Delft University of Technology

Faculty of Electrical Engineering, Mathematics and Computer Science
Network Architectures and Services

Evolving Properties of Growing Networks

Jingwei Wu

1385364

Committee Members:

Supervisor:	Prof.dr.ir. P.F.A. Van Mieghem
Mentor:	ir. J. M. Hernández
Member:	Prof.dr.ir. P.F.A. Van Mieghem
	Dr.ir. F. A. Kuipers
	Dr. G. Hooghiemstra
	ir. J. M. Hernández

9th November, 2009

M.Sc. Thesis No: PVM 2009-060

Evolving Properties of Growing Networks

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE
in
ELECTRICAL ENGINEERING
by

Jingwei Wu
born in Shanghai, China

Network Architectures and Services Group
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology

Abstract

Complex networks describe a wide range of systems and structures in the world. Any real network can be modeled as graph, expressed by an adjacency matrix or list. In many complex networks, when a graph of a certain type grows in size, its properties are expected to change. Each complex network presents specific topological features which characterize its individual properties and are influenced by the dynamics of processes executed on the network. The analysis of complex networks therefore relies on the use of measurements capable of expressing the most relevant topological features. Therefore, understanding and analyzing the properties of different sized graphs is a challenging topic in the research field.

The objective of the thesis is to understand the evolving properties of growing networks. Therefore it focuses on comparison of topological metrics with different number of nodes and links. Growing graphs will be approached by two different schemes: preferential link attachment and random link attachment. Several common types of graph models are involved in the thesis. And we also consider different real-world network examples.

With the analysis and comparison of numerical simulation results, we want to understand the changing tendency of topological metrics for evolving networks. In final, the thesis reveals different crucial factors affecting the evolving properties of growing network and concludes evolving properties based on both empirical and analytical results.

Key words: complex network, growth, topological metrics, preferential attachment, random attachment, topological characteristics, evolution process

Acknowledgement

This project was carried out at the Network Architectures and Services Group of Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, the Netherlands during the period from December 2008 to October 2009.

First of all, I would like to acknowledge my supervisor, Prof. dr. ir. P. F. A. Van Mieghem. His pioneering research work provides the necessary basis for the thesis.

I would like to appreciate my mentor, ir J. M. Hernández, for his patient guidance and academic support.

Finally, I would like to thank my family for their love and support through my life.

List of Contents

1. Introduction 12
   1.1. Background 12
   1.2. Motivation 13
   1.3. Thesis outline 14
2. Graph Theory Review 15
   2.1. Network and graph theory 15
   2.2. Topological measures 16
      2.2.1. Node degree 16
      2.2.2. Average node degree 16
      2.2.3. Probability distribution of node degree 17
      2.2.4. Hopcount 17
      2.2.5. Diameter 17
      2.2.6. Clustering coefficient 17
      2.2.7. Giant component 17
      2.2.8. Assortativity coefficient 18
      2.2.9. Centrality measurements 19
3. Models and Construction 21
   3.1. Network models 21
      3.1.1. Random graph of Erdős-Rényi 21
      3.1.2. Scale-free graph of Barabási-Albert 22
      3.1.3. Small world model of Watts and Strogatz 24
      3.1.4. Regular lattice graph 24
      3.1.5. Special graph models: ring graph and line graph 26
   3.2. Data sets of real networks 26
   3.3. Graph evolution process 29
      3.3.1. Growth procedure 29
      3.3.2. Link attachment schemes 30
   3.4. Experiment setup 31
4. Evolving Properties of Growing Graph Models 33
   4.1. Classification and model settings 33
   4.2. Scale-free graph of Barabási-Albert 36
      4.2.1. Node degree 37
      4.2.2. Average hopcount 38
      4.2.3. Clustering coefficient 39
      4.2.4. Assortativity coefficient 40
      4.2.5. Betweenness centrality 41
   4.3. Random graph of Erdős-Rényi 43
      4.3.1. Node degree 43

4.3.2. Average hopcount 45
4.3.3. Clustering coefficient 46
4.3.4. Assortativity coefficient 47
4.3.5. Betweenness centrality 48
4.4. Regular lattice graph 49
4.4.1. Node degree 49
4.4.2. Average hopcount 50
4.4.3. Clustering coefficient 51
4.4.4. Assortativity coefficient 52
4.4.5. Betweenness centrality 53
4.5. Small world model of Watts and Strogatz 54
4.5.1. Node degree 55
4.5.2. Average hopcount 55
4.5.3. Clustering coefficient 56
4.5.4. Assortativity coefficient 57
4.5.5. Betweenness centrality 58
4.6. Special models: line graph and ring graph 59
4.6.1. Node degree 59
4.6.2. Average hopcount 60
4.6.3. Clustering coefficient 61
4.6.4. Assortativity coefficient 61
4.6.5. Betweenness centrality 62
5. Evolving Properties of Growing Real Networks 64
5.1. Classification and model settings 64
5.2. Social-biological network: LesMis 66
5.3. Technological network: ElectrC_s208 68
5.4. Linguistic network: WordAdj 70
6. Conclusion and Future Work 73
6.1. Conclusion 73
6.2. Future work 76

List of Figures

Figure 1.1 an example of complex network: Arpanet

Figure 3.1.1 A random graph of Erdős-Rényi contains 130 nodes and 215 links

Figure 3.1.2 A scale-free graph of Barabási-Albert contains 130 nodes and 215 links

Figure 3.1.3 A small world model of Watts and Strogatz with 64 nodes, k = 2 and rewiring probability p = 0.1

Figure 3.1.4 A two-dimensional regular lattice graph with 10*10 nodes

Figure 3.1.5 An example of a two-nearest-node ring graph with 16 nodes and an example of line graph with 5 nodes

Figure 4.1 Legend in different figures

Figure 4.2.1 (a) Left: Evolving node degree, in scale free graph (b) Right: Evolving node distribution, in scale free graph

Figure 4.2.2 Evolving hopcount in scale free graph

Figure 4.2.3 Evolving hopcount in scale free graph

Figure 4.2.4 Evolving assortativity coefficient in scale free graph

Figure 4.2.5 (a)Left: Evolving node betweenness in scale free graph, (b)Right: Evolving link betweeness, in scale free graph

Figure 4.2.5 (c) Evolving node betweenness distribution in scale free graph

Figure 4.3.1 (a) Left: Evolving node degree, in random graph (b) Right: Evolving node distribution, in random graph

Figure 4.3.2 Evolving hopcount in random graph

Figure 4.3.3 Evolving clustering coefficient in random graph

Figure 4.3.4 Evolving assortativity coefficient in random graph

Figure 4.3.5 (a) Left: Evolving node betweenness in random graph, (b)Right: Evolving link betweeness, in random graph

Figure 4.4.1 (a) Left: Evolving node degree, in lattice graph (b) Right: Evolving node

distribution, in lattice graph

Figure 4.4.2 Evolving hopcount in lattice graph

Figure 4.4.3 Evolving clustering coefficient in lattice graph

Figure 4.4.4 Evolving assortativity coefficient in lattice graph

Figure 4.4.5 (a) Left: Evolving node betweenness in lattice graph, (b)Right: Evolving link betweeness, in lattice graph

Figure 4.4.5 (c) Evolving node betweenness distribution in lattice graph

Figure 4.5.1 (a) Left: Evolving node degree, in small world model (b) Right: Evolving node distribution, in small world model

Figure 4.5.2 Evolving hopcount in small world model

Figure 4.5.3 Evolving assortativity coefficient in small world model

Figure 4.5.4 Evolving assortativity coefficient in small world model

Figure 4.5.5 (a) Left: Evolving node betweenness in small world model, (b)Right: Evolving link betweeness, in small world model

Figure 4.5.5 (c) Evolving node betweenness distribution in small world model

Figure 4.6.1 (a) Left: Evolving node degree, in line graph (b) Right: Evolving node distribution, in line graph

Figure 4.6.1 (c) Left: Evolving node degree, in ring graph (d) Right: Evolving node distribution, in ring graph

Figure 4.6.2 (a) Left: Evolving hopcount in line graph (b) Right: Evolving hopcount in ring graph

Figure 4.6.3 (a) Left: Evolving clustering coefficient in line graph (b) Right: Evolving clustering coefficient in ring graph

Figure 4.6.4 (a) Left: Evolving assortativity coefficient in line graph (b) Right: Evolving assortativity coefficient in ring graph

Figure 4.6.5 (a) Left: Evolving node betweenness in line graph (b) Right: Evolving node betweenness in ring graph

Figure 4.6.5 (c) Left: Evolving link betweenness in line graph (d) Right: Evolving link betweenness in ring graph

Figure 4.6.5 (e) Left: Evolving node betweenness distribution in line graph (f) Right:

Evolving node betweenness distribution in ring graph

Figure 5.2 (a) Left: Evolving node degree, in LesMis network (b) Right: Evolving node distribution, in LesMis network

Figure 5.2 (c) Left: Evolving average hopcount, in LesMis network (d) Right: Evolving clustering coefficient, in LesMis network

Figure 5.2 (e) Left: Evolving assortativity coefficient, in LesMis network (f) Right: Evolving node betweenness, in LesMis network

Figure 5.2 (g) Left: Evolving node degree distribution, in LesMis network (h) Right: Evolving link betweenness, in LesMis network

Figure 5.3 (a) Left: Evolving node degree, in ElectrC_s208 network (b) Right: Evolving node distribution, in ElectrC_s208 network

Figure 5.3 (c) Left: Evolving average hopcount, in ElectrC_s208 network (d) Right: Evolving clustering coefficient, in ElectrC_s208 network

Figure 5.3 (e) Left: Evolving assortativity coefficient, in ElectrC_s208 network (f) Right: Evolving node betweenness, in ElectrC_s208 network

Figure 5.3 (g) Left: Evolving node degree distribution, in ElectrC_s208 network (h) Right: Evolving link betweenness, in ElectrC_s208 network

Figure 5.4 (a) Left: Evolving node degree, in WordAdj network (b) Right: Evolving node distribution, in WordAdj network

Figure 5.4 (c) Left: Evolving average hopcount, in WordAdj network (d) Right: Evolving clustering coefficient, in WordAdj network

Figure 5.4 (e) Left: Evolving assortativity coefficient, in WordAdj network (f) Right: Evolving node betweenness, in WordAdj network

Figure 5.4 (g) Left: Evolving node degree distribution, in WordAdj network (h) Right: Evolving link betweenness, in WordAdj network

1

1. Introduction

1.1. Background

Many real systems and structures in our lives can be regarded as different types of complex networks. Nowadays our society relies on large networks more strongly than ever. People increasingly use wired and wireless networks to communicate with others in their daily lives. One's personal relationship with other people in society constitutes a social network. In other research areas such as neuroscience, a neural network describes a population of physically interconnected neuron or a group of disparate neurons. Figure 1.1 shows the distribution map of ARPANET which is a pioneered wide-area computer networking in 1980s.

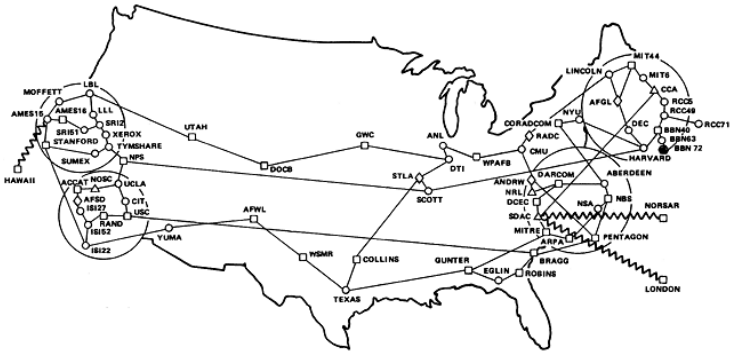


Figure 1.1 an example of complex network: ARPANET

Because of large size and the complexity of their interactions, the difficulty in representing a network partially lies in their topology and topological properties. Fortunately, graph theory has been well established and developed in the last several decades. A network is modeled as a graph described by an adjacency matrix or list. The advantages are their generality and flexibility for representing virtually or natural networks, including those undergoing dynamical change of topology [1]. The analysis

of complex networks therefore relies on the use of measurements capable of expressing the most relevant topological features. It helps us to research into topological features of a network by different types of topological measures. It gives us insight of networks and helps us to analyze the properties of them.

On the other hand, researchers are able to collect and process large data collection of networks since the computing ability increases dramatically in the last few years.

Therefore, contemporary complex network research consists of not only graph theory but statistical analysis as well. Regarding the converging development, lots of new measures and concepts have been put forward recently [2]. The research of the complex networks focus on not only the topological properties of stationary data sets of graphs but also their dynamical evolution.

1.2. Motivation

A network can be either stationary or dynamic. When a dynamic network of a certain class grows in size, its topological properties may even change [3]. We are interested in exploring how the properties change with the growth of graph and whether those changing tendencies are related to their types or other parameters.

In this thesis, we focus on dynamic evolution of network topology. This project presents a survey of different topological measures such as hopcount, clustering coefficient, degree distribution, assortativity coefficient and betweenness centrality. The statistical simulation results are derived from the simulation result of both graph models and data sets of real networks.

Thus, the objective of the present thesis is to better understand the evolving properties of dynamic networks when networks are growing in size during the evolution process. With the empirical data result from simulation, we analyze and compare the properties of metrics with different number of vertices and edges. The procedure of graph growth will be approached by growing graph with two different schemes: preferential link attachment and random link attachment. We repeat every evolution process for large number of cycles in each simulation to eliminate the influence of stochastic error. We also analyze the simulation results with the help of graph theory. We hope that it gives us insight on how topological metrics evolve with the number of vertices and edges.

1.3. Thesis outline

The first chapter addresses the background and the motivation of this thesis. The second chapter reviews the fundamentals of graph theory first. Then several topological metrics expressing the evolving properties of growing graph in the statistical simulation are proposed. The third chapter first introduces four typical types of graph models and two other special types. After that, we also briefly introduce the data sets of real networks used in the simulation. The last part of the chapter states the details on the construction of the models in the project.

The fourth chapter contains the statistical simulation result of six graph models. Analysis of evolving properties of graph is presented for each simulation. The fifth chapter shows the simulation results from the data sets of some real networks. The final chapter makes a conclusion and finally proposes the future work related to the thesis.

2

2. Graph Theory Review

2.1. Network and graph theory

In this section, we introduce fundamentals of graph theory. A network can be represented as a graph. A graph, G is usually defined as a set of edges E connected by a set of vertices V. The number of edges and the number of vertices in the network are denoted by N and L similarly. The notation of a graph therefore is denoted by $G(N,L)$. The numbers of N and L are usually assumed to be finite. A full mesh graph consists of N nodes where every node connects every other node directly and the total number of links is $L = N(N\text{-}1)/2$ which increases quadratically with the size of graph N.

If a link connects node i and j, those two nodes are called adjacent or neighboring nodes of graph G. A graph is connected if there exists at least one path between any pair of nodes in the graph and otherwise it is a disconnected graph.

The topological structure of a graph of N nodes can be represented by an adjacency matrix denoted by A. Adjacency matrix A is an $N*N$ real matrix. Each element a_{ij} of A is either one or zero depending on whether there exists a link between node i and j or not. Element a_{ij} is one if there is a link between node i and j, else $a_{ij} = 0$. The adjacency matrix A describes the interconnection pattern of a graph. Weights may be assigned to links in a graph which is called weighted graph. Links may also be given a direction and the graph can be directed. More than one link between a certain pair of nodes forms parallel links. There is no way of showing parallel links by the adjacency matrix A. When one link connects to the same node with its both ends, we have a self loop $a_{ii} = 0$.

In addition to adjacency matrix, adjacency list is an alternative representation which is similar to related data structure to represent a graph. In an adjacency list, the links sequences for each node are stored by representing the corresponding pair of

nodes. Adjacency list is high efficient in sparse graph, since it just contain links which really exist between pairs of nodes.

A simple graph is unweighted, undirected and any self loop or parallel link is prohibited. In this thesis, we restrict all the graphs to simple graphs unless they are stated otherwise and we also assume that all the graphs generated are connected. Therefore, the adjacency matrix of a simple graph is a diagonal matrix and elements on its main diagonal a_{ii} are equal to zero.

2.2. Topological measures

There are a large number of different topological metrics to characterize topological properties of a graph [4]. Some of the metrics are correlated to others to some extent, which means redundancy may lie between some metrics [5].

In this section, several key metrics are introduced. We use them to investigate and analyze the evolution of graph of different sizes and types in this thesis.

2.2.1. Node degree

The information of node degree is one of the most important characteristics of a graph. The degree of node *i* is the number of its direct neighboring nodes in the graph:

$$d_i = \sum_{j=1}^{N} a_{ij} \qquad (2.1)$$

It is obvious that node degree obeys the basic law below since each link builds up a connection between two nodes.

$$\sum_{j=1}^{N} d_j = 2L \qquad (2.2)$$

2.2.2. Average node degree

The average node degree of a graph evaluates the number of neighboring nodes a node has on average. It is defined as

$$E[D] = \frac{1}{N}\sum_{i=1}^{N} d_i = \frac{2L}{N} \qquad (2.3)$$

Especially, for a connected graph, the average node degree is lower bounded by

2-2/N when the graph contains no cycles and upper bounded by N-1 when the graph is full mesh.

2.2.3. Probability distribution of node degree

The probability distribution of node degree denoted by Pr[D=k] is the probability distribution function of node degree over the entire graph. It expresses the fraction of nodes in a graph with degree k Different types of graph are usually featured by different probability distribution of node degree which is a crucial metric in graph theory.

2.2.4. Hopcount

The shortest path is the shortest one among all existing paths between a pair of nodes. The hopcount between a pair of nodes denoted by H_{ij} and equals the number of hops in the shortest path which connects node i and j. The average hopcount of a graph denoted by *E[H]* and equals the average value of the hopcount of all shortest paths.

2.2.5. Diameter

Diameter is the maximum value among all shortest paths between all the pairs of nodes in the graph.

2.2.6. Clustering coefficient

The clustering coefficient $c_G(i)$ is defined as the ratio of the number of links y connecting the d_i neighbors of node i over the total potential d_i (d_i-1)/2 links.

$$c_G(i)=\frac{2y}{d_i(d_i-1)} \qquad (2.4)$$

The clustering coefficient characterizes the density of connections of node *i.* Another definition is the ratio between the number of triangles that includes node i and the number of triangles that would include node i if all neighboring nodes of i connect each other. It evaluates how nodes interconnect in the graph and expresses strength of local robustness. The average clustering coefficient C_G is the average value of clustering coefficients of all nodes in the graph. We use average clustering coefficient C_G in this thesis.

2.2.7. Giant component

A graph may be dismantled into isolated components or nodes if the link density

is too low. The largest connected cluster is called giant component in a disconnected graph. In a full connected graph, the giant component contains all the nodes through the whole graph. In this thesis all the graphs are connected unless stated otherwise.

2.2.8. Assortativity coefficient

Nodes have different types in some graphs. Correlations between nodes are found in the mixing pattern in many types of networks. For instance, in social networks, nodes with high degree tend to link high degree nodes as well. This phenomenon is called assortative mixing or assortativity. While in technological and biological networks, high degree nodes tend to connect low degree nodes, which show disassortative mixing, or dissortativity.

Unfortunately, the node degree distribution doesn't provide any details of the interconnection pattern between nodes in the graph. Assortativity coefficient r provides information whether a node attach to other nodes that are with similar type or not. There are several different ways to define and normalize the coefficient. In this thesis, we use the definition given by Newman [6]. The assortativity coefficient r is essentially the Pearson correlation coefficient between pairs of nodes.

Consider thea graph $G(N,L)$ with degree distribution $\Pr[D=k]$, where $\Pr[D=k]$ is the probability that a randomly chosen node in the graph has degree k.

Now if a node is reached by following a randomly chosen link in the graph, the degree of this chosen node is not distributed according to $\Pr[D=k]$. The reason is that more links ends at a node of high degree. As a result, the degree distribution for the node at the end of a randomly chosen link is in proportion to $k\Pr[D=k]$ instead of $\Pr[D=k]$. In this thesis, we define the remaining degree—the number of links leaving the node other than the one we arrived along. This number is one less than the total degree and hence is distributed in proportion to $(k + 1)\Pr[D=k+1]$.Thus, we define $\Pr[Q = k]$ as the remaining degree distribution,

$$\Pr[Q = k] = \frac{(k+1)\Pr[D = k+1]}{\sum_j j \Pr[D = j]} \qquad (2.5)$$

We define $\Pr[Q_x = k, Q_y = j]$ as the joint probability distribution of the remaining degrees of the two nodes x and y at either end of that randomly chosen link.

The assortativity coefficient is normalized to lie between -1 and 1 by dividing it by its variance

$$\sigma_q^2 = \sum_k (k^2 \Pr[Q_x = k]) - (\sum_k k \Pr[Q_x = k])^2 \qquad (2.6)$$

Then the assortativity coefficient r is defined as

$$r = \frac{1}{\sigma_q^2} \sum_{jk} jk(\Pr[Q_x = k, Q_y = j] - \Pr[Q_x = k]\Pr[Q_y = j]) \qquad (2.7)$$

If r is positive, it indicates a correlation between nodes of similar degree. If r is negative, it indicates correlation between nodes of different degree. Especially, when r = 1, the graph is said to have perfect assortative mixing patterns, while r = -1 the graph is totally disassortative. If r = 0, the graph is neither assortative nor disassortative. Assortative networks are known to percolate more easily while disassortative ones are more robust against node attack.

2.2.9. Centrality measurements

A node or a link plays an important role in a graph if a large number of shortest paths pass through it. The betweenness centrality of a node or a link is a measure to quantify its importance of a node or a link and defined as

$$B = \sum_{ij} \frac{\sigma(i, v, j)}{\sigma(i, j)} \qquad (2.8)$$

where $\sigma(i,j)$ is the number of shortest path between node i and j, and $\sigma(i,v,j)$ is the number of shortest paths connecting node i and j through node v or link v.

In the simulation of our work, betweenness just refers to numerator of B_u defined above. We denote node betweenness as B_n and denote link betweenness as B_l. The betweenness centrality estimates the potential traffic load on a selected node or link in a graph [7]. The average node betweenness and the link betweenness are denoted by $E[B_n]$ and $E[B_l]$ respectively in this thesis.

There exists a relation between average link betweenness $E[B_l]$ and average hopcount $E[H]$. If H_{ij} denotes the number of hops in the shortest path from node i to node k, the total number of hops in all shortest path in $G(N,L)$ is $\sum_{i=1}^{N} \sum_{k=i+1}^{N} H_{ik}$ which is equal to $\sum_{l=1}^{L} B_l$. Taking the expectation of them both gives the equation below:

$$E[B_l] = \frac{\binom{N}{2}}{L} E[H] \geq E[H] \qquad (2.9)$$

with equality only for the full mesh graph [14].

3

3. Models and Construction

3.1. Network models

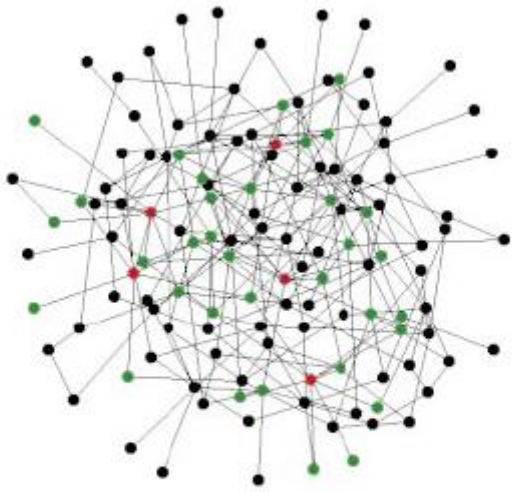
Many types of graph models are established in graph theory to model complex networks. Each type of graph model has its specific features. In this section, four most typical types of graph models are proposed. They are, random graph of Erdős-Rényi, the scale-free graph of Barabási-Albert, small world model of Watts and Strogatz and regular lattice graph. Besides those famous models above, two special graph models: ring graph and line graph are also introduced. We simulate and analysis these six types of models to explore the evolving properties of graph models in this thesis.

3.1.1. Random graph of Erdős-Rényi

The random graph model of Erdős-Rényi developed by Erdős and Rényi in the late 1940s separately is the first studied model of complex network [8]. Even today, it is still one of the most common models in the research of complex networks.

There are two most frequently occurring models. The first class of random graphs denoted by *Gr(N,L)* with *N* nodes and *L* links can be generated by starting with *N* disconnected nodes and zero links. Then the network is constructed by adding L links which are chosen randomly and independently from the total *N(N-1)*/2 potential links.

An alternative way to define the random graph model, denoted by $G_p(N)$, is to assume that any pair of nodes in the graph is connected with the probability *p*. Then the number of links in a random graph is not determinate but with the expectation $E[L] = pN(N-1)/2$. The average node degree of the graph is $E[D] = (N-1)\,p$.



Random Graph

Figure 3.1.1 A random graph of Erdős-Rényi contains 130 nodes and 215 links

The distribution of the degree D of an arbitrary node in the random graph is shown to be a binomial distribution:

$$\Pr[D=k]=\binom{N-1}{k}p^k(1-p)^{N-1-k} \qquad (3.1)$$

For large N and $Np = z$,the binomial distribution resembles the Poisson distribution with parameter z, see for the function details in section 4.3.1.

A random graph may be either connected or disconnected. A disconnected random graph consists of isolated clusters since the link density p is too low to form a connected graph. One notable aspect in random graph is that there exists a critical probability p_c at which the graph is almost connected. For large N, the critical threshold is $p_c \sim logN/N$ above which a giant cluster emerges which almost covers the entire graph.

3.1.2. Scale-free graph of Barabási-Albert

The degree distribution Pr[D=k] of scale-free graph decays as a power-law for large k , Pr[D=k] ~ $k^{-\gamma}$, where γ is the scaling exponent and is different for different types of graphs. Since Pr[D=k] is free of its characteristic scale for large N, graph with such degree distribution is called scale-free graph. An interesting feature in scale

free graph model is that small number of nodes is connected with a large number of nodes. Those nodes with significant node degree in the scale-free graph are called hubs through which a routine between two nodes usually passes. Therefore the average hopcount is relatively low in general.

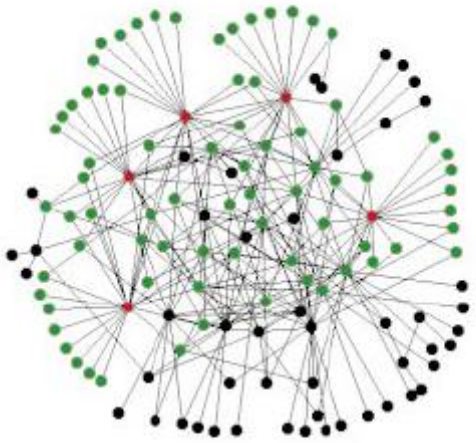


Figure 3.1.2 A scale-free graph of Barabási-Albert contains 130 nodes and 215 links

The Barabási–Albert model is a famous algorithm for generating scale-free networks by using growth and preferential attachment mechanism [9]. The model starts with a set of m_0 nodes. Each time one new node accompanied with m links is added to the present graph. Every new node links m different nodes in present graph to construct m new pairs of links.

The selection of m nodes in present graph obeys the linear preferential attachment rule, which means the probability of the new node i linking with an existing node j is proportional to the node degree of j,

$$P = \frac{d_j}{\sum_s d_s} \qquad (3.2)$$

where d_j is the degree of node i, $\sum_s d_s$ is the total degree in the network.

Obviously, the higher the node degree, the higher the probability is with which that node attach new nodes during the graph evolution with preferential attachment. This feature makes scale free network vulnerable to attack due to its inhomogeneous connectivity distribution [10].

3.1.3. Small world model of Watts and Strogatz

In a small world graph model, most nodes can be reached from other nodes using only a small number of links. Many real networks such as social network and telecommunication network exhibit the small world property. The famous six degrees of separation, discovered by Stanley Milgram [11] indicated the concept that everyone in the world is connected to everyone else by only six separations of acquaintances.

Watts and Strogatz model is the most popular model to construct on a ring random graph with small world property [12]. It starts with a lattice graph and each node connects to its k nearest neighbors. In addition, each link is also randomly rewired with probability p. The rewiring process adds $p\ N\ k\ /\ 2$ long-rang links to the graph object.

When $p = 1$, the graph is a regular ring lattice graph with lots of triangles and large distance. When $p = 0$, the graph becomes a total random graph, with small distance and few triangles. For a more normal case $0 < p < 1$, the generated graph lies in an intermediate status with both short distance and lots of triangles between those two extreme status above. The degree distribution of the Watts and Strogatz model are also influenced by the change of p.

For an intermediate value of p, the graph is a small world model which is highly clustered like regular lattice graph while it has small path length like a random graph.

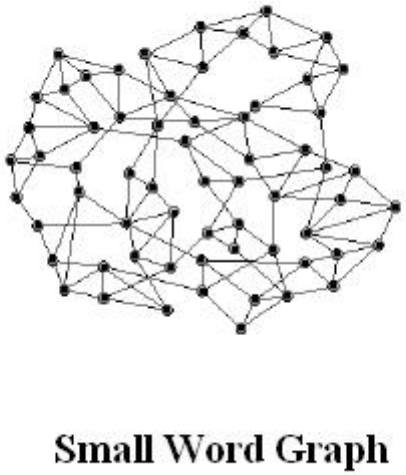
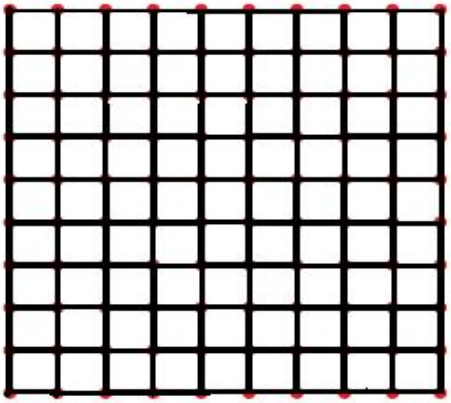


Figure 3.1.3 A small world model of Watts and Strogatz with 64 nodes, k = 2 and rewiring probability p = 0.1

3.1.4. Regular lattice graph

The regular lattice graph is a graph with N nodes on the regular grid structure. Only adjacent nodes are linked directly. Every node has equivalent distance to its

neighboring nodes.



Lattice Graph

Figure 3.1.4 A two-dimensional regular lattice graph with 10*10 nodes

If a two-dimensional lattice graph lies on a rectangular grid of size *mn*, the number of nodes in this lattice graph is $N=mn$. It is easy to verify the average degree:

$$E[D] = 4 - 2\frac{m+n}{mn} \qquad (3.3)$$

And the expected value of the hopcount is

$$E[H] = \frac{m+n}{3} \qquad (3.4)$$

In this thesis, we restrict regular lattice graph to two-dimensional full mesh lattice graph on a square grid of size n^2. The number of nodes in the graph is $N = n^2$ and the number of links in full mesh graph is $L=2n\ (n-1)$. For a dense lattice graph with $p = 1$, the expected mean degree is

$$E[D] = 4 - 2(n+n)/(n*n) = 4 - 4/n \qquad (3.5)$$

and the expected hopcount is

$$E[H] = (n+n)/3 = 2n/3 \qquad (3.6)$$

The hopcount increases polynomial with respect to the growth of graph size N.

3.1.5. Special graph models: ring graph and line graph

Besides the four typical types of graph models introduced in the previous sections, two special graph models, the ring graph and the line graph are also studied in this thesis to expand the scope of the graph models.

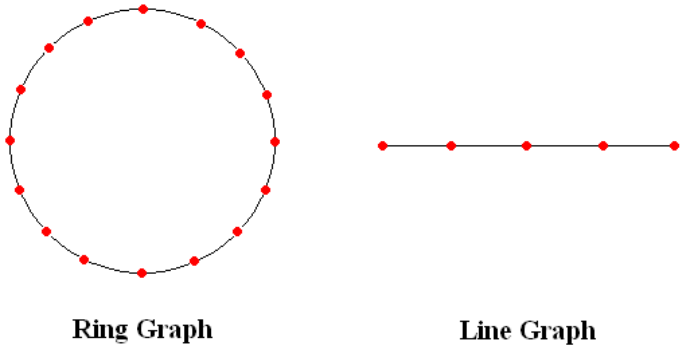


Figure 3.1.5 An example of a two-nearest-node ring graph with 16 nodes and an example of line graph with 5 nodes

◆ Ring graph

Ring graph can be regarded as one extreme state in construction of a small world graph of Watts and Strogatz with the link rewiring probability, p =0. A ring graph is called k-nearest ring graph in which each node is restricted to connect its nearest neighboring nodes in both direction. The topological pattern looks like a ring. It's easy to know that the number of links is equal to the number of nodes and the degree of each node is always two in a two-nearest ring graph.

◆ Line graph

Line graph is the other special model. It can be regarded as a graph in which a single link is cut off from a two-nearest ring graph. Therefore the number of links is one less than the number of nodes in a line graph.

3.2. Data sets of real networks

Besides different types of graph models introduced in section 3.1, we also use different public available sources of real networks in simulation. The type of the real

network has a wide range including social-biological networks, technological networks and Linguisticic networks. Short description of each real network is given in this section.

All the real networks have been preprocessed. We remove their parallel links, self loops, ignoring link weights, and ignore the directedness. One important consideration is that the number of nodes includes the disconnected components. In most cases the size of the largest connected component spans over the 90% of the nodes, but that is not always the case.

◆ America football network

This is the network of American football games between Division IA colleges during regular season fall 2000, as compiled by M. Girvan and M. Newman. Nodes represent different football teams in the association and links represent games between them in a season. In this network, there are 115 nodes and 613 links.

◆ ARPANET80 network

ARPANET is the short form of Advanced Research Projects Agency Network which pioneered wide-area computer networking and very much laid the foundations of the Internet, developing both the technical and social infrastructure of internetworking. ARPANET grew rapidly from construction of its first node in 1969. The ARPANET80 network plots the locations of the nodes on a base map of the outline of the United States in October 1980, around the middle of ARPANET's life. That network shows 71 nodes in the network and 86 connections between them. The nodes were the equivalent of today's routers. The most striking feature of this network is the large concentration of sites in California and the Northeast of the US, with a scattering of nodes in the interior. ARPANET had a few nodes outside of the continental USA. There was a node in Hawaii and two in Europe. Figure 1.1 in the first chapter is the distribution map of ARPANET in 1980s.

◆ C. elegans neural network

Caenorhabditis elegans is a free-living, transparent roundworm with around 1 mm in length, living in temperate soil environments. Research into the molecular and developmental biology of C. elegans was begun in 1974 and it has since been used extensively as a model organism. Celegansneural network is a biological network representing the neural network of C. Elegans with 297 nodes and 2148 links. That is pretty famous and commonly used real network example in network research.

◆ Electronic Circuits Networks

Three variants with different size of Electronic Circuits networks are selected from ISCAS89 Sequential Benchmark Circuits. The three variants of networks are denoted by ElectrC_s208 with 122 nodes and 189 links, ElectrC_s420 with 252 nodes and 399 links and ElectrC_s838 with 512 nodes and 819 links respectively. Each node represents a logical operation implemented physically. Links between them relate their inputs or outputs.

◆ Gnutella networks

Gnutella is a file sharing network. In recent years, it becomes one of the most popular file sharing networks on the Internet. A Gnutella crawler is a software program used to gather statistic information on the Gnutella file sharing network, such as the number of users, the market share of different clients and the geographical distribution of the user base. Four different crawls networks are available and denoted by Gnutella1 with 737 nodes and 803 links, Gnutella2 with 1568 nodes and 1906 links, Gnutella3 with 435 nodes and 459 links and Gnutella4 with 653 nodes and 738 links.

◆ Les Miserables network

Les Miserables network is a social network consisting of 77 co-appearance network of characters in the novel Les Miserables. The social network is composed of 77 nodes and 254 links

◆ Word adjacency network

The word adjacency network is an adjacency network of common adjectives and nouns in the novel David Copperfield by Charles Dickens. The network contains 112 nodes and 425 links.

◆ Western Europe Railway networks

The networks consist of railway infrastructure at two different level of one Western-European country. In the first data set the nodes in the network are service points such as railway stations and the links are the rails connecting them. There are 697 nodes and 785 links in 10 components among which the largest connected component contains 689 nodes. The other data set is the zoomed in level of the previous one. This level data contains 8730 nodes and 11350 links in 4 components. In our simulation, only the network of the first level is simulated.

◆ Western Europe Power Grid

The networks include power-grid infrastructure at three different levels of one

city-area in Western Europe. The largest network has 9168 nodes and 10147 links in 35 components. The second one includes 3690 nodes and 4206 links in 20 components. The third network contains 756 nodes and 786 links in 33 components. The largest component contains 296 nodes while the second largest component has 96 nodes. In our simulation, only the third network is simulated.

3.3. Graph evolution process

The evolution process of a growing graph is composed of two basic components: growth procedure and link attachment scheme. In this section, we introduce details of them used in the project.

3.3.1. Growth procedure

The objective of the thesis is to research and analyze the changing tendency of properties when different types of dynamic graphs grow in size continuously. The evolution process in the simulation consists of two basic components: growth and attachment.

A graph grows in size when new nodes are attached to the original graph. New links are inserted to connect new nodes and some existing nodes each time new node is generated. Since all the initial graphs are connected simple graphs, in each step of the construction during the evolution, the construction guarantees that the graph at any intermediate state is connected. In the simulation, we start with the initial graph $G\ (N, L)$ and set the target graph $G'\ (N', L')$, where $N' > N$, and $L' > L$. Nodes and links will be recursively added to the initial graph G in a certain fashion to increase the size of the graph. When N matches N', and L matches L', the simulation stops.

We repeat this evolution process with two different link attachment schemes for a large number of times in simulation. Let $\Delta N = N' - N$ and $\Delta L = L' - L$ be the number of new nodes and links attached to graph during the growth respectively. We define $\Delta N/N$ to be the increase ratio of graph size. Since L', L, N and N' can be arbitrary values, the number of links $m = \Delta L/\Delta N$ each time attached to a node newly generated may be a non-integer. We develop a mechanism to solve the problem during the evolution process in the simulation. First we calculate the great common divisor of ΔL and ΔN, denoted by $gcd\ (\Delta L, \Delta N)$. Then the whole evolution process of a graph is divided into $gcd(\Delta L, \Delta N)$ sub-procedures. Let b_n be the remainder when ΔL dividing ΔN, and denote $\Delta N / gcd(\Delta L, \Delta N) - b_n$ as a_n. In each sub-procedure, there are in total $b_n + a_n$ nodes newly attached to graph G and the first b_n nodes of them are generated with b_l links which are attached to b_l existing nodes and the rest a_n nodes connects to a_l existing nodes with a_l links which is one link less as an offset. The

weighted mean value $(a_l * a_n + b_l * b_n) / (a_n + b_n)$ is equal to $\Delta N / gcd(\Delta L, \Delta N)$.

The sequence of newly added nodes with a_l or b_l links can be distributed randomly instead of in a fixed order. But both of the mechanisms introduce some perturbation to the evolution process. In the thesis, we choose the fixed order mentioned above. In this growth scheme, at least one link should be generated when a new node is attached to the previous graph in the growth process. And therefore graph at any intermediate state is guaranteed to be connected.

During the growth procedure from initial graph G to target graph G', there are total $(\Delta N+1)$ states including initial state, target state and $\Delta N-1$ intermediate states. In each state, the intermediate graph is denoted by $G''(N'', L'')$. We calculate and record the topological metrics of $G''(N'', L'')$ in each state during the evolution in the simulation. We repeat the evolution process for hundreds of times at the same settings so that enough information is collected to evaluate general evolving properties of a growing graph and eliminate stochastic error.

3.3.2. Link attachment schemes

The other crucial component in the evolution process is the link attachment scheme. When a new node is generated during the evolution process, it is a key issue for that newly generated node to follow a certain rule to be attached to existing nodes in the present graph [13]. In this project, we propose two common link attachment schemes in the evolution process. They are preferential link attachment and random link attachment.

◆ **Preferential link attachment**

Preferential attachment means that a new node i is connected with an existing node j by inserting a new link a preferential attachment scheme. That preferential attachment rule can be related to any criteria. In our simulation, the probability of the new link attachment between node i and j is proportional to the degree of node j.

$$\Pr[a_{ij} = 1] = \frac{d_j}{\sum_n d_n} \qquad (3.7)$$

In general, node j with higher node degree in present graph has higher probability to link new generated node i, and vice versa, which displays 'the rich get richer' principle. Preferential attachment is also the attachment scheme to construct scale-free graph of Barabási-Albert, which possesses a power law degree distribution.

Let $n(u) \sim e^{-bu}$ denote the total number of nodes in graph at time u. The

preferential attachment establishes an exponential growth of a graph since it satisfies the condition $\frac{dn(u)}{du} = bn(u)$. It is proved that the exponential growth of a graph results in power-law node degree distribution in the target graph [14].

◆ **Random link attachment**

In the project, we also use random attachment as an alternative link attachment scheme. A new node *i* is connected randomly instead of preferentially to some of the existing nodes in a growing graph. In other words, a new node links a node from all nodes available with equal probability.

$$\Pr[a_{ij} = 1] = \frac{1}{N''} \qquad (3.8)$$

where *N"* refers to the total number of nodes in graph *G" (N", L")* before node are attached.

As is discussed before, when a new node is attached during the growth procedure, more than one link may be attached to a newly generated node. It depends on parameter setting. Within process of link attachment to one newly generated node, the former links connecting node does not influence the probability value in equation (3.7) until the intermediate state ends. Links are attached to the existing nodes with the same probability within one interim graph state. This regulation applies in both preferential and random attachment.

3.4. Experiment setup

During the graduation work, a C++ program was implemented to construct the experiment platform. In addition, the programming includes the Boost Graph Library which is a specific library appling in graph theory [15]. Codes are prepared for each type of graph model separately. In the simulation, we first set the essential parameter values including the number of nodes and links in the initial and the target graph object. Other parameters also need to be set depending on the type of initial graph. When the graph grows with size *N*, the evolution starts. In each interim state, value of each topological metric is recorded. Growth procedure of graph with each parameter setting is repeated for at least several hundreds of times to collect unbiased information.

After that, Matlab is used to process the original data result derived from the simulation. We plot the value of each measure in each state *G"(N",L")*. Since

simulation is repeated for hundreds of cycles, we calculate the average value and its standard deviation. As the graph grows in size, curve is formed by evolving value in each state.

The topological measures we selected to investigate their properties in the simulation are average clustering coefficient C_G, average hopcount $E[H]$, average assortativity coefficient r, average node degree $E[D]$, average node betweenness $E[B_n]$ and average link betweenness $E[B_l]$. In addition, node degree probability distribution $\Pr[D=k]$ and betweenness distribution are plotted in five states to give an insight of its evolving tendency. The five states consist of the initial and target state and three interim states by equivalent interpolation.

The model construction and algorithm constructed above is able to model graphs with different sizes. In theory, the number of nodes N and links L in initial graph object $G(N,L)$ can be quite large. In reality, due to the processing and memory capacity of devices, trade-offs are made between scale of simulated graphs and operation time. Based on the empirical results in simulation, we restrict both N and L to the scale of several hundreds. Therefore, the project focuses on graph of that scale.

4

4. Evolving Properties of Growing Graph Models

A large number of different graph models are established to model real networks in graph theory. Each type of model has its own characteristic topology and properties. In this chapter, we study into the evolving properties of growing graph models.

4.1. Classification and model settings

In our project, six types of graph models are applied to research evolving properties of graph models. The six different types contain four popular types of graph models and two special graph models. They are:

- Scale-free graph of Barabási-Albert
- Random graph of Erdős-Rényi
- Two-dimensional lattice graph
- Small world model of Watts and Strogatz
- Special graph models: ring graph and line graph

Five topological measures are calculated throughout the entire period of simulation. They are listed below:

- Average node degree: *E[D]*
- Average hopcount: *E[H]*
- Clustering coefficient of graph: C_G
- Assortativity coefficient: *r*

- Average node betweenness $E[B_n]$ and average link Betweenness $E[B_l]$

In addition, node degree probability distribution Pr[D=k] and node betweenness distribution in five states during the growth procedure are also given to offer more details of corresponding topological measures..

Two link attachment schemes introduced in Chapter 3 apply to the growth procedure:

- Preferential link attachment
- Random link attachment

The size of the graphs is restricted to the scale of hundreds of nodes because of the processing condition mentioned in chapter 3. We set the same number of nodes in initial and target states in all the models. Since the number of nodes in two-dimensional lattice graph is a square of an integer, set N = 225 in initial graph $G(N,L)$ and N' = 400 in target graphs $G'(N',L')$ respectively. As is introduced in Chapter 3, the construction and topology of graph varies with different types and parameter settings. It is not possible to guarantee that in the different graphs the number of L is kept equal. Instead, we try to keep the number of links between different graph models as close as possible to each other by adjusting related parameters in each model.

In the two-dimensional lattice graph, L is only dependent on N, since we assume that the rewiring probability is equal to one.

$$L = 2(N - \sqrt{N}) \qquad (4.1)$$

The number of nodes N is equal to 225, square of 15. Therefore, the number of links in initial lattice graph is equal to 420.

In Small world model of Watts and Strogatz, L is related to N and the k nearest neigbhors, given by

$$L = \frac{N \times k}{2} \qquad (4.2)$$

We assume construction of small world model lies between random graph and four-nearest-neighbor (k = 4) ring graph. Therefore, there are 450 links in initial graph of small world model. The link rewiring probability p is independent of L and is set to 0.09.

There are k links attached to a new node when a scale-free graph of Barabási-Albert is constructed. L in such scale-free graph results from N and the value k.

$$L \simeq kN \qquad (4.3)$$

There are 446 links in a scale-free graph with 225 nodes if two links are added each time a new node generated. So we set k to two in the simulation.

The situation in random graph of Erdős-Rényi is different from those. Since link connection probability is used to define the random graphs model, denoted by $G_p(N)$. The average number of links L $G_p(N)$ is:

$$E[L] = p\frac{N(N-1)}{2} \qquad (4.4)$$

We set the link connection probability p to 0.017. Therefore $E[L]$ = 428.40 and $E[D] = Np = 3.82$.

Last, regarding two special graph models, line graph and ring graph, there is no parameter available to adjust. Therefore L is fixed to 224 in line graph and 225 in ring graph.

Thus, L in four initial graph models ranges from 420 to 450. Node and link number settings are listed and divided by model types in Table 4.1.

Type of graph	L	L' (LPN =1)	L' (LPN =2)	L' (LPN =4)
Lattice	420	595	770	1120
Scale-free	446	621	796	1146
Small world	450	625	800	1150
Random	428	603	778	1128
Line graph	224	399	574	924
Ring graph	225	400	575	925

Table 4.1 Node and link number settings of graph models in simulation

Each time a new node is generated during the evolution process, we denote LPN as the number of links connecting the newly generated node with some existing nodes in the present graph. LPN plays a key role in the evolution process and influences the evolving performance of a dynamic graph.

We set LPN to different values and carry out simulations using two link attachment schemes to calculate topological measures of graph models. When LPN = 1, it is a special case that only one link connects a new node with present graph once a new node generated. For a more general case, LPN is an arbitrary integer and even a non-integer.

In this chapter, we set LPN to 1, 2 and 4 to analyze changing tendencies of those measures respectively. We also try studying the influence of LPN on the topological features of graph models. The target number of link L'' with different LPN settings in

each type of graph models is listed in Table 4.1.

With combinations of different parameter settings, the total number of simulations becomes large and there are several possible criterions to classify the simulation results. We analyze the simulation result of five topological measures in each graph. The results lie in a space of three-dimension: different types of graph models, different attachment schemes and different ratio *LPN*.

Therefore, we first classify the simulation results by model types. Then the simulation results within each typical model are further divided into two groups according to link attachment schemes. With in each group, we present the variation tendency of five topological measures under the condition of different *LPN* settings during the graph growth procedure.

Below we explain the legend of curves in different colors in figures in this chapter and chapter 5. Curves in red, green and blue in upper figure refer to the condition: *LPN* = 1, 2 and 4 respectively, in the evolution by preferential link attachment. Curves in cyan, magenta and yellow in lower figure refer to the condition: *LPN* = 1, 2 and 4 respectively in the evolution by random link attachment. An exception is the degree distribution and node betweenness distribution figures. In those figures, blue curves indicate initial distributions and red curves indicate target distributions.

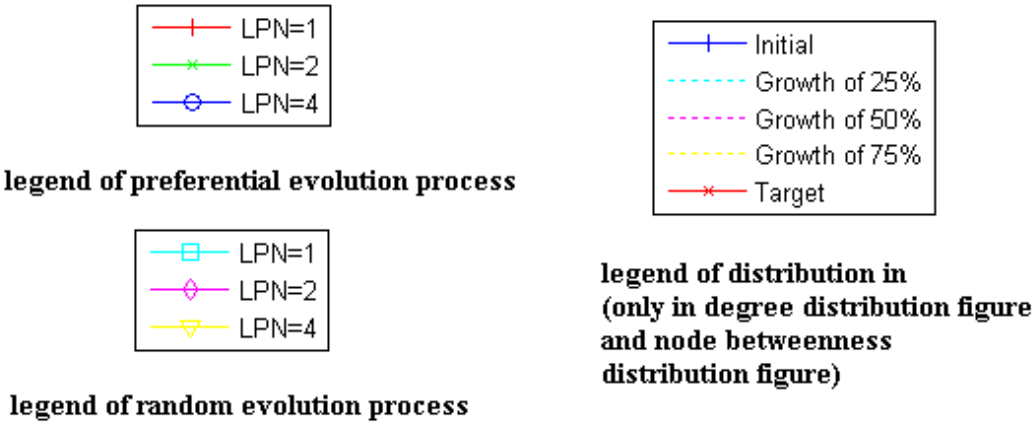


Figure 4.1 Legend in different figures

4.2. Scale-free graph of Barabási-Albert

In this section, we study the evolving properties of scale-free graphs of Barabási-Albert. During the construction of the scale-free graph of Barabási-Albert, each new node is attached to two existing nodes in the present graph. The initial scale-free graph *G(N,L)* grows from 225 nodes and 446 links to 400 nodes and the

corresponding number of links in target graph at different *LPN* listed in Table 4.1. In the simulation, each graph grows by both link attachment schemes.

4.2.1. Node degree

Figure 4.2.1 displays the changing tendency of average node degree *E[D]* in evolving graph object. Since *E[D] = 2L / N*, *E[D]* is independent of attachment scheme chosen in simulation. The blue curve in upper sub-figure of figure 4.2.1(a) is equal to the yellow curve in the lower sub-figure and so are the others. The changing tendency of *E[D]* by the preferential attachment is exactly the same as that growing by random attachment with the same *LPN* setting. Once the attachment scheme is given, changing tendency of *E[D]* is influenced by the *LPN* setting. *LPN* is constant during each growth procedure from *G(N,L)* to *G'(N',L')*. *N'* (and *L'*) in target graph *G'(N',L')* is composed of *N* (and *L*) in initial graph *G(N,L)* component and the number of new node (and new link) added during the growth procedure. Therefore, *E[D]* of graph at any intermediate regime is an interpolation changing with growing size.

$$E[D] = \frac{2(L+\Delta L)}{N+\Delta N} = 2\frac{(L+LPN\times\Delta N)}{N+\Delta N} \qquad (4.5)$$

In the simulation, initial scale free models start with 225 nodes and 446 links. We compare the simulation results in the condition of preferential attachment. The average node degree is 446*2 / 225 = 3.96 and all the three curves in upper figure of Figure 4.2.1(a) start at 3.96.

In the case of *LPN* = 1, the red curve declines to 3.11 monotonously. In the case of *LPN* = 2, the green curve steadies around 3.98. In the case of *LPN* = 4, the blue curve increases to 5.73 with higher absolute changing range than that in case of *LPN* = 1. The simulation results obey equation (4.5).

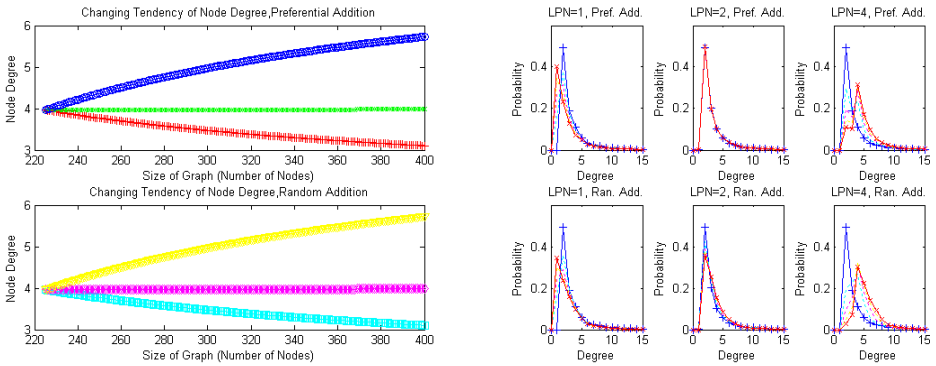


Figure 4.2.1 (a) Left: Evolving node degree, in scale free graph (b) Right: Evolving node distribution, in scale free graph

When the graph grows with a large increase, L (and N) in initial graph $G(N,L)$ is no longer a dominant component in the calculation of the average node degree by equation (4.5). Then $E[D]$ approaches the value double LPN as more nodes are added to present graph object during the evolution.

Figure 4.2.1 (b) plots changing tendency of degree distribution Pr[D=k] from initial state to target state. We first focus on the special case of LPN = 2 by preferential attachment. Simulation results show a feature that the probability distributions of target graph and all the intermediate graphs overlap with the initial one. In chapter 3, we define that every new node attached to two present nodes to construct a scale free graph in our project. The evolution process in this condition can be regarded as an extension of the construction of scale-free graph. The behavior in simulation result at LPN = 2 by preferential attachment demonstrates that the probability distribution of node degree in the scale-free graph is independent of its size.

The blue curve in Figure 4.2.1 shows Pr[D=2] peaks at 0.49. Nodes of two degree dominate in initial graph. We define every new node connects two present nodes under the construction of graph model. But new added nodes also change the connectivity of their adjacent node. The initial distribution follows a power-law tail for large k.

On the other hand, in the case of LPN = 2 by random attachment, Pr[D=k] is no longer distributed like a power law as N increases continuously. The random attachment scheme violates the initial distribution.

At a given LPN, LPN-connected nodes continuously added to graph object throughout the evolution. Therefore Pr[D=k] increases rapidly at k = LPN and LPN connected nodes tend to dominate in the graph as N increases.

4.2.2. Average hopcount

Curve of average hopcount $E[H]$ starts around 3.5. In Figure 4.2.2, $E[H]$ increases monotonously at LPN = 1 and 2 by both attachment schemes. On the other hand, at LPN = 4, $E[H]$ decays slightly by preferential attachment while $E[H]$ almost doesn't change by random attachment. In both attachment schemes, curve at higher LPN lies lower than those at lower LPN. High value of LPN increases the link the density, and therefore high LPN affects hopcount positively. Whether LPN is high enough to decreases $E[H]$ in a graph depends on the types of attachment and internal topological pattern of a graph.

At the same LPN, the $E[H]$ curve by preferential attachment is lower than that by random attachment. New generated nodes are more likely to attach highly connected

nodes in scale free graph if other condition remains the same. We know such nodes with high connectivity are called hubs which constitute the backbone of the graph structure. It usually builds up a path between two less connected nodes to save the length of hopcount. Regarding random attachment, since new nodes attached to existing nodes with equal probability, the average hopcount is higher in graph growing with same LPN.

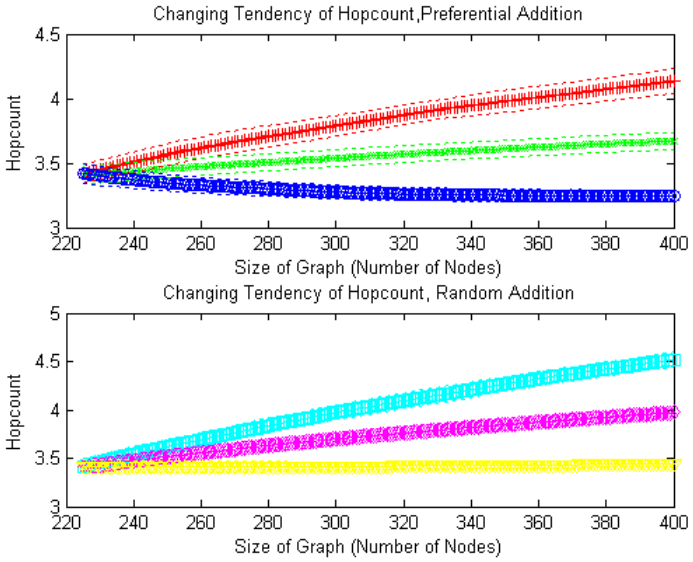


Figure 4.2.2 Evolving hopcount in scale free graph

4.2.3. Clustering coefficient

In figure 4.2.3, we observe that the average clustering coefficient C_G in the initial graph is low. We attach two links to new generated node in the construction of scale free graph. It makes the link density low, which results in low average clustering coefficient in initial graph. Low C_G reveals that neighboring nodes are seldom interconnected in graph object.

All the curves of clustering coefficient descend with growth of size N. It indicates evolution process with both attachment schemes deteriorate the performance of C_G. Only if a new node connects, present nodes which are already interlinked with others in the graph object, the newly introduced node increases C_G. Otherwise C_G decays as the graph grows. Since the number of nodes in graph is much larger than the maximum value of LPN in the evolution process, the previous situation seldom happens. Therefore, the evolving C_G decreases as N increases.

Lastly, we observe that the green curve decreases with least magnitude since we have illustrated that the growth procedure with $LPN = 2$ and preferential attachment is

extension of construction of scale free graph defined in this thesis. The evolving properties of C_G influenced least by the evolution process at $LPN = 2$ by preferential attachment.

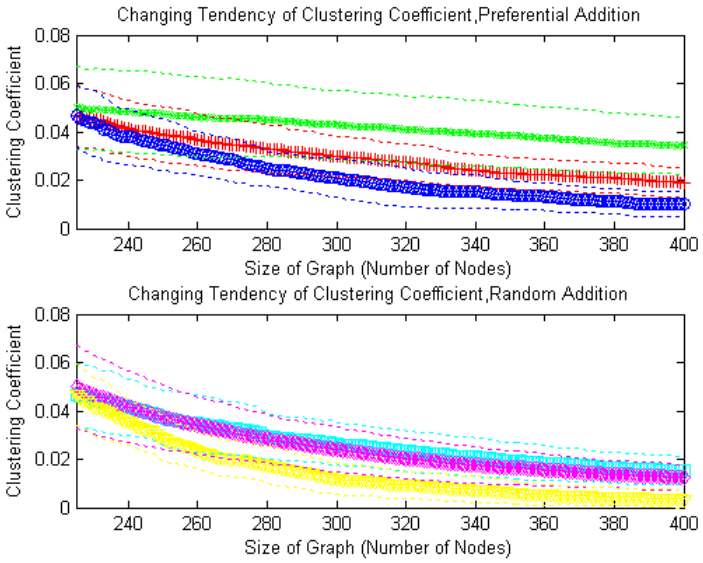


Figure 4.2.3 Evolving hopcount in scale free graph

4.2.4. Assortativity coefficient

The assortavity coefficient of initial scale free graph is around -0.15, which shows the initial scale free graph is disassortative.

A feature of scale free graph is that small number of nodes with high degree connects a large number of nodes of low degree nodes. That feature makes the scale free graph disasssortative mixing.

By preferential attachment, new nodes tend to connect to existing nodes with high degree. In Figure 4.2.4, character of disassortative mixing in graph is weakened slightly. All the target graphs are still disassortative.

By contrast, when the graph grows by random link attachment, disassortativity gets more weakened with increasing N and ends around -0.05. Random link attachment between newly generated nodes and existing nodes confuses the interconnection pattern in the initial graph. Therefore it weakens the disassortative mixing character more heavily.

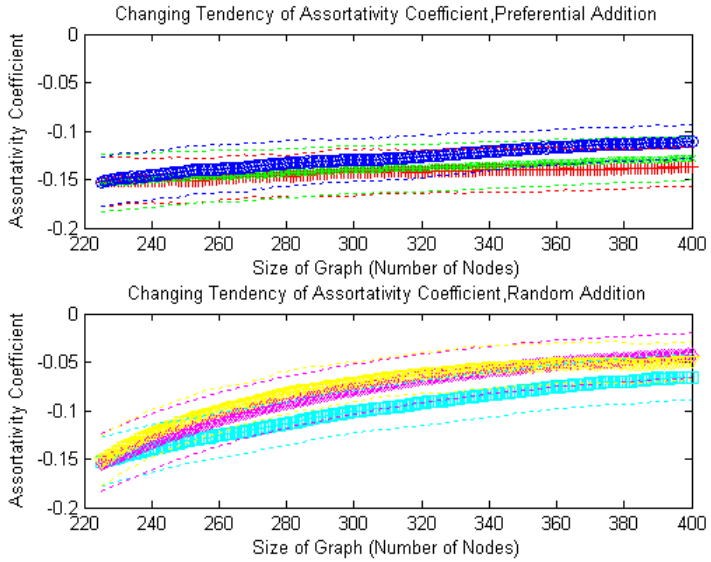


Figure 4.2.4 Evolving assortativity coefficient in scale free graph

4.2.5. Betweenness centrality

In this section, we discuss the evolving properties of node betweenness and link betweenness centrality in scale free graph object.

In chapter 2, we claim betweenness in simulation is defined as the total number of shortest paths instead of ratio. We first present the general performance of evolving betweenness and then discuss the distribution of node betweenness as we deal with node degree in section 4.2.1.

Figure 4.2.5(a) and (b) exhibit the general performance of evolving betweenness by calculating the average betweenness of node and link.

The average betweenness generally stays at low level throughout the entire evolution compared to the situation of other types of graph. The curves of node betweenness B_n start around 250 and link betweenness B_l start round 200. In all simulations, both B_n and B_l increases linearly. Since the number of nodes increases as well, we can not conclude that the evolving betweenness increases before the results of different size graphs are normalized in a proper way.

Therefore, in this section we also plot the node betweenness distribution to analyze how betweeness centrality evolves. Since the evolving property of node betweenness is quite similar to that of link betweeness not only by definition but also from the simulation result, we only analyze the distribution of node betweenness further. We calculate the node betweenness value of each node in graph in five states

during the whole evolution period. And then we sort those betweenness values in ascending order. The average value of node betweenness distribution in blue curve in Figure 4.2.5(c) is equal to the value of the starting point plotted in Figure 4.2.5(a) and (b). Obviously, Figure 4.2.5(c) contains more information.

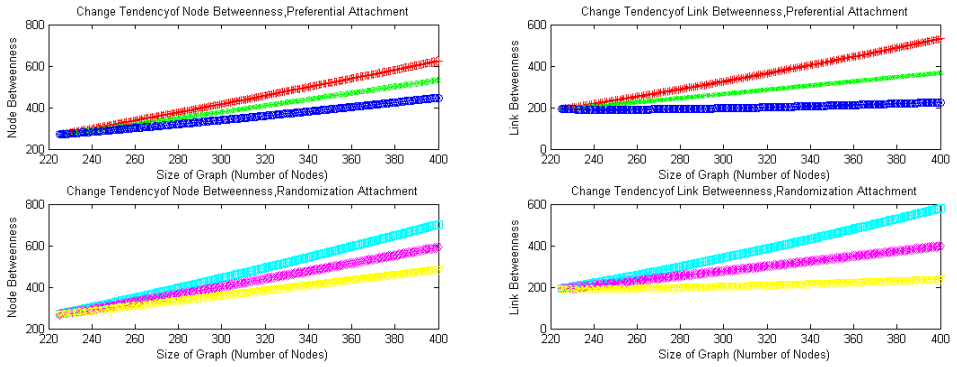


Figure 4.2.5 (a)Left: Evolving node betweenness in scale free graph, (b)Right: Evolving link betweeness, in scale free graph

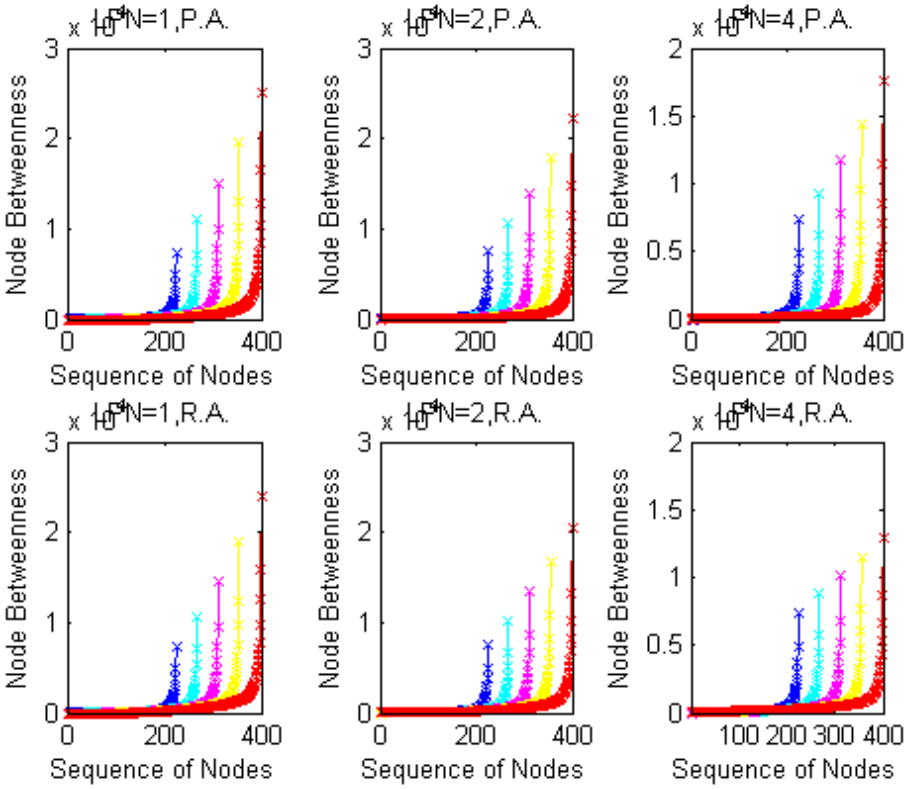


Figure 4.2.5 (c) Evolving node betweenness distribution in scale free graph

Figure 4.2.5 (c) shows that betweenness value of most nodes is low. By contrast, we observe that the tail of curve ascends sharply in each figure. It exhibits the phenomenon that a small number of nodes have high B_n in scale free graph.

Considering the topology of scale free graph, those nodes are highly connected and regarded as hubs through which large number of lower connected nodes create shortest path.

Then we deduce that most of nodes in the scale free model are of low degree in Figure 4.2.1(b) and of low betweenness in Figure 4.2.5(c).

Here we compare the evolving tendency of betweenness distribution at influence of different link attachment schemes. By preferential evolution, the maximum values of the curves are higher than those by random evolution. For instance, the maximum value of B_n in target graph of in upper-middle subfigure of Figure 4.2.5(c) is around 1800 while in lower-middle subfigure the maximum is only around 13000. We find that the preferential attachment enhances the importance of highly connected nodes in the field of traffic load during the evolution process.

4.3. Random graph of Erdős-Rényi

In this section, we investigate and analyze evolving properties of Erdős-Rényi random graphs.

We implement class $G_p(N)$ to define the random graph by setting the number of nodes N and link connection probability p. We set p to 0.017 so that the expectation of number of links $E[L]$ in the initial graph is equal to 428. The link connection probability p we set is just a little bit larger than the critical threshold p_c above which a giant cluster emerges which almost covers the entire graph. It can not guarantee every generated random graph is fully connected. In addition, we check full connectivity of initial random graphs once it is generated. Graph is regenerated if the initial graph is not full connected.

4.3.1. Node degree

Since the average node degree is independent of graph types and is only related to L and N, the evolving properties of average node degree in random graph is similar to that in scale free graph discussed in section 4.2.

In addition, the average node degree in random graph can be rewritten in terms of link connection probability p.

$$E[D] = \frac{2L}{N} = (N-1)p \qquad (4.6)$$

In this section, we mainly analyze the degree distribution *E[D]*.

The distribution of the degree *D* of an arbitrary node in the random graph $G_p(N)$ follows a binomial distribution. Let the binomial distribution denoted by S_N.

$$\Pr[D=k]=\binom{N-1}{k}p^k(1-p)^{N-1-k}=S_N \qquad (4.7)$$

For large *N*,

i) if *p*N* is a constant, say *z*, then S_N tends to be a Poisson distribution peaked at *z* and decaying exponentially.

$$S_N(k)\simeq\frac{z^k e^{-z}}{k!} \qquad (4.8)$$

ii) if *p* is a constant and independent of *N*, then the central limit theorem states that:

$$\frac{S_N-Np}{\sqrt{Np(1-p)}}\simeq\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \qquad (4.9)$$

tends to be a Gaussian distribution for large *N*.

In summary, the limit distribution of S_N depends on how *p* varies with increasing *N*.

In Figure 4.3.1(b), degree distribution Pr[*D*=*k*] in the initial random graph is plotted in blue curve. We know the distribution follows a binomial distribution. All six curves of Pr[*D*=*k*] largely vary in the evolution process. In target state, none of them present a binomial distribution any more.

In the condition of the same *LPN*, the tail of curve with the preferential evolution drops more rapidly than that in random evolution. Preferential evolution consists of growth and preferential link attachment. They are also two basic ingredients in the construction of scale free graph. The target graph has compound features of both random graph and scale free graph. Regarding degree distribution, we observe all the three curves in red look like power-law tails in upper sub-figure in Figure 4.3.1(b).

The difference between preferential and random evolution is different link attachment schemes during the growth procedure. The random evolution also includes the ingredient of growth. Therefore, the random evolution is believed to present some

properties of scale free graph partially. We find in the lower sub-figure, the degree distribution in target graph is neither a binomial nor a Poisson distribution. The three curves in red all have low tails which lie higher than those with preferential evolution in the upper sub-graph.

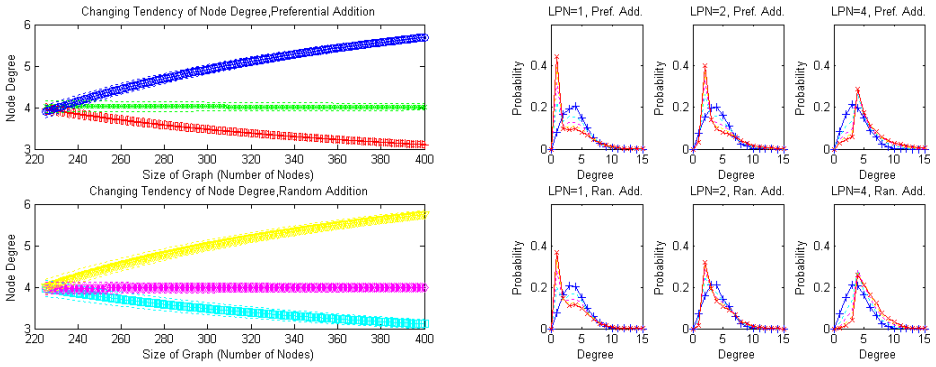


Figure 4.3.1 (a) Left: Evolving node degree, in random graph (b) Right: Evolving node distribution, in random graph

4.3.2. Average hopcount

Average hopcount *E[H]* in the classical random graph is expected to be larger than that in scale free graph due to different topological interconnections. There is no such node with high connectivity in random graph. Instead, every node links other nodes with equal link connection probability in random graph. Once the initial graph grows by preferential link attachment, newly generated nodes prefer to link high connected nodes in the graph object after the evolution begins. It alters the original degree distribution and creates some shortcuts in the graph as soon as graph evolution begins.

By definition, each node is connected to *E[D]* nodes on average in random graph. After h hops, $E[D]^h$ can be reached. And all nodes are reached when $E[D]^h \simeq N$, which implies that the average hopcount *E[H]* in random graph is approximately:

$$E[H] \simeq \frac{\log(N)}{\log(E[D])} \qquad (4.10)$$

Hofstad, Hooghiemstra and Van Mieghem [16], Hooghiemstra and Van Mieghem [17] provide a better approximation which is beyond the scope of the thesis. Thus, we use (4.10) for its simplicity and get the average hopcount $E[H] = 4.05$ by calculation. The growth of average hopcount is expected to be logarithmic in N in random graph.

Based on the simulation result, the curve of average hopcount *E[H]* in random graph with 225 nodes and 428 links starts at 4.09 , which is very close to the analytic result 4.05 according to approximation equation (4.10). *E[H]* in random graph is higher than that in scale free graph with 225 nodes and 446 links. For *LPN* = 1 and 2, *E[H]* increases with a high magnitude in random evolution than that in preferential ones. The reason has been stated in section 4.2. For *LPN* = 4 by preferential attachment, *E[H]* decreases at high decreasing rate at the beginning. Then the absolute value of derivative of it decays as *N* increases.

In contrast to the evolving behavior in random graph, there are some highly connected nodes in the scale free graph and its *E[H]* is lower because of the difference in topological pattern. We find no such phenomenon occurs in Figure 4.2.2.

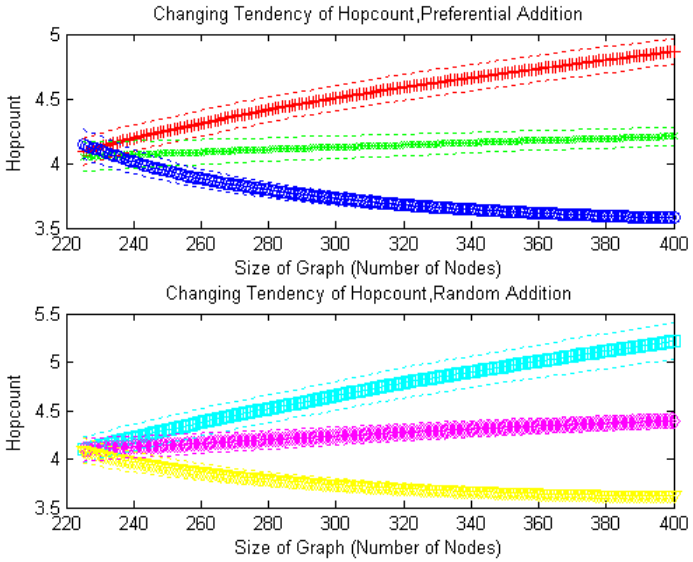


Figure 4.3.2 Evolving hopcount in random graph

4.3.3. Clustering coefficient

If we consider a node and its adjacent neighbors in a random graph, the probability that two of these adjacent neighbors are connected is equal to the probability that two randomly selected nodes are connected. As a consequence, the average clustering coefficient is

$$C_{G\_rand} = \frac{E[D]}{N} \simeq p \qquad (4.11)$$

From the empirical results in Figure 4.3.3, the average clustering coefficient approaches zero during the whole evolution.

On one hand, the connection probability set in random graph model is 0.017 in order to keep similar number of links in different graph types. As a result, the link density is low. On the other hand, we find the average clustering coefficient C_G is much lower than that in scale free graph in section 4.2.3. The difference lies in the topological interconnection between random and scale free graph.

The evolving properties of C_G in random graph are simple. When initial random graph grows by either link attachment, almost no neighboring nodes are interconnected in graph object. The local correlation is very weak in random graph and during its evolution process as well.

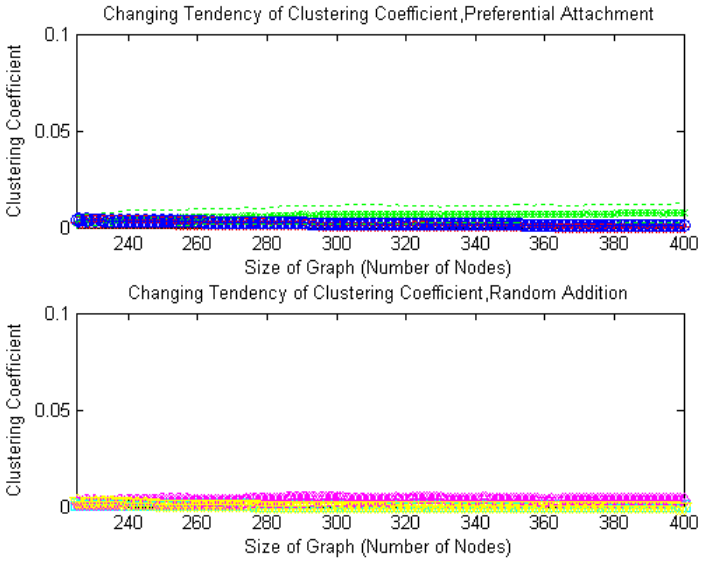


Figure 4.3.3 Evolving clustering coefficient in random graph

4.3.4. Assortativity coefficient

A random graph is neither assortative nor disassortative according to the definition of assortativity coefficient. Links connect pair of nodes randomly in random graph. Thus, the assortativity coefficient r is a random variable with an expectation of 0.

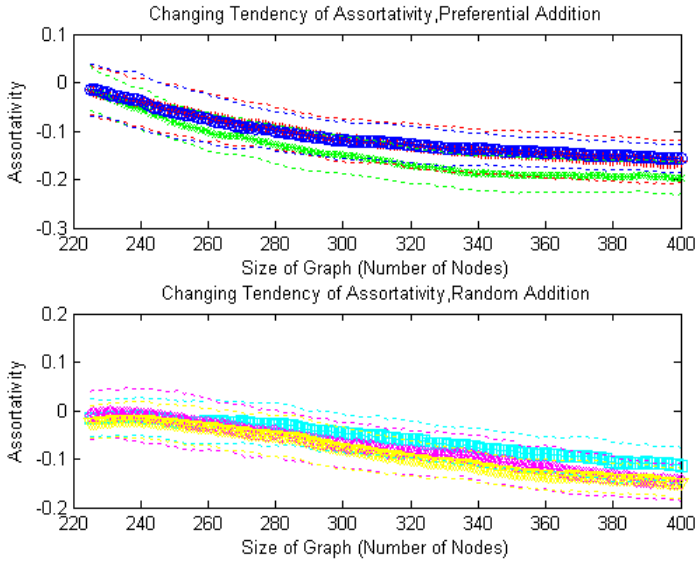


Figure 4.3.4 Evolving assortativity coefficient in random graph

In Figure 4.3.4, evolving assortativity coefficient decreases from 0 and then *r* gets negative. From section 4.2.4, we know the evolving assortativity coefficient remains around -0.15 in scale free graph during the whole evolution process. In the upper subfigure of Figure 4.3.4, after declining from zero, assortativity coefficient finally is stabilized around -0.17. We deduce that the graph object tends to possess partial original attribute of scale free graph as initial random graph grows by preferential attachment. By comparison, *r* decreases as well but only reaches -0.14 at the end of evolution process by random attachment.

4.3.5. Betweenness centrality

The average value of betweenness in Figure 4.3.5(a) and (b) is low and the situation is similar to that in figure 4.2.5(a) and (b).

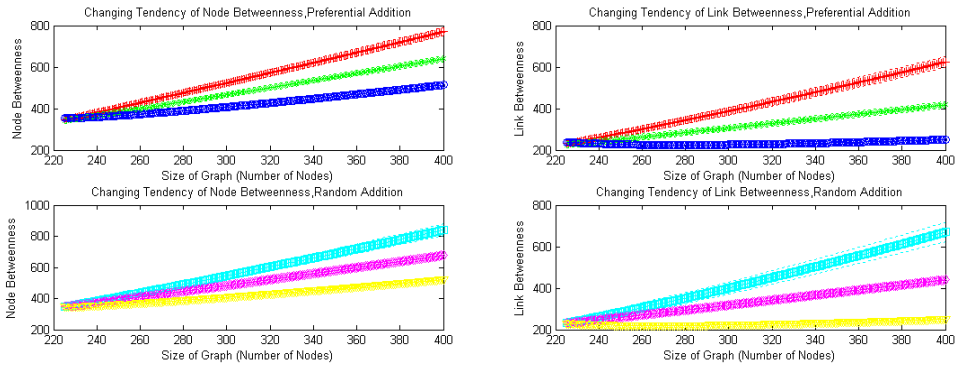


Figure 4.3.5 (a) Left: Evolving node betweenness in random graph, (b)Right: Evolving link betweeness, in random graph

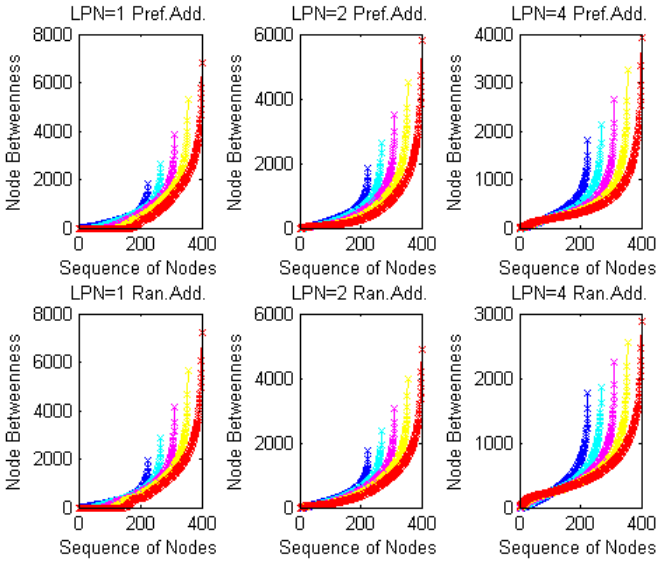


Figure 4.3.5 (c) Evolving node betweenness distribution in random graph

The distribution of node betweenness differs a lot although their average performance seems to be alike. Betweenness distrbution of scale free graph is unlike those in Figure 4.3.5(c) where most nodes are of low betweenness. In Figure 4.3.5(c), the sorted distribution increases gradually and the maximum value is lower. During the evolution from initial random graph, no such nodes like hubs exist. Nodes contribute to sustain the traffic load in random graph.

4.4. Regular lattice graph

The regular lattice graph model is a graph with 225 nodes on the regular grid structure. We assume that links are interlinked with the probability of 1. In other words, node connects all its four nearest nodes available in lattice graph.

The initial lattice graph generated is of 225 nodes and 420 links with determinate topology at the beginning of evolution process.

4.4.1. Node degree

The topology of a lattice graph is fixed once the number of nodes is provided. Due to this character, it's possible for us to investigate deeply into its details. From equation (3.5), we get the expression of average node degree in lattice graph $E[D] = 4 - 4/n$ where $n*n = N$. In addition to $E[D]$, the degree distribution $\Pr[D=k]$ is also deterministic in lattice graph. There are only three degree values in lattice graph. In

our lattice graph model, four nodes on the corner of lattice grid are of degree two. 52 nodes on the boundary of graph except the corner have 3 links. The rest 169 nodes connecting all four nearest nodes constitute most of nodes and dominate in the lattice graph. There is not any node with high degree in the lattice graph due to its topology.

In Figure 4.4.1(b), the curve of degree distribution in target graph is similar to each other in the condition of the same *LPN* setting.

We first discuss the situation when the first link is built up in the evolution process. In preferential evolution, each four-degree node have probability of 4/(420*2) = 0.0048 to each time a new links added, while each three-degree and two-degree node has probability of 0.0036 and 0.0024 respectively in initial lattice graph.

The first generated link has more than 0.048*169 = 81.2% to attach an existing three-degree node. On the other hand, each node has equivalent probability equal to 0.044 in initial state with random link attachment. The first link also has high probability of more than 75.1% to attach an existing four-degree node.

Limited number of links added to graph will not influence the dominance situation of four-degree node since the size of graph is relative large. Each perturbation changes the graph topology little so that probability of four-degree node decreases rapidly until nodes of three degree no longer dominate in the graph.

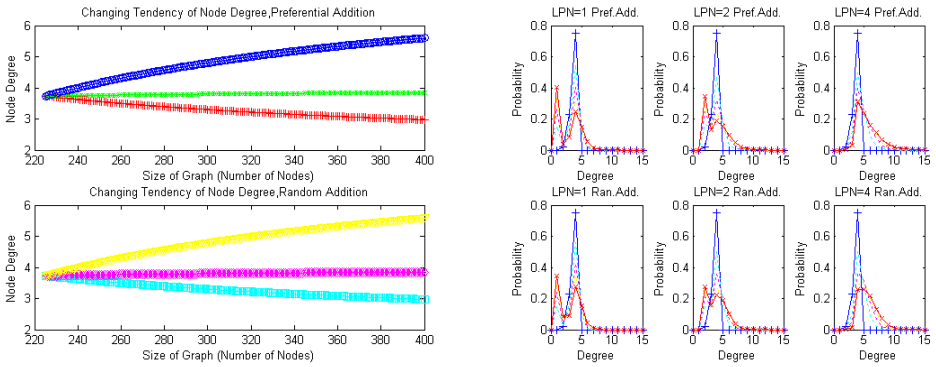


Figure 4.4.1 (a) Left: Evolving node degree, in lattice graph (b) Right: Evolving node distribution, in lattice graph

4.4.2. Average hopcount

Equation (3.6) gives the expression of average hopcount $E[H] = 2n/3$ in lattice graph. It implies that the growth of $E[H]$ in square lattice graph is proportional to increasing N. The average hopcount $E[H]$ of initial graph is equal to 10 by using equation (3.6). Average hopcount is high in lattice graph since each node only

connects its nearest node in each direction. There is not any other link established between nodes. One extreme case is the hopcount between two nodes on diagonal corners equals 28.

As a result, lots of shortcuts are created to connect widely separated parts of lattice graph when new nodes and links are added to graph at $LPN = 2$ or 4. In Figure 4.4.2 we find derivative of average hopcount curve is high at the beginning of evolution. The first several shortcuts decrease *E[H]* sufficiently at the beginning of the evolution process. As the graph continues to grows, *E[H]* drop down and the shortcut-effect introduced by new added links is gradually weakened.

For $LPN = 1$, new node fails to create any short cut between present nodes in the graph. New node first passes through its exclusive neighboring node to connect the graph. The curve of average hopcount increases slightly with fluctuation of small magnitude during both types of evolution. Particularly, the value of average hopcount by random attachment always lies a little higher than that with the same graph size N during preferential evolution.

Preferential attachment scheme makes new generated node more likely to connect high degree nodes in the graph but this effect is not obvious for initial lattice graph with large size. The reason has been state in section 4.4.1.

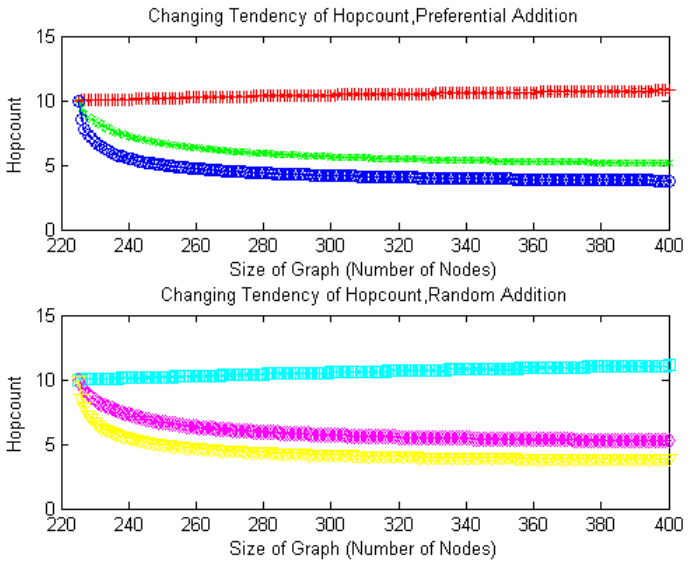


Figure 4.4.2 Evolving hopcount in lattice graph

4.4.3. Clustering coefficient

Because of the characteristic topological pattern in lattice graph, the clustering

coefficient of each node is zero in initial graph. In the condition of $LPN = 1$, the average clustering coefficient remains zero during such evolution.

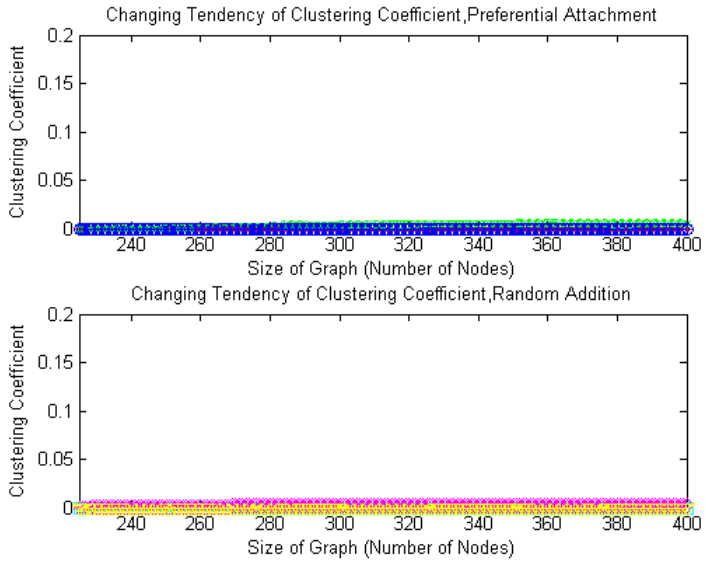


Figure 4.4.3 Evolving clustering coefficient in lattice graph

For $LPN = 2$ and 4, the number of nodes is relatively large even compared with the highest LPN we set in the simulation. It seldom happens that a new added node connect two or more than two nodes happen to be adjacent. The evolving C_G of graph positively approach zero during the whole evolution process. This behavior is similar to that in random graph discussed in section 4.3.4. The simulation result displayed in Figure 4.4.3 also demonstrates our analytical results.

4.4.4. Assortativity coefficient

Most nodes connect their neighboring nodes with the same degree in lattice graph. Lattice graph is obvious highly assortative. The assortativity coefficient r is equal to 0.61 in initial lattice graph. The number of nodes with four degree covers more than 75% of total nodes in initial graph. For $LPN = 4$ by both evolving attachment, the evolving average value decreases least at given growing size N.

The topological structure in ordered pattern is impaired gradually during the evolution and therefore the assortative mixing property deteriorates as nodes added continuously to graph and then disappears. As more nodes added to graph, r finally become negative.

We conclude that the attribute of assortativity in lattice graph alters when graph grows by either attachment scheme.

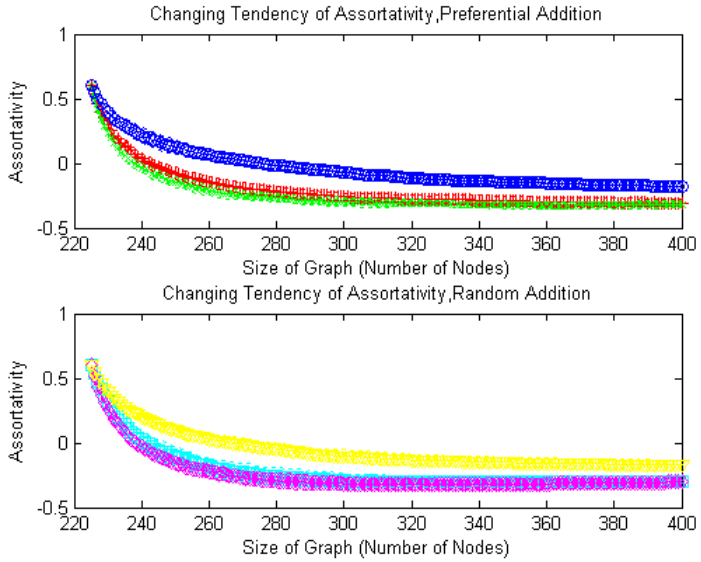


Figure 4.4.4 Evolving assortativity coefficient in lattice graph

4.4.5. Betweenness centrality

Figure 4.4.5(c) shows a feature that the sorted node betweenness distribution (in blue curve) increases almost linearly. The reason is that nodes located in the central part of the initial lattice graph play the most important role in traffic load due to the topology of lattice graph. Its importance is weakened linearly as the location of nodes moves towards in all directions.

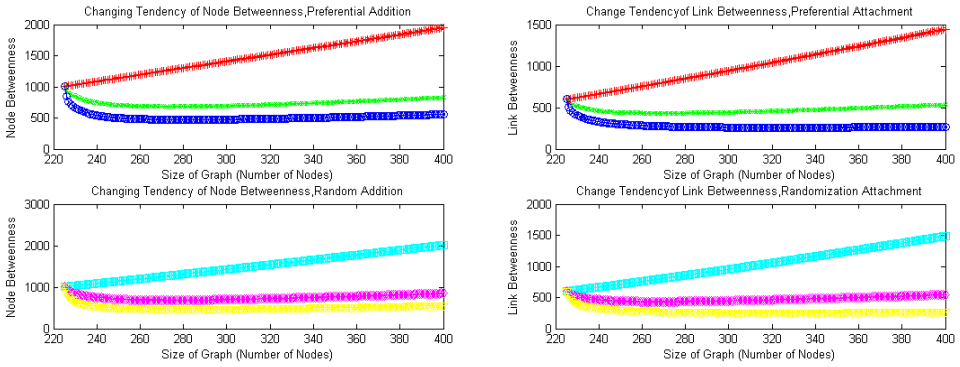


Figure 4.4.5 (a) Left: Evolving node betweenness in lattice graph, (b)Right: Evolving link betweeness, in lattice graph

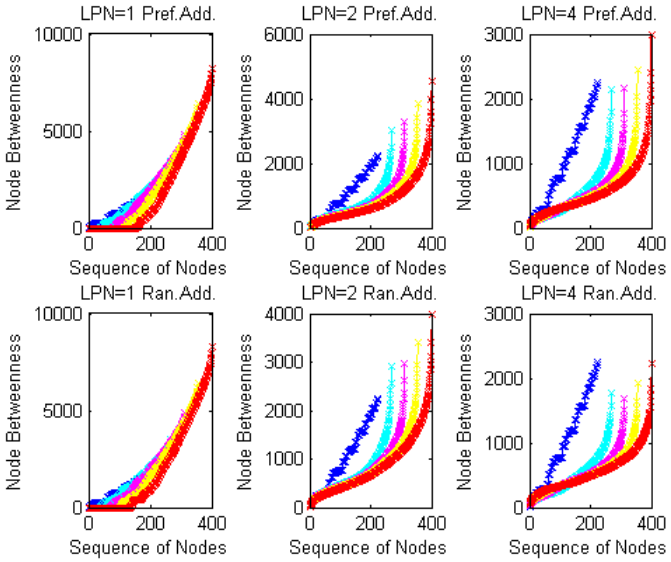


Figure 4.4.5 (c) Evolving node betweenness distribution in lattice graph

As we discuss above, the evolution built up lots of shortcuts in the graph if newly generated nodes connect more than one existing nodes during the evolution. For $LPN = 2$ and 4, we observe that the curves of average betweenness decrease dramatically as soon as evolution starts and then the curves keep steady as N continues to increase.

For $LPN = 1$, curves of the average betweenness increase. New nodes are attached to graph object with one link only. In upper left subfigure of figure 4.4.5(c) and lower left subfigure of figure 4.4.5(c), the flat part of red curve representing the betweenness distribution of target graph is introduced by the new added nodes during the evolution process. It covers approximately 170 nodes and the number is closed to the number of nodes added during the evolution. The new introduced nodes do not contribute much to betweenness centrality since they just have one link to attach the graph object.

4.5. Small world model of Watts and Strogatz

In this section, we discuss the evolving properties in small world model of Watts and Strogatz.

We assume construction of small world model lies between random graph and four-nearest node ring graph. The link rewiring probability p is set to 0.09. There are 450 links and 225 nodes in initial graph of small world model.

4.5.1. Node degree

Degree distributions Pr[D=k] of all targets graphs plotted in Figure 4.5.1(b) resemble those in initial lattice graph in Figure 4.4.1(b). With low link rewiring probability p = 0.09 in the simulation, the evolution of Pr[D=k] reserves mostly the evolving property of degree distribution in lattice graph. We set k-nearest neighbor number to 4 in ring graph in the construction of small world model. Therefore, more than 70% nodes are 4 connected in initial small world model.

As nodes are added to the graph object, the dominance situation of four-degree node is gradually weakened. In target state, Pr[D=k] peak at k = LPN since a large number of k connected nodes added to graph throughout the evolution process.

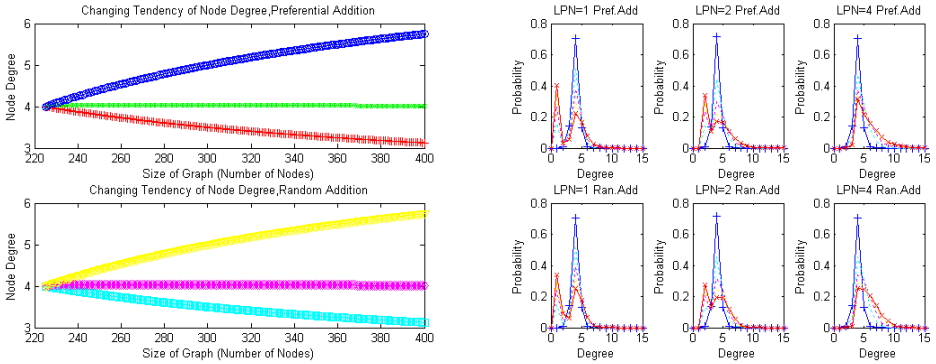


Figure 4.5.1 (a) Left: Evolving node degree, in small world model (b) Right: Evolving node distribution, in small world model

4.5.2. Average hopcount

The hopcount in initial small world model is plotted in Figure 4.5.2. The small world model lies in an intermediate regime between random graph and ring graph. With link rewiring probability 0.09, the average hopcount $E[H]$ is equal to 6.47 in initial small world model. It is lower than $E[H]$ in lattice graph at given graph size. In chapter 3, we explain that there is a change in the scaling of the average hopcount $E[H]$ as the fraction p increases. The rewiring process introduces pkN / 2 long-range links in the graph.

Limited long-rang links are introduced to graph due to p = 0.09. For LPN = 2 and 4, $E[H]$ descend with high decreasing rate at the beginning of evolution process. New links added still create some shortcuts during the evolution process. Since more and more shortcuts added as graph continues to grow, decreasing ratio of $E[H]$ declines. This feature is similar to that in lattice graph discussed in section 4.4.2.

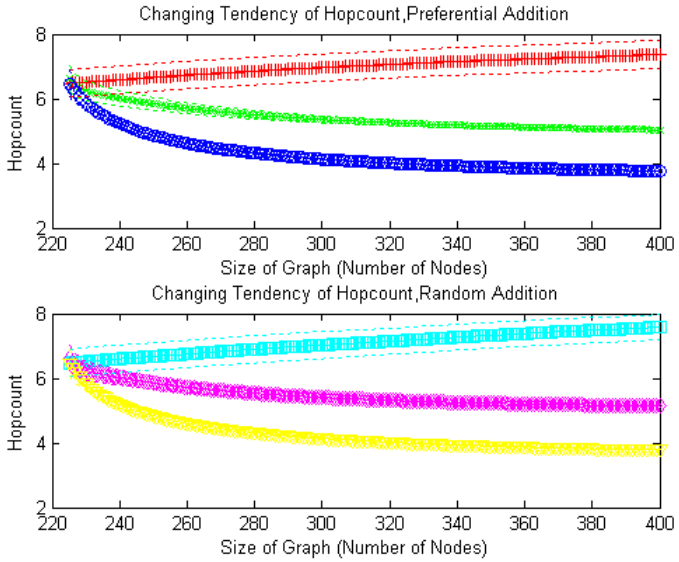


Figure 4.5.2 Evolving hopcount in small world model

4.5.3. Clustering coefficient

The proposal for constructing small world models is to create a type of graph with both high cluster as in lattice ring graph and low average hopcount as in random graph. In simulation, C_G is equal to 0.90 on average in initial small world graph.

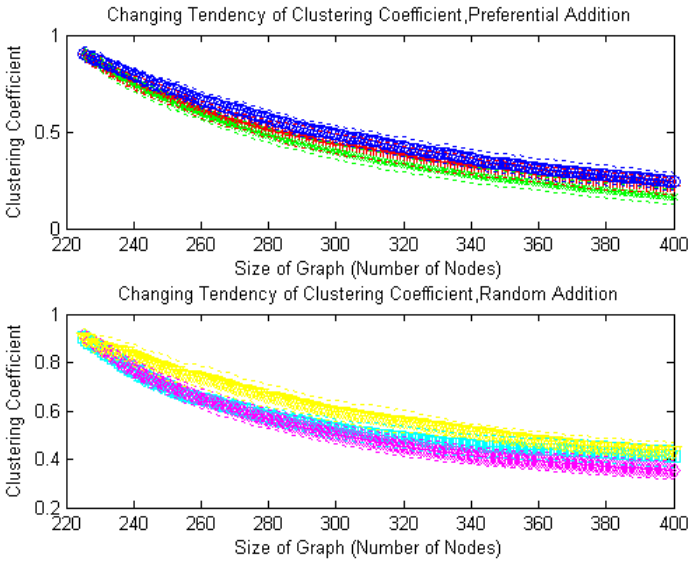


Figure 4.5.3 Evolving clustering coefficient in small world model

The initial model remains highly clustered because of the low rewiring

probability p. The evolving tendency resembles that in lattice graph in section 4.4.3. The high clustered property deteriorates as graph grows. The cause of decreasing C_G is the same as that in previous models. Newly generated nodes are unlikely connected to nodes already interlinked in present graph object. As a consequence, C_G decreases throughout the entire evolution process.

4.5.4. Assortativity coefficient

The small world model of Watts and Strogatz is a random graph model with small world characteristics and an abundance of short loops. Therefore, there is no correlation between node degrees in small world model as in random graph.

In the simulation, the assortativity coefficient r is equal to -0.025 in initial small world model. Figure 4.5.4 shows that r begins to decrease and the graph becomes disassortative when small world model start to grow. The descending trend from zero resembles the evolving properties of r in random graph in Figure 4.3.4.

Further, we observe curves at $LPN = 4$ decrease least. For $LPN = 1$ and 2, r falls rapidly first and then tends to be stable gradually after 100 nodes added. From the simulation results, there is no obvious difference in evolving r between two link attachment schemes.

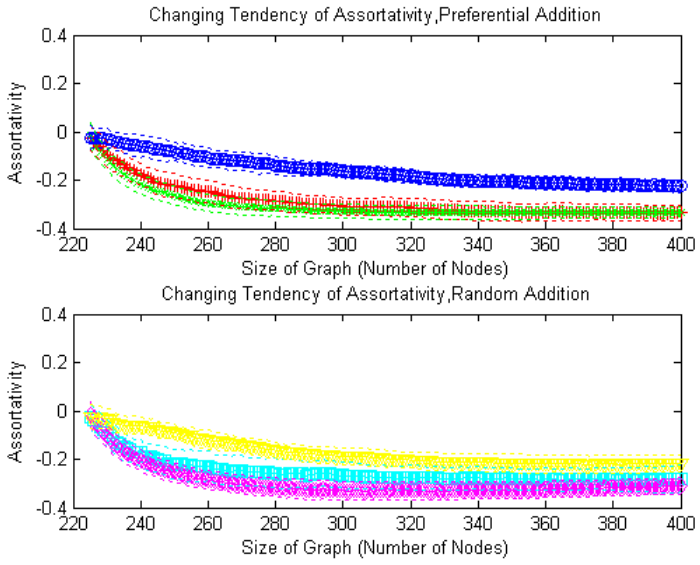


Figure 4.5.4 Evolving assortativity coefficient in small world model

4.5.5. Betweenness centrality

The average betweenness of initial small world model lies higher than those in random and scale free graph but much lower than that in lattice graph. In Figure 4.5.5 (a), curves at $LPN = 4$ keep stable as N increases. This indicates that the average betweenness falls down relatively. In Figure 4.5.5 (b), the curves of distribution also conform to the situation. At $LPN = 4$, maximum node betweenness of target distribution (in red curve) is even lower than the maximum value of initial distribution. It means less number of shortest paths pass across the node of the highest betweenness in the target graph. From the simulation result, the sorted distribution in target graph increases with lower increase ratio. We know shortest path between widely separated nodes largely relies on limited long range links in small world model. As we discuss in section 4.5.2, during the evolution process, new links create shortcuts and they relieve traffic load on nodes lying on the end of long range links in initial graphs.

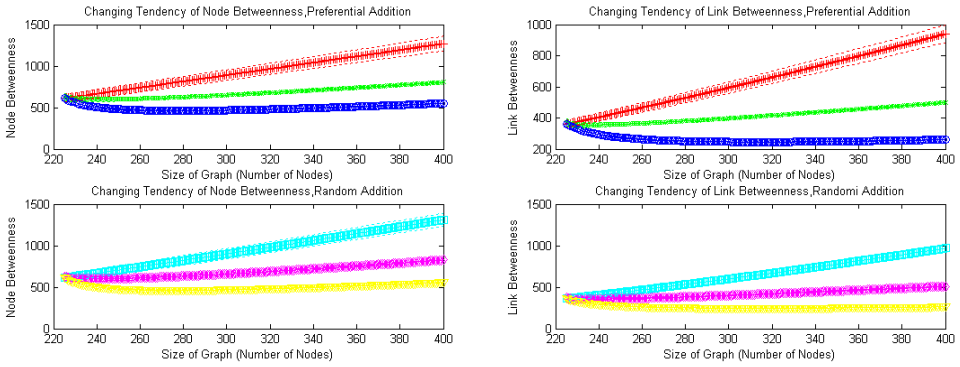


Figure 4.5.5 (a) Left: Evolving node betweenness in small world model, (b)Right: Evolving link betweeness, in small world model

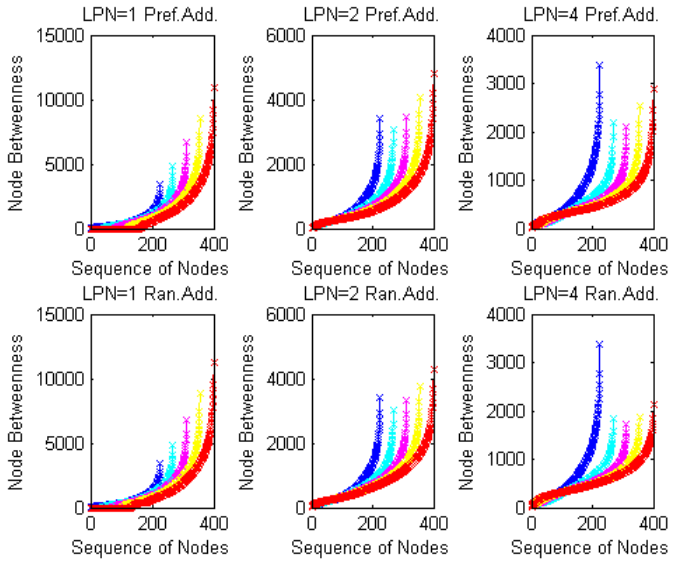


Figure 4.5.5 (c) Evolving node betweenness distribution in small world model

4.6. Special models: line graph and ring graph

After investigating the simulation results of four types of graph models, in this section we try to explore the evolving properties of two simple but special models: line graph and ring graph.

Both these two models are generated with deterministic topology so that the number of links can not be adjusted to approach those in typical models closely. Due to particular topology in these two models, some interesting features in some topological measures are revealed. We briefly discuss both types of graphs together in this section.

4.6.1. Node degree

There is one more link in the ring graph than that of the line graph at a given number of nodes during the evolution. As a consequence, the evolving tendencies of average node degree between two types of graphs are almost the same.

Since the number of link and node in the graph object at any state are much larger than one, difference in the number of link hardly generates any discrepancy of degree distributions by either link attachment. Particularly in the simulation of ring graph, the average node degree stays at 2 at $LPN = 1$.

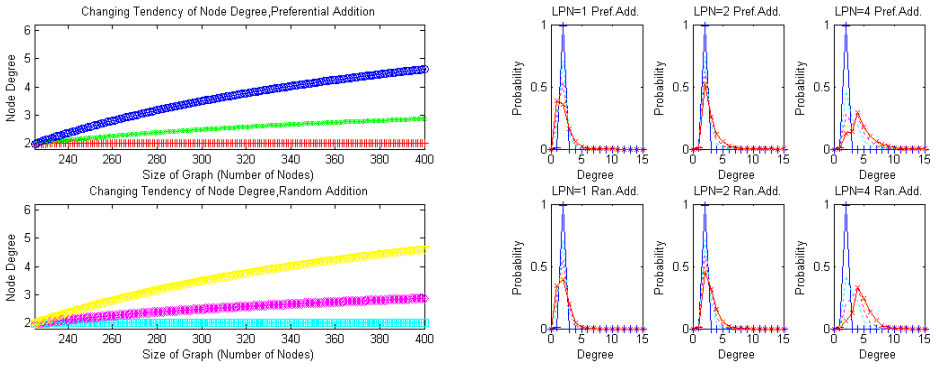


Figure 4.6.1 (a) Left: Evolving node degree, in line graph (b) Right: Evolving node distribution, in line graph

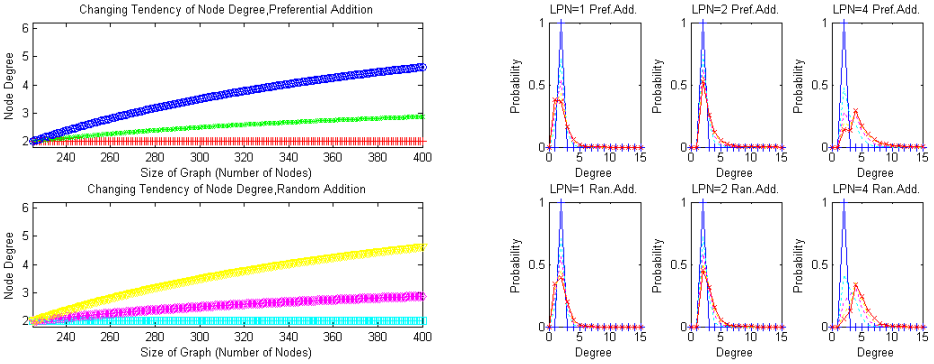


Figure 4.6.1 (c) Left: Evolving node degree, in ring graph (d) Right: Evolving node distribution, in ring graph.

4.6.2. Average hopcount

Regarding hopcount, the behavior resembles that in lattice graph in figure. It is obvious that the average hopcount in line graph is higher than that in ring graph. When *LPN* is larger than one, shortcuts built up between in the previous graph. With larger *LPN*, more shortcuts created each time a new node generated and therefore the curves fall suddenly at the beginning. For *LPN* = 4 in upper subfigure in Figure 4.6.2(a), after the first nodes with four links attached to the line graph, the average hopcount drops from 75.33 to 43.58. These four links almost decrease the average hopcount by 42%. After the second set of four links attached, the average hopcount falls to 27.7. After first several steps of growth, the curve falls gently and finally is stabilized.

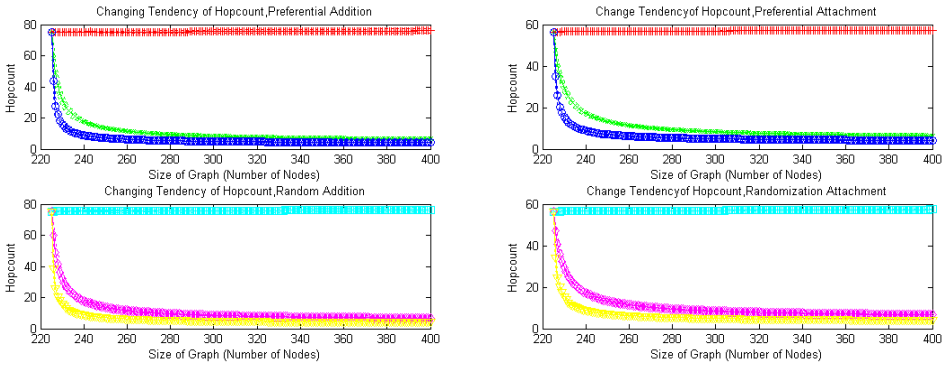


Figure 4.6.2 (a) Left: Evolving hopcount in line graph (b) Right: Evolving hopcount in ring graph

4.6.3. Clustering coefficient

The clustering coefficients in both types of graphs are near to zero during the evolution. It is the same reason as for lattice graph discussed in section 4.4.3.

The probability of connecting two adjacent nodes is quite low during the evolution with either attachment scheme.

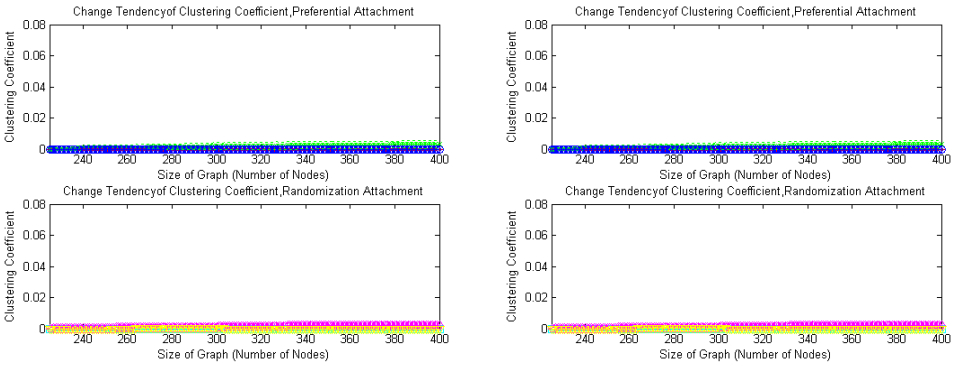


Figure 4.6.3 (a) Left: Evolving clustering coefficient in line graph (b) Right: Evolving clustering coefficient in ring graph

4.6.4. Assortativity coefficient

The value of assortativity coefficient in initial ring graph is undefined since both numerator and denominator are equal to zero. Each node in ring graph defined in this chapter connects its two nearest neighbors. As a result, the variance in the denominator is zero and the difference in the numerator is also zero.

For convenience, we remark it as zero though the initial ring graph is not neutral mixing. But the undefined situation doesn't occur in the simulation of line graph.

The assortativity coefficient r of line graph is equal to -0.0045 which indicates the line graph slightly disassortative.

The initial line graph contains 224 links and 225 nodes. Only two nodes at the end of the line connect their neighbors of different degree. The rest 222 pairs of nodes are with the same degree. The little discrepancy in topology results in large difference in values of assortativity coefficient.

Most nodes in the initial line graph and all nodes in the initial ring graph are of two degree. The evolving property of assortativity coefficient largely depends on value of LPN. For $LPN = 2$, the evolving assortativity coefficients decrease from

near -0.045 in line graph and undefined point in ring graph to -0.2. For *LPN* = 4, the coefficient is always positive and it becomes negative for *LPN* = 1.

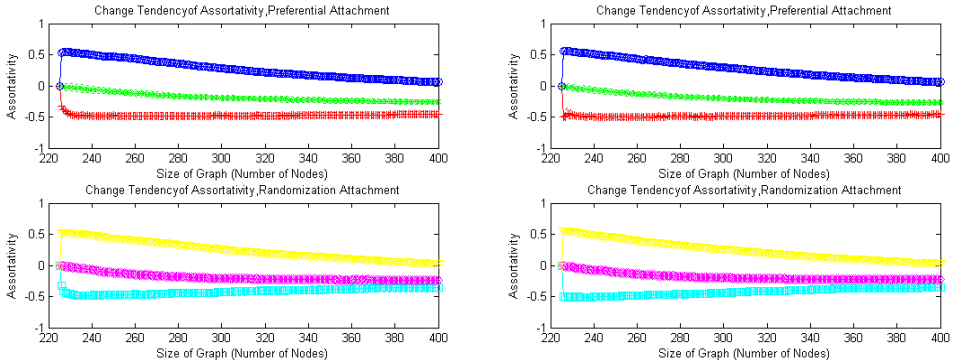


Figure 4.6.4 (a) Left: Evolving assortativity coefficient in line graph (b) Right: Evolving assortativity coefficient in ring graph

4.6.5. Betweenness centrality

The character of average betweenness is similar to that in the simulation of lattice graph with the same cause.

We focus on the distribution of node betweenness between two models. In the initial line graph, we observe that the sorted distribution increases approximately linearly and bends at the end of its tail. In initial ring graph, the curve is a parallel line since each node in the ring graph lies in the same status. For *LPN* more than one, the evolving node betweenness distributions tend to be a regular distribution as in the typical models.

For the special case of *LPN* = 1, since new nodes almost do not share the capacity of traffic load in the graph, most newly added nodes have low node betweenness. In the simulation of line graph, the rest part of the curve still keeps linear increasing with distortion.

In the simulation of ring graph, the curve is discontinuous at a certain number *N"* in the node sequence, where *N"* is approximately equal to the number of nodes added to the graph from the beginning of evolution.

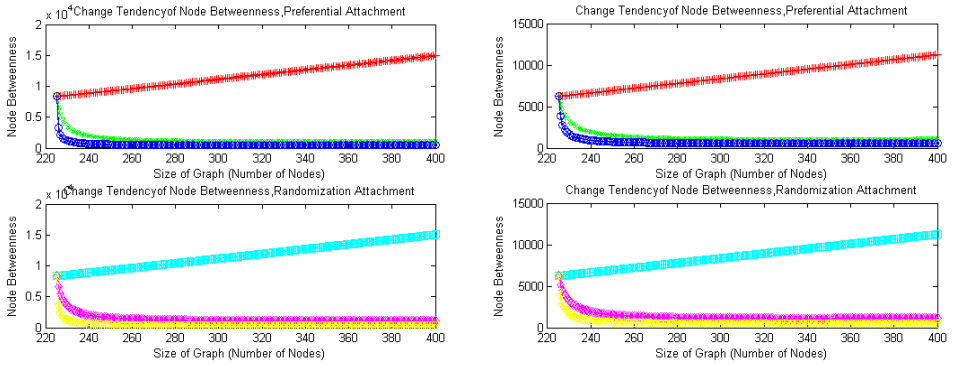


Figure 4.6.5 (a) Left: Evolving node betweenness in line graph (b) Right: Evolving node betweenness in ring graph

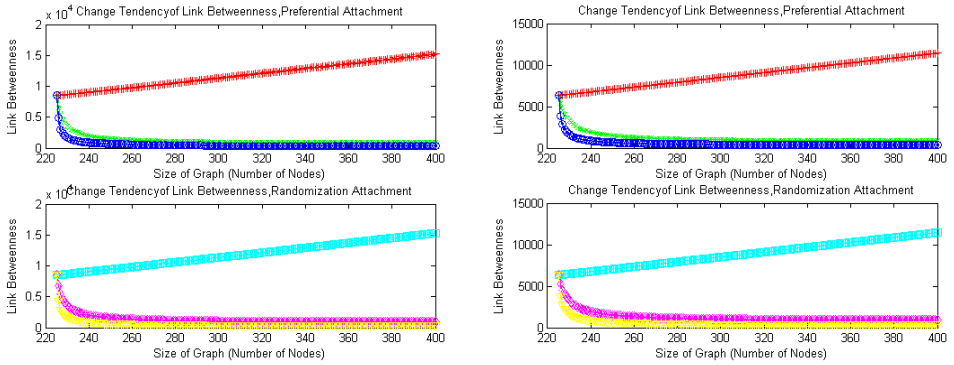


Figure 4.6.5 (c) Left: Evolving link betweenness in line graph (d) Right: Evolving link betweenness in ring graph

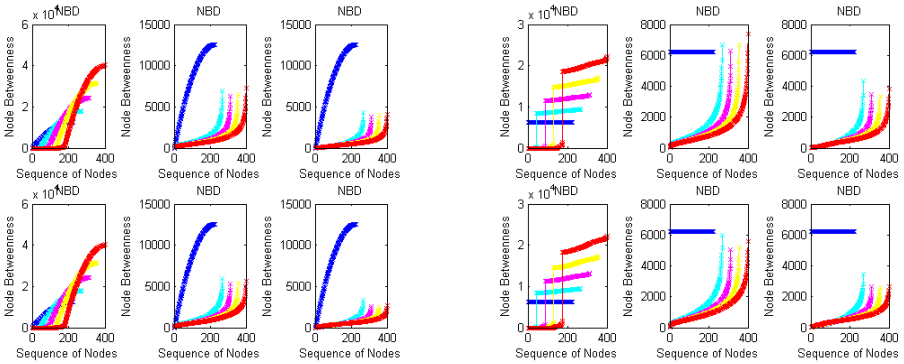


Figure 4.6.5 (e) Left: Evolving node betweenness distribution in line graph (f) Right: Evolving node betweenness distribution in ring graph

5

5. Evolving Properties of Growing Real Networks

5.1. Classification and model settings

In addition to the six graph models discussed in chapter 4, we simulate fifteen different data sets of real networks. The types of the real networks cover a wide range among social networks, biological networks, technological networks and Linguistic networks. The details of them have been introduced in chapter 3.

In this chapter, we present the simulation result of them and investigate their evolving properties. Due to the condition addressed in chapter 3, the number of nodes of fifteen selected real networks is constricted to the scope of several hundreds. The number of links is less than several thousands since none of the real networks have a significant link density.

Large increase ratio of the graph size results in loss of characteristics of the initial graph. The increase ratio of the number of links and nodes are both set to 20% in the growth procedure. In other words, we select link density as criterion and fix it during the entire procedure.

In table 5.1(a), details of number of links and nodes in both initial and target graphs of all the fifteen networks are given. The difference ΔN and ΔL are rounded to integers and are listed in the table 5.1(b). The quotient *LPN* which is the number of links added each time a new node is generated during the growth procedure. The value of *LPN* is approximately equal to the link density.

In chapter 4, we show that the value of *LPN* influences the evolving properties of growing graphs and we classify the simulation results by *LPN*. For real networks, we

also divide data sets of them into two groups by *LPN*.

We observe that values of *LPN* range between 1 and 2 in two-thirds of real networks. The values of *LPN* are larger than 2 in the rest one-third of total real networks.

name of file	L	N	L'	N'
ARPANET80	86	71	103	85
LesMis	254	77	305	92
WordAdj	425	112	510	134
American_football	613	115	736	138
ElectrC_s208	189	122	227	146
Florida	2075	128	2490	154
DigitalMaterial	269	187	323	224
ElectrC_s420	399	252	479	302
celegansneural	2148	297	2578	356
Gnutella3	459	435	551	522
ElectrC_s838	819	512	983	614
Gnutella4	738	653	886	784
W_EU_railway_infrastructure_level2	785	697	942	836
Gnutella1	803	737	964	884
W_EU_power_grid_level3	786	756	943	907

Table 5.1 (a) the number of links and nodes in initial and target graphs

name of file	$\Delta L = L'-L$	$\Delta N = N'-N$	LPN
ARPANET80	17	14	1.21
LesMis	51	15	3.4
WordAdj	85	22	3.86
American_football	123	23	5.35
ElectrC_s208	38	24	1.58
Florida	415	26	15.96
DigitalMaterial	54	37	1.46
ElectrC_s420	80	50	1.6
celegansneural	430	59	7.29
Gnutella3	92	87	1.06
ElectrC_s838	164	102	1.61
Gnutella4	148	131	1.13
W_EU_railway_infrastructure_level2	157	139	1.13
Gnutella1	161	147	1.10
W_EU_power_grid_level3	157	151	1.04

Table 5.1 (b) the difference of links and nodes in initial and target graphs

Each network has its characteristic properties and topology. Brief introduction of each graph has been listed in chapter 3. In this chapter, we classify them into three types. They are:

- Technological network
- Social-biological network
- Linguistic network

We choose one real network in each type to discuss in this chapter. Each selected network is presented by simulation results and we analyze the evolving properties during the evolution. We also have to deduce their topology by derived result since not much information of topology in each real network is provided.

5.2. Social-biological network: LesMis

Les Miserables network is a social network containing 77 co-appearance network of characters in the novel Les Miserables.

LesMis network grows from 77 nodes and 254 links to 92 nodes and 305 links Each new node is generated with 3.4 links on average during the evolution. In the simulation, we add either 3 or 4 links each time a new node added. The growth mechanism has been introduced in chapter 3. The average node degree holds around 6.6 during the evolution. The degree distribution in initial graph in Figure 5.2.1 shows that one-degree nodes occupy 22% of total nodes in the initial graph. Further, more than half of the nodes in initial graph have less than five adjacent nodes while some nodes have moderate degree. Several nodes have quite large degree more than 20. These characters who know other persons in the novel are usually supposed to be leading characters. During the evolution, Pr[k=3] and Pr[k=4] increase evidently because LPN = 3.4. Several high connected nodes tend to link more nodes during the preferential evolution and their node degree increase to around 40 in target graph. Such situation doesn't occur in random evolution. The initial average hopcount is only 2.65. In that novel, private relation between any two characters can be established by no more than two intermediate characters on average. It is natural for social network to own the small world properties and short average hopcount. The average hopcount holds around 2.65 in the evolution by preferential attachment and rises to 2.7 by the other attachment. High degree nodes work as hubs in the graph and relieve the increase in hopcount when nodes added.

The average clustering coefficient starts at 0.34 which is high in comparison to

other graphs in simulation, partially because of its link density. During the evolution process, the coefficient descends evenly to less than 0.18. The initial graph is disassortative and its assortativity coefficient r is -0.17. r increases towards zero during the evolution process. In particular, by random link attachment, that trend is apparent. The evolving behavior of assortativity coefficient r resembles that in scale free graph in Figure 4.2.4. Regarding the average betweenness, we find the latter part of the node betweenness curve ascends with steep slope. About ten key nodes in the graph always sustain large number of shortest path. It demonstrates the existence of hubs in the topological structure of LesMis network.

In conclusion, some evolving properties in LesMis network also exist in scale free graph and small world model from the simulation result.

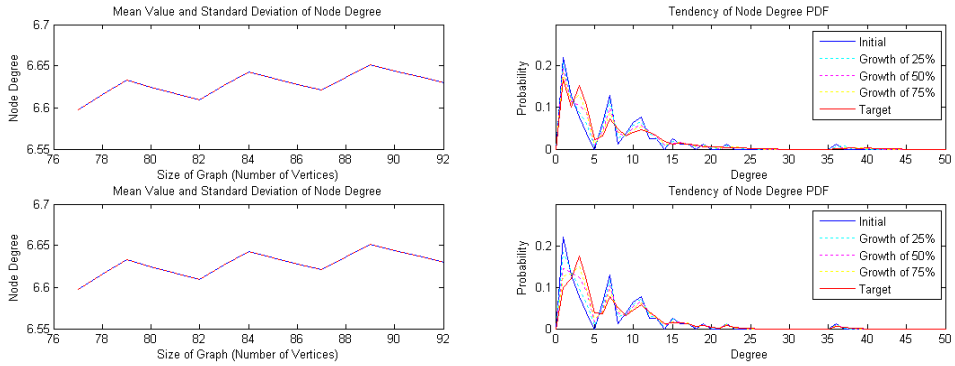


Figure 5.2 (a) Left: Evolving node degree, in LesMis network (b) Right: Evolving node distribution, in LesMis network

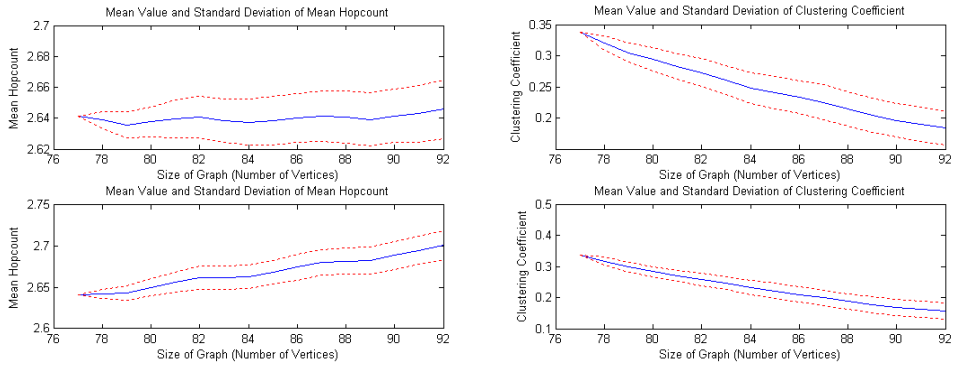


Figure 5.2 (c) Left: Evolving average hopcount, in LesMis network (d) Right: Evolving clustering coefficient, in LesMis network

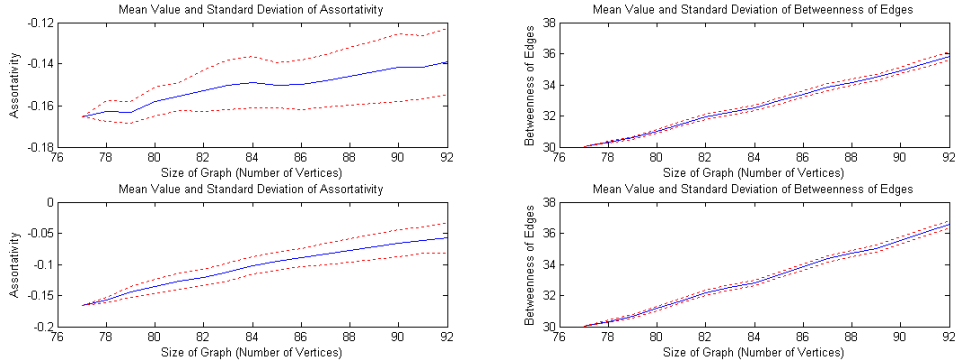


Figure 5.2 (e) Left: Evolving assortativity coefficient, in LesMis network (f) Right: Evolving node betweenness, in LesMis network

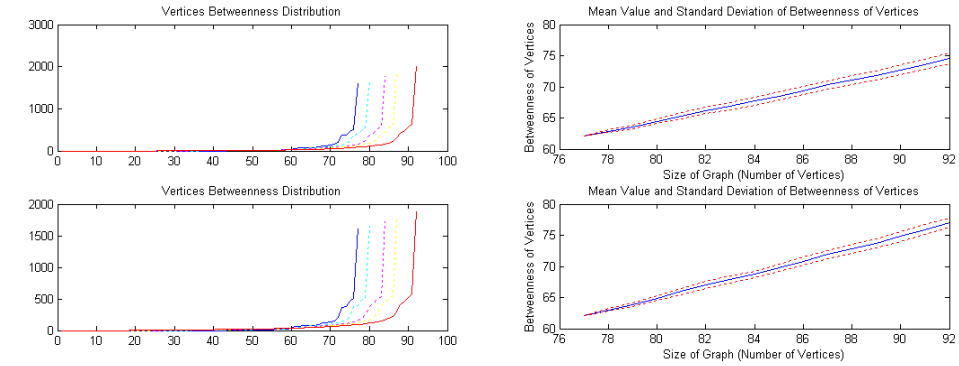


Figure 5.2 (g) Left: Evolving node degree distribution, in LesMis network (h) Right: Evolving link betweenness, in LesMis network

5.3. Technological network: ElectrC_s208

The circuit network ElectrC_s208 with link density 3.1 is composed of 122 nodes and 189 links. Each new node added to graph accompanied with 1.58 links during the evolution.

The degree distribution looks like binomial as in random graph and it differs from that in social network. The degree distribution becomes smoother as the size of the graph increases. Pr[k=3] falls down gradually. In compensation, probability values of other degree increase slightly. From simulation result, no obvious difference is observed between two attachment strategies.

A significant feature in the circuit network is that the average hopcount falls down when nodes and links added to graph. We deduce that nodes only interlinked to nodes nearby in the network. The backbone topological structure in this network may

resemble those in lattice. On the other hand, the average hopcount is not as high as that in line graph or ring graph.

The assortativity coefficient r starts around 0 so that the circuit network is initially neutral. Then r gets negative as graph continues to grow. The evolving tendency is also found in random graph and small world model.

From the simulation result of sorted node betweenness distribution, the latter part of the curve representing initial sorted distribution distributed approximately linearly. In chapter 4, we find out the properties that sorted distribution of node betweenness is distributed linearly in lattice graph. In three intermediate states and target state, the maximum value of node betweenness almost keeps the same during the evolution process. Since the size of graph increases, it means the importance of those nodes of high betweenness value is relatively weakened.

We find similar properties between the circuit network ElectrC_s208 and the lattice graph.

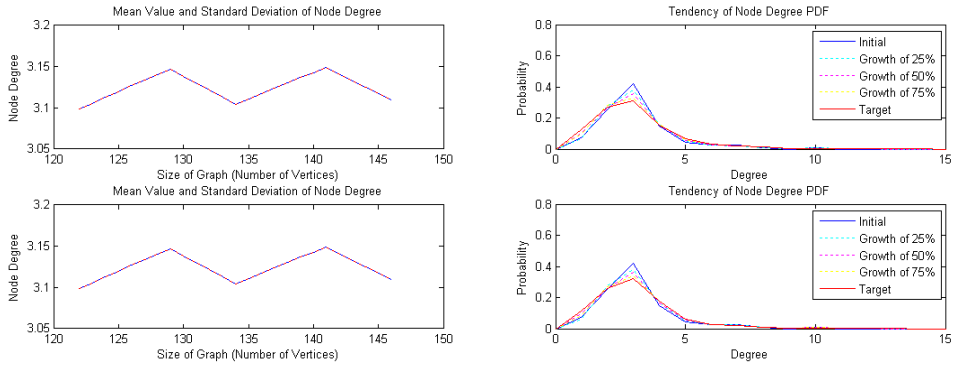


Figure 5.3 (a) Left: Evolving node degree, in ElectrC_s208 network (b) Right: Evolving node distribution, in ElectrC_s208 network

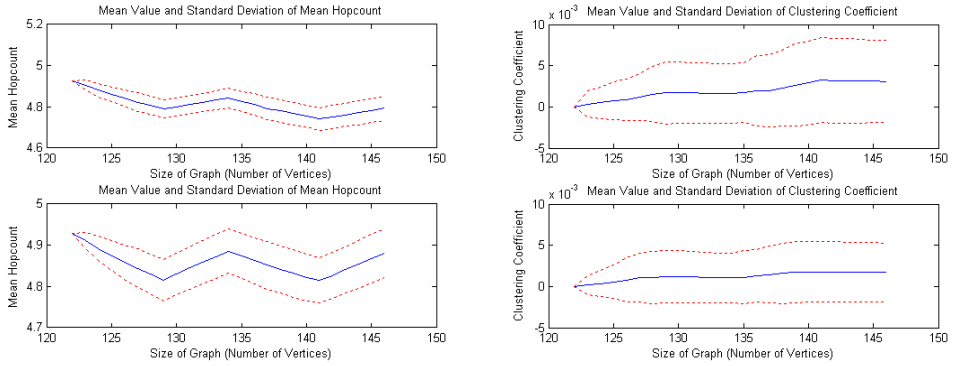


Figure 5.3 (c) Left: Evolving average hopcount, in ElectrC_s208 network (d) Right: Evolving clustering coefficient, in ElectrC_s208 network

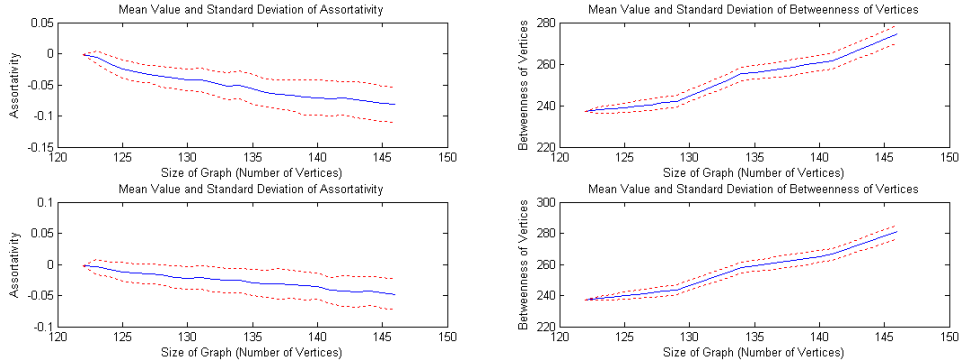


Figure 5.3 (e) Left: Evolving assortativity coefficient, in ElectrC_s208 network (f) Right: Evolving node betweenness, in ElectrC_s208 network

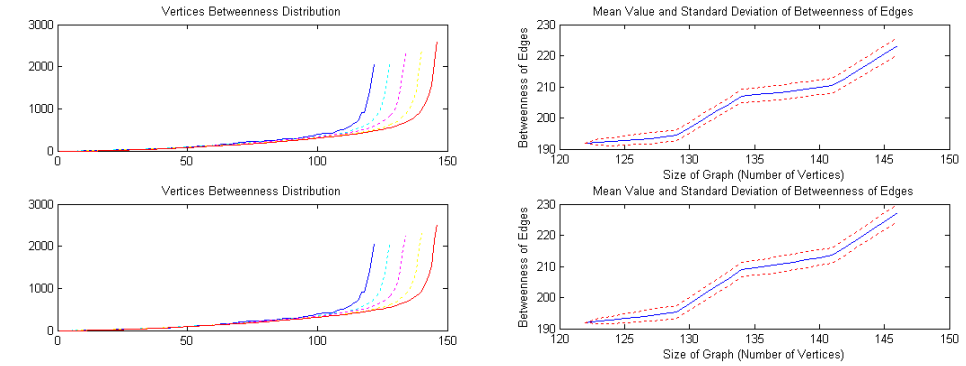


Figure 5.3 (g) Left: Evolving node degree distribution, in ElectrC_s208 network (h) Right: Evolving link betweenness, in ElectrC_s208 network

5.4. Linguistic network: WordAdj

The word adjacency network grows from 122 nodes to 146 nodes. Each time 3.86 links on average are added to newly attached node during the evolution process.

The degree distribution is similar to the degree distribution of LesMis network. In Figure 5.4(b), most nodes have low degree. Some nodes have moderate degree and several nodes have high degree. As graph grows, Pr[k=4] increases and it becomes to be highest after evolution. And the high node degree gets higher probability after evolution. The average hopcount is also low and increases with small magnitude. The average clustering coefficient is initially 0.28, It decreases once the evolution begins and finally reaches 0.09. Regarding the assortativity coefficient, it is -0.13 in initial Linguistic network and then increases when evolution begins. Particularly, in preferential evolution, it only rises to -0.13 while in random evolution, the coefficient

descends with apparent trend. The average betweenness of the Linguistic networks increases.

The Linguistic network WordAdj owns some similar properties to those in social network we discussed.

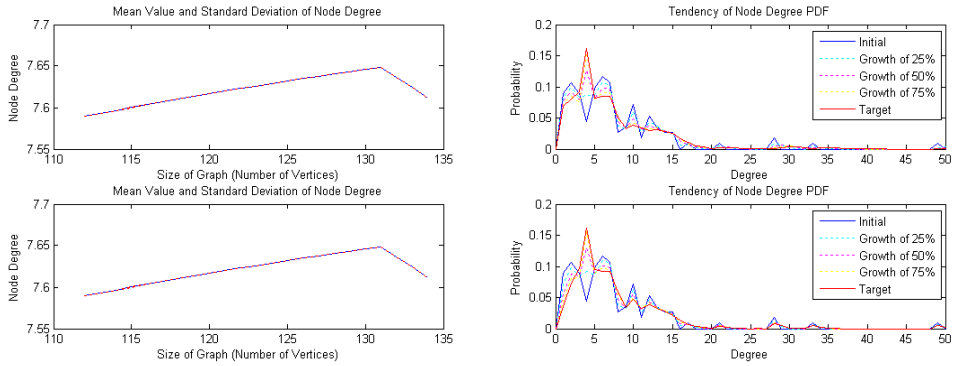


Figure 5.4 (a) Left: Evolving node degree, in WordAdj network (b) Right: Evolving node distribution, in WordAdj network

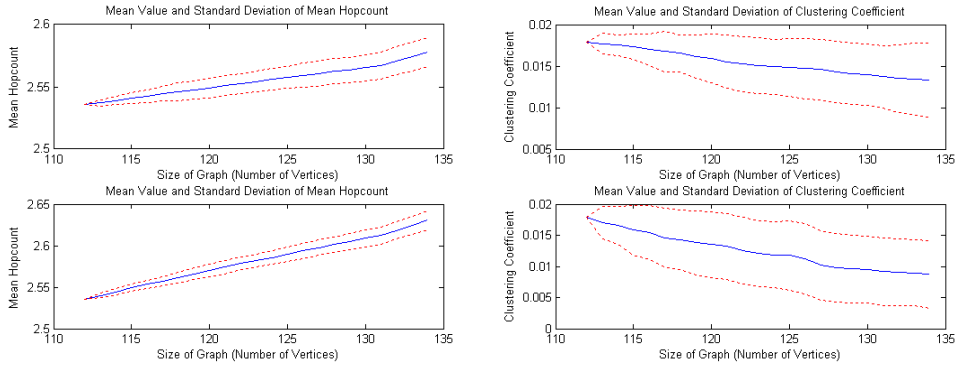


Figure 5.4 (c) Left: Evolving average hopcount, in WordAdj network (d) Right: Evolving clustering coefficient, in WordAdj network

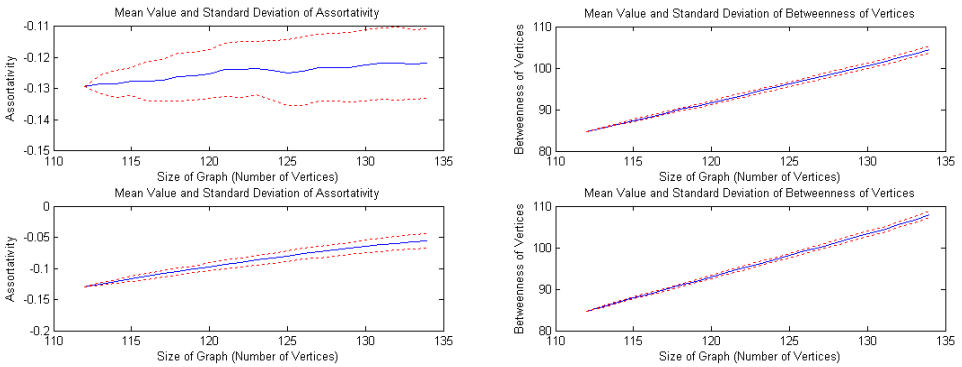


Figure 5.4 (e) Left: Evolving assortativity coefficient, in WordAdj network (f) Right: Evolving node betweenness, in WordAdj network

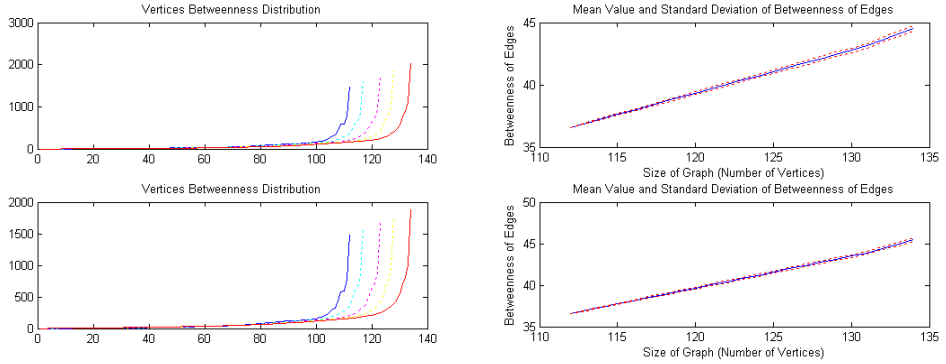


Figure 5.4 (g) Left: Evolving node degree distribution, in WordAdj network (h) Right: Evolving link betweenness, in WordAdj network

6

6. Conclusion and Future Work

6.1. Conclusion

In this thesis, we have implemented both statistical simulation and analytic method based on graph theory to study evolving properties of growing networks. We here have established six types of graph models and introduce fifteen data sets of real networks. Each growth procedure in the simulation is combined with two different link attachments. The number of links accompanied with a newly generated node, denoted by *LPN*, is set at different values in each simulation. We run a large number of simulations on both graph models and real networks. Simulation results of five topological measures were derived at the end of each simulation.

Based on empirical and analytic results, we conclude that the way properties of growing networks evolve is related to different aspects among which several crucial factors have remarkable influence. $LPN = 1$ is an irregular condition only for the purpose of analysis in thesis since each new node attached to present graph by a single link.We discuss the evolving properties conditioned on *LPN* larger than one.

Below we present the conclusions for the different topological metrics.

- Node degree.

The average node degree is directly related to the present number of links and nodes in a graph object. It is independent of the attachment scheme and the graph type. The evolving properties of *E[D]* is influenced by the *LPN* setting during the evolution . Therefore, *E[D]* is for the intermediate regime given by the closed formula (4.5) is an interpolation changing with growing size *N*.

Regarding the degree distribution, every graph has its characteristic distribution in initial states. Obviously, Pr[$D=LPN$] increases for the reason that *LPN*-connected

nodes are added to graph object continuously throughout the evolution process. During the simulation, a graph grows from 225 nodes to 400 nodes and its ratio increases 77.78% during the simulation. In most cases, nodes with *LPN* degree dominate in the target graph. We have explained that the growth procedure by a preferential link attachment is equivalent to the construction of scale free graph. The tail distributions of some target graph after preferential evolution decays more rapidly than that after random evolution. We deduce that the degree distributions in other types of target graphs obtain partial attributes in scale free graph by preferential evolution.

◆ Average hopcount

The evolving properties of average hopcount are mainly related to their graph types. We classify the six graph types into two classes.

In the first class, the average hopcount is polynomial with respect to N in such type of initial graphs. Lattice graph, line graph and ring graph are attributed to this class. The topological patterns of those initial graphs are totally changed during the evolution process. Graphs in the first class have high *E[H]*value. For *LPN* greater than one, the *E[H]* drops rapidly at the beginning of the evolution process since shortcuts are established in the graph. As N increases, the shortcut-effect introduced by newly added links is gradually weakened. *E[H]*declines and finally is almost stabilized by approaching a certain value.

In the second class, the average hopcount is relatively low. *E[H]*is logarithm in N in the initial random graph. *E[H]* in scale free graph is even lower than that in random graph. Small world model has small characteristic path length like random graph. During the evolution, the changing tendencies of *E[H]* in those graphs are relatively moderate and largely depend on *LPN*. A large enough *LPN* is still able to further decrease *E[H]*.

For both classes, we observe that different link attachment schemes do not apparently influence the changing tendency of *E[H]*.

◆ Clustering coefficient

The clustering coefficient in lattice graph, line graph and ring graph are all zero due to its topology. The clustering coefficient in random graph approaches zero since the connection probability p is small in real cases. We mainly discuss the following two types. In scale free graph, C_G is still small but is larger than that in random graph of the same size. The value of C_G in small world model is high since the link rewiring probability is usually set to a small value.

Only if a new node connects *LPN* present nodes which are already interlinked with others in the graph object, the process increases C_G. Otherwise C_G decays as graph grows. *LPN* is usually much smaller than *N* in the evolution process. Therefore, the evolving C_G in scale free and small world model decrease as *N* increases during the evolution process.

◆ Assortativity coefficient

The assortativity coefficient is influenced by compound factors which seem to be complex. Therefore we conclude the evolving assortativity coefficient by the observations of the empirical results.

In general, we find that the graph deviates from the original way it is mixed in initial states after evolution process by either attachment scheme. For neural graphs with $r = 0$, the graphs such as small world model and random graph tend to be disassortative since continuous *LPN*-connected nodes break the original mixing way. And assortative mixing property in lattice graph also deteriorates once the evolution process begins. The feature of hubs makes it disassortative in scale free graph. The evolution process also weakens the feature of interconnection and its property deteriorates.

◆ Betweenness centrality

The node and link betweenness are highly correlated to each other. A direct relation between $E[B_l]$ and $E[H]$ is given by equation (2.9).

We also divide the graph into groups. As for ring graph, line graph, lattice graph and small world model, shortcuts established by newly added links largely decrease the variance of the betweenness distribution during the evolution process. More nodes tend to sustain some amount of traffic load. By contrast, the importance of nodes which have high betweenness in initial state is relatively weakened after evolution.

For the scale free graph, the situation is opposite. And for random graph, the situation lies between those two situations above.

In summary, the parameter *LPN* together with the number of nodes *N* and links *L* affects the properties evolved in a growing graph directly. A large value of *LPN*

Regarding graph models, the evolving properties of them vary from type to type. The scheme of link attachment leads to some discrepancies in properties as well.

Regarding real networks, we conclude that evolving properties of real networks are partially combined with known attributes in some typical graph models. Therefore the evolving properties and a real network itself [18] can be partially described by combination of known attributes of related graph models. In research of different types of real networks, we make use of this method to deduce the topological structure of real network. Further, we analyze their evolving properties by derived properties of particular graph models. We find that a real network can not be only modeled by a pure type of graph model. In addition, some evolving properties of real network still fail to be explained by combination of known attributes in different models base on simulation results.

6.2. Future work

We select limited graph models and topological measures in this thesis. More graph models topological measures may be added in the future work.

In the project, we set link density low when we construct graph models. Thus, in some condition, some topological measurement results such as clustering coefficient are no satisfied. We may modify some settings such as *LPN* to establish more proper models

Regarding the growth process, we build up a mechanism in chapter 3 to deal with the situation when *LPN* is not integer. That mechanism introduces some perturbation into the simulation. We should try to find a better method to minimize the perturbation in simulation.

Due to the processing capacity of device, we restrict the graph size N to less than a thousand. In reality, many networks have much larger size. We can expand the size of graph in the simulation and discover a more efficient algorithm.

A real network may be partially assumed as a combination of different types of graphs. But there are still some evolving properties that can not be explained. We should explore the unknown part and try to describe a real network by a better expression.

Bibliography

[1] Reka Albert and Albert-Laszlo Barabasi, "*Statistical Mechanics of Complex Networks*", Reviews of Modern Physics, Volume 74, p.47-94, January 2002

[2] L.da F.Costa, F.A. Rodrigues, G. Travieso and P. R. Villas Boas, "*Characterization of Complex Networks: A Survey of Measurements*", Advanced in Physics, Volume 556, Issue 1, p. 167 – 242, January 2007 ,

[3] P. Van Mieghem, H. Wang, C. Doerr, J. Martin Hernandez, D. Hutchison, and M. Karaliopoulos, "*A Framework for Network Robustness*", Internal report

[4] Sergei Maslov, Kim Sneppen and Alexei Zaliznyak, "*Detection of Topological Patterns in Complex Networks: Correlation Profile of the Internet*", Physica A, Volume 333, p. 529-540, 2004

[5] A. Jamakovic, S. Uhlig, and I. Theisler, "*On the Relationships between Topological Metrics in Real-world Networks*", Proc. of the 4th European Conference on Complex Systems (ECCS'07), Dresden, Germany, October 2007

[6] M.E.J. Newman "*Assortative Mixing in Networks*", Physical Review Letter 89, p.208701, 2002

[7] Priya Mahadevan, Dmitri Krioukov, Marina Fomenkov, Bradley Huffaker, Xenofontas Dimitropoulos, kc claffy and Amin Vahdat "*The Internet AS-Level Topology: Three Data Sources and One Definitive Metric*", ACM SIGCOMM Computer Communication Review (CCR), v.36, n.1, p.17-26, 2006

[8] P. Erdos and A. Renyi, "*On the Evolution of Random Graphs*", Publication of the Mathematical Institute of the Hungarian Academy of Sciences, 5: 17-61, 1960

[9] R. Albert and A.-L. Barabási, "*Emergence of Scaling in Random Networks*", Science, Vol. 286, 509, 1999

[10] Reka Albert, Hawoong Jeong, Albert-Laszlo Barabasi, "*Error and Attack Tolerance of Complex Networks*", Nature, Vol. 406, 27 July 2000

[11] S. Milgram, "*The Small World Problem*", Psychology Today, 1, 61, 1967

[12] Duncan J. Watts and Steven H. Strogatz, "*Collective Dynamics of 'Small-World'*

Networks", Nature Vol 393, 4 June 1998

[13] Alina Beygelzimer, Geoffrey Grinstein, Ralph Linsker and Irina Rish, "*Improving Network Robustness by Edge Modification*", Physica A, Volume 333, p. 529-540, 2005

[14] P. Van Mieghem, "*Performance Analysis of Communications Networks and Systems*", Cambridge University Press, 2006

[15] Jeremy G. Siek, Lie-Quan Lee and Andrew Lumsdaine, "*Boost Graph Library: The User Guide and Reference Manual*", Addison-Wesley Professional, 2001

[16] R. Van der Hofstad, G. Hooghiemstra and P. Van Mieghem, 2005, "*Distances in Random Graphs with Finite Variance Degrees*", Random Structures and Algorithms, Vol. 27, No. 1, p. 76-123, August, 2005

[17] G. Hooghiemstra, and P. Van Mieghem, *"On the Mean Distance in Scale Free Graphs"*, Methodology and Computing in Applied Probability (MCAP), Vol. 7, p. 285-306, 2005

[18] R. Hekmat, *"Ad-hoc Networks: Fundamental Properties and Network Topologies"*, Springer, October 2006.