이학석사학위 논문

연결선 수 분포의 전체 형태가 척도 없는 연결망의 안정성에 미치는 영향

The effect of the entire degree distribution on the robustness of scale-free networks

2019년 2월

인하대학교 대학원

물리학과(이론 및 물성물리 전공)

김 홍 경

이학석사학위 논문

전체 연결선 수 분포의 전체 형태가 척도 없는 연결망의 안정성에 미치는 영향

The effect of the entire degree distribution on the robustness of scale-free networks

2019년 2월

지도교수 이 덕 선

이 논문을 석사학위 논문으로 제출함

인하대학교 대학원

물리학과(이론 및 물성물리 전공)

김 흥 경

The effect of the entire degree distribution on the robustness of scale-free networks

by

Heung Kyung Kim

A THESIS

Submitted to the faculty of

INHA UNIVERSITY

in partial fulfilment of the requirements

for the degree of

MASTER OF SCIENCE

Department of Physics

February 2019

이 논문을 김흥경의 석사학위논문으로 인정함.

2018년 12 월 3 일

주심 이 재 우 (인)

부심 이 덕 선 (인)

위원 김 기 식 (인)

Abstract

In this thesis, we investigate the robustness of the scale-free networks in the configuration model with focus on the effect of the functional form of the degree distribution. The original configuration model has a limitation; the mean degree, equal to twice of the total number of links to the number of nodes, is not controllable but fixed for given degree exponent. So, it's difficult to see the growth of the configuration-model networks. We overcome this limitation by introducing a real-valued parameter controlling the density of the minimum-degree nodes. The robustness of this generalized configuration-model scale-free networks is quantified by the largest connected component (LCC) as a function of the number of links per node. The results for the static model are considered as well for comparison. We consider various network size and degree exponents. An abnormal behavior of the LCC is observed in the configuration model. The LCC grows linearly from the beginning with increment of links until a specific point for a wide range of the degree exponent. We find that the linear growth is terminated when isolated nodes disappear. We compare the density of isolated nodes, the degree distribution, the cluster size distribution, and the network structure between the configuration model and the static model. We find that the main difference is in the degree

distribution. The configuration model's degree distribution maintains their power-law form in almost the whole range while the static model fails to keep the power-law form in the small-degree region. We take analytical approach to explain the origin of this phenomenon.

Contents

Figures

Table

Table 4.1.1 The number of links per node K and the model parameter x_{min} at which the linear growth of the LCC ends for various degree exponents γ and system size $N = 10^6$.

Introduction

Most systems consist of numerous elements which are linked to one another. They can be represented as a network consisting of nodes and links connecting them. Network systems have high complexity, but it's possible to analyze them in terms of network properties. Among them, network robustness is considered as one of the most important properties to understand the stability of networked systems. So many researchers have studied the network robustness and the way to improve it.

There are several representative networks. We usually classify networks by their degree distribution. The random network is one of the prototypical network models. In this model, all node pairs have the same probability to be linked, so the degree distribution has form of Poisson distribution. The random network model was proposed by Paul Erdős and Alfréd Rényi [1]. It was a trigger of the field which is called network science. But the model has a limitation. It is not appropriate to explain networks in the real world. They have different properties and features from those of the Erdős-Rényi networks. To overcome this limitation, many researchers had investigated real networks in the real world and made new models. Scale-free network is a successful result of their efforts [2].

Scale-free network is a network whose degree distribution follows a power law.

$$P(k) \sim k^{-\gamma}$$

Scale-free networks have been discovered in various fields including the Internet, biology, social-science, and physics [3,4]. Scale-free network and network science have drawn much attention. The most remarkable difference between random networks and scale-free networks is that scale-free networks have high robustness characterized by the size of the largest connected component (LCC). Callaway and Neuman found the concept of network robustness as clusters' size, especially the largest one [5]. Many nodes have very high degrees in scale-free networks compared with random networks, which is responsible for high robustness of scale-free networks. Subsequently, it had been wondered how to understand the LCC in a given network and many models for scale-free networks were proposed.

R. Albert and A.-L. Barabási, who had interest in the topology of real networks and found scaling property of them, introduced the Barabási-Albert (BA) model[4] based on the preferential attachment. They also studied the robustness of scale-free networks through percolation theory [4]. They showed that the LCC is maintained even with a small number of links under random removal but also it is highly vulnerable under the attack of high-degree nodes [4]. Yiping Chen and his colleagues also showed similar results; networks are broken when attack is concentrated on high-degree nodes or high-centrality nodes [6]. The effect of random removal has been studied extensively[7], and also the analytical approach is well developed [6]. There is a recent study focusing on not node deletion but the cascades of link failure [8].

Static model[9] and configuration model[10] are major models for scale-free networks. They display commonly the asymptotic power-law behavior of the degree distribution. The configuration model is generated with a degree distribution from the outset while the static model is generated with a fixed number of links and probability of connecting nodes. The static model is analytically solvable. As we want to see the development of SF networks as links are added for a fixed number of nodes, the static model's process is appropriate for the study. On the other hand, the configuration model is not well understood as much as the static model regarding the emergence of the giant connected component under link addition. For given degree exponent, the mean degree, equal to twice the number of links per node, is fixed, represented in terms of the Riemann zeta functions. So, the mean degree isn't manageable towards the value that we want. Although some revised versions of the configuration model were proposed, with the integer-valued minimum degree varied [11], it still had problems. The minimum degree couldn't go below 1.0, preventing the mean degree from fully controllable. In this thesis, we propose a new version of the configuration model networks with the mean degree fully controllable. We vary the fraction of small-degree nodes by introducing a real-valued parameter. This is just a slight modification of the algorithms to construct the configuration model scale-free networks in python. Through this process, we are able to investigate the growth of the LCC in the configuration model. The result obtained for the static model networks are compared with our results. The configuration model shows unusual phenomena. It has the giant component from the beginning of network growth, which grows linearly with the number of links until a specific number of links per node is reached. This contrasts with the static model. This result means that a giant component can exist even in very sparsely networked systems. Also, that means that we can improve robustness and stability of scale-free networks by changing the fraction of small-degree nodes. If we could find out the governing rule of these phenomena, we will be able to keep the functionality of networks damaged severely.

The early generation and linear growth of the LCC are seen first in simulations. We adjust the minimum degree controlling parameter (x_{min}), which will be introduced later, from 0.01 to 2 with interval 0.01. We made 50 ensembles for network size $\mathrm{N} = 10^6$, and 200 ensembles for network size $\mathrm{N} = 10^5$ and 10^4. We study analytically and numerically the size of the LCC in our configuration model. Numerical analysis helps us understand why these phenomena happen and we present analytical results to explain them.

2. Theoretical Background

Network systems are universal from the World Wide Web to social relationships [12], power grid [12], and even in biological cells. Elements are connected one another and interact. Therefore, representing a complex system as a network of nodes and links can be useful for understanding properties of the system [1,2]. Local connection properties are used to characterize the structural features of model or real-world networks. Here we introduce some of them as well as the frequently-used terminologies.

Degree

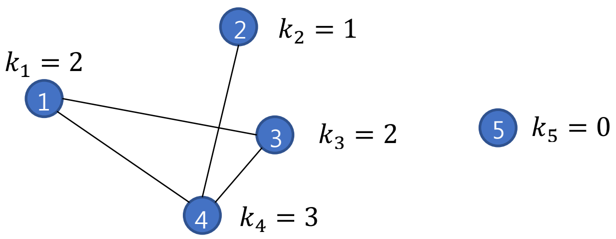

Figure 2.1 Degrees of nodes in a network.

The degree of a node is the number of links attached to it [12]. The degree of node i is usually denoted by k_i. The sum of the degrees over nodes is equal to twice the total number of links L,

$$\sum_{i=1}^{N} k_i = 2\mathrm{L}. \tag{2.1}$$

Excess degree

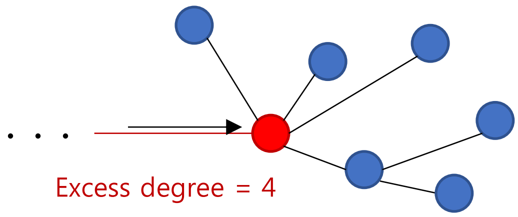

Figure 2.2 Excess degree of a node in a network.

If you choose a link randomly and follow that in one direction, you will meet a node at the end of the link. The number of links of the node except for the link you followed is called its excess degree. Namely, the excess degree of node i is equal to $k_i - 1$. The probability distribution $q(k)$ can be written in terms of $P(k)$ as

$$q(k) = \frac{(k+1)\,P(k+1)}{\langle k \rangle}. \tag{2.2}$$

Degree distribution

The degree distribution is the probability that a randomly-chosen node has degree k [12], denoted usually by $P(k)$.

Hub

Hub is a node which has a large degree.

Its influence is expected to be bigger than other small-degree nodes.

It can be a popular bridge when something moves from a node to another node.

Largest Connected Component (LCC)

A path is a sequence of adjacent links, and a connected component is a set of nodes any pair of which are connected by a path. The largest connected component is considered as a representative characteristic of the network structure. The LCC is called the giant component if its size S is comparable to the total number of nodes N,

$$\lim_{n \to \infty} \frac{S}{N} > 0. \tag{2.3}$$

Erdős–Rényi (ER) network

It's one of the random graph models [1]. This model starts with a fixed number of nodes. Suppose we have N nodes. Each distinct pair of nodes are connected with probability p. This procedure is repeated for all pairs of nodes. The mean number of links $\langle \mathrm{L} \rangle$ is proportional to the probability as

$$\langle \mathrm{L} \rangle = \frac{\mathrm{N}(\mathrm{N}-1)}{2} p. \tag{2.4}$$

Scale-free (SF) network

Scale-free network is a network whose degree distribution follows a power law [2]

$$\mathrm{P(k)} \sim k^{-\gamma}. \tag{2.5}$$

But It does not fully explain the meaning of 'scale-free'. Scale-freeness refers to a property of real networks. Real-world networks usually have $2 < \gamma < 3$. In this range of the degree exponent γ, the variance of the degree $\langle k^2 \rangle = \sum_k k^2 P(k)$ diverges. Therefore, degree has no defined variance. Hence the terminology 'scale-free'.

3. Model

3.1 Original configuration model for SF networks

We describe in this section the original configuration model for scale-free networks. The degree sequence is first generated to construct a network. Suppose that one is going to make a SF network with N nodes, and degree exponent γ.

i) Assign indices to nodes (i = 1,2,3, ⋯ ,N).

ii) Generate a random integer following $P(k) = \frac{k^{-\gamma}}{\sum_{1}^{\infty} k^{-\gamma}} = \frac{k^{-\gamma}}{\xi(\gamma)}$ for each node where $\xi(\gamma)$ is the Riemann zeta function $\xi(\gamma) = \sum_{i=1}^{\infty} i^{-\gamma}$.

iii) Choose two link stubs of distinct nodes and connect them.

iv) Repeat iii) until no isolated link stub is left.

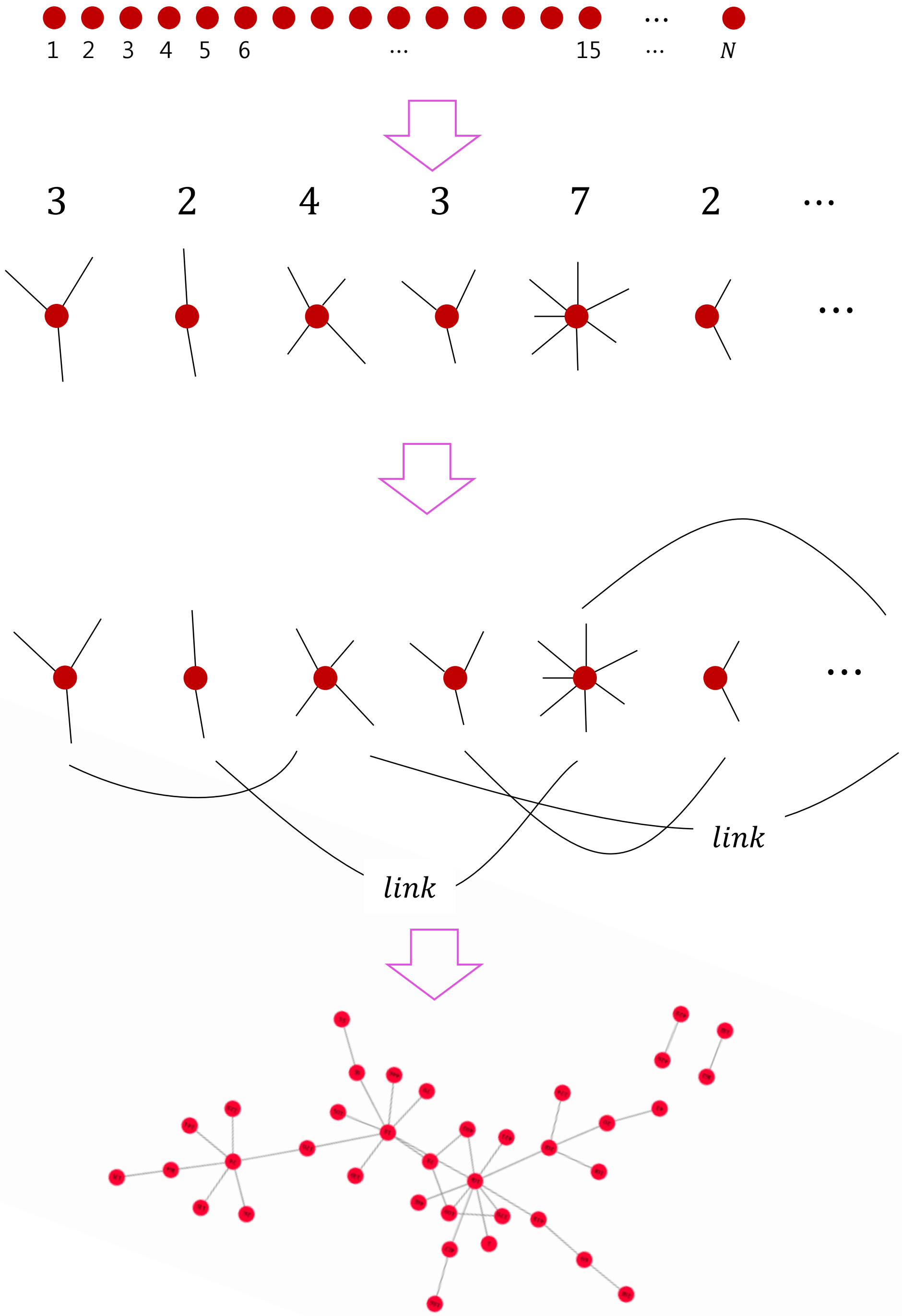

Figure 3.1.1 The process of composing the configuration-model SF network.

3.2 Static model for SF networks

In the static model for SF networks [9], each node has its probability to be chosen for connection with other nodes. Tuning the heterogeneity of the node selection probability, one can obtain a network of power-law degree distribution. This corresponds to a generation of ER network model. Let us generate a SF network with N nodes, L links, and degree exponent γ. Then the following procedures should be executed.

i) Assign indices to nodes ($\mathrm{i} = 1,2,3, \cdots, \mathrm{N}$).

ii) Choose and connect two nodes q and r probability p_q, p_r. The probability is given by

$$p_i = \frac{i^{-\frac{1}{\gamma-1}}}{\xi(\frac{1}{\gamma-1})}.$$

iii) Repeat ii) until the total number of links readies L.

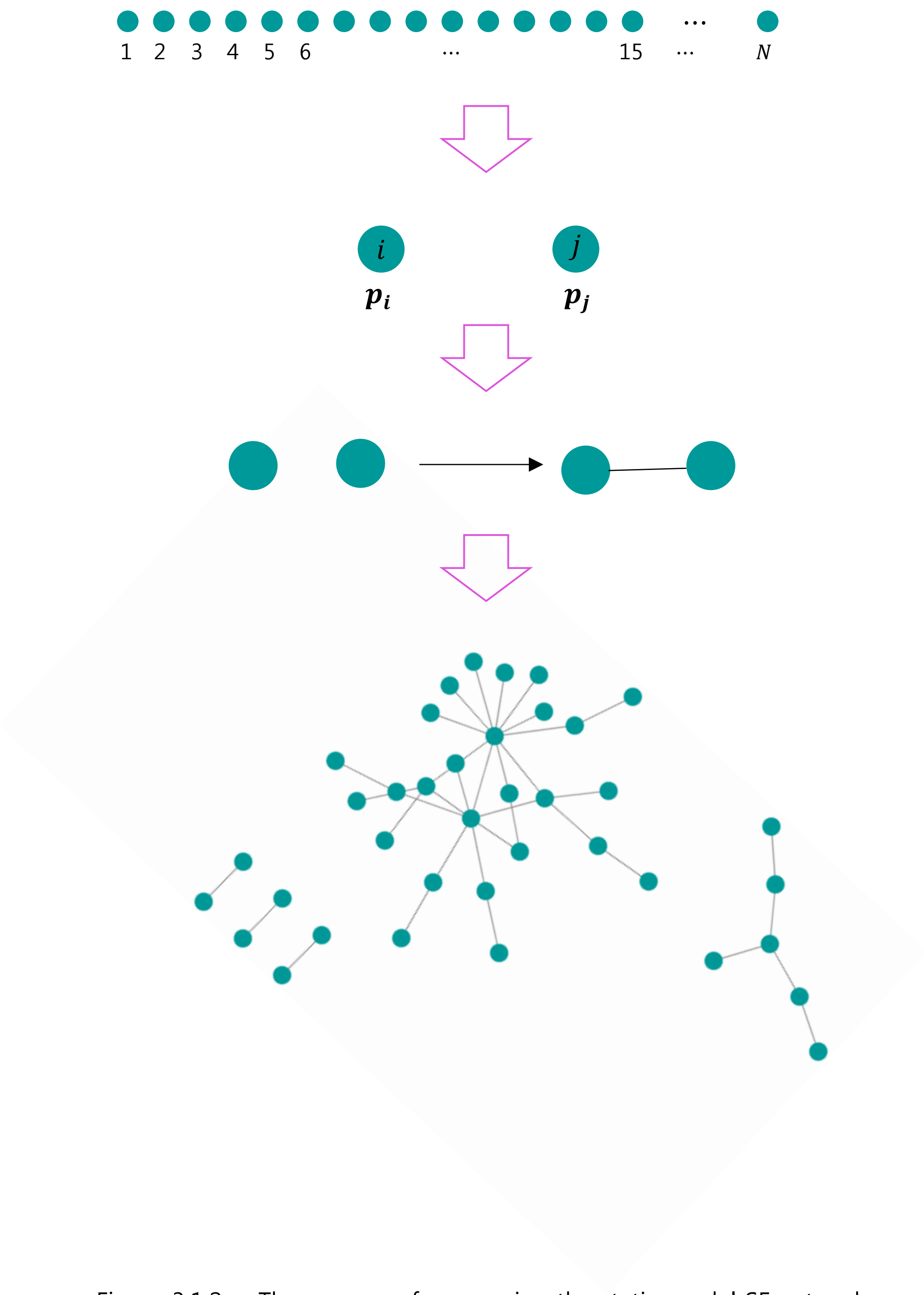

Figure 3.1.2 The process of composing the static-model SF network.

3.3 Our configuration model

In this thesis, we use a slightly modified version of the configuration model. We generate the degree sequence as follows. First, we extract a real-valued random x as

$$x = \left(x_{min}{}^{-(\gamma-1)} + (1-r)(x_{max}{}^{-(\gamma-1)} - x_{min}{}^{-(\gamma-1)}\right)^{-\frac{1}{\gamma-1}}, \qquad (3.3.1)$$

where r is a random number between 0 and 1. x is a random number between x_{min} and x_{max} follows a power-law distribution.

$$p(x) = \begin{cases} \left(\frac{\gamma}{x_{min}}\right)\left(\frac{x}{x_{min}}\right)^{-\gamma} & (x_{min} < x < x_{max}) \\ 0 & (x < x_{min} \ or \ x > x_{max}) \end{cases} \qquad (3.3.2)$$

We round down the number as $k = [x]$ because degree k is an integer. This is the way of determining the degree of a node. We iterate the above steps until all nodes have their degrees. We also check whether the sum of all degrees is an even number. Networks can have an unlinked edge if the sum is odd.

The minimum-degree controlling parameter, x_{min} is settled from 0.01 to 2 with an interval 0.01. x_{max}, the maximum-degree controlling parameter, is normally set to $\sqrt{N}$. But in this study, we set it differently to make it equal to that of the static model SF network.

$$x_{max} = 2\frac{\mathrm{L}}{\mathrm{N}}\frac{\gamma-2}{\gamma-1}\sqrt{\mathrm{N}} \tag{3.3.3}$$

It's the parameter used in Eq. (3.3.1). To do so, we need a reference value of L. We take it from the networks generated with $x_{max} = \sqrt{\mathrm{N}}$. We use Eq. (3.3.3) to compare various obtained results between the configuration model and the static model without the influence of different maximum degrees. Actually, x_{max} doesn't affect much the obtained results.

In simulations, we compose an ensemble of networks for N nodes, parameter x_{min}, and degree exponent γ. As the number of links per node K increases monotonically with x_{min}, the network ensemble can be considered as prepared for given N nodes, $\mathrm{L} = \mathrm{NK}$ links, and γ.

In this version of the configuration model, the degree distribution is given by

$$P(k) = \int_k^{k+1} P(x)dx = \begin{cases} \frac{(k+1)^{1-\gamma}-k^{1-\gamma}}{x_{max}{}^{1-\gamma}-x_{min}{}^{1-\gamma}} & (k \geq [x_{min}]), \\ 0 & otherwise. \end{cases} \tag{3.3.4}$$

Let us consider two situations when x_{min} is smaller than 1.0 and same or larger than 1.0 to evaluate the mean degree $\langle k \rangle = \sum kP(k)$.

i. $\boldsymbol{x_{min}}$ **< 1.0**

$$P(k=0) = \int_{x_{min}}^{1} P(x)\,dx = \frac{1 - x_{min}{}^{1-\gamma}}{x_{max}{}^{1-\gamma} - x_{min}{}^{1-\gamma}}$$

$$\approx\ 1 - x_{min}{}^{\gamma-1}, \tag{3.3.5}$$

$$P(k>0) = \int_k^{k+1} P(x)dx = \frac{(k+1)^{1-\gamma} - k^{1-\gamma}}{{x_{max}}^{1-\gamma} - {x_{min}}^{1-\gamma}}$$

$$\approx {x_{min}}^{\gamma-1}(k^{1-\gamma} - (k+1)^{1-\gamma}), \quad (3.3.6)$$

where we neglect ${x_{min}}^{1-\gamma}$.

The mean degree $\langle k \rangle$ is evaluated as

$$\langle k \rangle = \sum_{k=1}^{\infty} kP(k) = {x_{min}}^{\gamma-1} \sum_{k=1}^{\infty} \{k^{2-\gamma} - k(k+1)^{1-\gamma}\}$$

$$= {x_{min}}^{\gamma-1} \sum_{k=1}^{\infty} \{k^{2-\gamma} - (k+1)^{2-\gamma} + (k+1)^{1-\gamma}\}$$

$$= {x_{min}}^{\gamma-1}(1 + \textstyle\sum_{k=1}^{\infty}(1+k)^{1-\gamma})$$

$$= {x_{min}}^{\gamma-1}\xi(\gamma-1) \quad (3.3.7)$$

where ξ is the Riemann zeta function.

ii. $\boldsymbol{x_{min} \geq 1.0}$

$$P(k=0) = 0,$$

$$P(k=1) = \frac{2^{1-\gamma} - {x_{min}}^{1-\gamma}}{{x_{max}}^{1-\gamma} - {x_{min}}^{1-\gamma}}, \quad (3.3.8)$$

$$P(k \geq 2) = \frac{(k+1)^{1-\gamma} - k^{1-\gamma}}{{x_{max}}^{1-\gamma} - {x_{min}}^{1-\gamma}}. \quad (3.3.9)$$

The mean degree $\langle k \rangle$ is obtained similarly to the $x_{min} < 1$ case,

$$\langle k\rangle = \sum_{k=1}^{\infty} kP(k) = 1 - \left(\frac{x_{min}}{2}\right)^{\gamma-1} + x_{min}{}^{\gamma-1}\sum_{k=2}^{\infty}(k^{2-\gamma} - k(k+1)^{1-\gamma})$$

$$= 1 + x_{min}{}^{\gamma-1}(\xi(\gamma-1)-1). \qquad (3.3.10)$$

These results show how critical x_{min} is. All the equations depend on x_{min} influencing the connectivity of networks. Next, we consider the link-based degree distribution $\frac{kP(k)}{\langle k\rangle}$.

i. $\boldsymbol{x_{min}}$ **< 1.0**

$$\frac{kP(k)}{\langle k\rangle} = \frac{k\frac{(k+1)^{1-\gamma}-k^{1-\gamma}}{x_{min}{}^{1-\gamma}}}{x_{min}{}^{\gamma-1}\xi(\gamma-1)} = \frac{k\{(k+1)^{1-\gamma}-k^{1-\gamma}\}}{\xi(\gamma-1)} \qquad (3.3.11)$$

ii. $\boldsymbol{x_{min} \geq 1.0}$

$$\frac{kP(k)}{\langle k\rangle} = \frac{\frac{(k+1)^{1-\gamma}-k^{1-\gamma}}{x_{min}{}^{1-\gamma}}}{1+x_{min}{}^{\gamma-1}(\xi(\gamma-1)-1)} = \frac{k\{(k+1)^{1-\gamma}-k^{1-\gamma}\}}{x_{min}{}^{1-\gamma}+\xi(\gamma-1)-1} \qquad (3.3.12)$$

Comparing the two expressions for $\langle k\rangle$, we find a remarkable difference. x_{min} doesn't appear in Eq(3.3.11). This means that the excess degree distribution remains invariant as the number of links per node increases. As we will show, this invariance of $\frac{kP(k)}{\langle k\rangle}$ leads to an anomalous behavior of the giant component. Then, we observe the growth of the LCC in each network and take average over the networks in the ensemble.

4. Simulation results

4.1 Linear growth phase

In simulations of our configuration model, the LCC is shown to appear from the beginning and it grows linearly with the number of links per node $\mathrm{K}(=\frac{\mathrm{L}}{\mathrm{N}})$ until a specific point. This is well shown in Fig. 4.1.1; the LCC grows linearly with K until $\mathrm{K} \approx 0.81$ when $\gamma = 3.0$ and it changes its behavior after that point. It is contrasted to the growth of the LCC in the static model. In Fig. 4.1.2, the static model's LCC is quite small for K very small and it abruptly grows after passing a threshold. The failure to form a giant component for $\mathrm{K} \approx 0$ is a finite size effect [13]. If the size of static-model networks goes to infinity, a giant component would appear from the beginning.

Figure 4.1.3 shows the growth of the LCC in the configuration model for various degree exponents γ. The linear growth exists for $\gamma \leq 3.5$, but the range and the slope decrease with increasing the degree exponent. The range of K and x_{min} showing the linear growth is given in Table(4.1.1). We find that x_{min} values where the linear growth ends are always 1. This is related to the fraction of isolated nodes, which will be discussed below.

When $\gamma = 3.7$ and 4.0, the size of the LCC which is denoted as S is close to zero for K small as the growth of the LCC in the static model. The larger the degree exponent is, the wider the region of K showing $\mathrm{m}(=\frac{\mathrm{S}}{\mathrm{N}}) \approx 0$ is.

The point where such linear growth ends is where $x_{min} = 1.00$. For $x_{min} \geq 1$, $k_{min} = [x_{min}] = 1$, i.e., every node has at least one link. This means that when there is no isolated node, the linear growth does not hold. We measure how the proportion of isolated nodes, $P(k = 0)$, changes with the number of links per node K and present the result in Fig. 4.1.4. $P(k = 0)$ decreases linearly with K. This is observed for all γ showing the linear growth of m. The end points of these linear lines are the same as that terminating the linear growth of the LCC. This leads us to think that the density of isolated nodes affects the formation and growth of the LCC. We show the results for the static model in Fig. 4.1.5. $P(k = 0)$ in the static model behaves as a function of K smoothly, showing that the static model network has isolated nodes in the whole range of K that we consider.

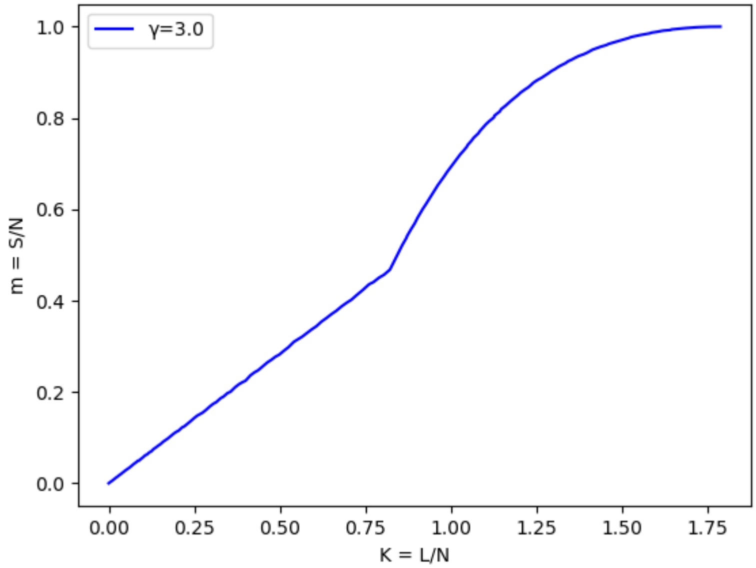

Figure 4.1.1 The relative size on $m = \frac{S}{N}$ of the LCC versus the number of links per node $K = \frac{L}{N}$ in the configuration model for $N = 10^6$ and $\gamma = 3.0$.

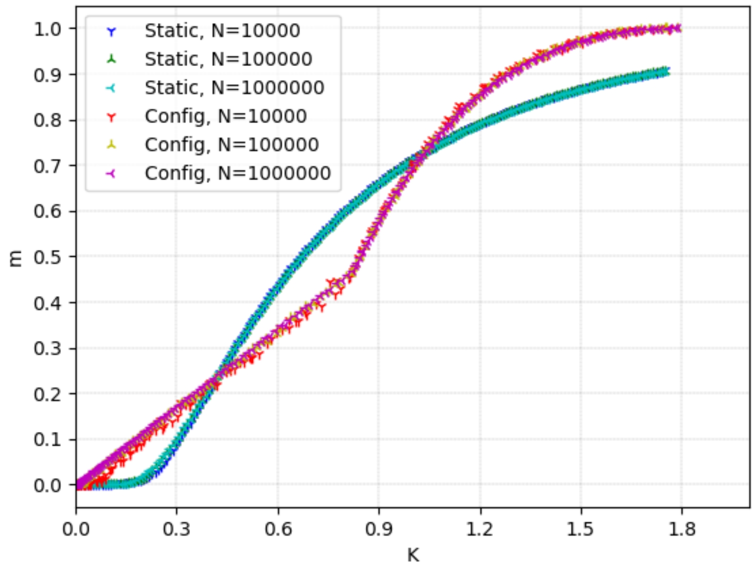

Figure 4.1.2 Growth of the LCC in the static model and the configuration model for $\gamma = 3.0$.

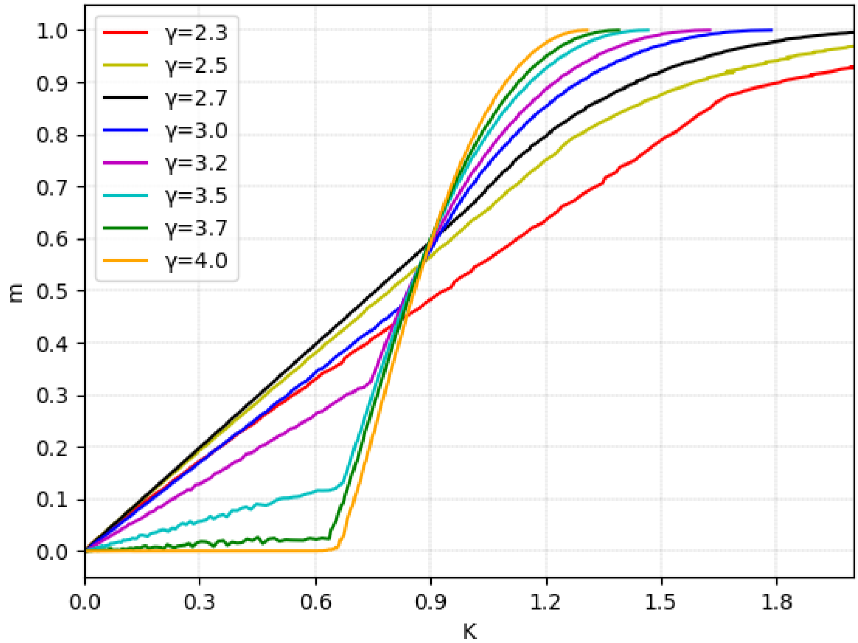

Figure 4.1.3 Growth of the LCC for various degree exponents and $N = 10^6$.

γ	K value where **linear line ends**	x_{min} value
2.3	**1.674**	**1.00**
2.5	**1.255**	**1.00**
2.7	**1.015**	**1.00**
3.0	**0.822**	**1.00**
3.2	**0.745**	**1.00**
3.5	**0.672**	**1.00**
3.7	**0.636**	**1.00**
4.0	**0.601**	**1.00**

Table 4.1.1 The number of links per node K and the model parameter x_{min} at which the linear growth of the LCC ends for various degree exponents γ and system size $N = 10^6$.

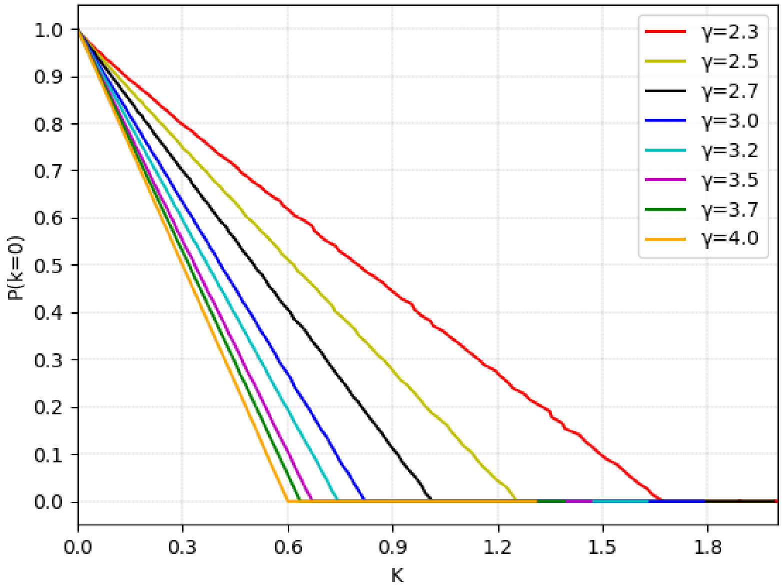

Figure 4.1.4 Density of $P(k=0)$ isolated nodes as a function of the number of links per node K for various degree exponents γ and $\mathrm{N}=10^6$.

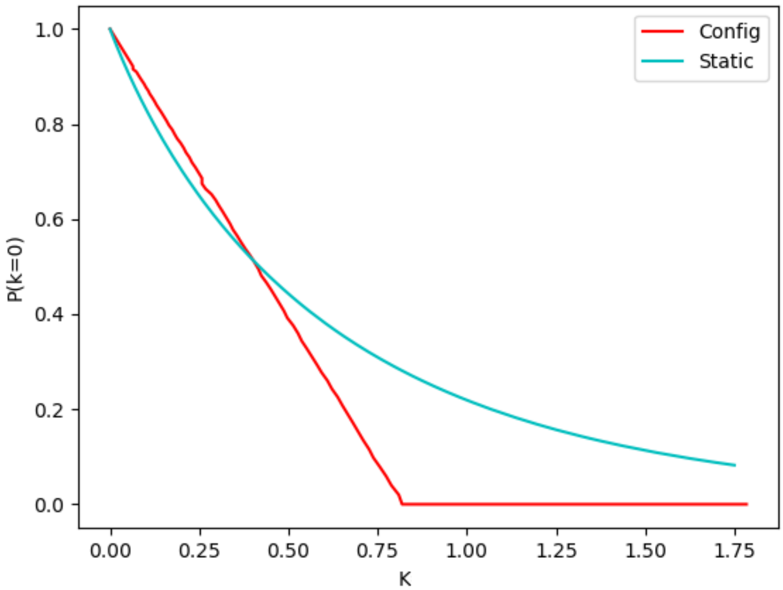

Figure 4.1.5 Density of isolated nodes in the configuration model and the static model for $\mathrm{N}=10^6$ and $\gamma=3.0$.

4.2 Degree distribution

To deepen our understanding, we investigate not only isolated nodes but also the degree distribution. We can understand the structure of the whole network by the degree distribution. The degree distribution is the ensemble-averaged one. For $\gamma = 3.0$, we choose two values of K with which the difference of the size of the LCC between the two models is quite large, $\mathrm{K} = {}^{\mathrm{L}}/_{\mathrm{N}} = 0.2$ and 0.7.

Firstly, Fig. 4.2.1 shows the degree distributions of the two models for $\mathrm{K} = 0.2$ and $\gamma = 3.0$. The fraction of high-degree nodes in the configuration model is much larger than that in the static model. On the contrary, the fraction of low-degree nodes is little bit smaller than the static model. It's a reasonable result to explain the size difference of the LCC between the two models. For $\mathrm{K} = 0.2$, the configuration model's LCC is much larger than the static model as you see in Fig. 4.1.2. So, high-degree nodes like hubs should be more abundant in the configuration model than in the static model. Although isolated nodes exist more in the configuration model as we can see Fig. 4.1.5, it's not critical as the difference of the fraction of high-degree nodes.

When $\mathrm{K} = 0.7$, the fraction of high-degree nodes in both models becomes similar. But difference is seen in the region of small degree. The fraction of isolated nodes in the static model is higher than the configuration model as we can see in Fig. 4.1.5. Also, in Fig. 4.2.2, $P(k = 2\sim8)$ is larger in the static model, making the degree distribution change its functional form for small k. The static model has a larger LCC at this value of K. Therefore, we can learn that a change in the proportion of small-degree nodes may affect the size of the LCC. For deeper understanding, we investigate the network structure and the cluster size distribution in the following section.

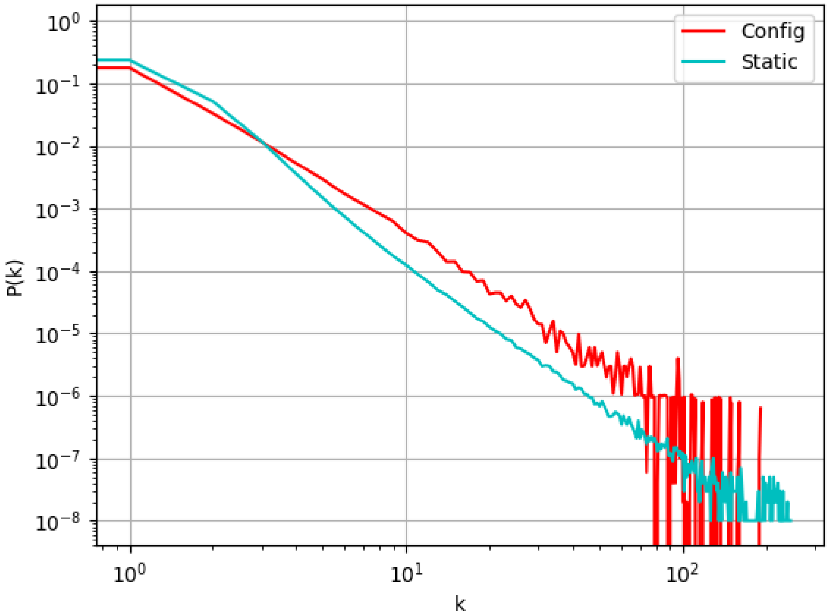

Figure 4.2.1 The degree distribution of for $K = 0.2$, $N = 10^6$, and $\gamma = 3.0$.

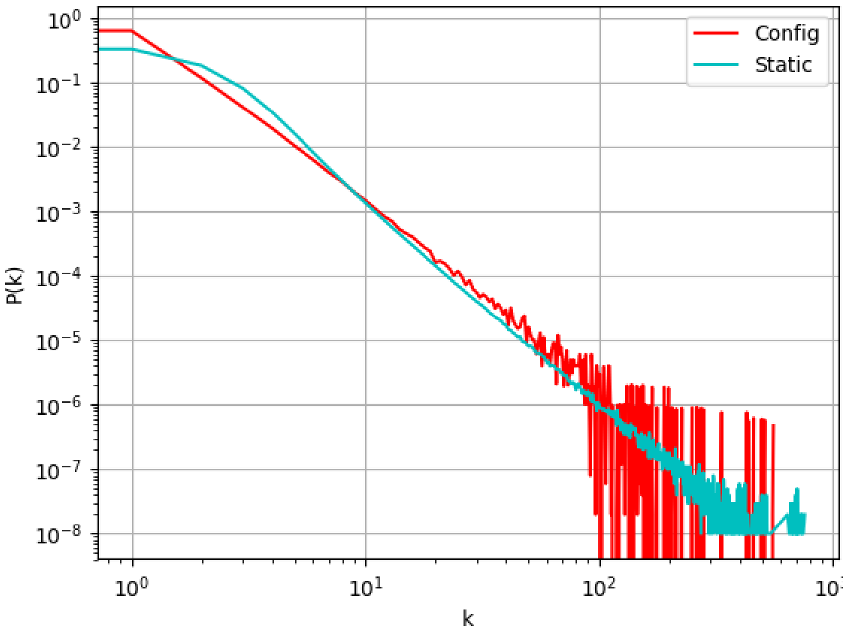

Figure 4.2.2 The degree distribution for $K = 0.7$, $N = 10^6$, and $\gamma = 3.0$.

4.3 Network structure

We visualize the network structure by using the network drawing tool 'Cytoscape' [15]. We draw small networks because large networks are too dense and complex to grasp. So, we make networks with $\mathrm{N} = 500$, $\gamma = 3.0$ in the configuration model and the static model. We use a layout 'degree sorted circle'. The layout gives us a whole view of the whole network structure. All figures in this section show one of networks among the ensemble.

Figure 4.3.1 shows a configuration-model network and Fig. 4.3.2 shows a static-model network when $\mathrm{K} = 0.2$, and $\gamma = 3.0$. The LCC and some large clusters are seen in the configuration model and more in quantity compared with the static model. This follows the tendency that we observed in Fig. 4.2.1. In these figures, isolated nodes are not included. So, we can notice that the number of isolated nodes in the configuration model is larger than the static model as we see in Fig. 4.1.5.

Figures 4.3.3 and 4.3.4 show the network structures of the two models for $\mathrm{K} = 0.7$, and $\gamma = 3.0$. The size of LCC is larger in the static model in contrast to the case of $\mathrm{K} = 0.2$. Both the number of isolated nodes and the number of degree-one nodes is larger in the configuration model than in the static model. It also consistent with the results in Figs 4.1.5 and 4.2.2.

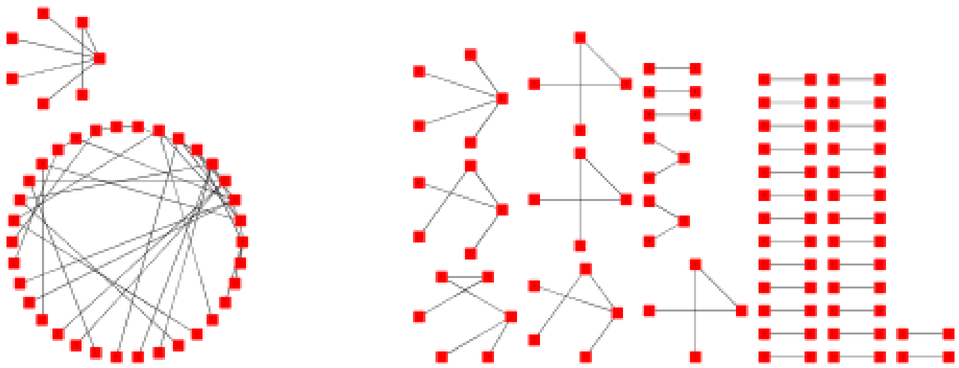

Figure 4.3.1 The structure of a configuration-model network for $K = 0.2$, $N = 500$ and $\gamma = 3.0$.

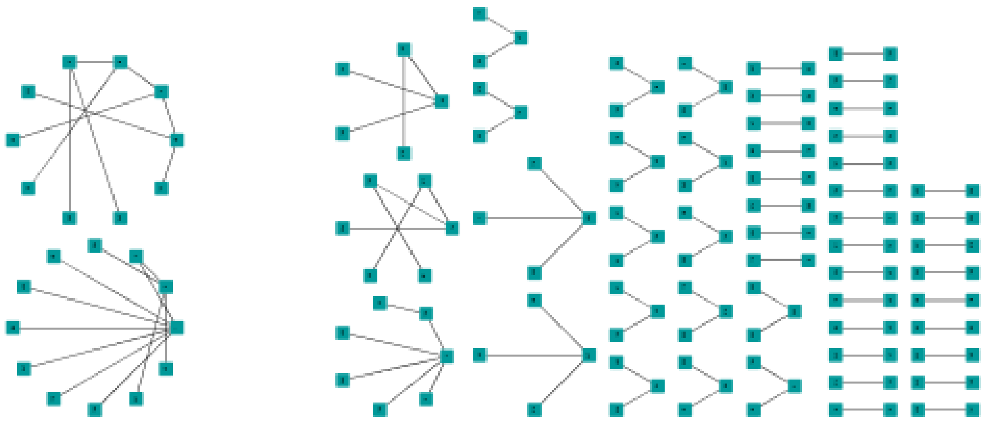

Figure 4.3.2 The structure of a static-model network for $K = 0.2$, $N = 500$ and $\gamma = 3.0$.

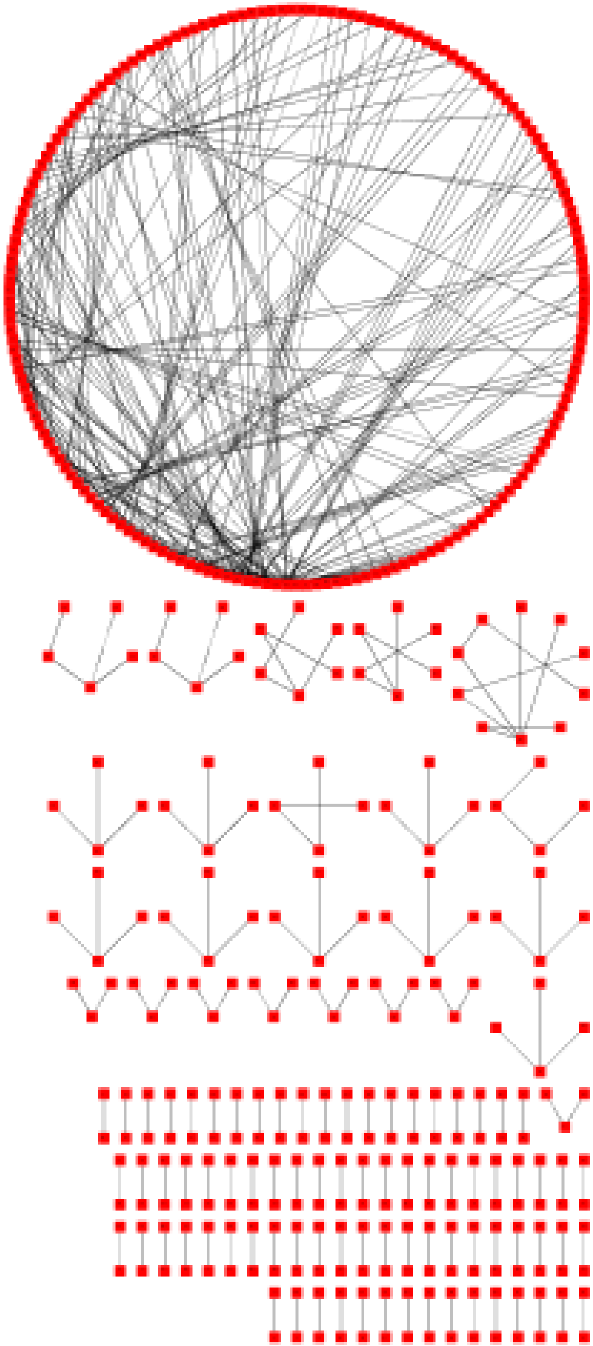

Figure 4.3.3 The structure of a configuration-model network for $K = 0.7$, $N = 500$ and $\gamma = 3.0$.

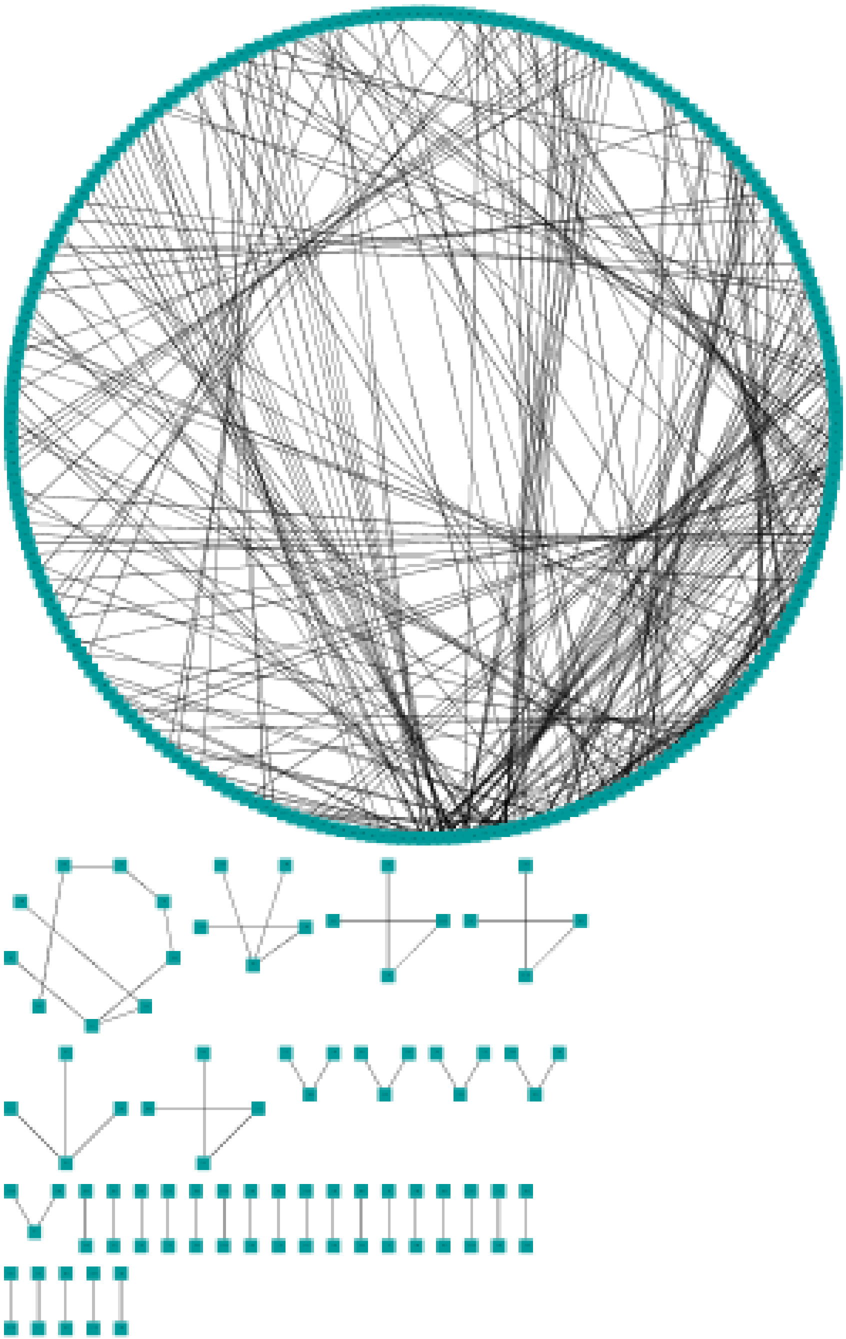

Figure 4.3.4 The structure of a static-model network for $K = 0.7$, $N = 500$ and $\gamma = 3.0$

4.4 Cluster-size distribution

The distribution of the size of connected components or clusters except the LCC is a key characteristic of network structure. We present the cluster-size distribution $P(s) = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}} \delta_{s_i,s}$ where s_i is the size of the cluster node i belongs to. Figure 4.4.1 shows cluster-size distribution of the two models when $\mathrm{K} = 0.2$, $\mathrm{N} = 10^6$, and $\gamma = 3.0$. A remarkable feature is the concentration of $P(s)$ at very small s in the configuration model compared with the relatively broad distribution in the static model. The existence of the giant component in the configuration model makes other clusters very small. The configuration model's maximum-cluster size except the LCC is about 60, however in the static model, it is about 1200.

Figure 4.4.2 shows the result for $\mathrm{K} = 0.7$. As the LCC of the static model grows larger than the configuration model, $P(s)$ is broader in the configuration model. The maximum-cluster size except the LCC is larger in the configuration model. These results are consistent with the network structures seen in Figs. 4.3.3 and 4.3.4

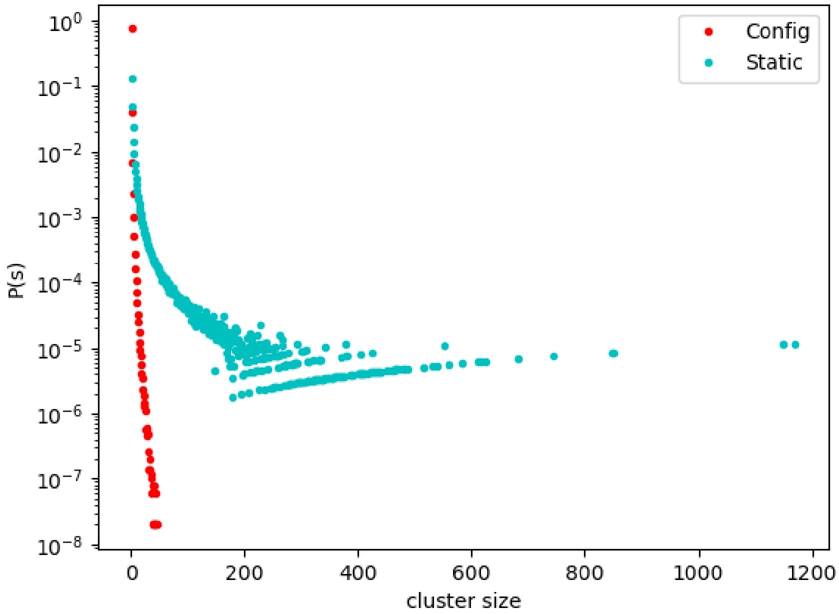

Figure 4.4.1 The cluster size distribution for $K = 0.2$, $N = 10^6$, and $\gamma = 3.0$.

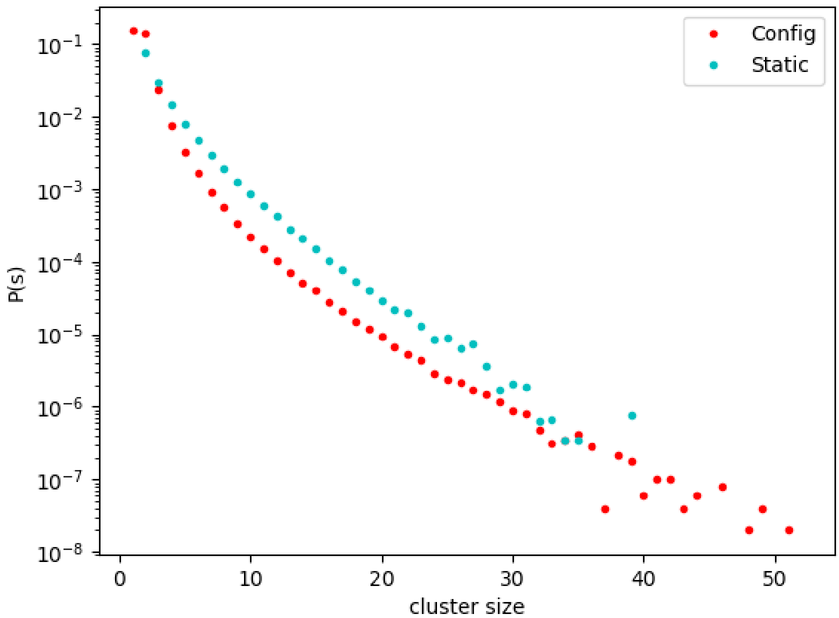

Figure 4.4.2 The cluster size distribution for $K = 0.7$, $N = 10^6$, and $\gamma = 3.0$.

5. Analytic results

We studied the process of the growth of networks and observed differences between two models. The numerical results suggest that the entire degree distribution affects the formation of the giant component. To better understand this, we derive the size of the giant component analytically by using the branching process approach [5].

5.1 Branching Process approach

Each node in a branching tree gives birth to k daughter nodes with branching probability $q(k)$ and the distribution of the size of this branching tree can be analytically obtained [14]. It can approximate the cluster-size distribution, which can be used to obtain the size of the giant component [5]. Let us consider the following:

π_s : the probability that a randomly chosen node belongs to a size-s cluster

ρ_s : the probability that a randomly chosen link leads to a size-s cluster

π_s is equal to $P(s)$ studied in Sec. 4.4, which is used in this chapter to avoid confusion with the degree distribution $P(k)$.

Then, the relative size of the LCC size is represented as

$$\mathrm{m} = \frac{\mathrm{S}}{\mathrm{N}} = 1 - \sum_{s=1}^{\mathrm{S}-1} \pi_s = 1 - h_0(1), \tag{5.1.1}$$

where S is the size of the LCC and we also defined two generating functions of π_s and ρ_s, respectively, as

$$h_0(z) = \sum_{s=1}^{\mathrm{S}-1} \pi_s z^s, \tag{5.1.2}$$

$$h_1(z) = \sum_{s=1}^{\mathrm{S}-1} \rho_s z^s. \tag{5.1.3}$$

For a given network, we have the following degree distributions:

$P(k)$: the probability that a node has k degree or the degree distribution

$q(k)$: the probability that a link leads to $k+1$ degree node or the excess degree distribution

The generating function of $P(k)$ and $q(k)$ are defined as follows:

$$g_o(z) = \sum_k P(k) z^k, \tag{5.1.4}$$

$$g_1(z) = \sum_k q(k) z^k. \tag{5.1.5}$$

Now we represent the cluster-size distribution π_s in a recursive way as

$$\boldsymbol{\pi_s} = \sum_k P(k) \sum_{s_1} \cdots\cdots \sum_{s_k} \rho_{s_1} \cdots\cdots \rho_{s_k} \, \delta_{(s_1+\cdots+s_k\,,\, s-1)}, \tag{5.1.6}$$

Where k is the degree of selected root node and s_i is the size of the cluster reached by the ith link of the root node.

Using Eq. (5.1.6) in Eq. (5.1.2), we obtain

$$h_0(z) = \sum_{s=1}^{S-1} \sum_k P(k) \sum_{s_1} \cdots \sum_{s_k} \rho_{s_1} \cdots \rho_{s_k}\, \delta_{(s_1+\cdots+s_k\,,\, s-1)}\, z^s$$

$$= z \sum_k P(k) \sum_{s_1} \cdots \sum_{s_k} \rho_{s_1} \cdots \rho_{s_k}\, z^{s_1+\cdots+s_k}$$

$$= z \sum_k P(k) \left[\sum_{s'} \rho_{s'} z^{s'} \right]^k$$

$$= z \textstyle\sum_k P(k)[h_1(z)]^k$$

$$= z\, g_0(h_1(z)). \tag{5.1.7}$$

Similarly, we obtain the self-consistent equation

$$h_1(z) = z\, g_1\big(h_1(z)\big). \tag{5.1.8}$$

From Eq. (5.1.1), one can obtain $\mathrm{m} = \frac{\mathrm{S}}{\mathrm{N}}$ by inserting $\mathrm{z} = 1$ in Eqs (5.1.7) and (5.1.8). Denoting $h_1(1)$ by u, we obtain the following equations:

$$h_0(1) = g_0(u), \tag{5.1.9}$$

$$u = g_1(u). \tag{5.1.10}$$

Solving Eq. (5.1.10) for u and using the solution in Eq. (5.1.9), we finally obtain the size of the LCC m. Figure 5.1.1 shows one example of finding numerically the solution to Eq. (5.1.10), in which $y = u$ and $y = g_1(u)$ intersect at the solution.

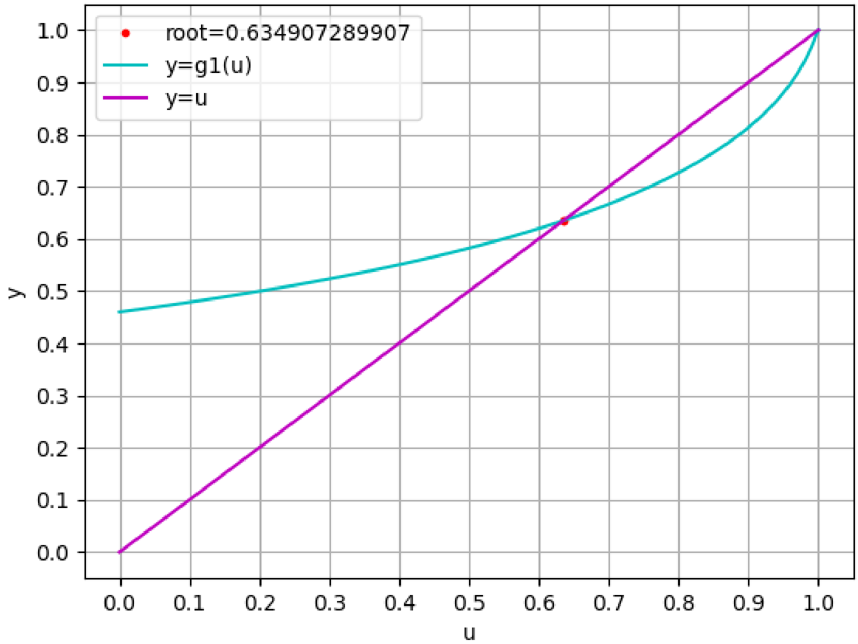

Figure 5.1.1 An example of solving Eq. (5.1.10) for u by plotting $y = g_1(u)$ and $y = u$. Red dot indicates the solution.

There is always a trivial solution $u = 1$ to Eq. (5.1.10). If there is only trivial solution, it makes $h_0(1) = 1$ and that indicates there is no giant component, $\mathrm{m} = 0$. We find that there is always a non-trivial solution $u < 1$ for all x_{min} in case of $\gamma = 3.0$. Figure 5.1.2 is the plots of Eq. (5.1.10) for $x_{min} < 1.0$. Figure 5.1.3 shows the plots for $x_{min} > 1.0$.

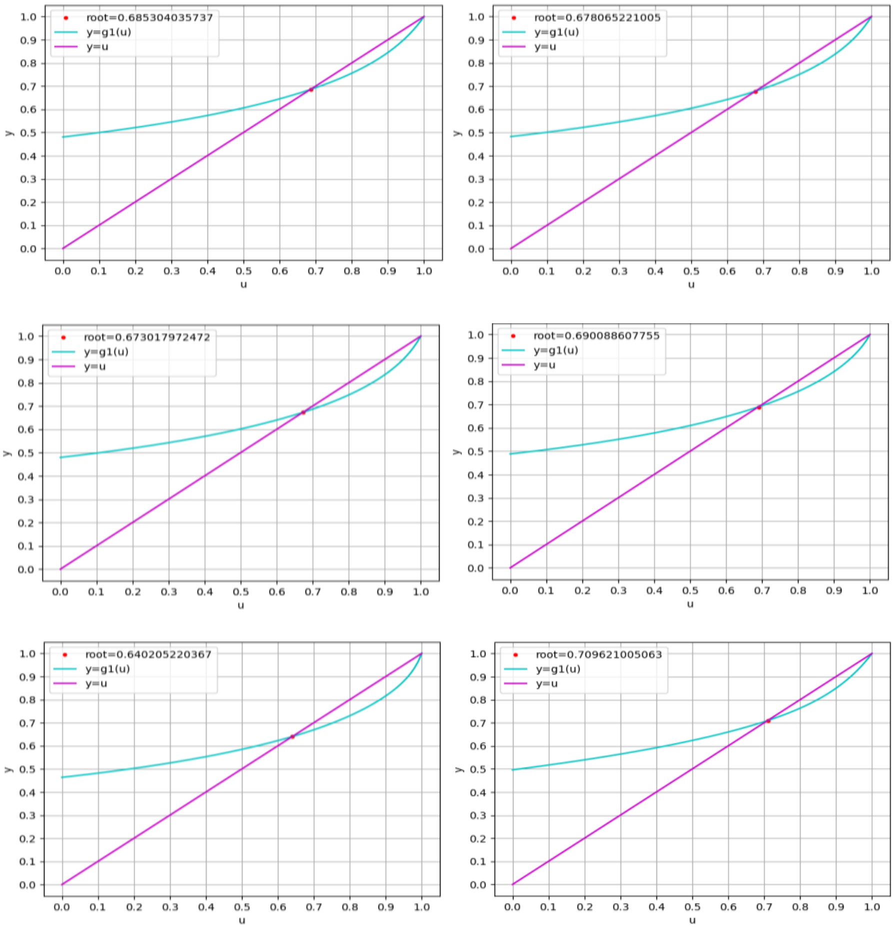

Figure 5.1.2 Plots of $g_1(u)$ and u for $x_{min} = 0.55, 0.6, 0.65, 0.7, 0.75, 0.8$ from left top to right bottom. The intersection points give the solution u to Eq. (5.1.10).

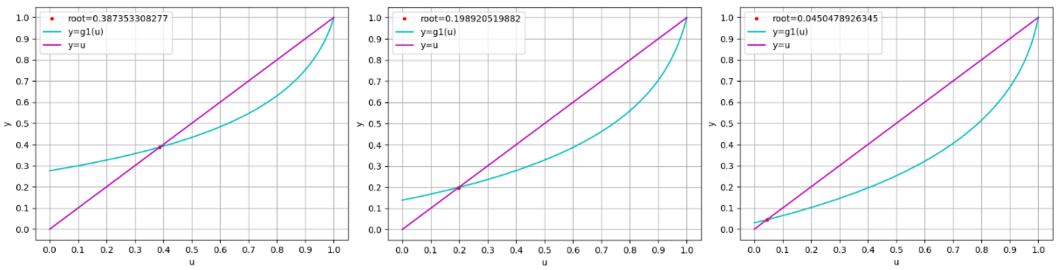

Figure 5.1.3 Plots of $g_1(u)$ and u for when $x_{min} = 1.3$,1.6, 1.9 from left to right.

The analytic solution to Eq. (5.1.9) agrees well with the simulation results as shown in Fig. 5.1.4. The linear growth of m with K is found to be due to the constancy of u in a range of K. Figure 5.1.5 shows that the solution u decreases rapidly from 1 to about 0.65 and then has almost the same value until $\mathrm{K} \approx 0.82$, which is the same as the end point of the linear growth of the LCC for $\gamma = 3.0$. In contrast, in Fig 5.1.6, static model doesn't show such stagnated region except for the beginning, where $u = 1$ and there is no giant component.

The invariance of u suggests that the function $g_1(u)$ or the excess degree distribution $q(k)$ does not change. So, we checked the degree distributions from $\mathrm{K} = 0.1$ to 0.8 with interval 0.1. Figures 5.1.6 and 5.1.7 show the degree distributions for different K in both models. A remarkable feature is that the configuration model's degree distributions have similar slopes in the whole range of k for different $\mathrm{K} = {}^{\mathrm{L}}/_{\mathrm{N}}$ while those of static models have variance in slope and curvature. Figures 5.1.8 and 5.1.9 show the excess degree distributions. In Figure 5.1.8, the configuration model's excess degree distributions $\frac{kP(k)}{\langle k \rangle}$ are all identical while those of the static model do not. This is the origin of the constant u for $0 < \mathrm{K} \leq 0.82$ and the linear growth of m with K.

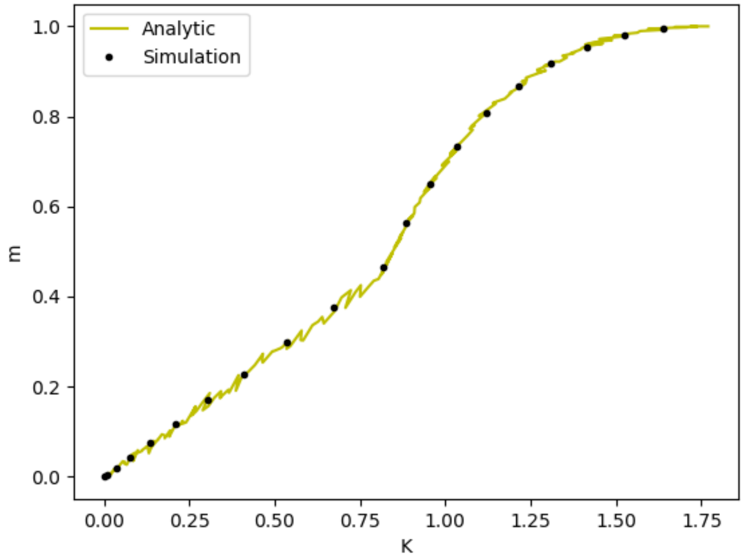

Figure 5.1.4 Plot of the relative size of the giant component from simulations (points) and analytic solutions (line) for $\gamma = 3.0$.

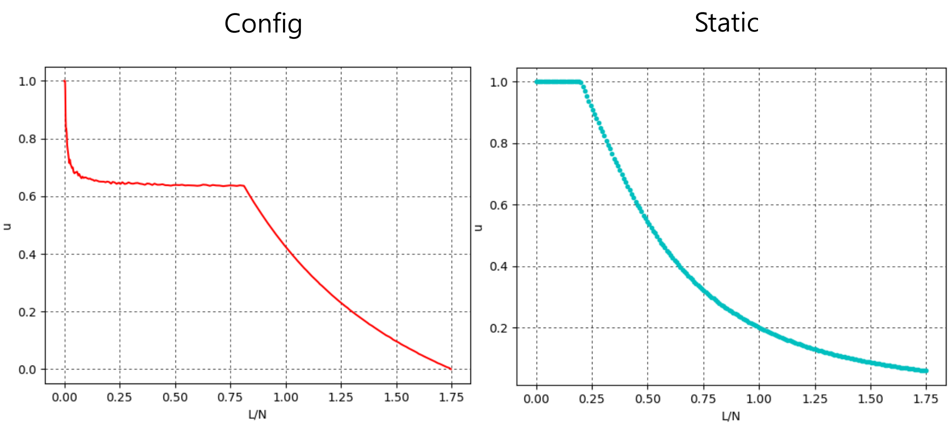

Figure 5.1.5 Plot of u versus $K = L/N$ in the configuration model (left) and the static model (right) for $N = 10^6$ and $\gamma = 3.0$.

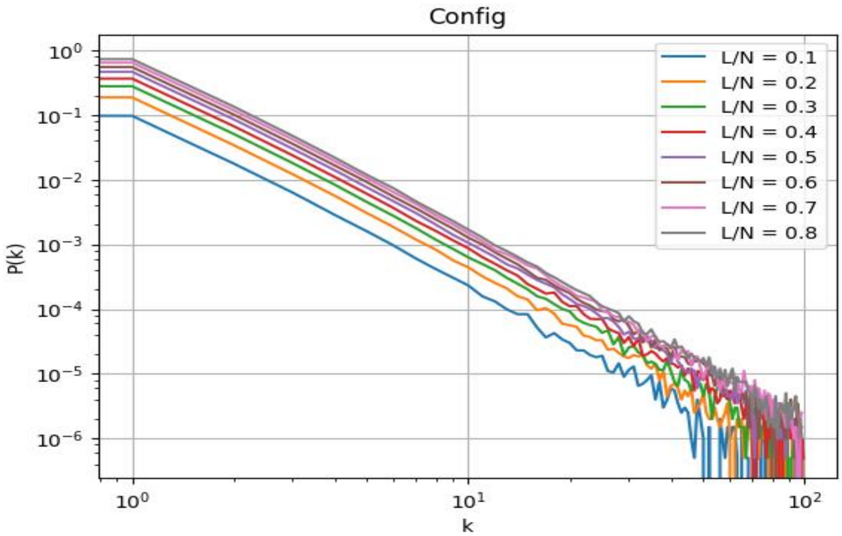

Figure 5.1.6 Degree distribution of the configuration model for $\mathrm{N} = 10^6$, $\gamma = 3.0$, and various $\mathrm{K} = \mathrm{L/N}$.

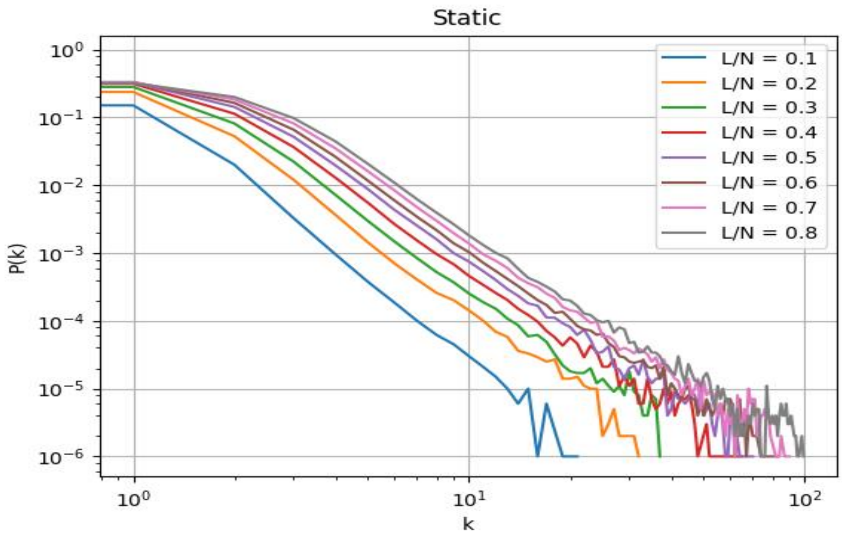

Figure 5.1.7 Degree distribution of the static model for $\mathrm{N} = 10^6, \gamma = 3.0$, and various $\mathrm{K} = \mathrm{L/N}$.

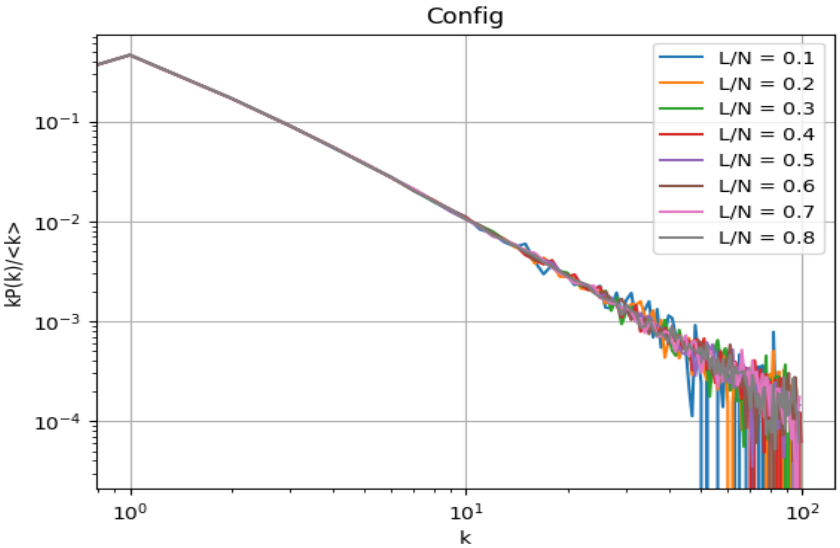

Figure 5.1.8 The link-based degree distribution $\frac{kP(k)}{\langle k \rangle}$ of the configuration model for $\mathrm{N} = 10^6, \gamma = 3.0,$ and various $\mathrm{K} = \mathrm{L/N}$.

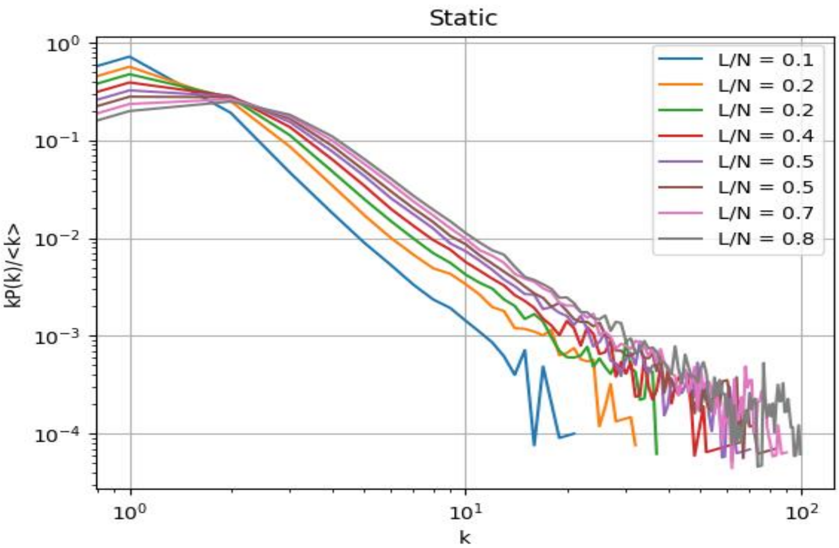

Figure 5.1.9 The link-based degree distribution $\frac{kP(k)}{\langle k \rangle}$ of the static model for $\mathrm{N} = 10^6, \gamma = 3.0,$ and various $\mathrm{K} = \mathrm{L/N}$.

6. Summary

Numerous networks in the real world have degree distribution in power-law form classified as scale-free networks. There are two representative models embodying the SF networks, the configuration model and the static model. The network robustness can be characterized by observing the largest connected component (LCC). In this thesis, we introduce a generalized version of the configuration model which overcomes the mean-degree limitation of the original version by introducing a real-valued parameter controlling the minimum-degree nodes. We find abnormal phenomena in the growth pattern of the LCC: The LCC grows linearly with increasing the number of links per node until a specific point. It is different from the LCC in the static model. To understand the origin of this phenomenon, we investigate the degree distribution, the density of isolated nodes, the cluster size distribution, and the network structure in comparison with the static model. We find that the origin lies in the different behaviors of the degree distribution in the small-degree region between the two models.

Then we present analytical solutions. We obtain the size of the LCC by using the generating function method. When the minimum degree is less than 1, the generating functions of the link-based degree distribution is invariant against the variation of the minimum degree parameter. All these peculiar features come from the degree distribution of the configuration model. The degree distribution of the configuration model rarely changes its power-law form, except for shift in its position as the mean degree is increased. This underlies the formation of the giant component even for quite small mean degree.

References

[1] P. Erdős and A. Rényi, “On the evolution of random graphs”, Publ. Math. Inst. Hung. Acad. Sci. 5, 17 (1960).

[2] A.-L. Barabási and R. Albert, “Emergence of scaling in random networks”, Science 286, 5439 (1999).

[3] A.-L. Barabási, R. Albert, H. Jeong, and G.Bianconi, “Power-law distribution of the world wide web”, Science 287, 2115 (2000).

[4] A.-L. Barabási and Marton Posfai, Network Science (Cambridge university press, 2016).

[5] Duncan S. Callaway, M. E. J. Newman, Steven H. Strogatz, and Duncan J. Watts, “Network Robustness and Fragility: Percolation on Random Graphs”, Rhys. Rev. Lett. 85, 5486 (2000).

[6] Yiping Chen, Gerald Paul, Reuven Cohen, Shlomo Havlin, Stephen P. Borgatti, Fredrik Liljeros, and H. E. Stanley, “Percolation theory applied to measures of fragmentation in social networks”, Phys. Rev. E 75, 046107 (2007).

[7] Reuven Cohen, Keren Erez, Daniel ben-Avraham, and Shlomo Havlin, “Resilience of the internet to random breakdowns”, Rhys. Rev. Lett. 85, 4626 (2000).

[8] Yeqian Feng, Bihui Sun, and An Zeng, “Cascade of links in complex networks”, Phys. Lett. A 381 (2017).

[9] K. -I. Goh, B. Kahng, and D. Kim, “Universal behavior of load

distribution in scale-free networks”, Phys. Rev. Lett. 87, 278701 (2001).

[10] Michele Catanzaro, Marián Boguñá, and Romualdo Pastor-Satorras, “Generation of uncorrelated random scale-free networks”, Phys. Rev. E 71, 027103 (2005).

[11] H.-H. Yoo and D.-S. Lee, “Relevance of the Minimum Degree to Dynamic Fluctuation in Strongly Heterogeneous Networks”, Journal of the Korean Physical Society, 72, 748 (2018).

[12] M. E. J. Newman, Networks: An Introduction (Oxford university press, 2010).

[13] D.-S. Lee, K.-I. Goh, B. Kahng, and D. Kim, “Evolution of scale-free random graphs: Potts model formulation”,Nucl. Phys. B 696, 351 (2004).

[14] R. Otter, “The multiplicative process”, Ann. Math. Statist.20, 206 (1949).

[15] Micheal E. Smoot, Keiichiro Ono, Johannes Ruscheinski, Peng-Liang Wang, and Trey Ideker, “Cytoscape2.8: new features for data intedgration and network vuisualization”, Bioinformatics.27, 431-432 (2011).