



Faculty of Electrical Engineering, Mathematics and Computer Science
Network Architectures and Services (NAS) Group

The Strength of Preference

Yuan Ren

1386441

Committee members:

Supervisor: Prof. dr. ir. P. F. A. Van Mieghem
Mentor: Huijuan Wang
Member: dr. ir. Fernando A. Kuipers
dr. ir. Anthony Lo
dr. Frank Vallentin

Date: 21 July 2009

MSc. Thesis No: PVM 2009 – 058

Copyright ©2009 by Y. Ren

All rights reserved. No part of the material protected by this copyright may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without the permission from the author and Delft University of Technology.

The Strength of Preference

THESIS

submitted in partial fulfillment of the
requirements for the degree of

Master of Science

In

Computer Engineering

By

Yuan Ren



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

Acknowledgment

My master thesis is done at the Network Architectures and Services (NAS) Group, Faculty Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, where I worked from November 2008 to July 2009. During the period, many people gave me their suggestions and help. I would like to express my acknowledgment to them.

First of all, I would like to express my appreciation to my supervisor Prof. dr. ir. P. F. A. Van Mieghem and my daily mentor ir. Huijuan Wang. Without your guide and help, I could not accomplish my thesis. I really enjoy the discussing time with you, which can always bring me new ideas and important suggestions. I also would like to thank ir. Javier M. Hernandez. Your clear explanation solved my problems, thus I can continue the analysis.

Secondly, I would like to thank my friends. When I was in trouble and felt upset, it's you to cheer my up and encouraged me to move on.

Finally, I would like to express my special thanks to my parents. Without your support, I could not reach here.

This thesis is dedicated to my parents.

Abstract

Growth and preferential attachment are the two ingredients of the scale-free network. Based on the Barabási-Albert scale-free model, we construct a more general model by replacing the linear preferential attachment with the nonlinear preferential attachment. We introduce different networks by controlling a parameter β which decides the preferential attachment of our model. We try to find the influence of β on the network structure and property.

To study the influence, we investigate our model in three directions: topological characteristics, correlation of topological measures and attack vulnerability.

In topological characteristics section, we specify the structural measures and the spectral measures to analyze the influence of β . These measures are the degree distribution, the average hopcount, the average clustering coefficient, the assortativity coefficient and the eigenvalues of the adjacency matrix and the Laplacian matrix.

In correlation section, we calculate the correlation coefficients between the four measures: the algebraic connectivity, the spectral radius, the average clustering coefficient and the average hopcount to study their relationship.

In attack vulnerability section, we use two kind node attack strategies to attack the networks based on the cluster-focused attack and the global attack respectively. Then we analyze the size of the largest cluster, the spectral radius and the algebraic connectivity of the largest cluster in the network.

Key words: complex network, growth, preferential attachment, topological characteristics, correlation, attack vulnerability

Table of Contents

Acknowledgment	5
Abstract	7
1. Introduction	13
1.1 Motivation.....	13
1.1.1 Model description	13
1.1.2 Study direction	13
1.2 Graph measures.....	15
1.2.1 Structural measures	15
1.2.2 Spectral measures	16
1.3 Complex network.....	16
1.3.1 Erdős-Rényi (ER) random graph	17
1.3.2 Small-world graph model	17
1.3.3 Scale-free graph model	18
2. Model	19
2.1 Model introduction.....	19
2.2 Model construction	19
3. Topological characteristics in relation to β	21
3.1 Degree distribution $\Pr[D = k]$	21
3.1.1 Degree distribution for a given β	21
3.1.2 Degree distribution in relation to β	24
3.2 Average hopcount H	28
3.3 Average clustering coefficient c	29
3.4 Eigenvalues λ_i of the adjacency matrix	32
3.4.1 Eigenvalues λ_i of the adjacency matrix	32
3.4.2 Spectral radius λ_1	34
3.5 Eigenvalues μ_i of the Laplacian matrix	35
3.5.1 Eigenvalues μ_i of the Laplacian matrix	35
3.5.2 Algebraic connectivity μ_{N-1}	37
3.6 Assortativity coefficient r	38
3.7 Summary	40

4. Correlation of topological measures in relation to β	43
4.1 Definition	43
4.2 Analysis	44
4.4 Summary	47
5. Attack vulnerability in relation to β	49
5.1 Introduction	49
5.2 Related work	49
5.3 Attack model	50
5.3.1 Attack object	50
5.3.2 Attack strategy	50
5.3.3 Analysis measurement	51
5.4 Analysis of the largest cluster	51
5.4.1 The size of the largest cluster	51
5.4.2 Spectral radius of the adjacency matrix	53
5.4.3 Algebraic connectivity of the Laplacian matrix	54
5.4.4 Comparison between the cluster-focused attack and the global attack	56
5.5 Summary	56
6. Conclusions and Future work	59
6.1 Conclusions	59
6.2 Future work	60
Appendix	61
A. Graphs with given node connectivity	61
B. Results from linear algebra	62
C. Figures	63
C.1 Topological measures	63
C.2 Simulation results of the network under the global attack	64
References	67

Table of Figures

Figure 2.1 The connectivity of nodes in $G^*(m=3, N=20, \beta \rightarrow +\infty)$	20
Figure 3.1 The degree distribution of $G^*(m=3,5, N=200,400,800, \beta=1)$	22
Figure 3.2 The degree distribution of $G^*(m=3,5, N=200,400,800, \beta=0)$	23
Figure 3.3 The degree distribution of $G^*(m=3,5, N=200,400,800, \beta=-1)$	24
Figure 3.4 The degree distribution of $G^*(m=3, N=800, \beta=-8,-4,-2,-1,0,1,2)$	26
Figure 3.5 The curve fitting of the degree distribution of $G^*(m=3, N=800, \beta=1)$	27
Figure 3.6 The curve fitting of the degree distribution of $G^*(m=3, N=800, \beta=0)$	27
Figure 3.7 The average hopcount distribution of $G^*(m=3, N=800, \beta=-8,-4,-2,-1,0,1,2)$	28
Figure 3.8 The mean value of the average hopcount of $G^*(m=3, N=800)$	29
Figure 3.9 The average clustering coefficient distribution of $G^*(m=3, N=800, \beta=2)$.30	
Figure 3.10 The average clustering coefficient distribution of $G^*(m=3, N=800,$ $\beta=-8,-4,-2,-1,0,1)$	31
Figure 3.11 The mean of the average clustering coefficient of $G^*(m=3, N=800)$	31
Figure 3.12 The spectrum distribution of the adjacency matrix of $G^*(m=3, N=800,$ $\beta=2)$	33
Figure 3.13 The spectrum distribution of the adjacency matrix of $G^*(m=3, N=800,$ $\beta=-8,-4,-2,-1,0,1)$	33
Figure 3.14 The curve fitting of the spectrum tail part of $G^*(m=3, N=800, \beta=1)$	34
Figure 3.15 The spectral radius distribution of $G^*(m=3, N=800, \beta=-8,-4,-2,-1,0,1,2)$	35
Figure 3.16 The spectrum distribution of the Laplacian matrix of $G^*(m=3, N=800,$ $\beta=2)$	36
Figure 3.17 The spectrum distribution of the Laplacian matrix of $G^*(m=3, N=800,$ $\beta=-8,-4,-2,-1,0,1,2)$	36
Figure 3.18 The algebraic connectivity distribution of $G^*(m=3, N=800,$ $\beta=-8,-4,-2,-1,0,1,2)$	37
Figure 3.19 The average of the algebraic connectivity of $G^*(m=3, N=800)$	38
Figure 3.20 The assortativity coefficient distribution of $G^*(m=3, N=800,$ $\beta=-8,-4,-2,-1,0,1,2)$	40

Figure 3.21 The average of measures of $G^*(m=3, N=800)$	41
Figure 4.1 the relationship between correlation coefficient $corr$ and β	46
Figure 5.1 The size of the largest cluster in relation to the number of the cluster-focused random failures	52
Figure 5.2 The size of the largest cluster in relation to the number of the cluster-focused degree-based attack	52
Figure 5.3 The spectral radius of the largest cluster in relation to the number of the cluster-focused random failures	53
Figure 5.4 The spectral radius of the largest cluster in relation to the largest cluster size when the network is under the cluster-focused degree-based attack	54
Figure 5.5 The algebraic connectivity of the largest cluster in relation to the largest cluster size when the network is under the cluster-focused random failures	55
Figure 5.6 The algebraic connectivity of the largest cluster in relation to the largest cluster size when the network is under the cluster-focused degree-based attack	55
Figure C.1 The curve fitting of the degree distribution on ccdf of $G^*(m=3, N=800, \beta=1)$	63
Figure C.2 The average of the spectral radius of $G^*(m=3, N=800)$	63
Figure C.3 The average of the assortativity coefficient of $G^*(m=3, N=800)$	63
Figure C.4 The size of the largest cluster in relation to the number of the global random failures	64
Figure C.5 The spectral radius of the largest cluster in relation to the largest cluster size when the network is under the global random failures	64
Figure C.6 The algebraic connectivity of the largest cluster in relation to the largest cluster size when the network is under the global random failures	64
Figure C.7 The size of the largest cluster in relation to the number of the global degree-based attack	65
Figure C.8 The spectral radius of the largest cluster in relation to the largest cluster size when the network is under the global degree-based attack	65
Figure C.9 The algebraic connectivity of the largest cluster in relation to the largest cluster size when the network is under the global degree-based attack	65
Figure C.10 The algebraic connectivity of the largest cluster in relation to the number of global degree-based attack	66

Introduction

Robustness of a network measures the ability of the network withstanding the changes of the network structure. The study on the robustness of real world networks can give us some insights on how to construct and maintain a relatively stable network in the future.

1.1 Motivation

1.1.1 Model description

Barabási and Albert first used the two network features: growth and preferential attachment to construct a network model named Barabási-Albert (BA) model. Growth means the number of nodes in a network is continuously growing; preferential attachment decides which existing nodes will have relationship with new adding nodes. These two features are already found in most real world networks.

We construct a general model based on the BA model. In our model, the preferential attachment is expanded from the linear one of the BA model to the nonlinear one of ours. We control the preferential attachment by using a parameter β . Thus different networks can be introduced by different values of this parameter.

1.1.2 Study direction

The network features can tell us the structure and the property of a network. It also can tell us which features could be good measures for the robustness of the network. We investigate the features of our model in the following three directions:

- Topological characteristics

Network structure is the base of a network. It decides the function of the network. The study on topological characteristics is to find the inherent structural property of

the network and the relationship between the structural property and the network function. For example, what topological property is related to the robustness of a network against the decomposition; what topological property has relationship with the information propagation and virus spreading of the network. After knowing this, we can go further to study how to enhance the robustness of the network, make the information propagation easier or let the virus spreading more difficult. Besides this, we also can use the inherent structural property to classify the networks. For example, the degree distribution of most real world networks follows a power law, which means the degree distribution has no relationship with the network size. These networks are classified into a network family named the scale-free network. The classification of networks makes the network study more effective.

In this paper, we study the influence of the parameter β on the structural property of our model through the following measures: the degree distribution, the average hopcount, the average clustering coefficient, the eigenvalues of the adjacency matrix and the Laplacian matrix (see the definition of these measures in Chapter 1.2).

- Correlation of topological measures

A network can be represented by a set of topological measures. In the set, some measures are highly correlated to others. In other words, the correlation coefficient between two measures is big. The set of measures to characterize a network can be reduced by choosing the measures that are relatively independent from each other. Thus, we deem it crucial to understand the correlation between topological measures. Correlation indicates the linear relationship between two measures and it strongly relies on the network topology. Using the correlation coefficient, we observe the change of the correlations between any pair of measures when the network structure changes along with the parameter β .

Based on the simulation results obtained in topological characteristics section, we calculate the correlation coefficients between any pair of the four topological measures: the algebraic connectivity, the spectral radius, the average clustering coefficient and the average hopcount to investigate the influence of β on the correlations between these measures (see the definition of these measures in Chapter 1.2).

- Attack vulnerability

Attack means the removal of nodes or links from the network. After attacking, the network may segment into several disconnected clusters. The study on attack vulnerability is a way to measure the robustness of the network. Different network structures will have different behaviors when the network is under different attack

strategies. Through the study we hope to know how to build a robust network against attack in the future.

In this paper, we examine the behaviors or the characteristics of different structures of our model when subject to various node removal attacks.

1.2 Graph measures

Graph measures are numerical expressions of the network topological property. This paper is focused on the study of the network property, so we use the graph measures to investigate the network structure and the network function of our model.

A graph G is denoted as $G = (N, L)$, where N is the number of nodes in the network, and L is the number of links. A complete graph or full mesh $G = K_N$ is a graph which has a link between every pair of nodes. This graph consists of N nodes and $L = \frac{N(N-1)}{2}$ links.

In a network the links can be weighted and directed. The link weight shows how strong the relationship two nodes have, and the direction gives the information about which is the source node and which is the destination node. In this paper we only consider the unweighted and undirected networks, which contain no self-loops.

1.2.1 Structural measures

- *Degree*: The degree d_i of a node i is the number of direct neighbors that the node i has in the network. The degree d_i satisfies the inequality $0 \leq d_i \leq N-1$.
- *Shortest path*: In the unweighted networks, the shortest path is the minimum number of hops that a source node needs to reach a destination node.
- *Hopcount*: The hopcount H is originally defined as the number of hops on the path between a source node and a destination node. Since in a network there may be several paths between a pair of nodes, so in this paper we define the hopcount as the number of hops on the shortest path between any pair of nodes. The average hopcount in a network is the average value of the shortest paths between all pairs of nodes.
- *Diameter*: Diameter H_{max} is the largest hopcount in the network.
- *Clustering coefficient*: The clustering coefficient characterizes the connection density in the neighborhood of a node i . It is defined as the ratio of the number of links y connecting the neighbors of node i over the total possible links $d_i(d_i-1)/2$ [22]:

$$c_i = \frac{2y}{d_i(d_i - 1)}$$

The average clustering coefficient is the average value of the clustering coefficients of all nodes in the network.

1.2.2 Spectral measures

- *Adjacency matrix*: The adjacency matrix A is a symmetric $N \times N$ matrix. The diagonal element of the matrix $a_{ii} = 0$ and the nondiagonal element a_{ij} is 1 or 0 depending on whether the two nodes are connected.

$$a_{ij} = \begin{cases} 0 & \text{if } i = j \\ 1 & \text{if } i \neq j \text{ and node } i \text{ is adjacent to node } j \\ 0 & \text{otherwise} \end{cases}$$

The set of all N adjacency eigenvalues $\lambda_N \leq \lambda_{N-1} \leq \dots \leq \lambda_1$ is called the adjacency spectrum of the graph G . The largest eigenvalue λ_1 of the adjacency matrix is defined as the *spectral radius*.

- *Laplacian matrix*: The Laplacian matrix Q is a symmetric $N \times N$ matrix. The diagonal element of the matrix q_{ii} is the degree d_i of node i and the nondiagonal element q_{ij} is -1 or 0 depending on whether the two nodes are adjacent.

$$q_{ij} = \begin{cases} d_i & \text{if } i = j \\ -1 & \text{if } i \neq j \text{ and node } i \text{ is adjacent to node } j \\ 0 & \text{otherwise} \end{cases}$$

The set of all N Laplacian eigenvalues μ_i is called the Laplacian spectrum of a graph G , and they are all real and nonnegative [5], which means the eigenvalues satisfy $\mu_N = 0 \leq \mu_{N-1} \leq \dots \leq \mu_1$. The second smallest eigenvalue μ_{N-1} of the Laplacian matrix is denoted as *algebraic connectivity*.

1.3 Complex network

Many real world networks, such as the World Wide Web, biological networks, social networks, business networks etc, are complex networks which are networks with substantial non-trivial topological features. For example, the human web has the small shortest path and the high clustering coefficient features. The citation network has the power-law degree distribution feature.

Until now, several models have been introduced to capture the features of real

world networks. Among them, the mostly studied models are the Erdős-Rényi random graphs, the small-world model and the scale-free model.

1.3.1 Erdős-Rényi (ER) random graph

The Erdős-Rényi (ER) random graph [2, 15, 21] may be the earliest and most investigated random network model. An Erdős-Rényi (ER) random graph with N nodes can be constructed in the following steps: first we begin with N nodes. Then a link exists between each node pair with the probability p . We denote the random graph by $G_p(N)$. The ER random graphs have a small average hopcount and a low clustering coefficient.

The degree D of an arbitrary node in the ER model has a binomial distribution:

$$\Pr[D = k] = \binom{N-1}{k} p^k (1-p)^{N-1-k} \approx \frac{z^k e^{-z}}{k!} \quad (1)$$

where z is the average node degree of the graph: $z = E[D] = (N-1)p$. In (1) the second term is the Poisson approximation for large N . So the degree distribution of the ER random graph has an exponential tail.

One greatest discovery of the ER random graph is that there exists a critical link density $p_c = \ln N / N$. When $p < p_c$, the graph consists of disconnected clusters; when $p > p_c$, a large cluster contains almost all nodes in the graph.

1.3.2 Small-world graph model

The small-world graph model is a graph which has a large number of nodes and a relatively short path between any pair of nodes. The small-world property can be captured by the small average hopcount and the high clustering coefficient. It has been demonstrated that many real world networks have such small-world property, such as social networks, gene networks etc. One famous theory of the small-world property is “Six degrees of separation” (also referred to as the “Human Web”). It refers to the idea that anyone can connect to any other on the Earth through at most six persons.

Watts and Strogatz (WS) model [16] is a random graph with the small-world property. The construction of the WS model follows the random rewiring procedure. We start from a ring lattice with N nodes and L links per nodes. Ring lattice is a network with N nodes placed on a circle structure. There are links between each node and its nearest and next-nearest neighbors. Then we randomly rewire each link with the probability p_r . When $p_r = 1$ the graph is a random graph and when $p_r = 0$ the graph is a regular lattice.

The degree distribution of the WS model depends on p_r . When $p_r \rightarrow 1$ the degree distribution is Poisson distribution, so it also has an exponential tail similar as the ER model.

1.3.3 Scale-free graph model

A scale-free graph is a graph whose degree distribution follows a power law with an exponent independent of the graph size N for large N . Hence the probability of a node in the graph having a degree k is $\Pr[D=k] \sim k^{-\gamma}$, where k is the node degree, γ is the scaling exponent and is typically in the range $2 < \gamma < 3$. Many real world networks appear to be scale-free, such as the World Wide Web, some social networks etc.

The first model introduced as a scale-free network is the Barabási-Albert (BA) model [1, 2]. The BA model begins with m_0 nodes. Then at each time step we add a new node with m links which will be connected to the m ($m \leq m_0$) different nodes which are already present in the network. The m different nodes are chosen by the preferential attachment, which indicates the probability Π of a node to be picked out depends on the degree of this node:

$$\Pi(d_i) = \frac{d_i}{\sum_j d_j}$$

where d_i is the degree of node i , $\sum_j d_j$ is the total degree in the network.

The degree distribution of the BA model follows a power law $\Pr[D=k] \sim k^{-\gamma}$ with $\gamma = 3$ [2] and is independent of m with larger N . The BA model indicates the "rich-get-richer" phenomenon [1] that the nodes with high degree have a large chance to be attached with new adding nodes. In real world, there are some examples supporting the feature, such as the citation patterns of the scientific publications with the exponent $\gamma = 3$ [9], the movie actors collaboration network with $\gamma = 2.3 \pm 0.1$ [1, 2] etc.

The clustering coefficient of the BA model follows approximately a power law $C \sim N^{-0.75}$. The larger the N is, the smaller the clustering coefficient C is.

Model

2.1 Model introduction

The Barabási-Albert model is the first model that uses preferential attachment to construct a network. Since the BA model was proposed, many researchers have studied this model. Some similarity between the BA model and the real world network has been found, such as the power law distribution, the “hub” which gathers many links from other nodes etc. Then P.L.Krapivsky and S.Redner [3, 4] raised an extensional model on the ground of the BA model. Replace the linear preferential attachment of the BA model with the nonlinear one $\Pi(d_i) = \frac{d_i^\beta}{\sum_j d_j^\beta}$ to construct the networks, where β changes from 0 to $+\infty$.

Until now these models all only consider the positive β . We are wondering what the network structure and property will be if β is negative. Thus we have our model that uses the nonlinear preferential attachment to construct the networks:

$$\Pi(d_i) = \frac{d_i^\beta}{\sum_j d_j^\beta}$$

where β is in the range $[-\infty, +\infty]$. The parameter β is named as *the strength of preference*. Given different values of β , we will have different network topologies.

2.2 Model construction

Our model $G^*(m, N, \beta)$ is constructed in the following steps, based on growth and preferential attachment:

- 1). Start from a complete graph with m nodes.
- 2). At every time step we add a new node with m links which will be attached to m different nodes which are already existed in the network. A network constructed like this has been proved to be *m -connected* (see the proof in appendix A). The m different

existed nodes are decided by the preferential attachment:

$$\Pi(d_i) = \frac{d_i^\beta}{\sum_j d_j^\beta}, \text{ where } \beta \in (-\infty, \infty)$$

3). Repeat step 2 until the total number of nodes in the graph is N .

Through the preferential attachment rule we can get a rough tendency of probability Π along with β . When $\beta > 0$, the larger the degree of a node is, the higher the probability it has to be connected to. On the contrary, when $\beta < 0$, the smaller the degree of a node is, the higher the probability it has to be attached to.

Obviously when $\beta = 1$ the network is the Barabási-Albert scale-free network, in which the probability Π is linearly related to the node degree. When $\beta = 0$, it is a network called growing exponential network [7], in which all the existing nodes have the same probability to be chosen. In other words, the nodes to be connected to are picked out randomly. When β approaches to $+\infty$, the network is unique. All the new adding nodes will be attached to the first m nodes. Fig.2.1 is an example of our model when $\beta \rightarrow +\infty$. In this figure, the bigger nodes are the original m nodes and the smaller nodes are the newly added nodes.

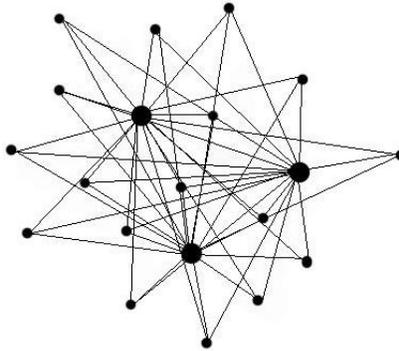


Figure 2.1 The connectivity of nodes in G^* ($m=3, N=20, \beta \rightarrow +\infty$)

3

Topological characteristics in relation to β

Topological property is the inherent property of a network. Through the study of the structural and spectral measures, we can capture the topological characteristics of a network. In this section, we carry out the simulations to calculate the structural measures and spectral measures of our model when $\beta = -8, -4, -2, -1, 0, 1, 2$. Then we analyze these measures to see the influence of β on the network structural features. For each class of $G^*(m, N, \beta)$ we iterate 10000 times to build 10000 different networks to obtain good statistical property of the class of graphs.

3.1 Degree distribution $\Pr[D = k]$

3.1.1 Degree distribution for a given β

Before our analysis on different preference strength β , we investigate the degree distribution of G^* with fixed β , but tunable m and N .

3.1.1.1 $\beta = 1, m = 3, 5, N = 200, 400, 800$

The class of graphs with $\beta = 1$ is the same as the BA model with $m_0 = m$.

It has been demonstrated that the degree distribution of the Barabási-Albert model has a power-law tail $\Pr[D=k] \sim k^{-\gamma}$ with the scaling exponent $\gamma = 3$ independent of m when N approaches to infinite [1, 2]. So the degree distribution of our model with $\beta = 1$ should also follow a power law and have an exponent close to 3. As Fig.3.1 shown, the curves have a power-law tail and the exponent γ has nearly no relationship with m since $\gamma = 2.61$ is hold by the five graphs.

According to the research result of A.L.Barabási and R.Albert, the exponent γ of the BA model is 3 which is large than our model 2.61. Two reasons can explain the bias. First, it is caused by the influence of N on γ . The larger the N is, the closer the γ is to 3. In our model the largest N is 800, if we let N be larger, the exponent γ will be closer to 3. Second, the exponent γ is obtained directly from the curve fitting on the

probability distribution function (pdf) of the degree. Actually, the pdf is less precise than the ccdf to capture the power-law behavior of the degree [21]. The ccdf is the complementary cumulative distribution function. It is defined as $F(x) = \Pr(X > x) = 1 - F(x)$. The exponent γ' of pdf can be computed from the exponent ν of ccdf by the equation $\gamma' = \nu + 1$. We use ccdf to calculate the degree distribution and get $\nu = 1.82$ (See the curve fitting on ccdf in Fig.C.1). Thus the computed exponent γ' of the degree distribution is $\gamma' = \nu + 1 = 2.82$, which is closer to 3 than the one obtained directly by the curve fitting on the pdf.

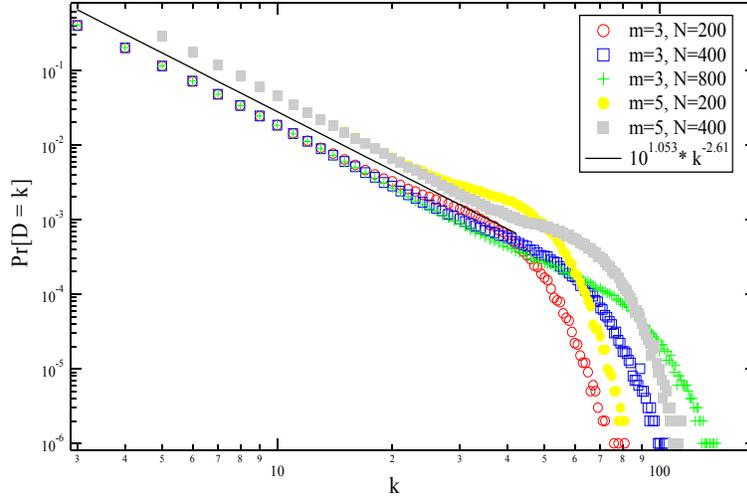


Figure 3.1 The degree distribution of $G^*(m=3,5, N=200,400,800, \beta=1)$

3.1.1.2 $\beta = 0, m = 3,5, N = 200,400,800$

The attachment rule of the graph with $\beta = 0$ can be rewritten as $\Pi(d_i) = \frac{1}{\sum_j d_j^0}$,

where i, j are the node i and j in the network, and $\sum_j d_j^0$ is the total number of nodes in the present network. It indicates that the probability of a node to be chosen is independent of the node degree. For all the nodes in the present network, the probability Π are the same. In other words, the nodes to be attached are chosen randomly.

Since A.L.Barabási and R.Albert proposed the Barabási-Albert model, they have done some simulation to demonstrate that the scale-free feature only presents when $\beta = 1$ [1]. And it has been proved that when $\beta = 0$ this model is not a scale-free network because of the absence of preferential attachment [1, 2].

The graph with $\beta = 0$ is named as growing exponential network [7]. It has been explained that the degree distribution of the model when $\beta = 0$ is exponential [7]. And further, the numerical representation of the degree distribution has been derived as $\Pr[D = k] = \frac{e}{m} \exp(-\frac{k}{m})$ by using the mean-field approximation [2]. It indicates that the exponent of the degree distribution tail is inversely proportional to m . As Fig.3.2 shown, the larger the m is, the smaller the exponent of the degree distribution is, which matches the analysis.

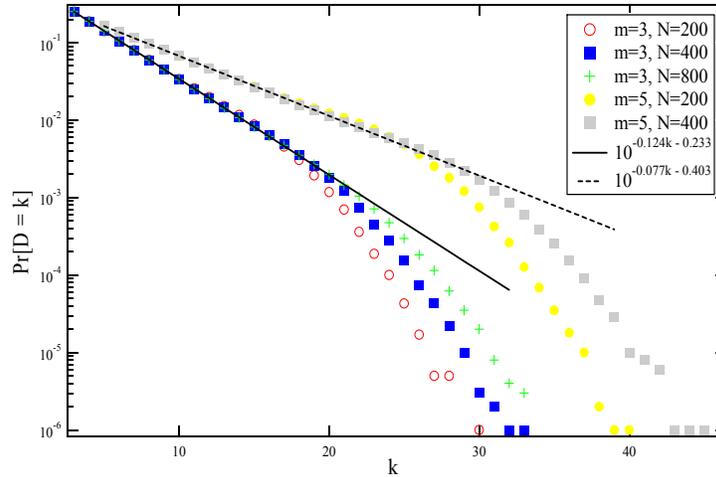


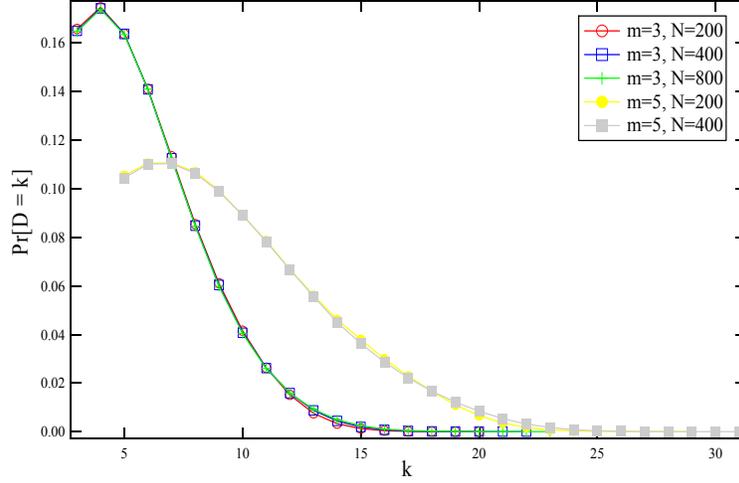
Figure 3.2 The degree distribution of $G^*(m=3,5, N=200,400,800, \beta=0)$

When $\beta = 0$ and $m = 1$, our model becomes the uniform recursive tree (URT). A URT is a random tree structure, in which any two nodes are connected by exactly one path. When constructing the URT, at each time step, a new node is attached uniformly to one existed node in the network. The degree distribution of the URT has an exponential tail and the exponent of the tail is $\ln 2$ [22].

3.1.1.3 $\beta = -1, m = 3,5, N = 200,400,800$

Different from the model with $\beta = 1$, when $\beta = -1$, the smaller degree of a node is, the larger chance it has to connect to a new adding node. Because in this network the nodes with the small degree m take the majority, so approximately we can think that the nodes to be connected to are chosen randomly from the most nodes with the small degree m .

The simulation results of the model with $\beta = -1$ are shown in Fig.3.3. The degree distribution does not follow a power law or have an exponential tail. A surprising phenomenon is that the networks with the same m have nearly the same degree distribution, no matter what N is. This indicates that the degree distribution is mainly related to m , and is independent of N .


 Figure 3.3 The degree distribution of $G^*(m=3,5, N=200,400,800, \beta=-1)$

In this section, we studied our model on fixed β and tunable m and N . Through the simulation results, we find that the degree distribution is always independent of N as long as N is large. When $\beta = -1, 0$, the degree distribution is related to m and when $\beta = 1$, the degree distribution is independent of m .

3.1.2 Degree distribution in relation to β

3.1.2.1 Related work with $\beta > 0$

P.L.Krapivsky and S.Redner have studied the nonlinear form $\Pi(k) \sim k^{-\beta}$ of the probability that a node is attached to a newly added node, where β is tunable in the range $[0, +\infty]$ [3, 4]. The model used by them is a specified one of our model that $m_0 = m = 1$ and $\beta \geq 0$. A graph constructed like this is in fact a tree structure. A tree is a graph in which any two nodes are connected by exactly one path. In other words, a tree has $N-1$ links and no loops exist in a tree.

They used rate-equation approach
$$\frac{dN_k}{dt} = \frac{1}{M_\beta} \left[(k-1)^\beta N_{k-1} - k^\beta N_k \right] + \delta_{k1} \quad (2),$$

where $M_\beta = \sum k^\beta N_k(t)$, to study the degree distribution $N_k(t)$ with time evolution. The degree distribution $N_k(t)$ is defined as the average number of nodes which have degree k at time step t . The first term $((k-1)^\beta N_{k-1}) / M_\beta$ in (2) counts for the probability that a node with degree $k-1$ to be connected to a new node and then its degree increases to k . The second term $(k^\beta N_k) / M_\beta$ is the probability of a node degree increasing from k to $k+1$. The term M_β is the β moment of $N_k(t)$ and it provides the proper normalization. The term δ_{k1} accumulates the continuous introduction of new nodes which do not have any incoming links but only one outgoing link.

P.L.Krapivsky and S.Redner divided the values of β into two parts to study the model separately. They are $0 < \beta < 1$ and $\beta > 1$.

1). Sublinear case for $0 < \beta < 1$.

In this regime, M_β and N_k grow linearly with time, satisfying $M_\beta = \mu t$ where $1 \leq \mu \leq 2$ and $N_k = n_k t$. Substitute M_β and N_k into Eqs.(2) we have the degree distribution $\Pr[D = k] = \frac{\mu}{k^\beta} \prod_{j=1}^k (1 + \frac{\mu}{j^\beta})^{-1}$, whose asymptotic behavior is

$$\Pr[D = k] \sim \begin{cases} k^{-\beta} \exp[-\mu(\frac{k^{1-\beta} - 2^{1-\beta}}{1-\beta})] & \frac{1}{2} < \beta < 1 \\ k^{(\mu^2 - 1/2)} \exp[-2\mu\sqrt{k}] & \beta = \frac{1}{2} \\ k^{-\beta} \exp[-\mu\frac{k^{1-\beta}}{1-\beta} + \frac{\mu^2}{2}\frac{k^{1-2\beta}}{1-2\beta}] & \frac{1}{3} < \beta < \frac{1}{2} \end{cases} \quad (3)$$

As shown in Eqs.(3) whenever β decreases below a value $1/x$, where x is an arbitrary positive integer, an additional term arises exponentially in $\Pr[D=k]$.

2). Superlinear case for $\beta > 1$.

When $\beta > 1$ the network exhibits a “winner takes all” phenomenon: a single node will get almost all the links from other nodes. From Fig.3.4 we can see the phenomenon clearly that most nodes of the model with $\beta = 2$ have the degree m , which means most links are owned by a few nodes.

In this regime there is no analytical solution for degree distribution. But when $\beta > 2$, through a discrete time process we can calculate the probability of the initial

node gathering all links in the network from $\Pi(d_i) = \prod_{j=1}^{\infty} \frac{1}{1 + j^{1-\beta}}$.

3.1.2.2 $\beta \rightarrow +\infty$

In the model when $\beta \rightarrow +\infty$, the m original nodes gather all the links from other nodes in the network, so the degree $d_{i_{\beta \rightarrow +\infty}}$ has only two values:

$$d_{i_{\beta \rightarrow +\infty}} = \begin{cases} N-1 & \text{if the node is one of the original } m \text{ nodes} \\ m & \text{if the node is one of the } N-m \text{ adding nodes} \end{cases}$$

Thus the degree distribution is
$$\begin{cases} \Pr[D = 3] = \frac{N - m}{N} \\ \Pr[D = N - 1] = \frac{m}{N} \end{cases}.$$

3.1.2.3. $\beta = -8, -4, -2, -1, 0, 1, 2$

Through the preferential attachment rule we know that when $\beta < 0$ the probability $\Pr(d_i)$ increases as d_i decreases. Since the newly added nodes often have the smallest degree, so when $\beta < 0$, the existing nodes to be attached to are mostly picked out among these former added nodes. Thus with β decreasing, the existed nodes to be attached can be seen as randomly picked out from the nodes with the smallest degree, so the connections of new adding nodes should scatter in these former added nodes.

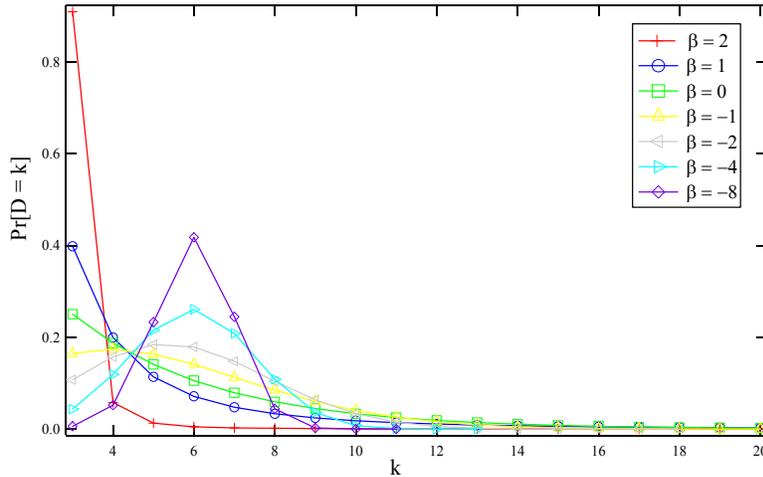


Figure 3.4 The degree distribution of $G^*(m=3, N=800, \beta=-8, -4, -2, -1, 0, 1, 2)$

From the above analysis we can deduce that the degree which corresponds to the peak should increase as β decreases. As Fig.3.4 shown, the simulation results match our analysis. Particularly, it seems the degree possessed by most nodes stays at 6 when $\beta < -2$. In fact, when $\beta \rightarrow -\infty$ our model is a regular graph. A regular graph is a graph in which all the nodes have the same degree. When constructing the network, adding one node with m links into the network will make the total degree of the network increase of $2m$. Thus when N is infinite, the average degree of the network is $2m$. In our model it is 6. So combined with the tendency of the degree distribution shown in Fig.3.4, we can deduce that when $\beta \rightarrow -\infty$ our model is a regular graph.

The model with $\beta = 0$ is a special case that it has no preferential attachment. It have been indicated in [1, 2, 7] that the model when $\beta = 1$ should follow a power law with the exponent $\gamma = 3$, and when $\beta = 0$ the model should have an exponential tail.

The curve fitting results are shown in Fig.3.5 and Fig.3.6 respectively. The exponent of the numerical result of the model with $\beta = 1$ does not exactly equal to 3 but it also can be accepted because the total node number N in our simulation is not big enough and we use the curve fitting on pdf directly to get γ . The result of the model with $\beta = 0$ is in good agreement with [1, 2, 7].

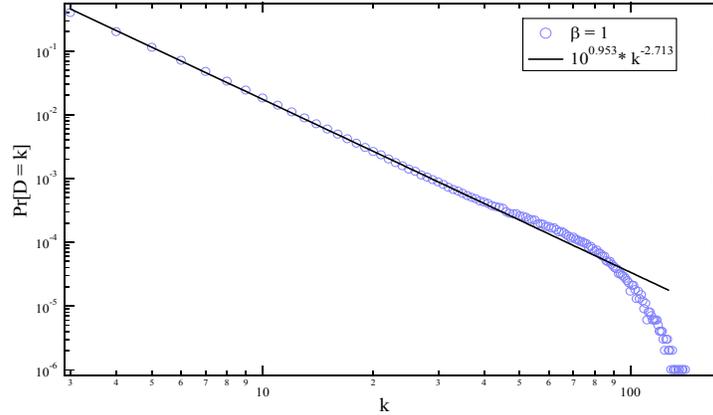


Figure 3.5 The curve fitting of the degree distribution of $G^*(m=3, N=800, \beta=1)$

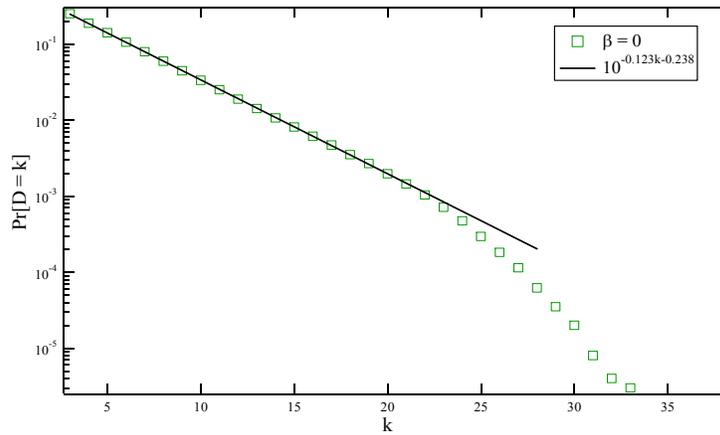


Figure 3.6 The curve fitting of the degree distribution of $G^*(m=3, N=800, \beta=0)$

In this section we studied the degree distribution of our model in relation to β . We observed that when $\beta \geq 0$ the degree possessed by most nodes is m but the percentage of degree m decreases as β decreases; when $-2 \leq \beta < 0$ the degree possessed by most nodes increases as β decreases; when $\beta < -2$ the smaller the β is, the larger the percentage of degree 6 is. Through the trend of the degree distribution when $\beta < -2$ we deduce that our model is a regular graph when $\beta \rightarrow -\infty$.

3.2 Average hopcount H

3.2.1 $\beta \rightarrow +\infty$

In this network, all the newly added nodes are attached to the m original nodes and the m nodes are completely connected to each other. Hence any source node can reach to any destination node in at most 2 hops. When the size N of the network is large, the diameter and the average hopcount of the network are

$$H_{\max, \beta \rightarrow +\infty} = E[H_{\beta \rightarrow +\infty}] = 2.$$

3.2.2 $\beta = -8, -4, -2, -1, 0, 1, 2$

As the analysis in the degree distribution, most links are attached to the m original nodes in these networks when $\beta \gg 1$. These networks can be seen as centralized networks where the m original nodes are the hubs; on the contrary, when $\beta < 0$ most links are attached to the nodes with the smallest degree which takes the majority in the network, so these networks can be regarded as distributed networks. Thus when β decreases from positive to negative, our model changes from centralized to distributed. This makes the number of hops needed to connect any two nodes increase accompanied by β decreasing from positive to negative.

We calculated the average of the hopcount between each pair of nodes in the network, thus for 10000 networks we have 10000 average hopcounts. The simulation result shown in Fig.3.7 indicates that the change of the hopcount agrees with our analysis.

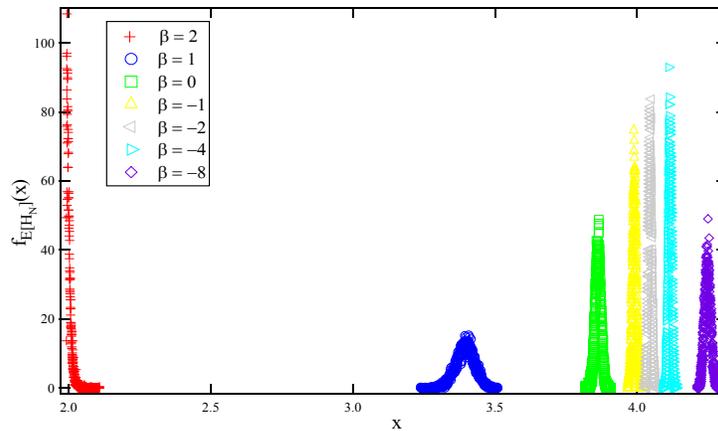


Figure 3.7 The average hopcount distribution of $G^*(m=3, N=800, \beta=-8, -4, -2, -1, 0, 1, 2)$

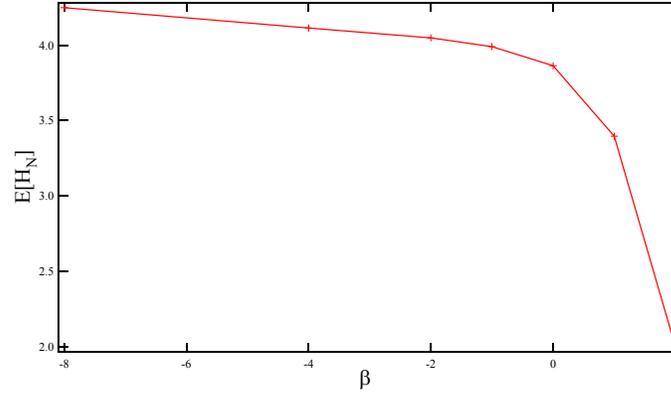


Figure 3.8 The mean value of the average hopcount of $G^*(m=3, N=800)$

The curves in Fig.3.7 show that the average hopcount distribution in G^* follows approximately a Gaussian distribution. Therefore we examine the mean value of the average hopcount of the class $G^*(m, N, \beta)$. The study on the mean value $E[H]$ could make the relationship between the average hopcount and β more clearly and numerically. From Fig.3.8 we find that the mean value of the average hopcount differs slightly when $\beta < 0$, but changes sharply for $\beta > 0$. Especially at $\beta = 2$ the mean value is close to the extreme situation that $E[H] = 2$. When β is larger than 2 and getting larger and larger, the network structure of our model is more and more close to the model when $\beta \rightarrow +\infty$. So combing with our simulation result we can deduce that $E[H] = 2$ for $\beta \geq 2$.

3.3 Average clustering coefficient c

The clustering coefficient represents the connectivity density between the neighbors of a node. It is a local structural property of the network.

3.3.1 $\beta \rightarrow +\infty$

In these networks, there are two types of node degree $d_{i, \beta \rightarrow +\infty}$. So we classify all the nodes into two classes to calculate the clustering coefficient of every node belonging to each class separately, then sum and average them to get the mean value which is the average clustering coefficient $c_{i, \beta \rightarrow +\infty}$ of the network.

- Class1: the m original nodes. Among the neighbors of any one node belong to class1, there are $(m-1) \cdot (N-m)$ links connecting the $m - 1$ original nodes to $N-m$ added nodes, and $(m-1) \cdot (m-2)/2$ link between the $m - 1$ original nodes. So for each node in this class, the clustering coefficient is

$$c_i = \frac{(m-1) \times (N-m) + (m-1) \times (m-2) / 2}{(N-1)(N-2) / 2}$$

• **Class2:** the $N-m$ added nodes. In this class, the neighbors of every node are the m original nodes, which fully connect to each other, thus the clustering coefficient of every node in this class is 1.

The sum of all nodes' clustering coefficient is

$$\sum_i c_i = m \cdot \frac{(m-1) \times (N-m) + (m-1) \times (m-2) / 2}{(N-1)(N-2) / 2} + (N-m) \cdot m$$

Thus the average clustering coefficient of this network is

$$c_{G_{\beta \rightarrow \infty}} = \left(m \cdot \frac{(m-1) \times (N-m) + (m-1) \times (m-2) / 2}{(N-1)(N-2) / 2} + (N-m) \cdot m \right) / N$$

When N is large, $c_{G_{\beta \rightarrow \infty}} = 1$

3.3.2 $\beta = -8, -4, -2, -1, 0, 1, 2$

Fig.3.9 and Fig.3.10 illustrate us that the average clustering coefficient of the model with $\beta = 2$ is much larger than the models with small β . This is because when $\beta = 2$ the neighbors of most nodes in the model are the m original nodes which are fully connected to each other. Hence the average clustering coefficient is high. When β is under 1, most newly added nodes in the network do not connect to the original m high degree nodes but distribute to the smallest degree nodes which have low possibility to connect to each other. Thus the average clustering coefficient is low for small β .

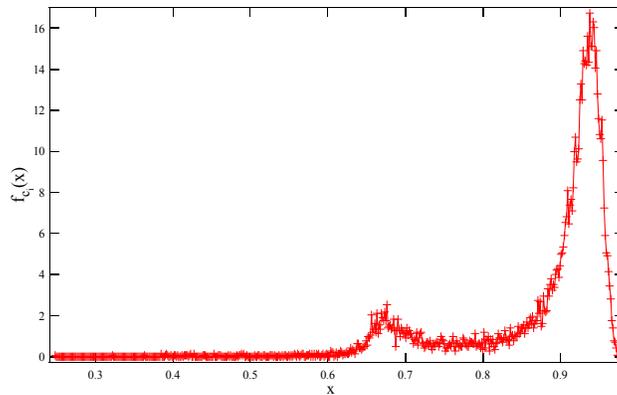


Figure 3.9 The average clustering coefficient distribution of $G^*(m=3, N=800, \beta=2)$

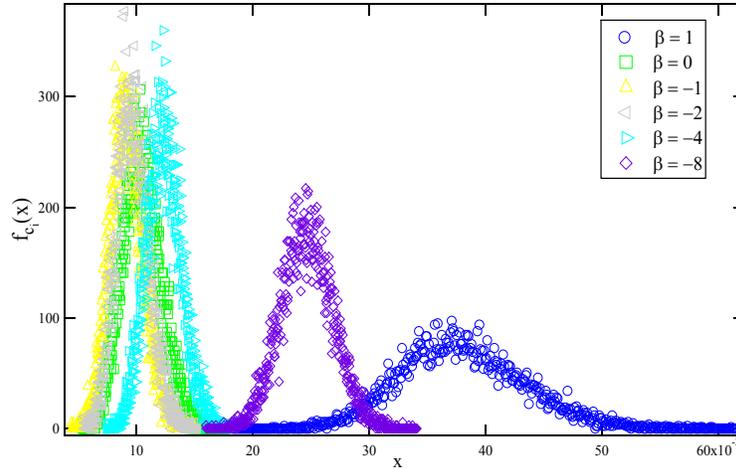


Figure 3.10 The average clustering coefficient distribution of $G^*(m=3, N=800, \beta=-8, -4, -2, -1, 0, 1)$

Also we study the mean value of the average clustering coefficient for each β , as shown in Fig.3.11.

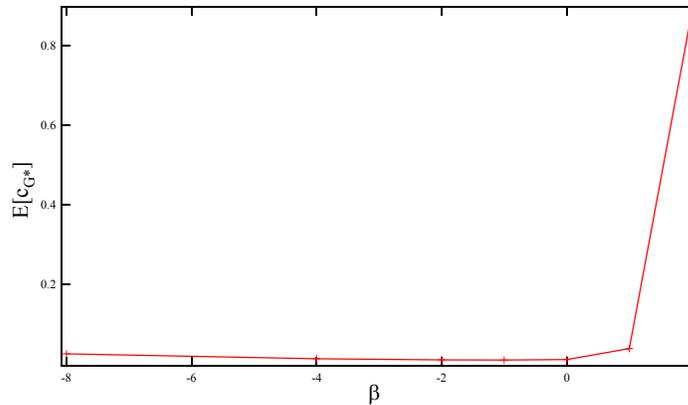


Figure 3.11 The mean of the average clustering coefficient of $G^* (m=3, N=800)$

In Fig.3.10 and Fig.3.11 we observe that when $\beta = 0$ the model has the lowest average clustering coefficient but the difference of the coefficients between the model with $\beta = 0$ and other models with $\beta < 0$ is not obvious. This can be explained by the preferential attachment rule. When $\beta = 0$ the model picks out the existed nodes randomly among all the nodes in the network; on the other hand, when $\beta < 0$ the models select the existed nodes from these smallest degree nodes also at random, so approximately we can consider when $\beta < 0$ the models choose nodes randomly. But it seems the range of the existed nodes with the highest probability to be chosen when $\beta < 0$ is smaller than the one when $\beta = 0$. This makes the model with $\beta = 0$ likely more distributed than other models with $\beta < 0$. Thus when $\beta = 0$ the model has the lowest average clustering coefficient. Meanwhile as β decreases under 0, the range of the existed nodes that have the high probability to be chosen is likely more and more small, so the average clustering coefficient increases a bit.

3.4 Eigenvalues λ_i of the adjacency matrix

3.4.1 Eigenvalues λ_i of the adjacency matrix

3.4.1.1 $\beta \rightarrow +\infty$

The adjacency matrix of $G_{\beta \rightarrow +\infty}(N)$ is $A_{G_{\beta \rightarrow +\infty}} = \begin{bmatrix} J_{3 \times 3} - I_{3 \times 3} & J_{3 \times (N-3)} \\ J_{(N-3) \times 3} & 0_{(N-3) \times (N-3)} \end{bmatrix}$ in which

$J_{m \times m}$ is a $m \times m$ matrix with all the elements equal to 1 and $I_{m \times m}$ is a $m \times m$ identity matrix, so the eigenvalues of $A_{G_{\beta \rightarrow +\infty}}$ can be calculated from the equation

$\det(A_{G_{\beta \rightarrow +\infty}} - \lambda I) = 0$, where $N > 3$.

$$\begin{aligned}
 \det(A_{G_{\beta \rightarrow +\infty}} - \lambda I) &= \det \begin{bmatrix} J_{3 \times 3} - I_{3 \times 3} - \lambda I_{3 \times 3} & J_{3 \times (N-3)} \\ J_{(N-3) \times 3} & (-\lambda) I_{(N-3) \times (N-3)} \end{bmatrix} \\
 &= \det((-\lambda) I_{(N-3) \times (N-3)}) \cdot \det((J_{3 \times 3} - (\lambda + 1) I_{3 \times 3}) - J_{3 \times (N-3)} \cdot (-\frac{1}{\lambda}) I_{(N-3) \times (N-3)} \cdot J_{(N-3) \times 3}) \\
 &= (-\lambda)^{N-3} \cdot \det(J_{3 \times 3} - (\lambda + 1) I_{3 \times 3} - (-\frac{N-3}{\lambda}) J_{3 \times 3}) \\
 &= (-\lambda)^{N-3} \cdot \det(\frac{\lambda + N - 3}{\lambda} J_{3 \times 3} - (\lambda + 1) I_{3 \times 3}) \\
 &= (-\lambda)^{N-3} \cdot (\frac{\lambda + N - 3}{\lambda})^3 \cdot \det(J_{3 \times 3} - \frac{(\lambda + 1) \cdot \lambda}{\lambda + N - 3} I_{3 \times 3}) \\
 &= (-\lambda)^{N-3} \cdot (\frac{\lambda + N - 3}{\lambda})^3 \cdot (-1)^3 \cdot (\frac{(\lambda + 1) \cdot \lambda}{\lambda + N - 3})^2 \cdot (\frac{(\lambda + 1) \cdot \lambda}{\lambda + N - 3} - 3) \\
 &= (-\lambda)^{N-4} \cdot (\lambda + 1)^2 \cdot ((\lambda - 1)^2 - 3N + 8)
 \end{aligned}$$

The spectral radius λ_1 of $G_{\beta \rightarrow +\infty}(N)$, which is denoted as the largest eigenvalue of the adjacency matrix, is $\sqrt{3N - 8} + 1$ when $N > 3$. The other eigenvalues when $\beta \rightarrow +\infty$ are either 0 or -1.

3.4.1.2 $\beta = -8, -4, -2, -1, 0, 1, 2$

Fig.3.12 and Fig.3.13 show the probability density function (pdf) of the eigenvalues of the adjacency matrix of G^* . Fig.3.12 is the eigenvalues of the model with $\beta = 2$ and Fig.3.13 is the eigenvalues of the models with other β .

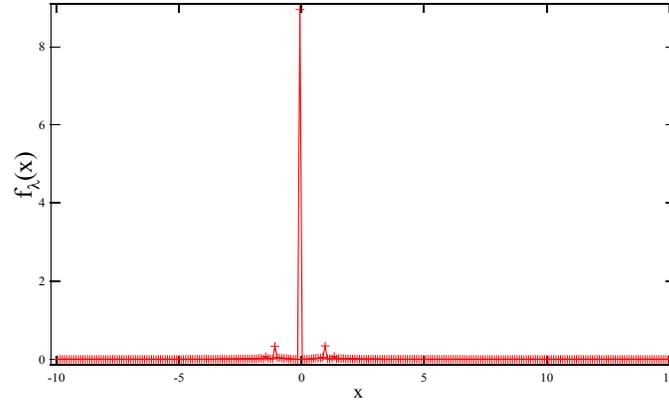


Figure 3.12 The spectrum distribution of the adjacency matrix of $G^*(m=3, N=800, \beta=2)$

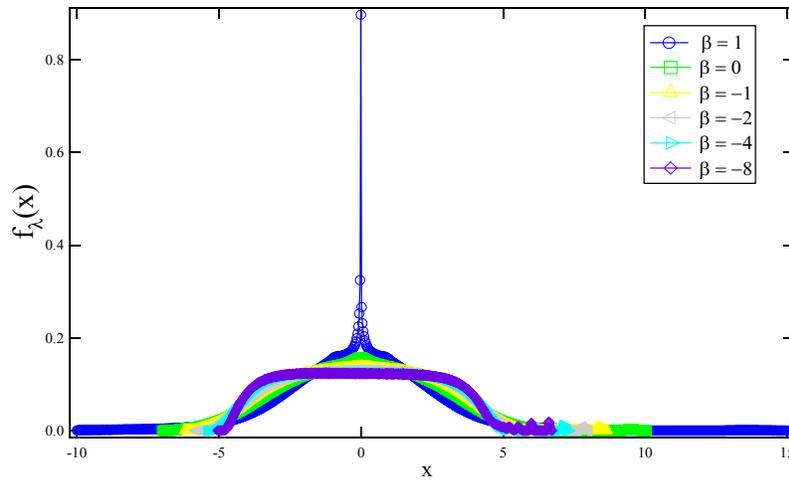


Figure 3.13 The spectrum distribution of the adjacency matrix of $G^*(m=3, N=800, \beta=-8, -4, -2, -1, 0, 1)$

These figures tell us that when $\beta = 2$ most eigenvalues of the model have the value 0 . We have deduced the spectrum of our model when $\beta \rightarrow +\infty$ in chapter 3.4.1.1: $N-4$ eigenvalues of the adjacency matrix should equal to 0 . When $\beta = 2$ the network structure of the model is close to the model when $\beta \rightarrow +\infty$. Hence the model with $\beta = 2$ should also have a similar spectrum as that with $\beta \rightarrow +\infty$. This explains why most eigenvalues of $G^*(\beta = 2)$ are equal to 0 .

When $\beta = 1$ the spectrum still has a peak at eigenvalues 0 , but the percentage of these eigenvalues is much less than the model with $\beta = 2$. Meanwhile the percentage of other eigenvalue increases. When $\beta \leq 1$, as β decreases, the peak at eigenvalue 0 disappears and the pdf of eigenvalues becomes more and more smooth. In other words, the probability density function of the eigenvalue approaches a uniform distribution. Compared with the model when $\beta = 2$, we also observe that the range of eigenvalues shortens as β decreases.

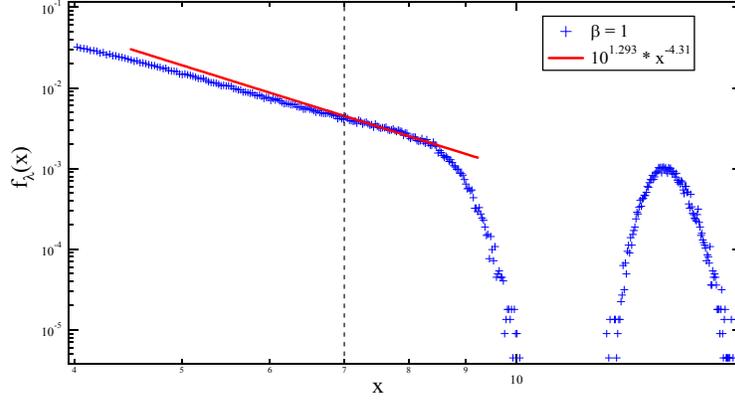


Figure 3.14 The curve fitting of the spectrum tail part of $G^*(m=3, N=800, \beta=1)$

Fig.3.14 is the curve fitting of the tail part of the spectrum when $\beta = 1$. It has been indicated in [13] that the eigenvalue distribution of a scale-free network follows a power law. In other words, if the degree distribution of a network has a power-law tail, then the eigenvalue distribution of this network also has a power-law tail. And it also has been demonstrated in [13] that in scale-free networks the exponent γ of degree distribution tail and the exponent δ of adjacency spectrum have the following relationship $\delta = 2\gamma - 1$ at large eigenvalues λ . So we do the curve fitting on these eigenvalues which locate in the range $\lambda \geq 7$ and then we obtain the exponent $\delta = 4.31$. Since we already have the degree distribution exponent $\gamma = 2.713$ in chapter 3.1.2.3, so we can calculate the theoretical exponent δ which should be around $\delta = 2\gamma - 1 = 4.426$. Thus we find that our simulation result is close to the theoretical analysis.

3.4.2 Spectral radius λ_1

Spectral radius is the largest eigenvalue of the adjacency matrix. It can be used to represent the capability of a network against the viruses spreading. The smaller the spectral radius is, the stronger the robustness of a network is against the spread of virus [12]. In other words, the information is more difficult to propagate. So the more centralized the network structure is, the weaker it is against the virus spreading. When β is large, most nodes are connected to the original m nodes which can be regarded as hubs in the network. In this circumstance, virus is easy to infect the hubs and then spread out. Thus compared with other 6 cases of β , the model with $\beta = 2$ should have the largest spectral radius. As Fig.3.15 and Fig.C.2 (Fig.C.2 is the average of the spectral radius) shown, when $\beta = 2$ the spectral radius of the model indeed has the largest value.

In Fig.3.15 we observe that not only the value, but also the range of the spectral radius gets larger as β increases. When β is smaller than 0, the spectral radius of these β models are very close. Meanwhile with β decreasing, the spectral radius is more and

more unanimous.

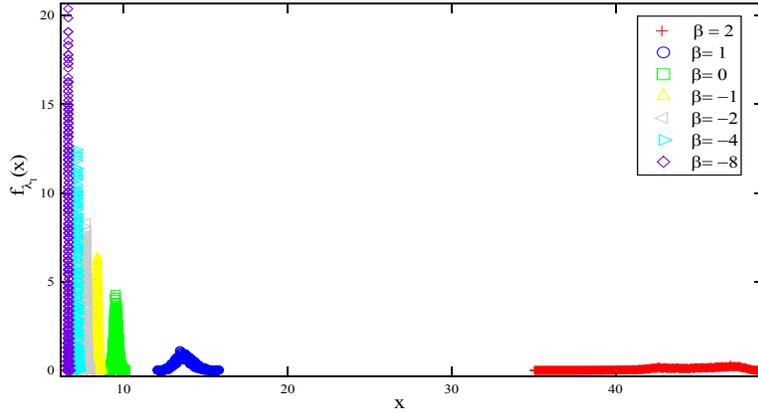


Figure 3.15 The spectral radius distribution of $G^*(m=3, N=800, \beta=-8, -4, -2, -1, 0, 1, 2)$

3.5 Eigenvalues μ_i of the Laplacian matrix

3.5.1 Eigenvalues μ_i of the Laplacian matrix

3.5.1.1 $\beta \rightarrow +\infty$

The Laplacian matrix Q can be calculated from $Q = \Delta - A$, where $\Delta = \text{diag}(d_i)$ in which d_i is the degree of node i ($i \in N$), A is the adjacency matrix of $G_{\beta \rightarrow +\infty}(N)$, written

as $A_{G_{\beta \rightarrow +\infty}} = \begin{bmatrix} J_{3 \times 3} - I_{3 \times 3} & J_{3 \times (N-3)} \\ J_{(N-3) \times 3} & 0_{(N-3) \times (N-3)} \end{bmatrix}$. The eigenvalues of the corresponding Laplacian

matrix $Q_{G_{\beta \rightarrow +\infty}} = \Delta_{G_{\beta \rightarrow +\infty}} - A_{G_{\beta \rightarrow +\infty}}$ are the solutions of $\det(Q_{G_{\beta \rightarrow +\infty}} - \mu I) = 0$. For $N > 3$,

$$\begin{aligned} \det(Q_{G_{\beta \rightarrow +\infty}} - \mu I) &= \det \begin{bmatrix} (N - \mu)I_{3 \times 3} - J_{3 \times 3} & -J_{3 \times (N-3)} \\ -J_{(N-3) \times 3} & (3 - \mu)I_{(N-3) \times (N-3)} \end{bmatrix} \\ &= \det((3 - \mu)I_{(N-3) \times (N-3)}) \cdot \det((N - \mu)I_{3 \times 3} - J_{3 \times 3} - J_{3 \times (N-3)} \cdot \frac{1}{3 - \mu} I_{(N-3) \times (N-3)} \cdot J_{(N-3) \times 3}) \\ &= (3 - \mu)^{N-3} \cdot \det((N - \mu)I_{3 \times 3} - \frac{N - \mu}{3 - \mu} J_{3 \times 3}) \\ &= (3 - \mu)^{N-3} \left(-\frac{N - \mu}{3 - \mu}\right)^3 \cdot \det(J_{3 \times 3} - (3 - \mu)I_{3 \times 3}) \\ &= -(3 - \mu)^{N-6} (N - \mu)^3 \cdot (3 - \mu)^2 \mu \\ &= -(3 - \mu)^{N-4} (N - \mu)^3 \mu \end{aligned}$$

The algebraic connectivity μ_{N-1} of $G_{\beta \rightarrow +\infty}(N)$, which is defined as the second smallest eigenvalue of the Laplacian matrix, is always 3 when $N > 4$ and is 4 for $N = 4$. The other eigenvalues when $\beta \rightarrow +\infty$ are either 0 or N .

3.5.1.2 $\beta = -8, -4, -2, -1, 0, 1, 2$

Fig.3.16 presents the eigenvalue distribution of the model when $\beta = 2$. Fig.3.17 shows the probability density function (pdf) of the eigenvalues ranging between 0 and 15 for the 7 cases of β .

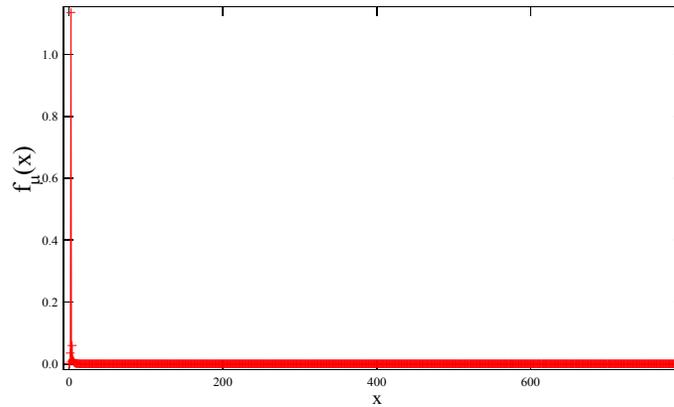


Figure 3.16 The spectrum distribution of the Laplacian matrix of $G^*(m=3, N=800, \beta=2)$

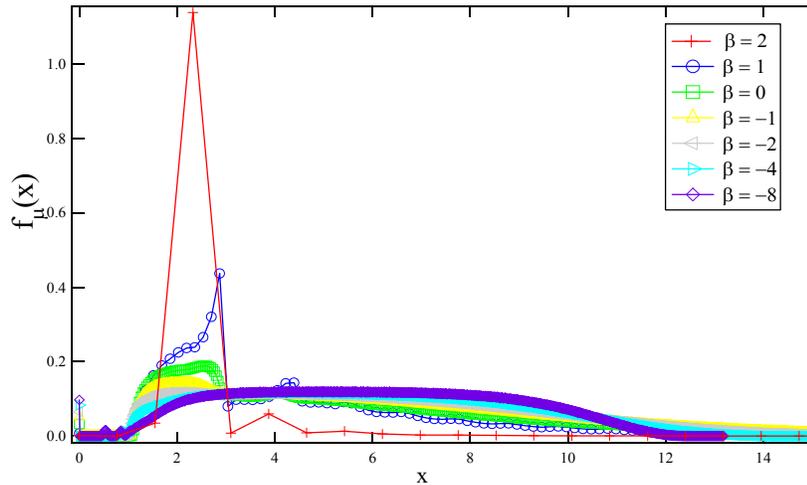


Figure 3.17 The spectrum distribution of the Laplacian matrix of $G^*(m=3, N=800, \beta=-8, -4, -2, -1, 0, 1, 2)$

According to the theoretical analysis, when $\beta = 2$ there should be a peak in the spectrum of the model since there should be a lot of eigenvalues have the value 3. But in Fig.3.17 this is not observed in our simulation result. This is caused by the

accumulative way of calculating the pdf. When $\beta = 2$ the eigenvalue range of the model is very large. So when we derive the pdf for the same number of bins, the model with $\beta = 2$ has a larger bin size than the models with other β . Compared the pdfs when $\beta = 2$ and $\beta = 1$, we can easily find that the bin size of the former is larger than that of the latter. So in this circumstance, the eigenvalues 3 would be accumulated into the bin whose range covers 3 but the mark of this bin may not directly points to eigenvalue 3.

The tendency of the spectrum of the Laplacian matrix is similar to that of the adjacency matrix. The peak corresponds to the eigenvalue 3 disappears and the curve becomes more and more smooth as β decreases, meanwhile the range of the eigenvalues also gets shorter.

3.5.2 Algebraic connectivity μ_{N-1}

The algebraic connectivity μ_{N-1} is the second smallest eigenvalue of the Laplacian matrix. It measures how a network constructs. The larger the algebraic connectivity is, the more difficult it is to cut a graph into disconnected components [11], which implies that the network is more robust with respect to connectivity.

When $\beta = 2$, most newly added nodes are attached to the original m nodes, so if one of the original m nodes loses its function, since the m nodes are full connected to each other, thus the information still can find an alternative way to propagate; if the broken node are one of the latter added nodes, as long as the broken node is neither the source node nor the destination node, there will be no influence on the information propagation. Thus the algebraic connectivity of the model with $\beta = 2$ should be larger than other β models.

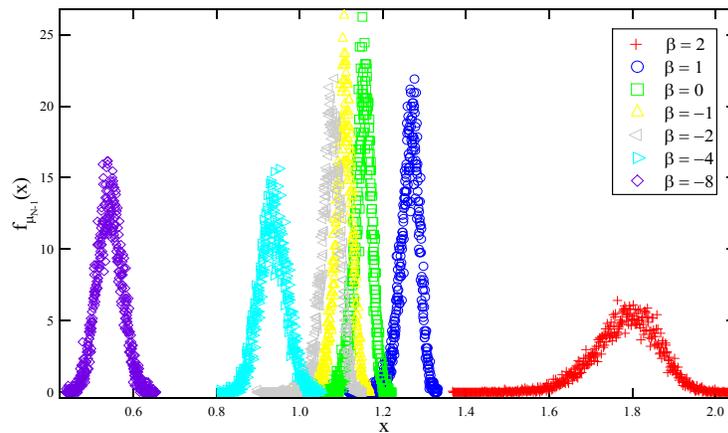


Figure 3.18 The algebraic connectivity distribution of $G^*(m=3, N=800, \beta=-8, -4, -2, -1, 0, 1, 2)$

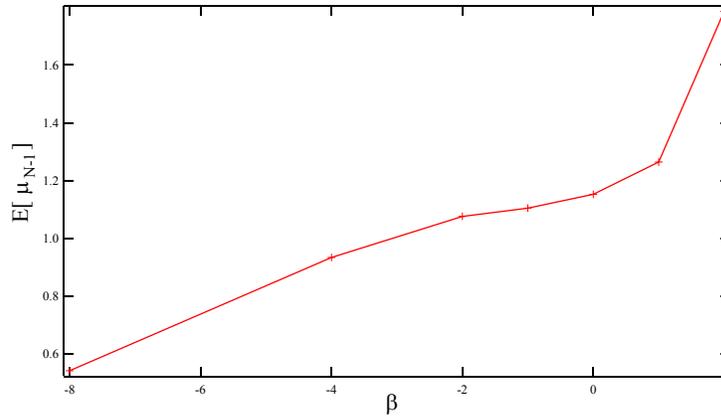


Figure 3.19 The average of the algebraic connectivity of $G^*(m=3, N=800)$

In Fig.3.18 and Fig.3.19 we notice that when $\beta = 2$ the algebraic connectivity of the model has not only the largest value but also has the largest value range. Fig.3.19 also shows us that the average algebraic connectivity is close when β is between -2 and 0 . Then it increases dramatically as β increases in the range $[0, 2]$.

3.6 Assortativity coefficient r

Assortativity refers to a preference for a network's nodes to attach to others in a similar or different way. For instance, the tendency that highly connected nodes tends to be connected with other high degree nodes is referred to as assortative mixing (assortativity); the tendency that high degree nodes tend to attach to low degree nodes is referred to as disassortative mixing (disassortativity). Usually we use assortativity coefficient to measure the assortativity of a network, which can be calculated from[6]:

$$r = \frac{L^{-1} \sum_l j_l k_l - [L^{-1} \sum_l \frac{1}{2} (j_l + k_l)]^2}{L^{-1} \sum_l \frac{1}{2} (j_l^2 + k_l^2) - [L^{-1} \sum_l \frac{1}{2} (j_l + k_l)]^2} \quad (4)$$

Where j_l, k_l are the degrees of the two nodes at the ends of the l -th link, with $l = 1, 2, \dots, L$. L is the total links of the network.

If the assortativity coefficient $r > 0$, the network is assortative; otherwise if $r < 0$, the network is disassortative. If $r = 0$, the network does not have assortative mixing feature.

3.6.1 $\beta \rightarrow +\infty$

Formula (4) indicates that the assortativity coefficient is decided by the degrees of the

two nodes at the ends of links. When $\beta \rightarrow +\infty$ there are two classes of node degrees, so we classify the links into two classes to calculate the coefficient respectively.

- Class1: one node with degree m and the other node with degree $N - 1$. This kind of links we have $m \cdot (N - m)$. The terms in (4) are

$$j_i k_i = (N - 1) \cdot m, j_i + k_i = (N - 1) + m, j_i^2 + k_i^2 = (N - 1)^2 + m^2.$$

- Class2: two nodes with the same degree $N - 1$. This kind of links there are $\frac{m \cdot (m - 1)}{2}$. The terms in (4) are

$$j_i k_i = (N - 1) \cdot (N - 1), j_i + k_i = (N - 1) + (N - 1), j_i^2 + k_i^2 = (N - 1)^2 + (N - 1)^2.$$

The total number of links in the network is $m \cdot (N - m) + \frac{m \cdot (m - 1)}{2}$. In our model $m = 3$, so the total number of links can be written as $m \cdot (N - m) + \frac{m \cdot (m - 1)}{2} = 3(N - 2)$. Substituting above terms and $m = 3$ into formula (4) we can get

$$r_{\beta \rightarrow +\infty} = \frac{36(N - 2)(4N^2 - 14N + 10) - 9(N^2 - 3N + 4)^2}{18(N - 2)[(N - 3)(N^2 - 2N + 10) + 2(N - 1)^2] - 9(N^2 - 3N + 4)^2}$$

So when N is large, $r_{\beta \rightarrow +\infty} = -1$.

3.6.2 $\beta = -8, -4, -2, -1, 0, 1, 2$

First we consider the model when $\beta = 0$. The probability for every existed node to be attached by new coming nodes is uniform $p = 1$. So the assortativity coefficient can be deduced from (4) as [6]: $r = \frac{p}{1 + 2p} = \frac{1}{1 + 2} \approx 0.33$. Then we consider the model when $\beta = 1$. The BA model has been proved in [6] that it has no assortative mixing property for $r = \frac{\log^2 N}{N} \approx 0$ when N is large.

The pdf of the assortativity coefficient in Fig.3.20 verifies that $r \rightarrow 0$ when $\beta = 1$ and $r = 0.33$ when $\beta = 0$. We also get the general tendency that the network is

disassortative when $\beta > 1$ and assortative when $\beta \leq 0$. When $\beta \leq 0$, the smaller the β is, the larger the assortativity coefficient the network has, which means the nodes in the network attach to others in a more similar way; that is to say, the low degree nodes will connect to other low degree nodes; on the contrary, when $\beta > 1$, the bigger β is, the nodes attach to others in a more different way. In other words, the highly connected nodes will be attached to the low degree nodes.

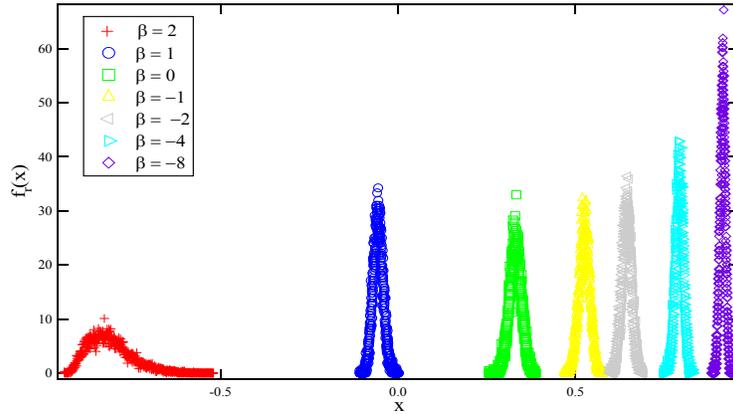


Figure 3.20 The assortativity coefficient distribution of $G^*(m=3, N=800, \beta=-8,-4,-2,-1,0,1,2)$

3.7 Summary

In this chapter we studied the structural measures and the spectral measures on our model when β changes from negative to positive to observe the influence of β on the topological measures. Especially, we deduced these measures of our model when $\beta \rightarrow +\infty$.

In the structural measures aspect, we studied the degree distribution, the average hopcount, the clustering coefficient and the assortativity coefficient; in the spectral measures, we studied the eigenvalues of the adjacency matrix and the Laplacian matrix.

When $\beta \rightarrow +\infty$ the first m nodes in the network collect all the links from the new added nodes. Thus the degrees in the network are either m or $N-m$, the average hopcount is 2 and the average clustering coefficient is 1 when N is large. The spectral radius is $\sqrt{3N-8}+1$ when $N>3$ and the algebraic connectivity is 3 when $N>4$. This network with $\beta \rightarrow +\infty$ is disassortative which means the nodes with big degree are attached to the nodes with small degree.

Besides the extreme situation of $\beta \rightarrow +\infty$, we specified 7 cases of β to study our model. In the structural measure aspect, the degree possessed by most nodes increases

as β decreases and it stays at 6 when $\beta < -2$. Through the trend of the degree distribution when $\beta < -2$ we deduce that our model is a regular graph when $\beta \rightarrow -\infty$. The same as the degree, the mean value of the average hopcount of the network decreases as β increases. When $\beta = 2$, the average hopcount is close to 2. In the spectral measures aspect, the peak at the eigenvalue 0 of the adjacency matrix disappears and the eigenvalues becomes uniform as β decreases; meanwhile the range of the eigenvalues also decreases. The Laplacian matrix is similar as the adjacency matrix. The peak at the eigenvalue 3 of the Laplacian matrix disappears and the range of the eigenvalues also decreases with β decreasing. When $\beta = 1$, the model has no assortativity; when $\beta > 1$ the model is disassortative and when $\beta \leq 0$ the model is assortative.

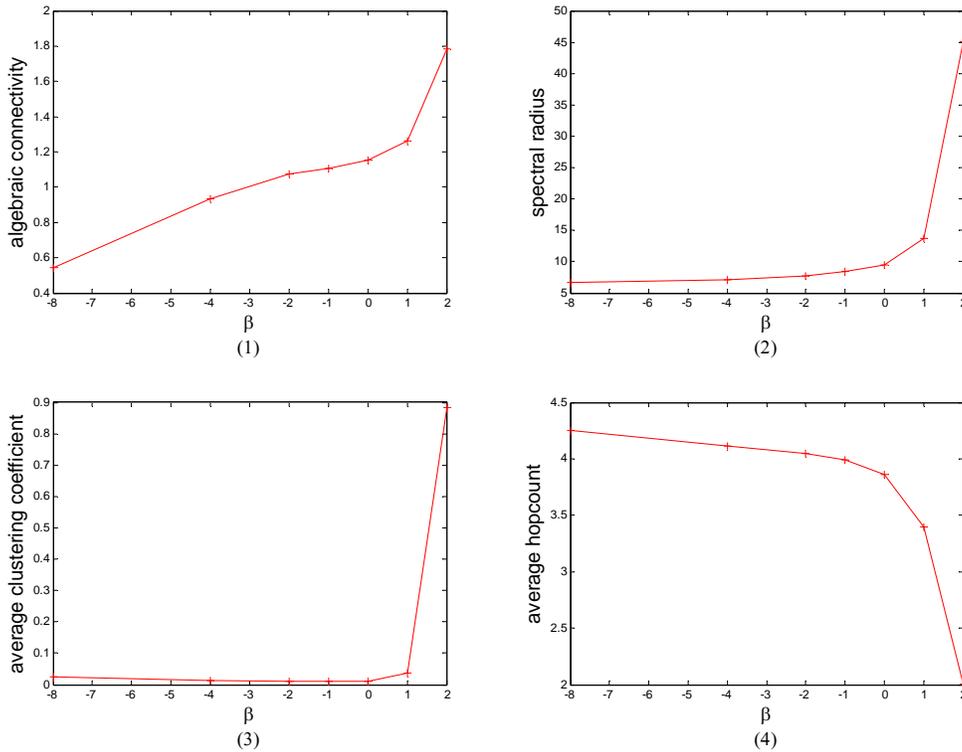


Figure 3.21 The average of measures of $G^*(m=3, N=800)$

(1).algebraic connectivity, (2).spectral radius, (3).average clustering coefficient, (4).average hopcount

Fig.3.21 is the comparison between the average values of the four measures: the algebraic connectivity, the spectral radius, the average clustering coefficient and the average hopcount. Through the comparison we can easily find that the algebraic connectivity is more sensitive than other three measures when β is in the range $[-8, 0]$. Thus the algebraic connectivity is a good measure for the network robustness when the network is in this β range.

4

Correlation of topological measures in relation to β

Topological measures of a network can be used to represent the network. The relationship between these measures may be strong or weak. We can choose the weakly related measures to characterize the network. Correlation is a way to observe the dependency between the measures. Since topological measures rely on the network structure, so the correlation between the measures also rely on the structure. In this section, we try to find the influence of β on the correlation of topological measures.

4.1 Definition

The correlation coefficient indicates the strength and direction of a linear relationship between two variables. It can be calculated by:

$$\begin{aligned} \text{corr}(X, Y) = \rho_{X, Y} &= \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E((X - \mu_X)(Y - \mu_Y))}{\sigma_X \sigma_Y} \\ &= \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}} \end{aligned}$$

Where μ_X, μ_Y are the expected values of X, Y and σ_X, σ_Y are the standard deviations of X, Y respectively. E is the expected value operator and cov is the covariance operator.

The closer the absolute value of the correlation coefficient is to 1, the stronger the correlation between the two variables is.

When correlation coefficient is used to observe the relationship between two variables, we need to consider not only the value but also the sign of the value. The strength of the correlation depends on the correlation coefficient value and the

direction depends on the sign of the value.

If the correlation coefficient is positive, it means the two variables change in the same trend: both of them increase or decrease approximately at the same time. On the contrary, if the coefficient is negative, the two variables are anti-correlated. In other words, the two variables change in a different way: one increases while the other decreases during the same time period.

Meanwhile we consider the influence of the value of the correlation coefficient. The larger the absolute value of the correlation coefficient is, the stronger the correlation between the two variables. There are two special cases: when the correlation coefficient equals to 1 or -1 . In both cases the two variables have the strongest correlation. When the coefficient equals to 1 , the two variables have an increasing linear relationship; when the coefficient equals to -1 , they have a decreasing linear relationship. When the coefficient is between -1 and 1 , its absolute value represents the weight of linear dependency between the two variables.

4.2 Analysis

In order to observe the influence of β on the correlation between the network measures, we choose four kind measures to calculate the correlation coefficient between any pair of two. The four measures are the algebraic connectivity μ_{N-1} , the average clustering coefficient $E[c]$, the average hopcount $E[H]$ and the spectral radius λ_1 of a network. Meanwhile we specify twelve different values of β from -16 to 8 in the calculation, in order to investigate the general dependency of the correlation of topological measures in relation to β . The correlation coefficient between any two measures computed based on 10000 networks are shown in Table 1.

To make the visualization easier, we use the following ranges to classify the correlation into four different levels [14]:

- no correlation: $0 \leq |corr| \leq 0.3$
- mild correlation: $0.3 \leq |corr| \leq 0.6$ (Marked with the underline font)
- significant correlation: $0.6 \leq |corr| \leq 0.9$ (Marked with the bold font)
- strong correlation: $0.9 \leq |corr| \leq 1$ (Marked with the underline and bold font)

From Table 1 we find that when $\beta = 8$, the correlation coefficients related to average hopcount are all 0 . This is because the correlation coefficient is meaningful only if the standard deviations of both variables are finite and nonzero. And when

$\beta = 8$, the average hopcount has the standard deviation $\sigma = 0$ because the average hopcount of the graphs all equals to 2, so the correlation coefficient related to the average hopcount is 0.

β	-16			-8			-4		
measures	2	3	4	2	3	4	2	3	4
1	-0.156	-0.816	<u>0.483</u>	-0.101	-0.823	<u>0.429</u>	-0.112	-0.670	<u>0.304</u>
2		0.430	<u>-0.354</u>		0.291	-0.274		0.294	-0.171
3			<u>-0.598</u>			<u>-0.486</u>			-0.257
β	-2			-1			0		
measures	2	3	4	2	3	4	2	3	4
1	-0.129	<u>-0.304</u>	0.006	-0.098	-0.282	-0.011	-0.102	-0.247	0.006
2		0.249	-0.094		0.151	-0.042		0.028	0.040
3			-0.083			-0.131			-0.323
β	0.5			1			1.5		
measures	2	3	4	2	3	4	2	3	4
1	-0.048	-0.215	0.017	0.025	-0.155	0.043	<u>0.328</u>	<u>-0.353</u>	<u>0.335</u>
2		-0.151	0.185		-0.717	0.719		<u>-0.904</u>	<u>0.928</u>
3			-0.525			-0.854			<u>-0.979</u>
β	2			4			8		
measures	2	3	4	2	3	4	2	3	4
1	0.090	<u>-0.351</u>	0.012	0.065	-0.100	<u>0.451</u>	0.012	0.000	0.647
2		<u>-0.437</u>	<u>0.457</u>		-0.011	<u>0.478</u>		-0.000	0.612
3			-0.716			-0.083			0.000

Table 1. Correlation coefficient between four measures: (1) algebraic connectivity, (2) average clustering coefficient, (3) average hopcount and (4) spectral radius

Fig.4.1 is the curve version of the simulation results. From this we can observe the trend more directly and clearly.

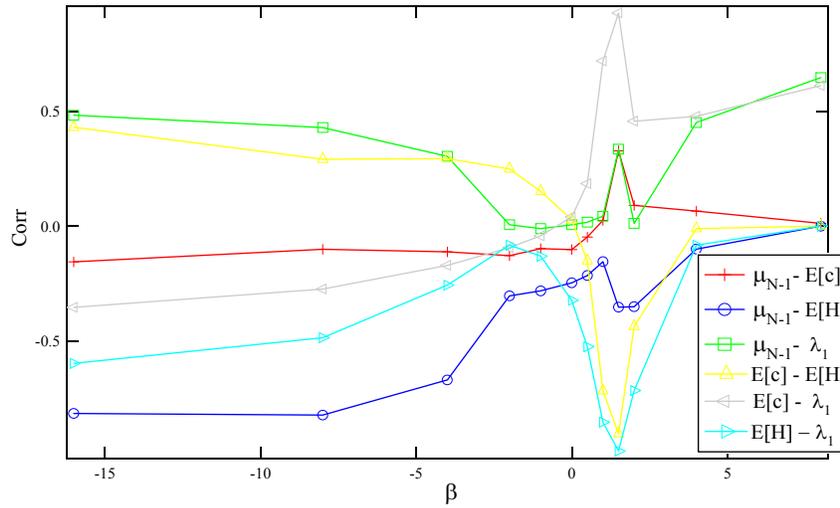


Figure 4.1 the relationship between correlation coefficient *corr* and β

The curves all change dramatically when β is in the range $[-1, 2]$. This is because correlation strongly relies on the network topology. In chapter 3 we have known that when $\beta < -1$ and $\beta > 2$, the network topologies are similar. Only when β is in the range $[-1, 2]$, the network topologies change greatly. In Fig.3.21 we know that the algebraic connectivity is more sensitive than other three measures. In other words, the change of the algebraic connectivity is big when β is in the range $[-8, 0]$, while other three measures change slightly in this β range. Thus in Fig.4.1 we can observe that when β is between -8 and 0 , the changes of the correlations related to the algebraic connectivity are relatively big, compared with their respective trend when β in the range $[-16, 0]$.

As illustrated in Chapter 3, expect for the average hopcount, other three measures are approximately positively correlated with β . Hence the correlation coefficients related to the average hopcount are mainly negative. The average clustering coefficient decreases first then increase as β increases. Correspondingly, the correlation coefficient between the average clustering coefficient and the average hopcount first is positive then becomes negative as β increases.

In Fig.4.1 we observe that the correlation coefficients between these three measures, which are the average clustering coefficient, the average hopcount and the spectral radius, reach a peak when $\beta = 1.5$. The average clustering coefficient is a local property of the network and the average hopcount is a global property. Thus when $\beta = 1.5$ we observe that the local property and the global property have the strongest relationship.

4.4 Summary

In the chapter we studied the correlation of topological measures of our model. We choose four measures to calculate the correlation coefficient between any pair of two. The four measures are: the algebraic connectivity, the spectral radius, the average clustering coefficient and the average hopcount. Through the correlation coefficient, we observe that the changes of the correlations related to the algebraic connectivity are relatively big when β is in the range $[-8, 0]$ since the algebraic connectivity is more sensitive than others in this β range. Apart from the algebraic connectivity, the correlations between other three measures change dramatically when β is in the range $[0, 4]$. Especially, the absolute values of the correlation coefficients between the three measures are larger than 0.9 when $\beta = 1.5$. This indicates that when $\beta = 1.5$ they have the strongest correlation with each other.

5

Attack vulnerability in relation to β

5.1 Introduction

The removal of nodes or links from the network may cause the segmentation of the network. Attack vulnerability denotes the changes of the network structure and property along with the removal of some nodes or links from the network. The study on attack vulnerability is one way to observe the robustness of the network. Based on these research results, we aim to gain some insights how to build a robust network against attack in the future.

5.2 Related work

Most studies have investigated the network performance of the Erdős-Rényi random networks, the Watts-Strogatz small-world networks, the Barabási-Albert scale-free networks and other developed models subject to different strategies of removing nodes or links [17, 18, 19, 20]. They examined the size of the giant component, the average hopcount and the clustering coefficient to understand the influence of a given attack strategy on the global and local property of a network.

These studies showed that when using recalculated degree-based attack strategy, the BA scale-free network is more vulnerable than the ER model and the WS model. The recalculated degree-based strategy is to choose a node with the highest degree to remove at each time and then recalculate all the nodes degrees after deleting every node. The reason why the BA model is more vulnerable is because the few nodes with a large number of connections in the BA model play a very important role in the network structure [18] and recalculated degree-based attack strategy attacks these nodes (hubs) at first. This makes the network easily to be cut into independent clusters. For the same reason, the ability of the BA model against the random failure strategy, which is to randomly choose a node to delete, is stronger than the degree-based attack strategy [18]. The network structure will not change a lot as long as the few high degree nodes in the BA model are not deleted. But the recalculated degree-based attack strategy is not so efficient to destroy the WS model compared to the BA model [20], and the ER model is the most robust when under the degree-based attack [18].

5.3 Attack model

Different structures of networks have different reactions to the same attack strategy. In this section, we study the attack vulnerability of our complex network models based on two kinds of attack objects: the cluster-focused attack and the global attack, and two kinds of node attack strategies: the random failure and the degree-based attack.

5.3.1 Attack object

There are two kinds of attack objects: the cluster-focused attack and the global attack.

- *Cluster-focused attack*: First we choose the largest cluster in the network as the attack object. Then we focus on this cluster to attack. If during the attack process, this cluster is split into two clusters or more, we choose the largest one to be the new attack object and then focus on this chosen one to attack until the cluster size is 1. Hence, any later attack object is always a sub-graph of an earlier object.

- *Global attack*: The attack object is the whole network. According to the attack strategy we choose nodes to remove from the network, no matter which cluster the removal nodes belong to. We repeat the removal steps until all the cluster size in the network are 1.

5.3.2 Attack strategy

We consider two kinds of attack strategies: the random failure and the degree-based attack.

- *Random failure strategy*: Every time, we choose a node randomly from the attack object to delete until the cluster size is 1.

- *Degree-based attack strategy*: Every time we choose a node with the highest degree from the attack object to delete. If there are several nodes with the same highest degree, choose one randomly among them. After deleting one node and clearing the connection between the deleted node and other nodes in the network, recalculate the degree of all the nodes in this attack object. Then repeat the steps mentioned above until the cluster has only 1 node.

5.3.3 Analysis measurement

Deleting some nodes from the network may make the network segmented. Thus one connected network will be divided into several unconnected clusters. The largest cluster can be a measure of the network function after attacking. So we study the largest cluster to observe the network performance under attack through the three aspects: the size of the cluster, the spectral radius and the algebraic connectivity of the largest cluster.

Our network model has $N = 800$ nodes. For the cluster-focused attack, we set every 10 nodes as an interval; for the global attack, we set every 20 nodes as an interval. We delete the nodes from the network one by one, and after deleting every interval number of nodes, we observe the property of the largest cluster.

5.4 Analysis of the largest cluster

First we put the emphasis on the network property analysis when the network is under the cluster-focused attack. Then we will compare the difference of the network property introduced by the cluster-focused attack and the global attack.

5.4.1 The size of the largest cluster

According to the present studies we know that the BA model is sensitive to the degree-based attack strategy. The size of the largest cluster should start with a very low value on the Y axis when $\beta \geq 1$, since the nodes with the highest degree have been deleted in the first interval, which makes the network segmented into many small clusters after removing the first few nodes. Thus the size of the largest cluster should decrease to be 1 after removing 20 or 30 nodes from the network. The larger the β is, the faster the stable status that all the clusters have only 1 node comes. Since $\beta \leq 0$ the size of the largest cluster should start with a high value on the Y axis.

For the random failure strategy, the behavior of our models with different β from -8 to 2 should not differ from each other too much because the models when $\beta \geq 1$ are only sensitive to the degree degree-based attack.

The simulation results are shown in Fig.5.1 and Fig.5.2. They present the largest cluster size in relation to the number of the node removals when the networks are under the random failure strategy and the degree-based attack strategy, respectively. In the following figures, the symbol N_{LC} represents the size of the largest cluster, while the symbol N_{rm} represents the number of nodes removed from the network. Thus in Fig.5.1 and Fig.5.2 the X axis N_{rm}/N is the ratio of the number of node removals to the total number of nodes N in the network; the Y axis N_{LC}/N is the ratio

of the number of nodes in the largest cluster to the total number of nodes N in the network.

Under the random failure strategy, the models with the 7 cases of β almost have the same behavior when the deleted nodes take under 30%. When the deleted nodes grow from 30% to 70%, the smaller the β is, the larger the largest cluster size is, but the difference of these models is very small. Once the removed nodes take up more than 70%, the advantage of the model with $\beta = 2$ is obvious. Other 6 cases of β become complete disconnected when the deleted nodes take around 75%, while the model with $\beta = 2$ persists until 90%.

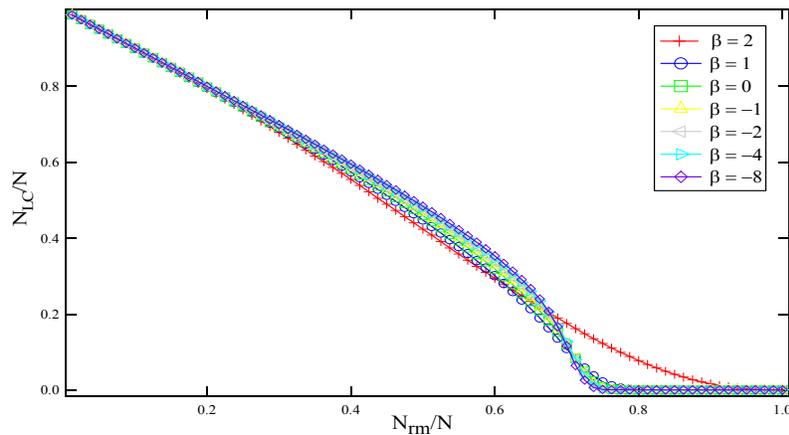


Figure 5.1 The size of the largest cluster in relation to the number of the cluster-focused random failures

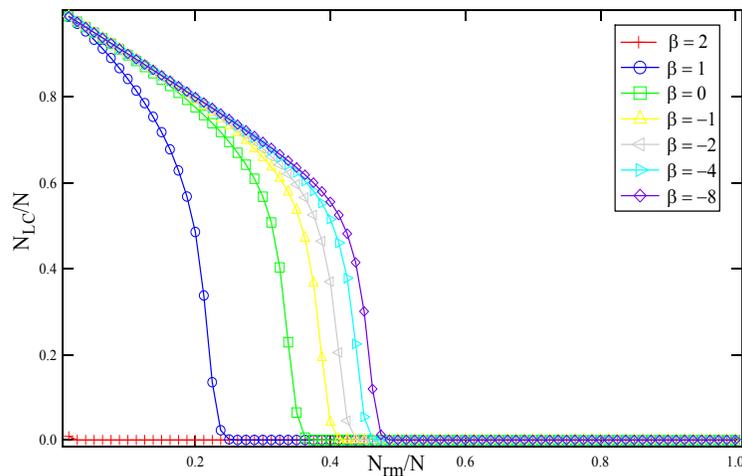


Figure 5.2 The size of the largest cluster in relation to the number of the cluster-focused degree-based attack

For the degree-based attack strategy, as we analyzed before, when $\beta = 2$ the size of the largest cluster starts at a low value and turns to be stable at the second interval. But

when $\beta = 1$ the model is different with our analysis that it starts at the same point with other models when $\beta \leq 0$. From Fig.5.2 we find that the smaller the β is, the longer time the largest cluster persists to break down.

Compared these two strategies, the degree-based attack strategy is more harmful to the network structures for all the 7 cases of β . Our network model under the degree-based attack strategy all break down earlier than the same model under the random select strategy.

5.4.2 Spectral radius of the adjacency matrix

Under the random failure strategy, the behavior of our model with a specific β can be observed directly in Fig.5.3: the larger the β is, the larger the spectral radius of the largest cluster is. A large spectral radius implies the information can propagate easily through the network. So does the virus. Thus if the information still can transmit along the network after several attacks, the network is robust. In Fig.5.3 we know that the spectral radius of the model with $\beta = 2$ is far larger than the models with other β , even the model with $\beta = 1$. So under the random failures, if we want to construct a network in terms of the capability of information propagation, we should let β be large.

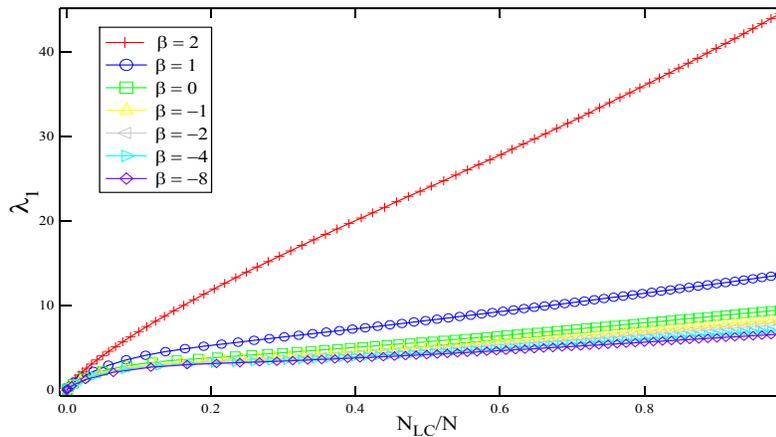


Figure 5.3 The spectral radius of the largest cluster in relation to the number of the cluster-focused random failures

We have calculated the spectral radius of our model with $\beta \rightarrow +\infty$ in Chapter 3.4.1.1. It is $\sqrt{3N - 8} + 1$ when $N > 3$. In Fig.5.3 we find that the curve of the model with $\beta = 2$ matches our analysis. This is because the network structure when $\beta = 2$ is close to that when $\beta \rightarrow +\infty$.

The behavior of the largest cluster under the degree-based attack strategy is more complicated than that under the random failure strategy. In Fig.5.4 we find that the

model with $\beta = 2$ already breaks down in the first few intervals. Apart from the model with $\beta = 2$, combined with Fig.C.2 we observe that the decreasing speed of the spectral radius of the largest cluster is proportional with the mean spectral radius of the network. In other words, if a network has a large spectral radius, then under the degree-based attack strategy, the spectral radius of the largest cluster decreases rapidly. Since the largest cluster size decreases to hold less than 50% nodes in the network, the spectral radius of the 6 cases of $\beta < 2$ are very close to each other.

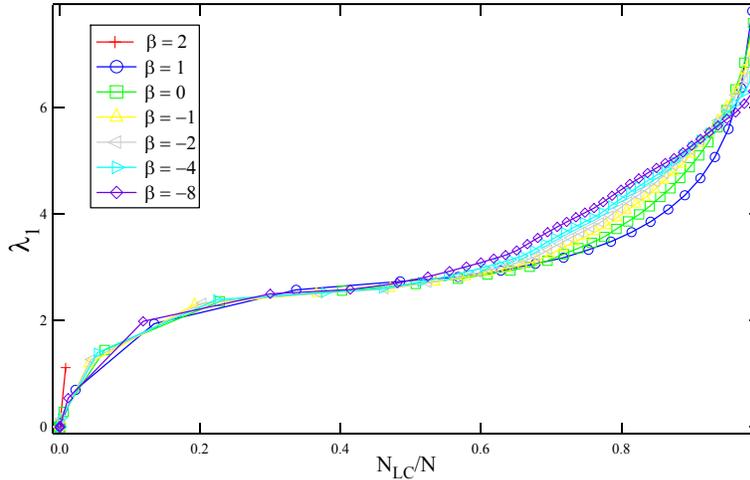


Figure 5.4 The spectral radius of the largest cluster in relation to the largest cluster size when the network is under the cluster-focused degree-based attack

5.4.3 Algebraic connectivity of the Laplacian matrix

The same as the spectral radius shown in Fig.5.3, under the random failure strategy the model with $\beta = 2$ has the best behavior in terms of the algebraic connectivity among all the 7 cases of β . From Fig.5.5 we know that when $\beta = 2$ the model has an absolutely large algebraic connectivity under the random failures. This indicates that when $\beta = 2$ the largest cluster of the model is hard to be cut into disconnected clusters. We also observe that the start point of the algebraic connectivity of the largest cluster on the Y axis increases as β increases. When β is in the range $[-4, 0]$, the difference between these models under the random failures is small. And since the largest cluster holds less than 70% nodes in the network, the behavior of the models with $\beta \leq 0$ becomes unanimous. In Fig.5.5 we find that all the 7 cases of β have a peak at the tail part of the algebraic connectivity curves. This is caused by the lack of the normalization of algebraic connectivity. In general, we do not compare the algebraic connectivity of the network with different size, since only one node may lead to the sharp change of the algebraic connectivity. For example, the algebraic connectivity of a complete graph with N nodes is $N-1$. But once we add a new node which will be attached to one of the existed nodes in the network, the algebraic connectivity of this network will reduce dramatically.

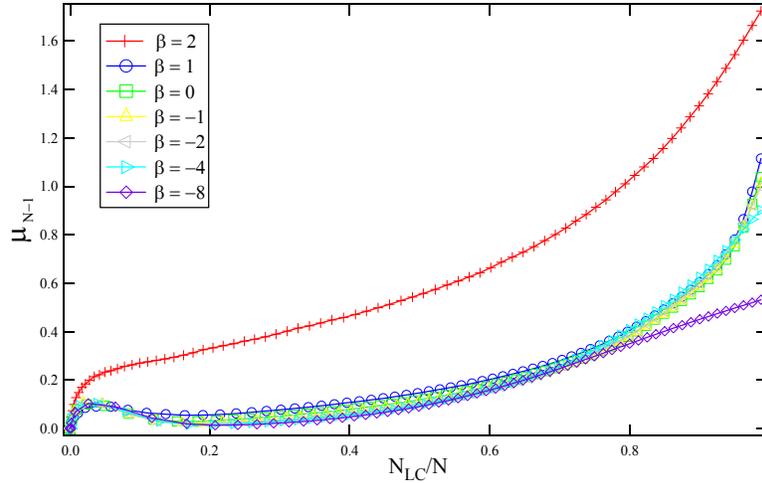


Figure 5.5 The algebraic connectivity of the largest cluster in relation to the largest cluster size when the network is under the cluster-focused random failures

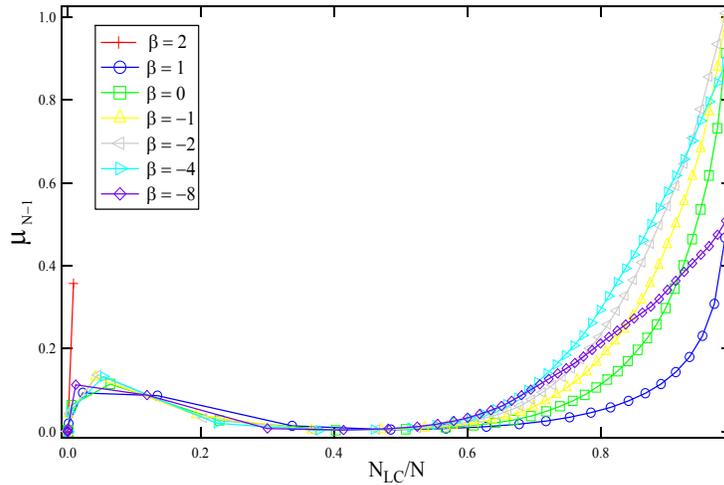


Figure 5.6 The algebraic connectivity of the largest cluster in relation to the largest cluster size when the network is under the cluster-focused degree-based attack

In Fig.5.6 we observe that under the degree-based attack strategy the model with $\beta=2$ has the smallest algebraic connectivity at each interval. Combined with Fig.3.19, the larger the mean algebraic connectivity of the network is, the faster the largest cluster breaks down. In other words, the larger the β is, the faster the algebraic connectivity of the largest cluster decreases.

The same as the networks under the random failure strategy, except the network with $\beta=2$, the networks with other 6 cases of β under the degree-based attack strategy also have a peak at the tail part of the algebraic connectivity curves for the absence of the normalization of algebraic connectivity.

5.4.4 Comparison between the cluster-focused attack and the global attack

In this part, we compare the simulation results when the network is under the cluster-focused attack and under the global attack to observe whether the different attack objects will cause different behaviors for the same model. The simulation results of the network under the global attack are shown in Appendix C.2.

From these simulation results we know that the attack vulnerability of the network under the cluster-focused attack and the global attack are similar. Differences can be observed only in the tail part of the curves. Compared with the network under the cluster-focused attack, when the network is under the global attack, the size of the largest cluster becomes 0 at a later time; the tail part of the spectral radius starts to decrease also at a later time and thus decreases more dramatically, the peak of the algebraic connectivity reaches higher when the network is under the global degree-based attack. This is because when the network is split into small clusters by node removals, the removed nodes chosen by the global attack may be not in the current investigated largest cluster. The global removal strategy postpones the time when the largest cluster breaks down.

For the absence of the normalization of the algebraic connectivity, there is also a peak at the tail part of the algebraic connectivity of the largest cluster when the network is under the global attack, as shown in Fig.C.6 and Fig.C.9. Especially, in Fig.C.10 we observe that the algebraic connectivity at the peak is 2 and the peak lasts for 2 or 3 intervals when the network is under the global degree-based attack. From this we can deduce that in this moment the network has several full connected clusters with the size 3.

5.5 Summary

In this chapter we studied the attack vulnerability of our network model for 7 cases of β based on the cluster-focused attack and the global attack when the networks are under the random failures and under the degree-based attack strategy. We choose three measures to study the vulnerability through the largest cluster in the network after attacking. The three measures are the size of the largest cluster, the spectral radius and the algebraic connectivity of the largest cluster.

In terms of the largest cluster size, we observed that when the network is under the random failures, the larger the β is, the later the network breaks down. In other words, the network is more robust. On the contrary, when the network is under the degree-based attack, the larger the β is, the earlier the network breaks down. And for all the 7 cases of our model, the degree-based attack strategy is more harmful to the

network than the random failures. Since the vulnerability relies on the network structure, so the behavior of the networks based on the global attack is similar to that based on the cluster-focused attack when using the same attack strategy.

In terms of the spectral radius and the algebraic connectivity of the largest cluster, when the network is under the random failures, the larger the β is, the larger the spectral radius and the algebraic connectivity are; on the other hand, when the network is under the degree-based attack, the larger the β is, the faster the spectral radius and the algebraic connectivity decrease. When the network is under the global degree-based attack, the algebraic connectivity of the largest cluster has a peak equals to 2 and lasts for 2 or 3 intervals at the tail part, which indicates that in this moment the network has several full connected clusters with the size 3.

6

Conclusions and Future work

6.1 Conclusions

In this paper we study the networks introduced by different preferential attachment rules which are controlled by the parameter β . We try to gain some insights on how to construct a robust network through the study of the influence of β on the network structure and property. We investigate these networks through three directions: topological characteristics, correlation of topological measures and attack vulnerability.

In topological characteristics section, we analyzed that when $\beta \rightarrow +\infty$, the first m nodes in the network gathers all the links from other newly added nodes; when $\beta \rightarrow -\infty$, the network is a regular graph in which all the nodes have the degree 6, and the network structure changes dramatically when β in the range $[-1, 2]$. From the simulation results, we observed the spectrum tendency of the adjacency matrix and the Laplacian matrix along with β . Besides this, we found the spectral radius and the algebraic connectivity both increase as β increases. Moreover, the algebraic connectivity is more sensitive than other measures we used in this section when β in the range $[-8, 0]$. Thus in this sense, the algebraic connectivity is a good measure for the network robustness when the network is in this β range.

In correlation section, we found that the correlation coefficients between the algebraic connectivity, the average clustering coefficient, the average hopcount and the spectral radius all change dramatically when β is between 0 and 4. Especially when $\beta = 1.5$, the absolute value of the correlation coefficients between the average clustering coefficient, the average hopcount and the spectral radius are the biggest, which indicates that these three measures are strong correlated with each other. When β is between -8 and 0, since the algebraic connectivity is more sensitive than other three measures, so the changes of the correlations related to the algebraic connectivity are bigger than the changes of the correlations between other three measures.

In attack vulnerability section, we observed that under the random failures, the network with large β is more robust in terms of the size of the largest cluster since the

network can persist for a longer time before break down. Meanwhile, the larger the β is, the larger the spectral radius and the algebraic connectivity of the largest cluster are. On the contrary, when the network is under the degree-based attack, the larger the β is, the faster the spectral radius and the algebraic connectivity of the largest cluster decreases. In other words, the larger the β is, the faster the size of the largest cluster becomes 0. The 7 cases of β proved that the degree-based attack is more harmful to the network than the random failures. And the behaviors of the network under the cluster-focused attack and the global attack are similar.

6.2 Future work

For the time limit, we only studied the three directions list above. There are still some measures or directions to investigate our model. This can be done in the future.

In topological characteristics section, we observed the spectrum trend of the adjacency matrix and the Laplacian matrix along with β . Then we can try to make use of this spectrum trend or try to find why the algebraic connectivity is sensitive when β in the range $[-8, 0]$.

In correlation section we observed that the spectral radius, the average clustering coefficient and the average hopcount are strong correlated to each other when $\beta = 1.5$. We can try to find why they have such strong correlations.

In attack vulnerability section we only used the node removal strategy to investigate the vulnerability of our model. In the future, we can use the link removal strategy to study the vulnerability of our model.

Appendix

A. Graphs with given node connectivity

Here is the original description and proof of theorems used in this paper, which have been proved by Huijuan Wang, Fernando Kuipers and Piet Van Mieghem.

Theorem 1 In a k -connected graph G with $N > k$ nodes, there are at least k node disjoint paths from a node to k other nodes.

Proof. It can be proved by contradiction. If there are only $x < k$ nodes disjoint paths, the graph G can be disconnected by removing x nodes, which contradicts that G is k -connected.

Theorem 2 Graph G starts from complete graph with m nodes. At each time step, we add a new node with m links that connect the new node to m different nodes that already present in G . The so constructed graph $G(N,L)$, where $N > m$,

$$L = \binom{m}{2} + (N - m)m, \text{ is } m\text{-connected.}$$

Proof. First, G initializes as $G = K_m$. After connecting a new node to the m nodes of K_m , $G = K_{m+1}$ is m connected.

Second, given $G(N,L)$ is k -connected, we are going to prove that $G(N+1,L+m)$ obtained by connecting a new node to m nodes in $G(N,L)$ is also k -connected. We name the new node as D and D is connected with m nodes A_1, A_2, \dots, A_m in $G(N,L)$.

From D to any of these m nodes $A_j, j \in [1, m]$, m node disjoint paths exist: the direct path $P^{(j)} = DA_j$ and $P^{(i)} = (DA_i) + P_{A_i A_j}$, where $1 \leq i \leq m$ and $i \neq j$, when the $m-1$ paths

$P_{A_i A_j}$ are node disjoint. Since $G(N,L)$ is k -connected, from nodes A_j to the $m-1$ node A_i ,

$1 \leq i \leq m$ and $i \neq j$, $m-1$ node disjoint paths $P_{A_i A_j}$ always exist according to Theorem 1.

Furthermore, between node D and any other node B in $G(N,L)$, we can always find m node disjoint paths: from D to the m nodes A_1, A_2, \dots, A_m via the m links, and from A_1, A_2, \dots, A_m to node B via the m node disjoint paths, which exist due to Theorem 1. Hence, between node D and any node in $G(N,L)$, at least m node disjoint paths exist. Since $G(N,L)$ is k -connected, between each node pair of $G(N,L)$, at least m disjoint paths exist. Therefore, $G(N+1,L+m)$ is k -connected.

B. Results from linear algebra

1. If $X_{m \times m} = (J - (\lambda + 1)I)_{m \times m}$, the inverse matrix of X is

$$X^{-1} = -\frac{1}{(\lambda + 1)(\lambda + 1 - m)}(J + (\lambda + 1 - m)I)_{m \times m}$$

2. $\det(J - xI)_{n \times n} = (-1)^n x^{n-1}(x - n)$

3. $\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det A \det(D - CA^{-1}B) = \det D \det(A - BD^{-1}C)$

4. Using $J_{k \times n} J_{n \times l} = nJ_{k \times l}$, we can calculate

$$\begin{aligned} Y &= J_{(N-m) \times m} X_{m \times m}^{-1} J_{m \times (N-m)} \\ &= -\frac{1}{(\lambda + 1)(\lambda + 1 - m)} J_{(N-m) \times m} (J_{m \times m} + (\lambda + 1 - m)I_{m \times m}) J_{m \times (N-m)} \\ &= -\frac{1}{(\lambda + 1)(\lambda + 1 - m)} (mJ_{(N-m) \times m} + (\lambda + 1 - m)J_{(N-m) \times m}) J_{m \times (N-m)} \\ &= -\frac{1}{(\lambda + 1)(\lambda + 1 - m)} (m^2 J_{(N-m) \times (N-m)} + m(\lambda + 1 - m)J_{(N-m) \times (N-m)}) \end{aligned}$$

C. Figures

C.1 Topological measures

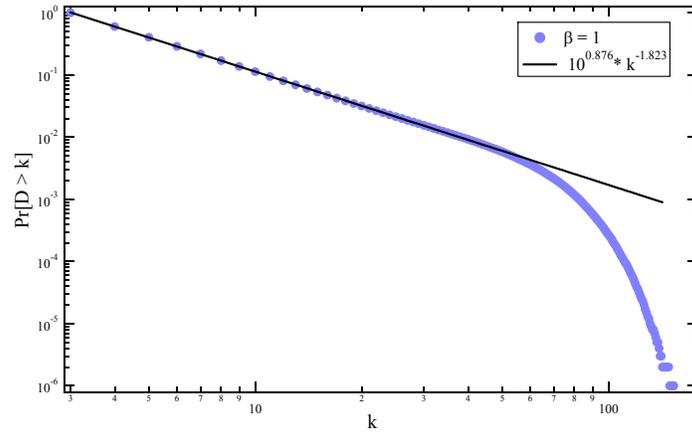


Figure C.1 The curve fitting of the degree distribution on ccdf of $G^*(m=3, N=800, \beta=1)$

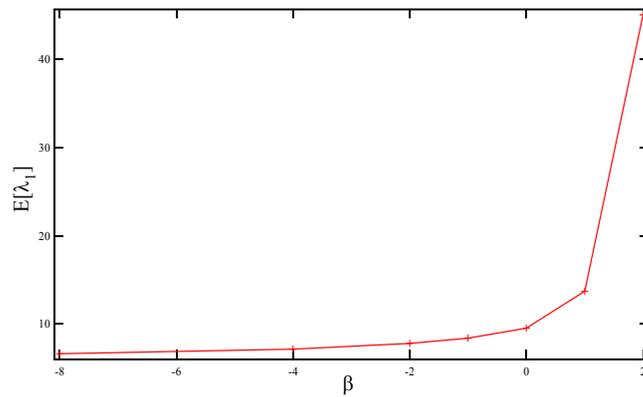


Figure C.2 The average of the spectral radius of $G^*(m=3, N=800)$

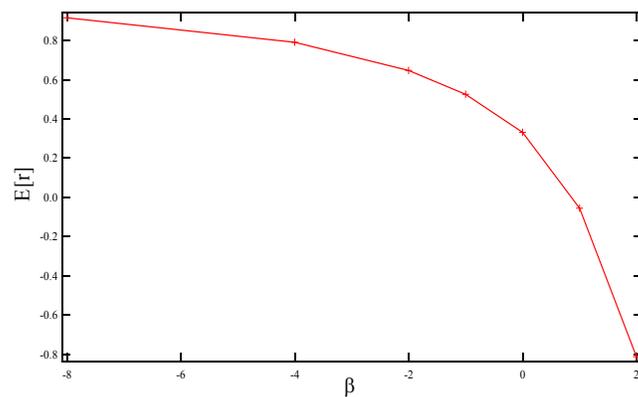


Figure C.3 The average of the assortativity coefficient of $G^*(m=3, N=800)$

C.2 Simulation results of the network under the global attack

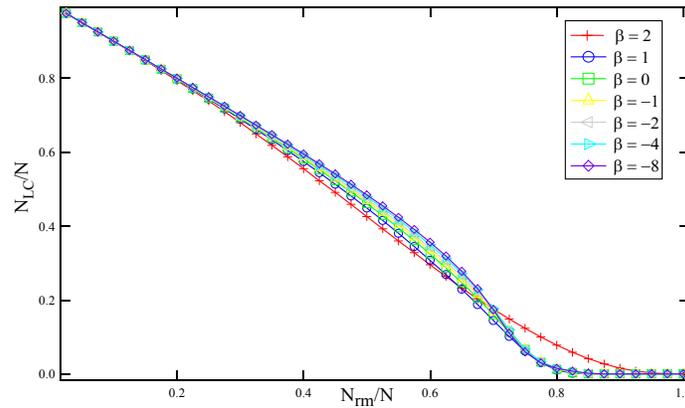


Figure C.4 The size of the largest cluster in relation to the number of the global random failures

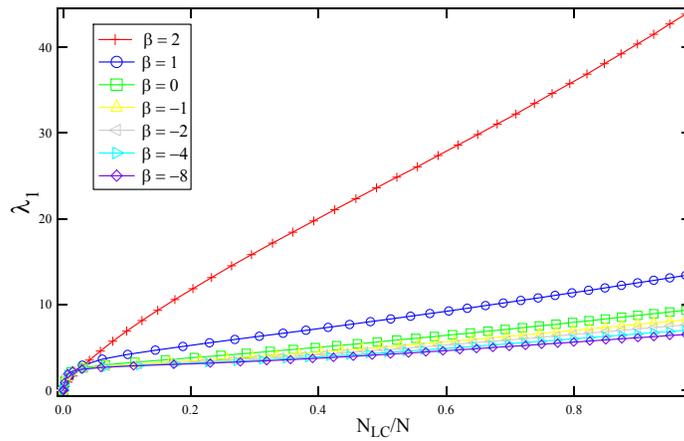


Figure C.5 The spectral radius of the largest cluster in relation to the largest cluster size when the network is under the global random failures

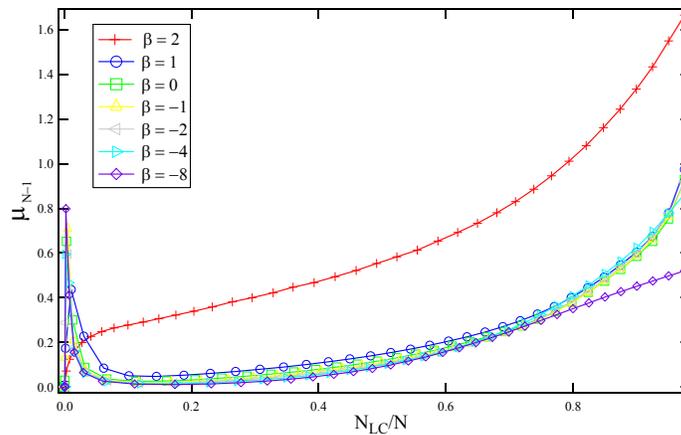


Figure C.6 The algebraic connectivity of the largest cluster in relation to the largest cluster size when the network is under the global random failures

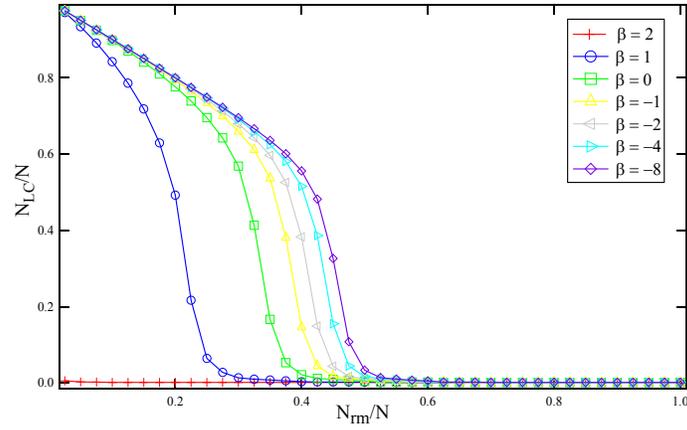


Figure C.7 The size of the largest cluster in relation to the number of the global degree-based attack

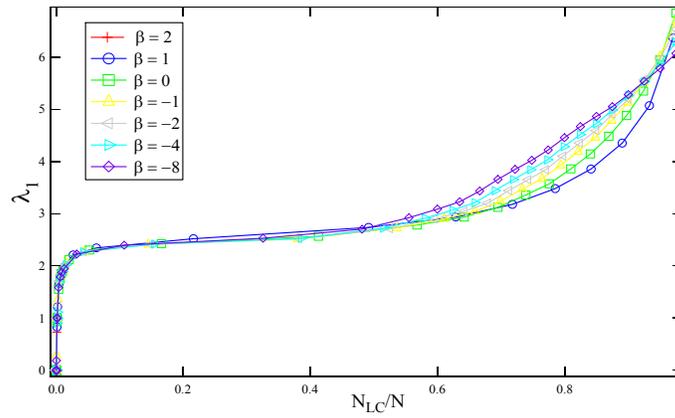


Figure C.8 The spectral radius of the largest cluster in relation to the largest cluster size when the network is under the global degree-based attack

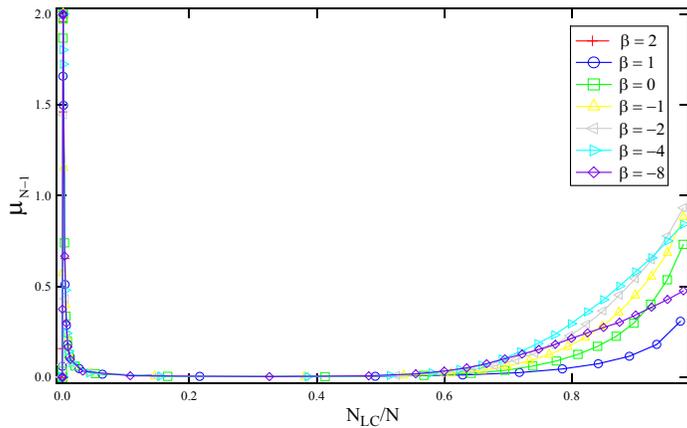


Figure C.9 The algebraic connectivity of the largest cluster in relation to the largest cluster size when the network is under the global degree-based attack

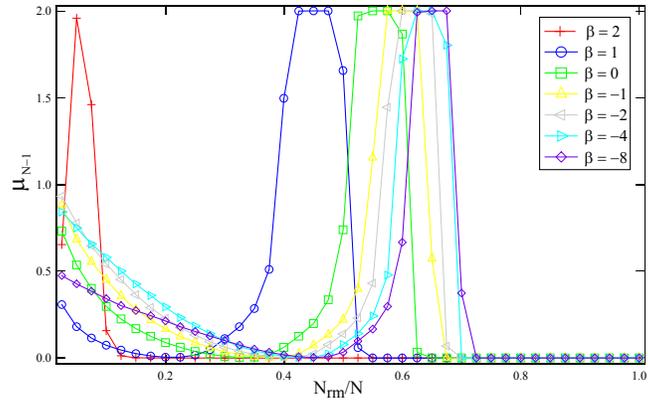


Figure C.10 The algebraic connectivity of the largest cluster in relation to the number of global degree-based attack

References

- [1] R. Albert and A.-L. Barabási, "*Emergence of scaling in random networks*", Science, Vol. 286, 509 (1999)
- [2] R. Albert and A.-L. Barabási, H.Jeong, "*Mean-field theory for scale-free random networks*", Physica A, volume 272, 173-187 (1999)
- [3] P.L.Krapivsky, S.Redner and F.Leyvraz, "*Connectivity of growing random networks*", Phys. Rev. Letter.85, 4629 (2000)
- [4] P.L.Krapivsky and S.Redner, "*Organization of growing random networks*", Phys. Rev. E 63, 066123 (2001)
- [5] B.Mohar, Y.Alavi, G.Chartrand, O.R.Oellermann and A.J.Schwenk, "*The Laplacian spectrum of graphs*", in "Graph Theory, Combinatorics and Applications", Vol. 2, 871-898 (1999)
- [6] M.E.J. Newman, "*Assortative mixing in networks*", Phys. Rev. Lett. 89, 208701 (2002)
- [7] S.N.Dorogovtsev, J.F.F.Mendes, "*evolution of networks*" (2003)
- [8] R. Albert and A.-L. Barabási, "*Statistical mechanics of complex networks*", Rev.Mod.Phys., volume 74, No.1 (2002)
- [9] S.Redner, "*How popular is your paper? An empirical study of the citation distribution*", Eur.Phys.J.B 4 2, 131-134 (1998)
- [10] S.N.Dorogovtsev, J.F.F.Mendes, "*Evolution of networks*", Oxford (2003)
- [11] M.Fiedler, "*Algebraic connectivity of graphs*", in "Czechoslovak Mathematical Journal", Vol. 23, 298-305 (1973)
- [12] Y.Wang, D.Chakrabarti, C.Wang and C.Faloutsos, "*Epidemic spreading in real networks: An eigenvalue viewpoint*", in "Proceedings of the 22nd Symposium in Reliable Distributed Computing" (2003)
- [13] S.N.Dorogovtsev, A.V.Goltsev, J.F.F.Mendes and A.N.Samukhin, "*Spectra of Complex Networks*", Phys. Rev. E 68, 046109 (2003)
- [14] A.Jamakovic, "*Characterization of complex networks*", (2008)

- [15] P. Erdős, A. Rényi, “*On the evolution of random graphs*”, Publication of the Mathematical Institute of the Hungarian Academy of Sciences, 5: 17-61 (1960)
- [16] D.J.Watts, S.H.Strogatz, “*Collective dynamics of 'small-world' networks*”, Nature 393, 440-442 (1998)
- [17] R.Albert, H.Jeong, A.-L.Barabási, “*Error and attack tolerance of complex networks*”, Nature (London) 406, 378(2000)
- [18] P.Holme, B.J.Kim, C.N.Yoon, S.K.Han, “*Attack vulnerability of complex networks*”, Phys. Rev.E 65, 056109(2002)
- [19] P.Crucitti, V.Latora, M.Marchiori, A.Rapisarda, “*Efficiency of scale-free networks-error and attack tolerance*”, Physica A 320, 622(2003)
- [20] R.Albert, I.Albert, G.L.Nakarado, “*Structural vulnerability of the North American power grid*”, Phys. Rev. E 69, 025103(R) (2004)
- [21] J.Martin Hernandez, T.Kleigerg, H.Wang, P.van Mieghem, “*A comparison of Topology Generators with Power Law Behavior*”, Proceedings of SPECTS, 484-493 (2007)
- [22] P.Van Mieghem, “*Performance Analysis of Communications Networks and Systems*”, Cambridge University Press (2006)