

Analysis and Study of Cascading Failures in Gene Network*

Wang Shudong¹, Shi Songtao², Ge Yanru³, Sun Longxiao¹, Xu Dashun⁴, Meng Dazhi^{1,2}

¹College of Information Science and Engineering, Shandong University of Science and Technology, Qingdao, Shandong 266510, China

²School of Software Engineering; College of Applied Science, Beijing University of Technology, Beijing 100124, China

³Department of Control Science and Engineering, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

⁴Department of Mathematics, Southern Illinois University Carbondale, Makandia 62958, USA
Email: wangshd2008@yahoo.com.cn, dzhmeng07@yahoo.com.cn

Abstract Genome forms gene networks in terms of complicated interactions to realize its functions. Further research for gene networks can help to comprehend and predict many unknown functions of genome. In this work, cascading failure models of weighted gene networks are built and the robustness of the models is also analyzed and discussed. Based on the data of normal and lung adenocarcinoma stages, by simulating and analysing the cascading failures of the two gene network models, that cascading failures occur more likely in the networks for adenocarcinoma experimental groups than the one of normal control group is discovered. In the numerical experiments, we notice that nine genes of experimental group and eight genes of control group are of very strong destructibility for the robustness of experimental and control networks respectively. The failures of these genes can lead to the collapse or paralysis of the whole network. Therefore, we conclude that these genes might play important roles in keeping normal level or developing lung adenocarcinoma of organisms. When applying the methods of modeling and analysing cascading failures in gene network to other diseases' data, biomedical scientists can be enlightened for understanding the mechanism of diseases and predicting the functions of significant genes.

Keywords Systems biology Gene network Cascading failure Network statistics

1. Introduction

Research on the biological functions of genome is a major issue in life science^[1-3]. The study for gene networks describing the complicated genetic interactions in genome is an important way to understand biological functions^[4-9]. So far, many methods have been proposed to build and analyze gene networks^[4]. Amy Hin Yan Tong et al.^[5] investigated the correspondence between the dense local neighborhoods in gene regulatory network and biological functions of yeasts. Mark Kittisopikul and Gürol M. Süel^[6] studied the biological significance of feed-forward loop motifs in gene network of *Escherichia coli* and discovered that most

* This work was supported by the National Natural Science Foundation of China (Grant Nos. 60874036, 60503002) and SDUST Research Fund.

feed-forward loops have two kinds of regulatory functions. Hallinan J. S. et al.^[7] analyzed the relationships among the motifs, feedback loops and their dynamical features in gene regulatory network. Bowers et al.^[8] proposed a computational approach-logic analysis of phylogenetic profiles to identify detailed relationships among genes or proteins on the basis of genomic data, which was applied into 4873 distinct orthologous protein families of 67 fully sequenced organisms, and identified 750, 000 triplets previously unknown logic regulatory relationships. Shudong Wang et al.^[9] constructed a logical network with 16 active genes of shoot in different external stimuli, and analyzed the dynamics of the logical network.

Now, the theoretical models of cascading failures and their mechanisms, prevention and control for various actual complex networks have been relatively deeply studied^[10-17]. For instance, R. Kinney et al.^[16] analyzed the cascading failures in the North American power grid. The results show that deliberate attacks can lead to a substantial decline in the transmission efficiencies of power grid, while random failures have nearly no influence. Ashley G. Smart et al.^[17] investigated the relationships between structure and robustness in the metabolic networks of *Escherichia coli*, *Methanosarcina barkeri*, *Staphylococcus aureus*, and *Saccharomyces cerevisiae* using a cascading failure model based on a topological flux balance criterion and found that metabolic networks are exceptionally robust compared to appropriate null models. But the reports are rare about cascading failures in gene network. This work investigates the influences of cascading failures on gene networks. Based on the documental data, by comparing cascading failures of gene networks for normal and lung adenocarcinoma groups, we discover control networks are quite robust than lung adenocarcinoma experimental ones. This indicates the change from normal organisms into lung adenocarcinoma ones may result from dysfunctions or gene mutations (considered as failures) of some genes. Through numerical experiments, we notice that failures of some genes in experimental and control groups can lead to collapse or paralysis of the whole network. These genes might play important roles in keeping normal level or developing lung adenocarcinoma in organisms.

This paper is organized as follows: the work background and existing methods for studying gene networks are introduced in the first part. The cascading failure model and its algorithm used in this work are presented in detail in the second part. The methods and main results of numerical experiments of cascading failure model are described based on two data sets of (normal) control group and (lung adenocarcinoma) experimental group in the third part. The obtained results are analyzed and discussed, and the possible corresponding biological significances are also pointed out in the fourth part.

2. The Model and Algorithm of Cascading Failure

In this research, we consider cascading failures of complex gene networks, so we treat genes no different from nodes of complex networks. We use the capacity-load in cascading failure model. Let $G = (V, E, W)$ be a complex (directed or undirected) gene network with node-set $V = \{1, 2, \dots, N\}$, edge-set E and weight-set W . Suppose w_{ij} is the weight from node i to j in complex gene network G . Then the edge-length from node i to j is defined as the reciprocal value $\frac{1}{w_{ij}}$ of w_{ij} . If $w_{ij} = 0$, then the edge-length from node i to j is ∞ . The

greater the weight between two nodes is, the lesser the edge-length is; the lesser the weight is, the greater the edge-length is, and vice versa. The shortest paths from node i to j are these paths corresponding to the smallest sum of edge-length in all the paths from node i to j . Obviously, the shortest paths from node i to j are not always unique. Suppose there exist

p shortest paths from node i to j . Then the load of any shortest path R is defined as $\frac{1}{p}$ of the product of all the weights in R , i.e. $\frac{\prod_{(i,j) \in R} w_{ij}}{p}$. The load L_j of node j is defined as the sum of the loads of all the shortest paths passing through node j . The capacity C_j of node j is proportional to its initial load L_j^0 , i.e. $C_j = (1 + \alpha)L_j^0, j = 1, 2, \dots, N$, where constant $\alpha \geq 0$ is a tolerance factor. If the load of a node is greater than its capacity, then it is called a failure node. After deleting node i , and causing s_i failure nodes (including node i), then s_i is defined as the size of cascading failure of node i and $d_i = \frac{s_i}{N}$ as the size-ratio of cascading failure. If $d_i \geq t_{cf}$, then the network breaks down, otherwise, the network doesn't have failure. This is a criterion of network failure, where t_{cf} is the threshold of network failure.

Let $sign1(i) = \begin{cases} 1, & d_i \geq t_{cf} \\ 0, & d_i < t_{cf} \end{cases}$. Then the percentage of failure nodes of the network

$$P = \frac{\sum_{i=1}^N sign1(i)}{N}; \text{ the largest size-ratio of cascading failure } R_{max} = \max \{d_i, i = 1, 2, \dots, N\};$$

the average size-ratio of cascading failure $R = \frac{\sum_{i=1}^N d_i}{N}$. Let $sign2(i) = \begin{cases} 1, & d_i \geq d \\ 0, & d_i < d \end{cases}$ (d is a variable parameter). Then the cumulative probability of size-ratio of cascading failure

$$P(d' \geq d) = \frac{\sum_{i=1}^N sign2(i)}{N}, \text{ which indicates the probability of size-ratio } d_i \text{ of cascading failure greater than } d.$$

Obviously, R_{max} , R and $P(d' \geq d)$ are the important parameters measuring the robustness or fragility of network.

Based on the above mentioned definitions and symbols, we present the algorithm (CFA) of cascading failure model as follows:

- ① Input the weight matrix of complex gene network $G = (V, E, W)$.
- ② Calculate initial load L_j^0 of node j and its capacity $C_j = (1 + \alpha)L_j^0, j = 1, 2, \dots, N. i = 1$.
- ③ Delete node i and its incident edges in the network.
- ④ Calculate the load of every node in the present network and compare the capacity with the load of every node. If the load is lesser than the capacity for every node in the present network, then go to ⑤, otherwise, delete every node and its incident edges whose load is greater than its capacity, go to ④.
- ⑤ If the size-ratio of cascading failure after deleting node i is greater than or equal to the threshold t_{cf} of network failure, then the network breaks down.
- ⑥ $i = i + 1$. If $i < N$, then go to ③.
- ⑦ Calculate the largest size-ratio of cascading failure R_{max} , average size-ratio of cascading

failure R , and the cumulative probability of size-ratio of cascading failure $P(d' \geq d)$.

3. The Methods and Results of Numerical Experiments

3.1 Data Sources

Data used in this work are from the results of lung adenocarcinoma network studied by Yuanyuan Zhang et al. (the detailed data sources can be seen in). We use the mutual information network and directed (or 1-order logic) network for control group (abbreviated as N) and lung adenocarcinoma experimental group (abbreviated as AC). In the mutual information network (directed weighted network), mutual information value (U value) is the weight of networks. We denote the weight matrices of mutual information (directed-weighted) network for control and experimental groups by $M_N(L_N)$ and $M_{AC}(L_{AC})$, respectively.

Table 1 The list of $P(d' \geq d)$ along with the change of the network threshold t_{net} in control and experimental networks respectively. When $d > 0.34$, $P(d' \geq d)$ is equal to zero

t_{net}	0.52		0.5		0.45		0.4	
d	AC	N	AC	N	AC	N	AC	N
0.01	1	1	1	1	1	1	1	1
0.02	1	1	1	1	1	0.0794	1	0.0714
0.03	1	0.1053	1	0.0870	0.1351	0.0794	0.1667	0.0714
0.04	0.2069	0.1053	0.1875	0.0870	0.1351	0.0794	0.1667	0.0714
0.05	0.2069	0.1053	0.1875	0.0870	0.1351	0.0794	0.1667	0.0714
0.06	0.2069	0.1053	0.1875	0.0870	0.1351	0.0794	0.1667	0.0714
0.07	0.1724	0.1053	0.1563	0.0870	0.1351	0.0794	0.1667	0.0714
0.08	0.1724	0.1053	0.1563	0.0870	0.1351	0.0794	0.1667	0.0714
0.09	0.1724	0.1053	0.1563	0.0870	0.1351	0.0794	0.1667	0.0714
0.1	0.1724	0.1053	0.1563	0.0870	0.1351	0.0794	0.1667	0.0714
0.11	0.1724	0.1053	0.1563	0.0870	0.1351	0.0794	0.1667	0.0714
0.12	0.1724	0.1053	0.1563	0.0870	0.1351	0.0794	0.1667	0.0714
0.13	0.1724	0.1053	0.1563	0.0870	0.1351	0.0794	0.1667	0.0714
0.14	0.1724	0.1053	0.1563	0.0870	0.1351	0.0794	0.1667	0.0714
0.15	0.1724	0.1053	0.1563	0.0870	0.1351	0	0.1667	0.0714
0.16	0.1724	0.1053	0.1563	0.0870	0.1351	0	0.1667	0.0286
0.17	0.1724	0.1053	0.1563	0.0870	0.1351	0	0.1667	0.0286
0.18	0.1724	0.1053	0.1563	0	0.1351	0	0.1667	0
0.19	0.1724	0.1053	0.1563	0	0.1351	0	0.1667	0
0.2	0.1724	0.1053	0.1563	0	0.1351	0	0.1667	0
0.21	0.1724	0.1053	0.1563	0	0.1351	0	0.1667	0
0.22	0.1724	0.0790	0.1563	0	0.1351	0	0.1667	0
0.23	0.1724	0.0790	0.1563	0	0.1351	0	0.1667	0
0.24	0.1724	0	0.1563	0	0.1351	0	0.1667	0
0.25	0.1724	0	0.1563	0	0.1351	0	0.1667	0
0.26	0.1724	0	0.1563	0	0.1351	0	0.1667	0
0.27	0.1724	0	0.1563	0	0.1351	0	0.1667	0

0.28	0.1724	0	0.1563	0	0.1351	0	0.1667	0
0.29	0.1724	0	0.0625	0	0.1351	0	0.1667	0
0.3	0.1724	0	0.0625	0	0	0	0.1667	0
0.31	0.1724	0	0.0625	0	0	0	0	0
0.32	0.069	0	0	0	0	0	0	0
0.33	0.069	0	0	0	0	0	0	0
0.34	0.069	0	0	0	0	0	0	0

3.2 The Methods and Results

3.2.1 The Results of Mutual Information Gene Network

In order to highlight the characteristics of the network structure, we analyze the changes of $P(d' \geq d)$ along with the network threshold t_{net} . When taking different network thresholds, we can obtain the mutual information networks with different coarse granularities. The corresponding weight matrices M_N and M_{AC} are the inputs in the above CFA algorithm. Obviously, the greater the network threshold t_{net} is, the coarser the granularity is, the more the lost information is and the computational complexity is relatively low; on the contrary, the lesser the network threshold t_{net} is, the finer the granularity is, the less the lost information is, but the computational complexity is relatively high. The detailed data and the changing curves are in Table 1 and Fig. 1. From Table 1 and Fig. 1, it is obvious that $P(d' \geq d)$ of experimental network is clearly higher than that of control one under any network threshold. This indicates the ratio of nodes in experimental network which can result in cascading failures is much greater than the one of control group. With the increasing of d , $P(d' \geq d)$ in control network reduces to zero earlier than in experimental one. In other words, taking certain appropriate d , control network has no failure while experimental network has more failures. Moreover, with the increasing of network threshold t_{net} , the platform value of $P(d' \geq d)$ of control network is 0.0714, 0.0794, 0.0870 and 0.1053 respectively, showing gradually increasing tendency. This indicates that with the decreasing of the numbers of nodes and edges, cascading failures are more likely to occur in the gene networks, namely: the robustness goes worse. The genes resulting in the cascading failures of control and experimental groups under all four network thresholds are NRAS, PIK3CA, MAPK9, TOP2A and FGF1, RET, WT1, TCL1A, HRK, respectively. The detailed situations can be seen in Table 2 and 3.

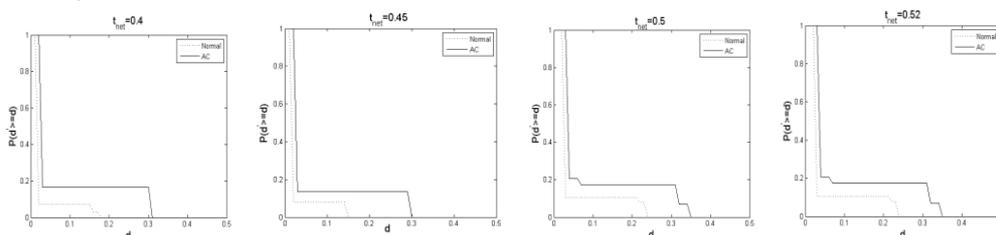


Fig. 1 Taking the network threshold t_{net} as 0.4, 0.45, 0.5, 0.52 respectively, the changing curves of $P(d' \geq d)$ along with d in control and experimental networks.

Table 2 The list of the relative greater size-ratio d_* of cascading failure of control network under different network thresholds t_{net} , where * denotes the gene. The following presentation is similar.

0.52	0.5	0.45	0.4
------	-----	------	-----

gene	d_* (%)						
NRAS	23.68	NRAS	17.39	NRAS	14.2857	NRAS	15.71
PIK3CA	23.68	PIK3CA	17.39	PIK3CA	14.2857	PIK3CA	17.14
MAPK9	21.05	MAPK9	17.39	MAPK9	14.2857	MAPK9	17.14
TOP2A	23.68	TOP2A	17.39	RBL1	14.2857	RBL1	15.71
				TOP2A	14.2857	TOP2A	15.71

Table 3 The list of the relative greater size-ratio d_* of cascading failure of experimental network under different network thresholds t_{net} .

0.52		0.5		0.45		0.4	
gene	d_* (%)						
FGF1	34.48	FGF1	31.25	FGF1	29.73	FGF1	30.95
RET	31.03	RET	28.125	RET	29.73	FGF2	30.95
WT1	31.03	WT1	28.125	WT1	29.73	HSPB2	30.95
TCL1A	34.48	TCL1A	31.25	TCL1A	29.73	RET	30.95
HRK	31.03	HRK	28.125	HRK	29.73	WT1	30.95
						TCL1A	30.95
						HRK	30.95

3.2.2 The Results of Directed Gene Network

To comprehensively measure the robustness and fragility of directed weighted gene network, we analyze the situations of R , R_{max} and $P(d' \geq d)$ with the changes of network thresholds t_{net} (Table 4 and Table 5). From Table 4 and 5, we discover that R , R_{max} and $P(d' \geq d)$ of experimental network are clearly greater than the ones of control network under any network threshold. This shows that cascading failures occur in experimental network more easily than in control one. The genes resulting in cascading failures of control and experimental networks under five network thresholds are BAD, ING1, RAF1, TRAF3 and ESR2, HSPB2, NOV, TAL1 respectively. The detailed situations can be seen in Table 6 and 7.

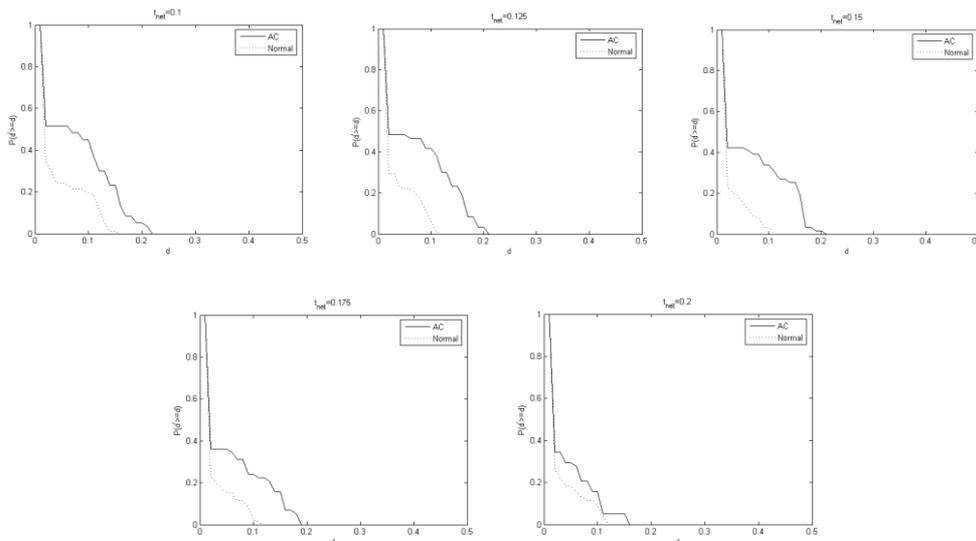


Fig. 2 Taking the network threshold t_{net} as 0.1, 0.125, 0.15, 0.175, 0.2 respectively, the changing curves of $P(d' \geq d)$ along with d in control and experimental networks.

Table 4 The list of R , R_{max} along with the change of network threshold t_{net} in control and experimental networks respectively.

t_{net}	Stage	No. of nodes	No. of edges	R	R_{max}
0.100	AC	60	487	0.1355	0.2167
	N	98	1124	0.0858	0.1531
0.125	AC	60	392	0.1385	0.2000
	N	95	887	0.0756	0.1158
0.150	AC	59	338	0.1390	0.2034
	N	90	700	0.0635	0.1000
0.175	AC	58	285	0.1281	0.1897
	N	86	560	0.0686	0.1163
0.200	AC	58	240	0.0888	0.1552
	N	77	446	0.0727	0.1169

Table 5 The list of $P(d' \geq d)$ along with the change of network threshold t_{net} in control and experimental networks respectively. When $d > 0.22$, $P(d' \geq d)$ is equal to zero.

t_{net}	0.1		0.125		0.15		0.175		0.2	
d	AC	N								
0	1	1	1	1	1	1	1	1	1	1
0.01	1	1	1	1	1	1	1	1	1	1
0.02	0.5167	0.3469	0.4833	0.2947	0.4237	0.2333	0.3621	0.2326	0.3448	0.2597
0.03	0.5167	0.2959	0.4833	0.2947	0.4237	0.2000	0.3621	0.1977	0.3448	0.2208
0.04	0.5167	0.2449	0.4833	0.2316	0.4237	0.1889	0.3621	0.1744	0.2931	0.1818
0.05	0.5167	0.2449	0.4833	0.2211	0.4237	0.1556	0.3621	0.1512	0.2931	0.1818

MAPK9	9.18								
-------	------	--	--	--	--	--	--	--	--

Table 7 The list of the relative greater size-ratio d_* of cascading failure of the experimental network under different network thresholds t_{net} , where * denotes the gene.

0.1		0.125		0.15		0.175		0.2	
gene	d_* (%)								
GLI2	21.67	FES	20.00	HSPB2	20.34	NOV	18.97	EXTL3	15.52
TP63	21.67	ROS1	20.00	NOV	18.64	WNT3	18.97	NOV	15.52
RET	20.00	HSPB2	18.33	ERG	16.95	TCL1A	18.97	ROS1	15.52
FES	18.33	WNT3	18.33	ESR2	16.95	EXTL3	17.24	ESR2	10.34
TAL1	18.33	TCL1A	18.33	EXTL3	16.95	ESR2	15.52	GLI2	10.34
ESR2	16.67	ESR2	16.67	FES	16.95	HSPB2	15.52	HSPB2	10.34
EXTL3	16.67	GLI2	16.67	CXCL3	16.95	TAL1	15.52	IL1A	10.34
NOV	16.67	CXCL3	16.67	TAL1	16.95	TP63	15.52	TAL1	10.34
E2F1	15.00	IL1A	16.67	WNT3	16.95	HRK	15.52	TCL1A	10.34
CXCL3	15.00	NOV	16.67	TP63	16.95				
HSPB2	15.00	TAL1	16.67	HRK	16.95				
ROS1	15.00								
WNT3	15.00								
TCL1A	15.00								

4. Conclusion and Analysis

In this research, we analyze and investigate the cascading failures in control and experimental networks. Through numerical experiments, we discover: under all the network thresholds, cascading failures occur in experimental networks more easily than in control ones for undirected and directed weighted gene networks. This indicates that the normal organisms are quite robust while diseased organisms are more fragile. In Table1, we notice that: with the increasing of network thresholds, the platform values of $P(d' \geq d)$ of control network are gradually increasing. This shows that with the decreasing of the numbers of nodes and edges, the robustness goes worse. In other words, with the increasing of the numbers of nodes and edges, the robustness goes better. This indicates the intrinsic reason of organisms functioning normally and stably maybe is that most genes play their own roles in organisms.

In the process of numerical experiments, we notice that failures of genes BAD, ING1, RAF1, TRAF3, NRAS, PIK3CA, MAPK9, TOP2A of control group and ESR2, HSPB2, NOV, TAL1, FGF1, RET, WT1, TCL1A, HRK of experimental group under all the network thresholds result in collapse or paralysis of the whole network. This provides some useful reference informations for the normal or diseased organisms. For example, activation of gene Bad may induce apoptosis in human lung adenocarcinoma cells^[18]. In other words, the failure of gene Bad leads to the defunctionalization of inducing the apoptosis of human lung adenocarcinoma cells and the organism might suffer from lung adenocarcinoma. Gene MAPK9 may enhance the stability of tumor suppressor p53 and its failure can reduce the stability of p53. Thus the organism might develop into cancer.

References

- [1] Motoki Shiga, Ichigaku Takigawa, Hiroshi Mamitsuka. Annotating gene function by combining expression data with a modular gene network. *Bioinformatics*, 2007, 23(13):i468-i478.
- [2] Marco Sardiello, Michela Palmieri, Alberto di Ronza, et al.. A gene network regulating lysosomal biogenesis and function. *Science*, 2009, 325(5939):473-477.
- [3] Roded Sharan¹, Trey Ideker. Modeling cellular machinery through biological network comparison. *Nature Biotechnology*, 2006, 24: 427-433.
- [4] Thomas Schlitt, Alvis Brazma. Present approaches to gene regulatory network modelling. *BMC Bioinformatics*, 2007, 8(Suppl 6) S9:1-22.
- [5] Amy Hin Yan Tong, Guillaume Lesage, Gary D. Bader, et al.. Global mapping of the yeast genetic interaction network. *Science*, 2004, 303(5659):808-813.
- [6] Mark Kittisopikul, Gürol M. Süel. Biological role of noise encoded in a genetic network motif. *PNAS*, 2010, 107 (30): 13300-13305.
- [7] Hallinan, J.S., Jackway, P.T. Network motifs, feedback loops and the dynamics of genetic regulatory networks. *Computational Intelligence in Bioinformatics and Computational Biology*, 2005. CIBCB '05, 1-7.
- [8] Bowers P M, Cokus S J, Eisenberg D, et al.. Use of logic relationships to decipher protein network organization. *Science*, 2004, 306(5706):2246-2249.
- [9] Shudong Wang, Yan Chen, Qingyun Wang, Eryan Li, Yansen Su, Dazhi Meng. Analysis for gene networks based on logical relationships. *Journal of Systems Science and Complexity*, 2010, 23(5): 999-1011.
- [10] Asavathiratham C. *The influence model: a tractable representation for the dynamics of networked Markov chains*. Elec. Eng. and Comp. Sci. Dept., MIT, 2000.
- [11] Dobson I, Chen J, Thorp J S, et al.. *Examining criticality of blackouts in power system models with cascading events*. Proceedings of 35th Hawaii International Conference on System Sciences, 2002, 63-72.
- [12] Moreno Y, Gómez J B, Pacheco A F. Instability of scale-free networks under node-breaking avalanches. *Europhys. Lett.*, 2002, 58(4):630-636.
- [13] Moreno Y, Pastor-Satorras R, Vázquez A, Vespignani A. Critical load and congestion instabilities in scale-free networks. *Europhys. Lett.*, 2003, 62:630-636.
- [14] Crucitti P, Latora V, Marchiori M. Model for cascading failures in complex networks. *Phys. Rev. E*, 2004, 69, 045104:1-4.
- [15] Xiaofan Wang, Xiang Li, Guanrong Chen. *The Theory and Application of Complex Networks*. Tsinghua Press, Beijing, 2006.4.
- [16] R.Kinney, P.Crucitti, R.Albert, V.Latora. Modeling cascading failures in the North American power grid. *Eur. Phys. J. B*, 2005, 46:101-107.
- [17] Ashley G. Smart, Luis A. N. Amaral, Julio M. Ottino. Cascading failure and robustness in metabolic networks. *PNAS*, 2008, 105(36):13223-13228.
- [18] Chen-Tzu Kuo, Ming-Jen Hsu, Bing-Chang Chen, et al.. Denbinobin induces apoptosis in human lung adenocarcinoma cells via Akt inactivation, Bad activation, and mitochondrial dysfunction. *Toxicol Lett*, 2008, 177(1):48-58.