real networks (Figure 3a, Youn et al. (2008))

The graphs shown in this figure highlight how the performance characteristics of different road traffic networks, when scrutinised individually, can vary substantially from each other in some respects but also share some broad commonalties. For example, the graphs of the Price of Anarchy for each city are clearly different, but there are also similarities in that they all appear to rise and then fall as demand increases. These graphs provide an illustration of how the theoretical studies described in section 2.4.3.2 - and even the numerical results shown in Figure 2.17 and Figure 2.18 - can obscure detailed features of how performance characteristics vary with respect to network structure.

2.5 Summary

This section summarises the main findings of this chapter under each of the three literature review questions that were posed in section 2.1.

1. **What are the key contributions and methodological approaches used in network science?**

Network science - and its antecedents in graph theory - has contributed many quantitative measures and methods that can be used to characterise the structural properties of networked systems, and has gone on to illustrate, through empirical studies, how many real world networked systems share similar structural characteristics. Network science has also proposed generative models for networks - such as the preferential attachment model, which produces scale-free networks - that suggest mechanisms that govern the formation of structure within networks. Finally, network science has also contributed a methodological approach that can be used to study how structure affects performance. Under this approach, several ensembles of synthetic networks are generated using canonical models from the

network science literature and the average performance characteristics, taken across the networks within each ensemble, are then compared.

Several limitations of the network science literature have also been identified. It has been noted that most empirical studies omit domain relevant information and have, thus far, focussed exclusively on the connectivity properties of networks. In addition, although numerical studies have shown - in the context of internet routing models - that the traffic flow and performance characteristics of networked systems do vary with network structure, it has also been noted that it is often unclear how different or indeed similar the network ensembles being compared are to each other. The results presented are also broad averages for network types generated by canonical models from the literature and do not show how performance characteristics vary between networks within each ensemble.

2. What structural properties have been shown to exist in supply and demand in road traffic networks?

Focussing first on network supply, most studies focus on urban areas. Empirical studies in network science have used two approaches to characterise the structural characteristics of road traffic networks: the so-called primal and dual approaches. Using the primal approach, in which nodes represent junctions and links represent road segments, empirical studies have shown that, at the microscopic level of nodes, links and cells, road traffic networks from urban areas from across the world share many similar structural characteristics. For example, with respect to their planarity, the angles formed between road segments at junctions and the radial distribution of nodes, link lengths and cell areas. In contrast, at the macroscopic level, empirical studies have also shown that there is significant variation between networks with respect to their size, density and connectivity. Under the dual approach, where nodes represent streets, network science has illustrated how urban road traffic networks can be shown to have both small-world and scale-free features, although it has been noted that the value of these findings for road traffic are unclear. Empirical studies under both approaches have been criticised for three additional reasons; firstly, that there is significant variation in the rules used to process raw network data, which introduces a degree of arbitrariness to their findings; secondly, that they all focus on connectivity properties and omit other important aspects of road supply such as link capacities and junction types[12]; and, thirdly, that it is not clear that such studies have captured the structural properties of the full range of network infrastructure that exists in urban areas from across the world.

[12] Note that this criticism can also be made of other empirical studies in network science that focussed on other networked systems.

Generative models for road traffic network infrastructure have also been proposed and have been shown to produce networks that are visually similar to networks in real urban areas and that also share some of the broad characteristics identified above. However, perhaps as a consequence of the focus of empirical studies on connectivity properties, none of these models include a representation of how other aspects of supply structure have developed. There is also little evidence to suggest that such models are able to mimic the evolutionary development of real road networks, which has been the subject of considerable study in transport geography.

In comparison with studies of network supply, empirical studies of the structural characteristics of travel demand are more limited in detail. This is largely due to difficulties in data collection, although availability should improve as data collection techniques based on new technologies become more widely available. Broadly, travel demand at various spatial scales and for different modes of transport, has been shown to have a structure that is highly heterogeneous and that is also polycentric in urban areas. However, it is noted that not one of these studies focuses exclusively on the pattern of travel in urban road traffic networks.

3. How have the effects of supply and demand structure on the performance characteristics of road traffic networks been studied thus far and what have such studies found?

The transportation and network science literature have used several different methods to study the performance characteristics of road traffic networks.

In the transportation literature, most empirical studies are location specific and focus on individual case studies, which therefore reveal little about how performance varies with respect to supply and demand structure. The transportation literature also contains literature of a theoretical nature in which upper bounds have been produced for the worst case value of one particular performance measure; the Price of Anarchy, across broad families of networks. However, such bounds do not reveal how performance may vary between networks within such families, i.e. between networks that may have very different structures of supply and demand. One notable exception to this pattern is the work of Levinson (2012) and Parthasarathi et al. (2012) who used observed data from a large number of different urban areas to search for correlations between network structure and network performance indicators, finding that cities with larger populations are typically more congested and have longer journey to work travel times.

A greater body of material for how performance varies with respect to supply and demand structure has been presented by numerical studies in the network science literature. These studies have used the same methodological approach as used by other studies in network

science to study jamming phenomena in the internet with the addition of traffic equilibrium modelling techniques to represent road traffic behaviour. Under this approach, such studies have generated large ensembles of synthetic networks using canonical models from the network science literature and have then compared the average performance of networks within each ensemble. Whilst this approach has illustrated that performance in road networks does indeed vary with respect to supply and demand structure, there are several methodological issues. Most prominently, these studies use structures of supply that are non-planar and therefore are not plausible for urban road traffic networks. They also do not provide any connections to the findings described under the second literature review question above. Similarly to the criticism made of network science studies of traffic flows in the internet, such studies also draw comparisons between networks for which the extent of structural similarity or dissimilarity is unclear. This makes it difficult to generalise their findings or to apply them to other families of networks.

The above points demonstrate that, although network science has established an approach for studying how network structure affects performance in networks, there is a need for further work to determine how this approach can be better applied in the specific context of road traffic networks. The purpose of Chapter 3 is to address this issue through the proposal of an investigative framework.

3 An Investigative Framework for Studying the Effects of Structure on Performance in Road Traffic Networks

3.1 Introduction

This chapter proposes an investigative framework for studying the effects of network structure on performance in road traffic networks, which addresses the deficiencies of existing approaches highlighted in chapter 2. This framework comprises an experimental part, in which numerical experiments are undertaken to study how one or more performance indicators vary with respect to specific aspects of network structure, and an analytical part, in which explanations are sought and developed to explain the patterns uncovered by the numerical experiments with the aim of establishing theory.

This chapter begins, in section 3.2, with a discussion of the main challenge that is faced by numerical experiments of the effects of structure on performance and of the approaches that have been used in existing literature to address this challenge. This discussion then feeds into and is used to justify the proposed investigative framework, which is presented in section 3.3.

3.2 The Main Challenge: Selecting Networks from the Search Space

The review in section 2.3 illustrated that investigations of the effects of structure on performance in road traffic networks have to contend with a huge, multi-dimensional search space of networks, which spans all possible configurations of supply and demand structure. This is particularly well illustrated by the columns in Table 2.4. On the supply-side, there is a huge, multi-dimensional space of possible infrastructure configurations; with respect to the numbers of nodes and links, their connection pattern and the functional form and associated parameters of how link travel costs are represented. Similarly, there is also a broad array of possible demand patterns; with respect to both the total amount and the distribution of demand between nodes in the supply network.

Given this high dimensionality, the main challenge faced by numerical investigations is of how to select network configurations for comparison from within this search space in such a way so as to provide useful insights into how network structure affects performance in road traffic networks. The selection of network configurations is important because it has a direct impact on the strength, generality and transferability of research findings.

There are two approaches to network selection that been used in existing literature: the synthetic networks approach of the network science papers, and the real-world data approach

of Parthasarathi and Levinson (2010) and Levinson (2012), which was described in section 2.4.3.1. The subsections that follow address each of these approaches in turn and then describe the approach that is proposed in this thesis.

3.2.1 Existing Approach 1: The Synthetic Networks Approach

As described in the reviews in sections 2.2.5 and 2.4.4, the network science approach to numerical experiments is to use a small number of canonical models from the network science literature, such as the small-world and scale-free network models described in section 2.2.3, to generate large ensembles of synthetic networks, and to then compare the performance characteristics of these different ensembles. One of the main issues with this approach, which has already been identified, is that it has been applied using models that typically generate non-planar supply networks and are, therefore, not plausible as representations of real road traffic networks. This method of selecting networks from the search space also has an additional flaw in that it provides only point-to-point comparisons between ensembles of particular network types, whose similarity or dissimilarity in structure is unclear.

To illustrate these points, Figure 3.1 provides a visual analogy of how networks are selected from the search space under this approach. In this figure, the square outline is used as a visual representation of the search space[13], and the dashed line demarks the boundary between planar and non-planar supply networks (which are known to form disjoint sets). Three ensembles of networks, each surrounded by dotted lines, can be seen in the non-planar region of the search space. These ensembles could represent, for example, scale-free, small-world and random networks.

[13] Figures in this chapter are used only to illustrate the differences between the different approaches to network selection. They are not intended to be and should not be interpreted as accurate representations of the search space.

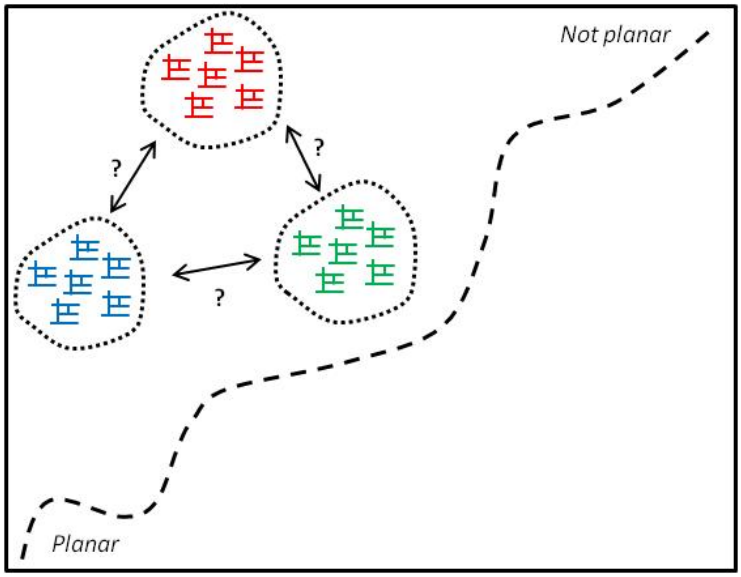

Figure 3.1 - Illustration of network selection for the network science approach

Under this analogy, typical road traffic networks would sit in the planar region of the search space, which contrasts with the location of the network ensembles used in existing network science studies. In addition, the ambiguity with respect to the unknown similarity or dissimilarity of network ensembles is represented by the arrows and question marks; i.e. the figure shows that it is unclear where these three ensembles sit in the search space in relation to each other. As an example from the literature, Wu et al. (2006) compares the performance of ensembles of scale-free, small-world and random networks but these comparisons give little insight into networks that do not neatly fit within these categories. As a result, this approach does not explain which aspects of supply or demand structure cause the differences in performance that have been observed. It is capable only of providing results of the form: networks of type 'A' perform better, on average, than networks of type 'B'.

3.2.2 Existing Approach 2: The Real-World Data Approach

An alternative network selection technique is demonstrated by Parthasarathi and Levinson (2010) and Levinson (2012). In this second approach, real data for a large sample of real urban road traffic networks are used, alongside regression techniques, to search for correlations between measures of supply and demand structure and indicators of network performance.

A visual analogy for this second approach is shown in Figure 3.2. In the context of selecting networks from the search space, this approach focuses on networks from the planar region and uses a larger sample of different network types than is typically used in the first approach, thereby providing greater coverage of the search space as is indicated. In selecting networks in this way, this approach also provides a large number of individual readings for a range of structural measures, which can each be paired with the value of a performance indicator,

thereby enabling correlations between these variables to be studied. For example, Levinson (2012) studied the correlation between fifty values of the typical journey to work travel time in fifty US cities and the population size of those fifty cities.

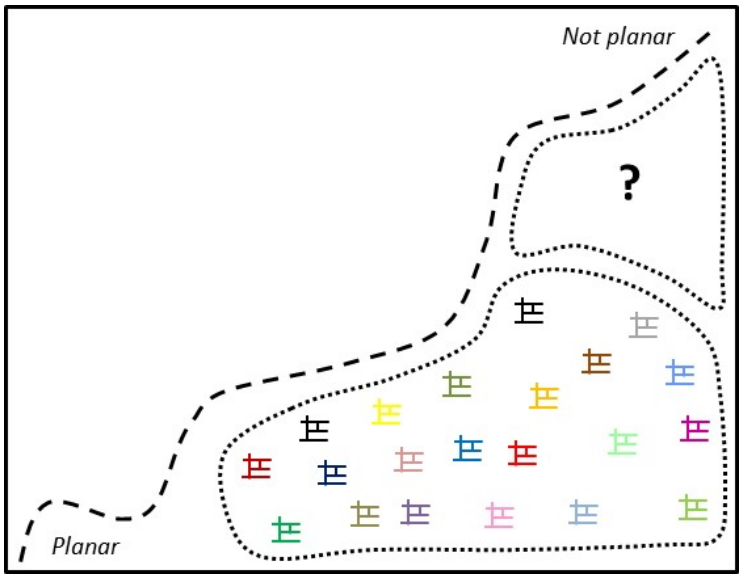

Figure 3.2 - Illustration of network selection for the Levinson (2012) approach

However, like the first approach, this approach also has drawbacks. Firstly, it is data intensive as it requires data on network supply, travel demand and network performance for each selected city, which may not be readily available or be of sufficient quality. For example, section 2.3 demonstrated that whilst data is broadly available for network supply (although it requires a significant amount of pre-processing), detailed data is not widely available for travel demand. For network performance data, Parthasarathi and Levinson (2010) and Levinson (2012) used publically available data from the Texas Transportation Institutes Urban Mobility Report (Schrank et al., 2012). However, such sources are limited to the indicators of interest to the organisation that collected the data and are also subject to noise as a result of the way in which the data has been collated and processed. Noise also occurs naturally in data collected from real-world systems, which can obscure relationships between variables.

Related to this first drawback, this approach is also restricted to networks for which the required data are available, which, in the context of the huge number of possible supply and demand configurations in the search space, are unlikely to span the entire range of possible structures of road traffic networks. This is represented by the question mark region in Figure 3.2. The networks in each sample are also likely to be very different to each other in several aspects of supply and demand structure. Overall, these drawbacks restrict the capability of regression analyses to identify which aspects of structure drive observed variations in performance. This could be a reason for the small R-squared values reported by both

Parthasarathi and Levinson (2010) and Levinson (2012) for the explanatory power of their statistical models.

3.2.3 The Proposed Approach: Fusion of Synthetic Networks and Real-World Data

The approach to network selection proposed in this thesis combines the benefits of flexibility and control in network generation, which is offered by the synthetic networks approach of network science, with the idea of studying an array of networks that span a range of real road traffic network structures, which is the basis of the regression analyses in the approach of Parthasarathi and Levinson (2010) and Levinson (2012). The proposed approach is to generate a spectrum of ensembles of synthetic networks, which provide a cross-section of the search space and in which only one aspect of network structure is varied, and to use a road traffic model to explore how performance indicators vary within each ensemble and with respect to the selected aspect of network structure. This approach to network selection is illustrated in Figure 3.3.

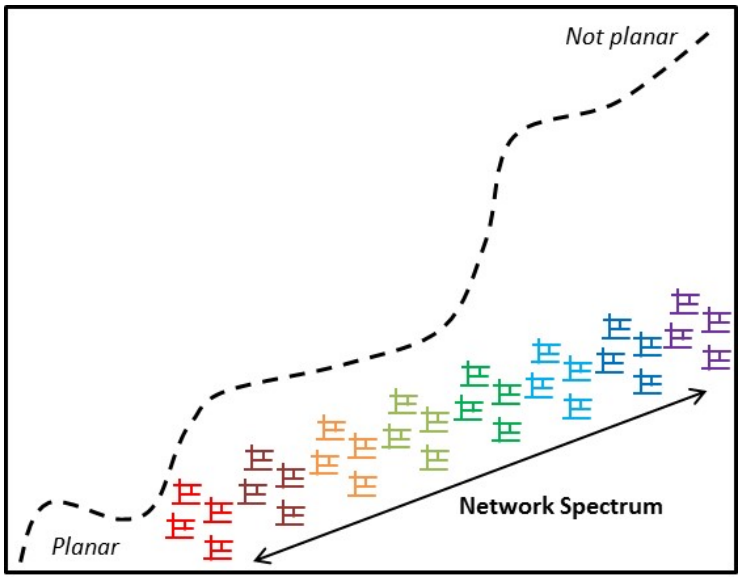

Figure 3.3 – Illustration of network selection for the approach proposed in this thesis

Under this approach, the aspects of structure that do not vary are fixed at values or in a configuration that is plausible for real road traffic networks, whilst the aspect of structure that varies does so to encompass a range of values for that structural feature that have been observed in real road traffic networks. A simple example of this approach would be a spectrum of network ensembles in which total demand is increased by a global demand multiplier, whilst network supply and the distribution of travel demand are fixed in plausible configurations.

In comparison with the previous approaches, illustrated in Figure 3.1 and Figure 3.2 above, this approach avoids point-to-point comparisons between separate categories of networks, which was a key flaw of the first approach, because one aspect of network structure varies

'continuously' across the spectrum of networks. This approach also avoids the problem of there being too many aspects of network structure that change at the same time, which, in the second approach, made it difficult to establish relationships between measures of structure and performance indicators. Like the second approach, this approach is still reliant on data for supply and demand structure in order to calibrate the generation of synthetic networks. However, unlike the second approach, this approach is not limited to a small number of networks nor does it require any additional data on network performance; the latter being produced by the road traffic model.

3.3 Statement of the Proposed Investigative Framework

Building on the discussion of the previous section, the investigative framework is now proposed. As stated previously, this framework includes both experimental and analytical aspects, which can be broken down into six steps:

Step 1. Identify a measurable aspect of network supply or demand structure and a network performance indicator to be studied.

Step 2. Design a model of road network generation that is capable of producing a spectrum of network ensembles, which span a range of values for the aspect of network structure to be studied and which are also *plausible* as representations of real road traffic networks.

Step 3. Calculate values of the performance indicator for the generated networks using an appropriate road traffic model.

Step 4. Create a graph of the calculated values of the performance indicator against a measure of the selected aspect of network supply or demand structure.

Step 5. Develop theory to explain the patterns shown in numerical results and, if possible, derive theoretical results to establish their generality and driving mechanisms.

Step 6. Document the preceding steps to include a complete description of the parameter settings used for supply and demand, making the results reproducible by other researchers.

The following remarks are made with respect to the steps in this framework:

- Empirical studies of the structure of road traffic networks, such as those described in section 2.3, are particularly useful for the first and second steps in this framework because they motivate interesting aspects of structure for investigations, and also help to define structural characteristics that generative models of road traffic networks should aim to replicate. The *plausibility* of networks produced by such models should be judged by the extent to which they replicate features observed in real road traffic networks.

- With reference to the third step, the framework is flexible and can accommodate any road traffic model of any level of complexity. The only conditions on its specification are that the model should be plausible for road traffic and appropriate for the specific context under study (e.g. urban or interurban travel), and that it should also be adequately described (i.e. step six).

- The graph produced in step four provides insight into how the selected aspect of network structure affects the selected performance indicator. This graph will also reveal the dispersion of network performance values when the mechanism for network generation has a stochastic component.

- The fifth step in the framework is important because it broadens the applicability of research findings. In this context, the preceding four steps in the framework could be seen as generating hypotheses, which are then investigated in more detail in the fifth step.

- The final step of the framework ensures that experiments and investigations can be reproduced by other researchers, avoiding unnecessary duplication of tests and also providing an intelligble foundation for future work. This documentation should include a description of the road traffic model and, where appropriate, provide data to support the stability and validity of model outputs. Inclusion of such data provides confidence that any presented differences in performance are real and not distorted by model noise. This is especially important when comparisons are to be made between a large number of different networks.

3.4 An Example Application of the Proposed Framework

The next three chapters of this thesis demonstrate an application of the proposed investigative framework. In accordance with the first step of the framework, and inspired by the empirical studies that were described in section 2.3, this application focusses on road traffic networks in an urban setting, for which empirical data is available, and investigates how performance varies with respect to four aspects of network structure: namely, the density of travel demand and the size, density and connectivity of network supply structure. Section 2.3.1.2.2 showed that there is considerable variation across urban areas in each of these structural characteristics, which distinguishes them as interesting features to be investigated.

With respect to performance, the investigations described in the chapters that follow study the effects of the four selected structural characteristics on two performance indicators: the average link Volume-to-Capacity ratio, which is a commonly used measure of congestion; and the Price of Anarchy, which was defined in section 2.4.1 and measures the inefficiency of the selfish behaviour of road users in comparison with socially optimal behaviour (Koutsoupias and Papadimitriou, 1999, Papadimitriou, 2001). Both of these measures have been used in previous studies from network science; see Table 2.4, and are also of interest to transportation

researchers. In particular, the Price of Anarchy is of interest because it can be used to measure the efficiency gains of first best road tolling schemes (Yang and Huang, 2005). An understanding of how this measure varies with respect to specific features of network structure would therefore be useful because it could help identify circumstances in which policy interventions (e.g. road pricing) designed to induce more efficient routing behaviour would be worth their costs of implementation (Mak and Rapoport, 2013).

The experimental part of the framework (steps two to four) is implemented for all eight combinations of performance indicators and aspects of network structure. The analytical part of the framework (step five) is implemented for only one of these combinations; the variation of the Price of Anarchy with respect to travel demand, although explanations for the other seven combinations are provided at the end of Chapter 5.

Chapter 4 describes the model of road network generation that was used and Chapter 5 describes the design and results of the numerical experiments that were then undertaken. These chapters therefore address the second, third and fourth steps of the investigative framework. Chapter 6 presents the analytical study of how the Price of Anarchy varies with respect to travel demand; thereby addressing the fifth step of the investigative framework.

4 A Model for Generating Spectrums of Synthetic Ensembles of Road Traffic Networks

4.1 Introduction

Having selected two performance indicators and four aspects of network structure to be investigated, the second step of the investigative framework requires a model of road network generation that is capable of producing a spectrum of ensembles of synthetic networks, which span a range of values of the selected aspects of network structure and which also produce plausible representations of real road traffic networks.

The generative model described in this section has three stages: the creation of a topological and geometric structure for each road traffic network in each ensemble, the generation of a travel demand structure, and the specification of a road traffic model. These stages are described in sections 4.2, 4.3 and 4.4 respectively. Section 4.5 then presents an example spectrum of synthetic networks generated by the model.

4.2 Network Model

Section 2.3.2 described two generative models for urban road traffic networks that have been proposed in the network science literature; the models of Barthelemy and Flammini (2009) and Courtat et al. (2011). Both models attempt to mimic an evolutionary process in which a network, starting from an initial seed network, is grown incrementally over time through the addition of new nodes, which each stimulate the growth of network links to connect them to the existing network.

Both Barthelemy and Flammini (2009) and Courtat et al. (2011) demonstrated that their models were able to reproduce many of the structural features observed in real urban road traffic networks, such as those described in section 2.3.1. Both models have a range of input parameters and settings that could be used or adjusted to produce a broad range of network types with different structural features, and both also include a stochastic component meaning that ensembles of networks that share similar structural features could be produced. As such, both of these approaches could be used within the second step of the investigative framework to generate ensembles of road traffic networks of different types.

However, as these models are not freely available to use and are not easy to replicate, a simpler model of network generation was used instead. This model is more similar in starting assumptions to the model of Erdös and Rényi (1959) than to the two models above because it

uses a predefined distribution of nodes in a domain to produce a candidate link set rather than adding nodes to a domain over time. This network model has three steps:

- **Step 1: Scatter n nodes randomly in a square domain, which is Akm^2 in size** - As uniformly randomly distributed nodes tend to occur in clusters, which would result in links of extremely short length, a rule is imposed that all nodes must be at least d_{min}km apart; i.e. a minimum link length.

- **Step 2: Construct the Minimum Weight Spanning Tree (MST) and Delaunay Triangulation on the node set generated in Step 1** - The Delaunay Triangulation for a set of nodes V is a triangulation of the node set that maximises the minimum angle of all triangles and contains the maximum possible number of links without violating planarity. Euler's formula shows that the maximum number of links in this graph is $3n - 6$ (Barthelemy, 2011). The MST is the graph of minimum total length that provides a path between every pair of nodes for a given node set. It is a subgraph of the Delaunay Triangulation and contains $n - 1$ links. The Delaunay Triangulation and MST define the candidate link set for the new network.

- **Step 3: Select $m \geq n - 1$ links from the Delaunay Triangulation to create a new network G defined on this node set, of which the first $n - 1$ links are from the MST** - The inclusion of the MST in G guarantees that there is at least one path between every pair of nodes, i.e. that the network is fully connected. The remaining $m - (n - 1)$ links are randomly selected from the remaining links in the Delaunay Triangulation.

The number of nodes n, the domain size A, the minimum link length d_{min} and the number of links m are input parameters to the model, which, when varied, generate networks with a broad variety of structures; see Figure 4.1 for examples. However, as the scattering of nodes in step 1 and the selection of links in step 3 are stochastic processes, this model produces a different network each time it is run. Several individual model runs can therefore be used to create ensembles of networks for a fixed set of input parameters, whose members therefore share similar structural features.

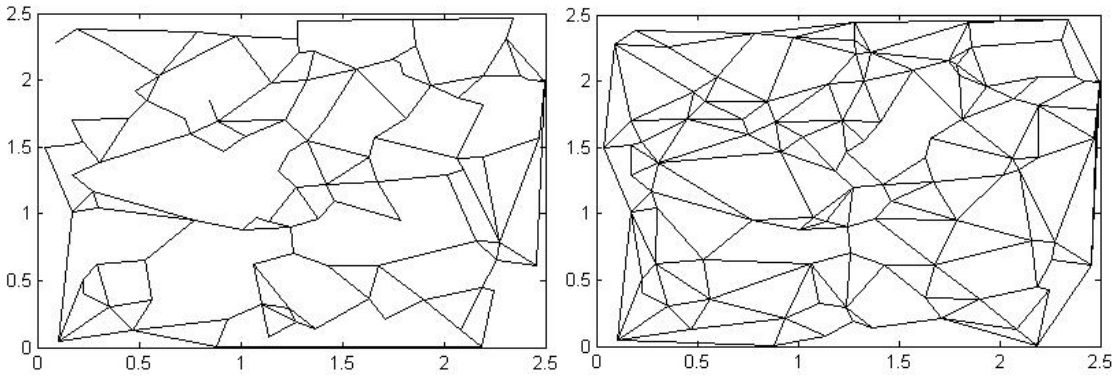

Figure 4.1 – Two example network realisations with input parameters $n = 100$, $A = 6.25$ and $d_{min} = 0.05$, with $m = 159$ (left) and $m = 229$ (right)

It is noted that the networks generated by this model do not replicate all of the structural properties of supply in urban road traffic networks that were identified in section 2.3.1. For example, Figure 4.1 illustrates that typical networks produced by this model have a wide range of angles between links at nodes, which contrasts with the pattern that is known to exist in real road traffic networks. The node distribution mechanism also produces a broadly uniform scattering of nodes across the domain, which contrasts with real urban road traffic networks for which the density of nodes has been shown to be high in the centre of an urban area but then decrease as distance from the centre increases.

However, the networks generated by this method are planar, which accords with the findings of empirical studies for real road traffic networks. The planarity of the networks produced by this model also makes them more plausible as representations of real urban road traffic networks than the networks used in the network science studies described in Table 2.4, which were typically non-planar.

4.3 Travel Demand Model

Section 2.3.2 highlighted that there is a lack of empirical data on the structure of travel demand in urban road traffic networks but that some studies, especially at the interurban level, have shown that demand is heterogeneously structured.

In light of this lack of data, the road network model described in this chapter adopts a simple assumption that travel demand is uniformly distributed across all node pairs in each network. Under this assumption, each node pair r is therefore assumed to have the same amount of demand q_r travelling between them. This approach was used by several network science studies described in Table 2.4. In order to ensure that a constant density of demand per km^2 ϱ_{dem} is maintained across network domains of different sizes, the total amount of demand in each network is defined as a function of the domain size A. The volume of demand per OD pair r is shown in equation (3).

$$q_r = \frac{A\varrho_{dem}}{n(n-1)} \quad (3)$$

Demand density ϱ_{dem} is an input parameter to the model. It is clear from equation (3) that this approach also includes implicit assumptions that demand density does not vary within the domain itself or as the size of the domain changes.

The generation of demand in this way is one simple way of ensuring that performance comparisons are fair between networks that serve domains of different sizes. For example, an even simpler assumption that could have been used is that each network domain produces a constant level of demand. However, it is questionable whether it would be fair to judge that a network in a large domain is 'better' ('worse') than a network in a small domain because it is more (less) congested under the same total amount of assigned demand.

4.4 Road Traffic Model

The final aspect of the model is a mechanism for the assignment of travel demand from section 4.3 to networks created by the model in section 4.2. Section 2.4.2 described several approaches that could be used to model flows on network links in road traffic networks. The model described in this section uses the User-Equilibrium (UE) and System Optimum (SO) road traffic models. Section 2.4.2 noted that although such techniques have deficiencies with respect to representing the dynamic nature of traffic flows (for example), such techniques are still widely used in transportation studies. A detailed technical description of this modelling approach is provided in Figure 4.2. It should be noted that, for simplicity and to reduce the number of variables, the effects of junction interactions are not included.

In addition to the specification of a network G and a travel demand matrix Q, the traffic models described in Figure 4.2 also required the specification of link cost functions c_i, which describe the cost (often travel time) to traverse a given link as a function of the volume of flow and its operational characteristics. The network model described in this chapter uses the BPR cost function (Bureau of Public Roads, 1964), which satisfies the conditions for existence and uniqueness set out in Figure 4.2 and which has the form shown in equation (4).

$$t_i = t_{i0}\left[1 + 0.15\left(\frac{x_i}{cap_i}\right)^4\right] \quad (4)$$

In equation (4), t_{i0} represents the travel time in free-flow conditions, x_i represents link flow and cap_i represents link capacity. Whilst this function has been criticised for providing a poor representation of traffic flows in urban areas, it also has the advantage of being simple and of requiring values for only a small number of parameters. It is also commonly used in

transportation literature; for example, most of the networks available for download from Bar-Gera (2001) use this cost function.

In general, the values of t_{i0} and cap_i depend on factors that include link length, the speed limit, the level of street frontage activity and the number of lanes. Guidance published for modelling practitioners by the Department for Transport in the UK recommends an assumption that links in central urban areas - where there is a speed limit of 30mph (48kph) - should have a maximum modelled capacity of $Q_i = 800$ vehicles per hour, per 3.65m lane (WebTAG, 2014). Adopting this form, for each link i, the network model described here assumes that $cap_i = Q_i \times k_i$, where k_i represents the number of lanes, and $t_{i0} = d_i/48$, where d_i represents link length in kilometres. The units of t_{i0} are therefore hours. The number of lanes on each link k_i and the lane capacity value Q_i are input parameters to the modelling process. The length d_i for each link is generated by the network model described in section 4.2.

Technical Description of the User Equilibrium (UE) and System Optimum (SO) Traffic Assignment Models

In this description, the topology of a traffic network is represented by a directed graph $G(V, A)$, comprising a set of nodes V and a set of directed links A. The costs of travel on each link $i \in A$ are represented by cost functions c_i. Travel demand is represented by an Origin-Destination (OD) vector Q with entries q_r denoting the volume of travel on OD movements $r = 1, \dots, R$ between pairs of nodes from V. Each OD movement r is served by a finite number $k = 1, \dots, \kappa_r$ of acyclic routes K^r. Using this notation, the UE principle described in section 2.4.2 can be characterised in mathematical notation as equation (5) (Patriksson, 1994).

$$\begin{gathered} f_k^r > 0 \Rightarrow C_k^r = \pi_r \\ f_k^r = 0 \Rightarrow C_k^r \geq \pi_r \\ \forall k \in K^r, \forall r = 1, \dots, R \end{gathered} \tag{5}$$

Here, f_k^r denotes the flow and $C_k^r = C(f_k^r)$ denotes the cost of travel on a route $k \in K^r$. The cost of travel on each route $k \in K^r$ is assumed to be the sum of link costs: $C_k^r = \sum_i c_i(x_i)\, \delta_{i,k}^r$, where $\delta_{i,k}^r = 1$, if link i is part of route $k \in K^r$, and zero otherwise. The $\delta_{i,k}^r$ terms form a link-path incidence matrix, which is denoted by Δ.

Subject to the above conditions and the assumption that link costs c_i are continuous, positive, separable and strictly increasing functions of link flows x_i, it can be shown that there exist unique link flows x_i^{UE} satisfying the UE conditions (5) and which solve (Sheffi, 1985):

$$\min_x \; z(x) = \sum_{i \in A} \int_0^{x_i} c_i(\omega) d\omega$$

subject to the constraints:

$$\sum_k f_k^r = q_r \; \forall r$$

$$x_i = \sum_r \sum_{k \in K^r} f_k^r \delta_{i,k}^r \; \forall i \qquad (6)$$

$$f_k^r \geq 0 \; \forall k \in K^r, \forall r = 1, \dots, R$$

Under the same conditions, unique link flows x_i^{SO} satisfying the SO principle also exist, and solve a minimisation program with the same constraints (6) but with objective function $\tilde{z}(x) = \sum_{i \in A} x_i c_i$. Under the assumption that link costs c_i are also differentiable, the SO objective function is equivalent to the UE objective function under a transformation of link costs $\tilde{c}_i = c_i + x_i \times dc_i/dx_i$. In comparison with c_i, the cost functions $\tilde{c}_i$ include the additional cost burden that each unit of flow imparts on all other units of flow on each link (Sheffi, 1985). Sheffi (1985, p71-74) refers to $\tilde{c}_i$ as *"marginal travel costs"* and defines the *"marginal total travel cost"* on a route $k \in K^r$ as $\tilde{C}_k^r = \sum_i \tilde{c}_i(x_i) \, \delta_{i,k}^r$.

Figure 4.2 - Technical Description of the User Equilibrium (UE) and System Optimum (SO) Traffic Assignment Models

4.5 An Example Spectrum of Synthetic Network Ensembles

Figure 4.3 shows an example spectrum of ensembles of road traffic networks generated by the model described in section 4. This spectrum was generated with $n = 100$ nodes, a domain size $A = 6.25\text{km}^2$, a minimum link length $d_{min} = 0.05\text{km}$ and numbers of links $m = 99, 114, 129, 144, 159, 174, 189, 204, 219, 234, 249$ and 264 respectively across the twelve network plots shown. Each individual network plot represents one run of the network model.

It can be seen that the connectivity of the networks increases incrementally across the spectrum shown in Figure 4.3. Indeed, these networks span a range of values of the meshedness measure between $M = 0$ for the first network and $M = 1$ for the final network in the spectrum. This figure therefore illustrates the benefits of drawing performance comparisons across the different network ensembles making up this spectrum because they can be linked directly to a quantitative measure of network structure. The networks it produces are also at least plausible for real road traffic networks, if not exhaustively accurate

as has been noted. This generative model therefore satisfies the requirements of the second step of the investigative framework described in the previous chapter.

The next chapter of this thesis presents the results of four numerical experiments, which use the model described in this section to investigate how two performance indicators vary with respect to four aspects of network supply and demand structure.

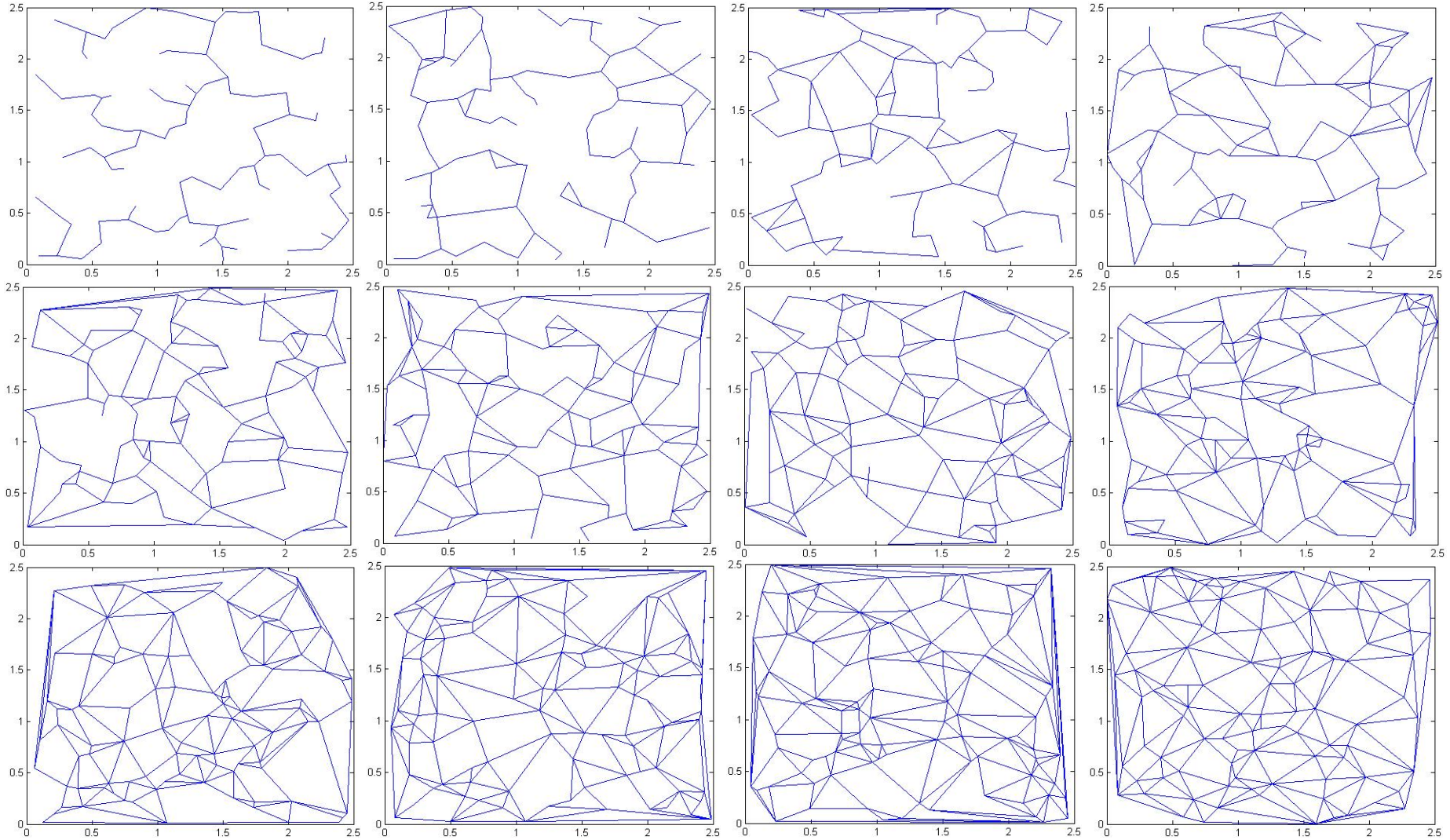

Figure 4.3 - An Example Spectrum of Synthetic Network Ensembles

5 Numerical Experiments of how Network Structure affects the Performance Characteristics of Road Traffic Networks

5.1 Introduction

This chapter describes and presents the results of the four numerical experiments specified at the end of chapter 3; thereby demonstrating the application of the third and fourth steps of the proposed investigative framework. These numerical experiments use the network model described in chapter 4 and focus on how two performance indicators; the average link V/C ratio and the Price of Anarchy, vary with respect to the density of travel demand and the size, density and connectivity of network supply structure.

The first section of this chapter sets out the parameter settings used to generate four spectrums of network ensembles across the four identified aspects of network structure and also provides a commentary on the size and complexity of the numerical experiments that were undertaken. The results of the numerical experiments are then presented in four subsections in section 5.3. A discussion of the results then follows in section 5.4.

5.2 Description of Experiments

5.2.1 Parameter Settings

The model of road network generation, described in chapter 4, has seven input parameters in total. Table 5.1 describes the values used for four of these parameters (A, n, m and ϱ_{dem}) to create spectrums of network ensembles across the four structural dimensions of interest. Table 5.1 also displays corresponding values for node density ϱ_n and meshedness M, which measure the density and connectivity of network supply structure.

In the first experiment, given the absence of empirical data on travel demand, demand density was varied across a broad range of values between $\varrho_{dem} = 1250$ and $\varrho_{dem} = 7950$. Whilst demand density was varied, the parameters that control the density and connectivity of network supply were fixed at average observed values for real urban road traffic networks, taken from Table 2.2 and Table 2.3 respectively. The domain size A and number of nodes n were also fixed in this experiment, although at much smaller values than the average observed values of $A = 296\text{km}^2$ and $n = 4713$ shown in Table 2.2. This was principally because a road traffic model run in a network of such size would have represented a significant computational burden in the computing environment used in this research; indeed, a single run of the road traffic model described in section 4.4, for a network with input parameters $A = 296$, $n = 4713$, $m = 7538$, $\varrho_{dem} = 2355$, $d_{min} = 0.05$, $k_i = 1$ and $Q_i = 800\ \forall i$, took

approximately 36 hours to find a UE link flow solution of sufficient precision. Repetition of this run time over a large number of network realisations would therefore have been impractical.

The second, third and fourth experiments, which explored the effects of network size, network density and network connectivity respectively, were setup in a similar fashion, with one aspect of network structure being varied whilst the remaining aspects remained unchanged. The parameters used in these experiments are described in Table 5.1. The range of network sizes in the second experiment was capped at $n = 500$ nodes in order to limit the computational burden of the experiments. The ranges of network density and network connectivity used in the third and fourth experiments encompassed the full range of observed values for real urban road traffic networks from Table 2.2 and Table 2.3 respectively. In each of the last three experiments, demand density was fixed at $\varrho_{dem} = 4350$ because, as will be shown, this value produced a reasonably congested (but not overly-congested) network and also the highest values of the Price of Anarchy measure.

Experiment Title	**Domain Size (Akm^2)**	**Num. of Nodes (n)**	**Node Density (ϱ_n)**	**Num. of Links (m)**	**Meshedness (M)**	**Demand Density (ϱ_{dem})**
1. Demand Density	6.25	100	16	158	0.3	1250, 1300, …, 7900, 7950
2. Network Size	1.25, 1.875, …, 30.625, 31.25	20, 30, …, 490, 500	16	30, 46, …, 782, 798	0.3	4350
3. Network Density	6.25	20, 25, …, 295, 300	3.2, 4, …, 47.2, 48	30, 38, …, 470, 478	0.3	4350
4. Network Connectivity	6.25	100	16	99, 104, …, 284, 289	0, 0.03, …, 0.95, 0.97	4350

Table 5.1 - Parameter Settings for Numerical Experiments

Values for the three remaining parameters (d_{min}, k_i and Q_i) were fixed in all four experiments. The minimum link length was set at $d_{min} = 0.05$km, which coincides with the shortest length of city blocks in the urban areas studied by Chan et al. (2011). The number of lanes was fixed at $k_i = 1$ for each network link, which was based on anecdotal evidence from the UK that one lane per link is the most common situation on urban roads. Finally, link capacity was fixed at $Q_i = 800$ for each network link, which is equal to the maximum value recommended by WebTAG (2014). Coupled with the assumption that every link has one lane, the implication of this assumption is that every network has only one road type in its road hierarchy.

5.2.2 Commentary

Due to the mathematical complexity of finding link flow solutions for the UE and SO traffic assignment models, solutions are usually derived using numerical methods. All of the numerical results presented in this chapter were derived using the Origin-Based Assignment (OBA) algorithm (Bar-Gera, 2002), using the OBA executable downloaded from Bar-Gera (2001). Each traffic assignment was solved to an Average Excess Cost no greater than 10^{-4}, which accords with the guidance on convergence provided by Boyce et al. (2004) and provides confidence that comparisons shown are true and not the result of model noise.

The results presented in the next section are based upon one hundred network realisations for each ensemble of parameter settings shown in Table 5.1. This corresponds to one hundred individual runs of the network model described in chapter 4 to create each network ensemble. The four spectrums described in Table 5.1 encompass 135, 49, 57 and 39 network ensembles respectively. When combined, the four experiments therefore included results from 28,000 network realisations and 56,000 traffic assignments under the UE and SO models.

All of the traffic assignments were undertaken on a remote desktop server, running the Windows Operating System with 64GB of RAM. The traffic assignment runs for each experiment took 107.75 hours, 421 hours, 67.5 hours and 9.5 hours respectively to complete. This amounts to an overall total run time of 605.5 hours, which is equivalent to approximately twenty-five days[14].

5.3 Results

The results for the two performance indicators are presented, in each experiment, as a sequence of boxplots for each network ensemble. This enables the dispersion of results across the networks within each ensemble to be presented. The tops and bottoms of each box represent the upper and lower quartiles respectively, and the band across the middle of each box represents the median value of the performance indicator. The whiskers extending out from the top and bottom of each box extend to the highest and lowest value data points that are within one times the interquartile range. Data points marked by circles represent network realisations in which the indicator value was greater than the upper quartile, or less than the lower quartile, by between 1.5 and 3 times the interquartile range. Data points marked by stars represent network realisations in which the indicator value was greater than the upper

[14] This figure is subject to the caveat that the remote desktop server that was used was shared with other researchers at the university, which may have increased run times.

quartile, or less than the lower quartile, by at least 3 times the interquartile range. Data points marked by circles and stars are commonly interpreted as outliers in statistical analyses.

5.3.1 Experiment 1: Demand Density

Figure 5.2 and Figure 5.3 present the variation of the average link V/C ratio and the Price of Anarchy as demand density is increased across the range $\varrho_{dem} \in [1250,7950]$. In these experiments, supply structure was fixed in each network ensemble, with domain size $A = 6.25\text{km}^2$, $n = 100$ nodes, a node density of $\varrho_n = 16$ and a meshedness value of $M = 0.3$. Two examples of such networks are shown in Figure 5.1.

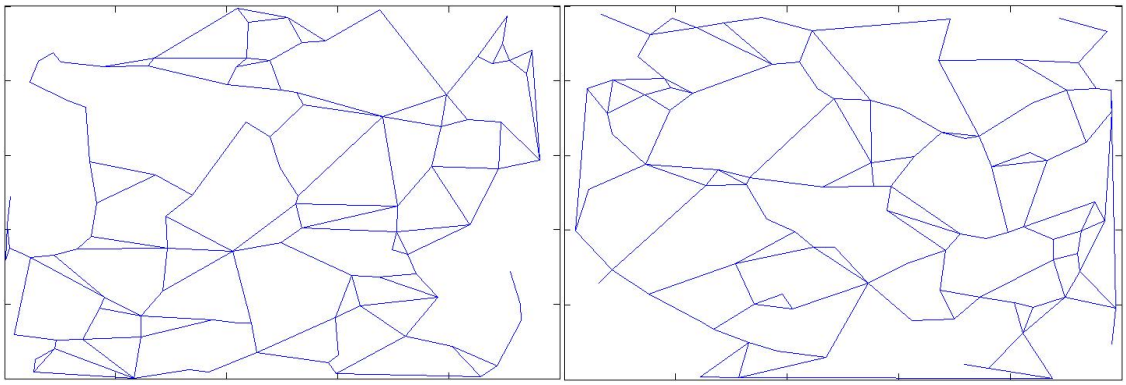

Figure 5.1 – Two Example Network Realisations from Experiment 1

With respect to the level of congestion, Figure 5.2 shows, perhaps unsurprisingly, that the average link V/C ratio increases monotonically as demand density increases. This increase also appears to be linear. It is hypothesised that this linearity is related to the averaging process within this performance measure.

It is also highlighted that the level of dispersion of performance values, across networks within each ensemble, increases as demand density increases, and that the distribution within most ensembles is positively skewed.

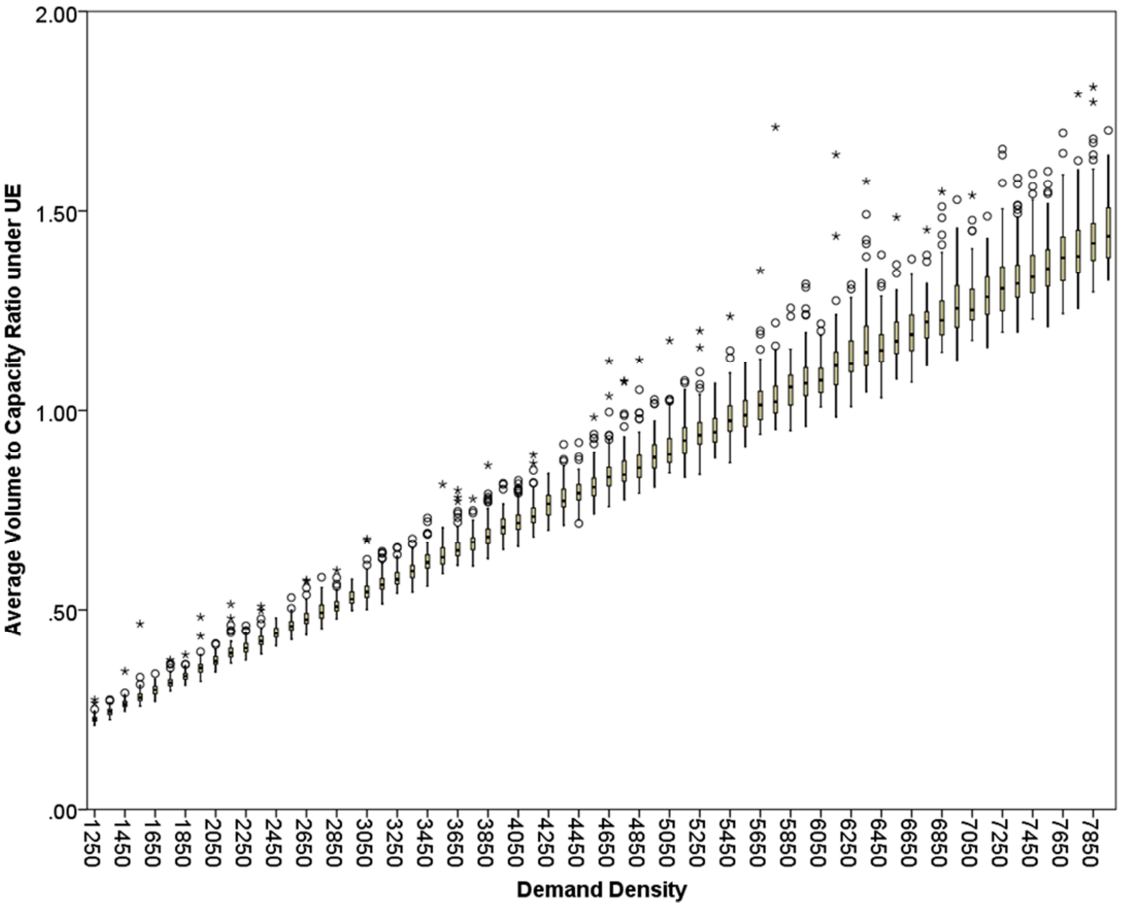

Figure 5.2 - Average link V/C ratio against Demand Density ϱ_{dem} ($A = 6.25$, $n = 100$, $\varrho_n = 16$, $m = 158$ and $M = 0.3$)

Figure 5.3 shows that the Price of Anarchy follows a unimodal pattern as demand density increases; with values initially increasing, before reaching a peak and then falling. The same broad pattern in values of the Price of Anarchy was uncovered by Youn et al. (2008) for increasing demand in random, scale-free, small-world and lattice networks. The dispersion of values of the Price of Anarchy within each ensemble appears to be much wider than is apparent in Figure 5.2, which suggests that the Price of Anarchy is more sensitive to stochastic variations in network structure. Figure 5.3 also shows that the level of dispersion within each ensemble initially increases as demand density increases, peaks and then decreases as values of the Price of Anarchy begin to fall.

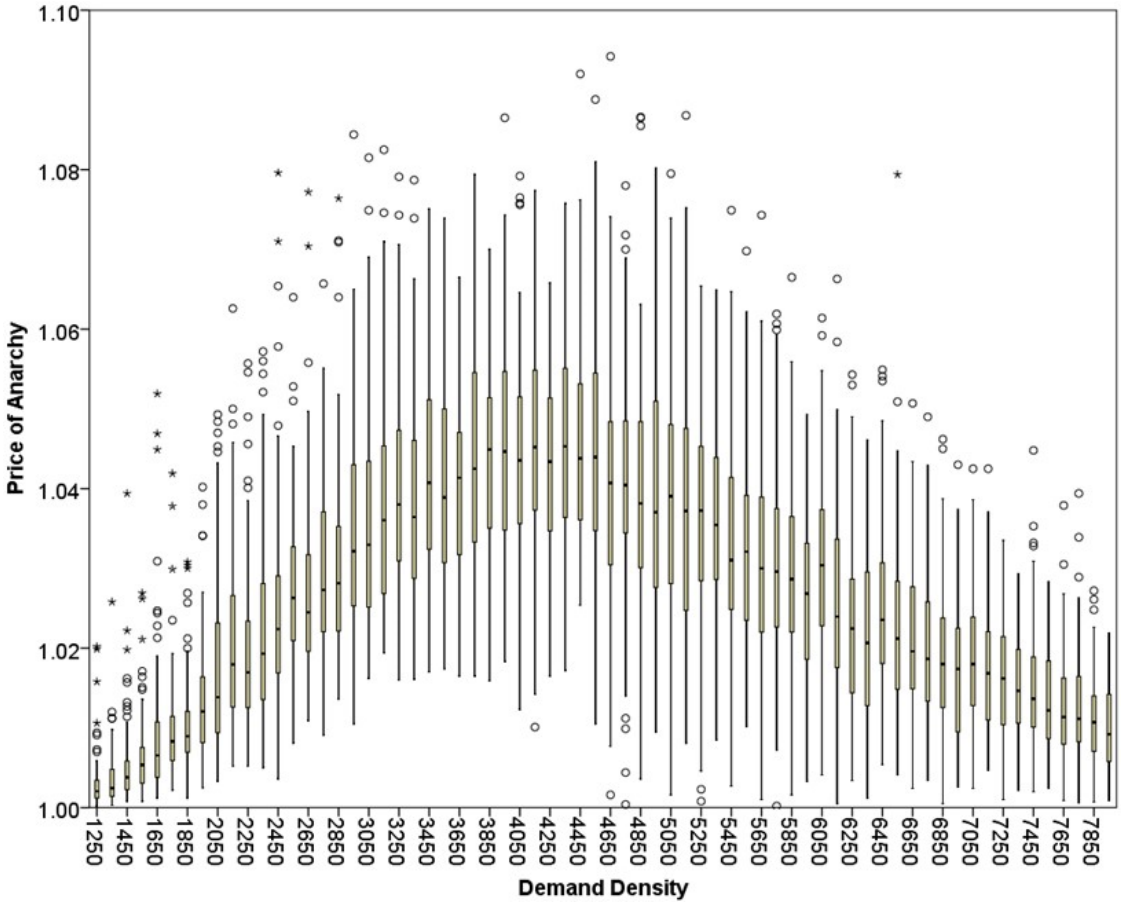

Figure 5.3 - Price of Anarchy against Demand Density ϱ_{dem} ($A = 6.25$, $n = 100$, $\varrho_n = 16$, $m = 158$ and $M = 0.3$)

5.3.2 Experiment 2: Network Size

Figure 5.5 and Figure 5.6 present the variation of the average link V/C ratio and the Price of Anarchy as network size is increased, between an ensemble of networks with $n = 20$ nodes in a domain size of $A = 1.25\text{km}^2$ and an ensemble of networks with $n = 500$ nodes in a domain size of $A = 31.25\text{km}^2$. Typical examples of networks at these two extremes and in the middle of the spectrum are shown in Figure 5.4. In each ensemble, node density was fixed at $\varrho_n = 16$, meshedness was fixed at $M = 0.3$ and demand density per km^2 was fixed at $\varrho_{dem} = 4350$. Note that although the *density* of travel demand was fixed, *total* travel demand still increases as the size of the network increases across this spectrum of networks; this is a consequence of the assumptions described in section 4.3. It therefore follows that the ratio of total travel demand $\sum_r q_r$ to total network supply $\sum_i m \times k_i \times Q_i$ is the same across each network ensemble in this network spectrum.

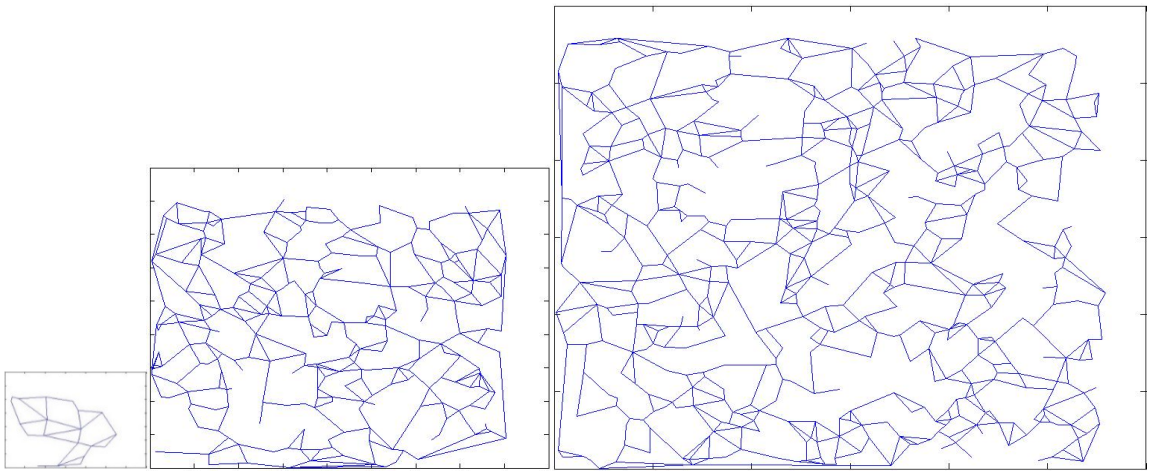

Figure 5.4 - Example Network Realisations from Network Ensembles at the lower end (left), in the middle (centre) and at the upper end (right) of the Network Spectrum in Experiment 2

With respect to congestion, Figure 5.5 shows that the average link V/C ratio increases monotonically as network size increases. It is also highlighted that the level of dispersion across networks within each ensemble increases as network size increases, and that the distribution within most ensembles is positively skewed.

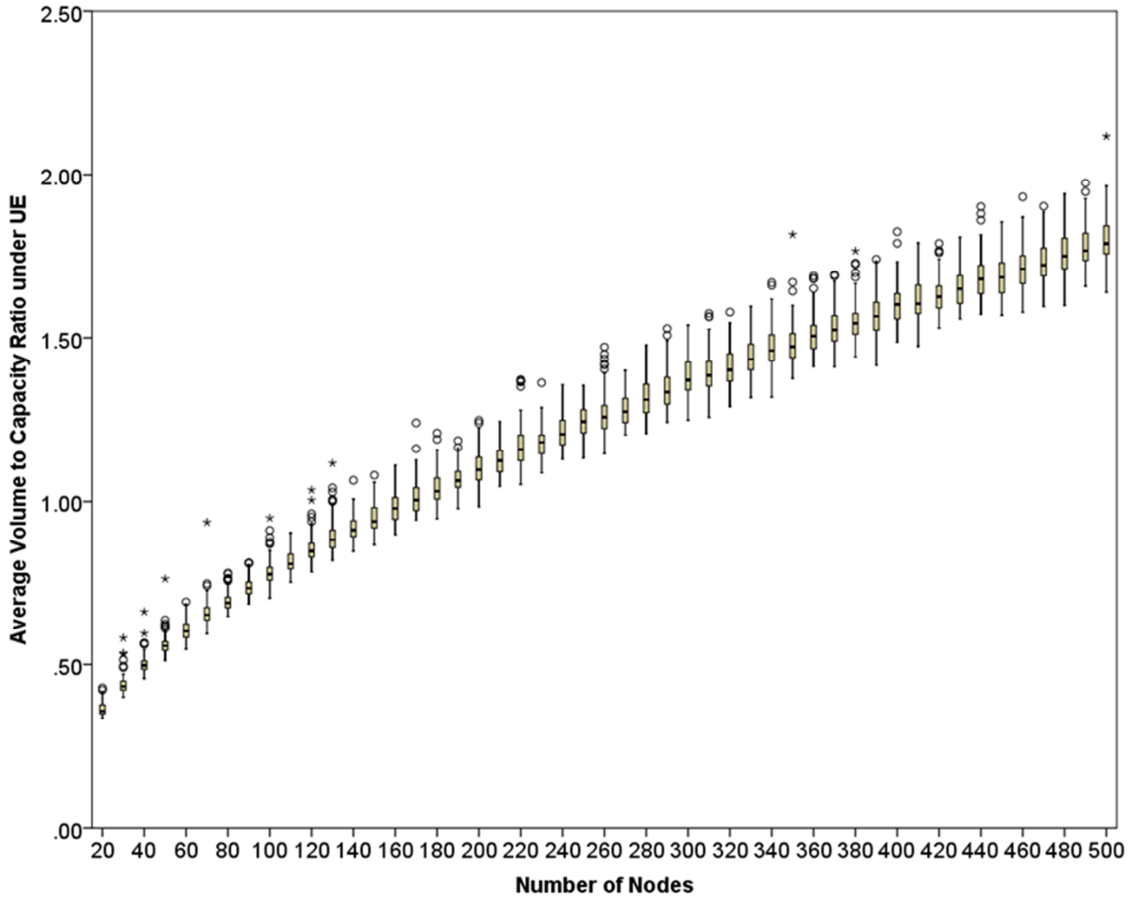

Figure 5.5 - Average link V/C ratio against Network Size ($\varrho_n = 16$, $M = 0.3$ and $\varrho_{dem} = 4350$)

Similarly to the results of the first experiment, Figure 5.6 shows that Price of Anarchy has a unimodal pattern of variation as network size increases. Again the dispersion of values of the Price of Anarchy within each ensemble is wider than is apparent for the average link V/C ratio. The level of dispersion also increases as network size increases and then gradually dissipates, mirroring the broad pattern in the median values of the Price of Anarchy.

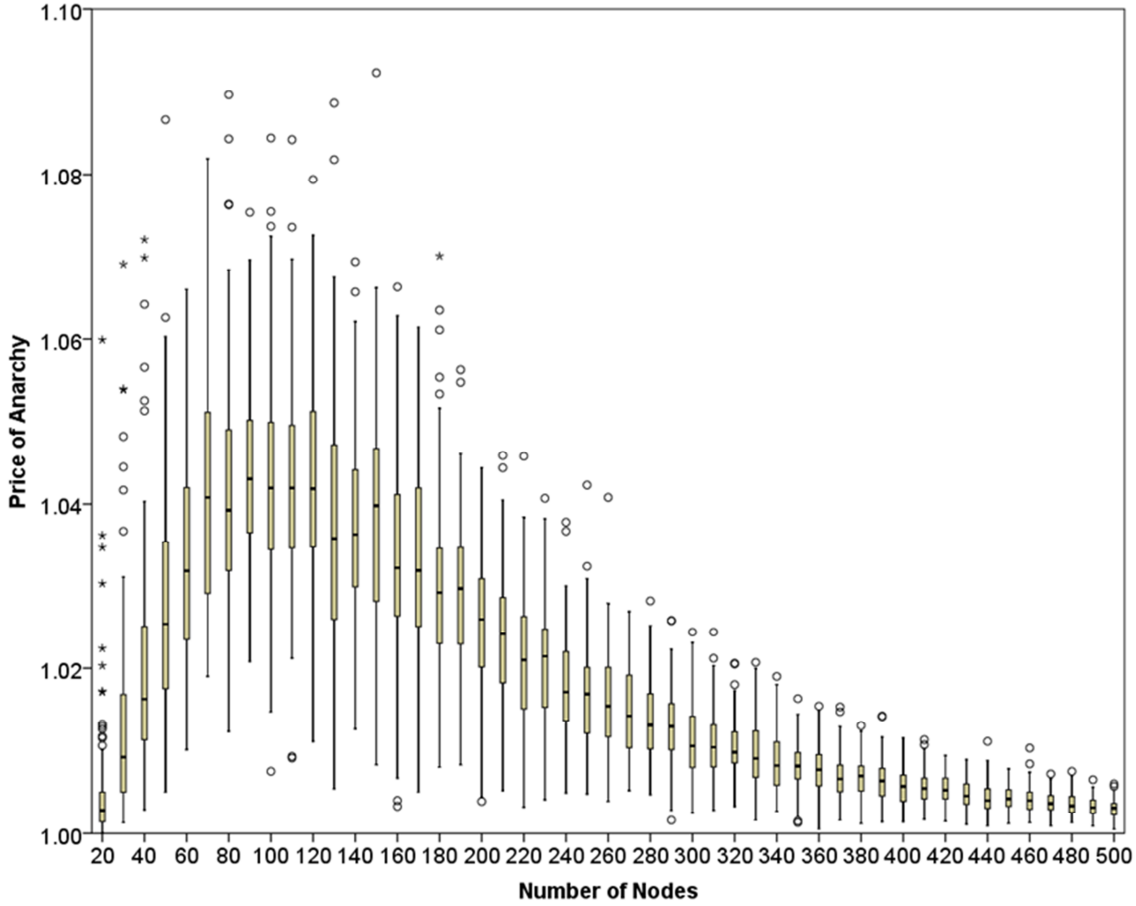

Figure 5.6 – Price of Anarchy against Network Size ($\varrho_n = 16$, $M = 0.3$ and $\varrho_{dem} = 4350$)

5.3.3 Experiment 3: Network Density

Figure 5.8 and Figure 5.9 present the variation of the average link V/C ratio and the Price of Anarchy as network density is increased, between an ensemble of networks with $\varrho_n = 3.2$ nodes per km^2 and an ensemble of networks with $\varrho_n = 48$ nodes per km^2. Typical network examples at the extremes and in the middle of this spectrum are shown in Figure 5.7. In each ensemble, the domain size was fixed at $A = 6.25$km^2, meshedness was fixed at $M = 0.3$ and demand density was fixed at $\varrho_{dem} = 4350$. In this experiment, in contrast to the second experiment, the total level of demand remains unchanged across the spectrum because the domain size is fixed. However, the amount of demand per OD pair falls because the number of OD pairs increases as the number of nodes increases.

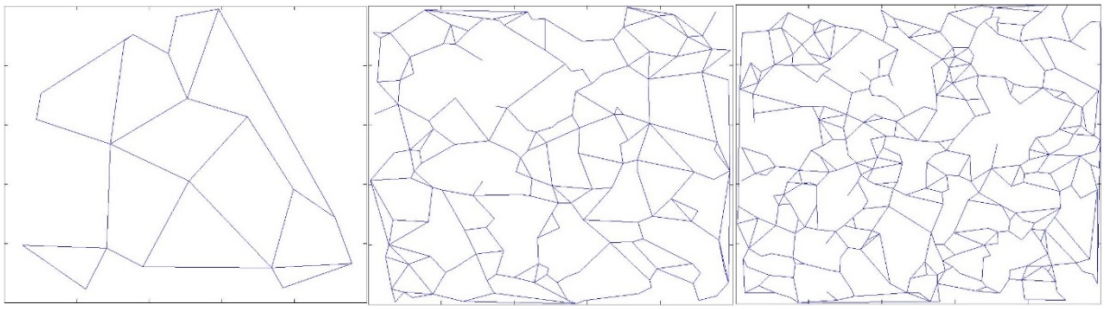

Figure 5.7 - Example Network Realisations from Network Ensembles at the lower end (left), in the middle (centre) and at the upper end (right) of the Network Spectrum in Experiment 3

With respect to the average level of congestion, Figure 5.8 shows that the average link V/C ratio falls as demand density increases. It is also noted that the levels of dispersion across the networks within each ensemble decrease as network density increases.

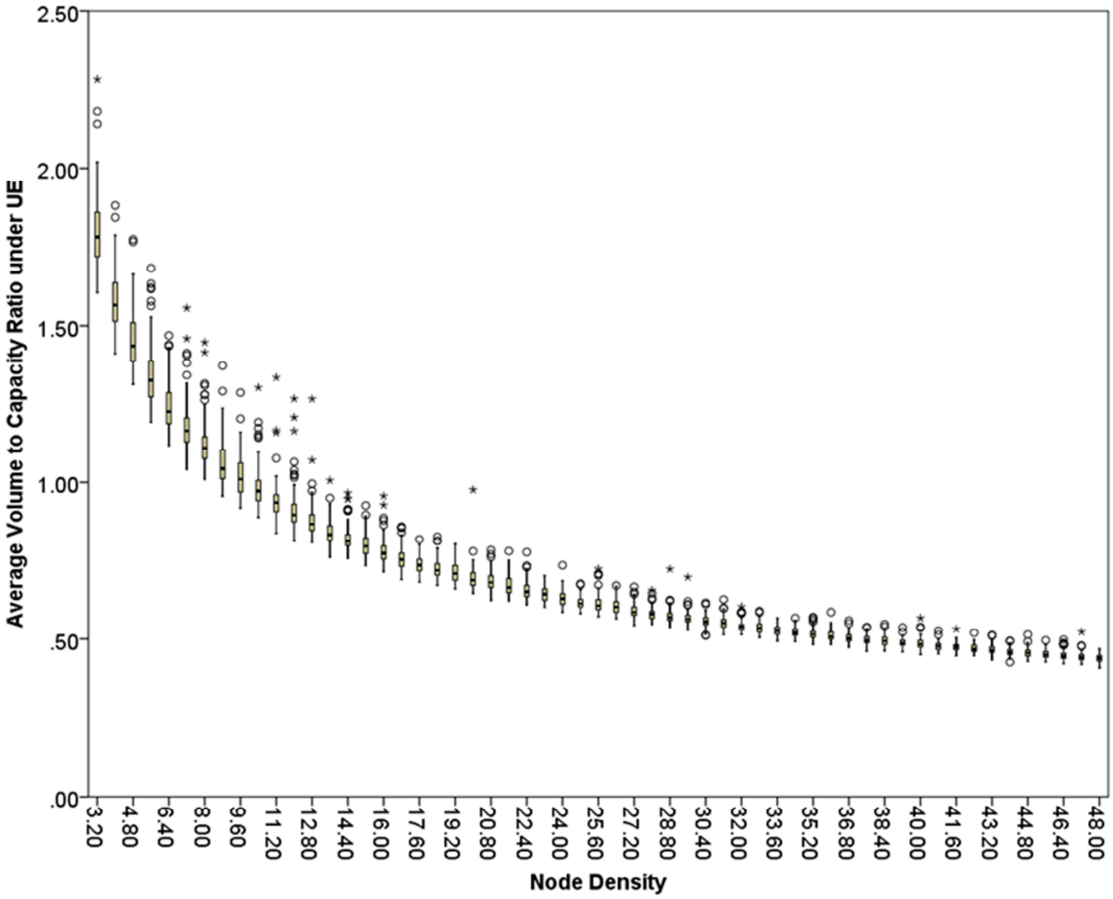

Figure 5.8 - Average link V/C ratio against Node Density ϱ_n ($A = 6.25$, $M = 0.3$ and $\varrho_{dem} = 4350$)

Similarly to the first two experiments, Figure 5.9 again shows a unimodal pattern in the Price of Anarchy as demand density increases. However, it is noted that the rate of decay in the Price of Anarchy is much shallower than is shown in either Figure 5.3 or Figure 5.6. The levels of dispersion of values of the Price of Anarchy within each ensemble initially increase, as node

density increases and then gradually dissipate, again mirroring the overall trend in the Price of Anarchy.

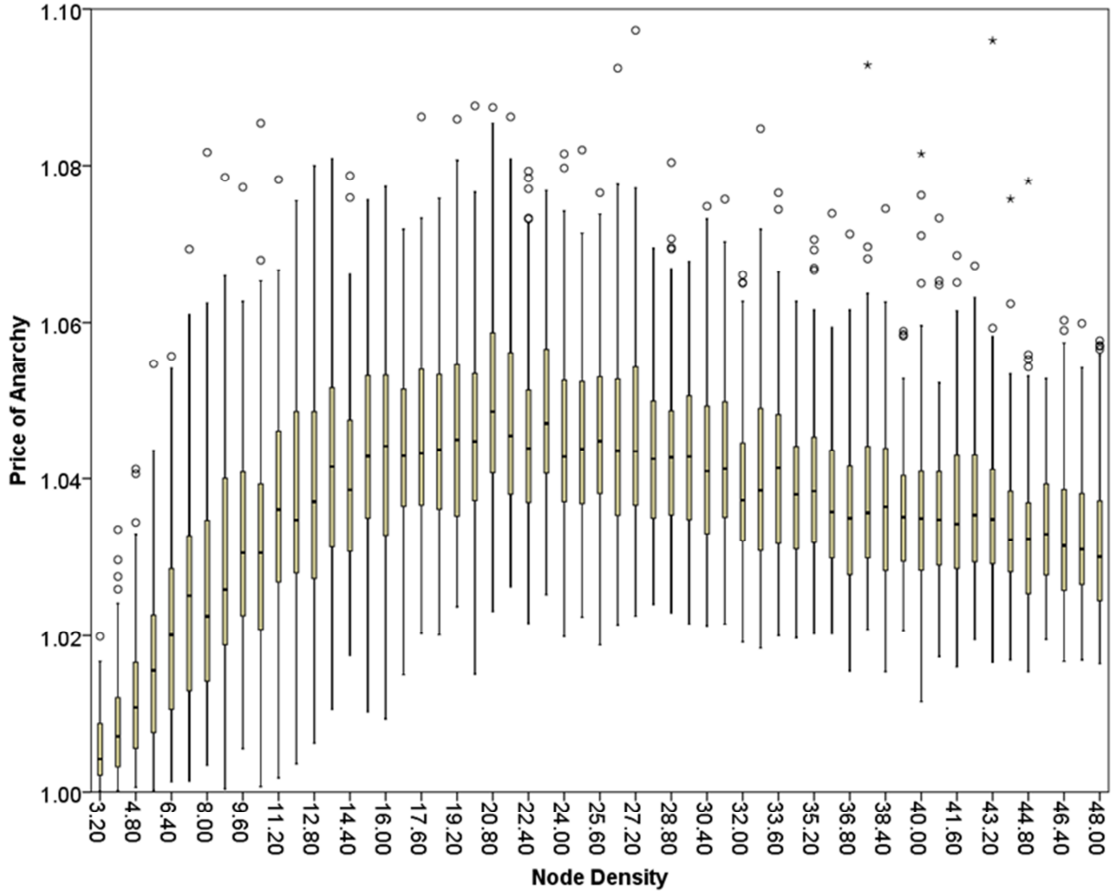

Figure 5.9 - Price of Anarchy against Node Density ϱ_n ($A = 6.25$, $M = 0.3$ and $\varrho_{dem} = 4350$)

5.3.4 Experiment 4: Network Connectivity

Figure 5.11 and Figure 5.12 present the variation of the average link V/C ratio and the Price of Anarchy as network connectivity is increased between an ensemble of networks with meshedness $M = 0$ and an ensemble of networks with meshedness $M = 0.97$. Typical network examples of these extremes and a network in the middle of the spectrum are shown in Figure 5.10. In each ensemble, the domain size was fixed at $A = 6.25\text{km}^2$, the number of nodes was fixed at $n = 100$ and demand density was fixed at $\varrho_{dem} = 4350$. In contrast to the second and third experiments, the total level of demand and the demand per OD pair both remain unchanged across the network spectrum in this experiment.

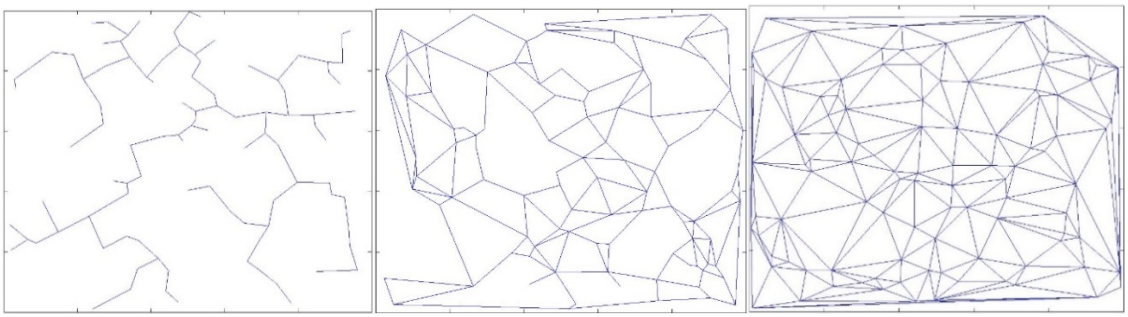

Figure 5.10 - Example Network Realisations from Network Ensembles at the lower end (left), in the middle (centre) and at the upper end (right) of the Network Spectrum in Experiment 3

With respect to the level of congestion, Figure 5.11 shows that the average link V/C ratio falls as network connectivity increases. Figure 5.11 also shows that the dispersion of congestion levels across network realisations is significantly different at different levels of connectivity; broadly, dispersion decreases as the level of network connectivity increases.

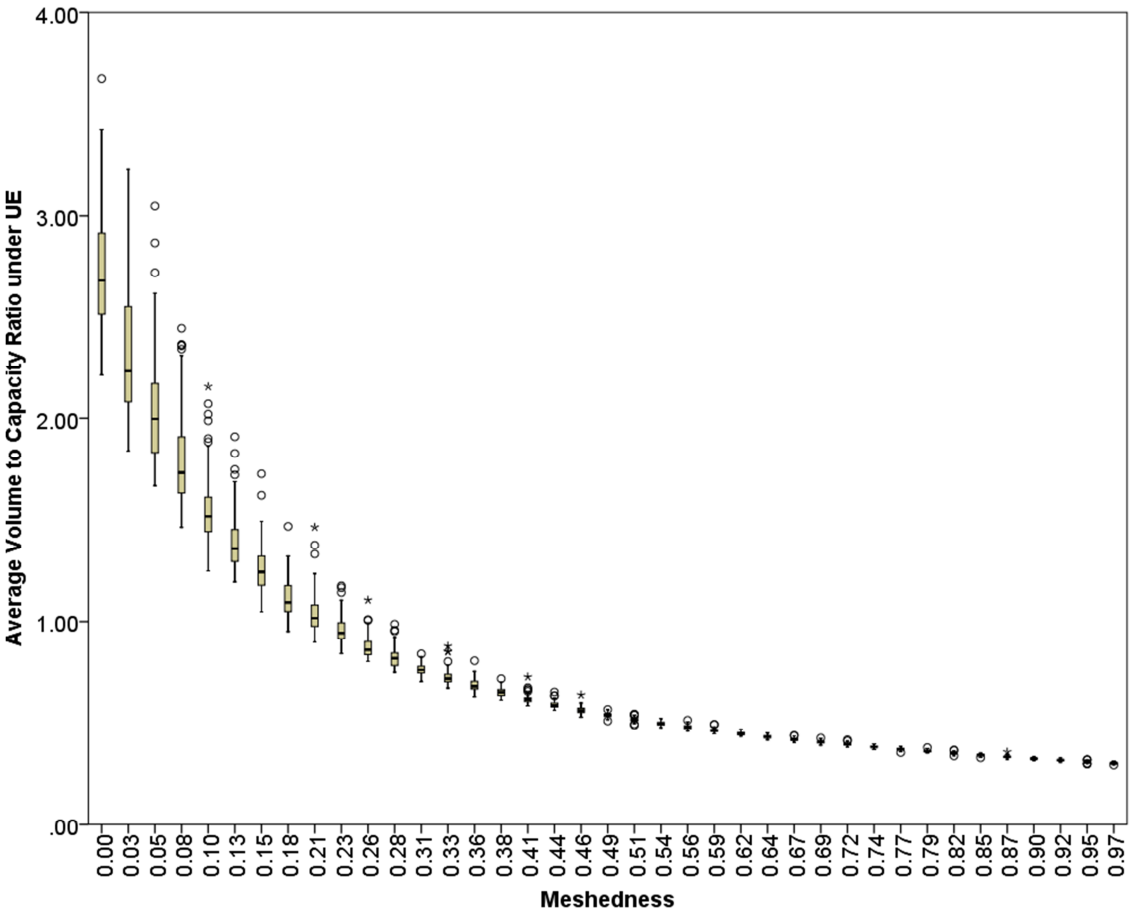

Figure 5.11 - Average link V/C ratio against Meshedness M ($A = 6.25$, $n = 100$, $\varrho_n = 16$ and $\varrho_{dem} = 4350$)

Similarly to the three previous experiments, the variation of the Price of Anarchy with network connectivity has a unimodal pattern, which reaches a peak median value at $M = 0.36$. As has also been shown previously, the level of dispersion of the Price of Anarchy across networks

within each ensemble is positively correlated with higher values of the Price of Anarchy across the spectrum.

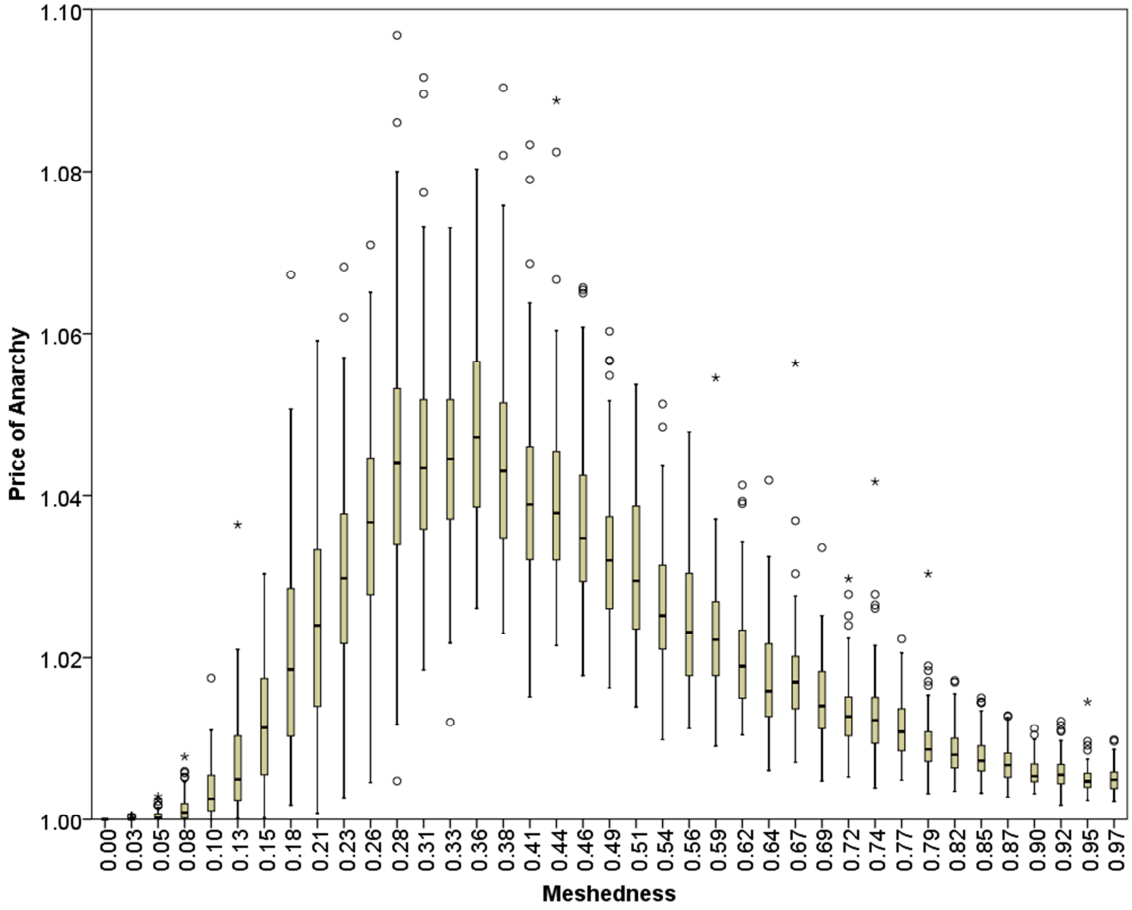

Figure 5.12 - Price of Anarchy against Meshedness M ($A = 6.25$, $n = 100$, $\varrho_n = 16$ and $\varrho_{dem} = 4350$)

5.4 Discussion

The results presented in section 5.3 illustrate that numerical investigations that follow the proposed experimental component of the investigative framework provide greater insight into how network structure affects performance than previous approaches. In particular, the results suggest the existence of clear relationships between specific aspects of network structure and performance indicators, which motivate further research questions around understanding the mechanisms that underpin the variations shown. Chapter 6 provides an example of such an analysis. The figures presented in section 5.3 contrast with the figures produced by previous studies in network science, such as those shown in Figure 2.17 and Figure 2.18, which, for the most part, did not present the variation of performance between networks within network ensembles and also did not provide any indication of how performance varies with respect to specific aspects of supply structure.

Turning to what the figures of the preceding section actually show, it can be seen that there are several similarities in the broad patterns exhibited by the two performance indicators

across the four experiments. For example, the first and second experiments both show that the average link V/C ratio increases monotonically with respect to increases in demand density and network size; whereas the third and fourth experiments show that this measure decreases monotonically with respect to increases in network density and network connectivity. All four experiments also show a unimodal pattern for the variation of the Price of Anarchy.

These similarities exist because all four experiments actually explore different ways of adjusting the balance between the total amount of travel demand and the total amount of network supply. In the first experiment, network supply remains unchanged as total demand increases. In the second experiment, total network supply and total travel demand increase together. Whereas in the third and fourth experiments, total demand is fixed whilst network supply is increased.

With respect to the average link V/C ratio, it therefore follows that the first, third and fourth experiments actually pick up on the same simple causal mechanism that governs the variation in this measure; this being that an increase in the ratio of total demand to total network supply leads to an increase in congestion, regardless of whether it is achieved by increasing demand or decreasing the amount of network supply, whilst a decrease in the ratio of total demand to total capacity leads to a decrease in congestion, regardless of whether it is achieved by decreasing demand or increasing capacity. This mechanism does not apply to the second experiment because total demand and total network supply increase at the same rate. The reason that the average link V/C ratio increases across this spectrum is that, as network size increases, the total volume of flow using routes that pass through the geometric centre of the domain also increases, and links in the geometric centre of the domain therefore become increasingly congested because their capacities remain fixed. Evidence for this effect is shown in Figure 5.13 and Figure 5.14, which display the spatial distribution of average link V/C ratios for the ensemble of networks generated from input parameters $n = 100$, $A = 6.25$, $m = 158$ and $\varrho_{dem} = 4350$. Note that this ensemble is a member of the network spectrums generated for each experiment presented in section 5.3. These plots were created by subdividing the domain into a grid of squares sized 0.1x0.1km, then calculating the average of all V/C ratios of those links in each network whose downstream nodes fall within each square.

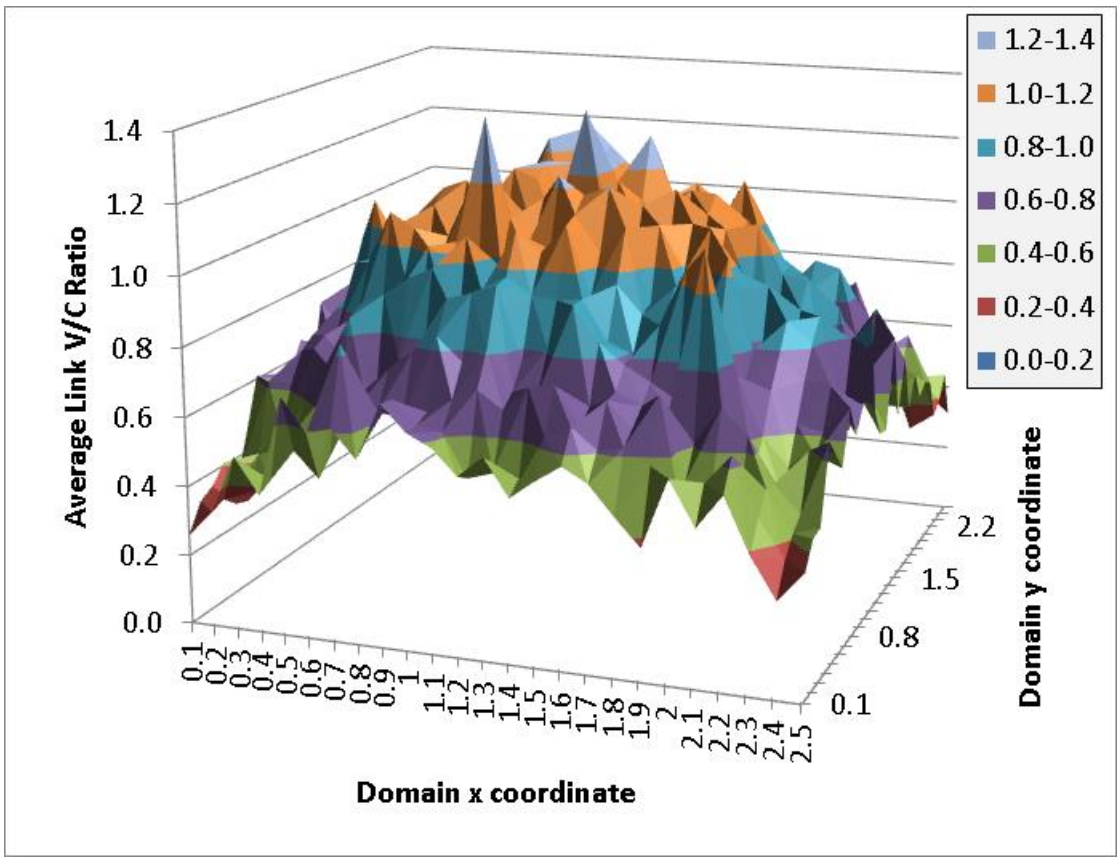

Figure 5.13 - Spatial Distribution of Average Link V/C Ratio - 3d view

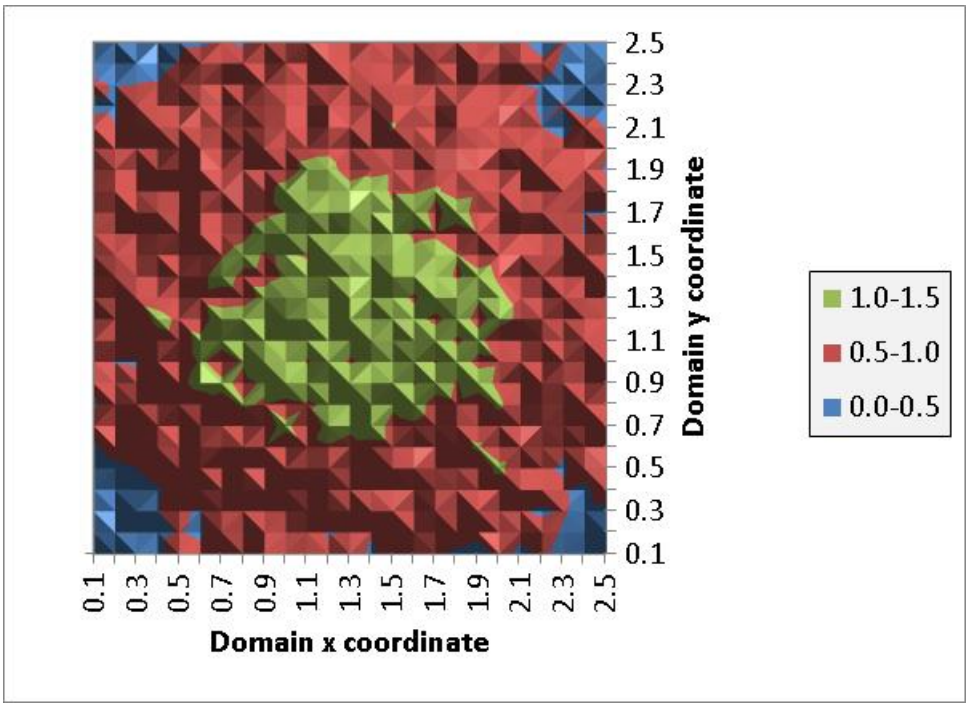

Figure 5.14 - Spatial Distribution of Average Link V/C Ratio - 2d view

Explanations for what happens to the average link V/C ratio in the right-hand limits of the four experiments, as the structural measure is increased beyond the limits explored numerically, are hampered by the computational restrictions of the computing environment in which the experiments were ran. However, it is possible to put forward some hypotheses.

To begin, consider the formula for the average link V/C ratio measure. A simple derivation, see equation (7), shows that this measure is dependent upon the sum of all link flows on network links divided by the number of links. This derivation works because capacity was assumed to be uniformly distributed across each network.

$$\text{Average Link V/C Ratio} = \frac{\sum_i \left(\frac{x_i}{cap_i}\right)}{m} = \frac{\sum_i \left(\frac{x_i}{800}\right)}{m} = \frac{1}{800} \times \frac{\sum_i x_i}{m} \qquad (7)$$

Within this formulation, it can also be seen that the sum of all link flows on network links is dependent upon total travel demand multiplied by the average number of used links per unit of demand. This statement holds because travel demand was assumed to be uniformly distributed in each network. These statements can be used to hypothesise what happens in the right-hand limits for each experiment.

In experiment 1, the domain size, number of nodes and number of links were fixed with respect to demand density. It follows from this that the denominator of equation (7) has a fixed value across the network spectrum. Turning to the numerator of equation (7), equation (3) shows that total travel demand increases linearly with respect to demand density in this experiment. The right-hand limiting behaviour is therefore dependent on how the average number of used links per unit of demand changes across the spectrum; a proxy for this behaviour is shown in Figure 5.15, which shows the average distance travelled per unit of demand. This figure shows that the distance travelled per unit of demand also increased across the spectrum, albeit slowly. This increase occurs because demand on shorter routes at lower demand levels is progressively pushed onto longer routes by congestion as demand increases. This trend of increasing distance travelled per unit of demand cannot continue because the network is finite in size. Therefore, in the limit of high demand, the distance travelled per unit of demand must plateau. At this point the numerator of equation (7) would continue to increase in a linear relationship with respect to demand density. It is therefore hypothesised that the linear trend shown in Figure 5.2 should continue as demand density increases.

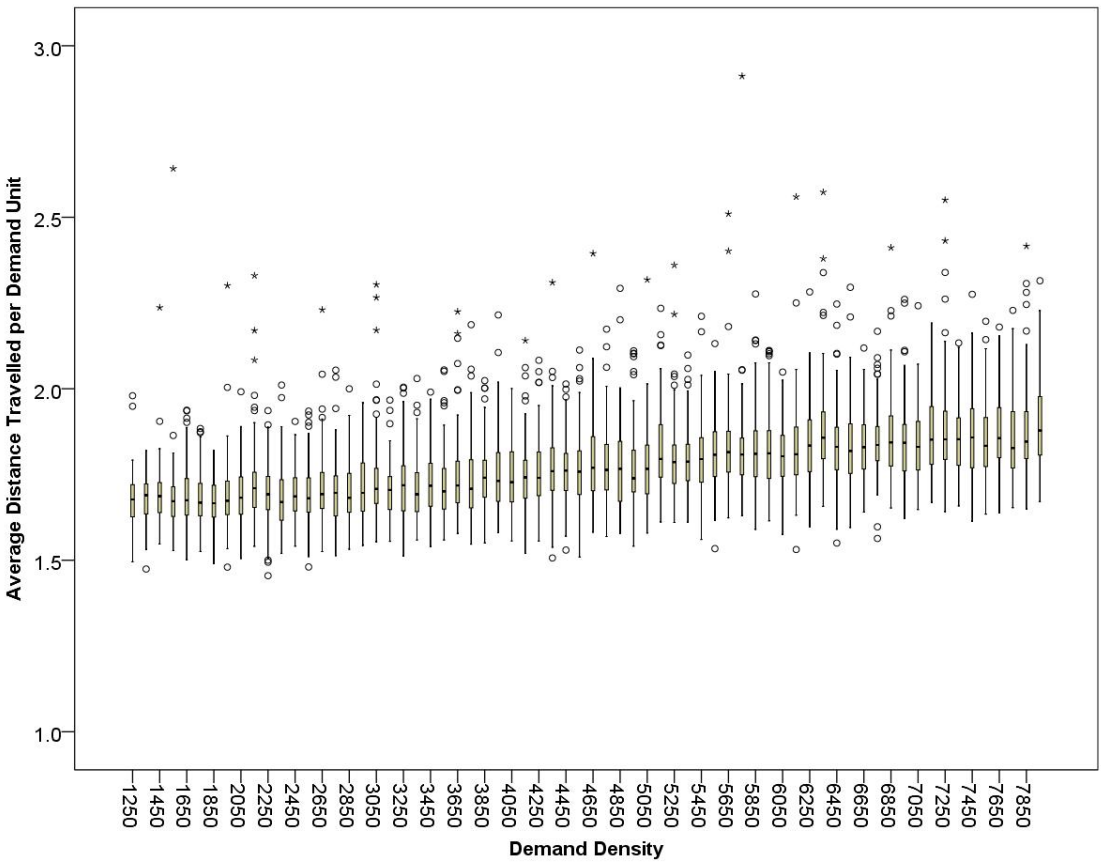

Figure 5.15 - Average distance travelled per unit of demand in experiment 1

In experiment 2, the domain size, number of links, demand density and, as consequence, total demand all increased linearly with respect to the number of nodes. It therefore follows that both the numerator and denominator of equation (7) have constituent parts that increase linearly with respect to the size of the network, albeit at different rates. The behaviour of the average link V/C ratio is therefore dependent upon how the average distance travelled per unit of demand varies with respect to network size; this behaviour is shown in Figure 5.16,which reveals an increasing relationship. However, it is obvious that this quantity would to continue to increase with respect to network size, precisely because the size of the network is increasing. A more interesting quantity to look at is the average distance travelled per unit of demand normalised by average shortest path length across the spectrum, which strips out network size effects from the measure. This quantity, shown in Figure 5.17, has a shallower gradient than average distance travelled per unit of demand and looks to be flattening out. Should this quantity indeed flatten out in the right-hand limit of large networks, this would mean that average distance travelled per unit of demand will increase only as a consequence of increases in network size. It would therefore follow from equation (7) that the average link V/C ratio will continue to increase in a linear fashion.

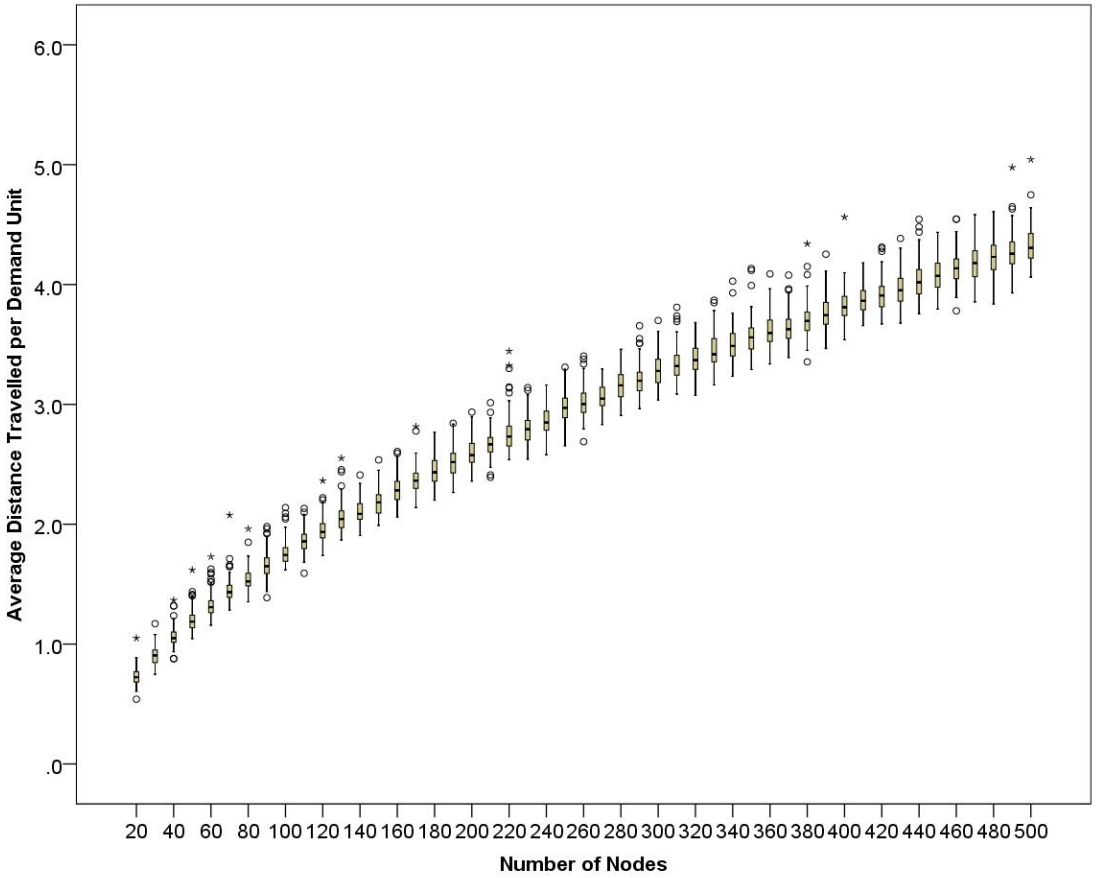

Figure 5.16 - Average distance travelled per unit of demand in experiment 2

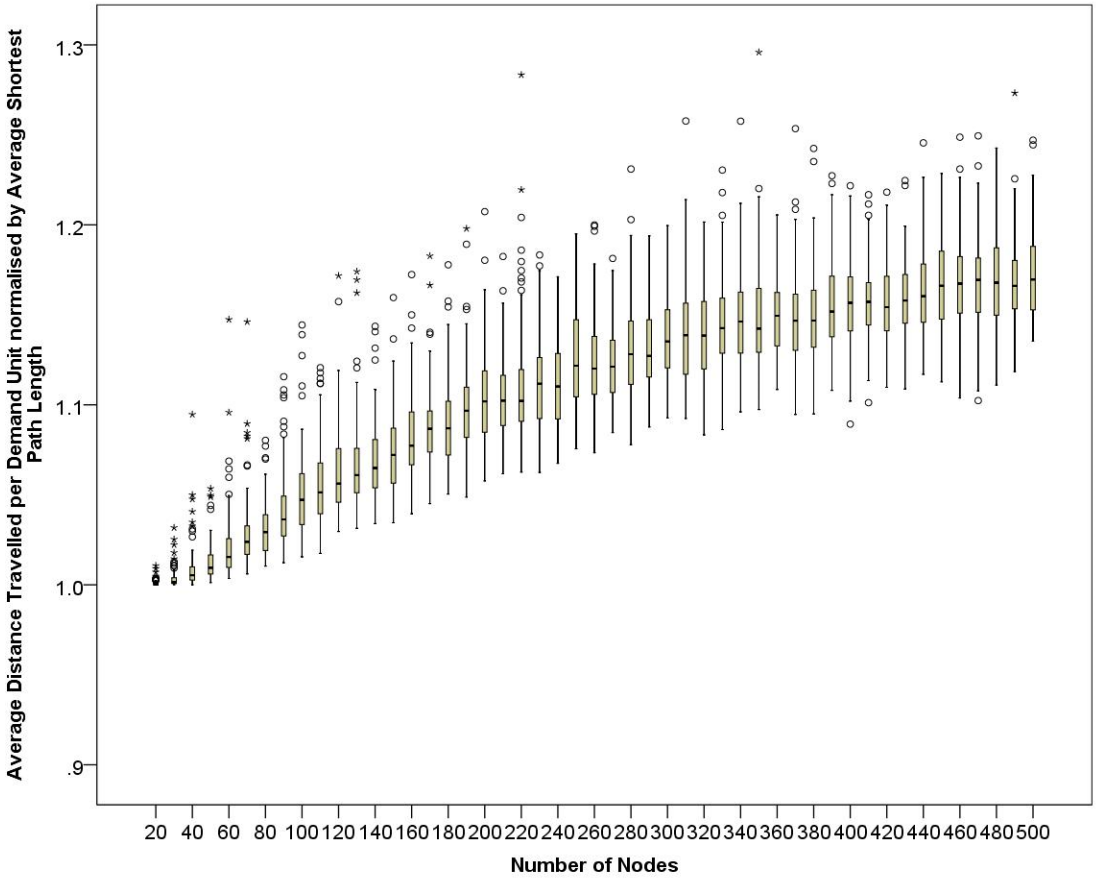

Figure 5.17 - Average distance travelled per unit of demand normalised by average shortest path length in experiment 2

In experiment 3, the domain size, demand density and, as a consequence, total demand were all fixed with respect to node density; whereas the number of nodes and the number of links increase linearly with respect to node density. Of the three constituent components of average link V/C ratio shown in equation (7), it therefore follows that the denominator increases linearly with respect to node density whilst total demand in the numerator is a fixed constant. Figure 5.18 shows that the average distance travelled per unit of demand falls slowly across this spectrum. As more nodes are added to the domain, it is clear that, in the right-hand limit, this quantity must level out because the shortest possible distance travel distance for any OD pair is that defined by the length of a straight line between those two nodes. It therefore follows that the numerator of equation (7) tends towards a constant in the limit of high node density, which means that the average link V/C ratio will therefore tend towards zero.

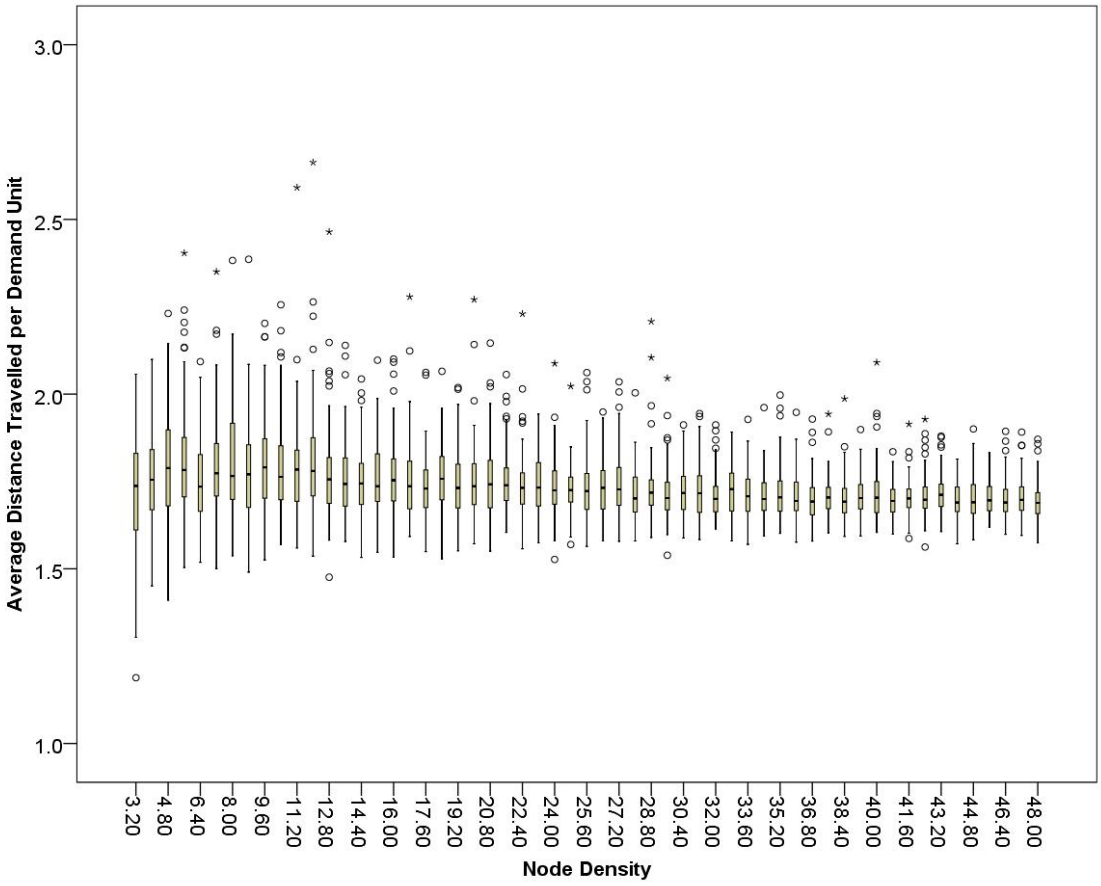

Figure 5.18 - Average distance travelled per unit of demand in experiment 3

In experiment 4, the planarity constraint of the network model means that the right-hand limit occurs at a meshedness value of $M = 1$, where the number of links $m = 3n - 6$. The full extent of the relationship between average link V/C ratio and connectivity is therefore shown in Figure 5.11.

Turning to the results for the Price of Anarchy, explanations for the unimodal patterns that are common to all four experiments are less obvious. However, these patterns do appear to be

strongly connected to how the level of congestion changes across each spectrum as the ratio of total demand to total network supply changes. In particular, comparisons of the graphs for the two performance indicators in each of the four experiments reveal that the peak regions of the Price of Anarchy coincide with values of the average link V/C ratio between approximately 0.5 and 0.8. It is also notable that the dispersion of values of the Price of Anarchy across networks within each ensemble is at its highest in those ensembles that produce the peak in the aforementioned unimodal pattern. This feature is also identifiable in the results of the numerical experiments of Youn et al. (2008), which were highlighted at the end of section 2.4.4 and which focus on the variation of the Price of Anarchy in three single OD sub-networks of the road networks in Boston, London and New York. The results for these experiments are shown in Figure 5.19.

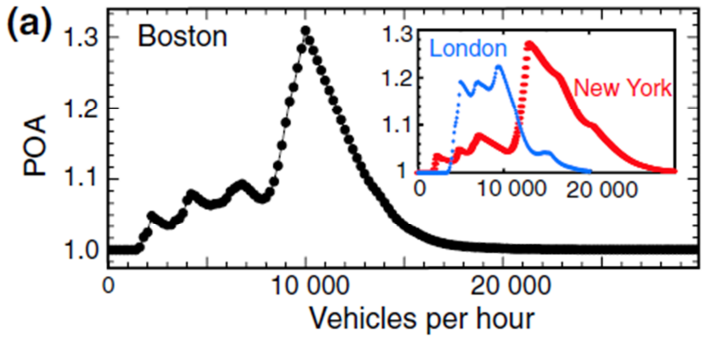

Figure 5.19 – Price of Anarchy against Demand for three real networks (Figure 3a, Youn et al. (2008))

In each city, it can be seen that there are broadly three identifiably distinct regions of behaviour: an initial region in which the Price of Anarchy is one; an intermediate region of fluctuations; and a final region of decay, which has a similar characteristic shape across all three networks. Yet, focussing on the detail of the individual graphs, the patterns for each city are obviously different. For example, the graphs for Boston and New York have single dominant peaks, which are both higher than the peak reached in London. Whereas, the graph for London remains closer to its maximum value for a longer interval of demand than in either Boston or New York. It is also evident that the Price of Anarchy is not a smooth function of demand; the peak in Boston is a prominent example of this feature. The similarities in the general behaviour of the Price Anarchy across the three cities as travel demand increases suggests that there may be common mechanisms that drive this variation. The differences between the three cities also suggests a reason for why features of the form shown in Figure 5.3 can appear when several Price of Anarchy graphs are grouped together.

The numerical experiments of Youn et al. (2008) and those presented in section 5.3 illustrate how the Price of Anarchy can take different values, at different levels of demand, in different road traffic networks. However Youn et al. (2008) does not provide an explanation for the

variation shown. The findings of theoretical studies highlighted in section 2.4.3.2, such as those of Roughgarden (2003) and Correa et al. (2008), are also of little explanatory use here because they reveal only the maximum value that the Price of Anarchy could reach across broad families of road traffic networks. An explanation for how the Price of Anarchy varies with travel demand is provided in chapter 6.

An additional feature of the results for the Price of Anarchy that is notable across all four experiments is of how its maximum value is quite low. Indeed, for the BPR cost functions used in these experiments, which have a maximum power $\beta = 4$, the upper bound for the Price of Anarchy of Roughgarden (2003), shown in equation (2), is approximately 2.15. This is significantly larger than the highest value of the Price of Anarchy of 1.05 across all of the results shown in section 5.3. This feature of the Price of Anarchy has also been identified in the literature; for example, Correa et al. (2008) cited two numerical studies by Jahn et al. (2005) and Qiu et al. (2006) in which values of the Price of Anarchy were also significantly lower than the upper bounds of Roughgarden (2003). The same is also true for the example from Youn et al. (2008) shown in Figure 5.19, where the largest value of the Price of Anarchy of approximately 1.3 is significantly smaller than the upper bound; derived using the results of Roughgarden (2003), of approximately 3.5 for such networks. An explanation for this feature of the Price of Anarchy is explored in chapter 7.

Before moving on to the next chapter, a final remark is made with respect to the realism of the results presented in this chapter. All of the figures presented for the average link V/C ratio measures show that many network ensembles have ratios greater than one. However, a V/C ratio greater than one does not make sense physically because it would imply significant queuing and potentially blocking back effects on other links. This highlights a limitation of the modelling approach that was used. An additional effect of congestion that is not represented in the modelling approach used here is that very high travel times also have feedback effects on the volume and configuration of travel demand; for example, because travellers choose to divert to other destinations, use a different mode or choose not to travel at all. The inclusion of such effects would change the patterns shown in the figures in the preceding sections. Indeed, it is hypothesised that, with the inclusion of such effects, the V/C ratio may reach a peak and then stabilise as travellers make other travel choices. This remark highlights that a more accurate model of road generation should include the generation of supply and demand structures and their feedback effects on each other. Such models do not currently exist in network science but they have been explored in the context of Land-Use and Transport Interaction models in transportation science; for example, see Wegener (2004).

6 Mechanisms that Govern the Variation of the Price of Anarchy with Travel Demand

6.1 Introduction

Motivated by the results of the numerical experiments presented in the previous chapter, this chapter explores how the Price of Anarchy varies with respect to travel demand; thereby addressing the fifth step of the investigative framework proposed in chapter 3.

More specifically, focussing on the general setting of traffic networks with multiple OD pairs and continuous, differentiable, separable and strictly increasing link cost functions, this chapter reveals the source of the variations shown in Figure 5.3 and Figure 5.19: namely, that as demand increases there are expansions and contractions in the set of routes (for each OD pair) that are of minimum cost under the UE model and of minimum marginal cost under the SO model. The different effects of these expansions and contractions on the Price of Anarchy are characterised through a series of theorems and conjectures. This chapter also shows, in a special case of road traffic networks that have *BPR-like cost functions*[15] of the form $c_i = a_i + b_i x_i^{\beta}$, that there is a systematic relationship between link flows under UE and SO, and that, consequently, there is also a systematic relationship between levels of demand at which expansions and contractions in UE and SO route sets occur. Finally, this chapter conjectures that in this special case, the Price of Anarchy has power law decay for large demand, which explains the similarities in the shape of the decays shown across the three networks in Figure 5.19.

The novelty of the material presented in this chapter is that it provides a thorough and rigorous explanation for the nature of how the Price of Anarchy varies with respect to travel demand, which has, thus far, been missing from numerical studies in network science, such as those described in section 2.4.4.

The first section of this chapter sets out additional mathematical preliminaries and notation that are necessary for the analysis that follows. These complement the mathematical descriptions of the UE and SO models set out in section 4.4. Section 6.3 then characterises the existence of expansions and contractions in minimum (marginal total) cost route sets under UE and SO and also proves that these are equivalent to expansions and contractions in the sets of

[15] Note that the cost functions used in chapters 4 and 5 are of this form.

links for each OD pair that have non-zero flow under the *condition of proportionality*[16]. This section also describes the systematic relationship between link flows under UE and SO for the special case described above. Section 6.4 presents theoretical results and conjectures, which characterise the effects of expansions and contractions in route sets on Total Network Travel Cost under SO, Total Network Travel Cost under UE and the Price of Anarchy. Section 6.5 then presents four numerical examples, which illustrate the theory of the preceding sections and also provide numerical evidence to support those theoretical results that are presented without proof.

6.2 Additional Mathematical Preliminaries and Notation

Recall that, in this thesis, a road traffic network is represented by a directed graph $G(V, A)$, comprising a set of nodes V and a set of directed links A for which the costs of travel on each link $i \in A$ are represented by cost functions c_i. Recall also that travel demand is represented by a vector Q with entries q_r denoting the volume of travel on OD movements $r = 1, \dots, R$, and that each OD movement is served by a finite number $k = 1, \dots, \kappa_r$ of acyclic routes K^r, each with flows f_k^r.

In this chapter, the cost of travel under UE on each route $k \in K^r$ is denoted $C_k^r = \sum_i c_i(x_i)\, \delta_{i,k}^r$, where the $\delta_{i,k}^r$ terms form a link-path incidence matrix, which is denoted Δ. Similarly, the *marginal total travel cost* under SO on a route $k \in K^r$ is denoted $\tilde{C}_k^r = \sum_i \tilde{c}_i(x_i)\, \delta_{i,k}^r$, where $\tilde{c}_i = c_i + x_i \times dc_i/dx_i$. The *minimum OD travel cost under UE*, for the r^{th} OD movement is denoted $\pi_r = \min_{k \in K^r} C_k^r$. Similarly the *minimum marginal total travel cost under SO* for the r^{th} OD movement is denoted $\tilde{\pi}_r = \min_{k \in K^r} \tilde{C}_k^r$. Finally, note that the set of links that comprise a route $k \in K^r$ is denoted $I_k^r = \left\{ i \in A \middle| \delta_{i,k}^r = 1 \right\} \subset A$.

In order to guarantee the existence and uniqueness of link flows under UE and SO, the following assumption is presumed to hold throughout this chapter:

Assumption A1: For each link $i \in A$ in a traffic network G, the cost function c_i is a continuous, twice differentiable, positive, separable and strictly increasing function of link flow x_i. It is also assumed that $d^2 c_i/dx^2 \geq 0$, $\forall i \in A$ to guarantee the existence of a unique SO solution.

In addition to link flows, the mathematical programs described in Figure 4.2 also guarantee the uniqueness of route costs under the UE and SO principles. However, route flows f_k^r are, in general, not unique. In fact, there are typically an infinite number of possible route flow solutions $F = \{f_k^r\}$ that satisfy the above constraints. A uniquely identifiable route flow

[16] The usefulness of this equivalence is explained in section 6.3.3

solution F^*, which is important for some of the analysis that follows, is that defined by the *condition of proportionality*, which was first proposed by Bar-Gera and Boyce (1999) and is defined as follows.

Definition 6.1: *"The condition of proportionality states that the same proportions apply to all travellers facing a choice between a pair of alternative segments (PASs), regardless of their origins and destinations, where a segment is defined as a sequence of one or more links"* (Bar-Gera et al., 2012).

This route flow solution has the useful property that *"any route that can be used under the UE conditions will be used"* (Bar-Gera et al., 2012). Lu and Nie (2010) have shown that route flows under the condition of proportionality vary continuously with respect to travel demand Q. As the SO problem can be transformed into an equivalent UE problem, it follows that there also exists a unique SO route flow solution, which is denoted by $\tilde{F}^*$, that satisfies the condition of proportionality. In networks with only a single origin, the route flow solutions F^* and $\tilde{F}^*$ can be derived from the approach proportions produced by the Origin-Based Assignment (OBA) algorithm (Bar-Gera, 2002, Bar-Gera et al., 2012). In networks with multiple origins, these route flow solutions cannot be derived using OBA; the Traffic Assignment by Paired Alternative Segments (TAPAS) algorithm can be used instead (Bar-Gera, 2010, Bar-Gera et al., 2012).

6.3 The Existence of Expansions and Contractions in Minimum Cost Route Sets

This section characterises how the set of routes for an OD movement, which are of minimum cost under UE, or minimum marginal total cost under SO, can *expand* or *contract* in response to a perturbation in travel demand. This section begins with two network examples to illustrate this behaviour, and then provides definitions and notation to characterise the different types of expansions and contractions that can occur in general traffic networks. It is then shown that, under the condition of proportionality, an expansion (contraction) in the minimum cost route set (under UE or SO), for an OD movement, is equivalent to an expansion (contraction) in the set of links that have non-zero flow for that OD movement.

In the special case of traffic networks with cost functions $c_i = a_i + b_i x_i^{\beta}$ for which all links share a common power β, it is shown that there is a systematic relationship between link flows under UE and SO, and that, consequently, there is also a systematic relationship between the levels of demand at which expansions and contractions occur in minimum cost route sets under UE and SO.

6.3.1 Illustrative Examples

6.3.1.1 Example 1: Expansions in the Minimum Cost Route Sets under UE and SO

Consider a traffic network of N parallel links, serving a single OD pair with increasing demand $q > 0$, and with affine link cost functions of the form $c_i = a_i + b_i x_i$, where $a_i, b_i > 0$ and $a_i < a_{i+1}\ \forall i = 1, \dots, N$. In such a network, under UE and at sufficiently low levels of demand q, all flow uses only the cheapest route, which is provided by link 1. This holds for all values of $q > 0$ for which:

$$c_1(x_1^{UE} = q) \leq c_2(x_2^{UE} = 0) \iff a_1 + b_1 q \leq a_2 \iff q \leq \frac{(a_2 - a_1)}{b_1}$$

For values of $q > (a_2 - a_1)/b_1$, link 2 activates and $c_1(x_1^{UE}) = c_2(x_2^{UE})$. Both links therefore carry flow at UE and the *set of minimum cost routes* comprises links 1 and 2. As q increases from this threshold the set of minimum cost routes remains unchanged provided:

$$\begin{aligned} &c_1(x_1^{UE}) = c_2(x_2^{UE}) \leq c_3(x_3^{UE} = 0) \Leftrightarrow a_1 + b_1 x_1^{UE} \leq a_3 \Leftrightarrow \cdots \\ &\Leftrightarrow a_1 + b_1\left(\frac{(a_2 - a_1) + b_2 q}{b_1 + b_2}\right) \leq a_3 \Leftrightarrow q \leq \frac{a_3 - a_2}{b_2} + \frac{a_3 - a_1}{b_1} \end{aligned} \quad (8)$$

For values of q above the threshold shown in equation (8), link 3 activates and $c_1(x_1^{UE}) = c_2(x_2^{UE}) = c_3(x_3^{UE})$; i.e. the set of minimum cost routes comprises links 1, 2 and 3.

As demand continues to increase, the minimum OD cost of travel continues to increase and further links become members of the minimum cost route set. This process continues until, at a sufficiently large level of demand, all links in the network belong to this set. It can be shown that under UE, for a given $M < N$, the set of minimum cost routes comprises M links for all values of q satisfying equation (9).

$$\sum_{i=1}^{M-1} \frac{a_M - a_i}{b_i} < q \leq \sum_{j=1}^{M} \frac{a_{M+1} - a_j}{b_j} \quad (9)$$

A similar pattern emerges under SO: increasing demand causes a sequence of links to be added to the set of minimum cost routes. Although, as travellers consider the marginal link travel costs $\tilde{c}_i$, rather than c_i, when choosing routes; it is the set of routes of minimum *marginal total* cost that changes. For the above parallel link network the cost transformation $\tilde{c}_i$ yields equation (10).

$$\tilde{c}_i = c_i + \frac{dc_i}{dx_i} x_i = (a_i + b_i x_i) + (b_i) x_i = a_i + 2b_i x_i \quad (10)$$

The pattern of changes in the minimum marginal total cost route set under SO can therefore be obtained by redefining $b_i := 2b_i$ in the above UE derivation. It follows that under SO, for a

given $M < N$, the set of routes that are of minimum marginal total cost comprises M routes for all values of q satisfying equation (11).

$$\sum_{l=1}^{M-1} \frac{a_M - a_l}{2b_l} < q \leq \sum_{m=1}^{M} \frac{a_{M+1} - a_m}{2b_m} \tag{11}$$

It follows from the above that as demand increases, the order in which routes become minimum cost under UE is exactly the same as the order in which routes become minimum marginal total cost under SO. This follows for general multiple OD networks from the cost function transformation $\tilde{c}_i$.

This example illustrates how the set of minimum cost routes under UE, and the set of minimum marginal total cost routes under SO, can expand in response to an increase in demand. This example could also be used to demonstrate that the sets of minimum cost routes under UE and SO can also contract. This could be achieved by starting with high demand q, such that all N links belong to the minimum cost route set, and by gradually decreasing q towards zero. The example that follows in section 6.3.1.2 demonstrates, perhaps counter-intuitively, that the set of minimum cost routes, under UE and SO, can also contract in response to an *increase* in demand.

6.3.1.2 Example 2: Contractions in the Minimum Cost Route Sets under UE and SO

Consider the five link traffic network shown in Figure 6.1, which serves two OD pairs $O \rightarrow D1$ and $O \rightarrow D2$ as shown. Further suppose that the five links have the following affine link cost functions: $c_1 = 2 + x_1$, $c_2 = 3 + x_2$, $c_3 = 9 + x_3$, $c_4 = 1 + x_4$ and $c_5 = 1 + x_5$; and that demand on the $O \rightarrow D2$ movement is fixed at $q_{O \rightarrow D2} = 1$. There are two routes for each OD pair: for $O \rightarrow D1$, the routes are link $\{1\}$ and links $\{2,4\}$; for $O \rightarrow D2$, the routes are links $\{2,5\}$ and link $\{3\}$.

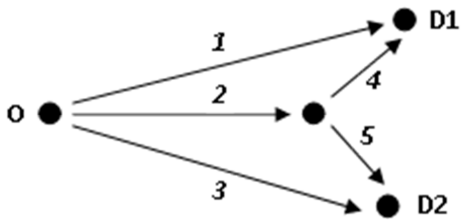

Figure 6.1 - Five Link Network with Two OD Pairs

Consider demand $q_{O \rightarrow D1}$ increasing from zero under SO. The variation of marginal total route costs under SO, for each of the four routes, with respect to $q_{O \rightarrow D1}$, is shown in Figure 6.2. In addition to providing further examples of expansions in the minimum cost route set; it can also be seen that, for $q_{O \rightarrow D1} < 11.5$, route $\{2,5\}$ is part of the minimum marginal total cost route set for OD movement $O \rightarrow D2$, but that, for demand $q_{O \rightarrow D1} > 11.5$, this route ceases to be a

member of this set. This example therefore demonstrates that the set of minimum marginal total cost routes under SO can contract due to an increase in travel demand. Furthermore, this example also demonstrates that the set of minimum marginal total cost routes for one OD pair; in this case $O \rightarrow D2$, can change due to an increase in demand on a different OD movement; in this case $O \rightarrow D1$. This latter observation demonstrates the potential complexity of possible dependencies that may exist between expansions and contractions on different OD movements.

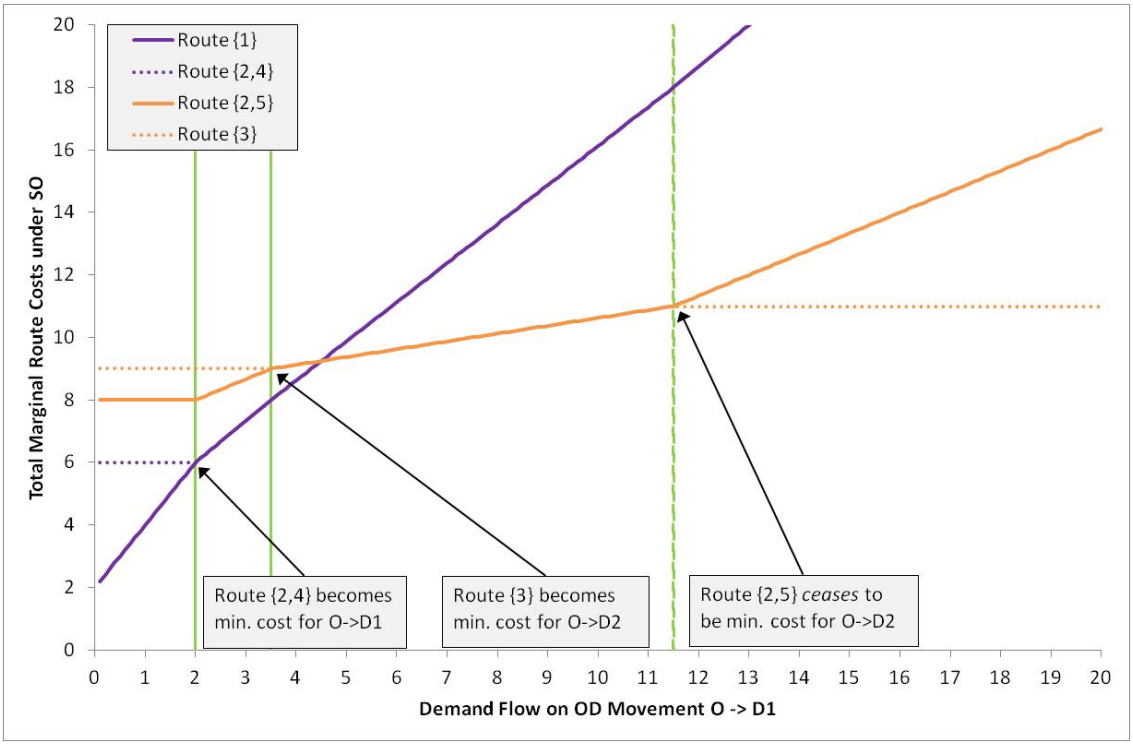

Figure 6.2 - Route Costs under SO against increasing demand on O->D1 for the network in Figure 6.1

It can be shown that exactly the same pattern of expansions and contractions also occurs under UE for this network example; although at different levels of demand $q_{O \rightarrow D1}$.

6.3.2 Definitions, Notation and Limiting Conditions

The examples presented in sections 6.3.1.1 and 6.3.1.2 illustrate that the set of minimum cost routes under UE, and the set of minimum marginal total cost routes under SO, for an OD movement, can expand or contract due to a perturbation in travel demand. The examples also demonstrate that an increase (or decrease) in demand on one OD movement has the potential to cause an expansion or a contraction in the route set of another OD movement. In section 6.4 it is shown that expansions and contractions in these sets, under UE and SO, have a significant influence on how the Price of Anarchy varies with travel demand. As such, the following definitions and notation are proposed in order to characterise these phenomena.

Definition 6.2: The set of *minimum cost routes under UE*, for an OD movement r, at a demand Q is defined as $K_{min}^{r} = \{k \in K^{r} | C_{k}^{r} = \pi_{r}(Q)\}$. To track changes in K_{min}^{r} with respect to perturbations in demand, a vector function $\Upsilon_{r}^{UE}(Q)$ is defined for each OD movement r, which has entries u_k for which $u_k = 1$, if $C_{k}^{r} = \pi_{r}(Q)$, and $u_k = 0$, if $C_{k}^{r} > \pi_{r}(Q)$.

Definition 6.3: A demand vector Q is defined as a *route transition point under UE* if there exist vectors $g, h \in \mathbb{R}^{R} \backslash \{0\}$ for which, for at least one OD movement r:

$$\lim_{\lambda_1 \to 0} \Upsilon_{r}^{\mathrm{UE}}(Q - \lambda_1 g) \neq \lim_{\lambda_2 \to 0} \Upsilon_{r}^{\mathrm{UE}}(Q + \lambda_2 h) \tag{12}$$

where $\lambda_1, \lambda_2 > 0$. Individual route transition points are denoted by η_{UE}, and the set of all such demand vectors for a given network G is denoted H_{UE}.

Route transition points are alternatively referred to as *degenerate* points of the UE problem in the Sensitivity Analysis literature (Josefsson and Patriksson, 2007, Patriksson, 2004)[17]. As shorthand, in the remainder of the paper, the limit on the left-hand side of equation (12) is referred to as $Q \to \eta_{UE}^{-}$ and the limit on the right-hand side of equation (12) is referred to as $Q \to \eta_{UE}^{+}$. Equivalent versions of definitions 6.2 and 6.3 are also defined for SO, with appropriate changes to superscripts and notation. For example, the set of minimum marginal total cost routes under SO, for an OD movement r, is defined as $\widetilde{K}_{min}^{r} = \left\{k \in K^{r} \middle| \tilde{C}_{k}^{r} = \tilde{\pi}_{r}(Q)\right\}$. Both K_{min}^{r} and $\widetilde{K}_{min}^{r}$ are uniquely defined under Assumption A1.

In the case of a network with only one OD movement, the notion of increasing/decreasing demand and the limits in definition 6.3 are very straightforward. However, in the multiple OD case, there are many possible directions of change "through" any particular demand vector in the R-dimensional space of OD demands, which yield a range of possible circumstances. For example, for each route transition point η, there could be several vectors g and h that satisfy the conditions of definition 6.3, and each g, h combination could represent either an increase, decrease or no change in travel demand on each OD movement r. It is also possible that the left-hand or right-hand limits in equation (12) could pass through another nearby route transition point. An example of such a situation is shown in Figure 9 of Josefsson and Patriksson (2007), in which there is curve of degenerate points in the UE problem. Finally, exactly at the route transition point itself, the vector function Υ_{r} could signify that there are expansions in the minimum cost route sets for one OD movement (or several OD movements), contractions in the minimum cost route set for a different OD movement (or several), and no change in the minimum cost route sets for a third OD movement (or several).

---

[17] This literature is covered in more detail in section 6.4.3.2.

A general theory of how route transition points affect the Price of Anarchy would include all of these possible circumstances, but this is a challenging goal. The theory presented in this chapter is therefore restricted to changes that occur at route transition points η when travel demand Q *increases* (for which a formal definition is provide in definition 6.4) and which also satisfy conditions C1-C3 below.

Definition 6.4: Consider two demand vectors $Q^1, Q^2 \in \mathbb{R}^R$ with $Q^i = [q_1^i, \dots, q_R^i]$. Demand is said to have *increased* from Q^1 to Q^2 if and only if $q_j^1 \leq q_j^2 \ \forall j = 1, \dots, R$, and $\exists j'$ for which $q_{j'}^1 < q_{j'}^2$.

Route Transition Point Conditions:

C1. Demand vectors g and h satisfy $g_r \geq 0$ and $h_r \geq 0, \forall r = 1, \dots, R$

C2. For the vectors g, h in C1, $\exists \lambda_1, \lambda_2 > 0$ such that $\forall \theta_1 \in [0, \lambda_1]$, $Q - \theta_1 g \notin \mathrm{H}$ and $\forall \theta_2 \in [0, \lambda_2]$ $Q + \theta_2 h \notin \mathrm{H}$

C3. At each route transition point η, either:

(i) $\forall r = 1, \dots, R$; for each $k \in K^r$, $\lim_{Q \to \eta_{UE}^-} u_k \leq \lim_{Q \to \eta_{UE}^+} u_k$ for entries u_k in Υ_r

(ii) $\forall r = 1, \dots, R$; for each $k \in K^r$, $\lim_{Q \to \eta_{UE}^-} u_k \geq \lim_{Q \to \eta_{UE}^+} u_k$ for entries u_k in Υ_r

Condition C1 is the most restrictive of the three conditions, as it excludes all cases in which a route transition occurs as demand decreases on one or more OD movements. Condition C2 excludes cases in which two route transition points are adjacent to each other. Condition C3 excludes cases at which there is an expansion in the minimum cost route set for at least one OD movement that occurs simultaneously with a contraction in the minimum cost route set for at least one different OD movement.

Before moving on to present theory of how route transition points affect the Price of Anarchy, the next sections describe an alternative characterisation of expansions and contractions, which is useful for numerical investigations, and also a systematic relationship between route transition points under UE and SO that exists for the special family of traffic networks highlighted in the introduction to this chapter.

6.3.3 An Alternative Characterisation of Minimum Cost Route Sets under UE and SO

The results that follow prove that, for each OD movement r, an expansion (contraction) in the set K_{min}^r or $\widetilde{K}_{min}^r$, is equivalent, under the condition of proportionality, to an expansion (contraction) in the set of links, under UE or SO, that have non-zero flow for that OD movement. These sets are referred to as the *Origin Specific Active Network* for an OD movement r and are formally defined in definition 6.5 as follows.

Definition 6.5: The *OD Specific Active Network under UE*, for an OD movement r, at a demand Q is the set $X_r^{UE} = \left\{ i \in A \,\middle|\, \exists k \in K^r, s.t. i \in I_k^r \;\&\; f_k^r > 0 \right\} \subseteq A$, where $f_k^r \in F^*$, the route flow solution that satisfies the condition of proportionality. To track changes in X_r^{UE} with respect to perturbations in demand, a vector function $\Phi_r^{UE}(Q)$ is defined for each OD movement r, which has entries v_i for which $v_i = 1$, if $\exists k \in K^r$ for which $i \in I_k^r$ and $f_k^r > 0$, and $v_i = 0$, if $\forall k \in K^r$ for which $i \in I_k^r$, $f_k^r = 0$.

An equivalent version of definition 6.5 is also defined for SO, with appropriate changes to superscripts and notation. The sets X_r^{UE} and X_r^{SO} are both uniquely defined under the condition of proportionality. Levels of demand at which these sets change are referred to as *link transition points*, which are formally defined in definition 6.6.

Definition 6.6: A demand vector Q is defined as a *link transition point under UE* if there exist vectors $g, h \in \mathbb{R}^R \backslash \{0\}$ for which, for at least one OD movement r:

$$\lim_{\mu_1 \to 0} \Phi_r^{\mathrm{UE}}(Q - \mu_1 g) \neq \lim_{\mu_2 \to 0} \Phi_r^{\mathrm{UE}}(Q + \mu_2 h) \tag{13}$$

where $\mu_1, \mu_2 > 0$. Individual link transition points are denoted by ω_{UE}, and the set of all such demand vectors for a given network G is denoted Ω_{UE}. Again, an equivalent version of definition 6.6 is also defined for SO.

This alternative characterisation of the changing nature with which demand is assigned to a traffic network is useful because it is often significantly easier to identify the set of active links under proportionality, for each OD movement, than it is to identify the set of minimum cost routes. This is because there are often many more routes than there are links, especially in large traffic networks, and the enumeration of routes is a computationally expensive procedure. Accordingly, these results are used in the examples in section 6.5 to track expansions and contractions in K_{min}^r and $\widetilde{K}_{min}^r$.

Proposition 6.1 and corollary 6.2 characterise the relationship between the sets K_{min}^r and X_r^{UE}, and the sets $\widetilde{K}_{min}^r$ and X_r^{SO} for an OD movement r.

Proposition 6.1: Consider a traffic network G for which Assumption A1 holds, and let $F^* = \{f_k^r\}$ represent the route flow solution under UE that satisfies the condition of proportionality. Suppose that Q represents a demand vector that is not a route transition point, i.e. $Q \notin \mathrm{H}_{UE}$. For a given OD movement r, further suppose that Q does not correspond to a level of demand at which X_r^{UE} changes. Then, for that OD movement r:

(i) A link $i \in X_r^{UE}$ if and only if $\exists k' \in K_{min}^r$ for which $i \in I_{k'}^r$.

(ii) A route $k' \in K_{min}^r$ if and only if $I_{k'}^r \subset X_r^{UE}$.

Part (i) describes how, for an OD movement r, the OD Specific Active Network under UE can be constructed from the set of minimum cost routes for the OD movement r. Part (ii) describes how, for an OD movement r, the set of minimum cost routes under UE can be constructed from the OD Specific Active Network for the OD movement r.

Proof: For parts (i) and (ii), the only if and if statements are addressed in turn.

(i) Only if statement: For a given link $i \in A$ suppose that $i \in X_r^{UE}$ for an OD movement r. Then by equation (6) $\exists k' \in K^r$ for which $f_{k'}^r > 0$. For this k', the UE conditions (5) imply that $C_{k'}^r = \pi_r$ and that therefore $k' \in K_{min}^r$.

(i) If statement: For a given link $i \in A$, suppose that $\exists k' \in K_{min}^r$ for which $i \in I_{k'}^r$. By definition 6.2, for this route k', it follows that $C_{k'}^r = \pi_r$. Under the condition of proportionality, a route flow solution F^* can be constructed for which $f_k^r > 0 \; \forall k$ for which $C_k^r = \pi_r$. By equation (6) this route flow solution provides that link i has positive flow for the OD movement r. It therefore follows that $i \in X_r^{UE}$.

(ii) Only if statement: For a given route $k' \in K^r$ suppose that $k' \in K_{min}^r$. Then, by definition 6.2, $C_{k'}^r = \pi_r$. Under the condition of proportionality, a route flow solution F can be constructed for which $f_{k'}^r > 0$. As $f_{k'}^r > 0$ and, by equation (6), all links $i \in I_{k'}^r$ contain the flow $f_{k'}^r$ as part of their summation, it follows that each such link i has positive flow for the OD movement r. In other words, $i \in X_r^{UE}, \forall i \in I_{k'}^r$, and therefore $I_{k'}^r \subset X_r^{UE}$.

(ii) If statement: For a given route $k' \in K^r$, suppose that $I_{k'}^r \subset X_r^{UE}$. Suppose, for a contradiction, that $k' \notin K_{min}^r$. Then by equation (5), $C_{k'}^r > \pi_r$. By starting assumption, all links $i \in I_{k'}^r$ carry flow for this OD, i.e. $i \in X_r^{UE}$. Hence each such link must lie on at least one route $k^* \in K^r \backslash \{k'\}$ for which $f_{k^*}^r > 0$ and hence, from equation (5), $C_{k^*}^r = \pi_r$. Therefore, each $k^* \in K_{min}^r$ and it follows, from the only if statement of part(ii), which has just been proven, that $I_{k*}^r \subset X_r^{UE}$.

Therefore both $I_{k'}^r \subset X_r^{UE}$ and $I_{k*}^r \subset X_r^{UE}$. Consider the pair(s) of alternative segments defined by the set of links $(I_{k'}^r \cup I_{k*}^r) \backslash (I_{k'}^r \cap I_{k*}^r) \subset X_r^{UE}$ i.e. both alternative segments (in each pair) are used. Under the condition of proportionality, it follows from Bar-Gera (2006) that *"for every used pair of alternative segments and every used route that contains one of the segments, there will be a similar used route containing the alternative segment"* (Bar-Gera et al., 2012). In this statement, the *"similar used route"* refers to a route that only differs from the *"used route"* in the pair of alternative segments; i.e. the *"used route"* and the *"similar used route"* overlap each other in the rest of their composition. This proportionality implies that $f_{k'}^r > 0$, which implies that $C_{k'}^r = \pi_r$. This contradicts the assumption that $k' \notin K_{min}^r$. ■

The equivalent statement of proposition 6.1 for SO is stated as follows.

Corollary 6.2: Consider a traffic network G for which Assumption A1 holds, and let $\tilde{F}^* = \{f_k^r\}$ represent the route flow solution under SO that satisfies the condition of proportionality. Suppose that Q represents a demand vector that is not a route transition point, i.e. $Q \notin \mathrm{H}_{SO}$. For a given OD movement r, further suppose that Q does not correspond to a level of demand at which X_r^{SO} changes. Then, for that OD movement r:

(i) A link $i \in X_r^{SO}$ if and only if $\exists k' \in \widetilde{K}_{min}^r$ for which $i \in I_{k'}^r$.

(ii) A route $k' \in \widetilde{K}_{min}^r$ if and only if $I_{k'}^r \subset X_r^{SO}$.

Proof: Traces that of proposition 6.1 with appropriate changes in notation from UE to SO. ■

The following results prove that the sets K_{min}^r ($\widetilde{K}_{min}^r$) and X_r^{UE} (X_r^{SO}) expand and contract at identical levels of demand.

Proposition 6.3: Consider a traffic network G for which Assumption A1 holds, and let $F^* = \{f_k^r\}$ represent the route flow solution under UE that satisfies the condition of proportionality. There is a one-to-one correspondence between route transition points η_{UE} and link transition points ω_{UE}.

Proof: This statement is proved by contradiction.

There are four cases to consider: i) $\exists \eta_{UE}$ corresponding to an expansion in K_{min}^r for which $\nexists \omega_{UE}$ corresponding to an expansion in X_r^{UE}, ii) $\exists \omega_{UE}$ corresponding to an expansion in X_r^{UE} for which $\nexists \eta_{UE}$ corresponding to an expansion in K_{min}^r, iii) $\exists \eta_{UE}$ corresponding to a contraction in K_{min}^r for which $\nexists \omega_{UE}$ corresponding to a contraction in X_r^{UE}, iv) $\exists \omega_{UE}$ corresponding to a contraction in X_r^{UE} for which $\nexists \eta_{UE}$ corresponding to an contraction in K_{min}^r. Proofs are provided for cases i) and ii); the proofs of iii) and iv) are similar.

Case i) Suppose, for a contradiction, that there exists an instance of demand η_{UE}, at which K_{min}^r expands for some OD movement r, but for which there does not exist a corresponding point ω_{UE}, at which X_r^{UE} expands for the same OD movement. Therefore, there is a perturbation of demand for which $\exists k \in K^r$, such that as $Q \to \eta_{UE}^-$, $k \notin K_{min}^r$, but that as $Q \to \eta_{UE}^+$, $k \in K_{min}^r$. It follows, from proposition 6.1(ii), that as $k \notin K_{min}^r$ as $Q \to \eta_{UE}^-$, $\exists i \in I_k^r$ for which $i \notin X_r^{UE}$. It also follows, from proposition 6.1(ii), that as $k \in K_{min}^r$ as $Q \to \eta_{UE}^+$, $i \in X_r^{UE}$, $\forall i \in I_k^r$. Hence $\exists i \in A$ that is added to X_r^{UE} at η_{UE}. This contradicts the starting assumption.

Case ii) Now suppose, for a contradiction, that there exists an instance of demand ω_{UE}, at which X_r^{UE} expands for some OD movement r, for which there does not exist a corresponding point η_{UE}, at which K_{min}^r expands for the same OD movement. Therefore, there is a perturbation of demand for which $\exists i \in A$, such that as $Q \to \omega_{UE}^-$, $i \notin X_r^{UE}$, but that as

$Q \rightarrow \omega_{UE}^{+}$, $i \in X_r^{UE}$. It follows, from proposition 6.1(i), that as $i \notin X_r^{UE}$ as $Q \rightarrow \omega_{UE}^{-}$, $k \notin K_{min}^r$, $\forall k \in K^r$ for which $i \in I_k^r$. It also follows, from proposition 6.1(i), that as $i \in X_r^{UE}$ as $Q \rightarrow \omega_{UE}^{+}$, $\exists k \in K_{min}^r$ for which $i \in I_k^r$. Hence $\exists k \in K^r$ that is added to K_{min}^r at ω_{UE}. This contradicts the starting assumption. ■

Corollary 6.4: Consider a traffic network G for which Assumption A1 holds, and let $\tilde{F}^* = \{f_k^r\}$ represent the route flow solution under SO that satisfies the condition of proportionality. There is a one-to-one correspondence between route transition points η_{SO} and link transition points ω_{SO}.

Proof: Traces that of proposition 6.3 with appropriate changes in notation from UE to SO. ■

6.3.4 A Systematic Relationship between UE and SO Link Flows and Route Transition Points

This section establishes two results for the special case[18] of traffic networks with link cost functions of the form $c_i = a_i + b_i x_i^{\beta}$ for which the coefficients $a_i, b_i > 0 \; \forall i$ and $\beta > 0$ is common to all links. This set of cost functions includes, but is not limited to, the well-known BPR cost function, which was used in the network model and numerical experiments described in chapters 4 and 5.

In this narrower context, it is proven, in theorem 6.5, that there is a systematic relationship between link flows under UE and SO. As a consequence of this, it is proven, in corollary 6.6, that there is also a systematic relationship between the levels of demand under UE and SO at which expansions and contractions occur in the minimum cost route sets K_{min}^r and $\widetilde{K}_{min}^r$. This systematic relationship can be observed in the parallel link example of section 6.3.1.1, for which the cost functions c_i belong to the cost function set considered here. A comparison of equations (9) and (11) reveals that the level of demand at which each route $k \in K^r$ is added to K_{min}^r is exactly half the level of demand at which the same route k is added to $\widetilde{K}_{min}^r$. Corollary 6.6 proves that this is indicative of a result that applies more generally to networks with multiple OD pairs.

Theorem 6.5: Consider a traffic network G that serves a demand matrix Q with entries $q_r > 0$, and that has cost functions of the form $c_i = a_i + b_i x_i^{\beta}$ $(a_i, b_i, \beta > 0)$, which satisfy Assumption A1. Let $x_i^{UE}(Q)$ and $x_i^{SO}(Q)$ denote UE and SO link flows respectively, which are defined as functions of the demand vector Q. Then, under these conditions, $\forall i \in A$:

---

[18] Elsewhere in the chapter link cost functions are assumed separable and monotonic.

$$x_i^{SO}\left(\frac{Q}{\sqrt[\beta]{\beta+1}}\right)=\frac{1}{\sqrt[\beta]{\beta+1}}x_i^{UE}(Q) \tag{14}$$

Proof: Consider a traffic network G, with demand matrix Q and link cost functions $c_i = a_i + b_i x_i^\beta$ has a UE link flow solution $x_i^{UE}(Q)$. The proof begins by noting that a different traffic assignment problem can be defined on G, with demand matrix $\hat{Q} = \lambda Q$ and link cost functions $\hat{c}_i = a_i + b_i(x_i/\lambda)^\beta$, which has a UE link flow solution $\hat{x}_i^{UE}(\hat{Q}) = \lambda x_i^{UE}(Q)$. In other words, the traffic assignment problem has been rescaled by λ.

Now consider the problem of finding an SO link flow solution $x_i^{SO}\left(Q/\sqrt[\beta]{\beta+1}\right)$ for a given road network G serving a demand matrix $Q/\sqrt[\beta]{\beta+1}$ with link costs $c_i = a_i + b_i x_i^\beta$ as defined in the left hand side of equation (14).

As noted in section 6.2, this problem is equivalent to finding a UE link flow solution $x_i^{UE}\left(Q/\sqrt[\beta]{\beta+1}\right)$ on G for a demand matrix $Q/\sqrt[\beta]{\beta+1}$ with transformed cost functions (Sheffi, 1985, p73):

$$\tilde{c}_i = c_i + \frac{dc_i}{dx_i}x_i = \left(a_i + b_i x_i^\beta\right) + \left(b_i\beta x_i^{\beta-1}\right)x_i = a_i + b_i(\beta+1)x_i^\beta$$
$$= a_i + b_i\left(\sqrt[\beta]{\beta+1}x_i\right)^\beta$$

Setting $\lambda = 1/\sqrt[\beta]{\beta+1}$ to simplify notation, this problem can be restated as one of finding a UE link flow solution $x_i^{UE}(\lambda Q)$ on G for a demand matrix λQ with cost functions $c_i = a_i + b_i(x_i/\lambda)^\beta$.

Applying the earlier scaling note, the UE link flow solution $x_i^{UE}(\lambda Q)$ in this restated problem is equivalent to a rescaled UE problem on G, which has link flow solution $\lambda\hat{x}_i^{UE}(Q)$ with demand matrix Q and link cost functions $c_i = a_i + b_i x_i^\beta$. However, this scaled problem is exactly the problem on the right hand side of equation (14), and it therefore follows that:

$$x_i^{SO}\left(Q/\sqrt[\beta]{\beta+1}\right)=\frac{1}{\sqrt[\beta]{\beta+1}}x_i^{UE}(Q) \quad \blacksquare$$

The following corollary describes the relationship between route transition points under UE and SO.

Corollary 6.6: Consider a traffic network G that serves a demand matrix Q with entries $q_r > 0$, and that has cost functions of the form $c_i = a_i + b_i x_i^\beta$ $(a_i, b_i, \beta > 0)$, which satisfy Assumption A1. Suppose that the condition of proportionality holds and that $f_k^{r,UE} \in F^*$ and $f_k^{r,SO} \in \tilde{F}^*$ represent the uniquely defined route flow for each route $k \in K^r$, under UE and SO

respectively. Let η_{UE} and η_{SO} represent instances of demand Q at which, for some OD movements r, the same routes $k \in K^r$ are added to or removed from the sets K_{min}^r and $\widetilde{K}_{min}^r$, respectively[19]. Then:

$$\eta_{SO} = \frac{1}{\sqrt[\beta]{\beta+1}}\eta_{UE} \tag{15}$$

Proof: Consider a given level of demand η_{UE} at which the set K_{min}^r expands for some OD movement r. Therefore, $\exists k \in K^r$, for some OD movement r, for which $k \notin K_{min}^r$ as $Q \to \eta_{UE}^-$, but for which $k \in K_{min}^r$ as $Q \to \eta_{UE}^+$. As route flows, under the condition of proportionality, are uniquely defined and vary continuously with respect to Q, it follows, from theorem 6.5, that $f_k^{r,UE}(Q) = \sqrt[\beta]{\beta+1} * f_k^{r,SO}\left(Q/\sqrt[\beta]{\beta+1}\right)$.

Now, if $k \notin K_{min}^r$ as $Q \to \eta_{UE}^-$, such that $f_k^{r,UE} = 0$ under the condition of proportionality, then it follows that $f_k^{r,SO} = 0$ and that $k \notin \widetilde{K}_{min}^r$ as $Q \to \left(1/\sqrt[\beta]{\beta+1}\right)\eta_{UE}^-$. In addition, if $\in K_{min}^r$ as $Q \to \eta_{UE}^+$, such that $f_k^{r,UE} > 0$ under the condition of proportionality, then it follows that $f_k^{r,SO} > 0$ and that $k \in \widetilde{K}_{min}^r$ as $Q \to \left(1/\sqrt[\beta]{\beta+1}\right)\eta_{UE}^+$. This implies that $\widetilde{K}_{min}^r$ expands under SO at $Q = \left(1/\sqrt[\beta]{\beta+1}\right)\eta_{UE}$, i.e. that $\exists \eta_{SO} = \left(1/\sqrt[\beta]{\beta+1}\right)\eta_{UE}$.

A similar argument works for when η_{UE} corresponds to a contraction of K_{min}^r, for some OD movement r. ■

It is important to note that corollary 6.6 does not predict the levels of demand at which the sets K_{min}^r or $\widetilde{K}_{min}^r$ will change; rather, it provides a method to identify the levels of demand at which, for example, K_{min}^r changes, given the levels of demand at which $\widetilde{K}_{min}^r$ changes.

6.4 The Variation of the Price of Anarchy with Travel Demand

This section presents theory that describes how the Price of Anarchy varies with travel demand. In order to provide motivation and context for this theory, this section begins, in section 6.4.1, by illustrating how the Price of Anarchy varies with travel demand in the network examples presented in section 6.3.1. Sections 6.4.2, 6.4.3 and 6.4.4 then present theory to describe the mechanisms that govern how the Price of Anarchy varies in general networks for low, intermediate and high levels of travel demand respectively.

---

[19] For such traffic networks, the existence of instances of demand, under UE and SO, at which the same routes $k \in K^r$ are added to or removed from K_{min}^r and $\widetilde{K}_{min}^r$ follows from the relationship described in theorem 6.5. In general traffic networks, this statement can be shown to follow from the SO cost transformation $\tilde{c}_i$.

For intermediate levels of demand, it was established in section 6.3 that, as travel demand Q changes, the sets K_{min}^{r} (X_r^{UE}) and $\widetilde{K}_{min}^{r}$ (X_r^{SO}) can expand and contract, for one or more OD movements r. The points at which these expansions and contractions occur were defined as route transition points and several types were identified. The theory presented in this section applies to route transition points that occur under increasing demand and which satisfy conditions C1-C3. These conditions were described in section 6.3.2.

The behaviour of the Price of Anarchy is dependent, by construction, on Total Network Travel Cost under SO (TTC^{SO}) and Total Network Travel Cost under UE (TTC^{UE}). This is important for the analysis that follows.

6.4.1 Illustrative Examples

6.4.1.1 Example 1: Parallel Link Network - Single Origin-Destination Pair Example

Recalling the example of section 6.3.1.1; consider increasing demand q in nine versions of a parallel link network with total links $N = 2,3,\dots,10$ and coefficients $a_i = i,\ b_i = 1$ for $i = 1,\dots,10$. Figure 6.3 displays the variation of the Price of Anarchy ρ_N for each of these nine networks, and also identifies the levels of demand under UE (green vertical lines) and SO (red vertical lines) at which expansions occur in the sets K_{min} and $\widetilde{K}_{min}$ respectively. These levels of demand correspond to those identified in equations (9) and (11) respectively.

For levels of demand q up to the first route transition point under SO, the Price of Anarchy is 1. Beyond this level of demand, Figure 6.3 illustrates, for each N, that levels of demand at which K_{min} expands coincide with all levels of demand at which the Price of Anarchy is non-differentiable. Furthermore, there is also a decrease in the gradient of the Price of Anarchy at each of these points. In contrast the Price of Anarchy appears to be differentiable at all levels of demand at which there is an expansion in $\widetilde{K}_{min}$. However, it is also evident, for each $M = 2,\dots,10$, that the graphs of ρ_{M-1} and ρ_M depart from each other at each of these points. This demonstrates that the new routes that are available in the M parallel link case have a material effect on the trajectory of the Price of Anarchy. Overall, Figure 6.3 suggests that expansions under UE lead to decreases in the Price of Anarchy whereas expansions under SO lead to increases in the Price of Anarchy.

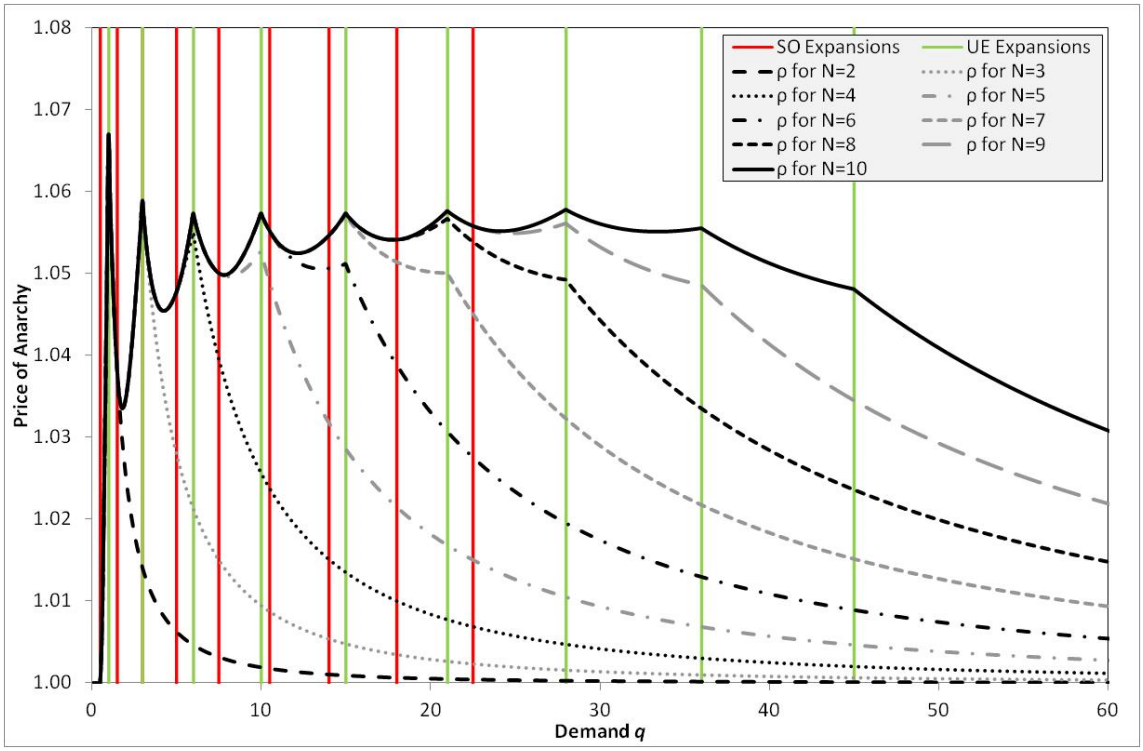

Figure 6.3 - The Variation of the Price of Anarchy against Demand in $N = 2, \dots, 10$ Parallel Link Network

As demand increases the Price of Anarchy eventually begins to decay back towards 1. The start of this decay coincides with the last route transition point under UE. An explicit formula for the Price of Anarchy in this region, for each network N, is shown in equation (16). This formula was derived analytically. The parameters α and γ are constants that depend on the coefficients a_i and b_i. This equation reveals that the leading order term of this decay is $O(1/q^2)$, which suggests that the similar characteristic shapes of decay, observed for the networks in Figure 5.19, are a systematic and more general feature of the behaviour of the Price of Anarchy for high demand.

$$\rho = 1 + \frac{1}{\alpha q^2 + \gamma q - 1} \tag{16}$$

6.4.1.2 Example 2: Five Link Network - Two Origin-Destination Pair Example

Now recall the five link network example of section 6.3.1.2, and consider increasing demand on OD movement $q_{O\to D1}$. The variation of the Price of Anarchy with demand $q_{O\to D1}$ is shown in Figure 6.4. The vertical lines signify levels of demand under UE (green) and SO (red) at which K_{min} and $\widetilde{K}_{min}$ expand (solid lines) and contract (dashed lines).

As was observed in Figure 6.3, this figures shows that the Price of Anarchy is one for all levels of demand $q_{O\to D1}$ up to the first route transition point under SO. Figure 6.4 also illustrates that at expansions in K^r_{min}, the Price of Anarchy is non-differentiable and that there is a decrease in

gradient; this is the same as the behaviour in Figure 6.3. Figure 6.4 also illustrates that the Price of Anarchy is non-differentiable at the single demand level corresponding to a contraction in K^r_{min}, and that this coincides with an increase in gradient. Under SO, the Price of Anarchy is differentiable at both points of expansion and also the point of contraction in $\widetilde{K}^r_{min}$; the former leads to an increase in the gradient of the Price of Anarchy whereas the latter leads to a decrease in the gradient of the Price of Anarchy. This example suggests therefore, that the effects of contractions in K^r_{min} and $\widetilde{K}^r_{min}$, on the Price of Anarchy, are the opposite of the effects of expansions in K^r_{min} and $\widetilde{K}^r_{min}$.

Finally, for demand beyond the final route transition point under UE, the Price of Anarchy again decays back towards 1. Although not included here, this rate of decay also satisfies $O(1/q^2)$ behaviour.

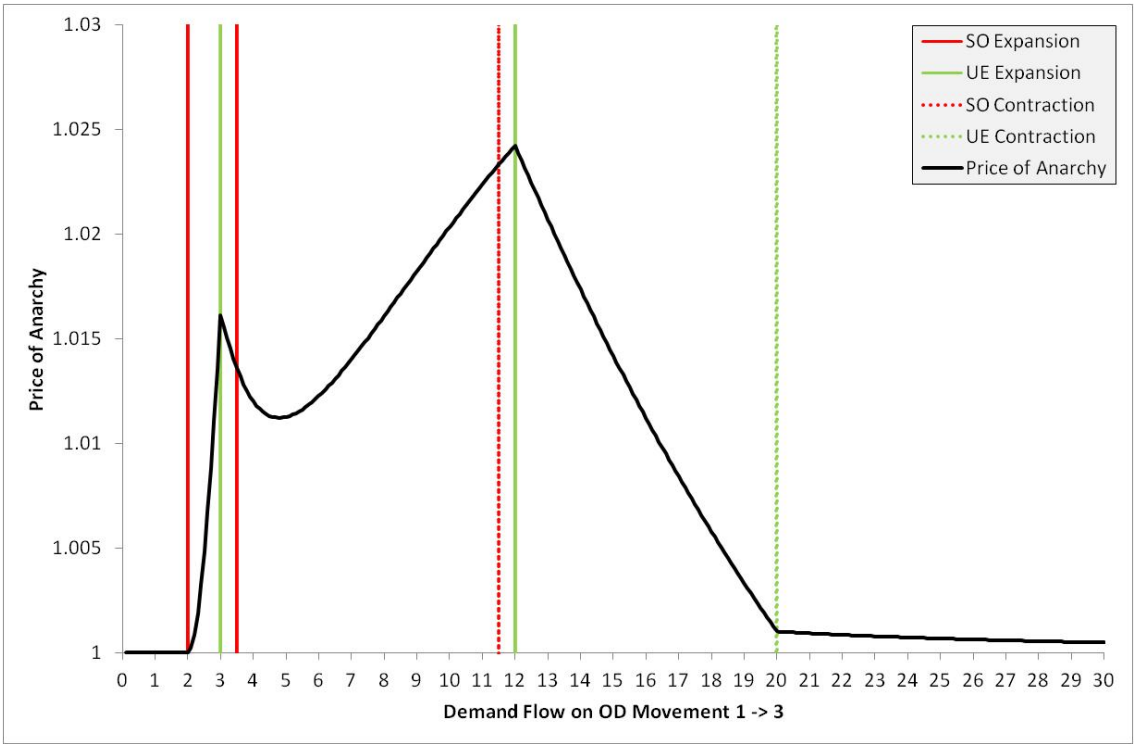

Figure 6.4 - The Variation of the Price of Anarchy against Demand in the Five Link Network of Figure 6.1

6.4.2 The Variation of the Price of Anarchy for Low Travel Demand

In traffic networks in which demand $q_r \to 0$ on all OD movements r, the cost of travel on each route $k \in K^r$, $\forall r$, is dictated by the free-flow travel cost component. In such cases, for such small levels of demand, the routes that are of minimum cost for each OD movement correspond to the shortest path or paths for each OD movement. This is true under both UE and SO; to see this, consider the cost function transformation $\tilde{c}_i = c_i + x_i * dc_i/dx_i$ for the SO problem. When $x_i \to 0$, the additional marginal cost term disappears and the cost of travel on each link is identical under UE and SO. In such cases, it follows that $x_i^{UE} = x_i^{SO}$ $\forall i \in A$, that

$TTC^{UE} = TTC^{SO}$ and that the Price of Anarchy $\rho = 1$. As demand q_r increases from zero, the shortest path(s) for each OD movement r still provide the minimum (marginal total) cost routes under UE and SO, provided that the second shortest routes have greater free-flow travel cost for each OD movement. The Price of Anarchy remains $\rho = 1$ until, for some OD movement r, the second shortest route in K^r becomes minimum cost, at which point there is a route transition point under SO. This discussion provides an explanation for the initial intervals of demand shown in Figure 5.19, Figure 6.3 and Figure 6.4.

6.4.3 The Variation of the Price of Anarchy for Intermediate Regions of Travel Demand

It is known that TTC^{UE} and TTC^{SO} are continuous and increasing functions of travel demand (Dafermos and Nagurney, 1984). As $x_i^{UE} = x_i^{SO}\ \forall i \in A$, for very low demand regions, it follows that TTC^{UE} and TTC^{SO} both increase at the same rates, at least until a route transition point occurs. This section describes the effects of route transition points, of the types described in conditions C1-C3, on the rates of change of TTC^{SO} (section 6.4.3.1), TTC^{UE} (section 6.4.3.2) and the Price of Anarchy (section 6.4.3.3).

6.4.3.1 The Sensitivity of Total Network Travel Cost under SO to Route Transition Points

The first result in this section proves that TTC^{SO} is also differentiable with respect to all demand Q, which, in particular, includes all demands $Q \in \mathrm{H}_{SO}$ that correspond to route transition points.

Proposition 6.7: Consider a traffic network G for which Assumption A1 holds. TTC^{SO} is differentiable with respect to all demand movements r for which $q_r > 0$.

Proof: Proof follows from the Envelope Theorem, which is stated as follows. For the constrained extremum problem:

$$V(z) = \max_{x_1, x_2, \ldots, x_n} f(x_1, x_2, \ldots, x_n, z)$$

$$\text{s.t.} g_j(x_1, x_2, \ldots, x_n, z) \geq 0 \text{ for } j = 1,2, \ldots, m$$

the Envelope Theorem states that, if the constraints satisfy the Slater condition and if $x_i(z)$ solve the first-order and complementary slackness conditions for the above problem, $\forall i$, then:

$$\frac{\partial V}{\partial z} = \frac{\partial f(x_1, x_2, \ldots, x_n, z)}{\partial z} + \sum_{j=1}^{m} \lambda_j \frac{\partial g_j}{\partial z}$$

where λ_j are Kuhn-Tucker multipliers. The Slater condition requires that there exists a point $(x_1, x_2, \ldots, x_n)$ for which $g_j(x_1, x_2, \ldots, x_n) > 0\ \forall j$.

The SO minimisation problem has objective function $\tilde{z}(x_1, x_2, \dots, x_n) = TTC^{SO}$ and is subject to constraints set out in equation (6). For this problem to satisfy the Slater condition, requires that there exists a link flow vector $(x_1, x_2, \dots, x_n)$ satisfying the equality constraints in (6) and which produces route flows $f_k^r > 0$, $\forall k \in K^r$, for all OD movements r. In other words a vector of link flows is required that satisfies the equality constraints and which produces positive route flows on all routes between all OD pairs. This can easily be achieved by setting $f_k^r = \gamma_k^r q_r$ where $0 < \gamma_k^r < 1$ and such that $\sum_k \gamma_k^r = 1 \; \forall r$, i.e. a link flow vector produced by assigning demand flows q_r to all routes $k \in K^r$ such that all routes receive a non-zero proportion of flow.

The SO minimisation problem therefore satisfies all of the conditions of the Envelope Theorem, which guarantees that the objective function $z(x_1, x_2, \dots, x_n, q_r)$ is differentiable with respect to demand q_r. As $\tilde{z}(x_1, x_2, \dots, x_n, q_r) = TTC^{SO}$, this guarantees that TTC^{SO} is a differentiable function of $q_r > 0$. ■

The next result considers the effect on TTC^{SO} of an increase in demand through a route transition point η_{SO} of the type described in condition C3(i); at which, for each OD movement r, either:

a) $\widetilde{K}_{min}^r$ remains unchanged as demand passes through η_{SO}; or

b) $\widetilde{K}_{min}^r$ expands to include one or more additional routes k.

Proposition 6.8: Consider a traffic network G with link path incidence matrix Δ and for which Assumption A1 holds. Let η_{SO} represent a route transition point satisfying conditions C1, C2 and C3(i); as described in a) and b) above. Denote the OD movements r that satisfy b) by r'. Label routes $k \in K^{r\prime}$ such that: for routes $k = 1, \dots, n_{r\prime}$, $k \in \widetilde{K}_{min}^{r\prime}$ for all demand values $Q \to \eta_{SO}^-$ and $Q \to \eta_{SO}^+$; and for routes $k = n_{r\prime} + 1, \dots, \kappa_{r\prime}$, $k \notin \widetilde{K}_{min}^{r\prime}$ for demand $Q \to \eta_{SO}^-$, but $k \in \widetilde{K}_{min}^{r\prime}$ for demand $Q \to \eta_{SO}^+$.

Suppose that $\hat{G}$ denotes an adjusted version of the network G, which has an identical link path incidence matrix $\hat{\Delta}$, except that for the OD movements r', all routes $k = n_{r\prime} + 1, \dots, \kappa_{r\prime}$ are omitted from $\hat{\Delta}$. Then for demand $Q \to \eta_{SO}^+$:

$$TTC_{\hat{G}}^{SO}(Q) > TTC_G^{SO}(Q)$$

Proof: Let $\Psi_G^*(Q) = \left\{\tilde{C}_k^r(Q) \middle| k \in \widetilde{K}_{min}^r, \; \forall r\right\}$ represent the unique set of route costs with minimum marginal total cost under SO, at demand Q. The route costs $\Psi_G^*(Q)$ are therefore associated with the vector of link flows $x_G^*(Q)$, which produce the minimum value of the objective function $\tilde{z}_G(Q)$ in the SO minimisation program defined in Figure 4.2. Note that $\tilde{z}_G(Q) = TTC_G^{SO}(Q)$.

By the starting assumptions; for demand levels $Q \to \eta_{SO}^{-}$, $\Psi_G^*(Q)$ uniquely minimises $\tilde{z}_G(Q)$, such that all routes $k = n_{r'} + 1, \dots, \kappa_{r'}$ for the OD movements r', satisfy $k \notin \widetilde{K}_{min}^{r'}$. Whereas, for demand levels $Q \to \eta_{SO}^{+}$, $\Psi_G^*(Q)$ uniquely minimises $\tilde{z}_G(Q)$, such that all routes $k = n_{r'} + 1, \dots, \kappa_{r'}$ for the OD movements r', satisfy $k \in \widetilde{K}_{min}^{r'}$. All other feasible route cost sets $\Psi_G(Q)$, satisfy $\tilde{z}\big(\Psi_G(Q)\big) > \tilde{z}\big(\Psi_G^*(Q)\big)$. In particular, all route cost sets $\Psi_G(Q)$, for demand levels $Q \to \eta_{SO}^{+}$, in which routes $k = n_{r'} + 1, \dots, \kappa_{r'}$ for the OD movements r', are restricted from $\widetilde{K}_{min}^{r'}$,, satisfy this condition.

Now consider the network $\hat{G}$. For demand levels $Q \to \eta_{SO}^{-}$, $\Psi_G^*(Q) = \Psi_{\hat{G}}^*(Q)$ and therefore $TTC_G^{SO}(Q) = TTC_{\hat{G}}^{SO}(Q)$. However, for demand levels $Q \to \eta_{SO}^{+}$, $\Psi_G^*(Q) \neq \Psi_{\hat{G}}^*(Q)$. This is because, in $\Psi_G^*(Q)$, the routes $k = n_{r'} + 1, \dots, \kappa_{r'}$ for the OD movements r', satisfy $k \in \widetilde{K}_{min}^{r'}$. Whereas, in $\Psi_{\hat{G}}^*(Q)$, the same routes are not in $\widetilde{K}_{min}^{r'}$ because they were omitted from the link path incidence matrix $\hat{\Delta}$ for $\hat{G}$ by starting assumption. However, $\Psi_{\hat{G}}^*(Q)$ is still feasible for the network G. It therefore follows that $\tilde{z}\left(\Psi_{\hat{G}}^*(Q)\right) > \tilde{z}\big(\Psi_G^*(Q)\big)$ for demand levels $Q \to \eta_{SO}^{+}$. This equation is equivalent to $TTC_{\hat{G}}^{SO}(Q) > TTC_G^{SO}(Q)$. ■

A visualisation of this result is provided in Figure 6.5. In this figure, it is assumed that at η_{SO}, the sets $\widetilde{K}_{min}^{r}$, for one or more OD movements r in a network G, expand such that the total number of minimum marginal total cost routes over all OD movements increases from N to $> N$. Under the terms of the assumptions of proposition 6.8, the traffic network $\hat{G}$ does not contain any of these additional routes. It can be seen, in Figure 6.5, that as $Q \to \eta_{SO}^{-}$, $TTC_G^{SO}(Q) = TTC_{\hat{G}}^{SO}(Q)$. However, at $Q = \eta_{SO}$, these functions diverge. $TTC_{\hat{G}}^{SO}$ represents what would have happened to TTC_G^{SO} if the routes that were added to the minimum cost route sets for the OD movements r, did not exist in G. As demand $Q \to \eta_{SO}^{+}$, TTC^{SO} does not continue to follow the trajectory that it was on for $Q \to \eta_{SO}^{-}$, for which there were N minimum cost routes in total; but instead shifts onto a *lower* trajectory for which there are $> N$ minimum cost routes in total, thereby slowing the rate of increase in TTC^{SO}.

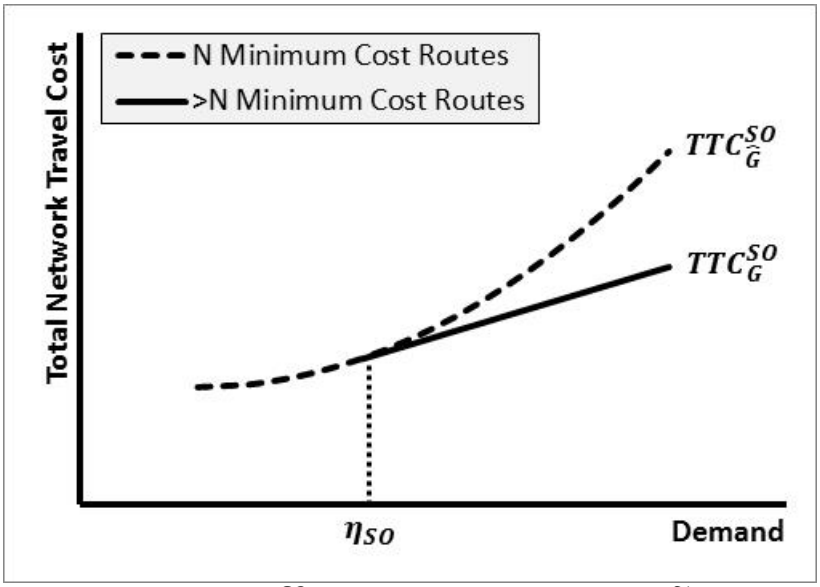

Figure 6.5 - The effect on TTC^{SO} of one or more expansions in $\widetilde{K}_{min}^{r}$, for some OD movements r

The final result of this subsection describes the effect on TTC^{SO} of an increase in demand through a route transition point η_{SO} of the type described in condition C3(ii); at which, for each OD movement r, either:

c) $\widetilde{K}_{min}^{r}$ remains unchanged as demand passes through η_{SO}; or

d) $\widetilde{K}_{min}^{r}$ contracts as one or more routes k are no longer of minimum marginal total cost.

Proposition 6.9: Consider a traffic network G with link path incidence matrix Δ and for which Assumption A1 holds. Let η_{SO} represent a route transition point satisfying the conditions C1, C2 and C3(ii); as described in c) and d) above. Denote the OD movements r that satisfy condition d) by r'. Label routes $k \in K^{r'}$ such that: for routes $k = 1, \dots, n_{r'}$, $k \in \widetilde{K}_{min}^{r'}$ for all demand values $Q \to \eta_{SO}^{-}$ and $Q \to \eta_{SO}^{+}$; and for routes $k = n_{r'} + 1, \dots, \kappa_{r'}$, $k \in \widetilde{K}_{min}^{r'}$ for demand $Q \to \eta_{SO}^{-}$, but $k \notin \widetilde{K}_{min}^{r'}$ for demand $Q \to \eta_{SO}^{+}$.

Suppose that $\hat{G}$ denotes an adjusted version of the network G, which has an identical link path incidence matrix $\hat{\Delta}$, except that for the OD movements r', all routes $k = n_{r'} + 1, \dots, \kappa_{r'}$ are omitted from $\hat{\Delta}$. Then for demand levels $Q \to \eta_{SO}^{-}$:

$$TTC_{\tilde{G}}^{SO}(Q) > TTC_{G}^{SO}(Q)$$

Proof: This proof uses similar arguments to those used to prove Proposition 6.8. ■

A visualisation of this result is provided in Figure 6.6. In this figure, it is assumed that at η_{SO}, the sets $\widetilde{K}_{min}^{r}$, for one or more OD movements r in a network G, contract such that the total number of minimum marginal total cost routes over all OD movements decreases from $> N$ to N. Under the terms of the assumptions of proposition 6.9, the traffic network $\hat{G}$ does not contain any of the routes that leave the minimum marginal total cost route set at η_{SO}. To

understand the implication of this graph, it is easiest to visualise what happens as demand decreases from the right hand side. It can be seen that for demand values $Q \to \eta_{SO}^{+}$, $TTC_{G}^{SO}(Q) = TTC_{\hat{G}}^{SO}(Q)$. At $Q = \eta_{SO}$, these functions diverge. For demand values $Q \to \eta_{SO}^{-}$, TTC^{SO} does not continue to follow the trajectory that it was on for $Q \to \eta_{SO}^{+}$, because, in the direction of decreasing demand, the set of routes of minimum marginal total cost expands from N to $> N$ routes in total, which leads to a lower value of TTC^{SO}. The effect of this behaviour, when considering increasing demand, is that as demand moves through a route transition point of type C3(ii), TTC^{SO} transfers onto a *higher* trajectory. There is, therefore, an acceleration in the rate of increase of TTC^{SO}.

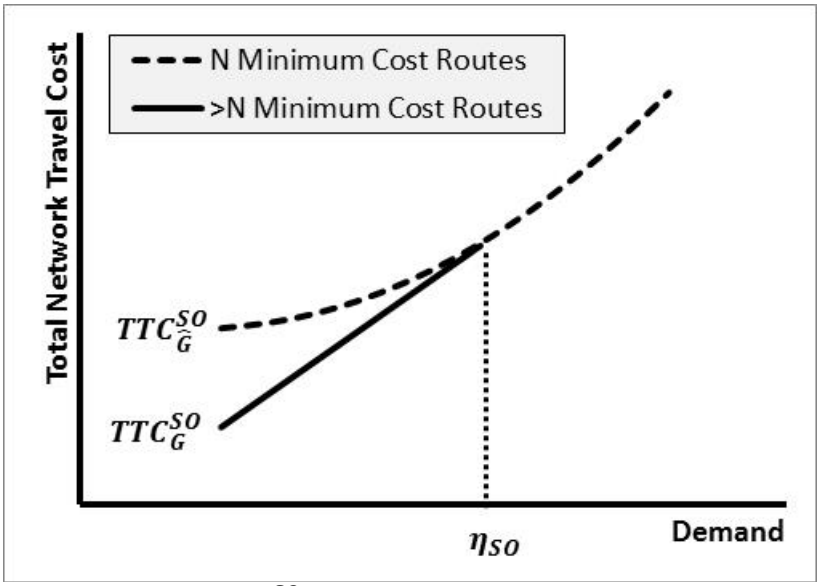

Figure 6.6 - The effect on TTC^{SO} of one or more contractions in $\widetilde{K}_{min}^{r}$, for some OD movements r

6.4.3.2 The Sensitivity of Total Network Travel Cost under UE to Route Transition Points

Similarly to section 6.4.3.1, this section begins by characterising the existence of derivatives of TTC^{UE} with respect to demand. These derivatives depend, by construction, upon the sensitivity of link flows x_i^{UE} with respect to increases in demand.

In the context of the UE traffic assignment problem, Patriksson (2004) provides a characterisation of the existence of directional derivatives and full derivatives of links flows. This is achieved through the derivation of a sensitivity problem, which yields directional derivatives of links flows provided that it has a unique solution. Josefsson and Patriksson (2007) built on Patriksson (2004) to show that, in traffic networks with separable link cost functions, a sufficient condition for the existence of a directional derivative of a link flow x_i^{UE}, is that the corresponding cost function c_i has a strictly positive derivative. For demands Q at which directional derivatives of link flows do exist, it follows from theorem 10 of Patriksson (2004) that full derivatives of those link flows also exist; provided it can be shown that

$\partial f_k^r/\partial q = 0$, $\forall k \in K^r$ for which $f_k^r = 0$ in every possible route flow solution F, and for any perturbation of demand. Within this statement, the derivatives $\partial f_k/\partial q$ must be consistent with the set of derivatives $\partial x_i/\partial q$, which uniquely solves the sensitivity problem.

The following result proves that TTC^{UE} is differentiable for $\forall Q \neq \eta_{UE}$. In this proof $K^r_{\sim min}$ denotes the set of routes $k \in K^r$ for which $C_k^r > \pi_r$.

Proposition 6.10: Consider a traffic network G for which Assumption A1 holds. $TTC^{UE}(Q)$ is differentiable $\forall Q \notin \mathrm{H}_{UE}$.

Proof: Suppose, for a traffic network G, that a demand $Q \notin \mathrm{H}_{UE}$ is given. By definition 6.3 it follows that $\lim_{Q\to\eta_{UE}^-} Y(Q) = \lim_{Q\to\eta_{UE}^+} Y(Q)$, for all trajectories of demand about η_{UE}. It therefore follows that there exists a neighbourhood of demand about Q for which, $\forall k \in K^r$ for each OD movement r, either (a) $k \in K^r_{min}$ for all Q in this neighbourhood, or (b) $k \in K^r_{\sim min}$ for all Q in this neighbourhood. In other words, the set K^r_{min} of routes that are of minimum cost, for each OD movement r, and the set $K_{\sim min}$ of routes that have costs strictly greater than minimum cost, for each OD movement r, do not change due to a small perturbation of demand.

By the UE conditions (5), it follows that $\forall k \in K^r_{\sim min}$ that $f_k^r = 0$, for each OD movement r. It also follows from the above argument that, for any small perturbation of demand, $f_k^r = 0$ will remain true. It consequently follows, from theorem 10 of Patriksson (2004), that link flows x_i^{UE} are differentiable for each link i for which $x_i^{UE} > 0$. As all link flows x_i^{UE} are differentiable functions of Q for all links i for which $x_i^{UE} > 0$, and all other links, for which there is no information about differentiability, have $x_i^{UE} = 0$, it follows that TTC^{UE} is differentiable at Q. This is by construction of TTC^{UE}, because it is a sum of products of differentiable functions. ■

The contrapositive result of proposition 6.10 is that all instances of demand Q, at which TTC^{UE} is not differentiable, must correspond to route transition points η_{UE}.

Conjectures 6.11 and 6.12 present claims for the behaviour of TTC^{UE} at route transition points of the types described in conditions C3(i) and C3(ii) respectively.

Conjecture 6.11: Consider a traffic network G for which Assumption A1 holds, and let η_{UE} represent a route transition point of type C3(i). Then:

$$\lim_{Q\to\eta_{UE}^-}\left(\frac{\partial}{\partial q}TTC^{UE}\right) > \lim_{Q\to\eta_{UE}^+}\left(\frac{\partial}{\partial q}TTC^{UE}\right)$$

Conjecture 6.12: Consider a traffic network G for which Assumption A1 holds, and let η_{UE} represent a route transition point of type C3(ii). Then:

$$\lim_{Q\to\eta_{UE}^{-}}\left(\frac{\partial}{\partial q}TTC^{UE}\right) < \lim_{Q\to\eta_{UE}^{+}}\left(\frac{\partial}{\partial q}TTC^{UE}\right)$$

The above conjectures are stated without proof. Numerical evidence supporting the truth of these conjectures can be found in the examples included in section 6.5. It is also noted that the directions of change in gradient are the same as the directions of change that have been proven for the SO case in propositions 6.8 and 6.9. It would therefore be counterintuitive if these results were not true in general. This is especially so in the case of traffic networks with cost functions of the form $c_i = a_i + b_i x_i^{\beta}$, for which theorem 6.5 proves that there is a systematic relationship between link flows under UE and SO.

It is remarked that proof of these conjectures is challenging because it is not possible to guarantee that the directional derivatives, stated in Conjectures 6.11 and 6.12, always exist (Josefsson and Patriksson, 2007). As an example, Josefsson and Patriksson (2007) remarked that directional derivatives cannot be guaranteed for the BPR cost functional form because it has zero cost derivative at zero flow. Given this difficulty, it is particularly noteworthy that TTC^{SO} is fully differentiable at all points of demand Q, including all route transition points η_{SO}, given the similarities that exist between the UE and SO models.

6.4.3.3 The Sensitivity of the Price of Anarchy to Route Transition Points

This section describes the implications of the results of sections 6.4.3.1 and 6.4.3.2 for the Price of Anarchy; starting with differentiability.

Corollary 6.13: Consider a traffic network G for which Assumption A1 holds. The Price of Anarchy is a differentiable function for all demand movements $q_r > 0$, for which $Q \notin \mathrm{H}_{UE}$.

Proof: Follows from propositions 4.1 and 4.4. ■

The results that follow describes the differing effects on the Price of Anarchy of route transition points of the types described in conditions C3(i) and C3(ii), under UE and SO. The results for the UE case are stated only as conjectures.

Theorem 6.14: Consider a traffic network G for which Assumption A1 holds.

(i) For a demand η_{SO}, which corresponds to a route transition point that satisfies condition C3(i):

$$\tilde{\rho}(Q) < \rho(Q), \forall Q \to \eta_{SO}^{+}$$

where $\tilde{\rho}$ represents a continuation of the trajectory of ρ for $Q \to \eta_{SO}^{-}$, into $Q \to \eta_{SO}^{+}$.

(ii) For a demand η_{SO}, which corresponds to a route transition point that satisfies condition C3(ii):

$$\tilde{\rho}(Q) > \rho(Q), \forall Q \to \eta_{SO}^{+}$$

where $\tilde{\rho}$ represents a continuation of the trajectory of ρ for $Q \to \eta_{SO}^{-}$, into $Q \to \eta_{SO}^{+}$.

Proof: Proof of (i) follows from proposition 6.8 and the fact that TTC^{SO} is on the denominator of ρ. Proof of (ii) follows from proposition 6.9, the associated discussion that followed and the fact that TTC^{SO} is on the denominator of ρ. ■

Conjecture 6.15: Consider a traffic network G for which Assumption A1 holds.

(i) At a demand η_{UE}, which corresponds to a route transition point that satisfies condition C3(i):

$$\lim_{Q \to \eta_{UE}^{-}} \left(\frac{\partial \rho}{\partial q} \right) > \lim_{Q \to \eta_{UE}^{+}} \left(\frac{\partial \rho}{\partial q} \right)$$

(ii) At a demand η_{UE}, which corresponds to a route transition point that satisfies condition C3(ii):

$$\lim_{Q \to \eta_{UE}^{-}} \left(\frac{\partial \rho}{\partial q} \right) < \lim_{Q \to \eta_{UE}^{+}} \left(\frac{\partial \rho}{\partial q} \right)$$

Proofs of parts (i) and (ii) of conjecture 6.15 will follow if conjectures 6.11 and 6.12 are true.

6.4.4 The Variation of the Price of Anarchy for High Travel Demand

As travel demand values q_r become larger, the network becomes saturated as the delay components of travel cost begin to dominate the free-flow component. In the network example in section 6.3.1.1, it was shown that expansions in the sets K_{min}^{r} and $\widetilde{K}_{min}^{r}$ eventually stop once demand reaches a sufficiently high threshold. This matches observations from numerical examples.

For the special case of traffic networks with cost functions of the form $c_i = a_i + b_i x_i^{\beta}$ $(a_i, b_i, \beta > 0)$, it is conjectured that, as demand Q continues to increase, the Price of Anarchy enters a region of decay that can be characterised by a power law. This characterisation is stated, without proof, in conjecture 6.16, and is illustrated in the numerical examples that follow in section 6.5.

Conjecture 6.16: Consider a traffic network G that serves a demand matrix Q with entries $q_r > 0$, and that has cost functions of the form $c_i = a_i + b_i x_i^{\beta}$ $(a_i, b_i, \beta > 0)$, which satisfy Assumption A1. Let ζ represent a global demand multiplier applied to the demand matrix Q. Then, as $\zeta \to \infty$, the leading order behaviour of the Price of Anarchy is $O\left(1/\zeta^{2\beta}\right)$.

6.5 Numerical Examples

This section presents four numerical examples, which provide illustrations of the theoretical results presented in sections 6.3 and 6.4; and also provide numerical evidence to support

conjectures 6.11, 6.12, 6.15 and 6.16, which were stated without formal proof. The first example in section 6.5.1 addresses the simplest scenario of the variation of the Price of Anarchy with increasing demand on a single OD pair. The second example in section 6.5.2 then presents a more complicated scenario; in which travel demand is increased, at different rates, on several OD pairs between a single origin and several destinations. The final two examples in section 6.5.3 then present two scenarios in which demand is uniformly increased on several OD pairs, between multiple origins and multiple destinations.

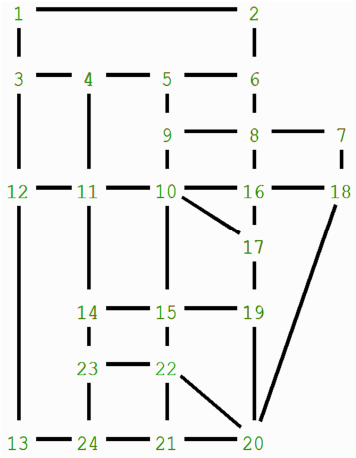

Figure 6.7 - Sioux Falls Network

The numerical examples in this section are based on the canonical test network of Sioux Falls[20], which is shown in Figure 6.7. This network comprises 24 nodes and 76 links, and the cost of travel c_i on each link i is represented by a BPR cost function with power $\beta = 4$, which is common to all links. Note that this network satisfies the conditions stated in Assumption A1, theorem 6.5 and corollary 6.6.

The results for each example are compiled from UE and SO traffic assignments undertaken at several discrete levels of travel demand. At each demand level j, travel demand q_r, on each OD movement r, is increased by a demand multiplier ζ_j^r; where $\zeta_j^r < \zeta_{j+1}^r \; \forall r, j$. This guarantees that demand is always increasing on each OD movement and therefore satisfies condition C1. As each traffic assignment is undertaken for discrete values of the demand multipliers ζ_j^r, it is not possible to identify the *exact* levels of demand at which each route transition point occurs. These levels of demand are therefore approximated in the analysis that follows by the first demand level j beyond the route transition point; this being the first level of demand at which it is possible to observe that either the minimum (marginal total) cost route set for an OD movement r or the OD specific active network for an OD movement r has

[20] Network and demand matrix files for Sioux Falls were obtained from Bar-Gera (2001).

changed. Each traffic assignment is calculated using the OBA algorithm, solved to an average excess cost of, at most, 10^{-9}.

6.5.1 Example 1: Increasing Demand in a Single Origin-Destination Pair Network

In this single OD pair scenario, the variation of the Price of Anarchy is studied as travel demand $q = 10$ is increased, using demand multipliers $\zeta_j = 1,2,\ldots,10000$, on the OD movement between node 20 and node 3 in the Sioux Falls network. Figure 6.8 displays the variation of the Price of Anarchy against travel demand q. The vertical lines in this figure signify levels of demand corresponding to route transition points η_{UE} and η_{SO}, at which the OD specific active networks X_1^{UE} (green lines) and X_1^{SO} (red lines), expand (solid lines) and contract (dashed lines). Recall that OD specific active networks provide an alternative characterisation for the minimum (marginal total) cost route set under UE and SO. This figure also displays graphs of the Price of Anarchy for 17 sub-networks (denoted ρ_1, ρ_2, etc), which correspond to the 17 different states of the active network, between route transition points, as demand increases.

Focussing on the graph for the full network, Figure 6.8 displays the same three identifiably distinct regions of behaviour of the Price of Anarchy that are evident in Figure 5.19: an initial region in which the Price of Anarchy is one, a period of fluctuations, followed by a decay back towards one. It can be seen that the Price of Anarchy varies smoothly $\forall q \in \mathrm{H}_{UE}$, which is consistent with corollary 6.13, and three of the four effects of expansions and contractions described in theorem 6.14 and conjecture 6.15 are clearly visible. For UE, at all points η_{UE} corresponding to an expansion of X^{UE}, the Price of Anarchy is non-differentiable and there is a decrease in the gradient of the Price of Anarchy, which provides numerical evidence to support conjecture 6.15(i). At the single point $\eta_{UE} \approx 38{,}000$, which corresponds to a contraction of X^{UE}, the Price of Anarchy is also non-differentiable and there is an increase in the gradient of the Price of Anarchy, which provides numerical evidence to support conjecture 6.15(ii). For SO, at all points η_{SO}, which correspond to an expansion of X^{SO}, the Price of Anarchy is smooth but transfers onto a higher trajectory than the Price of Anarchy for the sub-network that detaches, which illustrates theorem 6.14(i). The effect of a contraction in X^{SO} at a route transition point η_{SO} described in theorem 6.14(ii); for which there is a single point in this example at $\eta_{SO} \approx 25{,}000$, is less apparent.

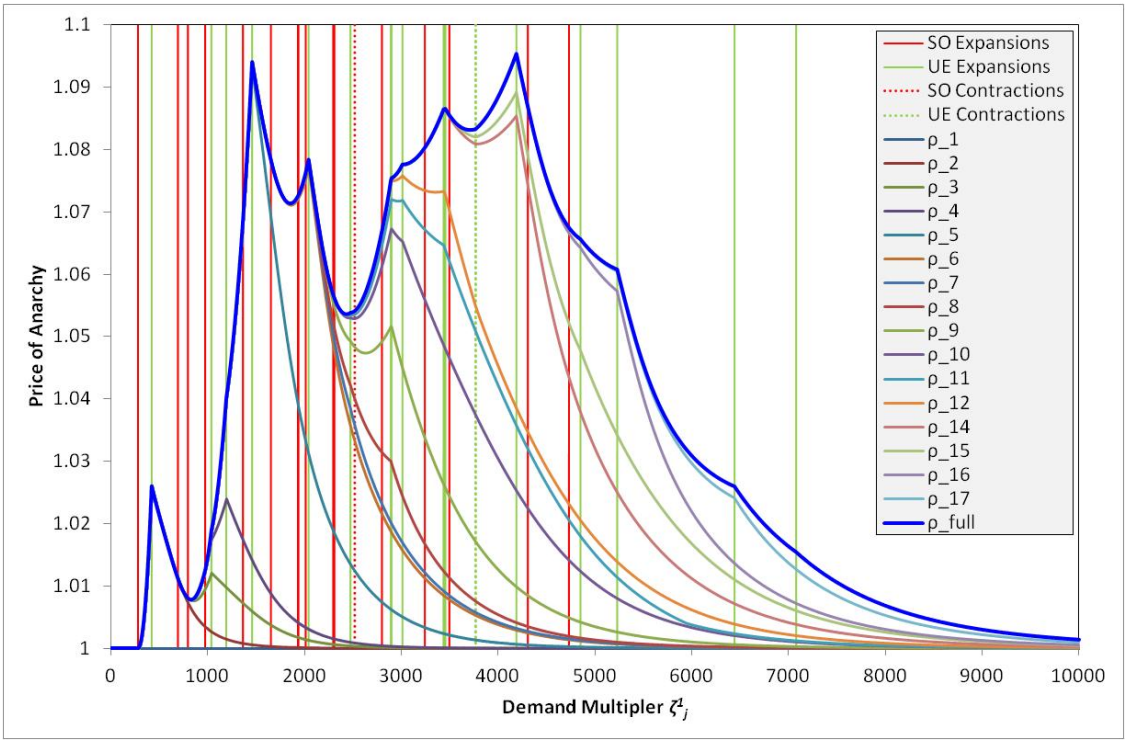

Figure 6.8 – The Variation of the Price of Anarchy against the Demand Multiplier ζ_j in Example 1

Turning to the systematic relationship between route transition points η_{UE} and η_{SO}, Table 6.1 lists the approximate levels of demand for each state of the active network as demand increases. The table also presents the value of η_{SO}/η_{UE} at each route transition point and shows the number of links that are active in each state of X_1^{UE} and X_1^{SO}. Given that, for $\beta = 4$, $1/\sqrt[\beta]{\beta+1} \approx 0.67$, the results in this table are consistent with the conclusions of corollary 6.6.

No.	Route Transition Points		$\frac{\eta_{SO}}{\eta_{UE}}$	Number of Active Links in X_1^{UE} & X_1^{SO}
	$\eta_{SO} = \zeta_j$	$\eta_{UE} = \zeta_j$		
1	1	1	-	6
2	285	426	0.6690	12
3	697	1,042	0.6689	14
4	800	1,196	0.6689	17
5	978	1,463	0.6685	19
6	1,368	2,046	0.6686	23
7	1,657	2,478	0.6687	28
8	1,936	2,895	0.6687	29
9	1,941	2,902	0.6688	30
10	2,016	3,015	0.6687	31
11	2,300	3,439	0.6688	32
12	2,313	3,458	0.6689	33

13	2,,520	3,769	0.6686	32
14	2,803	4,191	0.6688	34
15	3,246	4,853	0.6689	35
16	3,499	5,232	0.6688	36
17	4,309	6,443	0.6688	37
18	4,734	7,079	0.6687	38

Table 6.1 - Route Transition Points in Example 1

Finally, Figure 6.9 displays the decay rate of the Price of Anarchy for demand $q > 7{,}044$, which represents the level of demand of the final route transition point η_{UE}. This figure also plots a trend-line; calculated by Ordinary Least Squares regression, which shows that the decay in the Price of Anarchy is consistent with $O\left(1/\zeta_j^{2\beta}\right)$. Figure 6.9 also displays decay rates of the Price of Anarchy in adjusted versions of the Sioux Falls network for values of $\beta = 1{,}2{,}3$. The decay in each of these additional scenarios, from the point of the final route transition point η_{UE}, is also $O\left(1/\zeta_j^{2\beta}\right)$. These findings are consistent with conjecture 6.16.

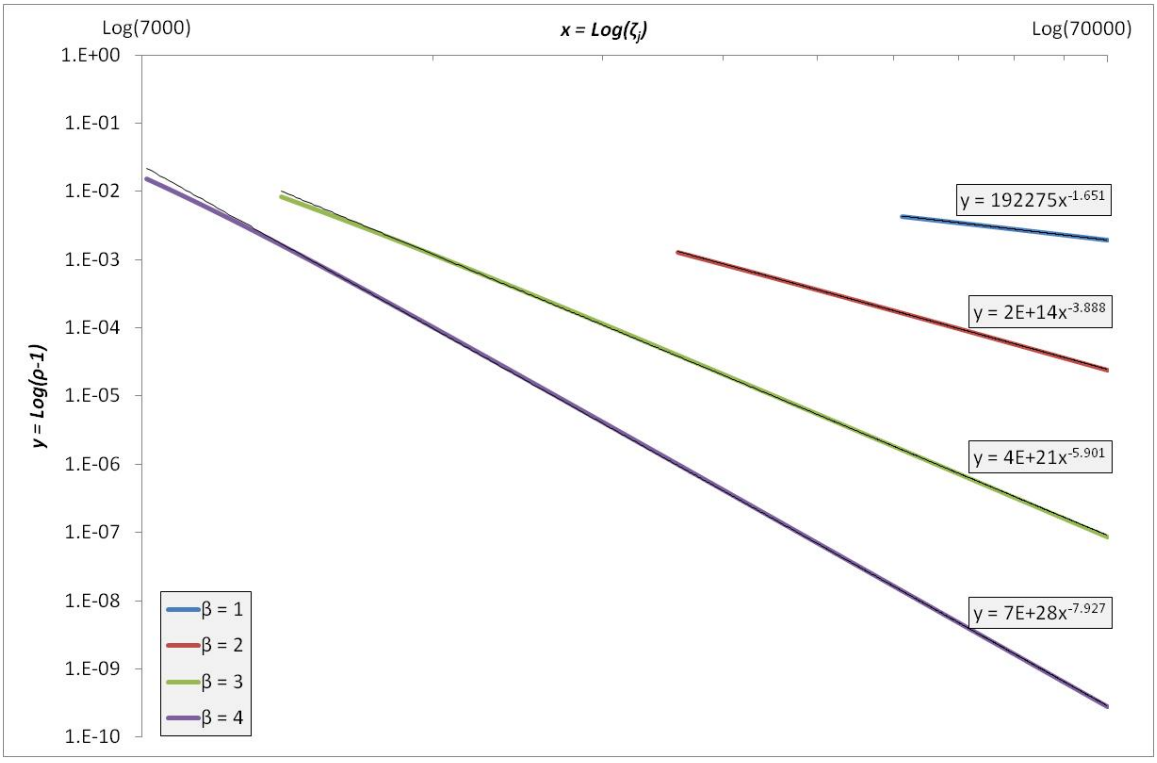

Figure 6.9 – Decay in the Price of Anarchy for High Demand in Example 1 for $\boldsymbol{\beta = 1}$, $\boldsymbol{\beta = 2}$, $\boldsymbol{\beta = 3}$, $\boldsymbol{\beta = 4}$

6.5.2 Example 2: Increasing Demand in a Multiple (One to Many) Origin-Destination Pair Network

In this multiple OD pair scenario, the variation of the Price of Anarchy is studied as travel demand is increased on 22 OD pairs; between a single origin at node 20 and destination nodes

$s_r = 1, 2, 3, \dots, 19, 21, 22, 23$ in the Sioux Falls network. The initial amount of demand q_r and the demand multipliers ζ_j^r are different for each of the 22 OD movements. This therefore represents a more complicated scenario than the single OD example that was explored in section 6.5.1. The initial amount of demand on each OD movement r is set at $q_r = 24 - s_r$. The demand multipliers ζ_j^r for each OD movement r are then set at $\zeta_j^r = j \times (1 + 0.01 s_r)$, with values of $j = 1,2, \dots, 2000$.

Figure 6.10 displays the variation of the Price of Anarchy against index values $j = 1,2, \dots, 400$ for the demand multipliers ζ_j^r. Similarly to Figure 6.8, the vertical lines in this figure signify levels of demand corresponding to route transition points η_{UE} and η_{SO}, at which one or more OD specific active networks X_r^{UE} and X_r^{SO} expand or contract. Even with the greater complexity of this example, Figure 6.10 provides further numerical evidence to support conjecture 6.15 and further illustrations of theorem 6.14. In particular, the increase in gradient of the Price of Anarchy at $\eta_{UE} = 180$ is much clearer than in Figure 6.8.

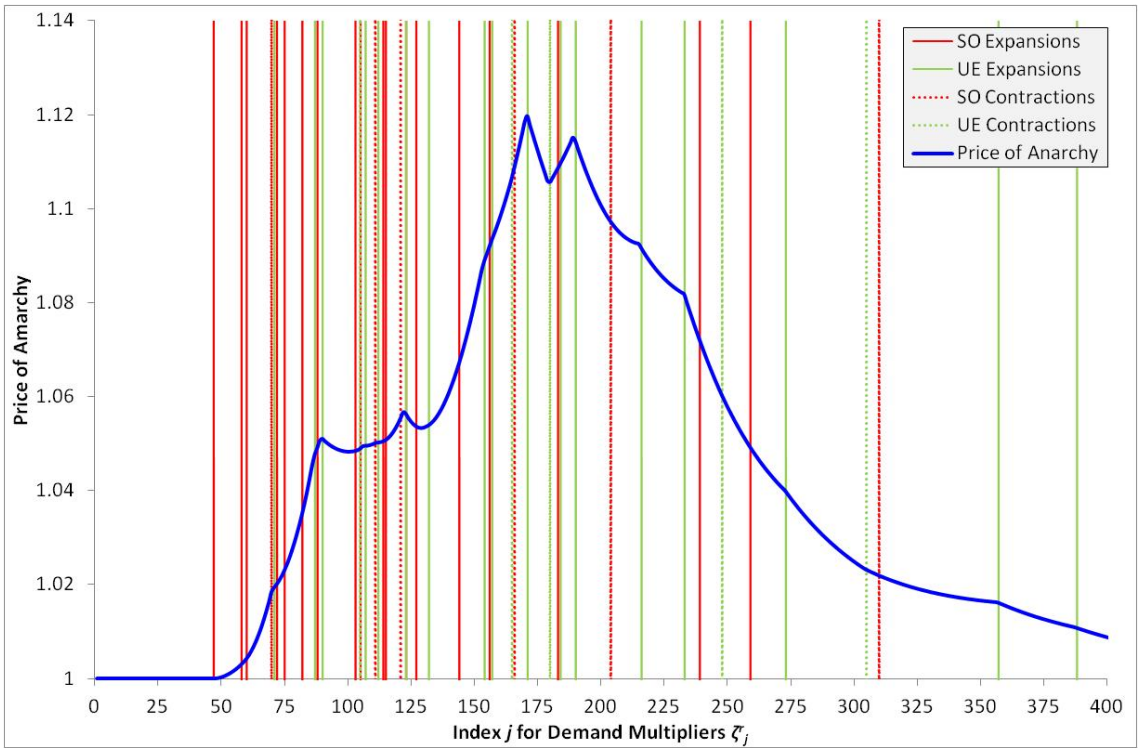

Figure 6.10 – The Variation of the Price of Anarchy against the Index j for Demand Multipliers ζ_j^r in Example 2

Turning to the systematic relationship between route transition points η_{UE} and η_{SO}, Table 6.2 lists the approximate levels of demand for each route transition point under UE and SO. In contrast to Table 6.1, the final column of this table displays the number of links with positive flow of the total of 38 links in Sioux Falls. These results are again consistent with the conclusions of corollary 6.6.

No.	Route Transition Points		$\frac{\eta_{SO}}{\eta_{UE}}$	Number of Active Links in $\cup_r X_r^{UE}$ & $\cup_r X_r^{SO}$
	$\eta_{SO} = j$	$\eta_{UE} = j$		
1	1	1	-	24
2	47	71	0.662	25
3	58	87	0.6667	26
4	60	90	0.6667	27
5	70	105	0.6667	26
6	72	107	0.6729	27
7	75	112	0.6696	28
8	82	123	0.6667	29
9	88	132	0.6667	30
10	103	154	0.6688	31
11	105	157	0.6688	32
12	111	165	0.6727	31
13	114	171	0.6667	32
14	115	171	0.6725	33
15	121	180	0.6722	32
16	123	184	0.6685	33
17	127	190	0.6684	34
18	144	216	0.6667	35
19	156	233	0.6695	36
20	166	248	0.6694	35
21	183	273	0.6703	36
22	204	305	0.6689	36
23	239	357	0.6695	37
24	259	388	0.6675	38
25	310	464	0.6681	37
26	418	625	0.6688	38

Table 6.2 - Route Transition Points in Example 2

Finally, Figure 6.11 displays the decay rate of the Price of Anarchy for demand indices $j > 625$, which represents the level of demand of the final route transition point η_{UE}. This decay is consistent with $O\left(1/j^{2\beta}\right)$ as is proposed in conjecture 6.16.

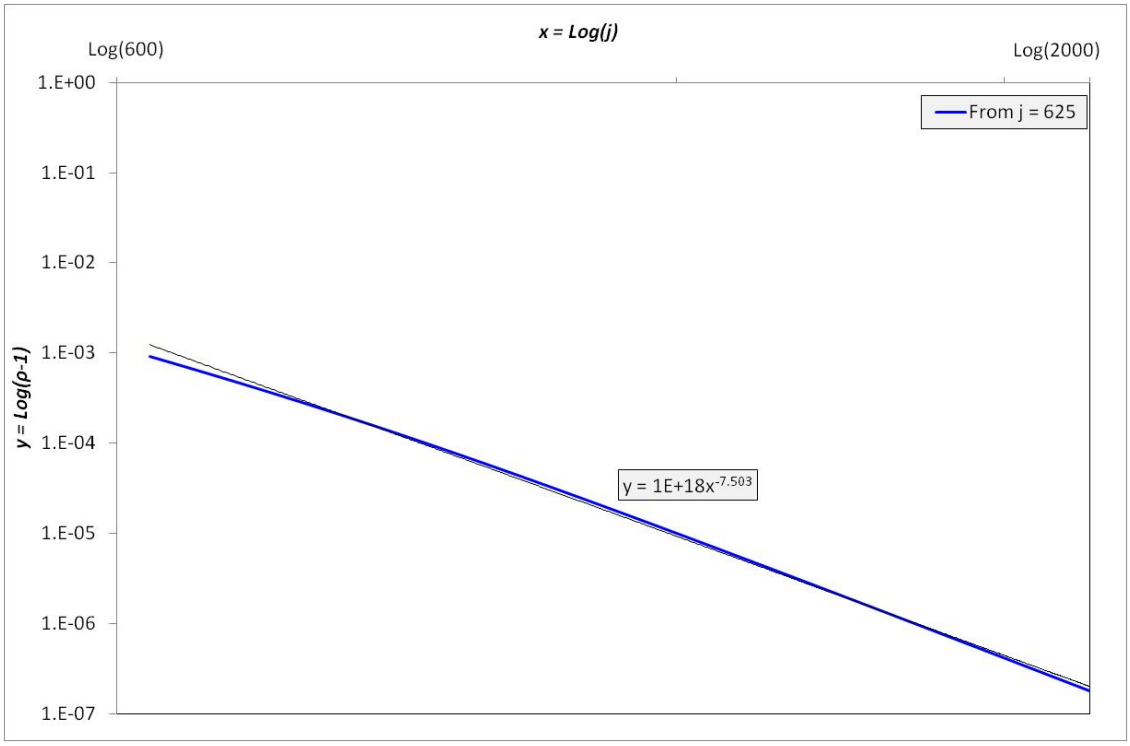

Figure 6.11 - Decay in the Price of Anarchy for High Demand in Example 2

6.5.3 Examples 3 and 4: Increasing Demand in a Multiple (Many to Many) Origin-Destination Pair Network

In addition to illustrating the theoretical results and conjectures of sections 6.3 and 6.4, the two examples in this section also illustrate challenges that exist in identifying route transition points in more complicated multiple OD networks.

6.5.3.1 Sioux Falls Network: Five Origin-Destination Pair Example

In this first multiple OD pair scenario, the variation of the Price of Anarchy is studied as travel demand is increased on five OD pairs $r = 1, \dots, 5$ in the Sioux Falls network: between node 20 and node 1; node 23 and node 2; node 20 and node 3; node 7 and node 13; and between node 1 and node 19. The initial amounts of demand on each OD movement were set at $q_1 = 23$, $q_2 = 14$, $q_3 = 17$, $q_4 = 18$ and $q_5 = 28$. The demand multipliers for each OD movement were identical, with values $\zeta_j^1 = \cdots = \zeta_j^5 = \zeta_j = 1,2, \dots, 8000$.

Figure 6.12 displays the variation of the Price of Anarchy against demand multipliers up to $\zeta_j = 1000$. Similarly to previous figures the vertical lines signify levels of demand corresponding to route transition points η_{UE} and η_{SO}. As OBA is unable to identify OD specific active networks in network examples with multiple origin nodes, the vertical lines represent only those route transition points at which an expansion (contraction) in an OD specific active network coincides with an expansion (contraction) in the overall active network, which is equivalent to $\bigcup_r X_r^{UE}$ and $\bigcup_r X_r^{SO}$. The overall active network is uniquely defined by link flows.

This demonstrates a limitation of using the OBA algorithm to identify changes in OD specific active networks for cases in which there are multiple origins. The consequence of this is that there may be route transition points that exist, which this method does not identify. Indeed, at $\zeta_j \approx 500$, there is a 'downward kink' in the graph of the Price of Anarchy, which suggests that there is a route transition point η_{UE} corresponding to the expansion of K^r_{min} for some OD movement r. The identification of route transition points, through observation of OD specific active networks, in this general case would require the TAPAS algorithm. The example in section 6.5.3.2 demonstrates an alternative approach to identifying route transition points, which uses route enumeration.

Despite this limitation of OBA, the behaviour of the Price of Anarchy at all other route transition points accords with the claims made in conjecture 6.15 and provides further illustrations of the statements in theorem 6.14.

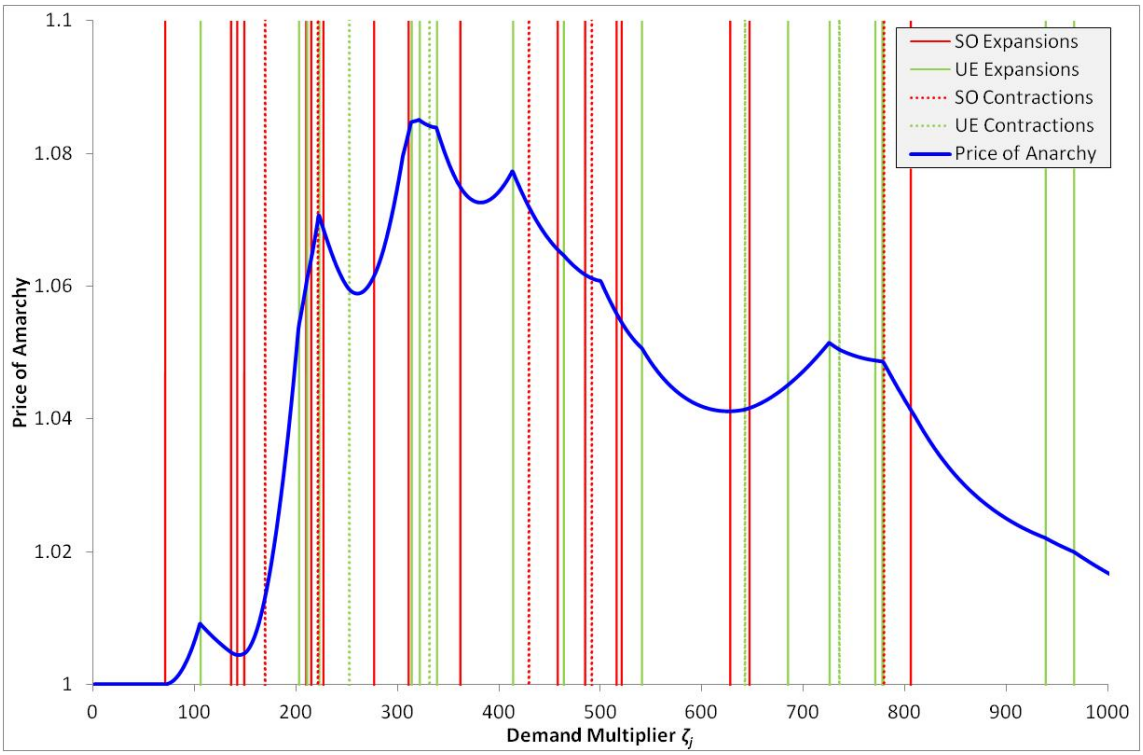

Figure 6.12 – The Variation of the Price of Anarchy against the Demand Multiplier ζ_j in Example 3

For each of the vertical lines in Figure 6.12, Table 6.3 lists the approximate levels of demand at which the overall active network changes as travel demand increases. Similarly to previous examples, these results are consistent with the conclusions of corollary 6.6.

No.	Route Transition Points		$\frac{\eta_{SO}}{\eta_{UE}}$	Number of Active Links in $\cup_r X_r^{UE}$ & $\cup_r X_r^{SO}$
	$\eta_{SO} = \zeta_j$	$\eta_{UE} = \zeta_j$		
1	1	1		21

2	71	106	0.6698	24
3	136	203	0.67	37
4	142	211	0.673	42
5	149	223	0.6682	46
6	170	253	0.6719	41
7	210	314	0.6688	46
8	215	322	0.6677	48
9	222	332	0.6687	46
10	227	339	0.6696	49
11	277	414	0.6691	50
12	311	464	0.6703	52
13	362	541	0.6691	56
14	430	643	0.6687	55
15	458	685	0.6686	56
16	485	726	0.668	57
17	492	736	0.6685	56
18	516	771	0.6693	57
19	521	778	0.6697	58
20	521	779	0.6688	59
21	628	939	0.6688	60
22	647	967	0.6691	61
23	780	1,166	0.669	60
24	806	1,205	0.6689	62
25	1,037	1,551	0.6686	63
26	1,428	2,136	0.6685	62

Table 6.3 - Route Transition Points in Example 3

Finally, Figure 6.13 displays the decay rate of the Price of Anarchy for values of the demand multiplier $\zeta_j > 2136$, which represents the level of demand of the final route transition point η_{UE}. This decay is consistent with $O\left(1/\zeta_j^{2\beta}\right)$ as is proposed in conjecture 6.16.

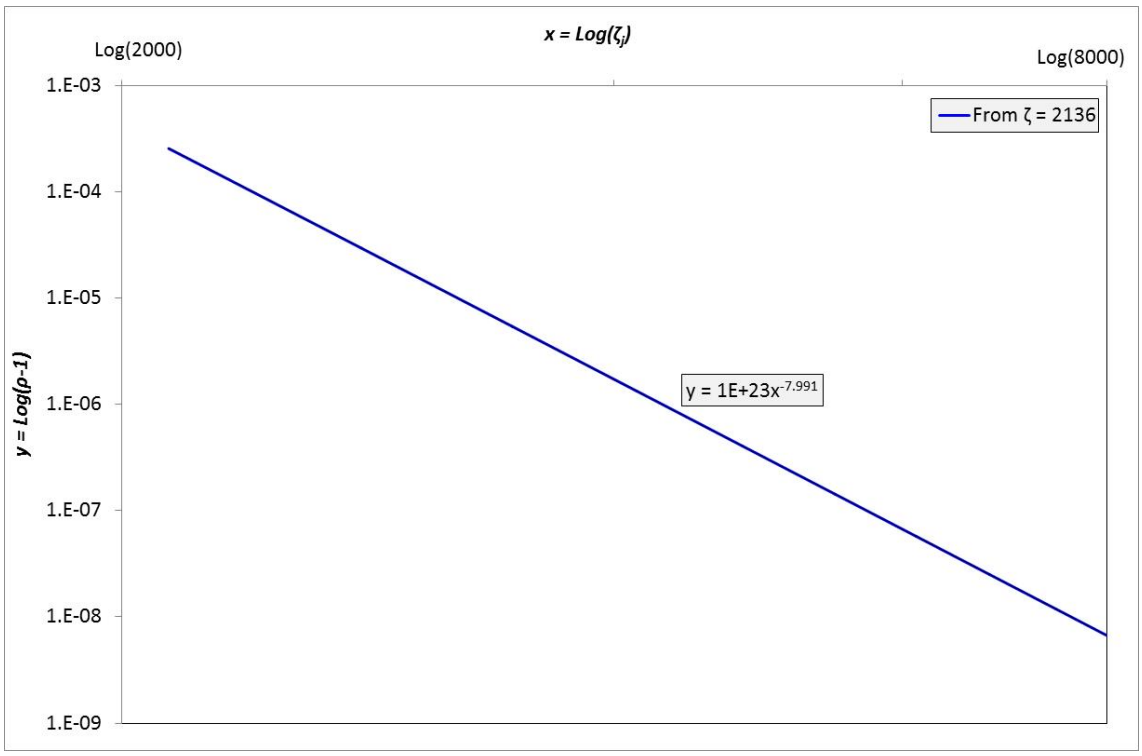

Figure 6.13 - Decay in the Price of Anarchy for High Demand in Example 3

6.5.3.2 Sioux Falls Network: 528 Origin-Destination Pair Example

In this second multiple OD pair scenario, the variation of the Price of Anarchy is studied as travel demand is increased in the Sioux Falls network, using the demand matrix file that is available at Bar-Gera (2001). This demand matrix contains 528 OD pairs. The initial amounts of demand on each OD movement are set at $q_r = 0.001q'_r$, where q'_r represents the value in the original matrix. Demand multipliers for each OD movement are then identical, with values $\zeta_j^1 = \cdots = \zeta_j^{528} = \zeta_j = 1{,}2, \ldots, 9000$.

Figure 6.14 displays the variation of the Price of Anarchy against demand multipliers up to $\zeta_j = 2000$. Similarly to Figure 6.12, this figure also uses vertical lines to signify levels of demand that correspond to route transition points η_{UE} and η_{SO} at which there is a change in the overall active networks $\bigcup_r X_r^{UE}$ and $\bigcup_r X_r^{SO}$. For this example, there are only two such route transition points, which Table 6.4 shows both satisfy the conclusions of corollary 6.6.

In order to better identify the full sets of route transition points H_{UE} and H_{SO}, an alternative methodology was employed in which, at each demand level, the number of routes were counted, for each OD movement r, that were within a tolerance 10^{-10} of the minimum (marginal total) cost route under UE and SO. This was inspired by the approach described in Bar-Gera (2006). This method identifies a total of 364 demand levels $\zeta_j \in [0{,}2000]$, which correspond to route transition points η_{UE} and η_{SO} for this network. As the inclusion of a vertical line for each of these points would make Figure 6.14 unintelligible, the difference

between the total numbers of minimum (marginal total) cost routes under UE and SO is plotted instead. Referred to as the *difference measure*, this measure is calculated, for each demand level j, as $\sum_r |\widetilde{K}^r_{min}(\zeta_j)| - \sum_r |K^r_{min}(\zeta_j)|$. Although this is a particularly coarse measure, it can be seen that it has a similar overall pattern to the Price of Anarchy (though with different magnitude). This measure, therefore, provides further numerical evidence to support the claims of conjecture 6.15 and the conclusions of theorem 6.14.

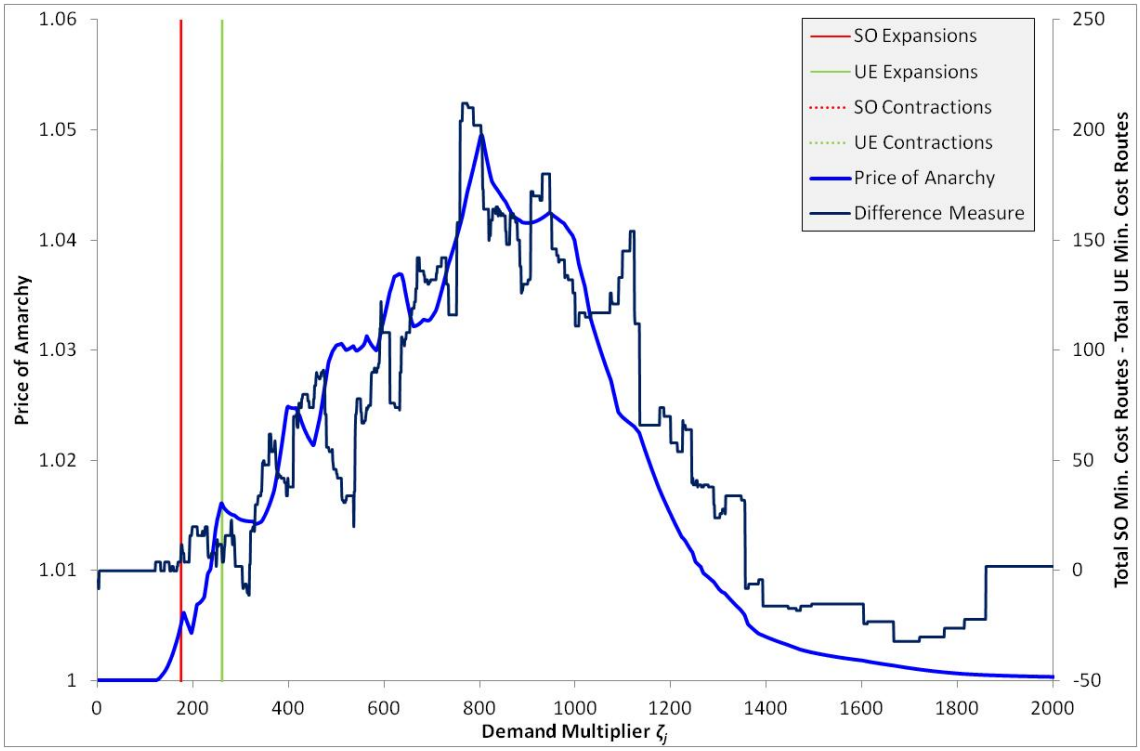

Figure 6.14 – The Variation of the Price of Anarchy against the Demand Multiplier ζ_j in Example 4

No.	Route Transition Points		$\frac{\eta_{SO}}{\eta_{UE}}$	Number of Active Links in $\cup_r X_r^{UE}$ & $\cup_r X_r^{SO}$
	$\eta_{SO} = \zeta_j$	$\eta_{UE} = \zeta_j$		
1	1	1		74
2	71	106	0.6698	75
3	136	203	0.67	76

Table 6.4 - Route Transition Points in Example 4

For values of $\zeta_j > 2000$, the difference measure becomes increasingly unstable as demand increases. This is an indicator that the level of convergence of 10^{-9} eventually (and inevitably) becomes unable to clearly identify expansions and contractions because of the magnitudes of travel costs. For this example, it is therefore not possible to identify the exact level of demand at which the final region of decay in the Price of Anarchy begins. For this reason, Figure 6.15 displays the decay rate of the Price of Anarchy for values of the demand multiplier $\zeta_j > 934$,

which signifies the first point in Figure 6.14 at which the Price of Anarchy begins to steadily fall. It can be seen from this figure that the decay rate of the Price of Anarchy eventually becomes consistent with $O\left(1/\zeta_j^{2\beta}\right)$, as is proposed in conjecture 6.16, for values of $\zeta_j > 4036$.

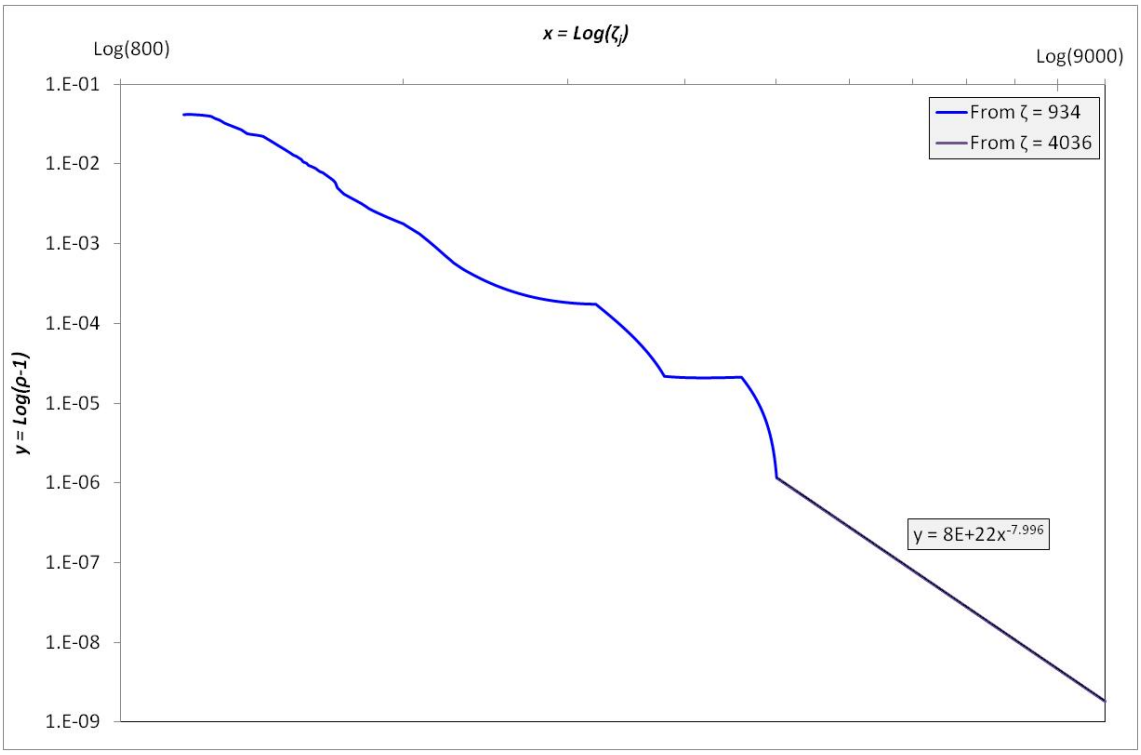

Figure 6.15 - Decay in the Price of Anarchy for High Demand in Example 4

7 Why values of the Price of Anarchy are small and an Alternative Measure for the Inefficiency of Selfish Routing

7.1 Introduction

It was noted in the discussion section at the end of chapter 5 that values of the Price of Anarchy across the network ensembles tested were consistently small, of the order of 1.05 or lower. This observation is also true of the numerical examples of section 6.5, for which the highest value achieved is of the order 1.12.

At face value, this would suggest that selfish routing under UE is relatively efficient in comparison with SO routing, and that, therefore, policy interventions, such as road pricing schemes, that are designed to induce more efficient routing behaviour would not be worthwhile because their benefits would be small in comparison with their costs of implementation. As an additional observation, it was also noted that values of the Price of Anarchy in the numerical experiments of chapter 5 and in numerical examples from the literature were significantly lower than the upper bounds that have been presented by Roughgarden (2003), for example. This raises questions of the usefulness of such upper bounds if they are so far removed from values of the Price of Anarchy observed in numerical studies.

This chapter explores the reasons behind these observations and goes on to propose an alternative measure of the inefficiency of selfish routing, which achieves higher values and also gets closer to the upper bounds of Roughgarden (2003). Section 7.2 shows that values are typically small in real traffic network examples because the Price of Anarchy measure has a sensitivity to the free-flow travel cost component. Section 7.3 then proposes a new measure of the inefficiency of selfish routing; called Price of Anarchy Delays, which does not suffer from this sensitivity. Section 7.4 proves that this measure is subject to the bounds of Roughgarden (2003) and section 7.5 then provides a numerical example of how this new measure varies with respect to travel demand in the numerical example of section 6.5.1.

7.2 Why are values of the Price of Anarchy small?

An examination of the formulation of Total Network Travel Cost within the definition of the Price of Anarchy reveals why values of the Price of Anarchy are small and why the Roughgarden (2003) bounds are not typically achieved for real road traffic networks.

To begin, first note that a route-based definition for the Price of Anarchy can be written as shown in equation (17).

$$\rho = \frac{TTC^{UE}}{TTC^{SO}} = \frac{\sum_r \sum_k f^r_{k,UE} C_k\left(f^r_{k,UE}\right)}{\sum_r \sum_k f^r_{k,SO} C_k\left(f^r_{k,SO}\right)} \tag{17}$$

To simplify notation, now consider a single OD network with K routes serving a demand q under the UE routing principle and suppose that, without loss of generality, routes $1, 2, \dots, \kappa_r$ are ordered such that route costs $C_1(0) \leq C_2(0) \leq \cdots \leq C_{\kappa_r}(0)$. It follows from the numerator of the route-based formulation shown in equation (17) that:

$$\begin{aligned} TTC &= \sum_{k=1}^{\kappa_r} f_k C_k(f_k) = \sum_{k=1}^{\kappa_r} f_k [C_k(0) + C_k(f_k) - C_k(0)] \\ &= \sum_{k=1}^{\kappa_r} f_k C_k(0) + \sum_{k=1}^{\kappa_r} f_k [C_k(f_k) - C_k(0)] \end{aligned} \tag{18}$$

Now as $C_1(0) \leq C_2(0) \leq \cdots \leq C_{\kappa_r}(0)$, define $\gamma_k = C_k(0) - C_1(0)$ for $k = 1, 2, 3, \dots, K$ to represent the additional free-flow costs of the longer routes $k = 2, 3, \dots, K$ for this OD pair, such that $C_1(0) \leq C_1(0) + \gamma_2 \leq \cdots \leq C_1(0) + \gamma_k$. Substituting the γ_k into the first term of TTC in equation (18) yields:

$$\begin{aligned} TTC &= \sum_{k=1}^{\kappa_r} f_k [C_1(0) + \gamma_k] + \sum_{k=1}^{\kappa_r} f_k [C_k(f_k) - C_k(0)] \\ &= \sum_{k=1}^{\kappa_r} f_k C_1(0) + \sum_{k=2}^{\kappa_r} f_k \gamma_k + \sum_{k=1}^{\kappa_r} f_k [C_k(f_k) - C_k(0)] \\ &= C_1(0) q + \sum_{k=2}^{\kappa_r} f_k \gamma_k + \sum_{k=1}^{\kappa_r} f_k [C_k(f_k) - C_k(0)] \end{aligned} \tag{19}$$

Equation (19) shows that TTC can be decomposed into a sum of: 1) the free-flow travel cost of routing all demand by the shortest path, 2) the additional free-flow travel costs incurred by those flows forced to use longer routes and 3) the travel delays due to congestion on all routes. Note also that all three of these cost components appear in both the numerator and denominator of the Price of Anarchy but that the first component in equation (19) is independent of the routing strategy; i.e. it takes the same value under both UE and SO. It is this free-flow cost component of TTC that makes the Price of Anarchy sensitive to free-flow travel costs in road traffic networks.

To illustrate this more clearly, consider the single OD 'lollipop' network example, shown in Figure 7.1, in which two routes serve a demand q from **O** to **D** for which the free-flow travel cost of the shortest path is equal to cost of travel on link 1. The terms a_1, a_3, b_2 and b_3 are positive coefficients.

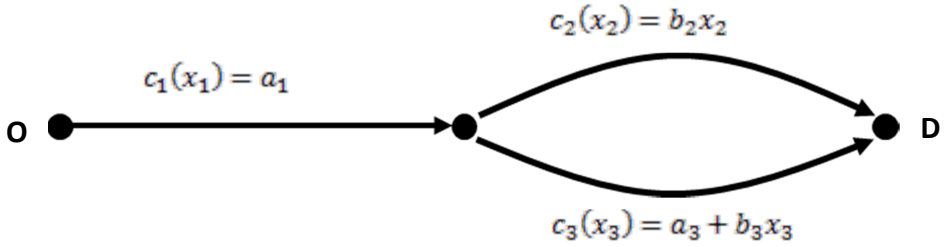

Figure 7.1 - Single OD 'Lollipop' Network

Combining equations (17) and (19), and setting link flows $x_1 = q$, $x_2 = x$ and $x_3 = q - x$ for some $x \in [0, q]$, the Price of Anarchy for this network can be written as follows:

$$\rho = \frac{TTC^{UE}}{TTC^{SO}} = \frac{a_1 q + [a_3(q - x_{UE})] + [b_2 x_{UE}^2 + b_3(q - x_{UE})^2]}{a_1 q + [a_3(q - x_{SO})] + [b_2 x_{SO}^2 + b_3(q - x_{SO})^2]} \quad (20)$$

For the first part of the journey from **O** to **D** all demand q uses link 1 because there are no alternatives. In the second part of the journey there is a choice of routes between links 2 and 3. The split of link flows x and $q - x$ between these two options is independent of the value of a_1 under both UE and SO. It is also clear from the formulation shown in equation (20) that the *absolute* difference in TTC between UE and SO (i.e. the numerator minus the denominator), and therefore the *absolute* benefit of rerouting, is independent of a_1 because the $a_1 q$ terms cancel out. However, as equation (20) shows, the Price of Anarchy is dependent on a_1 and, in particular, as $a_1 \to \infty$, $\rho \to 1$.

This dependence is illustrated by three numerical examples of the lollipop network shown in Figure 7.1, which have coefficients $a_3 = 2$, $b_2 = 1$ and $b_3 = 0.1$, with values of a_1 equal to 1, 3 and 5. The variation of the Price of Anarchy with travel demand q for these three examples is shown in Figure 7.2. This figure demonstrates how values of the Price of Anarchy fall as the free-flow cost component increases.

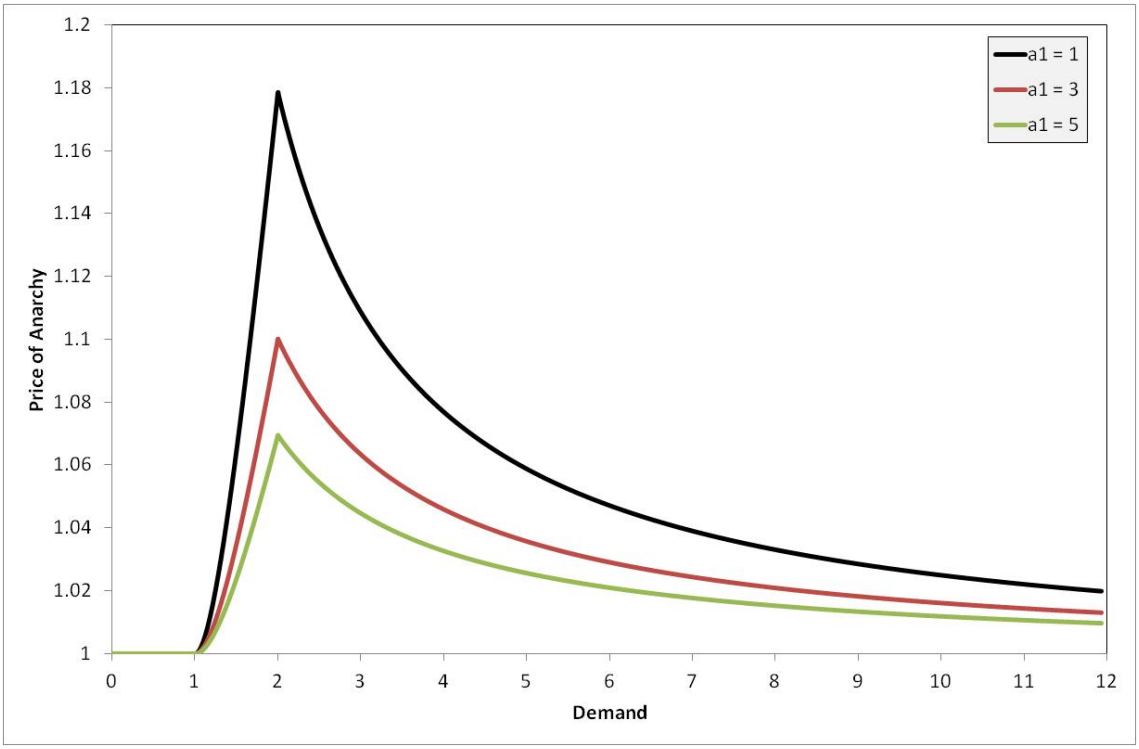

Figure 7.2 - The Variation of the Price of Anarchy with Travel Demand in Several Instances of the Lollipop Network shown in Figure 7.1

This example demonstrates that as the minimum travel cost that an individual traveller must pay to transit from their origin to their destination, which is the cost of the shortest path in free-flow conditions, becomes a larger proportion of the overall travel costs for the journey, the smaller the Price of Anarchy becomes. Therefore, whilst it is true that the Price of Anarchy captures the benefits of rerouting *relative* to the total cost of travel, it is argued that the Price of Anarchy masks the still potentially significant *absolute* benefits of rerouting. This is particularly significant in real road traffic networks because the free-flow component of cost typically represents a significant proportion of total travel cost (Correa et al., 2008). It also has implications for real models of road traffic networks in which constant cost centroid connectors are used to represent the cost of travel for travel flows entering a network from locations far outside the boundary of the network under study.

7.3 Price of Anarchy Delays: An Alternative Measure of the Inefficiency of Selfish Routing

In order to capture the absolute benefits of rerouting, this section proposes an alternative measure of the inefficiency of selfish routing called 'Price of Anarchy Delays'. This measure, which is defined below, isolates and compares the relative difference under UE and SO of only those cost components that are directly affected by the routing of flows, i.e. the second and third cost components in equation (19). In excluding free-flow costs; the first cost component

in equation (19), this new measure excludes the section of cost that is unavoidable to travellers and which cannot be altered by the routing strategy. Price of Anarchy Delays is defined as follows.

Definition 7.1: For a given road traffic network $G(N, A)$, with cost functions c and demand matrix Q, *Price of Anarchy Delays*; denoted ρ_d, is defined as:

$$\rho_d = \frac{TTC^{UE} - \sum_r C_1^r(0)q_r}{TTC^{SO} - \sum_r C_1^r(0)q_r} = \frac{\sum_{rs}\left(\sum_{k=2}^{\kappa_r} f_{k,UE}^r \gamma_k + \sum_{k=1}^{\kappa_r} f_{k,UE}^r \left[C_{k,UE}^r\left(f_{k,UE}^r\right) - C_{k,UE}^r(0)\right]\right)}{\sum_{rs}\left(\sum_{k=2}^{\kappa_r} f_{k,SO}^r \gamma_k + \sum_{k=1}^{\kappa_r} f_{k,SO}^r \left[C_{k,SO}^r\left(f_{k,SO}^r\right) - C_{k,SO}^r(0)\right]\right)}$$

Figure 7.3 shows how Price of Anarchy Delays varies as travel demand is increased in the same numerical examples displayed in Figure 7.2, in which the free-flow cost component varies from $a_1 = 1$ to $a_1 = 5$. Only one graph is visible because all three graphs actually overlap. This illustrates how Price of Anarchy Delays is independent of the free-flow travel cost component. A comparison of Figure 7.2 and Figure 7.3 also reveals that the values of Price of Anarchy Delays are significantly higher than values of the Price of Anarchy and are also much closer, at their maximum, to the upper bound of Roughgarden (2003), which is also shown by the dotted line in Figure 7.3.

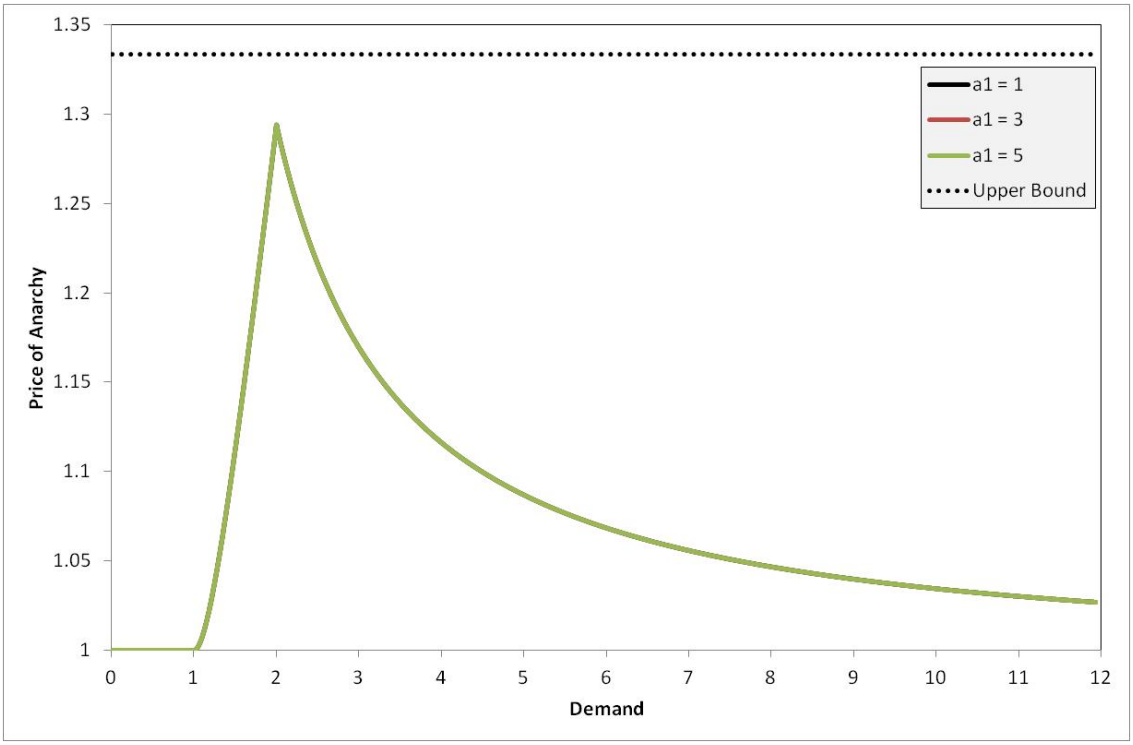

Figure 7.3 - The Variation of the Price of Anarchy with Travel Demand in the same instances of the Lollipop Network shown in Figure 7.2

This thesis does not propose that Price of Anarchy Delays is a better measure of the inefficiency of selfish routing than the Price of Anarchy, but rather that it provides an alternative point of viewpoint that complements the existing measure.

To visualise the additional insight that is provided by Price of Anarchy Delays consider the following real-world example; consider an individual commuter travelling from home to work for whom, on a normal weekday, under the assumption that all travellers act selfishly under the UE principle, such a journey takes 30 minutes of travel time. Now suppose that if travellers were to cooperate with each other, i.e. they were to choose routes under the SO principle, that the journey would take 20 minutes of travel time[21]. The Price of Anarchy for this commuter is therefore 1.5, i.e. the commuter experiences a 50% longer journey time because of selfish behaviour. However, now suppose that if the roads where empty, it would still have taken 10 minutes for the commuter to travel from home to work. This is time that the commuter could not have avoided and so cannot be altered by a change in routing behaviour, i.e. from UE to SO. Under UE the commuter therefore incurred 10 minutes of fixed travel time and 20 minutes of delays due to congestion. Under SO the commuter would still have incurred 10 minutes of fixed travel time but only 10 minutes of delays to due to congestion. Price of Anarchy Delays for this commuter is therefore 2. The differing insights provided by the Price of Anarchy and Price of Anarchy Delays for this example are, respectively, that the commuter experiences 50% more journey time due to selfish routing behaviour but 100% more delay.

The focus of Price of Anarchy Delays on the delays component of travel cost is particularly useful because, without resorting to significant infrastructure investments; e.g. by constructing new roads, the delay component is the principle part of travel time that a network manager can influence through policy interventions. Travel delays are used as a standard measure of network performance by public authorities. For example, in their 2012 Urban Mobility Report the Texas Transportation Institute used travel delay, aggregated across all road users and over a year, as a key performance indicator (Schrank et al., 2012). These arguments show how Price of Anarchy Delays could be of practical use to network managers.

7.4 An Upper Bound for Price of Anarchy Delays

Having established a new measure of the inefficiency of selfish routing, this section proves that the Roughgarden (2003) upper bounds for the Price of Anarchy also apply to Price of Anarchy Delays. However, before presenting this proof, an intermediate result is required that proves that Price of Anarchy Delays is invariant to the addition of a constant cost B_r to each route $k \in K_r$, for each OD movement r in a given road traffic network. This is the subject of proposition 7.1.

[21] This assumes that this individual traveller benefits from cooperative behaviour. Across all travellers some will benefit and some will lose out in comparison with the travel time they would have had under UE.

Proposition 7.1: Consider a traffic network G for which Assumption A1 of section 6.2 holds, with cost functions c_i and serving a given demand Q with entries $q_r > 0$. Suppose that route costs $C_k^{rs}(f_k)$ are subject to the following transformation, for some constant costs B_r:

$$\check{C}_k^r(f_k^r) = C_k^r(f_k^r) + B_r \tag{21}$$

Price of Anarchy Delays is invariant under the transformation of route costs shown in equation (21).

Proof: Proof of the proposition comes in two parts. It is first shown that link flows x_i are invariant to such a transformation of route costs. The transformation is then substituted into the definition of Price of Anarchy Delays to show the invariance.

To see that link flows x_i are invariant to such a transformation of route costs, suppose that centroid connectors are appended to each node in a given traffic network G such that each origin node in Q has a separate inbound connector for each destination node in Q. This creates exactly one centroid connector for each OD pair r, that each route $k \in K_r$ includes as part of its sequence of links but that which is also exclusive to that OD movement, i.e. no routes between any other OD movements use that centroid connector. The transformation of route costs can therefore be achieved by adding the additional constant costs B_r for each OD movement r onto the corresponding inbound centroid connector whilst leaving the remaining traffic network unchanged. As for the 'lollipop network' example shown in Figure 7.1, it is clear that link flows, under both UE and SO, are invariant to these additional constant costs.

Now consider the effect of the transformation $\check{C}_k^r(f_k^r) = C_k^r(f_k^r) + B_r$ on TTC for each OD movement r. Denote this cost as TTC_r. From equation (19), for each OD movement r under the route cost transformation, it follows that:

$$\begin{aligned}
\check{TTC}_r &= \check{C}_1^r(0)q_r + \sum_{k=2}^{\kappa_r} f_k^r\left[\check{C}_k^r(0) - \check{C}_1^r(0)\right] + \sum_{k=1}^{\kappa_r} f_k^r\left[\check{C}_k^r(f_k^r) - \check{C}_k^r(0)\right] \\
&= B_r q_r + C_1^r(0)q_r + \sum_{k=2}^{\kappa_r} f_k^r[(C_k^r(0) + B_r) - (C_1^r(0) + B_r)] \\
&\quad + \sum_{k=1}^{\kappa_r} f_k^r[(C_k^r(f_k^r) + B_r) - (C_k^r(0) + B_r)] \\
&= B_r q_r + C_1^r(0)q_r + \sum_{k=2}^{\kappa_r} f_k^r \gamma_k^r + \sum_{k=1}^{\kappa_r} f_k^r[C_k^r(f_k^r) - C_k^r(0)] \\
&= B_r q_r + TTC_r
\end{aligned} \tag{22}$$

This derivation is valid for both UE and SO routing. Therefore, combining equation (22) with definition 7.1 yields:

$$\begin{aligned}\widetilde{\rho_d} &= \frac{\widetilde{TTC}^{UE} - \sum_r \big(B_r + C_1^r(0)\big)q_r}{\widetilde{TTC}^{SO} - \sum_r \big(B_r + C_1^r(0)\big)q_r} \\ &= \frac{\sum_r \big[B_r q_r + TTC_r^{UE} - \big(B_r + C_1^r(0)\big)q_r\big]}{\sum_r \big[B_r q_r + TTC_r^{SO} - \big(B_r + C_1^r(0)\big)q_r\big]} \\ &= \frac{TTC^{UE} - \sum_r C_1^r(0)q_r}{TTC^{SO} - \sum_r C_1^r(0)q_r} \\ &= \rho_d\end{aligned} \tag{23}$$

Price of Anarchy Delays is therefore invariant to a transformation of route costs of the type defined by equation (21). ■

It is important to note that the proof does not place any condition on the values B_r must take; in particular, note that the result still remains true if $B_r < 0$ for one or more OD movements r. This is crucial for the proof of Theorem 7.2. The following result proves that Price of Anarchy Delays is subject to the same upper bounds as those proved by Roughgarden (2003) for the Price of Anarchy for road traffic networks with separable, polynomial form cost functions.

Theorem 7.2: Consider a traffic network G for which Assumption A1 of section 6.2 holds, with cost functions $c_i = a_i + b_i x_i^{\beta_i}$ $(a_i, b_i, \beta_i > 0)$ and serving a given demand Q with entries $q_r > 0$. Suppose that $p = \max_{i \in A} \beta_i$, then the following statement is true:

$$\rho_d \leq \left[1 - p(p+1)^{-(p+1)/p}\right]^{-1}$$

i.e. Price of Anarchy Delays is subject to the upper bounds proved by Roughgarden (2003) for the Price of Anarchy.

Proof: To see this, first note from equation (17) and definition 7.1 that in the special case where $C_1^r(0) = 0$ for all OD movements r, $\rho_d = \rho$. Price of Anarchy Delays is therefore subject to the same bounds as proved for the Price of Anarchy by Roughgarden (2003) in this case. Secondly, note that proposition 7.1 proves that for a general multiple OD traffic network with OD pairs r, Price of Anarchy Delays is invariant to a transformation of route costs $\check{C}_k^r(f_k^r) = C_k^r(f_k^r) + B_r$ for any constants B_r. Proof of the theorem follows by implementing the cost transformation $B_r = -C_1^r(0)\ \forall r$. ■

7.5 A Numerical Example in a Large Network

This section presents how Price of Anarchy Delays varies as travel demand is increased in the single OD network example described in section 6.5.1. Recall that in this example, travel

demand q increases on the OD movement between node 20 and node 3 in the Sioux Falls network. The variation of Price of Anarchy Delays with travel demand in this example is shown in Figure 7.4 alongside values for the Price of Anarchy taken from Figure 6.8.

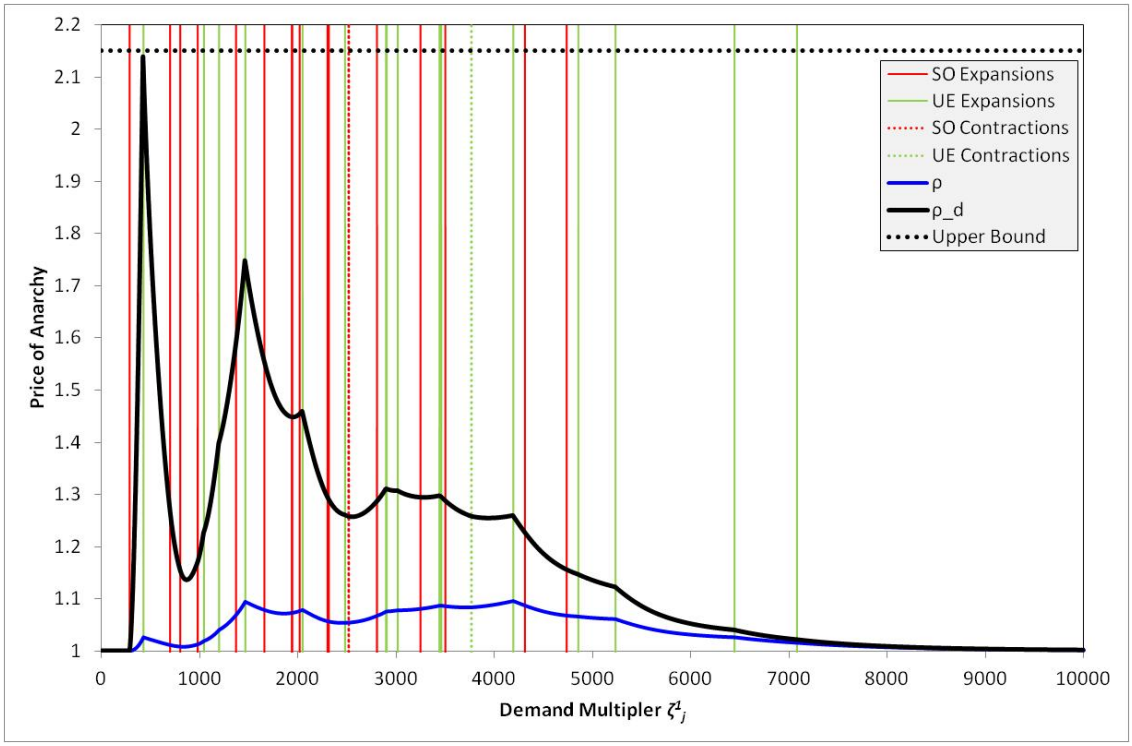

Figure 7.4 - The Variation of the Price of Anarchy ρ and Price of Anarchy Delays ρ_d against the Demand Multiplier ζ_j in Example 1 from section 6.5.1

As with the figures presented in section 6.5, the vertical lines in this figure represent levels of demand corresponding to expansions (solid lines) and contractions (dashed lines) in the minimum cost route sets under UE (green lines) and SO (red lines). Given the similarity of their definitions, it is unsurprising that Price of Anarchy Delays is subject to exactly the same mechanisms that govern the variation of the Price of Anarchy with travel demand that were characterised in chapter 6. However, it can also be seen that values of Price of Anarchy Delays are significantly higher than values of the Price of Anarchy, and also approach the upper bound of Roughgarden (2003), which is identified by the black dashed line.

8 Conclusions and Further Work

8.1 Introduction

This chapter summarises the main findings of the research presented in this thesis and evaluates the extent to which the aims and objectives presented in the introductory chapter have been achieved. Section 8.2 provides a summary of the main findings and original contributions made by this thesis. Section 8.3 then describes the limitations of the research presented and suggests how these could be addressed. Section 8.4 sets out a range of ideas for further research.

8.2 Summary of Main Findings and Original Contributions

The main goal of this thesis was to explore how contributions and methodological approaches from network science could be more appropriately and systematically applied to study how the performance characteristics of road traffic networks vary with respect to the structural properties of supply and demand. In the introductory chapter, it was argued that an understanding of how different structures of network infrastructure and travel demand combine to yield different performance characteristics would be useful for both transport policy and network design because such understanding could help identify how existing road traffic networks can be used more effectively (Mak and Rapoport, 2013), or how structural features, which yield desirable performance characteristics, could be built into the construction of new road traffic networks.

With this goal in mind, the research described within this thesis had two objectives:

1. To develop a systematic methodological approach, incorporating methods from network science, for investigations of how network performance varies with respect to the structural properties of supply and demand in road traffic networks, which is both generally applicable to a wide range of performance phenomena and also provides an intelligible foundation for further research.
2. To apply this methodology to identify and characterise relationships between one or more aspects of supply and demand structure in road traffic networks and one or more measures of network performance.

With reference to the first of these objectives, it was found, in chapter 2 of this thesis, that network science led numerical studies of the effects of network structure on performance have, thus far, used structures of network supply that are not plausible for road traffic

networks and also that the existing experimental approach used by these studies does not provide a coherent picture of how network structure affects performance. It was also highlighted that these studies have not provided explanations for their findings, and that this makes it difficult to generalise their conclusions to other families of networks.

In response to these deficiencies, this thesis has proposed an investigative framework, which comprises experimental and analytical components. The experimental part of the framework proposes a way of designing, conducting and recording the results of numerical experiments that focus on studying spectrums of ensembles of synthetic road traffic networks, which provide plausible representations of real road traffic networks and which also vary with respect to specific aspects of network structure. The analytical component of this framework then uses the results of the numerical component to develop explanations for observed variations.

This thesis has then gone to demonstrate the application of this framework to study how two performance indicators; the average link V/C ratio and the Price of Anarchy, vary with respect to the density of travel demand, and the size, density and connectivity of network supply structure. In so doing, a simple model of road network generation was proposed that is able to generate networks with a wide range of structural properties, including ranges of values that have been observed by empirical studies of real road traffic networks. Focussing specifically on how the Price of Anarchy varies as travel demand is increased in traffic networks, this thesis has then identified and characterised the effects of four mechanisms that govern this variation. These are, specifically, expansions and contractions in the sets of routes, for each OD movement, of minimum (marginal total) cost under UE and SO. In the special case of traffic networks with cost functions of the form $c_i = a_i + b_i x_i^{\beta}$, for which $a_i, b_i, \beta > 0$, this thesis has also proven that there is a systematic relationship between levels of demand under UE and SO at which expansions and contractions occur, and has conjectured that the Price of Anarchy has power law decay for large demand.

By demonstrating the application of this investigative framework, this thesis has demonstrated how this new approach is more comprehensive and systematic than previous studies in network science. Unlike previous numerical studies in network science, this approach takes advantage of the findings of empirical studies of network structure to motivate specific research questions. It also enables the generalizability of findings to be evaluated because performance phenomena are connected to specific aspects of network structure rather than to the names of the models used to generate networks. The methodological approach that has been developed therefore provides an intelligible foundation for further research, which was a key requirement of the first objective. By establishing a series of theoretical results on the

mechanisms that govern the variation of the Price of Anarchy with travel demand in chapter 6 and by providing an explanation for why values of the Price of Anarchy are typically small in numerical experiments in chapter 7, it is also evident that significant progress has been made under the second objective. This thesis has also proposed a new measure of the inefficiency of selfish routing; Price of Anarchy Delays, which complements the insight provided by the Price of Anarchy. Although outside the scope of the original objectives, this is nonetheless a substantive and original contribution.

8.3 Limitations and Suggested Refinements

The research presented in this thesis is subject to the following limitations.

Firstly, whilst the investigative framework presented in chapter 3 represents an advance on the approach utilised by existing studies in network science, it is also noted that the application of its numerical component is limited to the consideration of variation within only one aspect of supply or demand structure and therefore along only one structural dimension. As has been described, the structural characteristics of road traffic networks vary within a huge, multi-dimensional search space. Although the principles described within the investigative framework are sound, it is hypothesised that the approach could be enhanced by being seated within a broader statistical approach to the analysis of networks, similar to that demonstrated by Levinson (2012) and Parthasarathi et al. (2012).

Limitations are also identified with respect to the application of the investigative framework that has been demonstrated within chapters 4, 5 and 6 of this thesis.

It is noted that the model of road network generation included several simplifying assumptions that restricted its ability to reproduce several structural features that have been observed in real road traffic networks. For example, with respect to the distribution mechanism of nodes, the assumption of a single road type, the assumption of a uniform distribution of travel demand and the use of a road traffic model that did not include the effects of junction interactions, which are known to be particularly influential on travel times in urban areas. Simplifying assumptions were also made within the numerical experiments, particularly with respect to the fact that the networks used were considerably smaller than real road traffic networks. Some of these simplifying assumptions were necessary due to a lack of empirical data on several key aspects of supply and demand structure (for example, link capacities, junction types and the distribution of travel demand), and also in order to control the computational burden and, therefore, practicality of the numerical experiments.

It is suggested that the research presented within this thesis would benefit from a model of road network generation that incorporates the simultaneous development of supply and

demand through an evolutionary process, which incorporates feedback effects and also allows for a greater variety of traveller responses than simply route choice. On the supply side, the models of Barthelemy and Flammini (2009) and Courtat et al. (2011) provide attractive starting points that could be used to generate network connectivity structures that replicate more of the features observed in real road traffic networks. A more complex traffic model could also be used, which includes junction interactions and models the dynamic effects of traffic flow.

It should be noted however, that any increase in the complexity of the modelling used within the experimental part of the framework would almost certainly increase the computational burden of numerical experiments. This thesis has demonstrated that, even with experiments that include the simplifications highlighted as limitations above, this burden is not insignificant; recall that the numerical experiments presented in this thesis took approximately twenty-five days of continuous running to be completed. Addressing this computational burden is therefore a key objective for future research and it is suggested that research of the type described in this thesis could benefit from being undertaken in an alternative computing environment that is more suited to large scale calculation work, especially if more complicated models are to be used. Such a development would also enable bigger networks, perhaps on the scale of real road traffic networks, to be tested. In the context of these suggestions, a key point to note is that the investigative framework is sufficiently general that it could accommodate all of these proposed changes.

With reference to the application of the analytical component of the investigative framework in chapter 6, it should be noted that several of the theoretical results of chapter 6 were presented without proof. This statement applies to the results that described the effects on the Price of Anarchy of expansions and contractions in minimum cost route sets under UE, and also the power law decay in the Price of Anarchy for large demand. Although numerical evidence presented towards the end of chapter 6 supports the claims of these conjectures, it is noted that strict proofs are still required. The theory presented in this chapter was also restricted to particular types of demand movements and route transition points. These results could be further generalised by easing these restrictions; specifically, by allowing demand movements to move freely both up and down, allowing adjacent route transition points and also allowing simultaneous expansions and contractions in minimum cost route sets. This is likely to require further numerical work to explore what can be established in these more general settings.

8.4 Opportunities for Further Research

In addition to the suggested refinements made alongside the limitations highlighted in the preceding section, there are also several fertile areas and opportunities for further research, which can build upon the contributions made within this thesis.

Overall, it is the aspiration of the author that the investigative framework proposed by this thesis is an approach that can be used by other researchers in the transportation community to study how structure affects performance in road traffic networks. This thesis focussed on four aspects of network supply and demand structure but there are many other aspects that this thesis did not explore; for example the effects of different road hierarchical structures. It is also noted that the effects of the distribution of travel demand on network performance were not investigated by this thesis and remain largely unstudied across network science studies more generally. Researchers could also choose to focus on different aspects of performance and different performance indicators than those presented within this thesis. One particularly interesting idea to the author is that of exploring how network structure affects the propagation of congestion caused by high traffic volumes or traffic incidents within traffic networks, and also how quickly a traffic network can subsequently recover to a more normal level of service. This would require a road traffic model that includes junction interactions, blocking back effects and time dynamics (Snelder et al., 2012). It is also noted that other important aspects of performance phenomena could be investigated such as environmental indicators.

In order to apply the investigative framework to investigate how different aspects of supply and demand structure affect performance, further empirical studies of a much wider array of structural characteristics of road traffic networks are required. The literature review in chapter 2 identified that existing empirical studies of supply structure in network science have focussed almost exclusively on the topological and geometric properties of road traffic networks in urban areas. They have therefore not yet explored the structural characteristics of other features such as link capacities and junction types. Future empirical studies would also need to address the criticisms made, in section 2.3.1.4, of existing studies from network science with respect to how raw network data is processed and also of incorporating a much larger number of different areas for comparison. Open Street Map data offers a promising source of data for further work in this area; see, for example, Corcoran and Mooney (2013). Further empirical studies are also required on the demand-side, although it is noted that these may currently be somewhat inhibited by the difficulties of data collection.

From a broader perspective, it was noted in the scope of the research set out in chapter 1 that this thesis would not provide a comprehensive review of all past work in geography, spatial

science and urban studies. There is therefore a future research opportunity to explore how ideas and methodological approaches from these research areas relate to the methodological approach presented in this thesis. Finally, it is also highlighted that, with appropriate changes to the modelling approach, the proposed framework could also be applied to study the effects of network structure on performance for other transport modes such as urban rail and passenger transit systems.

9 References

ALBERT, R. & BARABASI, A. L. 2002. Statistical mechanics of complex networks. *Reviews of Modern Physics,* 74**,** 47-97.

ALBERT, R., JEONG, H. & BARABASI, A. L. 1999. Internet - Diameter of the World-Wide Web. *Nature,* 401**,** 130-131.

ALDERSON, D. L. 2008. Catching the "Network Science" Bug: Insight and Opportunity for the Operations Researcher. *Operations Research,* 56**,** 1047-1065.

ARENAS, A., DIAZ-GUILERA, A. & GUIMERA, R. 2001. Communication in networks with hierarchical branching. *Physical Review Letters,* 86**,** 3196-3199.

ARROWSMITH, D., DI BERNARDO, M., SORRENTINO, F. & IEEE 2005. Effects of variations of load distribution on network performance. *2005 Ieee International Symposium on Circuits and Systems.* New York: Ieee.

BAR-GERA, H. 2001. *Transportation network test problems* [Online]. Available: http://www.bgu.ac.il/~bargera/tntp/ [Accessed 01\02\ 2013].

BAR-GERA, H. 2002. Origin-based algorithm for the traffic assignment problem. *Transportation Science,* 36**,** 398-417.

BAR-GERA, H. 2006. Primal method for determining the most likely route flows in large road networks. *Transportation Science,* 40**,** 269-286.

BAR-GERA, H. 2010. Traffic assignment by paired alternative segments. *Transportation Research Part B-Methodological,* 44**,** 1022-1046.

BAR-GERA, H. & BOYCE, D. 1999. *Route flow entropy maximization in origin-based traffic assignment*.

BAR-GERA, H., BOYCE, D. & NIE, Y. 2012. User-equilibrium route flows and the condition of proportionality. *Transportation Research Part B-Methodological,* 46**,** 440-462.

BARABASI, A. L. & ALBERT, R. 1999. Emergence of scaling in random networks. *Science,* 286**,** 509-512.

BARRAT, A., BARTHELEMY, M., PASTOR-SATORRAS, R. & VESPIGNANI, A. 2004. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 3747-3752.

BARRAT, A., BARTHELEMY, M. & VESPIGNANI, A. 2005. The effects of spatial constraints on the evolution of weighted complex networks. *Journal of Statistical Mechanics-Theory and Experiment*.

BARTHELEMY, M. 2011. Spatial networks. *Physics Reports-Review Section of Physics Letters,* 499**,** 1-101.

BARTHELEMY, M. & FLAMMINI, A. 2008. Modeling urban street patterns. *Physical Review Letters,* 100.

BARTHELEMY, M. & FLAMMINI, A. 2009. Co-evolution of Density and Topology in a Simple Model of City Formation. *Networks & Spatial Economics,* 9**,** 401-425.

BATTY, M. 2004. A new theory of space syntax. *Syntax,* 44**,** p.36.

BATTY, M. 2008. The size, scale, and shape of cities. *Science,* 319**,** 769-771.

BAVELAS, A. 1948. A mathematical model for small group structures. *Human Organization,* 7**,** 16-30.

BOCCALETTI, S., LATORA, V., MORENO, Y., CHAVEZ, M. & HWANG, D. U. 2006. Complex networks: Structure and dynamics. *Physics Reports-Review Section of Physics Letters,* 424**,** 175-308.

BONDY, J. A. & MURTY, U. S. R. 2008. *Graph theory,* New York :, Springer.

BONO, F., GUTIERREZ, E. & POLJANSEK, K. 2010. Road traffic: A case study of flow and path-dependency in weighted directed networks. *Physica a-Statistical Mechanics and Its Applications,* 389**,** 5287-5297.

BOYCE, D., RALEVIC-DEKIC, B. & BAR-GERA, H. 2004. Convergence of traffic assignments: How much is enough? *Journal of Transportation Engineering-Asce,* 130**,** 49-55.

BUHL, J., GAUTRAIS, J., REEVES, N., SOLE, R. V., VALVERDE, S., KUNTZ, P. & THERAULAZ, G. 2006. Topological patterns in street networks of self-organized urban settlements. *European Physical Journal B,* 49**,** 513-522.

BUREAU OF PUBLIC ROADS 1964. *Traffic assignment manual for application with a large, high speed computer*, U.S. Dept. of Commerce, Bureau of Public Roads, Office of Planning, Urban Planning Division.

CACERES, N., ROMERO, L. M. & BENITEZ, F. G. 2013. Inferring origin-destination trip matrices from aggregate volumes on groups of links: a case study using volumes inferred from mobile phone data. *Journal of Advanced Transportation,* 47**,** 650-666.

CARDILLO, A., SCELLATO, S., LATORA, V. & PORTA, S. 2006. Structural properties of planar graphs of urban street patterns. *Physical Review E,* 73.

CASTILLO, E., MENÉNDEZ, J. M. & JIMÉNEZ, P. 2008. Trip matrix and path flow reconstruction and estimation based on plate scanning and link observations. *Transportation Research Part B: Methodological,* 42**,** 455-481.

CHAN, S. H. Y., DONNER, R. V. & LAMMER, S. 2011. Urban road networks - spatial networks with universal geometric features? A case study on Germany's largest cities. *European Physical Journal B,* 84**,** 563-577.

CHAU, C. K. & SIM, K. M. 2003. The price of anarchy for non-atomic congestion games with symmetric cost maps and elastic demands. *Operations Research Letters,* 31**,** 327-334.

CHOWELL, G., HYMAN, J. M., EUBANK, S. & CASTILLO-CHAVEZ, C. 2003. Scaling laws for the movement of people between locations in a large city. *Physical Review E,* 68.

CLAUSET, A., SHALIZI, C. R. & NEWMAN, M. E. J. 2009. Power-Law Distributions in Empirical Data. *Siam Review,* 51**,** 661-703.

CORCORAN, P. & MOONEY, P. 2013. Characterising the metric and topological evolution of OpenStreetMap network representations. *European Physical Journal-Special Topics,* 215**,** 109-122.

CORREA, J. R., SCHULZ, A. S. & STIER-MOSES, N. E. 2008. A geometric approach to the price of anarchy in nonatomic congestion games. *Games and Economic Behavior,* 64**,** 457-469.

COSTA, L. D., OLIVEIRA, O. N., TRAVIESO, G., RODRIGUES, F. A., BOAS, P. R. V., ANTIQUEIRA, L., VIANA, M. P. & ROCHA, L. E. C. 2011. Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Advances in Physics,* 60**,** 329-412.

COURTAT, T., GLOAGUEN, C. & DOUADY, S. 2011. Mathematics and morphogenesis of cities: A geometrical approach. *Physical Review E,* 83.

CURRY, L. 1964. The Random Spatial Economy: An Exploration in Settlement Theory. *Annals of the Association of American Geographers,* 54**,** 138-146.

DAFERMOS, S. & NAGURNEY, A. 1984. Sensitivity Analysis for the Asymmetric Network Equilibrium Problem. *Mathematical Programming,* 28**,** 174-184.

DE MONTIS, A., BARTHELEMY, M., CHESSA, A. & VESPIGNANI, A. 2007. The structure of interurban traffic: A weighted network analysis. *Environment and Planning B-Planning & Design,* 34**,** 905-924.

DERRIBLE, S. & KENNEDY, C. 2011. Applications of Graph Theory and Network Science to Transit Network Design. *Transport Reviews,* 31**,** 495-519.

DOYLE, J. C., ALDERSON, D. L., LI, L., LOW, S., ROUGHAN, M., SHALUNOV, S., TANAKA, R. & WILLINGER, W. 2005. The "robust yet fragile" nature of the Internet. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 14497-14502.

DUCRUET, C. & BEAUGUITTE, L. 2014. Spatial Science and Network Science: Review and Outcomes of a Complex Relationship. *Networks and Spatial Economics,* 14**,** 297-316.

DUMRAUF, D. & GAIRING, M. 2006. Price of anarchy for polynomial Wardrop games. *In:* SPIRAKIS, P., MAVRONICOLAS, M. & KONTOGIANNIS, S. (eds.) *Internet and Network Economics, Proceedings.* Berlin: Springer-Verlag Berlin.

ECHENIQUE, P., GOMEZ-GARDENES, J. & MORENO, Y. 2004. Improved routing strategies for Internet traffic delivery. *Physical Review E,* 70.

EDDINGTON, R. 2006. The Eddington Transport Study: Transport's role in sustaining the UK's productivity and competetiveness. *In:* TREASURY, H. (ed.). London.

ENGLERT, M., FRANKE, T. & OLBRICH, L. 2010. Sensitivity of Wardrop Equilibria. *Theory of Computing Systems,* 47**,** 3-14.

ERDÖS, P. & RÉNYI, A. 1959. On random graphs, I. *Publicationes Mathematicae (Debrecen),* 6**,** 290-297.

FALOUTSOS, M., FALOUTSOS, P. & FALOUTSOS, C. 1999. *On power-law relationships of the Internet topology,* New York, Assoc Computing Machinery.

FORTUNATO, S. 2010. Community detection in graphs. *Physics Reports-Review Section of Physics Letters,* 486**,** 75-174.

FREEMAN, L. C. 1977. A set of measures of centrality based on betweenness. *Sociometry,* 40**,** 35-41.

FREEMAN, L. C. 1979. Centrality in social networks: conceptual clarification. *Social Networks,* 1**,** 215-239.

GAO, S., LIU, Y., WANG, Y. L. & MA, X. J. 2013. Discovering Spatial Interaction Communities from Mobile Phone Data. *Transactions in GIS,* 17**,** 463-481.

GARRISON, W. & MARBLE, D. 1964. Factor-analytic study of the connectivity of a transportation network. *Papers of the Regional Science Association,* 12**,** 231-238.

GARRISON, W. L. & MARBLE, D. F. 1962. *The Structure of Transportation Networks.* Northwestern University.

GARRISON, W. L. & MARBLE, D. F. 1965. *A Prolegomenon to the Forecasting of Transportation Development.* Northwestern University.

GAVALDA, A., DUCH, J. & GOMEZ-GARDENES, J. 2012. Reciprocal interactions out of congestion-free adaptive networks. *Physical Review E,* 85.

GOODWIN, P. 2004. The economic costs of road traffic congestion. London, UK: UCL (University College London).

GUDMUNDSSON, A. & MOHAJERI, N. 2013. Entropy and order in urban street networks. *Scientific Reports,* 3.

GUTIERREZ, J. & GARCIA-PALOMARES, J. C. 2007. New spatial patterns of mobility within the metropolitan area of Madrid: Towards more complex and dispersed flow networks. *Journal of Transport Geography,* 15**,** 18-30.

HAN, D. & YANG, H. 2008. The multi-class, multi-criterion traffic equilibrium and the efficiency of congestion pricing. *Transportation Research Part E-Logistics and Transportation Review,* 44**,** 753-773.

HAN, D., YANG, H. & WANG, X. 2010. Efficiency of the plate-number-based traffic rationing in general networks. *Transportation Research Part E: Logistics and Transportation Review,* 46**,** 1095-1110.

HAN, D. R., LO, H. K. & YANG, H. 2008. On the price of anarchy for non-atomic congestion games under asymmetric cost maps and elastic demands. *Computers & Mathematics with Applications,* 56**,** 2737-2743.

HANSON, J. 1989. *Order and structure in urban space: a morphological history of the City of London.* University College London.

HAVLIN, S., KENETT, D. Y., BEN-JACOB, E., BUNDE, A., COHEN, R., HERMANN, H., KANTELHARDT, J. W., KERTESZ, J., KIRKPATRICK, S., KURTHS, J., PORTUGALI, J. & SOLOMON, S. 2012. Challenges in network science: Applications to infrastructures, climate, social systems and economics. *European Physical Journal-Special Topics,* 214**,** 273-293.

HILLIER, B. & HANSON, J. 1984. *The social logic of space,* Cambridge, Cambridge University Press.

HU, M. B., JIANG, R., WU, Y. H., WANG, W. X. & WU, Q. S. 2008. Urban traffic from the perspective of dual graph. *European Physical Journal B,* 63**,** 127-133.

HU, Y. H. & ZHU, D. L. 2009. Empirical analysis of the worldwide maritime transportation network. *Physica a-Statistical Mechanics and Its Applications,* 388**,** 2061-2071.

IQBAL, M. S., CHOUDHURY, C. F., WANG, P. & GONZALEZ, M. C. 2014. Development of origin-destination matrices using mobile phone call data. *Transportation Research Part C-Emerging Technologies,* 40**,** 63-74.

JAHN, O., MOHRING, R. H., SCHULZ, A. S. & STIER-MOSES, N. E. 2005. System-optimal routing of traffic flows with user constraints in networks with congestion. *Operations Research,* 53**,** 600-616.

JIANG, B. 2007. A topological pattern of urban street networks: Universality and peculiarity. *Physica a-Statistical Mechanics and Its Applications,* 384**,** 647-655.

JIANG, B. 2009. Street hierarchies: a minority of streets account for a majority of traffic flow. *International Journal of Geographical Information Science,* 23**,** 1033-1048.

JIANG, B. & CLARAMUNT, C. 2004. Topological analysis of urban street networks. *Environment and Planning B-Planning & Design,* 31**,** 151-162.

JIANG, B. & LIU, C. 2009. Street-based topological representations and analyses for predicting traffic flow in GIS. *International Journal of Geographical Information Science,* 23**,** 1119-1137.

JOSEFSSON, M. & PATRIKSSON, M. 2007. Sensitivity analysis of separable traffic equilibrium equilibria with application to bilevel optimization in network design. *Transportation Research Part B: Methodological,* 41**,** 4-31.

JUNG, W. S., WANG, F. & STANLEY, H. E. 2008. Gravity model in the Korean highway. *EPL (Europhysics Letters),* 81**,** 48005.

KALAPALA, V., SANWALANI, V., CLAUSET, A. & MOORE, C. 2006. Scale invariance in road networks. *Physical Review E,* 73.

KALUZA, P., KOLZSCH, A., GASTNER, M. T. & BLASIUS, B. 2010. The complex network of global cargo ship movements. *Journal of the Royal Society Interface,* 7**,** 1093-1103.

KANG, C. G., MA, X. J., TONG, D. Q. & LIU, Y. 2012. Intra-urban human mobility patterns: An urban morphology perspective. *Physica a-Statistical Mechanics and Its Applications,* 391**,** 1702-1717.

KANSKY, K. J. 1963. *Structure of Transportation Networks: Relationships between Network Geometry and Regional Characteristics,* Chicago, IL, University of Chicago Press.

KARAKOSTAS, G., KIM, T., VIGLAS, A. & XIA, H. 2011. On the degradation of performance for traffic networks with oblivious users. *Transportation Research Part B: Methodological,* 45**,** 364-371.

KOLBL, R. & HELBING, D. 2003. Energy laws in human travel behaviour. *New Journal of Physics,* 5.

KOUTSOUPIAS, E. & PAPADIMITRIOU, C. 1999. Worst-case equilibria. *In:* MEINEL, C. & TISON, S. (eds.) *Stacs'99 - 16th Annual Symposium on Theoretical Aspects of Computer Science.* Berlin: Springer-Verlag Berlin.

KURANT, M. & THIRAN, P. 2006a. Extraction and analysis of traffic and topologies of transportation networks. *Physical Review E,* 74.

KURANT, M. & THIRAN, P. 2006b. Layered complex networks. *Physical Review Letters,* 96.

LAMMER, S., GEHLSEN, B. & HELBING, D. 2006. Scaling laws in the spatial structure of urban road networks. *Physica a-Statistical Mechanics and Its Applications,* 363**,** 89-95.

LEVINSON, D. 2005. The Evolution of Transport Networks. *In:* BUTTON, K. & HENSHER, D. (eds.) *Handbook 6: Transport Strategy, Policy and Institutions.* Oxford: Elsevier.

LEVINSON, D. 2012. Network Structure and City Size. *Plos One,* 7.

LEVINSON, D. & EL-GENEIDY, A. 2009. The minimum circuity frontier and the journey to work. *Regional Science and Urban Economics,* 39**,** 732-738.

LEVINSON, D. & WU, Y. 2005. The rational locator reexamined: Are travel times still stable? *Transportation,* 32**,** 187-202.

LEVINSON, D. & YERRA, B. 2006. Self-organization of surface transportation networks. *Transportation Science,* 40**,** 179-188.

LIN, J. & BAN, Y. 2013. Complex Network Topology of Transportation Systems. *Transport Reviews***,** 1-28.

LIU, T.-L., CHEN, J. & HUANG, H.-J. 2011. Existence and efficiency of oligopoly equilibrium under toll and capacity competition. *Transportation Research Part E: Logistics and Transportation Review,* 47**,** 908-919.

LIU, T.-L., OUYANG, L.-Q. & HUANG, H.-J. 2007. Mixed Travel Behavior in Networks with ATIS and Upper Bound of Efficiency Loss. *Systems Engineering - Theory & Practice,* 27**,** 154-159.

LU, S. & NIE, Y. 2010. Stability of user-equilibrium route flow solutions for the traffic assignment problem. *Transportation Research Part B: Methodological,* 44**,** 609-617.

MAK, V. & RAPOPORT, A. 2013. The price of anarchy in social dilemmas: Traditional research paradigms and new network applications. *Organizational Behavior and Human Decision Processes,* 120**,** 142-153.

MAKSE, H. A., HAVLIN, S. & STANLEY, H. E. 1995. MODELING URBAN-GROWTH PATTERNS. *Nature,* 377**,** 608-612.

MARSHALL, S. 2005. *Streets and patterns,* London :, Spon.

MASUCCI, A. P., SERRAS, J., JOHANSSON, A. & BATTY, M. 2013. Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows. *Physical Review E,* 88.

MASUCCI, A. P., SMITH, D., CROOKS, A. & BATTY, M. 2009. Random planar graphs and the London street network. *European Physical Journal B,* 71**,** 259-271.

MILGRAM, S. 1967. The small world problem. *Psychology Today.*

MOKHTARIAN, P. L. & CHEN, C. 2004. TTB or not TTB, that is the question: a review and analysis of the empirical literature on travel time (and money) budgets. *Transportation Research Part a-Policy and Practice,* 38**,** 643-675.

MORENO, J. L. 1934. *Who Shall Survive?,* Beacon, NY, Beacon House.

NAGURNEY, A. & QIANG, Q. 2007. A network efficiency measure for congested networks. *Epl,* 79.

NEWMAN, M. E. J. 2002. Assortative mixing in networks. *Physical Review Letters,* 89.

NEWMAN, M. E. J. 2003. The structure and function of complex networks. *Siam Review,* 45**,** 167-256.

NEWMAN, M. E. J. 2010. *Networks: an introduction,* Oxford :, Oxford University Press.

NIEMINEN, J. 1974. On the centrality in a graph. *Scandinavian Journal of Psychology,* 15**,** 332-336.

OHIRA, T. & SAWATARI, R. 1998. Phase transition in a computer network traffic model. *Physical Review E,* 58**,** 193-195.

ORTIGOSA, J. & MENENDEZ, M. 2014. Traffic performance on quasi-grid urban structures. *Cities,* 36**,** 18-27.

ORTÚZAR, J. D. D. & WILLUMSEN, L. G. 2001. *Modelling transport,* Chichester, John Wiley.

PACIONE, M. 2005. *Urban geography: a global perspective,* Abingdon, Routledge.

PAPADIMITRIOU, C. 2001. Algorithms, games, and the internet. *Proceedings of the thirty-third annual ACM symposium on Theory of computing.* Hersonissos, Greece: ACM.

PARRY, K. & HAZELTON, M. L. 2012. Estimation of origin–destination matrices from link counts and sporadic routing data. *Transportation Research Part B: Methodological,* 46**,** 175-188.

PARTHASARATHI, P., HOCHMAIR, H. & LEVINSON, D. 2012. Network structure and spatial separation. *Environment and Planning B-Planning & Design,* 39**,** 137-154.

PARTHASARATHI, P. & LEVINSON, D. 2010. Network Structure and Metropolitan Mobility. *Working Papers.* University of Minnesota, Minneapolis.

PATRIKSSON, M. 1994. *The traffic assignment problem: models and methods*, VSP.

PATRIKSSON, M. 2004. Sensitivity analysis of traffic equilibria. *Transportation Science,* 38**,** 258-281.

PATUELLI, R., REGGIANI, A., GORMAN, S. P., NIJKAMP, P. & BADE, F. J. 2007. Network analysis of commuting flows: A comparative static approach to German data. *Networks & Spatial Economics,* 7**,** 315-331.

PERAKIS, G. 2007. The "Price of anarchy" under Nonlinear and asymmetric costs. *Mathematics of Operations Research,* 32**,** 614-628.

PORTA, S., CRUCITTI, P. & LATORA, V. 2006a. The network analysis of urban streets: A dual approach. *Physica a-Statistical Mechanics and Its Applications,* 369**,** 853-866.

PORTA, S., CRUCITTI, P. & LATORA, V. 2006b. The network analysis of urban streets: a primal approach. *Environment and Planning B-Planning & Design,* 33**,** 705-725.

QIU, L. L., YANG, Y. R., ZHANG, Y. & SHENKER, S. 2006. On selfish routing in Internet-like environments. *Ieee-Acm Transactions on Networking,* 14**,** 725-738.

REDNER, S. 1998. How popular is your paper? An empirical study of the citation distribution. *European Physical Journal B,* 4**,** 131-134.

REGGIANI, A., BUCCI, P. & RUSSO, G. 2011. Accessibility and Network Structures in the German Commuting. *Networks & Spatial Economics,* 11**,** 621-641.

RODRIGUE, J. P., COMTOIS, C. & SLACK, B. 2006. *The geography of transport systems,* Abingdon, Oxon, England ; New York :, Routledge.

ROTH, C., KANG, S. M., BATTY, M. & BARTHELEMY, M. 2011. Structure of Urban Movements: Polycentric Activity and Entangled Hierarchical Flows. *Plos One,* 6.

ROUGHGARDEN, T. 2003. The price of anarchy is independent of the network topology. *Journal of Computer and System Sciences,* 67**,** 341-364.

ROUGHGARDEN, T. & TARDOS, E. 2002. How bad is selfish routing? *Journal of the Acm,* 49**,** 236-259.

SABIDUSSI, G. 1966. The centrality index of a graph. *Psychometrika,* 31**,** 581-603.

SCHRANK, D., EISELE, B. & LOMAX, T. 2012. 2012 Annual Urban Mobility Report. Texas A&M University: Texas A&M Transportation Institute.

SEVTSUK, A. & RATTI, C. 2010. Does Urban Mobility Have a Daily Routine? Learning from the Aggregate Data of Mobile Networks. *Journal of Urban Technology,* 17**,** 41-60.

SHEFFI, Y. 1985. *Urban transportation networks: equilibrium analysis with mathematical programming methods,* Englewood Cliffs, N.J. :, Prentice-Hall.

SIENKIEWICZ, J. & HOLYST, J. A. 2005. Statistical analysis of 22 public transport networks in Poland. *Physical Review E,* 72.

SIMINI, F., GONZALEZ, M. C., MARITAN, A. & BARABASI, A. L. 2012. A universal model for mobility and migration patterns. *Nature,* 484**,** 96-100.

SNELDER, M., VAN ZUYLEN, H. J. & IMMERS, L. H. 2012. A framework for robustness analysis of road networks for short term variations in supply. *Transportation Research Part A: Policy and Practice,* 46**,** 828-842.

STOUFFER, S. A. 1940. Intervening Opportunities: A Theory Relating Mobility and Distance. *American Sociological Review,* 5**,** 845-867.

STRANO, E., NICOSIA, V., LATORA, V., PORTA, S. & BARTHELEMY, M. 2012. Elementary processes governing the evolution of road networks. *Scientific Reports,* 2.

STRANO, E., VIANA, M., COSTA, L. D., CARDILLO, A., PORTA, S. & LATORA, V. 2013. Urban street networks, a comparative analysis of ten European cities. *Environment and Planning B-Planning & Design,* 40**,** 1071-1086.

SUN, H. J., ZHANG, H. & WU, J. J. 2012. Correlated scale-free network with community: modeling and transportation dynamics. *Nonlinear Dynamics,* 69**,** 2097-2104.

TADIC, B., RODGERS, G. J. & THURNER, S. 2007. Transport on complex networks: Flow, jamming and optimization. *International Journal of Bifurcation and Chaos,* 17**,** 2363-2385.

TANG, M. & ZHOU, T. 2011. Efficient routing strategies in scale-free networks with limited bandwidth. *Physical Review E,* 84.

THOMSON, R. C. 2004. Bending the axial line: Smoothly continuous road centre-line segments as. *Proceedings 4th International Space Syntax Symposium, London, UK*.

TSEKERIS, T. & GEROLIMINIS, N. 2013. City size, network structure and traffic congestion. *Journal of Urban Economics,* 76**,** 1-14.

VON FERBER, C., HOLOVATCH, T., HOLOVATCH, Y. & PALCHYKOV, V. 2009. Public transport networks: empirical analysis and modeling. *European Physical Journal B,* 68**,** 261-275.

WAGNER, R. 2008. On the metric, topological and functional structures of urban networks. *Physica a-Statistical Mechanics and Its Applications,* 387**,** 2120-2132.

WARDROP, J. G. Some theoretical aspects of road traffic research. Proceedings of the Institute of Civil Engineers, Pt II, 1952. 325-378.

WATTS, D. J. & STROGATZ, S. H. 1998. Collective dynamics of 'small-world' networks. *Nature,* 393**,** 440-442.

WEBTAG 2014. TAG UNIT M3.1 - Highway Assignment Modelling. London: Department for Transport.

WEGENER, M. 2004. Overview of land-use transport models. *Handbook of transport geography and spatial systems,* 5**,** 127-146.

WU, J. J., GAO, Z. Y. & SUN, H. J. 2008a. Optimal traffic networks topology: A complex networks perspective. *Physica a-Statistical Mechanics and Its Applications,* 387**,** 1025-1032.

WU, J. J., GAO, Z. Y. & SUN, H. J. 2008b. Statistical Properties of Individual Choice Behaviors on Urban Traffic Networks. *Journal of Transportation Systems Engineering and Information Technology,* 8**,** 69-74.

WU, J. J., GAO, Z. Y., SUN, H. J. & HUANG, H. J. 2006. Congestion in different topologies of traffic networks. *Europhysics Letters,* 74**,** 560-566.

XIAO, F., YANG, H. & HAN, D. 2007. Competition and efficiency of private toll roads. *Transportation Research Part B: Methodological,* 41**,** 292-308.

XIE, F. & LEVINSON, D. 2009. Modeling the Growth of Transportation Networks: A Comprehensive Review. *Networks & Spatial Economics,* 9**,** 291-307.

YANG, H. & HUANG, H.-J. 2005. *Mathematical and Economic Theory of Road Pricing,* Oxford, Elsevier.

YANG, H., XU, W. & HEYDECKER, B. 2010. Bounding the efficiency of road pricing. *Transportation Research Part E-Logistics and Transportation Review,* 46**,** 90-108.

YASHIMA, K. & SASAKI, A. 2014. Epidemic Process over the Commute Network in a Metropolitan Area. *Plos One,* 9.

YOUN, H., GASTNER, M. T. & JEONG, H. 2008. Price of anarchy in transportation networks: Efficiency and optimality control. *Physical Review Letters,* 101.

ZAHAVI, Y. 1977. *The UMOT Model,* Washington D.C., The World Bank.

ZENG, H. L., GUO, Y. D., ZHU, C. P., MITROVIC, M., TADIC, B. & IEEE 2009. *CONGESTION PATTERNS OF TRAFFIC STUDIED ON NANJING CITY DUAL GRAPH,* New York, Ieee.

ZHANG, Y. Y., WANG, X. S., ZENG, P. & CHEN, X. H. 2011. Centrality Characteristics of Road Network Patterns of Traffic Analysis Zones. *Transportation Research Record*, 16-24.

ZHAO, L., LAI, Y. C., PARK, K. & YE, N. 2005. Onset of traffic congestion in complex networks. *Physical Review E,* 71.

ZHAO, X. M. & GAO, Z. Y. 2007. Topological effects on the performance of transportation networks. *Chinese Physics Letters,* 24, 283-286.

ZHU, Z. H., ZHENG, J. F., GAO, Z. Y. & DU, H. M. 2014. Properties of volume-capacity ratio in congested complex networks. *Physica a-Statistical Mechanics and Its Applications,* 400, 200-206.

10 List of Abbreviations

DfT	Department for Transport
ICN	Intersection Continuity Negotiation model
MST	Minimum Spanning Tree
OBA	Origin-Based Assignment
OD	Origin-Destination
PAS	Paired Alternative Segment
SO	System Optimum
TAPAS	Traffic Assignment by Paired Alternative Segments
TTC	Total Travel Cost
UE	User Equilibrium