The Cut-Size Property of Networks and Epidemics Processes

Yingli Ni

4515021

The Cut-Size Property of Networks and Epidemics Processes

by

Yingli Ni

in partial fulfillment of the requirements for the degree of

Master of Science

at the Delft University of Technology,

to be defended publicly on Tuesday August 29, 2017 at 13:30 PM.

Student number:	4515021
Project duration:	December 1, 2016 – August 29, 2017
Thesis committee:	Prof. dr. ir. Piet.Van Mieghem
	Prof. Marco Loog
	Prof. Wioletta. M. Ruszel
Daily supervisor:	Ir.Qiang Liu

This thesis is confidential and cannot be made public until December 31, 2017.

Delft
University of
Technology

Abstract

‘*I think the next 21st century will be the century of complexity*’ , as Stephen Hawking predicted at the beginning of 2000, this prediction is coming true gradually with the technology development. The development of science is going through a dramatical and historical change because of the Internet, computer, and all other technologies. There are more and more complex systems existing in nature and the human society, attracting attention of researchers, naming: the economic system, the social organizations, and the human brain. The study of complex systems is to understand the relationship between the system structure and functions. In order to characterize the behavior of a system and to infer its topology features, mathematical models of the network are derived and techniques are applied to represent a wide variety of complex systems. Therefore, this method based on graph theory and statistical physics, committing to properties of the complex system and specific features is the so-called network science.

This master thesis focuses on the particular problem about the cut-size property of different networks. The scope of networks is from trivial network models (e.g., random graph) to real networks (e.g., Power grid network and Facebook network). The Susceptible-Infected-Susceptible (SIS) epidemic model can describe the spreading processes of information or diseases on networks. Within this thesis, we explore the cut-size property of networks and seek the relations between the cut-size property and the spreading behaviors. Our result deduces the cut-size property and the relevant physical meanings of the real networks. In such a way, a deeper understanding of the cut-size would help researchers to obtain insights of real networks. Our results may also contribute to the study of control of dynamic processes on networks.

Keywords: Network science, the cut-size property, epidemic process, real networks

Preface

This master thesis is the final work during my graduation project as a master student at Network Architecture and Service (NAS) group at Delft University of Technology. I started my graduation project from December 2016 and finished it in August 2017. All information and details about my graduation thesis will represent in the following report.

My graduation project is under the supervision of professor P. van Mieghem and Qiang Liu. First of all, I would like to thank my supervisors for giving me this opportunity to work with them at NAS group. With their encouragement and understanding, I feel more confident in the theoretical research of network science and graph theory.

Especially, I would like to thank my daily supervisor, Qiang Liu, for all his patience, feedbacks, and advice. Thanks for your encouragement and all best suggestions.

Furthermore, I want to thank all my colleagues at NAS group, and all my friends in the Netherlands, thanks for your company and your support. All good moments we shared together are the best gift in my whole life.

Finally, I want to express my deepest gratitude for my parents, my grandmother, and my family. For the last 20 years, they gave constant encouragement, and they sacrificed a lot for me. No matter what kind of difficulty I met, they are always behind me with the deepest love.

父兮生我，母兮鞠我，拊我畜我，长我育我，顾我复我，出入腹我。

感谢我的母亲，您是我努力学习的榜样
感谢我的父亲，您给予我厚重如山的爱
感谢我的祖母, 您赐予我最好的疼爱和童年
Thank you 感谢

Yingli
Delft, August 2017

Contents

List of Figures

Glossary

List of Symbols:

G	Graph
A	Adjacency matrix
Δ	Diagonal matrix
Q	Laplacian matrix
d_{av}	Average degree
d_i	Degree of node i
p	Network density
λ_i	Eigenvalue of adjacency matrix
σ_i	Eigenvalue of laplacian matrix
SIS	Susceptible-Infected-Susceptible
SIR	Susceptible-Infected-Recovered
β	Infection rate
δ	Curing rate
τ	Effective infection rate
τ_c	Epidemic threshold
W_i	Node state
NIMFA	The N-Intertwined Mean-Field Approximation
C	Cut-set
K	Selected set
S	Complement set
k	Selected set size
$\overline{\eta}(G,k)$	Average interconnection constant

N_{cut}	Normalized cut-set
$\overline{N}_{cut}$	Average normalized cut-set
SSIS	Simulator of Susceptible-Infected-Susceptible
$\mathbf{y}$	Simulated metastable prevalence
$\tilde{y}$	Estimated metastable prevalence

Chapter 1

Introduction

1.1 Background

Nowadays, network science is gradually playing a central role in different disciplines such as: computer science, finance, biology, and sociology [1]. The key challenge of network science is to understand the relationship between the system structure and functions. To characterize system behavior and to infer topology functions, the networks are applied to represent a wide variety of complex systems. The elements of a system are represented as the network nodes, and the interactions are represented as the network links [2]. Most of the networks which represent systems neglect lots of information of the system and only leave the primary connections to reduce the complexity. This network approach provides an abstract and simplified structure. In addition, the advantages of the simple network representation are the ease to obtain meaningful results and efficiency to analyze complex problems.

Graph theory is a powerful mathematical technique for measuring and analyzing the properties of the networks [3]. The history of graph theory started in 1736 from the Königsberg seven-bridge problem which is solved by Leonhard Euler [4]. This pace-tracing problem is to find the way traveling through the Königsberg city by crossing each of the seven bridges once and only once. Euler proved that it is impossible to find the way, that satisfies the conditions, by drawing the abstract graph which only consists of the city as nodes and the bridges as links. This example is the beginning of the graph theory which shows the potential of the reduction approaches in the network science. Many practical problems can be represented by graphs only containing nodes and links then can be analyzed and solved by the graph theory.

Along with the rapid development of the network science in 21 century, there are lots of interesting properties of networks, like degree distribution, shortest path, and cluster coefficient, etc., which are studied to present insights of networks [11]. In 1959, Paul Erdős and Alfréd Rényi firstly introduced the Erdős–Rényi model as the basic network model for generating the random graph. In 1998, Duncan Watts and Steven Strogatz identified that most real networks have a small shortest path and a large clustering coefficient, which is the so-called small-world network [6]. After that, Albert-László Barabási firstly used the scale-free network to describe the variety of networks with a power-law degree distribution in 1999 [7]. All these impressive results indicate that most of the real networks have the similar structures and provide a solid support to study other interesting properties of real networks.

If a graph partitions into two sets, the number of the links between two partitions is the cut-size of this graph. Different partition methods and the size of the partition have the influence on the cut-size. Moreover, the cut-size can describe the interaction between these two partitions, which we think is meaningful to gain the insight of networks. Based on the strong interest on the cut-size of the networks, we explore the cut-size property of real networks in this thesis. Also, we intensively discuss the physical meaning of the cut-size and the relation between the cut-size and the epidemic process on networks. The networks in this thesis include real networks, for example, the Europe Internet road, Facebook, power grid network, etc., and also includes basic network models, for instance, Erdős–Rényi random network [5] and Barabási–Albert scale-free network [7] as mentioned above.

The epidemic process as a special example is critical to understand the dynamic process on networks [9], like computing virus, information propagation on the Internet and other diffusion processes. The spreading of the virus can be modeled by the epidemic processes on networks. The susceptible-Infected-susceptible (SIS) epidemic model is a basic model of Epidemiology [10]. In the SIS model, each node has two states: healthy (susceptible) state and infected state. The infected nodes can infect its healthy neighbors with the infection rate β and the infected node can be cured with the curing rate δ. There is an epidemic threshold τ_c which separates two phases of the epidemic process on networks. When the effective infection rate $\tau=\frac{\beta}{\delta}$ is above the epidemic threshold τ_c, the virus spreads and the prevalence will keep stable after a long period of time. The prevalence is the average fraction of the infected nodes. When the effective infection rate τ is lower than τ_c, the virus dies out and extinct almost exponentially fast. The analysis of the relations between epidemic behaviors and the interaction plays a significant role in the study of the epidemic process. Through the whole epidemic process, different time stages have the variable cut-size and the key to understand the virus spreading process lies in the cut-size property of the networks, which might provide a way to control the epidemic spread on networks [12].

1.2 Research Goals

The research goals for this thesis is intending to studying the property of the cut-size and find its the relation with other properties of the networks. The cut-size property of the networks is related to the epidemic process, and if we know the cut-size distribution, then we can approximate the cut-size at this time. The cut-size can represent the spreading ability of the networks. In addition, we can estimate the prevalence by using the cut-size property and compare the estimated prevalences with the simulated prevalence in the metastable states. How the degree affects the cut-size property is also a significant problem we would like to understand. The physical meaning of the cut-size of the real networks is another goal of this thesis. From the macro and micro perspective, the study of the cut-size can not only be applied to the epidemic process but can also be applied to other dynamic processes.

1.3 Thesis Organization

The thesis contains five main chapters.

Chapter 2 introduces the background knowledge of networks models and graph theory used in this thesis.

Chapter 3 studies the property of the cut-size, like the cut-size distribution, and the normalized cut-size, etc., by using different partition methods. We deduce the bounds of the cut-size and the normalized cut-size by the isoperimetric inequality and the isoperimetric constant. By using the partition method based on the degree, we discuss how the degree affects the cut-size property. Finally, all these derivations and calculation results are combined to make a conclusion of the cut-size property of real networks. Real networks data are collected from the KONECT network collection, which is a project to collect large network datasets of all types, compiled by the Institute for Web Science and Technologies at the University of Koblenz–Landau [22]. We use the network data of Facebook, Internet European road connection, email communication data [28] and other networks.

Chapter 4 presents the relationship between the cut-size and the Susceptible-Infected-Susceptible epidemic process, deriving the estimated prevalence via the cut-size property. Also, by comparing the difference between the simulated metastable prevalence and the estimated metastable prevalence, we can discuss the factors which influence the accuracy of the estimation.

Finally, **Chapter 5** concludes this thesis. We also discuss the future work and relevant topics.

Chapter 2

General Network Theory

In this chapter, we introduce the basic mathematical notation used throughout this thesis, including relevant metrics and networks.

2.1 Graph theory and notations

To study the networks, the graph theory is playing the critical role to solve problems of inter-disciplines. Many practical problems can be represented by graphs that only contain nodes and links. A graph consists of links and nodes, showing the relations between the elements without additional information. Within this thesis, graph theory is the base for studying the networks and for analyzing the cut-size property.

A network is a graph *G(N,L)* where N is the number of nodes in graph G, L is the set of links, and l is the number of the links. In this thesis, the graph is undirected, unweighted and connected. There are several basic notations represent the graph *G(N,L)* as follows.

The adjacency matrix A is a $N \times N$ matrix containing the graph structure information. The element a_{ij} denotes the link between node i and node j. An element a_{ij} is the binary number

$$a_{ij} = \begin{cases} 1 & \text{if } (i,j) \in L \\ 0 & \text{if } (i,j) \notin L \end{cases} \tag{2.1.1}$$

Therefore, A is real and symmetric matrix, and the eigenvalue of the adjacency matrix A are $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N$.

The degree of node i is the number of neighbors of node i which is given by

$$d_i = \sum_{j=1}^{N} a_{ij} \tag{2.1.2}$$

Another essential matrix of a graph is the Laplacian matrix Q

$$Q = \Delta - A \tag{2.1.3}$$

where Δ=diag(d_1,d_2,...,d_N) is the diagonal matrix computed from the degree of each node d_i . Both diagonal matrix and adjacency matrix are real and symmetric, therefore the Laplacian matrix Q is also symmetric and real with all eigenvalues nonnegative and satisfying the order $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_N$=0. Here the number of zero eigenvalues is the number of connected components of the graph. A connected graph only has one component and σ_N=0. The second smallest non-zero eigenvalue σ_{N-1} of Q is also known as the algebraic connectivity of the graph.

2.2 Basic network models

With the rapid evolvement of real networks, the complexity of networks topology is growing a lot, and the computation is infeasible because of the huge network size. These complex networks can not be presented as simple network models, for example, a random graph. Therefore, in order to analyze the problems on real complex networks, other essential network models, that are found in complex networks, are applied to analyze real networks. This section introduces the basic network models concerned in this thesis, including Erdős-Rényi random graph, small-world network, and the scale-free network.

Erdős-Rényi random graph

In 1959, Paul Erdős and Alfréd Rényi firstly introduced the Erdős Rényi network [5]. This network is commonly used to study the property of random graph [5, 20]. The ER random graph has N nodes, and each node pair is connected independently with the probability p. With the increasing of link probability p, the graph has more links. The ER graph is a complete graph when the probability p equals to 1. The average links number of the ER graph is $E[l] = \frac{N(N-1)p}{2}$ and the average degree is $d_{av} = (N-1)p$. The degree distribution is,

$$Pr(d_i = m) = \binom{N-1}{m} p^m (1-p)^{N-1-m} \tag{2.2.1}$$

If $N \to \infty$ and Np is a constant, then the degree distribution of random graph is a Poisson distribution with a bell shape

$$Pr(d_i = m) = \frac{(Np)^m e^m}{m!} \tag{2.2.2}$$

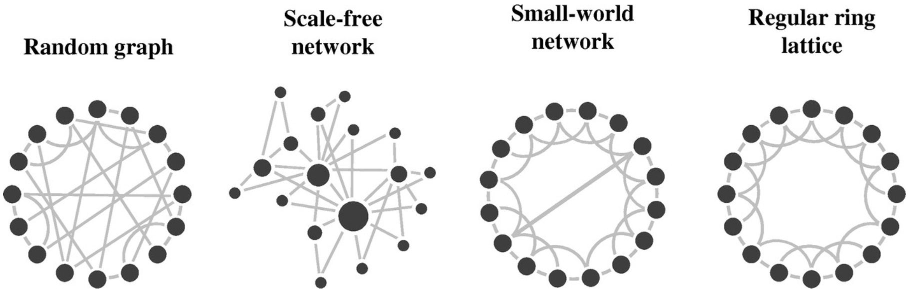

Figure 2.2.1: Example of random graph, scale-free network, small-world network [21].

Normally, the ER graph model is generated randomly. Any other properties observed in real networks cannot be reproduced by this model. Figure 2.2.1 illustrates different types of networks models, including the ER random graph, the scale-free network, and the small-world network. The first subgraph of the Figure 2.2.1 is the example of the ER random graph.

Small-world network

The small-worlds network illustrates the interesting phenomenon in real life that, for any two strangers, only six steps are needed to find their connection, which is the so-called *six degrees of separation* [16]. Since the first research by Watts–Strogatz [6], more and more real networks are found to have the small world properties. The small-world networks have a small average shortest path but a larger clustering coefficient, which makes small-world networks have the higher robustness to resist change and the shorter distance to relay information. The properties of the small-world network are between a random graph and a regular graph, whose each node has a constant degree due to the generating method. The third subgraph of Figure 2.2.1 shows an example of the small-world network.

Scale-free network

The scale-free network has a power-law degree distribution. In the scale-free network, most nodes have a very small degree and a few hubs have a very large degree. The degree distribution follows

$$p(k) \sim k^{-\gamma} \quad (2 < \gamma < 3) \tag{2.2.3}$$

The interests on scale-free networks started at the end of the 20 century. Albert-László Barabási defines the scale-free network to describe a series of networks follow a power-law distribution[17], and then the Barabási-Albert (BA) model becomes the most commonly used model to generate scale-free networks [18]. Before BA model, the heavy tail distribution is already found by Derek in citation networks [19]. Lots of real networks are also found following the power law degree distribution. The scale-free network model used in this thesis is generated via BA model. The generation of BA model starts from a few initial nodes and follows the degree-biased rule during each step. New nodes will be connected to existing nodes based on the degree. The nodes with a larger degree have a larger probability of connecting new nodes. The second subgraph of Figure 2.2.1 shows an example of the small-world network.

Except the network models mentioned above, some simple networks are also considered in this thesis as examples, including star graph, complete graph, and cycle graph. The star network with N nodes has one single central hub which is connected with $N-1$ leaf nodes. In a complete graph, all nodes are connected. The complete graph consists of $N(N-1)/2$ links and N nodes, and all nodes have the same degree $N-1$. The cycle network with N nodes only contain a single cycle, and every node has the degree 2. The above networks, especially the scale-free network and small-world network, can be seen as the bridge to promote the study of the cut-size property of real networks.

Chapter 3

The cut-size property of real networks

3.1 Overview

This chapter introduces the thesis contribution to the cut-size property of real networks, the bounds of the normalized cut-size and its average, the effect of degree on the normalized cut-size, and the physical meaning of the cut-size on specific real networks. This chapter firstly introduces the network datasets that are used in this thesis. In section 3.3, the definition and the mathematical properties of the cut-size are given and derived. Section 3.4 presents the result of the real networks by randomly partition methods. The physical meanings of the cut-size properties of the specific real network is analyzed and explained. The influence of the degree on the cut-size property by degree-based partition method is discussed in section 3.5. Section 3.6 concludes this chapter.

3.2 Dataset of real networks

Within this thesis, seven real networks are analyzed for studying the cut-size property. This section introduces these real networks. All these networks are collected from KONECT [22]. In addition, all selected networks are undirected and unweighted networks.

- US power grid network: This network is the power grid of the western United States, including N=4941 nodes and l=6594 links. The nodes represent power generators, transformers, and substations, and the links represent the voltage transmission lines between these power support stations [24]. Based on the paper by Duncan J. Watt [6], this network is a small-world network which has a large clustering coefficient and a small shortest path length. Table 3.2.1 shows the data of US power grid network.

- European E-road network: This network is the international road of the European cities, including N=1147 nodes that represent the cities, and l=1417 links represent the E-road within these cities. This network is neither a scale-free network nor a small-world network. This network has lots of communities based on different regions [28]. Figure 3.2.1 illustrates the structure of the European E-road network.

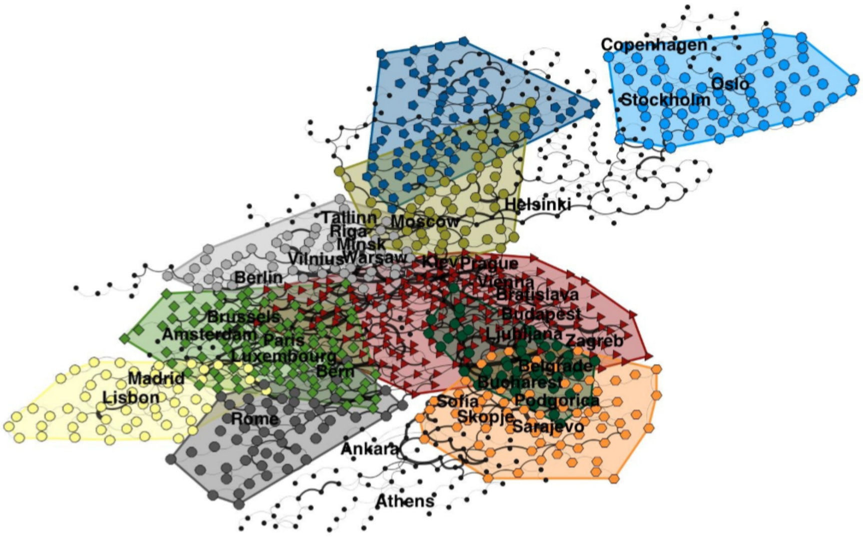

Figure 3.2.1: The European E-road network Community structure [28]

- Facebook network: This network is the Facebook user-to-user friendship network, consisting of N=2888 nodes that represent the users, and l=2981 links representing the friendship connection between users [25]. This network is an ego network which consists of the ego nodes and the directly neighborhoods. The Ego nodes have a larger degree and can be represented persons, groups, and organizations.

- Chicago network: This network is the transportation network of Chicago region with N=1467 transporting spots and l=1298 road connections between spots [26]. Although, in this network, the links are related to the traffic flow. Simplify the analysis, we assume this network is undirected and unweighted.

- Email communication network: This network is the email communication network at the University Rovira I Virgili in southern Spain. This network shows the evolution of social activities. This network contains N=1133 nodes and l=5451 links that represent the users and their communication, respectively [28]. Similar to the Facebook network, this network can reflect the human communication behaviors.
- Zachary karate club network: This network contains the N=34 nodes and l= 78 links that represent the members of the club and their relationships. This network has two communities because of the conflict between two teachers. Half of the members form a new club around one teacher, and members of another group find a new instructor or quit karate [29].
- Jazz musicians network: This network is the US Jazz musicians network at the beginning of 20th century. The network consists of 198 nodes and 2742 links, which denote the jazz musicians and the cooperation between musicians [30].

Network	**Nodes**	**Links**	**Network density**	**Cluster coefficient**	**Algebraic connectivity**	**Largest eigenvalue λ_1**
US power	4941	6594	0.00054030	0.103	0.00075921	7.4831
European E-road	1174	1471	0.0020579	0.0339	0.0011603	4.0104
Face-book	2888	2981	0.00071507	0.000359	0.0023771	27.803
Chicago transportation	1467	1298	0.0012071	0	0.0030897	4.9110
Email communication	1133	5451	0.0085002	0.166	0.33256	20.747
Zachary karate club	34	78	0.13904	0.256	0.46853	6.7257
Jazz Musician	198	2742	0.14059	0.52	0.57199	40.027

Table 3.2.1: Parameters of the real network involved

Table 3.2.1 presents the relevant properties. These datasets involve different types of real networks, like the social network, the human communication network, the infrastructure. All these network datasets are chosen because of the significant representatives of real complex networks. Therefore, the study of the cut-size property of these networks is meaningful.

3.3 Mathematical properties of the cut-size

In this section, we give the definition of the cut-size and mathematical properties.

A graph *G(N,L)* partitions into two node subsets, *K* and *S*. The cut-set is the set of links between two subsets. The cut-size is the number of links in the cut-set, which equals to

$$C = w^T Qw \tag{3.3.1}$$

where Q is the Laplacian matrix of graph G. The states vector as w=[W_1,W_2,...,W_N] and $W_i \in \{0,1\}$ is the Bernoulli random variable which represents the node partition that: if W_i =1, then node i belongs to the subset K, while if W_i =0, then node i belongs to the subset S. In this thesis, we define the nodes $v \in$ K as the selected nodes. The equation (3.3.1) gives a computable formula as the base to calculate the cut-size.

Figure 3.3.1 shows a particular example of a toy network which presents the relations between the cut-size and the epidemic process. The black nodes represent the infected nodes and the white nodes represent the healthy nodes. The cut-set is the set of red links connecting two subsets [12].

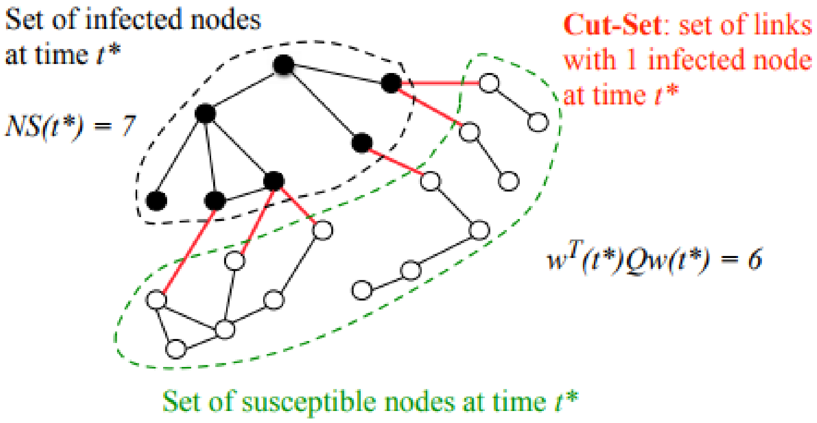

Figure 3.3.1: The epidemic state at moment t which shows three states part: a) the set of infected nodes at time t N(t)=7 as the black nodes, b) the set of susceptible nodes at time t N(t)=14 as the white nodes c) the cut-set, set of links with one end node in infected set d) the cut-size equals to C=6 [12].

3.3.1 The cut-size of simple networks

The star network consists of N nodes that only one central node is connected with N-1 leaf nodes. To study the cut-size distribution, we randomly partition the graph into two sets. Set K contains k nodes while another set S contains $N-k$ nodes. Each node has the same chance to be randomly selected into the set K with the probability $\frac{1}{k}$. The cut-size of each k has two different values according to the central node. If only leaf nodes are included within the selected set K, the cut-size equals to $C = k$ for each $k \in$ {1,...N-1} with the probability $\frac{N-k}{N}$. If the central node is included in the selected set k, the cut-size equals to $N-k$ for each $k \in$ {1,...N} with probability $\frac{k}{N}$. The probability is related to the network size N and the selected nodes number k if we know the selecting probability of the nodes.

A complete graph consists of $\frac{N(N-1)}{2}$ links and N nodes. All nodes are connected by links with the same degree *N-1*. For the complex graph, the cut-size is a function of the size of the selected set $|K| = k$. The cut-size is calculated as $(N-k)k$. Hence, for complete graph, the cut-size is determined when the number of the selected nodes k is known. Let us consider the cut-size of the cycle graph. The cycle graph consists of N nodes to conduct a single cycle. Every node of the cycle graph has the same degree 2. The cut-size distribution of a cycle graph is more complex than other that of a star graph or a complete graph, so it is hard to give the exact cut-size equation and probability for the cycle graph.

3.3.2 The normalized cut-size and its average

In this section, we discuss the average normalized cut-size derived by the average interconnection constant [38].

To derive the average normalized cut-size, we adopt the average interconnection constant which is introduced by Jasmina [38]

$$\overline{\eta}(G,k) = \frac{1}{\binom{N}{k}} \sum_{K \subset \{1,\ldots,N\}, |K|=k} \frac{e(K,S)}{|K|}, \tag{3.3.2}$$

where $e(K,S)$ denotes the number of links between the set S and the set K. The ratio $\frac{e(K,S)}{|k|}$ measures the connectivity between the set K and the set S. For each k, there exists $\binom{N}{k}$ combinations that lead to a very complex computation. Therefore, by averaging the ratio $\frac{e(K,S)}{|k|}$ over all $\binom{N}{k}$ combinations of the set with k nodes, we can obtain the average interconnection constant as

$$\overline{\eta}(G,k) = d_{av}\left(1 - \frac{k-1}{N-1}\right) = d_{av}\left(\frac{N-k}{N-1}\right) \tag{3.3.3}$$

where the d_{av} is the average degree of the graph. The average interconnection constant is determined for each k. The proof of the average interconnection constant (3.3.3), is given in Appendix A.

Recalling the definition of the cut-size C in previous section, the cut-size C varies with each k. To fairly compare the cut-size when k is from 1 to $N/2$, we define the normalized cut-size

$$N_{cut} = \frac{C}{|K||S|}, \quad K \subset \{1,...,N\} \quad 0 < k \leq \frac{N}{2} \tag{3.3.4}$$

If randomly select k nodes to formulate the set K, there are $\binom{N}{k}$ combinations leads to the varying cut-size and the varying normalized cut-size. Hence, we average the cut-size C over all combinations to obtain the average cut-size with k nodes as

$$E[C] = \frac{\sum_{K\subset\{1,...,N\},|K|=k} C}{\binom{N}{k}}, \quad K \subset \{1,...,N\} \tag{3.3.5}$$

where the average cut-size $E[C]$ is a constant for each k. Then, we calculate the average normalized cut-size for each k as

$$\overline{N}_{cut} = \frac{E[C]}{|K||S|}, \quad K \subset \{1,...,N\} \quad 0 < k \leq \frac{N}{2} \tag{3.3.6}$$

the average normalized cut-size is also a constant for each k. Because C=$e(K,S)$, the average interconnection constant is from (3.3.2)

$$\overline{\eta}(G,k) = \frac{1}{\binom{N}{k}} \sum_{K\subset\{1,...,N\},|K|=k} \frac{C}{|K|} \tag{3.3.7}$$

by substituting equation (3.3.5) into (3.3.7), we have

$$\overline{\eta}(G,k) = \frac{E[C]}{|K|} \tag{3.3.8}$$

by substituting equation (3.3.8) into (3.3.6), the average normalized cut-size is

$$\overline{N}_{cut} = \frac{\overline{\eta}(G,k)}{|S|} = \frac{d_{av}}{N-1} \tag{3.3.9}$$

where |S| equals $N-k$ and the ratio $\frac{d_{av}}{N-1}$ is the network density, which is the ratio between the number of links and the maximum possible number of links, i.e. $\frac{2l}{N(N-1)}$, where l is the number of the links of the graph.

As a result, the average normalized cut-size is derived, which is invariant for different value of k. For a graph, no matter how many nodes k are randomly selected to formulate the set K, the average normalized cut-size $\overline{N}_{cut}$ is fixed.

3.3.3 Bounds of the normalized cut-size

In this section, the bounds of the normalized cut-size is derived from the isoperimetric inequality [36, 37].

As introduced by Ganesh in [36], the so-called generalized isoperimetric constant of the graph $G(N,L)$ is

$$\eta(G,k) = \inf_{k\subset\{1,\ldots,N\},|K|\leq k} \frac{e(K,S)}{|K|}, \quad 0 \leq k \leq \frac{N}{2} \tag{3.3.10}$$

when $k=\frac{N}{2}$, the generalized isoperimetric constant $\eta(G,k)$ corresponds to the standard isoperimetric constant $\eta(G)$. For any graph $G(N,L)$, this standard isoperimetric constant follows the inequality

$$\eta(G) \geq \frac{\sigma_{N-1}}{2} \tag{3.3.11}$$

where σ_{N-1} is the algebraic connectivity, i.e. the smallest non-zero eigenvalue of the Laplacian matrix Q of the graph G. We replace the $e(K,S)$ in equation (3.3.10) by the cut-size C and then obtain

$$\eta(G,k) = \inf_{K\subset\{1,\ldots,N\},|K|\leq k} \frac{C}{|K|}, \quad 0 \leq k \leq \frac{N}{2} \tag{3.3.12}$$

by combining the equation (3.3.4) with equation (3.3.12), we can obtain

$$\eta(G,k) = \inf_{K\subset\{1,\ldots,N\},|K|\leq k} N_{cut}|S|, \quad 0 < k \leq \frac{N}{2} \tag{3.3.13}$$

when k=$\frac{N}{2}$, we can derive a lower bound of the normalized cut-size

$$N_{cut} > \frac{\sigma_{N-1}}{2N} \tag{3.3.14}$$

this low bound of the normalized cut-size is only determined by the eigenvalue σ_{N-1}.

Another lower bound and an upper bound can be derived by the isoperimetric inequality [37]. For a graph *G(N,L)* partitions into two sets *K* and *S*, the cut-size *C* between two sets satisfies:

$$\left| C - \frac{d_{av}|K||S|}{N-1} \right| \leq \frac{\theta}{N-1}\sqrt{|K|(N-|K|)|S|(N-|S|)} \tag{3.3.15}$$

where $\theta \geq max(|d_{av} - \sigma_i|)$ for $i \neq N$ and σ_i is the eigenvalue of Laplacian matrix Q of graph G that satisfies $\sigma_1 \geq ... \geq \sigma_{N-1} \geq \sigma_N$=0. For a connected graph, σ_{N-1} is the smallest nonzero eigenvalue. The inequality is simplified for a 2-partition by dividing |K| and |S| on both sides:

$$\left| \frac{C}{|K||S|} - \frac{d_{av}}{N-1} \right| \leq \frac{\theta}{N-1} \tag{3.3.16}$$

and then the first term in the left-hand side is the normalized cut-size and the second term is the average normalized cut-size. By substituting equation (3.3.4), we obtain bounds of the normalized cut-size

$$\left| N_{cut} - \frac{d_{av}}{N-1} \right| \leq \frac{\theta}{N-1}, \tag{3.3.17}$$

The parameter θ in (3.3.17) determines the upper bound and lower bound. For the Laplacian matrix Q, the diagonal element is degree of each nodes and the sum of the eigenvalues equals to the trace of Q. Thus, the average degree d_{av} is the average of all eigenvalues σ_i, and d_{av} is between σ_1 and σ_{N-1}. Therefore, $max(|d_{av} - \sigma_i|)$ equals to $max(|d_{av} - \sigma_1|, |d_{av} - \sigma_{N-1}|)$. The parameter θ is only related to the smallest nonzero eigenvalue σ_{N-1} and the largest eigenvalue σ_1, we have

$$\frac{d_{av} - \theta}{N-1} \leq N_{cut} \leq \frac{\theta + d_{av}}{N-1} \tag{3.3.18}$$

where θ=$max(|d_{av} - \sigma_1|, |d_{av} - \sigma_{N-1}|)$ determines bounds. Because $N_{cut} > 0$, a more precise inequality is derived as

$$\frac{max(d_{av} - max(|d_{av} - \sigma_1|, |d_{av} - \sigma_{N-1}|), 0)}{N-1} \leq N_{cut} \leq \frac{d_{av} + max(|d_{av} - \sigma_1|, |d_{av} - \sigma_{N-1}|)}{N-1} \tag{3.3.19}$$

This inequality provides a different lower bound of the normalized cut-size N_{cut} comparing to equation (3.3.14). If $|d_{av} - \sigma_1| > |d_{av} - \sigma_{N-1}|$, then the lower bound from (3.3.19) is zero and the bound (3.3.14) is tighter. If $|d_{av} - \sigma_{N-1}| > |d_{av} - \sigma_1|$, then the lower bound from (3.3.19) is tighter comparing to (3.3.14). Both bounds are related to the eigenvalue σ_1 or σ_{N-1} of the Laplacian matrix Q of graph G.

This section focuses on the mathematical property of the cut-size. The average normalized cut-size is derived, which is a constant determined by the network density or average degree. Meanwhile, the upper and lower bounds of the normalized cut-size are obtained by the isoperimetric inequality (3.3.19). The above knowledge about the cut-size provide the solid theoretical base for studying the cut-size property in real networks.

3.4 The cut-size of real networks

In this section, we discuss the cut-size property of several real networks which are introduced in section 3.2. The cut-size distribution and the theoretical bounds of the normalized cut-size are discussed in this section. The physical meaning of the cut-size for the specific real network is explained. In addition, the result of the random graph and the scale-free network are given in Appendix B as the supplement.

In order to obtain the cut-size and related properties, we calculate the cut-size by randomly partitioning the graph into two sets K with k nodes and S with $N-k$ node. If $W_i = 1$, then node i belongs to set K, otherwise, node i belongs to S. Based on the state of the network nodes, we formulate the state vector as w=$[W_1, W_2, ..., W_N]$. Then, by using equation (3.3.1), we calculate the cut-size of the set K. For each k, the random partition process is repeated for $m = 10^5$ times. By using equation (3.3.4), we can gain the maximum normalized cut-size, the minimum normalized cut-size as well as the average normalized cut-size, respectively. With k increase from 1 to N, the normalized cut-size of the whole graph is obtained.

3.4.1 Results of real networks

In this section, we present our results of real networks. The result of three real networks are shown in the following part of the section. There are some common properties observed from all the real networks. Results of the rest networks are given in Appendix B.

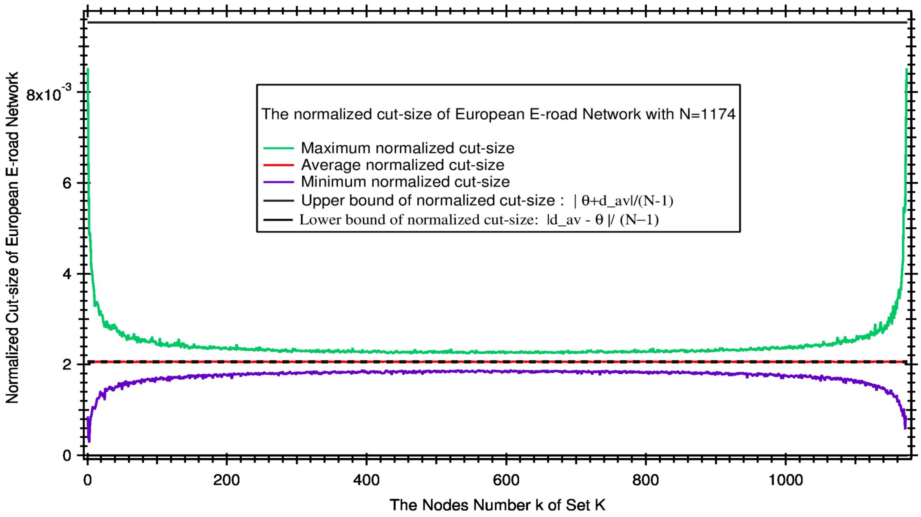

Figure 3.4.1: The normalized cut-size of the European E-road network

Figure 3.4.1 shows the result of the Europe E-road network. The red curve is the average normalized cut-size constant $\overline{N}_{cut}$ of this network and the black dash line is the network density of the E-road network. From equation (3.3.9), the average normalized cut-size is a constant related to the average degree and the network size. The average normalized cut-size equals to $\overline{N}_{cut} = 0.002$. The network density of the European E-road network is $p = 0.0020579$ which is shown in table 3.2.1. No matter how you many nodes are selected in set K, the average normalized cut-size is always a fixed number which equals to the network density. For each exact k value, the average cut-size can be derived based on this property. The green curve is the maximum normalized cut-size. When k=1, the maximum normalized cut-size has a large value because the distribution of the cut-size with only one node selected is actually the degree distribution. Moreover, with the increase of k, the maximum normalized cut-size approaches to the average normalized cut-size. However, the change slows down when approaching the average normalized cut-size.

For an increasing k, the number of new links added into the cut-size declines. The purple curve is the minimum normalized cut-size shown in Figure 3.4.1. The minimum normalized cut-size has an opposite changing trend in which the minimum cut-size continuously increases until approximately to the average normalized cut-size. When k is close to 1, the minimum normalized cut-size increases fast. Then the increasing speed decreases with k. Both the minimum and maximum normalized cut-size are close to the average normalized cut-size, indicating a centric phenomenon that the minimum and maximum cut-size near the mean cut-set size with the small variance. In addition, the changing rate of both the maximum and minimum normalized cut-size decreases. The two black curves as shown in Figure 3.4.1 represent the upper bound and lower bound for the normalized cut-size, which calculated by equation (3.3.19), respectively. These bounds are fairly loose comparing to the green and purple curves. However, the range of the normalized cut-size in the real network is much smaller as shown in Figure 3.4.1. Therefore, the theoretical bounds are inaccurate and can not represent the distribution of the normalized cut-size well since the value of the normalized cut-size concentrates to the average normalized cut-size.

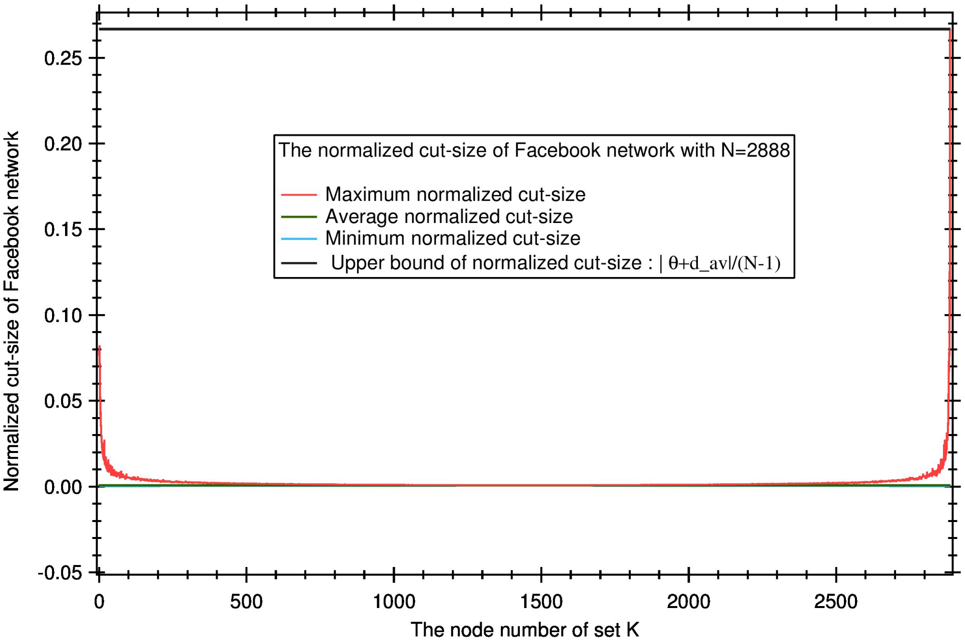

Figure 3.4.2: The normalized cut-size of the Facebook network

Figure 3.4.2 shows the result of the Facebook network. Facebook network contains different communities based on the users age, interests, gender or other user behaviors. The hub nodes present the nodes with lots of connections in a community, and the edge node present the nodes with few connections in a community. Comparing to the result of the Europe E-road network, the property of the normalized cut-size is entirely different. As shown in Figure 3.4.2, the red curve is the maximum normalized cut-size of the Facebook network. With a small number of the selected nodes, the maximum normalized cut-size rapidly reduces, which indicates that the Facebook network only contains a few nodes with a large degree. With the increase of k, the maximum normalized cut-size slowly approaches to the average normalized cut-size, which equals to $\overline{N}_{cut} = 0.00071$. The variation of the maximum normalized cut-size is huge, while the minimum normalized cut-size almost equals to the average normalized cut-size. The minimum normalized cut-size is small due to plenty of 1-degree nodes, and the connections between different subsets are weak. The real normalized cut-size has a smaller variation comparing with the theoretical upper and lower bounds derived by (3.3.19). From our result, relocating a high-degree node from one node subset to another one leads to a large change on the normalized cut-size comparing to relocating a small-degree node.

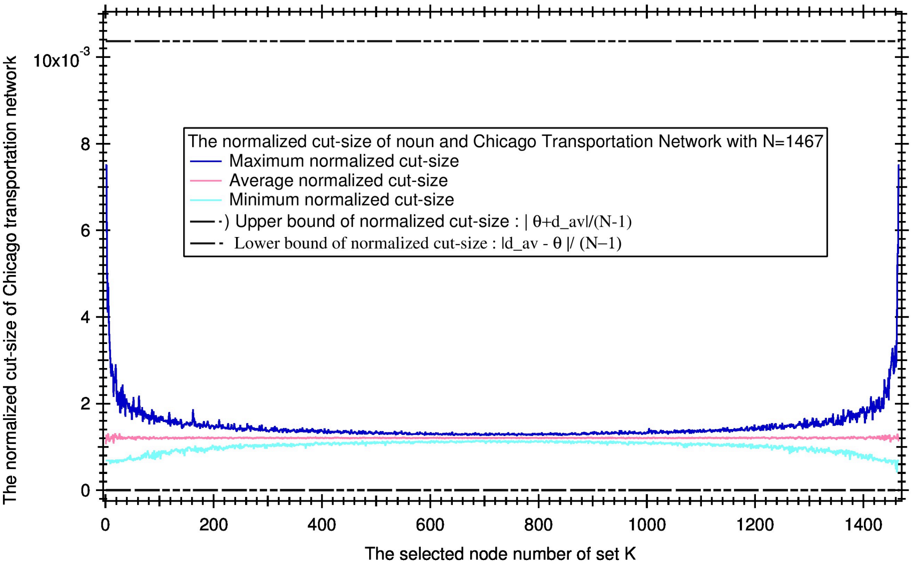

Figure 3.4.3: The normalized cut-size of the Chicago transportation network

Figure 3.4.3 illustrates the result of the transportation network of Chicago. The clustering coefficient of the Chicago network equals to 0, which means this network does not contain triangles. Similar to previous results, the pink curve presents the average normalized cut-size approximates to the network density where $\overline{N}_{cut} \approx p = 0.0012071$. The purple curve is the maximum normalized cut-size, which shows a fast decrease after more nodes selected, and then gradually approaches to the average normalized cut-size. The minimum normalized cut-size as shown as the green curve increases approximately to the average normalized cut-size. The maximum and minimum normalized cut-size have a centric trend that both of them converge to the average normalized cut-size. The black dash curves are theoretical upper and lower bounds based on the isoperimetric inequality. The difference between the maximum and minimum normalized cut-size is much smaller than the theoretical bounds.

The influence of the property of the cut-size can be explained with an example of Chicago transportation. Let us consider the traffic congestion problem on this network. If a congestion happens on a station, the congestion only influences the links which connect itself and stations with no congestion, since the buses can only drive to the station without congestion. If the traffic congestion happens randomly on the stations of the transportation network, the number of congested stations is small in the beginning. Then, the number of congested links can be very large if there are high-degree stations congested. The variation of the number of normalized congested links can be large. However, with the number of congested stations becomes large, for example, N/2, the change of the number of links with only one congested station is in a lower order, comparing to the product of the number of congested and non-congested stations. Then, the variation of the congestion is less influenced by the high-degree stations.

In this section, we present the results of real networks. There are some common properties observed from all real networks. According to these results, there are some general conclusions about the cut-size property for different real networks. 1) The normalized cut-size N_{cut} have a centric phenomenon that the range between the maximum and minimum normalized cut-size shrinks when randomly partition the network. The smallest variance achieves when th number of nodes in either partition approaches $k=\frac{N}{2}$. 2) With k approaching $N/2$, the value of the maximum and minimum normalized cut-size get closer to the average normalized cut-size $\overline{N}_{cut}$ which approximately equals to the network density. 3) The variation between the maximum and minimum normalized cut-size is much smaller than the theoretical bounds obtained by the isoperimetric inequality (3.3.18). But the bounds of the isoperimetric inequality cannot represent the properties of real networks. 4) By combing the average normalized cut-size and the set size k, the cut-size between two subsets in the graph can be approximated. All these results have a specific physical meaning for each real network.

3.4.2 The distribution of the cut-size

As discussed in the previous section, the normalized cut-size shows a con phenomenon, meaning that the range between the maximum and minimum normalized cut-size decrease with the number of nodes in each partition approaches to $\frac{N}{2}$. The maximum and minimum cut-size get closer to the average cut-size when K near $\frac{N}{2}$. The maximum normalized cut-size decrease approximately to the average normalized cut-size. The changing rate of the maximum normalized cut-size is related to the node degree in subset K. If a real network contains m hubs, then the maximum normalized cut-size decreases before k=m. If a real network has a lot of nodes with the similar degree, then the maximum normalized cut-size continuously decreases until k=$\frac{N}{2}$.

For a small k value nearby k=1, the cut-size distribution follows the original degree distribution. For instance, the cut-size distribution of a scale-free network follows a power-law distribution when k=1. The cut-size distribution approximately to the Gaussian distribution when $k \to \frac{N}{2}$. Because of the centric phenomenon, the maximum and minimum cut-size gradually change to the mean cut-size with a small variance. When k=$\frac{N}{2}$, the difference of the cut-size is minimal and the cut-size distribution resembles a Gaussian distribution. Figure 3.4.4 , Figure 3.4.5, Figure 3.4.6 and Figure 3.4.7 show the cut-size distribution with different k values of the European E-road network and the Facebook network.

Figures 3.4.4 and Figure 3.4.5 show the cut-size distribution of the Europe E-road network when choosing $k = 200$ and $k = \frac{N}{2}$. The blue dots represent the value of the cut-size and the red curve is a Gaussian distribution fitting curve. The parameters of the Gaussian fitting curve for Figure 3.4.4 are $\mu = 400.5$ and $\sigma = 20.83$, and for Figure 3.4.5 are $\mu = 709$, $\sigma = 26.52$ when k=$\frac{N}{2}$. Comparing to Figure 3.4.4 and Figure 3.4.5, the variances of the cut-size are similar, while the average cut-size increase with higher k. In addition, the variance of the normalized cut-size as shown in Appendix B.3.9, is smaller than 10^{-6}. The result proves the centric phenomenon of the cut-size.

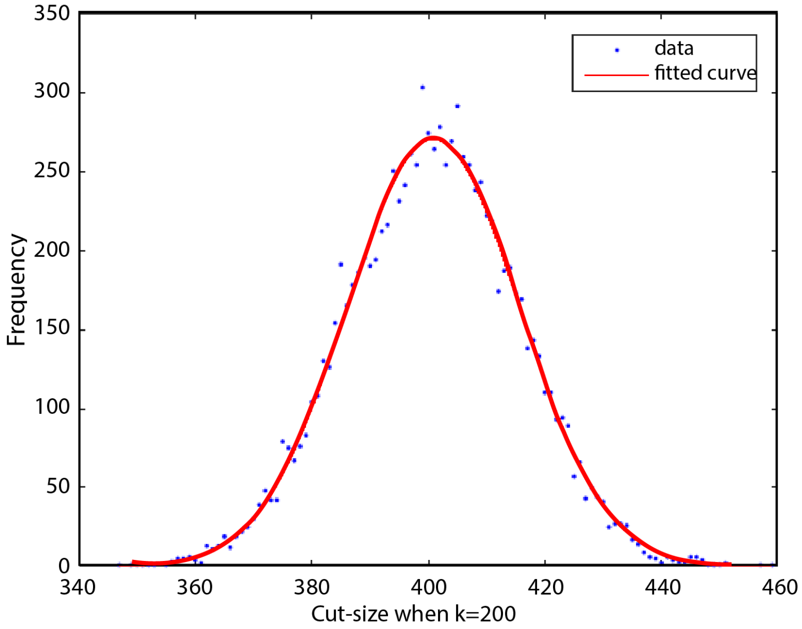

Figure 3.4.4: The cut-size distribution when k=200 for the European E-road network with N=1174

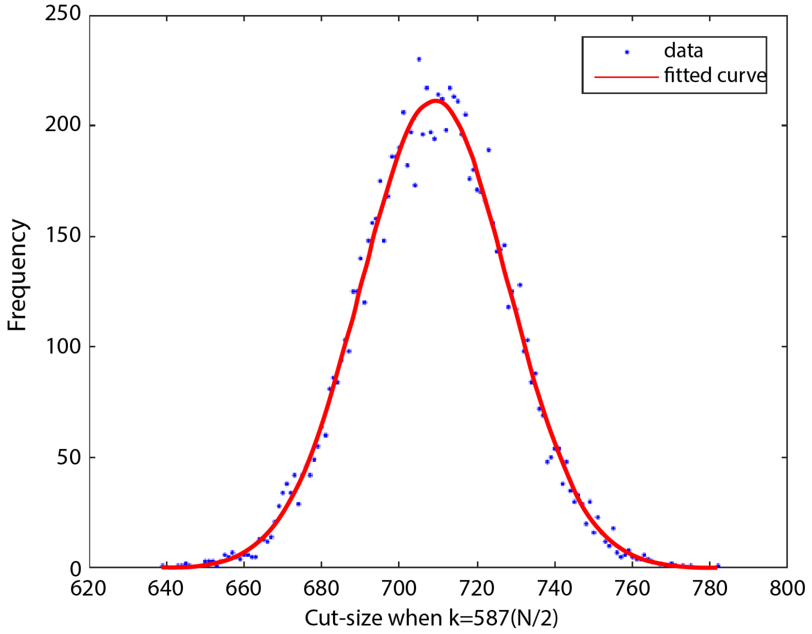

Figure 3.4.5: The cut-size distribution when k=$\frac{N}{2}$ for the European E-road network with N=1174

Figure 3.4.6 and Figure 3.4.7 show the cut-size distribution of Facebook network when $k=\frac{N}{2}$ and $k = 2000$. The blue dots represent the value of the cut-size and the red curve is a Gaussian distribution fitting curve. When $k=\frac{N}{2}$, the cut-size distribution resembles a Gaussian distribution

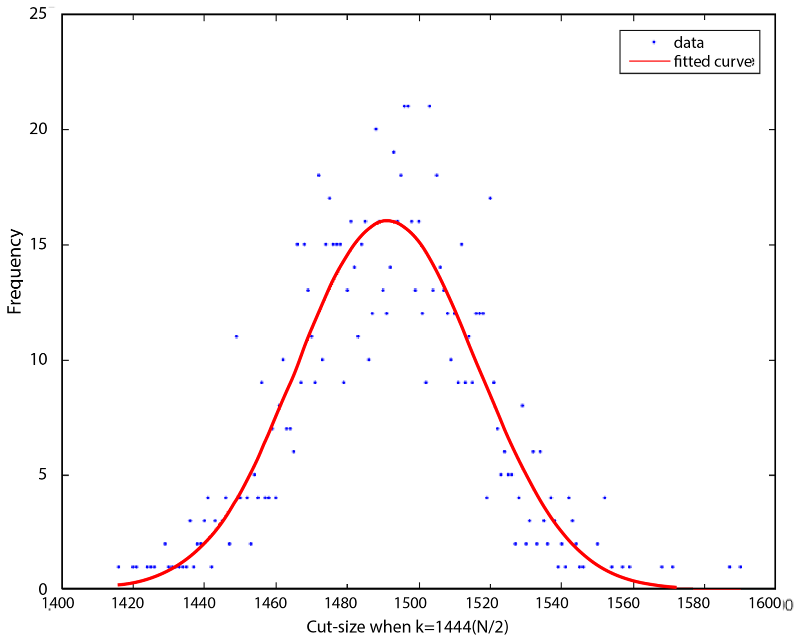

Figure 3.4.6: The cut-size distribution when k=$\frac{N}{2}$ for the Facebook with N=2888

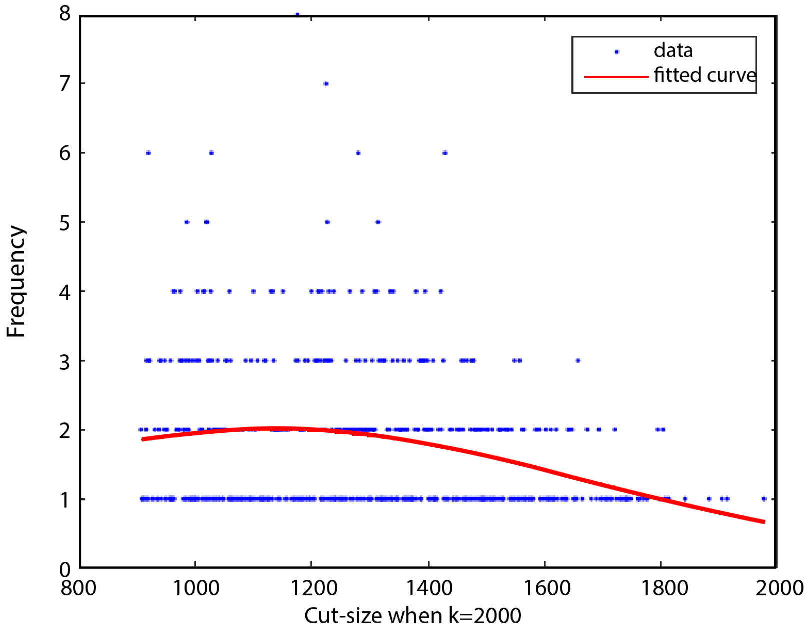

Figure 3.4.7: The cut-size distribution when k=2000 of the Facebook with N=2888

with $\mu = 1492$ and $\sigma = 32.02$. When k=2000, the cut-size distribution does not fit the Gaussian distribution with $\mu = 1201$ and $\sigma = 412.8$. The mean cut-size value when $k = 2000$ is smaller than the mean cut-size when k=$\frac{N}{2}$. The comparison between these two cut-size distributions illustrates that the distribution of the cut-size changes from a Gaussian distribution to degree distribution of the network with a higher k increase from $\frac{N}{2}$ to N. As shown in Appendix B.3.8, the variance between the normalized cut-size is smaller than $6\times\ 10^{-5}$, which also indicate the centric phenomenon of the cut-size.

3.4.3 Discussion

The normalized cut-size have a centric phenomenon, in which the maximum and minimum normalized cut-size get closer to the average normalized cut-size with k approaching $\frac{N}{2}$. The variance between the maximum and the minimum normalized cut-size is shrinking with the increase of k. Comparing to the theoretical bounds derived by the isoperimetric inequality (3.3.18), the range between the maximum and minimum normalized cut-size is much smaller than the loose theoretical bounds. Moreover, the average normalized cut-size is a constant related to the network density. The decreasing rate of the maximum normalized cut-size is related to the node degree in subset K. Also, for each k value, there exists a different cut-size distribution. The cut-size distribution closely resembles the Gaussian distribution when $K \approx \frac{N}{2}$.

3.5 The effect of degree on the cut-size property

In this section, the effect of degree is the main concern to study the normalized cut-size of real networks. The method to formulate the set K affects the normalized cut-size result.

3.5.1 The effect of degree

Based on the node degree, we select nodes to conduct the set K. The nodes with a higher degree have a higher probability to be chosen for set K, while the nodes with a lower degree have a lower probability to be selected for set K. The step to calculate the cut-size is given here. Firstly, we use the degree of each node to generate a node vector V which contains Nd_{av} elements. The number of node i appears d_i times in the node vector. Next, the nodes will be randomly selected within the node vector V. The selected nodes formulate the subset K in which most nodes have a larger degree. For each subset K, the cut-size is calculated by equation (3.3.1). The degree-based partition process repeats m= 10^5 times for each value of k to obtain the normalized cut-size of the networks.

Based on the above calculation method, there are some conclusions observed from all these results.

- The range between the maximum and minimum normalized cut-size shrinks with the increase k. The selection of high degree nodes within set K leads to a significant increase in the cut-size when k is small .
- The average normalized cut-size is not a constant and decreases with the increase k.
- The maximum normalized cut-size continuously declines with k. The minimum normalized cut-size increases when k is small and then decreases along with the average normalized cut-size.
- The nodes with the higher degree are playing more important role in the change of the normalized cut-size.

The results of three real networks are shown and discussed in the following section. Other networks are given in Appendix B.

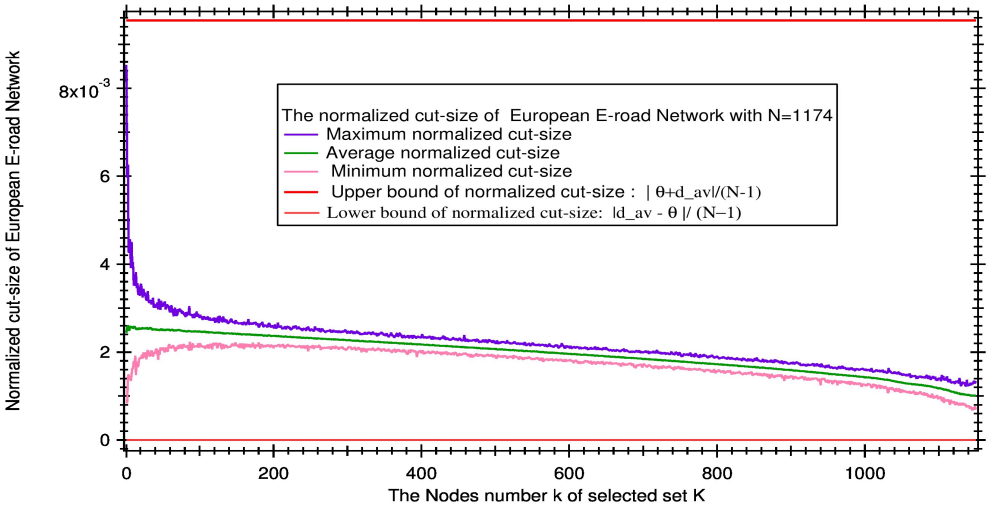

Figure 3.5.1: The normalized cut-size of the Europe E-road based on degree method

Figure 3.5.1 shows the normalized cut-size of the Europe E-road network based on the degree selection. Comparing to the figure 3.4.1, the result is entirely different. The green curve is the average normalized cut-size which is not a constant anymore and continuously decrease with k. The purple curve is the maximum normalized cut-size, which decreases significantly when $k< 50$

because the set K includes most of the high-degree nodes. When all high degree nodes are selected in set K, the decrease of the maximum normalized cut-size slows down. The pink curve is the minimum normalized cut-size, which increases approximately to the average normalized cut-size with a few lower-degree nodes. Then, the minimum normalized cut-size decreases along with the average normalized cut-size. The variance between the maximum and minimum normalized cut-size continuously shrinks. The normalized cut-size change largely before $k \leq 100$. After that, the normalized cut-size decrease approximately linearly with lower degree nodes. In addition, the red curves as shown in figure 3.5.1 represent the upper and low bounds derived from the isoperimetric inequality (3.3.18). A real example of the European E-road network is given here. If the news of the Netherlands toxic eggs [39] suddenly appear on the Internet, the cities with more neighbors, like Paris, Berlin, and Amsterdam, post this news. The neighbors of these central cities directly receive the news. After a period, all major cities with plenty of neighbors receive the news. But, the rural cities will receive the news with a delay.

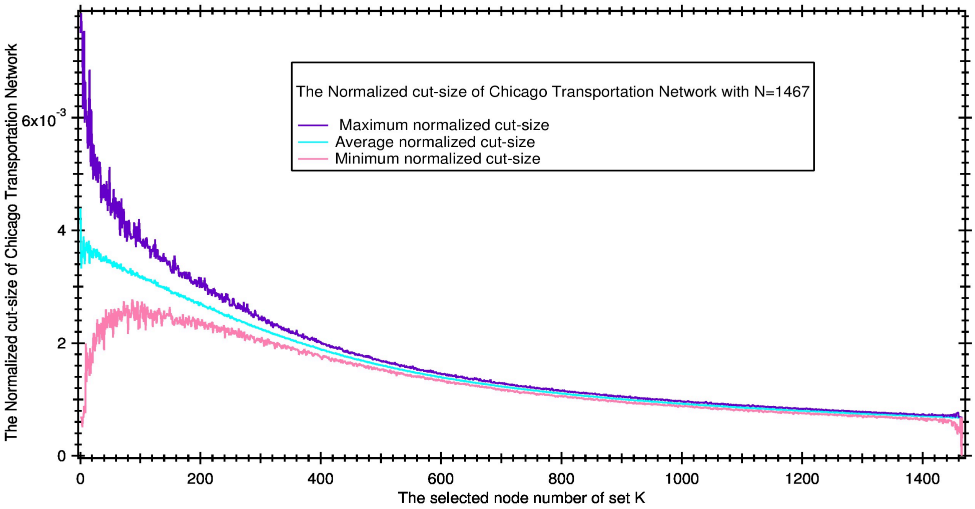

Figure 3.5.2: The normalized cut-size of the Chicago transportation based on degree method

Figure 3.5.2 illustrates the normalized cut-size of the Chicago transportation network. The purple curve is the maximum normalized cut-size which continuously decreases with k. The pink curve is the minimum normalized cut-size. The blue curve is the average normalized cut-size which has the decreasing trend within the whole process. Both the maximum and the minimum normalized cut-size have a centric phenomenon that approaching to the average normalized cut-size. The difference between the maximum and the minimum normalized cut-size decreases with the number of nodes in set K. Let us consider the traffic congestion on this network. If the congestion happened in the central station with a large degree, then the congestion will fast spread to other neighbors stations. The central station has the largest influence on creating congestion.

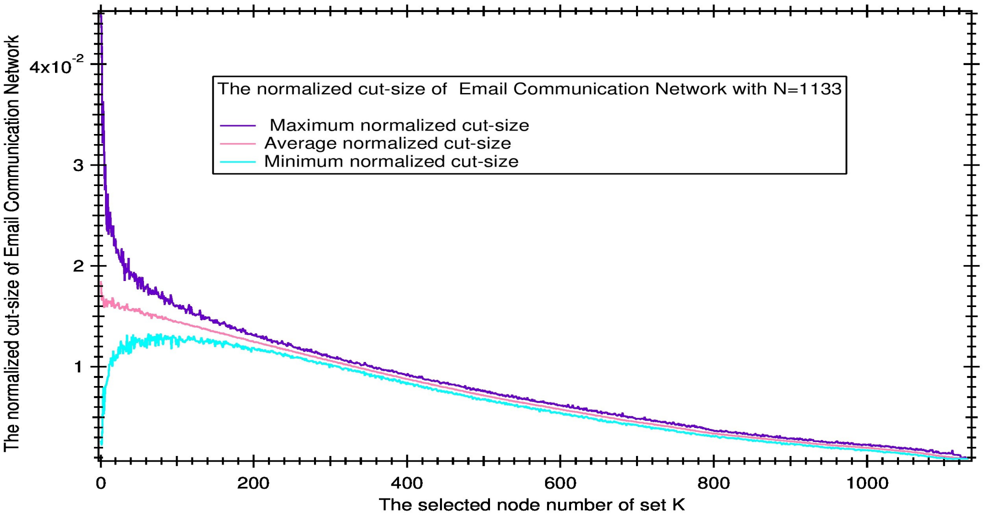

Figure 3.5.3: The normalized cut-size of the email communication based on degree method

Figure 3.5.3 shows the normalized cut-size of the email communication network. In this network, plenty of the nodes have the highest degree 11. The purple curve is the maximum normalized cut-size which decreases with k. The pink curve is the average normalized cut-size which also decreases with the increase k. The difference between the maximum and minimum normalized cut-size has the decreasing trend. When $k \approx 200$, this difference is quite small which indicates the maximum and the minimum cut-size are approaching each other. This network contains around 200 nodes with a high degree, and both the maximum and minimum normalized cut-size approximately equal to the average when k>200. Therefore, the larger-degree nodes have the large influence on the value of the normalized cut-size when partitioning the network.

Compared with the Figure 3.5.1, Figure 3.5.2, and Figure 3.5.3, the normalized cut-size changing trends are similar. The maximum normalized cut-size dramatical decreases with the number of the large-degree nodes in set K. When the low-degree nodes are selected to formulate set K, the decrease rate of the maximum normalized cut-size slows down. Furthermore, the average normalized cut-size continuously decreases with the increase k. The minimum normalized cut-size increases when k is approximately smaller than the number of the large-degree nodes, because of a similar reason in the maximum cut-size case. When k is approximately larger than the number of large-degree nodes, the minimum normalized cut-size gradually decreases. From the physical meaning perspective, the changing trend of the normalized cut-size is related to the degree distribution of each network. The Europe E-road has a few nodes with the highest degree and large amounts of nodes with a similar degree, so the decrease rate is slow after K=80. The Chicago transportation network contains plenty of nodes with the high degree, so the maximum normalized cut-size decrease fast to the average normalized cut-size when $k < 400$. The cut-size bursts at the beginning when more critical nodes are selected. Therefore, the nodes with high degree in real networks have a higher influence on the cut-size property when using degree-based partition method.

3.5.2 The effect of the hub

In this section, an example is given to show the effect of the hub nodes on the cut-size property. In this case, the network is partitioned into two sets based on the hub nodes, as a result of the minimum cut of the network.

We consider the Zachary karate club network [29] as the example to explain the effect of hub nodes for partitioning the network. This network has 34 members and 78 relations between members with a high clustering coefficient and a large network density. But, two teachers with the highest connections argue with each other, then this club splits into two parts. Half of the members form a new club around one teacher and other members of another group found a new instructor or quit the karate club. Only five nodes contain most of the links, so these five nodes are the hubs in the network. According to the degree of each node, we split these five hubs into two sets, set K and its complement set S. If the set K contains the large-degree nodes and most of its neighbors, then the number of links connecting nodes within K is large. The number of the links between two sets reaches a small size. If we want to obtain the minimum cut-size crossing two subsets, then we need to make sure to find the most connected part of the networks.

As shown on Figure 3.5.4, the nodes in Zachary karate network are partitioned into the red and

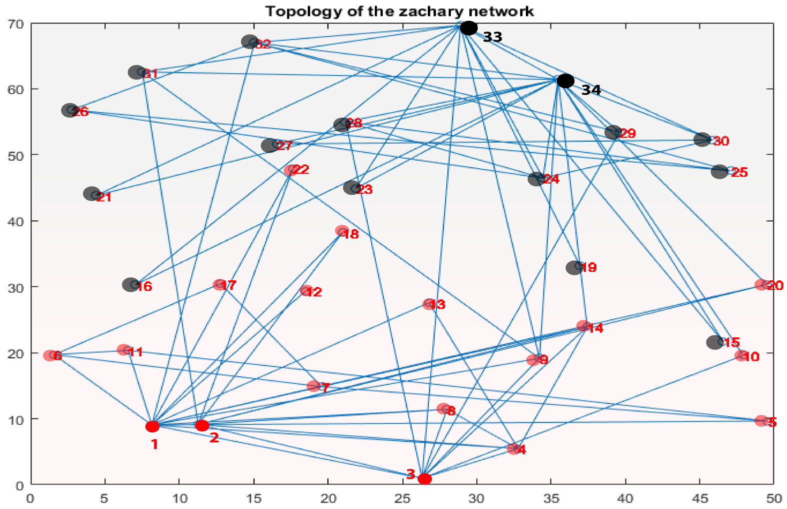

Figure 3.5.4: The two subsets of Zachary karate network

the black set. Nodes 1,2,3 are the most critical hubs in the red set. Most of the neighbors of node 1, node 2, and node 3 belong to the red set. Nodes 34 and node 33 are the critical hubs of the black sets and most of the neighbors of node 33 and node 34 also belong to the black set. These hubs decide the structure of the sets with the minimum cut. However, there still exists nodes which can belong to any sets with less importance. For instance, node 31 and node 8 have the same degree, connecting with both sets. These two nodes have less influence on the cut-size of this partition. In this case, the cut-size is the minimum value. We conclude that the hub nodes have the largest influence on the cut-size when partitioning the networks.

3.5.3 Discussion

Figure 3.5.1, Figure 3.5.2 and Figure 3.5.3 illustrate the normalized cut-size result based on the node degree. The method to select nodes in set K influences the cut-size. The number of nodes with high degree affect the change rate of the maximum and the minimum normalized cut-size. The range between the maximum and the minimum normalized cut-size is gradually shrinking when k is smaller than the number of nodes with high degree. Different from the maximum and minimum normalized cut-size, the average normalized cut-size decreases with k. Comparing to the theoretical bounds, the real normalized cut-size has a centric trend that concentrates around the average normalized cut-size. The theoretical bounds can not represent the real case well. In particular, the hub nodes have the largest influence on the change of the cut-size.

3.6 Conclusions

In this chapter, the average and the bounds of the normalized cut-size are derived by the average interconnection constant and the isoperimetric inequality, respectively. Based on the studying of all real networks, the general normalized cut-size results are obtained by applying two different methods of the graph partition. The normalized cut-size has a centric trend that closes to its average normalized cut-size, i.e. the normalized cut-size almost equals to the average normalized cut-size, and the average normalized cut-size is a constant which is related to the network density when the graph is randomly partitioned. Moreover,the cut-size distribution is a Gaussian distribution when $K=\frac{N}{2}$. When $k \approx 1$, the distribution of the cut-size is approximately the degree distribution of the networks. If the graph partition is based on degree, the normalized cut-size has a different result. Different from the case where the graph is randomly partitioned, our study shows that the average normalized cut-size decreases with the increase k, and the average normalized cut-size is not a constant anymore.

The difference between the maximum and the minimum normalized cut-size declines until a certain value of k. The number of nodes with a certain high degree is essential for the change of the normalized cut-size. What is more, the theoretical bounds derived by the isoperimetric inequality (3.3.18) can not represent the distribution of the normalized cut-size well since the value of the normalized cut-size concentrates to the average normalized cut-size. Also, the hub nodes have the largest influence when partitioning a graph.

Chapter 4

The cut-size property and the SIS epidemic process

4.1 Overview

This chapter discusses the application of the cut-size in the epidemic spreading process on networks. We provide an estimation of the prevalence in the metastable state. The average normalized cut-size is applied to the estimation. The simulated prevalence is obtained by the simulator of the Susceptible-Infected-Susceptible (SSIS). We compare the estimated prevalence with the simulation of the exact process. Then, we explain the comparisons result of real networks.

In section 4.2, we review the background knowledge of the SIS epidemic process on networks. In section 4.3, we estimate the metastable state prevalence by the cut-size property of the networks. In section 4.4, we compare the simulated prevalences of real networks and the estimated prevalences. Section 4.5 concludes this chapter.

4.2 Epidemic spreading on networks

The Susceptible-Infected-Susceptible (SIS) process is a basic epidemic model, in which each node has two states: susceptible (S) or infected (I). Another basic epidemic model is the Susceptible-

Infected-Recovered (SIR) model with three states transitions, in which two states transitions are same with SIS model, and the third state is, Recover state(R). An infected node can be cured or removed. The removed nodes are not involved in the spreading process after removing. More complex models are evolute from the SIR and SIS models, for instance, the Susceptible-Infected-Recovered-Susceptible (SIRS) model. The Susceptible-Infected-Recovered-Susceptible (SIRS) model has transition process from the removed state to susceptible state. The newborn susceptible node is added into the population when the node state change from recovered to susceptible $R \to S$ [48].

4.2.1 The SIS model and Markov chain

As a basic epidemic model, the Susceptible-Infected-Susceptible (SIS) model can be described by a Markov chain. In the SIS model, there are only two states of each node on networks [10]: susceptible (S) or infected (I). The susceptible (S) node can be infected with the infection rate β. The infected node can be cured with the curing rate δ. There exists two states transitions [10], $S \to$ I and $I \to S$. The transmission processes from S → I and I → S are independent Poisson processes.

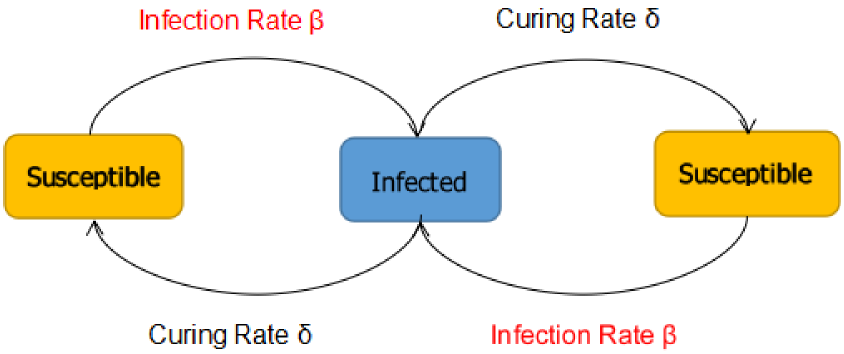

Figure 4.2.1: States transition process between infection and health for each node has one infected neighbor

Figure 4.2.1 illustrates the different states of node i and the state transition process in the SIS model. For each node, the node state at time t is a Bernoulli random variable $W_i(t) \in \{0,1\}$. The infected state of node i can be described as $W_i(t)=1$ with the infected probability $E[W_i(t)]=Pr[(W_i(t) =$

1]. The healthy state can be described as $W_i(t)$=0 with the healthy probability 1-$E[W_i(t)]$=$Pr[W_i(t) = 0]$. The curing process triggers directly for the infected node with the curing rate δ, while in the infection process, each infected neighbor of a healthy node contributes an infection rate β. The average period of the infected state and the healthy state of each node is $\frac{1}{\delta}$ and $\frac{1}{\beta}$,respectively. The time duration of being infected or healthy follows an exponential distribution. The infection rate of a healthy node i is $\beta\sum_{j=1}^{N} a_{ij}W_j(t)$, so the infection rate of node i is proportional to the number of the infected neighbors. A node with a large degree normally has a better chance of being infected [35].

Meanwhile, the effective infection rate of the network is $\tau = \frac{\beta}{\delta}$ which separates two phases of the epidemic process on networks. If the effective rate τ is above the epidemic threshold τ_c, then the epidemics break out with a non-zero fraction of infected nodes in the networks. If the effective infection rate τ is below the epidemic threshold τ_c, then leads to the virus dies out approximately exponentially fast. The epidemic threshold τ_c is the standard to judge where the phase changes from dies out phase to outbreak phase. The epidemic threshold τ_c is determined by the underlying networks.

The states transitions of SIS process are independent Poisson processes with the memoryless property. Since the process is Markovian, the process is only determined by the current network infection state. Hence, a continuous-time Markov chain with 2^N states can be used to describe the SIS process. The infinitesimal generator $Q_i(t)$ of each node is [14]

$$Q_i(t) = \begin{bmatrix} -q_i(t) & q_i(t) \\ \delta & -\delta \end{bmatrix} \tag{4.2.1}$$

where δ is the curing rate and $q_i(t)$=$\beta\sum_{j=1}^{N} a_{ij}W_j(t)$ [35].

Figure 4.2.2 shows one simple example of a complete graph with $N = 3$ nodes and uses Markov chain to express the transition of 8 states. A node with 1 presents the infected state and with 0 presents the healthy state.

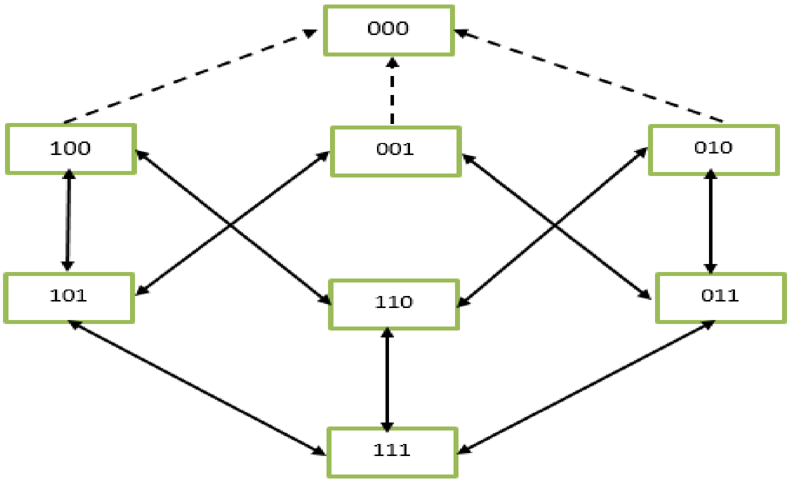

Figure 4.2.2: Exact Markov chain of SIS process

For a small graph, the Markov chain is available to describe the whole stage of SIS process. With the increasing of the network size N, the computation complexity increases exponentially, and the process is infeasible to be calculated for a large N. Therefore, the exact model is limited by the computation complexity. As the result, the approximation methods are needed to approximately solve the SIS process [14].

4.2.2 The Mean-field Approximation method

The N-Intertwined Mean-Field Approximation (NIMFA) [35] is a basic approximation method of the Markovian SIS epidemic process. NIMFA incorporates the adjacency matrix A, which contains topology information of the network. NIMFA is a reliable analytical approach to solve the SIS process, which can be applied to most networks with a high accuracy.

NIMFA approximates the random variable $q_i(t)$=$\beta\sum_{j=1}^{N} a_{ij}W_j(t)$ by the average $E[q_i(t)]$. The average $E[q_i(t)]$ is

$$E[q_i(t)] = \beta\sum_{j=1}^{N} a_{ij}E[W_j(t)] \approx \beta\sum_{j=1}^{N} a_{ij}v_j(t) \tag{4.2.2}$$

where $v_i(t) \approx E[q_i(t)]$ is the NIMFA infection probability of node i. The NIMFA governing equation of the infection probability is

$$\frac{dv_i(t)}{dt} = -\delta v_i(t) + \beta(1 - v_i(t)) \sum_{j=1}^{N} a_{ij} v_j(t) \tag{4.2.3}$$

the first term of the right-hand side represents the curing process that a infected node i with the infected probability $v_i(t)$ which is cured by the curing rate δ. The second term represents the infection process that a healthy node i with the healthy probability $(1 - v_i(t))$ which is infected by all infected neighbors with infection rate $\beta \sum_{j=1}^{N} a_{ij}$ [35].
The equation (4.2.3) can be written in a matrix form is

$$\frac{dV(t)}{dt} = (\beta A - \delta I)V(t) - \beta diag(v_i(t))AV(t) \tag{4.2.4}$$

where $V(t)$=$[v_1(t), v_2(t), ..., v_n(t)]^T$, the $diag(v_i(t))$ is a diagonal matrix whose diagonal elements are $V_i(t)$ for i=$1, 2, ..., N$. I is the identity matrix [35].

Based on the equation (4.2.4), the NIMFA provides a lower bound of the epidemic threshold τ_c as

$$\tau_c > \tau_c^{(1)} = \frac{1}{\lambda_1} \tag{4.2.5}$$

where λ_1 is the largest eigenvalue of the adjacency matrix A. The superscript (1) means the $\tau_c^{(1)}$ is a first order mean-field threshold [35].

4.2.3 SSIS

The simulator of SIS, which is a continuous-time Markov process simulator, is applied to obtain the prevalence of the epidemic process. By averaging the simulation result of the SIS process in the simulator with 10^5 times, the prevalence is obtained with enough accuracy. In the SSIS simulator, the infection and curing events randomly occur within the time line. Both processes are independent Poisson processes and the time interval of states transition follows exponential distributions. For an arbitrary time, the network state vector is $w(t) = [W_1(t), W_2(t), ..., W_N(T)]$. At time t=0, the initial number $Ny(0)$ of the infected nodes is set and we obtain the metastable prevalence $y(t)$ by running the simulator for enough time [33]. In this chapter, the SSIS simulator is applied to obtain the prevalence of real networks in the metastable state. In the simulation, we let all nodes infected initially to prevent the early large die-out probability, and thus, we obtain a relatively accurate metastable prevalence. Finally, the simulated prevalence $y(t)$ in the metastable state is compared with the prevalence estimated by the average normalized cut-size property.

4.3 The estimated prevalence

The prevalence $y(t)$ is the average fraction of infected nodes at time t. For each time t, the prevalence is determined by the topology of the network. The prevalence is hard to calculate because of the computational complexity. The infection rate β and the curing rate δ are independent of time, the fraction of the infection nodes is [12]

$$S(t) = \frac{1}{N}\sum_{i=1}^{N} W_i(t) \tag{4.3.1}$$

and the prevalence $y(t)$ is the expected fraction of the $S(t)$ equals

$$y(t) = E[S(t)] = \frac{1}{N}\sum_{i=1}^{N} Pr[W_i(t) = 1] \tag{4.3.2}$$

since $E[W_i(t)] = Pr[W_i(t) = 1]$] for the Bernoulli random variable. In the SIS process, the governing equation of the exact SIS prevalence is [12]:

$$\frac{dy(t)}{dt} = -y(t) + \frac{\tau}{N}E[w(t)^T Qw(t)] \tag{4.3.3}$$

where τ=$\frac{\beta}{\delta}$ is the effective infection rate, Q is the Laplacian matrix of the graph obtained and $w(t)$=[$W_1(t)$, ..., $W_N(t)$] is the network state vector. The term $E[w(t)^T Qw(t)]$ is the expected cut-size at t moment. From equation (4.3.3), the change rate of the prevalence equals to the difference between product of the expectation cut-size and $\frac{\tau}{N}$ and the prevalence y(t). This equation (4.3.3) implies that the prevalence is related to the expectation cut-size. We assume

$$\overline{N}_{cut} = \frac{E[w^T Qw]}{k(N-k)} \tag{4.3.4}$$

where k is the number of infected nodes, $(N-k)$ is the number of the healthy nodes, and here we approximate $Nk \approx y$ and the infected nodes are randomly distributed in the networks. Then, for $\frac{dy(t)}{dt} = 0$ in the metastable state, the estimated metastable prevalence $\tilde{y}$ is

$$\tilde{y} = 1 - \frac{1}{N\tau\bar{N}_{cut}} \tag{4.3.5}$$

If we know the adjacency matrix A of the network, then the prevalence and the epidemic threshold can be estimated by equation (4.3.5).

4.4 Simulation and explanation

In this section, the simulated metastable prevalence y of real networks is obtained by simulations, and the estimated prevalence $\tilde{y}$ is calculated by equation (4.3.5). The comparison between the simulated and the estimated metastable prevalence is given and discussed. The comparisons of four real networks are shown in this section. The result of other networks are given in Appendix B.

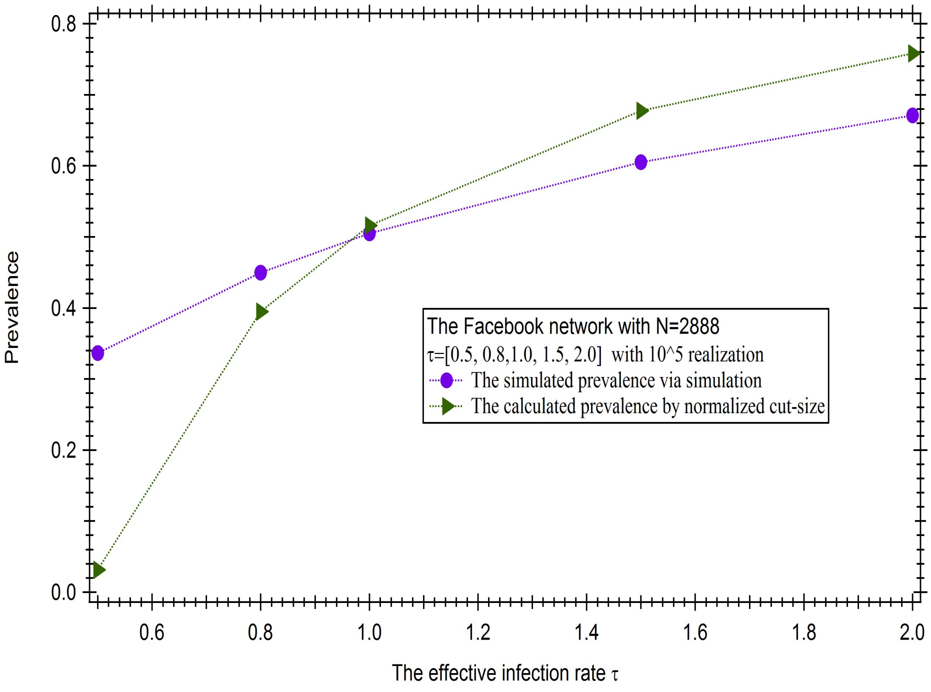

Figure 4.4.1: The Comparison between the simulated metastable prevalence y and estimated metastable prevalence $\tilde{y}$ of the Facebook network

Figure 4.4.1 shows the comparison between the simulated metastable prevalence y and the estimated metastable prevalence $\tilde{y}$ of the Facebook network. The simulation process is repeated 10^5 times. The green curve is the estimated metastable prevalence $\tilde{y}$ and the purple curve is the simulated metastable. When the effective infection rate τ is small, equaling to 0.5, the error between the simulation and the estimation is large. There is a crossing point when τ=1.0. At this point, the difference between the simulated metastable prevalence and estimated metastable prevalence is smallest. However, when τ=0.5, the estimated prevalence is almost zero which indicates that the infection cannot persist on the network. The simulated metastable prevalence $y = 0.34$ is higher than the estimated metastable prevalence.

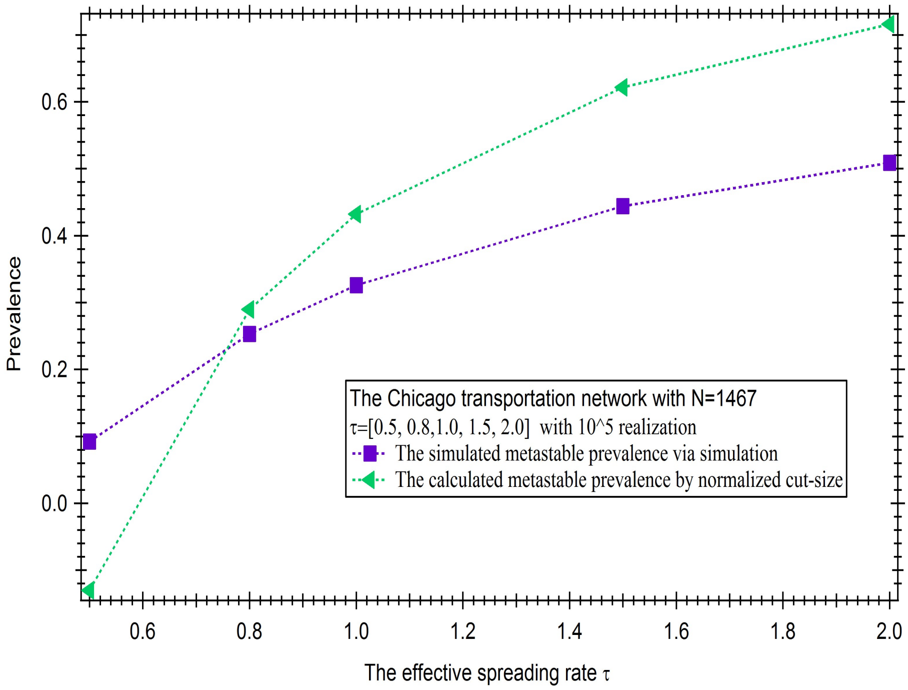

Figure 4.4.2: The Comparison between the simulated metastable prevalence y and estimated metastable prevalence $\tilde{y}$ of the Chicago transportation network

Figure 4.4.2 shows the comparison between the simulated metastable prevalence y and the estimated metastable prevalence $\tilde{y}$ of the Chicago transportation network. The simulation process is also repeated 10^5 times. The green curve is the estimated metastable prevalence $\tilde{y}$ and the purple curve is the simulated metastable prevalence y. The comparison shows one crossing point with the minimum difference when τ=0.8. For $\tau > 0.8$, the difference between the simulated and the estimated metastable prevalence is increasing. When τ=0.5, the estimated metastable prevalence is smaller than 0 which indicates presents no virus survivals under our estimation. The simulated metastable prevalence has the small value close to zero.

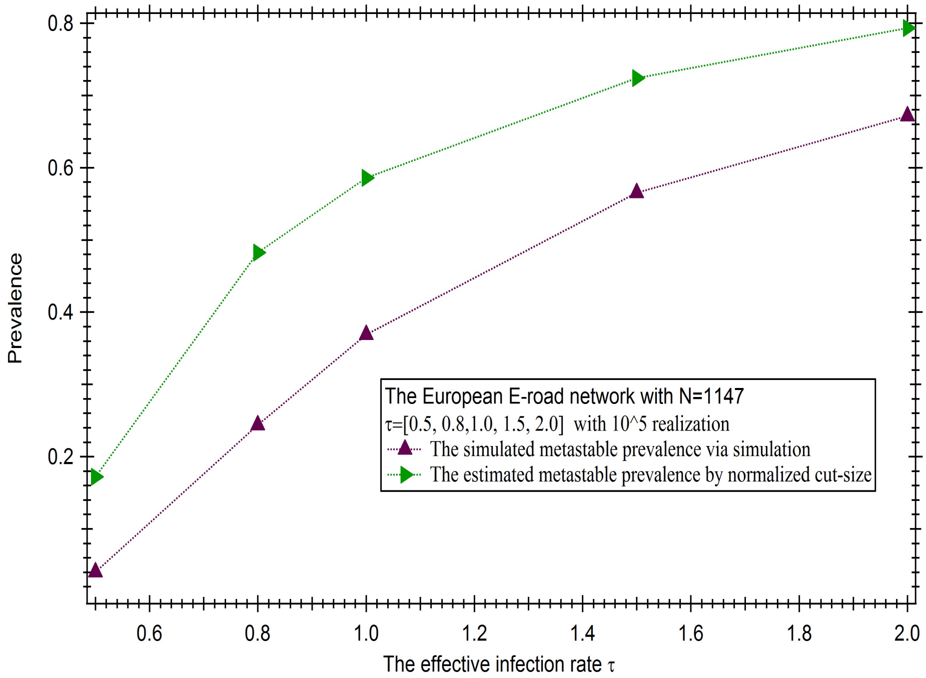

Figure 4.4.3: The Comparison between the simulated metastable prevalence y and estimated metastable prevalence $\tilde{y}$ of the Europe E-road network

Figure 4.4.3 shows the result of the European E-road network. The purple curve is the simulated metastable prevalence y, which is a smaller value compared with the estimated metastable prevalence $\tilde{y}$. The difference between the simulated prevalence y and the estimated metastable prevalence $\tilde{y}$ is the minimum value when τ=0.5. However, when τ=0.5, the simulated metastable prevalence approaches zero because the virus goes extinct on the networks. Hence, the effective infection rate τ=0.5 belows the epidemic threshold τ_c. With the higher effective infection rate $> \tau_c$, the virus is spreading through the network. Besides, with a larger effective infection rate τ, the estimated metastable prevalence is precise and closer to the simulate metastable prevalence.

Figure 4.4.4 shows the result of the Jazz musician network. The green line is the estimated

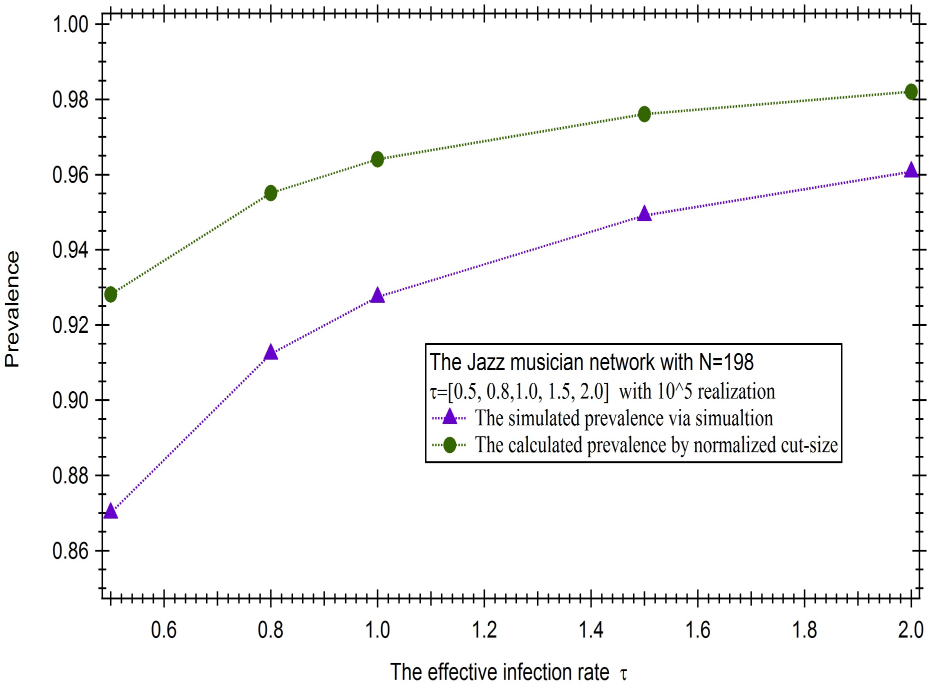

Figure 4.4.4: The Comparison between the simulated metastable prevalence y and estimated metastable prevalence $\tilde{y}$ of the Jazz musician network

metastable prevalence $\tilde{y}$ and the purple curve is the simulated metastable prevalence y. The difference between the simulated metastable prevalence y and the estimated metastable prevalence $\tilde{y}$ decreases which shows the estimated prevalence is precise for a larger effective infection rate τ. The estimated metastable prevalence gets closer to the simulated prevalence.

Figure 4.4.5 shows the comparison of the simulated metastable prevalence y and the estimated metastable prevalence $\tilde{y}$ of the Email communication network. The estimated metastable prevalence $\tilde{y}$ is larger than the simulated metastable prevalence. The difference between the simulated metastable prevalence y and the estimated metastable prevalence $\tilde{y}$ is decreasing with τ. The estimated prevalence is precise when the effective infection rate τ is larger, and the estimated prevalence is getting closer to the simulated metastable prevalence.

Figure 4.4.6 shows the comparison between the simulated metastable and the estimated metastable prevalence of the random ER network. The random network has a fix network size N=50 nodes and the link density p=0.1. The effective infection rate τ=[0.5,1,1.5,2,3,4,5]. The pink curve is the estimated metastable prevalence and the purple curve is the simulated metastable prevalence.

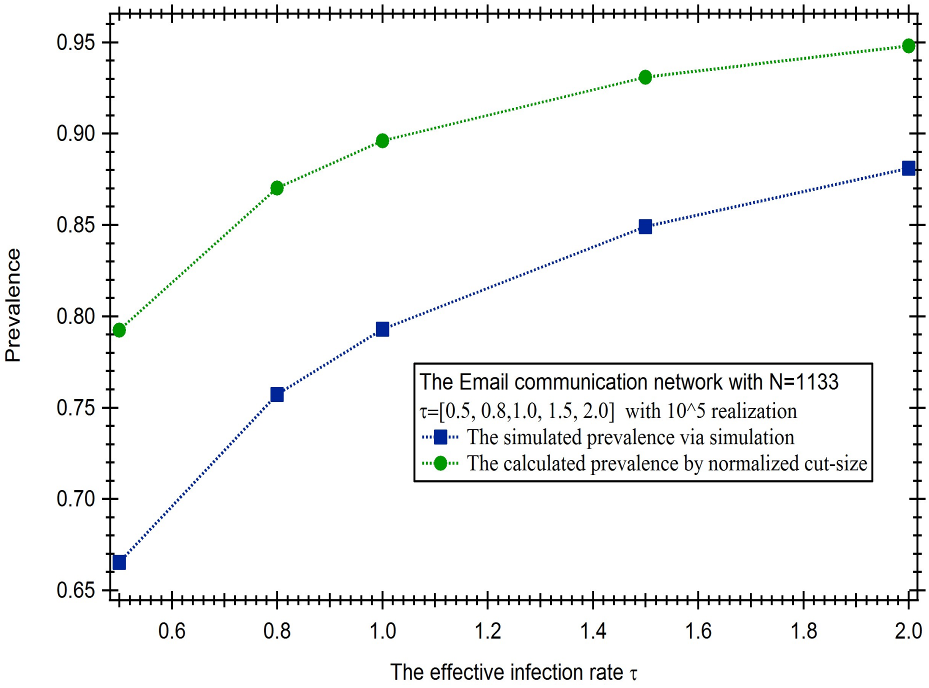

Figure 4.4.5: The Comparison between the simulated metastable prevalence y and estimated metastable prevalence $\tilde{y}$ of the Email communication network

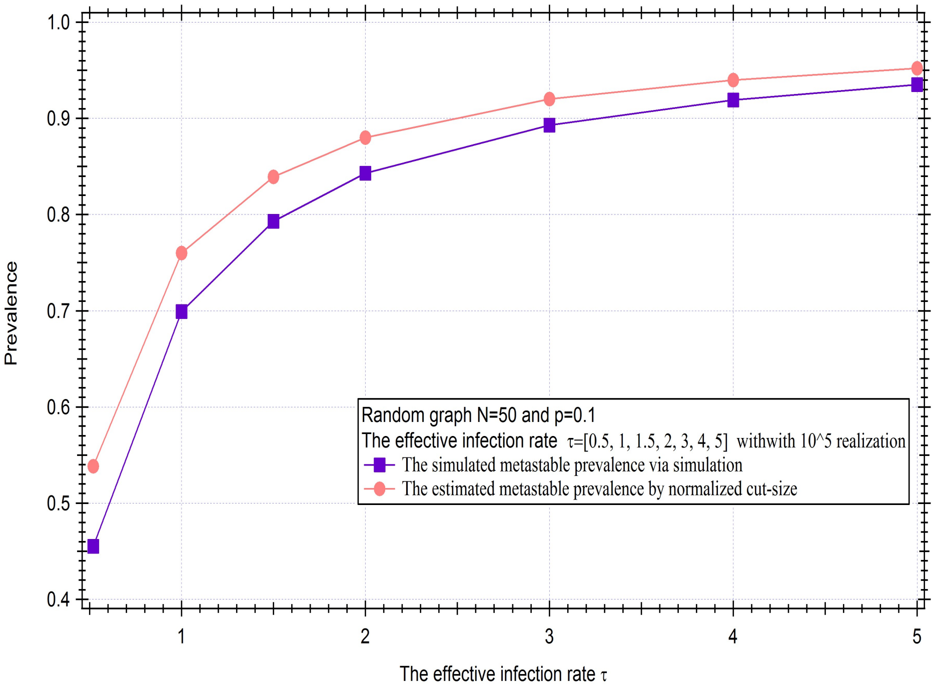

Figure 4.4.6: The Comparison between the simulated metastable prevalence y and estimated metastable prevalence $\tilde{y}$ of the random ER graph with N=50 and p=0.1

As shown in the figure, with the increase effective infection rate τ, the estimated prevalence get closer to the simulated prevalence. The difference decrease with a larger effective infection rate τ.

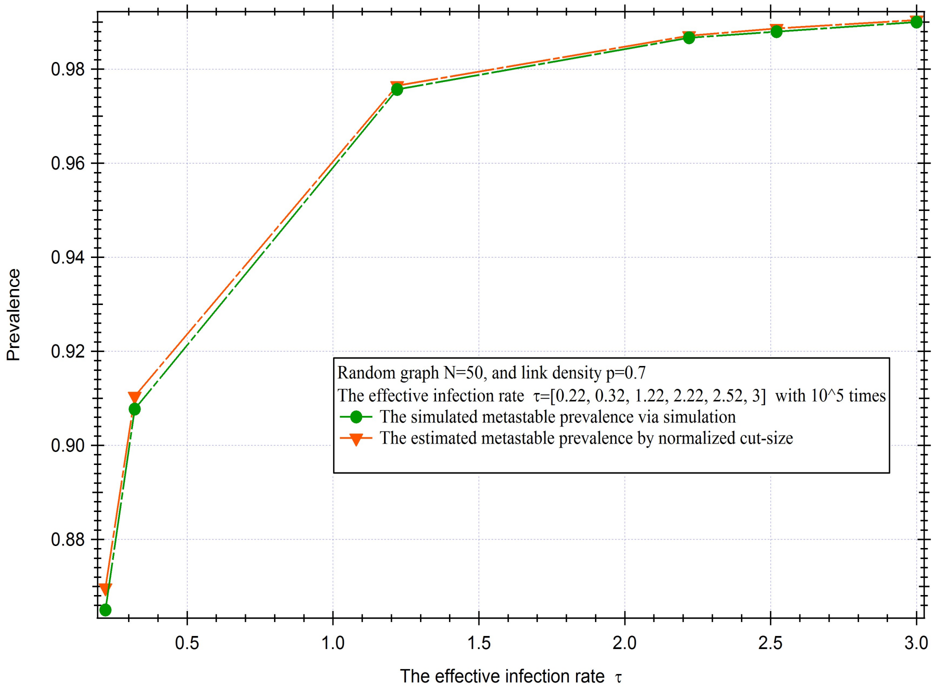

Figure 4.4.7: The Comparison between the simulated metastable prevalence y and estimated metastable prevalence $\tilde{y}$ of the random ER network with $N = 50$ and p=0.7

Figure 4.4.7 shows the comparison result of the random ER network with the same network size N=50 and a larger link density p=0.7. The effective infection rate τ=[0.22, 0.32, 1.22, 2.22, 2.52,3]. The red curve is the estimated metastable prevalence which has the higher value than the simulated metastable prevalence as shown as the green curve. Similarly, the difference between the estimated prevalence and the simulated prevalence declines with a larger effective infection rate τ. Comparing to the previous figure, the difference of the estimated prevalence and the simulated prevalence in Figure 4.4.7 is smaller than the difference in Figure 4.4.6 because of the higher link density p. Therefore, the link density p of the random network influence the difference between the estimated prevalence and the simulated prevalence.

Figure 4.4.8 shows the comparison of the estimated metastable prevalence and the simulated metastable prevalence of the ER network with N=50 and a fixed effective infection τ=0.22. In

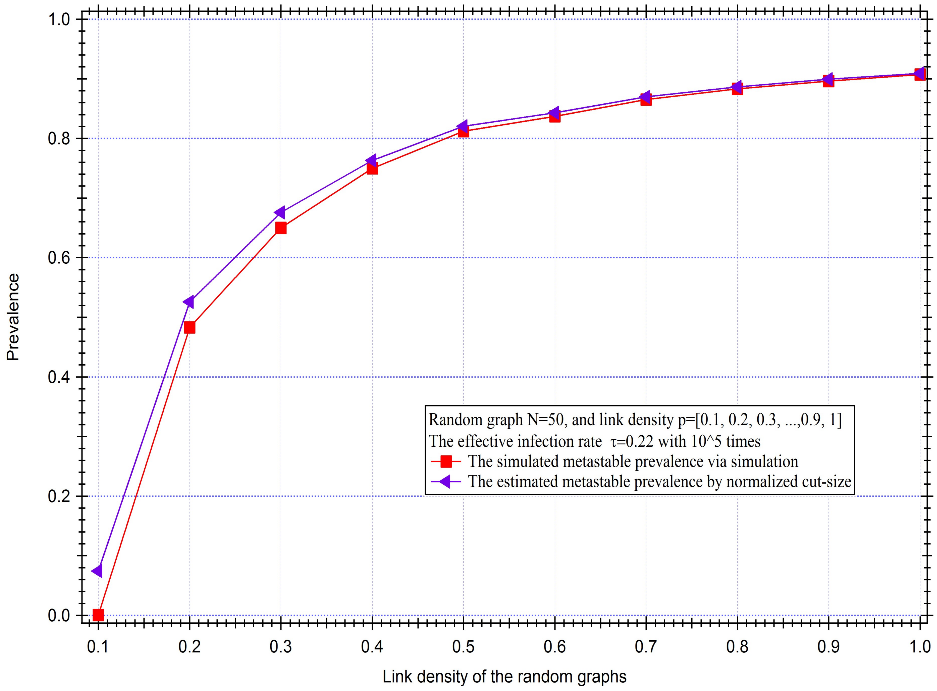

Figure 4.4.8: The Comparison between the simulated metastable prevalence y and estimated metastable prevalence $\tilde{y}$ of the random ER network with N=50 and τ=0.22

this figure, the link density continuously increases from $p = 0.1$ to $p = 1$. The purple curve is the estimated prevalence and the red curve is the simulated prevalence. When $p = 0.1$, the simulated prevalence close to zero which means the virus dies out in this network and the effective infection rate $\tau = 0.22 < \tau_c$. But, with the increase link density p, both of the prevalences are increasing, and the error between the simulated prevalence and the estimated prevalence decreases. The simulated and the estimated prevalence is almost zero with the link density $p > 0.7$.

According to the above results of the random network, we find the accuracy of the estimated metastable prevalence depends on the link density and the effective infection rate τ. With a larger link density, the difference between the estimated and the simulated prevalence is small. With a larger effective infection rate τ, the difference is small, and the estimated and the simulated prevalence are close.

Comparing to the results of the real networks, we find the difference between the estimated and the simulated prevalence of the homogeneous network is smaller. The estimated prevalence is more accuracy and the estimation is precise in homogeneous networks. However, for the real networks which are usually heterogeneous, the difference between the estimated and the simulated prevalence is larger. Besides, the estimated metastable prevalence is also affected by the effective infection rate and the network density. For example, based on the table 3.2.1, the network density (0.149) of the Jazz musician network is higher than the network density (0.000715) of the Facebook network. Then, by comparing the Figure 4.4.4 with the Figure 4.4.1, the error of estimation of the Jazz musician network is smaller than that of the Facebook network. Jazz musician network has a precisely estimated prevalence, and the higher network density leads to a precise estimated metastable prevalence.

4.5 Discussion

In this chapter, the normalized cut-size property is applied to the epidemic process to estimate the metastable prevalence without simulation. The estimated metastable prevalence is derived by equation (4.3.5). By comparing the simulated metastable prevalence and the estimated metastable prevalence, we obtain some conclusion observed from above comparison results. For a larger effective infection rate τ, the estimated metastable prevalence is precise and is close to the simulated metastable prevalence. For a larger network density, the difference of the simulated and the estimated prevalence is smaller. Besides, the accuracy of the estimated prevalence depends on the homogeneity of the networks. In a homogeneous network, for example, ER random network, the estimated prevalence is more accurate comparing to that in a heterogeneous network. Overall, the accuracy of the estimated prevalence depends both on the underlying networks and the effective infection rate τ.

Chapter 5

Conclusion

This chapter concludes the thesis work. Contributions and conclusions are summarized in section 5.1. In section 5.2, we introduce the possible future works.

5.1 Contributions

In this thesis, we aim to obtain initial understanding and explore more about the cut-size property on real networks. We study the cut-size property from different aspects to gain a general conclusion of the cut-size property of real networks. By using seven real networks, we obtain the normalized cut-size and analyze the physical meaning for real networks. Besides, we apply the cut-size property on the epidemic process to estimate the prevalence.

The thesis achieved the following contributions:

- We study the mathematical properties of the cut-size in networks. To fairly compare the cut-size for different formulations of network cut, we define the normalized cut-size and derive its average and bounds by the isoperimetric inequality. Based on the random partition method, we find that the average normalized cut-size which is a constant determined by the network density. The average cut-size of the network is verified by the results of real networks, while the mathematical bounds seem to be very loose. For the real networks, the normalized cut-size has a centric phenomenon that the normalized cut-size converges to its average. Therefore, the theoretical bounds can not represent the normalized cut-size well. What is more, based on the degree partition method, we find an entire difference result that the average normalized cut-size is not a constant anymore. The normalized cut-size still concentrate to the average normalized cut-size but decreases along with the average normalized cut-size. The larger-degree nodes have the largest influence on the normalized cut-size when partitioning the networks. Moreover, the theoretical bounds still cannot represent the normalize cut-size in degree-based partition method, because of the centric phenomenon of the normalized cut-size.

- Because of the centric phenomenon of the normalized cut-size, we apply the average normalized cut-size to the SIS epidemic process for estimating the prevalence in the metastable state. We find that for the homogeneous random network, the estimation is better and accurate. However, for the real networks which are heterogeneous, the deviation between the estimation and the simulation is large and inaccurate. Also, other factors have the influence on the accuracy of the estimated prevalence, for example, the network density and the effective infection rate τ. With a larger effective infection rate τ, the precisely estimated metastable prevalence can be obtained. With a larger network density, the estimated metastable prevalence is more accuracy. According to the above results, the cut-size property can be applied to predict the epidemic behaviors and understanding the epidemic process.

5.2 Future works

In this thesis, we used seven real networks to study the cut-size property. More real networks can be applied in the study of the cut-size property. Also, the networks used in this thesis are a part of the original networks. For example, the Facebook network has millions of users while our Facebook network in this thesis has only 2888 nodes. For further works, we plan to study larger networks and obtain their statistical property of the normalized cut-size, since currently there is rare research focuses on the cut properties of complex networks. The cut-size property is also related to the epidemic process. Hence, the study of the cut-size property may contribute to further understand and control of the dynamical processes in real complex systems.

Appendix A

A.1 Average interconnection constant

Assume a random node i with degree d_i is selected in set K, and it contains l_i links connecting node i to other nodes also located in set K. Then, the number of links that connect node i to set S equals to $d_i - l_i \geq 0$. From the equation (3.3.2), we have

$$\overline{\eta}(G,k) = \frac{1}{\binom{N}{k}} \sum_{(\forall K)K=(i_1,i_2,\ldots,i_k),K\subset\{1,\ldots,N\},|K|=k} \frac{\sum_{j=1}^{k} d_{i_j} - l_{i_j}}{k} \tag{A.1.1}$$

the combination number of the set K in which contains a random node i, equals to $\binom{N-1}{k-1}$. K=$(i_1, i_2, ..., i_k)$ with k nodes, and then for j=1,2...,k, d_{i_j} is the degree of the node i, where node i is the j element of the set K. l_{i_j} is the number of neighbors of a node i located in set k, where node i is the j element of the set K. Then, calculating the combination number of the varies $l_{i_j} = 0, 1, 2, ..., k$, we obtain

$$\overline{\eta}(G,k) = \frac{\binom{N-1}{k-1}\sum_{i=1}^{N} d_i}{k\binom{N}{k}} - \frac{1}{k\binom{N}{k}} \sum_{(\forall K)K=(i_1,i_2,\ldots,i_k),K\subset\{1,\ldots,N\},|K|=k} \sum_{j=1}^{k} l_{i_j} \tag{A.1.2}$$

Where the first term of the right hand equals to d_{av}. To calculate the second term, we set l_{i_j}=m is the number of the neighbors a node i located in set K. The combination of the set K, in which contains a node i and its m neighbors, equals to $\binom{N-1-d_i}{k-1-m}$ multiple the combination of a node i with its m neighbors $\binom{d_i}{m}$. For a large d_i, it is not always possible to contain any number of its neighbors in set K, hence the minimum number of the neighbors of a node i can be contained in set K is $m = max(0, m + d_i - N)$ and the maximum number is $m = min(d_i, k)$, then we have

$$\overline{\eta}(G,k) = d_{av} - \frac{\sum_{i=1}^{N} \sum_{max(0,m+d_i-N)}^{min(d_i,k)} \binom{N-1-d_i}{k-1-m} \binom{d_i}{m}}{k\binom{N}{k}} \tag{A.1.3}$$

Further simplifying the above equation and reach the average interconnection constant is

$$\overline{\eta}(G,k) = d_{av}(1 - \frac{k-1}{N-1}) = d_{av}(\frac{N-k}{N-1}) \tag{A.1.4}$$

where d_{av} is the average degree of graph and k is the number of nodes in the set K and N is the graph size.

Appendix B

In the appendix B, the rest cut-size results of real networks are shown.

B.1 The cut-size of real networks

From figure B.3.1 to figure B.3.7, these figures show the cut-size results of networks, including real networks result, random graph and scale-free graph result. These results are obtained by the random partition method, which support foregoing conclusions in chapter 3.4.

B.2 The cut-size of real networks based on degree

From figure B.3.10 to figure B.3.13, these figures show the cut-size property of real networks based on degree partition method. The results are similar to the ones that are analyzed in chapter 3.5, which give the solid supports to the conclusions of the cut-size property.

B.3 Prevalence comparison

From figure B.3.14 to figure B.3.23, these figures show the comparison between the simulated metastable prevalence and the estimated metastable prevalence of real networks. These results are the supplement to prove the conclusions in chapter 4.

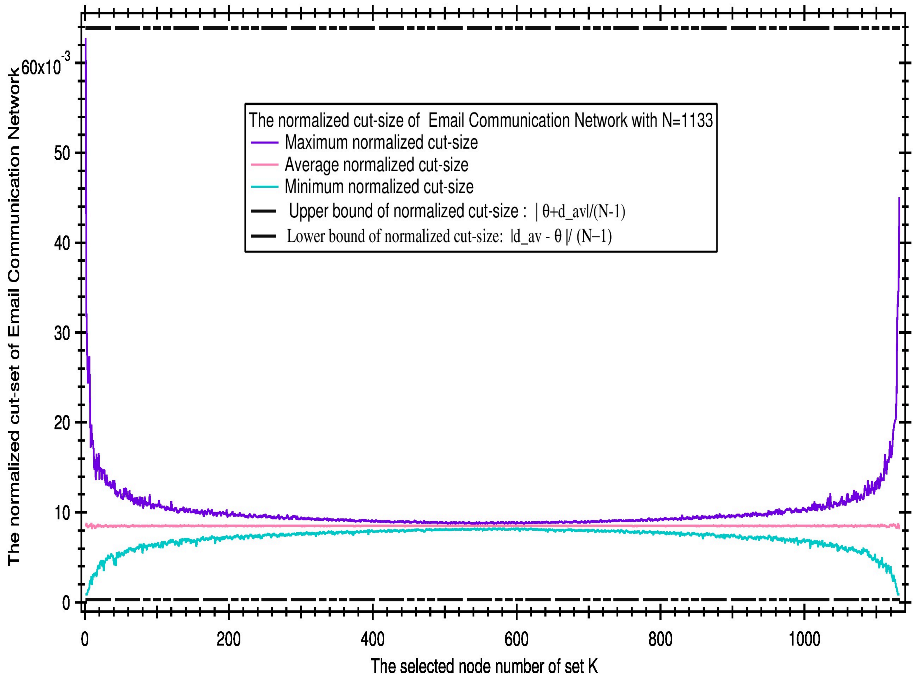

Figure B.3.1: The normalized cut-size of the Email communication network with N=1133.

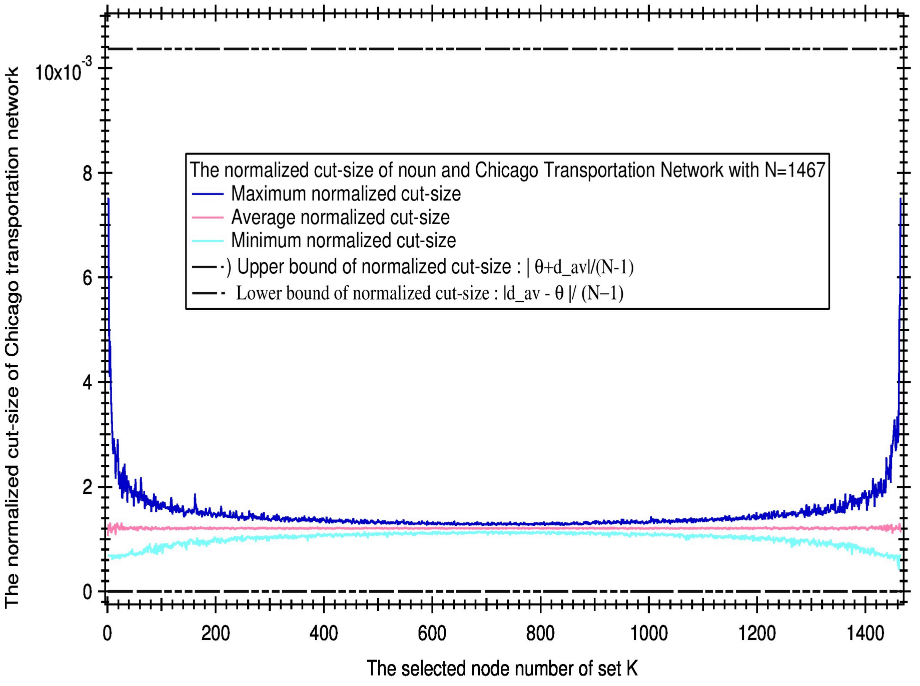

Figure B.3.2: The normalized cut-size of the Chicago transportation network with N=1467.

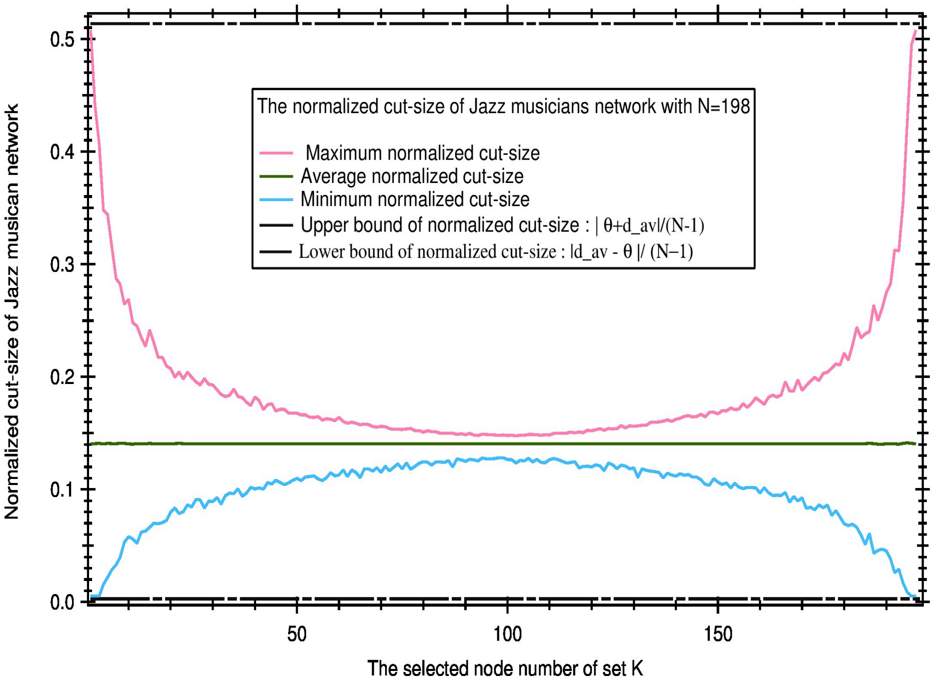

Figure B.3.3: The normalized cut-size of the Jazz musician network with N=198

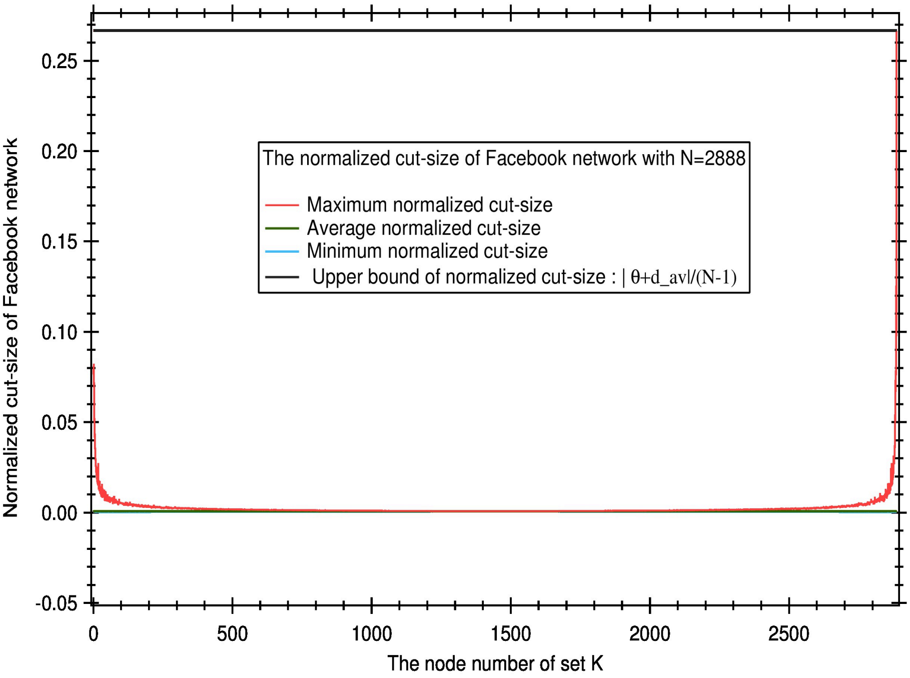

Figure B.3.4: The normalized cut-size of the Facebook network with N=2888

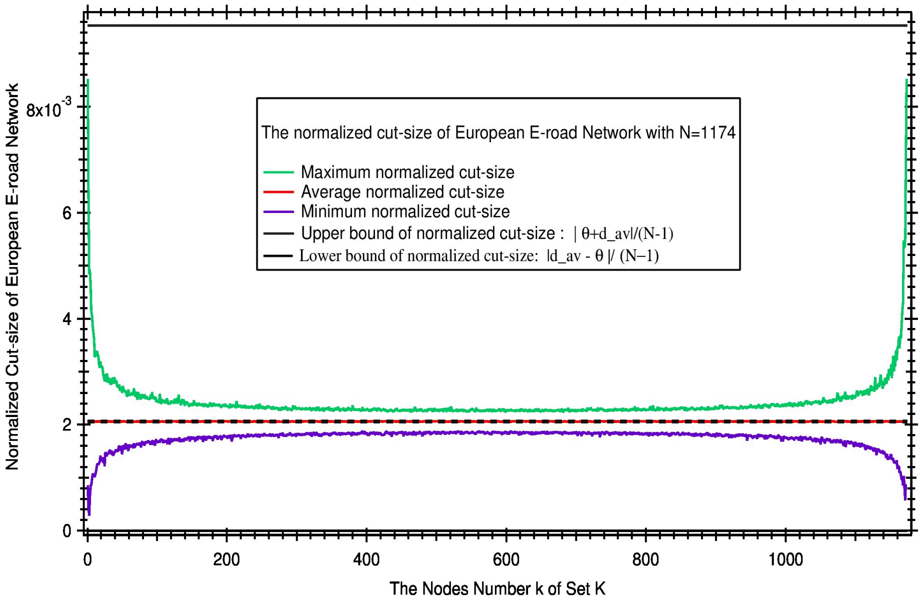

Figure B.3.5: The normalized cut-size of the Europe E-road network with N=1174

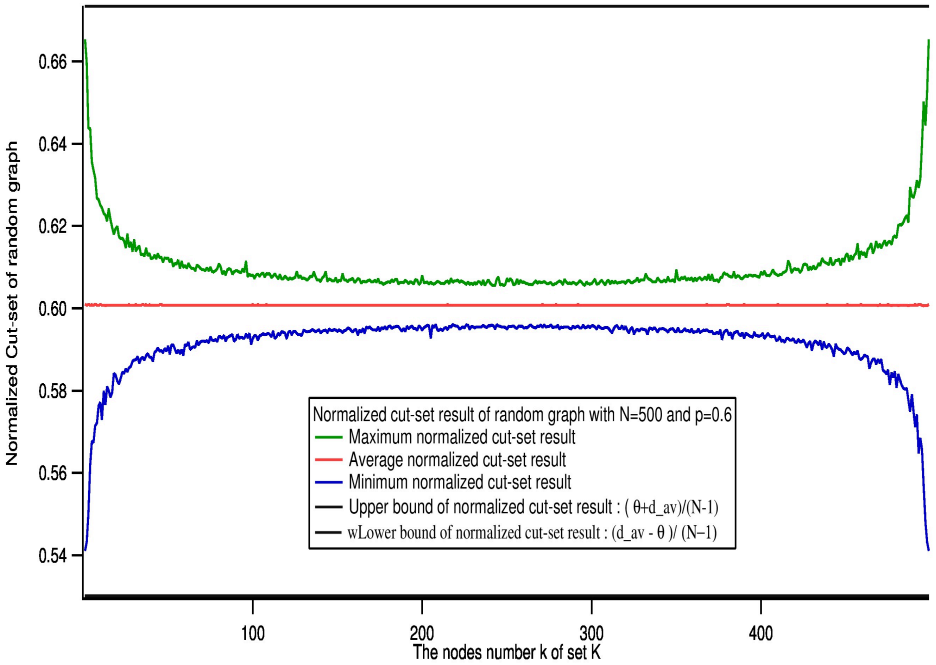

Figure B.3.6: The normalized cut-size of the random network with N=500

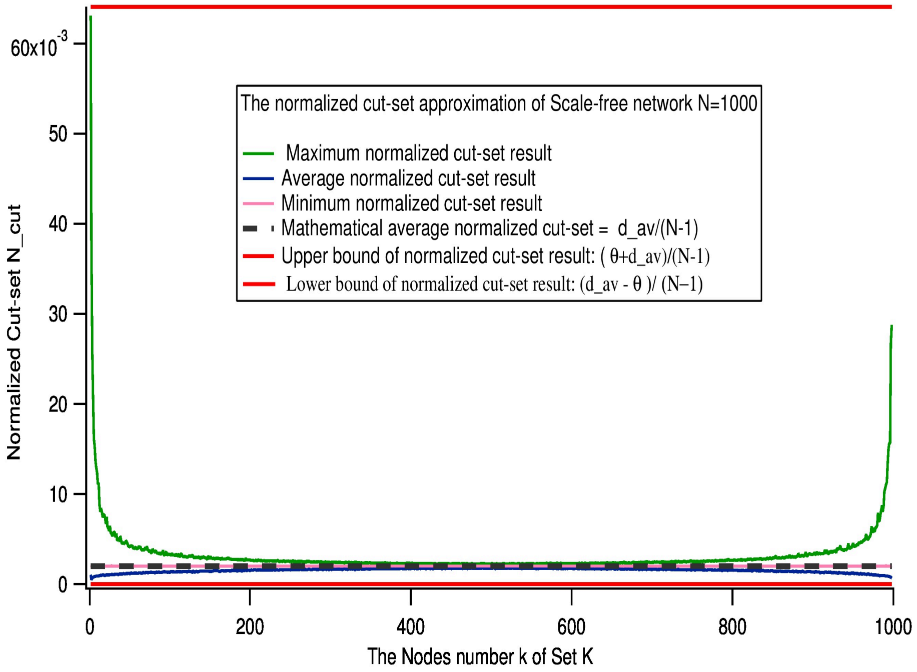

Figure B.3.7: The normalized cut-size of the Scale-free network with N=1000

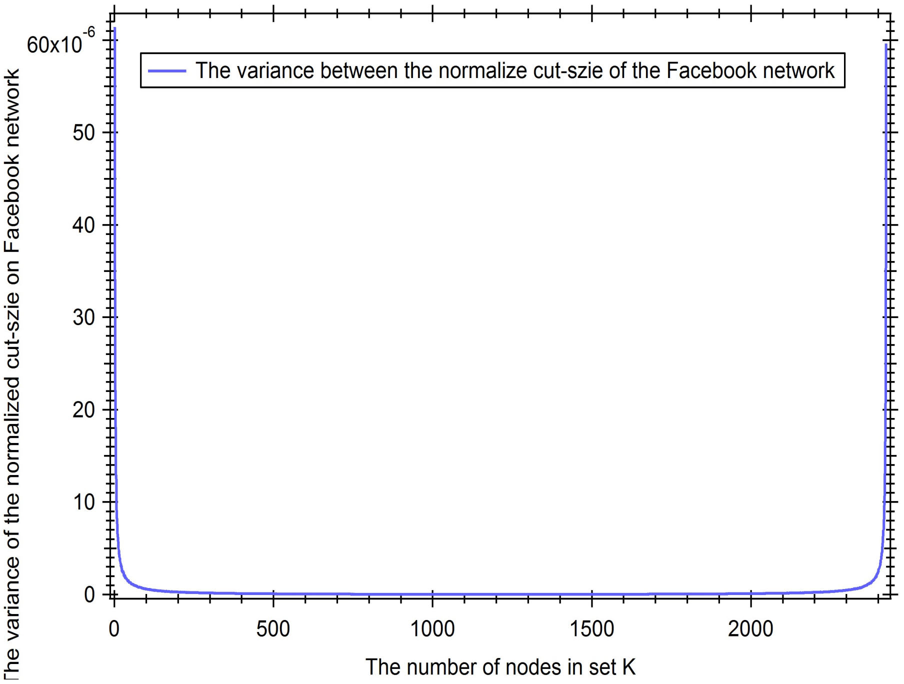

Figure B.3.8: The normalized cut-size variance of the Facebook network

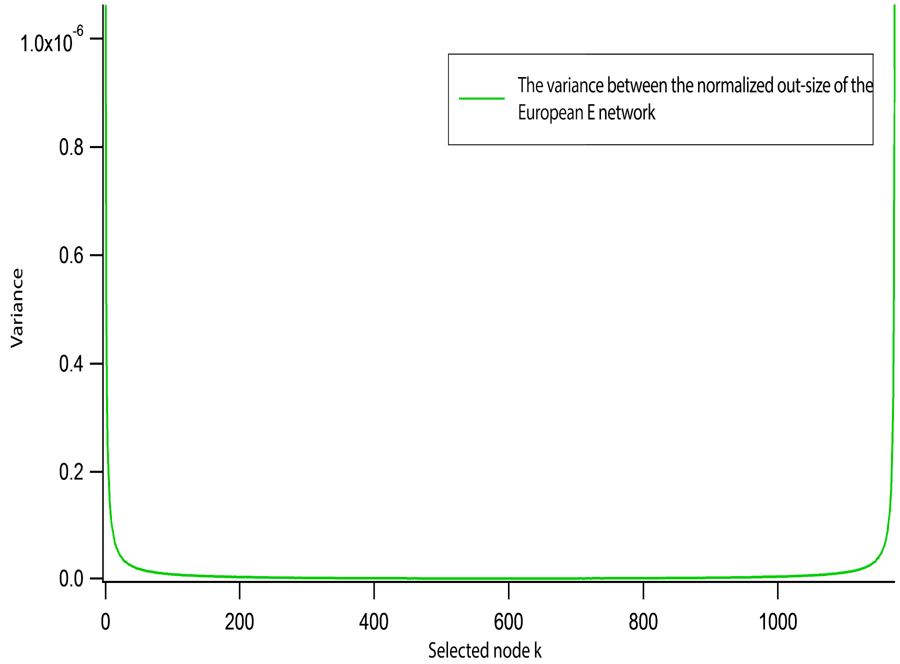

Figure B.3.9: The normalized cut-size variance of the European network

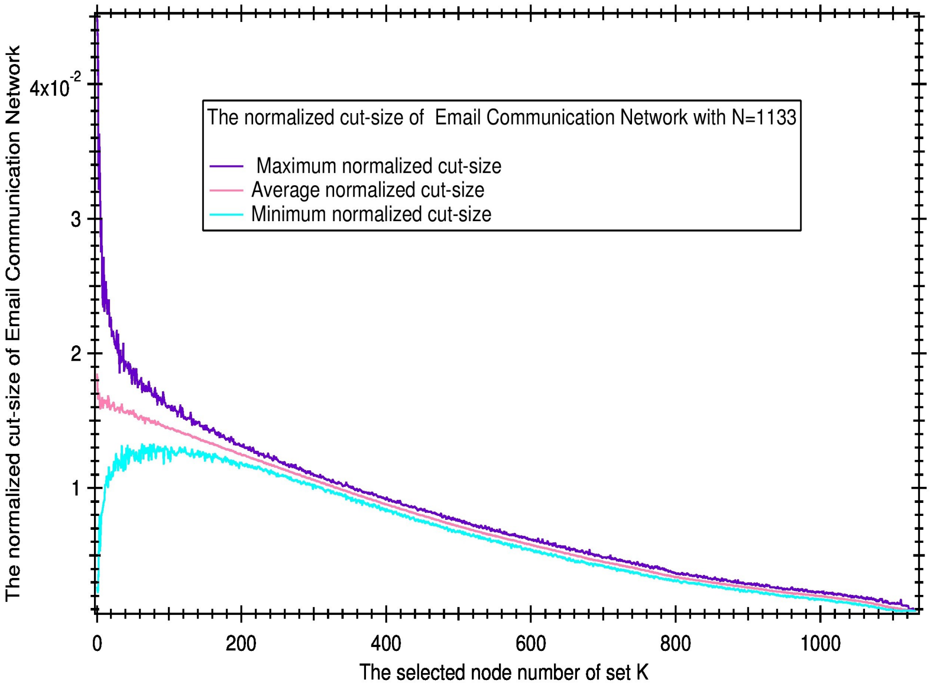

Figure B.3.10: The normalized cut-size of the email communication network based on degree

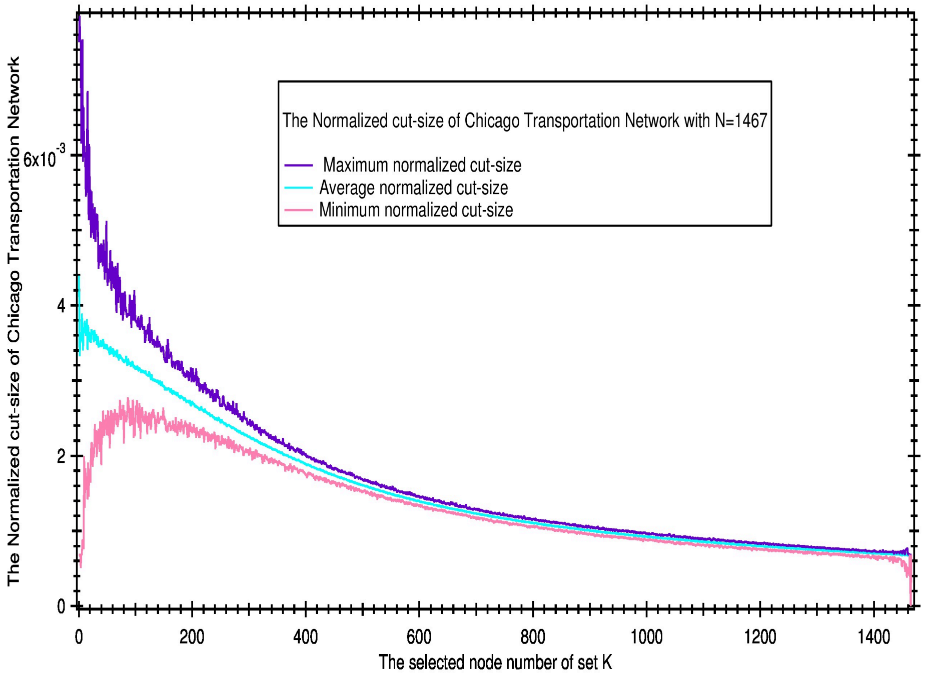

Figure B.3.11: The normalized cut-size of the Chicago transportation network based on degree

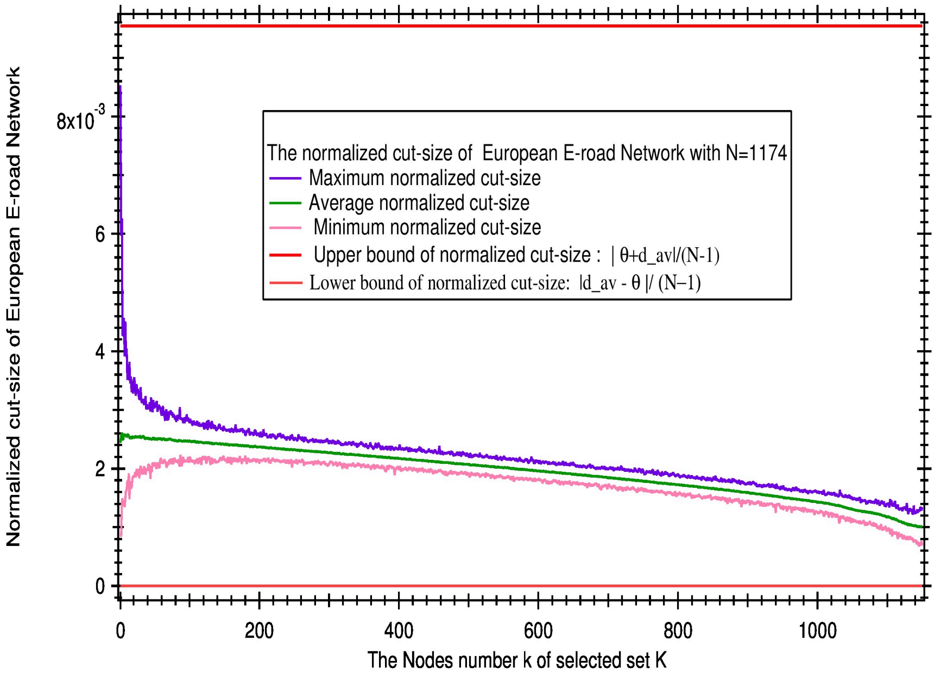

Figure B.3.12: The normalized cut-size of the Europe E-road network based on degree

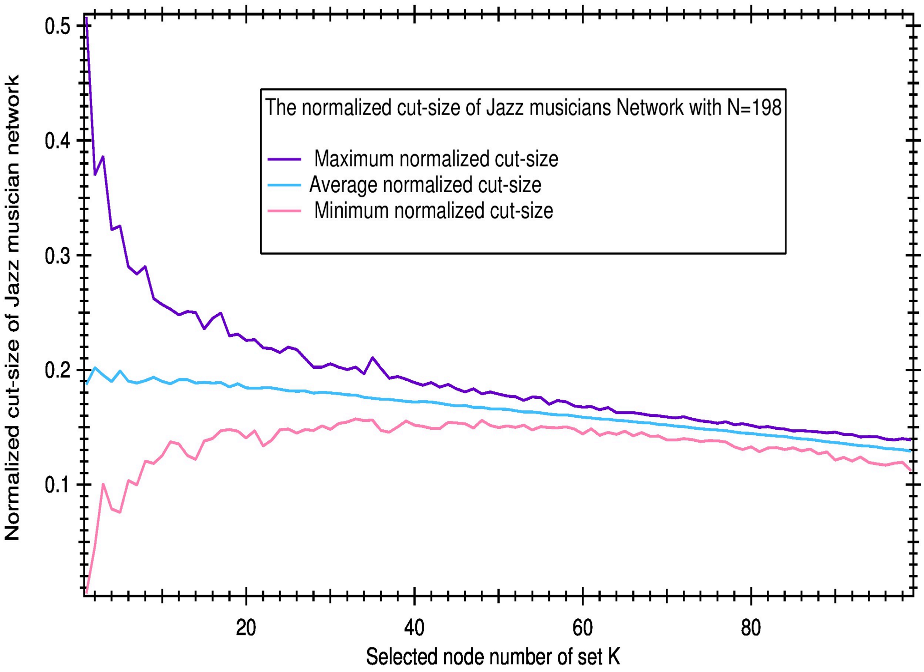

Figure B.3.13: The normalized cut-size of the Jazz musician network based on degree

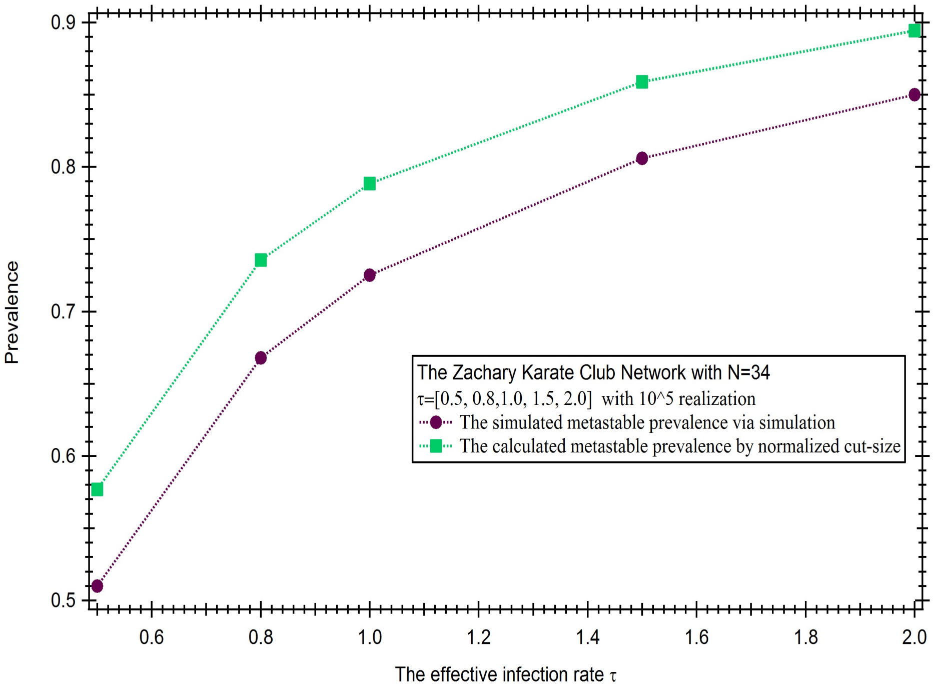

Figure B.3.14: The Prevalence comparison of the Zachary karate club network

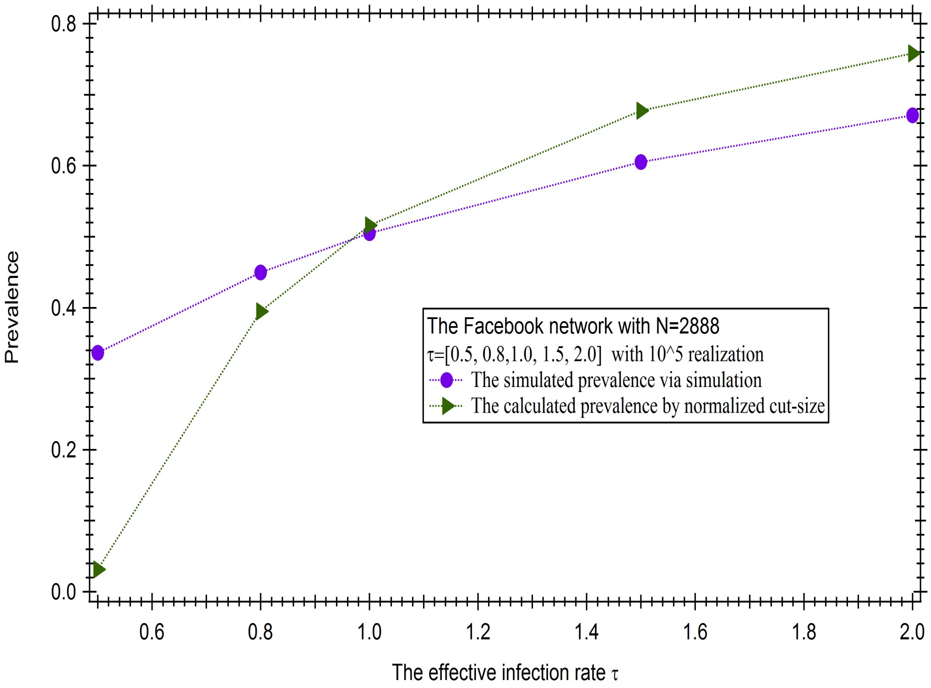

Figure B.3.15: The Prevalence comparison of the Facebook network

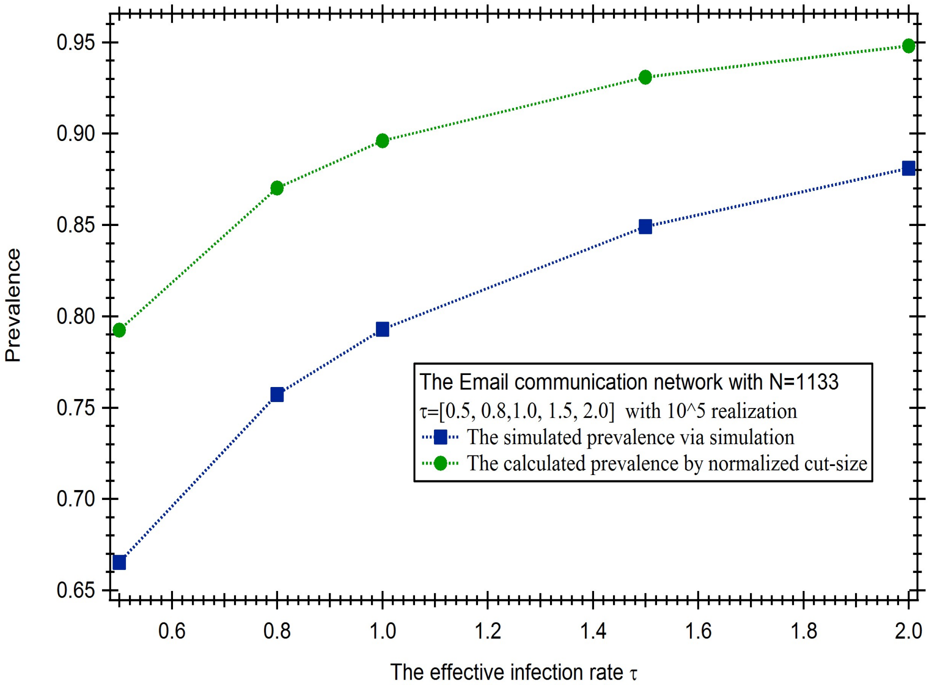

Figure B.3.16: The Prevalence comparison of the Email communication network

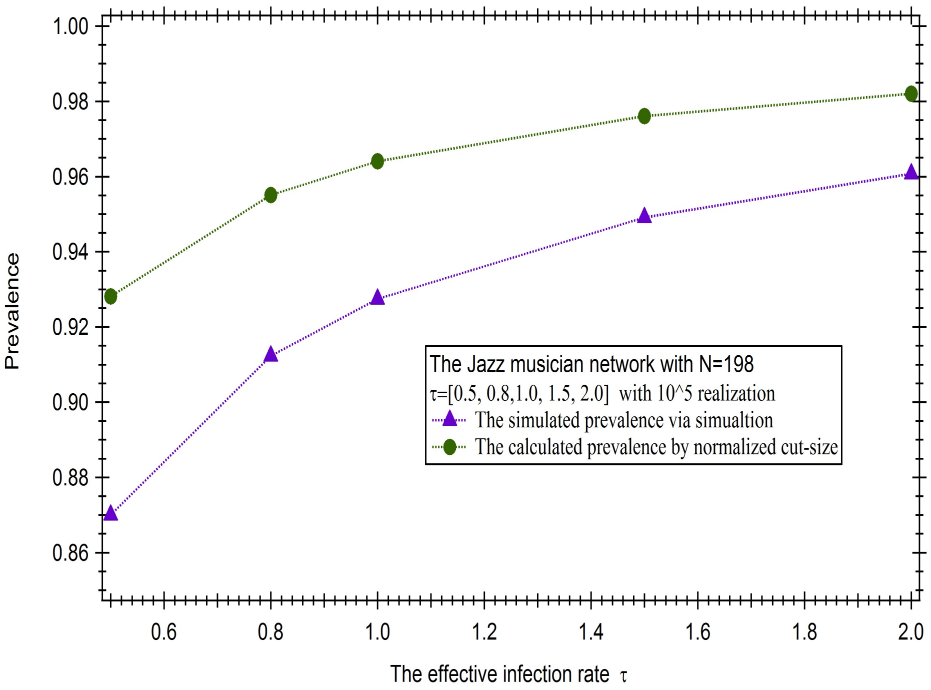

Figure B.3.17: The Prevalence comparison of the Jazz musician network

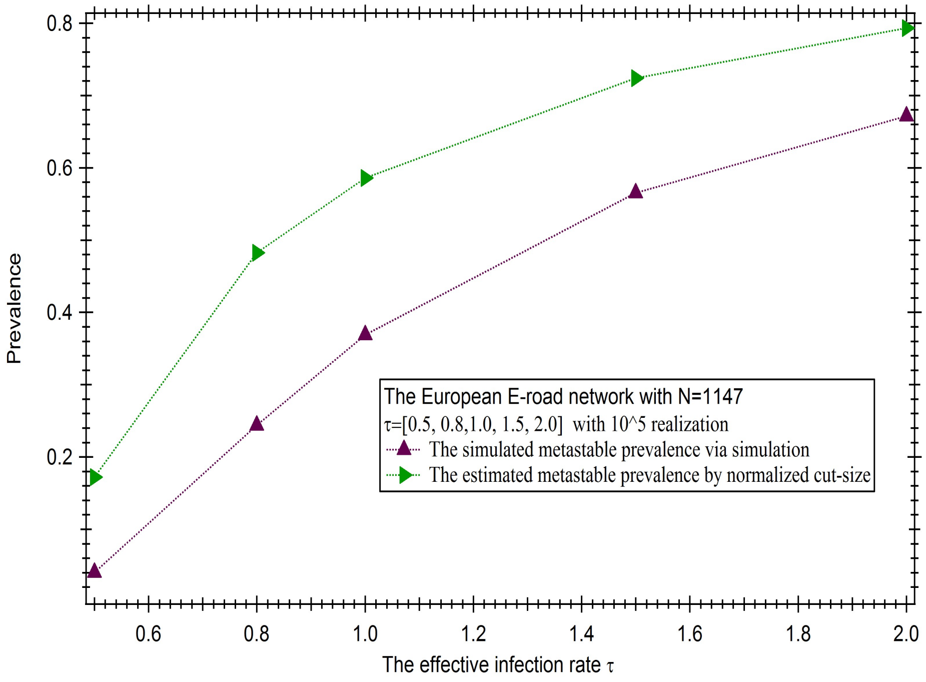

Figure B.3.18: The Prevalence comparison of the Europe E-road network

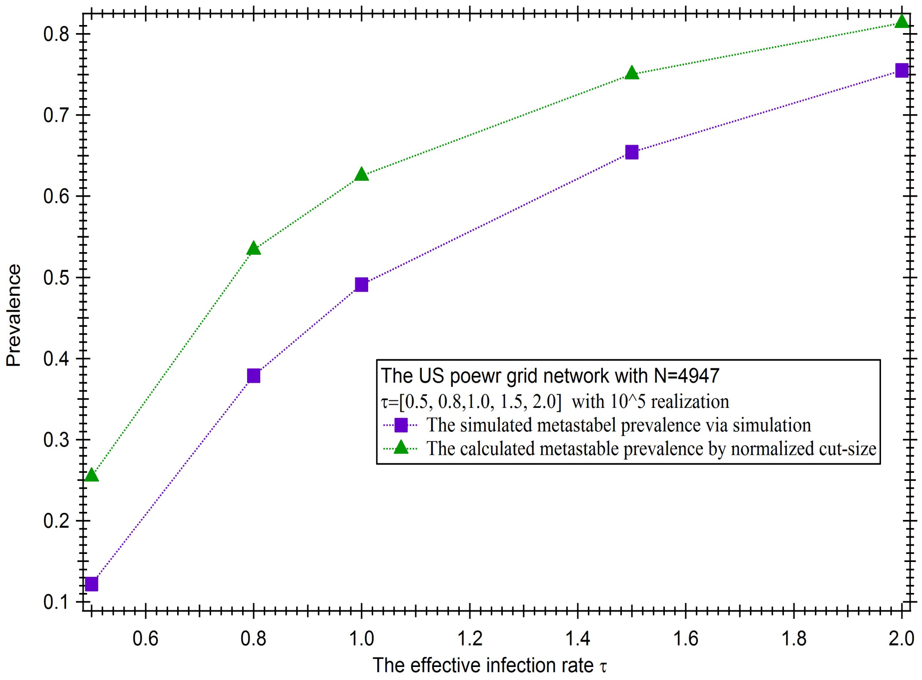

Figure B.3.19: The Prevalence comparison of US power grid network

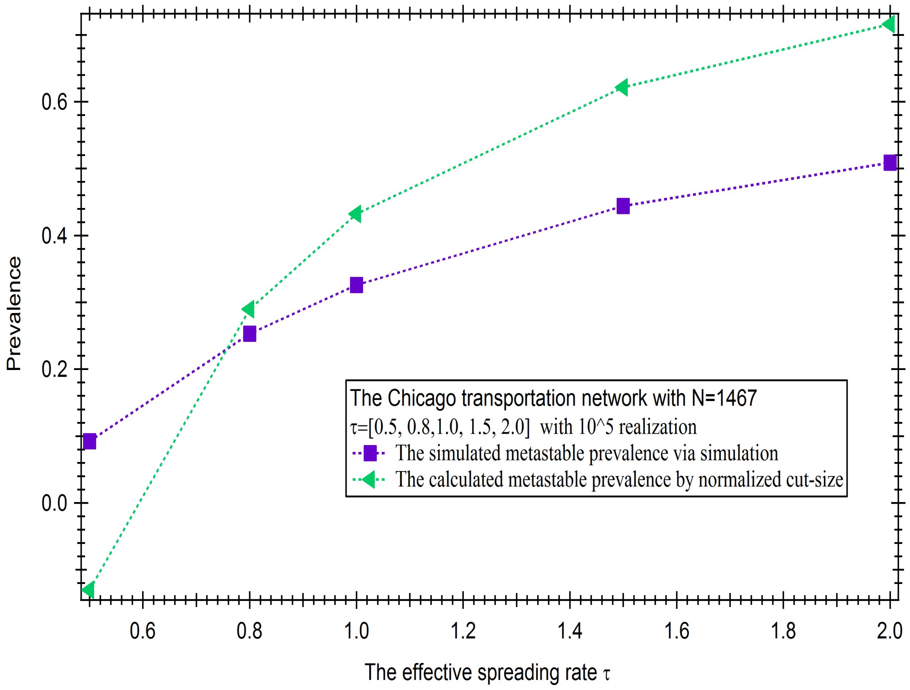

Figure B.3.20: The Prevalence comparison of the Chicago transportation network

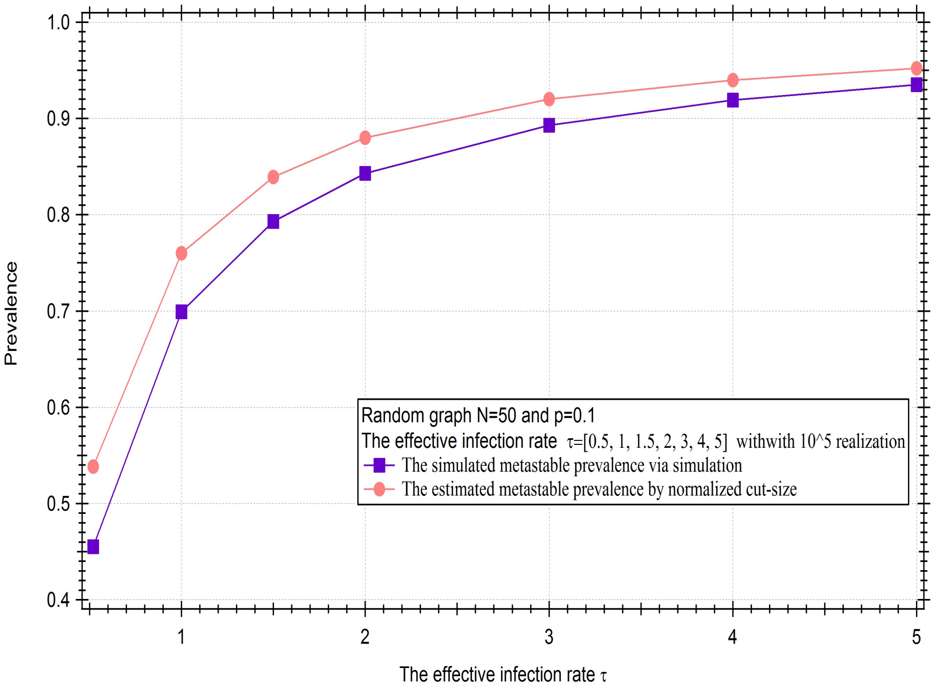

Figure B.3.21: The Prevalence comparison of random graph with link density *p*=0.1 for different effective infection rate τ

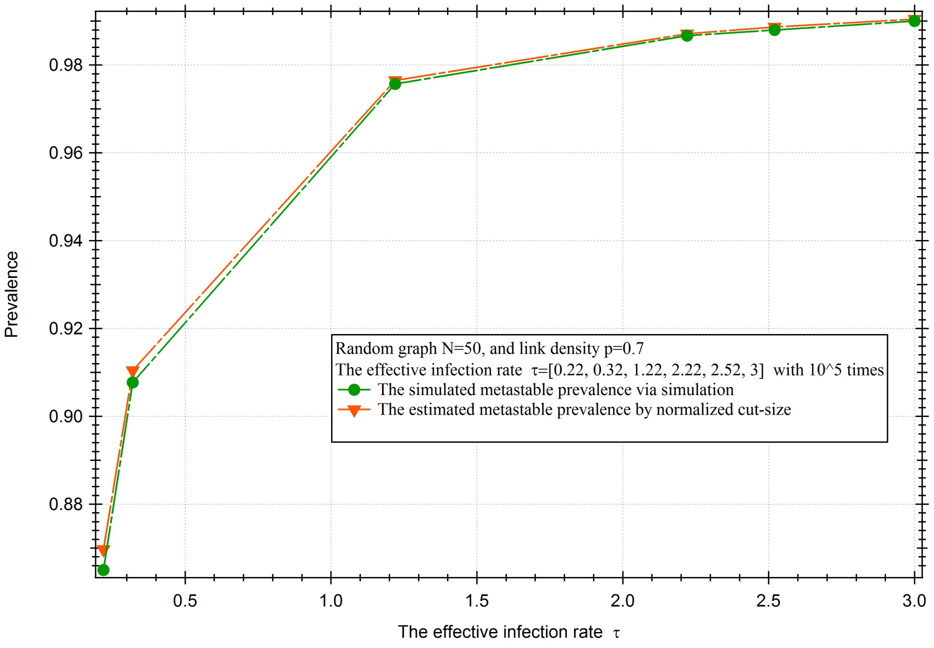

Figure B.3.22: The Prevalence comparison of random graph with link density p=0.7 for different effective infection rate τ

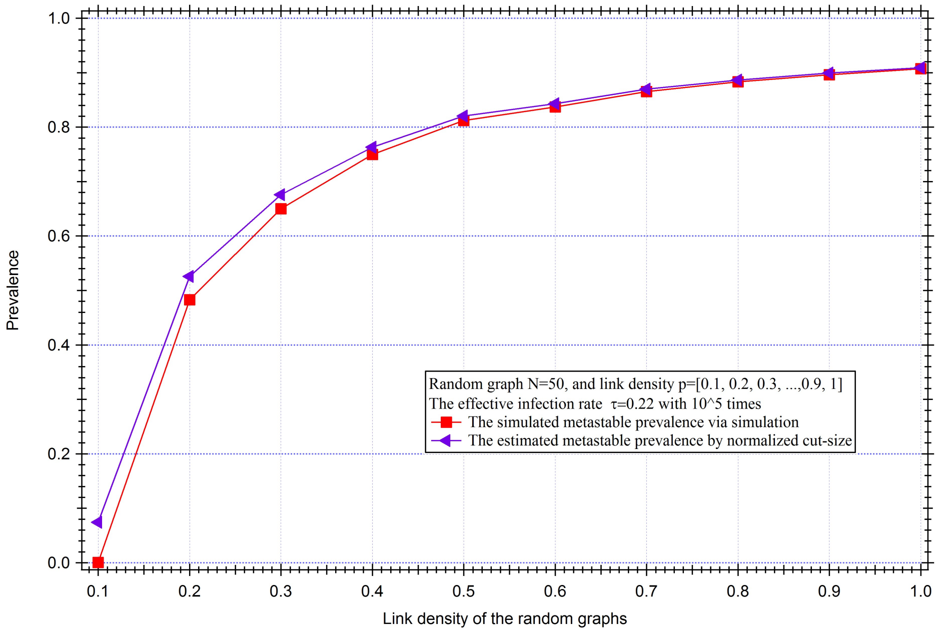

Figure B.3.23: The Prevalence comparison of Random graph with different link density

Bibliography

[1] F.Qing, W.Fan, Z.Gang et al, "New Interdisciplinary Science: Network Science", Progress in Physics (I), 27(3):239-343, 2007.

[2] M. Newman, Networks: an introduction. Oxford university press, 2010.

[3] D.Alderson , "Catching the "Network Science" Bug: Insight and Opportunity for the Operations Researcher". Operations Research, 56(5), pp.1047-1065, 2008.

[4] R. J. Wilson, "An eulerian trail through königsberg," Journal of graph theory, vol. 10, no. 3, pp. 265–275, 1986.

[5] P.Erdős, A.Rényi, "On Random Graphs". Publicationes Mathematicae, 6: 290–297, 1959.

[6] Duncan J. Watts and Steven H. Strogatz, "Collective dynamics of 'small-world' networks". 393(1):440–442, 1998.

[7] A.L.Barabásiand R.Albert. "Emergence of Scaling in Random Networks" . pp.509–512, 1999.

[8] N.Biggs, E.Lloyd, R. Wilson, "Graph Theory", 1736-1936, Oxford University Press, 1986.

[9] R.Pastor-Satorras, C.Castellano, P.Van Mieghem, and A. Vespignani, "Epidemic processes in complex networks", Reviews of modern physics, Vvol 87,no.3,P925, 2015.

[10] P. Van Mieghem, J.S. Omic and R.E. Kooij, "Virus Spread in Networks", IEEE Transactions on Networking, vol.17,no.1,pp.1-14, 2009.

[11] J. Watts, "THE 'NEW' SCIENCE OF NETWORKS", Annual Review of Sociology, 2004.

[12] P. Van Mieghem, "Universality of the SIS prevalence in networks". In: ArXiv e-prints, 2016.

[13] R. van de Bovenkamp, F Kuipers and P. Van Mieghem, "Domination-time dynamics in susceptible-infected-susceptible virus competition on networks", Physical Review E, Vol. 89, No. 4, p. 042818, 2014.

[14] P.Van Mieghem, Performance Analysis of Complex Networks and Systems. Cambridge, U.K.: Cambridge University Press, 2014. isbn: 9781107415874

[15] P. Van Mieghem. Graph Spectra for Complex Networks. Cambridge, U.K.: Cambridge University Press, 2011. isbn: 9780511921681

[16] J.Leskovec and E.Horvitz, "Planetary-Scale Views on an Instant-Messaging Network". arXiv:0803.0939 2007

[17] R.Albert; B,Albert, "Emergence of scaling in random networks". Science. 286 (5439): 509–512, 1999.

[18] R.Albert; B,Albert, "Statistical mechanics of complex networks". Reviews of Modern Physics. 74 (1): 47–97 2002

[19] D. Solla Price, D. J. "Networks of Scientific Papers". Science. 149 (3683): 510–515, 1965.

[20] M. E. J. Newman, S. H. Strogatz, D. J. Watts, "Random graphs with arbitrary degree distributions and their applications". Physical Review E. 64 (026118), 2001.

[21] K,Anderson, S.Lee, "Impact of Social Network Type and Structure on Modeling Normative Energy Use Behavior Interventions". Journal of Computing in Civil Engineering, 28(1), pp.30-39, 2014

[22] Konect Network Dataset – KONECT, April 2017.

[23] Euroroad network dataset – KONECT, April 2017.

[24] Us power grid network dataset – KONECT, April 2017.

[25] Facebook (nips) network dataset – KONECT, April 2017.

[26] Chicago network dataset – KONECT, April 2017.

[27] U. rovira i virgili network dataset – KONECT, April 2017.

[28] L. Šubelj and M.Bajec, "Robust network community detection using balanced propagation". Eur. Phys. J. B, 81(3):353–362, 2011.

[29] Zachary karate club network dataset – KONECT, April 2017.

[30] Jazz musicians network dataset – KONECT, April 2017.

[31] C.Jeff, “A lower bound for the smallest eigenvalue of the Laplacian”, Princeton, N. J.: Princeton Univ. Press. pp. 195–199. MR 0402831, 1970.

[32] F. Chung, “Four proofs for the Cheeger inequality and graph partition algorithms”, Fourth International Congress of Chinese Mathematicians, pp. 331–349, 2010.

[33] Q. Liu and P. Van Mieghem, “Die-out Probability in SIS Epidemic Processes on Networks”, Fifth International Workshop on Complex Networks and their Applications, Milan, Italy, Nov 30 - Dec 2, 2016.

[34] Q. Liu and P. Van Mieghem, ”Evaluation of an analytic, approximate formula for the time-varying SIS prevalence in different networks,”Physica A, vol.471, pp. 325-336, 2017.

[35] P. Van Mieghem, “The N -Intertwined SIS epidemic network model”, Computing (Springer), Vol. 93, Issue 2, p. 147-169, 2011.

[36] A. Ganesh, L. Massoulie, and D. Towsley. “The effect of network topology on the spread of epidemics” . In: Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies, 1455–1466 vol. 2, 2005.

[37] F. Chung, “Discrete Isoperimetric Inequalities” . In: Discrete Mathematics and Theoretical Computer Science, 1996.

[38] J.Omic, “Epidemics in Networks: Modeling, Optimization and Security Games”. Delft: Next Generation Infrastructures Foundation, 2010.

[39] The Daily Meal, “Millions of Toxic Eggs Recalled in Europe After Insecticide Scare”. [online] Available at: https://www.thedailymeal.com/toxic-eggs-recall-pesticide-scandal-germany-netherlands-europe/8617, 2017.

[40] P. Van Mieghem, “Time evolution of SIS epidemics on the Complete Graph” . In: Delft University of Technology, report20170405, 2017.

[41] P. Van Mieghem, “Approximate formula and bounds for the time-varying susceptibleinfected-susceptible prevalence in networks” . In: Phys. Rev. E 93, p. 052312, 2016.

[42] K.W¸egrzycki P. Sankowski, “Why Do Cascade Sizes Follow a Power-Law?”, 2017.

[43] M. E. J. Newman, “The structure and function of complex networks”, SIAM Review 45, 167–256, 2003.

[44] M. E.J. Newman, “Finding community structure in networks using the eigenvectors of matrices”. Physical Review E, 74(3), 2006.

[45] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, “Complex networks: Structure and dynamics”. Physics Reports 424, 175–308, 2006.

[46] M. E.J. Newman, “Modularity and community structure in networks”. Proceedings of the National Academy of Sciences, 103(23), pp.8577-8582, 2006.

[47] M. E.J. Newman, “Communities, modules and large-scale structure in networks”. Nature Physics, 8(1), pp.25-31, 2011

[48] M. Sekiguchi and E. Ishiwata, “Global dynamics of a discretized SIRS epidemic model with time delay,” Journal of Mathematical Analysis and Applications, vol. 371, no. 1, pp. 195–202, 2010.