SMALL-WORLD CHARACTERISTICS IN GEOGRAPHIC, EPIDEMIC, AND VIRTUAL SPACES: A COMPARATIVE STUDY

A Dissertation

by

ZENGWANG XU

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2007

Major Subject: Geography

SMALL-WORLD CHARACTERISTICS IN GEOGRAPHIC, EPIDEMIC, AND VIRTUAL SPACES: A COMPARATIVE STUDY

A Dissertation

by

ZENGWANG XU

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:	
Chair of Committee,	Daniel Z. Sui
Committee Members,	Andrew G. Klein
	Anthony M. Filippi
	Chanam Lee
Head of Department,	Douglas J. Sherman

May 2007

Major Subject: Geography

ABSTRACT

Small-World Characteristics in Geographic, Epidemic, and Virtual Spaces:

A Comparative Study. (May 2007)

Zengwang Xu, B.E., Southwestern JiaoTong University;

M.S., Nanjing University

Chair of Advisory Committee: Dr. Daniel Z. Sui

This dissertation focuses on a comparative study of small-world characteristics in geographical, epidemic, and virtual spaces. Small-world network is the major component of the "new science of networks" that emerged recently in research related to complex networks. It has shown a great potential to model the complex networks encountered in geographical studies. This dissertation, in an attempt to understand the emergence of small-world phenomenon in spatial networks, has investigated the small-world properties in aforementioned three spaces.

Specifically, this dissertation has studied roadway transportation networks at national, metropolitan, and intra-city scales via network autocorrelation methods to investigate the distance effect on the emergence of small-world properties. This dissertation also investigated the effect of small-world network properties on the epidemic diffusion and different control strategies through agent-based simulation on social networks. The AS-Level Internet in the contiguous U.S. has been studied in its relation between local and global connections, and its correspondence with small-world characteristics.

Through theoretical simulations and empirical studies on spatial networks, this dissertation has contributed to network science with a new method – network autocorrelation, and better understanding from the perspective of the relation between local and global connections and the distance effect in networks. A small-world phenomenon results from the interplay between the dynamics occurring on networks and the structure of networks; when the influencing distance of the dynamics reaches to the threshold of the network, the network will logically emerge as a small-world network. With the aid of numerical simulation a small-world network has a large number of local connections and a small number of global links. It is also found that the epidemics will take shorter time period to reach largest size on a small-world network and only particular control strategy, such as targeted control strategy, will be effective on small-world networks.

This dissertation bridges the gap between new science of networks and the network study in geography. It potentially contributes to GIScience with new modeling strategy for representing, analyzing, and modeling complexity in hazards prevention, landscape ecology, and sustainability science from a network-centric perspective.

DEDICATION

To my mother

ACKNOWLEDGEMENTS

I am deeply grateful to Dr. Daniel Z. Sui for his vision, advice, practical guidance and patience in helping me accomplish this dissertation research. Dr.Sui's advice on research methodology will be a life-long treasure for my career development. I am also indebted to all my advisory committee members, Dr. Klein, Dr. Filippi and Dr. Lee. Without their comments and encouragements, this dissertation could not have been completed.

I also want to thank our department head, Dr. Sherman, for his advice, generous funding support, career development opportunities, and jokes. The geography department has been rapidly evolving the last five years, but the Aggieland spirit and culture maintain. Surrounded by this culture, my graduate study at Texas A&M has benefited from many faculty members and fellow graduate students. Their encouragement and friendship are greatly appreciated.

I would also like to take this opportunity to thank Dr.Wendy Patzewitsch and Jose Gavinha for their help and friendship. My special thanks also go to Kirk Stueve, Lei Wang, and all other friends.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

FIGURE Page

LIST OF FIGURES (Continued)

FIGURE Page

LIST OF FIGURES (Continued)

FIGURE Page

LIST OF TABLES

TABLE Page

CHAPTER I

INTRODUCTION

1.1 Background

Networks are the underlying structures for many complex systems in nature and society. A network structure is considered to be a set of vertices (or nodes) with connections among them. Many models of interaction or communication can be formulated in terms of networks, in which vertices represent individuals (such as cities connected by transportation, computers connected by wires, web pages linked by hyperlinks, and people linked by social relations) and edges (or connections) represent contacts between individuals (such as hyperlinks or social contacts) (Lloyd and May 2001).

Geographers and social scientists have studied networks extensively for several decades (Haggett and Chorley 1969). The focus of those studies has been on the internal structure and topological pattern of various networks. The structures of social networks refer to the relationships among the interacting social entities, regularities in those patterns, and the implications of these relationships (Wasserman and Fause 1994). Network analysis in geographical studies emphasizes the structure of networks as well,

---

This dissertation follows the style and format of the *Annals of the Association of American Geographers*.

including spatial structure and its dynamics (Haggett and Chorley 1969). Previous studies on networks were based primarily on a clear layout of network connection topology. Although the growth of networks can be as simple as adding and/or adjusting nodes and links, many networks become more and more complex as they evolve. For many complex networks such as the Internet, power grids, etc., it is no longer possible to know how many vertices they contain and how they are connected to each other. The traditional analytical methods based on the clear layout of network topology are no longer effective for understanding the increasingly complex networks (Newman 2003).

Recent years have witnessed a paradigm shift of network studies, from a focus on the internal structure to the collective dynamics (Watts and Strogatz 1998) or statistical mechanics of networks (Albert and Barabasi 2002). The new paradigm attempts to discover the universal laws or regularities in the structure and evolution of complex networks and the underlying mechanisms governing those laws. During the past several years, this shift of focus has stimulated the emergence of the so-called "new science of networks", which resulted in the explosive growth of both theoretical and empirical studies on networks and networked systems (Barabasi 2002; Dorogovstev and Mendez 2003; Dorogovtsev and Mendes 2002; Strogatz 2001; Watts 1999; Watts 2003; Watts 2004). Studies on small-world networks have been one of the highlights and major contributions to the new science of networks.

The small-world network has been discovered as a ubiquitous feature of many real-world networks, and is considered to be a potential model for many real-world complex systems (Amaral et al. 2000; Hayes 2000; Watts 2003). A network with small-world characteristics implies that there is a very short separation between any two vertices in the network no matter how complex it is.

This shift of focus could potentially shed new light on the study of networks from geographic perspectives, as networked systems also abound in geographical studies. Typical examples include transportation networks, social contact networks, and networks in Cyberspace. Although abundant studies on the characteristics of individual networks have been reported in the literature (Albert and Barabasi 2002; Dorogovstsev and Mendes 2002; Dorogovstsev and Mendes 2003; Newman 2003), few studies exist on the methodologies and manifestations of small-world characteristics on spatial networks. The goal of this dissertation is to fill such a void in the literature by reviewing recent advances in network sciences and incorporating them into a comparative study of networks in three spaces, i.e., geographic, social, and virtual.

1.2 Objectives

The goal of this dissertation is to contribute to the literature by discussing new methods and clearer understanding of small-world characteristics in spatial networks through conducting a comparative study on the small-world characteristics among geographical, epidemic, and virtual spaces. Specifically, this research aims to achieve the following three objectives:

1. Explore the small-world characteristics in geographical space through studying the roadway transportation networks at intra-city, metropolitan, and national scales. Network autocorrelation is used as a new approach to investigate the different manifestations of the small-world network on the transportation networks at different scales. The distance effects in spatial networks and their implication on the small-world phenomenon is investigated.

2. Explore the small-world characteristics in epidemic space through studying the epidemic processes taking place in social contact networks, where vertices represent individuals and edges represent possible contacts between individuals. The social contact networks of different structural properties are simulated by a series of networks that are "rewired" from a completely regular network to a completely random network. The same epidemic spreading on different networks will be simulated. Different control strategies will also be simulated. The effects of the small-world networks on the epidemic diffusion and control strategies are investigated.

3. Explore the small-world characteristics in virtual space through studying the Internet at the Autonomous Systems (AS) level in which vertices are ASes and edges are links between ASes. The relation between global and local in the AS-level Internet and its relationship to the emergence of the small-world network is investigated.

1.3 Significance

The focus in this dissertation - the small-world characteristics of networks - is widely regarded as one of the major advancements in the new science of networks, and it recently emerged as a synthesis of interdisciplinary studies on complex networks. Inspired by this emerging field, this dissertation contributes to network science with studies of small-world characteristics in spatial networks.

Most networks previously studied in the literature are aspatial and topology-dominated, such as scientific publication citation networks (Newman 2001), collaboration networks of movie actors, metabolic networks, etc. In addition, some networks with obvious geographic reference were only studied from a topological perspective, such as transportation networks (Jiang 2004) and the Internet (Calvert, Doar, and Zegura 1997). The networks to be studied in this dissertation are geographical networks or spatial networks, in which vertices and/or edges have well-defined locations, and the connection topology alone cannot characterize their primary property. For example, in transportation networks represented by graphs where vertices are intersections and edges are road segments, not only topology but also the spatial locations of intersections and the types of road segments are influential properties of the network. Many geographical concepts, such as distances and locations, matter greatly on networks with spatial dimensions, and to ignore it is to miss the most interesting feature of these networks

(Gastner and Newman 2006b, 2006a) . Many real-world networks are spatial networks, and the networks with spatial embeddings are more complex than connection topology alone can characterize. The study on manifestations and verification methods of the small-world in spatial networks is a necessary step towards the application of the small-world network in a real-world scenario.

In addition, this dissertation contributes to network science a new methodology for studying the statistical mechanics of complex networks. Existing methodologies based on either the structural properties (average path length and clustering coefficient) or the efficiency of networks have not proven effective on spatial networks. This dissertation provides a new methodology using network autocorrelation to study the small-world characteristics of networks.

The new network science could potentially shed light on the study of networks from a geographic perspective as well. Network studies in geography date back to the 1960s (Haggett and Chorley 1969; Kansky 1963), and most, if not all, of the studies are based on network topology. With the increasing complexity of networks, it is impossible for many networks to have a clear layout of their topology. The study of the collective dynamics or statistical mechanics of networks in network science has created new concepts for network study in geography. In addition, the goal of this study is also consistent with Network Geography (Batty 2003), which is a new branch of geography that aims to incorporate new developments in network science into the network study in geography, and focus not only the representation of static structures but also on the dynamics occurring on and within such structures.

The small-world network model, combining the properties of deterministic network models and stochastic network models, is considered the potential model for many real-world complex systems (Amaral et al. 2000). With a better understanding of the characteristics of the small-world concept on spatial networks, this study goes a step further by modeling the complex network in a geographical context. While the complex network method has been adopted as a research agenda in many disciplines (Amaral and Ottino 2004; Natalia 2004), this study points us toward the missing pieces in methods and tools in identifying the most promising opportunities for complex network study in geography.

Furthermore, one of the significant unsolved issues of the current generation of Geographical Information Systems (GIS) is its representation of geographical space by location in an absolute sense (Batty 2003). Although it can derive some basic spatial relations from the geometry, the current generation of GIS in essence is not able to embrace spatial relations or interactions, which poses a significant barrier to its continuous development as Geographical Information Science (GIScience) (Batty 2003). Practically, the current GIS is incapable of representing and analyzing dynamic processes, especially in a systematic way. Based on a network-centric view of geographical systems (Batty 2003), this dissertation in the long run contributes to

GIScience by challenging the current GIS to deal with the dynamics occurring on and within the networks. Expanding the small-world theory with spatial dimensions could potentially contribute to GIScience with new models for representing, analyzing, and modeling complexity from a network-centric perspective (Sui 2006).

1.4 Dissertation structure

The dissertation has five chapters. Chapter I introduces research background, objectives, and significance of the study. Chapter II reviews the literature on small-world networks. Chapter III presents the methodology, followed by results in Chapter IV. Chapter V contains a summary and conclusions.

CHAPTER II

LITERATURE REVIEW

2.1 Emergence of the new science of networks

The network (or graph) theory dates back to the works of two famous mathematicians: Euler and Erdös. Euler's work on the Königsberg bridge problem established the conceptual basis of network theory (Barabasi 2002). He formed the Königsberg bridge problem as a network study and identified the topology as a key issue of the network. Erdös' contribution to network theory is the definition and analysis of the random graph (Amaral and Ottino 2004). Thereafter, networks have been intensively studied in many disciplines including mathematics, physics, geography, sociology, etc. However, the networks studied mostly have a small number of vertices, and edges between vertices are clearly defined. The primary method of network analysts was to draw a clear layout of the network with actual vertices and edges, and to answer questions about the network structure by examining that layout (Newman 2003). The advances of computer science, particularly the computerization of data acquisition and increased computing power, have made it possible to investigate many large-scale complex networks(Watts 2003). With probably millions of vertices and edges, it is quite challenging to visualize those networks. Not surprisingly, research on networks has been shifting its focus from a reductionisic approach, which has been successful for analyzing small size networks' detailed internal structure, to studying large-scale statistical properties of complex networks in hopes of understanding networked systems as a whole (Albert and Barabasi 2002; Newman 2003).

The past several years have witnessed a surge of interdisciplinary research in network science in general and the small-world network in particular, in an attempt to discover the universal structural properties in real-world complex networks and understand the mechanism governing the emergence of these properties (Barabasi 2002; Watts 2003). In addition, the interplay between the dynamics occurring in networks and the structural properties of networks has also been studied extensively (Watts 1999a; Watts and Strogatz 1998). These inquiries have contributed to the emergence of a "new science of networks" (Barabasi 2002; Buchanan 2002; Dorogovstsev and Mendes 2002; Dorogovstsev and Mendes 2003; Watts 2003; Watts 2004) to synthesize the fast developing field across many disciplines, including non-linear dynamics, statistical physics, biological sciences, etc. (Watts 2004). The small-world network, as one of the major accomplishments in the new science of networks, occupies the most prominent place in contemporary thinking about complex networks (Albert and Barabasi 2002; Barabasi 2001, 2002; Newman 2003; Watts 1999b).

Recent empirical studies show that spatial networks, such as an airline network (Amaral et al. 2000), subway network (Latora and Marchiori 2002), railway network (Sen et al. 2003) urban street network (Jiang 2004), and Internet (Gorman and Kulkarni 2004), despite their distinct structures, are small-world networks in nature, though the small-world feature might manifest differently in each. Many real-world networks are spatial networks, in which nodes have well-defined spatial locations, and the nodes and/or links of some networks are physical constructs, such as the highway, subway, or computer networks. Many spatial networks are results of the compromise between minimal connectivity barriers and geographic distance barriers (Black 2001), and their topologies are constrained by geographic embeddings. In addition to connection topology, location, distance, and geography are all vital aspects of these spatial networks (Gorman and Kulkarni 2004). The dissertation, emphasizing the small-world characteristics in spatial networks, intends to bridge the gap between the new science of networks and spatial networks.

2.2 General review of small-world networks

2.2.1 The small-world phenomena of complex networks

The small-world effect in social networks was initially studied by Milgram (1967) through an experiment of passing letters among strangers; he found that it only took an average of six steps for a letter to get from one stranger to another. He thus conjectured that two randomly chosen persons in a human social contact network can be connected only by a very short chain of intermediate acquaintances; this has been subsequently verified and widely accepted and is referred to as the small-world effect (Newman 2003). It was revisited by Watts and Strogatz (1998) and Watts (1999a; 1999b), who demonstrated that the small-world effect is a ubiquitous feature in sparse and decentralized networks. The "six degrees of separation" idea is usually considered an example of the small-world phenomenon in social networks. Networks with the small-world properties are referred to as small-world networks.

The small-world effect on networks implies that there is a surprisingly short separation between two randomly selected vertices, no matter how large and complex the networks are. This has been investigated and verified in a variety of networks, including social networks, information networks, technological networks, and biological networks (Newman 2003). Furthermore, the small-world property of complex networks implies that the average separation of the network increases very slowly with the number of vertices (logarithmically or slower). It is considered one of the universal structural properties of many real-world complex networks.

2.2.2 The small-world network models

The small-world network model was discovered by Watts and Strogatz (1998) through a network "rewiring" process (readjusting an edge to a randomly chosen vertex), and they found that a regular network can be rewired to a small-world network lying somewhere between completely regular and completely random networks (Figure 2-1). The rewiring process introduces certain randomness into the connection topology of a regular network, and it can cause a dramatic increase of network connectivity while the clustering of the network remains. Therefore, a small-world network is both a highly clustered and highly connected network.

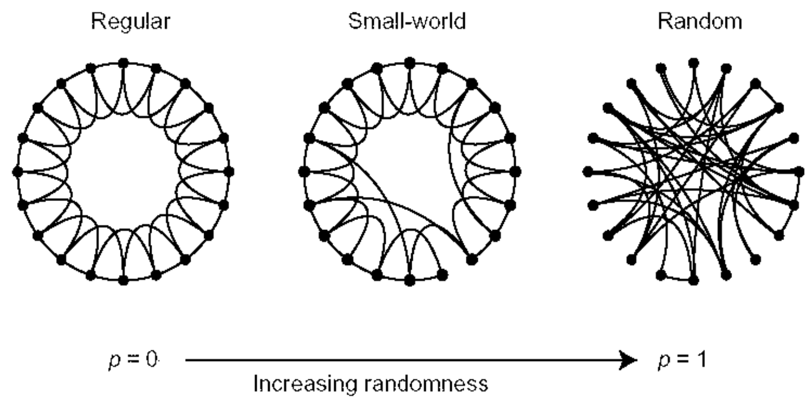


Figure 2-1: Regular, random, and small-world networks after (Watts and Strogatz 1998).

.

There are two small-world models in the literature – the WS model developed by Watts and Strogatz (1998) and the NW model developed by Newman and Watts (1999a; 1999b) and they were both discovered through a network rewiring processes. The rewiring process of the WS model, starting from a ring lattice with n vertices and k edges per vertex, rewires each and every edge of the regular network with probability p to a randomly chosen vertex. When the value of p changes, the regular network will be rewired, according to the changing value of p, to another network whose structural properties can be monitored by two measurements, i.e., the average path length L (the number of edges in the shortest path between two vertices and the average over all pairs of vertices) and the clustering coefficient C (the ratio between the existing edges among neighbors of a vertex and the possible edge in this neighborhood). The value of p

increases from 0 to 1. L and C are therefore functions of p. When $p = 0$, no edges in the completely regular network will be rewired, and the network has a large global separation ($L(0) \approx N/2k$, where N is the number of vertices in the network) and the network is "highly clustered"($C(0) \approx 3/4$). When $p = 1$, all edges in the completely regular network will be rewired, and the network has a small global separation (highly connected) ($L(1) \approx \ln(N)/\ln(k)$) and "poorly clustered" ($C(1) \approx k/N$). However, when the range of p is enlarged between 0 and 1, $C(p)$ of the network is much higher than its random limit $C(1)$, and $L(p)$ of the network is much lower than its regular limit $L(0)$ (Figure 2-2). The networks at these stages are highly clustered like regular networks, yet highly connected like random networks, and were named as small-world networks by Watts and Strogatz (1998) in reference to early works by Milgram and others (Milgram 1967; Travers and Milgram 1969).

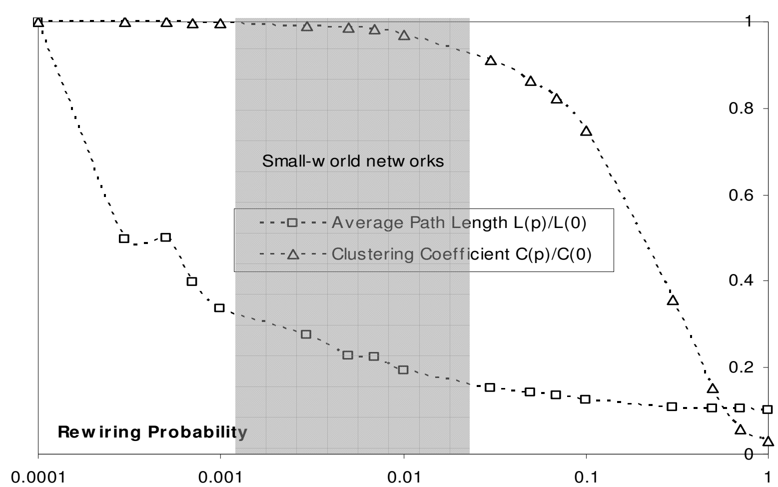


Figure 2-2: Variation of APL and CC of a ring lattice network ($n = 1000, k = 20$) rewired with probabilities increasing from 0 to 1, and small-world networks emerge approximately in the gray region.

Newman and Watts (1999a; 1999b) developed another small-world network model through a little bit different process, which is to add random links instead of rewiring. The rewiring process does not change the size of networks, while adding links does. These two models are one-dimensional; however, the same "rewiring or adding links" process can also be applied to the two-dimensional lattice to develop small-world models in two-dimensional space (Stoneham 1977).

The theoretical small-world network model has a unique connection topology combining regularity and randomness (Watts and Strogatz 1998). The connection topology of the network always affects its function (Strogatz 2001), and different assumptions on the connection topologies of networks lead to different network models. The topologies of network models are usually assumed either completely regular or completely random. Accordingly, they lead to either regular network models to random network models. The regular network models, in which each vertex is linked to exactly the same number of its neighbors, are highly clustered compared to the random network models, but are not highly connected. On the other hand, the random network models, in which vertices are linked randomly, are highly connected but not highly clustered. However, topologies of many real-world networks, such as transportation networks, Internet, scientific collaboration networks, and so on, are neither completely random nor completely regular. They are actually both highly clustered (high local connectivity) and highly connected (high global connectivity)(Amaral and Ottino 2004; Watts and Strogatz 1998). Therefore, neither random network models nor regular network models are able to capture all the structural properties of real-world networks. The small-world network model is able to fill such a void with properties of high global connectivity and high local clustering. The following table (Table 2-1) describes the comparison of three network models: connectivity vs. clustering.

Table 2-1: The comparison of random, regular and small-world networks.

Type of Network	Connectivity	Clustering
Random Network	Short global separation	No clustering
Regular Network	Long global separation	Highly clustered
Small-world Network	Short global separation	Highly clustered

2.2.3 Methods to detect the characteristics of a small-world network

Currently, two methods are commonly used to identify small-world networks. The first one is to calculate the average path length (APL) and clustering coefficient (CC) of a network and compare them with those of a completely random network of the same size (the size of a network is defined as the number of edges a network has)(Watts and Strogatz 1998). For a small-world network, its APL is close to that of a random network, while its CC is much higher. This method is suitable for a topology-dominated network in which the connection topology dominates the network structure. It has been used to identify most of the currently known small-world networks, such as the scientific collaboration network (Newman 2001c; Newman 2001a, 2001b), World Wide Web (Albert, Jeong, and Barabasi 1999), and electronic power grid (Watts and Strogatz 1998), etc. In other words, due to geographic distance barriers (Black 2001), the topologies of many real-world spatial networks are constrained by geographic embeddings and many real-world spatial networks are more complex than the topology alone can characterize. Therefore, focusing on mere topology does not take into account the topology constrained by the geographic embedding in the real-world spatial networks (Boccaletti et al. 2006).

The small-world network was determined initially by topological properties, i.e., average path length and clustering coefficient. However, the small-world effect does not limit its manifestations to mere topological properties. For example, Mathias and Gopal (2001) show that a small-world effect arises from the optimization of network connectivity and cost. For spatial networks, such as the different transportation networks covered in this dissertation, pure connection topology alone cannot characterize the major properties of the network. We thus argue that the small-world in these networks has manifestations different from purely topological properties, while the topological manifestation of the small-world effect leads to future investigations of more complex networks. This argument is empirically supported by recent literature. For example, the two papers by Latora and Marchiori (2002) and Gastner and Newman (2006a) demonstrate that the small-world effect could manifest in terms of network efficiency, which is not closely related to pure topological properties.

The second method is to calculate the network efficiency by taking into account the cost of the network construction and the traveling cost on a network (Gastner and Newman 2006a; Kaiser and Hilgetag 2004; Mathias and Gopal 2001). The concept of network efficiency dates back to Haggett and Chorley (1969), who argue that transport networks tend to reduce the total length of the route to be constructed while increasing the total transport cost, and the network achieves efficiency through balancing route construction cost and transportation cost.

After Latora and Marchiori (2001), the efficiency of networks can be defined as follows,

$$Efficiency = 1/N(N-1)\sum_{i \neq j} 1/d_{ij}$$

Where N is the number of the vertices, and d_{ij} is the shortest path distance between vertex i and j.

The global efficiency of a network can be normalized by the efficiency of a fully connected network of the same size. The local efficiency is defined as the average efficiency of the local sub-graphs. From the perspective of network efficiency, small-world networks are both globally and locally efficient (Latora and Marchiori 2002; Mathias and Gopal 2001).

This dissertation, based on the topological manifestation of the small-world effect, applies network autocorrelation statistics to spatial networks in attempt to discover the manifestation of small-world characteristics in spatial networks. The network autocorrelation method, which extends the Watts and Strogatz (1998) methodology of using a network rewiring process to discover the small-world network, employs network autocorrelation statistics to monitor the simulated network rewiring process, and then detects a signature corresponding to the emergence of small-world networks in network autocorrelation statistics. The network autocorrelation method is able to incorporate the distance effect in studying spatial small-world networks. It also goes a step further to better understand the small-world phenomenon as a result of the interplay between the dynamics and structure of spatial networks. By doing so, this study aims to contribute to the small-world literature a new method for identifying small-world networks in spatial networks and a better understanding of the small-world phenomenon taking placing on spatial networks.

2.3 Previous studies on networks in geographic, epidemic, and cyberspace

2.3.1 Geographic space

The transportation networks are one of the most influential networks in geographical space. The study of transportation networks is usually focused on the network structure (Haggett and Chorley 1969) and its influences on the distribution of human activities (Kansky 1963). For example, Peeters et al (1998) conclude that the grid network tends to foster a dispersed pattern of activities and the center of a radical network acts as an attractor. Many real-world networks are neither completely random nor completely regular, and they therefore conceptually have small-world characteristics (Newman 2003; Watts and Strogatz 1998). The transportation network, as an important real-world network in human society, has already attracted research attention to its small-world characteristics. Through a topological analysis on a functional graph representing the urban street networks in which vertices represent named streets and edges represent street intersections, Jiang (2004) finds that large urban street networks form the small-world networks if one only considers their street connection topology. The transportation network is actually a weighted network, i.e. the links of the network have different weights, such as the road classification and length of the road segments. Using a refined analysis to take into account the different weights of the links, Latora and Marchiori (2002) show that Boston's underground transportation system exhibits small-world behavior in terms of the efficiency of the network. Sen et al (2003) finds that small-world characteristics exist in Indian railway networks represented by a graph where stations are vertices and a train that stops at any two stations is as the link between the vertices. Many transportation networks, either urban street networks or railway, contain small-world characteristics, but they manifest in different ways. However, the existing small-world network studies have yet to consider the spatial dimensions of the transportation network, such as the distance effect.

2.3.2 Epidemic space

Epidemics diffuse through the human population by contact between infective individuals (those were infected and are able to infect others) and susceptible individuals (those are susceptible to be infected). The social network in which the vertices represent individuals and links represent the contacts is considered to be the underlying structure of the diffusion of many phenomena, such as fashions, rumors, innovation, information, and epidemics.

Two major types of epidemic models - compartmental and network models – have been developed to predict epidemic dynamics and inform possible intervention strategies (Keeling 2005). Compartment models classify the population into several epidemiological compartments, such as S (susceptible), E (exposed), I (infective), and R (recovered). People are all susceptible before the first infective is introduced. After being infected, individuals of the susceptible class can enter the exposed class E of those in the latent period, who are infected but not yet infectious. After the latent period ends individuals in class E enter class I of infective, and are capable of transmitting the infection. Infectious individuals could be removed by death or by entering the recovered class R, consisting of those with permanent infection-acquired immunity (Hethcote 2000). This epidemic process is also called the SEIR model. The SEIR model has a few variants, such as the SIR and SIS models.

With compartment models, it is possible to use nonlinear differential equations to model the temporal dynamics of disease propagation (Hethcote 2000). However, compartmental models assume, rather unrealistically, that infective are equally likely to infect any other susceptible individuals (Hethcote 2000; Newman, Jensen, and Ziff 2002; Newman 2002). In fact, epidemics diffuse through human population by contact between infective (those who were infected and are able to infect others) and susceptible people (those are susceptible to be infected); in other words, epidemics diffuse through human social contact networks, in which vertices represent individuals and links represent the contacts. An infective can only infect those he/she has contact with and no one else. Moreover, compartment models ignore the pattern of the disease transmission, which is crucial for the study of control strategies.

Epidemic modeling is experiencing a change from deterministic to stochastic and from compartmental to network paradigms (Koopman 2003). Aiming to increasing realism and heterogeneity, current studies of network modeling of epidemics have modified traditional quantitative models in two aspects: (1) replacing the homogeneity assumption with a network of connections among individuals; (2) considering different probabilities of infection in the connections (Newman 2002; Pastor-Satorras and Vespignani 2001, 2002). Newman et al (2000) and Newman (2002; 2002) show that a large class of standard epidemiological models can be solved in networks using ideas drawn from percolation theory. The small-world network, characterized by high connectivity and high clustering in its structural properties, is considered a potential model for studying social contact networks in which epidemics spread. In order to better understand small-world network models for epidemic modeling, it is necessary to investigate how structural properties of small-world networks affect epidemic dynamics and control strategies. This dissertation is going to fill such a void through studying one epidemic dynamic occurring in a series of networks that are rewired between a completely regular two-dimensional network and a completely random two-dimensional network.

Two major functions of epidemic modeling, for both compartmental and network models, are to predict the epidemic dynamics and inform the possible control strategies (Keeling 2005). The network modeling of epidemics, due to its capability to take into account the heterogeneities in transmission patterns, has provided a modeling platform to examine different control strategies. Several vaccination strategies have been investigated in the literature, i.e., mass, targeted, traced, and acquaintance vaccinations (Eames and Keeling 2003; Huerta and Tsimring 2002; Kaplan, Craft, and Wein 2002; Kiss, Green, and Kao 2005; Pastor-Satorras and Vespignani 2002; Tsimring and Huerta 2003). This dissertation is also going to examine the effect of small-world characteristics on different control strategies.

Vaccine development is usually the first step toward global immunization to combat epidemics. Even there is enough vaccine available for whole population, there is still a need for an improved understanding of how to conduct the vaccination process optimally (Anderson and May 1982, 1985). It is now widely accepted that adopting certain control strategies, rather than using them randomly or arbitrarily, increases the efficiency dramatically (Eames and Keeling 2003). The network modeling of epidemics, due to its capability to take into account heterogeneity in transmission patterns, has provided a modeling platform to examine different control strategies. The goal of control strategies is to reduce morbidity and mortality through reducing the number of susceptibles (by vaccination) and/or restricting transmission (by quarantine or isolation). To achieve this goal, several vaccination strategies have been investigated in the literature, such as mass, targeted, traced, and acquaintance vaccinations (Eames and Keeling 2003; Huerta and Tsimring 2002; Kaplan, Craft, and Wein 2002; Kiss, Green, and Kao 2005; Madar et al. 2004; Pastor-Satorras and Vespignani 2002; Tsimring and Huerta 2003).

A mass vaccination strategy is theoretically capable of preventing the diffusion of an epidemic, and it was used successfully in the eradication of smallpox in the late 1970s (Kaplan, Craft, and Wein 2002). However, this method is not practical economically, and adverse effects to certain population groups can be severe (Ferguson et al. 2003). Also, the mutation of certain pathogens could be faster than the development of new vaccine. In network models, mass vaccination is implemented through random immunization of a certain proportion of the vertices. The targeted vaccination in the network model is usually to immunize a certain proportion of the highest connected vertices (Pastor-Satorras and Vespignani 2002). Traced vaccination or the contact tracing strategy (Eames and Keeling 2003; Huerta and Tsimring 2002; Kiss, Green, and Kao 2005; Tsimring and Huerta 2003) traces the contacts of infected individuals. For the contract tracing strategy in this study, infected vertices are traced for a period of time while their immediate neighbors are checked for possible infection. During this period, the traced vertices will not infect others. After this period, the traced vertices are removed due to either treatment or isolation, and their susceptible neighbors are no long traced. Their infected neighbors will continue to be traced until all neighbors are susceptible. On the other hand, acquaintance vaccination was originally designed to select a random fraction of the vertices and look for a random acquaintance from each chosen vertex to immunize (Madar et al. 2004). Acquaintance vaccination is devised to select a random fraction of vertices and immunize their immediate neighbors until the expected vaccination proportion is reached.

2.3.3 Virtual space

The Internet is composed of millions of computers and related devices connected by various kinds of wires. Since the Internet topology has a strong bearing on management and performance issues, it has stimulated a great deal of studies, including topology modeling (Calvert, Doar, and Zegura 1997), statistical analysis of the topological properties (Vazquez, Pastor-Satorras, and Vespignani 2002), methods to extract the topology (Chang, Jamin, and Willinger 2001), etc. A better understanding of the Internet topology is essential to the construction of an accurate network emulation environment that could serve as a testing platform for routing protocols and traffic modeling(Amin et al. 2002; Mahadevan et al. 2005; Yook, Jeong, and Barabasi 2002). The topology of the Internet can be studied at either the router level or Autonomous System (AS) level. The physical layout of the Internet is a network of interconnected routers, and since each router belongs to an Autonomous System (AS), the Internet is also considered a network of interconnected ASes. At the AS–level, the Internet topology can be represented by a graph with ASes as nodes and AS peerings as links.

An Autonomous System (AS) is defined in the IETF document RFC 1930 (Hawkinson and Bates 1996) as follows: "The classic definition of an Autonomous System is a set of routers under a single technical administration, using an interior gateway protocol and common metrics to route packets within the AS, and using an exterior gateway protocol to route packets to other ASes. Since this classic definition was developed, it has become common for a single AS to use several interior gateway protocols and sometimes several sets of metrics within an AS. The use of the term Autonomous System here stresses the fact that, even when multiple IGPs and metrics are used, the administration of an AS appears to other ASes to have a single coherent interior routing plan and presents a consistent picture of what networks are reachable through it." ASes are the unit of routing policy in the modern Internet of exterior routing. Each AS has a unique AS number (ASN) to identify it on the Internet.

The study of AS-level topology of the Internet has significant implications for the technical, economic and regulatory aspects of the Internet's inter-domain routing system (CAIDA 2006 http://www.caida.org/analysis/topology/rank_as/). Besides the inferring algorithms(Gao 2001; Govindan and Tangmunarunkit 2000), methodology and validation of the AS-level topology (Mahadevan et al. 2005; Siamwalla, Sharma, and Keshav 1999; Zhang et al. 2005) has involved in studies on the modeling and properties of the Internet's large-scale topology(Chen et al. 2002; Falousos, Faloutsos, and Faloutsos 1999; Magoni and Pansiot 2001; Siganos et al. 2003; Yook, Jeong, and Barabasi 2002). Studies on the AS-level Internet have found that the physical layout of the routers and ASes form a fractal set that are determined by population density patterns around the globe(Yook, Jeong, and Barabasi 2002), which helps to better understand the fundamental driving force of Internet evolution. It also has found that the AS-Level Internet topology has shown the power-law distribution (Chen et al. 2002; Falousos, Faloutsos, and Faloutsos 1999; Siganos et al. 2003).

2.4 Summary

According to the empirical studies on social, information, technological, and biological networks, the small-world characteristics are found ubiquitous in complex networks in nature and society (Newman 2003).

The small-world network has a unique topology of high clustering and high connectivity. This topology will affect the function of networks and the processes taking place in the network according to an important understanding derived from the recent network studies (Albert and Barabasi 2002; Strogatz 2001). For instance, the topology of social networks affects the spread of the information and disease (Strogatz 2001), and the topology of transportation networks affects the distribution of human activities (Peeters, Thisse, and Thomas 1998).

The small-world network model is widely considered to be a great potential model for studying real-world complex networks (Granovetter 2003; Hayes 2000; Watts 2003).

The study of small-world networks on spatial networks will complement and expand recent small-world network studies for a better understanding of the application of small-world network models in real-world spatial networks. The theoretical small-world network model, which was discovered through studies on aspatial and topology-dominated networks, only takes into account the connection topology of the networks. In fact, many real-world networks are spatial networks, in which nodes have well-defined spatial locations, and the nodes and/or links of some networks are physical constructs, such as highway, subway, or computer networks. Furthermore, many spatial networks in the real world are results of the compromises between minimal connectivity barriers and geographic distance barriers (Black 2001), and their spatial dimensions are very important properties of the networks. In fact, many processes take place on spatial networks, such as epidemics diffusing through airline networks, viruses spreading over the Internet, and evacuations occurring on transportation networks. Study of the spatial dimensions of the small-world network is a necessary step to further its potential to model processes taking place in real-world spatial networks.

CHAPTER III

RESEARCH METHODOLOGY

3.1 Introduction

Network studies usually fall into three categories (Newman 2003): (1) study of different networks to find properties that can be used to characterize the structure and behavior of networked systems, and to suggest appropriate ways to measure these properties; (2) construct network models according to the common properties discovered; (3) study the dynamics occurring in the networks and their interplays with network structures. The study of small-world networks has been conducted along all three of these lines of inquires (Amaral et al. 2000; Newman 2001; Yook, Jeong, and Barabasi 2002). This study focuses on small-world network modeling and dynamics occurring in small-world networks.

With the rise of large-scale complex networked systems, most of the current network studies have been devoted to searching, characterizing, and understanding hidden regularities and patterns that are considered manifestations of underlying mechanisms governing the dynamics and evolution of these complex systems (Anonymous 2004). Many of these systems, in spite of their apparent complexities and randomness, exhibit characteristics of small-world networks, but different networks may manifest differently. For spatial networks, the distance effect and the relationship between local and global properties are two features that may affect the manifestations of small-world characteristics.

Since a network can be depicted as a graph with a set of vertices connected by a set of edges, graph representation of a network is the first step to studying the structural property of a network. In this study, the three types of networks discussed are represented as graphs. The transportation networks are represented by graphs in which the vertices are intersections and edges are road segments. The social contact network is represented by a graph in which the vertices are individuals and the edges are their contacts. The Internet is represented by a graph in which the vertices are Autonomous Systems (ASes) and the edges are connections (or peering) between ASes.

In the reminder of this chapter, section 3.2 introduces the basic methods of measuring the structural properties of small-world networks. These methods were initially used by Watts and Strogatz (1998) to identify small-world networks during a network rewiring process. This dissertation employs not only the same methods used by Watts and Strogatz (1998), but also other new methods, such as network autocorrelation (in section 3.3) and the relation between local and global properties (in section 3.5) to monitor the emergence of small-world networks during the network rewiring process. Section 3.2 introduces the methodology used to study the small-world effect on epidemic diffusion and control strategies.

3.2 Methods for detecting structural properties of small-world networks

A variety of networks' structural properties can be investigated based on their graph representation, including the small-world effect, clustering or transitivity, degree distribution, network resilience, mixing patterns, degree correlations, community structure, network navigation, etc.(Newman 2003).

In terms of the structural properties of networks, a small-world network has a highly connected and highly clustered topology. If only considering the connection topology, a small-world network has many clustered local links and a few global links among the clusters (Gorman and Kulkarni 2004). The mathematical characterization of the structural properties of the small-world network is based on two variables: average path length (APL) and clustering coefficient (CC). A network can be determined as a small-world network when its APL is close to and its CC is significantly greater than those of a random network of the same size. Following are the definition of the APL and CC.

3.2.1 Average path length (APL)

The average path length (APL) (Albert and Barabasi 2002; Newman 2003; Wang 2003; Watts and Strogatz 1998) or characteristic path length (Watts and Strogatz 1998) measures the typical separation between vertices in a network. It is defined as the number of edges in the shortest path between two vertices, averaged over all pairs of vertices (Watts and Strogatz 1998) to measure the typical topological separation between two vertices in a given network. It can also be defined as the shortest path distance between two vertices in the network, averaged over all pairs of vertices (Latora and Marchiori 2002). The average path length can be represented mathematically as follows.

$$APL(v_i) = \frac{1}{n}\sum_{j=1}^{n} Dist_{\min}(v_i, v_j) \tag{3-1}$$

where $APL(v_i)$ is the average path length of a given vertex v_i;

v_i and v_j are two vertices in the network;

n is the number of vertices in the network

$Dist_{\min}(v_i, v_j)$ is the shortest path distance between v_i and v_j, which could be the topological separation (the number of edges) or the geodesic separation (the physical length of the shortest path).

3.2.2 Clustering coefficient (CC)

The clustering coefficient (CC) (Watts and Strogatz 1998) measures how clustered the network is, or the cliquishness in a typical network neighborhood. In social networks, cliquishness represents the circles of friends or acquaintances in which every member knows every other member. The clustering coefficient can be precisely defined as the fraction of pairs of neighbors of a node that are also neighbors of each other (Wang and Chen 2003).

According to Wang and Chen (2003) and Watts and Strogatz (1998), the clustering coefficient of a given network can be defined mathematically as follows: given a vertex i in the network has k_i edges connecting it to k_i other vertices, in other words, the vertex i has k_i neighbors. There would be k_i $(k_i$ -1)/2 edges if all the neighbors fully connect each other. The clustering coefficient C_i of vertex i is then defined as the ratio between the number of edges E_i that actually exist among these k_i vertices and the total possible number of edges k_i $(k_i$ -1)/2, namely, C_i=2E_i/(k_i(k_i-1)). The clustering coefficient of the whole network is the average over all the vertices. The Figure 3-1 is an example for the calculation, in which the red node has four neighbors and there are actually four links in the neighborhood. If this neighborhood is fully connected, there will be six (4*(4-1)/2 = 6) links. Thus, the clustering coefficient of the red node is 0.67 (4/6).

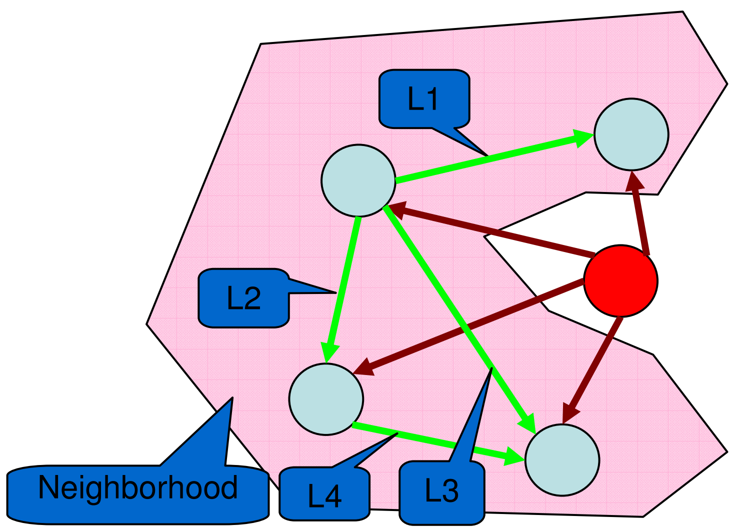


Figure 3-1: Clustering coefficient calculation in the neighborhood of a vertex.

The clustering coefficient measure above is for immediate neighbors of one vertex. It can be extended to the k-clustering coefficient (Jiang 2004), which takes into account the degree to which the k neighbors of a given node are interconnected. The k-clustering coefficient is defined as follows:

$$C^{(k)}(v_i) = \frac{2l_i^{(k)}}{m_i^{(k)}(m_i^{(k)} - 1)} \tag{3-2}$$

where, $l_i^{(k)}$ is the number of edges among k neighbors of vertex i and $m_i^{(k)}$ is the number of nodes within k neighborhood of vertex *i*.

Furthermore, the clustering coefficient can also extend to include edges. The lengths (or weights) of edges can also represent the strength (or cost) of the edges. Given the neighborhood in Figure 3-1, the average length of the edges in the neighborhood can also represent the degree of clustering as the following equation.

$$C_{edge} = 2\sum^{n_{edge}} l_i / (n(n-1)) \tag{3-3}$$

Where n_{edge} represents the number of actual edges in the neighborhood and n represents the number of vertices in the neighborhood.

3.3 Methodology for studying small-world networks in geographic space

3.3.1 Network autocorrelation

Spatial autocorrelation is the correlation of a variable with itself in different spatial locations. Spatial autocorrelation exists when there is any systematic pattern in the spatial distribution of a variable or when nearby or neighboring areas are more alike. A global spatial autocorrelation measure examines the correlation of a spatially distributed variable with itself in different locations throughout the entire area. A positive global spatial autocorrelation indicates that the value of the variable tends to be similar to its neighbors, while a negative spatial autocorrelation indicates the tendency of dissimilarity (Odland 1988). The best-known global spatial autocorrelation statistics are Moran's I and Geary's C(Cliff and Ord 1969, 1973). The local spatial autocorrelation measure scrutinizes the autocorrelation on a local scale. It is usually used to find the concentration (or clustering) patterns, i.e. hotspot or black spot. Flahaut et al (2003) used local spatial autocorrelation to find the black zones. The best-known local spatial autocorrelation statistics include the Getis-Ord G-statistics (Getis and Ord 1992; Ord and Getis 1995) and the Local Indicators of Spatial Association (LISA) (Anselin 1995).

Until recently, studies of spatial autocorrelation are primarily in the data linked with spatial points or areas. However, it is quite possible that the data linked to the vertices of a network has autocorrelation properties as well. The network autocorrelation, as a variant of spatial autocorrelation in the network context, has been studied in social networks (Dow, Burton, and White 1982), transportation networks, and flow systems (Black 1992, 2003; Black and Thomas 1998). Goodchild (1987) suggests that autocorrelation in the network context could exist in two situations: (1) the arc value represents a measurement of the similarity between the linked nodes, and (2) the arc has an actual value. For example, network autocorrelation exists for traffic accidents on road segments that are arcs in a transportation network.

Following Odland (1988), the Moran's *I* is defined as follows,

$$I = \frac{n}{\sum\sum \omega(d)_{ij}} \cdot \frac{\sum\sum \omega(d)_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum (x_i - \bar{x})^2}$$

where n is the number of vertices of the network, and the double summation indicates summation over all pairs of vertices.

x_i is the variable linked to each vertex i; in this paper, it is the average path length from vertex i to all other vertices.

x_j is the variable linked to each vertex j; in this paper, it is the average path length from vertex j to all other vertices.

$\bar{x}$ is the average of x_i over all vertices.

ω_{ij} is 1 if vertex i and vertex j are connected; 0 otherwise.

d is the lag distance (network distance, but not Euclidian distance)

The variance of Moran's *I*, under the assumption of randomization, is

$$Var(I) = \frac{N[(N^2 - 3N + 3)S_1 - NS_2 + 3S_0^2] - K[S_1 - 2NS_1 + 6S_0^2]}{(N-1)(N-2)(N-3)S_0^2} - \left(\frac{1}{N-1}\right)^2$$

$$S_0 = \sum_{i=1}^{N}\sum_{j=1}^{N} \omega_{ij}$$

$$S_1 = 1/2\sum_{i=1}^{N}\sum_{j=1}^{N} (\omega_{ij} + \omega_{ji})^2$$

$$S_2 = \sum_{i=1}^{N}\left(\sum_{i=1}^{N}\omega_{ij} + \sum_{j=1}^{N}\omega_{ji}\right)^2$$

$$K = \frac{1/N\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^4}{\left[1/N\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2\right]^2}$$

$$Z(I) = (I(d) - E(I(d)))\big/\sqrt{Var(I)}$$

Getis and Ord (1992) propose local G statistics to explore the concentration of value in a local neighborhood. In addition, they also propose a General G statistic to estimate the overall concentration of the spatially distributed variables as follows.

$$G(d) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_{ij}(d)x_i x_j}{\sum_{i=1}^{n}\sum_{j=1}^{n}x_i x_j}$$, j is not equal to i.

$$E[G(d)\} = W/[n(n-1)]$$

where $W = \sum_{i=1}\sum_{j=1}\omega_{ij}(d)$, j is not equal to i

The variance of G(d) (Cliff and Ord 1973 p70-71),

$$Var(G) = E(G^2) - [E(G(d))]^2$$

$$E(G^2) = \frac{B_0[\sum_{i=1}^{N} x_i^2]^2 + B_1\sum_{i=1}^{N} x_i^4 + B_2(\sum_{i=1}^{N} x_i)^2\sum_{i=1}^{N} x_i^2 + B_3\sum_{i=1}^{N} x_i\sum_{i=1}^{N} x_i^3 + B_4(\sum_{i=1}^{N} x_i)^4}{[(\sum_{i=1}^{N} x_i)^2 - \sum_{i=1}^{N} x_i^2]^2 N(N-1)(N-2)(N-3)}$$

where

$$B_0 = (N^2 - 3N + 3)S_1 - NS_2 + 3W^2$$

$$B_1 = -[(N^2 - N)S_1 - 2NS_2 + 3W^2]$$

$$B_2 = -[2NS_1 - (N+3)S_2 + 6W^2]$$

$$B_3 = 4(N-1)S_1 - 2(N+1)S_2 + 8W^2$$

$$B_4 = S_1 - S_2 + W^2$$

$$S_1 = \tfrac{1}{2}\sum_{i}^{N}\sum_{j}^{N}(\omega_{ij} + \omega_{ji})^2$$

$$S_2 = \sum_{i=1}^{N}(\sum_{j=1}^{N}\omega_{ij} + \sum_{j=1}^{N}\omega_{ji})^2$$

$$Z(G) = (G(d) - E(G(d)))\big/\sqrt{Var(G)}$$

The definitions of x_i, x_j, and ω_{ij} are the same as in Moran's *I*. This statistic measures overall concentration or lack of concentration of all pairs of (x_i, x_j) such that i and j are within d of each other (Getis and Ord 1992). In the network context, it is therefore a measure of the overall concentration pattern of the network. The higher General *G* value implies high values are more clustered; while the lower General *G* value implies low values are more clustered.

3.3.2 Using network autocorrelation to monitor the network rewiring process

As mentioned above, Watts and Strogatz (1998) use two indices, average path length (APL) and clustering coefficient (CC), to monitor changes in network structural properties during the network rewiring from a completely regular network to a completely random one. This study rewires the network using the same process used by Watts and Strogatz (1998). The rewiring process starts from the regular one-dimensional ring lattice network with 1000 vertices, each connected to the 20 nearest neighbors. For each vertex and the edge that connects it to its immediate neighbor in a clockwise direction, this edge is rewire to a vertex randomly chosen with certain probability $p(0-1)$, and a shortcut link is added only if there is no direct edge between them. Once this process is done to all the vertices in the network, the rewired network is introduced with randomness measured by $p(0-1)$. Repeating the same rewiring process with a different probability $p(0-1)$ will result in a network with different randomness. When $p=0$, no edge in the network will be rewired, and the network is still a regular network. When $p=1$, every edge in the network will be rewired, and the rewired network becomes a completely random network . With the probability increasing from 0 to 1, the network can be rewired into a series of networks between completely regular and completely random.

In addition to APL and CC, this study also uses the network autocorrelation indices - Moran's *I* and General *G* (as defined above) to monitor networks rewired from completely regular to completely random. For every vertex in the network, we calculate the average shortest network path length to all other vertices as an attribute of this vertex, which is a direct measure of the overall connectivity of the vertex in the network. The network autocorrelation statistics including Moran's *I* and General *G* statistics are thereafter calculated based on the attribute of each vertex, namely, the average shortest network path length of each vertex to all other vertices. As the network topology changes, the network autocorrelation indices will be recalculated right away to monitor the change. With the increase of the rewiring probability from 0 to 1, the network autocorrelation indices are used to monitor networks rewiring from a completely regular network to a completely random network. Results can be found in next chapter.

In the definition of clustering coefficient a neighborhood must be determined. A distance is required to determine the neighborhood and therefore to calculate the network autocorrelation statistics. A longer distance will determine a larger neighborhood, and a larger number of vertices will be counted into the calculation of either the clustering coefficient or the network autocorrelation statistics.

The variation of correlation distance in the calculation of network autocorrelation does not change the physical construction of spatial networks. However, it changes their connections logically. On the assumption that dynamics taking place on a vertex will immediately affect the vertices in its neighborhood, the size of neighborhood and therefore the distance will affect the interplay between network dynamics and the network structure. The network autocorrelation measures vary with the correlation distance, and therefore they can take into account the distance effect.

3.4 Methodology for studying networks in epidemic space

3.4.1 The basic characteristics of epidemics

Since many networks perform as substrates on top of which various dynamics take place, the structure of networks will affect the dynamics. For example, the structure of a social contact network will affect the epidemic process taking place in it. A small-world network model is used to simulate a social contact network, on which an epidemic process is simulated. The influence of a small-world network on epidemic diffusion and different control strategies is investigated.

The temporal dynamics of the epidemic diffusion in networks can be characterized by two features: Maximum Epidemic Size (MES) and the time to MES (TMES) (Ferguson et al. 2003; Shirley and Rushton 2005) (Figure 3-2). The MES measures the maximum number of infectives for an epidemic, and the TMES measures how soon the epidemic can reach its MES. Different epidemics may have different MES and TMES. The same epidemic spreading in different networks should have different MES and TMES, and their difference reflects the influence of the network structure on the epidemics taking place in the network.

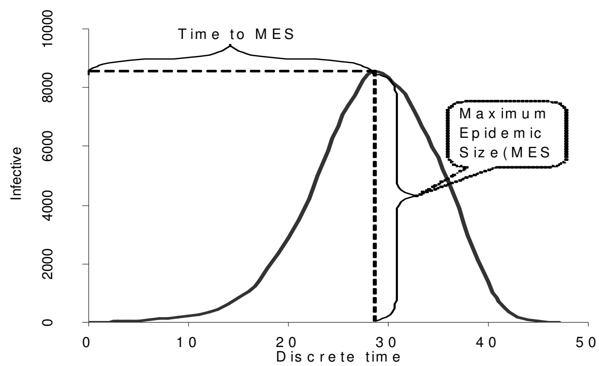


Figure 3-2: The two features of the temporal dynamics of epidemics on networks.

3.4.2 Epidemiological processes on social contact networks

People are classified as different epidemiological classes in traditional epidemiological models (Figure 3-3). The class M represents those with passive immunity. The class S represents susceptible individuals, who could transfer from class M or who are originally without passive immunity. When the transmission occurs from the infective to the susceptible, the susceptible enters the exposed class E in the latent period. The exposed are those who are infected but are not yet able to infect others. When the latent period ends, the exposed class E is able to transmit the infection, and then become the class of I, who is able to infect others. When the infectious period ends, the infective is recovered and becomes class R. Different epidemics have particular epidemiological processes, which could be MESIR, SIR, SIRS, or SIS, etc.

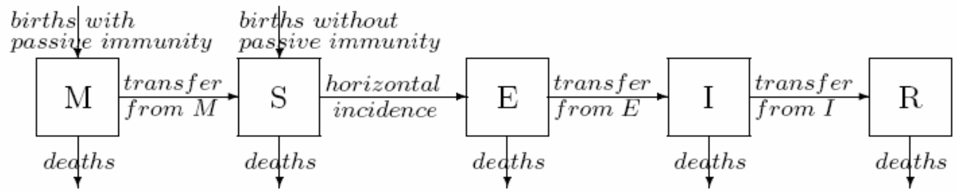


Figure 3-3: The transfer diagram for the MSEIR model with the passively immune class M, the susceptible class S, the exposed class E, the infective class I, and the recovered class R (Hethcote 2000).

This study expands the traditional epidemiological process to simulate an epidemic process in social contact networks. Social contacts in this study are represented by networks, in which the vertices represent individuals and the edges represent contacts between individuals. Each vertex in the network has five possible statuses: susceptible, infectious, recovered, vaccinated, and traced. All the vertices are susceptible until one randomly selected vertex is infected, and the infected vertex will infect its susceptible neighbors with certain probabilities. Infected vertices become recovered after a period of infection, and recovered vertices will remain recovered and immune to infection for the rest of their life cycle. In the contact tracing strategy, infected vertices become traced after a certain period of infection, and their neighbors will also be traced until all neighbors found are susceptible. Traced vertices become recovered if they were infective before tracing, and change back to susceptible if they were susceptible before tracing. Vaccination can only apply to susceptible vertices. Figure 3-4 is a schematic description of the possible change if the epidemic status of a vertex in its life cycle.

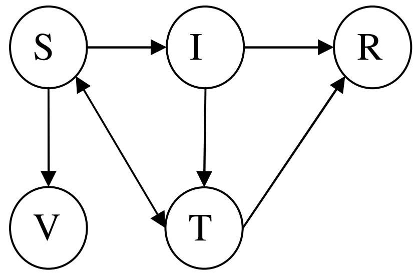


Figure 3-4: The changing map of possible epidemic statuses for a vertex in the social contact network (S=Susceptible, I=Infective, T=Traced, R=Recovered, V=Vaccinated).

3.4.3 Simulation of epidemic diffusion in networks

Each vertex in a network represents an individual who could have different epidemiological status, such as susceptible, infectious, or recovered. Each infected vertex could infect its susceptible neighbors with certain probabilities. In addition, the epidemiological status of each vertex also depends on the status of its neighbors. A susceptible vertex could only be infected by an infectious neighbor. The simulation process starts with a randomly selected vertex. Time proceeds by discrete steps, and the discrete time unit in this simulation could be considered a day. Each vertex has a discrete time counter $\gamma(t)$ to determine its epidemiological status. Each vertex in the network could have five possible epidemiological statuses during the simulation, i.e., susceptible ($\gamma(t)=0$), infective ($0 \le \gamma(t) < 12$), recovered ($12 \le \gamma(t)$), traced and vaccinated. The traced period and infectious period are arbitrarily set at 5 and 12 days, respectively. Periods could be changed, but this will not change the conclusions. The simulation is implemented by a C++ program coded according to the steps described in Figure 3-5.

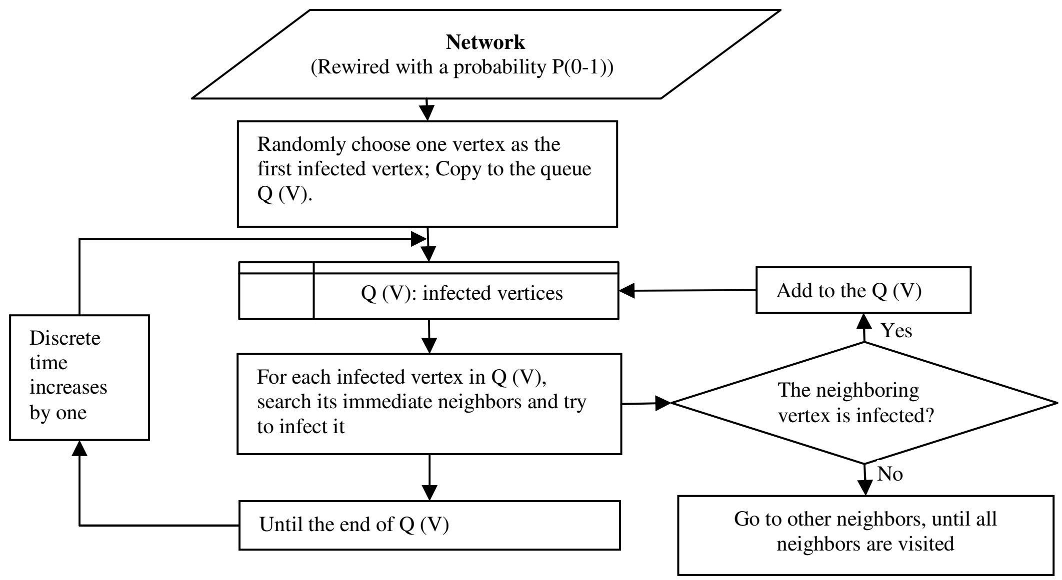


Figure 3-5: The flowchart of the simulation of epidemic diffusion in networks.

The control strategies of mass, targeted, traced, and acquaintance vaccinations are applied to the networks to examine how they affect the epidemic diffusion, and how networks with different structural properties affect these strategies. All the control strategies are implemented as introduced in Chapter II through changing the vertex status to vaccinated. A vaccinated vertex in the network will be immune and will not infect others for a certain period of time. The results will be shown in next chapter.

3.5 Methodology for studying small-world networks in virtual space

3.5.1 Local and global connections

Conceptually, the topology of a small-world network has a large number of local links and a few global links connecting local clusters together (Gorman and Kulkarni 2004).

The concept implies a relationship between local and global links in a network. All the links in a completely regular network are local links, and a large number of links in a completely random network should be global links. To investigate the relationship between local and global links, the changes of local and global links are simulated in a network rewiring process, in which a completely regular network (100×100) is rewired by probability (0-1). The rewiring process follows the procedure of Watts and Strogatz (1998) as described previously.

Three variables are used to monitor the changes of local and global links, including the Local Connection Index (LCI), Global Connection Index (GCI), and the ratio between GCI and LCI. LCI of a vertex is the number of its local links, and the local links could be links to immediate neighbors or links within a defined neighborhood. GCI of a vertex is the number of its global links, and the global links are the links to vertices outside the defined neighborhood. LCI or GCI of a network are the average LCI or GCI over all vertices in the network. Moreover, the Average Path Length (APL) and Clustering Coefficient (CC) are also used to monitor changes of the structural properties of networks and the emergence of small-world networks.

The following figure (Figure 3-6) shows the simulation results. All the variables, including APL, CC, LCI, GCI, and GCI/LCI, are normalized between 0 and 1. When the network is rewired by probabilities increasing from 0 to 1, small-world networks emerge when APL drops quickly while CC remains very high, which is consistent with results reported by Watts and Strogatz (1998). As rewiring probability increases from 0 to 1, LCI remains very high and GCI remains very low in the beginning, then LCI drops and GCI increases very quickly. However, the small-world networks, which are identified by high CC and low APL, have a high LCI and low GCI, or low ratio of GCI/LCI. When rewiring probabilities are greater than 0.1, LCI drops and GCI and GCI/LCI increase rapidly. This simulation resonates with the concept previously described, as small-world networks have a large number of local links and a few global links connecting local clusters.

3.5.2 Local and global degree distributions

Since every vertex could have local connections and global connections, the degree of a vertex can be classified as local degree and global degree. Therefore, the network degree distribution can be divided into local and global degree distributions. The following figures show the overall, local, and global degree distributions for a small-world network and a random network. Both of them are rewired from a 100×100 regular network, in which each vertex links its eight closest neighbors, and the rewiring probability for the small-world network is 0.01 and is 1 for the random network.

In global degree distribution of the small-world network (Figure 3-7), a large number of vertices have 0 degree, implying that these vertices have only local connections. The local degree distribution shows that a large number of vertices have reasonably high

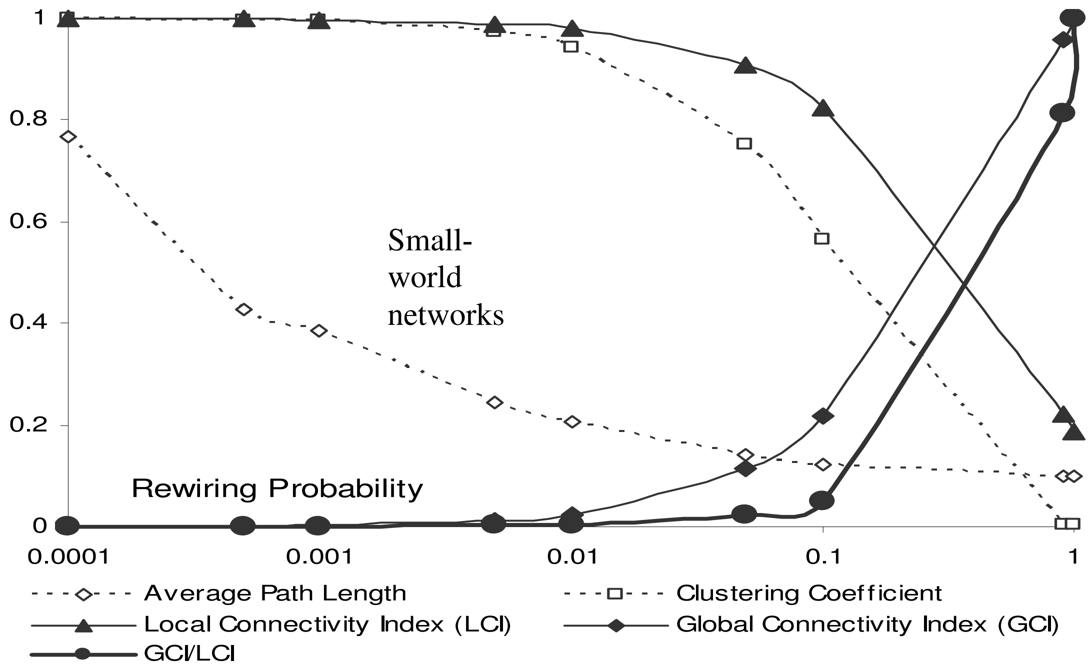
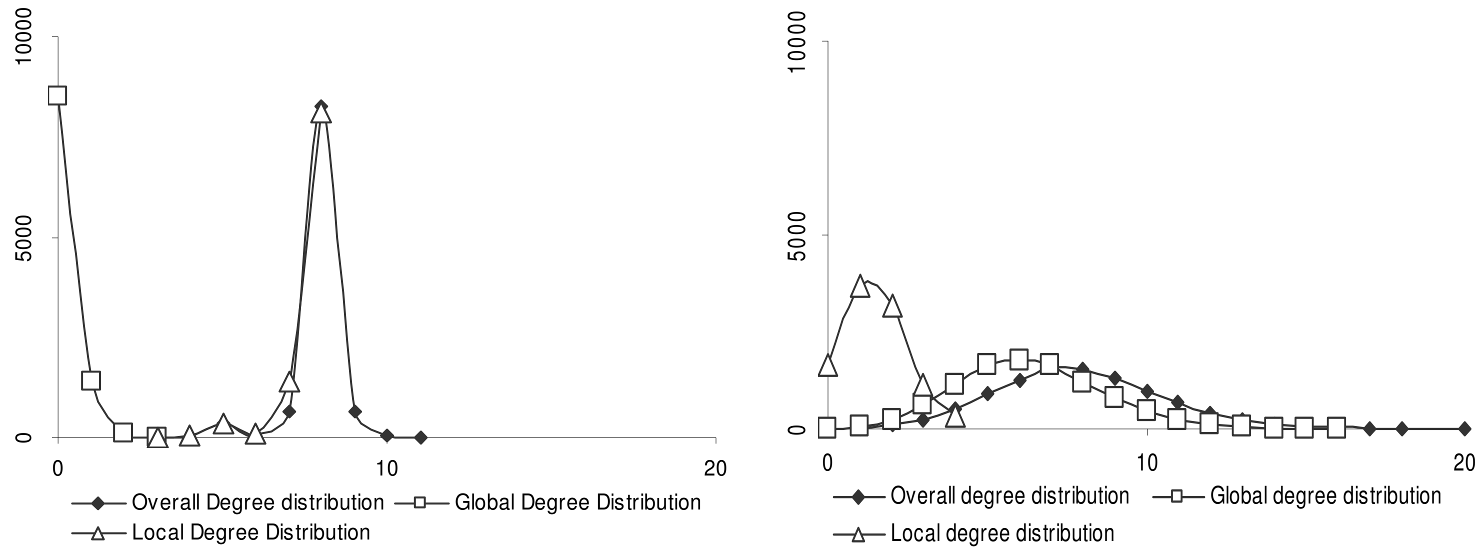


Figure 3-6: Variation of local and global connections when the network is rewired with probability from 0 to 1.

degrees, and no vertices have a higher degree than 10. Since both curves of local degree distribution and overall degree distribution are quite identical, the local degree distribution dominates the overall degree distribution of the small-world network. The conclusions derived from the degree distributions are quite consistent with the widely believed concept that a small-world network has a large number of local connections and only a few global connections linking the local clusters (Gorman and Kulkarni 2004). The degree distributions of the random network (Figure 3-7b) show that global connections dominate the network and a large number of vertices only have a few local connections.



(a) (b)

Figure 3-7: The local and global degree distributions for a small-world and random network.

CHAPTER IV

SMALL-WORLD CHARACTERISTICS IN GEOGRAPHIC, EPIDEMIC, AND VIRTUAL SPACES

4.1 Transportation networks in geographic space

A network autocorrelation approach is used to study the small-world phenomenon on transportation networks at the national, metropolitan, and intra-city levels. The network autocorrelation method, as shown in Chapter III, is used to investigate the structural properties of simulated networks evolving from completely regular to completely random. It is found thereafter that the emergence of the small-world networks corresponds to the convergent and low value of Moran's *I* and General *G* network autocorrelation statistics. This method is further tested on real-world transportation networks. It is found that when the lag distances used to calculate the network autocorrelation statistics reach a certain threshold Moran's *I* and General *G* become convergent and low, implying the distance effect in the logical connection of networks.

4.1.1 The transportation networks studied

Transportation networks at national, metropolitan and intra-city levels are studied. They are the U.S. interstate highway network (Figure 4-1a), the primary road network in the Houston-Galveston area (Figure 4-1b), and the Boston subway network (Figure 4-1c).

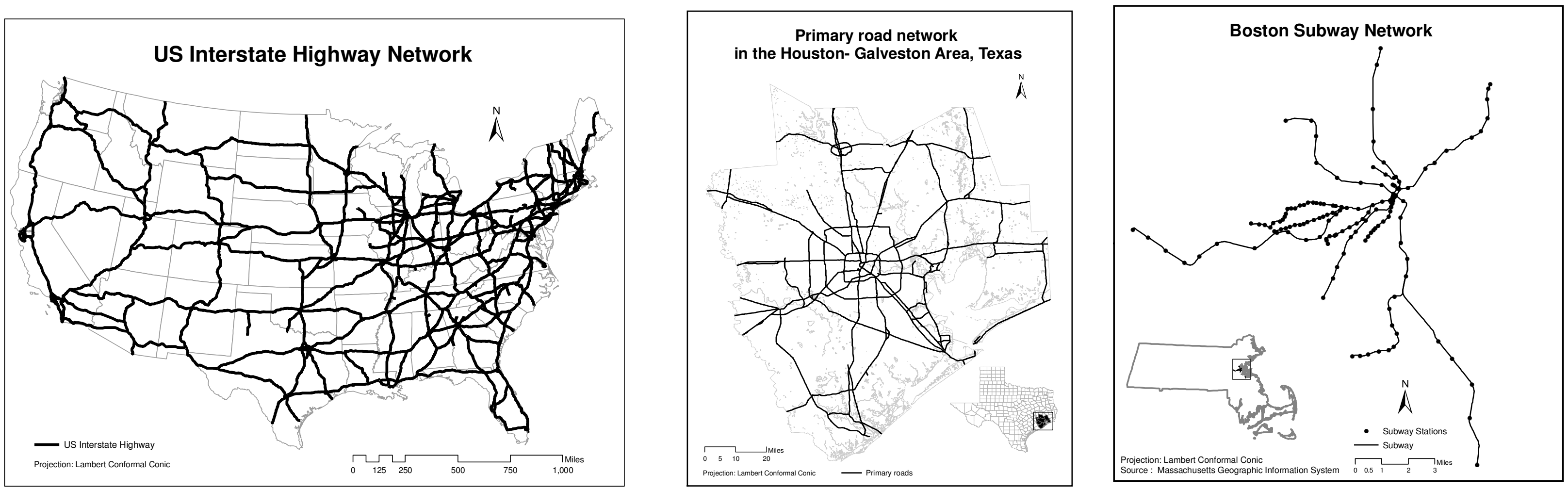
The U.S. interstate highway network has 1238 road segments (edges) and 832 intersections (vertices). The primary road network in the Houston-Galveston area has 370 road segments (edges) and 260 intersections (vertices). The Boston subway network has 148 segments (edges) and 139 stations (vertices).

For each and every vertex in the networks, its shortest network path length to all other vertices over the networks is computed as the variable to be studied by network autocorrelation statistics. It is the direct measure of the overall connectivity of the vertex. The network autocorrelation statistics - Moran's *I* and General *G* - are calculated according to the definitions in Chapter III.

4.1.2 Network autocorrelation analysis

4.1.2.1 Analysis on the simulated network

Two network autocorrelation statistics, Moran's *I* and General *G*, are used to monitor a simulated network rewiring process as described in the previous chapter. The simulation result (Figure 4-2) shows that the small-world network emerges in certain intermediate ranges of probability when the APL drops dramatically while the CC remains very high, which is consistent with the small-world phenomenon discovered by Watts and Strogatz (1998). In the meantime, it is found that the Moran's *I* statistic drops and the General *G* statistic increases, and both statistics tend to be convergent when the networks are rewired as small-world networks, and their values are relatively low. In other words, the small-world network has a special manifestation in their network autocorrelations, namely, their Moran's *I* and General *G* statistics tend to be close to each other, and their values are relatively low. Both network autocorrelations are calculated according to a variable of vertices and the average shortest network path length, which is a direct measure of the overall connectivity of the vertex. The global network autocorrelation, Moran's *I*, measures the overall similarity or dissimilarity in connectivity of the vertices in the network. It drops when networks are rewired from regular to random. The General *G* statistic measures the clustering of both high and low value connectivity of vertices in the network. Its ascendance from a regular to random network shows that vertices in the network change from the clustering of low connectivity to the clustering of high connectivity. The low values of General *G* and Moran's *I* imply that vertices with low connectivity are clustered and their global autocorrelation is low. It is consistent with the concept of network pattern underscoring the small-world network, i.e., the small-world network has a large number of low connected vertices and a few global connections linking local clusters (Gorman and Kulkarni 2004). The tendency of Moran's *I* and General *G* to be convergent implies that the small-world is a compromise between global connection and local clustering. For many real world networks, we can investigate the small-world properties through this signature in their network autocorrelations.



(a) (b) (c)

Figure 4-1: The transportation networks: (a) the U.S. interstate highway network; (b) the primary road network in Houston-Galveston Area; (c) the Boston subway network.

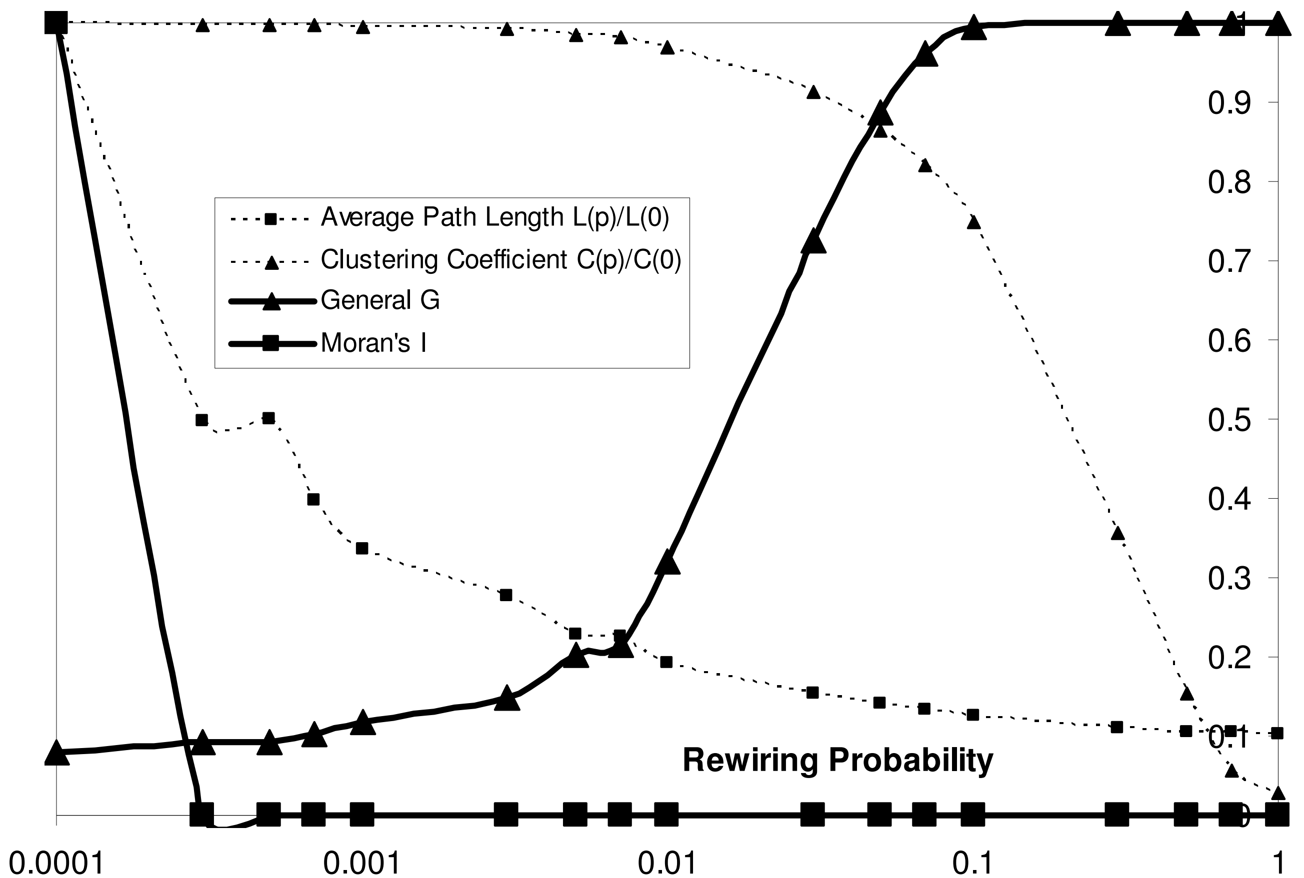


Figure 4-2: Network autocorrelation statistics Moran's *I* and General *G* for the series of network randomly rewired according to rewiring probabilities increasing from 0 to 1.

4.1.2.2 Analysis on real-world transportation networks

For all networks, both Moran's I and General *G* network autocorrelation statistics are calculated according to a series of ascendant lag distances to examine their trends. The results are shown in Figure 4-3. The network autocorrelation statistics, as reflected in the definitions of Moran's I and General *G*, are heavily affected by their lag distances. The lag distance actually determines the size of the neighborhood, namely, how many neighbors are going to be counted in computing network autocorrelation statistics. With the increasing of lag distances, variations of both network autocorrelation statistics are the same for all three networks. The Moran's *I* decreases while General *G* increases, and they become convergent when lag distances reach a certain thresholds (Figure 4-3). When lag distances are lower than thresholds, their increases cause the two network autocorrelation statistics to be identical. When lag distances are larger than thresholds, their increases cause the tendency of the two network autocorrelation statistics to depart (Figure 4-3). Table 4-1 shows the value of thresholds in lag distances and their corresponding autocorrelation statistics for the three networks.

It is interesting that Moran's *I* and General *G* are converging towards approximately similar values (0.32, 0.34 and 0.35) for all three networks. They are the closest points between Moran's *I* and General *G* and relatively low, which implies that the connectivity of vertices has relatively low global autocorrelation and vertices with low connectivity are more clustered than high-connectivity vertices. It enhances the conclusion from the simulated results in Chapter III, and implies that networks at these stages are small-world networks. Although conjecture on the existence of a possible identical convergent value between Moran's *I* and General *G* in different real-world transportation networks is interesting, it needs further investigate statistically and empirically, and it is beyond the scope of this study.

Table 4-1: Thresholds of lag distances for all networks and the converging values of network autocorrelation statistics (Moran's *I* and General *G* are identical).

Networks	**Lag distance (miles)**	**The converging value between Moran's *I* and** General *G*
Interstate hwy	839	0.35
Primary hwy in HGAC	46	0.32
Boston Subway	5.6	0.34

As mentioned before, each and every vertex in all three networks has an attribute, i.e., the average shortest network path length from the vertex to all other vertices in the networks. The attribute of a vertex reflects its overall connectivity in the network. The analysis of semivariance on this attribute studies how changes of lag distances affect this attribute. It is shown that, when lag distances are far less than the thresholds, their changes will not make much difference to semivariances (Figure 4-4), and even when they are close to the thresholds their effects on semivariance are still very small. Figure 4-4 also shows that the threshold distances are all far less than the correlation distance estimated by the variogram models; they are therefore reasonable distances for the correlation calculations. Since the lag distance determines which vertices will be counted as neighbors, on the assumption that dynamics occurring on a vertex will

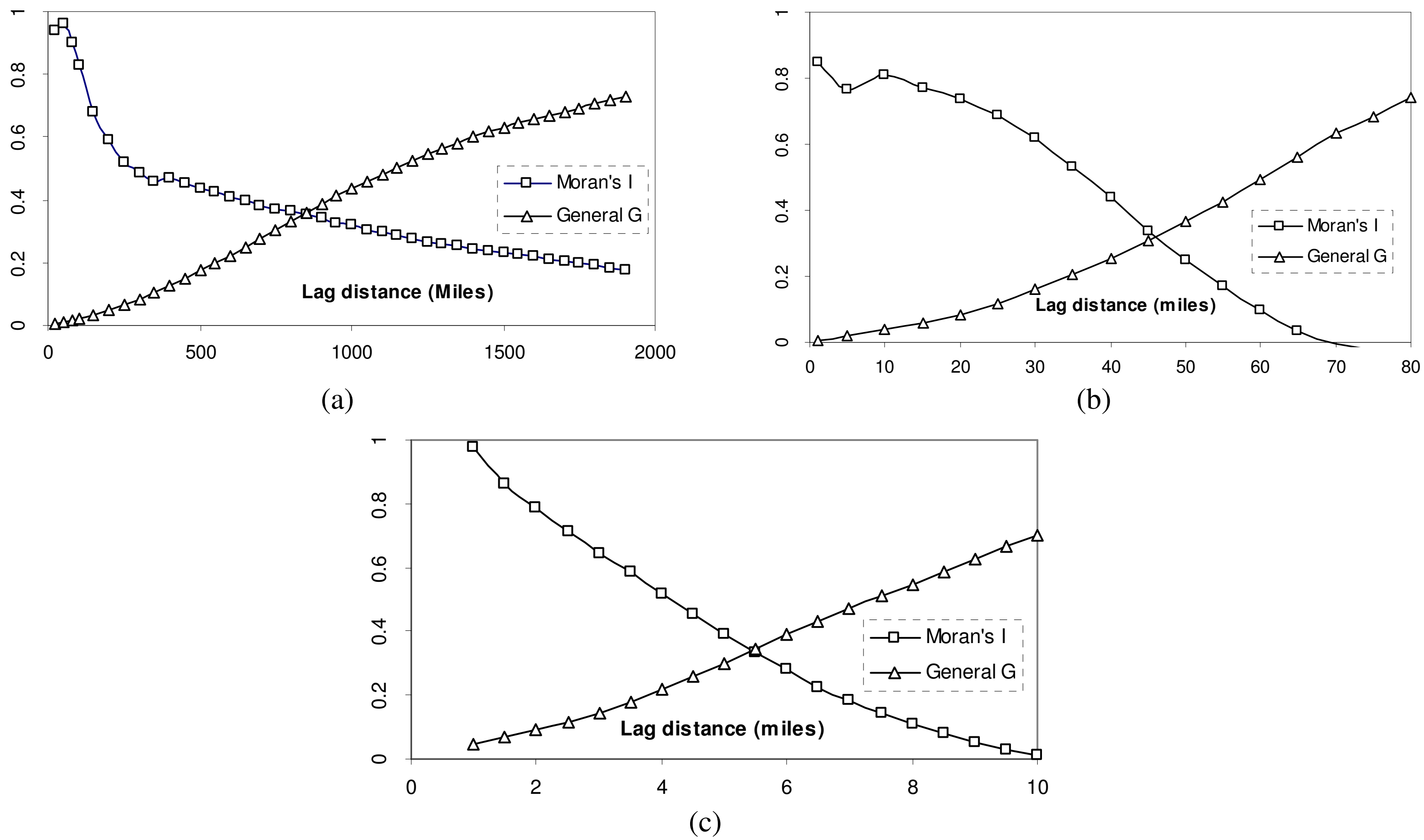


Figure 4-3: Variations of network autocorrelations (Moran's *I* and General *G*) with lag distances for networks: (a) the U.S. interstate highway network; (b) the primary road network in Houston-Galveston area; (c) the Boston subway network.

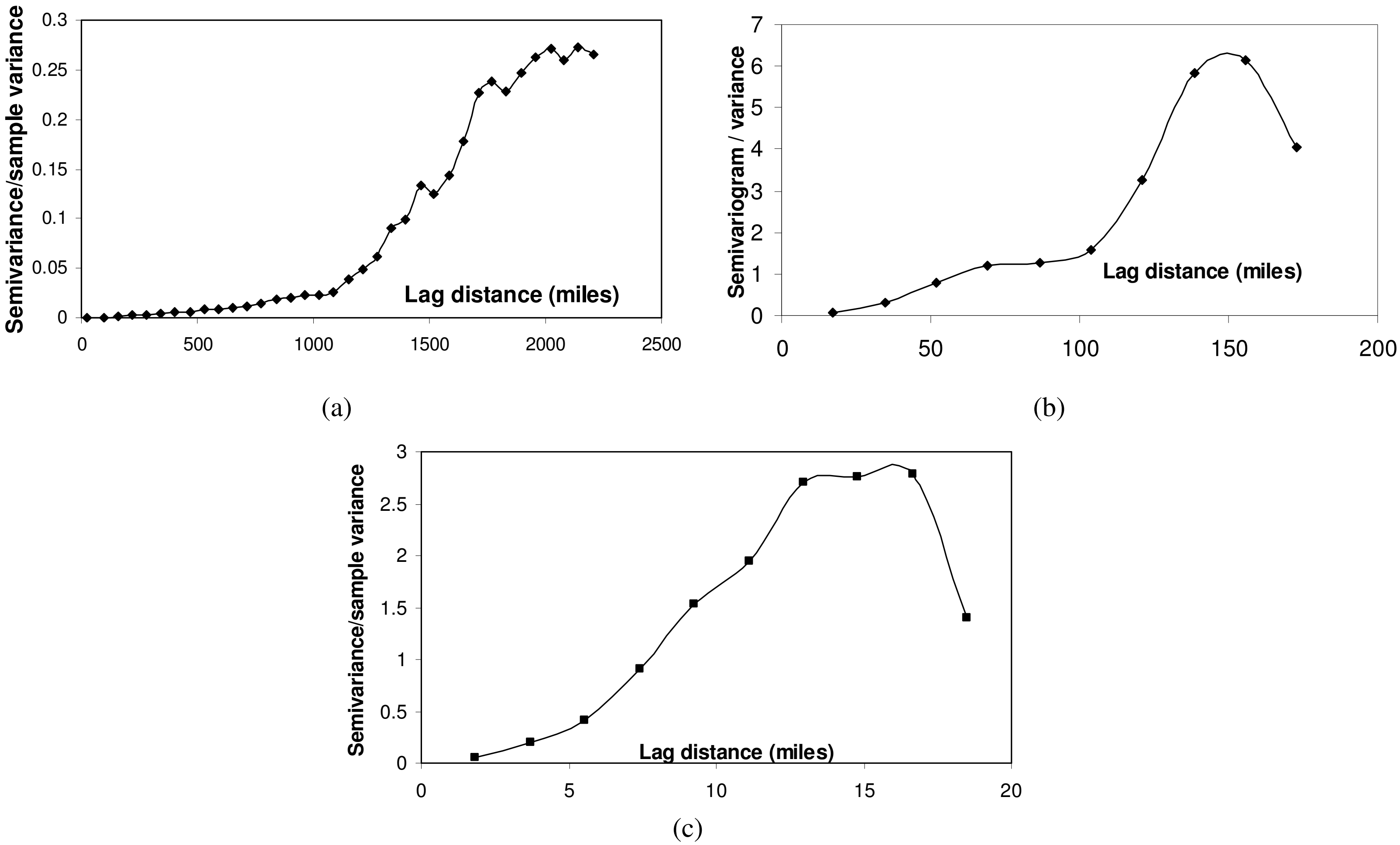


Figure 4-4: The variograms of the average shortest network path length of (a) U.S. interstate highway network, (b) Houston-Galveston highway network, and (c) the Boston subway network.

immediately affect its neighboring vertices (or vertices that are within its neighborhood), the lag distance can be considered as the influencing range of a vertex. It also can be put in this way: if dynamics taking place in vertices in the network have the potential to affect vertices within a distance close to the threshold distance, it will spread extremely rapidly, as if the dynamics are spreading on small-world networks. Although the physical topological structure of the networks is not small-world network, certain dynamics spreading on the networks could logically have small-world effects when their influencing distances are close to a threshold (in the case of interstate highway network, the threshold is 839 miles). Therefore, the small-world phenomena is the consequence not simply derived from the network topological structure, but the interplay between dynamics taking place in networks and the structure of the networks.

Similar analysis and conclusions were also conducted for the primary road network in the Houston-Galveston area, and the Boston subway network. We thus suggest that the primary road network in the Houston-Galveston area, tends to show small-world characteristics when the lag distance is close to 46 miles. For the Boston subway network, it exhibits small-world characteristics when the lag distance is close to 5.4 miles. The Boston subway network was also studied by means of network efficiency (Latora and Marchiori 2002). Although our conclusion is consistent with that of Latora and Marchiori (2002), this paper shows a quantitative method and takes into account the spatial patterns and organization of the network. The three empirical case studies illustrate that these three networks all have small-world properties, but those only emerge when the influencing distance of the dynamics taking place in the networks are close to a certain threshold.

Nevertheless, the Moran's *I* and General *G* statistics for the three networks are approximately the same when the lag distance threshold is reached (Table 4-1). If the convergence of Moran's *I* and General *G* statistics indicates the presence of a small-world network as shown by the simulation results, the networks reach their "smallest" when Moran's *I* and General *G* are identical. Watts and Strogatz (1998) only demonstrate that networks within a certain range of rewiring probability show small-world characteristics. There exists no criterion to identify the "smallest" world network that is supposed exist theoretically. With the increase of the lag distance, the Moran's *I* decreases and the General *G* statistic increases, which indicates that the spatial pattern of the network logically becomes more connected and concentrated. Moreover, at a critical lag distance where the two statistics converged, the network becomes the "smallest" network at a critical status of the compromise between connectivity and concentration (or clustering).

The results are consistent with Barthelemy's (2003) discovery that when distance effect is important, the connectivity distribution has a cut-off depending on the node density. Many small-world networks are often woven with a hierarchical modularity where the hubs play the important role of bridging the many small community clusters into a single, integrated network. The threshold lag distance identified by the network autocorrelation approach is another indication of hierarchical modularities among different spatial networks. Barabasi et al (2003) showed that the scale-free and high clustering of real networks are the consequences of a hierarchical organization, implying that small groups of nodes form increasingly large groups in a hierarchical manner, while still capable of maintaining a scale-free topology.

4.1.3 Summary and discussion

Despite recent voluminous literature on complex networks, there still exist few effective quantitative measures to identify the small-world characteristics of a network. The network autocorrelation approach we explored in this study can fill in this void. Traditional methods to evaluate small-world features have been proven ineffective for spatial networks (Latora and Marchiori 2002). The network autocorrelation approach used in this study has provided a new means to examine the small-world properties in spatial networks.

In this study, it is found that spatial networks on all three of these scales have displayed small-world characteristics when the correlation lag distances reaching certain thresholds make the Moran's *I* and General *G* network autocorrelation statistics convergent and low. The physical structure of the networks does not change, but the different lag distances, which determine the size of neighborhood, logically change their connections. It implies that small-world phenomenon occurring in networks results from the interplay between the network structure and the dynamics taking place on the network.

The proposed method is applied to evaluate the small-world characteristics of transportation networks at three different scales. Each network shows the small-world phenomena when its lag distance reaches a certain threshold. At the national level, it shows that the U.S. interstate highway network shows small-world properties when the lag distance is close to 839 miles. At the metropolitan level, the primary road network in the Houston-Galveston area, exhibits small-world features when the lag distance is close to 46 miles. At the intra-city level, the Boston subway network displays small-world characteristics when the lag distance is close to 5.4 miles.

The lag distance determines the number of vertices that will be considered neighbors of a vertex in a network. Therefore, on the assumption that dynamics taking place in a vertex will affect its neighbors immediately, the lag distance implies the influential distance of a vertex in networks. Although the spatial networks do not change physically, the changing lag distance logically changes the size of neighborhood, and therefore implicitly changes the connection of the networks. Once the influencing distance reaches the threshold, the dynamics taking place in the networks will spread through the network as they would spread on a small-world network. This method can also reveal the vulnerability of a network to certain dynamics that have influencing distances close to the threshold of the networks.

An example could be the travel occurring on transportation networks. The desired intra-city travel distance for most people should be far less than the threshold (for example, 46 miles in the Houston-Galveston area), otherwise it logically makes the network a small-world network where traffic jams could quick spread. The emergency evacuation could fail in the Houston-Galveston area, if the evacuation requires a traffic capacity well beyond the 46-mile design capacity. It implies that small-world phenomenon occurring on the roadway transportation networks results from the consequence of the interplay between the network structure and the travel behavior dynamics taking place on the network, not as a result from the pure connection topology of the network.

Moran's *I* and General *G* statistics of a spatial small-world not only tend to be convergent, but also tend to converge at a pretty much identical value among the three real world networks we tested (Table 4-1). Further investigation on this phenomenon should be done either statistically or empirically, but it is beyond this study.

4.2 Social contact networks in epidemic space

Epidemics spread among the human population through the social contact network that is simulated by a small-world network model in this study. A completely regular 2-D network is rewired by different rewiring probabilities (0-1) to produce a series of networks, including small-world networks. By comparing the characteristics of the same epidemic process taking place in these networks, the effect of the small-world network on epidemic diffusion is analyzed. The effectiveness of different epidemic control strategies, including mass vaccination, acquaintance vaccination, targeted vaccination and contact tracing, is examined. It was found that epidemics taking place in small-world networks take a relatively shorter time to reach large epidemic sizes. The effectiveness of control strategies varies with the randomness and vaccination proportion of networks. A targeted vaccination is the most effective strategy for small-world networks.

4.2.1 The emergence of small world networks in the two-dimensional network

Watts and Strogatz (1998) initially discovered the small-world network model through rewiring a completely regular network by certain probabilities (0-1). During the rewiring process, the small-world network was found within somewhere between the completely regular network and completely random network, and hence the small-world network model has combinative properties of the high clustering of regular networks and the high connectivity of random networks. Although the first small-world network model (Watts and Strogatz 1998) was discovered from a one-dimensional network (a ring lattice network), the same rewiring process can be applied to networks in a two-dimensional context.

The small-world networks in this study are obtained by rewiring a regular two-dimensional lattice using the same rewiring process used by Watts and Strogatz (1998).

The rewiring process generates a series of networks, including the small-world networks. The same epidemic process and different control strategies are simulated on these networks so that we can compare the effect of networks with different structural properties on epidemic diffusion and control strategies.

The rewiring process is monitored by two structural properties of networks, the Average Path Length (APL) and the Clustering Coefficient (CC). Using the same definitions as Watts and Strogatz (1998), the APL is the average number of edges in the shortest path between two vertices and average over all pairs of vertices in the network, and the CC of a vertex is the ratio between the existing edges and all possible edges among its neighbors. The CC of the whole network is the average of the CC over all vertices. The rewiring process starts from a regular two-dimensional lattice network with 100×100 vertices, each one connected to its eight nearest neighbors, except the vertices along the boundary have five nearest neighbors and the vertices at the four corners only have three nearest neighbors. For every vertex, all its edges are rewired to a vertex randomly chosen with probability $\phi(0-1)$, and a shortcut link is added only if there is no direct edge between them. With the probability increasing from 0 to 1, a series of networks are obtained by rewiring the network from completely regular to completely random (Figure 4-6). Consistent with Watt and Strogatz (1998)'s results, small-world networks emerge somewhere between a completely random network and a completely regular network when the APL drops dramatically while the CC remains very high (Figure 4-5).

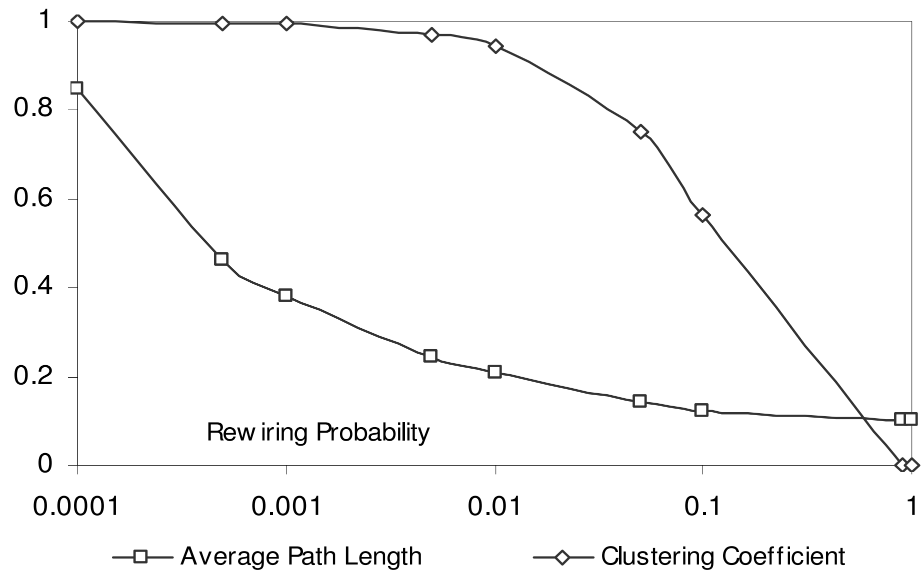


Figure 4-5: The variation of the structural properties (average path length and clustering coefficient) of the networks rewired by probability $\phi(0-1)$.

In the two-dimensional network, each vertex represents an individual who could have a different epidemiological status, such as S (susceptible), I (Infectious), or R (recovered). Each infected individual could only infect its neighbors linked by the network. The status of each vertex in the two-dimensional network will change as time progresses. Some individuals are infected while others are cured; the epidemic diffuses in the space represented by a two-dimensional network.

4.2.2 The small-world effect on the spatial diffusion of epidemics

The same epidemic dynamics is simulated on a series of network models rewired with different rewiring probability (0-1). Simulation results show that the connection topology severely affects the epidemic dynamics taking place in the networks. The MES

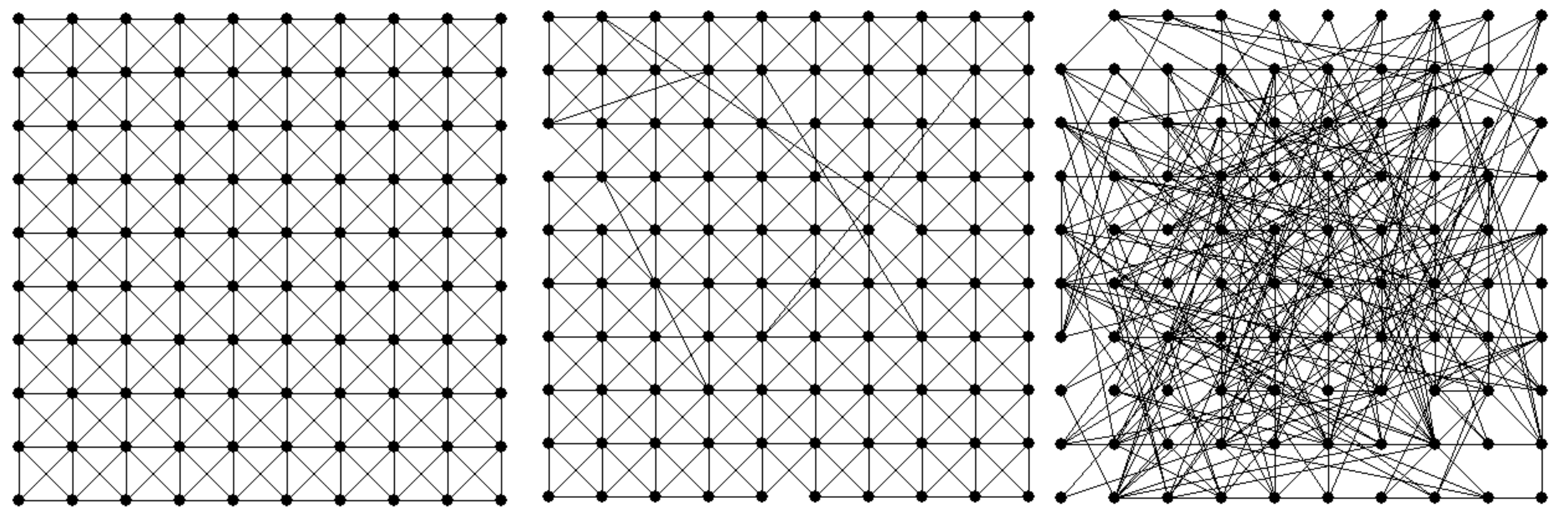
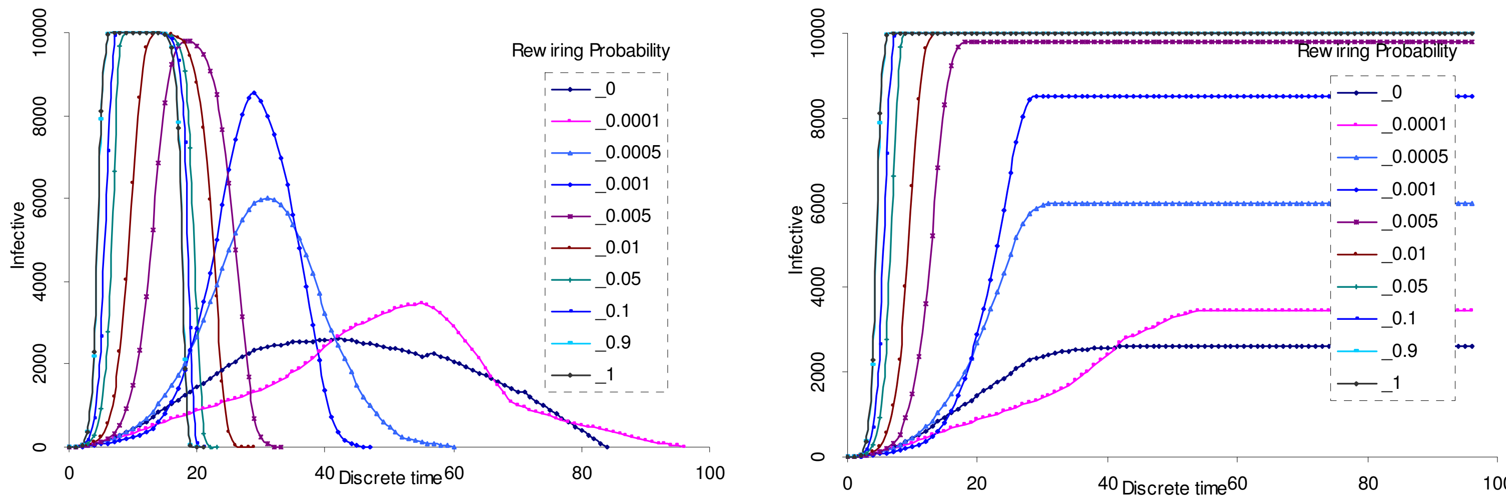


Figure 4-6: A simplified version of the lattice network with 10×10 vertices rewired from the completely regular network ($\phi = 0$) to the completely random network ($\phi = 1$). The small-world networks ($0 < \phi < 1$) are emerged somewhere between the completely regular network ($\phi = 0$) and the completely random network ($\phi = 1$) when the APL drops dramatically and the CC remains very high.

(defined in the previous chapter) increases with the rewiring probability, and the TMES (defined in the previous chapter) is negatively related to the rewiring probability (Figure 4-7). When the epidemic occurs in a completely regular network (rewiring probability of 0), it can only affect 26% of the vertices; the epidemic can infect the whole network if its rewiring probability is greater than 0.01. In completely random networks, the epidemic has the largest MES and takes the shortest time to reach it (Figure 4-7).

Networks rewired by different rewiring probabilities have different structural properties. As mentioned before, Watts and Strogatz (1998) used two variables, Average Path Length (APL) and Clustering Coefficient (CC), to monitor changes in the structural properties of networks and detect the emergence of small-world networks. In this paper, we use these two variables to not only detect the emergence of the small-world networks, but also investigate how they affect the MES and the TMES. When network rewiring probability increases a small amount, such as from 0 to 0.001, the APL of the network drops dramatically while the CC remains very high; at the same time, the MES has a significant increase and the TMES drops dramatically (Figure 4-8). The MES has a sound linear relation with its TMES ($y = -159.77x + 11618, R^2 = 0.8575$) (Figure 4-9) which implies that epidemics generally take a shorter time to reach a relatively larger maximum epidemic size. The introduction of a small amount of random links to the regular network will cause a huge increase in the epidemic size, and the epidemic will take a very short time to reach its largest size. For example, the MES reaches 36% when the rewiring probability is 0.0001, but increases to 85% when the rewiring probability



(a) (b)

Figure 4-7: The epidemic dynamics of infective (a) and cumulative infective (b) on networks of different rewiring probability (0-1).

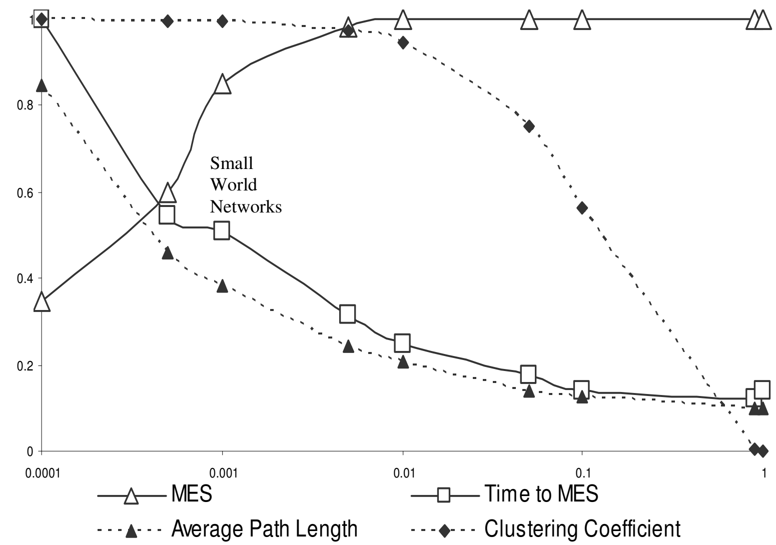


Figure 4-8: The variation of MES and TMES on networks of different rewiring probabilities.

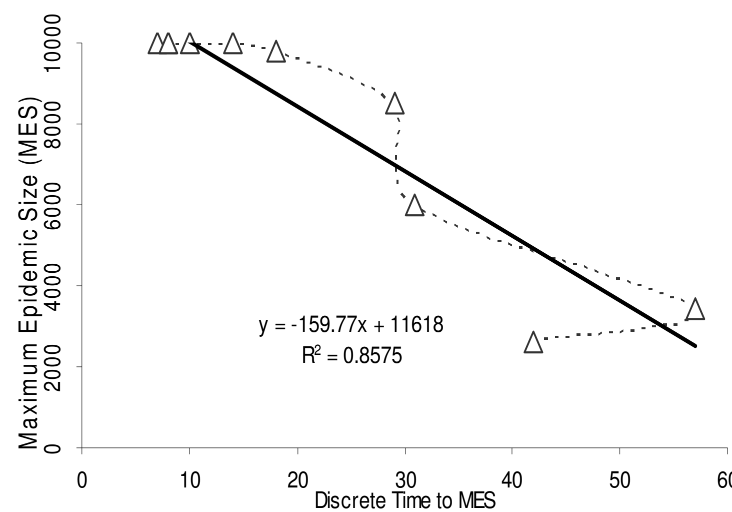


Figure 4-9: It generally takes shorter time to reach a larger MES.

increases to 0.001. The epidemic can spread to the whole network when the rewiring probability is greater than 0.01. This trend tells us that epidemics occurring on small-world networks will take shorter times to reach large epidemic sizes.

Varying values for APL and CC also affect the epidemic dynamics differently (Figure 4-10 and Figure 4-11). The APL has linear relationships with both MES ($MES = -0.8923APL + 1.1257, R^2 = 0.9601$) and TMES ($TMES = 0.8667APL + 0.0805, R^2 = 0.8757$) (Figure 4-10). A high APL corresponds to a low MES and a high TMES. Unlike APL, CC has completely different relations with both MES and TMES (Figure 4-11). The MES and TMES rapidly reach extreme status when the CC is still very high (for example 0.9 or higher), but when CC is lower than 0.9, the corresponding MES and TMES do not change significantly (Figure 4-11). When the network is rewired to the status of small-world networks, i.e., the network has high CC and low APL, the epidemics have a large MES and small TMES.

4.2.3 The small-world effect on control strategies

Following Zanette and Kuperman (2002), we use the maximum infection rate to compare the effectiveness of different control strategies. The maximum infection rate is defined as the ratio between the maximum infective population and total population.

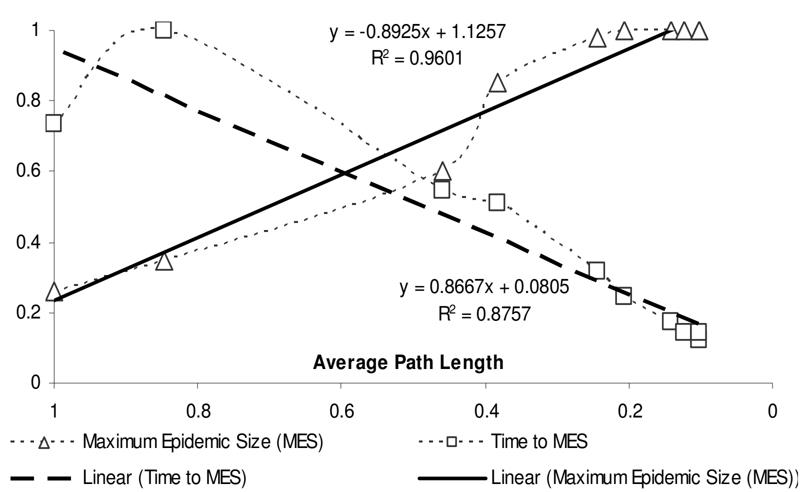


Figure 4-10: The relations between average path length and MES and TMES.

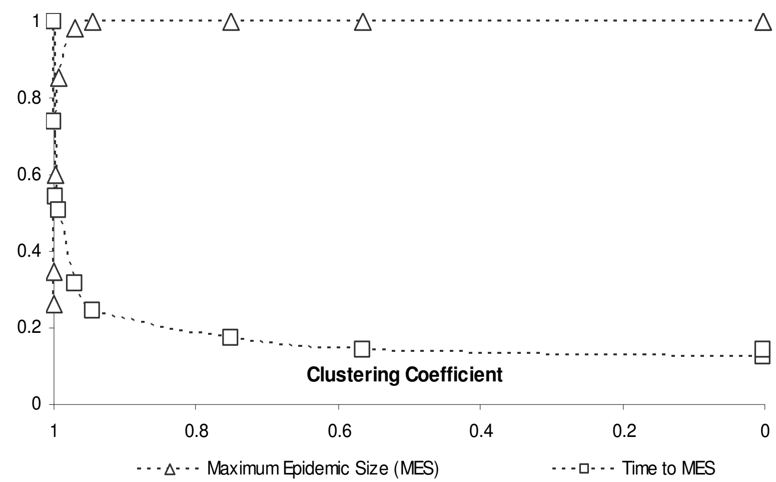
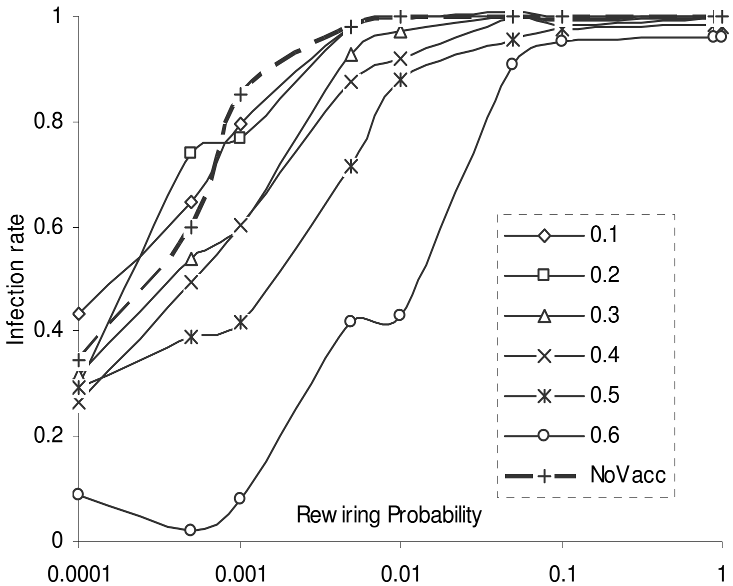


Figure 4-11: The relations between clustering coefficient and MES and TMES.

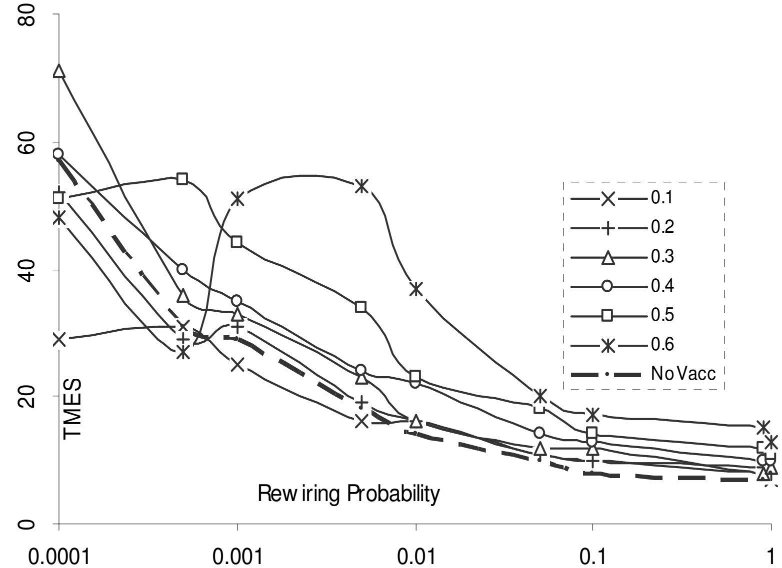
As shown in Figure 4-12 (a) (c) (e) and Figure 4-13, all control strategies become less effective with the increasing of the rewiring probability, since the maximum infection rates have dramatic increases for all control strategies as the rewiring probability increases. When rewiring probabilities are greater than 0.01, the maximum infection rates could reach 1, implying that all susceptible individuals could be infected. We also investigate the sensitivity of vaccination strategies on different proportions of vaccination. However, we only investigate the vaccination proportion less than or equal to 50% (except mass vaccination for 60%) since the simulated epidemics could be much more localized when the vaccination proportion is higher than 50%, due to our simulation starting with only one randomly selected vertex in the network. For mass vaccination and acquaintance vaccination, the maximum infection rates decrease with the increasing of the vaccination proportion. Targeted vaccination has very similar low infection rates when vaccination proportions are 0.1, 0.2 and 0.3, but has a huge variation in infection rate when vaccination proportions are 0.4 and 0.5. Since targeted vaccination immunizes highly connected vertices first, a small vaccination rate could be more effective as shown in figure 11(e). Comparing Figure 4-12 (a) and Figure 4-12 (c), we found that acquaintance vaccination is more effective than mass vaccination, particularly when the vaccination proportion is high, such as 0.5. Moreover, acquaintance vaccination performs better than mass vaccination when the rewiring probability is high, for example, when rewiring probability is 0.1, and for 50% vaccination rate the maximum infection rate is 0.822 for acquaintance vaccination (Figure 4-12(c)), but 0.9772 for mass vaccination (Figure 4-12(a)). The time to reach maximum epidemic size (TMES) for all the control strategies decreases with the increasing of the rewiring probability (Figure 4-12 (b), (d) and (f)). In general, the higher proportion of vaccination corresponds to longer TMES. However, when rewiring probability is very low, TMES may have a huge variation.

The purpose of contact tracing, or traced intervention, is to trace the neighbors of infected individuals. Tracing stops when all neighbors of the infected individuals are susceptible and no one else is infective. After a period of tracing, susceptible vertices will return to susceptible, and infected vertices will change to removed. The infection rate of contact tracing increases dramatically when the rewiring probability increases. The TMES decreases with the increasing of the rewiring probability. Figure 4-13 shows the variation of the MES and TMES of contract tracing while increasing the rewiring probability, and their comparison with the MES and TMES of the epidemics without any intervention.

In order to study the performance of different control strategies on low and high proportion of vaccination, we compare the effectiveness of control strategies when 10% and 50% susceptible are vaccinated. As shown in the Figure 4-14(a), when 10% of the susceptible are vaccinated, the targeted vaccination is the most effective. Contact tracing is more effective than mass and acquaintance vaccinations. Mass and acquaintance vaccinations show almost identical results; however, the acquaintance

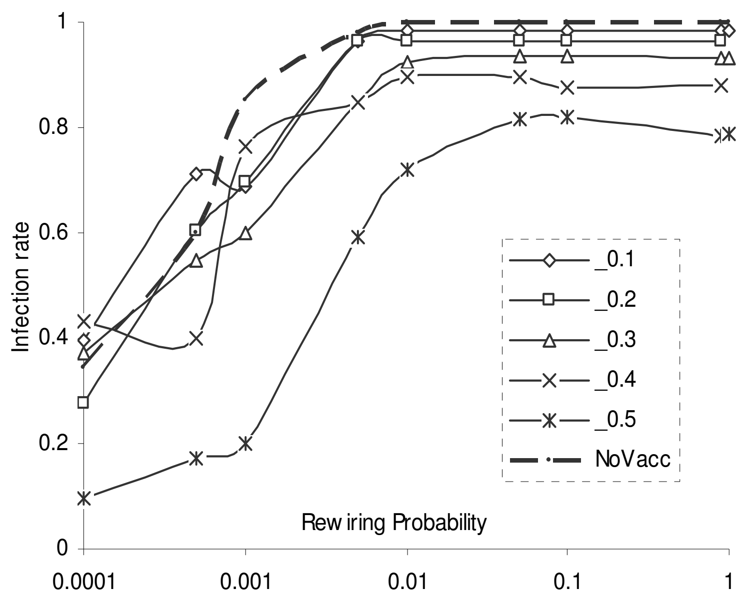


(a) Mass vaccination

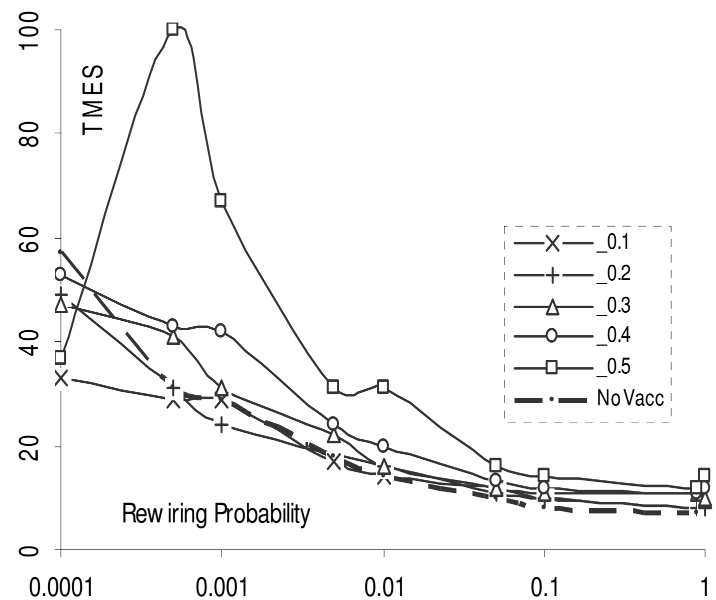


(b) Mass vaccination

Figure 4-12: The variation of the maximum infection rate and TMES with rewiring probability under different control strategies.

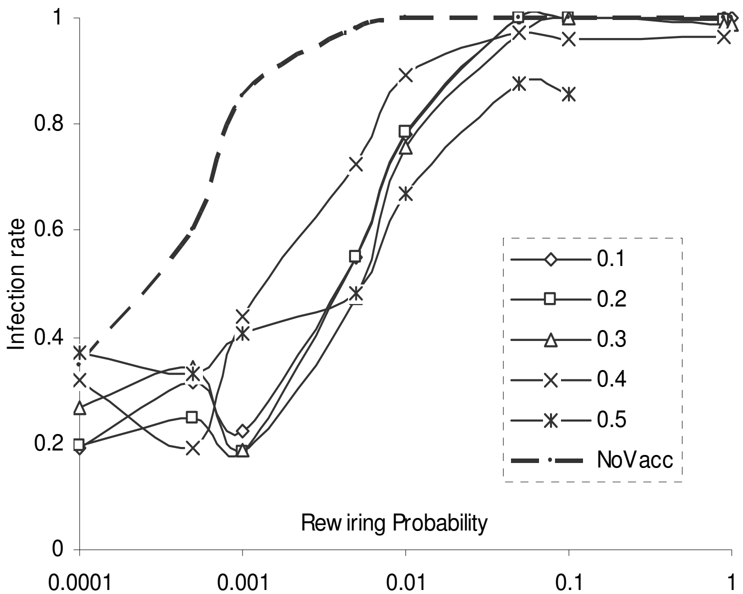


(c) Acquaintance vaccination

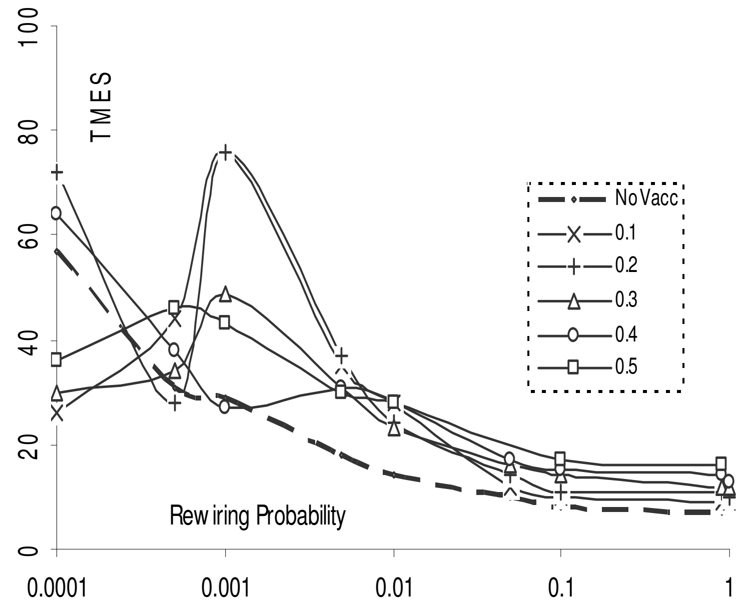


(d) Acquaintance vaccination

Figure 4-12 Continued.



(e) Targeted vaccination



(f) Targeted vaccination

Figure 4-12 Continued.

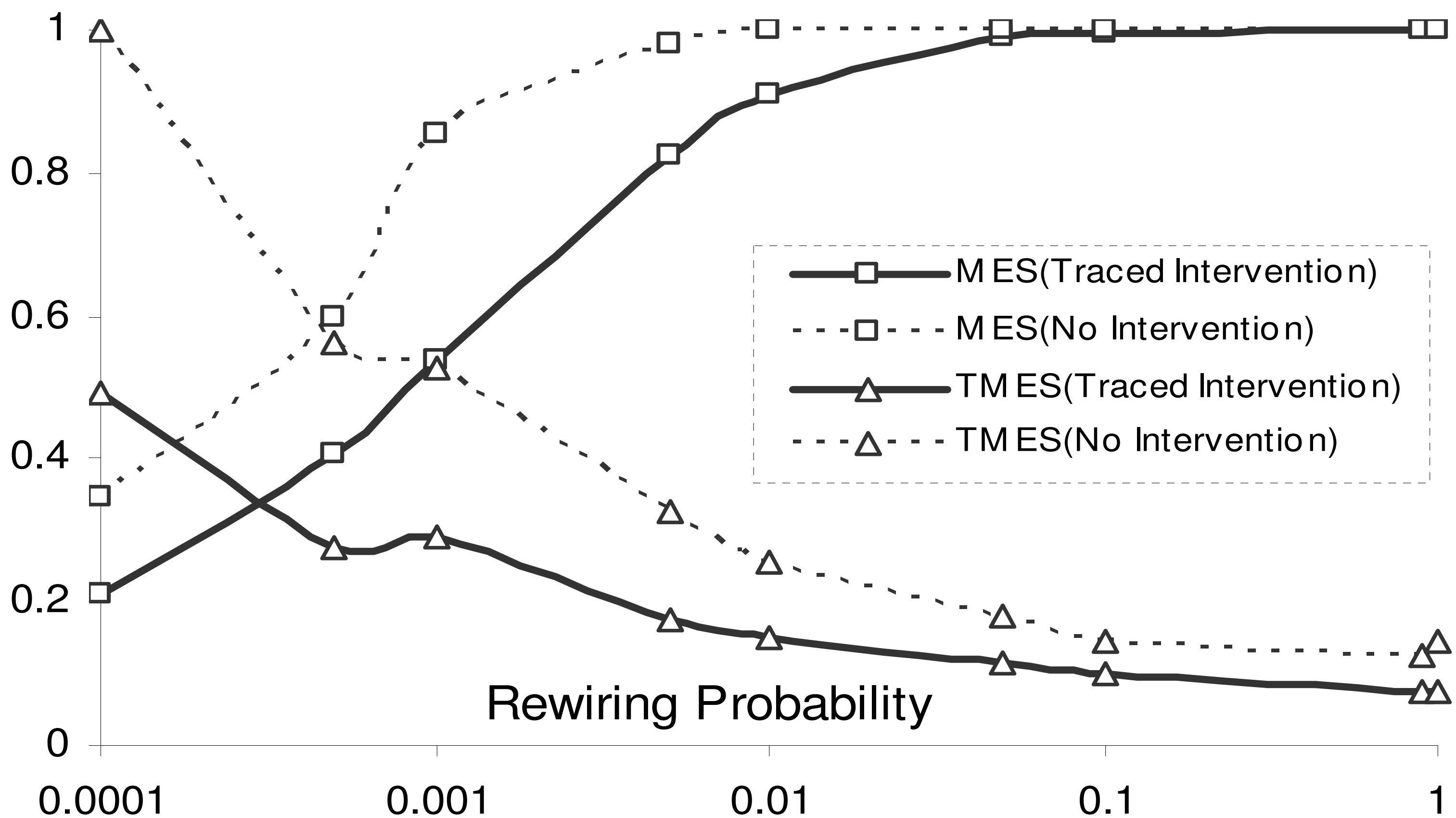


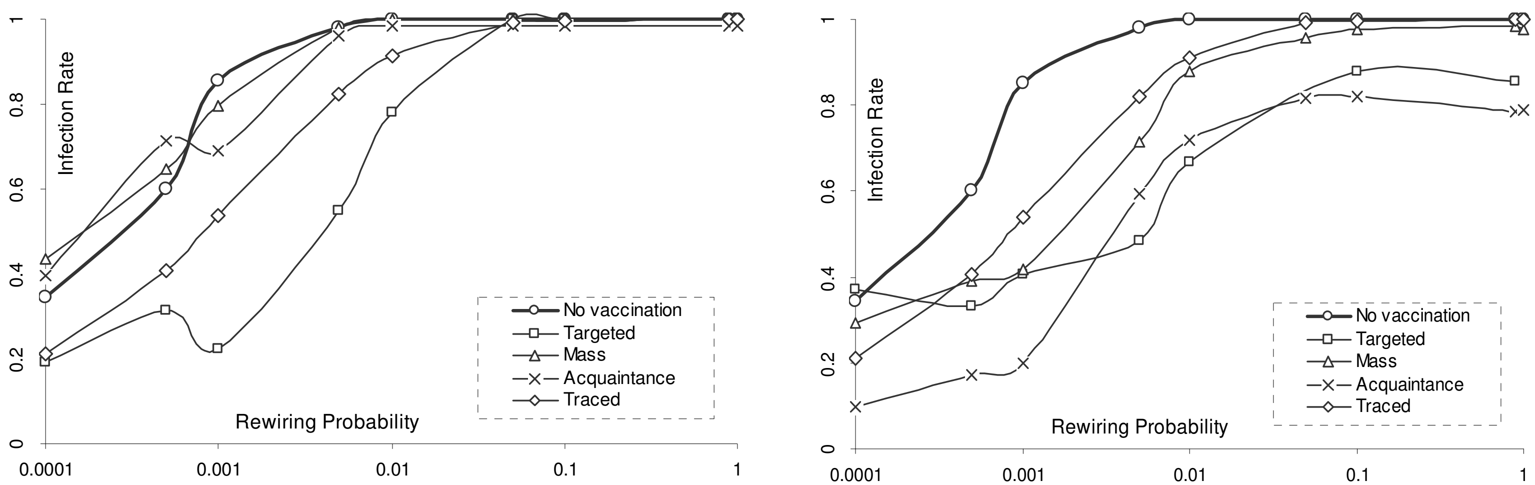
Figure 4-13: The variation of MES and TMES under the contact tracing strategy with the rewiring probabilities.

vaccination has more variation when rewiring probability is low. As shown in the Figure 4-1(b), when 50% susceptible are vaccinated, acquaintance vaccination is the most effective whether networks are rewired by low or high probabilities. Targeted vaccination is the most effective when networks are rewired by medium probabilities. Since completely random networks and completely regular networks are both homogenous, rewiring networks with either high or low probability tends to make them more homogeneous. Targeted vaccination is not effective on homogenous networks when these networks have low or high rewiring probabilities.

The effectiveness of control strategies is affected by the randomness and vaccination proportion of networks. Table 4-2 summarizes the most effective control strategies.

Table 4-2: The effective control strategies for networks with different vaccination proportion and randomness.

<table>
<tr><td colspan="2" rowspan="2">The effective control strategies</td><td colspan="2">Vaccination proportion</td></tr>
<tr><td>High</td><td>Low</td></tr>
<tr><td rowspan="3">Network randomness (rewiring probability)</td><td>High</td><td>Acquaintance</td><td>Acquaintance</td></tr>
<tr><td>Small-world</td><td>Targeted, acquaintance</td><td>Targeted</td></tr>
<tr><td>Low</td><td>Acquaintance</td><td>Targeted, Traced</td></tr>
</table>



(a) (b)

Figure 4-14: Comparison of the control strategies on networks with 10% (a) and 50% (b) of the vertices vaccinated.

4.2.4 Summary

Following and expanding the original methodology (Watts and Strogatz 1998), this study has conducted the simulation of the emergence of small-world networks betweencompletely regular and completely random networks in a two-dimensional lattice. On top of that, the effects of the small-world network on epidemic dynamics and different control strategies have also been investigated.

It is found that changes in the two structural properties of the small-world network, i.e. average path length and clustering coefficient, have significantly different effects on epidemics taking place in the network. The epidemic dynamics can be characterized by two properties, i.e., the maximum epidemic size and the time to maximum epidemic size. The average path length of the networks has linear relations with both properties; a small average path length corresponds to a large maximum epidemic size and a short time to reach the maximum epidemic size. The maximum epidemic size and the time to maximum epidemic size can rapidly peak when the clustering coefficient is still very high. When the clustering coefficient is lower than 0.9, it does not affect the epidemics significantly. The small-world network is highly connected (small average path length) and highly clustered (high clustering coefficient) network. Epidemics taking place on the small-world networks therefore take a relatively shorter time to reach large epidemic sizes.

The effectiveness of control strategies is affected by the perturbation of the network connections and the vaccination proportion of networks. Among the different control strategies simulated, the targeted control strategy is the most effective for a small-world network, while acquaintance vaccination is effective when a high proportion vaccination is involved.

4.3 AS-level Internet graph in virtual space

There are two major networks related to the virtual space, i.e., the World Wide Web (WWW), which is a network of Web pages containing information, linked together by hyperlinks from one page to another (Huberman 2001), and the Internet, which is a physical network of computers (or routers) linked together by optical fiber and other data connections. The Internet can be represented at router level or AS-level (defined in Chapter II). This study only investigates the AS-level Internet in which the vertices are ASes and the edges are the connections (or peerings) between ASes.

4.3.1 The AS-level Internet graph for the contiguous U.S.

The Internet AS-Level graphs can be extracted from three sources : (1) traceroute measurements; (2) BGP (Border Gateway Protocol); and (3) the WHOIS database (Mahadevan et al. 2005). Traceroute (www.traceroute.org) is a program that captures a sequence of IP hops along the forward path from the source to a given destination. As a part of the Macroscopic Topology Project, CAIDA (the Cooperative Association for Internet Data Analysis) provides a tool, skitter (http://www.caida.org/tools/measurement/skitter/), to collect the traceroute-based Internet topology measurements on a daily basis. CAIDA posts the adjacency matrix of the Internet AS-level graph computed daily from the skitter measurement (http://www.caida.org/tools/measurement/skitter/as_adjacencies.xml). The original skitter data can be obtained from CAIDA upon request for research purposes. This study uses a contiguous U.S. subset of the AS-level graph processed by Mahadevan et al (2005) on the basis of 31 days graphs for the month of March 2004. The processing includes filtering AS-sets (Rekhter 1994), multi-origin ASes(Mao et al. 2003), and private ASes (Hawkinson and Bates 1996a) from each daily graph, and discarding the indirect links, and then merging the daily graph to form one graph available to download from the following link: http://www.caida.org/analysis/topology/as_topo_comparisons. This graph, among others, is the one most closely reflecting the topology of actual Internet traffic flows(Mahadevan et al. 2005). Although not used in this study, BGP and the WHOIS database are important data sources commonly used to infer the AS-level Internet topology (Mahadevan et al. 2005). BGP is the protocol used for routing among ASes in the Internet (Rekhter 1994). The University of Oregon RouteViews project (http://www.routeviews.org) collects and archives both static snapshots of the BGP routing tables and updates from November 1997 to present. The WHOIS database (http://www.irr.net), which is a domain name registration database, has collected a wide range of information useful to network operators, including topological information (Mahadevan et al. 2005).

The AS-level Internet graph obtained does not have geographical coordinates. According to the Autonomous System Number (ASN) of each AS, geographical coordinates of most ASes can be found from the file of aslonglat.txt (http://www.caida.org/tools/utilities/netgeo/NGAPI/index.xml) or the NetGeo Server(http://netgeo.caida.org/perl/netgeo.cgi). The graph used in this study is a subset of the original graph to only cover the contiguous U.S., and has 3445 vertices and 8933 edges (Figure 4-15).

4.3.2 Relationship between local and global Internet connections at the AS-level

4.3.2.1 Local and global connections

As mentioned before, a classic method to identify the small-world network is to compare the APL and CC of a network to those of a same-size random network. The network can be identified as a small-world network if its APL is close to or smaller than that of the random network and its CC is much higher than that of the random network. In order to examine the AS-level Internet, a random network is simulated as the same size of the AS-level Internet graph, and the comparison results are shown in Table 4-3. The AS-level Internet graph has an APL of 2.849442, which is smaller than 5.081155 of the

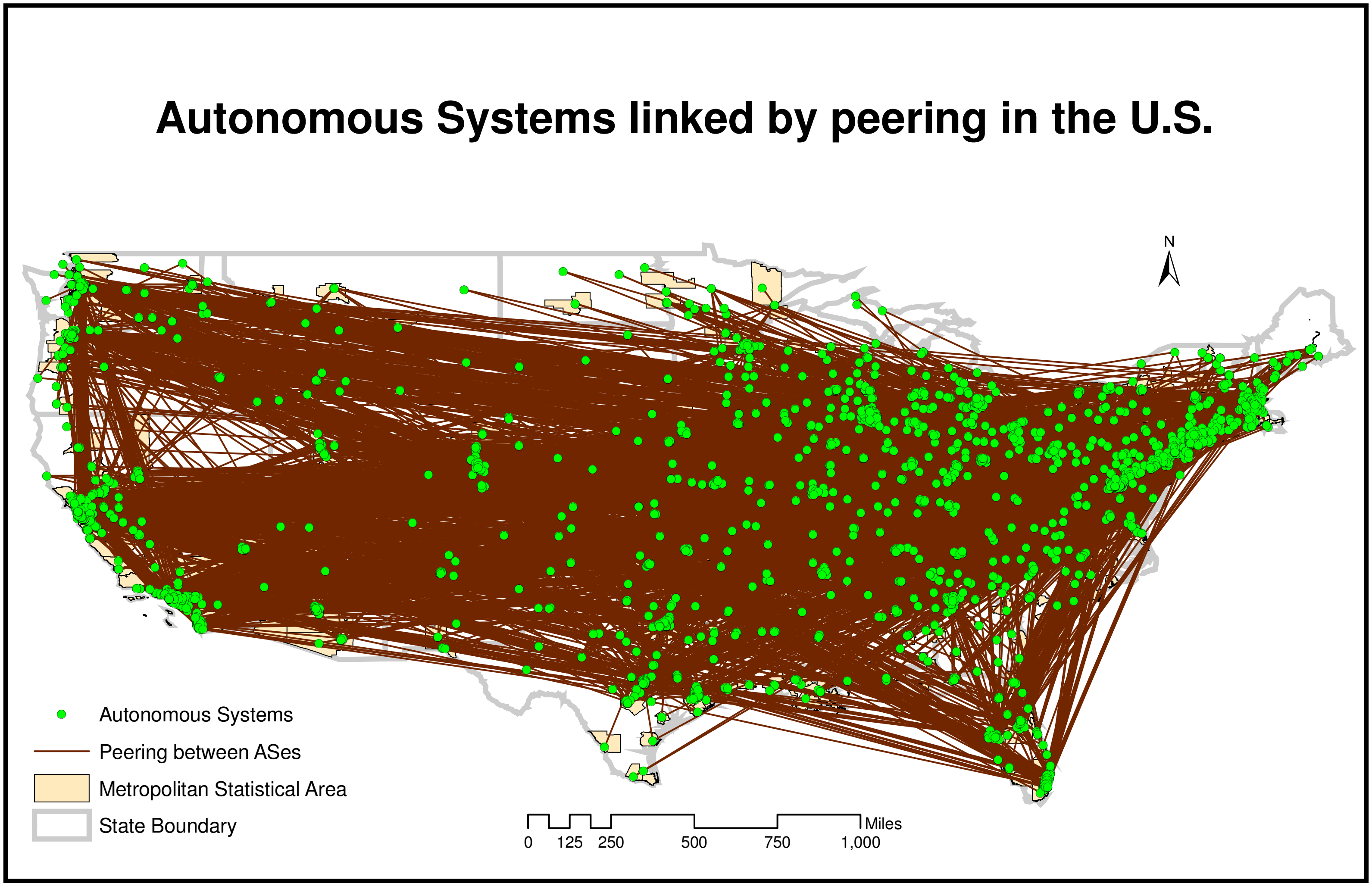


Figure 4-15: The spatial layout of AS-Level Internet graph for the contiguous U.S.

random network; it also has a CC of 0.514056 that is much larger than 0.001761 of the random network. Therefore, the AS-level Internet graph is a small-world network.

Table 4-3: Comparison between the AS-level Internet and a same-size random network in terms of Average Path Length (APL), Clustering Coefficient (CC), Local Connectivity Index (LCI), and Global Connectivity Index (GCI).

	APL	**CC**	**LCI**	**GCI**
AS-level Internet	2.849442	0.514056	0.499855	4.686212
Random network of the same size	5.081155	0.001761	0.163088	5.021474

As concluded by the simulation results in Chapter III, small-world networks should have higher LCI and lower GCI than those of a completely random network. The ASes can be geo-referenced according to their ASN and then overlapped with Metropolitan Statistical Area (MSA) (Figure 4-16) to decide the MSA each AS belongs in. The ASes that do not belong to any MSA will be assigned a special code. The connections between ASes can be classified as local connections and global connections. Local connections for an AS are those within the same MSA, and global links are those that reach out to different MSAs. Based on these definitions, the AS-level Internet graph has a LCI of 0.499855 that is higher than 0.163088 of the completely random network, and a GCI of 4.686212 that is lower than 5.021474 of the completely random network. Therefore, it has the properties of a small-world network. But the AS-level Internet graph has a high ratio between GCI and LCI of 10.7 (4.686212/0.499855=10.7), which implies that its global links are 10 times larger than local links. This is different from the topological concept of small-world networks, which states that a small-world network has a large number of local links and just a few global links connecting the local clusters. It will be further investigated through studying local and global degree distributions of the AS-level Internet graph.

4.3.2.2 Local and global degree distribution

The AS-level Internet graph has a skewed distribution or fat tail (Figure 4-17). A large number of ASes has a very low degree and a few ASes has a very high degree. A power function can be used to fit this distribution ($P(k) = 132.73k^{-0.9884}, R^2 = 0.6488$). A comparison of degree distributions between the AS-level Internet graph and a same-size random network is shown in the following figure (Figure 4-17). Compared with the random network, the AS-level Internet has a larger number of vertices of low overall degree and fewer vertices of high overall degree. However, the overall degree does not distinguish local or global connections.

The overall degree of a vertex can be classified into local degrees and global degrees according to whether the connections are local or global. A large number of vertices have zero local degrees in both networks (Figure 4-18), which implies that a large number of vertices only have global connections. The global degree distributions are quite the same as the overall degree distributions for both networks (Figure 4-18), which implies that the global degree distributions dominate the degree distributions of both networks. Although there is a significant difference in overall degree distributions

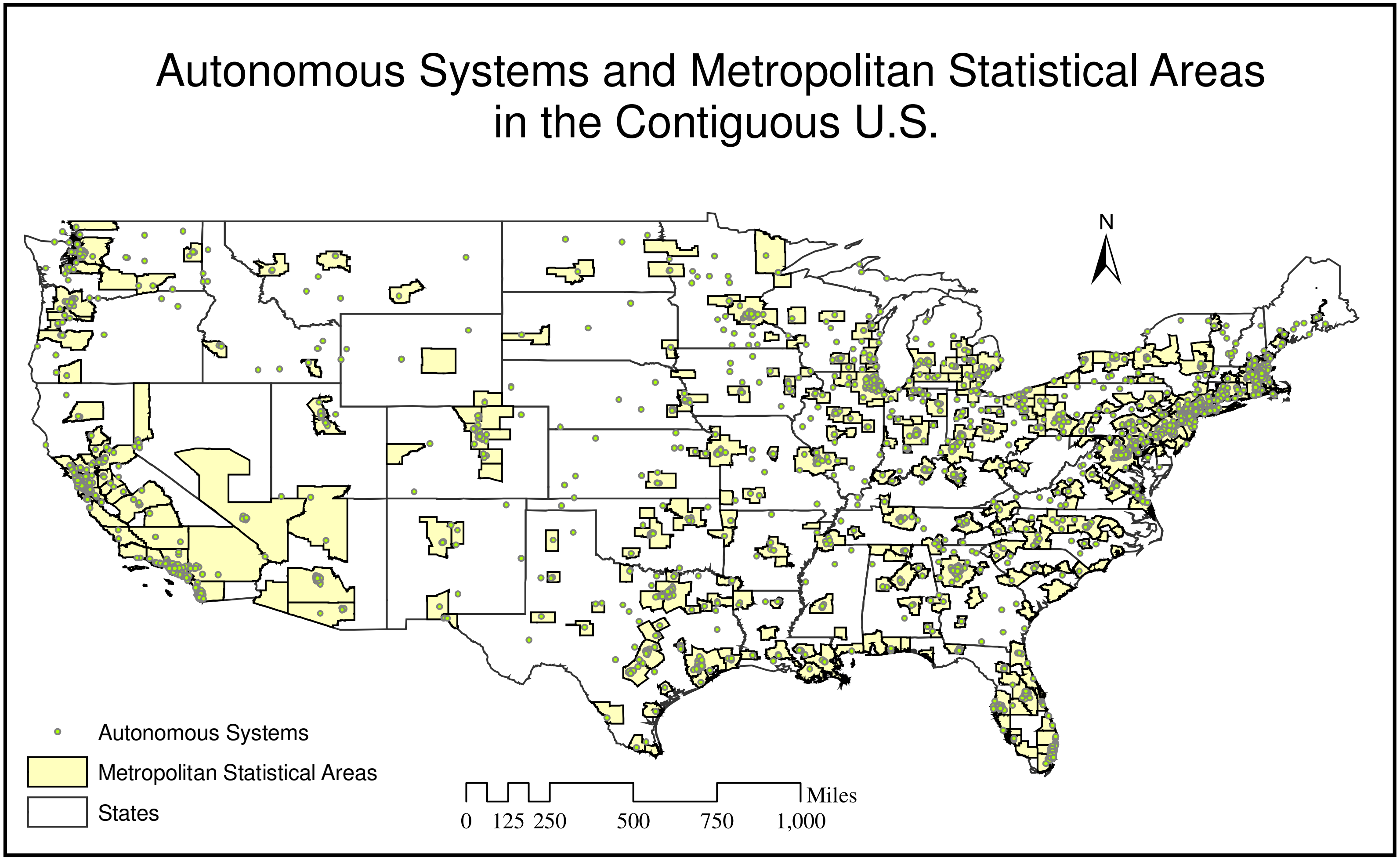


Figure 4-16: Geographic location map of autonomous systems (AS) and metropolitan statistical areas (MSA) in the contiguous U.S.

between those two networks (Figure 4-17), the AS-level Internet graph is close to the random network in two aspects (1) both of them have a significant large number of vertices that only have global connections; (2) both their global degree distributions dominate the overall degree distributions.

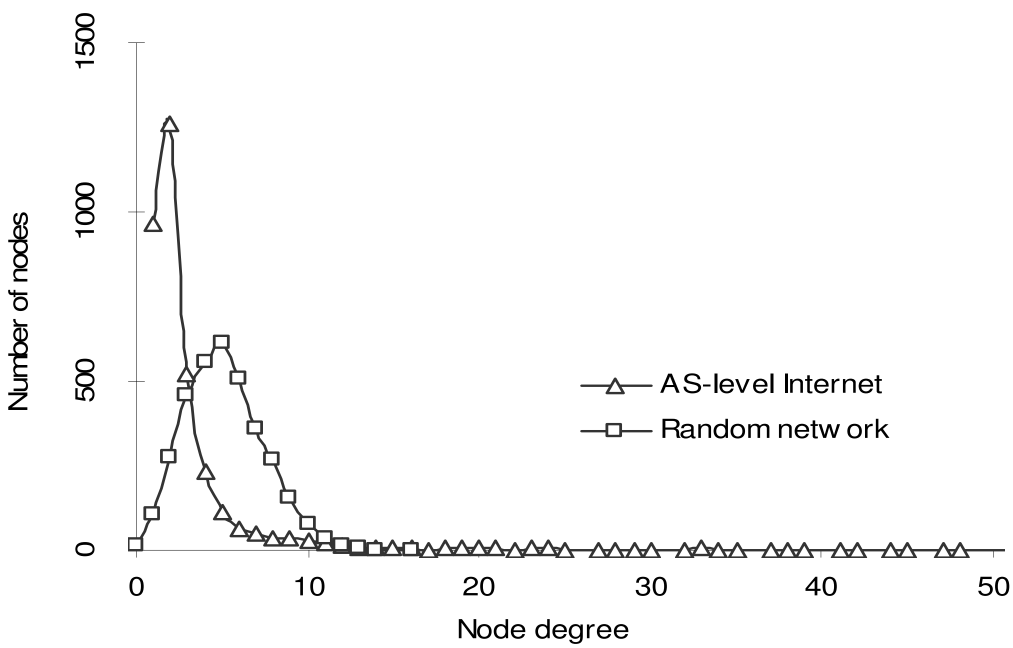


Figure 4-17: Comparison of overall degree distributions between the AS-level Internet graph and a same-size random graph.

Therefore, the AS-level Internet graph can be considered a small-world network, but has closer characteristics to a random network. It is a small-world network with more global connections.

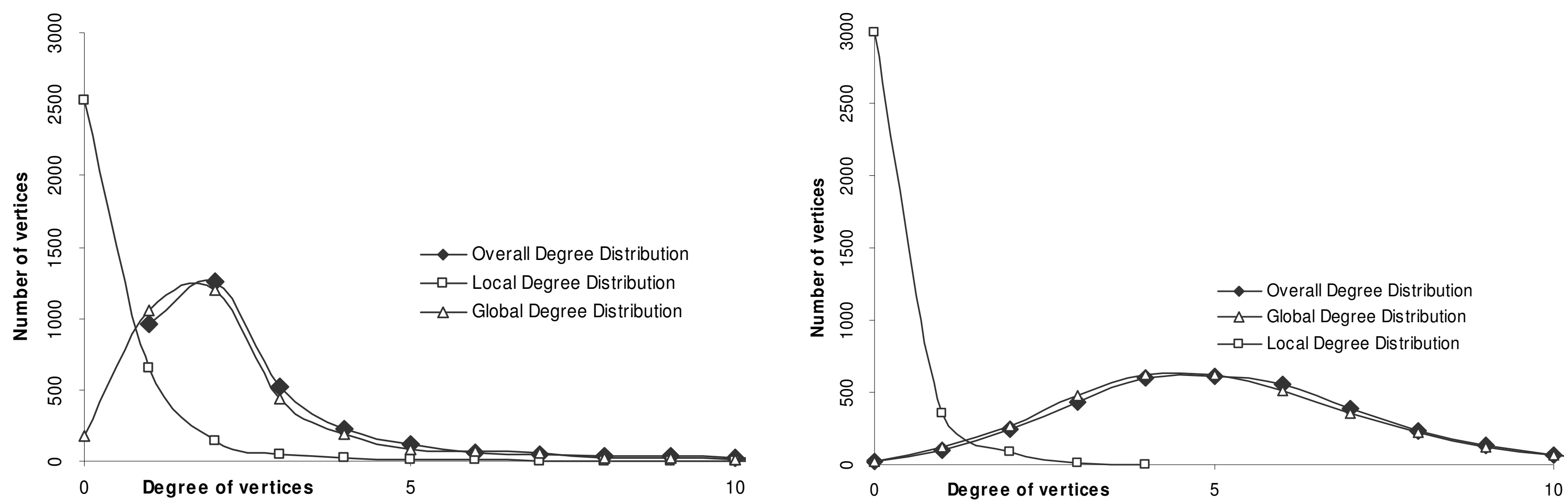
To explore the local and global connections, ASes with a degree of larger than 50, which means these ASes have more than 50 links to other ASes, are extracted from the AS-level Internet graph (Table 4-4). It shows that the global connection (GC) has dominated most of the high degree vertices.

Table 4-4: Local and global connections of ASes with degree larger than 50.

ASN	Degree	LC	GC	ASN	Degree	LC	GC
701	949	63	886	38	148	4	144
1239	738	35	703	7911	147	4	143
7018	701	5	696	2828	145	10	135
668	663	6	657	6347	113	10	103
3356	582	19	563	6395	113	6	107
209	536	40	496	11537	98	1	97
3549	305	14	291	2548	93	1	92
2914	298	3	295	2152	78	5	73
3561	246	0	246	6517	71	4	67
6461	241	22	219	5650	70	9	61
174	231	21	210	19029	63	1	62
7132	200	5	195	19262	59	32	27
4323	193	5	188	4200	58	1	57

4.3.3 Summary

The relationship between local and global connections is investigated for the Internet at the AS-level. This relationship, as discussed in Chapter III, shows that small-world networks have a high local connectivity index (LCI) and a low global connectivity index (GCI), which also correlates the relationship between local, global and overall degree distributions in small-world networks (i.e., the local degree distribution of the small-world network dominates its overall degree distribution and its global degree distribution has a fat tail). It reflects the underlying concept of small-world networks, i.e., the small-



(a) (b)

Figure 4-18: The global, local, and overall degree distributions of the AS-level Internet graph (a) and a random network of the same size (b).

world network has a large number of vertices with mostly local connections and a few global connections linking the local clusters.

The AS-level Internet graph has a higher clustering coefficient and a lower average path length than those of a same-size random network (Table 4-3). It therefore is a small-world network according to the definition of Watts and Strogatz (1998). However, it has a high GCI and low LCI, its global degree distribution dominates the overall degree distribution, and the local distribution has a fat tail, which are properties of random networks. Thus, the AS-level Internet graph possesses characteristics of a small-world network, but resembles more of a random network. It is a small-world network with more global connections.

CHAPTER V

SUMMARY AND CONCLUSIONS

5.1 Summary

The small-world phenomenon discovered by Milgram (1967) demonstrated the existence of a surprisingly short separation in social contact networks, which resonates well with the "it is a small world!" exclamation over the discovery of a common acquaintance between two strangers. The small-world network model initially developed by Watts and Strogatz (1998) formalized the dynamics of complex networks in a rigorous mathematical manner, indicating that the existence of small-world networks often lies within the intermediate range of a network rewiring process from a completely regular to a completely random network. Furthermore, small-world networks are often characterized by a topology with the high connectivity of random networks and high clustering of regular networks. The high connectivity of a small-world network changes very slowly with the size of networks. Empirical studies on a variety of networks show that the small-world characteristics are ubiquitous features for disperse and decentralized networks. However, the studies on small-world networks focused predominantly on aspatial and topology-dominated networks. Many real-world networks are spatial networks whose topologies are constrained by geographic embedding. For spatial networks, not only topology but also distance and geography are vital properties of the networks.

The dissertation, with an attempt to better understand the spatial dimension of the small-world networks, has contributed to the literature with new methods and a better understanding of small-world characteristics in spatial networks. The small-world characteristics in geographical, epidemic, and virtual spaces have been studied.

For geographical space, small world properties of transportation networks at national, metropolitan, and intra-city scales were investigated by a network autocorrelation approach. It is the first time in the literature of network science that the network autocorrelation statistics, Moran's *I* and General *G*, are used to detect features of a small-world network while a network is being rewired from a completely regular to a completely random network. The calculation of the network autocorrelation statistics is based on a direct measure of network connectivity for each and every vertex in the network, i.e., the average shortest path length of the vertex to all other vertices in the network. During the rewiring process for a simulated network, it is found that Moran's *I* and General *G* statistics tend to converge to similar values and have relatively low values when features of small-world networks start to emerge. This discovery is further investigated through calculating the network autocorrelation statistics of three transportation networks, i.e., the U.S. interstate highway network, the primary road network in Houston-Galveston area, and the Boston subway network. Since the network autocorrelation statistics are severely affected by the lag distances in the calculation, it is found that Moran's *I* and General *G* for the three networks are convergent and relatively low when the lag distances reach certain thresholds. The lag distance actually determines the size of neighborhood for a vertex, or how many vertices would be counted in the calculation. The change of lag distance in network autocorrelation calculations therefore implies the change of the logical connection of networks, and different lag distance implies different logical connection topology of the networks. On the assumption that dynamics taking place in a vertex will immediately affect all its neighboring vertices, the lag distance for the calculation of network autocorrelation statistics can be interpreted as the influencing distance of dynamics occurring in the vertex. When the influencing distance of dynamics taking place in the networks reaches certain thresholds, the small-world phenomenon starts to emerge in the networks. Obviously, the small-world phenomenon is often a result of the interplay between network dynamics and network structures, not simply the static network structure alone.

For epidemic space, the small-world effects on the epidemic diffusion and control strategies are investigated by simulating an epidemic process spreading in a series of networks rewired from a completely regular network by probabilities (0-1). The epidemic process is characterized by two properties, i.e., the maximum epidemic size and the time to reach it. The analysis of the characteristics of an epidemic process spreading in different networks found that the small-world characteristics - connectivity and clustering of networks - have significantly different effects on these epidemic properties. The connectivity of the networks has linear relations with both properties - a high connectivity corresponds to a large maximum epidemic size, and a short time to reach the maximum epidemic size. The maximum epidemic size and the time to maximum epidemic size can rapidly peak when the clustering coefficient is still very high. When the clustering coefficient is lower than 0.9, it does not affect the epidemics significantly. Since the small-world network is a highly connected (small average path length) and highly clustered (high clustering coefficient) network, epidemics taking place on the small-world networks therefore take a relatively shorter time to reach large epidemic sizes.

The effectiveness of four different control strategies - mass, targeted, traced, and acquaintance vaccinations - are also investigated. The effectiveness of control strategies is measured by the maximum infection rate. Comparison of different control strategies demonstrates that the effectiveness of control strategies is affected by the randomness and vaccination proportion of networks. The targeted control strategy is the most effective for a small-world network, while acquaintance vaccination is effective when it involves a high proportion vaccination.

For virtual space, Internet connection at the Autonomous Systems (AS) level is studied by analyzing the relationship between local and global connections and degree of distributions. The global connectivity index (GCI) and local connectivity index (LCI) are used to characterize the global and local connections in the network. The GCI, LCI, local and global degree distributions are used to monitor a network rewiring process where a regular network is rewired by probabilities (0-1). It is found that the manifestation of small-world networks is characterized by a very high LCI and a very low GCI, or low ratio of GCI/LCI, compared to those of random networks of the same size. It is consistent with the widely reported speculation that the topological structure of a small-world network should have a large number of local links and a few global links connecting local clusters together.

The local and global degree distribution between a small-world network and a same-size random network is compared. It is found that the local degree distribution of the small-world network dominates its overall degree distribution, and its global degree distribution has a fat tail, i.e., a large number of vertices have 0 or very low global connections and only a few vertices have high global connections. On the other hand, the global degree distribution of the random network dominates its overall degree distribution, and its local degree distribution has a fat tail, i.e., a large number of vertices have 0 or very low local connections and only a few vertices have high local connections.

Analysis of the local and global connections and degree distributions on the AS-level Internet connection demonstrates that the AS-level Internet has a high GCI and relatively low LCI, its global degree distribution dominates the overall degree distribution, and its local degree distribution has a fat tail, which are the properties of random networks. However, its APL closes to and its CC is much greater than those of the same-size random network, and the AS-level Internet is therefore a small-world network. Since Watts and Strogatz (1998) showed that a series of networks rewired by a region of probability between 0 and 1 within somewhere between completely regular and completely random network are small-world networks, some small-world networks will be closer to regular networks, while others are closer to random networks. The AS-Level Internet evident has certain features of a small-world network but is closer to be a random network.

5.2 Conclusions

The primary objective of this dissertation is to contribute to the literature new methods and better understandings of small-world characteristics in spatial networks by conducting a comparative study of the small world characteristics among geographical, epidemic, and virtual spaces. In this dissertation, a network rewiring process is monitored by Moran's *I*, General *G*, GCI, LCI, and local and global degree distribution to examine manifestations of the small-world phenomenon in spatial networks. The geographic underpinnings, the distance effect, and relation between local and global have been investigated regarding their correspondences with the emergence of small-world network features.

In conclusion, just as its acknowledged manifestation in network efficiency, the small-world phenomenon has different manifestations in spatial networks. In terms of the network autocorrelation, small-world networks tend to have converging and relatively low Moran's *I* and General *G* in their overall connectivity. The small-world phenomenon results from the interplay between dynamics taking place in networks and the network structure. An epidemic spreading in a small-world network will take relatively shorter time to reach its maximum size, and the targeted control strategy is the most effective in this case. For the relationship between local and global connections, small-world networks have a large number of local connections and a few global links. The local degree distribution dominates the overall degree distribution.

The small-world characteristics were initially discovered from the structural properties of the connection topology of networks. The connection topologies of the three networks studied are different and representative in their relevant spaces. The topology of the roadway transportation networks is explicitly constrained by geographical embedding. The social contact networks have a dynamic connection topology. The AS-level Internet has a more skewed distribution in connection topology. Despite these differences, the small-world characteristics are their common feature, but manifest differently as mentioned above. The small-world characteristics do have a special effect on the dynamics occurring in the network, such as the epidemics spreading in the social contact network.

The small-world network has been considered to have great potentials to model real-world network systems, including a large number of spatial networks. However, the original small-world network study focuses on the topology-dominated networks and ignores the spatial dimensions of networks. In attempt to better understand the small-world characteristics in spatial networks, this dissertation has extended the small-world network model with spatial considerations, and has potentially contributed to GIScience with new models for representing, analyzing, and modeling complexity from a network-centric perspective.

5.3 Remaining issues and future research

Networks are underlying structure for many systems. This study emphasizes networks in three spaces commonly encountered in geographic study. Although many studies exist on many different empirical networks, there should be more effort to study different empirical networks to find universal regularities among the increasingly complex networks.

This dissertation has focused the issue of spatial dimensions in the small-world networks. It is still a challenge for the small-world network to be embedded fully in spatial networks to model the dynamics (such as epidemics) occurring in spatial networks (such as an airline network). And since randomness exists in the topology of small-world networks, it remains a challenging task to use small-world networks to model the spatial diffusion of occurring along geographical networks.

The small-world network model has the potential to be the substrate on top of which the discrete simulation (such as agent-based models) can perform. Future study can be fruitful by linking the agent-based model with the small-world network to study the complex dynamics, such as epidemics and travel behavior in networks.

Modeling dynamic process is still a challenge to the current generation of geographic information systems. Future study will address this unsolved issue in GIScience from a network-centric perspective.

REFERENCES

Albert, R., and A.-L. Barabasi. 2002. Statistical mechanics of complex networks. *Review of Modern Physics* 74 (1):47-97.

Albert, R., H. Jeong, and A.-L. Barabasi. 1999. Diameter of the World-Wide Web. *Science* 401:130-131.

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

Amaral, L. A. N., A. Scala, M. Barthelemy, and H. E. Stanley. 2000. Classes of small-world networks. *Proceedings of the National Academy of Sciences* of *the USA* 97 (21):11149-11152.

Amin, V., Y. Ken, W. Kevin, M. Priya, K. Dejan, C. Jeff, and B. David. 2002. Scalability and accuracy in a large-scale network emulator. In *Proceedings of the 5th Symposium on Operating Systems Design and Implementation*. Boston, MA: ACM Press.

Anderson, R. M., and R. M. May. 1982. Directly transmitted infections diseases: control by vaccination. *Science* 215 (4536):1053-1060.

———. 1985. Vaccination and herd immunity to infectious diseases. *Nature* 318 (6044):323-329.

Anonymous. 2004. Virtual round table on ten leading questions for network research. *The European Physical Journal B* 38:143-145.

Anselin, L. 1995. Local indicators of spatial association - LISA. *Geographical Analysis* 27 (2):93-115.

Barabasi, A.-L. 2001. The physics of the Web. *Physics World* 14 (7):33-38.

———. 2002. *Linked: the new science of networks*. Cambridge, MA: Perseus Press.

Barabási, A., E. Ravasz, and Z. Oltvai. 2003. Hierarchical organization of modularity in complex networks. In *Statistical mechanics of complex networks*, eds. R. Pastor-Satorras, M. Rubi and A. Diaz-Guilera, 46-64. Berlin: Springer.

Barthelemy, M. 2003. Crossover from scale-free to spatial networks. *Europhysics Letters* 63 (6):915-921.

Batty, M. 2003. Network geography: relations, interactions, scaling and spatial processes in GIS. Working paper at the Center for Advanced Spatial Analysis, University College London.

Black, W. R. 1992. Network autocorrelation in transport network and flow systems. *Geographical Analysis* 24 (3):207-222.

———. 2001. An unpopular essay on transportation. *Journal of Transport Geography* 9:1-11.

———. 2003. *Transportation: A geographical analysis*. New York: The Guilford Press.

Black, W. R., and I. Thomas. 1998. Accidents on Belgium's motorways: a network autocorrelation analysis. *Journal of Transport Geography* 6 (1):23-31.

Boccaletti, S., V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang. 2006. Complex networks: Structure and dynamics. *Physics Reports* 424 (4-5):175-308.

Buchanan, M. 2002. *Small world - uncovering nature's hidden networks*. London, UK: Weidenfeld & Nicolson.

Calvert, K. L., M. B. Doar, and E. W. Zegura. 1997. Modeling internet topology. *Communications Magazine,IEEE* 35 (6):160-163.

Chang, H., S. Jamin, and W. Willinger. 2001. Inferring AS-level Internet topology from router-level path traces. In *Proceedings of SPIE ITCom 2001*, Denver,CO. 4526: 19-24

Chen, Q., Chang, Hyunseok, Govindan, R., Jamin, S. 2002. The origin of power laws in Internet topologies revisited. 2: 608-617 In *Proceedings of INFOCOM 2002*.

Cliff, A. D., and J. K. Ord. 1969. The problem of spatial autocorrelation. In *Studies in regional science*, ed. A.J.Scott, 25-55. London: Pion, Ltd.

———. 1973. *Spatial autocorrelation*. London: Pion.

Dorogovstev, S. N., and J. F. F. Mendez. 2003. *Evolution of networks: from biological nets to the Internet and WWW*: Oxford University Press.

Dorogovstsev, S. N., and J. F. F. Mendes. 2002. Evolution of networks. *Advances in Physics* 51 (4):1079-1187.

Dow, M. M., M. L. Burton, and D. R. White. 1982. Network autocorrelation: a simulation study of a foundational problem in regression and survey research. *Social Networks* 4:169-200.

Eames, K. T. D., and M. J. Keeling. 2003. Contact tracing and disease control. *Proceedings of the Royal Society B : Biological Sciences* 270 (1533):2565-2571.

Faloutsos, M., P. Faloutsos, and C. Faloutsos. 1999. On power-law relationships of the Internet topology. 251-262. In *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*. Cambridge, MA, United States: ACM Press.

Fang, W., and L. Peterson. 1999. Inter-AS traffic patterns and their implications. Paper read at Global Telecommunications Conference, 1999. GLOBECOM '99, at Rio de Janeireo, Brazil 3:1859-1968.

Ferguson, N. M., M. J. Keeling, W. J. Edmunds, R. Gani, B. T. Grenfell, R. M. Anderson, and S. Leach. 2003. Planning for smallpox outbreaks. *Nature* 425:681-685.

Flahaut, B., M. Mouchart, E. San Martin, and I. Thomas. 2003. The local spatial autocorrelation and the kernel method for identifying black zones: a comparative approach. *Accident Analysis and Prevention* 35:991-1004.

Gao, L. 2001. On inferring autonomous system relationships in the internet. *IEEE/ACM Trans. Netw.* 9 (6):733-745.

Gastner, M. T., and M. E. J. Newman. 2006a. Shape and efficiency in spatial distribution networks *Journal of Statistical Mechanics: Theory and Experiment* 2006 (1):1742-5468.

———. 2006b. The spatial structure of networks. *The European Physical Journal. B* 48 (22):247-252.

Getis, A., and J. K. Ord. 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24 (3):189-205.

Goodchild, M. F. 1987. *Spatial autocorrelation*, c*oncepts and techniques in modern geography No.48*. Norwich UK: Geo Abstracts Ltd.

Gorman, S. P., and R. Kulkarni. 2004. Spatial small world: new geographic patterns for an information economy. *Environment and Planning B* 31:273-296.

Govindan, R., and Tangmunarunkit, H. 2000. Heuristics for Internet map discovery. In INFOCOM 2000. IEEE, 26-30 Mar 2000, Tel Aviv, Israel. 3: 1371-1380.

Granovetter, M. 2003. Ignorance, knowledge, and outcomes in a small world. *Science* 301 (5634):773-774.

Haggett, P., and R. J. Chorley. 1969. *Network analysis in geography*. New York: St. Martin's Press.

Hawkinson, J., and T. Bates. 1996. *Guidelines for creation, selection, and registration of an Autonomous System (AS)*, Available http://www.rfc-editor.org/rfc/rfc1930.txt, Accessed 03-18-2007

Hayes, B. 2000. Graph theory in practice: part I. *American Scientist* 88:9-13.

Hethcote, H. W. 2000. The mathematics of infectious diseases. *SIAM Review* 42 (4):599-653.

Huberman, B. A. 2001. *The laws of the Web*. Cambridge, MA: MIT Press.

Huerta, R., and L. S. Tsimring. 2002. Contact tracing and epidemics control in social networks. *Physical Review E* 66:056115.

Jiang, B. 2004. Topological analysis of urban street networks. *Envronment and Planning B* 31:151-162.

Kaiser, M., and C. C. Hilgetag. 2004. Spatial growth of real-world networks. *Physical Review E* 69 (3):036103.

Kansky, K. 1963. *Structure of transportation networks*. Chicago, IL: University of Chicago Press.

Kaplan, E. H., D. L. Craft, and L. M. Wein. 2002. Emergency response to a smallpox attack : The case for mass vaccination. *Proceedings of the National Academy of Sciences of the USA* 99 (16):10935-10940.

Keeling, M. J. 2005. Models of foot-and-mouth disease. *Proceedings of the Royal Society B : Biological Sciences* 272:1195-1202.

Kiss, I., D. Green, and R. Kao. 2005. Disease contact tracing in random and clustered networks. *Proceedings of the Royal Society of London, Series B: Biological Sciences* 272 (1570):1407-1414.

Koopman, J. 2003. Modeling infection transmission. *Annual Review of Public Health* 25 (12):1-24.

Latora, V., and M. Marchiori. 2001. Efficient behavior of small-world networks. *Physical Review Letters* 87 (19):198701.

———. 2002. Is the Boston subway a small-world network? *Physica A* 314:109-113.

Lloyd, A. L., and May, Robert M.,. 2001. How viruses spread among computers and people. *Science* 292 (5520):1316-1317.

Madar, N., T. Kalisky, R. Cohen, D. ben-Avraham, and S. Havlin. 2004. Immunization and epidemic dynamics in complex networks. *The European Physics Journal B* 38:269-276.

Magoni, D., and J. J. Pansiot. 2001. Analysis of the autonomous system network topology. *SIGCOMM Comput. Commun. Rev.* 31 (3):26-37.

Mahadevan, P., D. Krioukov, M. Fomenkov, B. Huffaker, X. Dimitropoulos, K. Claffy, and A. Vahdat. 2005. Lessons from three views of the Internet topology: CAIDA technical report NO. CAIDA-TR-2005-02. Available http://www.caida.org/publications/papers/2005/tr-2005-02/. Accessed 03-18-2007

Mao, Z. M., J. Rexford, J. Wang, and R. H. Katz. 2003. Towards an accurate AS-level traceroute tool. *In Proceedings of ACM SIGCOMM*. 365-379.

Mathias, N., and V. Gopal. 2001. Small worlds: how and why. *Physical Review E* 63:02117.

Milgram, S. 1967. The small world problem. *Psychology Today* 1 (1):60-67.

Natalia, A. 2004. Transportation network topologies, Technical Memorandum, NASA/TM-2004-213259.

Newman, M. E. J. 2001a. Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E* 64 (1):016131.

———. 2001b. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E* 64 (1):016132.

———. 2001c. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the USA* 98:404-409.

———. 2002. Spread of epidemic disease on networks. *Physical Review E* 66 (1):016128.

———. 2003. The structure and function of complex networks. *SIAM Review* 45:167-256.

Newman, M. E. J., Jensen, I., and Ziff, R.M. 2002. Percolation and epidemics in a two-dimensional small world. *Physical Review E* 65:021904.

Newman, M. E. J., C. Moore, and D. J. Watts. 2000. Mean-field solution of small-world networks. *Physical Review Letters* 84:3201-3204.

Newman, M. E. J., and D. J. Watts. 1999a. Renormalization group analysis of the small-world network model. *Physics Letters A* 263:341-346.

———. 1999b. Scaling and percolation in the small-world network model. *Physical Review E* 60:7332-7342.

Odland, J. 1988. *Spaital Autocorrelation*. Ed. by G. I. Thrall. Vol. 9, Scientific Geography. Newbury Park,CA: SAGE Publications, Inc.

Ord, J. K., and A. Getis. 1995. Local spatial autocorrelation statistics: distributional issues and an application. *Geographical Analysis* 27 (4):286-306.

Pastor-Satorras, R., and A. Vespignani. 2001. Epidemic dynamics and endemic states in complex networks. *Physical Review E* 63:066117.

———. 2002. Immunization of complex networks. *Physical Review E* 65 (3):036104.

Peeters, D., J.-F. Thisse, and I. Thomas. 1998. Transportation networks and the locations of human activities. *Geographical Analysis* 30 (4):355-371.

Rekhter, Y., and Li, T. *A Border Gateway Protocol 4 (BGP-4)* Available http://tools.ietf.org/html?rfc=1654. Accessed: 03-18-2007

Sen, P., S. Dasgupta, A. Chatterjee, P. A. Sreeram, G. Mukherjee, and S. S. Manna. 2003. Small-world properties of the Indian railway network. *Physical Review E* 67 (3):036106.

Shirley, M. D. F., and S. P. Rushton. 2005. The impacts of network topology on disease spread. *Ecological Complexity* 2:287-299.

Siamwalla, R., R. Sharma, and S. Keshav. 1999. *Discovering Internet Topology* Available http://www.cs.cornell.edu/skeshav/papers/discovery.pdf. Accessed: 03-18-2007

Siganos, G., M. Faloutsos, P. Faloutsos, and C. Faloutsos. 2003. Power laws and the AS-level internet topology. *IEEE/ACM Trans. Netw.* 11 (4):514-524.

Stoneham, A. K. M. 1977. The small-world problem in a spatial context. *Environment and Planning A* 9:185-195.

Strogatz, S. H. 2001. Exploring complex networks. *Nature* 410 (6825):268-276.

Sui, D. Z. 2006. *Geography and the small world: challenges for GIScience*. Available http://www.geoplace.com/ME2/dirmod.asp?sid=119CFE3ACE2A48319AA7DE6A39B80D66&nm=News&type=Publishing&mod=Publications%3A%3AArticle&mid=8F3A7027421841978F18BE895F87F791&tier=4&id=2C8290C5803945EE8B9C48D0D3330D2B. Accessed: 1-20-2007.

Travers, J., and S. Milgram. 1969. An experimental study of the small world problem. *Sociometry* 32 (4): 425-443.

Tsimring, L. S., and R. Huerta. 2003. Modeling of contact tracing in social networks. *Physica* . 325: 33-39.

Vazquez, A., R. Pastor-Satorras, and A. Vespignani. 2002. *Internet topology at the router level and autonomous system level*. Available http://arxiv.org/abs/cond-mat/0206084. Accessed : 03-18-2007.

Wang, X., and Chen, Guanrong,. 2003. Complex network: small-world, scale-free and beyond. *IEEE Circuits and Systems Magazine* First quarter: 6-20.

Wasserman, S., and K. Fause. 1994. *Social network analysis: methods and applications*. Cambridge: Cambridge University Press.

Watts, D. J. 1999a. Networks, dynamics, and the small-world phenomenon. *The American Journal of Sociology* 105 (2):493-527.

———. 1999b. *Small worlds: the dynamics of networks between order and randomness*. Princeton, NJ: Princeton University Press.

———. 2003. *Six degree: the science of a connected age*. New York: W. W. Norton & Company.

———. 2004. The"new" science of networks. *Annual Review of Sociology* 30 (1):243-270.

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

Yook, S.-H., H. Jeong, and A.-L. Barabasi. 2002. Modeling the Internet's large-scale topology. *Proceedings of the National Academy of Sciences of the USA* 99 (21):13382-13386.

Zanette, D. H., and M. Kuperman. 2002. Effects of immunization in small-world epidemics. *Physica A* 309:445-452.

Zhang, B., R. Liu, M. Daniel, and L. Zhang. 2005. Collecting the internet AS-level topology. *SIGCOMM Comput. Commun. Rev.* 35 (1):53-61.

VITA

Name:	Zengwang Xu
Address:	Department of Geography, Texas A&M University 3147 TAMU, College Station, Texas 77843-3147
Email Address:	xuzw@hotmail.com; zengwang@gmail.com

Education:	B.E., Photogrammetry & Remote Sensing, Southwestern JiaoTong University, China, 1995 M.S., Cartography & GIS, Nanjing University, China, 1998 Ph.D., Geography, Texas A&M University, 2007