Exploring the topological patterns of urban street networks from analytical and visual perspectives

Junjun Yin

February 2009

Master’s Thesis in Geomatics
Abstract

Research interests in the studies of complex systems have been booming in many disciplines for the last decade. As the nature of geographic environment is a complex system, researches in this field are anticipated. In particular, the urban street networks in the Geographic Information System (GIS) as complex networks are brought forth for the thesis study. Meanwhile, identifying the scale-free property, which is represented as the power law distribution from a mathematical perspective, is a hot topic in the studies of complex systems. Many previous studies estimated the power law distributions with graphic method, which used linear regression method to identify the exponent value and estimate the quality that the power law fits to the empirical data. However, such strategy is considered to cause inaccurate results and lead to biased judgments. Whereas, the Maximum Likelihood Estimation (MLE) and the Goodness of fit test based on Kolmogorov-Smironv (KS) statistics will provide more solid and trustable results for the estimations. Therefore, this thesis addresses these updated methods exploring the topological patterns of urban street networks from an analytical perspective, which is estimating the power law distributions for the connectivity degree and length of the urban streets. Simultaneously, this thesis explores the street networks from a visual perspective as well. The visual perspective adopts the large network visualization tool (LaNet-vi), which is developed based on the k-core decomposition algorithm, to analyze the cores of the urban street networks. By retrieving the spatial information of the networks from GIS, it actually enables us to see how the urban street networks decomposed topologically and spatially. In particular, the 40 US urban street networks are reformed as natural street networks by using three “natural street” models.

The results from analytical perspective show that the 80/20 principle still exists for both the street connectivity degree and length qualitatively, which means around 20% natural streets in each network have a connectivity degree or length value above the average level, while the 80% ones are below the average. Moreover, the quantitative analysis revealed the fact that most of the distributions from the street connectivity degree or length of the 40 natural street networks follow a power law distribution with an exponential cut-off. Some of the rest cases are verified to have power law distributions and some extreme cases are still unclear for identifying which distribution form to fit. The comparisons are made to the power law statement from previous study which used the linear regression method. Moreover, the visual perspective not only provides us the chance to see the inner structures about the hierarchies and cores of the natural street networks topologically and spatially, but also serves as a reflection for the analytical perspective. Such relationships are discussed and the possibility of combining these two aspects are pointed out. In addition, the future work is also proposed for making better studies in this field.

Keywords: MLE, KS, power law, k-core, GIS
# Table of Contents

1. INTRODUCTION .........................................................................................................................1
   1.1 Background .................................................................................................................................1
   1.2 Aims .............................................................................................................................................2
   1.3 Structure of the thesis ..................................................................................................................3

2. LITERATURE REVIEW .............................................................................................................4
   2.1 Complex system studies and small world explorations ..............................................................4
      2.1.1 Complex systems ....................................................................................................................4
      2.1.2 Small world and scale free networks ..................................................................................5
   2.2 Power law distributions studies ..................................................................................................6
      2.2.1 Power law distributions ........................................................................................................6
      2.2.2 Heavy tailed alike distributions ............................................................................................8
   2.3 Conventional approaches and newly updated strategies ............................................................9
   2.4 Emergence of GIS in complex system study .............................................................................10
   2.5 Natural street models ..............................................................................................................11
   2.6 Network visualization for GIS ..................................................................................................12
      2.6.1 K-core decompositions .........................................................................................................13
      2.6.2 K-core decomposition layout in LaNet-vi ...........................................................................14
      2.6.3 The benefits of using K-core decomposition layout in LaNet-vi ........................................15
   2.7 Chapter Summary .....................................................................................................................15

3. A NATURAL STREET PERSPECTIVE FOR ANALYTICAL AND VISUAL STUDIES ..... 16
   3.1 Proposed methodology for the study .......................................................................................16
   3.2 Three “natural street” models ..................................................................................................17
   3.3 Analytical strategies ...................................................................................................................20
      3.3.1 Estimation the exponent from power law distribution using maximum likelihood estimation (MLE) ..........................................................20
      3.3.2 Goodness of fit .......................................................................................................................22
   3.3 Visual strategies .....................................................................................................................27
      3.3.1 Procedures for k-core decompositions and graphical representations ..............................27
3.3.2 Illustration of the graphical representation from LaNet-vi ............................................................. 27
3.3.3 Combinations with LaNet-vi and GIS ............................................................................................ 28
3.4 Chapter summary ............................................................................................................................ 28
4. THE 40 US URBAN STREET NETWORKS FOR CASE STUDY ..................................................... 30
4.1 Datasets and software ...................................................................................................................... 30
4.2 Generating the natural streets ........................................................................................................... 31
  4.2.1 Checking isolations ....................................................................................................................... 31
  4.2.2 Transforming the datasets into segment based urban streets .................................................... 32
  4.2.3 Three natural street models ......................................................................................................... 32
  4.2.4 Extracting the topological interrelationships and geometric properties .................................... 33
4.3 Analytical perspective: power law estimations ............................................................................... 34
  4.3.1 Qualitative exploration of the 40 US urban street networks ...................................................... 34
  4.3.2 Power law estimation for parameters: $x_{\min}$ and $\alpha$ ................................................................ 34
  4.3.3 Goodness of fit 1: P value test ...................................................................................................... 35
  4.3.4 Goodness of fit 2: Alternative distribution competing test .......................................................... 35
4.4 Visual Perspective: Network visualization for GIS ......................................................................... 36
  4.4.1 K-core decomposition in LaNet-vi ............................................................................................... 36
  4.4.2 K-core decomposition retrieving back to GIS ............................................................................ 37
4.5 Chapter Summary ............................................................................................................................. 38
5. RESULTS ............................................................................................................................................. 39
5.1 Analytical perspective ....................................................................................................................... 39
  5.1.1 Qualitative overview of the structure of 40 urban street networks .......................................... 39
  5.1.2 Power-law fit ............................................................................................................................... 40
  5.1.3 Goodness of fit 1: P value test ...................................................................................................... 41
  5.1.4 Goodness of fit 2: Alternative distributions competing test .................................................... 43
  5.1.5 Overall comparisons about the exponent $\alpha$ from the case study ...................................... 45
  5.2.1 k-core decompositions from LaNet-vi ....................................................................................... 46
  5.2.2 Retrieving model of k-core decomposition in GIS ..................................................................... 47
6. CONCLUSIONS AND FUTURE WORK.................................................................................. 49

6.1 Analysis and discussions ........................................................................................................... 49

6.1.1 Qualitative overview for the structure of the natural street networks .................................. 49
6.1.2 Quantitative analysis: Power law estimations and alternative distribution tests ................. 49
6.1.3 Comprehensive comparisons with previous study................................................................. 50
6.1.4 Network visualizations based on k-core decompositions ..................................................... 50
6.1.5 Visualizations for k-core decompositions in GIS ................................................................. 51
6.1.6 Combination of analytical and visual perspectives............................................................... 51

6.2 Conclusions ............................................................................................................................... 51

6.3 Future work ............................................................................................................................... 52

6.3.1 Problems to solve ................................................................................................................. 52
6.3.2 Future development ............................................................................................................... 52

REFERENCES ............................................................................................................................... 53
### List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 1.1</td>
<td>A brief flow chart illustrating the structure of the thesis</td>
</tr>
<tr>
<td>Fig. 2.1</td>
<td>Rewiring procedure from a regular lattice to a random network connection (a) ( p = 0 ) for a regular network (b) ( p = 0.5 ) for a scale-free network (c) ( p = 1 ) for a random network</td>
</tr>
<tr>
<td>Fig. 2.2</td>
<td>Two examples of Poisson distribution curves</td>
</tr>
<tr>
<td>Fig. 2.3</td>
<td>Heavy tail example using the US 1999 population data</td>
</tr>
<tr>
<td>Fig. 2.4</td>
<td>A synthetically generated example showing the “cut-off” tail of power law distribution</td>
</tr>
<tr>
<td>Fig. 2.5</td>
<td>(a) The histogram of the empirical data from a case study (b) The same histogram but in log-log scale plot (c) The histogram using logarithmic binning method in log-log scale plot (d) Cumulative histogram of empirical data in the log-log scale</td>
</tr>
<tr>
<td>Fig. 2.6</td>
<td>Conventional strategy for estimating the power law and getting the exponent parameter</td>
</tr>
<tr>
<td>Fig. 2.7</td>
<td>(a) A simple network with 16 arcs and 15 nodes (b) resolved into 7 natural streets (Jiang et al., 2008)</td>
</tr>
<tr>
<td>Fig. 2.8</td>
<td>Typical maps in GIS (a) maps for statistic analysis in GIS (b) thematic map with street names for illustration</td>
</tr>
<tr>
<td>Fig. 2.9</td>
<td>(a) Network visualization of topological relationships for streets in Satra, Sweden (b) Preferential Attachment network layout</td>
</tr>
<tr>
<td>Fig. 2.10</td>
<td>K (0, 1, 2 &amp; 3)-core of the simple graph annotated with colors</td>
</tr>
<tr>
<td>Fig. 2.11</td>
<td>Annotations for the layout from LaNet-vi</td>
</tr>
<tr>
<td>Fig. 3.1</td>
<td>A brief flow chart showing how the studies in GIS would be implemented</td>
</tr>
<tr>
<td>Fig. 3.2</td>
<td>Simple stroke tracking illustration</td>
</tr>
<tr>
<td>Fig. 3.3</td>
<td>Illustration for the definition of deflection angle in current thesis and Jiang (2007)</td>
</tr>
<tr>
<td>Fig. 3.4</td>
<td>Illustration for optimization for stroke tracking algorithm</td>
</tr>
<tr>
<td>Fig. 3.5</td>
<td>The maximum distance for the CDF of hypothesized data against the CDF of empirical data</td>
</tr>
<tr>
<td>Fig. 3.6</td>
<td>(a) exponential distribution (b) log-normal distribution (c) stretched exponential (Weibull) distribution (d) power law distribution with an exponential cut-off (e) power law distribution with the black solid line indicating the hypothesized power law should be</td>
</tr>
<tr>
<td>Fig. 3.7</td>
<td>Two k-core decomposition representations for different network revelations</td>
</tr>
<tr>
<td>Fig. 4.1</td>
<td>The proposed flow chart for the case study of 40 US urban street networks</td>
</tr>
<tr>
<td>Fig. 4.2</td>
<td>Converting streets into segments</td>
</tr>
<tr>
<td>Fig. 4.3</td>
<td>A simple example showing how a stroke generated from the segments</td>
</tr>
<tr>
<td>Fig. 4.4</td>
<td>(a) Power law fit for a set of sample data with a small value of ( x_{\text{min}} ) (b) Power law fit with a bigger ( x_{\text{min}} ) compared with the previous one</td>
</tr>
<tr>
<td>Fig. 4.5</td>
<td>(a) A simple netfile example, (b) The corresponding plot from Pajek</td>
</tr>
<tr>
<td>Fig. 4.6</td>
<td>(a) The working environment in LaNet-vi (b) The 2-D layout for k-core decomposition</td>
</tr>
<tr>
<td>Fig. 4.7</td>
<td>(a) A simple illustration for join the node ID with street ID (b) Rendering the map with k-core value in GIS</td>
</tr>
<tr>
<td>Fig. 5.1</td>
<td>Power law fit for SBF (threshold angle 45) for length</td>
</tr>
<tr>
<td>Fig. 5.2</td>
<td>Power law fit for SBF (threshold angle 45) for connectivity</td>
</tr>
<tr>
<td>Fig. 5.3</td>
<td>K-core decomposition representations from topological perspective in LaNet-vi</td>
</tr>
<tr>
<td>Fig. 5.4</td>
<td>The k-core decompositions with retrieved spatial information in GIS</td>
</tr>
</tbody>
</table>
List of Tables

Table 5.1 The percentage of streets that have a connectivity degree or length above the average .......................... 39

Table 5.2 The parameters: exponent (α), x_{min}, P value from the power law estimation for the SBF natural street network ................................................................................................................... 42

Table 5.3 Alternative distribution test against power law for the 40 natural street networks (length) .................. 43

Table 5.4 Alternative distribution test against power law for the 40 natural street networks (connectivity) ........................................................................................................................................ 44

Table 5.5 Comparisons between the previous studies and the thesis work about the estimated exponent α ........................................................................................................................................... 45

Table 6.1 Statistics the percentages over 40 the natural road networks whose length or degree above its average value ........................................................................................................ 49
Acknowledgements

I would like to express my gratitude to all those who gave me the possibility to complete this thesis. I am deeply indebted to my supervisor Prof. Bin Jiang whose help, stimulating suggestions and encouragement helped me in all the time of research for and writing of this thesis.

My former colleague Dr. Sijian Zhao from the Hong Kong Polytechnic University, I want to thank him for all his help, support and valuable hints. Especially I am obliged to Prof. Aaron Clauset, for his patience answering my questions about the issues related to power law distributions. Dr. César A. Hidalgo, his patient advices and suggestions on solving the problem of KS statistics.

Especially, I would like to give my special thanks to my parents whose support and love enabled me to complete this work.
1. Introduction

1.1 Background

Over the last decade, researches for exploring and studying the complex systems have been prevalent. With respect to the term “complex system”, it is a system consists of a variety of sub-components which interact with each other and hence make the whole one have some properties that are not easily captured from each individual. Beyond the scope of this definition, exploring the secret of the complex system, seeking the inner pattern beneath the common phenomena, and paving the way revealing how the nature works, is becoming an enthusiastic challenge for researchers. One common feature of the complex system usually refers to its nonlinearity, as is described by the term “dynamic systems” (Pavard and Dugdale, 2008). However, the appearances and representations of the existing systems are showing no direct clues about how the systems are structured and how the inner mechanisms work. Due to these remarkable dynamics, there rarely existed a fixed or universal model which could fit the patterns once and for all. Even though, with consistent studies and explorations of the complex systems, numerous discoveries of the patterns that characterized the systems were found. Taking “small world” theory for instance, which is more popularly known as “six degree separation” (Albert and Barabási, 2002), it has been widely captured in many systems and networks. For example: the collaboration network from IMDB (the Internet Movie Data Base), the collaboration network with Paul Erdös (ER model for short) (Erdös and Rényi, 1959), etc. These systems were found to share a “small world” structure. Such fact not only shows the possibility to seek patterns to figure out the complex systems, but also brings a lightening vision for the potential to find the universality.

To get a better understanding of how the systems are structured and how the patterns affect the running of the systems, we shall not only see through the existing common phenomena, but also touch the essential root, which is shifting from qualitative estimation to quantitative analysis. To reach this goal, it is necessary to set up certain mathematical models and adopt scientific expressions to explain these structures and properties. It is actually a fashion that “mathematic” talks. The “scale free” property, which was explicitly talked about by Albert and Barabási (2002), characterized itself by “power law” distribution from a mathematical point of view. Due to the importance of “scale free” property for the studies of complex systems, some say, fundamental properties for understanding the process beneath the dynamics (White et al., 2008), this thesis will illustrate in details in latter chapters.

Taking retrospect to the terms mentioned above: the specific term “network” has been mentioned frequently. It does draw our attention to the current research strategies for the complex systems. The new trend of the studies of the complex system not only paid attention to studied agents (events) to sense each individual’s characteristics, but also concerned on the relationships among them, which extended the studies to a topological level. With the help of “graph theory”, the complex systems can be extracted as vertices and links, and hence they are represented topologically. The essence of using networks to model the complex system together with small world coupling will show significant advantages to enhance such as computational power and synchronizability (Watts and Strogatz, 1998). Indeed, the current frontier research in complex systems orientates to the analysis of the properties of the “complex networks”.

Currently, many methods and approaches are available to analyze and establish a model that may reveal the patterns. Nevertheless, one of their common procedures is to analyze the patterns of distributions from the empirical data and hence conclude how the systems structured and how the agents behaved. Such studies can be done from both qualitative and quantitative aspects. From a mathematical perspective, being able to analyze the distributions will serve charactering the patterns of the complex systems and uncovering the myth how the system structured and worked.

Among a variety of well known distributions from the studies of complex systems, power laws (distribution) have been found and studied in numerous areas, such as ecology, finance, physics and other disciplines (White and Green, 2008). The power law distributions can actually be treated as the fundamental process, with which the dynamical systems used charactering the complex systems themselves (White and Green, 2008). It can be simply described as an implication that small occurrences have much (extremely) more possibilities to take place compared with the large instances within certain systems (Adamic, 2000). Although the studies on power laws have had quite a long history, there are still some puzzles on confirming the expected power laws. Mostly, it is because the
improper mathematical test applied which will end up with drawing significant biased conclusions for getting the power laws (Clauset et al., 2007). For instance, simple graphical methods based on linear regression method were used frequently (Goldstein et al., 2004), which were not suitable and solid strategies to fit the power law distributions to the empirical data. Technically, you have to estimate the power law distributions through firmly conducted mathematical process. Shalizi (2007) suggested that the declaration of finding a power law distribution should be based on strict mathematical test and verification. Obviously, this carries an entire new wave of re-examining the power law distributions and brings a fresh atmosphere for a new roll of research. Clauset et al. (2007) proposed an updated strategy by adopting maximum likelihood estimation (MLE) and goodness of fit test to accomplish the estimation and validation. The detailed procedures will be shown in following chapters.

Note that, when complex systems are involved for studying, most of the time, they refer to large scale systems or networks, such as the studies of World Wide Web (Barabási and Albert, 1999). In this case, the datasets are inescapably huge, which simply makes a chaos for human perception. Moreover, the information visualization is extremely important for understanding what the inner structure of the complex system all about. The emergence of visual analytics is actually to meet such requirement which will support the analytical approaches with interactive and effective visual analysis process (Aigner et al., 2007). The visual analytics approaches essentially combine the analytical reasoning with visual perception and cognitions to explore the myth of the complex system. Facing the massive data in the study of complex systems, the visualization of the inner structure will help understand the implications from the analytical approaches. Therefore, with respect to the rising research interests promoting the visual analytics, this thesis will attempt to address the studies of complex systems from both analytical and visual perspectives.

1.2 Aims

When the studies of complex systems involve geospatial data, the emergence of GIS will extrude prominent advantages dealing with such datasets, handling the computations and visualizing the graphical acquirements etc. Since “the nature of the geographic environment, as is a complex system” (Jiang, 2005), the GIS will not only provide the geometry information of the individuals, but also capable modeling the system from a topological level (Jiang and Claramunt, 2004). Jiang (2007) examined the topological patterns over a large sample of 40 US urban street networks and declared to have identified the existences of power laws. The work was based on generating the “natural streets” to reproduce the street networks (Jiang and Liu, 2007). The interesting part is that, Jiang et al. (2008) implemented two other types of natural street models, which had some modifications of the version in 2007. More important, the mentioned work adopted graphical methods, linear regression in particular, to estimate the power law parameters, which is believed to bring biased judgment. Therefore, this thesis shall re-examine the identical 40 US urban street networks from a natural street network perspective. Moreover, the revised work will not be limited to the topological relationships (connectivity degree) but appended the spatial information of the geometry features (length); and all three natural street models will be applied to produce the natural street networks. More specifically, it is from an analytical perspective to study the topological patterns of the natural street networks. In detail, the updated strategies using MLE and goodness-of-fit for power law estimation will be promoted for the analytical studies. The comparisons with the previous studies will be made to re-examine the previous conclusions.

On the other hand, simultaneous with the analytical perspective, this thesis shall study the natural street networks from a visual perspective as well. It will take advantages of GIS, since GIS is known for its capability of “Geo-Visualization”. The GIS techniques and applications will not only help visually discover and seek the patterns but also will show direct and intuitionist results. In contrast with the conventional visualization approaches, in which only different themes of color or symbols were used to categorize the map for recognition, this thesis will perform the network visualization approach based on k-core decomposition algorithm to provide knowledge and fancy visualization effects the same time. It is actually the initiative for promoting the visual analytics: to create a knowledge creating environment and provide interactive support to the analytical approach (Aigner et al., 2007; Ribarsky and Dill, 2008). In particular, with the help of GIS, the visualization will associate spatial information to create more meaningful maps for the complex systems.
To sum up, the aim of the thesis is to study the topological patterns of urban street networks from analytical and visual perspectives in GIS. In particular, the urban street networks will be reformed from three natural street models. The analytical perspective will adopt updated methodologies detecting and validating the power laws. Meanwhile, the visual perspective will make use of the network visualization together with GIS visualization strategies featuring the complex system studies with “maps”.

1.3 Structure of the thesis

The remainder of this thesis is organized as follows. Chapter 2 will be the literature review concerning two main research interests: analytical and visual strategies for the studies of complex systems. Whereas, chapter 3 will illustrate the methodologies and methods for the urban street networks starting from a natural street perspective for the studies of complex systems in GIS, in which the comparisons with conventional approaches and strategies will be shown as well. In chapter 4, the case studies over the 40 US urban street networks with detailed descriptions of procedures will be put into scene. Hence, chapter 5 will show the results from both analytical and visual perspectives. Finally in chapter 6, the conclusions over such a study will be made and the future work will be discussed from different perspectives. Figure 1.1 illustrates the flow chart of the thesis briefly.

![Fig. 1.1 A brief flow chart illustrating the structure of the thesis](image-url)
2. Literature review

In this chapter, as is outlined in the thesis structure, the scholarly literature reviews will mainly focus on two parts: complex system studies with particular concern on power law distribution and its related topics; and the network visualization approaches for complex systems which are based on k-core decomposition algorithm. For getting comprehensive understandings, the structure of this chapter is organized in a top-to-bottom manner, which starts from understanding the theory of complex system to its specific phenomena, power law distribution in particular. Within the procedures, some confusion among the current researches will be pointed out and corresponding solutions and strategies will be updated to refresh the atmosphere for power law analysis. Meanwhile, with respect to the visual analysis part, current visualization approaches for both complex networks and conventional GIS applications will be illustrated and compared. The feasibility of implementing the visualization by combining the functional GIS visualization will be gradually stripped into scene.

2.1 Complex system studies and small world explorations

2.1.1 Complex systems

During the last decade, the investigations on exploring the complex systems are becoming prevalent. The relevant researches are brought into a variety of fields, such as physics, social science, ecology, economics and even anthropology etc. What are the complex systems? It is the premise for the studies of the complex systems. A rough impression of the complex system is that a system consists of a variety of components and these components interact with each other following certain hidden pattern; the changes of some components would impact the whole system with certain behaviors. More specifically, according to the introduction (University of Michigan, 2008): a system or phenomenon will be considered as a “complex system” if it has following characteristics:

1. Agent-based: the study of complex system is essentially to study the behavior of the agents in the particular environment or phenomena. The properties of the system are the collections of the characteristics and behaviors that the agents perform.
2. Heterogeneous: it means that the behaviors and the characters of the agents are not all the same. They can share some common principles or run under the same disciplines, instead of being identically the same as clusters of colon copies. Therefore, they are fulfilled with more or less unique characteristics, for instance, human mobility (Marta et al., 2008; Brockmann et al., 2006).
3. Dynamic: one of the important properties that a complex system has is dynamic. It means the behaviors of the agents vary from time to time. The state of the system is not constantly stable. Once the agents in the environment adapt themselves, they would get affected by the other ones and the balance would be broken once again. This short balance is referred to “self-organized critical” state. The changes are usually nonlinear and disordered. A typical case for example is the size of the earthquake (Gutenberg and Richter, 1954).
4. Feedback: since the agents are interacted with others, the changes of the agents usually took place because of the affection from the activities of the others.
5. Organization: the inner structure of the complex system is usually in a hierarchy fashion. The agents’ properties thus can be categorized into different hierarchy level showing different pattern. A relevant case study is from Jiang (2008) about the streets and traffic analysis.
6. Emergence: the study of each individual agent about its behavior and characteristic will be massive. However, from the point of macro level, the emergence of the pattern will be patent. The Wisdom of Crowds (Surowiecki, 2004), from the view of social science illustrated it impressively.

Be based upon these definitions and the descriptions about the complex systems, the outline is portrayed. Apparently, a well designed spacecraft is not a complex system; even it does have numerous parts, they do not have the characteristics mentioned above. It can be only called “complicated systems” (Amaral and Ottino, 2004). This helps us understand that a complex system is
not merely a sum of various individuals but a collection featured with certain patterns because of the affections from all individuals.

2.1.2 Small world and scale free networks

Small world is one of the common properties that many complex systems share. It was first captured from the famous “small world” experiment, which is also known as “six degree of separation” (Newman et al., 2006). In that particular experiment, volunteers were trying to contact with any other (the one no matter they did or did not know) with mail. It revealed that the average steps for a person to reach another person are only “six steps” away. The computer based simulation with the help of graph theory can actually simulate the process, in which the studied objects within the environment are represented as nodes and links for their interacting relationships. Because of the simplified process of the complexity, it inspires the studies on how the networks are structured based on the topological patterns.

A small world network is an abstracted representation of complex system with small world properties. In physics and mathematics, the agents (individuals) are represented as nodes within the systems where the number of these agents is large and they sparsely distribute throughout the entire system. However, starting from one node to reach another one usually only takes a few steps, which was exhibited in the small world experiment already. Actually, a variety of cases of complex systems and phenomena were well modeled and characterized by using small world networks (Watts, 1999). Erdős and Rényi (1959) found that the distance between any pair of nodes in a random graph scales with the total number of the nodes which was related to the notion of “scale-free” (Albert and Barabási, 2002). Watts and Strogatz (1998) actually built a rewiring simulation according to the “Erdős and Rényi model” by setting up a randomness value p to control the form of the rewiring network. The process is shown in Figure 2.1, where p = 0 represents regular while p = 1 stands for a pure random network. The p value controls the possibility for the node to be connected to another in the lattice. As we can learn from the illustration, the status of connectivity among the nodes fluctuates dramatically when p value increases.

\[ k \sim \lambda^\gamma \]

Fig. 2.1 Rewiring procedure from a regular lattice to a random network connection (a) p = 0 for a regular network (b) p = 0.5 for a scale-free network (c) p = 1 for a random network

The study of degree distribution for the nodes in the networks will reflect the structural patterns (Barabási and Albert, 1999). The term “degree” is equivalent to “connectivity” of the nodes which refers to the number of neighbours connected to them. They studied many candidate networks, for instance the World Wide Web (Albert, Jeong and Barabási, 1999), and measured them in form of probability distribution function (PDF) from mathematical perspective. With respect to the rewiring network, the PDF was defined as \( P(k) \), where \( k \) was the degree of node \( n_k \). In random network, each pair of nodes was picked up randomly, and according to the statistics, the study showed that majority nodes in the network had a degree value around \( \bar{k} \), where \( \bar{k} \) was defined as the average degree of the entire network. The studied PDF followed a Poisson distribution but with a peak at \( k = \bar{k} \) and from the peak the distribution dropped down dramatically. The actual form of Poisson distribution plot is
shown in Figure 2.2, in which the value on the horizontal axis represents the degree of the nodes, while value of the vertical axis represents the probability that such a degree existed in the network. The red and green curves are two cases of Poisson distributions. The form varies because of the inner structure, for instance, the peak of green occurred around 3, while the one for the red curve appeared at 1. The green one means the majority nodes in the green case clustered with a degree value of 3. While, the circumstance for the red case that the majority of nodes only have an average degree of 1 which means the network structure is more diverse than the green one and it usually indicates that the nodes in the network were self-organized under certain rule rather than randomness (Barabási and Albert, 1999).

![Fig. 2.2](image)

Random networks are just extreme cases of small world networks with a randomness value of 1. Actually, the PDFs from some very large empirical networks, such as the World Wide Web (Albert et al., 1999), showed significant differences against Poisson distribution form, but with a “power law” tail. Networks with such statistical properties are referred as “scale-free” networks, which have power law characteristics (Barabási and Albert, 1999). The degree distributions is represented in a mathematical form of $P(k) \sim k^{-\alpha}$ or more detailed in Eq. 2-1:

$$P(k) = ak^{-\alpha} + o(k^{-\alpha})$$  \hspace{1cm} (2-1)

Here in the equation, $k$ represents the degree (or any objects for measurement), $a$ is a constant value to normalize the function and $\alpha$ is the exponent or the scaling parameter. In particular, $o(k^{-\alpha})$ is an asymptotically small function of $k$.

In practical, the “tail” may indicate many candidate distribution forms: exponential, log-normal, stretched exponential or power law distribution with exponential cut-off etc. Such mentioned distributions all look like the “fat tail” or equivalent to “heavy tail” distribution (Adler et al., 1998), because of a long and fast dropping tail which seems “fat” and “heavy”. Therefore, when studying the complex networks and trying to identify which distributions they might follow, selecting the best candidate distribution is very important along with the studies of PDF.

2.2 Power law distributions studies

2.2.1 Power law distributions

Before introducing the power law distributions, we first take a look at its mathematical form, which was illustrated in Eq. 2-1. The parameter $\alpha$ gathers all the extractions, which shapes the slop of the
distribution in a log-log plot (Clauset et al., 2007). From the mathematical representation of power law distribution, if we scale both sides of the formula with logarithm, which is shown in Eq. 2-2:

$$\log(P(k)) = -\alpha \log k + \log a$$  \hspace{1cm} (2-2)

Here, in the logarithmic scale, the function is now a linear representation of the distribution function, which means in the log-log plot the power law will be expected as a straight line with a slope of $\alpha$. This is actually the first direct and vivid impression about power law for the new learners to know about it.

Power laws usually characterized the phenomena of numerous occurrences in the nature, figuring the pattern of these occurrences and intending to reveal the principle or order (which explains the term “law”) underlying. In sociology and biology, it can be adopted to explain “How nature works” (Bak, 1996) and in economics where the “80/20” principle has been told (Koch, 1999), although not through exact mathematic analysis but qualitatively. In complex system studies, it is more from the mathematical perspective, in which the probability distributions were studied. Typical study cases are the Levy flights in ecology analysis of animal movements (Barthelemy et al., 2008), the frequency and size of earth quake size (Gutenberg and Richter, 1954) e.g. These studies had shown the common characteristics that power law distribution indicated that the big events (occurrences) happened rarely, while the small ones were widely spread throughout the systems (Adamic and Huberman, 2002). The studies of the probability distributions triggered the new interests in exploring the complex systems, including the phenomena in physics, social science, ecology, economics e.g. (Shalizi, 2007).

Power law distribution is a general term describing the particular mathematical form. In particular, there are also some other terms used under specific circumstances. For instance, Zipf, a PDF studied by George Kingsley Zipf (Adamic, 2000) which ranked the frequency of occurrence is shown in Eq. 2-3

$$P(r) \sim r^{-\alpha}$$  \hspace{1cm} (2-3)

Here $r$ represents the $r_{\text{max}}$ largest size. In contrast, Pareto distribution measures the events in different manner. Instead of ranking the size of occurrences, it measures the probability that how many events have an occurrence that are greater than $x$, whose mathematical expression is called the cumulative distribution function (CDF) showing in Eq. 2-4:

$$P(X > x) \sim x^{-\alpha}$$  \hspace{1cm} (2-4)

Applying this method of measuring the occurrences in the nature, it usually states that events bigger than a certain value occurs rarely, whereas smaller ones are common. It is more widely used compared to the Zipf method, since the Zipf methods can only be used to measure the discrete data, for instance, the frequency of works in articles, the population size in the country etc. In contrast, the Pareto method can actually measure both discrete and continuous type of data. According to the comparisons among Zipf, Pareto and Power laws by Adamic (2002), a pure power law also known as zeta distribution or discrete type of Pareto distribution (Goldstein et al, 2004), stands for an independent expression neither ranking the size nor counting the cumulative distribution but measuring exact the possibility that how many events have an exact value of $x$. For ease of illustration, it is shown in Eq. 2-5:

$$P(X = x) \sim x^{-\lambda}$$  \hspace{1cm} (2-5)

This particular expression is still in PDF manner. After being converted to CDF, the new equation is illustrated in Eq. 2-6:

$$P(X = x) \sim x^{-\lambda} = x^{-(\alpha+1)}$$  \hspace{1cm} (2-6)

Here, as it can be seen, the exponent for PDF is 1 less than the one for CDF. However, it is highly recommended by some researchers (Clauset et al., 2007, Newman, 2006) that using CDF for analysis will reduce the redundant error for estimating the exponent value compared with using PDF.
2.2.2 Heavy tailed alike distributions

The ideal power law distribution will be shown as a straight line in the log-log scale plot. In fact, when it is shown in a normal scaled plot, it has a shape with a long dropping tail, which is so called fat tail or heavy tail. Taking the CDF plot about the US population data of the year 1999 for example, the percentage of cities that have larger populations drops very fast along the tail. Although it is a sign for hidden power law distribution, it will still be categorized by the term “heavy tailed” distribution firstly, for other potential heavy tailed distributions will also have such kind of shape: log-normal, exponential, stretched exponential and even power law with a “cut-off” (Crovella et al., 2008, Anderson, 2006). The detailed mathematical representations will be illustrated in the latter chapters.

![Fig. 2.3 Heavy tail example using the US 1999 population data](image)

In theory, scale-free property will follow an exact power law distribution if the large networks will grow infinitely (Barabási and Albert, 1999). However, in most practical cases, the distributions do not follow the power law along the straight line in log-log plot over the entire data range but with a “cut-off”, which also refers to “truncated power law distributions” (Watts, 2003). The representation of such “cut-off” is illustrating in Figure 2.4.

![Fig. 2.4 A synthetically generated example showing the “cut-off” tail of power law distribution](image)

Here in Figure 2.4, Figure a represents an exact power law distribution in straight line. In contrast, Figures b contains a power law distribution with a faster dropping tail. Both the horizontal and vertical axes are in log-log scale. Apparently, the upper tail in Figure b drops much faster than its original trend which makes it as so-called a “cut-off” or “truncated” power law (Watts, 2003). In this case, since the distribution may not follow the power laws exactly over the entire data region, this actually gives a chance for the other potential distributions that may be a better fit for the data. These heavy tailed phenomena were widely captured in the complex system studies, such as: Internet connections, some of the stochastic process (Adler et al., 1998), levy flight for light (Barthelemy et al., 2008). All the mentioned heavy tailed distributions treated as the competitions against the power law distribution for being a preferred fit will be brought forth in the thesis studies.
2.3 Conventional approaches and newly updated strategies

Along with the discovery of small world phenomenon and scale free properties from the study of complex system, detecting power laws is becoming a hot pursuit. In the beginning, due to the limitations of the availability of proper mathematic strategy, some of these discoveries were based on visionary judgment by simply plotting the PDF created by histogram of data in a log-log plot to see whether it is a straight or approximately straight line. Such strategy was considered with poor quality in most cases (Newman, 2006). Newman (2006) also introduced the conventional measurement for the power laws, whose methods used histograms to calculate the frequency from dataset and segregating the data into equal bins to build to distribution function. As is known, the big events rarely happen, while the bin size is all the same throughout the whole range. Hence, the frequency of big events will be smaller leading to the noise (fluctuation) gathering at the tail. Instead of using even bins to make the histogram, the logarithmic bins were promoted, for instance, the logarithmic bins are 1, 1.2, 1.4 etc. When the width of bin grows, it would reduce the gathering of fluctuations and smooth the curve. To eliminate the affection from the fluctuations and avoid throwing away the tail, which may be rich of valuable information, cumulative histogram method to build a CDF was promoted. Cumulative histogram method will embrace the distribution from low level to higher level without having extreme small values. Actually, it is mostly used in the recent researches (Shalizi, 2007). The actual process of using the conventional binning method is illustrated from the Figures 2.5:

![Histograms and Cumulative Histograms](image)

**Fig. 2.5** (a) The histogram of the empirical data from a case study (b) The same histogram but in log-log scale plot (c) The histogram using logarithmic binning method in log-log scale plot (d) Cumulative histogram of empirical data in the log-log scale

Here in Figure b, the noises are significant around the tail, while the logarithmic binning method produces smoother distribution curves over the plot. In comparisons with previous two, cumulative histogram produced even better plot, in which the fluctuations are hardly seen at the tail.

Even though, a straight line alike curve does not guarantee it to be a power law. Many paper published adopted the approaches by making linear regression fitting over the distribution in the log-log scale, concluding whether it fits a power law or not by determining the significance of the correlation R square value and hence getting the slope value of the linear regression equation as the exponent value for the power law. As is shown in Figure 2.6, it is the mentioned convention method applied for the year 1990 US population data. The process was based on the CDF of the empirical data. Since the R square value is large enough (0.98), therefore, the slope value 1.8 is addressed as the exponent value for the power law fit. In particular, the topological analysis over the 40 US urban street networks by Jiang (2007) for power law estimations was performed in the same way.
Unfortunately, it is considered to lead to biased judgment and therefore cannot be trusted, for the linear regression approaches is based on assumption models, which assume the datasets came from a certain distribution model (Shalizi, 2007; Goldstein et al., 2004; Clauset et al., 2007). Apparently, it is not suitable to estimate the distributions. To fit the probability distributions, they proposed the Maximum Likelihood Estimation (MLE) to measure the exponent value of the power law, which will offer solid and trustable results.

Measuring the exponent $\alpha$ is treated as the fundamental process for the power law analysis (Newman, 2006). Since each unique power law distribution of the complex systems or phenomena, $\alpha$ value serves characterizing the shape of power law distribution and exposing the particular pattern the complex system may follow (Goldstein et al. 2007). Meanwhile, it is same important for validating the distribution to be a true power law as the estimating process. To test how well a power law fits the empirical data, Clauset et al. (2007) also proposed the “goodness of fit” test approaches by using “Kolmogorov-Smirnov” (KS) statistics. Furthermore, with respect to the fact that the heavy tailed distributions may have chance to be a better fit, the “goodness of fit for alternative distribution test” will be carried out for a robust confirmation. Many resent researches adopted the new approaches of using MLE and KS statistics for analyzing the power law distributions, for instance, Marta et al. (2008) studied the human mobility, Jiang et al. (2008) examined the human movement by taxi e.g. The initiative of revising the power law estimations in this thesis actually is in cope with the wave. The detailed procedures shall be illustrated in the coming chapters.

### 2.4 Emergence of GIS in complex system study

“The dynamic spatial redistribution of individuals is a key driving force of various spatiotemporal phenomena on geographical scales” (Brockmann et al., 2007). It exactly points out the important role that spatial redistribution plays. Conventional GIS approaches have encountered many limitations to interpret the interrelationships of the networks. When the networks evolve, those approaches do not have the capability to model such dynamics as they are essentially geometric representations (Jiang, 2005), and therefore the interrelationships were not well presented. As suggested by Jiang and Claramunt (2004), a topological network as the representation of the geographic environment shall be constructed modeling the dynamics. Jiang (2007) examined the connectivity degree over the 40 US urban street networks. The scale free properties were captured within these networks by concluding the power law existence. Meanwhile, a qualitative overview was about the 80/20 principle that nearly 80% streets had degree value around the average, whereas only 20% were above. After applying the Zipf power law ranking the connectivity degree and putting in the log-log scale plot, the “universality and peculiarity” within these street networks were stated. With respect to “university”, it refers to the common existence of power law distribution in each data set; while for “peculiarity”, the nested power law distributions were captures, which means the systems were attributed into two sub systems with two straight lines for power laws in the log-log scale plot.

In addition, due to the availability of various spatial-temporal datasets, such as the GPS data for tracking taxi navigation (Jiang et al., 2008), cell phone location data for tracking human mobility (Marta et al., 2008) and GPS data for detecting patterns of human movement (Rhee, 2007) e.g. These
GIS applications are not only capable of showing the geometric representation such as coordinates and distances etc. but also offering rich information for pattern findings, for instance, the speed, turning angle, the travel distance etc. Benefiting from these, the emergence of GIS in complex system study is a fresh wave. The thesis intends to verify the existence of power laws in GIS environment and explores the possibility of applying the analytical studies for the GIS characteristics.

2.5 Natural street models

Although, GIS is described and nominated as powerful as mentioned above, it still encounters embracement when handling the analysis of the complex networks from time to time. Taking the 40 US urban street networks for instance, due to the large number of street features, there is no doubt that not all the features are fulfilled with names. Yet, revised work can be carried out by referring to Google maps or other database to fill these blanks manually, not saying it would not work, but definitely time and energy consuming. This is also the limitation the so-called “Named Street” (Jiang and Liu, 2007) and other GIS analysis approaches have to face to. The notation of “stroke” (Thomson and Brooks, 2001) is proposed bursting from the idea that curvilinear can be drawn based on the good continuation, which means in specific that the generation of strokes through the network contains no dramatic abruption in direction at the junctions. It adopts the ideal of human perception of nature, “Gestalt principle” in psychology in particular. For instance, seen from different angle, the well known picture “Rubin vase” can be actually as a beautiful vase or two faces against each other. The streets generated under such principle are referred to “natural streets” (Jiang et al., 2008). For the ease of illustrations, the actual process can refer to Figure 2.7, the curves in Figure a can be naturally perceived as 7 smooth curves in Figure b. In Figure a, it is a simple network with 16 arcs and 15 nodes. The nodes are representing the term “junctions”, while the arcs between each pair of junctions are defined as “segments”. By tracing each segment based on the good continuation (by determining the deflection angle between each pair of segments) at each junction under a proper threshold angle, the natural streets generated are shown in Figure b. There are 7 reformed streets, each individual one consists of segments following the rule mentioned above.

![Fig. 2.7](image)

The threshold of angle can be specified conforming to the requirements for different analysis approaches. Jiang (2007) adopted 70 degree in their analysis of the 40 US urban streets. However, for human based applications, 45 degree is recommended assuming human choosing the roads when navigating in the street networks (Jiang et al., 2008). Such “natural streets” have been studied and claimed to have prominent advantages for the GIS based analysis in predicting the traffic flow and optimizing the design of the street networks (Jiang and Liu, 2007; Jiang et al., 2008).

The natural street model consists of three different tracking algorithms: Every-best-fit (EBF), Self-best-fit (SBF) and Self-fit (SF) (Jiang et al., 2008). All these three tracking algorithms were implemented in GIS by reforming the geometry features conforming to the mentioned principles. In the previous studies of the topological patterns of the 40 US urban street networks, the SBF natural street model with a threshold angle of 70 degree was applied. The work revealed some valuable information for analyzing the patterns of the 40 US urban street networks and showed advantages over the opponent “Named Streets” model. The later related researches using such model for traffic prediction (Jiang et al., 2008), showed a correlation value around 0.85 indicating a good level of
similarity between the natural streets and ideal ones. However, because in SBF natural street model, the starting feature is randomly chosen and the recursive process is involved, which means once a single natural street is generated, those segments inside of it would have no chance to be merged with others outside of the street, even the deflection angle is even better (smaller). Logically, this is not fair for the segment features just because of their order stored in the database. Therefore, the EBF natural street tracking algorithm was implemented and the discussion about the performance of EBF and SBF was carried out (Jiang, et al., 2008). Another tracking algorithm: the SF natural street model was also attached to show the inner differences among the three tracking algorithms. The detailed descriptions and methods for the three will be illustrated in the latter chapters.

Among the three models, the SBF natural street model is claimed to have the most prominent performance for predicting the traffic flow (Jiang et al., 2008). However, it does not necessarily mean that it is the best tracking algorithms among the three from every aspect. Therefore, this thesis will revise the previous study use the same SBF tracking algorithm with a threshold angle of 45. Moreover, the angle of 70 will be applied again to eliminate potential bias. Furthermore, the other two tracking algorithms will also be implemented in this thesis exploring the possibility of getting better performance and seeking differences.

2.6 Network visualization for GIS

The spatial is special. GIS has been claimed to have prominent power in visualization with its inborn capability of cartography functionalities. Facing the frontier of the complex system analysis, GIS should also play its own role by releasing its advantages in this field. To get a rough impression of conventional (or should say, ordinary) maps, two of them are exhibited in Figure 2.8. Figure a shows the population change over 1990 and 1999 for the states of USA with diagrams. Figure b is a normal map showing the names of the streets in Sätra, Sweden. You can get the spatial information easily about where it is and what happened in that place. Even though, you are puzzled about the interrelationships. Taking Figure a for instance, how the population distributed, what is the model describing the changes over all the states. And this is the visual analysis trying to deal with. However, it is not a simple process associating the statistics to each element in the map. The term “map” is not just a noun, GIS makes it a verb, an action to model the world, which is visually to unfold the inner structures of the complex networks. This is what this thesis intends to do, to “map” the complex systems in GIS.

Fig. 2.8 Typical maps in GIS (a) maps for statistic analysis in GIS (b) thematic map with street names for illustration

The representations of interrelationships within the complex systems can actually take advantages of using graph theory. For a graph (G), the objects inside are represented with vertices (V) and the interconnections are represented with links or arcs (L) (Vladimir and Matja, 2002). As is shown in Figure 2.9 for a brief illustration, Figure a represents the network visualization of the topological patterns of the streets in Sätra, Sweden, whose geometry representation is shown in Figure 2.8. Clearly, we can identify the “hubs” which are the most connected streets distinguished by the size of
the nodes and the “sub-hubs” can be seized accordingly. The example in Figure b is called “Preferential Attachment model” (Wilensky, 2005) serving as a model simulating how these “hubs” and “sub-hubs” arise in internet and the connectivity degree distribution inside of it has a power law characteristic (Albert and Barabási, 2002). The sense of illustrations here, from another perspective indicates that the complex network study can be digged in by using graph theory from the topology level, and the visualizations have made the outcomes visible and accessible easier.

![Fig. 2.9](image)

**Fig. 2.9** (a) Network visualization of topological relationships for streets in Satra, Sweden (b) Preferential Attachment network layout  
(Source a: Jiang (2005); b: Netlogo module library: network for Preferential Attachment model)

Actually, with the development of graph theory, a variety of approaches are available to accomplish the mission of network visualization. Moreover, among them, the visualization approach detecting cores of the networks based on k-core decomposition algorithm is brought forth, which is among few concepts that will slice the complex network into decompositions besides the connectivity relationships (Batagelj and Zaversnik, 2002).

### 2.6.1 K-core decompositions

The notion of “cores” is among few concepts which will show rich and meaningful decompositions of the large scaled complex networks (Garey and Johnson, 1979). It is usually explained with the help of graph theory: assume the graph $G = (V, L)$, where $V$ represents the vertices within the graph, and $L$ is the link or arc for their linkage. Then, the sub-graph $H = (C, L|C)$, where $C$ represents the partial vertices with $C \subseteq V$, is so called a k-core or a core of order $k$ (Batagelj and Zaversnik, 2002). Moreover, if every degree of the vertex within the sub-graph $H$ is greater than $k$, such sub-graph $H$ is called the maximum sub-graph. To get better understanding of the k-core structure, two other important terms shall be introduced: “main core” and “core number”. The former one corresponds to the core of the maximum order, while the latter one stands for the highest order of a core that contains the particular vertices. An example of a simple graph showing k-core decompositions is presented in Figure 2.10, the $k$ (0, 1, 2 & 3)-cores are identified according to the explanations above, where the black stands for value 0, blue for 1, green for 2 and red for 3. Apparently, within each shell (circle), the least degree value of the vertices is corresponding to the core number, which is the same value of the shell.
Extended from the simple illustration follows two important properties of the core structure:

1. The cores are in nested manners, which means sub-graph $H_i$ with higher core order is nested within the sub-graph $H_{i+1}$ with lower core order.

2. Since the isolated vertex exists (the one has a degree of 0), it means the sub-graphs do not have to be fully connected with each other.

The procedure for calculating the $k$-core decompositions will be illustrated in the latter chapters. However, the detailed implementations of the $k$-core decompositions for visualization are apparently beyond the scope of the thesis. Thanks to the contribution from A-Hamelin et al. (2005), the algorithm is successfully embedded as a function in their “Large scale network visualization tools” (LaNet-vi for short). The $k$-core decomposition for the large scale network or graph is represented in a two-dimensional layout featured with some of their topological and hierarchical properties shown.

### 2.6.2 $k$-core decomposition layout in LaNet-vi

The LaNet-vi is designed with the capability to visualize the large scale network in a form of two-dimensional layout that embeds some of its topological and hierarchical properties. Moreover, the detailed information about how the graphic representations were generated and how the layouts were drawn can refer to A-Hamelin et al. (2005). Although the inner mechanisms and detailed procedures on how the layouts are generated will not be emphasized, the properties and parameters attached in the layouts will be explained as well to understand what the layouts tell us. The example from LaNet-vi annotated with the properties and parameters shown in Figure 2.11 will make it easier to understand the corresponding meanings and information.
The following list extracted from A-Hamelin et al. (2005) will commentate each term respectively to understand the rich information in the figure:

1. **Degree**: The value of degree is the connectivity of the vertex from the entire network. In particular, the size of the vertices in the layout corresponds to the logarithmic scale of their original connectivity degree value.

2. **Coreness**: The value of coreness is not equivalent to degree; it indicates that the vertices within the sub-graph $H_i$ have an at least degree value $K$.

3. **Degree-Coreness Correlation**: The nodes with high degree do not necessarily exist in the high core, because, they may only connect to the vertices with low degree.

4. **Color scale**: The color is used to distinguish the level of coreness, starting from the violet representing the minimum value of coreness value to maximum $K_{max}$ with color red.

5. **Edges**: According to the graph theory, the edges represent the linkage within the nodes. For visualization effect, the edges from the k-core decomposition algorithm are randomly sampled from the original edges sets. One can control the value of proportion to show the edges in the layout.

6. **Component**: The decompositions of the networks can be segregated into multi-components (the right) due to the fraction of its inner structure.

With those terms introduced, the results from the LaNet-vi in the latter chapters will be identified and analyzed conforming to such mentioned basis.

### 2.6.3 The benefits of using K-core decomposition layout in LaNet-vi

The related studies using LaNet-vi for both visualization and analysis of real large networks and computer generated networks were performed by A-Hamelin et al. (2005). In particular, the “preferential attachment model” is adopted in studying the growth of network known as Barabási-Albert (BA) model in which the connectivity degree of the nodes follows a power law distribution (Albert and Barabási, 1999). Such network was visualized in LaNet-vi k-core decompositions, in which the layout could actually “see” the diversity of the distribution of connectivity degree. Other real networks, such as the World Wide Web (WWW) and the airports networks from International Air Transportation Association (IATA) (A-Hamelin et al., 2005) etc. were also visualized in LaNet-vi, and they had studied the structures of these networks such as the hierarchies. Therefore, using the network visualization for complex system studies will provide more possible ways for exploring the hidden myth. Moreover, combining the network visualization approaches with functionalities of GIS may produce surprising results and hence extend GIS to the cross-subject studies.

### 2.7 Chapter Summary

In this chapter, the literature review swept the definitions and some characteristics of the complex system by introducing its related topics, the properties of the small world and scale free networks in particular. With respect to the study of scale-free networks, the studies about the mathematical distribution: heavy-tailed and power law distributions are also introduced and illustrated, within in which the conventional approaches were compared versus the updated strategies. Following the review of the common studies, the emergence of GIS into the complex system studies is proposed and previous researches related are refereed. Meanwhile, the network visualization for GIS based on the k-core decomposition algorithm implemented in LaNet-vi is also brought forth to be scheduled for the thesis study.
3. A natural street perspective for analytical and visual studies

As is outlined in the previous chapters, this thesis will start from a natural street perspective to accomplish the studies with analytical and visual strategies: power law estimations based on updated methods and network visualization based on k-core decompositions algorithm respectively. Such studies will be performed in the GIS platform to bring GIS into the complex system studies. Regarding to the previous studies for the topological patterns of the 40 the urban street networks (Jiang, 2007), in which the SBF natural street model was applied, the power law existence was claimed by using the Zipf method and the exponent values were measured with linear regression method. Even power laws may do exist; there are high risks that these results are biased because of the improper methods used. Invoked by the updated strategies detecting and estimating the power law distributions introduced by Clauset et al. (2007), for the analytical part of this thesis, it will re-examine the probability distributions for the 40 US urban street networks for both connectivity degree and length, which measures the topological and the geometrical patterns respectively.

The essential goal of the visual studies of the street networks is to visually see the properties such as the structure about how these networks decomposed based on their interrelationships. The large network visualization tool (LaNet-vi) developed by A-Hamelin et al. (2005) will measure such complex properties based on the k-core decomposition algorithm. Working together with the GIS suite, it makes a spatial visualization of the complex network. This chapter will illustrate how the methodologies and methods applied in detail.

The remainder of this chapter is organized as follows: A brief flow chart will propose how the analytical and visual studies shall be carried out. The corresponding methods will hence be illustrated respectively. In addition, those conventional methods which were thought to cause inaccurate results or should be improved will be listed as comparisons to make the proposed methods more valid and reasonable.

3.1 Proposed methodology for the study

GIS is capable extracting the interrelationships of the street networks according to their spatial dependence. This is also how the topological network can be derived. Once the topological network is ready, the initial information, in particular, the connectivity degree and the geometrical length will be available through certain functions. The remaining work would be transported to the analytical and visual studies of the complex systems. A brief flow chart showing the proposed methodology is in Figure 3.1.
3.2 Three “natural street” models

There are actually three types of natural street models. The actual process is based on good continuation by determine the deflection angle between each pair of segments: the segment with a small deflection angle which is chosen to continue is naturally concatenated with the current one to form a natural street, the one with a bigger deflection angle will be treated as an independent street which still waits for been connected to others. The detailed procedures are illustrated as followings.

3.2.1 Self-best-fit (SBF) natural street model

Using natural street model will conquer the problems posed by named street model in case of loosing streets information due to the large scale of the street networks. Moreover, adopting the natural street model for predicting the traffic flow shows a significant correlation (Jiang et al., 2008), which implies the effectiveness using natural street model to re-produce the street networks. Jiang (2007) implemented the natural street model with a predefined threshold deflection angle of 70 degree in GIS, which means two elements with a deflection angle under 70 degree will be considered being merged as a natural street; otherwise, there is no possibility for action. Such natural street tracking process is base on the “segment based” street networks, in which a segment is the partial street between two junctions. The tracking algorithm uses a recursive process: for each segment, the tracking starts from
its two ends and concatenate the one of its connected segment at the junction by choosing the smallest deflection angle. Of course, the value of such deflection angle should be smaller than the threshold value. The new tracking will start from two ends of the new segment and will not stop while the deflection angle is less than the threshold angle value. Once it stops, a new natural street is generated. Here lists a simple natural street tracking illustration in Figure 3.2 showing how the process is done.

![Fig. 3.2 A simple stroke tracking illustration](image)

In Figure 3.2, the small street network consists of 10 segments $S_i$ and corresponding junctions $J_i$, namely, in which the arrows represent the tangent extension of the segments. The tracking process begins with a random choice of segment with a predefined threshold deflection angle of 45 degree and goes from the two directions of its ends. Therefore, a new natural street is generated as $<S_0, S_1, S_3, S_6>$. These segments within the natural street will not be considered any more in the following tracking process and the recursive process will be applied to the remaining segments until all the natural streets are generated. Such process is named as Self-best-fit natural street model (Jiang et al., 2008), as each segment considers itself to choose the other segments to continue, which seems more “selfish”.

This particular tracking process mentioned above actually has optimized the original model from Jiang (2007), in which the deflection angle was not defined as the angle between tangent extensions of two segments. The following illustration in Figure 3.3 will reveal the differences.

![Fig. 3.3 Illustration for the definition of deflection angle in current thesis and Jiang (2007)](image)

In Figure 3.3, the deflection angle in current thesis is $\alpha$, which is the angle between the tangent extensions of the segment $S_0$ and $S_1$, while the one in Jiang (2007) was derived by linking the two ends
first to form a new straight line and measured the angle in between. For instance, \( \alpha \) for \( S_0 \) and \( S_1 \). Apparently, the current definition is more reasonable.

Referring to the natural street process applied in the Swedish highway networks for predicting the traffic flow (Jiang et al., 2008), the deflection angle chose 45 degree instead of 70. The authors claimed that 45 degree is more close to the nature of human perception and it did show advantages over the other angles in predicting the traffic flow. Hence, 45 degree is adopted in this thesis as a default threshold value.

3.2.2 Every-best fit (EBF) and Self-fit (SF) natural street models

Note that the above tracking algorithm starts from choosing a random segment from the database. Such randomness will generate confusing natural streets under certain circumstances, which we can get impression from the following illustrations. An experimental tracking process is shown in Figure 3.3: in Figure a, there are 4 segments \( S_i \) and 5 junctions \( J_j \) namely. Since the initial selecting segment to start is random, for instance, \( S_1 \) in this case. The first decision to merge the segments is at \( J_3 \). In this example, \( \alpha \) is smaller than \( \beta \), thus the new natural street is \(<S_1, S_3>\) tagged with number 1 (the red solid line). The recursive process subsequently generated two other natural streets number 2 and 3 (the blue and green dot line respectively). The layout of natural streets will be dramatically different if \( S_2 \) would have been chosen to start the process, as is shown in Figure c, \( S_2 \) and \( S_4 \) will be concatenated to form a new natural street: number 2 (the blue dot line), while \( S_1 \) and \( S_3 \) will form another one: number 1 (the red solid line). Considered the basis of concept of good continuation, the natural streets in Figure 3 seem much more rational and reasonable than the ones in Figure b.

![Fig. 3.4 Illustration for optimization for stroke tracking algorithm](image)

To avoid such awkward, Jiang et al. (2008) implemented the tracking algorithm not only considered the smallest deflection angle between the current segment \( S_i \) and its connected segments at the junction, but also involved making the comparisons of the deflection angles among \( S_i \) and it connected segments despite of the directions. The modification procedure can be explained as following: assume the deflection angle \( \alpha_1 \) is the smallest angle among \( S_i \) and its connected segments at a certain junction, and the deflection angle \( \alpha_2 \) is the smallest angle among the current segment and its connected ones if ignoring the directions, both of them are less than the threshold value. If \( \alpha_1 > \alpha_2 \), the segment which has \( \alpha_1 \) with \( S_i \) will then not be merged with the \( S_i \) anymore, instead, the one has \( \alpha_2 \) with \( S_i \) will be joined thanks to its better continuation with \( S_j \).

The above tracking process will generate so called Every-best-fit natural street networks (Jiang et al., 2008). It makes sure that every segment in the street network gets its best fitted segments to form the “natural streets” with respect to good continuity at the same time and it is not “selfish” at all.
Hence, such process will produce a unique “natural street network”, which means only one street network will be produced no matter how many times the tracking process will be applied. However, there are still limitations. For instance, it takes place when the directions are omitted. Whether it is wrong or not, no one can actually judge due to the self-organized criticality (Bak, 1996). Jiang et al. (2008) also stated the form of “natural streets” followed a self-organized process. In particular, the Self-fit natural street model deals with the circumstance that among the deflection angles which are less than the threshold, the tracking algorithm randomly chooses one no matter whether it is the smallest or not and select the corresponding segment to continue. After they examined a variety of Swedish highway networks, they found the SBF based natural streets perform a significant correlation to the traffic data with a value above 0.8. They argued that it is a sensitivity issue in the formation of natural streets: even the SBF model is kind of “selfish” as it only chooses the best fit segment for itself. No one really knows how the streets in nature formed, whether completely organized by urban planning or completely natural producing? One example can refer to application to river networks using the similar process (Thomson and Brooks, 2001). In that case, apparently the directions should not be omitted, because the water flows with directions. And such studies about the traffic flow using natural street models are really rare, based on which we cannot conclude that the SBF is the best one for producing the natural streets neither is EBF. So, this is the real dilemma to implement the best natural street model. However, to verify which model is the best one for producing the best natural streets is beyond the scope of this thesis. In fact, the analytical study about the properties of the natural street networks is actually an essential issue. To make comprehensive comparisons verifying the previous studies about the US urban street networks, the SBF model will be adopted with the threshold value of 45 and 70. Meantime, the EBF and SF model will be applied as well to see the differences.

3.3 Analytical strategies

3.3.1 Estimation the exponent from power law distribution using maximum likelihood estimation (MLE)

Practically, we seldom know the distribution of the empirical data is drawn from a power law distribution. Ordinarily, once we have the distributions of the data, we shall use a series of distribution models to fit them but not using the linear regression method. The common distribution models such as Gaussian distribution model, log-normal distribution model e.g. will be applied to extract corresponding parameters. In particular, for the power law distribution, the essential goal of the estimation is to extract the exponent $\alpha$ in Eq. 3-1, as it shapes the power law distribution and indicates a certain order for each specific complex system

$$P(x) \sim x^{-\alpha}$$  \hspace{1cm} (3-1)

As mentioned above, the most common approach is to observe if the distribution of the empirical data lies in the logarithmic scale as a straight line, which is shown in Eq. 3-2:

$$\log(P(x)) \sim -\alpha \log x$$  \hspace{1cm} (3-2)

However, it is necessary to mention the conventional strategy using the linear regression method estimating the power laws. First, make a histogram of the empirical data as the PDF or use the rank to order the data to build a CDF, and then fit such distribution functions using the linear regression which is based on least-square methods. The slope of the linear regression fit will be treated as the exponent value $\alpha$ and the correlation value $R$ will imply how well it fits to the power law distribution. Unfortunately, there are some fatal errors involved in such strategy (Clauset et al., 2007; Goldstein et al., 2004; etc.). Clauset et al. (2007) pointed out the limitations in detail, for instance, using histogram or log-binning method will cause a lot of noises around the tail fluctuating up and down. Secondly, the linear regression method is based on assumptions, and hence it is not suitable for distributions. Last but not the least, the fit from linear regression usually does not satisfy the estimation of probability
distributions, for instance, the normalization procedure. Long words short, the current strategy of using the linear regression method will cause inaccurate estimation for the power law distribution and hence cannot be trusted.

The new strategy adopting the Maximum likelihood Estimation (MLE) which is claimed to provide solid and accurate estimation of distribution functions is widely used for the probability distribution studies of the complex system, such as Marta et al. (2008), Jiang et al. (2008) etc. The following mathematical illustrations will help understand how the MLE method would be applied and how the parameters would be extracted. The illustrations are based on some scholars’ work (Goldstein et al., 2004; Newman, 2006; Clauset et al., 2007; White et al., 2008).

The power law distributions actually consist of two circumstances due to the data type: continuous power laws and discrete ones. Continuous power law deals with the continuous real variables, such as distance and speed while the discrete ones take care of discrete sets of values, such as building heights, population etc. The following algorithm and equation deductions are referred to Newman (2006) and Clauset et al. (2007), who had given explicit descriptions.

First, let us focus on the continuous one. Assume the quantity x has a continuous power law distribution. Hence, the probability distribution function \( P(x) \) has such a mathematical property that

\[
P(x) = p(x)dx = \Pr(x \leq X < x + dx) = Cx^{-\alpha}dx
\]  

(3-3)

Here X represents the quantity from the empirical dataset, C serves as the normalization constant, and dx is the interval. The Eq. 3-3 should integrate to 1 from the entire data range. Obviously, Eq. 3-3 will diverse when x trends to 0; therefore, there must be a threshold value \( x_{\text{min}} \) starting from which the distribution will integrate and follow a power law distribution form. Hence, the normalization constant C can be calculated as following

\[
1 = \int_{x_{\text{min}}}^{\infty} p(x)dx = C \int_{x_{\text{min}}}^{\infty} x^{-\alpha}dx = \frac{C}{1-\alpha} [x_{\text{min}}^{-\alpha+1}]_{\text{min}}
\]  

(3-4)

From Eq. 3-4, we can see the value of \( \alpha \) should be greater than 1, otherwise Eq. 3-4 will diverge. Practically, the value of \( \alpha \) is between 2 and 3 (Newman, 2006). So in case \( \alpha > 1 \), the C value is

\[
C = (\alpha - 1)x_{\text{min}}^{\alpha - 1}
\]  

(3-5)

Therefore, the original power law expression can be transformed as this

\[
p(x) = \frac{\alpha - 1}{x_{\text{min}}} \left( \frac{x}{x_{\text{min}}} \right)^{-\alpha}
\]  

(3-6)

Given an n number empirical dataset \( x_i \), the probability that it is most likely drawn from the Eq. 3-6 is

\[
p(x | \alpha) = \prod_{i=1}^{n} \frac{\alpha - 1}{x_{\text{min}}} \left( \frac{x_i}{x_{\text{min}}} \right)^{-\alpha}
\]  

(3-7)

Such probability is the term likelihood that data is draw from a power law distribution. As suggested by Newman (2006) and Clauset et al. (2007), the likelihood is transformed into logarithmic form for ease of calculation, which is

\[
\zeta = \ln p(x | \alpha) = \ln \prod_{i=1}^{n} \left( \frac{\alpha - 1}{x_{\text{min}}} \left( \frac{x_i}{x_{\text{min}}} \right)^{-\alpha} \right) = \sum_{i=1}^{n} \left[ \ln(\alpha - 1) - \ln x_{\text{min}} - \alpha \ln \frac{x_i}{x_{\text{min}}} \right]
\]
Since we wish to get the maximum likelihood value, we can set \( \frac{\partial \xi}{\partial \alpha} = 0 \), thus we get the estimated \( \alpha \) value as

\[
\alpha = 1 + n \left[ \sum_{i=1}^{n} \ln \frac{x_i}{x_{\min}} \right]^{-1} \tag{3-9}
\]

This is how \( \alpha \) has been extracted hereto from the MLE method. The scholars also examined the way to estimate the standard deviation of its errors; however, this thesis is not intended to illustrate it in detail.

In case of discrete type of power law distributions, since the interval value \( dx \) is an integer, the expression will be

\[
p(x) = \Pr(X = x) = Cx^{-\alpha} \tag{3-10}
\]

Applying the same criteria in continuous type, there also exists \( x_{\min} \), thus we find

\[
p(x) = \frac{x^{-\alpha}}{\zeta(\alpha, x_{\min})} \tag{3-11}
\]

Where \( \zeta(\alpha, x_{\min}) \) is called the Hurwitz zeta function (Newman, 2006) in the form as this

\[
\zeta(\alpha, x_{\min}) = \sum_{n=0}^{\infty} (n + x_{\min})^{-\alpha} \tag{3-12}
\]

Applying the same MLE process, we will finally get the exponent \( \alpha \) as below

\[
\alpha = 1 + n \left[ \sum_{i=1}^{n} \ln \frac{x_i}{x_{\min} - \frac{1}{2}} \right]^{-1} \tag{3-13}
\]

Note that, the exponent \( \alpha \) for the discrete type is an approximation, since the distribution will not integrate from its data ranges. Clauset et al. (2007) compared the estimation for continuous power law and discrete one and suggested that if \( x_{\min} \geq 6 \), these two estimations are approximately the same giving \( \alpha \) for both the continuous and discrete types.

So far, the estimation for the exponent \( \alpha \) using MLE is done. Base on above flow of inductions, we believe the MLE will provide solid and accurate estimation for \( \alpha \) in the initial step, and the following methods will consolidate the statements for power law estimations.

### 3.3.2 Goodness of fit

Kolmogorov-Smirnov Statistics
The actual processes for discreet and continuous data type are slightly different. Although the discreet circumstance cannot integrate from the entire data range, the goodness of fit testing for both types is almost the same. Therefore the latter illustrations will not be separated for each circumstance.

The process of goodness of fit is based on “Kolmogorov-Smirnov” statistics (KS for short). It is used to test if distributions of the hypothesized and empirical datasets are from the same distribution form. However, there do exist other methods which have the similar functions, for instance, the Chi Square goodness fit test. Massey (1951) suggested KS test would provide better result for estimating the errors compared to other approaches. The KS test will provide an alpha value representing the level to reject the hypothesized distribution (the expected best fit distribution), maximum value 1 to accept and 0 to reject. The actual measurement is based on CDFs by measuring the maximum distance from the hypothesized distribution against the distribution from the empirical data (D value). The mathematical expression is shown in Eq. 3-14, where \( S(x) \) is the CDF of the empirical data while \( F(x) \) represents the CDF of the hypothesized distribution.

\[
D = \max \left| S(x) - F(x) \right|
\]

(3-14)

In particular, Figure 3.5 will illustrate where D exists.

![Fig. 3.5 The maximum distance for the CDF of hypothesized data against the CDF of empirical data](image)

Here in the simple illustration, the black solid line represents the CDF of hypothesized data, while the dashed line represents the CDF of the empirical one and the black vertical line represents the maximum distance between these two distributions.

**Estimate the lower bound** \( x_{\text{min}} \)

In this context, the lower bound \( x_{\text{min}} \) was pre-assumed. Actually, it is vital important to estimate \( x_{\text{min}} \) properly: if it is too small, it may contain the non-power law in the final estimation and that will lead to inaccurate result according to the estimation process. Conversely, if it is too large, the risk of losing part of the power law under \( x_{\text{min}} \) arises. Clauset et al. (2007) suggested using a way using KS statistics to get a proper \( x_{\text{min}} \) value. Referring to the definition of KS statistics, the judgment on whether two distribution forms match well depends on the maximum distance (D) in between. In particular, for power law distribution, since it starts from \( x_{\text{min}} \), therefore, the D value is calculated as Eq. 3-15:

\[
D = \max_{x \geq x_{\text{min}}} \left| S(x) - F(x) \right|
\]

(3-15)
To locate a proper $x_{\min}$ value, apparently, we are looking for a value which minimizes the D value in Eq. 3-15 (Clauset et al., 2007). It is a numerical process iterating from a certain small value from the empirical dataset and calculating the expected minimum D value, and hence the $x_{\min}$ will be seized. The validation work proving this process would provide a suitable $x_{\min}$ value was done by Clauset et al. (2007), however, the detailed process of that will not be addressed here.

**Goodness-of-fit 1: P value test**

In the practical experiments, there are usually a variety of results from the experiments under different circumstances. To identify if they follow the same statistic form, the method “bootstrap” is brought in for help. As introduced by Shanahan (2001), the fundamental idea is based on the “Central Limit Theorem”. Quoting the important assumption: “If random samples of n observations $y_1, y_2, \ldots, y_n$ are drawn from a population of finite mean $m$ and variance $s^2$, then when n is sufficiently large, the sampling distribution of the sample mean can be approximated by a normal density with mean $\mu = \frac{\sum y_i}{n}$ and standard deviation $\sigma \sqrt{\frac{1}{n}}$” Shanahan (2001). The “bootstrap method” expresses itself in a similar way but uniformly extracting series of data from the very same data source many times. By the term “many times”, no accurate number for it though, theoretically, as many as possible (the latter thesis study processed 1000 times) (Clauset et al, 2006; Shanahan, 2001). To apply the shrewd method in this thesis, we need to generate massive synthetic data mimicking the data from the hypothesized distribution models. Thanks to the random number generator, which is one of Clauset et al. (2007) suggested several models for the synthetic data generating. To generate the synthetic data that follows a power law, it shall follow the Eq. 3-16 below:

$$x = x_{\min}(1 - r)^{1/(\alpha - 1)}$$  \hspace{1cm} (3-16)

Here in Eq. 3-16, $r$ were random numbers uniformly distributed between 0 and 1, $\alpha$ is the exponent value from the MLE estimation process and $x_{\min}$ is the lower bound of the power law distribution. Given such parameters, the synthetic data generator will offer a series of data $x$, which follows the power law distribution. Clauset et al. (2007) also provide the algorithms to generate other synthetic data for other power law alike (heavy-tailed) distributions, such as exponential, stretched exponential and power law with an exponential cut-off e.g. for comparisons with power law. The detailed algorithms will be attached in the appendix.

Conforming to the theory of KS statistics and the idea of using “bootstrap” process, we disassemble the course into steps following: Apparently, the empirical data is separated into two parts based on its attributes, the lower part where $x < x_{\min}$ does not follow the power law distribution while the upper part where $x \geq x_{\min}$ is supposed to (Clauset et al., 2006). Therefore, assume that the total number of original data is $n$, and from the MLE estimation, there is $n_{\text{tail}}$ data with $x \geq x_{\min}$. For the data in $x \geq x_{\min}$, with the possibility $\frac{n_{\text{tail}}}{n}$, we generate the synthetic data using the very same random data generator and with the exact same scaling parameters: $\alpha$ and $x_{\min}$ by adopting the MLE method again to make sure getting the identical $\alpha$ and $x_{\min}$ value. For the lower part where $x < x_{\min}$, with the probability $1 - \frac{n_{\text{tail}}}{n}$, we choose an element from the particular data range uniformly at random. By applying the “bootstrap” process: for each loop within the process, we reuse exact data elements from the lower part ($x < x_{\min}$) but normalize them together with the ones for the upper part. By congregating and sorting the two synthetic parts, the preparation of generating the synthetic data for KS statistic test is ready.

Conforming to the KS statistics, the best fit for the hypothesized distribution to the empirical data distribution relies on the distance between the two distributions which is shown in the Eq.3-15:
To make solid validation of the goodness of fit, we generated enough synthetic datasets (1000) in this study. In each testing roll, under the process of ‘bootstrap’, we used the list of synthetic data q(x), starting from the minimum value q for the first time calculation of D, then the next process will strip to a new list excluding the minimum value to form a new list calculating the D value again until we got the entire list of D. We then store the minimum value of D for each round of calculation. The purpose for carrying out the specific calculation is to minimize the errors that would cause biased result for D value in each time calculation. After the 1000 times loop calculation, we eventually get a list the all D values. As we marked our primary D value from the distribution of empirical data against the best fit power law distribution, we count the number of $D_i > D_e$ and divide the number of total $D_i$. The proportion is the expected P value for the goodness of fit, whose formula is shown in Eq. 3-17.

$$P = \frac{\#(D_i > D_e)}{\#D_i}$$ (3-17)

As we can see from the equation, the bigger value of P indicates that the synthetic generated data under an exact exponent was not fitting as well as the empirical data; this reflects the fact that the hypothesized fit for the empirical data is at least not bad. Because of the solid support by using “bootstrap” process, such statement shall be trusted. On the other hand, if the P value is sufficient small, note there is no absolute threshold for this, for instance, less than 0.1 (Clauset et al., 2006) or less than 0.01 (Marta et al., 2008), the hypothesized power law distribution will be ruled out. As a matter of fact, such test will rule out the hypothesized fit if the P is significant small, while, a quite big P value only states the hypothesized fit is at least not a bad one.

**Goodness-of-fit 2: Alternative competitive distribution test**

For a small P value (less than 0.1, for instance), we can confidently rule out the distribution without worrying about losing a good fit. However, with respect the weaknesses of the P value test that it is good at ruling out the hypothesized distributions but not to rule them in (Clauset et al., 2007; Marta et al., 2008). Simply speaking, one can conclude on the basis of high P value that it is at least not a bad distribution or a candidate good fit. No one can guarantee the possibility; in this case, the alternative distribution tests are promoted to make a more solid conclusion (Clauset et al., 2007). According to the “heavy-tailed” distributions, so far, there are only several popular distributions well known: exponential, weibull, power law, log normal, and power law with cut-off. These mentioned distributions also hold a long tail in the normal scale plot and when turning to the log-log scale plot, they look very alike the practical power law distribution which are hard to be distinguished. As is shown in Figure 3.6, there are 4 examples of such distribution forms.

Here in the figures: Figure a contains an exponential distribution of the data, Figure b for a log-normal distribution, Figure c for a stretched exponential distribution which also refers to weibull distribution, and in Figure d, it is a particular case for comparing with the power law distribution in Figure e. In Figure e, it is the plot of the data which actually follows the power law distribution, and the black solid line indicates a perfect power law as a straight line. The emphasis on the power law distribution with an exponential cut-off is because of its mathematical expression in Eq. 3-18:

$$P(x) \sim x^{-\alpha} e^{-\lambda x}$$ (3-18)

The expression shows that it is similar with the power law distribution when x is sufficient small, since the rightmost part of the equation will not impact much. When x becomes larger, it will be showing a faster dropping tail, which is so called exponential cut-off (Adler et al., 1998; Goldstein et al., 2004). It is still one type of power law distributions referred as truncated power law distribution, and it has been widely captured in power law related studies (Rhee et al., 2007, Marta et al., 2008). Therefore, the existences of such distributions bring challenges for identifying the best fit model for the empirical data, and hence, the evaluation of the competence shall be carried out.
The essential idea of performing the alternative distribution test is to compare the likelihood from each distribution model by computing the ratio. The one with higher likelihood is to be a better fit (Clauset et al, 2007). Vuong (1989) proposed a method that can quantitatively estimate the estimate the standard deviation of the likelihood ratio. Assume we have two distributions in form as below:

\[
L_1 = \prod_{i=1}^{n} p_1(x_i) \quad \text{and} \quad L_2 = \prod_{i=1}^{n} p_2(x_i)
\]  

(3-19)

And then the ratio of the likelihood is calculated from Eq. 3-20:

\[
R = \frac{L_1}{L_2} = \prod_{i=1}^{n} \frac{p_1(x_i)}{p_2(x_i)}
\]  

(3-20)

Normally, as is suggested by Vuong (1989) and Clauset et al. (2007), the R is in logarithmic format for the status whether it is positive or negative will get the better fit easier. The logarithmic likelihood is expressed in Eq. 3-21:

\[
R_i = \sum_{i=1}^{n} \left[ \ln p_1(x_i) - \ln p_2(x_i) \right] = \sum_{i=1}^{n} \left[ \ell_i^{(1)} - \ell_i^{(2)} \right]
\]  

(3-21)

Here in the equation, \( \ell_i^{(1)} \) and \( \ell_i^{(2)} \) is the logarithmic likelihood of distribution models respectively. For the ease of concluding a better fit based on the R value, the normalization for the logarithmic likelihood ratio is necessary which is \( \frac{R}{\sqrt{n}\sqrt{\sigma}} \), where \( \sigma \) is the standard deviation value for logarithmic R. The ultimate decision will be based on R in case of sufficient positive or negative.
The goodness of fit for alternative distributions will take place when the P value test for power law distribution is not significant small, say larger than 0.1, otherwise the candidate power law distribution is ruled out already. The sign of R indicates which distribution is a favored one for the empirical data. In particular, since these candidate distributions are against the power law distribution, \( L \) represents the power law distribution always. In case of logarithmic scale, R value between -1 and 1 does not provide strong evidence which one would be favored, therefore, the judgment will based on p value only. Otherwise, if R is significantly smaller than -1, it implies the candidate distribution is more favored by the empirical data than the hypothesized power law distribution. Contrariwise, if R is significantly greater than 1, the power law distribution will be confirmed. It happens that more than one candidate distributions have R value either greater than 1 or less than -1; the comparisons among these candidate distributions should also be measured. Thanks to the calculation of R taking the power law distribution as a medium (each R is calculated involving power law distribution), such measurement will be based on the ratio of R values as well, the one with a larger absolute R value is supposed to be a better fit than the opponent.

3.3 Visual strategies

3.3.1 Procedures for k-core decompositions and graphical representations

To realize the visual strategies for study, this thesis will address it analyzing the decompositions of the natural street networks produced above, in particular, based on the k-core decomposition algorithm. The procedure for generating the k-core decompositions can refer to Batagelj and Zaversnik (2002). Simply saying, it is a “recursive cutting off” process, starting from the vertices with least degree value and identifying the shells which these vertices belong to. Once this has been done, these vertices will be cut off from the network and a higher level of network is ready for the same procedure. The entire process will be finished when the vertices within the particular k-core has the least degree greater than k, which means the innermost core is seized (A-Hamelin et al., 2005; Batagelj and Zaversnik, 2002). The k-core algorithm can thus be expressed as: the graph \( G = (V, L) \), where V for vertices and L for edges. Therefore the k-core can be obtained as \( k(v, u) = k(v, N(u)) \), by assuming the network in the form of \( N = (V, L, w) \), where w represents the function to assign the values for shells (Batagelj & Zaversnik, 2002). The definition of \( N(v) \) denotes the table for the neighbors of a particular vertex (v), with the property that \( N(v) \subseteq V \).

A-Hamelin et al. (2005) implemented the k-core decomposition algorithm in LaNet-vi with a two dimensional graphical layout, whose example has been mentioned in the above chapters. The detailed algorithms for calculating the k-core decompositions and drawing the graphics in LaNet-vi are not illustrated here, since this thesis only adopts it as a tool for network visualization analysis. Moreover, it offers a control panel for setting up the parameters to control the layout in the graphical representations, such as background, edges control value, graph color value etc., which will let the users require their different focuses. The layout from LaNet-vi will be capable of supplying rich information about the “fingerprints” of the large networks we are dealing with. For examples, the existence of hierarchy structure, the evaluations of the centrality for those networks.

3.3.2 Illustration of the graphical representation from LaNet-vi

Once the graphical representation of the k-core decompositions for the large networks is ready, the analysis based on the visual effects shall be subsequently on schedule. Therefore, the recognitions about the revelations from the layout are important for the further studies. As is shown in Figure 3.7, there are two types of layouts from LaNet-vi representing two different k-core decomposition structures respectively. Both figures have the legends using color starting from violet for lowest value of coreness and versa red for the highest. And the degree legend to the leftmost represents the size
logarithmic proportional to the original connectivity degree of nodes in the systems. The detailed information about these elements can be found in previous chapters.

Moreover, there is one significant difference between the two figures: for the figure to the left, it is a single component structured decomposition layout. In contrast, the one to the right has two existing components innermost. In the case of multiple components, the diameter of the each shell differs according to number of inner vertices, in particular seen from the figure: the one with more inner vertices has a large diameter. It is due to the process of k-core decomposition calculation, when the nodes are ruled out from the k-core, they do not belong to the same connected component which is also referred as fragmentation of the network (A-Hamelin et al., 2005). Therefore, the layouts provide us the opportunity to see whether the network is structured as single component or fragmented into multiple components, hence the topological pattern will be revealed.

### 3.3.3 Combinations with LaNet-vi and GIS

The layouts from LaNet-vi can tell the topological properties of the networks about how the elements (vertices) decomposed in the system. However, on the other hand, the actual geometry information and tags are eliminated. While sometimes, these sets of information are important indeed. To deal with such dilemma, the feasibility of retrieving the network visualization back to GIS is anticipated. Depending on the capability of ArcGIS desktop, which is a set of GIS software, the spatial information of each street is saved in the database, once the “ID” of any street is called, the corresponding information will be retrieved. Therefore, matching the ID of street in GIS and the ID of vertices in network visualization will connected these two applications together.

Once the retrieving process in available, both the topological and geometrical properties of each street are ready for further study. The corresponding k-core decompositions will actually be seen in the map, realizing the goal of mapping the complex networks and visually seeing the decompositions of the complex networks.

### 3.4 Chapter summary

In this chapter, it consists of the methodologies and methods for the studies of the topological and geometrical patterns of the natural street works from an analytical and visual perspective. The procedures for producing three natural street models are introduced for preparing the natural street networks as complex networks. MLE methods are promoted for estimating the power law distributions with comparisons of conventional approaches. The methods using goodness-of-fit tests for both the P value test and the alternative heavy-tailed distributions being a favored fit are prepared, which intends to make solid conclusions about the existence of power law distributions. On the other hand, the
procedure for getting the k-core decompositions from the complex networks is introduced, and the possibility of retrieving such topological patterns into GIS is also outlined.
4. The 40 US urban street networks for case study

In this chapter, the particular case study will be carried out based on the 40 US urban street networks. In particular, it will study the topological and geometrical patterns of the reformed natural street networks in an analytical and visual perspective. On one hand, we can study the complex properties of the street networks within the GIS environment and explore the possibility of network visualization based on the interrelationships discovering certain patterns of the networks. On the other hand, this particular case study will sever as a revised study for re-examining the topological patterns of the 40 US urban street networks in previous studies (Jiang, 2007). First of all, the proposed flow chart showing how the case study would be performed is shown as below:

**Fig. 4.1** The proposed flow chart for the case study of 40 US urban street networks

4.1 Datasets and software

The datasets adopted in this thesis were derived from the Topologically Integrated Geographic Encoding and Referencing (TIGER) database which is developed by the US Census Bureau in
shapefile format (ESRI, 2008). Jiang (2007) had experimented and analyzed the topological patterns over 40 sample datasets. In particular, regarding to the forthcoming comparisons between the previous work and the re-examination, it is the identical 40 US cities will be used in the case study. The order of figures is descending sorted by the size of the cities, so does the coming results. Note that, the geometry length is represented with degrees constrained by their original projections; however, it would not affect the analysis since the relative relations are not changed. The figures of those shapefiles are listed in appendix.

To carry out the investigations dealing with the datasets, a set of GIS software, ArcGIS 9.2 Desktop, is adopted as the platform. It will be used to generate the “nature streets” from the 40 US urban street networks, and export the topological relationships for analysis. The large network visualization tool (LaNet-vi) developed by Alvarez-Hamelin et al. (2006) will be adopted creating two-dimensional layouts representing the k-core decompositions of the natural street networks. Together with ArcGIS, the geometry and topological information will be attached into scene.

Since the mathematical process is extremely important to detect and verify the power law distributions, the analytical strategy will benefit from two sets of scientific computing software: Matlab and R. More specifically, Matlab will deal with detecting the existence of power laws and estimating the parameters, while R will perform the selection for a better fit against power law among several other candidate heavy-tailed distribution models.

4.2 Generating the natural streets

In each shapefile, the streets are represented as polylines with attributes of length and names. However, the information of names are incomplete, for instance there are 7699 streets out of total 225346 missing names. Therefore, it is an obstacle for the named street model, since the specific model will summary the structures and properties all based on the name attribute (Jiang et al., 2008). In this case, using natural street models for re-generating the street networks from the giving datasets is highly anticipated, regarding to its good performance from previous studies about the traffic studies (Jiang et al., 2008) and human mobility studies (Jiang et al., 2008). For this case study, the three natural street models will be applied respectively with a threshold of the deflection angle as 45 degree. To make fair comparisons with the previous studies, the threshold value of 70 degree will be applied as well. In particular, since the previous study also adopted the SBF natural street model and such model showed a best performance against the other two in traffic flow prediction (Jiang et al., 2008), hence, the SBF model is chosen for both analytical and visual studies. Nevertheless, the EBF and SF models will be produced as well but only for analytical studies.

4.2.1 Checking isolations

As is known that the natural street tracking process is a recursive process, there should be no isolated streets. Otherwise, the tracking will stop at any isolated street and terminate the entire process. Fortunately, most of the 40 US urban streets only contain few isolated streets; taking Los Angles and Chicago for examples, there are only 93 and 87 streets out of 225346 and 198149 total number (0.41% and 0.44%) respectively are isolated to the main part, which will hence cause little negative impact for the studies, especially compared with the named stroke model. The process of getting rid of the isolated streets is implemented in ArcGIS desktop described as below.

Input: complete street shape file
Output: integrated street shape file without isolations
The Main function
Sub Start ( )
Start with any street from the main part where streets Loop to search the intersected streets
If connected streets more than one
   Tag the chosen street with true
   Start from the chosen street and perform the recursive process
Otherwise
   Start with any street from the main part where streets Loop to search the intersected streets
   If connected streets more than one
      Tag the chosen street with true
      Start from the chosen street and perform the recursive process
   Otherwise
4.2.2 Transforming the datasets into segment based urban streets

The datasets are “clear” to perform the natural street tracking process, when they are free of isolated items. However, such tacking process will make every decision at the junction based on the deflection angle to choose which way to continue. In this case, the segment based street networks are needed. It can be actually done with the help of ArcGIS desktop using the Data Interoperability function, which will convert the intersections of streets into junctions and take every piece of street between the junctions as a segment. Figure 4.2 shows how they are represented in GIS.

![Fig. 4.2 Converting streets into segments](image)

Here in Figure a, the highlighted street is converted into segments in Figure b, in which the red points in the circle represent the junctions where streets intersect.

4.2.3 Three natural street models

The EBF natural street model will produce only one unique set of street network for the entire city. Because the criteria for making decision for each segment to continue with another one will not only consider the deflection angle but also concern the good continuation for each other segment at the junction. In contrast, the SBF and SF model will produce multiple sets of natural street networks due to randomly selecting a segment to start. To re-examine the previous work which was also a SBF process, synchronizing the procedure of the tracking process in this case study is necessary. It shall be accomplished by all starting from the top to button in the database. Actually, the only difference among the three models will only happen under the circumstance that there is a better fit. An extreme case will be that all three models will be identical if the streets are perfect naturally formed and there exist no randomness. However, the point is that the general tracking process is almost the same. The following pseudo code will show how the process implemented. The detailed information about all three models can refer to Jiang et al. (2008).

```
Input: Segment-based street shape file
Output: Stroke-based street shape file
The Main function
Sub Start ( )
Start for the first segment in the database
While (not last segment) do
  If (the current segment is processed) then
  End
End sub ( )
```
Go to next segment
Else
Change the status of that segment to be processed;
Get a new segment by calling function ‘SearchSegmentByPoint’ with the old segment and it’s ‘from’ direction as parameters;
Get another new segment by calling function ‘SearchSegmentByPoint’ with the last new constructed segment and it’s ‘to’ direction as parameters;
(Differences among three stroke models will be arising here)
Create a stroke with final constructed segment and add to the stroke-based shape file;
End
End while
End sub

The segments will hence be merged into natural streets correspondingly after each tracking process. Figure 4.3 shows how a natural street would be like.

![Fig. 4.3 A simple example showing how a stroke generated from the segments](image)

Here in Figure a, it is a simple segment based street network and the highlighted segments are ready for the natural street tracking process. In Figure b, the highlighted polylines represent a natural street from the segments generated from a certain natural street model.

### 4.2.4 Extracting the topological interrelationships and geometric properties

Generating the natural street networks is just the first step for the complex network studies in GIS environment, for the natural streets are merely geometrical representations, which are not capable of supplying sufficient information about interrelationships among the features. As mentioned above, a topological network modeling the GIS environment will serve to characterize the dynamics (Jiang and Claramunt, 2004). Thanks to the capability of ArcGIS desktop, we can program inside to get such topological relationships by examining how each natural street intersected with others. Meanwhile, the particular geometrical properties: the length of each natural street will be exported for further analysis. The following pseudo code shows how the work is performed.

**Input:** Stroke-based street shape file  
**Output:** Table for the records of interrelationships and geometric length

**The Main function**

```plaintext
Sub Start ( )
Get all the strokes;
Start for the first stroke in the database
While (not last stroke) do
    Count the number of the intersected strokes with the current one
    Note the records of the intersected strokes in array-list
    Get the geometry length of the current stroke
    Export these sets of information into two text file: one for connectivity values parallel with length value, one for interrelationships in net file format
End while
End sub
```
Here in the pseudo code, the connectivity degree value of each natural street is derived by counting the number of other streets that are intersected with it. And the geometrical length is in degree format, while the net files have the records summarizing the interrelationships, for instance, if street 1 intersects with street 2 and street 3, the records will be shown as “street1 street2” and “street 1 street3” in separate lines.

4.3 Analytical perspective: power law estimations

4.3.1 Qualitative exploration of the 40 US urban street networks

Before estimating the empirical data seeking the power law distributions, a qualitative exploration over the datasets is necessary to have an overall impression (connectivity degree and length of streets in particular). It will rely on determining if the 80/20 principle exists in the natural street networks, which is qualitatively see if the quantity of elements that have a value above the average make up around 20% of all, while the ones below the average are around 80%.

The average value of the quantities in the datasets is calculated first and it will separate the datasets into two parts: in this case study, streets with a connectivity degree or length value larger than the average value and the ones below the average. Afterwards, the proportion is summed up over all the 40 dataset respectively. This will allow us to have a brief overview on the structure of these 40 street networks and to foresee the potential for power law distribution existence, for the power law exists when the empirical data dramatically diverse (Cesar, 2008).

4.3.2 Power law estimation for parameters: \( x_{\text{min}} \) and \( \alpha \)

As mentioned above, the first step for estimating the power law distribution is to get a proper \( x_{\text{min}} \) and hence extract the exponent \( \alpha \) value. The \( x_{\text{min}} \) can be seized by using the numeric estimation method, which is: for each possible \( x_{\text{min}} \), we calculate the corresponding \( \alpha \) value by using the MLE method. The dataset above \( x_{\text{min}} \) is supposed to follow the power law distribution; hence the KS test will be expressed as is shown in Eq. 3-15. Therefore the then loop from all the possible \( x_{\text{min}} \), the one that will give minimum D value will be chosen as the expected value.

The source code implementing each individual function mentioned above is provided by Clauset et al. (2007). For this particular case study, there are two kinds of data types: discrete (connectivity) and continuous (length), the code in Matlab will automatically distinguish the data type and apply proper methods estimating \( x_{\text{min}} \) and \( \alpha \). In particular, we choose 1000 sample data uniformly from the empirical data for calculation each run time instead of accounting for the entire dataset to enhance the speed. As an output from the calculation, the CDF for the empirical data will be plotted in log-log scale and the power law fit line will be attached as well. Taking Figure 4.4 for example, in the log-log scale plot of Figure a, the blue circles represent the CDF for the empirical data while the black solid line represents the power law distribution with the exponent value of \( \alpha \). The sample data we used in this figure were generated by the random number generator with an \( \alpha \) value of 2.5. Since the data was synthetically generated, the power law fits it very well. Actually, the top end of the black solid power law line corresponds to the \( x_{\text{min}} \) in the horizontal axis, since the power law starts to fit from \( x_{\text{min}} \). Figure b serves as another example of power law fit but starts from a larger \( x_{\text{min}} \) for comparison with the one in Figure a.
4.3.3 Goodness of fit 1: P value test

Once $\alpha$ has been located, the shape of the distribution is fixed. However, we still have to estimate how well or bad the power law distribution fits the empirical data. As the P value test we mentioned above can only rule out the hypothesized distribution model if the P value is sufficiently small. To make a solid conclusion about the P value, 1000 time repetitions are performed for the “bootstrap” process. The difference for continuous and discrete type is slightly different when implementing the code in Matlab, whose pseudo code is shown below.

Input: 40 tables for the records of connectivity and street length
Output: exponent $\alpha$, $X_{\text{min}}$, P value and power law fit plot in log-log scale for each urban street

The Main function
Sub Start ( )
Get the directory of the 40 tables;
Start for the first table
While (not last table) do
Read the data into Matlab
Calculate the $X_{\text{min}}$ value and extracting $\alpha$ for both discrete and continuous type
Perform the p test repeating 1000 times using bootstrap method
Save the results to a new table
Plot the power law fit in log-log scale plot
Next table
End
End sub

4.3.4 Goodness of fit 2: Alternative distribution competing test

A pretty large P value does not guarantee the distribution of the empirical data is drawn from the power law distribution, for there are a series of competitive heavy tailed distributions, which have been mentioned above. Therefore, the alternative distribution test competing for being a more favored fit should be performed as well (Clauset et al., 2007). In this case study, regarding to the different data types, two series of alternative distributions are prepared to compete with the power law distribution. For continuous type: log-normal, exponential, stretched exponential (weibull), power law with cut-off against power law distribution (Pareto in this case), and for discrete type: Poisson, log-normal, discrete exponential, discrete stretched exponential and discrete power law with cut-off against power law distribution (Zeta distribution in this case), will compete with the power law distribution respectively.
The final judgment for choosing a favored distribution is based on the logarithmic R value introduced by Vuong (1989). The pseudo code illustrating such procedure is shown below.

Input: 40 tables for the records of connectivity and street length
Output: p value for possibility being better distribution and log ratio of likelihood from Vuong method

The Main function
Sub Start ( )
Get the directory of the 40 tables;
Start for the first table
While (not last table) do
Read the data into R
Calculate the $x_{min}$ value
Perform the power law fit and other hypothesized distributions for calculating the maximum likelihood
Calculate the logarithmic ratio of likelihood
Save the results to a new table
Next table
End
End sub

4.4 Visual Perspective: Network visualization for GIS

4.4.1 K-core decomposition in LaNet-vi

The network visualization based on the k-core decomposition algorithm is implemented as a software package in LaNet-vi, which is a set of software developed by A-Hamelin et al. (2005) for the sake of visualization for large scale networks. In this case study, the natural streets from the 40 US urban street networks are represented as vertices, while the interrelationships are represented by the netfile, which is a predefined file format for the software Pajek. The most important feature it has is that each interrelationship is represented as an edge with a pair of elements ID for its two ends. A simple example representing the interrelationships for 4 nodes is shown below.

```
*Vertices 4
1
2
3
4
*Edges
1 2
1 3
2 3
2 4
```

Fig. 4.5 (a) A simple netfile example, (b) The corresponding plot from Pajek

Here in Figure a, it is a simple netfile for 4 vertices and the records of connecting status. In Figure b, it is the plot for visualization. In particular, the above netfile only represents undirected inter-connections corresponding to the representations under *Edges in Figure a.

Moreover, the 40 US urban street networks are also in undirected format, since there is no direction involved about how these streets intersected with others. The netfiles for the 40 networks are extracted along calculating the connectivity degree values since that process will loop the entire network and make a record of the interrelationship for each street. The network visualization process is then carried out in LaNet-vi. To display the results more clearly, the parameters for setting up the links are set
sufficiently small; therefore, the links (edges) will not be shown in the final layout. Figure 4.6 will show how the process is carried out in LaNet-vi.

![Figure 4.6](image)

**Fig. 4.6** (a) The working environment in LaNet-vi (b) The 2-D layout for k-core decomposition

Here in Figure a, it is the environment for the network analysis workbench (LaNet-vi), and the parameter “Edges” for controlling the display of edges in the layout is set sufficient small, 0.001 in particular. While in Figure b, it is the k-core decomposition layout with the colorful legend showing the index for k-core decomposition and the white spots representing the size, which is logarithmic proportional to the original degree of the vertices. The format for layout in Figure b is configured in LaNet-vi, by specifying the color of background, the height and width of the image, etc.

### 4.4.2 K-core decomposition retrieving back to GIS

The network visualization will be helpful identifying how these decompositions distribute within the network and exposing the patterns about the hierarchy structure and diversity etc. However, these are based on the topological relationships where the spatial information is not involved. Thanks to the database in ArcGIS desktop, where the spatial information is stored, it is possible to retrieve the topological properties back to GIS by joining the same index value of the vertices in the topological network and the record ID value of the streets in GIS. In LaNet-vi, identifying the node index can be done by using the annotating process, which will attach an ID to each vertex identical to its corresponding street.

The annotated table with the vertex ID will be appended with the corresponding coreness value in LaNet-vi. After loading the annotated table into ArcGIS desktop, the “join” function will recognize the vertex ID and the street ID and match each pair together. The coreness value is therefore associated to each street. The last but not the least step is to visualize the k-core decomposition in ArcGIS desktop rendering the map as in LaNet-vi. This is the real meaning of featuring the complex network with maps. The join process is illustrated in Figure 4.7.
Fig. 4.7 (a) A simple illustration for join the node ID with street ID (b) Rendering the map with k-core value in GIS

Here in Figure a, it is the join process in ArcGIS desktop to associate the information of k-core decompositions to the streets, and Figure b is the visualization effect from coreness value rendering the map. From this basis, we can identify the network visualization with spatial information, which means we can actually see how these decompositions distributed and where the hierarchy structure located e.g. Together with the analytical approach, we may also “see” the diversity of the natural street networks as complex networks. Such possibility will be discussed in the following chapters.

4.5 Chapter Summary

In this chapter, the case study about the topological patterns of the natural streets from 40 US urban street networks is performed from analytical and visual perspectives. The study procedures follow the methodologies and methods mentioned above. The analytical approach consists of estimating power law distributions using MLE and performing goodness of fit for both P value test and alternative distribution test. Moreover, the detailed methods and procedures on how the work is done are described with either pseudo codes or full text. The visual approach uses the network visualization for the topological patterns based on the k-core decomposition algorithm in LaNet-vi and retrieves the spatial information in GIS.
5. RESULTS

In this chapter, the results drawn from the case studies carried out above will be brought forth. The study explores the topological patterns of 40 US urban street networks from analytical and visual perspectives. Moreover, three types of natural street networks are produced for the study framework.

On one hand, from the analytical perspective: be based on the connectivity degree and length of the natural street networks, the results about the power law estimation are separated into two categories because of the data types: discreet and continuous. Before digging in to seek the existences of power laws, qualitative overview of the structure of the data is performed by examining the existence of 80/20 principle. Then, the results for power law estimations will be illustrated respectively supported by the CDF in log-log scale plots with the hypothesized power law fit (a straight line in log-log scale). Moreover, the tables with the exponent α for shaping the power laws and the P value for the level of goodness of fit are prepared. In addition, the results about the alternative distributions selecting a more favored fit are drawn to consolidate the power law estimations and validations.

On the other hand, from the visual perspective: utilizing the k-core decomposition representations from LaNet-vi, we can actually see how the natural street networks decomposed by detecting the cores. The natural street networks then are interpreted by annotating the k-core values and retrieving the spatial information in GIS.

5.1 Analytical perspective

5.1.1 Qualitative overview of the structure of 40 urban street networks

The qualitative overview the structure of the natural street networks is evaluating the quantities that have a connectivity degree or length value above the average level respectively. Table 5.1 shows such results including the case from previous studies (Jiang, 2007) and the three types of natural street networks generated in this thesis.

<table>
<thead>
<tr>
<th>City/Partition</th>
<th>SBF 70</th>
<th>EBF 70</th>
<th>SBF 45</th>
<th>EBF 45</th>
<th>SF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Los Angeles</td>
<td>4(21%)</td>
<td>5(20%)</td>
<td>5(24%)</td>
<td>5(22%)</td>
<td>5(24%)</td>
</tr>
<tr>
<td>Phoenix</td>
<td>4(15%)</td>
<td>4(24%)</td>
<td>4(24%)</td>
<td>4(24%)</td>
<td>4(24%)</td>
</tr>
<tr>
<td>Chicago</td>
<td>5(18%)</td>
<td>6(25%)</td>
<td>6(26%)</td>
<td>6(25%)</td>
<td>6(25%)</td>
</tr>
<tr>
<td>Houston</td>
<td>4(19%)</td>
<td>5(28%)</td>
<td>5(26%)</td>
<td>5(28%)</td>
<td>5(26%)</td>
</tr>
<tr>
<td>Pasadena</td>
<td>4(20%)</td>
<td>5(28%)</td>
<td>6(24%)</td>
<td>5(29%)</td>
<td>6(25%)</td>
</tr>
<tr>
<td>San Diego</td>
<td>4(16%)</td>
<td>4(23%)</td>
<td>5(23%)</td>
<td>4(24%)</td>
<td>4(24%)</td>
</tr>
<tr>
<td>Hollywood</td>
<td>4(16%)</td>
<td>4(24%)</td>
<td>4(25%)</td>
<td>4(25%)</td>
<td>4(25%)</td>
</tr>
<tr>
<td>Dallas</td>
<td>5(16%)</td>
<td>5(23%)</td>
<td>6(25%)</td>
<td>5(24%)</td>
<td>6(26%)</td>
</tr>
<tr>
<td>Arlington 61</td>
<td>4(21%)</td>
<td>5(24%)</td>
<td>5(25%)</td>
<td>5(24%)</td>
<td>5(26%)</td>
</tr>
<tr>
<td>Las Vegas</td>
<td>3(22%)</td>
<td>4(22%)</td>
<td>4(22%)</td>
<td>4(22%)</td>
<td>4(22%)</td>
</tr>
<tr>
<td>Sunnyvale</td>
<td>5(16%)</td>
<td>4(24%)</td>
<td>4(25%)</td>
<td>4(24%)</td>
<td>4(24%)</td>
</tr>
<tr>
<td>St. Petersburg</td>
<td>4(19%)</td>
<td>5(28%)</td>
<td>4(25%)</td>
<td>5(28%)</td>
<td>4(25%)</td>
</tr>
<tr>
<td>Detroit</td>
<td>6(18%)</td>
<td>6(26%)</td>
<td>8(26%)</td>
<td>6(23%)</td>
<td>8(25%)</td>
</tr>
<tr>
<td>San Antonio</td>
<td>5(16%)</td>
<td>5(30%)</td>
<td>5(26%)</td>
<td>5(26%)</td>
<td>5(31%)</td>
</tr>
<tr>
<td>New York</td>
<td>7(21%)</td>
<td>8(26%)</td>
<td>8(25%)</td>
<td>8(25%)</td>
<td>8(25%)</td>
</tr>
<tr>
<td>Birmingham</td>
<td>4(18%)</td>
<td>4(26%)</td>
<td>5(26%)</td>
<td>4(27%)</td>
<td>5(26%)</td>
</tr>
<tr>
<td>Tacoma</td>
<td>3(20%)</td>
<td>4(21%)</td>
<td>5(23%)</td>
<td>4(21%)</td>
<td>5(24%)</td>
</tr>
<tr>
<td>Sterling Heights</td>
<td>4(16%)</td>
<td>4(25%)</td>
<td>5(26%)</td>
<td>4(25%)</td>
<td>5(26%)</td>
</tr>
<tr>
<td>Moreno Valley</td>
<td>4(15%)</td>
<td>4(23%)</td>
<td>5(23%)</td>
<td>4(23%)</td>
<td>5(23%)</td>
</tr>
</tbody>
</table>
**5.1.2 Power-law fit**

For the ease of illustration, the figures for power law fit for the SBF natural street model with a chosen angle of 45 degree are shown in Figure 5.1 and Figure 5.2 for length and connectivity respectively. In each figure, the results for the natural street networks of 9 top sized cities are presented; the rest will be shown in the appendix. Additionally, the other figures for the analysis of power law fit for the other two natural street models will also be shown in the appendix.

In each plot, both the horizontal and vertical axes are scaled with log10, which is known as the log-log plot. In particular, the number in the horizontal axis represents the actual value of the measured variables (connectivity and length), while the vertical axis consists of the value from the CDF, which is labeled with blue circles. Meanwhile, within each plot, the black dashed line (a straight line in log-log scale) represents the hypothesized power law fit. Even the power law fit cannot be confirmed merely based on what we see from the plots, it does visually offer general impression how well the power law fits the empirical datasets.

<table>
<thead>
<tr>
<th>City</th>
<th>Partition (Length)</th>
<th>Partition (Connectivity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Louisville</td>
<td>4(25%) 5(27%)</td>
<td>4(26%) 5(27%)</td>
</tr>
<tr>
<td>Dayton</td>
<td>4(20%) 5(30%)</td>
<td>5(27%) 5(30%)</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>6(18%) 7(19%)</td>
<td>7(19%) 6(24%)</td>
</tr>
<tr>
<td>Greensboro</td>
<td>4(16%) 4(25%)</td>
<td>4(25%) 5(25%)</td>
</tr>
<tr>
<td>Gray</td>
<td>5(16%) 5(22%)</td>
<td>6(25%) 5(22%)</td>
</tr>
<tr>
<td>Little Rock</td>
<td>4(21%) 5(30%)</td>
<td>5(26%) 5(31%)</td>
</tr>
<tr>
<td>Spokane</td>
<td>5(20%) 6(26%)</td>
<td>7(26%) 6(26%)</td>
</tr>
<tr>
<td>Bakersfield</td>
<td>4(18%) 5(26%)</td>
<td>5(22%) 5(26%)</td>
</tr>
<tr>
<td>Newark</td>
<td>5(17%) 5(23%)</td>
<td>6(26%) 5(24%)</td>
</tr>
<tr>
<td>Saint Paul</td>
<td>5(18%) 6(25%)</td>
<td>7(23%) 6(25%)</td>
</tr>
<tr>
<td>Fort Wayne</td>
<td>4(18%) 4(25%)</td>
<td>6(24%) 5(27%)</td>
</tr>
<tr>
<td>Corpus Christi</td>
<td>4(21%) 5(21%)</td>
<td>5(24%) 5(27%)</td>
</tr>
<tr>
<td>Anaheim</td>
<td>3(21%) 4(34%)</td>
<td>4(24%) 4(34%)</td>
</tr>
<tr>
<td>Columbus</td>
<td>4(20%) 5(31%)</td>
<td>5(27%) 5(31%)</td>
</tr>
<tr>
<td>Norfolk</td>
<td>5(21%) 5(27%)</td>
<td>5(28%) 5(29%)</td>
</tr>
<tr>
<td>Topeka</td>
<td>5(17%) 5(23%)</td>
<td>6(26%) 5(24%)</td>
</tr>
<tr>
<td>Beaumont</td>
<td>5(20%) 6(26%)</td>
<td>6(26%) 6(27%)</td>
</tr>
<tr>
<td>Arlington</td>
<td>5(16%) 5(30%)</td>
<td>5(25%) 5(31%)</td>
</tr>
<tr>
<td>Laredo</td>
<td>6(21%) 7(26%)</td>
<td>7(26%) 7(26%)</td>
</tr>
<tr>
<td>Fullerton</td>
<td>4(16%) 4(24%)</td>
<td>4(24%) 4(26%)</td>
</tr>
<tr>
<td>Average</td>
<td>4(18%) 5(23%)</td>
<td>5(26%) 5(24%)</td>
</tr>
</tbody>
</table>

Partition: the percentage of streets that have value above the average level;
C and L: connectivity degree and length respectively;
L*: measuring unit is in degree;
**: results from previous studies

![Figure 5.1](image1)

![Figure 5.2](image2)

(1. Los Angles) (2. Phoenix) (3. Chicago)
Fig. 5.1 power law fit for SBF (threshold angle 45) for length

Fig. 5.2 Power law fit for SBF (threshold angle 45) for connectivity

5.1.3 Goodness of fit 1: P value test

The figures shown above are just like the “treasure maps”, tempting for the researchers to believe that they have found the power laws (the treasure). However, the visionary impressions can cause delusion, whereas, mathematics will not. The goodness of fit will act as the pushpins to locate the coordinates of
the “real treasure” by ruling out the unqualified distributions. In general, unless these power law-alike distributions pass the goodness of fit (P value passes the threshold value), otherwise they are merely some cute plots within the log-log scale chart.

In Table 5.2, listed are the parameters for the power law distributions estimation for the natural street networks generated from the 40 US urban street networks, which are in a descending order corresponding to the size of the cities. The exponent $\alpha$, $x_{\min}$ and the P value are ready for the first-round judgment. In particular, this table serves the analysis for SBF natural street network with a threshold angle of 45 degree. The results for other types of natural street networks will be shown in the appendix respectively.

**Table 5.2** The parameters: exponent (\(\alpha\)), $x_{\min}$, P value from the power law estimation for the SBF natural street network

<table>
<thead>
<tr>
<th>City</th>
<th>Length (continuous type)</th>
<th>Connectivity (discrete type)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha_i$</td>
<td>$x_{\min}$</td>
</tr>
<tr>
<td>Los Angeles</td>
<td>2.43</td>
<td>19.7</td>
</tr>
<tr>
<td>Phoenix</td>
<td>2.44</td>
<td>9.1</td>
</tr>
<tr>
<td>Chicago</td>
<td>2.53</td>
<td>7.4</td>
</tr>
<tr>
<td>Houston</td>
<td>2.60</td>
<td>10.1</td>
</tr>
<tr>
<td>Pasadena</td>
<td>2.36</td>
<td>3.1</td>
</tr>
<tr>
<td>San Diego</td>
<td>2.71</td>
<td>10.9</td>
</tr>
<tr>
<td>Hollywood</td>
<td>2.68</td>
<td>8.3</td>
</tr>
<tr>
<td>Dallas</td>
<td>2.59</td>
<td>17.5</td>
</tr>
<tr>
<td>Arlington100</td>
<td>2.57</td>
<td>6</td>
</tr>
<tr>
<td>Las Vegas</td>
<td>2.47</td>
<td>9.5</td>
</tr>
<tr>
<td>Sunnyvale</td>
<td>2.56</td>
<td>9.6</td>
</tr>
<tr>
<td>St.Petersburg</td>
<td>2.59</td>
<td>6.9</td>
</tr>
<tr>
<td>Detroit</td>
<td>2.69</td>
<td>17.7</td>
</tr>
<tr>
<td>San Antonio</td>
<td>2.44</td>
<td>9.7</td>
</tr>
<tr>
<td>New York</td>
<td>3.00</td>
<td>34.1</td>
</tr>
<tr>
<td>Birmingham</td>
<td>2.55</td>
<td>6.6</td>
</tr>
<tr>
<td>Tacoma</td>
<td>2.47</td>
<td>6.9</td>
</tr>
<tr>
<td>Sterling Heights</td>
<td>2.56</td>
<td>8.5</td>
</tr>
<tr>
<td>Moreno Valley</td>
<td>2.61</td>
<td>16.9</td>
</tr>
<tr>
<td>Louisville</td>
<td>2.60</td>
<td>8.7</td>
</tr>
<tr>
<td>Dayton</td>
<td>2.57</td>
<td>10.6</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>2.40</td>
<td>4.8</td>
</tr>
<tr>
<td>Greensboro</td>
<td>2.51</td>
<td>7.4</td>
</tr>
<tr>
<td>Gray</td>
<td>2.59</td>
<td>14</td>
</tr>
<tr>
<td>Little Rock</td>
<td>2.66</td>
<td>7.5</td>
</tr>
<tr>
<td>Spokane</td>
<td>2.35</td>
<td>5.7</td>
</tr>
<tr>
<td>Bakersfield</td>
<td>2.61</td>
<td>8.2</td>
</tr>
<tr>
<td>Newark</td>
<td>2.65</td>
<td>6</td>
</tr>
<tr>
<td>Saint Paul</td>
<td>2.41</td>
<td>10.1</td>
</tr>
<tr>
<td>Fort Wayne</td>
<td>2.55</td>
<td>8.6</td>
</tr>
<tr>
<td>Fremont</td>
<td>2.82</td>
<td>20.7</td>
</tr>
</tbody>
</table>
5.1.4 Goodness of fit 2: Alternative distributions competing test

The P value can be used to rule out the unqualified distributions by determining the significance of its value. In case that it is small, say, less than 0.1, it shall rule out the hypothesized distribution; otherwise keep it for further verifications. Such verification addresses the alternative distributions competing test to pick one candidate distribution being a more favored fit to the empirical dataset. Table 5.3 and Table 5.4 are showing the results from such competitions for length and connectivity respectively. Note that, the one in discrete case (connectivity) has Poisson distribution as a candidate value. In case that it is small, say, less than 0.1, it shall rule out the hypothesized distribution; otherwise keep it for further verifications. Such verification addresses the alternative distributions competing test to pick one candidate distribution being a more favored fit to the empirical dataset. Table 5.3 and Table 5.4 are showing the results from such competitions for length and connectivity respectively. Note that, the one in discrete case (connectivity) has Poisson distribution as a candidate one compared with the one in continuous case (length).

Table 5.3 Alternative distribution test against power law for the 40 natural street networks (length)

<table>
<thead>
<tr>
<th>City</th>
<th>Power law</th>
<th>Log-normal</th>
<th>Exponential</th>
<th>Stretched exp</th>
<th>Power law + cutoff</th>
<th>Support for power law</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>LR</td>
<td>P</td>
<td>LR</td>
<td>P</td>
<td>LR</td>
</tr>
<tr>
<td>Los Angeles</td>
<td>0.07</td>
<td>-1.67</td>
<td>0.23</td>
<td>747.96</td>
<td>0.00</td>
<td>-1.81</td>
</tr>
<tr>
<td>Phoenix</td>
<td>0.79</td>
<td>-0.62</td>
<td>0.44</td>
<td>215.83</td>
<td>0.00</td>
<td>-0.70</td>
</tr>
<tr>
<td>Chicago</td>
<td>0.06</td>
<td>-3.92</td>
<td>0.07</td>
<td>427.99</td>
<td>0.00</td>
<td>-4.31</td>
</tr>
<tr>
<td>Houston</td>
<td>0.00</td>
<td>0.06</td>
<td>0.39</td>
<td>1492.02</td>
<td>0.00</td>
<td>2.22</td>
</tr>
<tr>
<td>Pasadena</td>
<td>0.00</td>
<td>0.08</td>
<td>0.40</td>
<td>1657.52</td>
<td>0.00</td>
<td>2.58</td>
</tr>
<tr>
<td>San Diego</td>
<td>0.16</td>
<td>-2.61</td>
<td>0.17</td>
<td>359.68</td>
<td>0.00</td>
<td>-2.78</td>
</tr>
<tr>
<td>Hollywood</td>
<td>0.66</td>
<td>-0.54</td>
<td>0.45</td>
<td>94.17</td>
<td>0.00</td>
<td>-0.62</td>
</tr>
<tr>
<td>Dallas</td>
<td>0.22</td>
<td>-0.38</td>
<td>0.56</td>
<td>636.44</td>
<td>0.00</td>
<td>-0.28</td>
</tr>
<tr>
<td>Arlington 61</td>
<td>0.62</td>
<td>-1.08</td>
<td>0.33</td>
<td>428.29</td>
<td>0.00</td>
<td>-1.16</td>
</tr>
<tr>
<td>Las Vegas</td>
<td>0.00</td>
<td>-0.10</td>
<td>0.78</td>
<td>994.54</td>
<td>0.00</td>
<td>0.26</td>
</tr>
<tr>
<td>Sunnyvale</td>
<td>0.05</td>
<td>-0.25</td>
<td>0.65</td>
<td>507.21</td>
<td>0.00</td>
<td>-0.15</td>
</tr>
<tr>
<td>St.Petersburg</td>
<td>0.00</td>
<td>-0.02</td>
<td>0.90</td>
<td>567.21</td>
<td>0.00</td>
<td>0.33</td>
</tr>
<tr>
<td>Detroit</td>
<td>0.07</td>
<td>-0.99</td>
<td>0.42</td>
<td>267.57</td>
<td>0.00</td>
<td>-1.02</td>
</tr>
<tr>
<td>San Antonio</td>
<td>0.44</td>
<td>-0.02</td>
<td>0.89</td>
<td>247.79</td>
<td>0.00</td>
<td>0.11</td>
</tr>
<tr>
<td>New York</td>
<td>0.15</td>
<td>-16.57</td>
<td>0.00</td>
<td>191.93</td>
<td>0.00</td>
<td>-17.20</td>
</tr>
<tr>
<td>Birmingham</td>
<td>0.07</td>
<td>-0.22</td>
<td>0.67</td>
<td>353.10</td>
<td>0.00</td>
<td>-0.17</td>
</tr>
<tr>
<td>Tacoma</td>
<td>0.00</td>
<td>-15.24</td>
<td>0.00</td>
<td>539.42</td>
<td>0.00</td>
<td>-16.68</td>
</tr>
<tr>
<td>Sterling Heights</td>
<td>0.07</td>
<td>0.02</td>
<td>0.20</td>
<td>414.22</td>
<td>0.00</td>
<td>1.11</td>
</tr>
<tr>
<td>Moreno Valley</td>
<td>0.00</td>
<td>-7.62</td>
<td>0.01</td>
<td>321.20</td>
<td>0.00</td>
<td>-8.39</td>
</tr>
<tr>
<td>Louisville</td>
<td>0.26</td>
<td>-0.08</td>
<td>0.80</td>
<td>346.19</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Dayton</td>
<td>0.21</td>
<td>-0.10</td>
<td>0.76</td>
<td>328.92</td>
<td>0.00</td>
<td>-0.02</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>0.01</td>
<td>-5.26</td>
<td>0.04</td>
<td>14.48</td>
<td>0.14</td>
<td>-5.69</td>
</tr>
<tr>
<td>Greensboro</td>
<td>0.34</td>
<td>-1.64</td>
<td>0.23</td>
<td>138.23</td>
<td>0.00</td>
<td>-1.82</td>
</tr>
<tr>
<td>City</td>
<td>Power law</td>
<td>Poisson</td>
<td>Log-normal</td>
<td>Exponential</td>
<td>Stretched exp</td>
<td>Power law + cutoff</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------</td>
<td>---------</td>
<td>------------</td>
<td>-------------</td>
<td>---------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Gray</td>
<td>0.04</td>
<td>-1.01</td>
<td>0.37</td>
<td>293.68</td>
<td>0.00</td>
<td>-1.10</td>
</tr>
<tr>
<td>Little Rock</td>
<td>0.03</td>
<td>-4.71</td>
<td>0.05</td>
<td>129.67</td>
<td>0.00</td>
<td>-5.16</td>
</tr>
<tr>
<td>Spokane</td>
<td>0.12</td>
<td>-0.06</td>
<td>0.84</td>
<td>63.71</td>
<td>0.00</td>
<td>-0.05</td>
</tr>
<tr>
<td>Bakersfield</td>
<td>0.00</td>
<td>-0.40</td>
<td>0.57</td>
<td>698.53</td>
<td>0.00</td>
<td>-0.33</td>
</tr>
<tr>
<td>Newark</td>
<td>0.09</td>
<td>-3.54</td>
<td>0.09</td>
<td>104.46</td>
<td>0.00</td>
<td>-3.90</td>
</tr>
<tr>
<td>Saint Paul</td>
<td>0.00</td>
<td>-21.74</td>
<td>0.00</td>
<td>392.56</td>
<td>0.00</td>
<td>-23.56</td>
</tr>
<tr>
<td>Fort Wayne</td>
<td>0.08</td>
<td>-0.37</td>
<td>0.53</td>
<td>96.25</td>
<td>0.00</td>
<td>-0.43</td>
</tr>
<tr>
<td>Fremont</td>
<td>0.02</td>
<td>-2.51</td>
<td>0.18</td>
<td>210.56</td>
<td>0.00</td>
<td>-2.66</td>
</tr>
<tr>
<td>Corpus Christi</td>
<td>0.92</td>
<td>-0.45</td>
<td>0.51</td>
<td>64.03</td>
<td>0.00</td>
<td>-0.51</td>
</tr>
<tr>
<td>Anaheim</td>
<td>0.55</td>
<td>-0.44</td>
<td>0.51</td>
<td>31.53</td>
<td>0.00</td>
<td>-0.53</td>
</tr>
<tr>
<td>Columbus</td>
<td>0.31</td>
<td>-0.21</td>
<td>0.68</td>
<td>136.06</td>
<td>0.00</td>
<td>-0.22</td>
</tr>
<tr>
<td>Norfolk</td>
<td>0.17</td>
<td>-0.88</td>
<td>0.42</td>
<td>110.83</td>
<td>0.00</td>
<td>-0.93</td>
</tr>
<tr>
<td>Topeka</td>
<td>0.21</td>
<td>-0.59</td>
<td>0.50</td>
<td>123.53</td>
<td>0.00</td>
<td>-0.64</td>
</tr>
<tr>
<td>Beaumont</td>
<td>0.22</td>
<td>-0.63</td>
<td>0.42</td>
<td>26.99</td>
<td>0.00</td>
<td>-0.73</td>
</tr>
<tr>
<td>Arlington 100</td>
<td>0.06</td>
<td>-0.17</td>
<td>0.68</td>
<td>45.26</td>
<td>0.00</td>
<td>-0.20</td>
</tr>
<tr>
<td>Laredo</td>
<td>0.00</td>
<td>-5.75</td>
<td>0.03</td>
<td>2.19</td>
<td>0.75</td>
<td>-6.56</td>
</tr>
<tr>
<td>Fullerton</td>
<td>0.12</td>
<td>-0.48</td>
<td>0.51</td>
<td>29.71</td>
<td>0.00</td>
<td>-0.54</td>
</tr>
</tbody>
</table>

LR: logarithmic ratio of likelihood
1*: power law + cutoff
2*: moderate

Table 5.4 Alternative distribution test against power law for the 40 natural street networks (connectivity)
5.1.5 Overall comparisons about the exponent $\alpha$ from the case study

The overall comparisons between the previous studies and the revised work from this thesis are performed based on examining the exponent $\alpha$ for the connectivity degree. The results for all the natural street networks: EBF, SBF and SF with a threshold angle of 45 and 70 degree generated from this thesis work and the one from the previous study are listed in Table 5.5.

Table 5.5 Comparisons between the previous studies and the thesis work about the estimated exponent $\alpha$

<table>
<thead>
<tr>
<th>City</th>
<th>Previous study</th>
<th>SBF 70*</th>
<th>EBF 70*</th>
<th>SBF 45**</th>
<th>EBF 45**</th>
<th>SF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Los Angeles</td>
<td>2.0</td>
<td>3.16</td>
<td>3.16</td>
<td>3.21</td>
<td>3.21</td>
<td>3.14</td>
</tr>
<tr>
<td>Phoenix</td>
<td>2.1</td>
<td>3.01</td>
<td>2.99</td>
<td>3.01</td>
<td>3.01</td>
<td>3.08</td>
</tr>
<tr>
<td>Chicago</td>
<td>2.0</td>
<td>2.81</td>
<td>2.68</td>
<td>2.86</td>
<td>2.82</td>
<td>2.68</td>
</tr>
<tr>
<td>Houston</td>
<td>2.1</td>
<td>2.85</td>
<td>2.84</td>
<td>2.88</td>
<td>2.89</td>
<td>3.05</td>
</tr>
<tr>
<td>Pasadena</td>
<td>2.1</td>
<td>2.92</td>
<td>2.91</td>
<td>2.96</td>
<td>2.95</td>
<td>2.9</td>
</tr>
<tr>
<td>San Diego</td>
<td>2.3</td>
<td>3.1</td>
<td>3.05</td>
<td>2.57</td>
<td>2.56</td>
<td>2.51</td>
</tr>
<tr>
<td>Hollywood</td>
<td>2.1</td>
<td>2.94</td>
<td>2.98</td>
<td>3.1</td>
<td>3.06</td>
<td>2.98</td>
</tr>
<tr>
<td>Dallas</td>
<td>2.0</td>
<td>2.77</td>
<td>2.69</td>
<td>2.72</td>
<td>2.67</td>
<td>2.74</td>
</tr>
<tr>
<td>Arlington Taxes</td>
<td>2.0</td>
<td>3.02</td>
<td>2.88</td>
<td>3.15</td>
<td>3.11</td>
<td>3.01</td>
</tr>
<tr>
<td>Las Vegas</td>
<td>2.2</td>
<td>2.87</td>
<td>2.84</td>
<td>2.89</td>
<td>2.87</td>
<td>2.97</td>
</tr>
<tr>
<td>Sunnyvale</td>
<td>2.2</td>
<td>2.78</td>
<td>2.81</td>
<td>2.84</td>
<td>2.82</td>
<td>2.86</td>
</tr>
<tr>
<td>St. Petersburg</td>
<td>2.1</td>
<td>2.94</td>
<td>2.95</td>
<td>3</td>
<td>3.07</td>
<td>3.04</td>
</tr>
<tr>
<td>Detroit</td>
<td>1.9</td>
<td>2.69</td>
<td>2.53</td>
<td>2.58</td>
<td>2.57</td>
<td>2.73</td>
</tr>
<tr>
<td>San Antonio</td>
<td>2.0</td>
<td>2.86</td>
<td>2.83</td>
<td>2.9</td>
<td>2.9</td>
<td>2.87</td>
</tr>
<tr>
<td>New York</td>
<td>1.7</td>
<td>2.67</td>
<td>2.66</td>
<td>2.67</td>
<td>2.67</td>
<td>2.67</td>
</tr>
<tr>
<td>Birmingham</td>
<td>2.2</td>
<td>2.94</td>
<td>2.88</td>
<td>3.05</td>
<td>2.97</td>
<td>3.03</td>
</tr>
<tr>
<td>Tacoma</td>
<td>2.4</td>
<td>2.79</td>
<td>2.81</td>
<td>2.86</td>
<td>2.87</td>
<td>2.83</td>
</tr>
<tr>
<td>Sterling Heights</td>
<td>2.2</td>
<td>2.97</td>
<td>2.95</td>
<td>3.01</td>
<td>3</td>
<td>2.95</td>
</tr>
</tbody>
</table>

1* power law + cutoff
2* moderate
P: results from P value test
<table>
<thead>
<tr>
<th>City</th>
<th>2.0</th>
<th>2.1</th>
<th>2.2</th>
<th>2.3</th>
<th>2.4</th>
<th>2.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moreno Valley</td>
<td>2.2</td>
<td>3.02</td>
<td>3.01</td>
<td>3.09</td>
<td>3.1</td>
<td>3.01</td>
</tr>
<tr>
<td>Louisville</td>
<td>2.2</td>
<td>2.92</td>
<td>2.89</td>
<td>2.91</td>
<td>2.92</td>
<td>3.07</td>
</tr>
<tr>
<td>Dayton</td>
<td>2.1</td>
<td>2.22</td>
<td>2.22</td>
<td>2.22</td>
<td>2.23</td>
<td>2.22</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>1.8</td>
<td>2.86</td>
<td>2.89</td>
<td>2.91</td>
<td>3.16</td>
<td>2.93</td>
</tr>
<tr>
<td>Greensboro</td>
<td>2.2</td>
<td>2.77</td>
<td>2.71</td>
<td>2.82</td>
<td>2.81</td>
<td>2.75</td>
</tr>
<tr>
<td>Gray</td>
<td>2.0</td>
<td>2.68</td>
<td>2.66</td>
<td>2.73</td>
<td>2.72</td>
<td>2.71</td>
</tr>
<tr>
<td>Little Rock</td>
<td>2.1</td>
<td>2.93</td>
<td>2.85</td>
<td>2.93</td>
<td>2.93</td>
<td>2.97</td>
</tr>
<tr>
<td>Spokane</td>
<td>2.0</td>
<td>2.99</td>
<td>2.99</td>
<td>3.01</td>
<td>3.02</td>
<td>2.99</td>
</tr>
<tr>
<td>Bakersfield</td>
<td>2.1</td>
<td>2.7</td>
<td>2.67</td>
<td>2.74</td>
<td>2.73</td>
<td>2.7</td>
</tr>
<tr>
<td>Newark</td>
<td>2.0</td>
<td>3.01</td>
<td>3.09</td>
<td>3.13</td>
<td>3.12</td>
<td>3.09</td>
</tr>
<tr>
<td>Saint Paul</td>
<td>1.9</td>
<td>2.41</td>
<td>2.49</td>
<td>2.47</td>
<td>2.52</td>
<td>2.41</td>
</tr>
<tr>
<td>Fort Wayne</td>
<td>2.2</td>
<td>2.87</td>
<td>2.88</td>
<td>2.79</td>
<td>2.85</td>
<td>2.9</td>
</tr>
<tr>
<td>Fremont</td>
<td>2.2</td>
<td>2.42</td>
<td>2.35</td>
<td>2.42</td>
<td>2.41</td>
<td>2.34</td>
</tr>
<tr>
<td>Corpus Christi</td>
<td>2.0</td>
<td>2.89</td>
<td>2.84</td>
<td>2.88</td>
<td>2.89</td>
<td>2.97</td>
</tr>
<tr>
<td>Anaheim</td>
<td>2.3</td>
<td>2.82</td>
<td>2.8</td>
<td>2.94</td>
<td>2.92</td>
<td>2.8</td>
</tr>
<tr>
<td>Columbus</td>
<td>2.0</td>
<td>2.59</td>
<td>2.57</td>
<td>2.64</td>
<td>2.63</td>
<td>2.59</td>
</tr>
<tr>
<td>Norfolk</td>
<td>1.9</td>
<td>2.52</td>
<td>2.5</td>
<td>2.57</td>
<td>2.56</td>
<td>2.89</td>
</tr>
<tr>
<td>Topeka</td>
<td>2.0</td>
<td>3.11</td>
<td>3.1</td>
<td>3.17</td>
<td>3.15</td>
<td>3.04</td>
</tr>
<tr>
<td>Beaumont</td>
<td>1.9</td>
<td>2.87</td>
<td>2.87</td>
<td>2.94</td>
<td>2.94</td>
<td>2.83</td>
</tr>
<tr>
<td>Arlington</td>
<td>2.0</td>
<td>2.94</td>
<td>2.94</td>
<td>2.97</td>
<td>2.97</td>
<td>2.96</td>
</tr>
<tr>
<td>Laredo</td>
<td>1.7</td>
<td>2.73</td>
<td>2.71</td>
<td>2.74</td>
<td>2.74</td>
<td>2.69</td>
</tr>
<tr>
<td>Fullerton</td>
<td>2.3</td>
<td>2.81</td>
<td>2.85</td>
<td>2.87</td>
<td>2.85</td>
<td>2.8</td>
</tr>
</tbody>
</table>

*: the chosen threshold angle is 70 degree  
**: the chosen threshold angle is 45 degree

5.2 Visual perspective

5.2.1 k-core decompositions from LaNet-vi

What can we expect from the visual perspective for the study of the 40 US urban street networks? As shown in Figure 5.3, 9 samples out of total 40 images illustrate the k-core decompositions for the 40 SBF natural street networks with a chosen threshold angle of 45 degree. For the sake of simplicity for visualization, the links are shrunk as slim as it could be in each plot. Meanwhile, there are two items provided: the color legend, which starts from the lowest value (violet) to the highest (red) corresponding to the coreness value; and the white dot legend indicating the size of the nodes that are logarithmic proportional to their original degree. The results for the rest of the cities will be shown in the appendix.
5.2.2 Retrieving model of k-core decomposition in GIS

One major purpose for extracting the topological information of these 40 complex networks is to understand how the urban street networks structured and self-organized. Contrariwise, spatially view such properties by identifying the exact location for each natural street in GIS will full feature such concern. Retrieving the decomposition information back to GIS will realize the goal. As is shown in Figure 5.4, the retrieving models displaying as conventional maps are produced in ArcGIS desktop based on the annotations of the k-core decomposition values. For the ease of illustration, some typical natural street networks are selected.
Fig. 5.4 The k-core decompositions with retrieved spatial information in GIS
6. Conclusions and future work

In this chapter, the conclusions for the case study are brought forth together with the proposed future work. Before drawing the conclusions, the analysis and discussions about the results from the case study is performed ahead from analytical perspective and visual perspective respectively. Meanwhile, the possibility of combing these two perspectives for the studies of complex systems is discussed. Based on the analysis and discussions, the conclusions are made. The future work is hence proposed to improve the strategies used in this study and extend the study for further applications.

6.1 Analysis and discussions

6.1.1 Qualitative overview for the structure of the natural street networks

Over the entire 40 natural street networks generated from the US urban street networks, about 20% natural streets have a connectivity degree or length value greater than the average. Actually, such percentage varies from case to case around 20%, which is shown in the results part. Note that, in the previous study, such percentage for length was not estimated. Table 6.1 shows the statistics, which summarized the averages of the percentages over the 40 natural street networks. As mentioned above, the previous study also adopted the SBF natural street model with a threshold angle of 70 degree to generate the street networks but differed in the defining the deflection angle. More specifically, the natural street networks from the SBF natural street model with a deflection angle of 70 degree in this thesis work followed the identical process as the one in the previous study. However, the average connectivity degree value is 1 more than the one from previous studies. Apparently, the definition for the deflection angle in this thesis makes more sense, which can also refer to natural street models for the Swedish highway studies (Jiang et al., 2008). Nevertheless, such percentages are quite consistent with an approximation to the 80/20 principle, which shows around 80% streets have a connectivity degree or length value below the average, while the rest 20% have the value above the average. Practically, the distribution of such dataset will appear as a long-tail in the normal scale plot, which is a heavy-tailed alike distribution and hence the power law estimation shall be scheduled.

Table 6.1 Summary of the percentages that are above the average for the 40 natural street networks

<table>
<thead>
<tr>
<th></th>
<th>SBF 70</th>
<th>EBF 70</th>
<th>SBF 45</th>
<th>EBF 45</th>
<th>SF</th>
</tr>
</thead>
<tbody>
<tr>
<td>**</td>
<td>C</td>
<td>C</td>
<td>L</td>
<td>C</td>
<td>L</td>
</tr>
<tr>
<td>Average</td>
<td>4(18%)</td>
<td>5(23%)</td>
<td>5(26%)</td>
<td>5(24%)</td>
<td>5(26%)</td>
</tr>
</tbody>
</table>

*: C for connectivity and the L for length respectively,
**: Results from previous study

6.1.2 Quantitative analysis: Power law estimations and alternative distribution tests

Regarding to the log-log plots drawn with the CDF of each natural street network and the power law fit (a straight line), visionary judgment will cause delusions. Because almost every plot shows a pretty straight slope and most part of it fits the power law distribution well. This is what graphic methods do, the linear regression methods in particular (Goldstein et al., 2004). Indeed, it is a method based on the assumption that the data came from a certain distribution and hence it is not suitable for the estimations for distributions (Clauset et al., 2007). It is also the exact reason this thesis adopts MLE and goodness of fit to estimate the power law distributions. Here, we intend to make a tough criterion by setting the significance of P value to 0.1. From this basis, we count the number of the power law estimations that have a P value less than 0.1, which is 11/40 for the estimations of length and 14/40 of the ones for connectivity. If we confirm the premise of using 0.1 as the criteria, technically, those
distributions shall be ruled out to be fitted by power law distributions. In other words, if the power law-alike distributions cannot pass the criteria, they are merely some cute plots within log-log scale. Note that, it is the summary for SBF natural street networks with a threshold angle of 45 degree.

So far, the distributions of data from the rest cases, which passed the P value tests, are now the candidates for being fitted by the power law distributions. Yet, the other heavy-tailed distributions may have chance to be a favored fit against power law distribution. Such competition can be judged from the alternative distribution test. As is shown in the results, 31/40 cases show that power law distribution with an exponential cut-off is a favored fit for the case of length. The rest 9 cases are tagged with “moderate”, which means power law distribution is still a favored fit in the competitions. In this case, we have to cross check the P value to make a final decision. Actually, only 2/9, St. Petersburg and Pasadena in particular, are with a P value less than 0.1, which means, none of these mentioned distributions is a good fit to them. For the rest of the 7 cases, so far, we can make a decision that they follow a power law distribution. After applying the same manner to the case of connectivity degree, 20/40 cases chose power law distribution with an exponential cut-off as a favored fit. For the rest 20 cases, 9/20 of them cannot select a good fit, in particular, listed as: Las Vegas, Sunny Vale, Sterling Heights, Birmingham, Dayton, Fort Wayne, Columbus, Corpus Christi and Fullerton namely. And the last 11 cases are believed to follow a power law distribution.

6.1.3 Comprehensive comparisons with previous study

The previous study examined the topological pattern of the 40 US urban street networks based on the natural street networks with a threshold angle of 70, although the definition of deflection angle in that study is considered improper. However, due to the fact that the US urban street networks are more “grid-like” structure, natural street model may still work. Nevertheless, the author also claimed the universality and peculiarity about the power laws they found. In particular, they believed that “the universal power law is a patent signature of self-organizing cities” and some of the networks have two power law exponents (Jiang, 2007). However, despite of affection from the deflection angle, most of the exponent values extracted from the previous study using linear regression method are smaller than the ones from the MLE method. Although, for the discrete type power law estimation, MLE will produce slightly bigger exponent (Clauset et al, 2007), the exponent values are quite consistent no matter what natural street model used. Moreover, the dual power law exponents did not exist in the thesis study but follow a form of power law distribution with an exponential cut-off, Philadelphia and Laredo in particular. Regarding to the mechanism of MLE, \(x_{\text{min}}\) is pre-calculated and it will separate the distribution into two sections. From the overall estimation procedure, MLE will only consider the upper part (the tail), while the lower part will be left. Therefore, there may be dual-exponents for power law estimation. The example can refer the study of human mobility (Jiang et al., 2008).

6.1.4 Network visualizations based on k-core decompositions

The network visualizations were realized based on the interrelationships of the natural street networks, and were implemented in the LaNet-vi based on the k-core decomposition algorithm. It differs with the applications that only consider the connectivity of the network. As we can see from the results, those natural streets with big connectivity value (the nodes with big size) are not necessarily in the inner-most core. Another impression we can get form the images is that these natural street networks are not highly clustered, but diversely distributed with significant decrease in vertex number, which may indicates the hidden power laws. The last but not the least, some networks are fractured into multi-component structured decompositions. It implies the fact that not all the cities are booming from the center and spreading to all directions like the ER model. Instead, there are multiple sites (components in the figures) that are equally clustered as the center of the city being “self-organized”.

6.1.5 Visualizations for k-core decompositions in GIS

It is not easy to perceive the “invisible hands” which created the urban street networks (Jiang, 2007), and it is hard to capture or predict the evolution of the cities because of the dynamics (Portugali, 2000), but we can at least “see” how the streets distribute spatially in k-core decomposition forms. With the help of the GIS, the k-core decompositions can retrieve spatial information from the features. We can identify which parts are the center cores of the city and why these natural streets with big connectivity degree do not exist in the inner-most cores. For instance, the single component structured street networks developed from the inner most core and spread to all directions, while the multi-component structured networks have two or more than two centers, at which the streets clustered. When multiple components exist, we can identify where these cores belong to. In GIS, the information for each natural street about connectivity degree, length and k-core value are stored in the same platform. It is able to examine whether there is strong correlation between these attribute, which may be implemented in future work.

6.1.6 Combination of analytical and visual perspectives

The analytical perspective and visual perspective can actually be combined together to make one as a reflection of the other. This possibility is illustrated from following basis:

1. From the visualization of k-core decompositions, we can actually see, these natural street networks are not evenly distributed, but diversely spread through the entire networks, which show the existences of power laws.

2. With respect to power law distribution with an exponential cut-off, the “cut-off” happens when big values in the dataset are extremely rare compared with situations under power law distributions. From the decomposition perspective, it is more likely to be happening in the networks, which are single component structured and pretty tightly clustered. And for the multi-component structured networks, it still inclines to have a power law distribution with cut-off, but when the gap between small and big value are not that big, take Moreno Valley for instance, the power law will have bigger possibility fitting the data.

3. The decompositions of the ones, whose distributions do not follow a power law or a power law with a cut-off, are mostly sparsely distributed in the layouts.

However, there are always exceptions. A lot of work should be implemented furthermore. Long words short, it is already quite enthusiastic to visually see how the complex networks decomposed and know what kind hidden order are “controlling” their forms.

6.2 Conclusions

This thesis explored the topological patterns of the 40 US urban street networks from analytical and visual perspectives. Three natural street models were implemented to produce the natural street networks from the 40 US urban street networks. From the study of this thesis, we found the 80/20 principle exists for the overall structures of the street networks, which shows around 80% streets have a connectivity degree or length above the average, while around 20% streets have such attributes below the average. Furthermore, the distributions of the street networks were studied for power law estimations by adopting MLE and goodness of fit test. To eliminate the risk that the other heavy tailed distributions may be a better fit, the alternative distribution tests were implemented as well. From this basis, we conclude that most of distributions of the street connectivity degree and length of the street networks follow a power law distribution with an exponential cut-off; some of the rest have a power law distribution; while only some extreme cases still did not pick any distribution to fit. As we can learn from the mathematical form of power law distribution with an exponential cut-off, it is an extreme case of power law distribution; therefore, the general power law is still a “universal” signature of the urban street networks. From the visual perspective, by studying the decompositions of the street network, we performed the network visualizations for the street networks, which are based on k-core decomposition algorithm. It enables us to see the topological patterns of the urban street networks.
Moreover, by retrieving the spatial information from the street networks from GIS, it is possible to locate the decompositions spatially, which means we can see how the street networks decomposed and structured in the geographic maps. Last but not the least, it is possible to combine these two perspectives together for better study, each perspective can serve as a reflection for the other one. With proper strategy, the study of complex systems in GIS using from these two perspectives will provide us more surprising results.

6.3 Future work

6.3.1 Problems to solve

Even the study of this thesis has already been performed, there still remain some problems. For examples:

Just because SBF natural street model performed the best in Jiang et al. (2008) for predicting the traffic flow that does not necessarily mean it is the best model for generating the natural streets. In fact, the process of SBF randomly selecting a start segment and this randomness is not controllable, thus it will produce massive sets of natural streets. Apparently, the related research is anticipated to dig into the real mechanism of urban street networks and thus establish a more reasonable natural street model.

From the analytical perspective, the power law distribution with an exponential cut-off selected from the alternative distribution test for a favored fit does not guarantee it to be a good fit to the datasets. It can only be confirmed as the best fit among the distributions we introduced. As we did not perform the P value test to the power law distribution with an exponential cut-off, there might be chance to rule out the unqualified distributions just like the P value test for power law distribution. Meanwhile, even the power law distribution with an exponential cut-off was confirmed; the corresponding parameters were not extracted. Therefore, the P value test for power law distribution with an exponential cut-off and goodness of fit test should be implemented to make a full study of the distributions from complex systems. Indeed, such work should be applied to any heavy-tailed distribution mentioned above.

6.3.2 Future development

Despite the problems mentioned above, here we list some proposed work for future development:

As the entire study was based on the 40 US street networks, we are constrained within the particular architecture of the street networks which is “grid-like” (Jiang, 2007) and seems more or less under urban planning with human interference. Since we concern the self-organized characteristic of the complex systems, the future work ought to take some other street networks for consideration, for examples, the historical street networks. Hence there might be some other interesting discovers.

The study of this thesis can actually be applied for the concern of urban design. For instance, as suggested by Jiang (2008), the 80/20 principle can advise the urban planners to emphasis the 20% streets that are important to handle the traffic flow. This is also one direction to link the study of street networks to practical projects.

The visual effects from the k-core decomposition are impressive. It tells the structures of the complex systems and co-operate with GIS to spatially represent these information. Meanwhile, it supports the analytical studies. However, the level of interactive communication between these two perspectives is still pretty low, and regard to the current visual analytics applications, real-time, fast responding and high level interactive communication capability are really the trend. Therefore, to develop a better combine research strategy with analytical and visual perspectives to study the topological patterns of urban street networks are the courses the future work should head for.
REFERENCES


Hill, L. (2006), GIS support in DRL. Social Science Network & Data Service, 9, issue 1


Surowiecki J. (2004). The wisdom of crowds: Why the many are smarter than the few. (London: ABACUS)


Vuong, Q. H. (1989). Econometrica, 57, 307


Appendix

Here in the appendix are the supplementary figures and tables for this thesis. Figure 1 is the shapefiles for the 40 US urban street networks studied in this thesis. Figure 2 is the power law estimations for the case of length of the natural street networks with a threshold angle of 45, which is shown as the CDF plots in log-log scale with a fit line of power law distribution (the black line). Figure 3 corresponds to the Figure 2 but for the connectivity degree of the street networks. Following the Table 1, it consists of the exponent ($\alpha$), $x_{\text{min}}$ and $P$ value from the power law estimations with MLE method, and it is for the cases of Every-best-fit (EBF) natural street networks with a threshold angle of 45 and Self-fit (SF) natural street networks. The Figure 4 in the end is the k-core decomposition layouts from LaNet-vi for the case of natural street networks with a threshold angle of 45.
Fig. 1 Shapefiles for the 40 US urban street networks

(Source: http://www.census.gov/geo/www/tiger/ & www.hig.se/~bjg/PhyADData/data_Publication)
Pr(X ≥ x)
Fig. 2 Power law fit for Self-best-fit (threshold angle 45) for with CDFs of length in log-log plot


(31. Fremont)  (32. Corpus Christi)  (33. Anaheim)

(34. Columbus)  (35. Norfolk)  (36. Topeka)


(40. Fullerton)

Fig. 3 Power law fit for Self-best-fit (threshold angle 45) for with CDFs of connectivity degree in log-log plot

Table 1 Power law estimations for exponent ($\alpha$), $x_{\text{min}}$ and P value for EBF and SF natural street networks

<table>
<thead>
<tr>
<th></th>
<th>EBF 45 C</th>
<th>EBF 45 L</th>
<th>SF C</th>
<th>SF L</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>$x_{\text{min}}$</td>
<td>P</td>
<td>$\alpha$</td>
<td>$x_{\text{min}}$</td>
</tr>
<tr>
<td>2.72</td>
<td>4</td>
<td>0.29</td>
<td>2.59</td>
<td>13.9</td>
</tr>
<tr>
<td>3.21</td>
<td>4</td>
<td>0.00</td>
<td>2.43</td>
<td>17.9</td>
</tr>
<tr>
<td>2.67</td>
<td>5</td>
<td>0.00</td>
<td>2.60</td>
<td>17.5</td>
</tr>
<tr>
<td>2.92</td>
<td>9</td>
<td>0.37</td>
<td>2.58</td>
<td>12.1</td>
</tr>
<tr>
<td>2.85</td>
<td>7</td>
<td>0.16</td>
<td>2.53</td>
<td>8.6</td>
</tr>
<tr>
<td>2.92</td>
<td>5</td>
<td>0.02</td>
<td>2.62</td>
<td>12.3</td>
</tr>
<tr>
<td>3.10</td>
<td>4</td>
<td>0.03</td>
<td>2.61</td>
<td>16</td>
</tr>
<tr>
<td>2.82</td>
<td>5</td>
<td>0.00</td>
<td>2.57</td>
<td>10.4</td>
</tr>
<tr>
<td>2.82</td>
<td>6</td>
<td>0.84</td>
<td>2.46</td>
<td>4.9</td>
</tr>
<tr>
<td>3.16</td>
<td>4</td>
<td>0.00</td>
<td>2.39</td>
<td>4.8</td>
</tr>
<tr>
<td>2.97</td>
<td>4</td>
<td>0.00</td>
<td>2.61</td>
<td>7.5</td>
</tr>
</tbody>
</table>
2.94  4  0.37  2.65  6  0.00  2.83  4  0.91  2.68  10  0.02
2.57  5  0.01  2.90  35.8  0.41  2.73  9  0.00  3.15  35.5  0.02
2.89  7  0.81  2.69  12.4  0.14  2.97  8  0.26  2.71  11.8  0.02
2.52  8  0.03  2.49  10.6  0.00  2.41  6  0.13  3.79  32.2  0.11
3.06  6  0.85  2.68  9  0.15  2.98  5  0.00  2.62  7.9  0.00
2.74  3  0.08  2.39  6.3  0.00  2.69  3  0.01  2.44  7  0.00
3.15  4  0.00  2.55  8.2  0.12  3.04  4  0.01  2.60  12.8  0.08
3.02  4  0.11  2.27  3.2  0.00  2.99  4  0.24  2.22  3  0.00
2.93  5  0.27  2.66  7.5  0.42  2.97  6  0.00  2.56  7.1  0.01
3.07  7  0.09  2.58  7.8  0.57  3.04  6  0.28  2.54  6.7  0.00
2.41  3  0.00  2.24  5.8  0.00  2.34  3  0.01  2.32  6.2  0.00
3.00  4  0.16  2.55  7.8  0.36  2.95  4  0.11  2.39  5.1  0.00
2.87  5  0.44  2.47  6.9  0.15  2.83  5  0.19  2.46  6.6  0.06
2.81  6  0.25  2.57  8.2  0.04  2.75  5  0.00  2.77  12.1  0.00
2.56  4  0.00  2.96  19.4  0.14  2.89  10  0.17  2.92  19.2  0.02
2.95  4  0.02  2.35  3.1  0.00  2.90  4  0.09  2.37  5.4  0.01
2.73  4  0.16  2.59  7.8  0.02  2.70  4  0.00  2.68  9.4  0.00
2.63  4  0.04  2.20  4.3  0.00  2.59  4  0.00  2.17  4.4  0.00
2.90  4  0.27  2.51  6.1  0.00  2.87  4  0.14  2.49  6.7  0.07
2.82  5  0.03  2.63  7.9  0.00  2.72  4  0.00  2.63  8.3  0.00
2.87  4  0.85  2.49  9.7  0.54  2.97  7  0.60  2.50  5.7  0.00
3.01  3  0.05  2.41  5  0.11  3.08  4  0.16  2.45  3.4  0.06
3.11  5  0.02  2.51  4.9  0.06  3.01  5  0.04  2.47  5.5  0.01
3.12  7  0.76  2.62  5.5  0.04  3.09  7  0.09  2.57  5.7  0.00
2.85  5  0.53  2.52  6.7  0.39  2.80  5  0.53  2.44  6.2  0.00
2.56  4  0.00  2.77  11.3  0.49  2.51  4  0.00  2.82  12.2  0.43
2.89  6  0.71  2.54  6.7  0.00  3.05  8  0.18  2.65  9.9  0.02
2.23  3  0.00  2.15  4.4  0.00  2.22  3  0.01  3.12  18.2  0.08
2.97  4  0.23  2.53  8.5  0.14  3.03  5  0.37  2.57  6.5  0.16

45: Threshold angel for natural street model
C: Connectivity degree
L: Length
(10. Las Vegas)
(11. Sunnyvale)
(12. St.Petersburg)
(13. Detroit)
(14. San Antonio)
(15. New York)
Fig. 4 K-core decompositions in 2 dimensional for SBF natural street networks with a threshold angle of 45