國立政治大學理學院應用物理研究所

碩士論文

Graduate Institute of Applied Physics, College of Science

National Chengchi University

Master Thesis

計算大尺度複雜網路 ：以競賽網路及電力網路為例

Computational large-scale complex networks : competition network and power grid

劉彥宏

Liu, Yen-Hung

指導教授 ： 蕭又新

Advisor : Shiau, Yuo-Hsien

中華民國一〇一年七月

July,2012

Abstract

This thesis can be divided into two parts. In the first part, we review some basic properties of the complex networks. The most important features are: small world networks and scale-free degree distribution. Then, we introduce three complex models : BA model, EBA model, and W-S small world model. Next, we analyze a real data—CTTC network to test if it has the features we have mentioned above. By the EBA and BA model simulations, we try to illustrate why there are some similarities between the simulations and real data, but they are still so different in most of aspects.

In the second part, we review the definitions of the topology and reliable efficiency of a network structure. Next, we discuss two cascading failure model based on different definitions of load of a transmission line in a power grid. Finally, we use three different ways: topology efficiency vulnerability, cascading failure triggered by betweenness overload, and cascading failure triggered by the transient dynamics overload to test the vulnerability of edges in an assuming power grid. The cascading failure triggered by the transient dynamic overload can be viewed as a simplified power flow model. We sort the most vulnerable edges in three different ways. By this, we can observe the difference of the vulnerability analysis based on the complex network and the characteristic of the power transmission..

Keywords: small world, scale-free degree distribution, complex network, vulnerability, cascading failure

摘要

這篇論文主要可以分成兩個部分。第一部分，我們整理了關於複雜網路的初步研討。最重要的特性有：小世界網路、無尺度度分布。並且介紹了三種模型：BA 模型、EBA 模型，以及 W-S small world model。接者對於一份實際的社會網路資料—台灣業餘桌球選手對戰網路，做網路的結構分析，試驗其是否具有上述的兩種特性。透過兩種可以模擬出無尺度度分布特性的模型：BA 以及 EBA 模型。我們藉由這兩種模型模擬的結果，以及和競賽網路的比較，試者去闡述模型與理論間為何有些相似，卻又如此不同。並討論了賽制設計對於結構的影響。

在第二部分裡，我們回顧了一些對於網路的拓樸性效率以及可靠度效率的研討，並且討論了兩種不同負載定義下的連鎖故障行為。最後我們使用其中三種方法：拓樸性效率脆弱性、參與中間度(betweenness)過載引發的連鎖性故障行為，以及電力網路的動態電流變化造成的連鎖性故障，對於一個假想的電網做傳輸線的弱點排序。其中由動態電流過載(transient dynamic overload)造成的連鎖性故障可以視為一個簡化後的電力動態網路模型，藉由這三者間排序的不同，我們可以看到複雜網路分析以及基於電力網路傳輸特性所模擬的結果差異。

關鍵字 ： 小世界、無尺度度分布、複雜網路、脆弱性分析、連鎖故障行為

Content

1 Introduction 1

1.1 Background 1

2 About the network 5

2.1 Math and tools in the network 5

2.1.1 the basic structures of networks 5

2.1.2 Adjacency matrix 6

2.1.3 Degree and average degree 8

2.1.4 Path and Shortest path 9

2.1.5 Clustering Coefficient 10

2.1.6 Betweenness 11

2.1.7 Components and the largest component 12

2.2 Random network 13

2.2.1 clustering coefficient 13

2.2.2 degree distribution 14

2.2.3 shortest path and the largest component 15

2.2.4 Implementation of random network : shuffle method and configuration model 17

2.3 Important properties in the real networks 18

2.3.1 high clustering , small world, the small world quotient (Q) 19

2.3.2 Power-law degree distribution 20

2.4 The models 22

2.4.1 Watts and Strogatz small world model 22

2.4.2 Barabási–Albert model(BA model) 25

2.4.3 Extended BA model (EBA model) 29

2.5 Case study I: CTTC competition network and model simulations 32

2.5.1 Materials 32

2.5.2 Analysis of real data..33

2.5.3 BA model simulation 37

2.5.4 extend BA model simulation...40

2.5.5 discussion...43

3 Vulnerability and cascading failure of a power grid..47

3.1 Efficiency..47

3.1.1 topology efficiency 47

3.1.2 reliability efficiency 48

3.1.3 efficiency vulnerability 49

3.2 Cascading failures...50

3.2.1 Betweenness overload...50

3.2.2 transient dynamics 54

3.3 case study II: the imaginary British power grid......................................62

3.3.1 efficiency vulnerability 62

3.3.2 the cascading failure triggered by the betweenness overload......64

3.3.3 the cascading failure triggered by the transient dynamics overload 66

3.4 Discussion...68

4 Conclusion and future work...70

References...72

Figure 1 : Königsberg bridges problem and its network structure.................2

Figure 2 : the basic structures of networks. 5

Figure 3 : this is an undirected and unweighted network. There are three shortest path between A and C. Two of them pass through B. We give each shortest path a weight $1/3$, and the pair A and C contribute $2/3$ to the betweenness of B...12

Figure 4 : two components in a network..12

Figure 5 : the fraction of the largest component in a random network........16

Figure 6 : a search in a random network from a source...............................17

Figure 7 : “stubs” and configuration model..18

Figure 8 : A power law distribution is plotted in two different scales. 21

Figure 9 : degree distributions for(A) the movie actor network. (B) the internet(www) network. (C) the power grid network 22

Figure 10 : Random rewiring procedure for interpolating between a regular ring..23

Figure 11 : the W-S small world model 24

Figure 12 :degree distribution in the Watts-Strogatz small world model 25

Figure 13 : the degree distribution of model A at time step = 150000.........28

Figure 14 : the degree distribution of model B. N=1000. edge number = 1000..28

Figure 15 : the degree distribution of model B. N=1000. edge number = 150000...29

Figure 16 : the three different degree distributions generated by different parameter values in EBA model, time step = 100000..........................31

Figure 17 : the CTTC tournament design 32

Figure 18 : CTTC networks degree distribution for every year...................35

Figure 19 : The scatter diagram between k_i and CC_i from 2004 to 2011.....36

Figure 20 : The histogram for the CCi from 2004 to 2011 36

Figure 21 : degree distribution of BA model simulations............................37

Figure 22: The BA model simulation scatter diagram between ki and CCi 39

Figure 23 : The histogram for the CCi from 2004 to 2011 in BA model simulation...39

Figure 24 : extend BA model simulation power law 40

Figure 25 : the CC-k diagrams for the extend BA model simulations.........42

Figure 26 : The histograms of the clustering coefficients of the extend BA model simulations 43

Figure 27 : the “overlap” of three tournaments 45

Figure 28: the small world from “efficiency” view 48

Figure 29 : V^* test for the 2006 CTCC amateur table-tennis players competition network. 50

Figure 30 : the flow chart the cascading failure triggered by the betweenness-overload..51

Figure 31 : Cascading failures in the random network (homogeneous networks) $\bar{k}=3$, $N=5000$. The network is generated by the shuffle method. 52

Figure 32 : Cascading failures described in a scale-free network. The network is generated by using BA model. The nodes number N = 5000, $m_0=m=3$ 53

Figure 33 : Cascading failure failure in the western U.S. power grid.. 53

Figure 34 : A small power grid. 54

Figure 35 :　the imaginary British power grid..56

Figure 36 : the load variation of edge D with the evolution of time............57

Figure 37 : the load variations of three different edges after the removal of edge D...58

Figure 38 : the flow chart of the cascading failures triggered by the transient dynamics of a power grid. 59

Figure 39 : the cascading failures triggered by transient dynamics of the power grid. The data is from [3] 60

Figure 40 : the the group number variations of the removal of single edge 61

Figure 41 : the ten most vulnerable edges according to th topological vulnerability evaluation 63

Figure 42 : the locations of the ten most vulnerable edges according to the cascading failures triggered by the betweenness overloads in the imaginary British power grid. 65

Figure 43 : the locations of the ten most vulnerable edges according to the cascading failures triggered by the transient dynamics overloads .67

Figure 44: the two different cascading behaviors the two strategy of the removal of edges...68

Table 1 : In [3], Duncan J. Watts and Steven H. Strogatz list the average shortest path lengths and the clustering coefficients for three empirical datas. 20

Table 2 : the scale-free properties of three different network 21

Table 3 : the network sizes and the power-law exponent of the CTTC competition networks........34

Table 4 : CTTC network properties analysis 34

Table 5 : BA model simulation result 38

Table 6 : the extend BA model simulation results 40

Table 7 : the small world properties of the EBA simulation........41

Table 8 : the ten most vulnerable edges according to the topological vulnerability evaluation 63

Table 9 : the ten most vulnerable edges in the imaginary British power grid according to the cascading failures triggered by the betweenness overloads........65

Table 10 : the ten most vulnerable edges according to the cascading failures triggered by the transient dynamics overloads........67

1 Introduction

1.1 Background

During the past two decays, the “network science” has attracted many attentions from mathematician, physicist, and sociologist. All the structures we mentioned below are structures existing in nature and human activity. Internet, a collection of computers linked by data connections. Food web, depicts feeding connections in an ecological community. Social network, people are connected if one knows another one. Lexical networks, the words are linked if they exist in a sentence. Neural network, neurons are connected by synapses. Basically, all these structures can be viewed as a combination of the individuals, and there are interactions between those individuals. “Network” is a science of a simplified representation to reduce a structure capturing only the topology properties. In this manner, we can identify their characteristics and simulate their behaviors in many different conditions.

The history of network can be traced back to Königsberg Bridge Problem which was solved by Leonhard Euler. There were seven bridges across the river which is through the city of Königsberg. The problem is “does any single path crosses seven bridges exactly once which is called Eulerian path exist?”. Figure 1 shows the map and the network structure of the problem.

Figure 1(a) is a map of 18th century Königsberg. Figure 1(b) is a simplified pattern of Königsberg. Figure 1(c) is the network structure of Königsberg Bridge Problem. This figure is from [1]

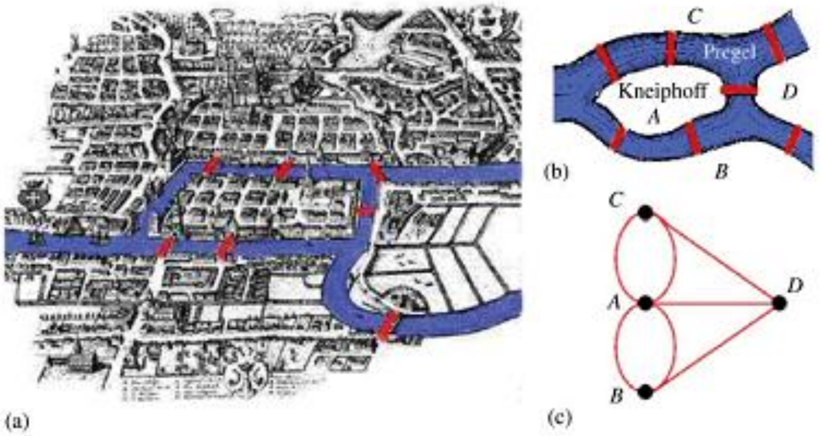

Figure 1 : Königsberg bridges problem and its network structure

In Figure 1(c), we can see the network structure of the Königsberg Bridge Problem. The solution of this problem can be simplified as below: the Eulerian path traverses each edge once. These kinds of paths should enter and leave the nodes which are passing through except the source node and the end node. It means there can be at most two nodes having odd numbers of degree in network language. However, all the nodes in Figure 1 have odd degrees so that there is no Euler path existing in the structure.

The mathematical tools used to solve this problem is consider to be the first theorem in *graph theory* which is used to described the network structures by the researchers who study networks nowadays. These basic mathematical tools of network will be described reference in 2.1 and 2.2. These descriptions are mostly from [2]. In this manner, the well-known Erdős–Rényi model of random graph was developed by Paul Erdős and Alfréd Rényi. The properties of the random graph do not match the real data . We will discuss these differences in 2.3.

There are some modern features of the science of network indicated in [1]., they are :

- Focus on the real world network and concern the theoretical and empirical questions

- View networks as system evolving in time according dynamical rules

The example is like the “small world experiment”[3] in Sociology which is a empirical study. In 1998, Duncan J. Watts and Steven H. Strogatz published their famous paper, “Collective dynamics of small world networks”[4], and the model they presented is the first model realizing the properties of the small world networks via a simple dynamical rule. We will explain it clearly in2.4.1. The other properties the researchers found in empirical data of networks is “scale-free degree distribution.” In 1999, Reka Albert and Albert-Laszlo Barabasi gave the first model in their paper “Emergence of Scaling in Random Networks” to reproduce the observed scale-free degree distribution in real data. The model is called BA model in the community of network science. The model is composed of a network structure and two generic mechanisms which we will discuss this more clearly in 2.4.2.

There is another research domain about the robustness of the complex structure. It was found that the scale-free networks have a higher degree of error tolerance than random network (Erdős–Rényi model), but error tolerance comes at a price in that the scale-free networks are extremely vulnerable to attacks (that is, the removal of some nodes that play important roles in the network connectivity. For this purpose, we can remove the nodes having highest degree or betweenness.)[5, 6]. There are many infrastructures composed of network structures, such as telecommunications, gas and water supply, transportation, and power grid. Power grid can be represented as a network of *n* nodes and *k* edges. Nodes are generators, substations and transformers and the edges are transmission lines. The power grid is vulnerable to natural disasters and physical attacks. It means the nodes and edges having a high probability to be remove from the power grid because they fail. We will review previous research about the vulnerability and cascading failure in power grids and apply them on a assuming power grid from [7].

1.2 Purpose

In this thesis, we review the history of the complex network and test that does CTTC network have the properties we observed in many real network data. Further, we apply three methodologies test the robustness of a imaginary power grid.

The thesis can be divided into two parts: (1) analysis of CTTC network (2) vulnerability analysis of power grid. In (1), The purpose is to explain the structure of CTTC network and compare the real data with two models: BA model and EBA model. In (2), we apply the topology efficiency vulnerability and two cascading models in network science to see if we can identify the most vulnerable lines in an assuming power grid.

2 About the network

2.1 Math and tools in the network

2.1.1 the basic structures of networks

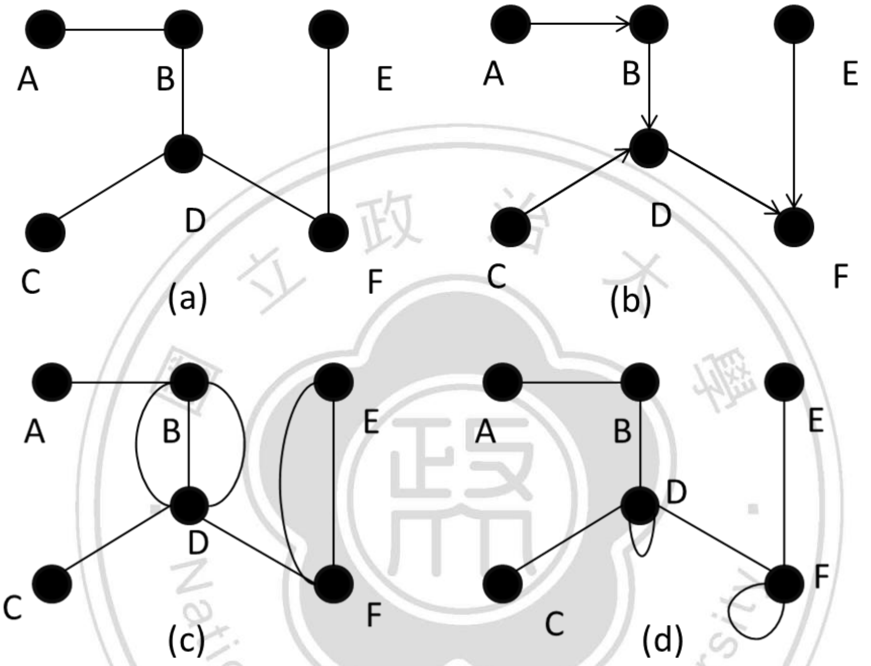

Figure 2 : the basic structures of networks.

In the Figure 2(a), the small network is consisted of five nodes and five edges. In mathematical language, we can label these nodes by integers from1 ... n. For the example above, we can replace A for 1, B for 2, and C for 3…etc. In this way, we can denote an edge between nodes i and j by (i, j). The whole network can be represented by the value of n and a list of all the edges. In the figure 2.1(a), the network has $n = 6$ nodes and edges $(1,2), (2,4), (3,4), (4,6)$ and$(5,6)$. In (a), the edges have no directions. In fact, we can assign directions to edges to describe relations in networks. Like (b), the components in the network are all the same with the one in (a),but the edges have directions. The arrows denote the directions of the edges. In (c), there are multi-edges between nodes B,D and nodes E,F. In(d), the circles on nodes D and nodes F are called self-edges.

2.1.2 Adjacency matrix

The adjacency matrix $\boldsymbol{A}$ of a graph is the matrix with elements A_{ij} such as

$$A_{ij} = \begin{cases} 1 & \textit{if there is an edge between node } j \textit{ and node } i \\ 0 & \textit{otherwise} \end{cases} \tag{1}$$

For instance, the adjacency matrix of the network in Figure 2(a) is

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

There are two things about the network we need to notice, first, the diagonal matrix elements are all zero and second, the matrix is symmetric. The reason of the first property is that there is no self-edge in the network. The self-edge means an edge start and end at the same node. And the reason of the second property is the network is undirected, which means if there is an edge between i and j, there is an edge between j and i

For a directed network

$$A_{ij} = \begin{cases} 1 \; if \;\; there \; is \; an \; edge \; from \; nod \\ 0 \quad otherwise \end{cases} \tag{2}$$

For the network in Figure 2 (b), which is to say

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

In Figure 2(c), we may see two different cases called multiedges and self-edges. In adjacency matrix, a multiedge is represented by setting the corresponding matrix element A_{ij} equal to the multiplicity of the edge. For instance, in (c), the adjacency matrix is

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 3 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 & 2 & 0 \end{pmatrix}$$

If there is a self-edge on a node i, the corresponding diagonal element A_{ii} is equal to 2. This is because every self-edge has two origins and ends. Like all non-self-edges appear twice in the adjacency matrix, an edge connecting node i and node j means A_{ij} and A_{ji} are 1. If we count edges equally, self-edges will appear twice. The adjacency matrix of the network in Figure 2(d) is

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 2 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 6 \end{pmatrix}$$

2.1.3 Degree and average degree

The degree of a node in a network means the number of edges connected to it. Degree of node i is denoted by k_i. In an undirected network, the degree can be represented in terms of the adjacency matrix as

$$k_i = \sum_{j=1}^{n} A_{ij} \tag{3}$$

Furthermore, in an undirected network, each edge had two ends and the sum of degrees is equal to the number of the ends of the edges in the network. if there are m edges , then

$$m = \frac{1}{2}\sum_{i=1}^{n} k_i \tag{4}$$

In the research of the network, there is an important value called the “average degree” of a network. The average degree $\bar{k}$ is defined as

$$\bar{k} = \frac{1}{n}\sum_{i=1}^{n} k_i \tag{5}$$

In a directed network, degree is divided into two kind, one is in-degree, and the other is out-degree. The in-degree is the number of ingoing edges connected to a node and the out-degree is the number of the outgoing edges connected to a node. Similar to the undirected network, in terms of the adjacency matrix, the in-degree and the out-degree can be written as

$$k_i^{in} = \sum_{j=1}^{n} A_{ij} \qquad k_j^{out} = \sum_{i=1}^{n} A_{ij} \tag{6}$$

The number of edges m in a directed network is equal to the total number of ingoing ends of edges or the total number of outgoing ends of edges at all nodes. That is

$$m = \sum_{i=1}^{n} k_i^{in} = \sum_{j=1}^{n} k_j^{out} \tag{7}$$

2.1.4 Path and Shortest path

A path in a network is a sequence of nodes such that every successive pair of nodes in the sequence is connected by an edge in the network. In an unweighted and undirected network, the length of a path is equal to how many edges is the path composed with? Consider two nodes *i* and *j*. If there is a path from *j* to *i* via some node *k*, then the product $A_{ik}A_{kj}$ is equal to 1, and the length of the path is 2. In this manner, the total number of paths of length 2 from j to i, via any other vertex can be written as

$$N_{ij}^{(2)} = \sum_{k=1}^{n} A_{ik} A_{kj} = [A^2]_{ij} \tag{8}$$

In Figure 2(a), which path between nodes A and F is the shortest path if the network is undirected and unweighted? Obviously, the path $A \to B \to D \to F$ is the shortest path between nodes A and F. The shortest path between nodes *i* and *j*, also called geodesic path, can be described as the smallest value of *l* such that $[A^l]_{ij}$>0. We can define a "shortest path matrix " ***d*** for a network. The element d_{ij} denotes the shortest path from node *i* to node *j*. Depending on the network is weighted or unweighted, there are several different algorithms to compute the shortest path.

Because in the study of the real cases in this paper, we focus on the unweighted and undirected networks only, only the BFS algorithms is adopted to calculate the shortest path and the largest component number in this paper.

The average shortest path length L between all pairs can be defined as

$$L = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}}{n(n-1)} \tag{9}$$

The shortest average path length can be called the average path length, or path length for simplicity.

2.1.5 Clustering Coefficient

Before we talk about the clustering coefficient, recall a relation in mathematics. For example, if a = c and c = b then it follows a = b also, because “ = ” is transitive. In social networks, we can ask a question like “What is the probability of my friend A also knows other one of my friends B” And we can view the question as the transitivity of the network. Because, in a network, the relation if people know each other can be expressed by if the nodes are connected by edges, we can quantify the transitivity as follow:

- *global clustering coefficient*

For an entire network, we can define a value to measure the probability. The value is called the global clustering coefficient C_g.

$$C_g = \frac{\text{number of closed paths}}{\text{number of paths of le}} \tag{10}$$

It is to say

$$C_g = \frac{\text{number of triangles}}{\text{number of paths of length two}} \tag{11}$$

- *local clustering coefficient*

For a single node in a network, the local clustering coefficient is

$$C_l = \frac{\text{numbers of pairs of neighbors of i that are connected}}{\text{number of pairs of neighbors of i}} \qquad (12)$$

2.1.6 *Betweenness*

Betweenness of a node is the total number of data packets passing through the node when every pair of nodes send and receive a data packet along the path connecting the pair[8]. These data could be the messages, news, information, or rumors on a social network[2]. We can simply assume these data always take the shortest path through the network. Then the betweenness B_i of a node *i* is given by

$$B_i = \sum_{s \neq t} n_{st}^i \qquad (13)$$

For a network, there are more than one shortest path between a pair, the betweenness of node *i* can be written as

$$B_i = \sum_{s \neq t} \frac{n_{s \neq t}^i}{g_{st}} \qquad (14)$$

This means we assign each shortest path between the pair a weight equal to the inverse of the number of the shortest paths between the pair. And then the betweenness of node *i* is the sum of the weights of all shortest paths passing through the node. There is a simple example illustrated in Figure 3. There are three shortest path between A and C. Two of them pass through B. We give each shortest path a weight $1/3$, and the pair A and C contribute $2/3$ to the betweenness of B.

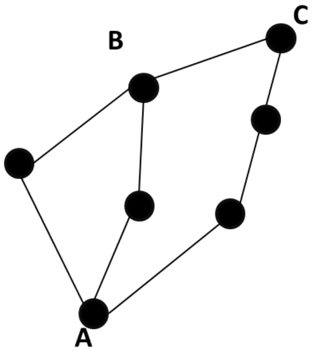

Figure 3 : this is an undirected and unweighted network. There are three shortest path between A and C. Two of them pass through B. We give each shortest path a weight $1/3$, and the pair A and C contribute $2/3$ to the betweenness of B.

2.1.7 Components and the largest component

In Figure 4, the network is divided into two groups. There is no path from any node in the left group to any node in the right. For example, there is no path from the node labeled ***1*** to the node labeled ***2.*** A network like this is called disconnected. Conversely, if there is a path for each pair of nodes in the network, it is connected. In Figure 4 the two small groups are the components of the whole network.

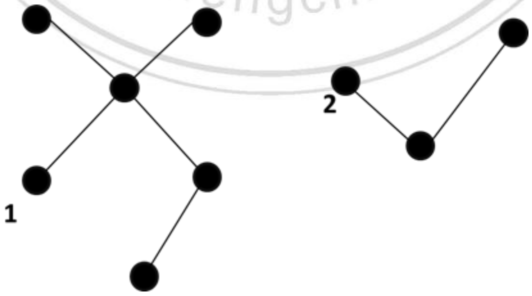

Figure 4 : two components in a network

The largest component is the component having the most nodes in the network.

2.2 Random network

A random network is a network in which there are some parameters taking fixed values, but in other respects, the network is random. For example, if we fix the number of the edges *m* and the number of the nodes *n*. That is to say, we take n nodes and place m edges among them randomly. This kind of random network model is usually referred to by its mathematical expression *G(n, m)*. Because it is hard to calculate the properties of the random network in this kind of expression, the other expression is often adopted by the science and mathematics communities. The model is called *G(n, p).* In this model, the number of nodes, *n*, and the probability of edges existing between nodes, *p*, are fixed. Paul Erdős and Alfréd Rényi published a series of papers about the model in the 1950s. The model is often called the "Erdős–Rényi model". The first property we can calculate is the average degree $\bar{k}$. It is obvious that the average number of the nodes for a node *i* will connect is $p(n-1)$. That is to say

$$\bar{k} = p(n-1) \tag{15}$$

2.2.1 *clustering coefficient*

Recall in 2.1.5, the clustering coefficient in a social network can be viewed as the probability that two neighbors of a node are also neighbors of each other. And this relation is called transitivity. For a random network case, the probability that there is an edge between two neighbors of a node is equal to the probability that there is an edge between any two chosen pairs. As we know, the latter is equal to p.

$$C_g^{rand} = p = \frac{\bar{k}}{n-1} \tag{16}$$

2.2.2 degree distribution

Degree distribution p_k is the probability that a node connects other k nodes in a network. The probability of a node connects other particular k nodes but not to other $n-k-1$ nodes is $p^k(1-p)^{n-k-1}$. There are C_{n-1}^k ways to choose the k nodes in other $n-1$ nodes. So the degree distribution p_k is given by

$$p_k = C_{n-1}^k p^k (1-p)^{n-k-1} \tag{17}$$

From (15), we have

$$p = \frac{\bar{k}}{n-1} \tag{18}$$

If the size of the network n is large enough to make p extremely small, then we can use the first order Taylor Series Expansion on the $\ln(1-p)^{n-k-1}$ term. It means

$$\ln(1-p)^{n-k-1} = (n-k-1)\ln(1-\frac{\bar{k}}{n-1}) \simeq -(n-k-1)\frac{\bar{k}}{n-1} \simeq -\bar{k} \tag{19}$$

$$(1-p)^{n-k-1} \simeq e^{-\bar{k}} \tag{20}$$

And once again, for large n,

$$C_{n-1}^k = \frac{(n-1)!}{(n-1-k)!k!} = \frac{(n-1)^k}{k!} \tag{21}$$

Combine Equation(18)(20)(21), Equation(17) can be written as

$$p_k = C_{n-1}^k p^k (1-p)^{n-k-1} = \frac{(n-1)!}{(n-1-k)!k!} p^k e^{-\bar{k}} = \frac{(n-1)^k}{k!}\left(\frac{\bar{k}}{n-1}\right)^k e^{-\bar{k}}$$

$$= \frac{\bar{k}^k}{k!} e^{-\bar{k}} \tag{22}$$

As we can see, this is actually a Poisson distribution.

2.2.3 shortest path and the largest component

Here we only show the estimation about the average shortest path in a random network[2]. First, the size of the largest component in the random network can be computed exactly as the network size $n \to \infty$. Assume u is the fraction of nodes in the random network that do not belong to the largest component. We can view u as the probability a node i not being connected to the largest component via any other $n-1$ nodes. There are two conditions for the node j in the other $n-1$ nodes. The first is node i does not connect to node j, and the second is node i connects to node j, but j does not belong to the largest component. For node i and node j, the probability can be written as $1-p+pu$, But now we consider node j is chosen from the other $n-1$ nodes, so now we have

$$u=(1-p+pu)^{n-1} \tag{23}$$

By using Eq (18)

$$u=\left[1-\frac{\bar{k}}{n-1}(1-u)\right]^{n-1} \tag{24}$$

Take logs of both sides and use the first order Taylor Series Expansion for large n

$$\ln u \simeq -(n-1)\frac{\bar{k}}{n-1}(1-u) = -\bar{k}(1-u) \tag{25}$$

And then

$$u=e^{-\bar{k}(1-u} \tag{26}$$

let L be the fraction of the nodes in the largest component, that is

$L=1-u$,

$$L=1-e^{-\bar{k}L} \tag{27}$$

we take differential of both sides

$$1 = \bar{k}e^{-\bar{k}L} \tag{28}$$

As we can see, equation (28) implies that equation (27) has a solution $L \neq 0$ if $\bar{k} > 1$.

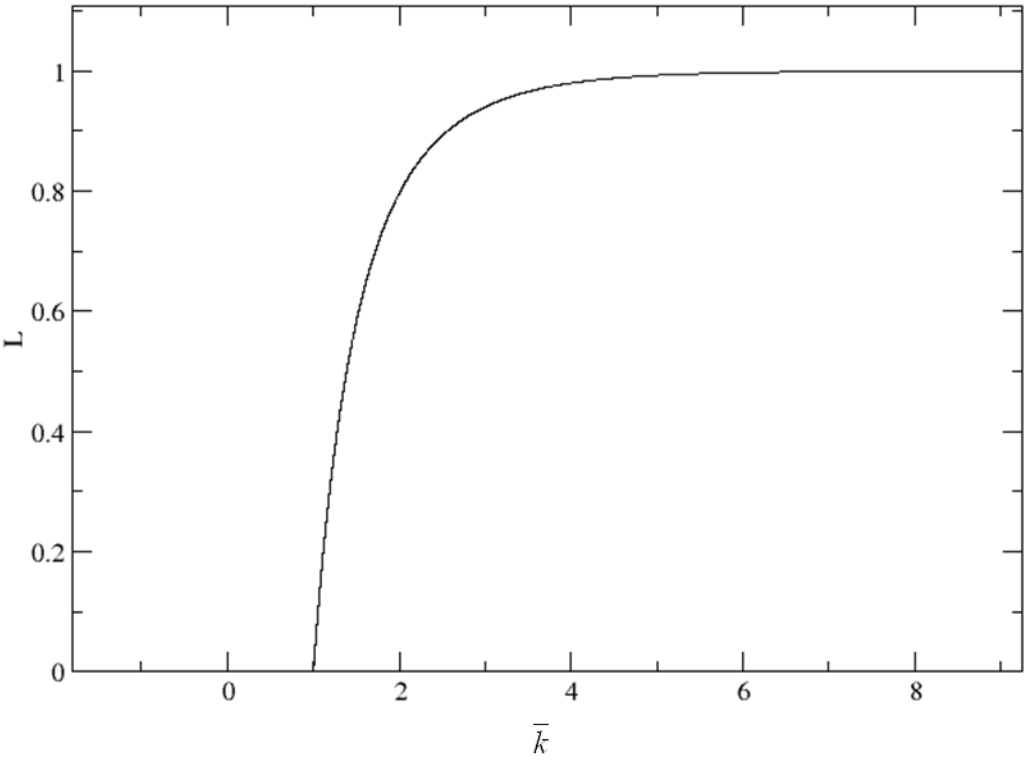

Figure 5 : the fraction of the largest component in a random network

Next, imagine the search from a node in a random network like we illustrated in Figure 6. If there are u nodes in the *nth level* , then the number of nodes outside the *nth* level connected by a node in the *nth* level in a random network *G(n, p) is*

$$p(n-u) = \bar{k}\frac{n-u}{n-1} \simeq \bar{k} \tag{29}$$

As $n \to \infty$, the approximation will be valid. And it implies the number of nodes in the next level will be $\bar{k}$ times than in level which the search is on. That is to say the number of nodes s steps away from a randomly chosen node in the random network is $\bar{k}^s$. We can use $\bar{k}^s \simeq n$ to estimate the average shortest path in the random network.

$$s \simeq \frac{\ln n}{\ln \bar{k}} \qquad (30)$$

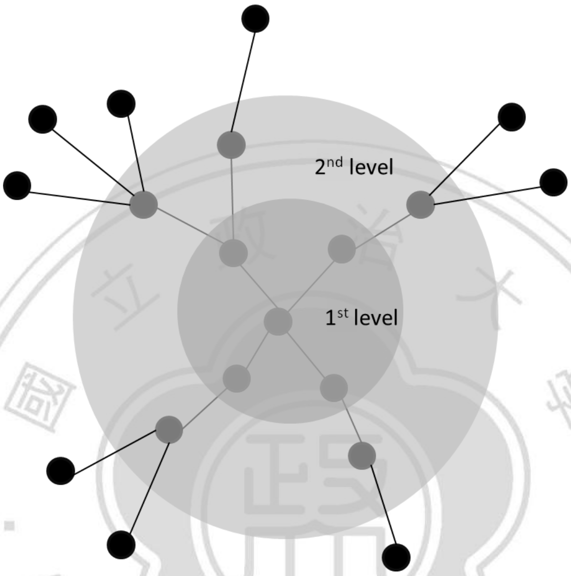

Figure 6 : a search in a random network from a source.

2.2.4 Implementation of random network : shuffle method and configuration model

✧ *Shuffle method*

Assume the node number of the random network is n and the edge number is m. Hence we know there are $n(n-1)/2$ locations which we can choose to put the edges in. Next, we choose an empty location between a pair in the network and put an edge between the pair until there are m edges in this network. This is what we call the shuffle method.

✧ *Configuration model*

In [9], the authors proposed a method called “configuration model” to creat networks having any kind of degree distribution. There are n nodes in a network, and we want the “degree sequence ”(i.e. the degrees of all the nodes) obey the degree distribution p_k we desire. All we need to do is very simple. There are N nodes and no edges in the network in a network. Next, we put stubs on all of the nodes like Figure 7. The numbers of each node are the same with the degree sequence (generated from the degree distribution we want). Then we choose a pair of the stubs randomly and connect them until all the stubs are chosen. The network we want will be formed. Of course, the sum of the stub numbers must be even.

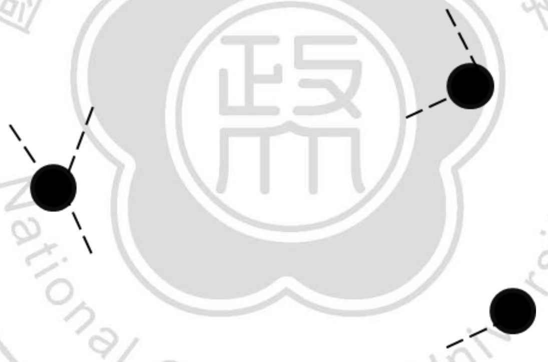

Figure 7 : “stubs” and configuration model

2.3 Important properties in the real networks

In the section 2.2, we list three properties in the random networks. But the real networks have some difference properties comparing with the random network.

2.3.1 *high clustering , small world, the small world quotient (Q)*

Recall that in the section 2.2.1, the global clustering coefficient in a random network can be computed as $C_g^{rand} = \bar{k}/(n-1)$. If we use this formula to calculate the global clustering coefficient in real network data, the values will be far smaller than the values via using Eq (11). The history of the small world problem can be traced back to sixty years ago. In the 1950s, Kochen and de Sola Pool wrote a paper, which was eventually published in 1978, which tackled what is known today as the "small world" problem[10]: If two persons are chosen randomly from a population, what are the probability that they would know each other. In other words, how long a chain of acquaintances would be required to connect them? In 1967, Stanley Milgram designed an experiment to answer the problem[3]. In brief, Milgram sent packages to 196 people in Omaha, Nebraska, with a request that they forward them to the intended recipient, but he provided no address. Someone receiving the package were asked to send the packages to friends, acquaintances who they felt might send the packages to the right destination, until the package is received by the intended recipient. The result revealed that about five and a half intermediaries were needed for the packages received by the intended recipient. This told us the "small world" property in the real networks. Actually, most of the real networks show this two properties. In [4] the authors list the average shortest paths and the clustering coefficients of the empirical data(Table 1).

Table 1 : the average shortest path lengths and the clustering coefficients of three empirical data.

	Sactual	Srandom	Cactual	Crandom
Films actors	*3.65*	*2.99*	*0.79*	*0.00027*
power grids	*18.7*	*12.4*	*0.08*	*0.005*
C.elegans	*2.65*	*2.25*	*0.28*	*0.05*

If a network has this two kind properties, then it is called the small work network. There is a simple index called "small world quotient"[11] to quantize these phenomena.

$$Q = \frac{C(p)}{C_{random}} \Big/ \frac{L(p)}{L_{random}}$$

The higher the small quotient is, the more small world properties it has.

2.3.2 *Power-law degree distribution*

From the data published by Taiwan government, we can know the average male height of the ages from 19 to 30 years old during the years from 2005 to 2008 was 172.4 cm. in this example, we can say the average height is the typical value which individual measurements are centered. It means there is some variation around the value, but we can't find any person who is 1000 cm or 10 cm. But not all values we measure are peaked around a typical value, for example, a power law distribution.

A power law distribution has the form

$$p(x) = Cx^{-\gamma} \tag{31}$$

The constant γ is called the exponent of the power law. we make a data to obey the power law distribution and plot the distribution in Fig. 2-8. For the data, $C = 1$, $\gamma = 2$, and we plot the numbers in a normal histogram. But if we want to reveal the power law distribution, it is better to plot the numbers in a log-log scale like we did in the right of Fig 2-8. As we can see, the slope of the straight line is $-\gamma$.

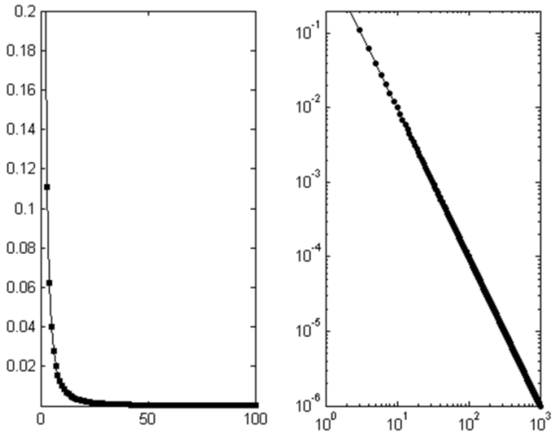

Figure 8 : A power law distribution is plotted in two different scales.

In 2.2.2, the degree distribution p_k in a Erdős–Rényi model is actually a Poisson distribution, but in the real complex network data, the researchers found most of them having power-law degree distribution with different γ values. In[12], the authors list the scale-free exponent γ of three networks (Table 2), and draw them(Figure 9).

Table 2 : the scale-free properties of three different network

	N	$\bar{k}$	$-\gamma$
movie actors	*21250*	*28.78*	*2.3*
internet(www)	*325729*	*5.46*	*2.1*

power grid	*4941*	*2.67*	*4*

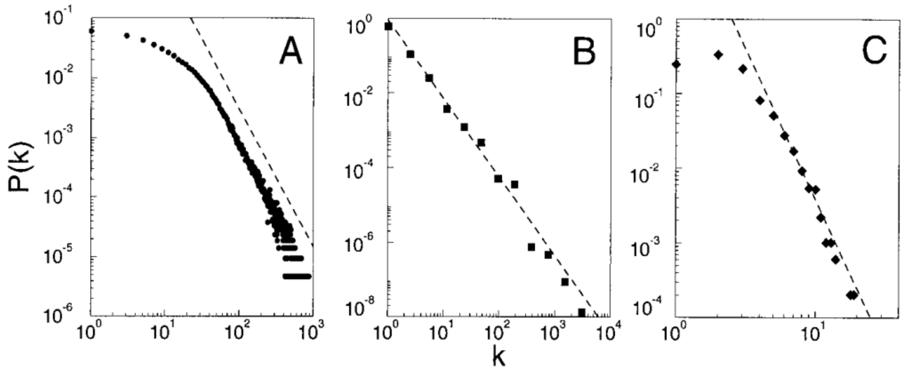

Figure 9 : degree distributions for(A) the movie actor network. (B) the internet(www) network. (C) the power grid network

The scale-free property is common in complex research but not universal[13]. For example, the electric power grid of Southern California and the airport network in the world have a form called "exponential decay" which decays faster than the power-law distribution.

2.4 The models

2.4.1 Watts and Strogatz small world model

In 1998, Duncan J. Watts and Steven H. Strogatz devised a method for converting a fully connected ordered network into a fully connected random network[4]. This method is called random rewiring. The researchers began with the simplest ordered network structure: a lattice ring. There are *n* nodes in the lattice ring, and each of them connects with its *k* nearest neighbor by undirected and unweight edged. In a clockwise sense, they choose a node and the edge connecting it to its nearest neighbor. With probability *p*, they reconnect this edge to a node chosen randomly over the ring, consider each node in turn until one lap is completed. Next, they consider the edges connecting nodes to their second-nearest neighbors. clockwise, and randomly rewire each of the edges with probability p. they continue this process proceeding outward to more distant neighbors after each lap, until every edge in the original lattice ring has been concerned once. Figure 10 shows how the graph changes with different rewiring probability. For $p = 0$, it is obvious the network is regular, and for $p = 1$ the network is a random network.

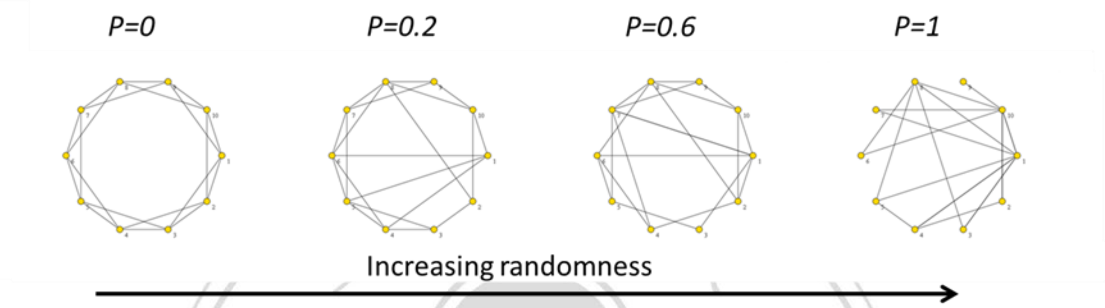

Figure 10 : Random rewiring procedure for interpolating between a regular ring lattice and a random network

In Figure 11, we show the random rewiring procedure how to convert a ring of nodes from an ordered network into a random one. For a small variation of the random rewiring probability p., the value of the shortest path length falls rapidly to one typical value of a random network while the network still retains a high clustering coefficient. In this range of p, the network has small shortest path length and large clustering coefficient, and the coexistence of these two properties is what makes a network a small world.

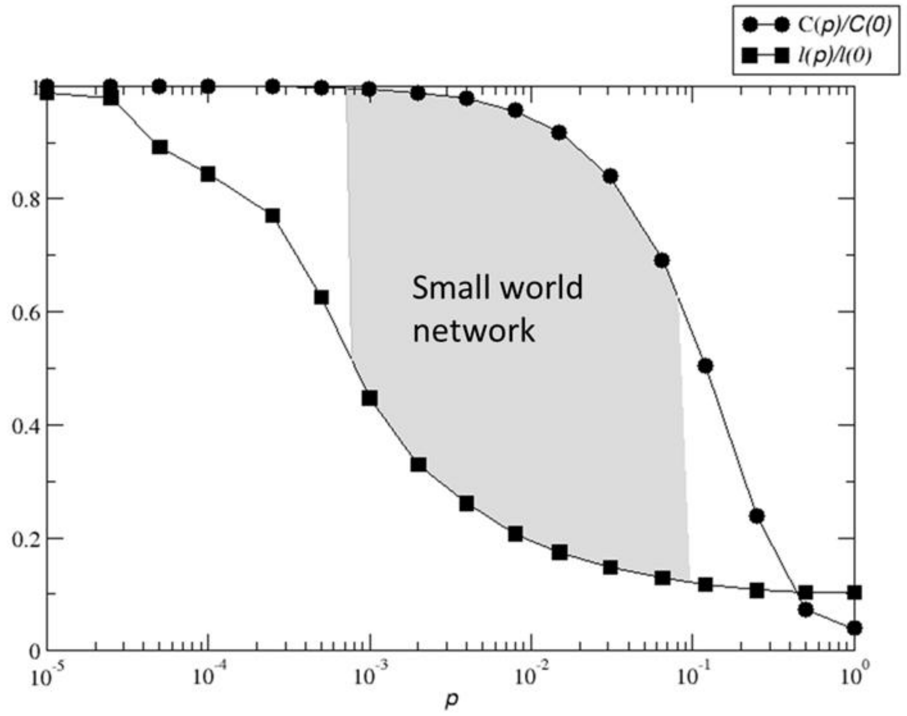

Figure 11 : the W-S small world model

In , we plot the degree distribution of the Watts-Strogatz small world model to see if the distribution obeys the power law. In the first, the random rewire probability $p = 0$, and all the degree is a fixed value 10. When p is getting higher, p is still peaked around $k = 10$ but it gets broader. Via the observation, we can realize the Watts-Strogatz model doesn’t reveal the power-law degree distribution existing in many real network data(Figure 12).

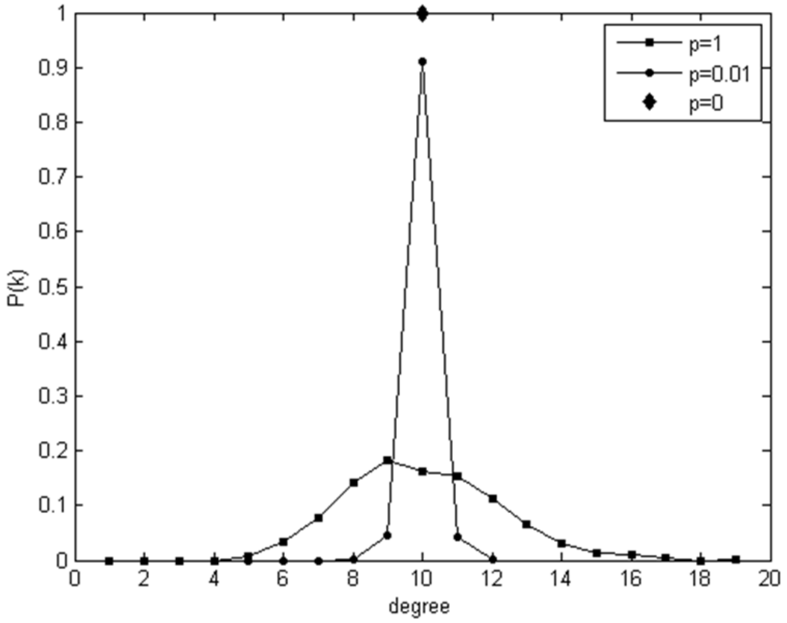

Figure 12 :degree distribution in the Watts-Strogatz small world model

2.4.2 Barabási–Albert model(BA model)

In 1999, Albert-László Barabási and Réka Albert published their famous paper “Emergence of Scaling in Random Networks.”. In this paper, they declared there are two generic mechanisms would make the power-law degree distribution happen. These two generic mechanisms are written as follow Networks expand continuously via the addition of the new nodes and new nodes connect preferentially to the existing nodes have higher degree.

In this model, there are m_0 nodes at the beginning. At every time step, a new node is added into the network. The new node connects m existing nodes in the network. And the probability of the new node connecting the old node i is

$$\prod(k_i) = \frac{k_i}{\sum_j k_j} \tag{32}$$

Where $\sum_j k_j$ is the sum of all the nodes degree at that time step.

Furthermore, consider a node i after t time steps with a degree $k_i(t)$ when the total number of edges is mt. When a new node is added to the network, the probability that it is joined to i is $\frac{mk_i}{\sum_j k_j} = \frac{mk_i}{2mt}$, by taking expectations, the equation transform into

$$k_i(t+1) - k_i(t) = \frac{mk_i}{2mt} = \frac{k_i}{2t}$$

then

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t}$$

By using the initial condition, as $t = t_i, k_i(t_i) = m$, t_i means the time when the node i is added to the network . We get the solution

$$k_i(t) = m(\frac{t}{t_i})^{0.5}$$

From the equation we can realize the degree of the older nodes with a smaller t_i grow faster than the degree of young nodes. Take one step ahead, the probability that a node i has a degree smaller than k,

Using the equation above, it become

$$P[k_i(t) < k] \; = P\left[t_i > \frac{m^2 t}{k^2}\right] = 1 - P\left(t_v \leq \frac{mt^2}{k^2}\right)$$

$$= 1 - \frac{m^2 t}{k^2(t + m_0)}$$

The last equality assumes that nodes are added at equal time intervals to the network, so the density $(t_v) = {}^{1}\!/_{(m_0 + t)}$.

Recall that

$$P(k) = \frac{\partial P[k_i(t) < k]}{\partial k} \tag{33}$$

Which leads to the solution

$$P(k) = \frac{2m^2 t}{(t + m_0)} k^{-3}$$

Over long time duration it becomes a stationary solution

$$P(k) = \frac{2m^2}{k^3} \tag{34}$$

This is obviously a power law distribution with scale-free exponent $\gamma = 3$. Furthermore, the authors of [12] indicate that the scale-free degree distribution will be invalid for any absence of Growth of the network and preferential attachment. They use two models described below to show this.

Model A keeps the growing character of the network, but preferential attachment is eliminated by assuming a new node is connected with equal probability to any node in the network. That is to say

$$\prod(k_i)=\frac{1}{m_0+t-1} \tag{35}$$

The authors of [12] find this model will generate an exponential degree distribution. We realize model A and plot the degree distribution for different $m \text{ and } m_0$ values in Figure 13.

Model B assumes there are N nodes and no edges in the network at first. At each time step, they randomly choose a node and connect it to another node in the network with probability

$$\prod(k_i)=\frac{k_i}{\sum_j k_j} \tag{36}$$

They find that the degree distribution $p(k)$ is not stationary. We realize model B and plot the degree distribution of the network for different time steps in Figure 14 and Figure 15.

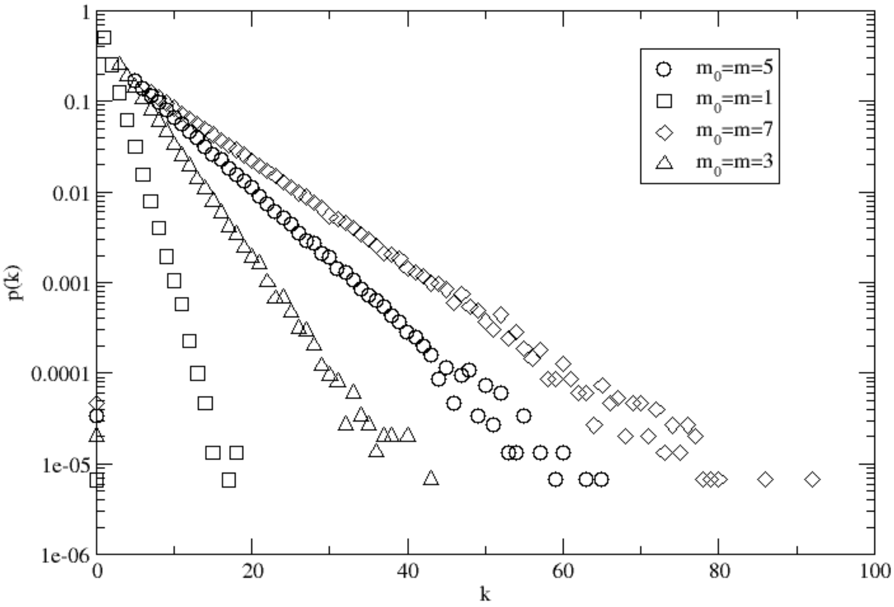

Figure 13 : the degree distribution of model A at time step = 150000.

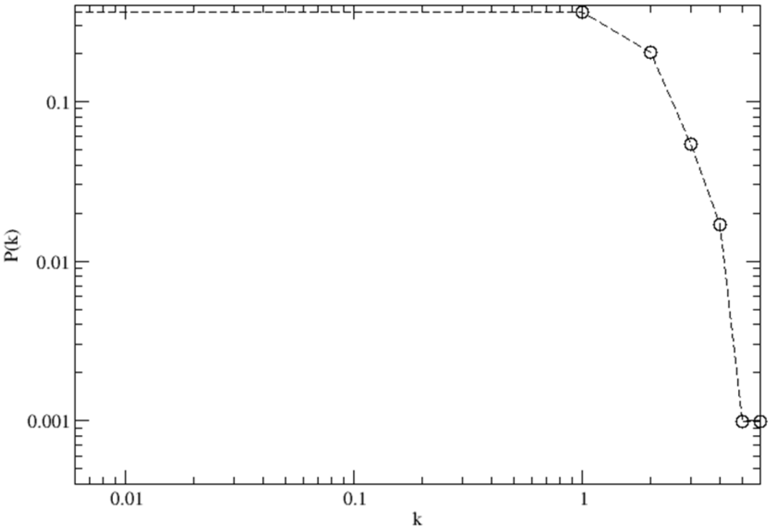

Figure 14 : the degree distribution of model B. N=1000. edge number = 1000

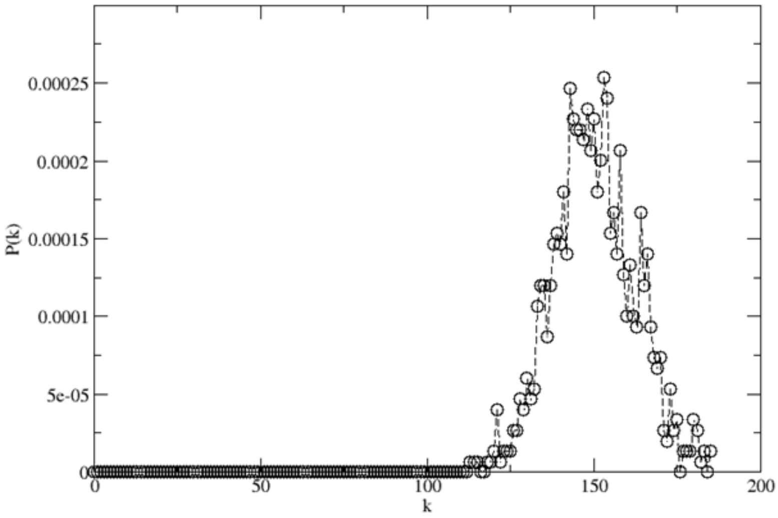

Figure 15 : the degree distribution of model B. N=1000. edge number = 150000

2.4.3 Extended BA model (EBA model)

In 1999, Réka Albert and Albert-László Barabási presented a model[14] to adjust the shortcomings of the BA model they brought up earlier. There are more parameters to describe the evolution of a complex network generation. In the paper, the authors called it a extended model. That's why the following researchers call it the extended BA model. Recall the BA model mechanism, in every time step, the new node is added to the network, and connects to m existed nodes. In extended BA model, there are three different possible situations during each time step. In EBA model, there are m_0 isolated nodes in the initial condition. At each time step, one of these operations below could happen:

✧ *Addition of new edges:* With probability p, m edges are added to the network. The one end of the new edge is selected randomly, and the other end is selected with probability

$$\Pi\left(k_i\right)=\frac{k_i+1}{\sum_j k_j+1} \tag{37}$$

reflecting the fact that new edge preferentially point to popular nodes .

✧ *rewiring of old edges* : with probability q, m edges are rewired. The edges are selected randomly by a node i and a edge l_{ij} connecting to the node are selected randomly. Next the edges are removed, and a new edge $l_{ij'}$ replacing the old edge. The new edge connect the node i and node j'. The node j' is chosen with probability $\Pi(k_{i'})$.

✧ *Addition of new nodes :* With probability $1-p-q$, a new node is added to the network. The new node connects to m nodes already present in the network. The node i which the new node connect is selected with probability $\Pi(k_i)$.

To explain the generic mechanisms of the network evolution, the undirected edges are used which means the edges are non-directional. In the model the probabilities p can be varied in the interval $0 \le p < 1$, and q can be varied in the interval $0 \le q < 1-p$,since $0 \le (p+q) < 1$. By using the continuum theory and assuming k_i changes continuously. The degree distribution has a power-law form

$$P\left(k\right)\propto[k+\alpha(p,q,m)]^{-\gamma(p,q,m)} \tag{38}$$

where

$$\alpha\left(p,q,m\right)=A\left(p,q,m\right)+1 \tag{39}$$

$$\gamma\left(p,q,m\right)=B\left(p,q,m\right)+1 \tag{40}$$

$$A(p,q,m)=(p-q)(\frac{2m(1-q)}{1-p-q}+1) \tag{41}$$

$$B(p,q,m)=\frac{2m(1-q)+1-p-q}{m} \qquad (42)$$

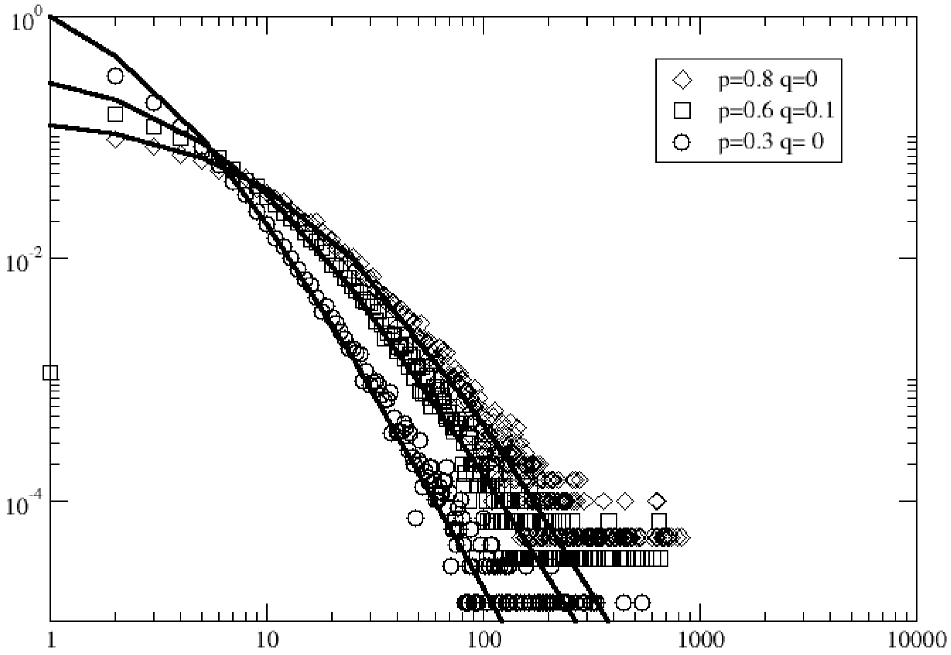

Figure 16 : the three different degree distributions generated by different parameter values in EBA model, time step = 100000.

In Figure 16, we plot the different degree distribution with different parameter.

2.5 Case study I: CTTC competition network and model simulations

2.5.1 Materials

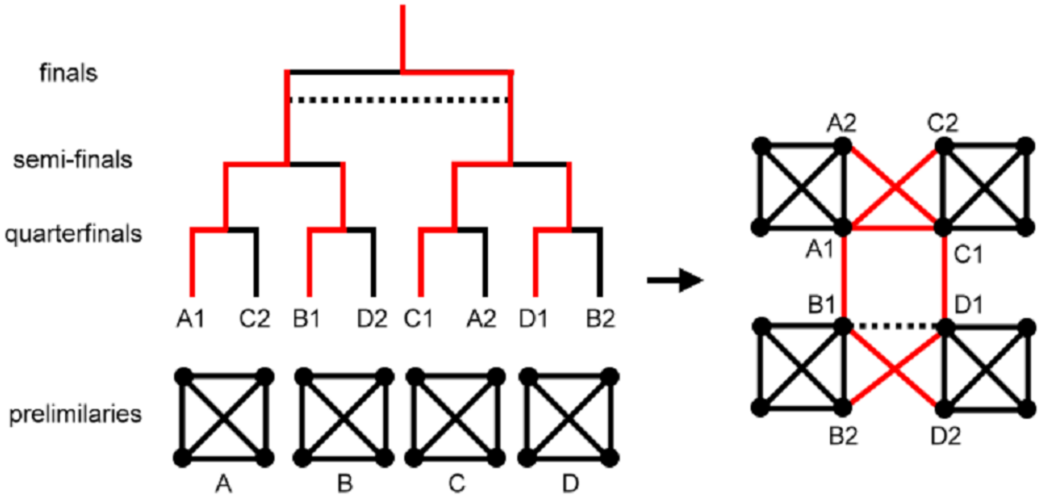

Figure 17 : the CTTC tournament design

The datasets are from CTTC (http://www.cybertabletennis.com/portal/) in Taiwan. CTTC has tried to promote table-tennis rating games for amateur players in Taiwan over the last decade. In the left, the game is divided into the round-robin and single elimination stages. In the right, it is shown that the network is formed by the competition design. The competition games in each tournament include round-robin and single elimination (i.e., knockout) stages. As we mentioned above, there are 8 players in the first round of the single elimination stage. Only the winners can enter the next round. The left part in Figure 17 illustrates the tournament design including round robin, quarterfinals, semi-finals, and final. Note that the number of groups and the number of players in a group may be different for different tournaments. Owing to the justice, directors will keep the same number of participants in a preliminary group.

2.5.2 Analysis of real data

In 2.5, the clustering coefficient will be remark as CC to substitute for C . In Table 3 and Table 4 , we illustrate the scale-free and small world properties for the CTCC table tennis competition networks every year. We can see the node number of the largest component node number (LC size), scale-free exponent γ for the degrees of the nodes are at least larger than 5. We plot the degree distribution in Figure 18, we can see the degree distribution obeys power-law distribution. The local and global clustering coefficients, average path length are also computed and listed. Furthermore, we compute the small world quotient (Q). It is obvious to find that both local and global small world quotients are much larger than 1 for every year. Thus competition networks display well-optimized efficient structures between table-tennis players. In addition, local Q is larger than global Q because of CC_{global} with a lower value compared to that of CC_{global}, and PL is quite similar for both cases of competition and random networks. The higher transitivity for competition networks rather than that of random networks is due to the preliminaries based upon the round-robin match. All players are fully connected cliques in the same group (Figure 17), which will cause local and global clustering coefficients with much higher values. Concerning on the previous researches about movie actor collaboration network [15], $\mathrm{CC}_{\mathrm{local}}$, PL, and γ is 0.79, 3.65, and 2.3 ± 0.1, respectively. Thus we may say that movie actor collaboration network exhibits a more well-optimized efficient structure compared to that of the CTTC network.

Table 3 : the network sizes and the power-law exponent of the CTTC competition networks

Year	*Network Size*	γ	$\gamma_{k_i \geq 5}$
2004	*3166*	*1.96*	*2.19*
2005	*5901*	*1.85*	*2.00*
2006	*7622*	*1.77*	*1.90*
2007	*9193*	*1.78*	*1.90*
2008	*9001*	*1.75*	*1.87*
2009	*8650*	*1.68*	*1.77*
2010	*8802*	*1.74*	*1.88*
2011	*4979*	*1.79*	*2.08*

Table 4 : CTTC network properties analysis

Year	LC size	Clustering Coefficient (CC)			Path Length (L)		Small World Q	
		Local	Global	Random*	Actual	Random*	Local	Global
2004	3166	0.2394	0.1297	0.0017	5.55	5.00	126.4	68.4
2005	5901	0.2387	0.1154	0.0013	4.83	4.48	169.2	82.2
2006	7622	0.2598	0.1105	0.0012	4.69	4.24	190.8	80.6
2007	9193	0.2727	0.1123	0.0012	4.42	4.09	215.7	89.2
2008	9001	0.3454	0.1104	0.0013	4.29	3.91	235.6	75.3
2009	8650	0.3757	0.1190	0.0016	4.11	3.75	216.8	68.6
2010	8802	0.3472	0.1149	0.0015	4.23	3.78	204.8	68.1
2011	4979	0.4176	0.1484	0.0022	4.39	3.79	162.2	57.4

* The average of one hundred simulations

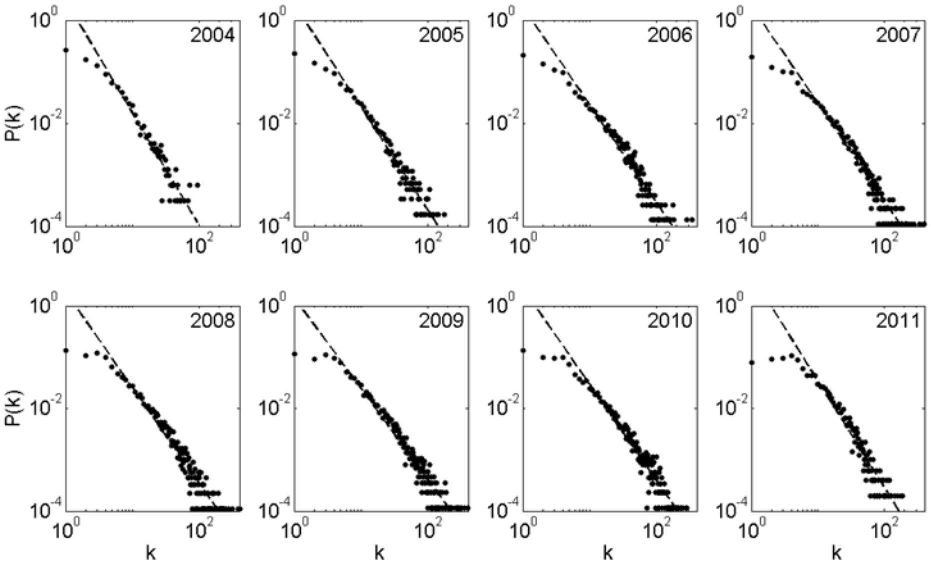

Figure 18 : CTTC networks degree distribution for every year

The scatter diagrams for CC_i versus k_i from 2004 to 2011 are illustrated in Figure 19. In general, a smaller (larger) k_i is corresponding to a higher (lower) CC_i. The smallest k_i with $CC_i = 1$ represents those players were losers who never played single elimination games. In contrary, excellent players should have much more connections with other players. Thus a lower CC_i for excellent players can be expected. According to the results shown in Fig. 3, we plot the histogram of CC_i for competition networks in Figure 21 and the expected U-shape-like spectra are observed from 2004 to 2011.

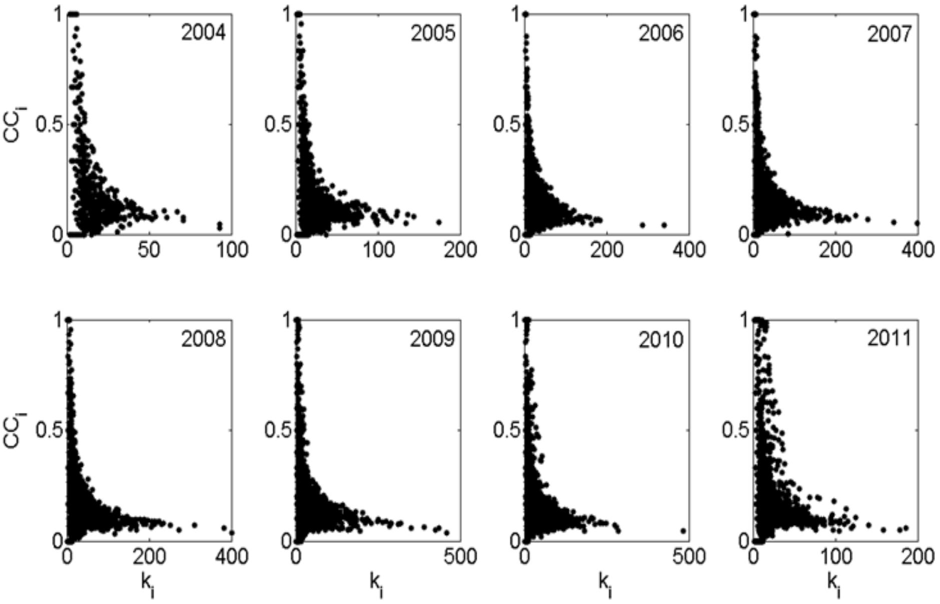

Figure 19 : The scatter diagram between k_i and CC_i from 2004 to 2011

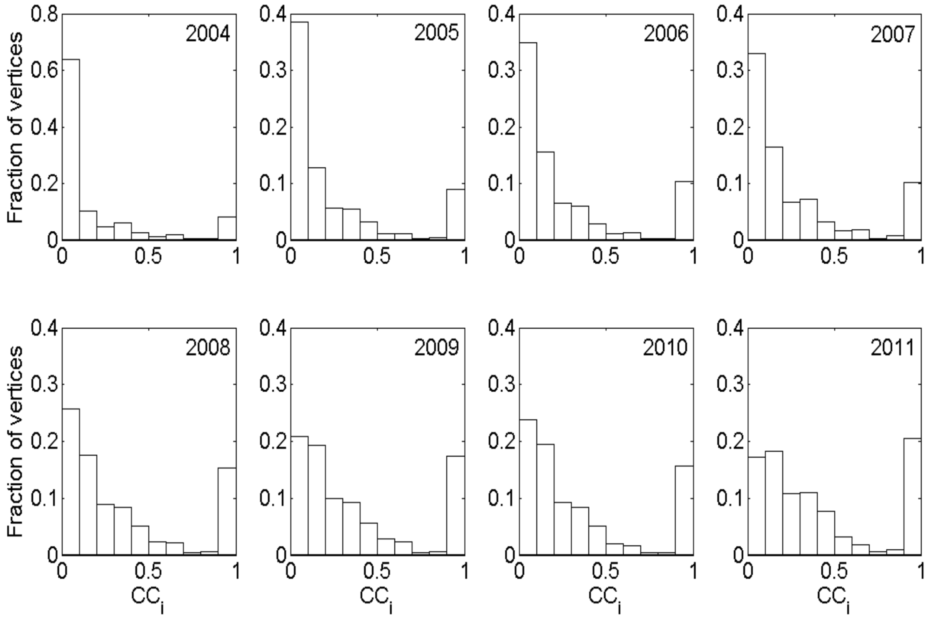

Figure 20 : The histogram for the CCi from 2004 to 2011

2.5.3 BA model simulation

In this section, we use the mechanism of the BA model to simulate the CTTC competition networks every year. That is to say, we create networks having the same sizes with the largest component of the real data, and we use different m values to make the average degree $\bar{k}$ of these networks approximate the real data. Figure 21 illustrates the degree distributions for the BA model simulation networks, and it is obvious to find power law degree distribution from double-logarithm plots.

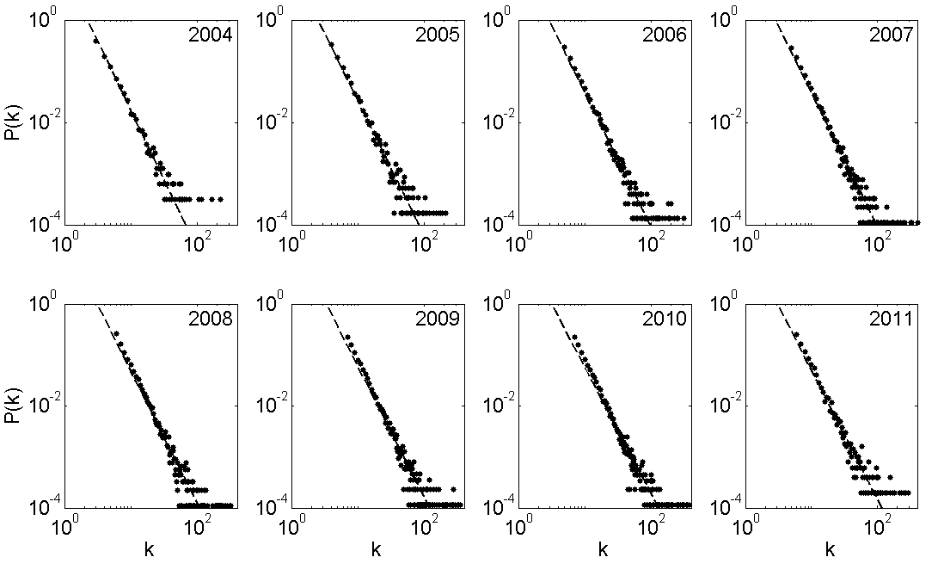

Figure 21 : degree distribution of BA model simulations

In Table 5, we list the simulation results and the small world quotient (Q). The average values of γ are about 2.63 ± 0.09 which is larger than that of the competition network (i.e. $\gamma_{k_i} = 1.98 \pm 0.2$). Both local/global Q are larger than 1 for every year. However, all the small world quotients are much smaller than the real datasets. In addition, local Q is larger than global Q because of $\mathrm{CC}_{\mathrm{global}}$ with a lower value compared to that of $\mathrm{CC}_{\mathrm{local}}$ and average path length is quite similar for both cases of BA model simulations and random networks. Based upon results shown in Table 4 and Table 5 , we may conclude that BA networks cannot well explain the results of table-tennis competition networks. The small world Q in BA networks is much lower than that of the competition networks. The main reason is resulted from local and global clustering coefficients with much lower values in BA networks.

Table 5 : BA model simulation result

Year	l size	$\langle k \rangle$	γ_{fit}	Clustering Coefficient			Path Length		Small World Q	
				Local	Global	Random*	BA model	Random*	Local	Global
2004	3166	6	2.72	0.010	0.0038	0.0020	3.89	4.60	5.91	2.25
2005	5901	8	2.63	0.009	0.0061	0.0017	3.74	4.41	6.24	4.23
2006	7622	10	2.62	0.008	0.0061	0.0013	3.59	4.13	7.08	5.40
2007	9193	10	2.61	0.007	0.0054	0.0011	3.64	4.22	7.38	5.69
2008	9001	12	2.59	0.008	0.0064	0.0014	3.47	3.92	6.46	5.16
2009	8650	14	2.57	0.009	0.0078	0.0017	3.31	3.72	5.95	5.16
2010	8802	14	2.48	0.009	0.0076	0.0016	3.32	3.72	6.30	5.32
2011	4979	12	2.54	0.013	0.0100	0.0027	3.30	3.69	5.38	4.14

* The ensemble average of one hundred simulations

Figure 22 shows the scatter diagrams for CC_i versus k_i. The maximum CC_i obtained from 2004 to 2011 of BA model simulations is below 0.4 which indicates no fully connected cliques can be observed in BA networks.

Figure 23 illustrates the histograms of CC_i for BA networks, and L-shape-like spectra are observed for every year.

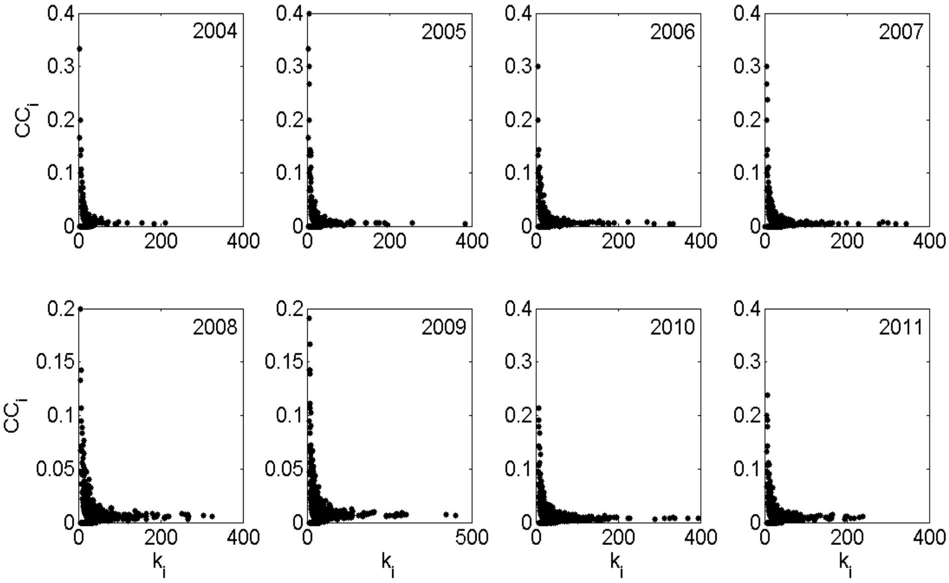

Figure 22: The BA model simulation scatter diagram between ki and CCi

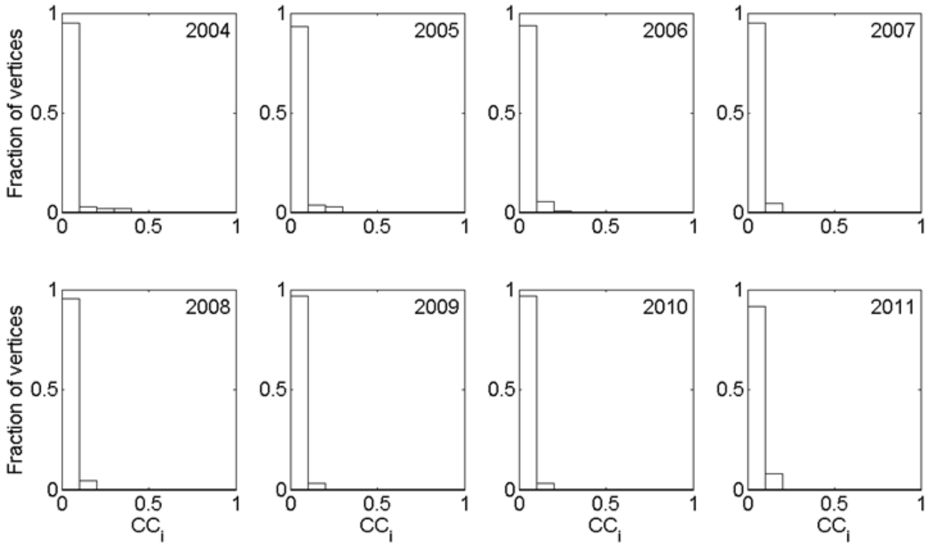

Figure 23 : The histogram for the CCi from 2004 to 2011 in BA model simulation

2.5.4 extend BA model simulation

In the extend BA model simulations for the CTTC networks, we tune the parameters to make the power law exponent γ approximate the values in CTTC networks.

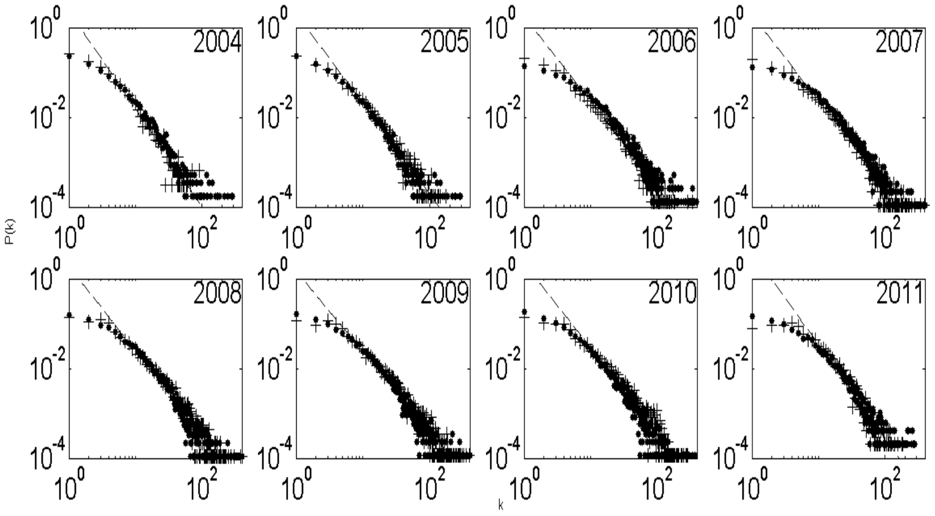

Figure 24 : extend BA model simulation power law

Table 6 : the extend BA model simulation results

year	*size*	$\overline{k}_{real}$	$m_0 = m$	p	q	γ_{fit}	$\overline{k}$
2004	*3166*	*5.33*	*1*	0.22	0.53	*2.36*	*3.81*
2005	*5901*	*7.71*	*1*	0.32	0.56	*2.21*	*7.31*
2006	*7622*	*9.41*	*1*	0.36	0.58	*2.13*	*9.89*
2007	*9193*	*10.84*	*1*	0.36	0.58	*2.18*	*13.35*
2008	*9001*	*12.11*	*1*	0.33	0.60	*2.08*	*11.45*
2009	*8650*	*13.65*	*1*	0.29	0.64	*2.08*	*11.73*
2010	*8802*	*13.32*	*1*	0.32	0.60	*2.09*	*13.89*
2011	*4979*	*11.13*	*1*	0.42	0.50	*2.06*	*12.21*

In Table 7, the global and local clustering coefficients are higher than the values in BA model simulations, and so are the small world quotients. But they are still much smaller than the values of CTTC networks.

Table 7 : the small world properties of the EBA simulation

year	Clustering Coefficient (C)			Path Length (L)		Small World Q	
	Local	Global	Random*	Actual	Random*	Local	Global
2004	0.0180	0.0170	0.0020	4.42	4.60	9.37	8.85
2005	0.0300	0.0230	0.0017	3.84	4.41	20.27	15.54
2006	0.0410	0.0350	0.0013	3.60	4.13	36.18	30.89
2007	0.0410	0.0320	0.0011	3.59	4.22	43.81	34.20
2008	0.0310	0.0260	0.0014	3.54	3.92	24.52	20.56
2009	0.0330	0.0280	0.0017	3.46	3.72	20.87	17.71
2010	0.0260	0.0230	0.0016	3.55	3.72	17.03	15.06
2011	0.0450	0.0390	0.0027	3.49	3.69	17.62	15.27

In Figure 25, we see a major difference between EBA and BA mode simulations is that there are some nodes in EBA simulations having local clustering coefficients larger than 0.4 .

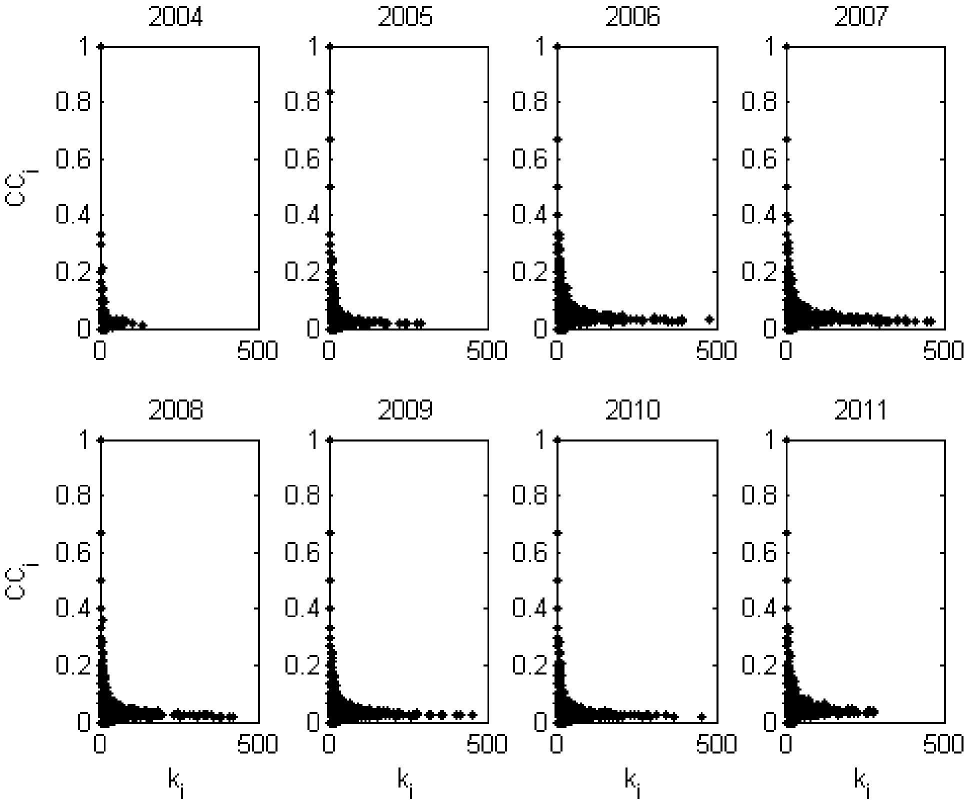

Figure 25 : the CC-k diagrams for the extend BA model simulations

The histograms of the local clustering coefficients for EBA model(Figure 26) simulations shows the L-shape-like spectra which is similar to what we observe in BA model simulation. The only difference is there are some fractions of nodes having higher local clustering coefficients than BA model simulations.

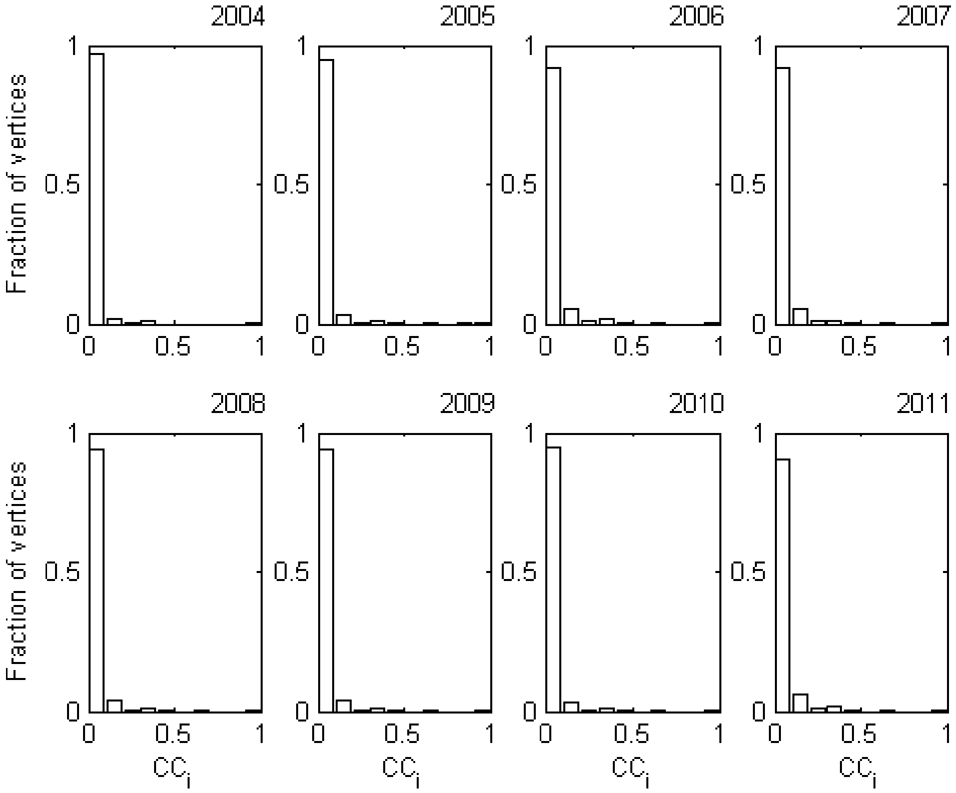

Figure 26 : The histograms of the clustering coefficients of the extend BA model simulations

2.5.5 discussion

In BA model simulations, the preferential attachment is a key mechanism to generate the scale-free degree distribution (Figure 21). However, in competition networks scaling crossover phenomena are observed in Figure 18. In the high k regime it represents players won more games, thus who had more opportunities to compete with more opponents. We call this phenomenon as “strong-get-stronger”. Compared to social competitive systems, “rich-get-richer” is a well-accepted result, and which indeed reflects the result of preferential attachment.

On the other hand, the EBA model simulations have higher clustering coefficients than the BA model simulations have, and the scale-fee exponent γ are closer to the value of CTTC networks than BA model simulations. We think that is because the additions of new edges are more contrast to the reality in CTTC networks. And the rewiring of old edges also contributes on the increase of the clustering coefficients.

However, in CTTC networks the preliminary is a special design for generating the group winner, which enforces players to join the round-robin match. Therefore, the concept of “strong-get-stronger” cannot be applied to the round-robin match. Owing to that, it can be expected that this tournament design will disturb the scale-free degree distribution in the low k regime. The critical degree (i.e., as k=5) shown in Figure 18 represents the maximum number of players in the round-robin match should be equal to 6. And this is why we consider $\gamma_{k_i \geq 5}$ is a more suitable parameter to explain the preferential attachment in CTTC networks. Moreover, the preliminary will make high transitivity (clustering) between players and the small world Q would be much larger than 1.

Round-robin and knockout matches constitute competition networks. In network topology round-robin matches will make players fully connected. Thus, there are small regular networks embedded in competition networks. Knockout matches are for finding excellent players with the consideration of shortening the competition time. Topological connections from knockout matches would be different from regular connections in round-robin matches. High clustering and low degree can be observed in round-robin matches. On the contrary, knockout matches will result in low clustering and high degree. In Figure 20, U-shape-like histograms well reflect these two designs for table-tennis competitions.

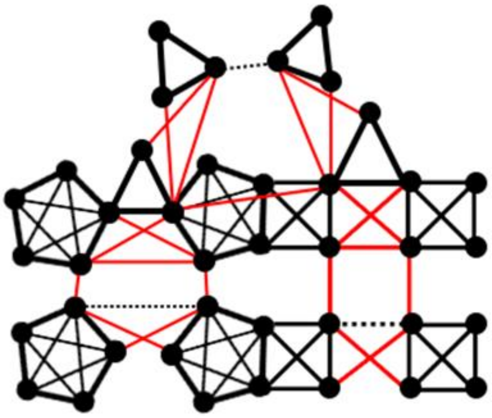

Figure 27 : the “overlap” of three tournaments

Figure 27 schematically illustrates the network topology including three tournaments based upon our datasets. The numbers of players in the round-robin match for these three tournaments are 3, 4, and 5, respectively. Thus small regular networks including triangle, square, and pentagon patterns can be observed. When group winner and runner up are generated, knockout matches will come up. Hence the dynamics of “strong-get-stronger” proceeds till to the final competition. Besides, the overlap players inmeans that these players have high interest in table-tennis competitions. The overlap players shown in these three tournaments are crucial for the network topology. Because of them, a giant/optimized network can be expected.

The network topology created by round-robin matches is very similar as that of the artist collaboration network[11]. However, C_{global} is, respectively, equal to 0.345 ± 0.029 and 0.120 ± 0.013 for the artist collaboration network and the competition network. The lower clustering coefficient in our case is mainly resulted from the dynamics of “strong-get-stronger” in knockout matches, which will reduce the clustering effect. It shall be noted that the overlap artists in the artist collaboration network reveals these artists are quite famous, thus the preferential collaboration as well as a giant/optimized network can be expected. In the point of view in topology, we may understand the overlap players in the competition network play the same role as the overlap artists in the artist collaboration network. Therefore, there are two different kinds of preferential attachment in human-made competition networks. One is the “strong-get-stronger” process to find excellent players in a tournament, which is named inner-tournament preferential attachment. The other is inter-tournament preferential attachment because of the connection of different tournaments through overlap players. Owing to that, the topology of table-tennis competitions is a network with different communities, where groups of nodes within the same community connections are dense, but between different communities connections are sparser. So far, many social and biological networks with community structure have been reported [16-18]). More recently, Li and Maini (2005) proposed an evolving network model with community structure based on the inner-community preferential attachment and inter-community preferential attachment mechanisms [19]. A scale-free distribution ($\gamma = 3$) in the whole k regime can be generated from their theoretical results as well as numerical simulations. Therefore, their community model still cannot well explain the topology of table-tennis competitions in Taiwan.

3 Vulnerability and cascading failure of a power grid

3.1 Efficiency

3.1.1 topology efficiency

The concept of "efficiency" of a network is a measure of how efficiently it exchanges information..[20] The efficiency ε_{ij} is the information exchanging from node *i* to node *j* in a network *G* can be defined to be inversely proportional to the shortest path distance, i.e $\varepsilon_{ij} = 1/d_{ij}$. When there is no path from node i to node j, $\varepsilon_{ij} = 0$, because of $d_{ij} = +\infty$. The efficiency of the network can be expressed by the average efficiency of the total pairs in the network, that is to say:

$$E(G)_{glob} = \frac{\sum_{i \neq j \in G} \varepsilon_{ij}}{N(N-1)} = \frac{\sum_{i \neq j \in G} {}^{1}\!/_{d_{ij}}}{N(N-1)} \tag{43}$$

The efficiency of a network is usually called the *global efficiency*. That is why we denote it by E_{glob}. Next, we consider how efficient of the information exchanging is between the neighbors of node *i* when *i* is removed. If a sub network G_i is formed by the neighbors of node *I*, then

$$E(G_i) = \frac{\sum_{n \neq m \in G_i} \varepsilon_{nm}}{k_i(k_i - 1)} \tag{44}$$

then, we can define E_{loc} as follow

$$E_{loc} = \frac{1}{N} \sum_{i \in G} E(G_i) \tag{45}$$

the local efficiency reveals how much the network *G* is fault tolerant, thus it shows how efficient the information exchanging is between the neighbors of i when i is removed. If we use Watts-Strogatz model to observe these quantities, then we will find the global and local efficiency change with the variation of the random rewire probability *p,* like Figure 28.

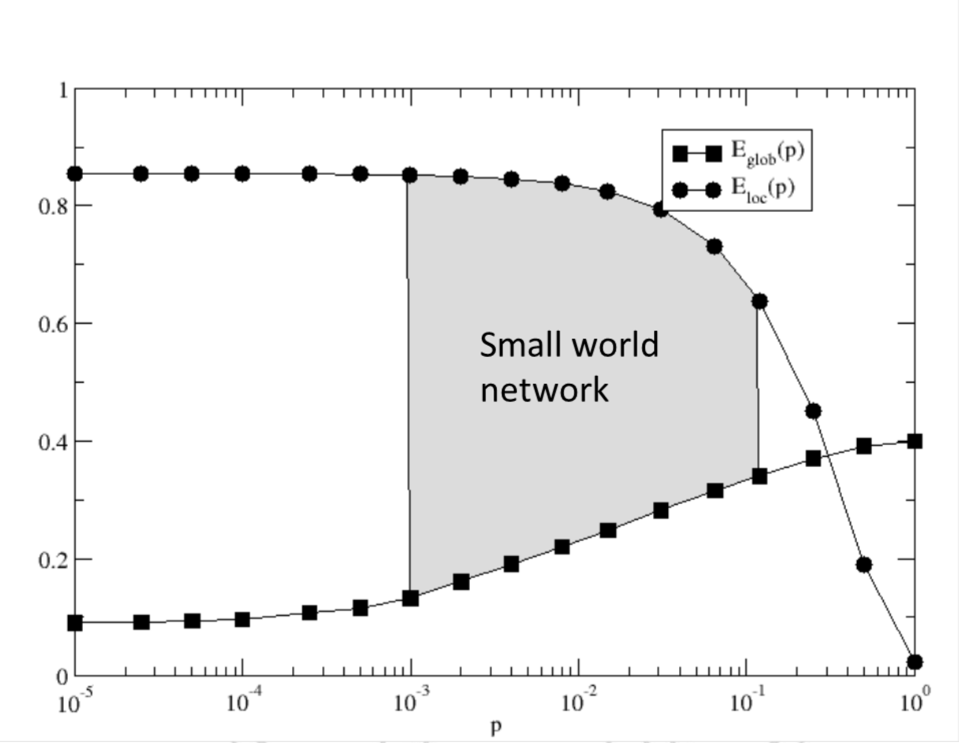

Figure 28: the small world from “efficiency” view

3.1.2 reliability efficiency

If we consider the transmission reliability between pairs of node i and j in a network structure, such as a power grid, then the shortest reliable path $\{d_{ij}\}$ can be computed as[21]

$$d_{ij} = \min_{\gamma_{ij}} \left(\frac{1}{\prod_{mn \in \gamma_{ij}} P_{mn}} \right) \qquad (46)$$

$\{P_{mn}\}$ is the transmission reliability between node i and j and $0 \le p_{ij} \le 1$. γ_{ij} is the path which connects nodes i and j. Note that $1 \le d_{ij} < \infty$, the lower value corresponding to the existence of a perfectly reliable path connecting i and j (no failure will occur in the links involving this path, that is to say, $p_{mn} = 1 \ \ \forall mn \in \gamma_{ij}$) and the upper value corresponding to the situation of no paths connecting i and j efficiency vulnerability, that is to say, there is at least failure in all connections from i to j. By Taking the shortest reliable path $\{d_{ij}\}$ into (43) and (45), we can calculate the local and global reliability efficiency of the network.

3.1.3 efficiency vulnerability

If we remove some edges or nodes in a network *G*, then the value of d_{ij} will change. So does the global efficiency $E(G)_{glob}$. In this manner, we can measure how great the removals impact the structure of the whole network *G* by using the value called vulnerability V^* [21].

$$\mathrm{V}^* = \frac{E_{glob}(G) - E_{glob}(G^*)}{E_{glob}(G)} \tag{47}$$

G^* is the new network resulting from the removals of some edges or nodes in the network *G*. For example, Figure 29 shows the 2006 CTCC amateur table-tennis players competition network, the network can be viewed as two subgroup connected by an edge from node 2307 to node 5872. If we remove an edge in turn and compute the V^* value for every new graph. The V^* value of the edge from node 2307 to node 5872 is the highest one in the 2006 CTCC amateur table-tennis players competition network. As we can see, the efficiency vulnerability is a useful index to find the most important (vulnerable) edge in a large network. In this case, we only calculate the topology efficiency.

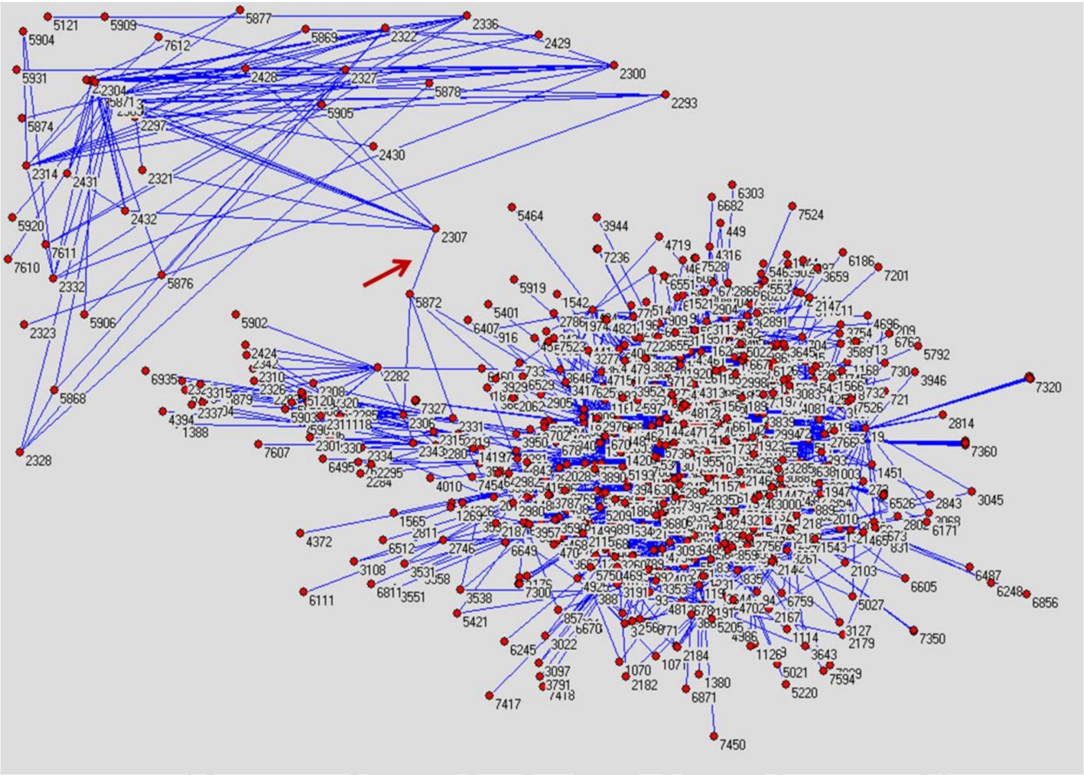

Figure 29 : V^{*} test for the 2006 CTCC amateur table-tennis players competition network.(This figure is draw by Pajek.)

3.2 Cascading failures

3.2.1 Betweenness overload

For a given network *G*, assume that at each time step, one unit of the relevant quantity, like energy, information etc., is exchanged between all pairs of nodes and transmitted along the shortest path connecting them. The load L_i of a node is the total number of shortest path passing through it, that is to say , the betweenness. And the capacity is defined as follow [6],

$$C_i = (1+\alpha)L_i \qquad \text{i = 1,2,......N} , \quad \alpha \geq 0 \qquad (48)$$

α is called the "tolerance parameter", which reflect the ability of a node to handle the additional increases of the load.

If we remove a node *j*, then the redistribution of currents will occur. If the new load of node *j*, $L_j^{'}$ becomes larger than the capacity of node j, this will make node *j* fail. The failure will lead to a new redistribution of the load in the network. As a result, subsequent failures can occur, the process is what we call the "cascading failure."[6]

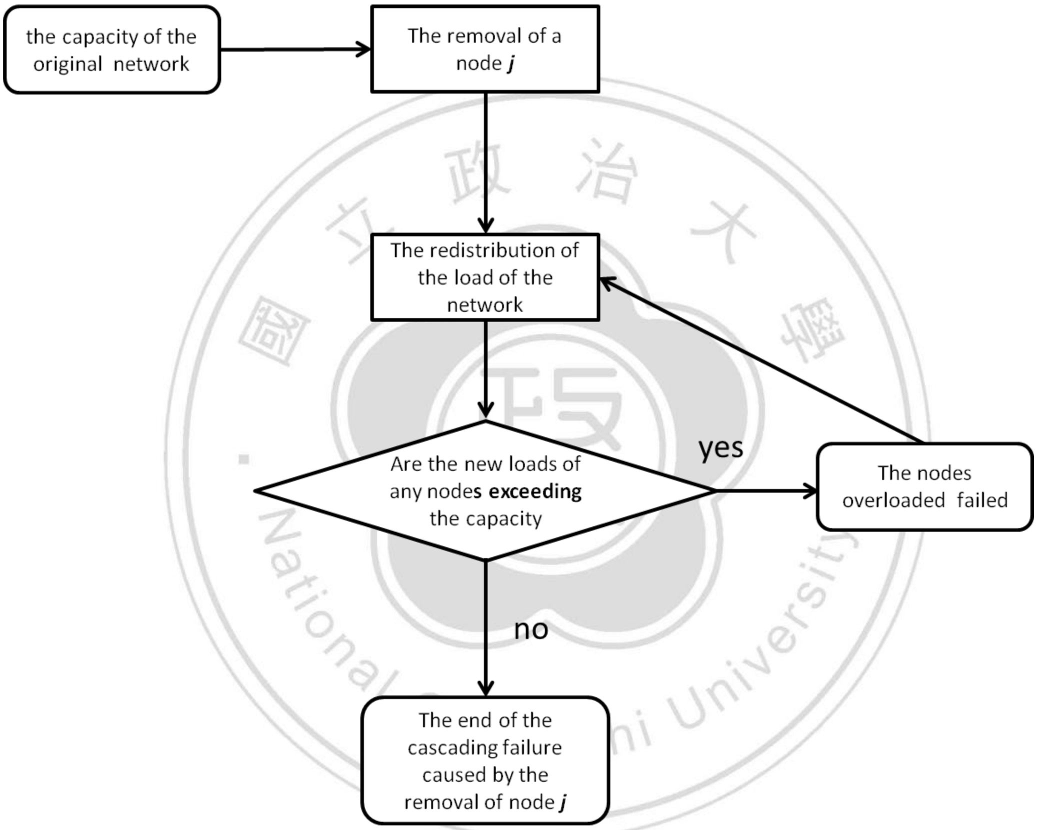
the capacity of the original network
The removal of a node *j*
The redistribution of the load of the network
Are the new loads of any nodes **exceeding** the capacity
yes
The nodes overloaded failed
no
The end of the cascading failure caused by the removal of node *j*

Figure 30 : the flow chart the cascading failure triggered by the betweenness-overload

There are three strategies of the removals of the nodes in [6]. They are the removal of nodes chosen at random, the removal among those with largest degrees (attack), and the removal among those with largest load (attack). We plot the different cascading failure of the random network (homogeneous network) in Figure 31, scale-free network in Figure 32, and the western U.S. power grid in Figure 33.

Among these three figures, each curve corresponds to the average over five triggers and ten realizations for the attack on the nodes having the largest betweenness (load) and degrees. For random removals of nodes, the curve corresponds to the average over 50 triggers.

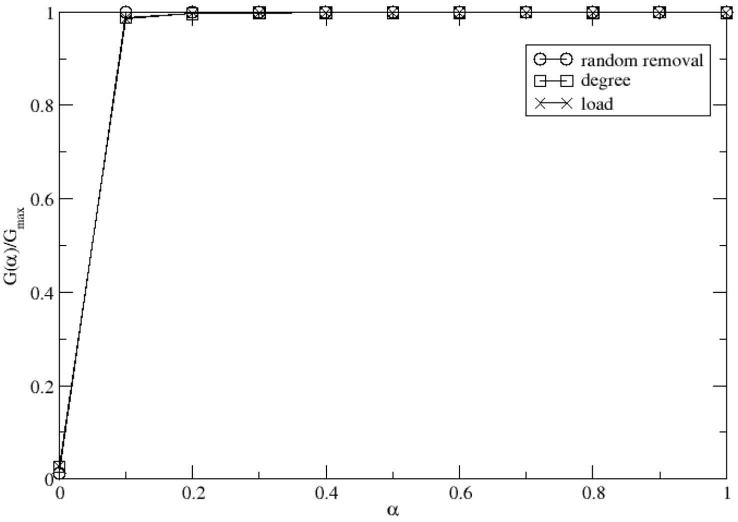

Figure 31 : Cascading failures in the random network (homogeneous networks) $\bar{k} = 3$, $N = 5000$. The network is generated by the shuffle method.

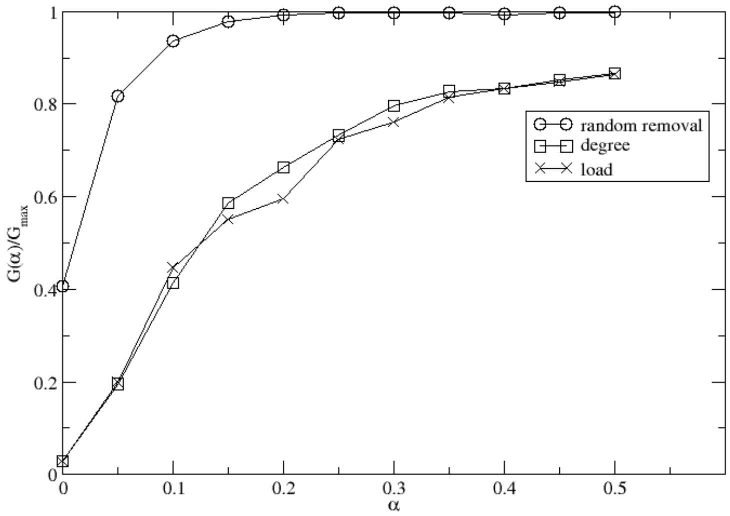

Figure 32 : Cascading failures described in a scale-free network. The network is generated by using BA model. The nodes number N = 5000, $m_0 = m = 3$.

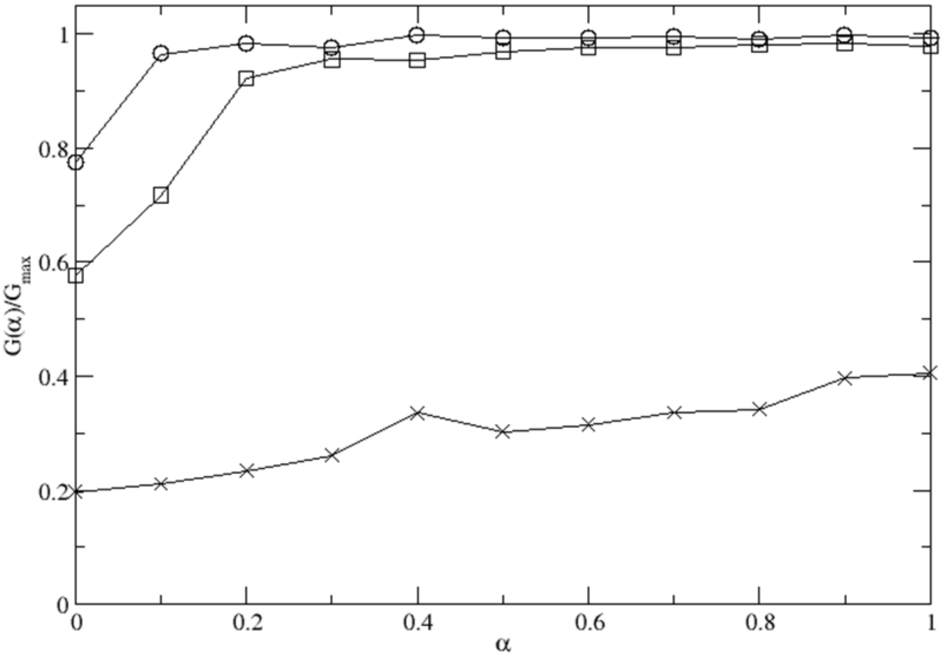

Figure 33 : Cascading failure failure in the western U.S. power grid. The legends are as defined in Figure 32

These results conform with [6] : The random networks appear to be more robust against attacks(removal among the nodes having the largest degrees or load) than the networks having a power-law degree distribution(Figure 30 and Figure 33) . The damage caused by the removals among the nodes having the largest load or degree is much larger than that by random removal, as shown in Figure 32.

3.2.2 transient dynamics

Another cascading failure model containing the time-dependent adjustments has been presented.[7] Assume there is a power grid consisting of N nodes, and some of them are generators, transmission stations, and utilities like the small power grid in Figure 34.

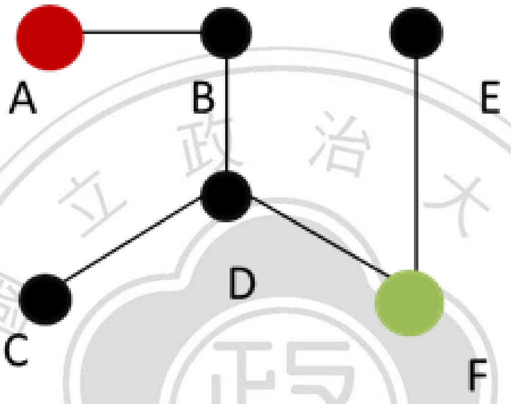

Figure 34 : A small power grid.

Node A is the generator which is the current source and the green node F is the utility where the current sinks. The other nodes in the small power grid are transmission stations. The question will be how to describe the electric current flowing in the power grid. Image the current flows from node A and then passes through node B to node D. The current on node D will be divided into two parts, one will follow the direction to the utility node F, and the other will flow into the transmission station node C. This behavior can be described as follow. The current on an edge from *j* → *I* at time *t* can be expressed as below[22]

$$C_{ij}(t) = W_{ij}\frac{C_j(t)}{\sum_k W_{kj}} = T_{ij}C_j(t) \tag{49}$$

The factor $W_{ij}\big/\sum_k W_{kj} = T_{ij}$ is the fraction of the current on node j outflows to node i. So the total current on node i at time *t+1* can be written as the sum of all the currents on the edge connecting the neighbors of node i.

$$C_i(t+1) = \sum\nolimits_{j=1}^{N} T_{ij} C_j(t) + j^{\pm} \quad , \; T_{ij} = \frac{W_{ij}}{W_i} \quad , \quad W_i = \sum\nolimits_{i=1}^{N} W_{ij} \qquad (50)$$

$j^{\pm}$ is the possible source($j^{\pm}$>0) or sink($j^{\pm}$<0) term. The load of an edge between node i and j can be defined via

$$L_{ij}(t) = C_{ij}(t) + C_{ji}(t) \qquad (51)$$

These allow us to study the wavelike spreading of the redistribution of the load in the power grid. For example, if we remove edge D in the imaginary British power grid[7], we will see different perturbations of the load on node A, B ,and C. The imaginary British power grid assume the topology of the UK high-voltage power transmission grid consisting of 120 nodes (generators, utilities, and transmission stations), and 160 edges (transmission lines) like Figure 35. In addition, the 10 red nodes in the network are the generators, the 10 green nodes are the utilities.

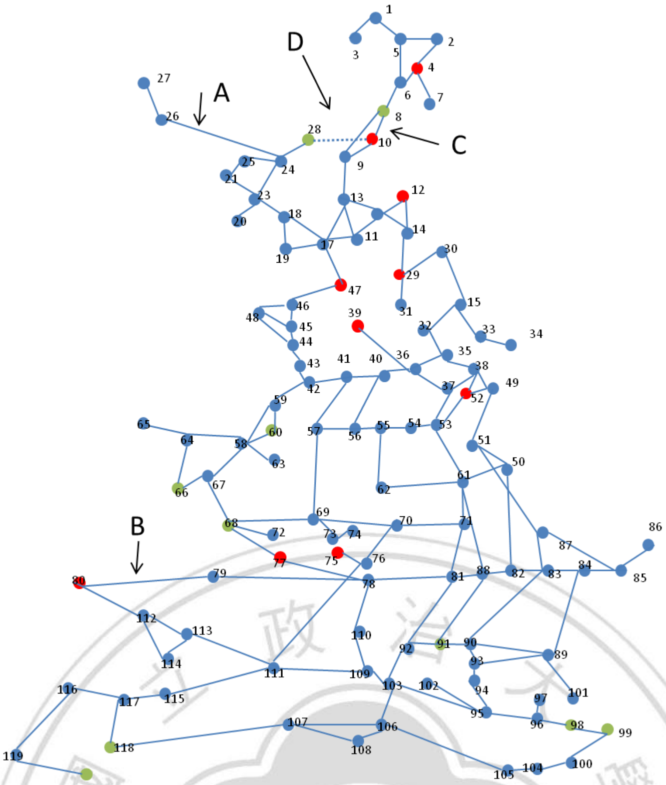

Figure 35： the imaginary British power grid

At the first (time=0), the currents on all edges and nodes are zero. With the evolution of time, the currents flow from the source nodes and distribute farther and farther from the source nodes, finally sink on the sink nodes. In this imaginary power grid, all the edges are unweighted. The currents (and the loads) on every nodes and edges will finally reach a stationary state which means the currents or loads will be constants. The load of the edge between node *i* and *j* when it is stationary is denoted as L_{ij}^{S} . In Figure 36, we plot the stabilizing process of edge D. The normalized edge load is defined as $L_{ij}(t)/L_{ij}^{S}$

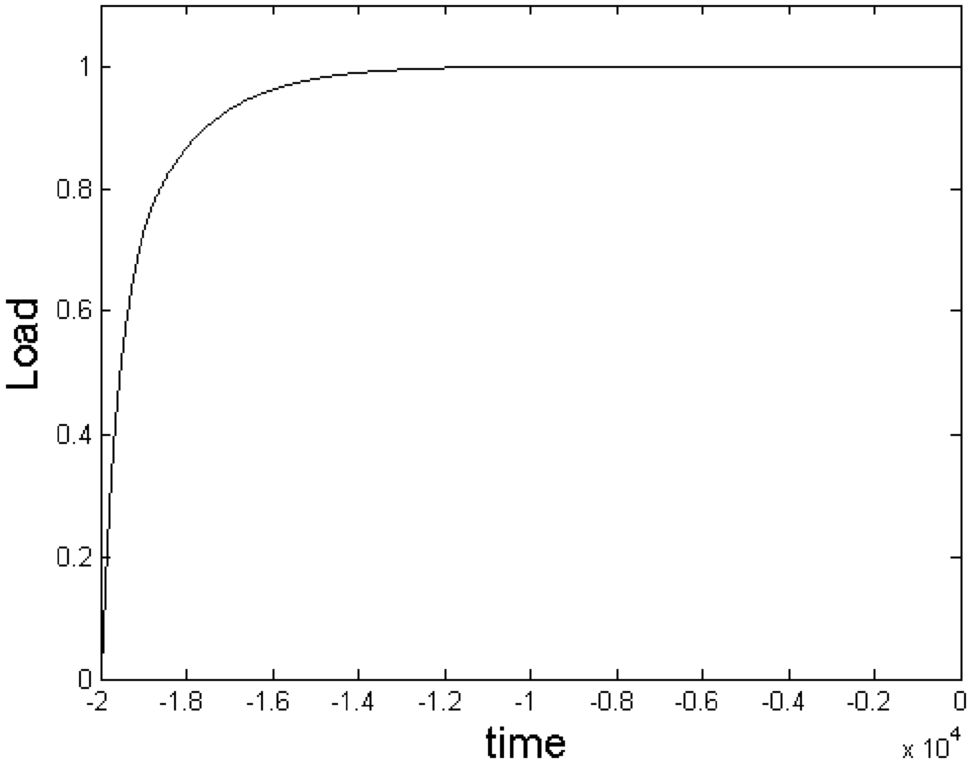

Figure 36 : the load variation of edge D with the evolution of time

Next, we consider if we remove edge D, the flow of the currents on the network will redistribute like we plot in Figure 37.As time<0, the loads on each edge are stationary, and the removal happen at time=0. At time=0 , the loads on every edges are defined as L_{ij}^{S} , and we can use normalized edge load $L_{ij}(t)/L_{ij}^{S}$ to observe the load variation behaviors on every edge. The variation behaviors of the load on each edge are very different (Figure 37). It is observable that after 300-400 time steps away from the time when the removal happened, the loads will be almost stationary again.

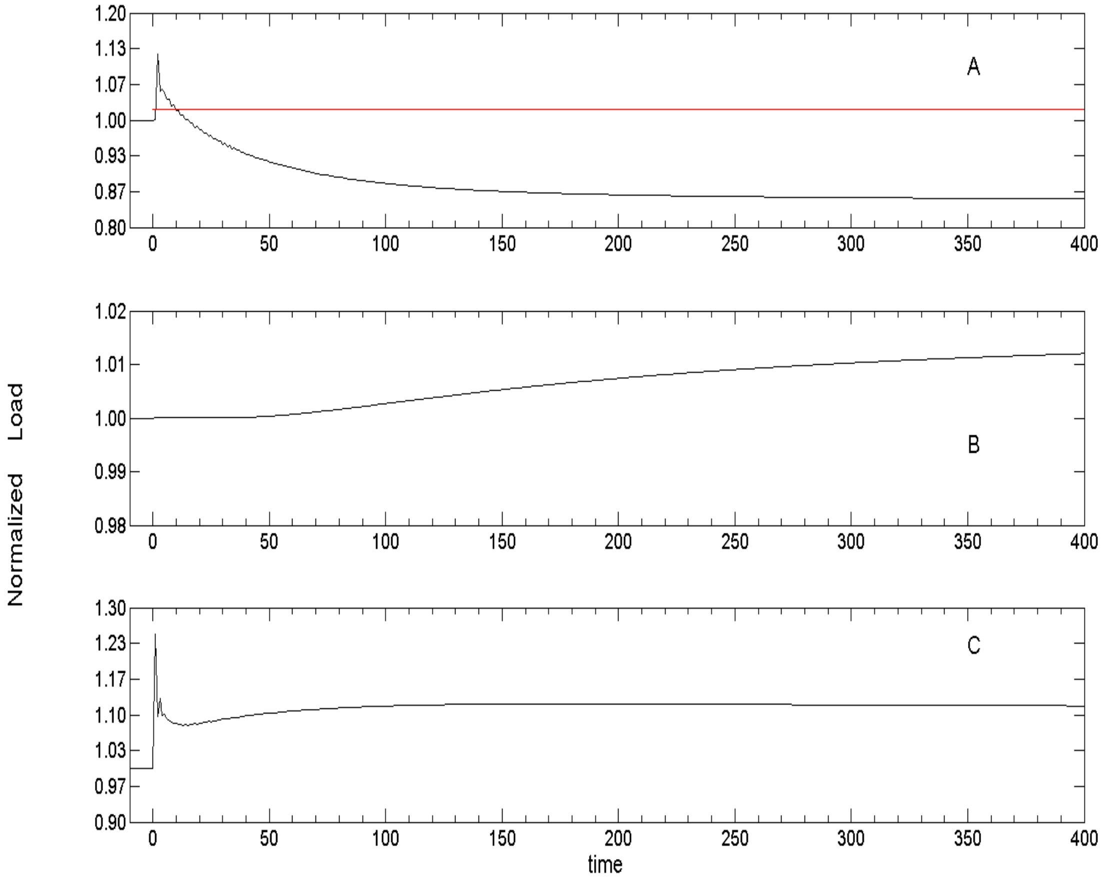

Figure 37 : the load variations of three different edges after the removal of edge D

As time t=0, we can define the edges capacities as[6]

$$C_{ij} = (1+\alpha)L_{ij}^{S} \tag{52}$$

α is which we call tolerance parameter. It reflects the ability of the edge to bear the load. Now we concern the overload situations that may occur before the new stationary state is reached. We define a new parameter τ , which is called "overload time", such that an edge will fail if the time duration of the edge overload exceeds a time period τ . More detail, for edge A we mentioned in Figure 37, if the value of the tolerance parameter α is 0.01, then we can define the capacity value is 1.01 which is the red line in Figure 37. The overload lasts for about 10 time steps. In this situation, if τ is equal to 50, the edge will not fail, because the overload time is not long enough to make it happen. In the opposite, if the value of τ is only 5, edge A fails. In this manner, we can model the cascading failures like Figure 30, and the only difference is the definitions of the load based on the betweeness centrality and the transient dynamics of the power grid. The process of the cascading failure triggered by a removal of an edge in the transient dynamics model is drawn in Figure 38.

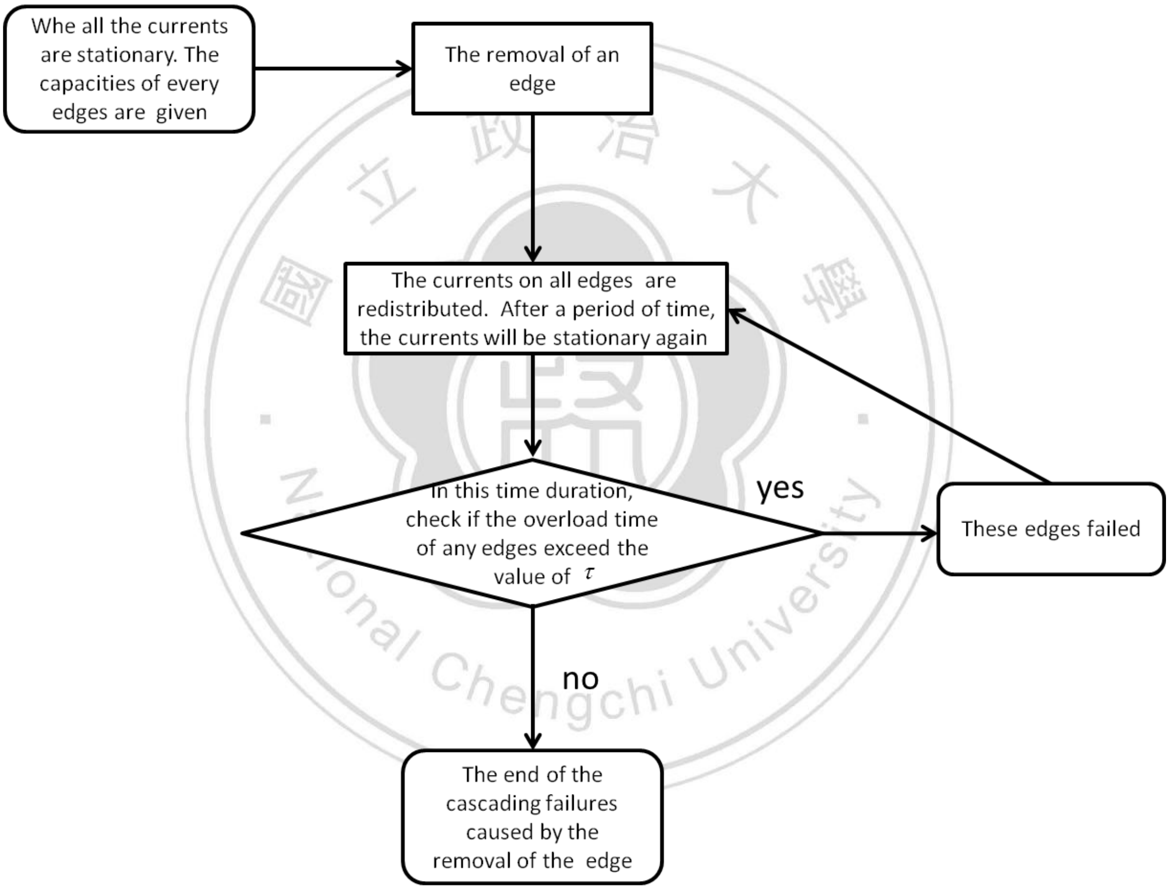
Whe all the currents are stationary. The capacities of every edges are given
The removal of an edge
The currents on all edges are redistributed. After a period of time, the currents will be stationary again
In this time duration, check if the overload time of any edges exceed the value of τ
yes
These edges failed
no
The end of the cascading failures caused by the removal of the edge

Figure 38 : the flow chart of the cascading failures triggered by the transient dynamics of a power grid.

Furthermore, we can evaluate the effect of the cascading failures by studying the fraction of edges remaining in the giant component $G(\alpha)/G_{\max}$.

In [7], the authors use this cascading failure model to test the robustness of the western US power grid[4]. The edges are assigned weights, drawn from a uniform distribution on the interval $[1,10]$, and 100 generators and 100 utility nodes $|j^{\pm}| = 10^{-8}$ are assigned randomly. The results were obtained by averaging over all possible removal of single edge. In Figure 39 we plot the cascading failures behaviors in western US power grid for different tolerance parameters α, and different curves represent the cascading failures behaviors for different values of τ .

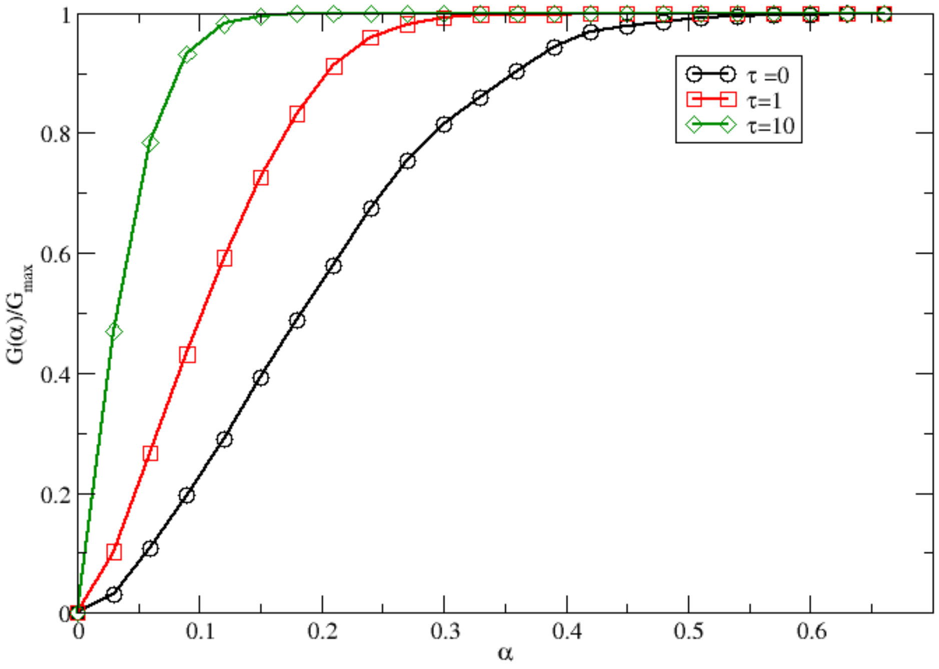

Figure 39 : the cascading failures triggered by transient dynamics of the power grid. The data is from [4]

As the tolerance parameter $\alpha = 0.1$, we observe the cascading failure triggered by the transient dynamic model for different removals of edge, and plot some of them in Figure 39. In most of the removals, the value of $G(\alpha)/G_{\max}$ will be extremely small corresponding a complete collapse, and the other removals will make the size of the power grid almost unchanged. We plot the largest component number variations of the removal of single edge in Figure 40.

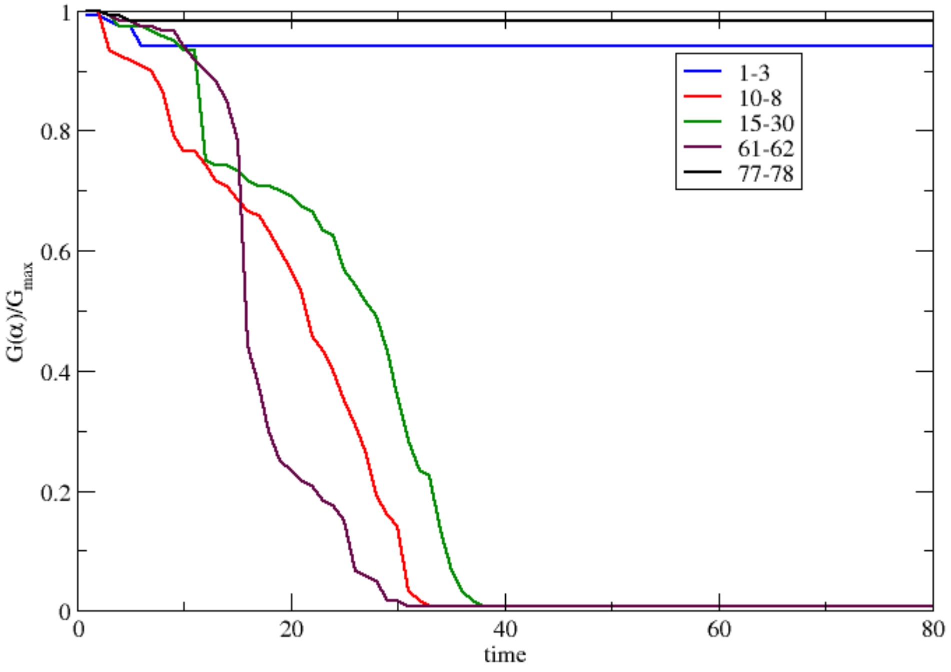

Figure 40 : the largest component number variations of the removal of single edge

This is a cascading failure for a specified weight distribution. For ranking the vulnerability of edges and excluding the influence of the distribution of the weight, we average the cascading failure triggered by the removal of an edge under different distribution of weight on the interval $[1,10]$. That is to say, for every realizations, the weights of every edges are generated randomly on the interval $[1,10]$, and the cascading failures triggered by the removal of each edge, $G(\alpha)/G_{\max}$, are recorded. Then the average largest group node number will be the realization average of $G(\alpha)/G_{\max}$.

3.3 case study II: the imaginary British power grid

In a power grid network, the nodes represent generators, transformers and substations, and edges represent high-voltage transmission lines between them. Now we have three different ways to test the vulnerability for a power grid. They are efficiency vulnerability, betweenness-cascading failures, transient-dynamics cascading failures. We still use the assuming British power grid as the example. For the three different ways to test the vulnerability of the power grid, we focus on the attack or errors of the transmission lines, that is to say, the removals of the edges. As a result, we sort the 10 most vulnerability edges for each way.

3.3.1 efficiency vulnerability

In this case, we neglect the properties of a node, that is to say, no matter what roles (sources or utilities) the nodes play, we only care about the global topological efficiencies variation after each removal of edge. We calculate the V^* values for each removal of edge and rank them to find the most vulnerable edges in the imaginary British power grid. We list the tem most vulnerable edges in Table 8 and plot them in Figure 41.

Table 8 : the ten most vulnerable edges according to the topological vulnerability evaluation in the assuming British power grid

V^*	starting node ——	arriving node
5.92%	6	8
3.92%	32	35
3.69%	42	43
3.47%	15	32
3.27%	116	117
3.23%	43	44
3.06%	69	73
2.93%	17	47
2.71%	46	47
2.69%	15	30

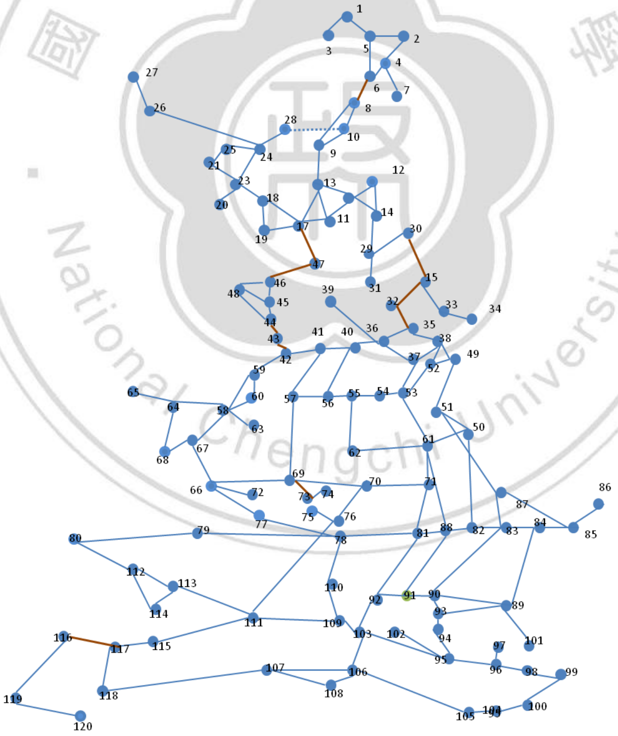

Figure 41 : the ten most vulnerable edges according to the topological vulnerability evaluation in the assuming power grid

3.3.2 the cascading failure triggered by the betweenness overload

We remove one of the edges in the imaginary British power grid and use the cascading failure model discuss in 3.2.1 to see if the edge removed played an important role in the power grid. For this purpose, after each cascading failure triggered by an removal of edge has come to its end, we record the node number of the largest component of this power grid. We list the ten most vulnerable edges in this kind of cascading failure in Table 9 and plot them in Figure 42. The tolerance parameter $\alpha = 0.1$.

Table 9 : the ten most vulnerable edges in the assuming British power grid according to the cascading failures triggered by the betweenness overloads.

the largest group node number after cascading failure	starting node	arriving node
13	40	41
14	111	115
19	11	16
19	89	90
20	36	40
21	35	36
22	50	61
23	8	10
23	58	64
23	82	83

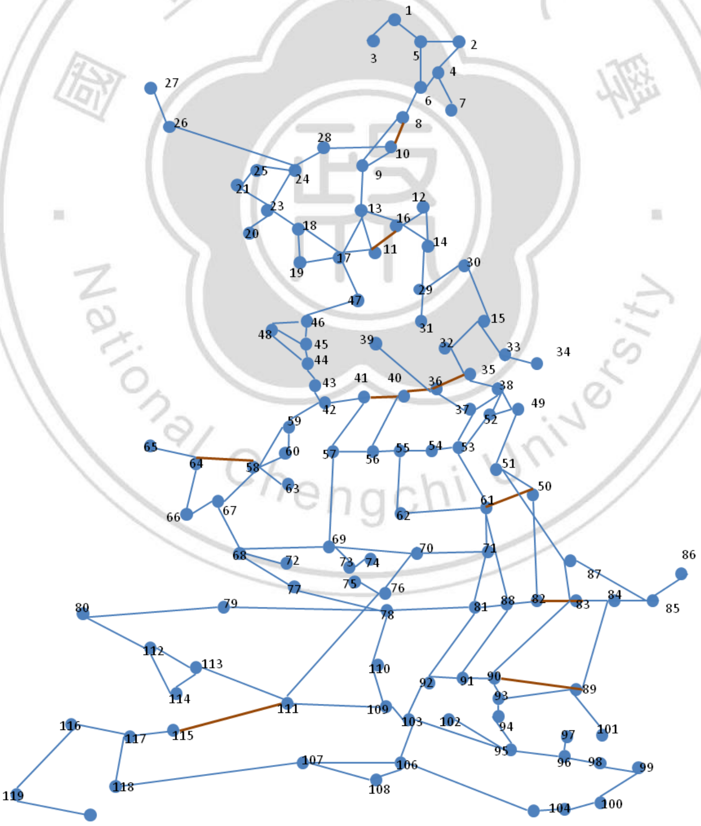

Figure 42 : the locations of the ten most vulnerable edges according to the cascading failures triggered by the betweenness overloads in the assuming British power grid.

3.3.3 the cascading failure triggered by the transient dynamics overload

We use the method in 3.2.2. The realizations of weight is 2000 to exclude the influence of the distribution of weight. The tolerance parameter $\alpha = 0.1$ and the overload time τ =1. The ten most vulnerable edges are listed in Table 10 and we plot their locations in Figure 43.

Table 10 : the ten most vulnerable edges according to the cascading failures triggered by the transient dynamics overloads in the assuming British power grid

the average largest group node number	starting node——	arriving node
4.26	44	43
4.42	32	15
5.19	104	100
5.45	105	104
5.74	106	105
6.54	43	42
8.17	35	32
8.28	117	115
8.73	30	15
9.83	111	115

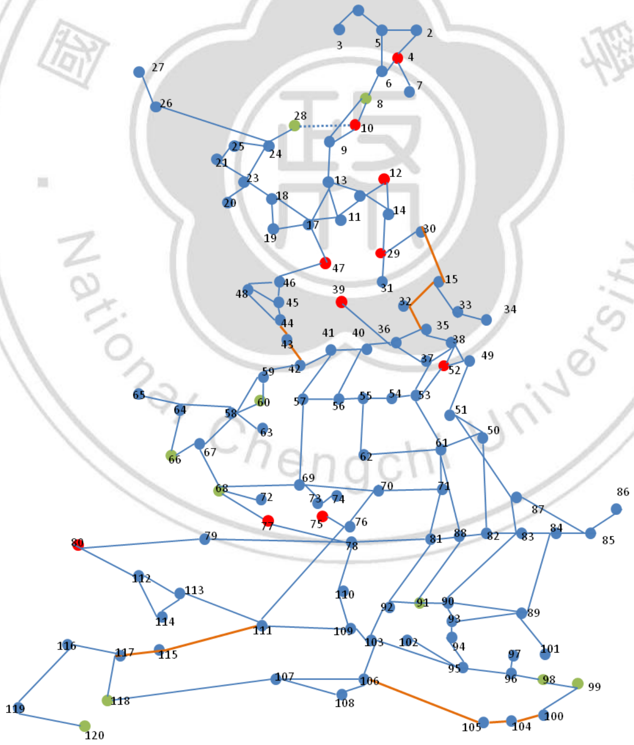

Figure 43 : the locations of the ten most vulnerable edges according to the cascading failures triggered by the transient dynamics overloads in the assuming British power grid

3.4 Discussion

That is reasonable for the three methodologies show different results. They view the vulnerability of a power grid in different concepts. But they show the reliability or vulnerability of a power grid can be study by complex network analysis. For efficiency vulnerability, we think the most vulnerable lines are the removals of the edges will make the network be divided into two parts or make the shortest path between many pairs increasing a lot.

For the cascading failure triggered by the betweenness overload, we can measure the average fraction of the largest component after different kind removal of edges, like[6].

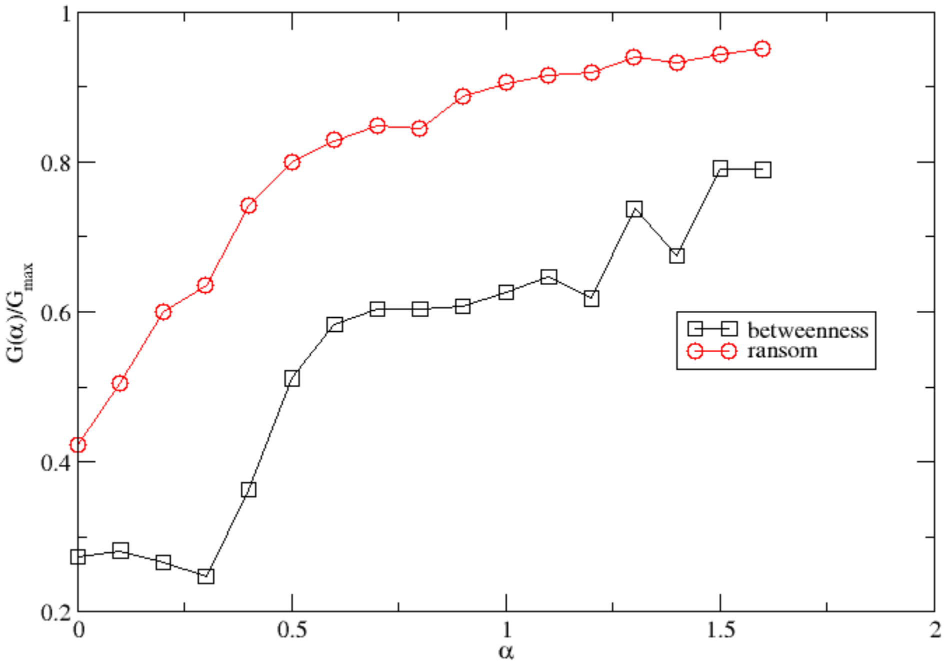

Figure 44: the two different cascading behaviors the two strategy of the removal of edges

In Figure 44, the red line represents the group change of 500 random removals of edges. The black line is the attack on the edges having high betweenness. Every realization there are five triggers, and the edges are chosen from the 20 edges having highest betweenness. For different α values, there are 100 realizations. In Figure 44, the cascading failure triggered by the removals of high betweenness edges are larger than the random removals. We think the behaviors of the two lines indicate that the removals of the nodes having high betweenness play an important role in triggering cascading failure in this power grid.

The cascading failure triggered by the transient dynamics overloads model is the most complicated case among the three methodologies. The factors which affect the cascading failure behaviors include the properties of the nodes which the edges connect, the locations of the edges, the tolerance parameter α , overload time τ . Although the size of the assuming network is small, but we show how to analyze the vulnerability of edges by using three different aspects.

4 Conclusion and future work

In case study I, we use two models: BA model and EBA model to simulate the CTTC network. By comparing real data and the simulation results, we find the impacts of the tournament design on the network structures. But there is still much analysis we hope to do in the future. For example, in the case study, the networks are all undirected and unweighted, thus these are all simple topology network structures. The information they carry is limited. If we can define the direction and the strength of an edge in CTTC networks, we can know more about the CTTC.

In case study II, we apply three different methodologies on an assuming power grid, and sort the 10 most vulnerable edges in 3.3. As we can see, the result will be different because we take different parameters into account in three different ways. This may indicate two things. First it will be more realistic to use complex network for vulnerability analysis by taking into consideration about actual electrical parameters. For example, the electrical distance is composed of the power transmission distribution factor and impedance[23],and there will be more 'realistic' analysis of vulnerability of power grid[23, 24]. Second, the further research should focus on how to merge the different insights of the different approaches, and combining with very realistic modeling(including physical laws and system dynamics)[21]. For example, The authors of [7] declared that the simple dynamical approach gives insights into the systems in which network topology is combined with flow, conservation of flow, and distribution laws. The cascading failure triggered by the transient dynamic model is simpler than the fully realistic state-of-the-art simulation, that is, power grids that include capacities, inductors, power generation, etc. The fully realistic state-of-the art simulation will spent a lot of time and they are very expensive. In this aspect, the transient dynamic model could quickly and economically give an overview of the system for more detailed and realistic simulations.[7] This kind of idea was also proposed in [21]. In the future, these differences between these methodologies should be discussed further. Furthermore, in our analysis we only discuss the vulnerability of single edge, but in real world some edges could fail at the same time. This kind of vulnerability should be studied in the future.

References

[1] M.E.J. Newman, A.-L.s. Barabási, D.J. Watts, The structure and dynamics of networks, Princeton University Press, Princeton, 2006.

[2] M.E.J. Newman, Networks : an introduction, Oxford University Press, Oxford ; New York, 2010.

[3] J. Travers, S. Milgram, Experimental Study of Small World Problem, Sociometry, 32 (1969) 425-443.

[4] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature, 393 (1998) 440-442.

[5] R. Albert, H. Jeong, A.L. Barabasi, Error and attack tolerance of complex networks, Nature, 406 (2000) 378-382.

[6] A. Motter, Y.-C. Lai, Cascade-based attacks on complex networks, Physical Review E, 66 (2002).

[7] I. Simonsen, L. Buzna, K. Peters, S. Bornholdt, D. Helbing, Transient Dynamics Increasing Network Vulnerability to Cascading Failures, Physical Review Letters, 100 (2008).

[8] K.I. Goh, B. Kahng, D. Kim, Universal Behavior of Load Distribution in Scale-Free Networks, Physical Review Letters, 87 (2001).

[9] E.R.C. Edward .A. Bender, The asymptotic number of labeled graphs with given degree sequences, Journal of Combinatorial Theory, Series A, 24 (1978) 296-307.

[10] I.d.S. Pool, M. Kochen, Contacts and Influence, Social Networks, 1 (1978) 5-51.

[11] B. Uzzi, J. Spiro, Collaboration and creativity: The small world problem, Am J Sociol, 111 (2005) 447-504.

[12] A.L. Barabasi, R. Albert, Emergence of scaling in random networks, Science, 286 (1999) 509-512.

[13] S.H. Strogatz, Exploring complex networks, Nature, 410 (2001) 268-276.

[14] R. Albert, A.L. Barabasi, Topology of evolving networks: local events and universality, Phys Rev Lett, 85 (2000) 5234-5237.

[15] D.J. Watts, S.H. Strogatz, Collective dynamics of 'samll-world' networks, Nature, 343 (1998) 440-442.

[16] S. Redner, How popular is your paper? An empirical study of the citation distribution, THE EUROPEAN PHYSICAL JOURNAL B, 4 (1998) 131-134.

[17] D.M. Pennock, G.W. Flake, S. Lawrence, E.J. Glover, C.L. Giles, Winners don't take all: Characterizing the competition for links on the web, Proceedings of the National Academy of Sciences of the United States of America, 99 (2002) 5207-5211.

[18] O. Sporns, D.R. Chialvo, M. Kaiser, C.C. Hilgetag, Organization, development and function of complex brain networks, Trends Cogn Sci, 8 (2004) 418-425.

[19] C. Li, P.K. Maini, An evolving network model with community structure, JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL, 38 (2005) 9741-9749.

[20] V. Latora, M. Marchiori, Efficient Behavior of Small-World Networks, Physical Review Letters, 87 (2001).

[21] I. Eusgeld, W. Kröger, G. Sansavini, M. Schläpfer, E. Zio, The role of network theory and object-oriented modeling within a framework for the vulnerability analysis of critical infrastructures, Reliability Engineering & System Safety, 94 (2009) 954-963.

[22] I. Simonsen, Diffusion and networks: A powerful combination!, Physica A: Statistical Mechanics and its Applications, 357 (2005) 317-330.

[23] S. Arianos, E. Bompard, A. Carbone, F. Xue, Power grid vulnerability: a complex network approach, Chaos, 19 (2009) 013119.

[24] X.Y. Ajendra Dwivedi, Peter Sokolowski, Identifying Vulnerable Lines In a Power Network using Complex Network Theory, IEEE International Symposium on Industrial Electronics (ISlE 2009), (2009).