UNCOVERING ROAD NETWORK STRUCTURE THROUGH COMPLEX NETWORK ANALYSIS

HONG ZHANG

Ph.D

THE HONG KONG POLYTECHNIC UNIVERSITY

2011

The Hong Kong Polytechnic University
Department of Land Surveying & Geo-Informatics

Uncovering Road Network Structure through Complex Network Analysis

Hong Zhang

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

October, 2010

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_______________________________(Signed)

__________Hong Zhang_________(Name of student)

Abstract

Studies on road networks have received intensive interdisciplinary attention during latest several years for two reasons. The first is that the road network is a common and easily accessible spatial complex system. The second is that a road network has a close relationship with human life and city evolution. Current studies cover almost every aspect of a road network, for instance, road feature extraction from image data, road map generalization and schematization, road selection and optimization, traffic simulation and statistical analysis of road networks. However, little attention has been devoted to structure of a road network. This project attempts to study the properties of the structure of road networks in various aspects, i.e. scale-free, mixing patterns, small-world, hierarchy and structural fractal.

By "scale-free" it is meant that the distribution of the node connectivities in a network follows a power law. In study of the scale-free, the revised Kolmogorov-Smirnov statistics was adopted to replace conventional least square method as it provides more reliable and robust result for a power law distribution. In the study of mixing patterns of road connectivities, the profile of connectivity correlation probability was firstly employed to visualize the results. By "small-world" it is meant that any two nodes in a network can be connected by a relatively short chain. In examination of the small-world structure, two measures, i.e., clustering coefficient and characteristic path length were introduced. In study of hierarchical structure, the ego network analysis rooted in social science was introduced to define the order of each road so as to construct the hierarchical structure of a road network. The ego network was then improved to become weighted ego network by assigning a weight to each link in the network. By "structural fractal" it is meant that the structures of an object at different scales look self-similar. To examine this structure, the Maximum Excluded Mass Burning (MEMB) algorithm was employed. Traffic and other socio-economic data were collected to explore their relationships with these structural properties.

Through these studies, it has been found that (1) road networks are scale-free and the estimated exponents of power law distributions of their road connectivities range from 2.11 to 3.50 with an average of 2.69; (2) shortcuts that make a road network easily accessible are created by mixing pattern in a scale-free road networks, and the profile of connectivity correlation probability is valid to visualize the mixing pattern of road connectivities; (3) road networks are small-world with an average length of the shortest paths (between any two roads) of eight. Results also indicate that a more developed region has a more mature road network to meet its transportation demands; (4) road networks are hierarchical in terms of the opportunities and constrains of each road in flow transmission. Ego network analysis is effective in formation of the hierarchical structure of a road network, and the weighted ego network analysis improves the performance; (5) road networks are structural fractals with fractal dimensions ranged from 2.94 to 4.90. Results indicate that the more complex the structure of the road network is, the more developed the region will be.

The findings shed new light on the organizational principles and dynamical mechanisms of road networks. They reveal that, like other complex adaptive systems, a road network develops in a self-organized way by maximizing its ability of flow transmission in a limited geographical space. In addition, this study help investigate the socio-economic meanings of an urban structure. It provides empirical guides for future planning and policymaking for regional development and transportation design.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Zhilin Li, for his time, knowledge, talent, patience, encouragement and support. Without his continuous guidance and help, it would not be possible to complete the thesis. Prof. Li guided me in finding the joy and overcoming the difficulties of research. I owe a lot to him. I'm also extremely grateful to my ex-supervisor, Prof. Bin Jiang, who accepted me as his first Ph.D student and opened the door of research for me. I was deeply impressed by his hard work, eagerness to learn and meticulous scholarship, which taught me to be a real researcher. Thanks for his time and support. I want to show my gratitude to Prof. Makse at the Benjamin Levich Institute of City College of New York, and Prof. Xinsheng Wang at Hubei University for their helps.

I would like to take this chance to acknowledge the academic and administrative staff members in the Department of Land Surveying and Geo-informatics for their support of this study. I owe special thanks to Prof. Yongqi Chen, Prof. Xiaoli Ding, Prof. Lilian Pun, Prof. Wu Chen and Dr. Bruce King.

Thanks are also due to my colleagues and friends. I'm indebted to Dr. Linguo Yuan, Dr. Shijian Zhao, Dr Xu Yu, Dr. Hernán Rozenland, Mr Qi Zhou, Mr Peng Ti, Ms Hui Sun, Mr Jue Wang and Mr. Xiong Feijiang for their priceless comments and helps on this study.

I especially want to thank Mr Zhanghai Shi for his unwavering encouragement and support during the study. I would also like to thank my family for their encouragement and understanding.

Finally, I am thankful to The Hong Kong Polytechnic University for offering the Ph.D scholarship, the U.S. Census Bureau for providing the TIGER data, the Autotoll of Hong Kong Ltd for providing the traffic flow data. Thanks to Batagelj and Mrvar for the Pajek software.

Table of contents

List of figures

List of tables

Chapter 1 Introduction

A network is a collection of nodes joined together in pairs by links (Newman 2010). Many objects of interest in social, physical, economic and geographical sciences can be treated as networks. For instance, a road network can be represented by an abstract connectivity graph where nodes are named streets and links representing the intersection relationship between the two corresponding named streets. As this study aims to show, thinking of a road network as a kind of abstract graph or network can lead to new and useful insights, especially for understanding the universal organizational principles and evolutionary rules of road networks. This Chapter provides an overview of the role of road networks in urban systems, human life and complex network science. Emphasis is placed on the introduction of complex network analysis to road network study. This approach and its potential applications are only minimally represented in the literature and so are in need of investigation.

1.1 Background and motivation

Roads wind their ways in the city with the complexity of plant roots. To some extent, road networks define urban spatial patterns (Forman and Alexander 1998). The entire road system has long been treated as a magic combination of morphology, structure and function, principally for two reasons. On one hand, like the blood circulation system of the human body, the road system conveys information, energy and various flows to support the normal operation and growth of the city. On the other hand, the process of designing and developing a road network is like writing a melody. Each road with its individual shape, length and name forms the basic musical note. The "notes" of unique positions and roles work together and make a harmonious song. In this sense, the structure and pattern of a road network is both an art and a science.

1.1.1 The role of road networks in urban systems and human life

As one of the five elements composing the image of a city (Lynch 1960), roads combine together to form the skeleton of a city. Roads or paths serve as both traffic channels in large scale spaces and social relationship bridges in small scale spaces. For instance, strangers navigate in the city by path selection and interact with the city by sight-seeing around the road. Citizens extend their physical communication spaces with the help of roads. In fact, roads comprise around 20-30% of land in urban areas

(Chester *et al*., 1996). Few features of urban infrastructure are more prominent than local roads, highways and railways.

Urban morphology is also greatly influenced by its road network (Salingaros 2003). The influences can be observed everywhere. For instance, urban spatial expansion (Roberto *et al*., 2002, Fan *et al*., 2009), urban development (Snellen *et al*., 2002, Dong *et al*., 2007, Hillier 2009) and land use change (Kim and Sohn 2002, Chirapiwat 2005).

The close relationship between a road network and human behaviors has also been recognized by researchers. Fruitful results of research into this topic can be found in many fields, especially in the space syntax community. For instance, it is found that morphology and pattern affect route selection (Dalton 2003), human movement (Hillier and Iida 2005), air pollution (Croxford *et al*., 1996), crime (Hillier and Sahbaz 2008), location of amenities (Rocco and Nes 2005), residential property (Chiaradia *et al*., 2009), housing and real estate development (Matthews and Turnbull 2007, Baran *et al*., 2008).

1.1.2 Road network as a spatial complex system

Apart from the important roles a road network plays in an urban system and human life, a road network itself is a complex system. A complex system is a system composed of interconnected parts that as a whole exhibit one or more properties not obvious from each individual part (Joslyn and Rocha 2000). The often-cited examples of complex systems are ant colonies, the immune system, biological cells, the Internet, the World Wide Web, economic markets and human social networks. A road network is also a complex system due to the fact that roads intersect with each other in various manners and an entire road network demonstrates certain universal properties.

Conventional approaches for complex systems include artificial intelligence, control theory and chaos theory. Recently, network theory has become an essential ingredient in the study of complex systems (Amaral and Ottino 2004). A network is a simplified representation that reduces a system to an abstract structure capturing only the basics of connection patterns and little else (Newman 2010). A road network as a complex system can also be represented by a kind of graph or abstract network (see Figure 1.1).

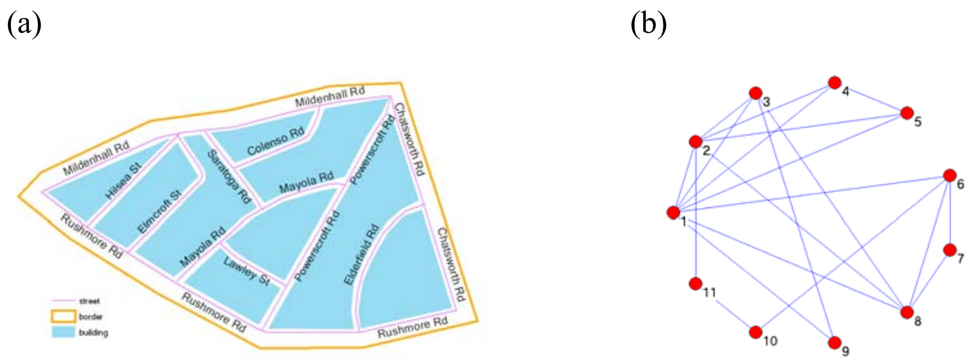


Figure 1.1 A sampled road network and one of its graph representations

Figure 1.1(b) is only one of the graph representations of a road network in Figure 1.1(a). It is noted from Figure 1.1 that unlike other kinds of highly abstract complex systems, a road network is a spatial complex system, because that a road network is closely embedded in the geographic space. In a road network, each road intersection occupies a unique position in the earth surface, and each road segment represents a real-life physical connection. In fact, the geometric information has long been considered in the representation and modeling of road networks. For instance, a weight can be given to each link in Figure 1.1(b) according to the length of the geometric distance between any two roads. Certain morphological measures such as the number of turns, the sum of the deflection angles were also taken into account in road network analyses.

Researchers found that as spatial complex systems, road networks demonstrate certain universal properties. For instance, a road network is neither a tree nor a grid, but a combination of them (Alexander 1965). A road network is a geometric fractal (Batty and Longley 1994). In addition, it is found that the connection patterns of roads in a road network have great influences on the function of the entire road network. In other words, the connection patterns of roads in a road network affect the traffic flow accommodation, information, navigability and searchability of the road network.

Stated briefly, the motivation for this work grew out of the realization of the important role of a road network in the urban system and human life, and recent theoretical advances in complex system.

1.2 Road network study: the state of the art

People principally utilize roads to move around a region. Civil engineers design and construct roads; psychologists are interested in how people select routes; physicists are curious about the organizational principles of road networks; geographers study the underlying structure and pattern of road networks; and cartographers are concerned with multi-scale representation and schematic mapping of the road networks etc. Despite the differences in emphasis and interest, these specialists share a common interest, that is, they wish the road network to be navigable in perception and robust in transportation.

1.2.1 Representations of road networks

A road network can be represented in various ways. As illustrated in Figure 1.2, three main approaches have been developed to road network representation. The approaches include cognition-based representation (e.g. sketch map, axial line and stroke), geometry-based representation (e.g., projected shape file of a road network) and structure-based representation (e.g., primal graph and dual graph).

- repersentation of a road network
  - Cognition-based
    - Cognitive map
    - Sketch map
    - Axial line
    - Characteristic points
    - The Intersection Continuity Negotiation Line
    - Stroke
  - Semantic : Named streets
  - Geometry-based: Projected shapefile
  - Structure-based
    - Primal graph
    - Dual graph

Figure 1.2 A classification scheme of representations of a road network

1.2.2 Modeling and analysis of road networks

A series of studies based on these representations have been carried out on road networks. Current studies cover almost every aspect of a road network. Amongst them, the main research topics for modeling and analyzing road networks are briefly given as follows:

- Road network modeling: 3D road network modeling, artificial intelligence modeling for human route selection, agent-based transportation modeling
- Road network analysis: road feature extraction from image data, road map generalization and schematization, route selection and optimization, and traffic simulation of road networks

In the work of modeling and analysis, techniques from various disciplines are adopted, including mathematics (graph theory, set theory, fractal geometry), computer science and artificial intelligence (genetic algorithm, simulated annealing, machine learning, artificial neural networks, ant intelligence), geographical information science (spatial analysis, spatial relations), cartography (multi-scale representation, simplification, thematic mapping, visualization), physics (information theory, ensemble theory, Monte Carlo method), and network science (degree, path, component and community).

1.2.3 Universalities shared by urban street networks

Investigation into the universalities shared by road networks is of great significance. On one hand, it helps us understand the organizational rules and evolutionary principles of an urban system; on the other hand, it can provide technical support for urban design and transportation planning.

It is however noted that most of current studies on the properties of road networks focused on geometry, morphology and pattern. Little attention has been paid to road network structure. Regarding the approaches involved in exploring the properties shared by road networks, most of them concern spatial analysis, pattern recognition, mathematical and statistic analyses and computer-aided modeling. So far little effort has been devoted to complex network analysis. Studies on urban street networks and small-area road networks have explored certain regularities shared by road networks as follows:

- Self-organized

Self-organization is the process where a structure or pattern appears in a system without a central authority or external element imposing it through planning. It is a fundamental property of open and complex systems. For instance, structural phase transition in physics, liquid crystals in chemistry, bird/fish flocking behavior in biology and herd behavior in human society.

In geographic domain, researchers' interests in self-organization have been triggered by two seminal studies, i.e., Haken's (1983) theory of synergetics and Prigognine's dissipative structures (Nicolis and Prigogine 1977). Road networks are found to be self-organized as they demonstrate some universal structural or morphological properties irrespective of different sizes and morphologies. The typical universalities for road networks include fractal, hierarchical, small-world and scale-free properties.

- Fractality

The term fractal was coined by Mandelbrot (1967) over forty years ago to define the shapes or behaviors that have similar properties at all levels of magnification or across all times. Fractals are ubiquitous in nature and society from snowflakes and trees to phase transitions and critical phenomena (Stanley 1971, Feder 1988). A large body of literature can be found about the fractality of geometric objects (Mandelbrot 1982, Muller 1987, Batty and Longley 1994, West et al., 1999, Shen 2002, Abbott 2006, Mcnally and Mazza 2010), because fractals were first discovered in geometric objects, e.g. coastal lines and snowflakes.

In the geographical domain, investigations into the fractality of the geometric nature (or form) of geographical objects (e.g. cities and road networks) and its relationship with certain socio-economic and environmental issues have been carried out (Batty and Longley 1994, Batty and Xie 1999, Guimera 2003, Lu and Tang 2004, Batty 2005, 2008). The fractal has also become a tool for map generalization (Goodchild 1980, Muller 1987, Butternfield 1989, Longley and Batty 1989, Li and Wu 2005).

- Hierarchical

In the context of road networks, hierarchy means different orders, priorities and advantages for roads. There are various ways to define road hierarchies. For instance, functional class (Kulash 2006), centrality (Jiang and Claramunt 2004), geometric length (Thomson 2006), travel-time (Lämmer et al., 2006), flow dimension and flow capacity (Jiang 2008a), connectivity (Jiang 2008b), and spatial knowledge (Tomko et al., 2008). However, most measures remain at the level of theoretical orientations, and for this reason, real-life data need to be collected to evaluate them.

- Small-world

It is found that any two people in the world can be connected by a short chain of intermediate acquaintances, or "six degrees of separation" proposed by Milgram (1967). This phenomenon in social science was then extended to general networks by Watts and Strogatz (1998). A small-world network is characterized by high local clustering and good global accessibility. Many social and natural networks demonstrate a small-world structure. For instance, the scientific co-authorship network (Newman 2001), the World Wide Web (Barabási et al., 2000), actor networks (Amaral et al., 2000, Strogatz 2001), neural networks (White et al., 1986) and power networks (Watts and Strogatz 1998).

In geography, a large number of transportation systems has been found to be small-world including subway networks (Latora and Marchiori 2002), railway networks (Sen et al., 2003), airline networks (Guimera et al., 2005) and urban street networks (Buhl et al., 2006, Porta et al., 2006a, 2006b, Gao et al., 2007, Jiang 2007, Xu and Sui 2007)

- Scale-free

In network science, a scale-free network refers to the network with a power law degree (or node connectivity) distribution. The scale-free structure has been observed in a host of networks. Urban street networks are also found to be scale-free in terms of geometric length and road connectivity (Jiang 2007).

It is noted that by urban street networks, we mean the road networks embedded in the largest urban land use parcel in a city (Jiang 2007). Despite all the above efforts, there are still some knowledge gaps. For instance, most studies on road networks have focused on urban street networks. This focus is a narrow one and only captures part of the urban system. Other limitations to current road network study are listed as follows:

- the status of the direct connection of each road is discussed, while the connections between neighbouring roads are ignored;
- only single scale is investigated, but the structural property of road networks at different scales is still a challenge and
- many measures remain at the level of theoretical orientations, and so it is still necessary to explore the possible applications of the measures.

1.3 Scope and objectives of this study

This project attempts to investigate the universal organizational principles of road networks. More precisely, this project aims

- to conduct a systematic investigation into the structural properties of road networks in various aspects, i.e. scale-free, mixing patterns, small-world, hierarchy and structural fractal (fractality); and
- To explore the possible association of these properties with social-economic development.

To achieve these objectives, complex network analysis is adopted or developed. In such analysis, the original road network was first converted into an abstract connectivity graph based on the concept of stroke and the intersection relationship between strokes. Then four types of structural analyses have been carried out, i.e., scale-free, mixing pattern, small-world, hierarchical and structural fractal. Socio-economic data including traffic flow, population and housing units are obtained to explore the possible applications of the measures or approaches. A framework for this study is built as follows:

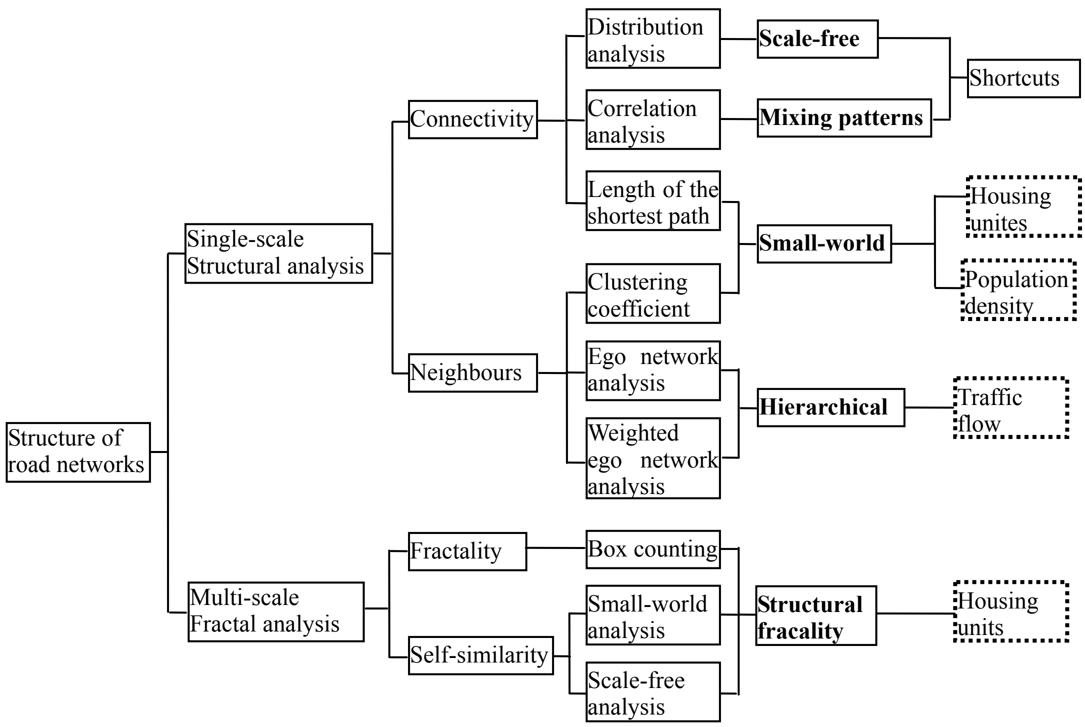


Figure 1.3 The framework for investigating the structure of a road network by complex network analysis

In Figure 1.3 measures and approaches for studying the structure of a road network are represented by solid rectangles. The structural properties detected are highlighted in bold. The theoretical models are evaluated by socio-economic data in dash rectangles.

The contributions of this study are summarized as below:

- It provides a novel way to visualize the mixing pattern of a road network has been introduced, and such property has been first time related to scale-free property by shortcuts;
- It points out the “eight degrees of separation” phenomenon in road network structure;
- It proposes improvements of conventional ego network analysis;
- It provides efficient approaches to form the hierarchical structure of road networks; and finally,
- It discovers that road networks are structural fractals. The self-similar structure may make a road network robust and easily navigable.

It is noted that to provide more reliable results, a large sample of general road networks have been collected. The findings of the five kinds of analyses help bridge current knowledge gaps in road network study and blur the borders of many disciplines.

1.4 Structure of the study

The study comprises eight Chapters. A short preview of the contents of these chapters is provided below:

Chapter 2 is a survey of current representations and analyses of road networks. Regarding the representation, three main types of representations for road networks are identified, which are cognition-based representation, geometry-based representation and structure-based representation. Special attention is given to the concept of stroke and stroke-based connectivity graph. With regard to road network analysis, five main research topics and major techniques involved for each topic are reviewed.

Then four types of structural analyses have been carried out, i.e., scale-free, mixing pattern (Chapter 3), small-world (Chapter 4), hierarchical (Chapter 5 and 6) and structural fractal (Chapter 7). The scale-free and small-world properties are considered because that they have been given the most attention in network science, and they are closely related to the hierarchy, navigability, and robustness of road networks (Kleinberg 2000, Kalapala et al. 2006; Song et al. 2006). Scale-free and mixing pattern analysis focus on the status and preference of the connectivity of each road, which may brings a road network with shortcuts. To examine whether and to what extent the shortcuts exist at both local and global level, the small-world analysis including local clustering and global average shortest path length is carried out.

The scale-free and small-world analyses indicate that the importance of a road can be inferred by combination of connectivity and clustering status among its neighbours. Therefore, ego network analysis, which considers both of the two factors, is introduced to form a hierarchical road network. And this approach is further improved to be weighted ego network analysis by considering preferential attachment and “degree of 1 effect”.

It is found that 50 general road networks of different natural, social and historical origins all show scale-free, small-world and hierarchical structural properties. It is thus natural to make a hypothesis that the road networks are structurally fractals. We first examine the structural fractal property of road networks by box-counting, and then study the self-similarity in the structures of road networks by small-world and scale-free analyses.

Finally, Chapter 8 concludes the study. The main discoveries and contributions of the research are pointed out. Limitations are explained and future work is proposed. The electronic auxiliary material related to this study is attached in the appendix.

Chapter 2 Representation and analysis approaches for road networks

This Chapter provides a brief review of the methodologies for representing and analyzing road networks. Before starting the review, the notion of representation should be made clear. A representation is considered propositional or image-like as a result of how it interacted informationally with the representation agent. Two factors are considered during the representation process, i.e. content and format (Davies 2002). Content is about what you want to represent, and format focuses on how you represent the object(s). The aim of representation is to visualize the relationships and patterns of given data.

2.1 Representation approaches for road networks

A good collection of representation approaches have been proposed in the past to capture the morphology, structure and function of road networks. Generally speaking, there are three main types of representation approaches for road networks, i.e., cognition-based representation, geometry-based representation and structure-based representation.

2.1.1 Cognition-based representation

The cognition-based representation of geographic environments is developed because of the fact that people utilize an internal model generated by what s/he observes to gain useful information from the outside environment. Human cognition or visual perception is therefore the first and paramount consideration in this approach. However, human cognition and perception is a very complex process and it is difficult to interpret the forming process and working mechanism of the internal model.

The first work aimed at dealing with the challenge was the rats experiment carried out by Tolman (1948). In Tolman's experiment, the behaviors of rats, and by analogy human's behaviors, in different environments were observed. Tolman hypothesized that there is a map-like representation, in the human mind which guides human daily movement and he named this the "cognitive map". Tolman's seminal work initiated a lengthy stream of research on understanding how humans perceive and structure information from the environment, and how they use this knowledge to perform several tasks (Carr and Schissler 1969, Horton and Reynolds 1971, Siegel and White 1975, Golledge 1978, Aragones and Arredondo 1985, Banai 1999).

Based on the above efforts, Lynch (1960) firstly and systematically proposed that humans have the aid of five elements in their outside environment to form an image of the city. The five elements are paths, edges, districts, nodes and landmarks. Lynch's theory emphasized the role of visual perception, while Portugali (2000) proposed that both visual and non-visual modes of sensation and information collectively transfer the very large spatial entities to an abstract internal representation in the human mind.

The results of research into the principles and processes of human cognition constructed foundations for different representation models that interpret human cognition of the geographical environment (Golledge 2002). Amongst them, the sketch map has been one of the most frequently used approaches for cpaturing human cognition of urban space. A sketch map is usually drawn by an interviewer to express his/her impression of a geographical environment (see Figure 2.1).

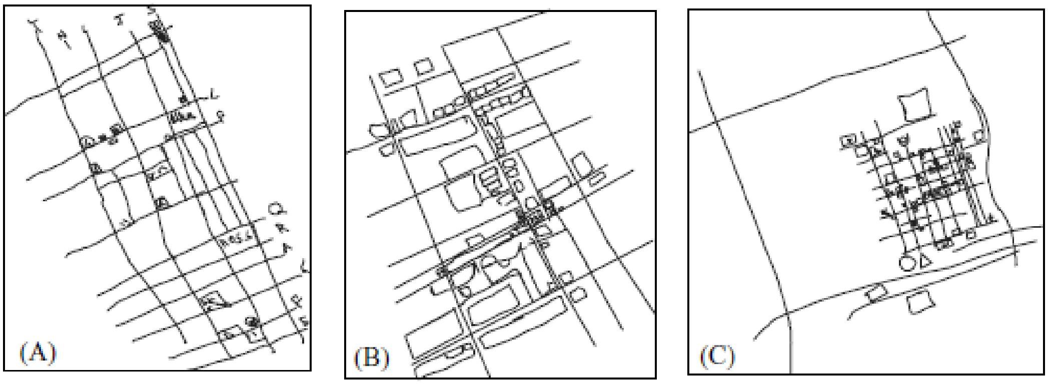


Figure 2.1 Three kinds of sketch maps in central Toronto (Huynh *et al*., 2008)

As illustrated in Figure 2.1, the sketch map intuitively reflects human perception and cognition of geographical environments including road networks. However, a sketch map is usually too simple and rough. To solve this problem, Hillier and his colleagues (Hillier and Hanson 1984, Hillier *et al*., 1983, Hillier 1999) developed an approach of axial map to represent space by a set of axial lines. The term axial line is defined to be the longest line that one can draw from a randomly selected starting point in the open space, and an axial map is thought to be composed of the smallest number of axial lines which pass through the open space and are intersected (see Figure 2.2).

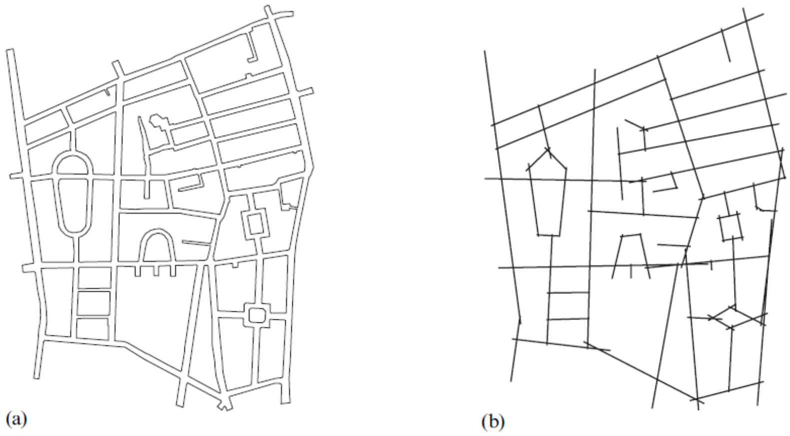


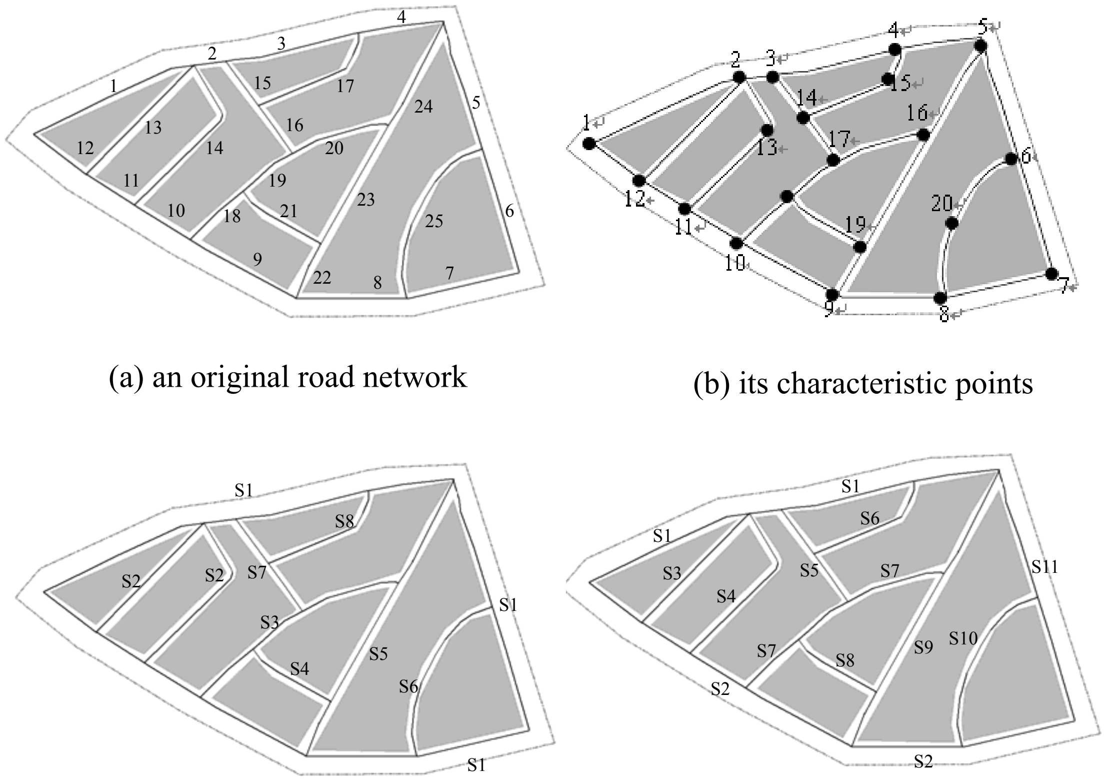
Figure 2.2 Barnsbury road network and its axial map (Hillier and Hanson, 1984)

The formation of an axial map and its quantitative measures build the foundation of space syntax theory, which sheds important light on understanding the art and science of architecture at different scales. However, the production of an axial map greatly depends on the experience of map producers. It is therefore not always reliable.

Apart from the sketch map and the axial map, there are other ways to represent a road network based on human cognition and perception (see Figure 2.3). For instance, characteristic points, which are important points guiding human navigation (Jiang and Claramunt 2002), the Intersection Continuity Negotiation (ICN) line (Portal *et al.*, 2006b) and strokes based on the Gestalt principle of good continuity (Thomson and Richardson 1995, Thomson 2003).

The characteristic point utilizes points that are key in the description of the morphology of a road segment (see Figure 2.3(b)), e.g. the ends of a road and the turning points on a road segment. The ultimate goal of the characteristic point is to sketch the configuration of space. This method effectively depicts human route-searching behavior in the real world and possesses certain advantages over the axial map (Jiang and Claramunt 2002), but it ignores other information such as the length and connectivity of individual road. In addition, the point is a zero-dimensional expression in geometry. Thus is not reasonable to take it to represent the two-dimensional or three-dimensional road network.

These limitations have been addressed by two further approaches. One is the Intersection Continuity Negotiation (ICN) model (Porta *et al.*, 2006b) based on human's preference to go straight at intersections (see Figure 2.3(c)); another is to concatenate segments based on the Gestalt Principle of good continuation of perception, which states that graphic elements that suggest a continued visual line will tend to be grouped together (Bastoky 2010). The concatenated long line is called a stroke (Thomson and Richardson 1995, Thomson 2003).



(a) an original road network

(b) its characteristic points

(c) Its ICN lines

(d) Its strokes

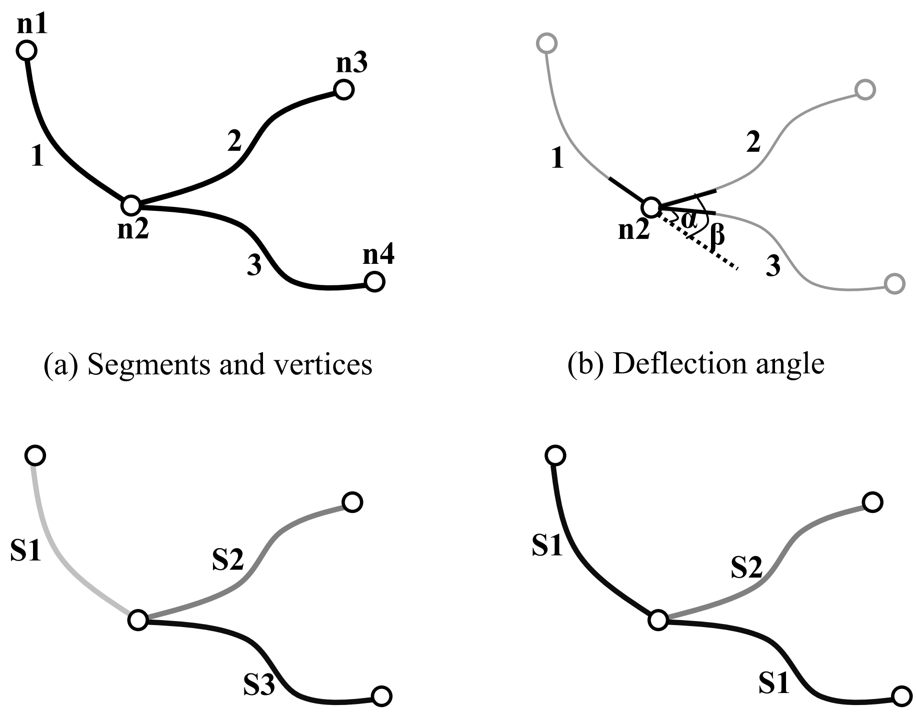
Figure 2.3 A sampled road network and its representations

Both ICN line and strokes indicate that road segments should be concatenated for valid representation. However, the stroke is easily obtained and more intuitive, and has been frequently used in map generalization and structural analysis (Edwardes and Mackaness 2000, Zhang 2005). In the present study, the stroke has therefore been adopted to represent a road network. The following paragraphs explain details of the construction of strokes.

A number of criteria have been developed to extract strokes. Amongst them, deflection angle has been most frequently used. It extracts strokes based on segment negotiation at the given junction. As illustrated in Figure 2.4(a), if junction n2 is chosen as the starting junction, the three original road segments 1, 2 and 3 (black lines) intersecting at junction n2 will negotiate with each other to form strokes. As a road arc is actually composed of a series of very small linear segments with given intervals, for each road arc, its nearest linear segment to junction n2 is chosen and the deflection angles α and β between these linear segments are obtained (see Figure 2.4(b)).

A threshold δ of the deflection angle needs to be defined before the extraction of strokes. In this example, since α is smaller than β, there are only two possible results of strokes in accordance with the value of δ as below:

- if $\alpha > \delta$, three individual strokes S1, S2 and S3 are formed for each road segment 1, 2 and 3 (see Figure 2,4(c)); and
- if $\alpha \le \delta$, two strokes are formed, that is, road segment 1 is connected to road segment 3 to form stroke S1, and road segment 2 itself forms another stroke S2 (see Figure 2.4(d)).



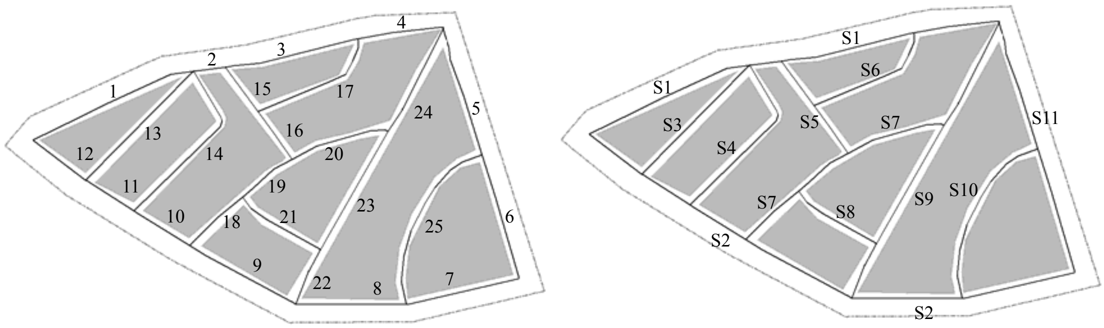
(a) Segments and vertices (b) Deflection angle

(c) Strokes for $\alpha > \delta$ (d) strokes for $\alpha \le \delta$

Figure 2.4 The construction of strokes from individual segments

Regarding the value of δ, previous studies suggested that any value between degree 30 and degree 75 is reliable (Jiang *et al.*, 2008). In this study, the value of δ

is set as degree 60. To better illustrate stroke extraction, Figure 2.5 shows how the strokes (Figure 2.5(b)) were obtained from segments (Figure 2.5(a)) for a sampled road network. In this example, the original 25 road segments were concatenated into 11 strokes.



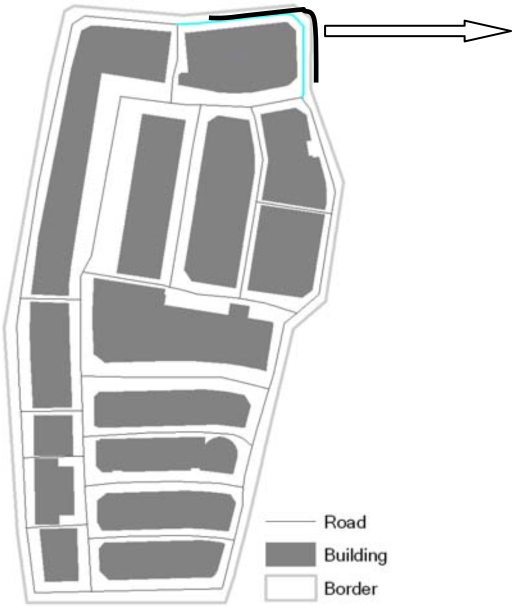
(a) Segments in a notional road network (b) Strokes extracted from (a)

Figure 2.5 A sampled road network and its strokes

Put briefly, the cognition-based representation has been an important tool to represent a road network based on human perception and cognition. However, it still has certain limitations. For instance, construction of a cognitive map is labour-intensive and subjective; the derivation of an axial map is time consuming, especially for a large road network. More importantly, some information is ignored in cognition-based approach, for instance, the location, shape and connection properties. It is therefore necessary to develop new representation approaches that take into account this missing information.

2.1.2 Geometry-based representation

Geometry-based representation focuses on geometric aspects of geographic objects, e.g., shape, size, direction and position (Li and Huang 2002). Compared with cognition-based representation, geometry-based representation has two obvious advantages. On one hand, precise geometric or geographic coordination is significantly considered; on the other hand, distance or proximity is a major concern. Geometrical representations therefore capture the embedding of a road network in geographical space. In fact, in the geometry-based representation, a road is typically represented by a polyline or a collection of polylines, where such polyline(s) capture(s) the centreline of part of a road (see Figure 2.6).



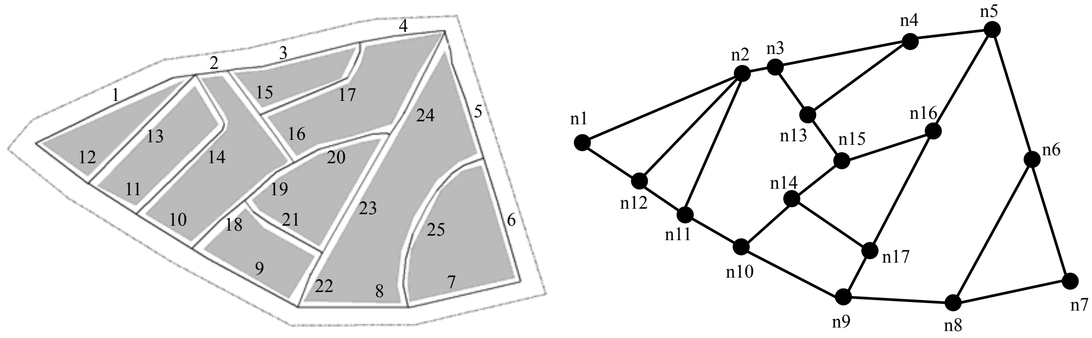
FID	0
Shape	Polyline
Length	161.231116 meters
Vertex coordinate	(**814839.6, 829071.1**) (814855.5, 829071.4) (814926.6, 829079.4) (814931.3, 829077.9) (814939.6, 829069.1) (814939.3, 829016.2) (**814938.9, 829012.4**)

Figure 2.6 A road in a housing estate and its geometric representation

As illustrated in Figure 2.6, geometric representation of road networks provides information on shape, coordination and thus morphology of each road segment. The highlighted coordinates in bold are the starting point and ending point of the road segment. The main disadvantages of geometry-based representation are that it requires a large space to store the information and can not intuitively reflect the structure of road networks.

2.1.3 Structure-based representation

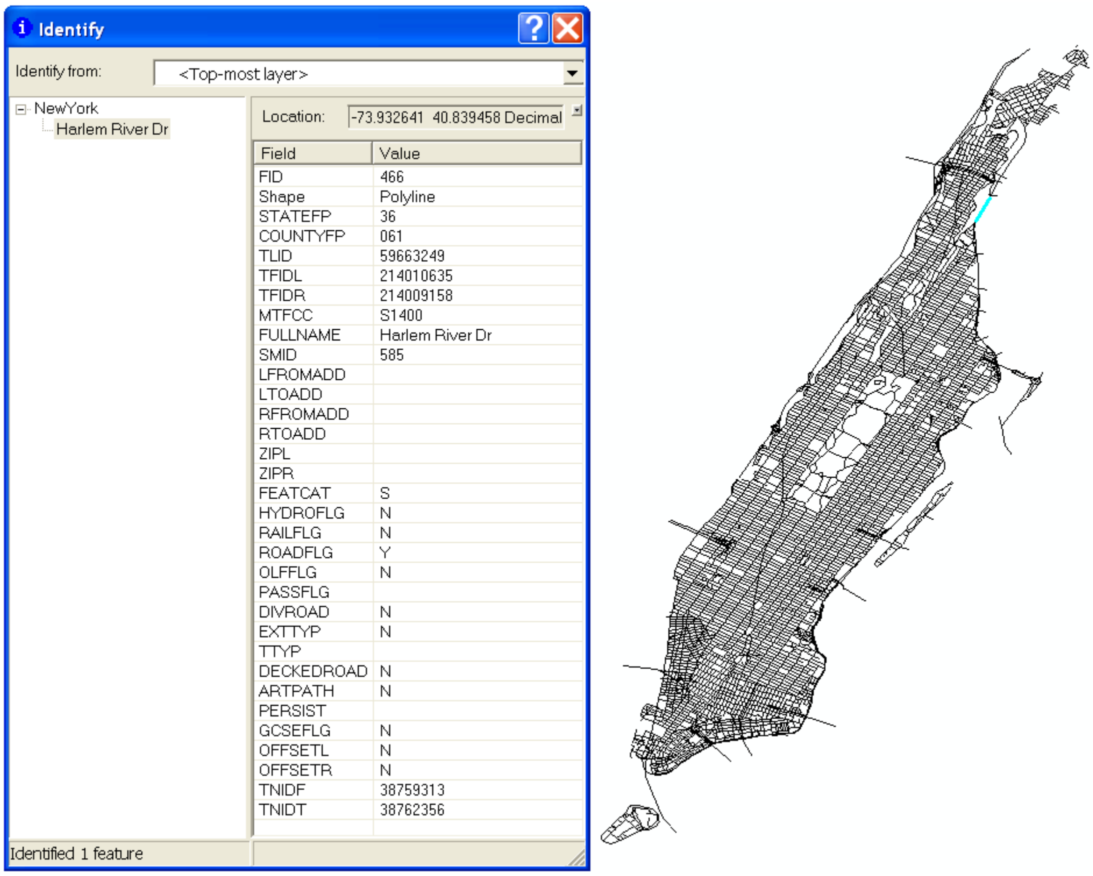
In the structure-based representation, a road network is represented by a graph with nodes, links and weights. In some cases, direction is also considered. Generally speaking, two kinds of structure-based representations have been developed, i.e. primal representation and dual representation. In the primal representation, road intersections are turned into nodes and road segments into edges. Such a representation is intuitive and has a strong connection to the geographic dimension (Porta *et al.*, 2006a). Figure 2.7 gives an example of the primal graph of a sampled road network.



(a) a sampled road network (b) its primal graph

Figure 2.7 A sampled road network and its primal graph

A large number of real-life road networks are actually stored in such a primal form. For example, the TIGER (Topologically Integrated Geographic Encoding and Referencing) dataset of U.S. Census Bureau is built in accordance with the rule that road centrelines are embedded between nodes (see Figure 2.8).


Identify
Identify from: <Top-most layer>
NewYork
Harlem River Dr
Location: -73.932641 40.839458 Decimal

Field	Value
FID	466
Shape	Polyline
STATEFP	36
COUNTYFP	061
TLID	59663249
TFIDL	214010635
TFIDR	214009158
MTFCC	S1400
FULLNAME	Harlem River Dr
SMID	585
LFROMADD	
LTOADD	
RFROMADD	
RTOADD	
ZIPL	
ZIPR	
FEATCAT	S
HYDROFLG	N
RAILFLG	N
ROADFLG	Y
OLFFLG	N
PASSFLG	
DIVROAD	N
EXTTYP	N
TTYP	
DECKEDROAD	N
ARTPATH	N
PERSIST	
GCSEFLG	N
OFFSETL	N
OFFSETR	N
TNIDF	38759313
TNIDT	38762356

Identified 1 feature

Figure 2.8 The attributes of a road in TIGER data

As illustrated in Figure 2.8, the three fields named “TLID”, “TNIDF” and “TNIDT” mean “permanent edge ID”, “From node identifier” and “To node identifier” respectively.

In the dual representation of road networks, a node can represent various things, from road segment to named street, from stroke to axial line. Table 2.1 illustrates the common structure-based representations of a sampled road network of London (see Figure 2.9). It is noted that different representations convey different complexity and information of the road network. The advantages and shortcomings of dual representation of road networks have been often remarked (Batty 2004, Jiang and Claramunt 2004, Ratti 2004, Porta *et al*., 2006b).

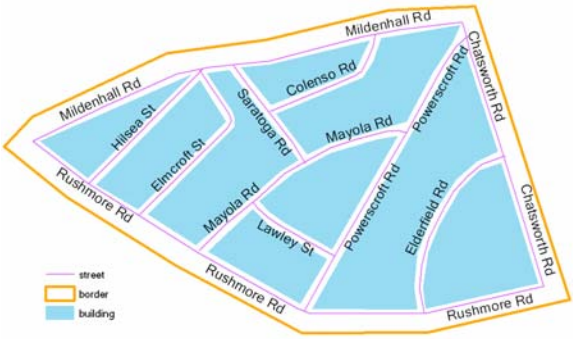


Figure 2.9 A sampled road network in London

Table 2.1 Different representations of a sampled road network

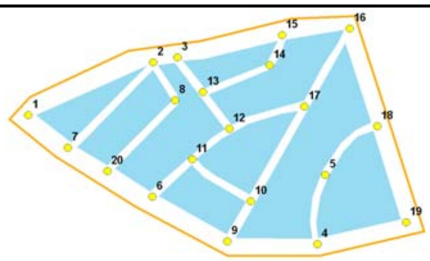
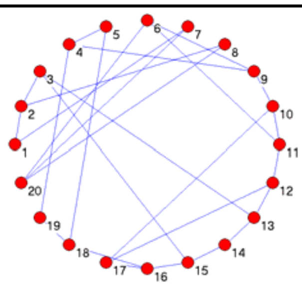
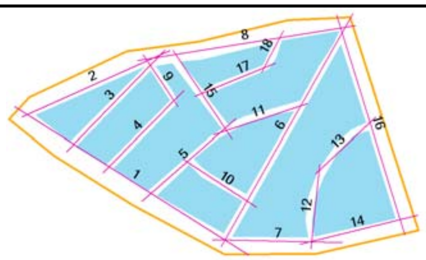
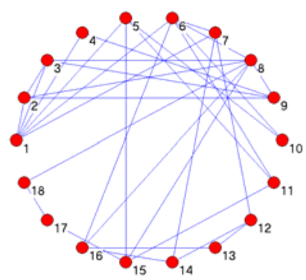
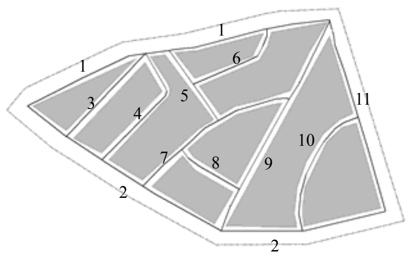
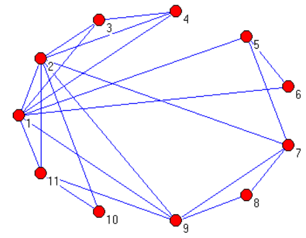
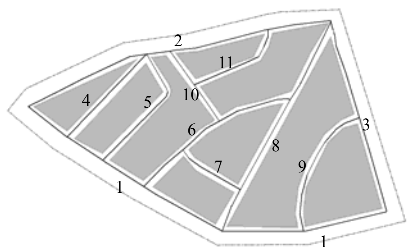
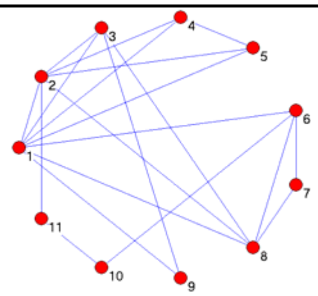
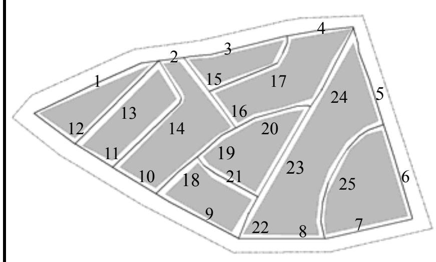
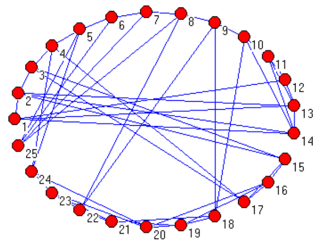
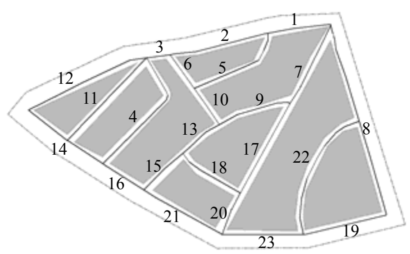
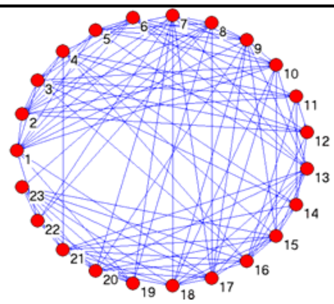
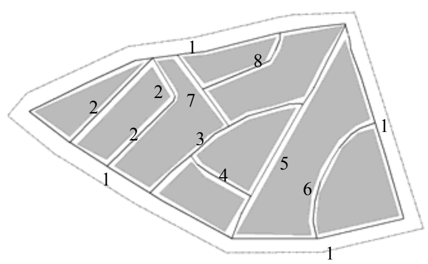
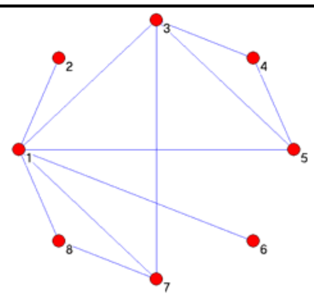
Graph / Object	Generalized elements	Graph representation
characteristic points		
axial line		

Table 2.1 Different representations of a sampled road network (cont'd)

Graph / Object	Generalized elements	Graph representation
strokes (deflection angle: 60 degree)		
named street		
segment		
alternative chain (based on axial line)		
Intersection Continuity Negotiation (ICN)		

Besides the three main representation approaches, there are other ways to represent a road network, e.g. by its semantic information. The named street is a typical example of such an approach, in which spatially adjacent road segments sharing the same name are linked together to become a named street. However, the name information of a road network is usually incomplete, varied with history and affected by culture. The named street is therefore a less than reliable technique.

Stated briefly, a good collection of representation approaches have been developed to capture human perception and cognition of road networks. Morphology, structure and function of road networks have also been considered. Each of the approaches has its advantages and weaknesses. A wide range of studies has been carried out on road networks based on these representations.

2.2 Analysis approaches for road networks

Road networks have received intensive interdisciplinary attention over the past several years. Researchers from various disciplines have quite different concerns. For instance, mathematicians and physicists focused on the statistical regularities that most road networks seem to share (Rosvall *et al*., 2005b, Cardillo *et al*., 2006, Crucitti *et al*., 2006, Porta *et al*., 2008, Boguñá *et al*., 2009), researchers from computer and information sciences are interested in the information of road network (Travençolo and Costa 2008, Wagner 2008), geographers have paid more attention to the relationship between the properties of road networks and human behaviors (Hillier and Hanson 1984, Hillier and Iida 2005, Matthews and Turnbull 2007, Baran *et al*., 2008, Hillier and Sahbaz 2008). Four main research topics for road network analyses which have gained most attention are identified in the following subsections.

2.2.1 Road feature extraction from image data

Automatic road feature extraction from remote sensed imagery has been an active research area in road network study. A number of semiautomatic and automatic methods and algorithms have been developed for road feature extraction over the past 40 years. The development of road feature extraction has experienced changes from low resolution images to middle and high-resolution images, from extraction based on single data source to multisource images fusion extraction, from single scale to multi-scale or multi-resolution, from pixel-based to feature-based or object-based.

Road feature extraction in general consists of four steps (Gruen and Li 1995), i.e., road sharpening, road finding, road tracking and road linking. To extract road features from image data, the geometric, photometric, topological, functional and contextual properties of roads are considered (Vosselman and Knecht 1995). Sometimes, visual information and processing should also be taken into account, especially in computer vision community (Marr 1982).

Regarding the algorithms and approaches for road feature extraction, current literature provides six approaches: template matching, knowledge-based approach, snakes, dynamic programming approach, ridge (/valley) detection, mathematical morphology, Hough transform, multi-scale analysis and image segmentation (Wang *et al.*, 2009). Certain algorithms rooted in computer science, pattern recognition, artificial intelligence and mathematics have also been adopted to recognize roads from images (Anil and Natarajan 2010).

Extraction of road features from images, then, is a very important research topic, especially when human society is at the ever increasing development stage. However, the advance of efficiency and accuracy of road feature extraction still depends on the improvement of image resolution and the development of image data classification technology. Greater research attention should be devoted to this topic.

2.2.2 Road map generalization and schematization

To meet the requirements of automatic road map updating and map service displaying on small-format devices such as PDA, a series of models and algorithms has been developed for road map generalization and schematization (Mackaness 1995, Thomson and Richardson 1995, Li and Choi 2002).

As proposed by Li (2006), there are at least six key operations/processes for the generalization of a network (for example, roads, rivers or other linear features), i.e., classification, selection, elimination, simplification, typification and symbolization. A review of the algorithms for scale-driven generalization of individual line features, line smoothing and line transformation can be found in Li’s book (2006).

Many approaches have been developed to support road map generalization. For instance, Bjøke (2005) has attempted to introduce information theory to road map generalization; Thomson (2006) proposed that strokes based on “good continuation grouping principle” from perception theory could be applied to concatenate road segments into strokes. A further consideration is that, models developed in information science and artificial intelligence have also been powerful tools in road selection at different scales, for instance, self-organizing maps (Jiang and Harrie 2004), machine learning (Mustiere 2005), genetic algorithm (Deng *et al*., 2005), and artificial neural network (García Balboa and Ariza López 2008).

Unlike map generalization, a schematic map is generated to simplify the content and graph and remove the noisy information of the original map. The London tube map is a common example of schematic maps or topological maps. Road map schematization includes processes such as shape simplification of road features, road replacement, displacement and resizing of road map. Based on various criteria, constraints or processing steps have been proposed for the schematization of general road networks (Elroi 1988, Avelar and Muller 2000, Scott and Rodgers 2005, Li and Dong 2010). To date, road map schematization still partly depends on a cartographer's professional experience.

2.2.3 Route selection and optimization

Route selection and optimization have become interdisciplinary research topics as they reflect the relationship between human cognition and the geographical environment. Current research on this topic is variously focused on algorithm design, route selection criteria and human navigation behaviour.

Many algorithms have been developed to find the optimal paths between origins and destinations based on the least-cost. The routing could be either along an existing network or across a continuous landscape (Zhan 1997, Church and Murray 2009). As for routing along a network, most algorithms are designed based on arc-node network model, including Dijkstra's algorithm, Bellman-Ford algorithm, k-shortest path algorithm, A* shortest path algorithm and simplest path algorithm. Compared with routing along a network, routing across a continuous landscape is very complex as it involves a huge solution space. In other words, there are an infinite number of possible routes between origins and destinations on a continuous space (Zhang and Armstrong 2008). The approaches rooted in artificial intelligence and computer science have therefore been utilized to solve this problem (Yu *et al*., 2003, Sharma *et al*., 2006, Li *et al*., 2009).

As noted by researchers, many factors govern route selection, such as cost, distance, number of turns, trip purpose, prior learning and personal experience etc. A detailed list of parameters for route selection can be found by COMSIS Corporation (1995) and Golledge (1995). In addition, a series of models have been proposed to explore human behavior of route selection including logit and probit route choice model, multiple criteria decision making model, random utility model and instrumental rationality model. It is worthy of note that recently consumer behavior theories have also been introduced to help understand human route selection behavior (Morikawa *et al.*, 2005).

The insight obtained from route selection give hints for route optimization, especially for transport design. With different goals, route optimization usually needs to take into account accessibility, proximity and land-use changes. Route optimization therefore is a Multiple Criteria Decision Making (MCDM) problem especially for multimodal transportation. The incorporation of artificial intelligence, calculus of variations, dynamic programming, linear programming, enumeration and genetic algorithm (GA) for route optimization has become increasingly important for route optimization (Kang *et al.*, 2009, Richter 2009).

2.2.4 Traffic simulation of a road network

The growing ubiquity of vehicle traffic in everyday life has generated considerable interest in models of traffic simulation, and a large body of research in the area has appeared in the last 60 years. Traffic simulation tries to solve the problem: given the road network, the behavior model, and initial car states, how does the traffic in the system evolve? Answers to this problem help explore specific phenomena, such as jams and unstable, stop-and-go traffic patterns, and to evaluate network configurations to aid real-world traffic engineering.

Numerous techniques have been developed to simulate traffic on a road network, i.e., traffic simulation at a micro scale. These techniques include agent-based modeling, cellular automata, genetic algorithm and mathematical modeling for continuous flows (Aw and Rascle 2000, Zhang 2002, Willemsen *et al.*, 2006). For all these approaches, the general considerations are the geometric, kinematics and dynamic constrains on the cars. Reviews on traffic simulation are provided by Tango *et al.*, (2007), Kesting *et al.*, (2008) and Kotusevski and Hawick (2009). A factor particularly worthy of mention is that the development of traffic sensors (cameras, road sensor and Global Position System) makes car tracking and realistic recording of street traffic possible, which aid the design of complex rules governing traffic behavior.

Unlike transportation engineers who focus on traffic simulation at a micro scale, physicists tend to model traffic as a continuous flow and often use formulations that are inspired by gas-kinetic flow or hydrodynamic flow equations. For instance, Helbing and his colleagues (2001) have presented a gas-kinetic traffic equation for macroscopic traffic simulation, and Morphet (2010) introduced concepts rooted in thermodynamics to transportation systems. A detailed introduction to macroscopic traffic simulation is provided by Gilkerson *et al*., (2005). This class of simulation methods is known as macroscopic traffic simulation.

A third class of traffic simulation is the mesoscopic method. The mesoscopic simulation is currently at an early stage and tries to keep balance between detail (microscale) and scalability to larger networks (macroscale). In the mesoscopic simulation, traffic is treated as continuum and Boltzmann-type mesoscale equations are adopted to simulate traffic dynamics (Prigogine and Andrews 1960, Wang *et al*., 2005, Willemsen *et al*., 2006).

Traffic simulation of road networks, then, integrates multidisciplinary methodologies to capture vehicle behaviors on a road network. The pattern and information identified from traffic simulation can be further used as a benchmark for measures in road network analysis.

2.3 Measures for characterizing a road network

Perhaps the most fundamental and important question in road network study is how to characterize a road network. The knowledge obtained from this topic will greatly advance the development of all the other research topics discussed in this Chapter. Here measures for characterizing a road network are classified into two types, i.e., geometric measures and topological measures.

2.3.1 Geometric measures

- **Distance**

The distance of a route is defined as the sum of segment lengths, while the distance between two adjacent road segments is half the sum of their lengths.

- **Angle change**

The angle change along a route is the total number of the angular changes along the route. In Figure 2.10, the angle change between S and b is $w(\theta) + w(\pi - \phi)$. The value of angle may be standardized for simplicity, for instance, $w(0) = 0$ and $w(\pi / 2) = 1$.

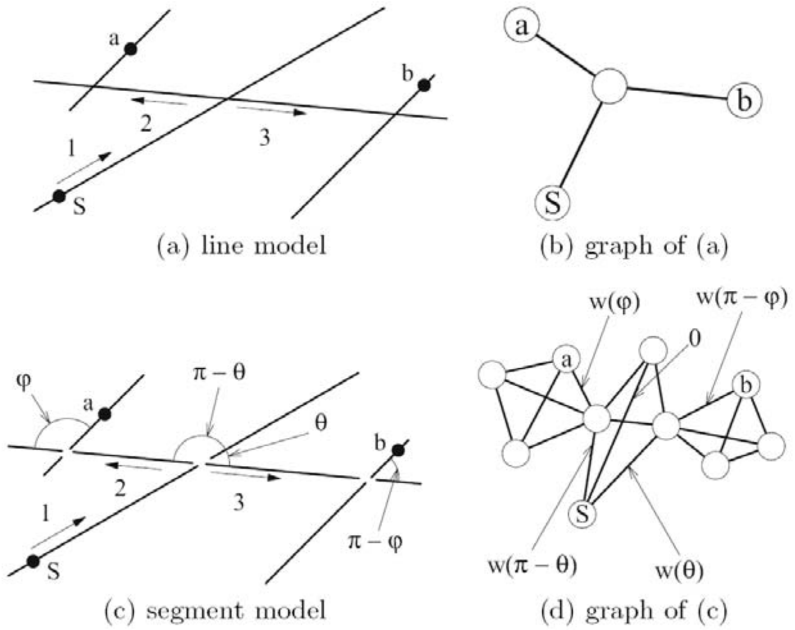


Figure 2.10 A sampled road network and its disaggregated model (Hillier and Iida 2005)

- **Metric reach**

Metric reach is defined as the total length of the streets which are reached from a given starting point in a street system. From the given starting point, one can take all available directions and move freely along streets until a given distance threshold is reached. In this process, each line segment should be counted only once.

- **Road density**

Road density can be expressed in various ways. The conventional definition of road density is the ratio of the total length of all road segments in a given region to the area of the region.

- **Mesh density**

Mesh density is a concept proposed by Chen *et al*., (2009) to quantify the local variations of road density and thus for road map generalization. A mesh in the road network is a naturally closed region that does not contain any other region. The mesh density is the perimeter of a mesh divided by the area of the mesh.

- **Fractal dimension**

Fractal dimension provides a quantified description for the irregularity and complexity of the shapes and patterns of an object (Rodriguez-Iturbe and Rinaldo 2001). The most commonly used approach for calculating fractal dimension is the box-counting , which is defined as below:

$$D = \lim_{r \to 0} \frac{\log N(r)}{\log(1/r)} \tag{2.1}$$

Where, r is the size of the box and $N(r)$ is the number of boxes of size r that intercept at least one point of the object.

Many other models have been built on the basis of distance, shape and density. For instance, the distance decay model, the inverse distance weighting (IDW) model, gravity model and logistic regression model. In addition, conventional spatial analysis including buffer analysis, Voronoi diagram and Moran's I contribute significantly to geometric modeling of road networks.

Despite all the efforts discussed so far, there remain limitations to geometric measures and modeling. For instance, it can not capture the structure and pattern of a road network. Geometric measures also show limitations in describing the changes and dynamics of road networks, because they are constrained by geometric distance and coordinate system. To solve these problems, structural measures have been developed.

2.3.2 Structural measures

- **Number of turns**

The number of turns refers to the number of changes of direction that have to be taken on the route from road A to road B. For instance, the fewest turn from S to b is only one in Figure 2.10.

- **Connectivity**

Connectivity (k) of a road is the total number of the other roads that are intersected with the road.

- **Control value (Ctrl)**

Control value (ctrl) measures the degree of control from a road to its neighbouring roads. The control value of road i is determined as follows:

$$ctrl_i = \sum_{j=1}^{k_i} \frac{1}{k_j} \tag{2.2}$$

Where k_i is the total number of roads that are directly connected to road i, and k_j is the connectivity of the *j*th road that is directly linked to road i.

- **Depth**

The depth (d_{ij}) between two roads is the minimum number of steps needed to be taken between the two roads.

- **Mean Depth**

The mean depth (MD_i) of road i is the average number of the steps in the shortest path from road i to all the other roads in the road network. It can be defined as follows:

$$MD_i = \frac{\sum_{j=1}^{n} d_{ij}}{n-1} \tag{2.3}$$

Where, n is the total number of roads in a road network.

- **Integration**

Integration describes the degree to which a node is integrated or segregated from an entire system (global integration) or from the local neighbours within several steps (local integration). The integration of a node is measured by either Relative Asymmetry (RA) or Real Relative Asymmetry (RRA) as follows:

$$RA_i = \frac{2(MD_i - 1)}{n-2} \quad and \quad RRA_i = \frac{RA_i}{D_n} \tag{2.4}$$

Where

$$D_n = \frac{2\left\{n\left(\log_2^{((n+2)/3-1)}\right)\right\}}{(n-1)(n-2)} \tag{2.5}$$

- **Intelligibility**

Intelligibility is the correlation between the local and global parameters.

The above measures are most frequently used by the space syntax community. Recently, the emerging science of networks offers an extensive set of measures for road network analysis. Before introducing the measures, key concepts about network and complex network analysis should be explained. A network is composed by a collection of nodes joined together in pairs by links (Newman 2010). Many objects of interest in the physical, biological, social and geographical can be thought of as networks. For instance, in the friendship network, each person is a node and there is a link between two persons if they are friends. Some real-life networks are very complex partly because of their huge sizes, various link directions and weights and dynamic evolution. Analytical approaches developed for such kind of networks are called complex network analysis, which often lead to new and useful insights. The most frequently used measures in complex network analysis for road networks are centrality, shortest path and clustering coefficient.

- **Centrality**

It is found that in many social networks, there are certain people or organizations that are central. Those people or organizations usually have advantageous positions and thus have better access to information and better opportunities to spread information (Nooy *et al*., 2005). Several measures have been proposed to quantify the centrality of a node. Amongst them, the most popular ones are degree centrality or connectivity, closeness centrality and betweenness centrality (Freeman 1979).

Before introducing the principles of the above three measures, some definitions and notations of **graph theory** should be introduced. In graph theory, a graph G is formed by a set of vertices (or nodes) $V(G)$ and a set of edges (or links) $E(G)$ that connect pairs of vertices (or nodes). Two nodes which are incident with a common link are adjacent, and two distinct adjacent nodes are neighbours. The set of neighbours of a node i in a graph G is denoted by $N_G(i)$ (Bondy and Murty 2008).

The **connectivity** (or **degree**) of a node i in a graph G, denoted by $k_G(i)$, is the number of nodes that are directly connected to node i. In the following sections, we use N_i and k_i in place of $N_G(i)$ and $k_G(i)$ for simplicity. Mathematically, the connectivity of a node i is defined as:

$$k_i = \sum_{j \in N} e_{ij} \tag{2.6}$$

Where,

$$e_{ij} = \begin{cases} 1, \text{ if } i \neq j \text{ and node } i \text{ and node } j \text{ are directly connected by a link} \\ 0, \text{ if } i \neq j \text{ and node } i \text{ and node } j \text{ are not directly connected} \\ 0, \text{ if } i = j \end{cases} \tag{2.7}$$

Another measure of centrality is **closeness centrality**, which measures the mean steps from a node to all the other nodes in the graphas follows:

$$C_i = \frac{n-1}{\sum_{j=1}^{n} d_{ij}} \tag{2.8}$$

Where, n is the total number of nodes in the graph, and d_{ij} is the length of the shortest path between node i and node j.

Degree and closeness centrality measure the reachability of a node in a graph, while **Betweenness** centrality measures the extent to which a node lies on paths between other nodes as follows:

$$B_i = \sum_{st} \frac{n_{st}^{i}}{g_{st}} \tag{2.9}$$

Where g_{st} is the total number of shortest paths from node *s* to node *t* and

$$n_{st}^{i}=\begin{cases}1, & \text{if node } i \text{ lies on the shortest path from node } s \text{ to node } t\\ 0, & \text{else}\end{cases}$$

- **Clustering coefficient**

The clustering coefficient (C) is developed to describe the link probability of the neighbours of a node. If the neighbours of a node are fully connected, the clustering coefficient of the node will be 1. In sharp contrast, if there are none connections between the neighbours of the node, the clustering coefficient for a node will be 0.

Mathematically, the clustering coefficient C_i for a node i in an undirected graph G is defined as the ratio of the existing links between the neighbours of node i to all the possible number of links between the neighbours of node i. In an undirected graph, the link e_{pq} are considered identical to link e_{qp}. Therefore, if a node i has k_i neighbours, the maximum links between the neighbours should be $\frac{k_i(k_i-1)}{2}$, and the clustering coefficient of node i is defined as follows:

$$C_i=\frac{2\left|\left\{e_{pq}\right\}\right|}{k_i\left(k_i-1\right)}\quad\left(p,q\in N_i,e_{pq}\in E\right)\tag{2.10}$$

Where N_i is the set of neighbours of a node i, E is the set of links in graph G. The value of e_{pq} is obtained as:

$$e_{pq}=\begin{cases}1, & \text{if } p\neq q \text{ and node } p \text{ and node } q \text{ are directly connected by a link}\\ 0, & \text{if } p\neq q \text{ and node } p \text{ and node } q \text{ are not directly connected}\\ 0, & \text{if } p=q\end{cases}\tag{2.11}$$

The clustering coefficient for the entire graph G is the average value of the clustering coefficients of all the nodes in the graph as follows:

$$C\left(G\right)=\frac{1}{n}\sum\nolimits_{i=1}^{n}C_i\tag{2.12}$$

Where, n is the total number of nodes in graph G.

- **Characteristic path length**

The characteristic path length (L) is a global measure of the typical separation between any two nodes in the graph. L is defined as the average value of the lengths of the shortest paths between any two nodes in a graph G , i.e.,

$$L = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}}{n(n-1)} \tag{2.13}$$

Where, d_{ij} is the length of the shortest path between node i and node j .

Based on the structural measures, it is found that urban street networks are scale-free and small-world. Despite all these efforts, there are still a number of limitations related to existing road network study. The main disadvantages are presented below:

- many measures are only limited to the direct connection of the road, while the status of the neighbours of the road and their influences on the function of the road network are ignored;
- existing research on road networks is limited to single scale, i.e., the original road map. It is of interest to examine the structural properties of road networks at different length scales; and
- most current measures remain at the level of theoretical orientations, and so it is necessary to explore the possible applications of the measures.

Chapter 3 Mixing pattern and shortcuts in scale-free road networks

In previous chapter, a brief review of road network study has been given. In this chapter, we will start investigating road network structure from connectivity with the aid of complex network analysis.

A scale-free network is defined as the network in which the distribution of the node connectivities follows a power law. Many real-life networks are scale-free. In other words, they are dominated by a relatively small number of hubs, and such networks are found to be predictable, robust to random attack and ruled by fundamental rules. In the context of road networks, investigating into the scale-free property of road networks will give significant implications for developing a robust road system.

This chapter attempts to investigate the scale-free structure and mixing pattern of a road network. The two structural properties of a road network help produce an easily accessible road network with shortcuts. To be specific, three questions will be addressed in this part of study: firstly, how to examine the scale-free property in the structure of a road network? Secondly, how to visualize the mixing pattern in road connectivities? The third question asks: how does mixing pattern in a scale-free road network affect the function of the network?

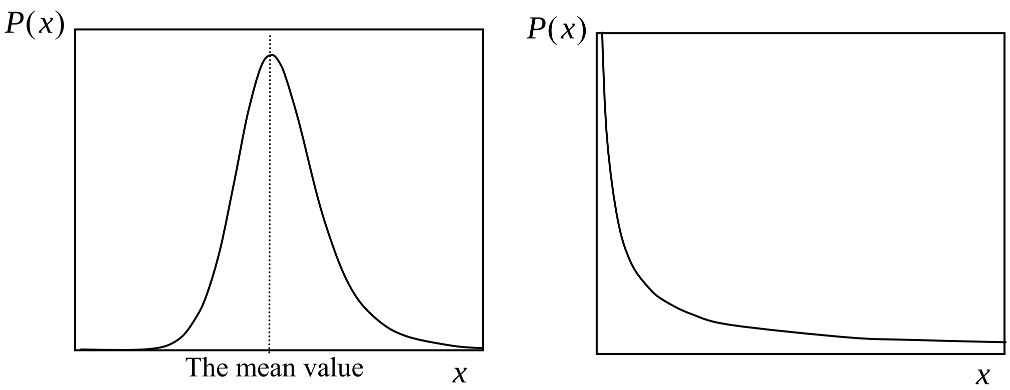
The first question can be answered by introducing a reliable and robust statistics to discern and characterize the power law distribution of road connectivities. Because that a scale-free network is defined by the power law distribution of its node connectivities. In the context of a road network, the connectivity of a road is computed by the total number of roads that are intersected with it.

The second question is about visualization of the results of the mixing pattern analysis. This question can be answered by the application of the profile of connectivity correlation probability to each road, and visualizing the results by an image. To find the solution to the third question, the distance between any two roads in a road network is investigated, because flow transmission is thought of as the most important function of a road network.

3.1 Concepts: scale-free and mixing pattern

By “scale-free” it is meant that the distribution of the node connectivities in a network follows a power law. Recently, power law distribution has drawn increasing attention as many real-life networks are found to follow such a distribution in the node connectivities. The power law distribution has been and still is a major tool for characterizing a scale-free network.

Before introducing the measures of a scale-free network, the basic notion of a power law distribution needs to be introduced. Two kinds of distributions are widely observed in human daily life, which are normal distribution and power law distribution. As illustrated in Figure 3.1(a), a normal distribution is the state in which the variability of investigated quantities is similarly distributed before and after the mean value of the quantities. In the normal distribution, the mean value can be used to describe the investigated quantity. For instance, if height is the investigated quantity, it is reasonable to say that an adult Hong Kong male is around 175cm tall as no one deviates very far from this height. By contrast, a power law distribution demonstrates a long tail pattern with few extreme values of the investigated quantity (see Figure 3.1(b)). In a power law distribution, the mean value is not entirely reliable since it can be greatly influenced even by a few extreme values. For instance, it is not a useful statement to say that the wealth of an adult Hong Kong male is around 10 million dollars, because the reality is that few people are millionaires or billionaires, and most people are actually not worth 10 million dollars.



(a) normal distribution (b) power law distribution

Figure 3.1 A normal distribution and a power law distribution

The scale-free structure is important in the characterization of a network structure. However, the way each node connects with other nodes is still a challenge. In real-life networks, different linking preferences have been observed, e.g. for many networks, a node preferentially links to other nodes that possess similar node connectivities. An example is a social relationship network, where popular people want to make friends with other popular people. Thus "the rich get richer". By contrast, in other networks like cellular networks, a node tends to connect to other dissimilar nodes. In the context of road networks, despite it is empirically observed that roads with different functional classes are interconnected with each other, the mixing pattern of road connectivities appear to date not have been investigated.

3.2 Design of experimental tests of scale-free structure and mixing pattern

Before examining the scale-free structure of road networks and to visualize the mixing pattern of road connectivities, an original road map needs to be first converted into an abstract graph with nodes and links. Since the primal representation of road networks has serious limitations when applied to the description of the structure of a road network, a kind of dual graph is employed to represent a road network in this part of study.

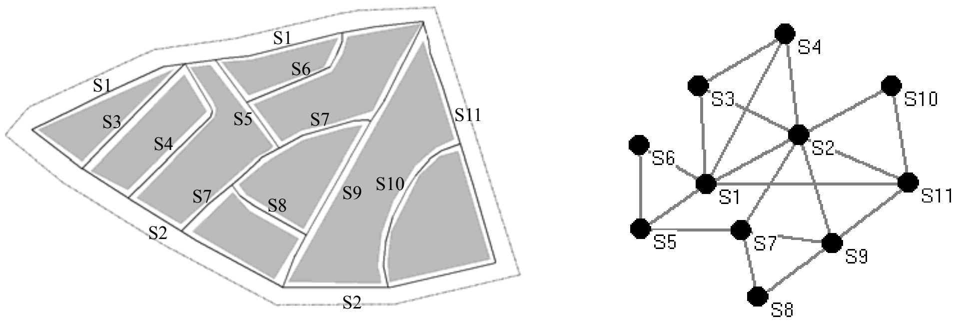
3.2.1 Data sources and data processing

The latest road networks of the 50 most populous counties in the USA are involved in this part of study. The data are downloaded from the website of the U.S. Census Bureau and are recorded in Topologically Integrated Geographic Encoding and Referencing (TIGER) format. In the TIGER format, each road segment has a unique ID and the information conveying "From node ID" and "To node ID". Such information is vital for the following data processing:

- Step 1: generate a segment-based road map and remove isolated road segments;
- Step 2: build strokes from the segment-based road map;
- Step 3: derive an abstract graph from the final map layer; and
- Step 4: calculate the connectivity of each stroke and examine the power law

distribution of road connectivities.

As listed in step 2, the stroke-based graph representation has been adopted to represent an original road map. A stroke is composed by concatenating road segments in accordance with the Gestalt principle of good continuation of perception (see Chapter 2 for a full discussion). In the final stroke-based graph, each node represents an individual stroke, and there is a link between two nodes in the graph if the two corresponding strokes are intersected with each other in the original map (see Figure 3.2).



(a) A stroke-based road network (b) a graph representation of (a)

Figure 3.2 A stroke-based road network and its graph representation

By representing an original road map as an abstract graph, the connectivity of each node (stroke) can be obtained. Connectivity is the number of nodes that are directly connected to the studied node (see Chapter 2 for more details). Taking node S1 in Figure 3.2(b) as an example, its connectivity is 6 as there are six nodes directly connected to it. The six nodes are S2, S3, S4, S5, S6 and S11. In a similar fashion, connectivities of all the other nodes (strokes) can be obtained.

3.2.2 Revised Kolmogorov-Smironov statistics for discerning and characterizing scale-free networks

In various disciplines, power law distribution is a major tool for characterizing a scale-free network. Generally speaking, a quantity x obeys a power law if it is drawn from a probability distribution as follows:

$$P(x) \propto x^{-\alpha} \tag{3.1}$$

Where, α is the exponent or scaling parameter of the power law distribution.

Discerning and characterizing a power law distribution is usually complicated as it has a tail (see Figure 3.1(b)). The tail representing large but rare events is very difficult to be measured especially for the large fluctuations. Drawing a histogram of x may help, but the definition of the interval is subjective to some extent. The most frequently used approach for investigating a power law data is the least-squares fitting. However, this approach has two significant weaknesses as below:

- it can not tell whether the power law fit is a good match to the data or not; and
- it produces unsatisfactory and even inaccurate estimation of the exponent of a power law distribution

To solve these problems, Clauset and his colleagues (2009) have recently presented a statistical framework to improve the discerning and characterizing of a power law distribution. These researchers use a revised Kolmogorov-Smirnov (KS) statistics to test the hypothesis of a power law distribution. Their approach combines maximum-likelihood fitting methods with goodness-of-fit tests, and results of applying their approach to a large number of real-life data clearly indicate that the revised KS statistics is reliable and robust in defining and describing a power law distribution.

The revised KS statistics is adopted in this part of study to discern and characterize scale-free road networks. As mentioned in the previous section, stroke connectivities can be obtained from the stroke-based connectivity graph. The Cumulative Distribution Function (CDF) of stroke connectivities can then be computed according to the following Equation:

$$P\left(K \geq k\right) = mk^{-\alpha} \tag{3.2}$$

Where, k is the connectivity of a stroke, m is a constant, and α is the exponent or scaling parameter of the distribution. In practices, not all values of k obey the power law distribution, the minimum threshold $\widehat{k}_{\min}$ thus needs to be set and the estimated exponent $\widehat{\alpha}$ of the power law distribution is defined as below:

$$\widehat{\alpha} = 1 + n\left[\sum_{i=1}^{n} \ln \frac{k_i}{\widehat{k}_{\min}}\right]^{-1} \tag{3.3}$$

Where, k_i is an observed stroke connectivity, and n is the total number of strokes with a connectivity of $k_i \geq \widehat{k}_{\min}$. The value of $\widehat{k}_{\min}$ is vital for characterizing a power law distribution. Considering the drawbacks of least-squares fitting, the revised KS statistics are employed to estimate the value of $\widehat{k}_{\min}$. In this approach, the value of $\widehat{k}_{\min}$ is set to be the one that minimizes the maximum distances between the CDFs of the studied road network and the fitted model.

Mathematically, the value of $\widehat{\alpha}$ reflects the degree of the variance of the investigated quantity, or heterogeneity of the quantity. In the context of road networks, a larger $\widehat{\alpha}$ means more heterogenous stroke connectivities.

3.2.3 Connectivity correlation probability for visualizing mixing pattern

As introduced in previous section, mixing pattern has drawn increasing attention in network science. Existing literature provides a large number of approaches to capture the mixing patterns in a complex network (Pastor-Satorras *et al.*, 2001, Maslov and Sneppen 2002, Newman 2002, 2003a). However, most of the measures are rough to some extent and are not visualizable especially for large networks. Regarding road networks, the visualization of the mixing pattern of road connectivities appears to date not have been investigated yet.

In this part of study, the approach developed by Maslov and Sneppen (2002) is employed to investigate the mixing pattern of stroke connectivities since it yields visualizable results. In the Maslov-Sneppen approach, Connectivity Correlation Probability (CCP) is used to describe the mixing pattern of node connectivities in a complex network as below:

$$Z\left(k_1,k_2\right)=\frac{P\left(k_1,k_2\right)-\overline{P_r\left(k_1,k_2\right)}}{\sigma_r\left(k_1,k_2\right)} \tag{3.4}$$

Where, $P\left(k_1,k_2\right)$ is the probability that two nodes with connectivity of k_1 and k_2 are directly connected in a network. $P_r\left(k_1,k_2\right)$ is the same probability but in a randomized version of the studied network, and $\sigma_r\left(k_1,k_2\right)$ is the standard deviation of a series of $P_r\left(k_1,k_2\right)$.

To obtain the value of $\sigma_r\left(k_1,k_2\right)$, a large number of random counterparts of each road network are built. Two principles underpinned the construction of a random counterpart are: the total number of nodes and links are the same as the graph representation of each road network and the distribution of node connectivities is the same as that of the graph representation of the investigated road network.

3.3 Scale-free structure of 50 sets of road networks

The revised KS statistics is applied to the 50 sets of road networks to examine their scale-free structure. The results and analysis of the experimental test are listed in this section.

3.3.1 Experimental results

The general statistics for stroke connectivities and the estimated exponents of their power law distributions of stroke connectivities are displayed in Table 3.1.

Table 3.1 The generic statistics for stroke connectivities of in road networks

Rank	County	State	n	m	$\langle k \rangle$	k_{max}	$\hat{k}_{min}$	$\hat{\alpha}$
24	Alameda	California	20006	61616	4.16	137	25 ± 9	3.50 ± 0.05
30	Allegheny	Pennsylvania	30211	81511	3.40	126	5 ± 3	2.70 ± 0.19
20	Bexar	Texas	26325	74314	3.65	190	7 ± 1	2.67 ± 0.04
26	Bronx	New York	2283	9999	6.76	119	17 ± 5	3.19 ± 0.36
17	Broward	Florida	20515	55607	3.42	151	7 ± 1	2.72 ± 0.06
15	Clark	Nevada	30554	77161	3.05	142	5 ± 1	2.86 ± 0.04
37	Contra Costa	California	20451	50507	2.94	135	11 ± 3	3.10 ± 0.23
2	Cook	Illinois	55422	192769	4.96	238	5 ± 6	2.40 ± 0.20
28	Cuyahoga	Ohio	13338	38964	3.84	241	4 ± 1	2.65 ± 0.04
9	Dallas	Texas	30274	99524	4.57	322	5 ± 1	2.48 ± 0.03
46	DuPage	Illinois	12286	33305	3.42	154	8 ± 1	2.81 ± 0.09
48	Erie	New York	10350	29457	3.69	148	4 ± 1	2.53 ± 0.06
39	Fairfax	Virginia	19118	44049	2.61	122	5 ± 1	2.73 ± 0.06
34	Franklin	Ohio	3227	7370	2.57	118	2 ± 0	2.46 ± 0.04
49	Fresno	California	19667	55313	2.65	186	5 ± 1	2.65 ± 0.04
40	Fulton	Georgia	14528	37352	3.14	180	5 ± 1	2.61 ± 0.06
3	Harris	Texas	73251	205408	3.61	328	5 ± 1	2.57 ± 0.02
33	Hennepin	Minnesota	15904	46404	3.84	164	8 ± 2	2.65 ± 0.10
32	Hillsborough	Florida	18277	50659	3.54	213	7 ± 1	2.60 ± 0.05
14	King	Washington	48215	121882	3.06	247	2 ± 4	2.34 ± 0.18
7	Kings	New York	2592	14575	9.24	150	40 ± 11	3.50 ± 0.43
1	Los Angeles	California	101590	293536	3.78	323	6 ± 1	2.46 ± 0.02
4	Maricopa	Arizona	79026	208780	3.28	374	5 ± 1	2.75 ± 0.03
8	Miami-Dade	Florida	28034	85653	4.11	277	5 ± 1	2.58 ± 0.03

(Rank=ranking in terms of population estimation of year 2009, n =total number of strokes (or nodes), m = number of stroke intersections (or links), $\langle k \rangle$ =average connectivity of the stroke, $k_{\max}$ = the maximum connectivity of the stroke of road networks, $\widehat{k}_{\min}$ = the lower bound to the power law behavior, $\widehat{\alpha}$ = the estimated exponent of the power law distribution)

Table 3.1 The generic statistics for stroke connectivities of in road networks (Cont'd)

Rank	County	State	n	m	$\langle k \rangle$	$k_{\max}$	$\widehat{k}_{\min}$	$\hat{\alpha}$
23	Middlesex	Massachusetts	31115	82221	3.28	179	4 ± 0	2.57 ± 0.02
44	Milwaukee	Wisconsin	7329	25998	5.09	153	7 ± 3	2.51 ± 0.12
45	Montgomery	Maryland	12193	31125	3.11	129	9 ± 2	2.94 ± 0.15
27	Nassau	New York	17393	51158	3.88	229	7 ± 2	2.77 ± 0.08
19	New York	New York	1249	6499	8.40	256	8 ± 3	2.41 ± 0.21
31	Oakland	Michigan	21312	55858	3.24	261	4 ± 0	2.70 ± 0.03
5	Orange	California	45006	113085	3.02	127	6 ± 1	2.75 ± 0.04
35	Orange	Florida	19902	50296	3.05	84	8 ± 2	3.10 ± 0.15
29	Palm Beach	Florida	20725	52514	3.06	306	5 ± 2	2.65 ± 0.05
21	Philadelphia	Pennsylvania	8186	33075	6.08	276	2 ± 0	2.11 ± 0.01
41	Pima	Arizona	26169	66329	3.07	178	5 ± 2	2.73 ± 0.09
47	Pinellas	Florida	16428	48076	3.85	214	5 ± 0	2.64 ± 0.03
10	Queens	New York	6465	27707	6.57	222	9 ± 4	2.63 ± 0.15
11	Riverside	California	44736	113275	3.06	112	5 ± 2	2.80 ± 0.09
25	Sacramento	California	21306	56587	3.31	249	5 ± 1	2.64 ± 0.05
38	Salt Lake	Utah	18178	46742	3.14	142	5 ± 0	2.78 ± 0.04
12	San Bernardino	California	52593	145004	3.51	269	7 ± 1	2.73 ± 0.06
6	San Diego	California	51022	130500	3.12	166	4 ± 2	2.53 ± 0.10
16	Santa Clara	California	27753	71752	3.17	190	4 ± 1	2.52 ± 0.03
50	Shelby	Tennessee	19808	53854	3.44	208	4 ± 1	2.60 ± 0.05
36	St.Louis	Missouri	20275	50432	2.97	111	4 ± 1	2.65 ± 0.05
22	Suffolk	New York	31743	87473	3.51	263	10 ± 2	2.74 ± 0.06
18	Tarrant	Texas	29109	86679	3.96	263	6 ± 1	2.59 ± 0.03
42	Travis	Texas	12886	36320	3.64	112	6 ± 1	2.70 ± 0.07
13	Wayne	Michigan	18116	66230	5.31	352	9 ± 3	2.48 ± 0.06
43	Westchester	New York	14808	40432	3.46	191	4 ± 1	2.53 ± 0.05

(Rank=ranking in terms of population estimation of year 2009, n =total number of strokes (or nodes), m = number of stroke intersections (or links), $\langle k \rangle$ =average connectivity of the stroke, $k_{\max}$ = the maximum connectivity of the stroke of road networks, $\widehat{k}_{\min}$ = the lower bound to the power law behavior, $\widehat{\alpha}$ = the estimated exponent of the power law distribution)

The power law distributions of stroke conenctivities of 50 road networks are illustrated in Figure 3.3. In Figure 3.3, the road connectivities and their cumulative distribution probabilities were log-log plotted to provide an intuitive visualization.

In Figure 3.3, circle points represent road connectivities and lines are model fitting results. According to Equation (3.2), the more points the line can fit, the more significant the power law distribution is. Thus the scale-free road network is more reliable. Another important piece of information that emerges is that the steeper the line, the more heterogenous the road connectivities are.

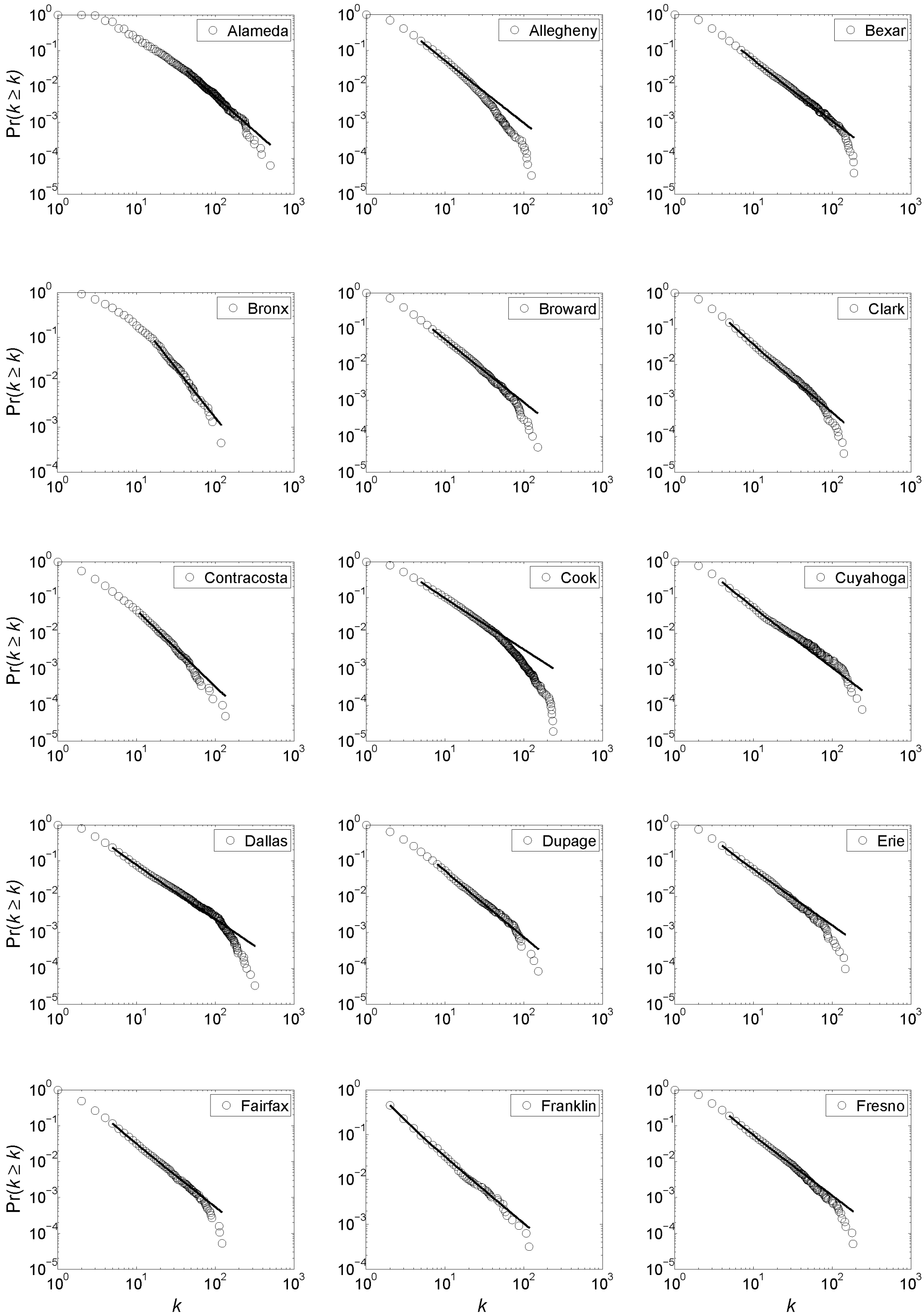


Figure 3.3 Power law distributions of the road connectivities in 50 road networks

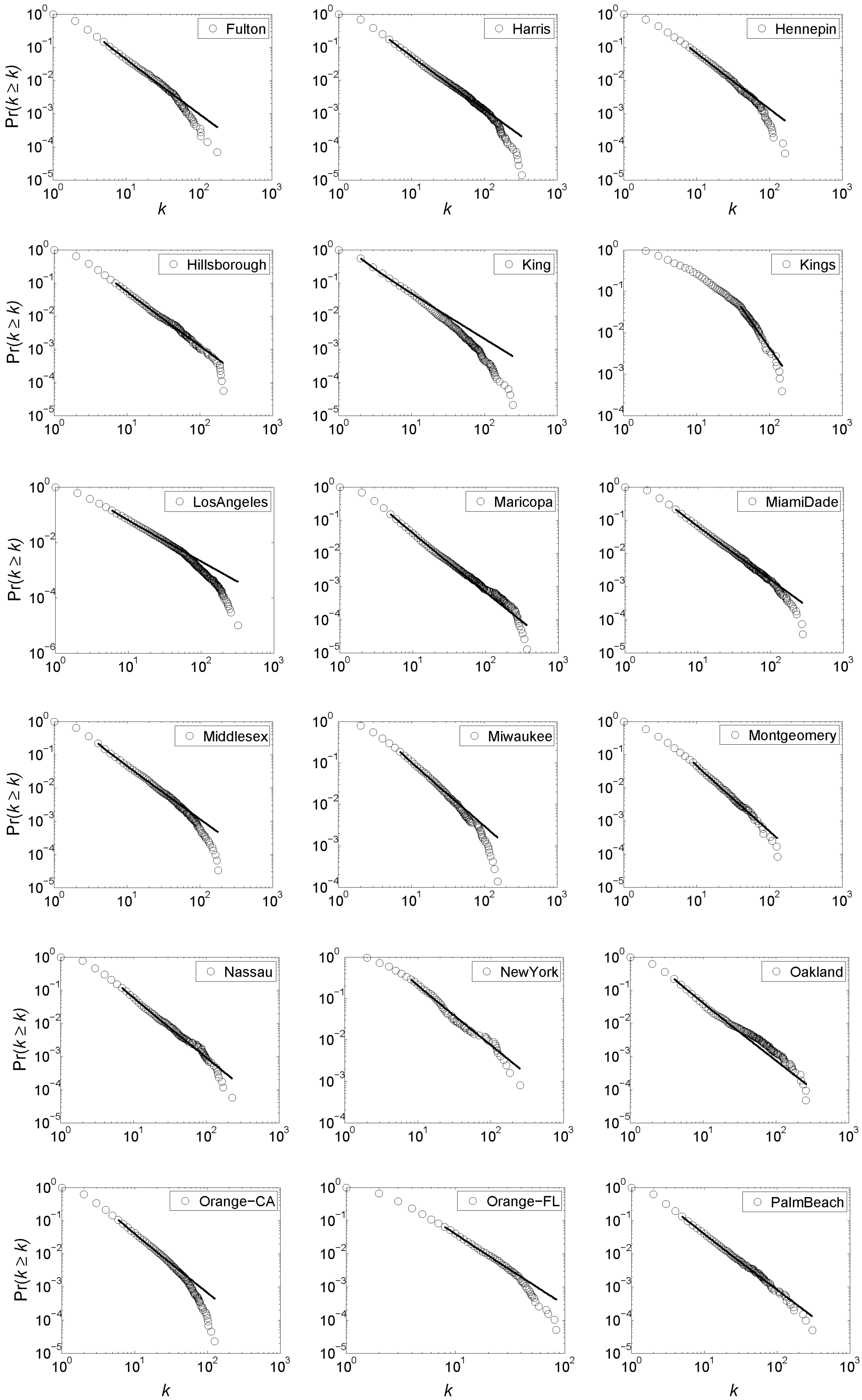


Figure 3.3 Power law distributions of the road connectivities in 50 road networks (cont'd)

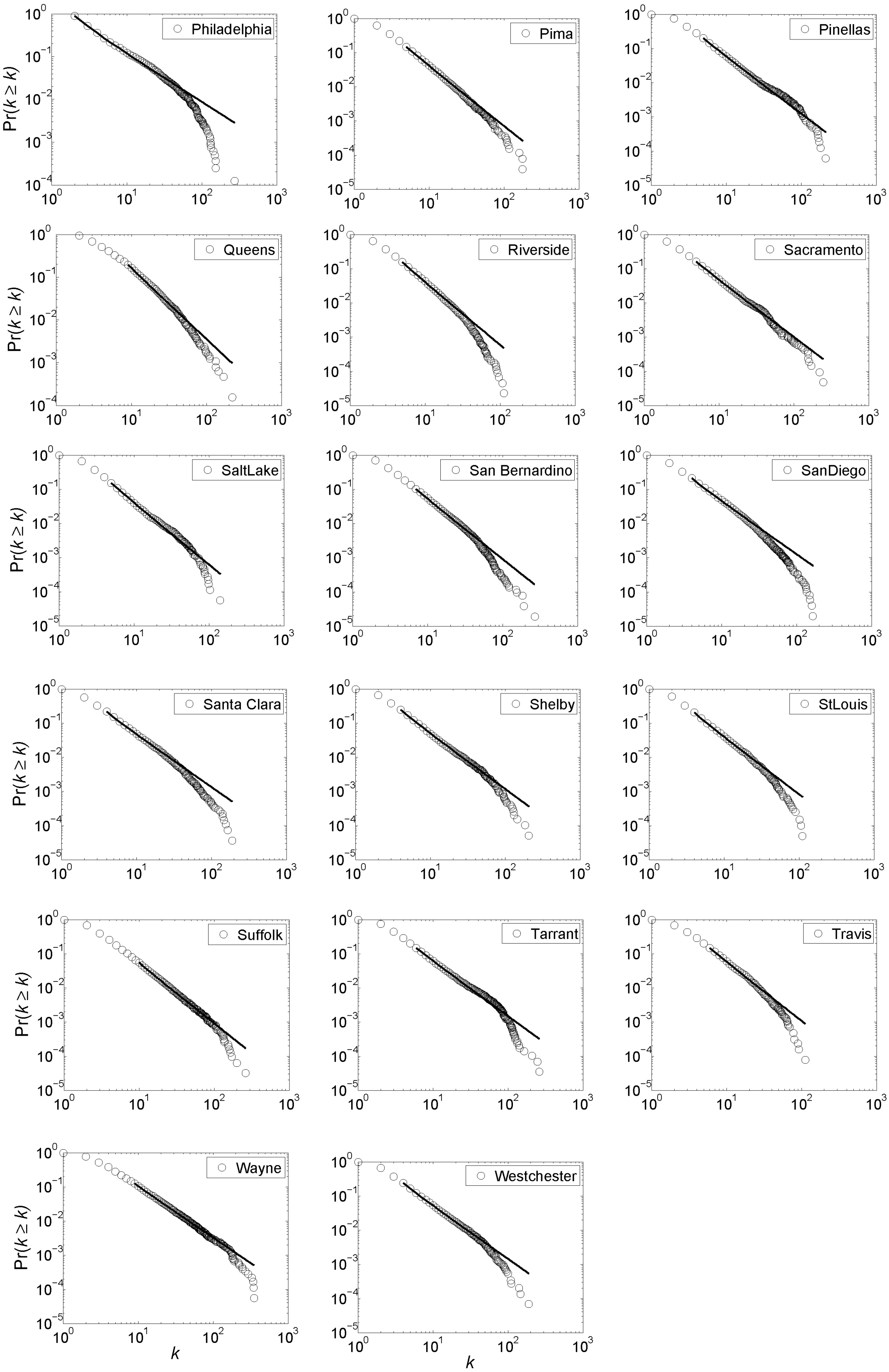


Figure 3.3 Power law distributions of the road connectivities in 50 road networks (cont'd)

3.3.2 Analysis of the results

As illustrated in Figure 3.3 and Table 3.1, the road connectivities of all the 50 road networks follow the power law distributions; that is, road networks are scale-free. It is also found that the values of the estimated exponents of the power law distributions range from 2.11 to 3.5 with an average of 2.69. 90% of the exponents are in the range of $2 < \hat{\alpha} < 3$ (see Figure 3.4).

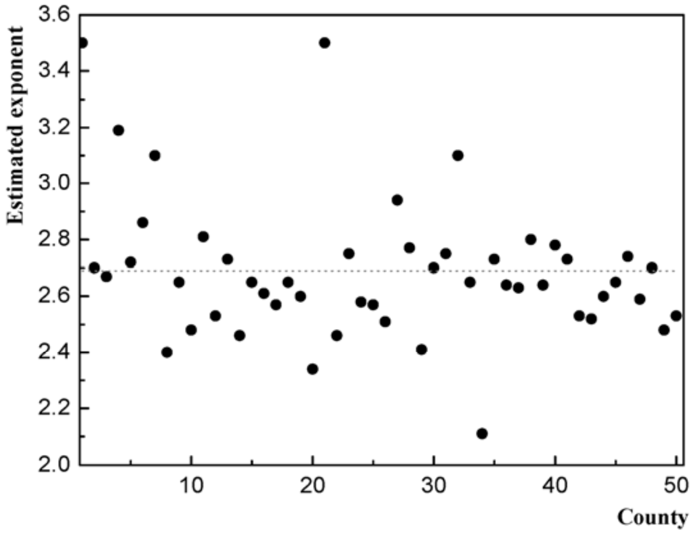


Figure 3.4 Variances of the power law exponents of 50 road networks

Figure 3.4 describes the variance of $\hat{\alpha}$. Intuitively, a larger $\hat{\alpha}$ corresponds to a steeper fitting line and indicates more heterogenous road connectivities in a road network. The results in this part of study are consistent with real-life conditions. For instance, New York County has a small value of $\hat{\alpha} = 2.41 \pm 0.21$. This is because New York County is flat and has a very regular road network and so the road connectivities in this region do not vary much. By contrast, Alameda County has a large value of $\hat{\alpha} = 3.50 \pm 0.05$. One possible reason for this is that there are numerous mountains and rivers in this region, which creates a large number of wandering and minor roads. The road network in Alameda County thus tends to be very heterogenous.

The scale-free structure of urban street networks has been studied by Jiang (2007) by least squares fitting. In Jiang's study, an urban street network is defined as a road network embedded in the largest urban land use parcel in a city. Comparing the results of this part of study with Jiang's results, it is found that values of $\hat{\alpha}$ for county road networks are slightly larger than those for urban road networks. This difference may be caused by regional disparity, e.g. in Jiang's study, the largest urban land use parcel is usually the Central Business District of the city. Roads in this area are usually important and are highly clustered. The road connectivities in this region do not deviate from each other very much, especially in a stroke-based representation. By contrast, in the USA, a county is a very large region which includes one or more cities. It is more likely to have long curved roads and many minor roads created by complex land forms. Roads in a county therefore tend to vary significantly in terms of connectivity.

3.3.3 Influence of scale-free property on road network structure

The power law distributions of road connectivities reflect the scale-free property of road networks. Such a property provides support for defining three hierarchies of a road network (Jiang 2008b): the top 1%, 20% and remaining 80% of streets ranked by road connectivity. The hierarchies are found to be in accordance with the pattern of real-life traffic flow in an urban street network.

Four county road networks are randomly selected to demonstrate the three hierarchies in scale-free road networks as illustrated in Figure 3.5.

(a) Bexar

Figure 3.5 Hierarchical structures of 4 road networks by stroke connectivity (— top 1%, — top 20%, — bottom 80%)

(b) Fresno

(c) Harris

(d) Oakland

Figure 3.5 Hierarchical structures of 4 road networks by stroke connectivity (cont'd)
(— top 1%, — top 20%, — bottom 80%)

Figure 3.5 shows that the top 1% of strokes are well-connected. They extend across the entire region and form a "traffic hub or traffic center" in the region. Further study of the connection between strokes with different hierarchies is carried out in the next section.

3.4 Mixing patterns in 50 sets of road networks

As introduced in Section 3.2, random counterparts are constructed to obtain the value of connectivity correlation probability. It is noted that the larger the number of random counterparts built, the more reliable the value of $Z(k_1,k_2)$ is. In this part of study, 100 random counterparts are built for each road network to obtain $P_r(k_1,k_2)$ and $Z(k_1,k_2)$. Results indicate that all the 50 road networks show mixing patterns in terms of connectivity.

Figure 3.6 visualizes the values of $Z(k_1,k_2)$ for stroke connectivities of the 4 road networks studied in previous section. In Figure 3.6, x and y axes represent stroke connectivities. Due to the fact that the values of road connectivities are not continuous and heterogenous, the values of road connectivities are logarithmically binned into 2 bins per decade to improve visualization. Interpolation technique is involved to produce the final images.

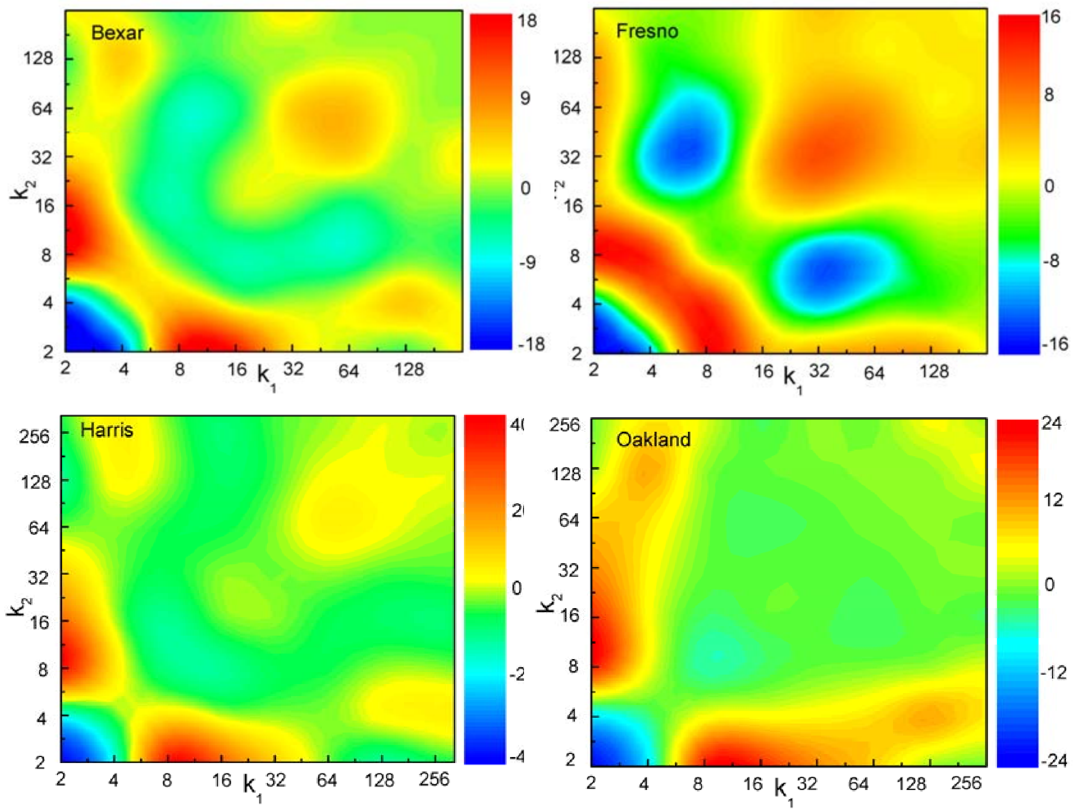


Figure 3.6 Correlation profiles of stroke connectivities in 4 road networks

In Figure 3.6, different colors represent various values of $Z\left(k_1,k_2\right)$. The warm colors like yellow and red mean that there is a positive trend in the connection between roads with given connectivities compared to random conditions. By contrast, blue and green colors indicate a repulsion trend in the connection between roads with given connectivities. Figure 3.6 clearly shows that in a road network, well connected roads tend to connect both to other well connected roads and to roads of low connectivities. Such a phenomenon is called a mixing pattern and its meaning in scale-free road networks will be studied further in the following section.

3.5 Shortcuts created by mixing pattern in scale-free road networks

Shortcuts are vital to improve the efficiency of flow transmission in a network. However, this question has been ignored to some extent in road network study, partly because that human's travel behaviors are complex in a large road network. In this section, topological distances between two roads are studied.

A hypothesis is proposed to tackle the last problem, that is, how does mixing pattern in a scale-free road network affect its function? The hypothesis states that the mixing patterns in a scale-free road network will create an easily accessible road network, or a road network with small average length of the shortest paths between any two roads.

The proposed hypothesis is based on both the empirical results discussed in the previous sections and human experiences. As shown in Figures 3.5 and 3.6, the top 1% of strokes form the skeleton of the entire road network. The top 20% of strokes act as bridges to connect minor roads to the top 1% of strokes. Such "bridges" are also observed in human daily life. For instance, people prefer to drive on highways to save time, but drivers sometimes need to jump from major roads to minor roads to arrive at their destination. This "jumping" process may well be done more frequently if the destination is very far away.

To test the hypothesis, the distances between any two roads in each road network are calculated. By using the term "distance", we mean the shortest topologic distance, or the minimum number of the steps taken from one road to the other road. The probability distributions of the lengths of the shortest paths between any two roads in each road network are shown in Figure 3.7.

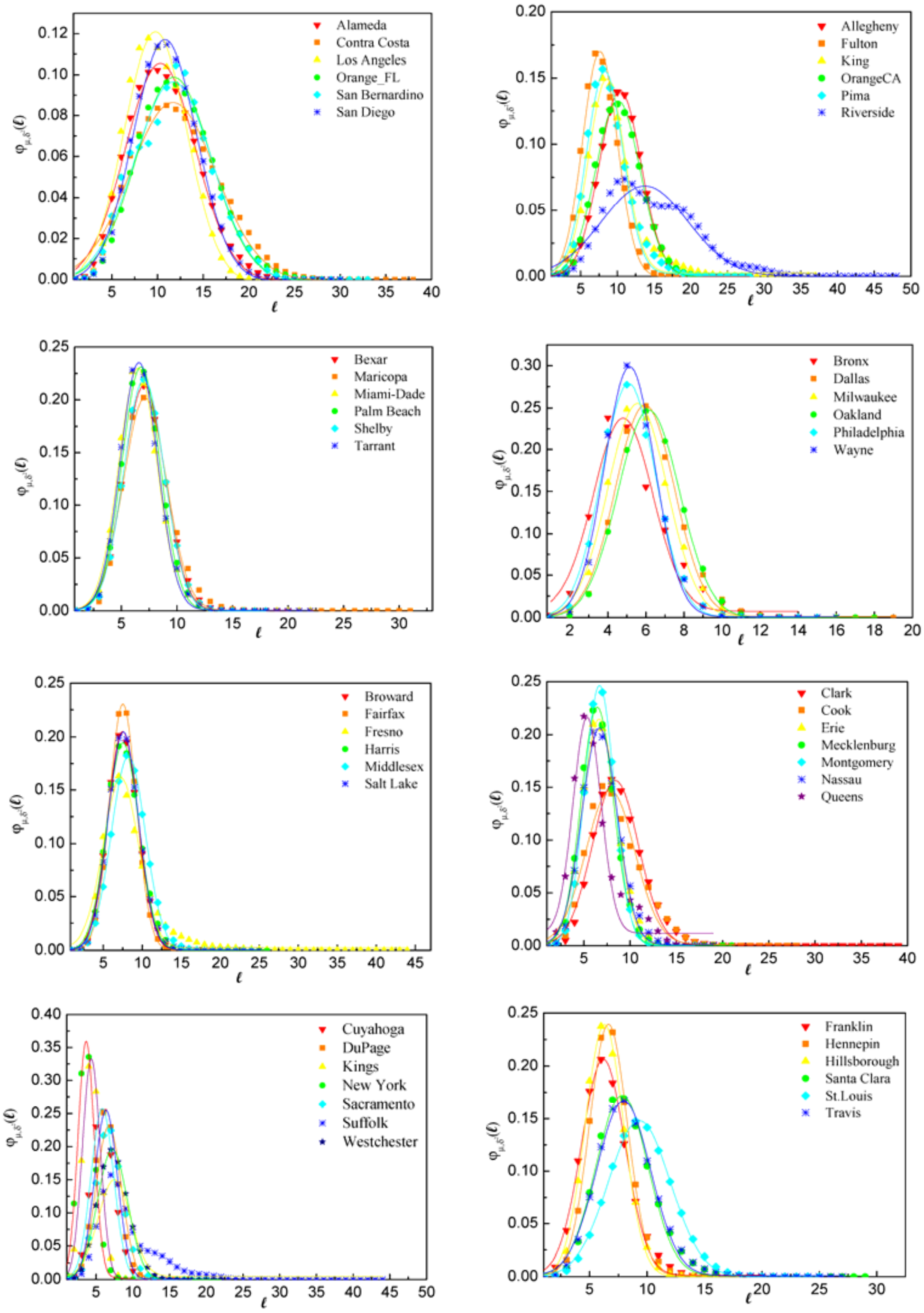


Figure 3.7 Distributions of the lengths of the shortest paths in 50 road networks

Results shown in Figure 3.7 imply that: (1) the average distance between any two strokes in a road network is very small. In our case, the length of the shortest paths between any two strokes range from 1 to 48. These values are indeed small compared to the massive sizes of the road networks; (2) the distributions of the length of the shortest paths in road networks follow a normal distribution, with a mean value of 2.6~12. These results indicate clearly that road networks are easily accessible with the aid of some strokes, which are actually shortcuts in the road network.

3.6 Summary

This part of study focuses on the shortcuts created by the mixing pattern in a scale-free road network. A scale-free network is a network whose distribution of node connectivities follows a power law. In order to examine the scale-free property of the structure of a road network, the revised Kolmogorov-Smirnov statistics was introduced to replace conventional least square method as it provides more reliable and robust statistics for power law distribution. The profile of connectivity correlation probability was firstly introduced to road network study to visualize the mixing patterns of road connectivities. At the end, the hypothesis that the mixing patterns in a scale-free road network create shortcuts has been tested.

Based on the limited experimental tests, it is found that (1) road networks are scale-free. The estimated exponents of power law distributions of road connectivities range from 2.11 to 3.50 with an average of 2.69; (2) shortcuts that make a road network easily accessible are created by the mixing pattern in scale-free road networks, and the profile of connectivity correlation probability is a valid way to visualize the mixing pattern of road connectivities.

Chapter 4 Small-world structure of road networks

In previous chapter, the connectivities of roads were analyzed from the aspect of road-road intersection probability and preferences. In this chapter, a further study on the connection of roads at both local and global levels will be given in the next Chapter.

By "small-world" it is meant that in the network, any two nodes can be connected by a relatively short chain. Many real-life road networks are small-world. And small-world is found to be an optimal way to organize a network and guides network evolution. Investigation the small-world property of road networks can help us understand how road networks organize locally and globally, and to further explore the influence of structural accessibility on urban development.

This Chapter attempts to investigate the small-world property of general road networks. Two questions are addressed. First, the small-world structure of road networks is examined. Second, the relationship between small-world structure of road networks and regional development is considered. Regional development is represented by population density and the housing unit density in the region in this part of study. It is hypothesized that, as a general trend, a more developed region with a high population density and housing unit density, will have a more accessible road network to serve its transportation demands.

4.1 Small-world network: concept and theory

Before examining the small-world structure of road networks, the basic concept and theory of small-world were explained in this section.

4.1.1 Six degrees of separation

Researchers' interests in small-world property were stimulated by a social phenomenon. That is, two people who had never met before surprisingly found that they shared a friend in common. Such a common social experience was first studied by Pool and Kochen in the mid-1950s (Pool and Kochen 1978). They attempted to explore the mathematical characteristics of social contacts. Pool and Kochen's study stimulated American psychologist Stanley Milgram, who carried out the first experimental study to detect accurately the number of degrees of separation in real-life social networks (Milgram 1967). Milgram aimed to calculate the average number of intermediates to get any two people connected. In his experiment, 96 people were randomly selected from the states of Kansas and Nebraska in the USA. Participants were asked to send a piece of mail to their acquaintances. The acquaintance who received the mail would then send the mail along to another acquaintance… the process kept going until the mail reached a designated “target” person in Boston in the USA. During the mail sending process, the information of the intermediaries was recorded. Based on this information, Milgram successfully counted the number of transfers in mail sending. Suprisingly, he found that the mails passed through only six people on average before arriving at their destination. After that, the small world problem has been attracted increasing attention from social science community.

Milgram's finding is both surprising and striking. The finding revealed that people lived in a small world. However, the sample in Milgram’s experiment was not big enough. Dodds and his colleagues (2003) at Columbia University then conducted a modern version of the experiment of social separation with the help of the computer and the Internet. In Dodds’s experiment, over 60,000 participants from 166 different countries were recruited and were given the identity of a target individual. The participants were instructed to send an e-mail message to someone they thought could get the message closer to the target. The process of e-mail sending kept going until it finally arrived to the target. The result of the internet experiment revealed a median chain length of between five and seven. Based on the two mail sending experiments, researchers proposed that people live in a small world.

4.1.2 Network rewiring experiment

The two seminal work of mail sending stimulated researchers’ interests in the modeling and characterizing a small-world network. Much theoretical and empirical work has been undertaken. The most famous one is the model built by Watts and Strogatz (1998). They constructed a series of networks including regular, partly random and completely random networks. All the networks have the same number of nodes and links but different pattern of links. As illustrated in Figure 4.1 (a), a regular graph is built in which each node has the same connectivity of two. A random network is built by randomly removing and reconnecting some links in Figure 4.1(a) but avoiding duplicated links. The process of changing links is called “rewiring”. Obviously, a series of random networks can be produced. Watts and Strogatz proposed that the degree of randomness could be adjusted using a probability p . By changing the value of p , the network can be 'tuned' between regularity ($p=0$) and disorder ($p=1$) as illustrated in Figure 4.1.

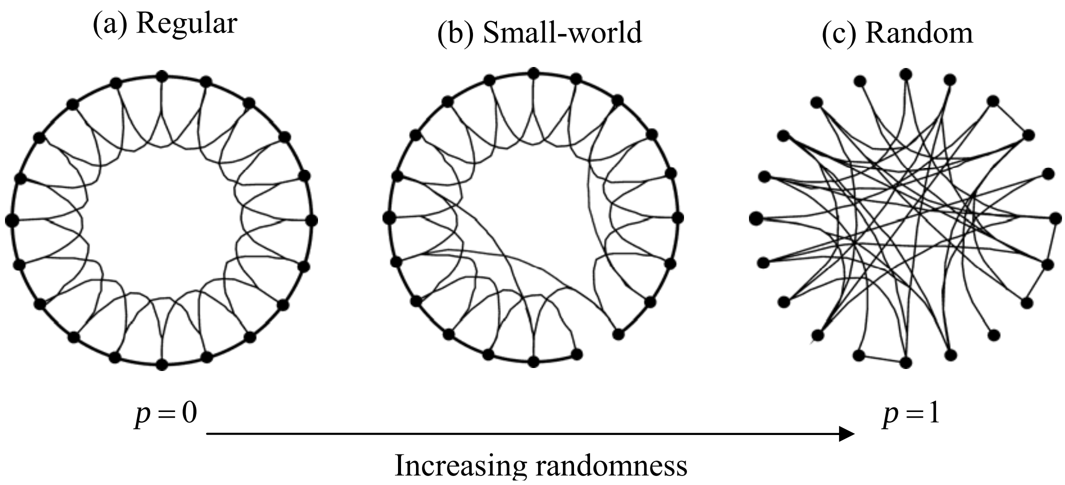


Figure 4.1 Network rewiring experiment (Watts and Strogatz 1998)

Figure 4.1(a) is a regular graph ($p=0$) as each node has the same connectivity of two. Unlike it, Figure 4.1(b) is a partly random network as only several links are rewired. Figure 4.1(c) is completely random because all links are rewired.

Comparing the structural properties of networks with different randomness, Watts and Strogatz (1998) noted that there is a special kind of network. The network behaves partly like regular networks and partly like random networks. The network is called small-world network, which can be produced by adjusting the randomness of link rewiring. Further studies indicated that such a network has high local clustering like regular networks, and good global accessibility like random networks.

Apart from the Watts-Strogatz model, there are other models for building a small-world network. For instance, instead of link rewiring, Monasson (1999) and Newman and Watts (1999) proposed shortcuts adding model. In such a model, a small-world network was built by adding shortcuts between two randomly chosen nodes.

4.2 Design of experimental tests on small-world road networks

Figure 4.1 provides an intuitive understanding of the small-world structure of networks. However, real-life networks are much more complex than that. Data processing is needed to apply the quantitative measures in small-world studies to road networks. The experimental test is designed to examine the small-world structure of road networks.

4.2.1 Data sources and data processing

The dataset involved in this study are the same as Chapter 3, i.e., the road networks of the 50 most populous counties in the U.S.. For each county, its total estimated population, the housing units of year 2008, and the land area of year 2000 are collected from the website of U.S. Census Bureau: http://quickfacts.census.gov/qfd/states/. The main data processing is outlined as follows:

- Step 1: generate a segment-based road map and remove isolated road segments;
- Step 2: build strokes (see Chapter 2) from the segment-based road map;
- Step 3: derive a topological connectivity graph from the final map layer;
- Step 4: build a random counterpart for the connectivity graph of each road network; and
- Step 5: calculate the two small-world measures for each road network and its random counterpart.

It should be noted that in step 2, a deflection angle of 60 degree was set to form strokes. In step 5, the random counterpart for each road network was built with the help of Pajek software (Nooy *et al.*, 2005). For each road network, its random counterpart was built to have the same total number of nodes and links as the given road network.

4.2.2 Measures for characterizing a small-world road network

Two measures have been developed by Watts and Strogatz (1998) to capture the small-world structure of networks. The two measures are clustering coefficient (C) and the characteristic path length (L). As introduced in Chapter 2, clustering coefficient (C) describes the link probability of the neighbours of a node, and characteristic path length (L) measures the separation between two nodes.

Mathematically, clustering coefficient (C) of a node i is defined as the ratio of the existing links between the neighbours of node i to all the possible number of links between the neighbours of node i.

$$C_i = \frac{2\left|\left\{e_{pq}\right\}\right|}{k_i\left(k_i - 1\right)} \quad \left(p, q \in N_i, e_{pq} \in E\right) \tag{4.1}$$

Where, N_i is the set of neighbours of a node i, E is the set of links in a graph G. The clustering coefficient of an entire network is the average value of the clustering coefficients of all the nodes in the network.

The characteristic path length (L) is defined as the average value of the length of the shortest paths between any two nodes in a graph G, i.e.,

$$L = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}}{n\left(n-1\right)} \tag{4.2}$$

Where, d_{ij} is the length of the shortest path between node i and node j.

A sampled road network and its connectivity graph are used to show the calculation of the clustering coefficient (C) and characteristic path length (L) as illustrated in Figure 4.2. Taking node S1 in Figure 4.2(b) as an example, there are six neighbours of it: node S2, S3, S4, S5, S6 and S11. Therefore, the connectivity of node S1 is 6, or, $k_{\mathrm{S1}} = 6$. In an undirected graph, the maximum number of links between the neighbours of node S1 is $k_{\mathrm{S1}} * (k_{\mathrm{S1}} - 1)/2 = 6*5/2 = 15$, when all the neighbours of node S1 are connected to each other. However, in fact, only 5 links are found in the neighbours of node S1, which are e_{S2S3}, e_{S2S4}, e_{S2S11}, e_{S3S4} and e_{S5S6} in Figure 4.2(b). According to Equation (4.1), the clustering coefficient for node S1 is 5/15=0.333. The clustering coefficients for all the other nodes in Figure 4.2(b) are listed in Table 4.1 as below.

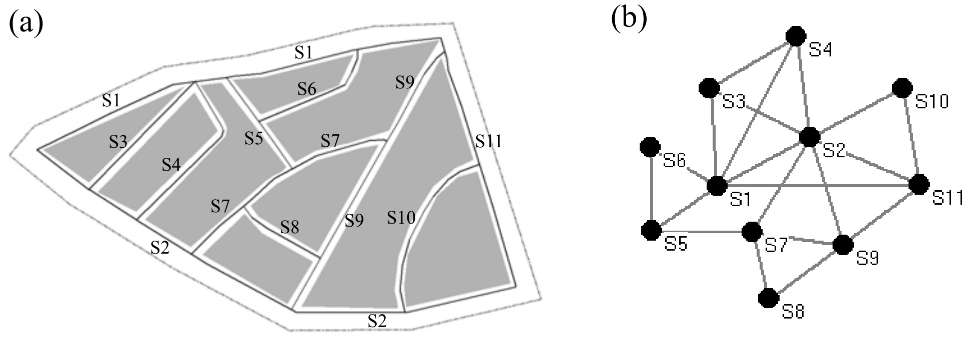


Figure 4.2 A sampled road network and its graph representation

Similarly, the clustering coefficient and the length of the shortest path from a node to all the other reachable nodes can be obtained. Results are shown in Table 4.1.

Table 4.1 The values of C and L for nodes in Figure 4.2(b)

Nodes	Neighbours	Clustering Coefficient	Length of the shortest path from a node to the other reachable nodes
S1	S2, S3, S4, S5, S6,S11	0.333	1+1+1+1+1+2+3+2+2+1=15
S2	S1, S3, S4, S7, S9, S10, S11	0.333	1+1+1+2+2+1+2+1+1+1=13
S3	S1, S2, S4	1	1+1+1+2+2+2+3+2+2+2=18
S4	S1, S2, S3	1	1+1+1+2+2+2+3+2+2+2=18
S5	S1, S6, S7	0.333	1+2+2+2+1+1+2+2+3+2=18
S6	S1, S5	1	1+2+2+2+1+2+3+3+3+2=21
S7	S2, S5, S8, S9	0.333	2+1+2+2+1+2+1+1+2+2=16
S8	S7, S9	1	3+2+3+3+2+3+1+1+3+2=23
S9	S2, S7, S8, S11	0.500	2+1+2+2+2+3+1+1+2+1=17
S10	S2, S11	1	2+1+2+2+3+3+2+3+2+1=21
S11	S1, S2, S9, S10	0.500	1+1+2+2+2+2+2+2+1+1=16

The clustering coefficient and characteristic path length for the sampled road network in Figure 4.2(b) are calculated as follows:

$$C(G) = \frac{1}{11} * (0.333*4 + 1*5 + 0.5*2) = \frac{7.332}{11} = 0.667 \tag{4.3}$$

$$L(G) = \frac{15+13+18*3+21*2+16*2+23+17}{11*(11-1)} = 1.782 \tag{4.4}$$

After introducing the concept and theory of small-world, next section will apply the two measures to road networks.

4.3 Examining the small-world structure of road networks

4.3.1 Experimental results

The results of C and L for 50 road networks and their random counterparts are listed in Table 4.2. It is found that all the 50 road networks show small-world property, with L pretty close to that of the random counterpart and clustering coefficient 1000 times as large as that of the random counterpart. More importantly, the mean value of L is only eight, which means there is an eight degrees of separation in road networks. This finding is novel to current road network study.

Table 4.2 The topological properties of 50 road networks

Rank	County	State	n	m	$\langle k \rangle$	L	C	L_{rand}	C_{rand}
24	Alameda	California	20006	61616	4.160	10.658	0.221	7.090	0.000146
30	Allegheny	Pennsylvania	30211	81511	3.396	10.405	0.169	8.459	0.000108
20	Bexar	Texas	26325	74314	3.646	7.175	0.163	7.960	0.000136
26	Bronx	New York	2283	9999	6.758	5.255	0.255	4.263	0.002428
17	Broward	Florida	20515	55607	3.421	7.614	0.193	8.096	0.000194
15	Clark	Nevada	30554	77161	3.051	8.840	0.128	9.172	0.000079
37	Contra Costa	California	20451	50507	2.939	12.239	0.145	9.114	0.000200
2	Cook	Illinois	55422	192769	4.956	8.435	0.176	6.990	0.000049
28	Cuyahoga	Ohio	13338	38964	3.841	5.972	0.162	7.122	0.000179
9	Dallas	Texas	30274	99524	4.575	6.160	0.284	6.940	0.000209
46	DuPage	Illinois	12286	33305	3.421	6.435	0.133	7.667	0.000152
48	Erie	New York	10350	29457	3.692	6.846	0.174	7.178	0.000152
39	Fairfax	Virginia	19118	44049	2.608	7.569	0.097	10.103	0.000125
34	Franklin	Ohio	3227	7370	2.567	6.436	0.130	8.301	0.000186
49	Fresno	California	19667	55313	3.625	8.241	0.150	7.762	0.000207
40	Fulton	Georgia	14528	37352	3.142	7.853	0.156	8.337	0.000123
3	Harris	Texas	73251	205408	3.608	7.709	0.198	8.801	0.000035
33	Hennepin	Minnesota	15904	46404	3.835	6.773	0.143	7.304	0.000318
32	Hillsborough	Florida	18277	50659	3.543	6.420	0.158	7.831	0.000144
14	King	Washington	48215	121882	3.056	9.416	0.116	9.642	0.000078
7	Kings	New York	2592	14575	9.245	4.391	0.179	3.768	0.002993
1	Los Angeles	California	101590	293536	3.779	9.980	0.143	8.766	0.000034
4	Maricopa	Arizona	79026	208780	3.284	7.421	0.163	9.525	0.000096
8	Miami-Dade	Florida	28034	85653	4.111	6.670	0.158	7.366	0.000164
23	Middlesex	Massachusetts	31115	82221	3.285	8.358	0.209	8.709	0.000086

Rank=ranking in terms of population estimation of year 2009, n =number of strokes (or nodes), m = number of stroke intersections (or links), $\langle k \rangle$ =average connectivity of strokes, L =characteristic path length of the road network, C = clustering coefficient of the road network, L_{rand} = characteristic path length of the random counterpart, C_{rand} = clustering coefficient of the random counterpart.

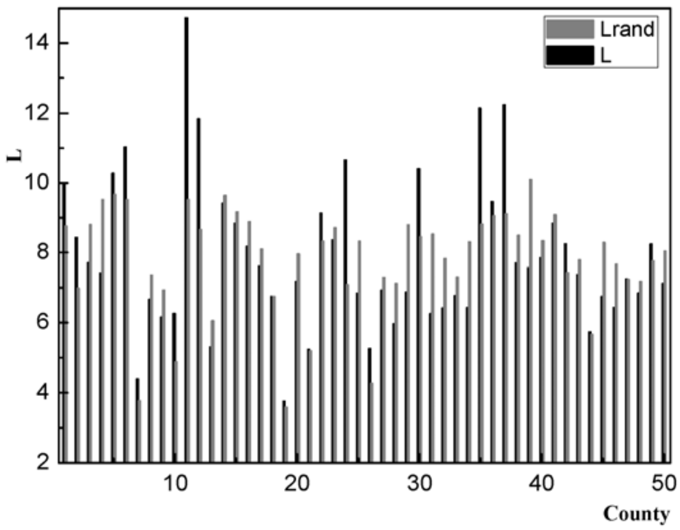
Table 4.2 The topological properties of 50 road networks (cont'd)

Rank	County	State	*n*	*m*	<*k*>	*L*	C	L_{rand}	C_{rand}
44	Milwaukee	Wisconsin	7329	25998	5.094	5.730	0.175	5.665	0.000790
45	Montgomery	Maryland	12193	31125	3.105	6.748	0.149	8.286	0.000452
27	Nassau	New York	17393	51158	3.882	6.940	0.189	7.296	0.000186
19	New York	New York	1249	6499	8.404	3.762	0.274	3.588	0.007894
31	Oakland	Michigan	21312	55858	3.242	6.256	0.164	8.530	0.000090
5	Orange	California	45006	113085	3.025	10.277	0.123	9.672	0.000059
35	Orange	Florida	19902	50296	3.054	12.145	0.171	8.814	0.000143
29	Palm Beach	Florida	20725	52514	3.068	6.879	0.146	8.792	0.000049
21	Philadelphia	Pennsylvania	8186	33075	6.080	5.237	0.206	5.184	0.000752
41	Pima	Arizona	26169	66329	3.069	8.837	0.143	9.083	0.000077
47	Pinellas	Florida	16428	48076	3.853	7.253	0.167	7.238	0.000201
10	Queens	New York	6465	27707	6.571	6.265	0.213	4.883	0.000782
11	Riverside	California	44736	113275	3.064	14.728	0.120	9.522	0.000075
25	Sacramento	California	21306	56587	3.312	6.841	0.155	8.327	0.000167
38	Salt Lake	Utah	18178	46742	3.143	7.707	0.117	8.493	0.000131
12	San Bernardino	California	52593	145004	3.514	11.836	0.150	8.663	0.000033
6	San Diego	California	51022	130500	3.115	11.024	0.146	9.514	0.000035
16	Santa Clara	California	27753	71752	3.171	8.179	0.154	8.887	0.000064
50	Shelby	Tennessee	19808	53854	3.437	7.127	0.207	8.042	0.000237
36	St.Louis	Missouri	20275	50432	2.975	9.467	0.154	9.057	0.000145
22	Suffolk	New York	31743	87473	3.511	9.130	0.155	8.326	0.000142
18	Tarrant	Texas	29109	86679	3.955	6.757	0.232	6.757	0.000048
42	Travis	Texas	12886	36320	3.637	8.249	0.185	7.435	0.000141
13	Wayne	Michigan	18116	66230	5.312	5.303	0.218	6.053	0.000229
43	Westchester	New York	14808	40432	3.461	7.373	0.213	7.793	0.000396

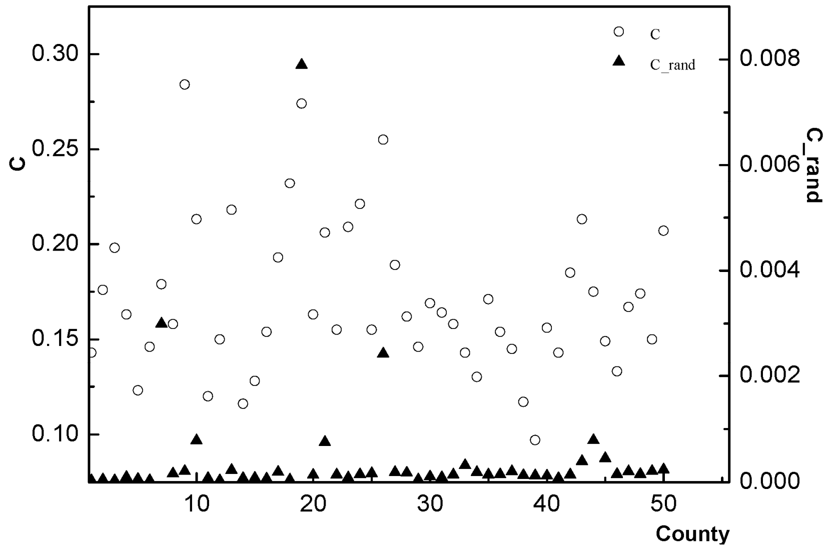
Rank=ranking in terms of population estimation of year 2009, n=number of strokes (or nodes), m = number of stroke intersections (or links), $\langle k \rangle$ =average connectivity of strokes, L =characteristic path length of the road network, C= clustering coefficient of the road network, *Lrand* = characteristic path length of the random counterpart, C_{rand} = clustering coefficient of the random counterpart.

The values of C and L for 50 road networks and their random counterparts are plotted in Figure 4.3. It is found that all the values of L are close to L_{rand}, but C deviates significantly from C_{rand}.

The results of clustering coefficients are consistent with those of primary studies on urban street networks (Buhl *et al.*, 2006, Porta *et al.*, 2006b, Gao *et al.*, 2007, Jiang 2007, Xu and Sui 2007). However, the results of characteristic path lengths are consistent with Jiang (2007) but deviate from those of Portal *et al.* (2006b) and Buhl *et al.* (2006). In previous study, the characteristic path lengths in the road networks are slightly greater than their random counterparts.



(a) Comparison of L and L_{rand}



(b) comparison of C and C_{rand}

Figure 4.3 Comparison of L and C for 50 road networks and their random counterparts

One reason lies in the possible isolated nodes in the random counterpart of a studied network. It is noticed that there are isolated nodes in the final random counterpart produced by Pajek software. And the larger the size of the studied network is, the more isolated nodes are in the random counterpart. As illustrated in Figure 4.4, x and y axes represent the total number of nodes in the studied networks and the percentage of the isolated nodes in their random counterparts respectively.

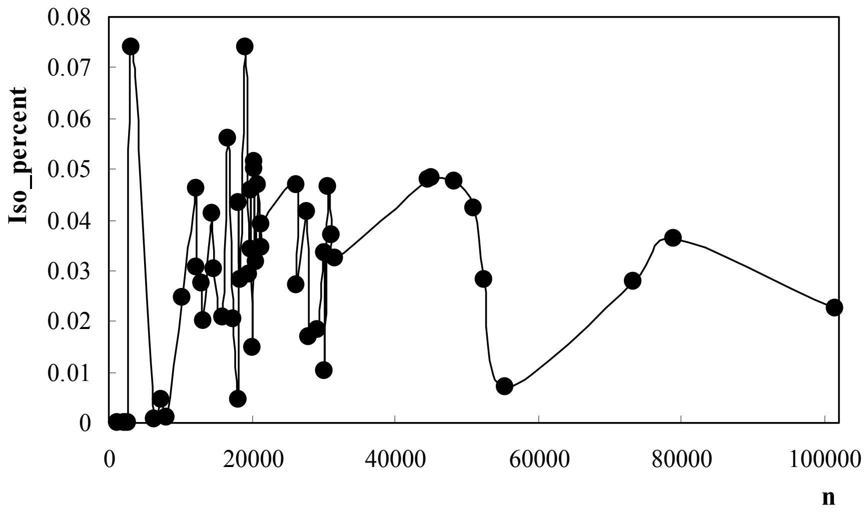


Figure 4.4 The percentages of the isolated nodes in the random counterparts of 50 road networks

Amongst all the 50 road networks, only three random counterparts of road networks do not have any isolated nodes. The three counties are New York County, Bronx County and Queens County. In fact, the three road networks also have the fewest strokes amongst all the 50 road networks. The percentages of the isolated nodes for the remaining 47 road networks range from 0.07% to 7.41%.

The existence of the isolated nodes affects the characteristic path length of the entire network as they influence the length of the shortest path between other nodes. The influence will be very significant if the isolated nodes are traffic hubs. For instance, if a traffic centre is out of operation, people may have to transfer more times than before to arrive to their destinations. Despite the influences of isolated nodes, all the 50 road networks show small-world structure, because that $L > L_{rand}$ or $L \approx L_{rand}$, and $C >> C_{rand}$.

4.3.2 Analysis of the results

This subsection attempts to explore the meaning of the small-world property of a road network to the development of the region. Characteristic path length (L) is adopted to study such a relationship because it is a global measure. Before starting analysis, it is necessary to iterate the small-world—regional development hypothesis. That is, a region with a higher population density or housing unit density signifies a more mature stage of development. And a more developed region has a more easily accessible road network, i.e., a shorter characteristic path length of its road network.

The results of the 50 road networks are illustrated in Figure 4.5. The close relationship between the characteristic path length and population density is obvious. So does that of housing unit density. In Figure 4.5, the five counties with the highest population density and housing unit density are New York county, Kings county, Bronx county, Queens county and Pennsylvania county respectively. All of them have short characteristic path lengths of $L < 7$.

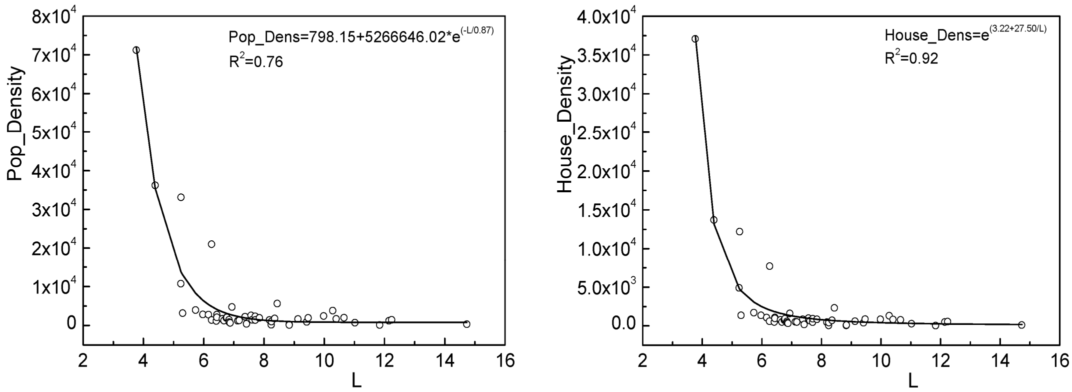


Figure 4.5 Relationships between the characteristic path length and population density and housing unit density

The function fittings for the variables are listed on the top right of Figure 4.5 and are plotted in dark lines. The results imply that: (1) the structure of the road network in a region influences the population density and housing unit density of the region; (2) the better accessibility of the road network, the higher the population density the region has, and vice versa. The principle holds true for housing unit density; (3) A developed region is equipped with a more mature transportation system than a less developed region. These results provide support for the proposed hypothesis in this part of study.

4.4 Summary

Small-world has been an important property for many social and natural networks. By "small-world" it is meant that any two nodes in a network can be connected by a relatively short chain. This part of study examines the small-world property of road networks with the help of two measures, i.e., clustering coefficient and characteristic path length. It is found that road networks are small-world with an average length of the shortest path (between any two roads) of eight. Results also indicate that a more developed region has a more mature road network to meet its transportation demands.

Chapter 5 Hierarchical structure of road networks with ego network analysis

In previous chapter, a study on the connection of roads at both local and global levels has been carried out. This chapter shifts attention to the hierarchical structure of road networks. As mentioned in Chapter 1, hierarchical structure is an important property of road networks especially for navigation and transportation. A number of approaches have been developed to define the hierarchies of roads as discussed in Chapter 2. However, most measures such as flow dimension and flow capacity (Jiang 2008a), still remain at a level of theoretical orientations. This part of study aims to explore a new technique for forming the hierarchical structure of a general road network, which can be applied to infer the distribution pattern of traffic flow. In this part of study, the ego network analysis developed in social science is firstly adopted to describe the hierarchical structure of a road network, which describes the status of each node in an entire network from the knowledge of the local network (or ego-centred network). This is a new application of ego network analysis to the road network.

5.1 Ego network: principles and measurements

Social science focuses on social structure and conceptualizes it as a network of social ties. Sociologists either examine the structure of a social group, or turn to each individual in the local network. The former is defined as the socio-centered approach while the latter has been known as an ego-centered methodology (Nooy *et al*., 2005). Of them, analysis based on ego-network has been identified as one of the three principal approaches for social capital analysis (Odella 2006). It has also been widely applied to other social networks, including computer conference network (Freeman 1982), academic collaboration network (Newman 2003b), citation network (Bar-Ilan 2006) and web questionnaire network (Vehovar *et al*., 2008).

5.1.1 Basic concepts of an ego network

An ego network is defined as a network consisting of a focal node (“ego”) and the nodes to whom ego is directly connected to (these are called “alters”) plus the ties, if any, among the alters (Everett and Borgatti 2005). For the sake of simplicity but without loss of generality, Figure 5.1 lists three simple ego networks, all of which are composed by one ego and two alters. To differentiate these two alters, we name them as alter1 and alter2 respectively.

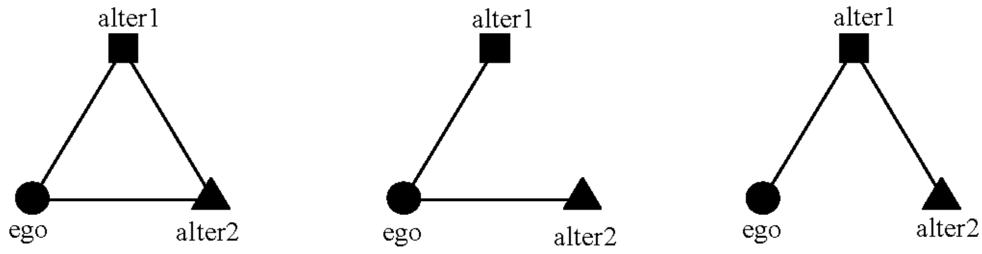


Figure 5.1 Three forms of ego networks

As illustrated in Figure 5.1, there are different ways for an ego to be connected to the alter(s). This means that an ego may have different constraints and opportunities in transfer of flow or information for different networks and thus have different behaviors. Figure 5.1(a) illustrates a complete ego network, in which the ego, alter1 and alter2 can directly communicate with each other. Their statuses in this network are equal. In other words, they have the same constraint and equal opportunity. Such a complete ego network will reduce the individuality of its members, because they share norms and information and tend to behave more like a group than individuals (Simmel 1950). Therefore, a complete ego network is very static and beneficial for all individuals.

However, most ego network is not complete. In Figure 5.1(b), alter1 and alter2 can not communicate with each other directly. The ego acts as an important intermediate in holding the communication between the other two alters. In other words, the ego has more opportunities to control the transfer of information in the entire network. In social science, this opportunity of control is known as the tertius gaudens, or the third who benefits, or the tertius strategy, in which the ego can induce and exploit competition or rivalry between the other two as they are not directly related (Simmel 1950, Nooy *et al*., 2005). In contrast to Figure 5.1(b) where ego stays in an advantageous position, in Figure 5.1(c), ego is severely controlled by alter1. In this case, ego relies on alter1 to communicate with others and thus has fewer opportunities.

An ego network is the basic unit in a complex social network in real life, as almost all kinds of networks can be decomposed to ego networks. In ego network analysis, each node in the network is in turn treated as an ego and all its immediate neighbours as alters, thus a massive network can be decomposed to a number of ego networks. A series of analyses can then be carried out to describe network structure based on various criteria such as size (or degree of ego), composition (e.g. homophily, homogeneity and quality) and structure (e.g. brokerage, efficiency, constraint and density).

5.1.2 Constraint and opportunity in the ego network

In the structure-based analysis, structural hole theory described by Burt (1995) has been treated as one of the most efficient ways to accurately define the positional status (i.e. ranking) of each node in an ego network. This theory was built on the hypothesis that, in a social network, an individual's advantage or power is based on his/her control over the spread of information, goods or services between his/her immediate neighbours, and the absence of a tie between either ego or alter and other alters would induce a structural hole—breakage in communication. Two measurements have been developed to quantitatively describe structural holes, i.e., constraint and aggregate constraint (Burt 1995).

In order to systematically introduce the definitions of constraint and aggregate constraint, some basic definitions and notions of the graph theory are given as follows.

A graph G is formed by a set of vertices (or nodes) $V(G)$ and a set of edges (or links) $E(G)$ that connect pairs of vertices (or nodes). A graph can be directed or undirected, weighted or unweighted, complete or incomplete, and connected or unconnected. In fact, Figure 5.1(a) is an undirected, unweighted but complete graph but Figures 5.1(b) and 5.1(c) are undirected, unweighted and incomplete.

As discussed in section 5.1.1, the ego in Figure 5.1(b) has more opportunities to control the transfer of information in the network than the egos in Figure 5.1(a) and Figure 5.1(c). In social science, the structural hole theory is involved to compute the opportunities for the ego to the spread of information in each case. In order to achieve this, the links between pairs of actors need to be directed and weighted. The weight of each link (p_{ij}) (also called link strength or proportional strength) from node i to its adjacent node j can be defined as the reciprocal of the degree (or connectivity) (k) of node i . Mathematically

$$p_{ij} = \frac{1}{k_i} \qquad (j \in N_i) \tag{5.1}$$

For instance, in Figure 5.2(a), the ego is connected to both alter1 and alter2, so its degree of connectivity is 2. The strengths (weights) of links from this ego to alter1 and to alter2 are both 1/2=0.5. The strengths (weights) of links from alter1 to ego and from alter1 to alter2 are identical as both the ego and alter2 have identical degree (or connectivity). Similarly, the strengths (weights) for links between alter1 and alter2 and between ego and alter2 are identical. However, in Figure 5.2(b), the degree of alter1 is only one, thus the weight of the link from alter1 to ego is 1 (=1/1) although the weight of the link from ego to alter1 is 0.5. A similar illustration of a directed and weighted graph for Figure 5.1(c) is given in Figure 5.2(c).

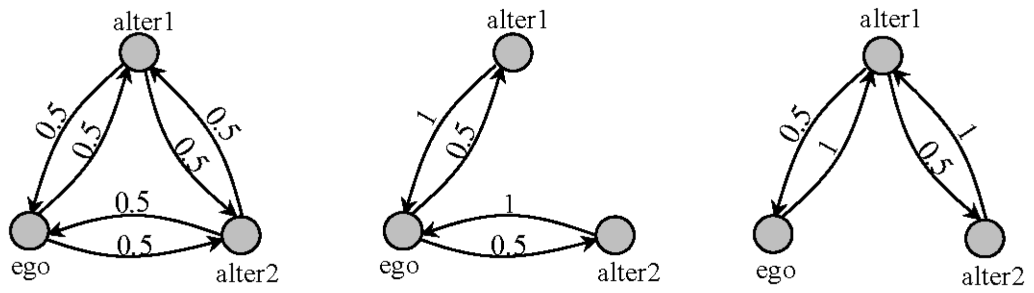


Figure 5.2 Graphs and proportional strengths of three kinds of ego networks

If any two nodes (node j and node q) , which are neighbours of node i , are directly connected, the indirect link strength ($p_{ij}^{'}$) from node i to node j is defined as:

$$p_{ij}^{'} = p_{iq} p_{qj} \qquad (j \in N_i, q \in N_i, q \neq j, q \neq i) \tag{5.2}$$

This is like the relationship between the ego and alter2 via alter1 as shown in Figure 5.2(a). In other words, in this case, the strength of the indirect link from ego to alter2 is 0.5×0.5=0.25. In this case, nodes i , j and q form a triangle.

The constraint (C_{ij}) of node i by node j is computed by the square of the sum of the direct link strength and the indirect link strength from node i to node j (Burt 1995):

$$C_{ij} = \left(p_{ij} + \sum p_{ij}^{'}\right)^2 = \left(p_{ij} + \sum_q p_{iq} p_{qj}\right)^2 \qquad \left(j \in N_i, q \in N_i, q \neq i, q \neq j\right) \tag{5.3}$$

It is noted that the larger the C_{ij}, the bigger the constraint of node i by node j, the smaller the opportunity for node i to spread information. In the case of Figure 5.2(a), the constraint from ego to alter1 is $(0.5 + 0.5 * 0.5)^2 = 0.5625$. On the other hand, the C values for the links from ego to alter1 in Figures 5.2(b) and 5.2(c) are respectively 0.25(=$\left(0.5\right)^2$) and 1(=$\left(1\right)^2$). This shows that in these three Figures, the ego has the biggest control over alter1 in Figure 5.2(b) (i.e. the ego-control network) and most constrained by alter1 in Figure 5.2(c), (i.e. the ego-passive network).

The constraint is for analysis at the link level (e.g. ego to alter). This is not enough because there might be more than one link connected to a node (e.g. ego). Therefore, a new measurement, called Aggregate Constraint$\left(AC\right)$, needs to be introduced to reflect the status of each node. AC_i is defined as the sum of the constraint of node i by all its neighbours.

$$AC_i = \sum_{s=1}^{m} C_{is} \qquad \left(s \in N_i, s \neq i\right) \tag{5.4}$$

Where, m is the size of the set of N_i. It is obvious that the larger the AC_i value, the bigger the constraint over node i by all its neighbouring nodes, and the smaller the opportunity for node i. The AC values for the ego in the Figure 5.1 are 1.125 (=0.5625+0.5625), 0.5 (=$\left(0.5\right)^2 + \left(0.5\right)^2$) and 1(=$\left(1\right)^2$). The results reveal that the ego in Figure 5.2(b) is least constrained by its neighbours and has most opportunity to spread information.

5.1.3 Centrality rank

It can be noted here that a large Aggregate Constraint $\left(AC\right)$ means a small opportunity for the given ego. This is inconvenient to use. Accordingly, a new measurement named Centrality Rank (CR) is introduced to make the relationship more intuitive, which is defined as follows:

$$CR_i = \frac{1}{AC_i} = \frac{1}{\sum_{s=1}^{m} C_{is}} \qquad \left(s \in N_i, s \neq i\right) \tag{5.5}$$

As the name suggests, CR is a measurement of the order of ego in the entire network. The higher the CR, the more important the node in flow transmission in the entire network.

Since CR focuses on structural holes instead of centre and periphery, it provides a different perspective for forming network structure and shows some advantages over traditional network measurements. Taking the degree (or connectivity) as an example, the degree of ego in Figure 5.1(a) and Figure 5.1(b) is both 2 although the statuses of egos are quite different. This is also the case if the clustering coefficient is considered. For example, the clustering coefficient of the ego in Figure 5.1(b) and Figure 5.1(c) is both 0. On the other hand, ego network analysis as explained in section 5.1.2 is able to differentiate these different structures and thus will be applied to the structure of hierarchy for a complex network, which is to be discussed in section 5.2.

5.2 Ego network for forming the hierarchical structure of road networks

In the previous section, the constraint and CR for the actors in an ego network are defined. This section applies the concept of ego network to a larger network, which consists of many ego networks. Figure 5.3 is an example of such a network.

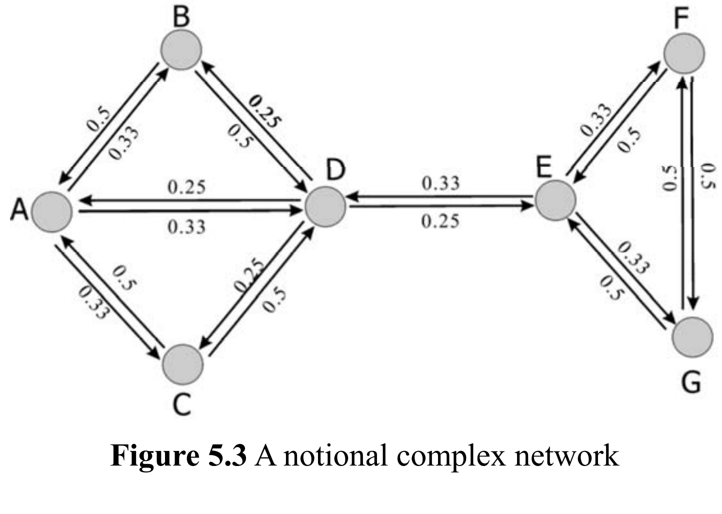


Figure 5.3 A notional complex network

We take node D in Figure 5.3 as an example to show the computation of the CR. The first step is to get the node-node connectivity relation, which is represented by an adjacency matrix A as follows:

$$A=\begin{array}{c} \\ v_A \\ v_B \\ v_C \\ v_D \\ v_E \\ v_F \\ v_G \end{array}\left[\begin{array}{ccccccc} v_A & v_B & v_C & v_D & v_E & v_F & v_G \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{array}\right] \tag{5.6}$$

By summing up each row of A, we obtain the degree k of each node (see matrix (5.7)) and the proportional strength matrix P (see matrix (5.8) as follows:

$$K=\begin{bmatrix} v_A & v_B & v_C & v_D & v_E & v_F & v_G \\ 3 & 2 & 2 & 4 & 3 & 2 & 2 \end{bmatrix} \tag{5.7}$$

$$P=\begin{array}{c} \\ v_A \\ v_B \\ v_C \\ v_D \\ v_E \\ v_F \\ v_G \end{array}\left[\begin{array}{ccccccc} v_A & v_B & v_C & v_D & v_E & v_F & v_G \\ 0 & 0.33 & 0.33 & 0.33 & 0 & 0 & 0 \\ 0.50 & 0 & 0 & 0.50 & 0 & 0 & 0 \\ 0.50 & 0 & 0 & 0.50 & 0 & 0 & 0 \\ 0.25 & 0.25 & 0.25 & 0 & 0.25 & 0 & 0 \\ 0 & 0 & 0 & 0.33 & 0 & 0.33 & 0.33 \\ 0 & 0 & 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 \end{array}\right] \tag{5.8}$$

The second step is to detect the direct and indirect links from node D to all its neighbours, and to sum up the squares of the link strength to obtain constraints $C\left(v_D\right)$ as follows:

$$C\left(v_D\right)=\left\{\begin{array}{cccc} v_A & v_B & v_C & v_E \\ \begin{bmatrix} v_D \to v_A \\ v_D \to v_B \to v_A \\ v_D \to v_C \to v_A \end{bmatrix} & \begin{bmatrix} v_D \to v_B \\ v_D \to v_A \to v_B \end{bmatrix} & \begin{bmatrix} v_D \to v_c \\ v_D \to v_A \to v_c \end{bmatrix} & \left[v_D \to v_E\right] \end{array}\right\}=\left\{\begin{array}{cccc} v_A & v_B & v_C & v_E \\ \begin{bmatrix} 0.25 \\ 0.25*0.5 \\ 0.25*0.5 \end{bmatrix} & \begin{array}{c} 0.25 \\ 0.25*0.33 \end{array} & \begin{array}{c} 0.25 \\ 0.25*0.33 \end{array} & 0.25 \end{array}\right\} \tag{5.9}$$

The aggregate constraint is the sum of $C\left(v_D\right)$:

$$AC\left(v_D\right)=C_{DA}+C_{DB}+C_{DC}+C_{DE}=\left(0.25+0.25\times0.5+0.25\times0.5\right)^2+\left(0.25+0.25\times0.33\right)^2$$

$$+\left(0.25+0.25\times0.33\right)^2+\left(0.25\right)^2=0.25+0.1111+0.1111+0.0625=0.5347$$

$$CR\left(v_D\right)=1/C\left(v_D\right)=1/0.5347=1.8702$$

Similarly the AC and CR for all other nodes can be computed as in Table 5.1.

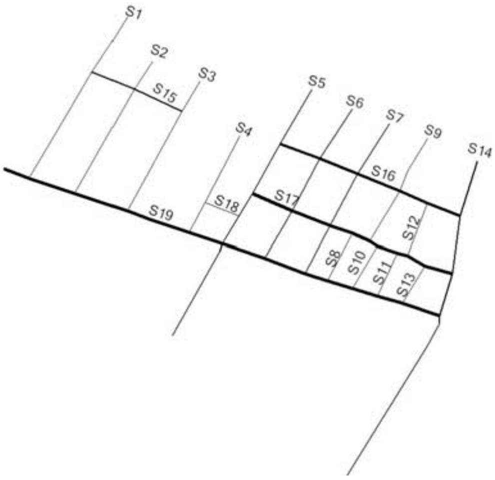
Table 5.1 The computation of relevant measures in a notional network

Nodes	Neighbours	Constraint	Aggregate Constraint	Centrality Rank
A	B, C, D	0.1736, 0.1736, 0.4444	0.7917	1.2631
B	A, D	0.3906, 0.4444	0.8351	1.1975
C	A, D	0.3906, 0.4444	0.8351	1.1975
D	A, B, C, E	0.25, 0.1111, 0.1111, 0.0625	0.5347	1.8702
E	D, F, G	0.1111, 0.25, 0.25	0.6111	1.6364
F	E	0.5625,0.444	1.0069	0.9931
G	E	0.5625,0.444	1.0069	0.9931

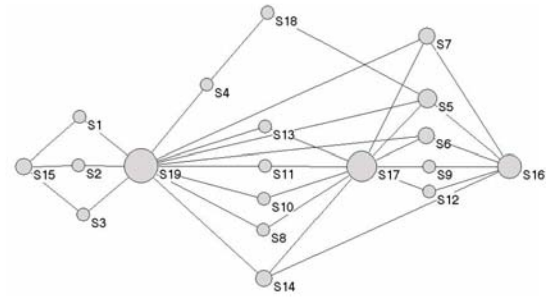
According to the definition of CR, node D is the most important node in the flow transmission, while node F and node G are the least important in the flow transmission. This result coincides well with Figure 5.3. However, Figure 5.3 is rather artificial. In this section, two sets of real-life data are used for the illustration of applying ego network analysis to road networks. One set is a sample of Los Angeles road network, which is very regular and grid-like; another is the Sätra road network, which is bell-shaped with many curved allies.

The most important step in applying ego network analysis to road networks is to form a natural road from road segments because the topological structure of segment-based road network is quite tedious (Kalapala *et al*., 2006). As mentioned in section 1, various approaches have been developed to representing and modeling of road networks. However, as noted by researchers, the application of axial line is limited by its nonautomatic derivation from road networks (Ratti 2004), while the named street is rather culture dependent and is always incomplete in street network databases (Jiang 2008b). By contrast, stroke (see Thomson 2003) has been one of the most frequently used approaches to represent road networks and is thus used in this part of study. A stroke is formed by concatenating a series of road segments based on the Gestalt principle of good continuity (see Chapter 3). We use the every-best-fit principle to derive strokes for a given road network (Jiang *et al*., 2008).

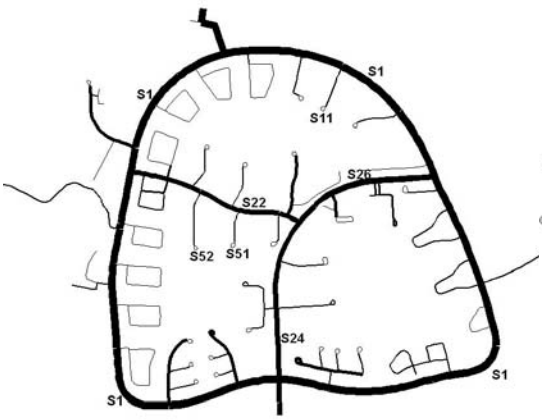
After deriving strokes (i.e. natural roads of a road network), a connectivity graph of the road network needs to be constructed. In such a graph, roads need to be represented by nodes and thus the intersections represented by links, as CR is for node (ego) in an ego network but one wants to use CR to weigh the importance of roads. In this way, a CR for each road in the network (i.e. each node of the graph) can be computed. The stroke representations and their corresponding connectivity graphs are illustrated in Figure 5.4.



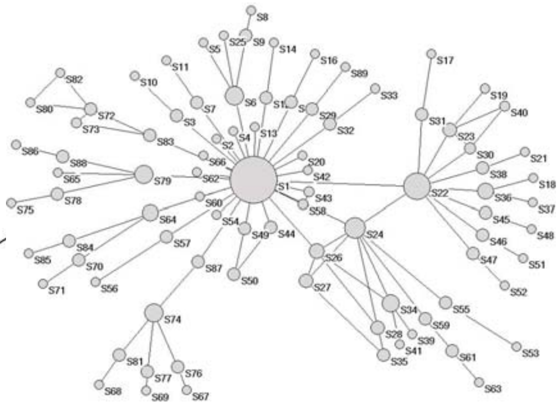
(a) Sampled LosAngeles road network



(b) Connectivity graph of (a)



(c) Sätra road network (Jiang, 2008a)



(d) Connectivity graph of (c)

Figure 5.4 Two real-life road networks and their graphs by ego network analysis

From these connectivity graphs, CR of all strokes (i.e. nodes in Figures 5.4(b) and 5.4(d)) are computed and visualized by size. In Figures 5.4(a) and 5.4(c), the thicker the line, the higher the CR of the stroke. In Figure 5.4(b) and Figure 5.4(d), the larger the node, the higher the CR of the corresponding stroke.

It is illustrated in Figure 5.4(a) that stroke S_{19} has the highest CR value, followed by stroke S_{17} and S_{16}. According to the definitions of Equations (5.4) and

(5.5), the fact that these three strokes have high CR values implies that there are many structural holes in the neighbours of the three strokes, and the three strokes play important roles in flow transmission in the entire network. These inferences are intuitively represented by Figure 5.4(b), where none of the neighbours of strokes S_{19}, S_{17} and S_{16} are directly connected to each other. The three strokes form the backbone to hold the integrity of the entire network. Indeed, other strokes all rely on these three strokes to communicate with others. For instance, the left part of the network relies on stroke S_{19} to reach the right part of the network, and vice versa. If we only look at the right part, strokes S_{17} and S_{16} serve as transferring ports between its neighbours. By contrast, all the longitudinal strokes, except for strokes S_5, S_6, S_7 and S_{14}, have the lowest CR values, which means they possess two important properties. Firstly, there are few structural holes in their ego networks and thus it is easy to navigate in such kind of ego networks. Secondly, they tend to be strongly constrained by their neighbours.

The same pattern is also found in Sätra road network. In Figure 5.4(c), stroke S_1, the bell-shaped stroke, has the highest CR value and thus plays a central role in the entire road network since the internal strokes within the ′bell′ relies on stroke S_1 to reach the outside strokes of the “bell”. Following stroke S_1 are strokes S_{22} and S_{24}, which also have relatively high CR values. This well confirms with the network pattern described in Figure 5.4(d), where the three strokes play a central role in flow transmission between strokes. In a similar fashion, in Figure 5.4(c), strokes with relatively low CR values are those antenniform rings at the end of straight strokes. They are not so important because they are relatively isolated.

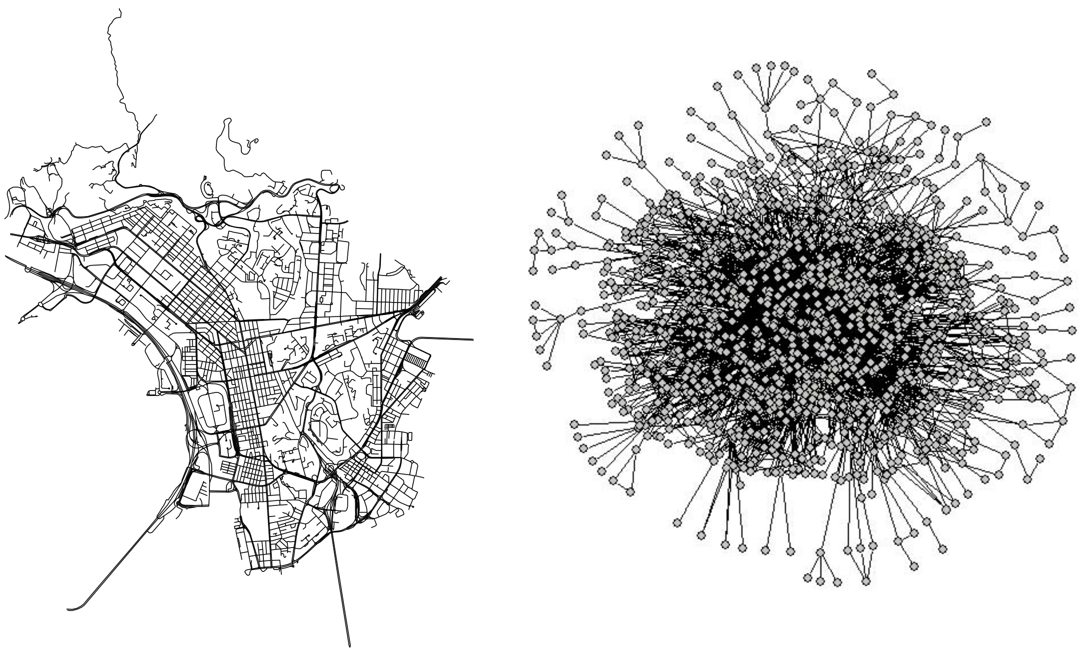
To sum up, in this section, we demonstrate how ego network and CR are applied to characterize the hierarchical structure of road networks. It has been demonstrated that CR, inspired by ego network analysis in social science, can also be used to weigh the importance of roads (strokes) in road network analysis.

It should be noted that some physical properties of the road itself, including width, built hierarchy (e.g. national highway, provincial highway and road etc.), number of lanes, and location of traffic lights, are not considered in our study.

5.3 Evaluation of ego-centered hierarchical road network

5.3.1 Design of experiment

To evaluate the applicability of ego network analysis for structuring hierarchical road network, some real-life data should be used for experimental testing. The dataset used in this part of study is the KSY road network of Hong Kong (Figure 5.5(a)), and its traffic flow data gathered by vehicles equipped with GPS receivers. The KSY district includes three sub-districts, i.e. Kowloon City, Sham Shui Po and Yau Tsim Mong, which are spatially adjacent. There are two main reasons for us to choose KSY district as our study area. One is that the land use pattern in this area is quite diverse and it is highly populated. Traffic flows are always high in this area. Another reason is that the GPS data gathered in this area is comparatively more complete than other districts. The GPS data includes the following information: vehicle_id, utc_date (year/month/date/hour/minute/second), latitude, longitude, speed, course and location. The GPS data is then registered to Hong Kong street network using GIS software. In this part of study, we used the traffic data as a benchmark for evaluation. Such a use is based on the assumption that roads with higher centrality rank values tend to be more important for the entire network and thus will have higher traffic flow.



(a) Original KSY district road (b) Its connectivity graph

Figure 5.5 The original road map of KSY district of Hong Kong and its connectivity graph

5.3.2 Data processing

We firstly convert the original road map into a connectivity graph (see Figure 5.5(b)). Then compute the centrality rank of each modeling unit, and sort them in ascending order.

The data processing can be described in a step-by-step fashion as follows:

- Step 1: generate a segment-based road map and remove isolated road segment.
- Step 2: build strokes from the segment-based road map.
- Step 3: derive traffic flow for each stroke.
- Step 4: produce a connectivity graph (see Figure 5.5(b)) and
- Step 5: compute centrality rank of each stroke in the road network

In the formation of stroke in each road network, a deflection angle threshold is set (degree 60 in this part of study). The final stroke-based KSY road network includes 1444 strokes derived from 6237 road segments. As illustrated in Figure 5.5(b), the connectivity graph of this network is very complex. The average daily traffic flow of each stroke is derived on the basis of CarID in mobility information and GPS points located with in the buffer of the stroke.

5.3.3 Results and analysis

The overall spatial patterns of the centrality rank value and traffic flow are respectively illustrated in Figure 5.6(a) and Figure 5.6(b). The classification systems in both cases are based on percentage. That is, the first class is top 2%, second class top 5%, third class top 10%, forth class top 15% and fifth class the rest. These two graphs reveal a high positive correlation, especially for strokes with high orders in the hierarchy. i.e., the higher the centrality rank value of the stroke, the more traffic flow the stroke accommodates.

The centrality rank value and traffic flow in strokes for each class are then represented separately in Figure 5.7. Comparing the centrality rank values on the left side with traffic flow on the right parts of Figure 5.7, it is found that the stroke hierarchies defined by CR conform well to the hierarchies emerged from real-life traffic flow. Especially for the top 15% strokes, the coherence of the two graphs is as high as 85% even without considering the boundary effect. It means that ego-centered analysis can be used to structure a road network.

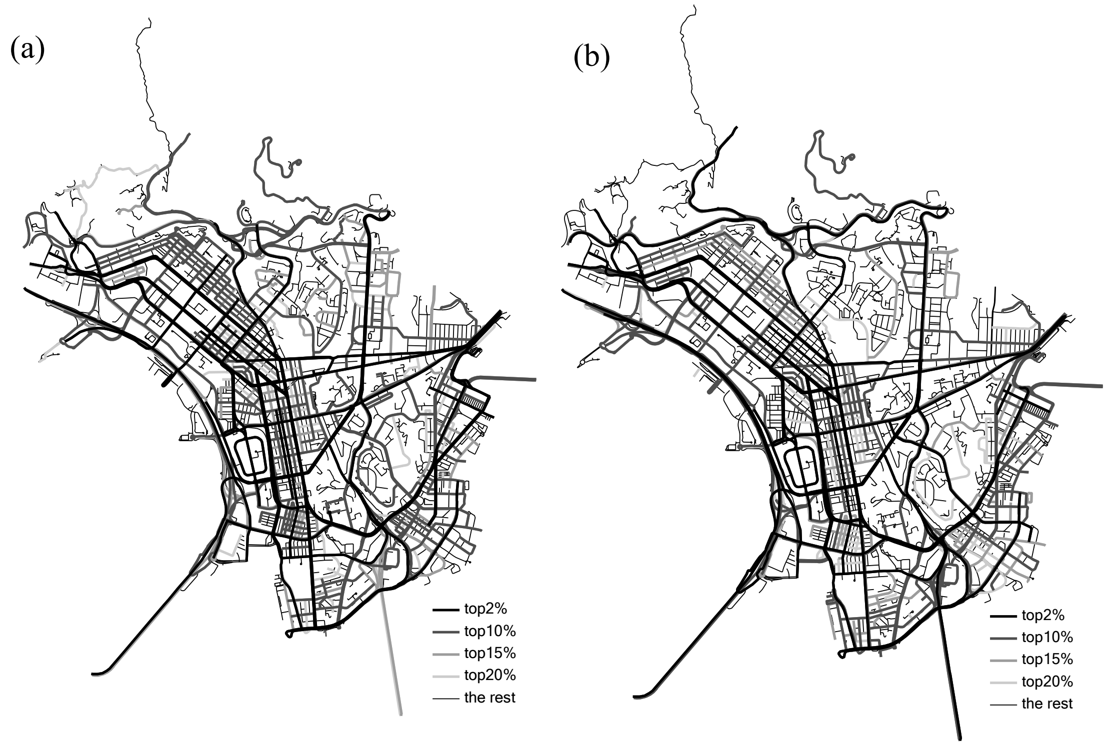
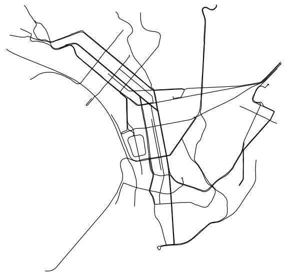


Figure 5.6 The spatial pattern of the KSY road network

It is noted from Figure 5.6 and 5.7 that there are also some differences between the specific strokes selected by *CR* and traffic flow. For instance, some strokes are important in transfer of flow but are not selected according to ego-centered analysis. We noted that most such kinds of strokes are located in the boundary of the study area. This is clearly a kind of boundary effect. In contrast, although some strokes are quite important in terms of topologic analysis, they may not be selected by driver or pedestrian in the navigation process, because of the complexity in land form, location of traffic lights, and/or other factors. For instance, in Figures 5.7(g) and 5.7(h), some strokes are important according to structural hole analysis but are not so frequently used by drivers. This may be due to the fact that almost all such kinds of strokes are located in Yau Yat Tsuen (which is a residential area) or park-clustered areas (e.g., Lok-Fu Park, Kowloon Tsai Park and Kowloon Walled City Park). It is quite possible that those roads are actually more important in transfer of pedestrian flow than vehicular flow.



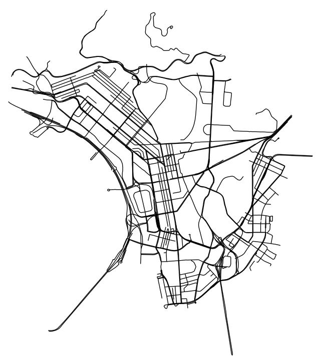
(a)The top 2% strokes selected by CR

(b)The top 2% strokes selected by traffic flow

(c)The top 10% strokes selected by CR

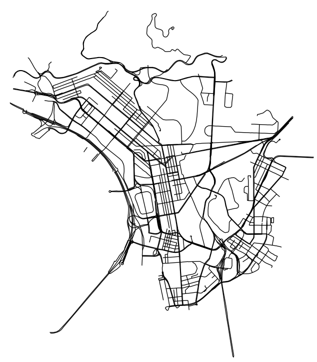
(d)The top 10% strokes selected by traffic flow

(e)The top 15% strokes selected by CR



(f)The top 15% strokes selected by traffic flow

(g)The top 20% strokes selected by CR



(h)The top 20% strokes selected by traffic flow

Figure 5.7 Strokes selected by CR or traffic flow of the KSY road network

The above differences induced by natural or social factors actually do not affect the efficiency and validity of CR to structure road networks. It is found that the top 2%, 5%, 10% and 15% strokes selected in terms of CR respectively accommodate 27%, 49%, 61% and 70% of the total traffic flow of the entire road network. The validity of CR lies in the fact that it well combines both geometric (e.g. length) and topologic properties (e.g. degree, closeness and clustering coefficient) of the strokes.

5.4 Discussion

The ego network is unlike space syntax, which has been widely used to investigate the relationship between structural properties of road networks and traffic flow distribution (Hillier and Hanson 1984, Hillier and Iida 2005, Parvin *et al.*, 2007). The space syntax relies on axial line to represent and model urban space. An axial line is derived on the basis of visual continuity. Space syntax investigates the integration of the entire road network from the aspect of the intersection relationship of an axial line to all its neighbours (both immediate and non immediate). The approach involved in this part of study is unlike that. On one hand, stroke is derived in terms of movement continuity, therefore, stroke well maintains the original form of road network; on the other hand, ego network analysis highlights the integrity and compactness of the local network. The concept of structural holes vividly describes the interrelation between one given node and all its immediate (and only immediate) neighbours.

This also inspires us to extent existing ego network analysis from the immediate neighbours or 1-step neighbours of a given node, to the k-step neighbours of a node. As local integration and global integration respectively considers axial lines that are within a given step distance (e.g. 3-step) to a given axial line and all axial lines, in a similar fashion, 1-step aggregate constraint can be extended to k-aggregate constrains. The 1-step ego network analysis only considers triad, the 2-step ego network takes into account both triads and quaternion, and the 3-step ego network analysis will involve triads, quaternion and even pentad in the ego network. In this way, the concept of local integration and global integration in space syntax can be reinterpreted. However, it should be noted that, the longer the topologic distance between a pair of node, the more nodes in this route, and the smaller the link strength between them (as $\prod_{i=1}^{n-1}\frac{1}{k_i}$ is approach to 0 when k and n is big enough). If the network is big enough, the link strength will approach 0.

5.5 Summary

In this part of study, we have introduced the concept of ego network analysis to form the hierarchical structure of a road network. A new measurement, named centrality rank, is built to assign an order value to each road in a road network. The effectiveness and validity of centrality rank have been evaluated by a real-life data -- the KSY road network. It is found that the road hierarchies defined by centrality rank are in excellent agreement with the spatial pattern of traffic flow distribution. Therefore, based on this limited evaluation, it may be concluded that ego network analysis is a promising means for the formation of hierarchy for a road network.

It is hoped that the ideas and methods presented in this part of study will be useful in the analysis of many other types of urban networks. Possible applications include map generalization, park location, population migration network, daily commuting network and other related areas. Future work is to be carried out to improve this approach. One of the possibilities is to make use of semantic information or searchability (Haken and Portugali 2003, Rosvall *et al.*, 2005a, Boguñá *et al.*, 2009) to weigh link strength because it is currently assumed that all the links from one node to its immediate neighbours have equal link strength (see Equation (5.1)) and it might not be the case for real-life networks.

Chapter 6 Hierarchical structure of road networks with weighted ego network analysis

The previous Chapter shows the feasibility and effectiveness of ego network analysis in ranking strokes in a road network. However, this approach has two significant limitations and should be improved. That is the focus of this chapter.

6.1 Weighted ego network analysis: theoretical basis

It is found that there are at listed two limitations for traditional ego network analysis, which are listed as follows:

- **The deviation of the definition of link intensity from reality**

As illustrated in section 5.1, the weight of each link (p_{ij}) or link strength from a given node i to all its neighbours are identical as defined by Equation (5.1). It obviously deviates from real-life conditions. For instance, in the social relationship network, one tends to communicate more frequently with friends who share lots of common interests with him/her but less frequently with those who are only bowing acquaintances of him/her.

- **The degree 1 effect**

The ′degree 1 effect′ means that those end nodes of dangles (i.e. with degree of 1) are assigned higher ranking orders than other more connected nodes, so lead to unreasonable structure of node hierarchy. It is best explained by an example shown in Figure 6.1. In this Figure, there are 4 nodes. The ranks or opportunities for these four nodes are $A > B > C = D$ according to conventional ego network analysis (i.e. the aggregate constraints of them are respectively 0.6111, 1, 1.0069 and 1.0069). This is obviously wrong, because the removal of node C will reduce the alternative connection chains from node A to node D, so does node D, while the removal of node B will not affect the flow transmission between other three nodes. This phenomenon is described as ′degree 1 effect′.

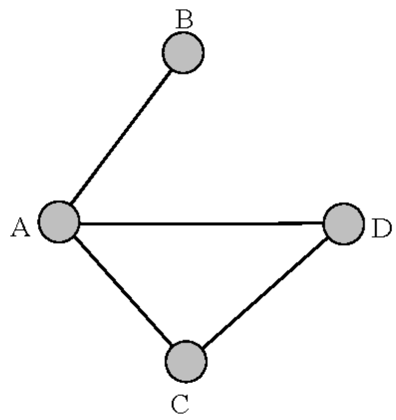


Figure 6.1 A sketch map to show the ′degree 1 effect′

6.2 Weighted ego network analysis: definitions

To overcome these disadvantages, some improvements have been made in this part of study. Firstly, a weight is assigned to each link. To determine the weight of a link, preferential attachment raised by Barabási and Albert (1999) in network science has been used, which states that a new node is more likely to link to other well-connected nodes. The weight of the strength of a link from node i to node j is defined as the ratio of the degree of node j to the sum of the degrees of all i's neighbouring nodes. Mathematically,

$$w_{ij} = \frac{k_j}{\sum_{s=1}^{m} k_s} \qquad \left(j \in N_i \right) \tag{6.1}$$

Where, N_i is the set of the neighbours of node i and m is the size of the set of N_i. The definition of total link strength, constraint, aggregate constraint and centrality rank in conventional ego network analysis need to be changed accordingly.

In fact, it is possible to define weight in different ways, such as using closeness, betweenness, clustering coefficient (Freeman 1979), and stroke length. However, the results are not as good as those obtained by Equation (6.1).

To get rid of the ′degree 1 effect′, we define the Aggregate Constraint (AC) of those nodes of degree 1 as infinite. This is due to the fact that nodes with degree 1 are usually completely controlled by its neighbour in flow transmission. Considering the fact that the ego network for different nodes may have different sizes, the final aggregate constraint of the node is then normalized, i.e. divided by the degree of the node. Finally, a new measure called Weighted Average Centrality Rank ($WACR$) is developed to analyze the opportunity of the node in the flow transmission in the weighted ego network, as follows:

$$WACR_i = \frac{1}{AAC_i'} = \frac{k_i}{AC_i'} = \frac{k_i}{\sum_{s=1}^{m}\left(w_{is} + \sum w_{iq}w_{qs}\right)^2} \qquad \left(s \in N_i,\ q \in N_i\ and\ q \neq s\right) \tag{6.2}$$

From Equation (6.2), it can be noted here the $WACR$ is determined by two factors. One is the degree of the nodes in the ego network and the other is the triads in its ego network. Therefore, $WACR$ combines the concepts of degree and clustering coefficient in traditional network analysis. Intuitively, a higher degree and fewer links in the neighbours of a node lead to a higher opportunity for this node because there will be more structural holes in the ego network of the node. More importantly, the larger the $WACR\left(v_i\right)$ value, the more important status the node stays in flow transmission in the entire network.

The average of the $WACR$ of the individual nodes is the weighted average centrality rank of the graph.

$$WACR\left(G\right) = \frac{\sum_{i=1}^{n} WACR_i}{n} = \frac{\sum_{i=1}^{n}\left(\frac{k_i}{\sum_{s=1}^{m}\left(w_{is} + \sum w_{iq}w_{qs}\right)^2}\right)}{n} \qquad \left(s \in N_i,\ q \in N_i\ and\ q \neq s\right) \tag{6.3}$$

6.3 Theoretical evaluation of weighted ego network analysis

To illustrate the improvements of weighted ego network analysis in characterizing a hierarchical road network, the two sample road networks discussed in Chapter 5 are used for comparison (see Figure 6.2).

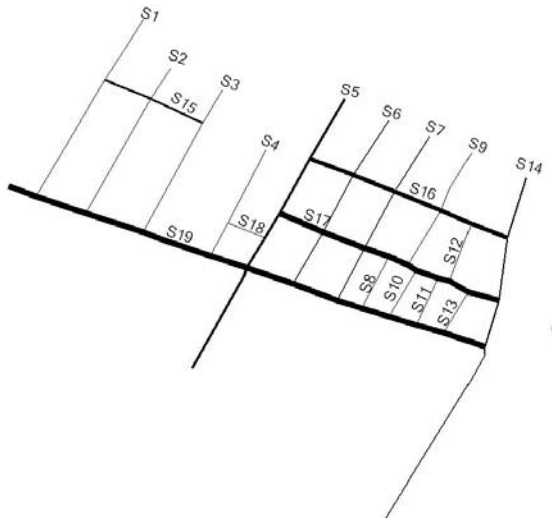
As shown in Figure 6.2(a) and 6.2(b), for the sampled Los Angeles road network, the ranks for all the strokes determined by weighted ego network analysis are listed as follows

$$S_{19} > S_{17} > S_{16} > S_5 > S_{15} > S_6 = S_7 = S_{14} > S_8 = S_{10} = S_{11} = S_{13} > S_9 = S_{12} > S_{18} > S_1 = S_2 = S_3 > S_4 ,$$

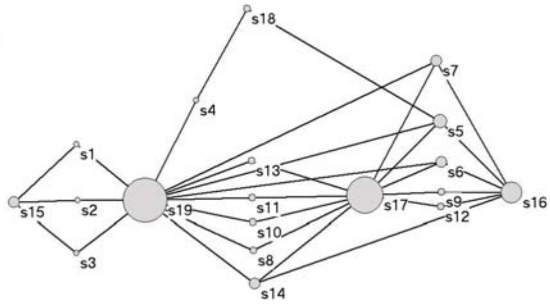
while the result from conventional ego network analysis is:

$$S_{19} > S_{17} > S_{16} > S_5 > S_6 = S_7 = S_{14} = S_{15} > all\ the\ rest .$$

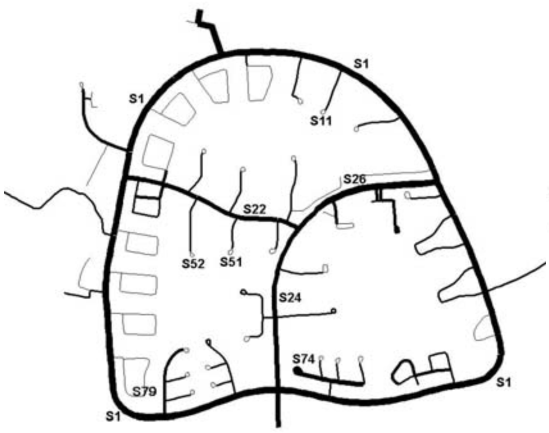
The former result is more reasonable and reliable. For instance, if there is a traffic jam in stroke S_{15}, stroke S_1, S_2 and S_3 can only rely on stroke S_{19} to transfer flows, while the traffic jam in stroke S_6, S_7 or S_{14} will not affect their interconnected neighbouring strokes so much.



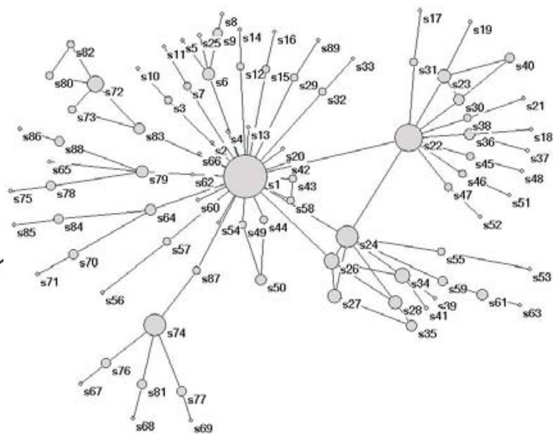
(a) Sampled LosAngeles road network



(b) Connectivity graph of (a)



(c) Sätra road network (Jiang 2008a)



(d) Connectivity graph of (c)

Figure 6.2 Two real-life road networks and their connectivity graphs by weighted ego network analysis

The same pattern is also found in Sätra road network as shown in Figure 6.2(c) and 6.2(d). Due to the limitation of space, details are skipped. Figure 6.2 clearly illustrates a more reliable result obtained by the weighted ego network analysis. As the analysis of the advantages here stays at a theoretical level, the following section will use real life traffic flow data as a benchmark to evaluate them.

6.4 Experimental evaluation of weighted ego network analysis

6.4.1 Data source and data processing

To examine whether the weighted ego network analysis is better than conventional ego network analysis in forming hierarchical a road network, the same dataset used in Chapter 5 was involved for experimental test, i.e., KSY road network of Hong Kong (see Figure 5.5(a)) and its traffic flow data gathered by vehicles equipped with GPS receivers. In this part of study, the traffic data is used as a benchmark for evaluation. Such a use is based on the assumption that roads with higher weighted average centrality rank values tend to be more important in transfer of traffic flow for the entire network.

As described in previous Chapters, before computing the $WACR$ value of each stroke, we should firstly convert the original road map into a connectivity graph as Figure 5.5(b) illustrated. The data processing is the same as that of ego network analysis. The calculation of WACR is achieved with the help of Matlab software.

6.4.2 Results and analysis

The overall spatial patterns of the CR, $WACR$ and traffic flow are respectively illustrated in Figure 6.3(a), 6.3(b) and 6.3(c). The classification systems in all cases are based on percentage. That is, the first class is top 2%, second class top 10%, third class top 15%, fourth class top 20% and fifth class the rest. The results show that they both perform well. i.e., the higher the CR or $WACR$ value of the stroke, the more traffic flow the stroke accommodates. It is also found out that $WACR$ produces more consistent result than CR from visual inspection of Figure 6.4.

(a) The spatial pattern of CR value of the KSY road network

(b). The spatial pattern of WACR value of the KSY road network

Figure 6.3 The spatial patterns of the KSY road network

(c) The spatial pattern of real-life traffic flow of the KSY road network

Figure 6.3 The spatial patterns of the KSY road network (cont'd)

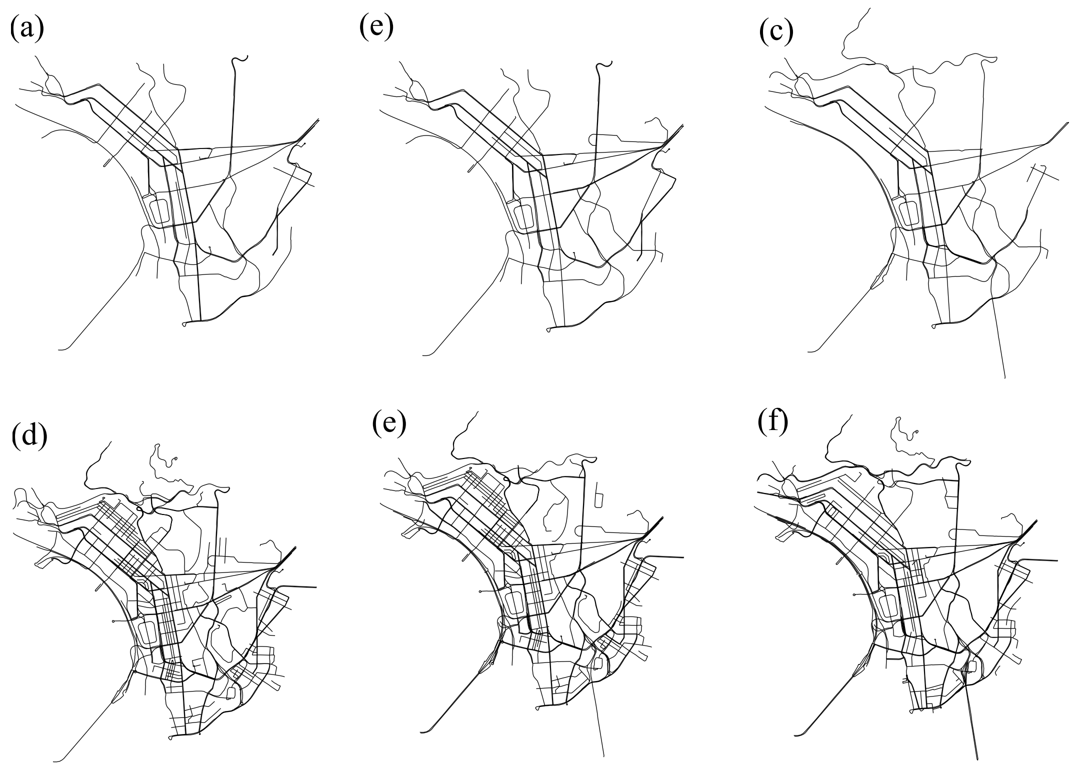


Figure 6.4 The top 2 (a-c), 10(d-f),15(g-i), and 20%(j-l) of strokes selected by CR, WACR, and traffic flow.

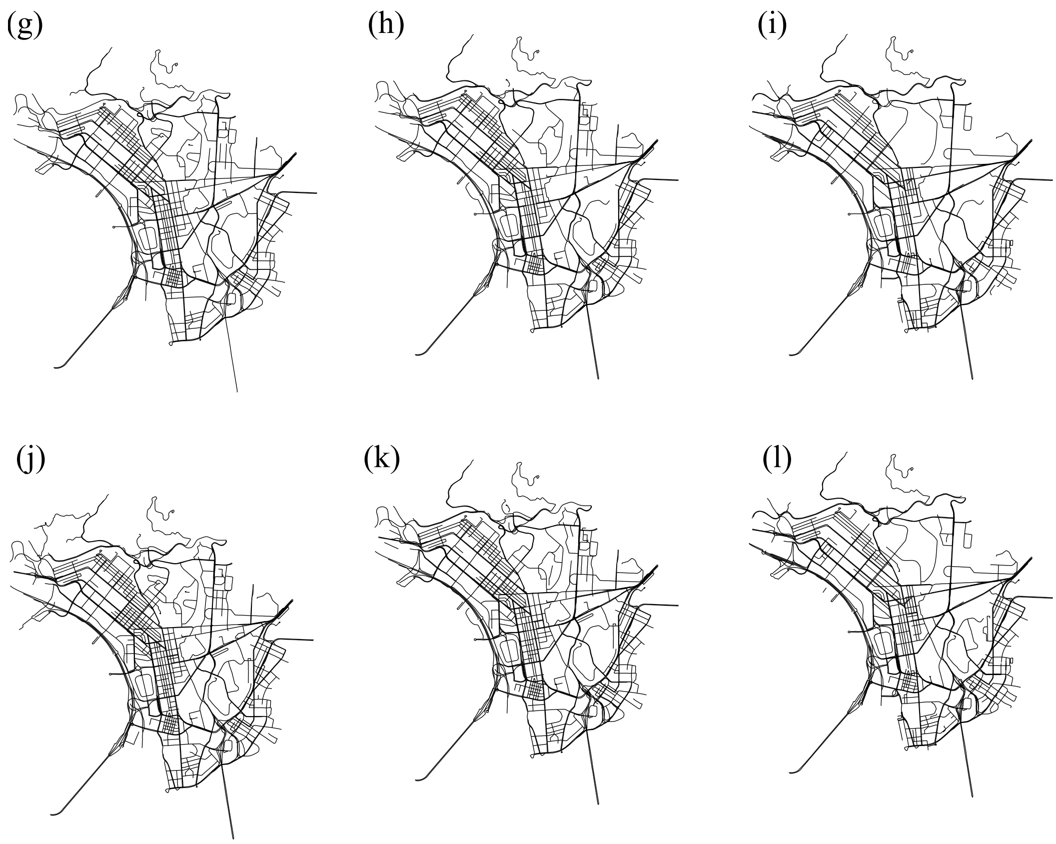


Figure 6.4 The top 2 (a-c), 10(d-f), 15(g-i), and 20 %(j-l) of strokes selected by CR, WACR, and traffic flow.

Figure 6.4 separately represents the first 4 classes of the KSY road network. In each row, the strokes of the same class selected by $WACR$ are closer to the spatial pattern of the corresponding traffic flow. The consistency (in terms of percentage) with traffic flow data (in terms of road segments) is listed in Table 6.1.

Table 6.1 The consistency with traffic flow by $WACR$ and CR

	Number of roadsegments		**Length of road segments**		**Traffic flow accommodated**	
Hierarchy	ego	wego[a]	ego	wego	ego	wego
Top 2%	82.1%	87.1%	79.0%	81.3%	34.1%	34.8%
Top 10%	89.9%	95.8%	83.3%	86.0%	61.4%	63.9%
Top 15%	92.5%	94.0%	87.8%	89.7%	69.7%	70.8%
Top 20%	92.6%	94.6%	89.9%	90.7%	76.3%	77.8%

[a] the weighted ego network

Table 6.1 shows that strokes selected by $WACR$ and CR both fit well to the traffic flow distribution pattern. However, the new method does improve the results by certain percentages, i.e. 5% (82.1% to 87.1%) for top 2% strokes, 6% (89.9% to

95.8%) for top 10%, 1.5% (92.5% to 94.0%) for top 15%, and 2% (92.6% to 94.6%) strokes. This improvement may be more significant for a larger network.

In fact, experiments have also been carried out to compare the hierarchical structure of the road network formed simply based on the length of strokes and those formed by ego networks. The former is far deviated from the real-life data.

Despite the high consistency, there are still some differences between the strokes selected by conventional/weighted ego network analysis and traffic flow. For instance, some strokes are important in flow transmission but are not selected according to ego-centered analysis. It is noted that a large majority of such kinds of strokes are located in the boundary of the study area. This is due to the boundary effect, which can be interpreted as the sensibility of $WACR$ or CR to the boundary conditions. i.e. some strokes located in the boundary are not important in terms of ego-centered network analysis as they usually have few neighbours, but they are important as external links and thus accommodate high traffic flow. In contrast, although some strokes are quite important in terms of structural analysis, they may not be selected by drivers or pedestrians in real life, because of land form, location of traffic lights, personal preference, land use pattern, natural attraction etc.

The above differences induced by natural or social factors actually do not affect the validity of $WACR$ and CR to structure road networks. It is found that the top 2%, 10%, 15% and 20% strokes selected by using $WACR$ accommodate 35%, 64%, 71% and 78% of the total traffic flow of the entire road network (see Table 6.1).

6.5 Summary

In this part of study, we have improved the conventional ego network analysis by assigning a weight to each link in the network. The effectiveness and feasibility of both ego network analysis and weighted ego network analysis have been evaluated by a real-life dataset -- the KSY road network. It is found out that both of them perform well in forming hierarchical structure of a road network, and weighted ego network analysis is more consistent with real-life traffic condition.

Based on this limited evaluation, it may be concluded that ego network analysis provides a promising means for forming hierarchical road networks, and proposed weighted ego network analysis is more robust and reliable.

The approach mentioned in our study has great potential to be applied to other areas. On one hand, the concept of weighted ego network analysis presented in this part of study provides a new tool to uncover the structural properties of real life networks, including social network, cellular network, scientific-citation network and grid network etc; on the other hand, as geographical environment can be described as a network of interacting objects, both the approach of conventional ego network analysis and weighted ego network analysis could shed lights on geographic science. Possible applications include map generalization, regional planning, traffic management, transportation planning, urban crime analysis and urban pollution.

It is noticed that until now all the structural analyses are limited to single scale, it is thus of interest to examine whether the structural properties identified from them hold true for multiple scales of a road network, which is the focus of next chapter.

Chapter 7 Fractality in the topological structure of road networks

The previous four types of complex network analysis are all related to single scale structure of road networks, i.e., the original road network. In this Chapter, attention will be devoted to the structure of a road network at several scales. That is, the fractal property in the structure of road networks will be examined.

7.1 Fractal: concept and mathematics

The term "fractal" comes from the Latin word "*frangere*", which means "to break". It is Mandelbrot (1967, 1982) who firstly used this word to refer to the complex, rough, irregular, and sometimes branching objects found in nature, and who first brought fractals to the attention of the world in his pioneering work "How long is the coast of Britain?" After that, the self-similar mystique of fractals has wormed its way into the physical, biological, and actuarial sciences, as well as psychology and abstract art (Mureika 2007).

Natural patterns including clouds, snow flakes, mountain ranges, and lightning etc. frequently appear irregular and complex, and are hard to measure by Euclidean geometry. Such natural objects are however found to display similar geometric characteristics over a range of scales. Mandelbrot (1967) then coined the term "fractals" to characterize those objects in which properly scaled portions are identical (in a deterministic or statistical sense) to the original object. Usually "fractal property" and "fractality" are used interchangeably. Mandelbrot also introduced the concept of fractal dimension to measure the irregularity of the object.

In his pioneering work, Mandelbrot (1967) asked "how Natural patterns including clouds, snow flakes, mountain ranges, and lightning etc. frequently appear irregular and complex, and are hard to measure by Euclidean geometry. Such natural objects are however found to display similar geometric characteristics over a range of scales. Mandelbrot (1967) then coined the term "fractals" to characterize those objects in which properly scaled portions are identical (in a deterministic or statistical sense) to the original object. Usually "fractal property" and "fractality" are used interchangeably. Mandelbrot also introduced the concept of fractal dimension to measure the irregularity of the object.

Mathematically, geometric fractal dimension (D_g) can be obtained from the slope of the log-log plot of the step-lengths and corresponding total lengths. Just as the mean describes normal data, the fractal dimension is the single most important descriptor of fractal objects, processes, and data (Brown, Witschey, and Liebovitch 2005). Intuitively, the value of D_g reflects the space-filling ability of an object and the complexity of the morphology of the object. An object with a higher value of D_g means a more complex or irregular shape and thus conveys more information (Mandelbrot 1967, 1982; Goodchild and Mark 1987; Batty 2005).

7.2 Box-counting algorithm for computing fractal dimension of a complex network

Most existing approaches have been developed for the computation of the fractal dimension of geometric object in Euclidian space. In this section, a box-counting algorithm for calculating the fractal dimension of the topological structure of a road network is outlined.

7.2.1 Box-counting for obtaining geometrical fractal dimension

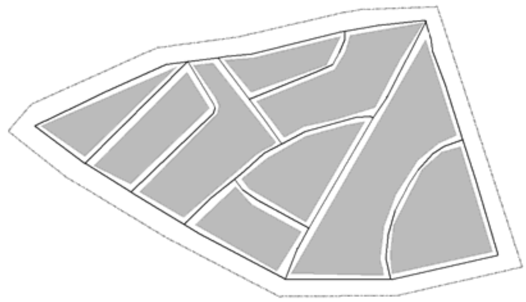
To facilitate the discussion of box covering in the structural space, some concepts and methods for box-counting in Euclidean space are introduced here.

In Euclidean space, there are a number of ways to calculate the fractal dimension, including Calliper, box-counting, pixel-dilation and mass-radius (Mandelbrot 1982, Peitgen and Saupe 1985). Amongst them, box-counting has been a primary way to compute the fractal dimension of self-similar objects in fractal geometry (Hausdorff 1919, Richardson 1961, Mandelbrot 1967, Feder 1988). The mathematical description of box-counting is briefly given below.

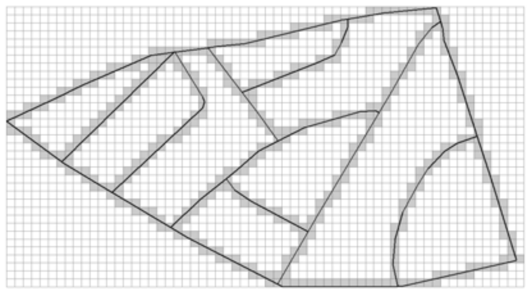
The box-counting refers to the process of covering the object with a regular gird of size S. By progressively reducing the value of S, one can obtain a series of corresponding numbers N_s. It is found that the number of the boxes required to cover a given space depends on the sizes of the boxes, i.e.

$$N_S \propto S^{-D} \tag{7.1}$$

Where, S is the lateral size of the boxes, N_s is the number of boxes required to cover a given object with boxes of size S (see Figure 7.1) and D is the box-counting fractal dimension.



(a) a notional road network



(b) box covering for (a)

Figure 7.1 A sketch map for calculating geometric fractal dimension

7.2.2 Fractal dimension: from geometric space to structural space

The above definition of the fractal dimension also holds true for the fractal dimension of a complex networks in the structural space. That is, one could tile a given complex network with a number of boxes with a specified topological distance. Now the issues are how to title the network in structural space, and how to define the topological distance.

Topological distance is defined by the order of neighbours. If two nodes are directly connected, they are first order neighbours (or immediate neighbours) and their topological distance will be 1. On the other hand, if two nodes are connected via another node, their topological distance will be 2. Suppose the given lateral size (in terms of topological distance) of a box is ℓ_B, the boxes required to cover a network N_B is then:

$$N_B \propto \ell_B^{-D_t} \tag{7.2}$$

N_B value to be obtained from the box-counting process is not unique and may be starting-point dependent. Therefore, in Equation (7.2), the minimum number of boxes required is counted for N_B.

In fact, to find the minimum number of boxes needed to cover a topological structure (or dual representation) of a road network is a challenging task although considerable attention has been recently devoted to this problem (Song *et al*., 2005, Goh *et al*., 2006, Song *et al*., 2007, Kim *et al*., 2007a, 2007b). In the next sub-section, a relatively reliable algorithm, called Maximum Excluded Mass Burning (MEMB) is described.

7.2.3 Computing structural fractal dimension of a complex network

To illustrate the principle of MEMB to tile a road network with a given box size, the sampled street network in Figure 7.1 is used as an example (Figure 7.2(a)). This road network is first converted to a graph (Figure 7.2(b)) and then the box-counting process can be applied to this graph.

From existing literature, it can be found that two algorithms have been widely used for the box-counting of a complex network. They are Compact-Box-Burning (CBB) and Maximum-Excluded-Mass-Burning (MEMB), both developed by Song *et al*. (2007). It has also been pointed out by themselves that the MEMB algorithm produces more reliable and deterministic result, especially for inhomogeneous networks. Therefore, it is adopted in this part of study. This algorithm can be downloaded from the link: http://lisgi1.engr.ccny.cuny.edu/~makse/soft_data.html.

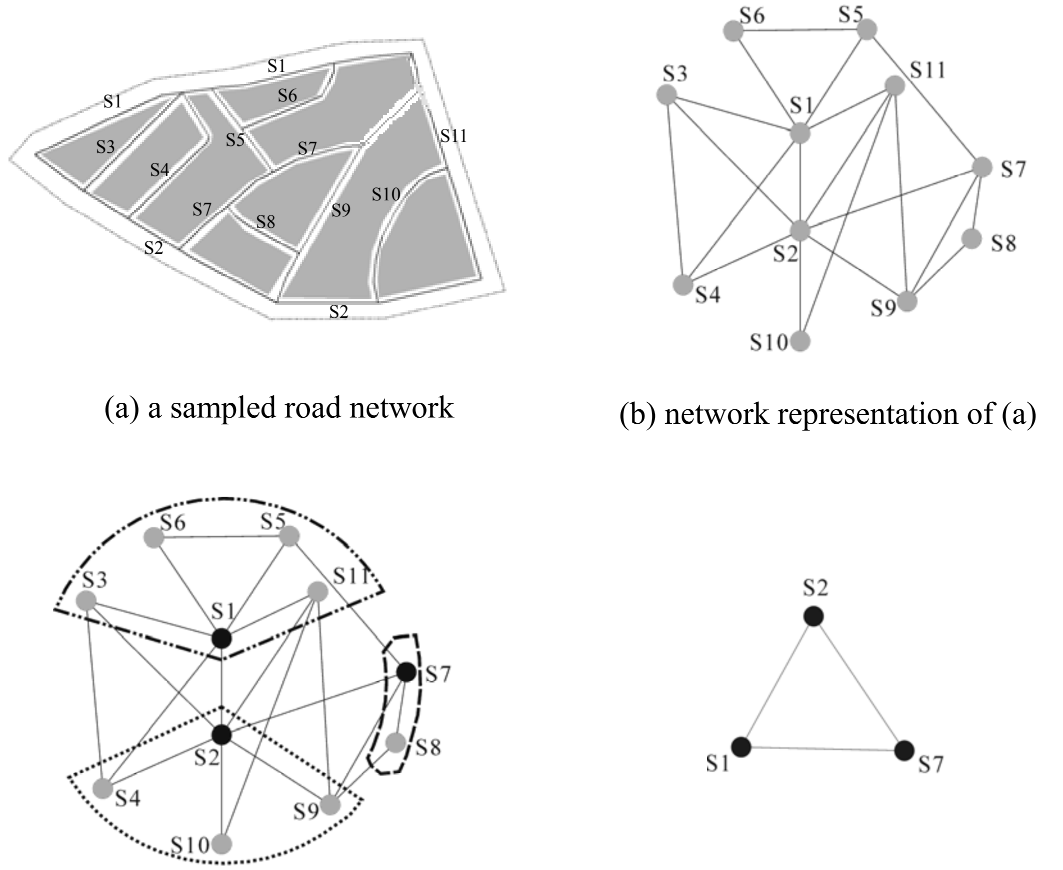
The MEMB algorithm attempts to find the minimum number of boxes needed to tile the network with the constraint that the maximum topological distance between any two nodes in a box can not exceed the box size ℓ_B.

As an example, a box with size $\ell_B = 1$ is applied to tile the road network (Figure 7.2(c)). The basic principles of MEMB algorithm are that

(1) each node belongs to only one box, and

(2) the distances between any given two nodes in a box should be smaller than $\ell_B + 1$.

The box-counting process starts from S1. According to the first principle, nodes S1 to S6 and S11 should be covered by this first box. However, nodes S2 and S4 are then excluded after considering the second principle. The S2 then becomes the next starting point. According to the first principle, all the remaining nodes except S8 should be covered by this box. However, after considering second principle, node S7 is excluded. In the end, three boxes are identified to tile this entire network (see Figure 7.2(c)).



(a) a sampled road network (b) network representation of (a)

(c) road network tiled by boxes with $\ell_B = 1$ (d) the renormalization graph of (c)

Figure 7.2 A sketch map for the calculation of structural fractal dimension

It should be noted that in Figure 7.2 (c), each box is depicted by different line styles and each black node means central node or burning seed of the box. To give a better visual impression, each box is replaced by a single node. This process is called renormalization in network science. After this, links between the nodes are built if there is at least one link between the boxes at the former stage. The result is a renormalized graph shown in Figure 7.2(d). It should be noted here that Figure 7.2 demonstrates the working process for only a simple road network, but real-life road networks are more complex than that.

By applying different sizes of boxes to tile the same road network, we can obtain a series of N_B and ℓ_B to calculate the fractal dimension of a road network. More details will be given in the following section.

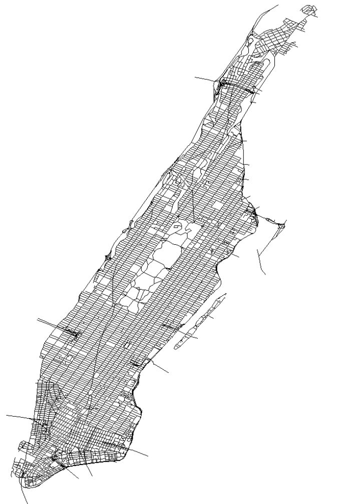
7.3 Procedure for computing fractal dimension of real-life road networks

The data studied in Chapter 3 and Chapter 4 have been used to explore the fractal property or fractality in the topological structure of road networks. The procedure from raw data to the calculation of the structural fractal dimension can be described in a step-by-step fashion as follows:

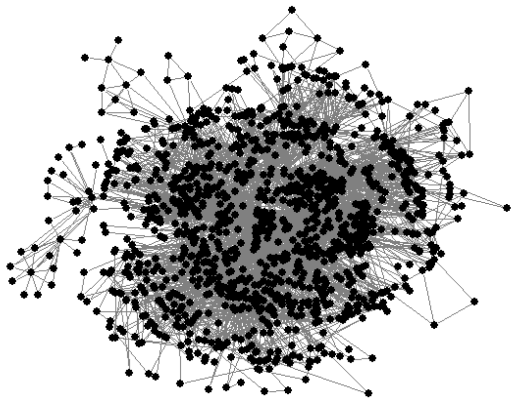
- Step 1: generate a segment-based road map and remove isolated road segments.
- Step 2: build strokes from the segment-based road map.
- Step 3: produce a topological connectivity graph of the final map layer.
- Step 4: obtain a series of ℓ_B and N_B for the connectivity graph of each road network
- Step 5: compute the structural fractal dimension of each road network

It should be noted that in step 2, a deflection angle of degree 60 was used to build strokes. The whole data processing is illustrated in Figure 7.3. In Figure 7.3, the road network of New York County was used as an example and boxes with different sizes of ℓ_B have been used. In Figure 7.3(b)-(f), boxes with sizes of 1, 3, 5, 7 and 9 are used. The number of boxes required for each case is recorded.

In Figure 7.3(g), x axis is box size (ℓ_B) in log scale and y axis is the number of boxes needed to tile the entire road network (N_B), also in log scale. The slope of the inclined straight line is the fractal dimension, which is obtained from the least-square fitting of (ℓ_B, N_B). From the log-log plot (see Figure 7.3(g)) of the number of boxes against box size, a value of 3.56 is obtained for the fractal dimension of this network. It is noted that New York County has the smallest number of strokes. Therefore, there are only four points involved in the calculation of the final fractal dimension. The left county road networks all involve much more points in the calculation of structural fractal dimension.

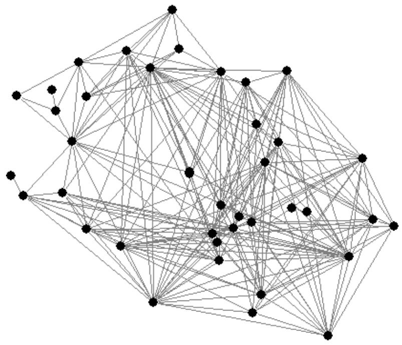


(a) original New York

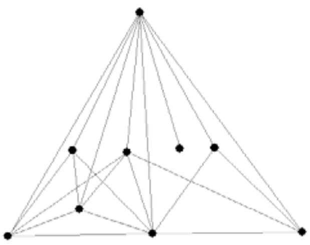


(b) $\ell_B = 1$ (N_B=1249)

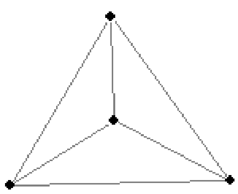
(c) $\ell_B = 3$ (N_B=191)



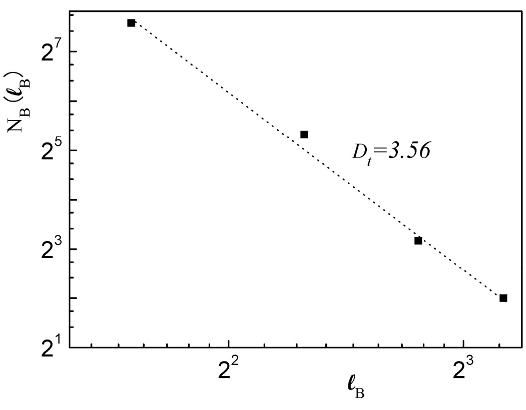
(d) $\ell_B = 5$ (N_B=40)



(e) $\ell_B = 7$ (N_B=9)



(f) $\ell_B = 9$ (N_B=4)



(g) Calculation of D_t

Figure 7.3 The road network in New York County and its renormalization

In the end, the grid-like New York road network is gradually renormalized to a single node. During this process, the backbone of the road network of New York County is emerged, which composed of a triangle, its barycenter and all the links (see Figure 7.3(f). The box-counting of New York road network at different stages is depicted in Figure 7.4. In the following figures, different colors represent different boxes.

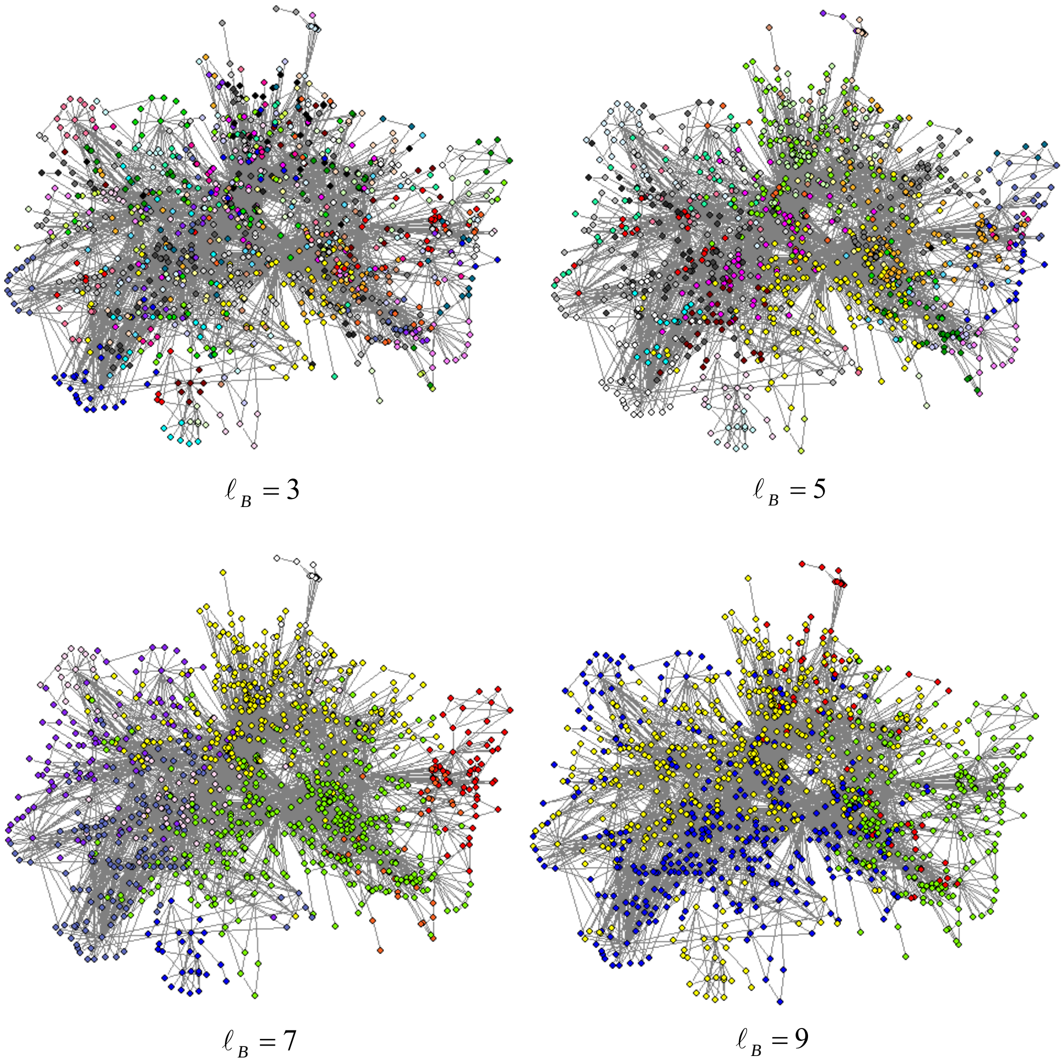


Figure 7.4 Box-counting for road network in New York county

7.4 Structural fractal in road networks

Following the principle procedures described in section 7.2, a thorough analysis on the fractality of the topological structure of 50 road networks has been conducted. Figure 7.5 lists the fractal scaling analyses for the topological structures of 50 road networks based on the MEMB approach. Actually all road networks tend to be fractal in their topological structures. In other words, the connectivity manners of strokes in all the 50 road networks are self-similar at different (length) scales.

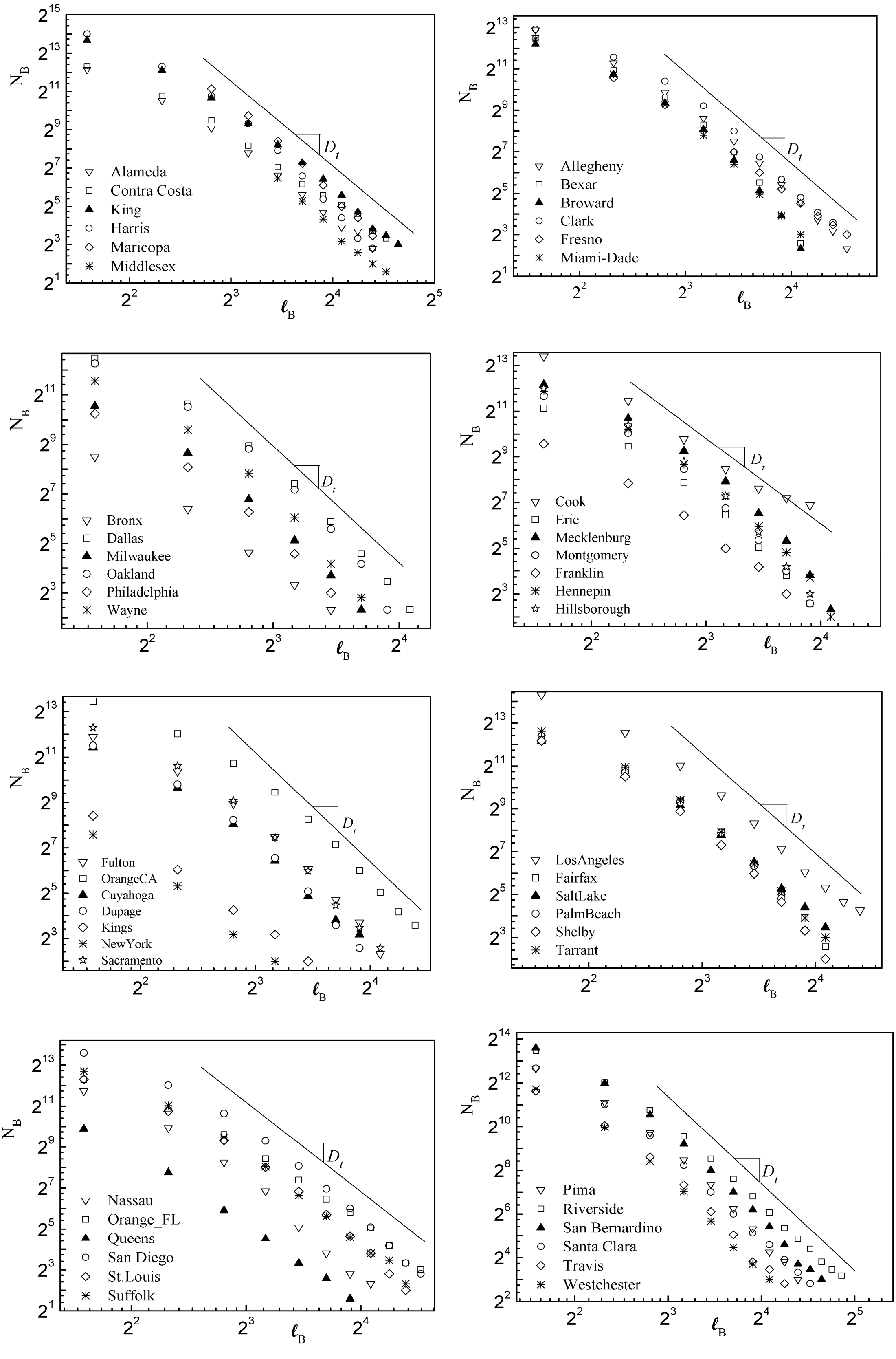


Figure 7.5 Computation of structural fractal dimensions of 50 road networks

The general network properties and final structural fractal dimensions for all the 50 road networks are listed in Table 7.1.

Table 7.1 50 road networks and their structural measures

Rank	County	State	*n*	*m*	<k>	*L*	D_t
24	Alameda	California	20006	61616	4.16	10.66	3.58
30	Allegheny	Pennsylvania	30211	81511	3.40	10.40	3.69
20	Bexar	Texas	26325	74314	3.65	7.18	4.25
26	Bronx	New York	2283	9999	6.76	5.25	3.32
17	Broward	Florida	20515	55607	3.42	7.61	3.89
15	Clark	Nevada	30554	77161	3.05	8.84	3.53
37	Contra Costa	California	20451	50507	2.94	12.24	3.42
2	Cook	Illinois	55422	192769	4.96	8.44	2.94
28	Cuyahoga	Ohio	13342	38968	3.84	5.97	3.72
9	Dallas	Texas	30274	99524	4.57	6.165	4.11
46	DuPage	Illinois	12286	33305	3.42	6.44	3.91
48	Erie	New York	10350	29457	3.69	6.85	3.68
39	Fairfax	Virginia	19118	44049	2.61	7.57	3.95
34	Franklin	Ohio	3227	7370	2.57	6.44	3.09
49	Fresno	California	19667	55313	3.62	8.24	3.72
40	Fulton	Georgia	14528	37352	3.14	7.85	3.83
3	Harris	Texas	73251	205408	3.61	7.71	4.72
33	Hennepin	Minnesota	15904	46404	3.84	6.77	3.84
32	Hillsborough	Florida	18277	50659	3.54	6.42	3.91
14	King	Washington	48215	121882	3.06	9.42	4.23
7	Kings	New York	2592	14575	9.24	4.39	4.34
1	Los Angeles	California	101590	293536	3.78	9.98	4.44
4	Maricopa	Arizona	79026	208780	3.28	7.42	4.90
8	Miami-Dade	Florida	28034	85653	4.11	6.67	3.87
23	Middlesex	Massachusetts	31115	82221	3.28	8.36	3.71
44	Milwaukee	Wisconsin	7329	25998	5.09	5.73	3.89
45	Montgomery	Maryland	12193	31125	3.11	6.75	3.91
27	Nassau	New York	17393	51158	3.88	6.94	3.74
19	New York	New York	1249	6499	8.40	3.76	3.56
31	Oakland	Michigan	21312	55858	3.24	6.26	4.20
5	Orange	California	45006	113085	3.03	10.28	4.14
35	Orange	Florida	19902	50296	3.05	12.15	3.45
29	Palm Beach	Florida	20725	52514	3.07	6.88	3.78
21	Philadelphia	Pennsylvania	8186	33075	6.08	5.24	3.82
41	Pima	Arizona	26169	66329	3.07	8.84	3.36
47	Pinellas	Florida	16428	48076	3.85	6.42	3.87

(Rank=ranking in terms of population estimation of year 2009, n=total number of strokes (or nodes), m= total number of stroke intersections (or edges), $\langle k \rangle$=average connectivity of the stroke, L= average length of the shortest paths of road networks, D_t= structural fractal dimension)

Table 7.2 50 road networks and their structural measures (cont'd)

Rank	County	State	*n*	*m*	<*k*>	*L*	D_t
10	Queens	New York	6465	27707	6.57	6.26	3.51
11	Riverside	California	44736	113275	3.06	14.73	3.77
25	Sacramento	California	21306	56587	3.31	6.84	3.87
38	Salt Lake	Utah	18178	46742	3.14	7.71	3.54
12	San Bernardino	California	52593	145004	3.51	11.84	4.24
6	San Diego	California	51022	130500	3.12	11.02	4.11
16	Santa Clara	California	27753	71752	3.17	8.18	3.35
50	Shelby	Tennessee	19808	53854	3.44	7.13	3.81
36	St.Louis	Missouri	20275	50432	2.97	9.47	3.45
22	Suffolk	New York	31743	87473	3.51	9.13	3.66
18	Tarrant	Texas	29109	86679	3.96	6.76	3.76
42	Travis	Texas	12886	36320	3.64	8.25	3.33
13	Wayne	Michigan	18116	66230	5.31	5.30	4.15
43	Westchester	New York	14808	40432	3.46	7.37	3.41

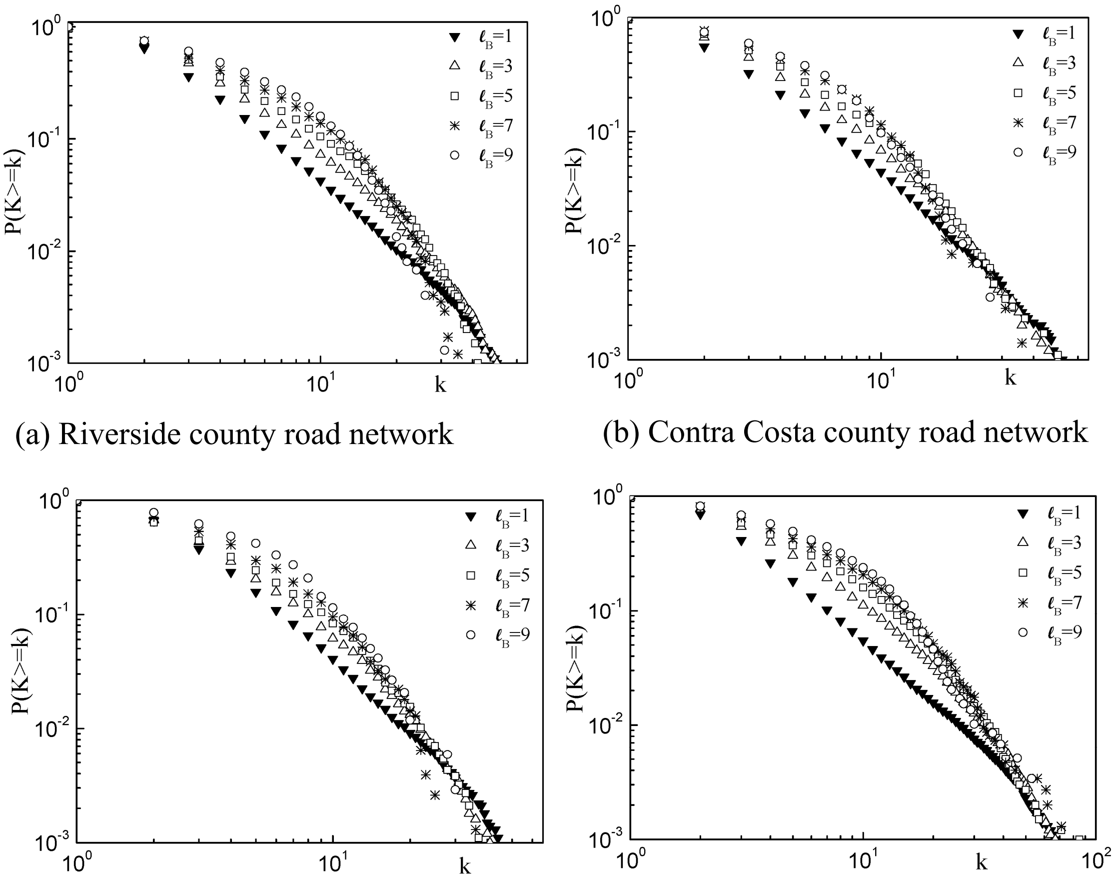
The fractal dimensions of the topological structures of 50 road networks range from 2.94 to 4.90. Amongst them, the Cook road network has the smallest structural fractal dimension of 2.94. One of the possible reasons is that Cook road network is principally composed of regular short roads, and there are many small and disperse hubs in its network representation. Therefore, the number of boxes needed to tile this network drops dramatically in the first five steps and decreases very slowly after that. For instance, N_B={10712, 2808, 874, 352, 195, 146, 118, 106, 102, 98, 96, 94}, corresponding to ℓ_B={3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 29}. This sharp dropping creates a small self-similar exponent in the log-log plot.

7.5 Structural self-similarity for road networks at different scales

As pointed out in the previous section, road networks are fractals in terms of their topological structure. It is then of interest to investigate the self-similarity of a road network at different scales. This is the topic of this section. In this part of study, two types of structural properties are studied, i.e. scale-free and small-world.

7.5.1 Scale-free road networks at different scales

It is found in Chapter 3 that all the 50 road networks are scale-free. This subsection focuses on examining the scale-free property of the structure of a road network at several scales. Figure 7.5 lists the distributions of the stroke connectivities in four county road networks. The four road networks are selected because that they need the most steps of renormalization. The *x* and *y* axes in Figure 7.6 correspond to stroke connectivities (*k*) and the cumulative probability distribution (*P(k>=k)*) respectively. In each graph, five sets of {*k*, *P(k>=k)*} values are represented, i.e. when ℓ_B=1 (solid inversed triangles), ℓ_B=3 (open inversed triangles), ℓ_B=5 (open squares), ℓ_B=7(open cross) and ℓ_B=9(open circle).



(a) Riverside county road network

(b) Contra Costa county road network

(c) Orange county (Florida) road network

(d) San Bernardino county road network

Figure 7.6 The power law distributions of the connectivities in 4 road networks at 5 scales

As described in Figure 7.5, there is a significant trend that road networks at the five scales all tend to be scale-free, despite certain disparity of the distributions of stroke connectivities at the given five length scales. For this disparity, there are two possible reasons. On one hand, as described in the box-counting approach, some details for the internal linkages of a box are lost during the renormalization process. On the other hand, the MEMB algorithm only employs boxes with odd numbers as size, i.e. ℓ_B=1, ℓ_B=3, ℓ_B=5, ℓ_B=7, etc. while the larger the box size, the simpler the structure for the renormalized networks.

7.5.2 Small-world road networks at different scales

As shown in Chapter 4, general road networks are small-world. However, the research is based on single scale, i.e., the original road network itself. In other words, one still doesn't know whether small-world structure preserves at several scales of a road network. This subsection deals with this challenge. Similarly, clustering coefficient and characteristic path length are employed to detect the small-world property. The final results for the four county road networks described in section 7.4.1 are listed in Table 7.2.

Table 7.3 The small-world structure of 4 road networks at five scales

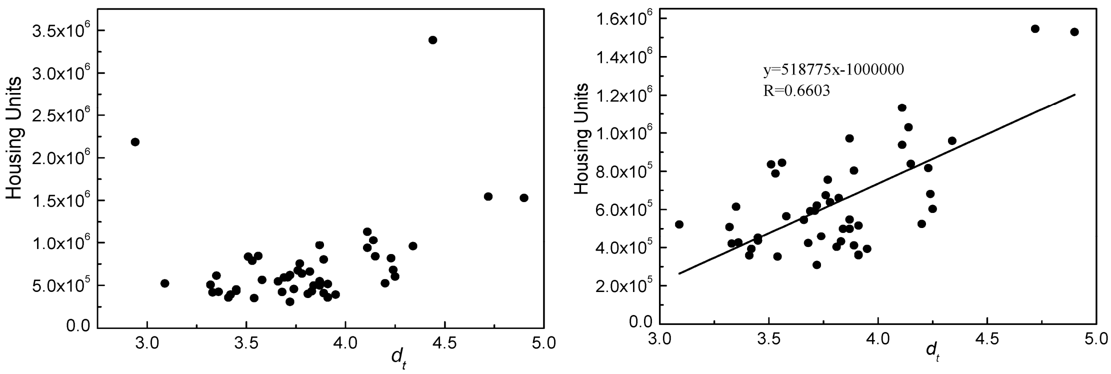
County	*scale*	*n*	*m*	*<k>*	*L*	*C*	L_{rand}	C_{rand}
Riverside	$\ell_B=1$	44736	113275	3.06	14.73	0.12	9.49	0.0001
	$\ell_B=3$	11215	54391	3.85	11.49	0.31	7.05	0.0004
	$\ell_B=5$	4101	21969	4.36	10.13	0.35	5.82	0.0011
	$\ell_B=7$	1730	9947	4.75	8.12	0.38	4.91	0.0025
	$\ell_B=9$	748	4539	5.06	7.11	0.38	4.18	0.0059
Contra Costa	$\ell_B=1$	20451	50507	2.94	12.24	0.14	9.18	0.0002
	$\ell_B=3$	5089	23465	3.61	9.93	0.26	6.70	0.0006
	$\ell_B=5$	1751	9029	4.16	8.16	0.35	5.42	0.0011
	$\ell_B=7$	714	3956	4.54	7.02	0.38	4.65	0.0037
	$\ell_B=9$	288	1596	4.53	5.60	0.37	3.87	0.0109
Orange_Florida	$\ell_B=1$	19902	50296	3.05	12.15	0.17	8.88	0.0002
	$\ell_B=3$	5004	22532	3.50	9.40	0.27	6.97	0.0008
	$\ell_B=5$	1866	8840	3.74	7.42	0.28	5.79	0.0017
	$\ell_B=7$	777	4029	4.18	6.08	0.32	4.77	0.0016
	$\ell_B=9$	340	1996	4.86	4.84	0.36	3.85	0.0126
San Bernardino	$\ell_B=1$	52593	145004	3.51	11.84	0.15	8.73	0.0001
	$\ell_B=3$	12360	71919	4.82	9.73	0.38	6.14	0.0003
	$\ell_B=5$	4034	26669	5.61	8.01	0.42	4.96	0.0019
	$\ell_B=7$	1488	10817	6.27	7.17	0.45	4.18	0.0051
	$\ell_B=9$	587	4478	6.62	5.80	0.45	3.60	0.0131

It can be found from Table 7.2 that, at the five length scales, the topological structure of all the four road networks show the small-world property, that is,

$L \geq L_{random}$ but $C >> C_{random}$. In fact, all the other 46 county road networks demonstrate small-world property consistently at different given scales.

7.6 Linking structural fractal with regional development

As discussed at the beginning of this Chapter, a substantial number of analyses have been conducted on the relationship between the geometric fractality of a geographical object and its socio-economic attributes. This section intends to explore the relationship between the structural complexity of a road network and the regional development. In this part of study, the structural complexity is described by the structural fractal dimension of the road network, and the regional development is reflected by total housing units (available from http://quickfacts.census.gov/qfd/index.html). The result is shown in Figure 7.7. When the two points with the highest housing units are treated as outliers, there is a positive relationship (r=0.66) between these two variables.



(a) original data (b) data removing the outliers

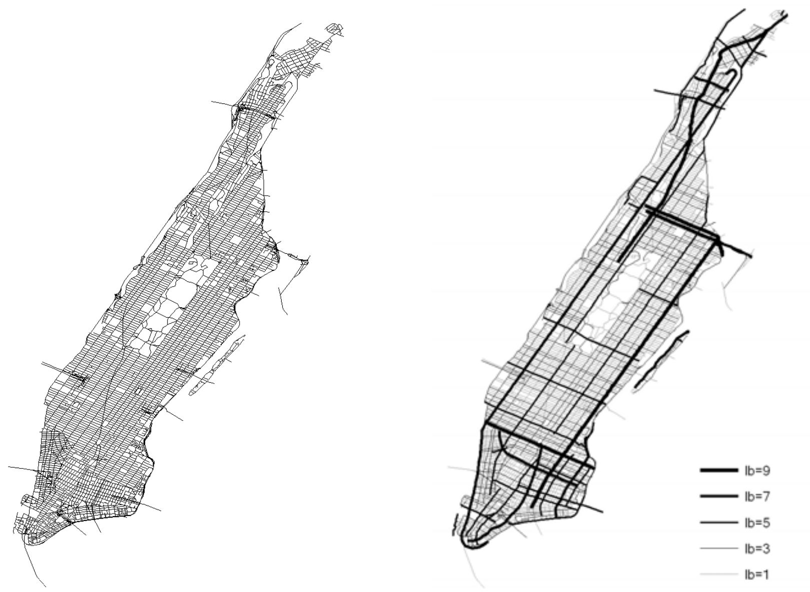
Figure 7.7 Relationship between housing units and structural fractality

The limited examination implies that structural fractal sheds certain light on describing the complex nature of the topological structure of a road network.

7.7 Discussion and summary

Primary research has shown that road networks are geometric fractals. This part of study points out that road networks are also structural fractals. The connectivity manners of roads at different scales look self-similar. At all given scales, in the topological structure of a road network, there are few roads that are far more well-connected than the other roads. And regardless of the differences in network size, the entire road network is easily accessible due to the small-world property of the road system. Based on our limited examination, it is found out that there is a close relationship between the complexity of the structure of a road network and its regional development.

The process of renormalizing a complex box may shed light on simulating urban growth from bottom up. As illustrated in Figure 7.8(b), the corresponding strokes of the central nodes at each renormalization step are highlighted. Obviously, the later stage the stroke survives at, the more important the stroke is in the road network. Indeed, by observing the structure of those central strokes and their connection statuses, one can obtain the information of the topological structure of the entire road network, which provides guidance for simulating urban dynamics.



(a) (b)

Figure 7.8 Road network in New York County and its skeletons at 5 scales

More importantly, according to Batty (2005) and Hillier (2009), there are ‘strong order and a pattern that emerges from the myriad of decisions and processes required for a city to develop and expand physically’. city is succedded in keeping a good balance between local and global organization to get a sustainable form. The discovery of structural fractality of road networks gives important hints to deal with this problem. In addition, as found by West *et al*. (1999), fractal-like networks can effectively endow life as the fractal structure or allometric scaling of organisms maximizes the internal efficiency, by minimizing the scaling of transport distances and times. In a similar fashion, if city is treated as a life system like human body.

Roads are blood vessels. The road system is constrained by the shape and volume of the city. Urban structure has to evolve in a fractal way in order to be robust in transportation and efficient to maximize its internal flow transmission. The scale-invariant structure including hierarchical branching and shortcuts, may help achieve this goal.

Further work will be carried out to investigate the relationship between structural fractal of road networks and socio-economic and environmental factors. Simulating city growth based on road skeleton and structural fractal is also of interest. Besides, study urban complexity and information from the aspect of structural fractality of urban space is interesting work.

Chapter 8 Conclusions and recommendations

8.1 Summary

Road network has been a central subject of urban study not only because that it forms the skeleton of an urban system, but also because that it plays an important role in human life. Recently, increasing attention has been devoted to road networks from the emerging new science of networks, partly because that a road network is a common and easily accessible spatial complex system. As this study has shown, investigating the structure of a road network with the aid of complex network analysis provides new and useful insight for understanding the universal organizational principles and evolutionary rules of road networks.

In order to investigate the structure of a road network, an original road map should be first converted into an abstract graph with nodes and links. Methodologies and measures rooted in various disciplines were then either employed or newly developed to investigate the structure of a road network. In total, five kinds of complex network analysis have been carried out and discussed as below:

The first analysis focuses on the shortcuts created by the mixing pattern in a scale-free road network. A scale-free network is a network in which the distribution of node connectivities follows a power law. In order to examine the scale-free property of road networks, the revised Kolmogorov-Smirnov statistics was introduced to replace conventional least square method as it provides more reliable and robust statistics for power law distribution. The profile of connectivity correlation probability was firstly introduced to road network study to visualize the mixing patterns of road connectivities. At the end, the hypothesis that the mixing patterns in a hierarchical road network create shortcuts has been tested.

The second analysis concerns the small-world structure of road networks. Small-world means that any two nodes in a network can be connected by a relatively short connection chain. In examination of the small-world structure, two measures, i.e., clustering coefficient and characteristic path length were introduced. The possible influences of small-world structure on regional development were also addressed.

The third part of study attempts to form the hierarchical structure of a road network. Ego network analysis rooted in social science was first employed to describe the opportunities and constraints of each node in the transfer of flow to the neighbour(s) of a node. A new measure named Centrality Rank (CR) was developed to quantify the order of each node. Ego network analysis was then applied to a sampled road network and two real-life road networks. At the end, traffic flow data gathered from Global Position System (GPS) data was utilized as a benchmark to evaluate the feasibility and validity of ego network analysis in forming a hierarchical road network.

The ego network was further improved to become a weighted ego network by assigning a weight to each link, as there are two significant limitations in conventional ego network analysis. A new measure named Weighted Average Centrality Rank (WACR) was developed to define the order of each road. The weighted ego network analysis was then applied to the same dataset as that of the ego network analysis. Results of the two approaches were compared.

The final part of study investigates the structural fractality and self-similarity in the topological structure of a road network. The Maximum-Excluded-Mass-Burning (MEMB) algorithm rooted in statistical physics was first utilized to examine the fractality in the topological structures of road networks. Socio-economic data were collected to explore the possible applications of structural fractal.

8.2 Conclusions

Based on the limited tests, the key conclusions are drawn as below:

(1) Road networks are scale-free. The estimated exponents of power law distributions of road connectivities range from 2.11 to 3.50 with an average of 2.69;

(2) shortcuts that make a road network easily accessible are created by the mixing pattern in scale-free road networks, and the profile of connectivity correlation probability is a valid way to visualize the mixing patterns of road connectivities;

(3) road networks are small-world with an average length of the shortest paths of eight. Results also indicate that a more developed region has a more mature system to meet its transportation demands;

(4) Ego network analysis is effective in the formation of the hierarchical

structure of road networks and the weighted ego network produces improved results. The orders of roads defined by the two approaches are well consistent with the real-life pattern of traffic flow, and

(5) road networks are structural fractals. The structural fractal dimensions of studied road networks vary from 2.94 to 4.90. Results indicate that a developed region tends to have a more complex road network.

Stated briefly, results obtained from this study show that road networks are self-organized with certain universal principles, e.g. scale-free, mixing pattern, small-world, hierarchical and fractal. Such structures help meet various traffic demands in the region, and maximize the ability of the traffic flow accommodation of the road network. Therefore, in practice, when building more roads is constrained by the geographical environment, a scientifically designed road network with certain topological properties helps meet the traffic demands of the region. More importantly, the complex network analysis introduced in this study is also suitable for other kinds of networks, for instance, economic cooperation network, telephone communication network, population migration network and airplane network.

8.3 Potential of this work

The potential of this work can be described as follows:

Firstly, this work provides a systematic framework to describe the structural properties of a road network. The same approach can be applied to other kinds of urban networks, e.g. economic cooperation networks, human migration networks and telephone networks.

Secondly, the findings of scale-free, small-world, hierarchical, structural fractal and self-similar properties provide support to the idea that a city is a Complex Adaptive System (CAS) and Self-Organizing System (SOS). Indeed, nature tend to organize in an efficient structure with the aid of small-world and scale-free properties at both single and multiple scales. This is because that in a scale-free road network, there are few roads with extremely large values of connectivities. Such roads form the backbone of a road network and play important roles in flow transmission. With the aid of small-world structure, a road network becomes compact and the efficiency of flow transmission is improved because a small-world road network has a short characteristic path length and a high local clustering coefficient. Regarding the fractal structure, it has been found to be the optimal way to build a sustainable city (Chen 2005), which reveals the strong order and pattern emerging from the myriads of decisions and processes required for a city to develop and expand physically (Batty 2005; Hillier 2009).

The third potential is in the simulation of the spatial-temporal dynamics of both urban systems and transportation networks. During recent years, a series of models have been built to simulate urban dynamics based on cellular automata, spatial epidemic, diffusion-limited aggregation, and dielectric breakdown models (Fotheringham, Batty, and Longley 1989; Batty and Longley 1994; Batty 2005). Results of these models show that urban systems form and evolve in a fractal way to maximally fill space. However, all these models are limited to geometric morphology and the structure of the systems is ignored to some extent. In this sense, geometrical fractal, structural fractal, small-world and scale-free properties should all be taken into account when simulate the growth of an urban system.

Lastly but not leastly, at a much higher level, structural fractality blurs the borders of many disciplines including statistics, physics, geometry, geography, computer science, social science, and network science. For instance, small-world property observed in social networks holds true for a geographical network at several scales; scale-free property developed in statistics has been used to characterize the structure of a road network; structural fractal dimension is a measurement of complexity, which can be compared to entropy in thermodynamics based on the probability distribution of the values of structural properties. To go a step further, West, Brown, and Enquist (1999) have claimed that the fractal geometry or allometric scaling of organisms is the fourth dimension of life. If this is true, structural fractality may be the fifth dimension of the life as such a structure tends to maximize the internal efficiency by forming hierarchies and minimize the traffic distance with shortcuts. To deal with this challenge, space, scale, and structure needed to be united in a framework of maximum entropy or maximization of utility, which is a new frontier of urban study (Batty 2010).

8.4 Limitations and future work

The main constraint of this study is that it only considered one graph representation of road networks, i.e., stroke. Other representations listed in Chapter 2 were not taken into account. Further evaluation of the representation approaches is of interest.

The second limitation is that the investigation on the structural fractality of road networks is at an early stage. More data are needed to be collected to detect the relationship between structural fractal dimension, information, entropy and regional development. Another future study will concern analyzing information transfer, energy consumption and the dynamic evolution of a city based on the structural properties discovered in this study.

References

Abbott A., 2006. Fractals and art: in the hands of a master. *Nature*, 439, pp.648-650.

Alexander C., 1965. A city is not a tree. *Architectural Forum*, 122(1+2), pp.58-62.

Amaral A. N., Scala A., Barthélémy M. and Stanley H. E., 2000. Classes of small world networks. *Proceedings of National Academy of Science of the United States of America*, 97(21), pp. 11149-11152.

Amaral L.A.N. and Ottino J.M., 2004. Complex networks: augmenting the framework for the study of complex systems. *The European Physical Journal B*, 38(2), pp.147-162.

Anil P.N. and Natarajan Dr.S., 2010. Automatic road extraction from high resolution imagery based on statistical region merging and skeletonization. *International Journal of Science and Technology*, 2(3), pp.165-171.

Aragones J. and Arredondo J., 1985. Structure of urban cognitive maps. *Journal of Environmental Psychology*, 5, pp.197-212.

Avelar S. and Muller M., 2000. Generating topologically correct schematic maps. *Proceedings of the 9th International Spatial Data Handling*, pp.4a.28-4a.35.

Aw A. and Rascle M., 2000. Resurrection of "second order" models of traffic flow. *SIAM Journal of Applied Math*, 60(3), pp.916-938.

Banai R., 1999. A methodology for the image of the city. *Environmentand Behavior B: Planning and Design*, 26(1), pp. 133-144.

Barabási A.-L. and Albert R., 1999. Emergence of scaling in random networks. *Science*, 286(5439), pp.509-512.

Barabási A.-L., Albert R., and Jeong H., 2000. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A: Statistical Mechanics and its Applications*, 281(1-4), pp. 69-77.

Baran P.K., Rodriacuteguez D.A. and Khattak A.J., 2008. Space syntax and walking in a new urbanist and suburban neighbourhoods. *Journal of Urban Design*, 13(1), pp. 5-28.

Bar-Ilan J., 2006. An ego-centric citation analysis of the works of Michael O.Rabin based on multiple citation indexes. *Information Processing and Management*, 42(6), pp. 1553-1566.

Bastoky G., 2010. *Fundamentals of Document Design for the Technical Writer*. http://www.bastoky.com/docdesign_home.htm (last accessed on 28th, October 2010).

Batty M., 2004. A new theory of space syntax. *CASA Working Paper*, Available online at: http://eprints.ucl.ac.uk/211/

Batty M. and Longley P. A., 1994. *Fractal Cities: A Geometry of Form and Function*. Academic Press, London. 394 pp.

Batty M. and Xie Y., 1999. Self-organized criticality and urban development. *Discrete Dynamics in Nature and Society*, 3(2-3), pp.109-214.

Batty M., 2005. *Cities and Complexity: Understanding Cities Through Cellular Automata, Agent-Based Models, and Fractals*. MIT Press, Cambridge. 589 pp.

Batty M., 2008. The size, scale, and shape of cities. *Science*, 319(5864), pp.769-771.

Bjøke J.T., 2005, Map generalization of road networks. In: *Visualisation and the Common Operational Picture*. pp. 17-1—17-8.

Boguñá M., Krioukov D. and Claffy K.C., 2009. Navigability of complex networks. *Nature Physics*, 5, pp. 74-80.

Bondy J.A. and Murty U.S.R., 2008. *Graph Theory*. Springer, New York. 651 pp.

Brown C. T., Witschey W. R. T. and Liebovitch, L. S. 2005. The broken past: fractals in Archaeology. *Journal of Archaeological Method and Theory*, 12 (1), pp.37-78.

Buhl J., Gautrais J., Reeves N., Sole R. V., Valverde S., Kuntz P. and Theraulaz G., 2006. Topological patterns in street networks of self-organized urban settlements, *The European Physical Journal B*, 49(4), pp.513-522.

Burt R.S., 1995. *Structual Holes, The Social Structure of Competition*. Harvard University Press (Cambridge, Massachusetts, and London, England: Harvard University), paperback edition, 324pp.

Buhl J., Gautrais J., Reeves N., Sole R. V., Valverde S., Kuntz P. and Theraulaz G., 2006. Topological patterns in street networks of self-organized urban settlements. *The European Physical Journal B*, 49(4), pp.513-522.

Buttenfield B.P., 1989. Scale-dependence and self-similarity in cartographic lines. *Cartographica*, 26(1), pp.79-100.

Cardillo A., Scellato S., Latora V. and Porta S., 2006. Structural properties of planar graphs of urban street patterns. *Physical Review E*, 73(6), pp. 066107.01-066107.08.

Carr S., and Schissler D., 1969. The city as a trip: perceptual selection and memory in the view from the road. *Environment and Behavior*, 1(1), pp.7-35.

Chen J., Hu Y. G., Li Z.L., Zhao R.L. and Meng L.Q., 2009. Selective omission of road features based on mesh density for automatic map generalization. *International Journal of Geographical Information Science*, 23(8), pp. 1013-1032.

Chen, Y. G. 2005. Fractal city and urban planning (In Chinese). *City Planning Review,* 29 (2), pp.33-41.

Chester L., Arnold Jr. and Gibbons C. J., 1996. Impervious surface coverage: the emergence of a key environmental indicator. *Journal of the American Planning Association*, 62(2), pp.243-258.

Chiaradia A., Hillier B. and Barnes Y. and Schwander C., 2009. Residential property value patterns in London: space syntax spatial analysis. In: *Proceedings of the 7th International Space Syntax Symposium*, Stockholm, Sweden, pp. 015.01-015.012.

Chirapiwat T., 2005. Street configurations and commercial and mixed-use land-use patterns: a morphological study of the northeastern region of bangkok to evaluate recent transportation and land-use plans. Ph.D. Thesis. University of Michigan.

Church R.L. and Murray A. T., 2009. *Business Site Selection, Location analysis and GIS*. Wiley, New York. 320 pp.

Clauset A., Rohilla C. and Newman M. E. J., 2009. Power law distributions in empirical data. *SIAM Review*, 51(4). pp.661-703.

COMSIS Coporation, 1995. Report submitted to Federal Highway Administration, U.S. Department of Transportation, DTFH61-95-C-00017.

Croxford B., Penn A. and Hillier B., 1996. Spatial distribution of urban pollution:

civilizing urban traffic. Science of The Total Environment, 189-190, pp. 3-9.

Crucitti P., Latora V. and Porta S., 2006. Centrality measures in spatial networks of urban streets. *Physical Review E*, 73(3), pp.036125.01-036125.05.

Dalton R.C., 2003. The secret is to follow your nose: route path selection and angularity. *Environment and Behavior*, 35(1), pp 107-131.

Davies J., 2002. Visual abstraction in analogical problem solving: a dissertation proposal, available online at: http://www.jimdavies.org/research/visual-analogy/proposal/proposal.html (accessed on 28th, October 2010)

Deng H. Y., Wu F., Zhai R. J. and Liu W.W., 2005. A generalization model of road networks based on genetic algorithm. *Geomatics and Information Science of Wuhan University*, 31(2), pp.164-167.

Dodds P. S., Muhamad R. and Watts D.J., 2003. An experimental study of search in global social networks. *Science*, 301(5634), pp.827-829.

Dong W., Wang B. and Duan Z., 2007. Expansion of ürümqi urban area and its spatial differentiation. *Science in China D (Earth Sciences)*, 50(z1), pp.159-168.

Edwardes A.J. and Mackaness W., 2000. Intelligent generalization of urban road networks. In: *GIS Research UK Conference*, York, UK, pp.81-85.

Elroi D.S., 1998. Designing a network line map schmatization software enhancement package. Proceedings of the 8th International ESRI User Conference, available online at: http://www.elroi.com/papers_ESRI_1988/ESRI88.pdf (accessed on 28th, October 2010)

Everett M. and Borgatti S.P., 2005. Ego network betweenness. *Social Networks*, 27(1), pp. 31-38.

Fan F.L., Wang Y.P., Qiu M.H. and Wang Z.S., 2009. Evaluating the temporal and spatial urban expansion patterns of Guangzhou from 1979 to 2003 by remote sensing and GIS methods. *International Journal of Geographical Information Science*, 23(11), pp.1371-1388.

Feder J., 1988. *Fractals*. Plenum Press, New York, 283 pp.

Forman R. and Alexander L., 1998. Roads and their major ecological effect. *Annual Review of Ecology and Systematics*, 29, pp.207-231.

Freeman L.C., 1979. Centrality in social networks: conceptual clarification. *Social Networks*, 1(3), pp. 215–239.

Freeman L.C., 1982. Centered graphs and the structure of ego networks. *Mathematical Social Science*, 3(3), pp. 291-304.

Gao Z.H., Li M.C., Chen Z.J., Li F.X. and Sun D., 2007. Research on small world characteristics of urban road network. *Geography and Geo-Information Science*, 23(4), pp.97-101.

García Balboa J. L. and Ariza López F. J. 2008. Generalization-oriented line classification by means of an artifical neural network. *Geoinformatica*, 12(3), pp. 289-312.

Gilkerson J., Li W. and Owen D. 2005. Traffic Simulation, available online at: www.mgnet.org/~douglas/Classes/cs521/**traffic/Traffic**2005.ppt (accessed on 28th, October 2010)

Golledge R.G., 1978. Representing, interpreting, and using cognized environments. *Papers of the Regional Science Association*, 41(1), pp.169-204.

Golledge R.G., 1995. Path selection and route preference in human navigation: A progress report. *Lecture Notes in Computer Science*, 988, pp. 207-222.

Golledge R.G., 2002. Cognitive Maps. In: *Encyclopedia of Social Measurement*, pp.329-339.

Goh K.-I., Salvi G., Kahng B. and Kim D., 2006. Skeleton and fractal scaling in complex networks. *Physical Review Letters*, 96(1), pp.018701.01-018701.04.

Goodchild M.F., 1980. Fractals and the accuracy of geographical measures. *Mathematical Geology*, 12(2), pp.85-98.

Goodchild M. F. andMark. D. M. 1987. The fractal nature of geographic phenomena. *Annals of the Association of American Geographers*,77 (2), pp.265-278.

Gruen A. and Li H.H., 1995, Road extraction from aerial and satellite images by dynamic programming. Journal of Photogrammetry and Remote Sensing, 50(4), pp.11-20.

Guimera R., Danon L., Diaz-Guilera A., Giralt F. and Arenas A., 2003. Self-similar community structure in organizations. *Physical Review E*, 68(6), pp. 065103.01-065103.04.

Guimera R., Mossa S., Turtschi A. and Amaral L., 2005. The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences of the United States of America*, 102(22), pp.7794-7799.

Haken H., 1983. *Synergetics, an Introduction: Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry, and Biology* (Tthird Revised and Enlarged Edition). Springer-Verlag: New York. 371pp.

Haken H. and Portugali J., 2003. The face of the city is its information. *Journal of Environmental Psychology*, 23(4), pp. 385-408.

Hausdorff F., 1919. Dimension and outer measure. *Mathematical Annuals*, 79, pp.157-179.

Helbing D., Hennecke A., Shvetsov V. and Treiber M., 2001. MASTER: macroscopic traffic simulation based on a gas-kinetic, non-local traffic model. *Transportation Research, Part B: Methodological*, 35(2), pp.183-211.

Hillier B., 1999. *Space is the Machine: A Configurational Theory of Architecture*. Cambridge University Press, Cambridge. 480 pp.

Hillier B., 2009. Spatial sustainability in cities: organic patterns and sustainable forms. In: *Proceedings of the 7th International Space Syntax Symposium*, Stockholm, Sweden, pp. k01.01-k01.20.

Hillier B. and Hanson J., 1984. *The Social Logic of Space*. Cambridge University, Cambridge. 296 pp.

Hillier B., Hanson J., Peponis J., Hudson J. and Burdett R., 1983. Space syntax. *Architects Journal*, 178(48), pp.67-75.

Hillier B. and Iida S., 2005. Network and psychological effects in urban movement. In: *Lecture Notes in Computer Science 3693*, pp.475-490.

Hillier B. and Sahbaz O., 2008. An evidence based approach to crime and urban design, or, can we have vitality, sustainability and security all at once? Bartlett *School of Graduate Studies Report,* Available online at: http://www.spacesyntax.com/Files/MediaFiles/Hillier%20Sahbaz_An%20evidence%20based%20approach_010408.pdf (last accessed on 28th, October 2010).

Horton F. E. and Reynolds D. R., 1971. .Effects of urban spatial structure on individual behavior. *Economic Geography*, 47, pp.36-48.

Huynh N.T., Hall G. B., Doherty S. and Smith W.W., 2008. Interpreting urban space through cognitive map sketching and sequence analysis. *The Canadian Geographer*, 52(2), pp.222-240.

Jiang B., 2007. A topological pattern of urban street networks: universality and peculiarity. *Physica A: Statistical Mechanics and its Applications*, 384(2), pp. 647-655.

Jiang B., 2008a. Flow dimension and capacity for structuring urban street networks. *Physica A: Statistical Mechanics and its Applications*, 387(16-17), pp. 4440-4452.

Jiang B., 2008b. Street hierarchies: a minority of streets account for a majority of traffic flow. *International Journal of Geographical Information Science*, 23(8), pp.1033-1048.

Jiang B. and Claramunt C., 2002. Integration of Space Syntax into GIS: new perspectives for urban morphology. *Transactions in GIS*, 6(3), pp.295-309.

Jiang B. and Claramunt C., 2004. A structural approach to the model generalization of an urban street network. *Geoinformatica*, 8(2), pp. 157-171.

Jiang B. and Harrie L., 2004. Selection of streets from a network using self-organizing maps. *Transactions in GIS*, 8(3), pp. 335-350.

Jiang B., Zhao S.J. and Yin J.J., 2008. Self-organized natural roads for predicting traffic flow: a sensitivity study. *Journal of Statistical Mechanics: Theory and Experiment*, 7. P07008.01- P07008.27.

Joslyn C. and Rocha L., 2000. Towards semiotic agent-based models of socio-technical organizations, Proc. AI, Simulation and Planning in High Autonomy Systems (AIS 2000) Conference, Tucson, Arizona, pp. 70-79.

Kalapala V., Sanwalani V., Clauset A. and Moore C., 2006. Scale invariance in road networks. *Physical Review E*, 73(2), pp. 026130.01-026130.06.

Kang M.W., Schonfeld P. and Yang N., 2009. Prescreening and Repairing in a Genetic Algorithm for Highway Alignment Optimization, *Computer-Aided Civil and Infrastructure Engineering*, 24(2), pp. 109-119.

Kesting A., Treiber M. and Helbing D. 2008. Agents for traffic simulation. Available online at: http://arxiv.org/PS_cache/arxiv/pdf/0805/0805.0300v1.pdf (accessed on 28th, October 2010)

Kim H.-K. and Sohn D.W., 2002. An analysis of the relationship between land use density of office buildings and urban street configuration—case studies of two areas in Seoul by space syntax analysis. *Cities*, 19(6), pp.409-418.

Kim J.S., Goh K.-I. and Kahng B., 2007a. A box-counting algorithm for fractal scaling in scale-free networks. *Chaos*, 17(2), pp.026116.01-026116.06.

Kim J.S., Goh K.-I., Kahng B. and Kim D., 2007b. Fractality and self-similarity in scale-free networks. *New Journal of Physics*, 9(6), pp.177.01-177.12.

Kleinberg, J. M. 2000. Navigation in a small world. *Nature*, 406, pp.845.

Kotusevski G. and Hawick K.A. 2009. A review of traffic simulation software, available online at: http://www.massey.ac.nz/~kahawick/cstn/095/cstn-095.pdf (accessed on 28th, October 2010)

Kulash W., 2006. Hierarchy of streets and roads. In: *Planning and Urban Design Stands*, pp.226-228.

Lämmer S., Gehlsen B. and Helbing D., 2006, Scaling laws in the spatial structure of urban road networks. *Physica A*, 363(1), pp. 89–95.

Latora V. and Marchiori M., 2002. Is the Boston subway a small-world network? *Physica A: Statistical Mechanics and its Applications*, 314(1-4), pp.109-113.

Li W.J. and Wu H.H., 2005. Fractal attenuation analysis of cartographic object's self-similarity on cartographic generalization (In Chinese). *Geomatics and Information Science of Wuhan University*, 30(4), pp.309-312.

Li X., He J.Q. and Liu X. P., 2009. Ant intelligence for solving optimal path-covering problems with multi-objectives. *International Journal of Geographical Information Science*, 23(7), pp. 839-857

Li Z.L., 2006. *Algorithmic Foundation of Multi-Scale Spatial Representation*. CRC Press, Boca Raton. 280 pp.

Li Z.L. and Choi Y. H., 2002. Topographic map generalization: Association of road elimination with thematic attributes. *The Cartographic Journal*, 39(2), pp. 153-166.

Li Z.L. and Dong W.H., 2010. A stroke-based method for automated generation of schematic network maps. *International Journal of Geographical Information Science*, Preprint. Available online at: http://www.informaworld.com/smpp/1377391149-98458866/content~db=all~content=a927307136~frm=titlelink (last accessed on 28th, October 2010)

Li Z.L. and Huang P., 2002. Quantitative measures for spatial information of maps, *International Journal of Geographic Information Science*, 16(7), pp.699-709.

Longley P.A. and Batty M., 1989. Fractal measurement and line generalization. *Computers and Geosciences*, 15(2), pp.167-183.

Lu Y.M. and Tang J.M., 2004. Fractal dimension of a transportation network and its relationship with urban growth: a study of the Dallas-Fort Worth area. *Environment and Planning B: Planning and Design*, 31(6), pp.895-911.

Lynch K., 1960. *The Image of the City*. MIT Press: Cambridge. 202 pp.

Mackaness W. A., 1995. Analysis of urban road networks to support cartographic generalization. *Cartography and Geographic Information Systems*, 22(4), pp. 306-316.

Mandelbrot B.B., 1967. How long is the coast of Britain? Statistical self-similarity and fractal dimension. *Science*,155(3775), pp. 636-638.

Mandelbrot B.B., 1982. *The Fractal Geometry of Nature*. W.H. Freeman and Company, New York. 498 pp.

Marr D., 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Freeman: San Francisco. 397pp.

Maslov S. and Sneppen K., 2002. Specificity and stability in topology of protein networks. *Science*, 296(5569). pp.910-913.

Matthews J.W. and Turnbull G. K., 2007. Neighbourhood street layout and property value: the interaction of accessibility and land use mix. *The Journal of Real Estate Finance and Economics*, 35(2), pp. 111-141.

Mcnally J. G. and Mazza D., 2010. Fractal geometry in the nucleus. *The EMBO Journal*, 29(1), pp.2-3.

Milgram S., 1967. Small-world problem. *Psychology Today*, 1(1), pp.61-67.

Monasson R., 1999. Diffusion, localization and dispersion relations on “small-world” lattices. *The European Physical Journal B*, 12(4), pp.555-567.

Morikawa T., Miwa T., Kurauchi S., Yamamoto T. and Kobayashi K. 2005. Driver’s route choice behavior and its implications on network simulation and traffic assignment. *Simulation Approaches in Transportation Analysis*, 31, part IV, pp.341-369.

Morphet R. 2010. Thermodynamic potentials and phase change for transport systems. *CASA working paper*. available online at: http://www.casa.ucl.ac.uk/working_papers/paper156.pdf (accessed on 28th, October 2010)

Muller J.C., 1987. Fractal and Automated Line Generalization. *The Cartographic Journal*, 24(1), pp.27-34.

Mureika, J. R., 2007. Beyond fractals. *Nature* 445, pp.262

Mustiere S., 2005. Cartographic generalization of roads in a local and adaptive approach: a knowledge acquisition problem. *International Journal of Geographical Information Science*, 19(8-9), pp. 937-955.

Newman M. E. J., 2001. Scientific collaboration networks: I. Network construction and fundamental results. *Physical Review E*, 64(1), pp. 016131.01-016131.08.

Newman M.E.J., 2002. Assortative mixing in networks. *Physical Review Letter*, 89(20), pp.208701.01-208701.04.

Newman, M.E.J. 2003a. Mixing patterns in networks. *Physical Review E*, 67(2), pp.026126.01-208701.013.

Newman M.E.J., 2003b, Ego-centered networks and the ripple effect. *Social Networks*, 25(1), pp.83-95.

Newman M.E.J., 2010. *Networks: an Introduction*. Oxford University Press, Oxford, NewYork, 772pp.

Newman M.E.J. and Watts D.J., 1999. Renormalization group analysis of the small-world network model. *Physics Letters A*, A 263(4-6), pp.341-346.

Nicolis G. and Prigogine I., 1977. *Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order through Fluctuations*. John, Wiley & Sons, New York. 512 pp.

Nooy W.D., MrvarA. and Batagelj V., 2005. *Exploratory Social Network Analysis with Pajek*. Cambridge University Press, New York. 334 pp.

Odella F., 2006. Using ego-network in surveys: methodological and empirical research issues. In: *International Conference on Network Science*, Indiana University, Bloomington, USA, Available online at http://vw.indiana.edu/netsci06/conference/Odella_Using.pdf (last accessed on 28th, October 2010).

Parvin A., Arlen Y. and Jia B., 2007. Multilevel pedestrian movement: does visibility make any difference? In: *Proceeding of the 6th International Space Syntax Symposium*, Istanbul, Turkey. Available online at: http://www.spacesyntaxistanbul.itu.edu.tr/papers%5Clongpapers%5C040%20-%20Parvin%20Ye%20Jia.pdf (last accessed on 28th, October 2010).

Pastor-Satorras R., Vázquez A. and Vespignani A., 2001. Dynamical and correlation properties of the Internet. *Physical Review Letter*, 87(25), pp.258701.01-258701.04.

Peitgen H-O. and Saupe D., 1985. *The Science of Fractal Images*. Springer-Verlag New York Incoporation, New York. 312 pp.

Pool I.D.S. abd Kochen M., 1978. Contacts and influence. *Social Networks*, 1(1), pp. 5-51.

Porta S., Crucitti P. and Latora V., 2006a. The network analysis of urban streets: a primal approach. *Environment and Planning B: Planning and Design*, 33(5), pp.705-725.

Porta S., Crucitti P. and Latora V., 2006b. The network analysis of urban streets: a dual approach. *Physica A: Statistical Mechanics and its Applications*, 369(2), pp. 853-866.

Porta S., Latora V., Wang F.H., Strano E., Cardillo A., Scellato S., Iacoviello V. and Messora R., 2008. Street centrality and densities of retail and services in Bologna, Italy. *Environment and Planning B: Planning and Design*, 36(3), pp.450-465.

Portugali J., 2000. *Self-Organization and the City*. Springer, Berlin. 352 pp.

Prigogine I. and Andrews F. C., 1960. A Boltzmann like approach for traffic flow.

Operations Research, 8(789), pp.789-797.

Ratti C., 2004, Space syntax: some inconsistencies. *Environment and Planning B: Planning and Design*, 31(4), pp.487-499.

Richardson, L.F. 1961. The problem of contiguity: an appendix of statistics of deadly quarrels. *General Systems Yearbook 6*, pp.139-187.

Richter K.-F. 2009**.** Adaptable path planning in regionalized environments, In: *Proceedings of the 9th International Conference on Spatial Information Theory*, pp. 453-470.

Roberto C., Maria C.G. and Paolo R., 2002. Urban mobility and urban form: the social and environmental costs of different patterns of urban expansion. *Ecological Economics*, 40(2), pp.199-216.

Rocco R. and Nes A. V., 2005. The location of advanced producer services and urban change: a space syntax approach. In: *Proceedings of the 5th International Space Syntax Symposium*, In: *Proceedings of 5th International Space Syntax Symposium*. pp. 805-808.

Rodriguez-Iturbe I. and Rinaldo A., 2001. *Fractal River Basins: Change and Self-Organization*. Cambridge University Press, Cambridge. 547 pp.

Rosvall M., Grölund A., Minnhagen P. and Sneppen K., 2005a. Searchability of networks. *Physics Review E*, 72(4), pp.046117.01–046117.08.

Rosvall M., Trusina A., Minnhagen P. and Sneppen K., 2005b. Networks and cities: an information perspective. *Physical Review Letters*, 94(2), pp.028701.01-028701.04.

Salingaros N.A., 2003. Connecting the fractal city. Keynote speech, 5th Biennial of towns and town planners in Europe, Barcelona, April 2003.

Scott J.M. and Rodgers P., 2005. Automatic metro map design techniques . *Proceedings of the XXII International Cartographic Conference (ICC2005)*. CD room.

Sen P., Dasgupta S., Chatterjee A., Sreeram P.A., Mukherjee G. and Manna S.S., 2003. Small-world properties of the Indian railway network. *Physical Review E*, 67(3), pp.036106.01-036106.05.

Sharma A., Vyas V. and Deodhare D., 2006. An algorithm for site selection in GIS based on swarm intelligence. In: *IEEE Congress on Evolutionary Computation*, pp. 1020-1027.

Shen G. Q., 2002. Fractal dimension and fractal growth of urbanized areas. *International Journal of Geographical Information Science*, 16(5), pp.419-437.

Siegel A. and White S., 1975. The development of spatial representations of large-scale environments. In: *Advances in Child Development and Behavior*, pp.9-55.

Simmel G., 1950. Individual and society. In: *The Sociology of Georg Simmel*, pp. 402-408.

Snellen D., Borgers A. and Timmermans H., 2002. Urban form, road network type, and mode choice for frequently conducted activities : a multilevel analysis using quasi-experimental design data. *Environment and Planning A*, 34(7), pp. 1207-1220.

Song C.M., Havlin S. and Makse H. A., 2005. Self-similarity of complex networks. *Nature*, 433, pp.392–395.

Song C.M., Havlin S. and Makse H. A., 2006. Origins of fractality in the growth of complex networks. *Natural Physics,* 2, pp.275-281.

Song C.M., Gallos L.K., Havlin S. and Makse H. A., 2007. How to calculate the fractal dimension of a complex network: the box covering algorithm. *Journal of Statistical Mechanics: Theory and Experiment*, 27(3), P03006.

Stanley H.E., 1971. *Introduction to Phase Transitions and Critical Phenomena*. Oxford Press, Oxford and New York. 336 pp.

Strogatz S.H., 2001. Exploring complex networks. *Nature*, 410, pp. 268-276.

Tango F., Montanari R. and Marzani S., 2007. Present and future of simulation traffic models. *Modeling Driver Behavior in Automotive Environments*, VIII, pp.400-427.

Thomson R.C., 2003. Bending the axial line: smoothly continuous road centre-line segments as a basis for road network analysis. In: *Proceedings of the 4th Space Syntax International Symposium*, pp. 50.01–50.10.

Thomson R.C., 2006. The ‘stroke’ concept in geographic network generalization and analysis. *Progress in Spatial Data Handling*, 11, 681–697.

Thomson R. C. and Richardson D. E., 1995. A graph theory to road generalization. In: *Proceedings of the 17th International Cartographic Conference*, pp. 1871-1880.

Tolman E.C., 1948. Cognitive maps in rats and men. *Psychological Review*, 55(4), pp.189-208.

Tomko M., Winter S. and Claramunt C., 2008. Experiential hierarchies of streets. *Computers, Environment and Urban Systems*, 32(1), pp. 41-52.

Travençolo B.A.N. and Costa L.da. F., 2008. Accessibility in complex networks. *Physics Letter A*, 373(1), pp. 89-95.

Vehovar V., Manfreda K.L., Koren G. and Hlebec V., 2008. Measuring ego-centered social networks on the web: questionnaire design issues. *Social Networks*, 30(3), pp. 213-222.

Vosselman G. and Knecht J. de., 1995. Road tracing by profile matching and Kalman filtering. *Proceedings Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Images*, pp. 265-274.

Wagner R., 2008. On the metric, topological and functional structures of urban networks. *Physica A: Statistical Mechanics and its Applications*, 387(8-9), pp. 2120–2132.

Wang H., Kearney J., Cremer J. and Willemsen P., 2005. Steering behaviors for autonomous vehicles in virtual environments. In *Proceedings of IEEE virtual Reality Conference*, pp.155-162.

Wang P.F., Wang L., Feng X.Z. and Xiao P.F., 2009. Review of road extraction from remote sensing images (In Chinese). *Remote Sensing Technology and Application*, 24(3), pp.284–290.

Watts D. J. and Strogatz S.H., 1998. Collective dynamics of small-world networks. *Nature*, 393, pp. 440-442.

West G. B., Brown, J. H. and Enquist B.J., 1999. The fourth dimension of life: fractal geometry and allometric scaling of organisms. *Science*, 284(5420), pp.1677-1679.

White J. G., Southgate E., Thomson J. N. and Brenner S., 1986. The structure of the nervous system of the nematode Caenorhabditis elegans. *Philosophical Transactions of the Royal Society of London*, 314(1165), pp. 1-340.

Willemsen P., Kearney J. and Wang H., 2006. Ribbon networks for modeling navigable paths of autonomous agents in virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, 12(3), pp.331-342.

Xu Z.W. and Sui D.Z., 2007. Small-world characteristics on transportation networks: a perspective from network autocorrelation. *Journal of Geographical Systems*, 9(2), pp.189-205.

Yu C., Lee J. and Munro-stasiuk M.J., 2003. Extensions to least-cost path algorithms forroadway planning. *International Journal of Geographical Information Scien*ce, 17(4), pp. 361-376.

Zhan F. B., 1997. Three fastest shortest path algorithms on real road networks: data structures and procedures. *Journal of Geographic Information and Decision Analysis*, 1(1), pp. 69-82.

Zhang H. M., 2002. A non-equilibrium traffic model devoid of gas-like behavior. *Transportation Research Part B: Methodological*, 36(3), pp.275-290.

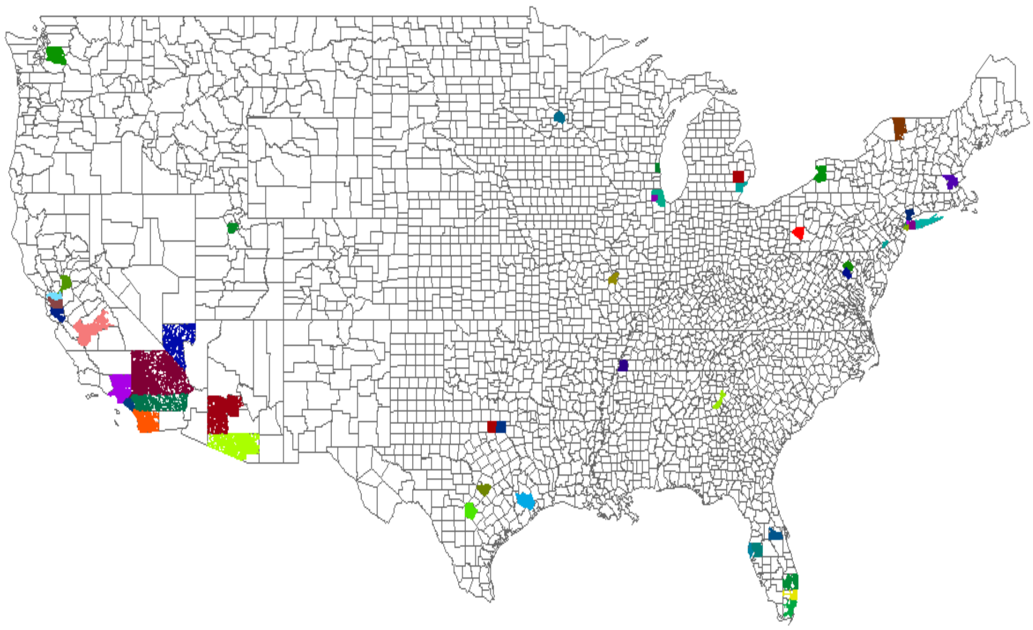
Zhang Q.N., 2005. Road network generalization based on connection analysis. In: *Developments in Spatial Data Handling*, pp.343-353.

Zhang X.D. and Armstrong M. P., 2008. Genetic algorithms and the corridor location problem: multiple objectives and alternative solutions. *Environment and Planning B: Planning and Design*, 35(1), pp. 148-168.

Appendix: 50 road networks in the USA

50 road networks of the most populous U.S. counties are listed as below. All the data are included in \attachment\Chapter3 directory on the attached CD-ROM. The auxiliary materials are provided in Tagged Image File Format (TIFF) files.

It should be noted that the TIGER road networks have been processed by steps descried in Chapter 4.2.1. i.e., remove isolated road segments and build strokes. It is interested that the shape of some road networks are like animals, for instance, Nassau road networks forms a bear shape, St.Louis road network is of a shape of horse, while the road network of Erie, Hennepin, Los Angeles and Middlesex are likes birds.

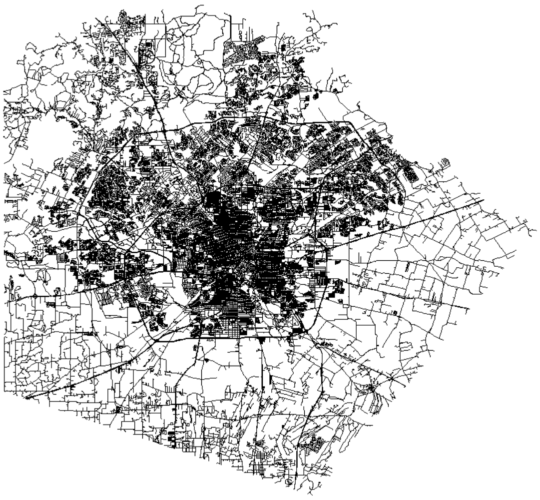


The geographical locations of the selected 50 counties in the USA

The 50 road networks

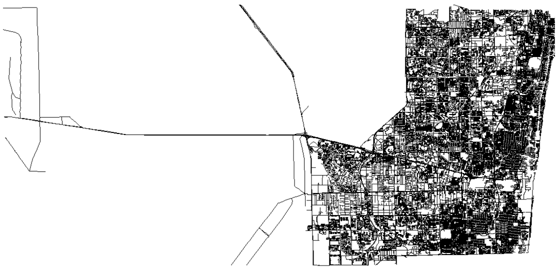
Alameda

Allegheny

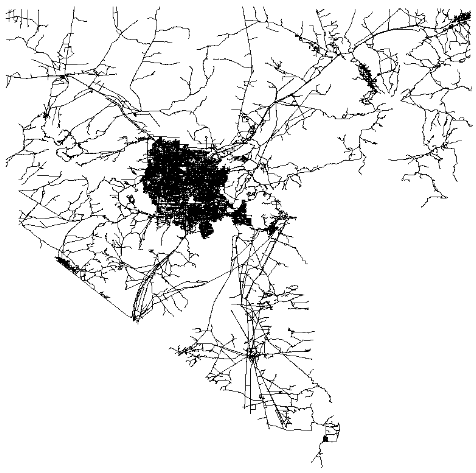


Bexar

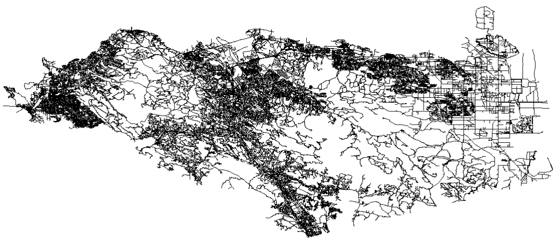
Bronx



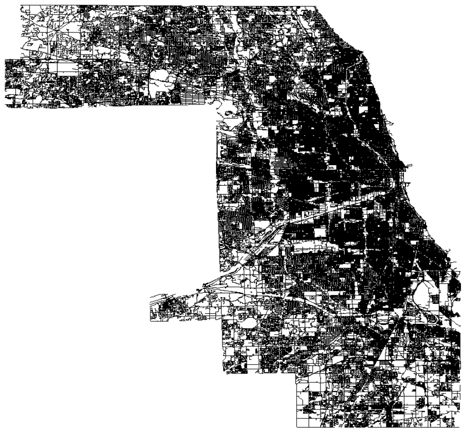
Broward



Clark

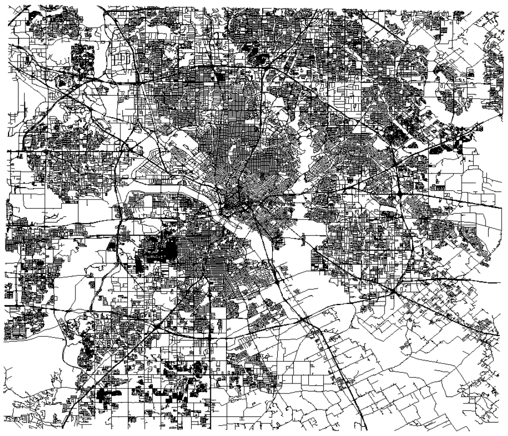


Contra Costa

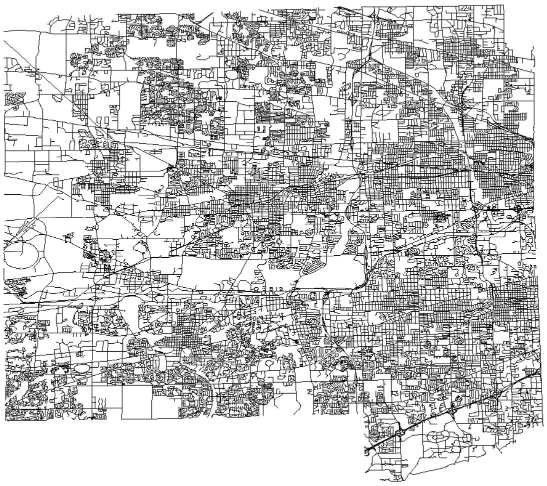


Cook

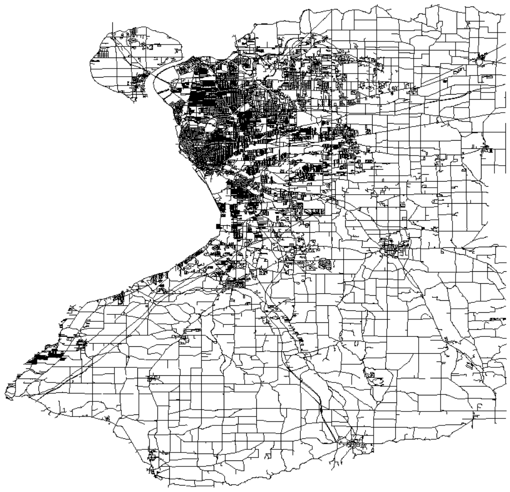
Cuyahoga



Dallas

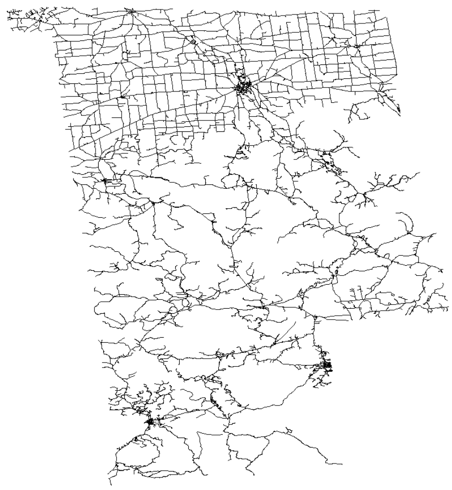


DuPage

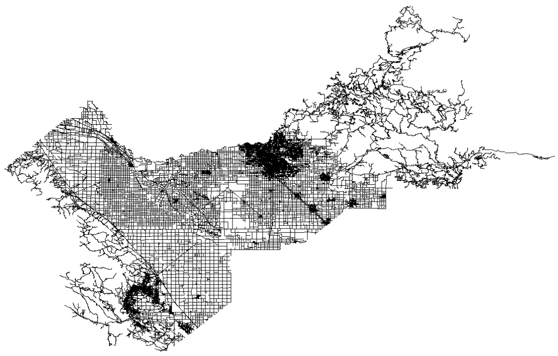


Erie

Fairfax

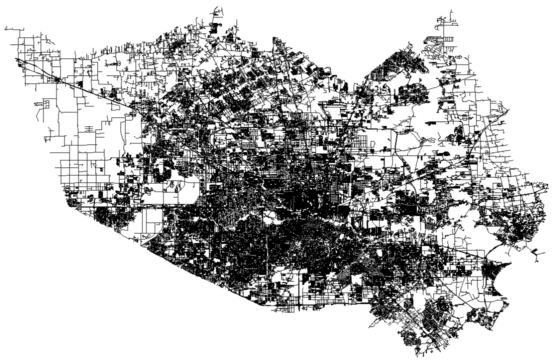


Franklin

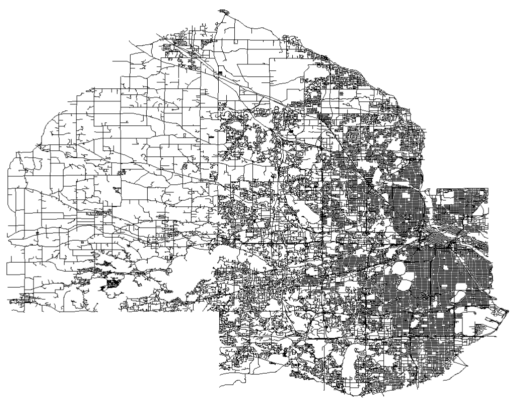


Fresno

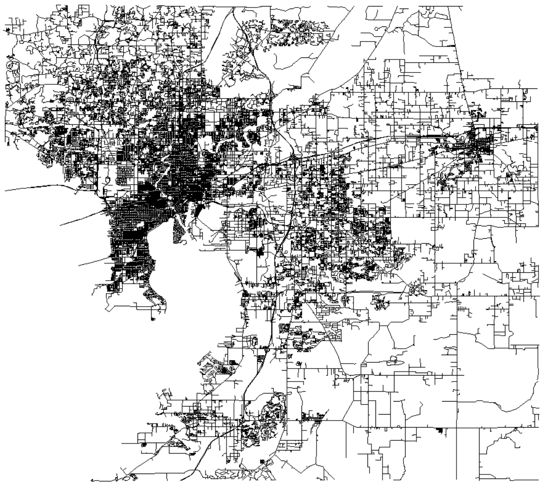
Fulton



Harris



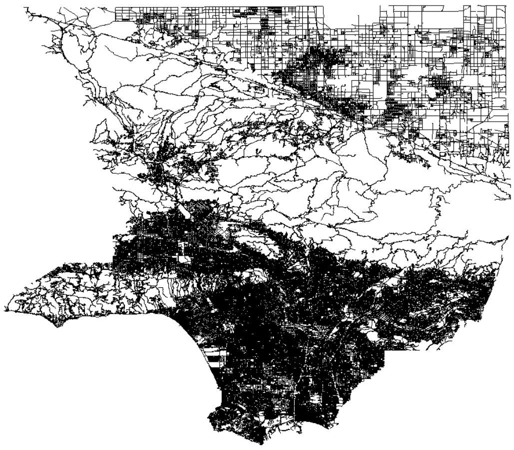
Hennepin



Hillsborough

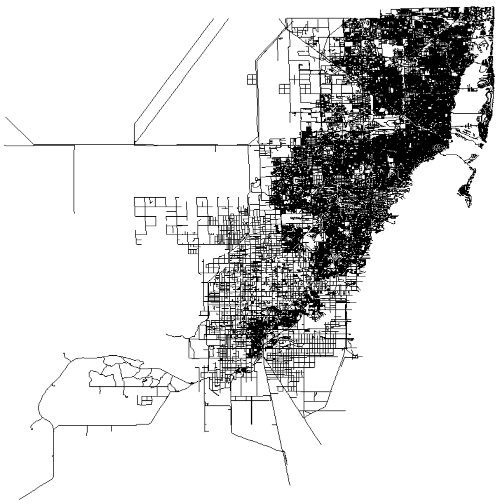
King

Kings

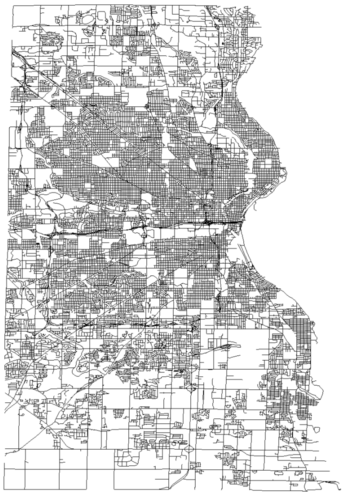


Los Angeles

Maricopa



Miami-Dade

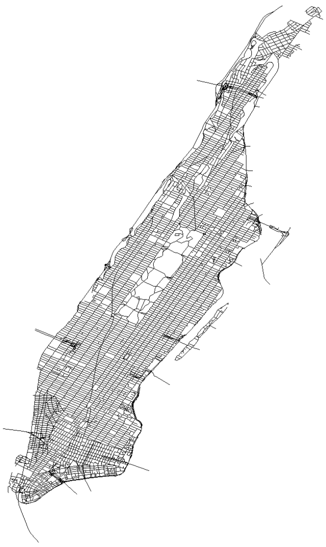


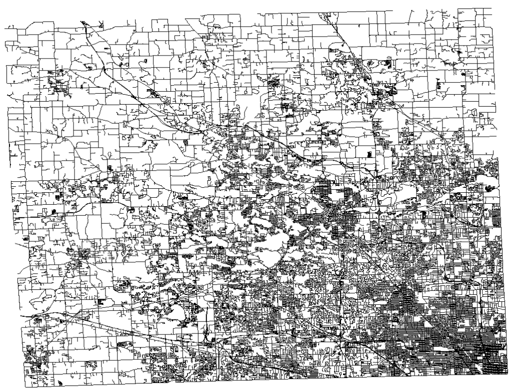
Middlesex

Milwaukee

Montgomery

Nassau

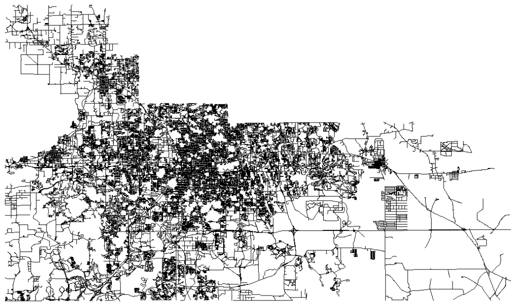




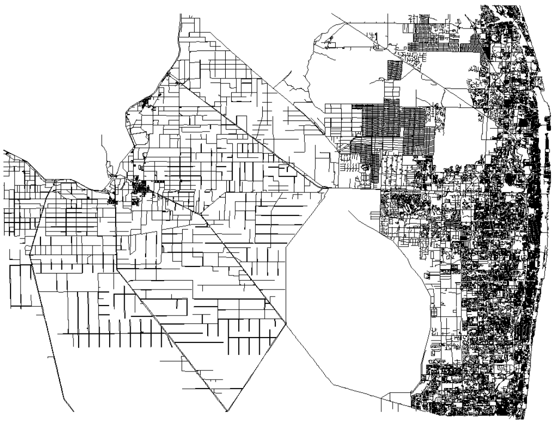
New York

Oakland

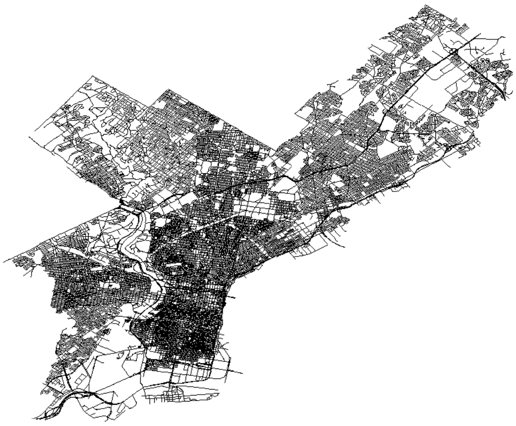
Orange_CA



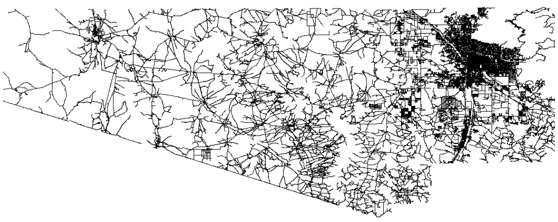
Orange_FL



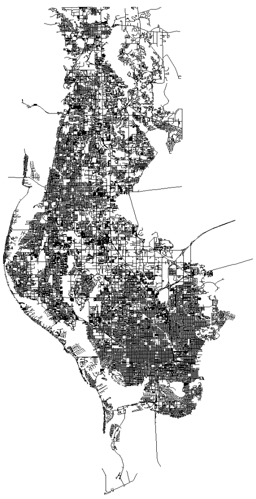
Palm Beach



Philadelphia

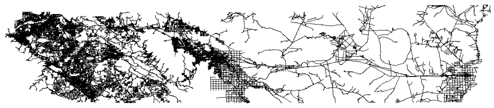


Pima

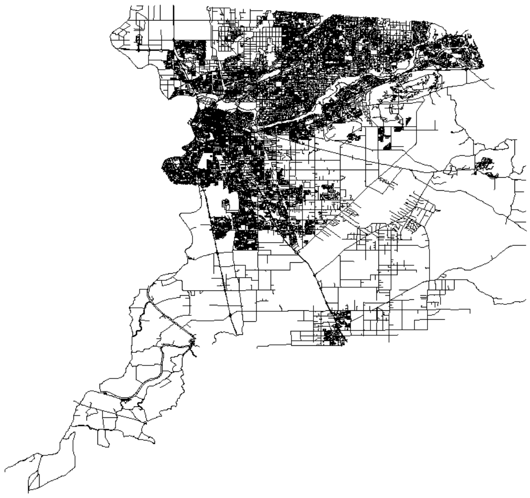


Pinellas

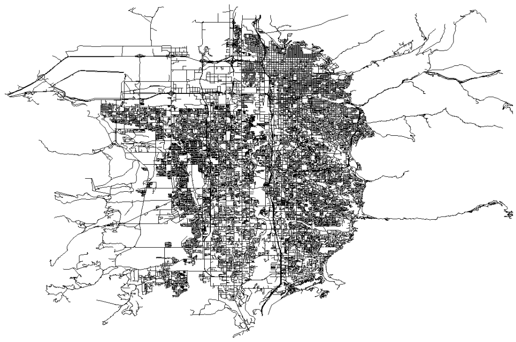
Queens



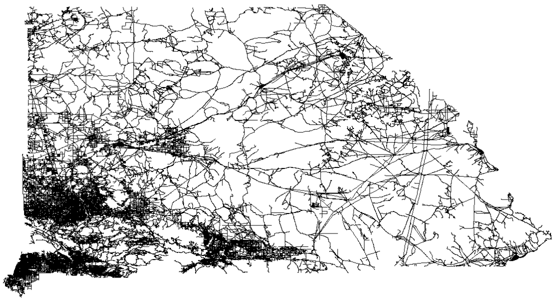
Riverside



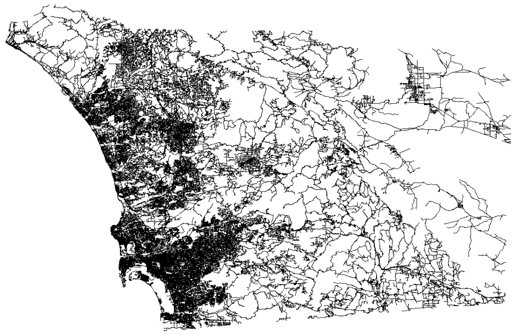
Sacramento



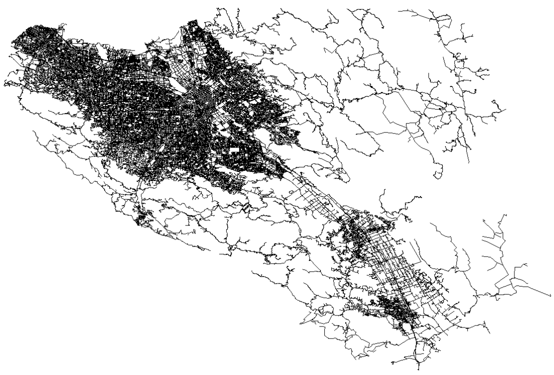
Salt Lake



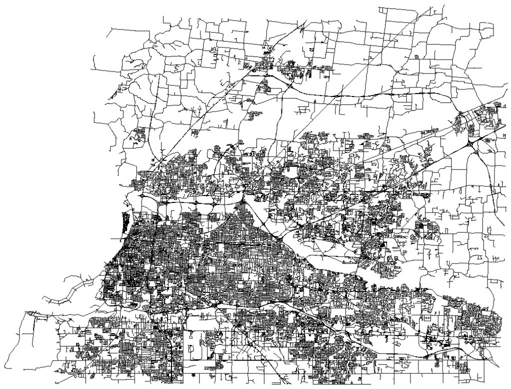
San Bernardino



San Diego

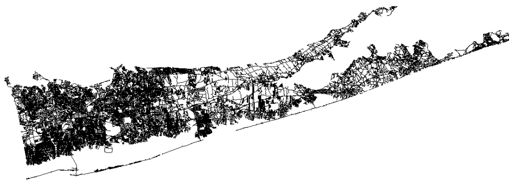


Santa Clara

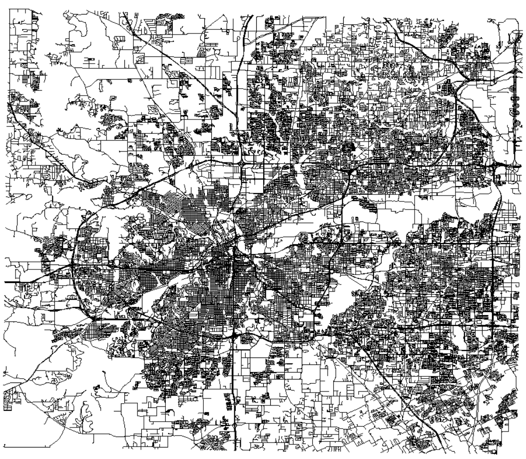


Shelby

St.Louis

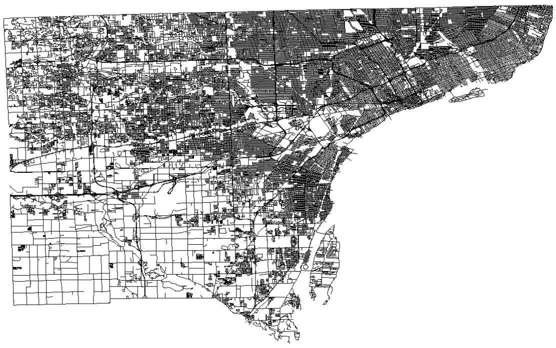


Suffolk



Tarrant

Travis



Wayne

Westchester