

BOSTON UNIVERSITY  
GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

**NETWORK THEORY  
AND  
ITS APPLICATIONS IN ECONOMIC SYSTEMS**

by

**XUQING HUANG**

B.A., University of Science and Technology of China, 2007  
M.A., Boston University, 2010

Submitted in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy  
2013

Approved by

First Reader

---

H. Eugene Stanley, Ph.D.  
Professor of Physics, William Fairfield Warren Distinguished Professor

Second Reader

---

Irena Vodenska, Ph.D.  
Assistant Professor of Administrative Sciences

*To the most wonderful parents in the whole world,*

*Mrs. Xiuxuan Xing and Mr. Ruquan Huang.*

# Acknowledgments

The successful completion of my doctoral degree requires the collaboration and contribution of many different people in their different ways in order to make this possible. It is with immense gratitude that I acknowledge the support and help of the following

First I would like to thank my thesis advisor, Professor H. Eugene Stanley, for accepting me into his scientific family, for unwavering financial support and perpetual encouragement. His work ethics and passion for pursuing scientific beauty has always stimulated me. His big heart for both science and people has promoted many cutting-edge interdisciplinary fields and encouraged countless collaborators and students. I am both proud and humbled to have worked with him.

I also want to thank Professor Shlomo Havlin, who has been kindly encouraging and guiding my research since the first day of my Ph.D. research. His support, guidance, advice throughout the research project, as well as his painstaking effort in proof reading the drafts, are greatly appreciated. Indeed, without his guidance, I would not be able to put the topic together.

I would also like to thank my collaborators, Dr. Fengzhong Wang, Professor Irena Vodenska, Professor Sergey Buldyrev, Professor Huijuan Wang, Dr. Jianxi Gao, Qian Li and Shuai Shao for interesting stimulating discussions and the pleasure and honor of publishing with them. I thank Bob Tomposki, Mirtha Cabello and Jerry Morrow for their crucial administrative help. I thank Erik Lascaris and Guoan Hu for all the technical support.

I am thankful to the Boston University Department of Physics faculty for all the re-

sources and the quality of education they have provided me. Thank you to my dissertation committee, Professor William Klein, William Skocpol, Emanuel Katz, Irena Vodenska and Eugene Stanley for their patience and help with the dissertation.

I must acknowledge all the friends who have supported me during my doctoral studies at Boston University. I thank Min Liu, Songbo Jin, Qian Li, Di Zhou and Zhiqiang Su for lifting me when I am down and celebrating with me when I am successful. There are not many six years in life, I will remember the happiness you brought to my life during these years. I also acknowledge all my classmates and colleagues in the lab for all these years.

Last but not least, none of this would ever have transpired without the unconditional love and support of my precious parents Mr. Ruquan Huang, Mrs. Xiuxuan Xing and my dearest fiancée Angela Tang. Their constant support, love and care and steadfast faith has gotten me where I am.

**NETWORK THEORY**  
**AND**  
**ITS APPLICATION IN ECONOMIC SYSTEMS**

(Order No.            )

**XUQING HUANG**

Boston University Graduate School of Arts and Sciences, 2013

Major Professor: H. Eugene Stanley, Professor of Physics

**ABSTRACT**

This dissertation covers the two major parts of my Ph.D. research: i) developing a theoretical framework of complex networks; and ii) applying statistical physics concepts and methods to quantitatively analyze complex systems.

In part I, we focus on developing theories of interdependent networks, which includes two chapters: 1) We develop a mathematical framework to study the percolation of interdependent networks under targeted attack and find that when the highly-connected nodes are protected and have lower probability to fail, in contrast to single scale-free (SF) networks where the percolation threshold  $p_c = 0$ , coupled SF networks are significantly more vulnerable, with  $p_c$  significantly larger than zero. 2) We analytically demonstrate that clustering, which quantifies the propensity for two neighbors of the same vertex to also be neighbors of each other, significantly increases the vulnerability of the system.

In part II, we apply concepts and methods developed in statistical physics to study economic systems, which also includes two chapters: 1) Centrality measures are widely studied in statistical physics to describe the importance of certain nodes in complex systems. We apply centrality measure concepts to study the influence of directors in the US corporate governance network, in which nodes represent directors and links between two directors represent their service on common company boards. 2) Using concepts of percolation and

cascading failure, we study the robustness of the interconnected banking system as a complex network. With empirical banks' balance sheet data in 2007 as input to the model, we find that our bipartite network model efficiently identifies a significant portion of the actual failed banks during the financial crisis between 2008 and 2011. The results suggest that complex network models could be useful for systemic risk stress testing of banking systems. We also find first-order-like phase transition in the banking system, which means that the banking system can experience an abrupt collapse under certain circumstances.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Random Networks and Properties . . . . .	3
1.2	Generating Function Method . . . . .	6
1.3	Interdependent Networks . . . . .	8
<b>2</b>	<b>Robustness of Interdependent Networks Under Targeted Attack</b>	<b>12</b>
2.1	Introduction . . . . .	12
2.2	Targeted Attack Model . . . . .	13
2.3	Interdependent Networks Model . . . . .	14
2.4	Mapping Method . . . . .	14
2.5	Results . . . . .	17
2.6	Summary . . . . .	21
<b>3</b>	<b>The Robustness Of Interdependent Clustered Networks</b>	<b>23</b>
3.1	Introduction . . . . .	23
3.2	Site Percolation of Single Clustered Networks . . . . .	25
3.3	Degree-Degree Correlation . . . . .	28
3.4	Percolation on Interdependent Clustered Networks . . . . .	30
3.5	Conclusion and Summary . . . . .	34
<b>4</b>	<b>Identifying Influential Directors in the United States Corporate Governance Network</b>	<b>36</b>

4.1	Introduction . . . . .	36
4.2	Data . . . . .	39
4.3	Directors' Network and Its Properties . . . . .	40
4.4	Influence Factor Measure . . . . .	41
4.5	Properties Of The Influence Factor . . . . .	43
4.6	Empirical Tests of Methods . . . . .	45
4.6.1	Test: Power and influence of female directors in the US corporate governance network . . . . .	45
4.6.2	Test: Power and influence of directors in the US corporate governance network for specific industries . . . . .	46
4.7	Conclusion . . . . .	48
<b>5</b>	<b>Cascading Failures in Bi-partite Graphs: Model for Systemic Risk Prop- agation</b>	<b>57</b>
5.1	Introduction . . . . .	57
5.2	Properties of Failed Banks. . . . .	59
5.3	Cascading Failure Propagation Model. . . . .	62
5.4	Empirical Test and Analysis. . . . .	64
5.5	Discussion . . . . .	70
5.6	Methods . . . . .	72
<b>6</b>	<b>Conclusion</b>	<b>76</b>
	<b>Bibliography</b>	<b>79</b>
	<b>Curriculum Vitae</b>	<b>91</b>

# List of Tables

5.1	Description of Commercial Banks - Balance Sheet Data (CBBSD) from Whar-	
	ton Research Data Services. . . . .	75

# List of Figures

1.1	Illustration of cascading failure in interdependent networks. . . . .	11
2.1	(a) Dependence of $p_c$ on $\alpha$ for SF single and interdependent networks with average degree $\langle k \rangle = 4$ . (b) Values of $p_c$ vs $\alpha$ for SF interdependent networks with different $\lambda$ and lower cut-off $m = 2$ . . . . .	18
2.2	Values of $p_\infty$ vs $p$ when both networks are initially attacked. . . . .	20
3.1	Size of giant component $g(p)$ in single networks with degree distribution Eq. (3.7) and average degree $\langle k \rangle = 4$ , as a function of $p$ , the fraction of remaining nodes after random removal of nodes. . . . .	27
3.2	Degree-degrees correlation as a function of the clustering coefficient. . . . .	30
3.3	(a) Size of mutually connected giant component as a function of cascading failure steps $n$ . (b) Size of giant component, $p_\infty$ , in interdependent networks as a function of $p$ . . . . .	32
3.4	(a) Size of giant component as a function of $p$ for fixed clustering coefficient. (b) Critical threshold $p_c$ as a function of average degree for different clustering coefficients. . . . .	33
4.1	Illustration of a subnetwork of the company-director bipartite network. . . . .	37
4.2	Illustration of a bipartite network and its “one-mode” projection. . . . .	50
4.3	Cumulative degree distribution of the network of directors. . . . .	51
4.4	Illustration of the distance between directors. . . . .	52

4.5	Demonstration of cumulative distribution of director's <i>influence factor</i> $I$ and $NIF \tilde{I}$ in different years. . . . .	53
4.6	Percentage of overlap of top directors according to influence factor between different realizaitons v.s. years. . . . .	54
4.7	Comparison between the <i>influence factor</i> $I$ , TCC and the existing centrality measures for a typical year 1999. . . . .	55
4.8	Comparison between the efficiency $\epsilon$ of the <i>influence factor</i> measure and other measures in identifying the most influential directors. . . . .	56
5.1	Bank-asset bipartite network model with banks as one node type and assets as the other node type. . . . .	60
5.2	Comparison of probability density functions (PDF) of weight of typical assets and equity ratios between all banks and FDIC listed failed banks for 2007. .	61
5.3	Fraction of survived banks after cascading failures as function of the initial loss of value of certain asset, with $\eta = 0$ . . . . .	65
5.4	ROC curves of the prediction of failed banks by our cascading failure model when the loans for construction and land development are initially shocked (top figures) and when the loans secured by nonfarm nonresidential properties are initially shocked (bottom figures). . . . .	67
5.5	ROC curves of predictions of failed banks by our cascading failure model when (a) loans secured by 1-4 family resid. properties and (b) agricultural loans are initially shocked respectively. . . . .	68
5.6	Survival rate of banks when asset 0 ( loans for construction and land development ) is initially shocked as function of one parameter with the other two parameters fixed. . . . .	70
5.7	Phase diagram for parameter $\alpha$ and $p$ , when $\eta = 0.26$ . . . . .	71

# List of Abbreviations

CBBSD	Commercial Banks-Balance Sheet Data
CDF	Cumulative probability Distribution Function
ER	Erdős Rényi
FDIC	Federal Deposit Insurance Corporation
FDL	Failed Bank List
IRRC	Investor Responsibility Research Center
NIF	Normalized Influence Factor
NOI	Number Of Iterations
PDF	Probability Distribution Function
ROC curve	Receiver Operating Characteristic curve
SF	Scale-Free
TCC	Total Capitalization of all the Companies

# Chapter 1

## Introduction

Networks are present in almost every aspect of our life, i.e. communication networks of telephones and cellular phones, the electrical power grid, computer communication networks in technological world; the network of friendship between individuals, working relations or common hobbies, and the network of business relations between persons and firms in the social and economic world.

This dissertation develops theoretical framework of complex networks and applies complex networks models to quantitatively analyze economics systems. This thesis is separated into two major parts: i) theoretical development of complex networks models, ii) application of complex networks models.

In part I, we focus on developing theories of interdependent networks. Due to technological progress, modern systems are becoming more and more mutually coupled and depend on each other to provide proper functionality [1–3]. For example, a power station network and a communication network are coupled, since nodes rely for power supply on the power stations, while the power stations depend for their control on the communication nodes [4]. While in the past the study of single networks has been dominant in the networks field, the question of robustness of interdependent networks has recently become of large interest [5–8]. In interdependent networks, nodes from one network depend on nodes from another network and vice versa. Consequently, when nodes from one network fail they cause nodes in the other network to fail too. When some initial failure of nodes happens,

this may trigger a recursive process of cascading failures that can completely fragment both networks. This part includes two chapters: 1) We develop a mathematical framework to study the percolation of interdependent networks under targeted-attack and find that when the highly connected nodes are protected and have lower probability to fail, in contrast to single scale-free (SF) networks where the percolation threshold  $p_c = 0$ , coupled SF networks are significantly more vulnerable with  $p_c$  significantly larger than zero. 2) We analytically demonstrate that clustering, which quantifies the propensity for two neighbours of the same vertex to also be neighbours of each other, significantly increases the vulnerability of the system.

In part II, we apply the complex network models to study economic systems. In the wake of the recent global financial crisis, increased attention has been given to the study of the dynamics of economic systems and to systemic risk in particular. The widespread impact of the current EU sovereign debt crisis and the 2008 world financial crisis show that as economic systems become increasingly interconnected, local exogenous or endogenous shocks can provoke global cascading system failure that is difficult to reverse and that cripples the system for a prolonged period of time. Thus policy makers are compelled to create and implement safety measures that can prevent cascading system failures or soften their systemic impact. Based on the success of complex networks in modelling interconnected systems, applying complex network theory to study economical systems has been under the spot light [9–14]. This part includes two chapters: 1) We study the US corporate governance network, in which nodes representing directors and links between two directors representing their service on common company boards, and proposed a quantitative influence measure to identify the most influential directors in the network. 2) We propose a bipartite networks model to simulate the risk propagation process during financial crisis. With empirical bank's balance sheet data in 2007 as input to the model, we find that our model efficiently identifies a significant portion of the actual failed banks reported by Federal Deposit Insurance Corporation during the financial crisis between 2008 and 2011. The results suggest that this model could be useful for systemic risk stress testing for financial

systems. The model also identifies that commercial rather than residential real estate assets are major culprits for the failure of over 350 US commercial banks during 2008 - 2011.

The rest of this introduction will first present some basic definitions of complex networks and its properties. Then we introduce the generating function method which is extensively used in this thesis for deriving percolation properties of networks. At last, we will introduce the framework and some developments in interdependent networks research.

## 1.1 Random Networks and Properties

The mathematical concepts describing networks are graphs. Graphs represent the essential topological properties of a network by treating the network as a collection of nodes and links among nodes. The number of links that a node is attached is defined as the degree of this node. Before 1960 the work on graph theory has mainly dealt with the properties of special individual graphs. In the 1960s, Paul Erdős and Alfred Rényi initiated a systematic study of random graphs [15–17]. Random graph theory is, in fact, not the study of individual graphs, but the study of a statistical ensemble of graphs. A property is said to exist for a class of graphs if the fraction of graphs not having it in the ensemble is of zero measure. An interesting characteristic of the ensemble is that many of its properties have a related threshold  $p_c(N)$ , such that the property exists, in the thermodynamic limit of  $N \rightarrow \infty$ , with probability 0 if  $p < p_c$ , and with probability 1 if  $p > p_c$ . This phenomenon is the same as the physical concept of a percolation phase transition. For example, the giant component of a network, i.e. a set of connected nodes, is such a property.

Two types of random graphs are mostly widely studied: Erdős Rényi networks (ER networks) and scale-free (SF networks) networks. The Erdős Rényi networks are firstly studied by Paul Erdős and Alfred Rényi in 1960s. It has been traditionally the dominant subject of study in the field of random graphs. A well-known result for the ER networks is that the degree distribution is Poissonian,

$$P(k) = e^{-z} z^k / k!, \quad (1.1)$$

where  $z = \langle k \rangle$  is the average degree. The critical threshold for a giant component to exist in ER networks is  $\langle k \rangle > 1$ . If  $\langle k \rangle < 1$  only small components exist, and size of the largest component is proportional to  $\log N$  [16]. Direct measurement of the degree distribution for real networks, such as the Internet [18, 19], WWW [20], e-mail networks [21], citations of scientific articles [22], metabolic networks [23], airline networks [24], neuronal networks [25] and many more, show that Poisson distribution does not apply. Instead, these networks exhibit a power-law degree distribution,

$$P(k) = ck^{-\gamma}, k = m, \dots, K, \quad (1.2)$$

where  $c \approx (\gamma - 1)m^{\gamma-1}$  is a normalization factor, and  $m$  and  $K$  are the lower and upper cutoffs for the degree of a node. The divergence of moments higher than  $(\gamma-1)$  is responsible for many of the anomalous properties attributed to scale-free networks.

Degree distribution alone is not enough to characterize the network. There are other quantities, such as the closeness (average distance between each pair of nodes), degree-degree correlation (between connected nodes), betweenness centrality and clustering coefficient. Since these properties are studied in this thesis, here we introduce the brief definition of these properties of networks.

## Distance

In a network, distance between two nodes is defined as the length of the shortest path between these two nodes. To find the distance from a origin node to all the other nodes, we just need to build a minimum spanning tree from the network. The distance from a node to the origin node is the number of steps spanning the tree when this node is reached.

## Degree-degree correlation

Degree-degree correlation is defined as the degree correlation of all pairs of linked nodes,

$$\rho_D = \frac{\sum_{\text{all links}} (k_l - \langle k \rangle)(k_r - \langle k \rangle)}{\sigma_k^2}, \quad (1.3)$$

where  $k_l$  and  $k_r$  are the degree of nodes at the two ends of a link,  $\langle k \rangle$  is the average degree of nodes in the network and  $\sigma_k$  is the standard deviation of degrees of nodes. When degree-degree correlation is larger than zero, that means high (low) degree nodes tend to connect with high (low) degree nodes. This case is called assortative. Otherwise, the system is disassortative.

### Betweenness centrality

Betweenness is a centrality measure of a vertex within a graph. Betweenness centrality quantifies the number of times a node is on the shortest path between two other nodes. The betweenness of a vertex  $i$  is dened to be the fraction of shortest paths between pairs of vertices in a network that pass through  $i$ . If, as is frequently the case, there is more than one shortest path between a given pair of vertices, then each such path is given equal weight such that the weights sum to unity. To be precise, suppose that  $g_i(st)$  is the number of geodesic paths from vertex  $s$  to vertex  $t$  that pass through  $i$ , and suppose that  $n_{st}$  is the total number of geodesic paths from  $s$  to  $t$ . Then the betweenness of vertex  $i$  is [26],

$$b_i = \frac{\sum_{s < t} g_i^{st}}{n(n-1)/2}, \quad (1.4)$$

where  $n$  is the total number of nodes in the network. Betweenness centrality can be regarded as a measure of the extent to which a node has control over information owing between others.

### Clustering

Clustering quantifies the propensity for two neighbours of the same vertex to also be neighbors of each other, forming triangle-shaped configurations in the network [27–29]. The clustering is high if two nodes sharing a neighbor have a high probability to be connected to each other. There are two common definitions of clustering. The first is global,

$$c = \frac{3 \times (\text{number of triangles in network})}{\text{number of connected triples}}, \quad (1.5)$$

where a “connected triple” means a single vertex with edges running to an unordered pair of other vertices. A second definition of clustering is based on the average of the clustering for single nodes. The clustering for a single node is the fraction of pairs of linked neighbors out of the total number of pairs of neighbors,

$$c_i = \frac{\text{number of triangles connected to vertex } i}{\text{number of triples centered on vertex } i}. \quad (1.6)$$

Then the clustering coefficient for the whole network is the average of  $c_i$ .

### Bonacich power centrality

Bonacich argued that a node’s centrality in a network should depend on three criteria: 1) the number of links to other nodes; 2) the intensity of the link; and 3) the centrality of those with whom one is linked. If we have the adjacent matrix  $M$  for a network, then the Bonacich centrality for a node  $i$  is defined as [30]

$$B_i = \frac{1}{\lambda} \sum_{j=1}^N M_{ij} B_j, \quad (1.7)$$

where  $N$  is the size of the network,  $\lambda$  is the largest eigenvalue of the matrix  $M$ .

Later, Bonacich interpret the power centrality [31] in another way, in terms of random walks that have a fixed probability of dying per step. The power centrality of vertex  $i$  is the expected number of times such a walk passes through  $i$ , averaged over all possible starting points for the walk.

## 1.2 Generating Function Method

Generating function method is very powerful and commonly used to describe random graphs with arbitrary degree distributions. Newman Strogatz and Watts comprehensively introduced this technique in ref. [32]. Here we introduce the results that we can use. Suppose that we have a unipartite undirected graph of  $N$  vertices, with  $N$  large. We denote the

generating function of  $G_0(x)$  for the probability distribution of node degree  $k$  as

$$G_0(x) = \sum_{k=0}^{\infty} P(k)x^k, \quad (1.8)$$

where  $P(k)$  is the probability that a randomly chosen node on the graph has degree  $k$ . The probability  $P(k)$  is given by the  $k^{\text{th}}$  derivative of  $G_0$  according to

$$p_k = \frac{1}{k!} \left. \frac{d^k G_0}{dx^k} \right|_{x=0}. \quad (1.9)$$

Thus the generating function encapsulates all the information contained in the discrete probability distribution  $P(k)$ .

If the distribution of a property  $k$  of an object is generated by a given generating function, then the distribution of the total of  $k$  summed over  $m$  independent realizations of the object is generated by the  $m^{\text{th}}$  power of that generating function. As an example, considering the generating function of the total degree of two nodes:

$$\begin{aligned} [G_0(x)]^2 &= \left[ \sum_k P(k)x^k \right] = \sum_{jk} P(k)P(j)x^{j+k} \\ &= P(0)P(0)x^0 + (P(0)P(1) + P(1)P(0))x^1 \\ &\quad + (P(0)P(2) + P(1)P(1) + P(2)P(0))x^2 \\ &\quad + (P(0)P(3) + P(1)P(2) + P(2)P(1) + P(3)P(0))x^3 + \dots \end{aligned} \quad (1.10)$$

It is easy to see that the coefficient of the power of  $x^n$  in this expression is precisely the sum of all products  $P(i)P(k)$  such that  $j+k=n$ , which correctly gives the probability that the sum of the degrees of the two vertices will be  $n$ .

If we start at a randomly chosen node and follow each of the edges at that vertex to reach the  $k$  nearest neighbors, then the nodes arrived at each have the distribution of remaining outgoing edges generated by  $G_1(x)$ . The generating function is

$$G_1(x) = \frac{G'_0(x)}{G'_0(1)} = \frac{1}{\langle k \rangle} G'_0(x), \quad (1.11)$$

where  $\langle k \rangle$  is the average node degree.

When a giant component exists in a network, the generating function formalism still works. And the size of the giant component ( $S$ ), which is the fraction of the graph occupied by the giant component, can be found by

$$\begin{aligned} S &= 1 - G_0(u), \\ u &= G_1(u). \end{aligned} \tag{1.12}$$

And the phase transition, when the giant component in a network first to emerge, is at the point

$$G'_1(1) = 1. \tag{1.13}$$

### 1.3 Interdependent Networks

Before 2010, after many years of intense study on networks, almost all work done has concentrated on the limited case of a single network which does not interact with other networks. Such situations rarely, if ever, occur in nature. Just as in the case of idealized gas, when interactions are present as in nature, new physical laws appear. In 2010, Buldyrev and etc. [7] firstly proposed a mathematical framework to study the robustness of interdependent networks. This mathematical framework is the basis of the theoretical research on interdependent networks in this thesis.

The interdependent networks model consists of two networks A and B, and assume that the functioning of a node in network A depends on the ability of one or more nodes in network B to supply a critical resource to the node in network A. Similarly, a node in network B depends on a set of nodes in network A. The networks can be connected in different ways; in the most general configuration one could specify the distributions of connections between the nodes from both networks. The networks can have the same, or different, typologies. The model can easily be extended to an arbitrary number of interacting networks each with its own specific topology and dependence on the other networks. For example, an interesting dependence for three interacting networks could be a circular dependency in

which the nodes in network B depend on network A for a resource, the nodes of network C depend on the nodes of network B for a resource and the nodes of network A depend on network C for resources.

The giant component and critical threshold of interdependent networks can be solved analytically using the apparatus of generating functions. We assume the nodes in two interdependent networks are one to one correspondent. If one node fails, the corresponding node fails. The generating functions of degree distribution of networks A and B are defined as

$$\begin{aligned} G_{A0}(x) &= \sum_k P_A(k)x^k \\ G_{B0}(x) &= \sum_k P_B(k)x^k. \end{aligned} \tag{1.14}$$

The generating functions of the underlining branching process, Eq.1.11, are

$$\begin{aligned} G_{A1}(x) &= G'_{A0}(x)/G'_{A0}(1), \\ G_{B1}(x) &= G'_{B0}(x)/G'_{B0}(1). \end{aligned} \tag{1.15}$$

Random removal of fraction  $1 - p$  of nodes will change the degree distribution of the remaining nodes [33]

$$\begin{aligned} G_{A0}(x, p) &= G_{A0}(1 - p(1 - x)), \\ G_{B0}(x, p) &= G_{B0}(1 - p(1 - x)), \\ G_{A1}(x, p) &= G_{A1}(1 - p(1 - x)), \\ G_{B1}(x, p) &= G_{B1}(1 - p(1 - x)). \end{aligned} \tag{1.16}$$

According to Eq. 1.12, the giant component of the network A after removing  $1 - p$  fraction of nodes is

$$p_A(p) = 1 - G_{A0}(f_A, p), \tag{1.17}$$

where  $f_A$  satisfies equation

$$f_A(p) = G_{A1}(f_A, p). \tag{1.18}$$

Next, we present the formalism for the cascade process step by step, as shown in fig. 1.1. On the first step, after the initial attack which removes  $(1 - p)$  fraction of nodes (the red part in network  $A$  in fig. 1.1) from networks  $A$ , the remaining network is named  $A_0$ , with a size of  $\psi'_1 = p$  fraction of the network  $A$ , number of nodes  $N_0 = pN$ . The orange part in fig.1.1 are the fraction of nodes detached from the giant component after the red part is removed. The giant component of the remaining network is thus of size  $\psi_1 = \psi'_1 p_A(\psi'_1)$ , number of nodes is  $N_1 = N_0 p_A(\psi'_1)$ . In the second stage of the cascading failure, the failed node in network  $A$  cause their corresponding nodes in network  $B$  to fail. Thus  $(1 - \psi_1)$  fraction of nodes in network  $B$  fail, which cause the green part disconnected from the giant component. The remaining fraction of network  $B$  is  $\phi'_1 = \psi_1$ , and the fraction of nodes in the giant component of network  $B$  is  $\phi_1 = \phi'_1 p_B(\phi'_1)$ , number of nodes  $N_2 = N_1 p_B(\phi'_1)$ . In the third step of the cascading failure, the nodes in the giant component of network  $A$  corresponding to the green part in network  $B$  is removed. The removal of these nodes from  $\psi_1$  causes the purple part in network  $A$  to be disconnected from the giant component, which leaves a giant component size of  $\psi_2$ . That is equivalent to the removal of the same fraction of nodes from  $A_0$ . To achieve the same giant component size  $\psi_2$ , the total number of nodes that must be removed from network  $A$  is  $(1 - p_B(\psi_1))N_0$  nodes from  $A_0$  plus the number of the initially attacked nodes  $(1 - p)N$  (green and red part in the right bottom figure in fig. 1.1), which causes the blue part to be detached from the giant component. Thus, the total number of nodes that must be removed from network  $A$  is  $(1 - pp_B(\psi_1))N$ . Hence the third-stage failure is equivalent to a random attack in which  $p$  is replaced by  $\phi'_1 = pp_B(\psi_1)$ . The following steps repeats the above procedure.

Following this approach, we can construct the sequence of giant components of networks,

$$\psi'_1 = p, \quad \psi_1 = \psi'_1 p_A(\psi'_1), \quad (1.19)$$

$$\phi'_1 = pp_A(\psi'_1), \quad \phi_1 = \phi'_1 p_B(\phi'_1), \quad (1.20)$$

$$\psi'_2 = pp_B(\phi'_1), \quad \psi_2 = \psi'_2 p_A(\psi'_2), \dots \quad (1.21)$$

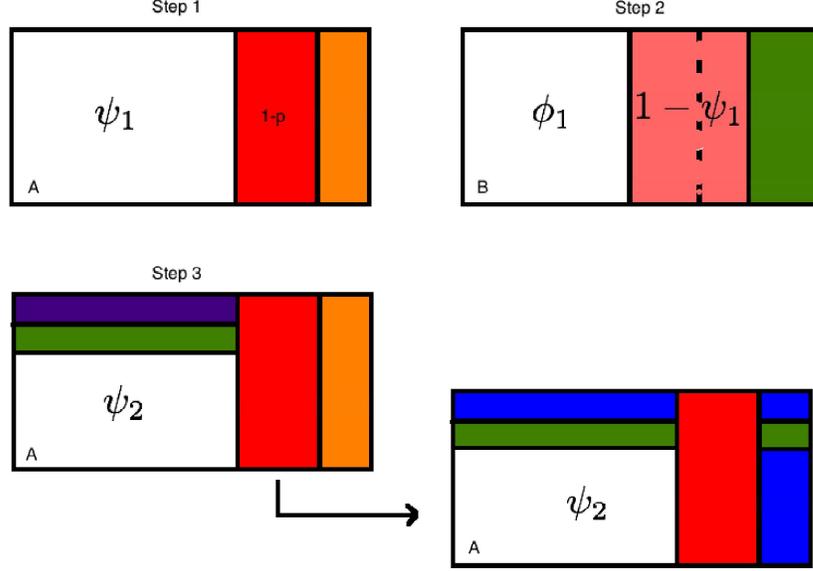


Figure 1.1: Illustration of cascading failure in interdependent networks.

At the end of the cascade process,  $\psi'_n = \psi'_{n+1}$ ,  $\phi'_n = \phi'_{n+1}$ . Thus we arrive to a system of

$$x = pp_A(y), y = pp_B(x), \quad (1.22)$$

which also gives equation

$$x = pp_A(pp_B(x)). \quad (1.23)$$

This equation can be solved graphically. After we find  $x$ , the size of the mutually connected giant component of the interdependent networks can be found by  $S = xp_B(x)$ . The critical condition when the giant component of the interdependent networks exists is  $1 = p^2 p'_A(pp_B(x)) p'_A(x)$ , where  $p'_A(x)$  is the derivative of function  $p_A(x)$ .

## Chapter 2

# Robustness of Interdependent Networks Under Targeted Attack

### 2.1 Introduction

Modern systems due to technological progress are becoming more and more mutually coupled and depend on each other to provide proper functionality [2–4]. For example, blackouts are usually caused by cascading failures between the power grid and its communication support system [4]. While cascade of failures in one network, e.g., overload failure, can cause dramatic damage to a system [34, 35], social disruptions caused by recent disasters, ranging from hurricanes to large-scale power outages and terrorist attacks, have shown that the most dangerous vulnerability is hiding in the many interdependencies across different networks [1, 36]. The question of robustness of interdependent networks has recently become of interest [5–8]. In interdependent networks, nodes from one network depend on nodes from another network and vice versa. Consequently, when nodes from one network fail they cause nodes in the other network to fail, too. When some initial failure of nodes happens, this may trigger a recursive process of cascading failures that can completely fragment both networks.

Recently, a theoretical framework was developed [7] to study the process of cascading failures in interdependent network caused by *random* initial failure of nodes. They show that due to the coupling between networks, interdependent networks are extremely vulnerable

to random failure. However, when we consider real scenarios, initial failure is mostly not random. It may be due to a *targeted attack* on important central nodes. It can also occur to low central nodes because important central nodes are purposely defended, e.g. in internet networks, heavily connected hubs are purposely more secured. Indeed, it was shown that targeted attacks on high degree nodes [37–41] or high betweenness nodes [42, 43] in *single* networks have dramatic effect on their robustness. The question of robustness of *interdependent* networks under targeted attack or defence has not been addressed.

## 2.2 Targeted Attack Model

In this chapter, we develop a mathematical framework for understanding the robustness of interdependent networks under initial targeted attack which depends on degree of nodes. The framework is based on a general technique we develop to solve targeted attack problems in networks by mapping them to random attack problems. A value  $W_\alpha(k_i)$  is assigned to each node, which represents the probability that a node  $i$  with  $k_i$  links is initially attacked and become inactive. We focus on the family of functions [40]

$$W_\alpha(k_i) = \frac{k_i^\alpha}{\sum_{i=1}^N k_i^\alpha}, \quad -\infty < \alpha < +\infty. \quad (2.1)$$

When  $\alpha > 0$ , nodes with higher degree are more vulnerable and those nodes are intentionally attacked, while for  $\alpha < 0$ , nodes with higher degree are defended and so have lower probability to fail. The case  $\alpha = 0$ ,  $W_0 = \frac{1}{N}$ , represents the random removal of nodes [7] and the case  $\alpha \rightarrow \infty$  represents the targeted attack case where nodes are removed strictly in the order from high degree to low degree. For the  $\alpha < 0$  case, nodes with zero degree should be removed before analysis begins. An important special case  $\alpha = 1$  corresponds to the acquaintance immunization strategy [44].

## 2.3 Interdependent Networks Model

Our model consists of two networks, A and B, with the same number of nodes,  $N$ . The  $N$  nodes in each network are connected to nodes in the other network by bidirectional dependency links, thereby establishing a one-to-one correspondence. The functioning of a A-node in network A depends on the functioning of the corresponding B-node in network B and vice versa. Within each network, the nodes are randomly connected with degree distributions  $P_A(k)$  and  $P_B(k)$  respectively. We begin by studying the situation where only network A is attacked. We initially remove a fraction,  $1 - p$ , of the A-nodes of network A with probability  $W_\alpha(k_i)$  (Eq.(2.1)) and remove all the A-links that connect to those removed nodes. As nodes and links are sequentially removed, network A begins to fragment into connected components. Nodes that are not connected to the giant component are considered inactive and are removed. Owing to the dependence between the networks, all the B-nodes in network B that are connected to the removed A-nodes in network A are then also removed. Network B also begins to fragment into connected components and only the nodes in the giant component are kept. Then network B spreads damage back to network A. The damage is spreaded between network A and B, back and forth until they completely fragment or arrive to a mutually connected component and no further removal of nodes and links occurs.

## 2.4 Mapping Method

The main idea of our approach is to find an equivalent network  $A'$ , such that the *targeted* attack problem on interdependent networks  $A$  and  $B$  can be solved as a *random* attack problem on interdependent networks  $A'$  and  $B$ . We start by finding the new degree distribution of network  $A$  after removing, according to Eq.(2.1),  $1 - p$  fraction of nodes but before the links of the remaining nodes which connect to the removed nodes are removed. Let  $A_p(k)$  be the number of nodes with degree  $k$  and  $P_p(k)$  be the new degree distribution

of the remaining fraction  $p$  of nodes in network  $A$ ,

$$P_p(k) = \frac{A_p(k)}{pN}. \quad (2.2)$$

When another node is removed,  $A_p(k)$  changes as

$$A_{(p-1/N)}(k) = A_p(k) - \frac{P_p(k)k^\alpha}{\langle k(p)^\alpha \rangle}, \quad (2.3)$$

where  $\langle k(p)^\alpha \rangle \equiv \sum P_p(k)k^\alpha$ . In the limit of  $N \rightarrow \infty$ , Eq.(2.3) can be presented in terms of derivative of  $A_p(k)$  with respect to  $p$ ,

$$\frac{dA_p(k)}{dp} = N \frac{P_p(k)k^\alpha}{\langle k(p)^\alpha \rangle}. \quad (2.4)$$

Differentiating Eq.(2.2) with respect to  $p$  and using Eq.(2.4), we obtain

$$-p \frac{dP_p(k)}{dp} = P_p(k) - \frac{P_p(k)k^\alpha}{\langle k(p)^\alpha \rangle}, \quad (2.5)$$

which is exact for  $N \rightarrow \infty$ . In order to solve Eq.(2.5), we define a function  $G_\alpha(x) \equiv \sum_k P(k)x^{k^\alpha}$ , and substitute  $f \equiv G_\alpha^{-1}(p)$ . We find by direct differentiation that [45]

$$P_p(k) = P(k) \frac{f^{k^\alpha}}{G_\alpha(f)} = \frac{1}{p} P(k) f^{k^\alpha}, \quad (2.6)$$

$$\langle k(p)^\alpha \rangle = \frac{f G'_\alpha(f)}{G_\alpha(f)}, \quad (2.7)$$

satisfy the Eq.(2.5). With this degree distribution, the generating function of the nodes left in network  $A$  before removing the links to the removed nodes is

$$G_{Ab}(x) \equiv \sum_k P_p(k)x^k = \frac{1}{p} \sum_k P(k) f^{k^\alpha} x^k. \quad (2.8)$$

Because network  $A$  is randomly connected, the probability of a link emanating from a remaining node is equal to the ratio of the number of links emanating from the remaining

nodes to the total number of links emanating from all the nodes of the original network:

$$\tilde{p} \equiv \frac{pN \langle k(p) \rangle}{N \langle k \rangle} = \frac{\sum_k P(k) k f^{k^\alpha}}{\sum_k P(k) k}, \quad (2.9)$$

where  $\langle k \rangle$  is the average degree of the original network  $A$ ,  $\langle k(p) \rangle$  is the average degree of remaining nodes before the links that are disconnected are removed. Removing the links which connect to the deleted nodes of a randomly connected network is equivalent to randomly removing a  $(1 - \tilde{p})$  fraction of links of the remaining nodes. Using the same approach as in ref. [33], one can show that the generating function of the remaining nodes after random removal of  $(1 - \tilde{p})$  fraction of links is equal to the original distribution of the network with a new argument  $z = 1 - \tilde{p} + x\tilde{p}$ . Thus the generating function of the new degree distribution of the nodes left in network  $A$  after their links to the removed nodes are also removed is

$$G_{Ac}(x) \equiv G_{Ab}(1 - \tilde{p} + \tilde{p}x). \quad (2.10)$$

The only difference in the cascading process under *targeted* attack from the case under *random* attack is the first stage where the initial attack is exerted on the network  $A$ . If we find a network  $A'$  with generating function  $\tilde{G}_{A0}(x)$ , such that after a random attack with  $(1 - p)$  fraction of removed, the generating function of nodes left in  $A'$  is the same as  $G_{Ac}(x)$ , then the targeted attack problem on interdependent networks  $A$  and  $B$  can be solved as a random attack problem on interdependent networks  $A'$  and  $B$ . We find  $\tilde{G}_{A0}(x)$  by solving the equation  $\tilde{G}_{A0}(1 - p + px) = G_{Ac}(x)$  and from Eq.(2.10),

$$\tilde{G}_{A0}(x) = G_{Ab}\left(1 + \frac{\tilde{p}}{p}(x - 1)\right). \quad (2.11)$$

Up to now, we have mapped the problem of cascade of failures of nodes in interdependent networks caused by an initial targeted attack to the problem of a random attack. Since the derivation of equations only depends on the generating function of network  $A$ , this approach can be generally applied to study both single networks with dependency links [46] and other more general interdependent network models, as long as the nodes in those networks are

randomly connected.

## 2.5 Results

Next we can apply the framework developed in Ref. [7],  $g_A(p) = 1 - \tilde{G}_{A0}[1 - p(1 - f_A)]$ , where  $f_A$  is a function of  $p$  that satisfies the transcendental equation  $f_A = \tilde{G}_{A1}[1 - p(1 - f_A)]$ . Analogous equations exist for network B. As the interdependent networks achieve a mutually connected giant component, the fraction of nodes left in giant component is  $p_\infty$ . The system satisfies the equations

$$\begin{aligned} x &= pg_A(y), \\ y &= pg_B(x), \end{aligned} \tag{2.12}$$

where the two unknown variables  $x$  and  $y$  satisfy  $p_\infty = xg_B(x) = yg_A(y)$ . Eliminating  $y$  from these equations, we obtain a single equation

$$x = pg_A[pg_B(x)]. \tag{2.13}$$

The critical case ( $p = p_c$ ) emerges when both sides of this equation have equal derivatives,

$$1 = p^2 \frac{dg_A}{dx}[pg_B(x)] \frac{dg_B}{dx}(x)|_{x=x_c, p=p_c}. \tag{2.14}$$

which, together with Eq.(2.13), yields the solution for  $p_c$  and the critical size of the giant mutually connected component,  $p_\infty(p_c) = x_c g_B(x_c)$ . In general, there is no explicit expression as a solution and  $p_c$  and  $x_c$  can be found numerically.

We now analyze the specific classes of Erdős-Rényi (ER) [15, 16] and scale-free (SF) [18, 47–49] networks. The lines in Fig. 2.1 represent the critical thresholds,  $p_c$ , for coupled SF networks with different  $\alpha$  obtained by solutions of Eq.(2.13) and Eq.(2.14), which are in excellent agreement with simulations. Several conclusions from Fig. 2.1 are as follows: (i) Remarkably, while  $p_c$  for a single SF network approaches to 0 quickly when  $\alpha$  becomes zero or negative (see also [40]),  $p_c$  for interdependent networks is non-zero for the entire range of  $\alpha$  (Fig. 2.1(a)). This follows from the fact that failure of the least connected

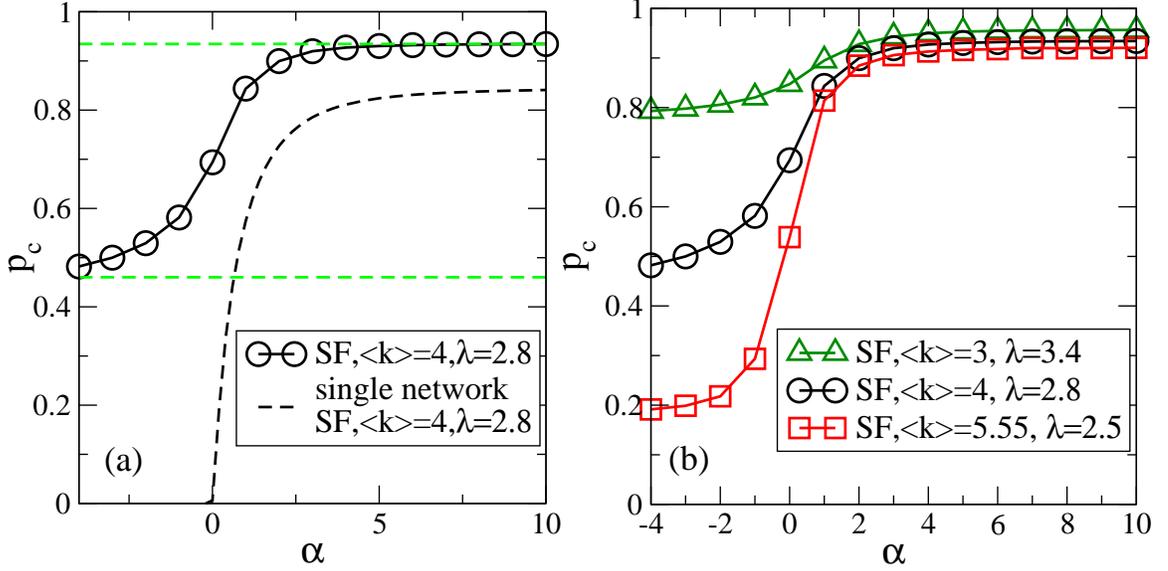


Figure 2.1: (a) Dependence of  $p_c$  on  $\alpha$  for SF single and interdependent networks with average degree  $\langle k \rangle = 4$ . The lower cut-off of the degree is  $m = 2$ . The horizontal lines represent the upper and lower limits of  $p_c$ . The black dashed line represents  $p_c$  for SF free network. (b) Values of  $p_c$  vs  $\alpha$  for SF interdependent networks with different  $\lambda$  and lower cut-off  $m = 2$ . The  $\lambda$  in the legends of both the graphs are approximate numbers.

nodes in one network may lead to failure of well connected nodes in the other network, which makes interdependent networks significantly more difficult to protect compared to a single network. (ii) targeted attacks ( $\alpha > 0$ ) and defense strategies ( $\alpha < 0$ ) are more effective for interdependent networks with broader degree distributions. In Fig. 2.1(b), comparing the lines of  $\lambda = 2.5$ ,  $\lambda = 2.8$  and  $\lambda = 3.4$  with  $m = 2$ , one can see that the lower is  $\lambda$  the more sensitive is  $p_c$  to the change of  $\alpha$ . Accordingly, robustness of interdependent networks with broader degree distributions decreases more under the same targeted attacks.

Simplified forms for  $G_{Ab}(x)$ ,  $G_{Ac}(x)$  and  $\tilde{G}_{A0}(x)$  from Eqs.(2.8),(2.10) and (2.11) exist when  $\alpha = 1$ ,

$$G_{Ab}(x) = \frac{1}{p} \sum_k P(k) f^k x^k = \frac{1}{p} G_{A0}(fx), \quad (2.15)$$

$$G_{Ac}(x) = \frac{1}{p} G_{A0}(f(1 - \tilde{p} + \tilde{p}x)), \quad (2.16)$$

$$\tilde{G}_{A0}(x) = \frac{1}{p} G_{A0}\left(\frac{\tilde{p}}{p} f(x-1) + f\right). \quad (2.17)$$

where  $G_{A0}(x)$  is the original generating function of the network A,  $f = G_{A0}^{-1}(p)$  and  $\tilde{p} = \frac{G'_{A0}(f)}{G'_{A0}(1)}f$ .

Explicit solutions of percolation quantities exist for the case of interdependent Erdős-Rényi networks, when  $\alpha = 1$  and both of the two networks are initially attacked simultaneously. The two networks originally have generating functions  $G_{A0}(x)$  and  $G_{B0}(x)$ . Initially,  $(1 - p_1)$  and  $(1 - p_2)$  fraction of nodes are targeted (according to Eq. (1) and  $\alpha = 1$ ) and removed from network A and B respectively. Similarly, we start by finding the equivalent networks  $A'$  and  $B'$  such that a fraction  $(1 - p_1p_2)$  of random initial attack on both of the networks has the same effect as  $(1 - p_1)$  and  $(1 - p_2)$  fraction of nodes are intentionally removed from network A and network B respectively. After removal of initially failed nodes and all the links that connect to the removed nodes, according to Eq.(2.16), the generating function of the nodes left in network A is

$$G_{Ac}(x) = \frac{1}{p_1}G_{A0}(f_1(1 - \tilde{p}_1 + \tilde{p}_1x)), \quad (2.18)$$

where  $f_1 \equiv G_{A0}^{-1}(p_1)$ ,  $\tilde{p}_1 \equiv f_1 \frac{G'_{A0}(f_1)}{G'_{A0}(1)}$ . Furthermore,  $(1 - p_2)$  fraction of the remaining A-nodes are randomly removed. Because each remaining A-node's corresponding B-node in network B has a possibility  $(1 - p_2)$  to be initially attacked, which leads to fail this A-node. The generating function of the nodes left in network A is

$$G_{Ad}(x) \equiv G_{Ac}(1 - p_2 + p_2x) = \frac{1}{p_1}G_{A0}(f_1 + \tilde{p}_1f_1p_2(x - 1)). \quad (2.19)$$

Now we can find the generating function of the equivalent network  $A'$  by  $\tilde{G}_{A0}(1 - p_1p_2 + p_1p_2x) = G_{Ad}(x)$ :

$$\tilde{G}_{A0}(x) = \frac{1}{p_1}G_{A0}\left(\frac{\tilde{p}_1}{p_1}f_1(x - 1) + f_1\right). \quad (2.20)$$

The same holds for network  $B'$ .

For ER networks, the generating function is  $G_0(x) = e^{\langle k \rangle(x-1)}$  [33], so  $f_1 = \frac{\ln(p_1)}{\langle k \rangle_1} + 1, f_2 = \frac{\ln(p_2)}{\langle k \rangle_2} + 1, \tilde{G}_{A0}(x) = \tilde{G}_{A1}(x) = e^{\langle k \rangle_1 f_1^2(x-1)}$  and  $\tilde{G}_{B0}(x) = \tilde{G}_{B1}(x) = e^{\langle k \rangle_2 f_2^2(x-1)}$ . From

Eq.(2.12),

$$\begin{aligned} x &= p_1 p_2 g_A(y) = p_1 p_2 (1 - f_A), \\ y &= p_1 p_2 g_B(x) = p_1 p_2 (1 - f_B), \end{aligned} \quad (2.21)$$

where

$$\begin{aligned} f_A &= e^{\langle k \rangle_1 f_1^2 y (f_A - 1)}, \\ f_B &= e^{\langle k \rangle_2 f_2^2 x (f_B - 1)}. \end{aligned} \quad (2.22)$$

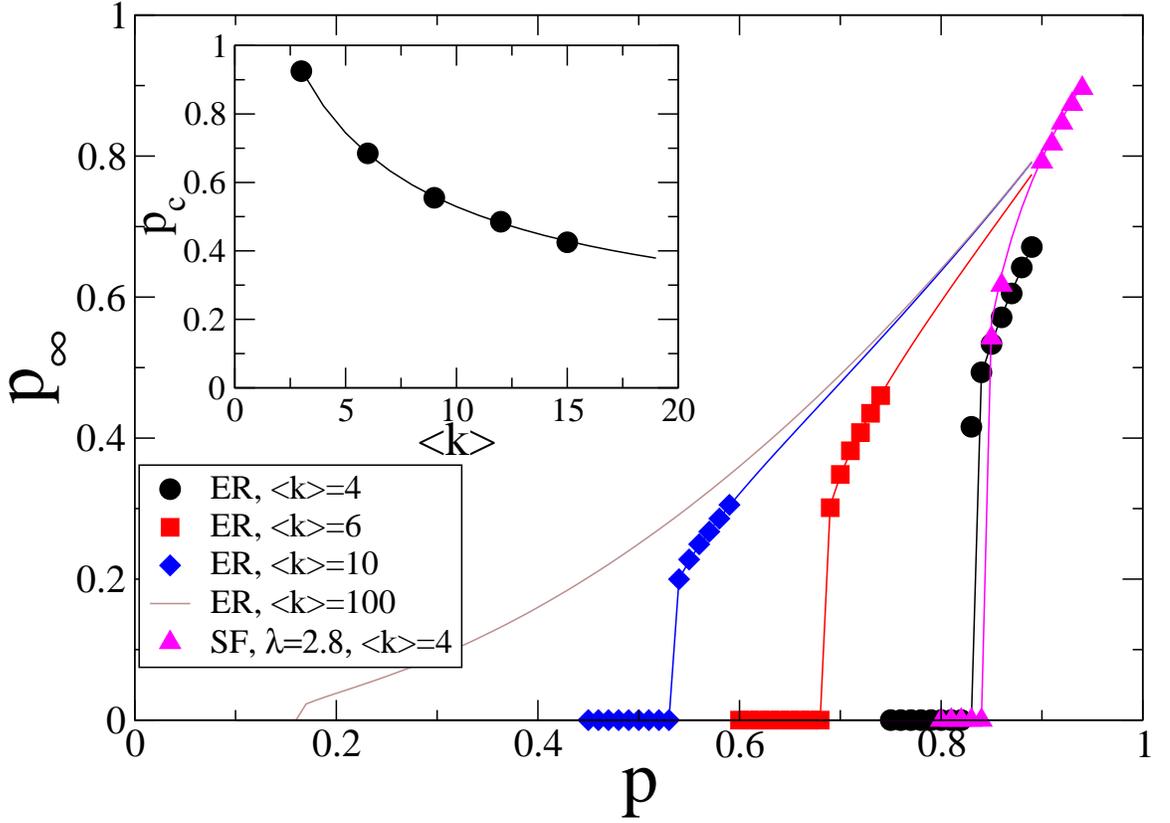


Figure 2.2: Values of  $p_\infty$  vs  $p$  when both networks are initially attacked. Both networks in the interdependent networks are ER or SF networks with the same average degree. The symbols represent simulation data ( $N = 10^6$  nodes). The solid lines are theoretical predictions, Eq.(2.23). The dashed line represents simulation data for interdependent scale-free networks with  $\lambda = 2.8$ ,  $\langle k \rangle = 4$ . All results are for  $\alpha = 1$ . Inset: Values of  $p_c$  vs average degree of ER networks. The symbols represent simulation data, while the line is the theory, Eq.(2.24).

In the case  $\langle k \rangle_1 = \langle k \rangle_2 = \langle k \rangle$  and  $p_1 = p_2 = p$ , we find that

$$p_\infty = p^2 (1 - e^{\langle k \rangle f^2 p_\infty}). \quad (2.23)$$

where  $f_1 = f_2 \equiv f = \frac{\ln(p)}{\langle k \rangle} + 1$ , and  $p_c$  satisfies relation:

$$\langle k \rangle p_c^2 f_c = 2.4554, \quad (2.24)$$

with  $f_c = \frac{\ln(p_c)}{\langle k \rangle} + 1$ . Fig. 2.2 shows that the simulation confirms well the theory. Compared to the case of random attack on one network, where  $p_c = 2.4554/\langle k \rangle$  [7], in Eq.(2.24), the factor  $f_c$  reflects the effect of targeted attack on high degree nodes to increase  $p_c$ . The term  $p_c^2$  in Eq.(2.24) is since we are initially attacking both networks simultaneously instead of only attacking one network. Indeed for the case of initial random attack on two networks simultaneously, from Eq.(2.21) and  $f_A = e^{\langle k \rangle_1 y (f_A - 1)}$ ,  $f_B = e^{\langle k \rangle_1 y (f_B - 1)}$  [7] we obtain  $\langle k \rangle p_c^2 = 2.4554$ .

## 2.6 Summary

In summary, we developed a theoretical framework for understanding the robustness of interdependent networks under targeted attacks on specific degree nodes. We introduce a method and show that targeted-attack problems in networks can be mapped to random-attack problems by transforming the networks which are under initial attack. It provides a routine method (if the random-attack case is solvable) to study the targeted-attack problems in both single networks and randomly connected and uncorrelated interdependent networks, i.e. (i) the case of three or more interdependent networks, (ii) the case of partially coupled interdependent networks, (iii) the case in which a node from network  $A$  can depend on more than one node from network  $B$ . By applying the method, we find that in contrast to single networks, when the highly connected nodes are defended ( $\alpha < 0$ ), the percolation threshold  $p_c$  has a finite non-zero value which is significantly larger than zero. For example, when the degrees of all nodes are known and nodes can only be damaged from lower degree to high degree ( $\alpha \rightarrow -\infty$ ),  $p_c \approx 0.46$  for coupled SF networks with  $\lambda = 2.8$  and  $\langle k \rangle = 4$  while  $p_c$  for the same single SF network is 0 (Fig. 2.1). The implications of the present study are dramatic. The current methods applied to design robust networks and improve the

robustness of current networks, i.e. protecting the high degree nodes, need to be modified to apply to interdependent network systems.

## Chapter 3

# The Robustness Of Interdependent Clustered Networks

### 3.1 Introduction

In a system of interdependent networks, the functioning of nodes in one network is dependent upon the functioning of nodes in other networks of the system. The failure of nodes in one network can cause nodes in other networks to fail, which in turn can cause further damage to the first network, leading to cascading failures and catastrophic consequences. For example, power blackouts across entire countries have been caused by cascading failures between the interdependent communication and power grid systems [4, 50]. Because infrastructures in our modern society are becoming increasingly interdependent, understanding how systemic robustness is affected by these interdependencies is essential if we are to design infrastructures that are resilient [1, 6, 51, 52]. Another example is the human organism is an integrated network where complex physiological systems, each with its own regulatory mechanisms, continuously interact, and where failure of one network can trigger a breakdown of the entire system [53]. In addition to research carried out on specific systems [54–62], a mathematical framework [7] and its generalizations [8, 63, 64] have been developed recently. These studies use a percolation approach to analyse a system of two or more interdependent networks subject to cascading failure [65, 66]. It was found that interdependent networks are significantly more vulnerable than their stand-alone counter-

parts. The dynamics of cascading failure are strongly affected by the structure patterns of network components and by the interaction between networks. This research has focused almost exclusively on random interdependent networks in which clustering within component networks is small or approaches zero. Clustering quantifies the propensity for two neighbours of the same vertex to also be neighbors of each other, forming triangle-shaped configurations in the network [27–29]. Unlike random networks in which there is very little or no clustering, real-world networks exhibit significant clustering. Recent studies have shown that, for single networks, both bond percolation and site percolation in clustered networks have higher epidemic thresholds compared to the unclustered networks [67–73].

Here we present a mathematical framework for understanding how the robustness of interdependent networks is affected by clustering within the network components. We extend the percolation method developed by Newman [67] for single clustered networks to coupled clustered networks. We find that interdependent networks that exhibit significant clustering are more vulnerable to random node failure than networks without significant clustering. We are able to simplify our interdependent networks model—without losing its general applicability—by reducing its size to two networks, A and B, each having the same number of nodes  $N$ . The  $N$  nodes in A and B have bidirectional dependency links to each other, establishing a one-to-one correspondence. Thus the functioning of a node in network A depends on the functioning of the corresponding node in network B and vice versa. Each network is defined by a joint distribution  $P_{st}$  (generating function  $G_0(x, y) = \sum_{s,t=0}^{\infty} P_{st} x^s y^t$ ) that specifies the fraction of nodes connected to  $s$  single edges and  $t$  triangles [67]. The conventional degree of each node is thus  $k = s + 2t$ . The clustering coefficient  $c$  is

$$\begin{aligned}
 c &= \frac{3 \times (\text{number of triangles in network})}{\text{number of connected triples}} \\
 &= \frac{N \sum_{st} t P_{st}}{N \sum_k \binom{k}{2} P_k}.
 \end{aligned} \tag{3.1}$$

### 3.2 Site Percolation of Single Clustered Networks

We begin by studying the generating function of remaining nodes after a fraction of  $(1-p)$  nodes is randomly removed from one clustered network. After the nodes are removed, we define  $t'_i$  to be the number of triangles of which node  $i$  is a part,  $d'_i$  to be the number of single edges that form triangles prior to attack, and  $n'_i$  to be the number of stand-alone single edges prior to attack. This network is thus defined by the joint distribution  $P_{n',t',d'}$ . The probability that a node has  $n'$  single edges from single edges is the sum of all the probabilities that nodes with more than  $n'$  single edges will have exactly  $n'$  edges remaining, which is  $Q_1(n') \equiv \sum_{s=n'}^{\infty} \binom{s}{n'} p^{n'} (1-p)^{s-n'}$ . Similarly, the probability that a node has  $t'$  triangles is the sum of all the probabilities that nodes with more than  $t'$  triangles will have exactly  $t'$  triangles remaining. Since the probability that a triangle will survive is  $p^2$ , the sum is  $Q_2(t') \equiv \sum_{t=t'}^{\infty} \binom{t}{t'} p^{2t'} (1-p^2)^{t-t'}$ . The probability that a triangle corner will have one edge broken is  $\frac{2p(1-p)}{1-p^2}$  and the probability that it will have both edges broken is  $\frac{(1-p)^2}{1-p^2}$ . Thus the probability that a node had  $d'$  single edges forming triangles prior to their destruction is  $Q_3(d') \equiv \binom{t-t'}{d'} \left[\frac{2p(1-p)}{1-p^2}\right]^{d'} \left[\frac{(1-p)^2}{1-p^2}\right]^{t-t'-d'}$ . Combining these three, we have the corresponding generating function

$$\begin{aligned}
G(x, y, z, p) &= \sum_{n',t',d'} P_{n',t',d'} x^{n'} y^{t'} z^{d'} \\
&= \sum_{n'=0}^{\infty} x^{n'} Q_1(n') \sum_{t'=0}^{\infty} y^{t'} Q_2(t') \sum_{d'=0}^{t-t'} z^{d'} Q_3(d') P_{s,t} \\
&= G_0(xp + 1 - p, yp^2 + 2zp(1-p) + (1-p)^2). \tag{3.2}
\end{aligned}$$

We define  $s' = n' + d'$  to be the total number of single links of a node after attack. The joint degree distribution after attack is  $P'_{s',t'}$  which satisfies  $P'_{s',t'} = \sum_{n'=0}^{s'} P_{n',t',d'}$ , with

$d' = s' - n'$ . The generating function of  $P'_{s',t'}$  is

$$\begin{aligned}
G_0(x, y, p) &= \sum_{s',t'} P'_{s',t'} x^{s'} y^{t'} \\
&= \sum_{s'=0}^{\infty} \sum_{n'=0}^{s'} \sum_{t'} P_{n',t',d'} x^{s'} y^{t'} \\
&= \sum_{n',d',t'} P_{n',t',d'} x^{n'} y^{t'} x^{d'} \\
&= G(x, y, x, p).
\end{aligned} \tag{3.3}$$

Therefore, the generating function of the remaining network after attack is

$$G_0(x, y, p) = G_0(xp + 1 - p, yp^2 + 2xp(1 - p) + (1 - p)^2). \tag{3.4}$$

The size of the giant component  $g(p)$  of the remaining network according to Ref. [67] is

$$g(p) = 1 - G_0(u, v^2, p), \tag{3.5}$$

where

$$u = G_q(u, v^2, p), \tag{3.6}$$

$$v = G_r(u, v^2, p),$$

and  $G_q(x, y, p) = \frac{1}{\mu} \frac{\partial G_0(x, y, p)}{\partial x}$ ,  $G_r(x, y, p) = \frac{1}{\nu} \frac{\partial G_0(x, y, p)}{\partial y}$  where  $\mu$  and  $\nu$  are the average number of single links and triangles per node, respectively.

As an example, consider the case when  $(1 - p)$  fraction of nodes are removed randomly from a network with doubly Poisson degree distribution

$$P_{st} = e^{-\mu} \frac{\mu^s}{s!} e^{-\nu} \frac{\nu^t}{t!}, \tag{3.7}$$

where the parameters  $\mu$  and  $\nu$  are the average numbers of single edges and triangles per vertex, respectively. According to Eq. (3.1), the clustering coefficient is  $c = \frac{2\nu}{2\nu + (\mu + 2\nu)^2}$ . Then,

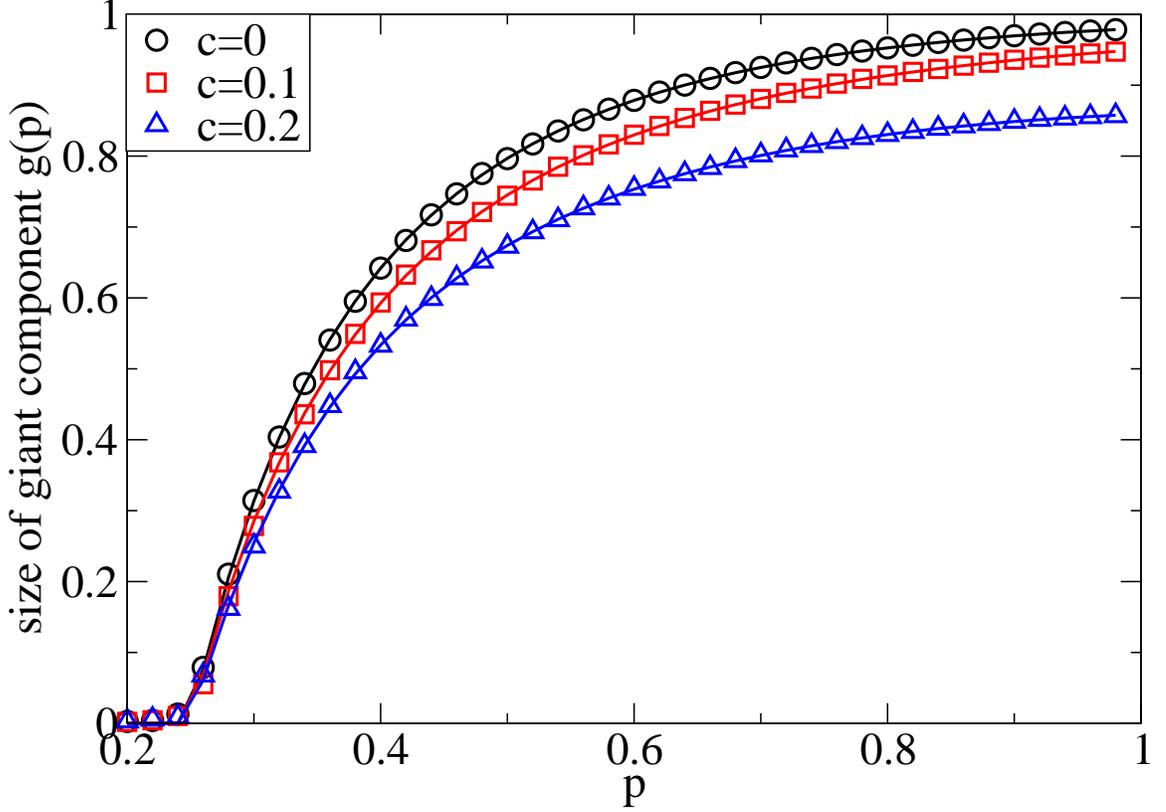


Figure 3.1: Size of giant component  $g(p)$  in single networks with degree distribution Eq. (3.7) and average degree  $\langle k \rangle = 4$ , as a function of  $p$ , the fraction of remaining nodes after random removal of nodes. Curves are from theory Eq. 3.8, symbols are from simulation.

$G_0(x, y) = e^{\mu(x-1)} e^{\nu(y-1)}$  and  $G_0(x, y, p) = G_q(x, y, p) = G_r(x, y, p) = e^{[\mu p + 2p(1-p)\nu](x-1)} e^{\nu p^2(y-1)}$ , and  $u = v = 1 - g(p)$ , leading to

$$g(p) = 1 - e^{[\mu p + 2p(1-p)\nu]g(p)} e^{\nu p^2(g(p)^2 - 2g(p))}. \quad (3.8)$$

This equation is a closed-form solution for the giant component  $g(p)$  and can be solved numerically. The critical case appears when the derivatives of the both sides of Eq. (3.8) are equal. That leads to the critical condition  $\langle k \rangle p_c = 1$ , which is independent of clustering. However the degree distribution of the doubly Poisson model changes as we keep the average degree and change the clustering coefficient. When the degree distribution is fixed, the critical threshold actually increases as clustering increases [70, 71]. Furthermore, Fig. 3.1

shows the resulting giant component as a function of  $p$ . Note that single networks with higher clustering have smaller giant components.

### 3.3 Degree-Degree Correlation

When constructing clustering in a network, it is usually impossible to avoid generating degree-degree correlations. To better understand the effect of clustering on degree-degree correlations, we present an analytical expression of degree correlation as a function of the clustering coefficient for a doubly Poisson-clustered network—see Eq. (3.7).

The degree-degree correlation [74] can be expressed as

$$\rho_D = \frac{N_1 N_3 - N_2^2}{N_1 \sum_{i=1}^N d_i^3 N_2^2} \quad (3.9)$$

where  $N_m$  is the total number of  $m$  hop walks between all possible node pairs  $(i, j)$  including cases  $i = j$ .

The generating function of the degree of a node in the network is  $\sum_{s,t=0}^{\infty} P_{st} z^{s+2t} = G_0(z, z^2)$ . Let  $q_{st}$  be the fraction of nodes with  $s$  single edges and  $t$  triangles that are reached by traversing a random single link, where  $s$  includes the traversed link and  $r_{st}$  is the fraction of nodes with  $s$  single edges and  $t$  triangles reached by traversing a link of a triangle,  $q_{st} = \frac{sP_{s,t}}{\langle s \rangle}$ ,  $r_{st} = \frac{tP_{s,t}}{\langle t \rangle}$ . Their corresponding generating functions are  $G_q(x, y) = \frac{1}{\langle s \rangle} \frac{\partial G_0(x, y)}{\partial x} x$  and  $G_r(x, y) = \frac{1}{\langle t \rangle} \frac{\partial G_0(x, y)}{\partial y} y$ . Moreover,  $N_3 = \sum_i \sum_j a_{ij} N_2(j)$ , where  $N_2(j)$  is the total number of two-hop walks starting from node  $j$ . The number of three-hop walks from a node  $i$  is equal to the total number of two-hop walks starting from all of its neighbours. Thus,  $N_3 = \sum_j k_j N_2(j)$ , where the number of two-hop walks starting from a node  $j$  with degree  $k_j$  will be counted  $k_j$  times in  $N_3$ . Equivalently,  $N_3 = N \sum_{st} (s + 2t) P_{s,t} N_2(s, t)$ , where  $N_2(s, t)$  is the number of two hop walks from a node with  $s$  single edges and  $t$  triangles. The generating function of the number of single edges and of triangles reached in two hops from a random node is  $G_2(x, y) = \sum_{st} P_{s,t} \cdot G_q^s(x, y) \cdot G_r^{2t}(x, y)$ . The generating function of the total number of links and of triangles reached within three hops starting from all nodes is

$G_3(x, y) = N \sum_{st} P_{s,t} \cdot (G_q(x, y))^{s(s+2t)} \cdot (G_r(x, y))^{2t(s+2t)}$ . The number  $N_k$  of  $k$ -hop walks can be approximated by its mean in a large network

$$\begin{aligned} N_1 &= N \langle k \rangle, \\ N_2 &= N \frac{\partial G_2}{\partial x} \Big|_{x=1, y=1} + 2N \frac{\partial G_2}{\partial y} \Big|_{x=1, y=1} \\ N_3 &= \frac{\partial G_3}{\partial x} \Big|_{x=1, y=1} + 2 \frac{\partial G_3}{\partial y} \Big|_{x=1, y=1} \end{aligned}$$

When both  $s$  and  $t$  follow a Poisson distribution,

$$\begin{aligned} G_0(x, y) &= e^{\mu(x-1)} e^{\nu(y-1)} \\ G_q(x, y) &= G_0(x, y)x \\ G_r(x, y) &= G_0(x, y)y. \end{aligned}$$

In this case,

$$\begin{aligned} N_1 &= N \langle k \rangle \\ N_2 &= N \langle k \rangle \left( \frac{\langle k \rangle}{1-c} + 1 \right) \\ N_3 &= (\langle k \rangle^3 + 2\langle k \rangle^2 + 4\nu\langle k \rangle + \langle k \rangle + 6\nu) N \\ \sum_{i=1}^N d_i^3 &= (\langle k \rangle^3 + 3\langle k \rangle^2 + (6\nu + 1)\langle k \rangle + 6\nu) N, \end{aligned}$$

which together with Eq. (3.9) leads to

$$\rho_D = \frac{c - c^2 - \langle k \rangle c^2}{1 - c + \langle k \rangle c - 2\langle k \rangle c^2}, \quad (3.10)$$

where  $c$  is the clustering coefficient, Eq. (3.1).

Figure 3.2 shows the relation between the degree correlation and the clustering coefficient  $c$  for a Poissonian network [see Eq. (3.7)], for two given average degrees ( $\langle k \rangle = 3$  and 4). The figure shows a positive degree-degree correlation across the entire range, which means the model is assortative [70]. The degree-degree correlation increases until  $c$  achieves half

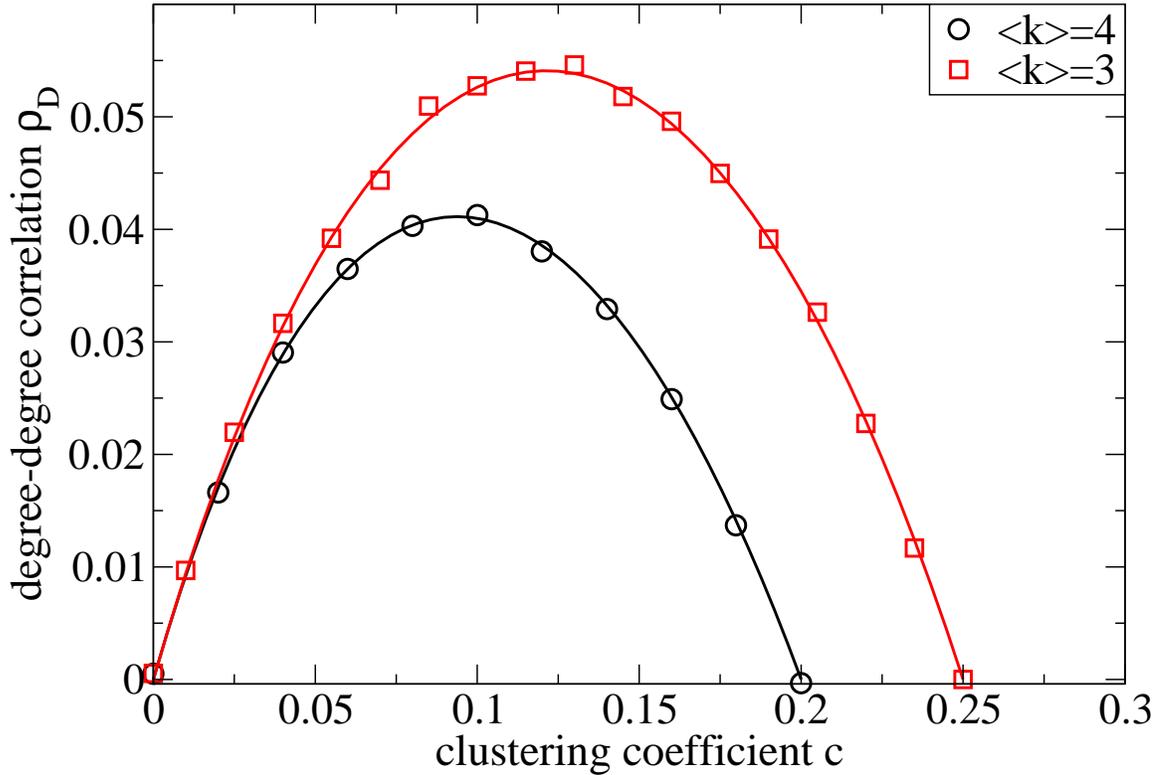


Figure 3.2: Degree-degrees correlation as a function of the clustering coefficient for Poisson network (Eq. (3.7)) with average degree  $\langle k \rangle = 3$  and 4. Curves are from theory (Eq. 3.10) and symbols from simulations.

of its maximum and then decreases to zero when  $c$  reaches its maximum. When  $c$  is 0 or the maximum, the nodes connect to either all single links or all triangles, respectively.

### 3.4 Percolation on Interdependent Clustered Networks

To study how clustering within interdependent networks affects a system's robustness, we apply the interdependent networks framework [7]. In interdependent networks A and B, a fraction  $(1 - p)$  of nodes is first removed from network A. Then the size of the giant components of networks A and B in each cascading failure step is defined to be  $p_1, p_2, \dots, p_n$ , which are calculated iteratively

$$\begin{aligned}
 p_n &= \mu_{n-1} g_A(\mu_{n-1}), n \text{ is odd,} \\
 p_n &= \mu_n g_B(\mu_n), n \text{ is even,}
 \end{aligned}
 \tag{3.11}$$

where  $\mu_0 = p$  and  $\mu_n$  are intermediate variables that satisfy

$$\begin{aligned}\mu_n &= pg_A(\mu_{n-1}), n \text{ is odd,} \\ \mu_n &= pg_B(\mu_{n-1}), n \text{ is even.}\end{aligned}\tag{3.12}$$

As interdependent networks A and B form a stable mutually-connected giant component,  $n \rightarrow \infty$  and  $\mu_n = \mu_{n-2}$ , the fraction of nodes left in the giant component is  $p_\infty$ . This system satisfies

$$\begin{aligned}x &= pg_A(y), \\ y &= pg_B(x),\end{aligned}\tag{3.13}$$

where the two unknown variables  $x$  and  $y$  can be used to calculate  $p_\infty = xg_B(x) = yg_A(y)$ . Eliminating  $y$  from these equations, we obtain a single equation

$$x = pg_A[pg_B(x)].\tag{3.14}$$

The critical case ( $p = p_c$ ) emerges when both sides of this equation have equal derivatives,

$$1 = p^2 \frac{dg_A}{dx}[pg_B(x)] \frac{dg_B}{dx}(x)|_{x=x_c, p=p_c},\tag{3.15}$$

which, together with Eq. (3.14), yields the solution for  $p_c$  and the critical size of the giant mutually-connected component,  $p_\infty(p_c) = x_c g_B(x_c)$ .

Consider for example the case in which each network has doubly-Poisson degree distributions as in Eq. (3.7). From Eq. (3.13), we have  $x = p(1 - u_A)$ ,  $y = p(1 - u_B)$ , where

$$\begin{aligned}u_A &= v_A = e^{[\mu_A y + 2y(1-y)\mu_A](u_A - 1) + \nu_A p^2 (v_A^2 - 1)}, \\ u_B &= v_B = e^{[\mu_B x + 2x(1-x)\mu_B](u_B - 1) + \nu_B p^2 (v_B^2 - 1)}.\end{aligned}$$

If the two networks have the same clustering,  $\mu \equiv \mu_A = \mu_B$  and  $\nu \equiv \nu_A = \nu_B$ ,  $p_\infty$  is then

$$p_\infty = p(1 - e^{\nu p_\infty^2 - (\mu + 2\nu)p_\infty})^2.\tag{3.16}$$

The giant component,  $p_\infty$ , for interdependent clustered networks can thus be obtained

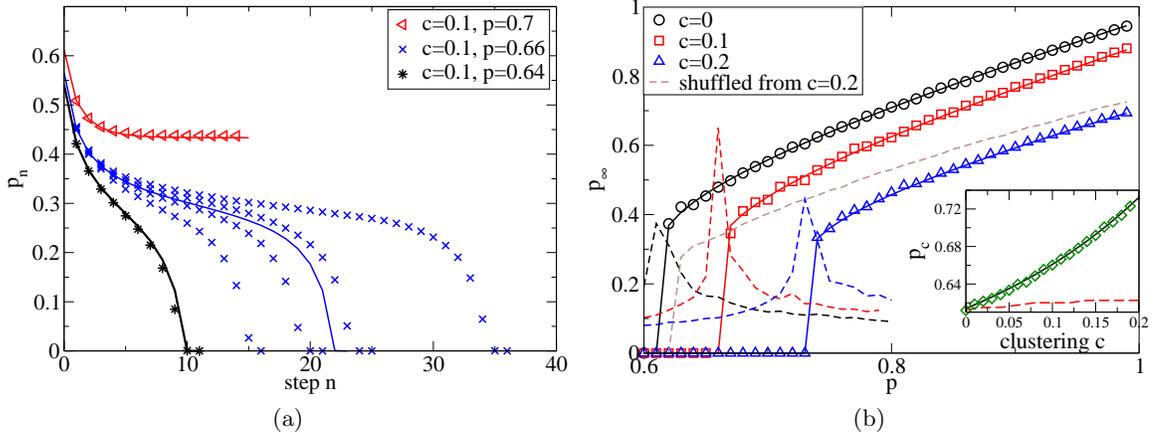


Figure 3.3: (a) Size of mutually connected giant component as a function of cascading failure steps  $n$ . Results are for  $c = 1$ ,  $p = 0.64$  (below  $p_c$ ),  $p = 0.66$  (at  $p_c$ ) and  $p = 0.7$  (above  $p_c$ ). Lines represent theory (Eqs. (3.11) and (3.12)) and dots are from simulations. Note that at  $p_c$  there are large fluctuations. (b) Size of giant component,  $p_\infty$ , in interdependent networks with both networks having clustering via degree distribution Eq. (3.7) and average degree  $\langle k \rangle = 4$ , as a function of  $p$ . Dashed lines are number of iterations (NOI) before cascading failure stops obtained by simulation. The star curve is for shuffled  $c = 0.2$  network, which keeps the same degree distribution but without clustering and without degree-degree correlation. Inset: Green squares and solid line represents critical thresholds,  $p_c$ , of interdependent networks as a function of clustering coefficient  $c$ . Red dashed line represents critical threshold of shuffled interdependent networks which originally has clustering coefficient  $c$ . The shuffled networks have zero clustering and degree-degree correlation, but has the same degree distribution as the original clustered networks. In all figures, symbols and dashed lines represent simulation, solid curves represent theoretical results.

by solving Eq. (3.16). Note that when  $\nu = 0$  we obtain from Eq. (3.16) the result obtained in Ref. [7] for random interdependent ER networks. Figure 3.3a, using numerical simulation, compares the size of the giant component after  $n$  stages of cascading failure with the theoretical prediction of Eq. (3.11). When  $p = 0.7$  and  $p = 0.64$ , which are not near the critical threshold ( $p_c = 0.6609$ ), the agreement with simulation is perfect. Below and near the critical threshold, the simulation initially agrees with the theoretical prediction but then deviates for large  $n$  due to the random fluctuations of structure in different realizations [7]. By solving Eq. (3.16), we have  $p_\infty$  as a function of  $p$  in Fig. 3.3b for a given average degree and several values of clustering coefficients and in Fig. 3.4a for a given clustering and for different average degree values. As the figure shows, when higher clustering within a

network is introduced, the percolation transition yields a higher value of  $p_c$  (see inset of Fig. 3.3b).

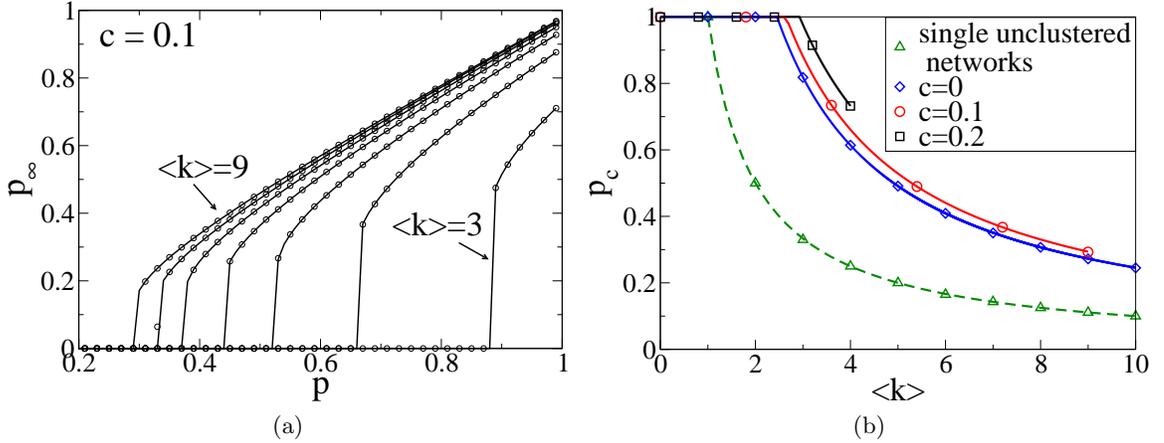


Figure 3.4: (a) Size of giant component as a function of  $p$  for fixed clustering coefficient  $c = 0.1$  and different average degrees. From right to left  $\langle k \rangle = 3, 4, 5, \dots, 9$ . (b) Critical threshold  $p_c$  as a function of average degree for different clustering coefficients. The solid curves are for interdependent networks and the dashed curve is for single networks. Symbols and curves represent simulation and theoretical predictions respectively.

When clustering changes in this doubly Poisson distribution model, degree distribution and degree-degree correlation also change. First, to address the influence of the degree distribution, we study the critical thresholds of shuffled clustered networks. Shuffled clustered networks have neither clustering nor degree-degree distribution but keep the same degree distribution as the original clustered networks. The brown dashed curve in Fig. 3.3b represents the giant component of interdependent shuffled clustered networks with original clustering  $c = 0.2$ . The figure shows that the difference in  $p_c$  between the  $c = 0$  network and the shuffled  $c = 0.2$  network is only 0.01, while the difference between the  $c = 0$  and the  $c = 0.2$  networks is 0.12. In addition,  $c = 0.2$  clustered networks has no degree-degree correlation (Fig. 3.2), which means the 0.12 shift of  $p_c$  is due to clustering and not to a change in degree distribution. We also show the critical thresholds of interdependent shuffled clustered networks as the red dashed line in the inset of Fig. 3.3b. Note that the change of degree distribution barely shifts the critical threshold. We next discuss the effect of the degree-degree correlation on the change of critical threshold. From Ref. [75], the degree assortativity alone

monotonously increases the percolation critical threshold of interdependent networks. Because in our case degree-degree correlation first increases and then decreases (see Fig. 3.2), while the critical threshold of interdependent networks increases monotonously as clustering increases (see inset of Fig. 3.3b), we conclude that clustering alone increases the value of  $p_c$ . Thus clustering within networks reduces the robustness of interdependent networks. This probably occurs because clustered networks contain some links in triangles that do not contribute to the giant component, and in each stage of cascading failure the giant component will be smaller than in the unclustered case.

We also study the effect of the mean degree  $\langle k \rangle$  on the percolation critical point. Figures 3.4a and 3.4b both show that, when clustering is fixed, the percolation critical point of interdependent networks decreases as the average degree  $\langle k \rangle$  of network increases, making the system more robust. Figure 3.4b also shows that a larger minimum average degree is needed to maintain the network against collapse without any node removal as clustering increases.

### 3.5 Conclusion and Summary

To conclude, based on Newman's single network clustering model, we present a generating-function formalism solution for site percolation on both single and interdependent clustered networks. We also derive an analytical expression, Eq. (3.10), for degree-degree correlation as a function of the clustering coefficient for a doubly-Poisson network. Our results help us better understand the effect of clustering on the percolation of interdependent networks. We discuss the influence of a change of degree distribution and the degree-degree correlation associated with clustering in the model on the critical threshold of interdependent networks and conclude that  $p_c$  for interdependent networks increases when networks are more highly clustered. In the clustering model we are using, high clustering is hard to reach, because of lack of higher order cliques than triangles. Recently, models which can have higher clustering and are analytically solvable were proposed [73, 76], which are important complements to the clustering model in this letter. We believe that with very high clustering, the relationship

between the robustness of interdependent networks and clustering would be similar: keeping the degree distribution the same, interdependent networks would be easier to break down under random failure when clustering of the networks are higher.

## Chapter 4

# Identifying Influential Directors in the United States Corporate Governance Network

### 4.1 Introduction

Corporate governance is important for developing company policies and assuring business growth and innovation. For successful industry positioning and competitiveness, companies tend to elect directors who are influential and well known in the business world as well as in the community. We assume that the influence of a director is reflected by the impact that a director can have on the whole industry. Usually an influential director functions as a model in the industry and can impose his/her philosophy to a wide range of companies. Many rankings of powerful and influential people have been created by business magazines based on interviews and public opinion. However, to the best of our knowledge, there are no quantitative studies conducted on influence of corporate directors.

The influential directors, according to research done by economists, are usually those who serve on many company boards, because they are more likely to be active in various policy planning organizations and form a leading edge of the “capitalist class” [77]. These directors also often constitute a vanguard of the corporate elite, typically they are often in the forefront of innovation, and well integrated in the community [78, 79]. In addition,

directors of large companies are considered to be more important than those who serve on small company boards, which is emphasized by magazines that create lists of the “most powerful people”.

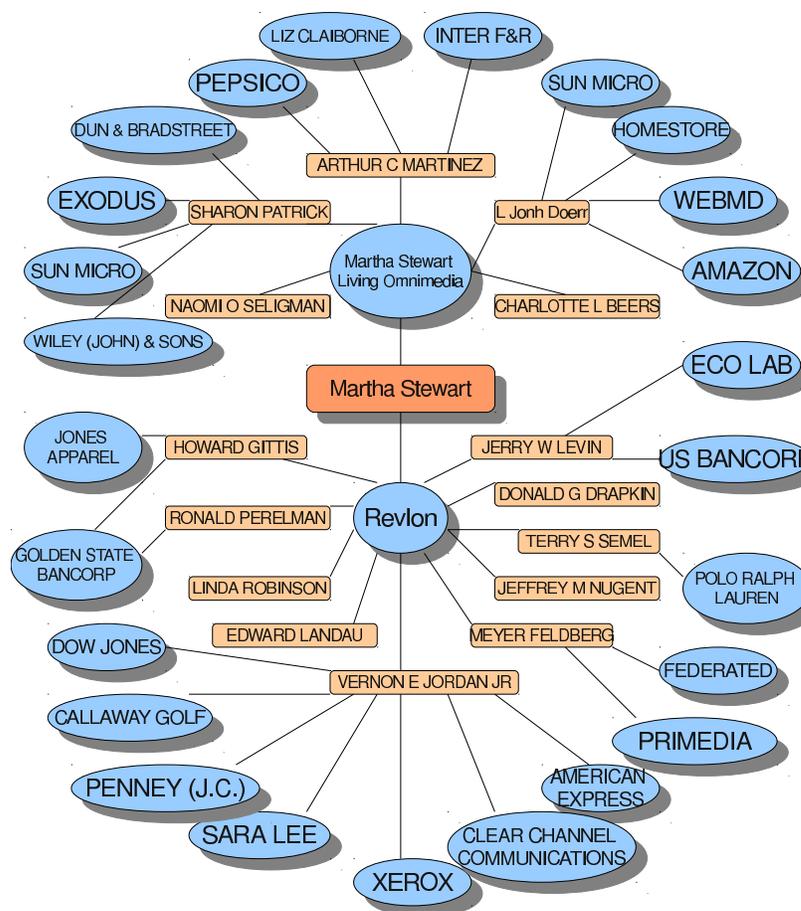


Figure 4.1: Illustration of a subnetwork of the company-director bipartite network. Nodes shown as squares represent directors, while nodes shown as circles depict companies. If a director serves on the board of a company, a link exists between them. This subnetwork includes Martha Stewart, the corporate boards on which she serves, and the other directors who serve on common boards with Martha Stewart including the additional companies on which boards these directors serve.

We define the total capitalization of all the companies (TCC) with which a director is

affiliated as a quantity that contains both the number and the size of the companies on which boards a director serves. We argue that the number and the size of the companies with which a director is affiliated does not entirely reflect the influence of a director. To illustrate this point, we show in Fig. 4.1 an example of Martha Stewart in the year 2001 when she was named the third most powerful woman in America by *Ladies Home Journal*. As illustrated in this figure, Martha Stewart was director of only two companies, *Revlon Corp.* and *Martha Stewart Living Omnimedia*. If we rank the directors by TCC, Martha Stewart would only be ranked in the bottom 16th percentile in the *Investor Responsibility Research Center* (IRRC) directors database. Fig. 4.1 shows that the directors who serve on the same boards as Martha Stewart are also affiliated with many large companies. We argue that Martha Stewart’s influence comes from her proximity to directors who serve on the boards of these large companies. This indicates that the relative position of a director (node) in the network contributes to this director’s influence, i.e. the influence of a director depends not only on his/her own characteristics but also on the characteristics of the other directors surrounding a specific director.

In this paper, we develop a systematic measure named *influence factor* (I), that incorporates both the topological and non-topological characteristics of directors in the network, to quantitatively study the influence of directors. In our approach, the influence of a director is based on the amount of information a director obtains from the other directors due to his/her position in the network. In director’s network, nodes represent directors and links between directors represent their service on common corporate boards. In such a network, we assume that directors acquire information from the companies with which they are affiliated, and spread this information to other directors through information-sharing between connected directors [80]. The *influence factor* measure is affected by common centrality measures taken from complex network theory such as degree [81], betweenness [82], closeness [83] and the capitalization of companies. According to the *influence factor* method, Martha Stewart is ranked in the 84th percentile in 2001, in contrast to only 16th percentile when ranked by TCC. We also find that Douglas Leone, who was named as one of the

top 10 venture capitalists in US by *Forbes*, is only in the 30th percentile when ranked by TCC. However, based on the *influence factor* method, he is ranked in the 90th percentile in 2001. We then statistically compare the *influence factor* method with TCC and common centrality measures, such as degree, betweenness, closeness, K-shell [84–86] and Bonacich centrality [30, 31, 87], which are usually assumed to be equivalent to influence [88–90]. We apply all these methods to identify the influential directors who are selected as powerful people by popular business opinion, such as “*Powerful Women in Business*” from *Fortune* magazine, “*Powerful People in Networking*” from *Networking World* magazine and “*100 Most Influential People in Finance*” from *Treasury and Risk Management* magazine. We find that for all three cases the *influence factor* method is consistently among the most efficient methods to identify the most influential directors.

## 4.2 Data

We build a network of directors based on the IRRC directors database [91] which contains information about approximately 1,600 US corporations and 10,000 directors per year from 1996 to 2006.

We compare our results with popular rankings from magazines including:

1) The ranking of “*50 most powerful women in business*” from *Fortune* for 9 years from 1998-2006. Each year, Fortune interviews industry experts, Wall Street analysts and executive recruiters to identify powerful women. The importance of these business women is broadly valued by revenues and profits controlled, their influence inside the company, the importance of the business in the global economy, and its impact on American culture. Each year the rankings include 50 female directors. We regard the same person in different years as a different entity. Thus, there are 450 entities for 9 years, of which 193 entities are included in the IRRC directors database.

2) The list of “*the most powerful people in networking*” from *Network World* for 9 years from 1997-2006 except 2001 (due to magazine policy change in 2001). From 1997 to 2000, the lists include 25 people and from 2002 to 2006, they include 50 people each year. Thus

there are 350 entities, in which 112 entities are included in IRRC directors database.

3) The list of “100 most influential people in finance” from *Treasury and Risk Management* for 4 years from 2003-2006. The lists include 400 entities, in which 47 are included in IRRC directors database.

Less than half of the influential people listed in these business magazines are included in the IRRC directors database, since many influential people selected by the magazines are not directors. In this article, we focus only on the power of influential directors.

### 4.3 Directors’ Network and Its Properties

For each year, we create a bipartite network of companies and directors [92] based on the IRRC database. As shown in Fig. 4.2, a node in this network represents alternatively a director or a company. A link between a director and a company represents the fact that the director serves on the board of the company. The largest connected cluster of the bipartite network includes over 80% of the companies and directors in the database, while the second largest cluster only contains less than 3% of companies and directors. Given this topology, we only study the largest cluster of the network. In a typical year, e.g. 1999, the largest cluster contains 1,528 companies and 11,116 directors. By projecting the bipartite network into one-mode [93], we create a director’s network (Fig. 4.2). The existence of a link between two directors means that they serve on at least one common board. Note that in this network, directors within one company’s board form a fully connected cluster. The overall network is constructed by attaching these fully connected clusters to each other.

Previous studies [92] analyze statistical properties of similar director’s networks, such as degree distribution. Consistently, Fig. 4.3 shows that the tail of the CDF of degree in our director’s network is exponential,  $exp(-k/k_c)$ , where  $k$  is the node degree of a director and  $k_c$  is the exponential decay parameter. Up to a degree of 8, all the CDFs for different years display a plateau as a consequence of the fact that 8 is the characteristic degree of directors for every year. This means that a large number of directors have degree around 8. Because 80% of directors serve on only one corporate board, the characteristic degree of directors

is also the characteristic size of the boards. Directors who serve on many corporate boards usually have large degree and their degree distribution is described by  $k_c$ . We find that  $k_c$  decreases over time from 1997 to 2006. Since the characteristic size of the boards is stable, this indicates a tendency for directors to serve on fewer boards [94].

## 4.4 Influence Factor Measure

We introduce a model to analyze the influence of directors by defining the *influence factor* for each director based on the level of information that the director can obtain from the entire network. The rationale for using the amount of information to value the influence of directors is that (i) information is a valuable commodity in corporate governance, and the more information the director has, the more valuable as a director she or he is; (ii) a director who has access to company information tends to be able to impose his or her influence on those companies, which indicates that the amount of information a director obtains reflects the level of the influence he or she has. These two points coincides with our view of influence of a director. The first point enables directors to impose their influence well and strongly. The second point enables directors to impact a wide range of the whole industry.

The *influence factor* model is defined as follows:

(i) Each company is considered a source of information, which can be obtained by the directors who serve on the company's board. The amount of information embedded in the company is valued by the market capitalization of the company, based on the fact that directors who can obtain information from and impose influence on large companies should be more influential than directors who are affiliated with a marginal company.

(ii) After information is obtained by directors, it flows in the director's network by information-sharing between directors who are connected.

In Fig. 4.4, we demonstrate, as an example, how the *influence factor* of director  $\mathbf{u}$  is calculated:

(i) We determine the amount of information  $w_j$  for each company by their market cap-

italization. If we choose to study the influence of directors within the technology industry, we set  $w_j$  for the companies from other industries to 0, e.g.  $w_A = 0$ ,  $w_D = 0$ , because **A** and **D** are financial companies.

(ii) Distances between each director and director **u** are calculated in the director's network as shown in Fig. 4.4 and we define  $d_j$  as the shortest distance between director **u** and those directors who serve on the board of company  $j$ .

(iii) We reduce the information of each company by  $r_j$  until it reaches director **u**.  $r_j$  is the information reduction rate per unit distance. Since this reduction rate differs from pair of directors to another pair depending on people's nature and relationship, we assume  $r_j$  to be some random number following certain possibility distribution. Without further knowledge of how people share information with each other we assume for simplicity a uniform distribution and choose  $r_j$  to be a random number between 0 and 1. All information relevant to company **E** can be accessed by director **y** (who sits on the board of **E**), but only a fraction,  $(w_E \cdot r_{E1})$ , is passed to director **x** (who sits with **y** on another board), and so on. Thus, the amount of information about company **E** that director **u** can access is  $S_{uE} \equiv w_E \cdot r_{E1} \cdot r_{E2} \cdot r_{E3}$ .

(iv) By adding the information of all the companies we find that the total information that passes through director **u** is  $S_u = S_{uA} + S_{uB} + S_{uC} + S_{uD} + S_{uE} + S_{uF}$ .

(v) Then the *influence factor* ( $I$ ) is calculated as the percentage of total amount of information that flows through director  $u$  by  $I_u \equiv S_u / \sum_j w_j$ .

In general, the *influence factor*  $I$  of a director  $i$  is defined as

$$I_i \equiv \frac{\sum_j w_j r_{j1} r_{j2} \dots r_{jd_j}}{\sum_j w_j}, \quad (4.1)$$

where  $w_j$  is the amount of information embedded in company  $j$  based on market capitalization,  $d_j$  is the shortest distance between director  $i$  and the directors in company  $j$ , which represents the number of intermediaries the information of company  $j$  has passed before it reaches director  $i$ , and  $r_j$  is the random information reduction rate. We obtain the *influence factor* of director **i** as the average of 50 random realizations of  $I_i$  calculated by Eq. (1).

The *normalized influence factor* (NIF) of each director is then defined as

$$\tilde{I}_i \equiv \frac{I_i - \langle I \rangle}{\sigma(I)}, \quad (4.2)$$

where  $\langle I \rangle \equiv \sum_{i=1}^n I_i/n$  is the annual average of the *influence factor* of all directors and  $\sigma(I)$  is the standard deviation of *influence factors* of directors for one year. A negative value of the *NIF* does not mean that a director has negative influence. Only the relative rank is meaningful, e.g. a director with *NIF*  $-0.5$  is more influential than a director who has *NIF*  $-1.1$ .

## 4.5 Properties Of The Influence Factor

For different years, as shown in Fig. 4.5(a), the *influence factor*  $I$  of directors for all the companies follows different cumulative distribution functions (CDF) because the sizes of the networks are different from year to year. Hence, the *influence factor* of directors are not comparable over the years. However, we find that the CDFs of the *NIF*  $\tilde{I}$  for different years collapse to a single curve, which means that there is a scaling mechanism for  $\tilde{I}$  for different years. As shown in Fig. 4.5(b), the scaled curve fits the complement cumulative function of the Gaussian function. This scaling property enables us to compare *NIF* of directors for different years.

As described in the previous section, the information reduction rate is a random parameter. However, we find that the rankings of directors given by different realizations are consistent with each other, as shown in Fig. 4.6. For each realizations, we calculate the influence factor of directors and choose the top 100 and 1000 directors out of around 10,000 directors each year. We then find the overlapping percentage of these top directors for every pair of realizations and plot the average overlapping ratio and error bar in the graph. Out of 10,000 directors each year, there is about 60% overlap for top 100 directors and more than 80% overlap for top 1,000 directors. This result justifies our approach since it shows that the microscopic detail of how much information is shared by a certain pair of directors

is not critical for the process of finding the most influential directors, instead the network property plays an important role.

The *influence factor* of a director is calculated based on a progressively-reduced information exchange process, which is relevant to director's network topological properties. We now compare the *influence factor* to *TCC* and to the other centrality measures such as *degree*, *K-shell*, *closeness*, *betweenness* and *Bonacich centrality*. Fig.4.7 shows that the *influence factor* is not significantly correlated with the centrality measures except closeness. The correlation between *influence factor* and closeness is not surprising because the obtained information for calculating the *influence factor* depends strongly on the distance between directors and the closeness measures the average distance from one director to all the other directors in the network. However, in addition to distance, the *influence factor* also depends on the capitalization of companies, which differentiates *influence factor* from closeness. This difference is reflected in the relative variance of *influence factor* with respect to closeness which is consistently larger than 10% as shown in the inset of Fig. 4.7(c). Indeed, when we test the methods empirically in Section VI, we find that this variance causes big difference between *influence factor* measure and closeness measure in identifying powerful people.

In addition, in Fig. 4.7, we show the effect of  $N_a$ , the number of companies with which a director is affiliated. The graphs show that directors with larger  $N_a$  tend to be more powerful by all measures. However, there is always a large overlap between directors with different  $N_a$ , which supports our argument that directors who serve on more corporate boards are not necessarily more powerful than those who serve on fewer boards.

As discussed in Section III, the director's network is comprised of many fully connected clusters. Because of this specific topology, we argue that *degree*, *K-shell*, and *betweenness* can not entirely reflect the influence of a director. The *degree* and *K-shell* of the directors depend largely on the size of the boards. If a board consists of a large number of directors, all the members of that particular board will have high degrees and will be present in the nucleus of the network by *K-shell* measure, even if, as an extreme example, that board is

isolated from the rest of the network. *Betweenness centrality* of a node is defined as the times that a node is on the shortest paths between all pairs of vertices. Because all directors who serve only on one board will not occur on the shortest paths between other directors and have zero betweenness, *betweenness centrality* does not distinguish the importance of people who are only affiliated with one company’s board.

The *influence factor* is a measure affected by both nontopological properties, such as the capitalization of the companies, and the topological properties of a director, such as degree, closeness, etc. Moreover, the *influence factor* method is useful when applied to a network consisting of many fully connected clusters. Below we test how efficient our *influence factor* measure is in identifying influential people compared to TCC and existing centrality measures.

## 4.6 Empirical Tests of Methods

In order to study the efficiency of a method in identifying the most influential directors from the IRRC database, we first define an efficiency coefficient  $\epsilon$  for each method as follows:

- (1) Rank the directors according to the method that we want to test.
- (2) Use the influential people lists made by popular business magazines as a benchmark.
- (3) Examine the percentage ( $p$ ) of people in the magazine lists who are included in the top  $q$  percents of people from the database ranking, i.e. we select 10% ( $q = 10\%$ ) of people who are ranked at the top of the database list and find that 30% ( $p = 30\%$ ) of the people in the magazine lists appear in the top 10% of people from the database.
- (4) Define the efficiency coefficient  $\epsilon \equiv p/q$ . The larger the  $\epsilon$ , the better the performance of the particular method is in identifying influential directors.

### 4.6.1 Test: Power and influence of female directors in the US corporate governance network

The ranking of “50 most powerful women in business” selects the 50 most powerful and influential businesswomen in the US every year according to the criteria of *Fortune*. Every

year, about 20 out of these 50 powerful women are included in our database. To improve the statistics, we increase the sample size by mixing the executive women of all years from 1998 to 2006, which is validated by the scaling relation for  $\tilde{I}$ , indicating that the *influence factor* of directors in different years are comparable (see Section V). This means that the same executives in different years are treated as different entities. We find that between 1998 and 2006, 193 entities out of 450 are included in our database.

We apply the *NIF*, *TCC*, and common centrality measures to identify the influential female directors selected by these rankings. We plot the values of the efficiency coefficient  $\epsilon$  versus  $q$  in Fig. 4.8(a) and  $p$  versus  $q$  in Fig. 4.8(b), showing that the *NIF* is more efficient in identifying powerful female directors in the network compared to the other centrality measures. Only the performance of *TCC* is comparable with the *NIF*. The top 10% powerful female directors identified by *NIF* from our database contain 40% of the directors who appear in the “50 most powerful women in business” list from *Fortune*.

#### 4.6.2 Test: Power and influence of directors in the US corporate governance network for specific industries

In addition to studying the influence of directors in the overall US corporate governance network, we also analyze the influence of directors over a certain group of companies by assigning zero weight to those companies that we do not want to consider, as described in section IV. Here we examine influential directors for the financial industry and the networking industry.

*Treasury and risk management* compiles annual rankings of the “100 most influential people in finance”, selecting powerful people from the financial industry. In our database, companies are categorized into different economic groups, which allows us to identify the companies that belong to the financial industry. We assign zero weight to the companies that are not in the financial economic groups to calculate the *influence factor* and *NIF* of directors for financial industry by Eqs.(4.1) and ( 4.2). We then calculate the efficiency coefficient  $\epsilon$  for each method and plot our results in Fig. 4.8(c) and Fig. 4.8(d). We see

that in the financial industry, only closeness provides similar performance to *NIF*, while the other centrality measures provide inferior performance to *NIF*. In addition, capitalization of companies, measured by *TCC*, is found not to be a determining factor for influential directors in the financial industry.

*Network World* publishes annual rankings of “*the most powerful people in networking*”. These lists include powerful people in networking and communication technology related industries. In the IRRC database, these industries correspond to technology and communication economic groups. To calculate the *influence factor* and *NIF* of each director for the networking industry, we assign zero weight to the companies outside of the technology and communication economic groups. We then calculate the efficiency coefficient  $\epsilon$  for each method and plot Fig. 4.8(e) and Fig. 4.8(f), showing that in the networking industry, lists made according to *TCC* match the magazine lists better than lists made according to *NIF*. However, the *NIF* is superior in identifying the powerful directors in networking industry compared to the centrality measures, such as closeness, betweenness, K-shell, degree and Bonacich centrality.

In summary, the *influence factor* measure is superior to degree, betweenness, K-shell and Bonacich centrality in identifying powerful directors for all cases. Closeness shows similar efficiency in identifying influential directors as the *influence factor* measure in the financial industry. The *TCC* is as efficient as the *influence factor* measure when studying female directors, while *TCC* is more efficient than the *influence factor* measure when the power of directors in the networking industry is analyzed. When considering the criteria for creating the most powerful people lists, the above results can be explained as follows: The “*100 most powerful people in finance*” list is made by interviewing executives, bankers, economists, technology vendors and consultants, so a director who has a shorter distance to all the other directors in the network is more likely to be known by the other directors in the network. The “*50 most powerful women in business*” list and the “*most powerful people in networking*” list emphasize the revenues and profits controlled by the directors and the importance of their businesses in the global economy. Thus a director who is affiliated

with large companies would be considered to be more important. Nevertheless, our results show that regardless of the approach used by magazines to create powerful people lists, our *influence factor* measure is always among the most efficient methods in identifying powerful people from these lists.

## 4.7 Conclusion

In this paper we have analyzed the power of directors in the US corporate governance network. To measure the influence of directors, we develop a new measure, the *influence factor*, that offers an objective and quantitative way of determining the power of directors. In our network, nodes represent directors and the links between two directors exist if the two directors serve on at least one common corporate board. We build this network of directors based on the *Investor Responsibility Research Center* directors database for the 11-year period between 1996 and 2006, and find that the director's network is comprised of many fully connected clusters. This network topology presents a challenge for the existing centrality measures to properly reflect the importance of a director. The *influence factor* method is based on an information-sharing process that propagates through the network, where the amount of information obtained by a director from other directors depends on the distance between the directors. The longer the distance between two directors is, the more intermediaries they have between them, hence the higher the information reduction rate is. In addition, the *influence factor* is also affected by the market capitalization of the companies with which directors are affiliated. Thus, the *influence factor* combines the topological and non-topological properties of directors in the network. This combination makes the *influence factor* more suitable for identifying influential people in the overall corporate governance network or specific industries compared to other centrality measures or TCC.

In addition to determining the *influence factor*, we also evaluate the *normalized influence factor* (*NIF*) of directors for different years and find a scaling relation between the *NIF* values which allows us to compare the influence of directors across the years. We then compare the

efficiency  $\epsilon$  of the *influence factor* in identifying powerful people with the efficiencies of other centrality measures and TCC, using popular magazine lists as benchmarks. Powerful people lists created by magazines reflect public opinions of directors. Hence they are appropriate to use as benchmarks when testing how well different measures reflect the influence of a director.

We find that contrary to commonly accepted belief that directors of large companies are most powerful, in some instances, influential directors do not serve on boards of large companies. We also find that the *influence factor* measure is consistently either the best or one of the two best methods in identifying the influential people listed in the “50 most powerful women in business” from *Fortune*, “powerful people in networking” from *Networking World*, and “100 most influential people in finance” from *Treasury and Risk Management*. In some cases, closeness and TCC are in competition with the *influential factor* method when the criteria for creating the most powerful people lists emphasize TCC or closeness. However the *influential factor* method is still a better choice to identify the influential directors overall because of its consistency of performance in all three cases regardless of the criteria used to create the powerful people lists.

Even though the amount of information that a director can access through the network may not be the single aspect when determining the influence of the director, the *influence factor* measure developed here properly reflects the influence of directors in the US corporate governance network, and can be a good quantitative and objective measure to identify influential directors in corporate networks.

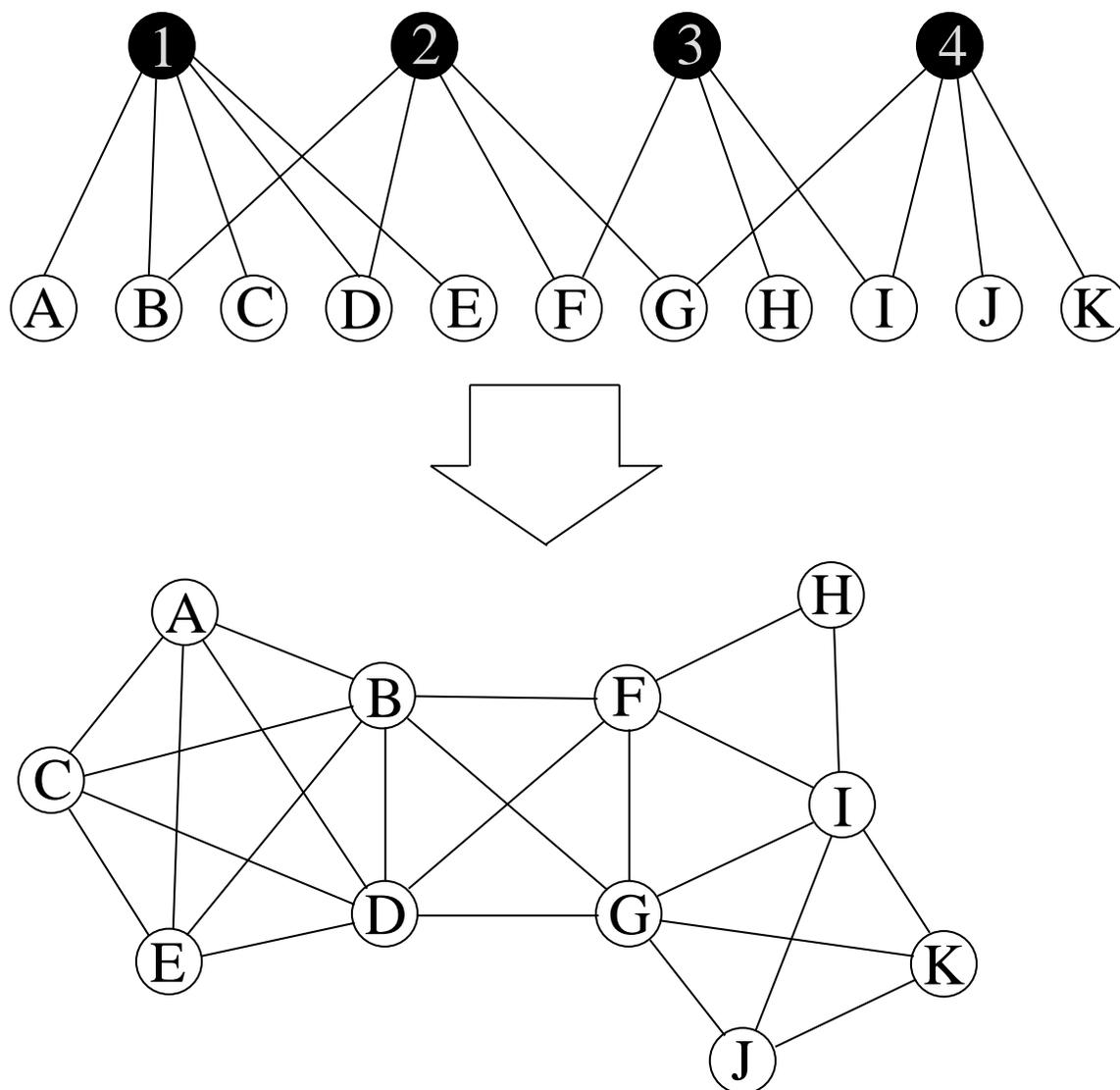


Figure 4.2: Illustration of a bipartite network and its “one-mode” projection [93]. Nodes labeled by numbers correspond to boards, nodes labeled by letters correspond to directors.

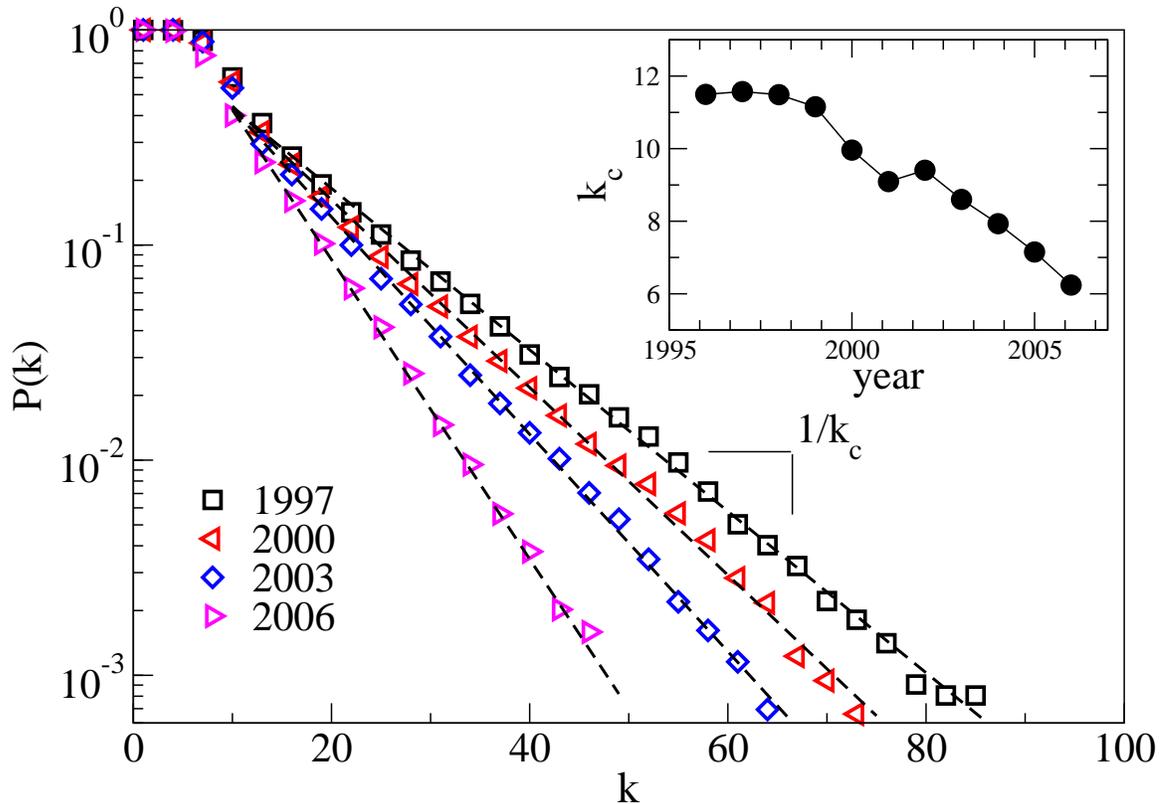


Figure 4.3: (Color online) Demonstration that the cumulative degree distribution function ( $P(k)$ ) of the network of directors for 4 typical years 1997, 2000, 2003, 2006 follows exponential distribution  $P(k) \propto \exp(-\frac{k}{k_c})$ . Note that  $k_c$  decreases as time evolves (inset graph) which means that directors tend to sit on fewer boards in more recent compared to earlier years. The CDF displays a plateau up to a degree of 8 as a consequence of the fact that 8 is the characteristic size of a board for all years studied.

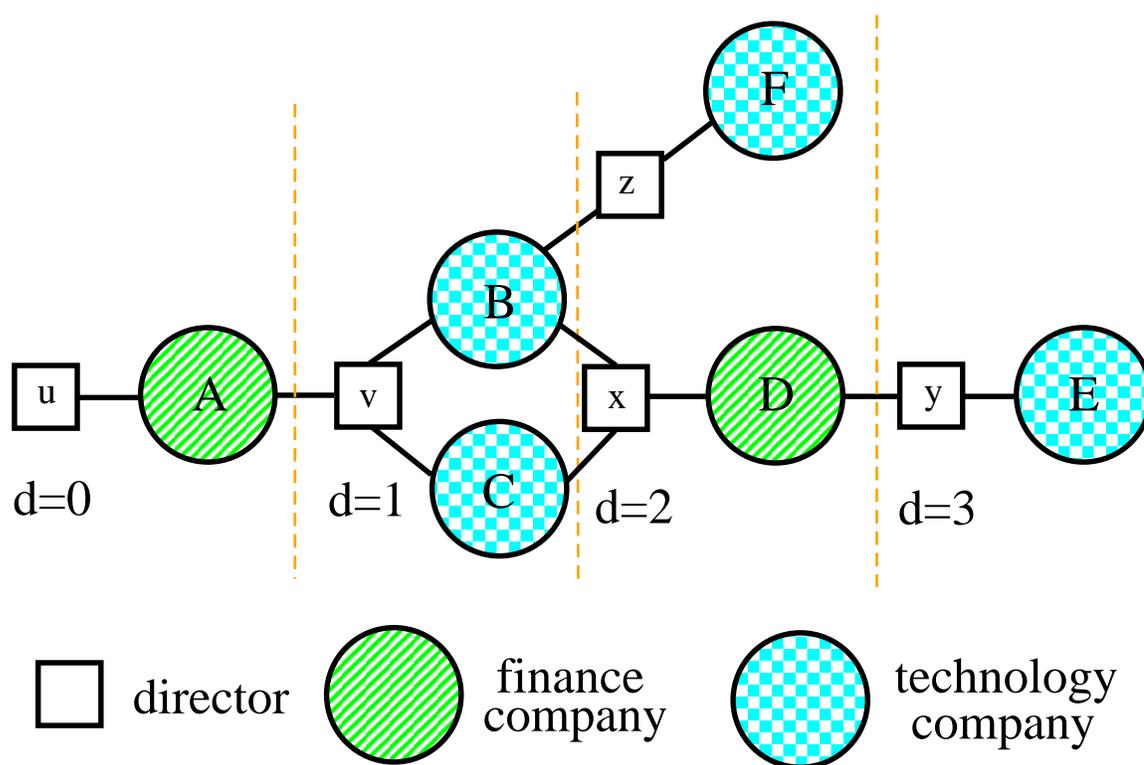


Figure 4.4: (Color online) Illustration of the distance between directors used when calculating *influence factor* ( $I$ ) of a director. The distance between two directors is defined as the number of intermediate companies on the shortest path between these two directors. Thus with respect to director  $u$ , director  $v$  has a distance 1, director  $x$  has a distance 2, etc.

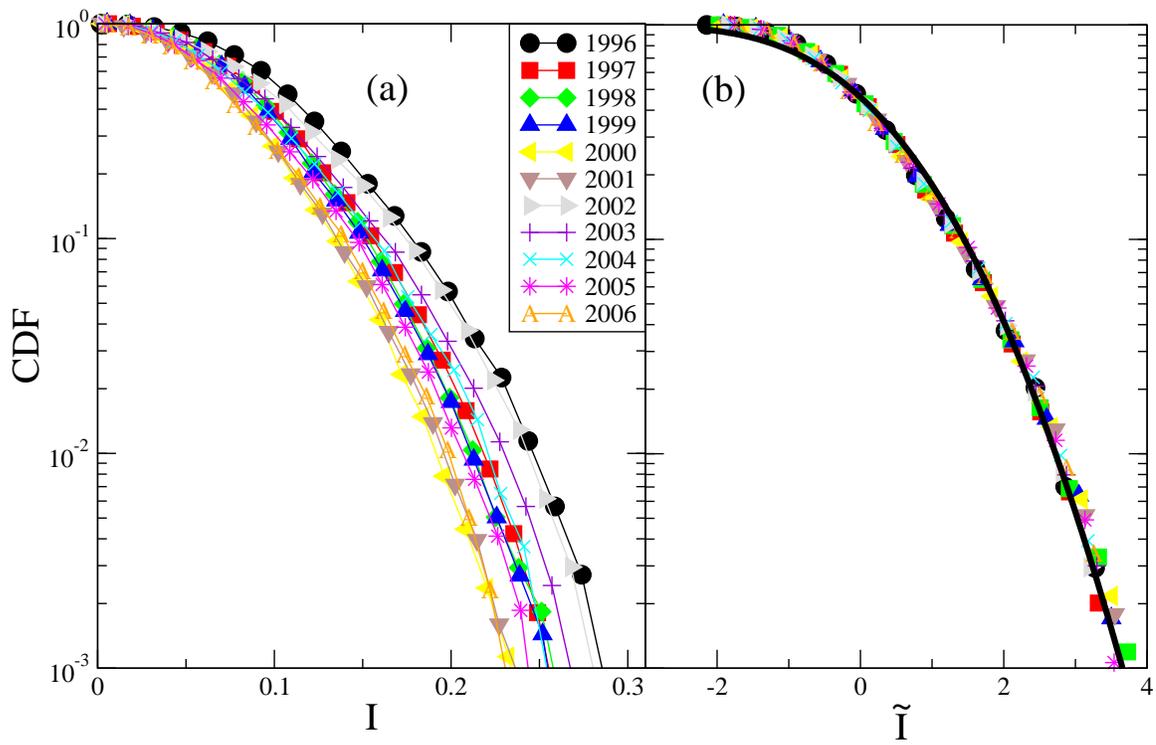


Figure 4.5: (Color online) Demonstration of cumulative distribution of director's *influence factor*  $I$  and *NIF*  $\tilde{I}$  in different years. (a) Cumulative distributions of *influence factor*  $I$  in different years shows dissimilar behavior. (b) Cumulative distributions of the *normalized influence factor*  $\tilde{I}$ , collapse onto a single curve, which indicates a scaling relation for  $\tilde{I}$ . The solid curve corresponds to the complement cumulative function of the Gaussian function with a mean  $-0.12$  and a standard deviation of  $1.22$ . The scaling relation makes  $\tilde{I}$  of directors in different years comparable.

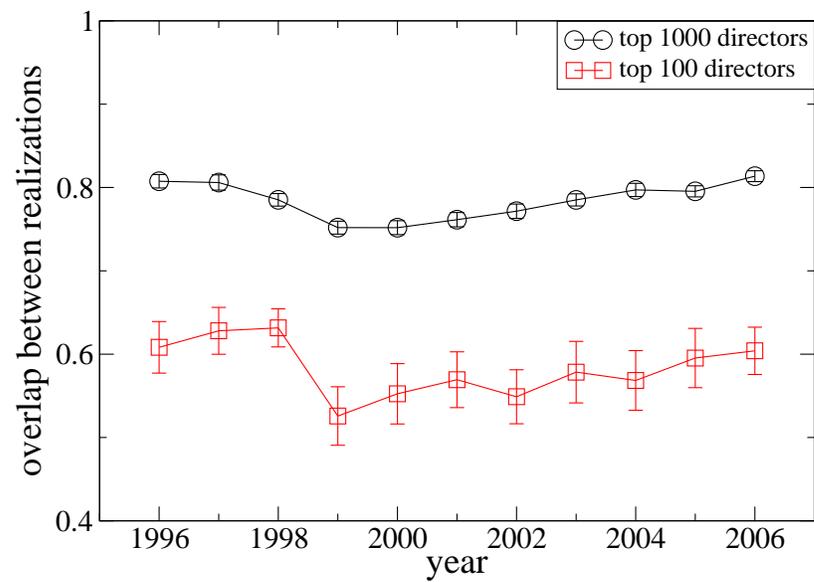


Figure 4.6: Percentage of overlap of top directors according to influence factor between different realizations v.s. years. For each realization, we calculate the influence factor for each director and choose the top 100 and 1000 directors out of around 10,000 directors each year. We find the overlapping percentage of these top directors between each pair of realizations then find the average and error bar of these overlapping percentages.

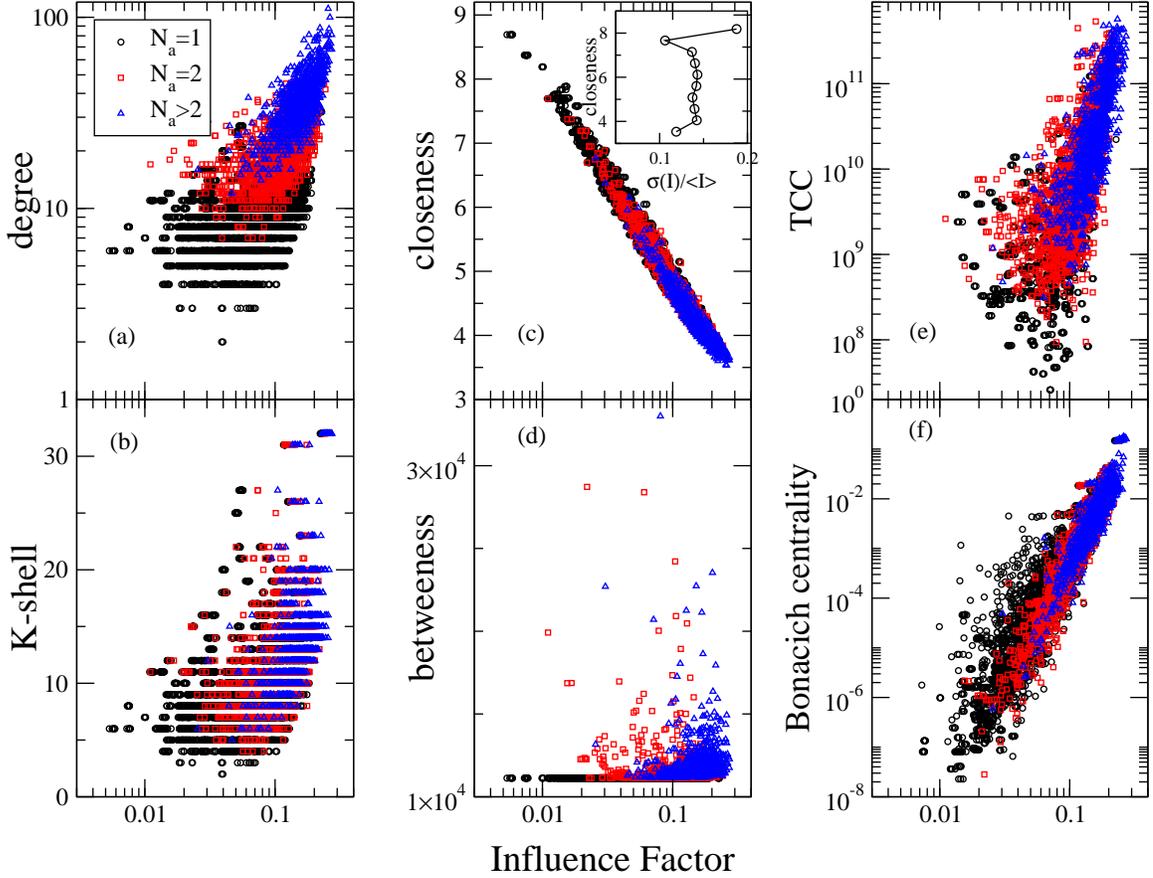


Figure 4.7: (Color online) Comparison between the *influence factor*  $I$ , TCC and the existing centrality measures for a typical year 1999.  $N_a$  is the number of companies with which a director is affiliated. We can see that directors with larger  $N_a$  tend to be more powerful by all measures. However, there is always a large overlap between directors with different  $N_a$ , which supports our argument that directors who serve on more corporate boards are not necessarily more powerful than those who serve on fewer boards. Moreover, we find (i) significant correlation between closeness and *influence factor*, (ii) some positive correlation between TCC, Bonacich centrality and *influence factor* and (iii) a low correlation between *influence factor* and degree, K-shell and betweenness. Inset: The relative variance of *influence factor* ( $\sigma(I)/\langle I \rangle$ ) with respect to closeness. Directors are divided into 10 bins according to their closeness, and variance  $\sigma(I)$  and average  $\langle I \rangle$  for each bin are calculated to plot this relative variance versus closeness graph. Typically the relative variance is around 14%.

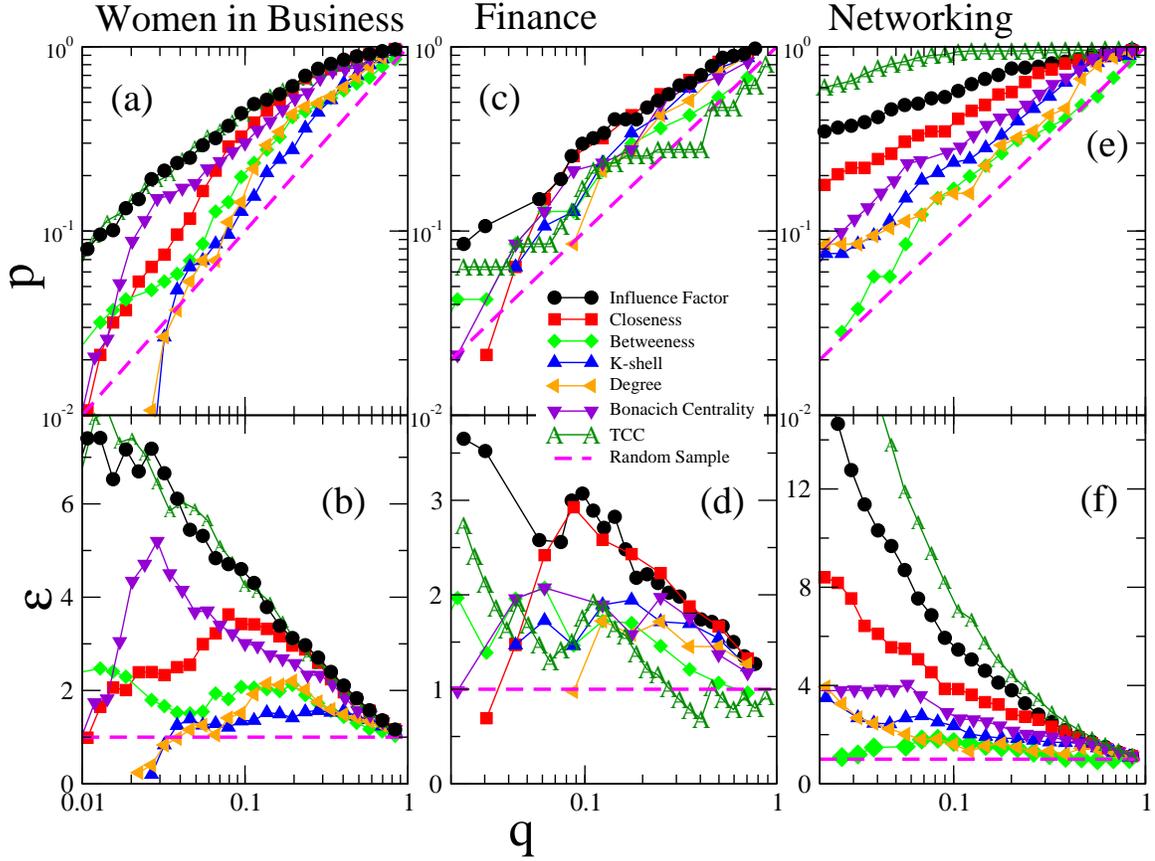


Figure 4.8: (Color online) Comparison between the efficiency  $\epsilon$  of the *influence factor* measure and other measures in identifying the most influential directors. We apply the *NIF*, TCC, closeness, betweenness, K-shell, degree and Bonacich centrality to identify the influential people listed by magazines. The threshold  $q$  is the top fraction of directors after they have been sorted by descending importance, e.g.  $q = 0.1$  for *influence factor* means selecting 10% of the directors with the highest  $\tilde{I}$  in our database;  $p$  is the fraction of directors in the magazines' powerful people lists who are included in the director's set selected from the IRRC database by threshold  $q$ . The dashed line ( $p = q$ ,  $\epsilon = 1$ ) is obtained when directors are randomly selected from the database instead of being ranked. The ratio  $\epsilon \equiv p/q$  represents how efficient a measure is in identifying powerful people listed by magazines from the IRRC directors database. Here we show three cases, “*Powerful Women in Business*” from *Fortune*, “*Influential People in Finance*” from *Treasury and Risk Management* and “*Powerful People in Networking*” from *Network World*. In the case of powerful women in business TCC is as efficient as the *influence factor* measure, in the financial industry closeness shows similar efficiency as the *influence factor* measure, while in the case of powerful people in networking TCC is more efficient than the *influence factor* measure.

## Chapter 5

# Cascading Failures in Bi-partite Graphs: Model for Systemic Risk Propagation

### 5.1 Introduction

There have been dramatic advances in the field of complex networks in recent years [7, 18, 27, 48, 49, 95]. The Internet, airline routes and electric power grids are all examples of networks in which connectivity between network components is essential.

Because of the strong connectivity, catastrophic cascading failure of nodes in networks can happen when the system is under a shock, especially if the shocked nodes represent hubs, or have high centrality measures in the network [6, 34, 39, 63, 96]. So, in order to minimize the systemic risk, these networks should be designed to be robust to external shocks. In the wake of the recent global financial crisis, increased attention has been given to the study of the dynamics of economic systems and to systemic risk in particular. The widespread impact of the current EU sovereign debt crisis and the 2008 world financial crisis show that as economic systems become increasingly interconnected, local exogenous or endogenous shocks can provoke global cascading system failure that is difficult to reverse and that cripples the system for a prolonged period of time. Thus policy makers are compelled to create and implement safety measures that can prevent cascading system failures or soften their systemic impact. Based on the success of complex networks in modelling interconnected systems, applying complex network theory to study economical

systems has been under the spot light [9–14].

There are two channels of risk contagion in the banking system, (i) direct interbank liability linkages between financial institutions and (ii) contagion via changes in bank asset values. The former, which has been given extensive empirical and theoretical study [97–101], focuses on the dynamics of loss propagation via the complex network of direct counterpart exposures following an initial default. The latter, based on bank financial statements and financial ratio analysis, has received scant attention. A financial shock that contributes to the bankruptcy of a bank in a complex network will cause the bank to sell its assets. If the market’s ability to absorb these sales is less than perfect, the market prices of the assets that the bankrupted bank sells will decrease. Other banks that own similar assets could also fail because of loss in asset value and increased inability to meet liability obligations. This imposes further downward pressure on asset values and contributes to further asset devaluation in the market. Damage in the banking network thus continues to spread, and the result is a cascading of risk propagation throughout the system [102, 103]. In this paper we model the risk contagion via changes in asset values in the banking system.

In the past 2008 financial crisis, 371 commercial banks failed between 1/1/2008 and 7/1/2011. The Failed Bank Lists from the Federal Deposit Insurance Corporation (FBL-FDIC) records the names of failed banks and the time when the banks failed. We use this list as an experimental benchmark to our model. The other dataset that we use is the US Commercial Banks Balance Sheet Data (CBBSD) from Wharton Research Data Services, which contains the amounts of assets in each category that the US commercial banks had on their balance sheets (see Method section for more detail). We use this dataset as an input to our model.

The contributions of this paper are as follows. We first analyse the properties of the failed banks from FBL-FDIC, examining the weights in specific assets as well as equity to asset ratios. We then construct a bipartite banking network that is composed of two types of nodes, (i) banks and (ii) bank assets. Link between a bank and a bank asset exists when the bank has the asset on its balance sheet. We also develop a cascading failure model

to simulate the crisis spreading process in the bipartite network. We then populate the model by the banks' balance sheet data (CBBSD) for 2007, and run the cascading failure model by initially introducing a shock to the banking system. We compared the failed banks identified by model with the actual failed banks from the FBL-FDIC from 2008 to 2011, and find that our model simulates well the crisis spreading process and identifies a significant portion of the actual failed banks. Thus, we suggest that our model could be useful to stress test systemic risk of the banking system. For example, we can test each particular asset or groups of assets influence on the overall financial system i.e. if the agricultural assets drop by 20% in value, we can study which banks could be vulnerable to failure, and offer policy suggestions to prevent such failure, such as requirement to reduce exposure to agricultural loans or closely monitor the exposed banks. Finally, we show that sharp transition can occur in the model as parameters change. The bank network can switch between two distinct regions, stable and unstable, which means that the banking system can either survive and be healthy or completely collapse. Because it is important that policy makers keep the world economic system in the stable region, we suggest that our model for systemic risk propagation might also be applicable to other complex financial systems, e.g., to model how sovereign debt value deterioration affects the global banking system or how the depreciation or appreciation of certain currencies impact the world economy.

## 5.2 Properties of Failed Banks.

To build a sound banking system network and systemic risk cascading failure model, we need to study the properties of the failed banks. The asset portfolios of commercial banks contain such asset categories as commercial loans, residential mortgages, and short and long-term investments. We model banks according to how they construct their asset portfolios (upper panel of fig. 5.1). For each bank, the CBBSD contains 13 different non-overlapping asset categories, e.g., bank  $i$  owns amounts  $B_{i,0}, B_{i,1}, \dots, B_{i,12}$  of each asset, respectively. The total asset value  $B_i$  and total liability value  $L_i$  of a bank  $i$  are obtained from CBBSD dataset. The weight of each asset  $m$  in the overall asset portfolio of a bank  $i$  is then defined

as  $w_{i,m} \equiv B_{i,m}/B_i$ . From the perspective of the asset categories, we define the *total market value* of an asset  $m$  as  $A_m \equiv \sum_i B_{i,m}$ . Thus the market share of bank  $i$  in asset  $m$  is  $s_{i,m} \equiv B_{i,m}/A_m$ .

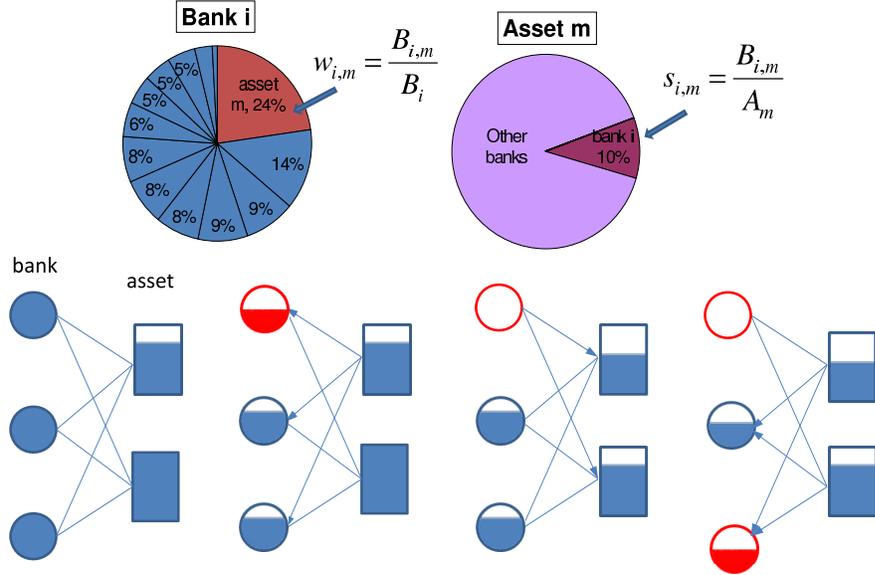


Figure 5.1: Bank-asset bipartite network model with banks as one node type and assets as the other node type. Link between a bank and an asset exists if the bank has the asset on its balance sheet. Upper panel: illustration of bank-node and asset-node.  $B_{i,m}$  is the amount of asset  $m$  that bank  $i$  owns. Thus, a bank  $i$  with total asset value  $B_i$  has  $w_{i,m}$  fraction of its total asset value in asset  $m$ .  $s_{i,m}$  is the fraction of asset  $m$  that the bank holds out. Lower panel: illustration of the cascading failure process. The rectangles represent the assets and the circles represent the banks. From left to right, initially, an asset suffers loss in value which causes all the related banks' total assets to shrink. When a bank's remaining asset value is below certain threshold (e.g. the bank's total liability), the bank fails. Failure of the bank elicits disposal of bank assets which further affects the market value of the assets. This adversely affects other banks that hold this asset and the total value of their assets may drop below the threshold which may result in further bank failures. This cascading failure process propagates back and forth between banks and assets until no more banks fail.

To study the properties of failed banks between 2008 and 2011, we focus on the weight of each bank's assets. For certain assets, we find that the asset weight distributions for all banks differ from the asset weight distributions for failed banks. Figures 5.2(a) and 5.2(c)

show that, unlike survived banks, failed banks cluster in a region heavily weighted with construction and development loans and loans secured by nonfarm nonresidential properties. Failed banks have less agricultural loans in their asset portfolios compared to survived banks (fig. 5.2(d)). These results confirm the nature of the most recent financial crisis of 2007–2011 in which bank failures were largely caused by real estate-based loans, including loans for construction and land development and loans secured by nonfarm nonresidential properties [104]. In this kind of financial crisis, banks with greater agricultural loan assets are more financially robust [105]. Figure 5.2(e) shows that failed banks tend to have lower equity to asset ratios, i.e. failed banks generally had higher leverage ratios than survived banks during the financial crisis of 2008–2011 [106].

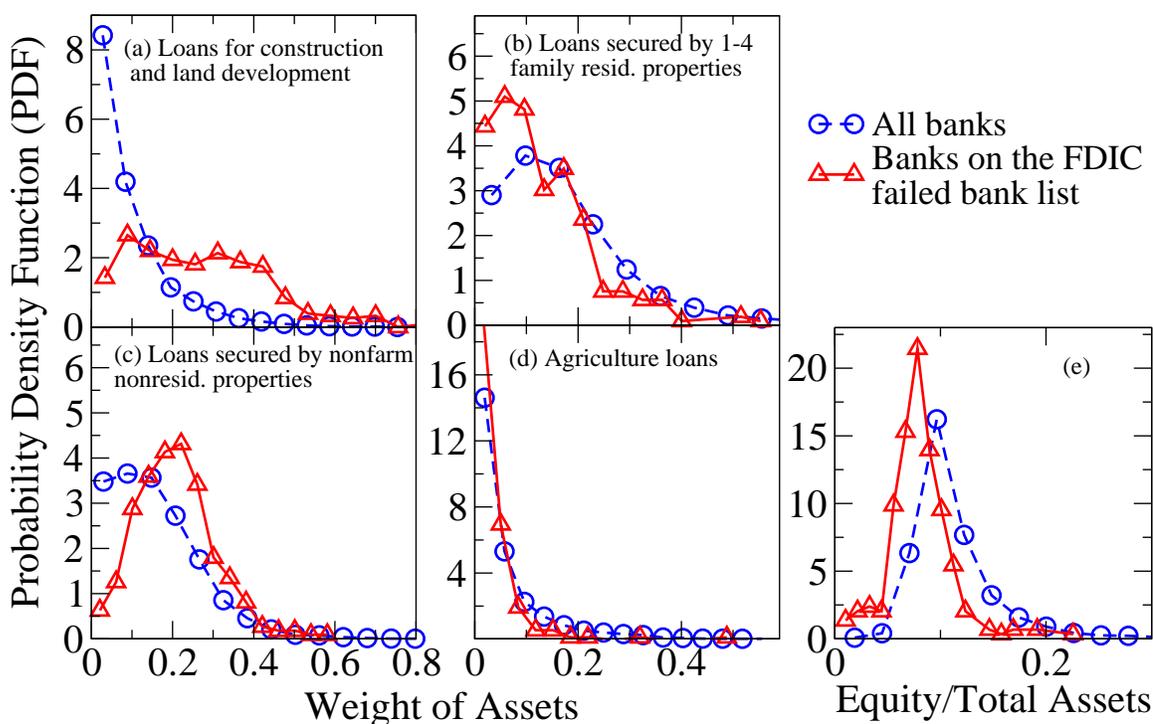


Figure 5.2: Comparison of probability density functions (PDF) of weight of typical assets and equity ratios between all banks and FDIC listed failed banks for 2007. (a) PDF of the weight of loans for construction and land development in banks' total asset. (b) PDF of the weight of loans secured by 1-4 family residential properties in banks' total assets. (c) PDF of the weight of loans secured by nonfarm nonresidential properties in banks' total assets. (d) PDF of the weight of agriculture loans in banks' total assets. (e) PDF of banks' equity to asset ratios. Blue circles curves represents PDFs of all banks, red triangles represents PDFs of those banks that are on the FDIC failed bank list.

### 5.3 Cascading Failure Propagation Model.

To study the systemic risk of the banking system as complex networks, we construct a cascading failure model based on the facts presented in the previous section.

We first build a bipartite network which contains two types of nodes, banks on one hand and bank assets on the other. Link exists between a bank and an asset when the bank has the asset on its balance sheet. No links between banks or between assets exist. To simulate the cascading failure process, we develop and apply the following model with three parameters  $p$ ,  $\eta$  and  $\alpha$  (illustrated in fig. 5.1):

1. We initially shock certain asset  $m$ , reducing the *Total Market Value* of asset  $m$  to  $p$  fraction of its original value,  $p \in [0, 1]$ . The smaller the  $p$  is, the larger the shock. When  $p$  is 0, the total market value of asset  $m$  is wiped out. When  $p$  is 1, no shock is imposed.
2. When the market deteriorates, each bank  $i$  that owns the shocked asset  $m$  will experience  $B_{i,m}(1 - p)$  reduction in value, where  $B_{i,m}$  is the amount of asset  $m$  that is on bank  $i$ 's balance sheet.
3. When the total asset value of a bank declines to a level below the level of promised payments on the debt, it causes distress or default. The total asset value that triggers an incidence of distress lies somewhere between the book value of total liabilities and short-term liabilities. In the corporate sector default analysis, Moody's Analytics used the sum of short-term debt, interest payments and half of long-term debt [107–109] as the distress barrier. However, in the past financial crisis, external aid from other financial institutes or from the government played a significant role in distorting this distress barrier, thus even when a bank's total value of assets was below its liabilities, the bank could still survive. We describe these combined effects using random number  $r$  that is uniformly distributed in range  $[0, \eta]$ , where  $\eta \in [0, 0.5]$  is a parameter controlling tolerance of a bank's assets being below its liabilities. We define the distress barrier to be  $(1 - r) \cdot L_i$ , such that a bank fails when  $B_i < (1 - r) \cdot L_i$ . For

such distress barrier with evolving randomness, the probability  $P(B_i, L_i)$  for a bank  $i$  to fail can be written as

$$P(B_i, L_i) = \begin{cases} 0 & \text{if } B_i \geq L_i \\ (L_i - B_i)/(\eta L_i) & \text{if } \eta \neq 0, L_i > B_i > (1 - \eta)L_i \\ 1 & \text{if } (1 - \eta)L_i > B_i \end{cases} \quad (5.1)$$

4. We assume that when a bank  $i$  fails, the overall market value of each asset  $m$  that the failed bank owns suffers  $\alpha B_{i,m}$  value deduction, where  $\alpha \in [0, 1]$  is a third parameter in the model that describes the market's reaction to a bank failure. The unit price of asset  $m$  becomes  $\frac{A_m - \alpha B_{i,m}}{A_m}$  of its original price. That is because the failed banks need to sell assets to meet their liabilities and the market's ability to absorb this sale is not perfect, which leads to price decrease of the affected assets. The loss of the market value of each asset  $m$  is proportional to  $B_{i,m}$ , the amount of asset  $m$  that the failed bank  $i$  owns. Depending on the liquidity of an asset,  $\alpha$  can be between 0 and 1. When an asset is extremely liquid, the market value of the asset will not be adversely affected by asset sales,  $\alpha = 0$ . When the market is extremely illiquid, then the value of asset could potentially have zero value. Thus the aggregated total market value of asset  $A_m$  will be reduced to  $A_m - B_{i,m}$ , which corresponds to  $\alpha = 1$ .
5. Further deterioration of asset values can then contribute to failure of more banks. Thus the damage in the bipartite network spreads between banks and assets bidirectionally until the cascading failure stops.

Usually financial crises start with a burst of economic bubbles. The correspondence of the model's initial shock parameter  $p$  in reality can be described as the drop of certain asset value at the beginning of a crisis. For example, when the dot-com bubble burst, the technology heavy NASDAQ Composite index lost 66% percents of its value, plunging from the peak of 5048 in March 10, 2000 to the 1720 in April 2, 2001.

## 5.4 Empirical Test and Analysis.

To empirically test our model, we introduce a shock into the banking system by reducing  $(1 - p)$  percentage of the value of a single asset  $m$ . We then monitor the progression of bank failure until the cascading process stops. We examine two distinct groups of banks 1) all the analyzed banks from CBBSD dataset, and 2) the banks from the FDL-FDIC failed bank list. We then study the fraction of banks that were identified as survived by our model in both groups. We plot both of these fractions versus the sizes of initial shocks in fig. 5.3, for parameter  $\eta = 0$ . The four plots correspond to four typical assets being initially shocked respectively. Figure 5.3(a) and figure 5.3(c) show that when the commercial real estate loans, i.e. loans for construction and land development and loans secured by nonfarm nonresidential properties, suffer initial shock respectively, the survival rate of the banks from the first group (all banks), according to our model, is distinctly above the survival rate of the second group of banks (FBL-FDIC failed banks list). This illustrates that when the commercial real estate loans are initially shocked, the model can identify the actual failed banks efficiently. Figures 5.3(b) and 5.3(d) show that when we impose initial shock on loans secured by 1-4 family residual properties or agricultural loans, the model does not clearly separate the two groups of banks. This result indicates that the commercial bank failures during the 2008 financial crisis stems from value deterioration of commercial real estate loans.

To quantitatively test the efficiency of the model in identifying failed banks, we use the receiver-operating-characteristic (ROC) curve analysis, which plots the fraction of true positives out of the positives and the fraction of false positive out of the negatives for a binary classifier system. ROC curve analysis is a standard method in signal detection theory as well as in psychology, medicine and biometrics [110]. We choose a parameter combination of  $p$ ,  $\eta$  and  $\alpha$  to run the model to determine which banks fail, and compare this prediction with the FDIC list of failed banks. The true positive rate is defined as the fraction of the actual failed banks that are also identified as failed in our model. The false positive rate is

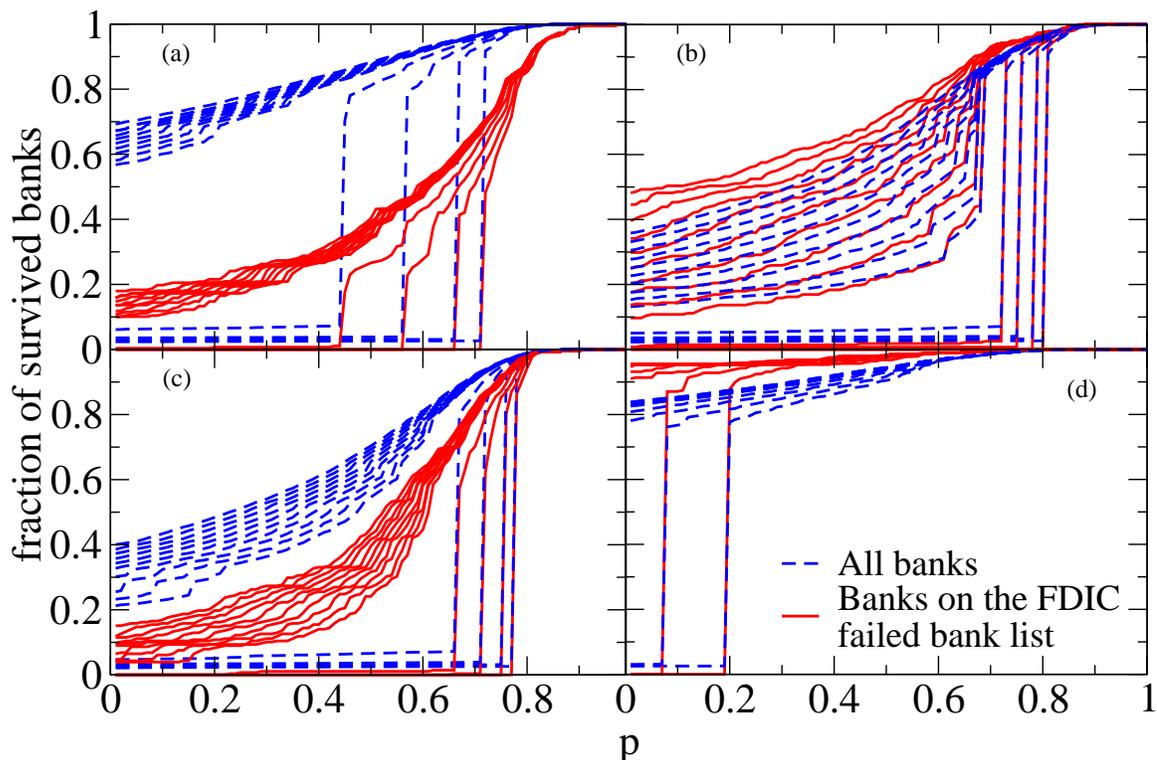


Figure 5.3: Fraction of survived banks after cascading failures as function of the initial loss of value of certain asset, with  $\eta = 0$ . Blue dashed lines represent the fraction of survived banks out of all banks, and the red solid lines represent the fraction of survived banks out of the 278 failed banks from FDIC failed bank list. The parameter  $\alpha$  is changed from 0 to 0.1 by 0.01 to produce 10 lines for each case. (a) Initial shock is imposed to loans for construction and land development. The red solid lines are significantly lower than the blue dashed lines separating clearly the failed banks from the set of all banks. (b) Initial shock is imposed to loans secured by 1-4 family resid. properties. The red solid lines and blue dashed lines are entangled. (c) Initial shock on loans secured by nonfarm nonresid. properties. The red solid lines are distinguishably lower than the blue dashed lines, similarly as in the case under (a). (d) Initial shock on agricultural loans. The red solid lines are slightly higher than the blue dashed line, not showing clear distinction between failed and non failed banks.

the fraction of banks that are not on the FDIC list of failed banks but are identified as failing by our model. Each point in the ROC curve corresponds to one parameter combination. A complete random guess would give points along the diagonal line from the left bottom to the top right corner. The more a point is above the diagonal line, the stronger predictive power the model has.

We firstly impose the initial shock to the construction and land development loans and plot the ROC curves in the top row of fig. 5.4. As fig. 5.4(a) shows, when the false-positive rate is below 0.2 we have a relatively high true-positive to false-positive ratio. For example, the four black dots in fig. 5.4(a) represent the false-positive rate and true positive rate pairs  $(0.06, 0.5)$ ,  $(0.1, 0.61)$ ,  $(0.15, 0.72)$  and  $(0.2, 0.78)$  respectively. The pair  $(0.06, 0.5)$  corresponds to the parameter combination  $(\alpha, \eta, p) = (0.14, 0.26, 0.6)$ , which means using this parameter combination, the model can identify 50% of the actual failed banks that are on the FBL-FDIC with cost of 6% false positive prediction. Overall, the ROC curve is bended well above the diagonal curve, which means the model captures a significant portion of the real-world behavior and has predictive power.

However, fig. 5.4(a) alone is not enough to justify our complex networks model as necessary model to describe the systemic risk in this banking system. If all of the actual failed banks owned a large amount of loans for construction and land development, then these banks will fail in the model in the first round of failure after this type of asset is initially shocked. In that case, we only need to look at the weight of this asset in the banks' portfolio to identify the failed banks. However, we find that the failure of banks does not only occur because of the initial shock to specific assets, but also because of the amplified damage by positive feed back in the complex banking network. The interdependency between banks and the complexity of network structure are crucial to this amplified damage in the system. To demonstrate our findings we conduct separately ROC curve analysis for the first-step prediction (bank failures caused directly by the initial shock on an asset) as well as for the consecutive-steps prediction (bank failures caused by a cascading failure process) as shown in figs. 5.4(b) and 5.4(c). We find that in addition to the first-step effective predictions, the consecutive-steps of the model further efficiently identify failed banks that can not be identified by the first-step (ROC curve is above the diagonal line). Fig. 5.4(d) further shows the number of failed banks correctly identified through the first and consecutive steps of the cascading failure simulation for the four parameter combinations selected from fig. 5.4(a) (black dots in the figures). In all four cases, the number of failed banks predicted by

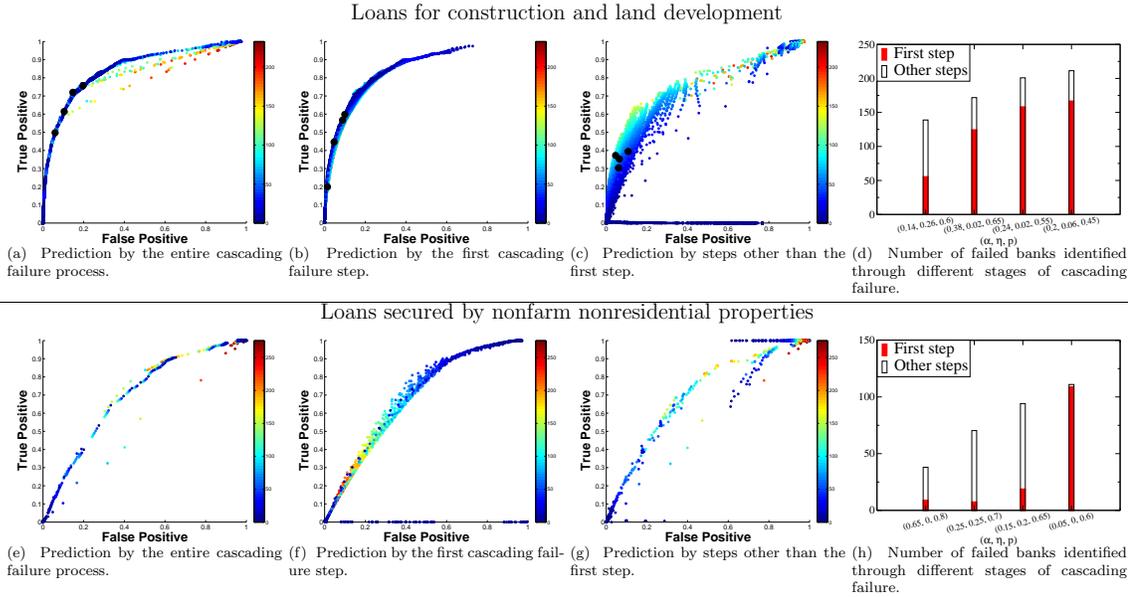


Figure 5.4: ROC curves of the prediction of failed banks by our cascading failure model when the loans for construction and land development are initially shocked (top figures) and when the loans secured by nonfarm nonresidential properties are initially shocked (bottom figures), based on 2007 data. Each point of the ROC curves corresponds to one combination of parameters  $(\alpha, \eta, p)$ . (a)(e) ROC curve of predictions made by the entire cascading failure process, (b)(f) of predictions made by the first cascading failure step and (c)(g) of predictions made by the other than the first cascading steps. The color of a dot represents the number of failed banks correctly identified by the model with the corresponding parameters combination. (d)(h) For fixed false positive rates of 5%, 10%, 15%, and 20%, we find parameter combinations with maximum true positive rates in fig. (a), and show the number of failed banks identified by the first step (red) and the number of failed banks identified by the other steps in the cascading failure process (white). The black dots in (a)(b)(c) show the positions of four combinations respectively.

the consecutive steps represents a significant fraction of the total number of failed banks identified. This result shows that some banks did fail only because of the the complex interconnections between banks in the system, which contributes to the risk contagion in the system. Thus, our model captures the complexity feature of the banking system and can offer prediction better than predictions made only based on balance sheet but without considering interactions between banks.

In addition to construction and land development loans, we also test our cascading failure model by simulating initial shock on other assets. The ROC curves in the bottom

row of fig. 5.4 show that the loans secured by nonfarm nonresidential properties, when initially shocked, have lower predictive power (smaller true-positive to false-positive ratio) compared to the case when initial shock is imposed on loans for construction and land development. ROC curve tests for assets of loans secured by 1-4 family residential properties and agricultural loans, as shown in figs. 5.5(a) and 5.5(b), exhibit curves that are almost diagonal, indicating that initial shocks on these two assets have no predictive power on the failure of the banks in the 2007–2011 financial crisis. A truly random behavior would render points along the diagonal line (the so-called line of no discrimination) from the bottom left to the top right corners.

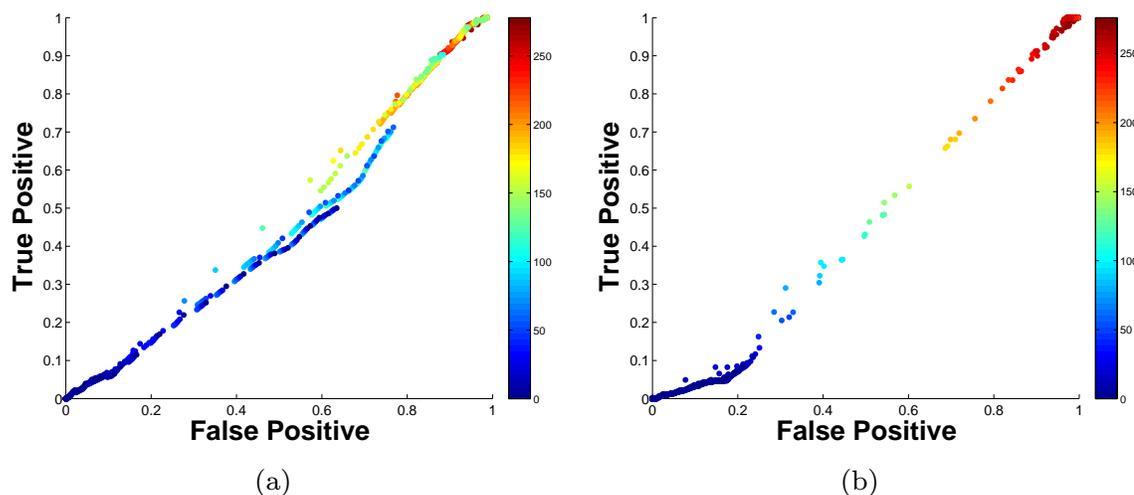


Figure 5.5: ROC curves of predictions of failed banks by our cascading failure model when (a) loans secured by 1-4 family resid. properties and (b) agricultural loans are initially shocked respectively. The straighter the ROC curve is, the closer it is to random case, meaning the less predictive power in regard to the failure of the commercial banks during the 2007-2011 financial crisis.

The above ROC curve results suggest that the construction and land development loans and the loans secured by nonfarm nonresidential properties were the two asset types most relevant in the failure of commercial banks during the 2007–2011 financial crisis. It is largely believed that the past financial crisis is caused by residential real estate assets. However, we do not find evidence that loans secured by 1-4 family residential properties are responsible for commercial banks failures. This result is consistent with the conclusion of ref. [104]

that the cause of the commercial banks failure between 2007-2011 were largely caused by commercial real estate-based loans rather than residential mortgages.

Our final exploration is of the percolation-like property exhibited by the bank-asset bipartite network. Complex networks usually exhibit percolation phase transitions. As the dependent parameter changes, the giant component of connected clusters in the network can drop to zero at the critical point. In the bank-asset bipartite network model we go beyond the giant component of connected clusters and study *all* survived banks. Thus, percolation theory can not be applied. However, we find that a percolation-like phenomenon also exists in this model. We study the number of survived banks after the cascading failure process, tuning one parameter and keeping the other two parameters fixed. We find that the number of survived nodes in networks can change dramatically with a small change of parameters. The parameter combination is chosen as the first example in figure 5.4(d),  $\alpha = 0.14$ ,  $\eta = 0.26$ , and  $p = 0.6$ . We show that the fraction of surviving banks changes smoothly as parameters  $p$  and  $\eta$  change (see figs. 5.6(a) and 5.6(c)). But as  $\alpha$  changes, the fraction of surviving banks changes abruptly at a critical point and displays a first-order-like abrupt phase transition (fig. 5.6(b)). We show that the first-order-like phase transition also exists for  $p$  and  $\eta$  for a certain parameter combination pool. As an example, we choose another parameter combination ( $\alpha = 0.35$ ,  $\eta = 0.2$ , and  $p = 0.6$ ). We show in the right panel of fig. 5.6 that a first-order-like phase transition exists for all three parameters, which means the system is at risk of abrupt collapse. Figure 5.6(d) shows that, when the initial shock parameter  $p$  for an asset is below a certain threshold, even if the other asset market values are undamaged, almost all banks default because the cascading failure of this single asset (construction and land development loans) significantly affects the overall financial system. Figure 5.6(e) shows that when the effect of bank failures on asset market values is sufficiently large, the whole banking system is at risk of collapse. Figure 5.6(f) shows that when  $\eta$  is large, i.e., when the bank distress barrier of default is more relaxed, the robustness of the system improves significantly. Thus, the bank-asset bipartite network behaves differently for different parameter combinations. Figure 5.7 plots the phase diagram

for this bank network. Two different regions exist for parameters  $p$  and  $\alpha$ . In region I, the bank network system is in a stable state, i.e., after cascading failure a significant number of banks will still survive. In region II, the cascading failure process contributes to the collapse of the entire bank network. Given that the bank network as a complex system exhibits these two distinct states, it is extremely important that policy makers institute rules that will keep the banking system in the stable region.

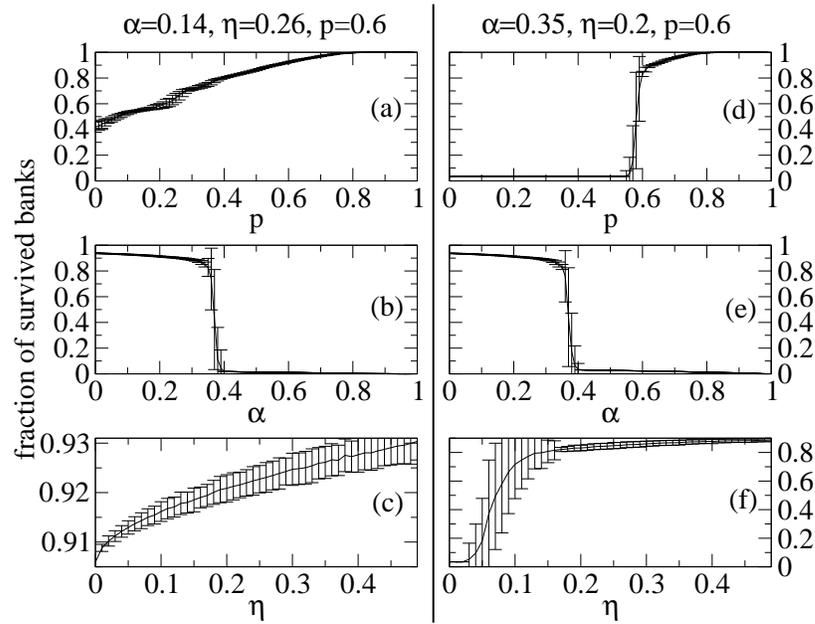


Figure 5.6: Survival rate of banks when asset 0 ( loans for construction and land development ) is initially shocked as function of one parameter with the other two parameters fixed. Average over 300 independent realizations with 95% confidence interval. Left panel: parameter combination  $\alpha = 0.14$ ,  $\eta = 0.26$  and  $p = 0.6$ ; right panel: parameter combination  $\alpha = 0.35$ ,  $\eta = 0.2$  and  $p = 0.6$ .

## 5.5 Discussion

In this paper, we develop a bipartite network model for systemic risk propagation and specifically study the cascading failure process in the banking system. We first study the properties of the defaulting banks during the 2007–2008 financial crisis, and find that they differ from the properties of the survived banks. We then construct a bipartite banking

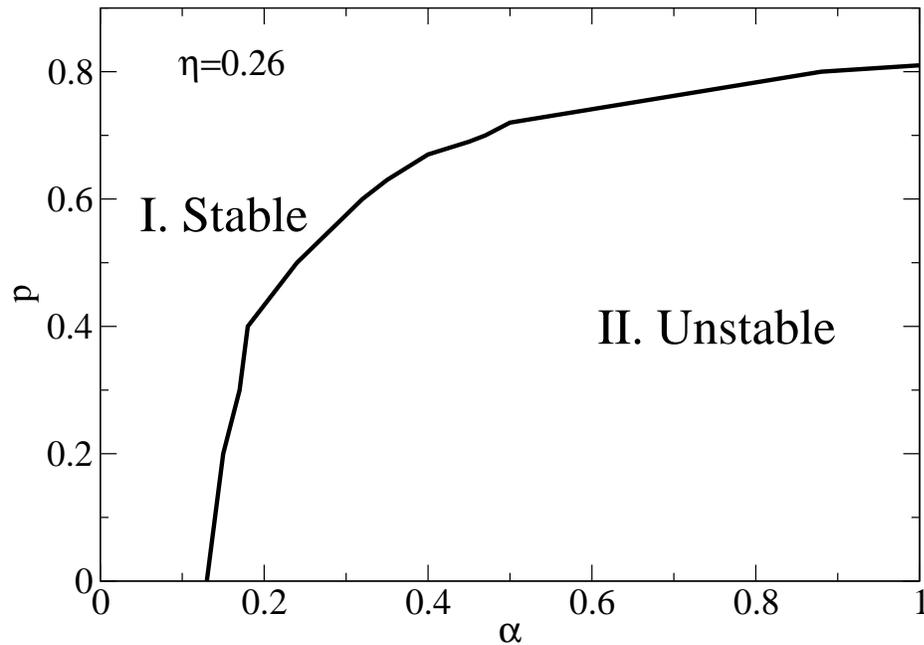


Figure 5.7: Phase diagram for parameter  $\alpha$  and  $p$ , when  $\eta = 0.26$ . The network is stable in region I. Significant part of banks in the network would still survive after cascading failure. In region II, almost all the banks in the network fail after cascading failure.

network that is composed of (i) banks on one hand and (ii) bank assets on the other. We also propose a cascading failure model to simulate the crisis spreading process in banking networks. We introduce a shock into the banking system by reducing a specific asset value and we monitor the cascading effect of this value reduction on banks and on other asset values. We test our model using 2007 balance sheet data by identifying the empirically failed banks between 2008 and 2011, and find through ROC curve analysis that our model simulates well the crisis spreading process and identifies a significant portion of the actual failed banks from the FDIC failed bank database.

Furthermore, studying the cascading failure of banks step by step shows that the complex structure of the bank network indeed contributes to the spreading of financial crisis, which makes a complex network model necessary in describing and predicting the behavior of the banking system. Thus, we suggest that our model could be useful to stress test systemic risk of the banking system. For example, we can stress test the model to predict which banks could be in danger and how many banks could fail if the agricultural assets drop

20% in value. We then offer policy suggestions such as requirement to reduce exposure to agricultural loans or closely monitor vulnerable banks. Then the model also indicates possible ways to mitigate the propagation of financial crisis. From the model we know that risk in the banking system propagates bidirectionally between assets and banks. Suppressing propagation either way could be very helpful to mitigating such catastrophes. The first way is to provide liquidity to the market, thus when distressed banks need to sell assets, the market will not overreact. The second way to curb systemic risk contagion is to ensure that banks are solvent and have healthy balance sheets, i.e. no excess leverage, higher capital requirements, appropriate levels of liquid assets, etc. in order to be able to absorb shocks to the asset value. Possible measures could be to pay a periodic fee to a supervising institution during non-crisis periods in exchange for obtaining emergency liquidity, as proposed by Perotti et al. [111].

Lastly, we show that as the parameters of the system change the bank network can switch between two distinct regions, stable and unstable, which are separated by a so-called phase transition boundary. We suggest that the bank network be understood in complex system terms and that its closeness to the phase transition boundary be diligently monitored in order to forestall system failure.

We suggest that our model for systemic risk propagation might be applicable to other complex systems, e.g., to study the effect of sovereign debt value deterioration on the global banking system or to analyze the impact of depreciation or appreciation of certain currencies on the world economy.

After this work was completed, we learned of the independent work of Caccioli et al. [112], also addressing the challenges of systemic risk due to overlapping portfolios. Their independent results are complementary to ours.

## 5.6 Methods

**Data Sets And Explanations.** We use two data sets in this paper. The first is the Commercial Banks - Balance Sheet Data (CBBSD) from Wharton Research Data Services [113]

for the time period 1/1/1976 to 12/31/2008, which contains the amounts of 13 specific assets and the total assets, total liabilities, and total equities for each bank. We enumerate the assets from 0 to 12 to simplify the problem and categorize the assets into real estate loans, other loans, and other assets. These assets are listed in Table 5.1. We study the data for the year 2007, which contains 7,846 US banks. All banks have total assets data, but 21,171 data spots out of the total  $7,846 \times 13 = 101,998$  data spots for specific assets are blank. For banks with complete data, it is confirmed that the total asset value equals the sum of individual asset. The absent data causes the sum of the individual assets to be lower than the total assets. Furthermore, in some cases, the sum of the individual assets can be smaller than the bank's total liabilities, which leads the banks to fail before any shock is introduced in the model. Thus we need to ensure that the sum of the individual asset values is equal to the total assets value, by allocating the difference between the total asset and available individual assets to the missing assets. If a bank has more than one missing asset, the distribution of the difference to the assets is proportional to the average amount of these assets on the balance sheets of other banks.

The step-by-step methodology is described as follows:

1. For each bank  $i$ , we calculate the weight  $w_{i,m} = \frac{B_{i,m}}{B_i}$  of asset  $m$  in the bank's portfolio.
2. We then calculate the average weight of each asset  $\langle w \rangle_m = \frac{\sum_i w_{i,m}}{N}$ , where  $N$  is the total number of banks.
3. From the total asset and known specific assets, we calculate the total amount for the unknown assets, which is  $(B_i - \sum_{\text{known assets}} B_{i,m})$ . We then distribute this total amount to each unknown asset by their average weight ( $\langle w \rangle_m$ ) ratios. For example, if a bank  $i$  lacks data on asset  $x$  and asset  $y$ , the amount of asset  $x$  is calculated as
$$B_{i,x} = (B_i - \sum_{m \neq x, y} B_{i,m}) \frac{\langle w \rangle_x}{\sum_{m=x,y} \langle w \rangle_m}.$$

The second dataset that we use is the Failed Bank List from the Federal Deposit Insurance Corporation (FBL-FDIC) [114], which shows that 371 banks failed during the 1/1/2008

– 7/1/2011 period and that only 27 banks failed during the 2000–2007 period. We use this representative dataset to empirically test our model for the 2008 financial crisis. Of the 371 banks in the FBL-FDIC dataset, 278 banks are included in the Commercial Banks - Balance Sheet Data dataset in 2007.

	index	Balance Sheet Asset Variables	Rows	$\langle w \rangle_m$
Real Estate Loans	0	Loans for construction and land development	6139	0.082
	1	Loans secured by farmland	5932	0.038
	2	Loans secured by 1-4 family residential properties	7553	0.167
	3	Loans secured by multifamily (>5) residential properties	5381	0.013
	4	Loans secured by nonfarm nonresidential properties.	7495	0.150
Other Loans	5	Agricultural loans	5167	0.041
	6	Commercial and industrial loans	3117	0.031
	7	Loans to individuals	7504	0.097
	8	All other loans	7049	0.171
	9	Obligations (other than securities and leases) of states and political subdivision in the U.S.	7559	0.046
Other Assets	10	Held-to-maturity securities	5924	0.003
	11	Available-for-sale securities, total	3445	0.004
	12	Premises and fixed assets including capitalized lease	7751	0.020

Table 5.1: Description of Commercial Banks - Balance Sheet Data (CBBSD) from Wharton Research Data Services. The third column represents the number of available rows of data of each asset for the year 2007 before completion. The total number of banks in 2007 in the CBBSD is 7846.  $\langle w \rangle_m$  is the average asset weight of banks.

# Chapter 6

## Conclusion

This thesis covers my research on the complex networks field during the past 5 years, including theoretical development of interdependent networks models and application of complex networks models to study economical systems.

In the modern society, our infrastructure systems are more and more coupled. Failure in one system will be non-linearly spread into other systems and cause dramatic cost to the whole society. To design such large robust interdependent systems or to protect the existing interdependent systems has become more and more difficult. Interdependent networks research offers deep insight to these issues. In the second chapter, we study the robustness of interdependent networks under targeted attack on specific degree nodes. We introduce a method and show that targeted-attack problems in networks can be mapped to random-attack problems by transforming the networks which are under initial attack. It provides a routine method (if the random-attack case is solvable) to study the targeted-attack problems in both single networks and randomly connected and uncorrelated interdependent networks, i.e. (i) the case of three or more interdependent networks, (ii) the case of partially coupled interdependent networks, (iii) the case in which a node from network  $A$  can depend on more than one node from network  $B$ . By applying the method, we find that in contrast to single networks, when the highly connected nodes are defended, the percolation threshold  $p_c$  has a finite non-zero value which is significantly larger than zero. For example, when the degrees of all nodes are known and nodes can only be damaged from lower degree to high degree,

$p_c \approx 0.46$  for coupled SF networks with  $\lambda = 2.8$  and  $\langle k \rangle = 4$  while  $p_c$  for the same single SF network is 0. The implications of the study are dramatic. It indicates that current methods applied to design robust networks and improve the robustness of current networks, i.e. protecting the high degree nodes, need to be modified to apply to interdependent network systems. Then in the third chapter, we study how clustering influences the robustness of interdependent networks. Since usually degree-degree correlation is inevitable when clustering is brought into network, we also derive an analytical expression, for degree-degree correlation as a function of the clustering coefficient. Such that we can study the influence of clustering alone on the robustness of interdependent networks. We conclude that  $p_c$  for interdependent networks increases when networks are more highly clustered. This occurs because clustered networks contain some links in triangles that do not contribute to the giant component, and in each stage of cascading failure the giant component will be smaller than in the unclustered case.

Complex networks models are widely studied because they offer tools to describe the complex systems in our reality. One group of the systems that can be successfully described by networks are the economical systems. The relationship developed between people and institutions during business practice naturally form a network, which can serve for the good, i.e. transfer information and influence, or serve for the bad, i.e. spread of crisis. In this thesis, we apply the complex networks model to describe two economical systems: i) board of directors' network; and ii) commercial bank network. In chapter 4, we analyze the power of directors in the US corporate governance network through complex networks methodology. To measure the influence of directors, we develop a *influence factor* measure, which offers an objective and quantitative way of describing how information and influence are transferred through the network, which enables us to determine the power of directors. Tested by influential people lists of popular magazines, we find that the *influence factor* measure is consistently either the best or one of the two best methods in identifying influential people. We find that contrary to commonly accepted belief that directors of large companies are most powerful, in some instances, influential directors do not serve on boards of large

companies. In chapter 5, we develop a bipartite network model for systemic risk propagation and specifically study the cascading failure process in the banking system. We first study the properties of the defaulting banks during the 2007–2008 financial crisis, and find that they differ from the properties of the survived banks. We then construct a bipartite banking network that is composed of (i) banks on one hand and (ii) bank assets on the other. We also propose a cascading failure model to simulate the crisis spreading process in banking networks. We introduce a shock into the banking system by reducing a specific asset value and we monitor the cascading effect of this value reduction on banks and on other asset values. We test our model using 2007 balance sheet data by identifying the empirically failed banks between 2008 and 2011, and find through ROC curve analysis that our model simulates well the crisis spreading process and identifies a significant portion of the actual failed banks from the FDIC failed bank database. We show that as the parameters of the model change the bank network can switch between two distinct regions, stable and unstable, which are separated by a so-called phase transition boundary. We suggest that our model for systemic risk propagation might be applicable to other complex systems, e.g., to study the effect of sovereign debt value deterioration on the global banking system or to analyze the impact of depreciation or appreciation of certain currencies on the world economy.

# Bibliography

- [1] Steven M Rinaldi, James P Peerenboom, and Terrence K Kelly. Identifying, understanding, and analyzing critical infrastructure interdependencies. *Control Systems, IEEE*, 21(6):11–25, 2001.
- [2] Jean-Claude Laprie, Karama Kanoun, and Mohamed Kaâniche. Modelling interdependencies between the electricity and information infrastructures. *Computer Safety, Reliability, and Security*, pages 54–67, 2007.
- [3] Stefano Panzieri and Roberto Setola. Failures propagation in critical interdependent infrastructures. *International Journal of Modelling, Identification and Control*, 3(1):69–78, 2008.
- [4] V Rosato, L Issacharoff, F Tiriticco, S Meloni, S Porcellinis, and R Setola. Modelling interdependent infrastructures using interacting dynamical models. *International Journal of Critical Infrastructures*, 4(1):63–79, 2008.
- [5] EA Leicht and Raissa M D’Souza. Percolation on interacting networks. *arXiv preprint arXiv:0907.0894*, 2009.
- [6] Alessandro Vespignani. Complex networks: The fragility of interdependency. *Nature*, 464(7291):984–985, 2010.
- [7] Sergey V Buldyrev, Roni Parshani, Gerald Paul, H Eugene Stanley, and Shlomo Havlin. Catastrophic cascade of failures in interdependent networks. *Nature*, 464(7291):1025–1028, 2010.

- [8] Roni Parshani, Sergey V Buldyrev, and Shlomo Havlin. Interdependent networks: Reducing the coupling strength leads to a change from a first to second order percolation transition. *Physical Review Letters*, 105(4):48701, 2010.
- [9] Robert M May, Simon A Levin, and George Sugihara. Complex systems: ecology for bankers. *Nature*, 451(7181):893–895, 2008.
- [10] Frank Schweitzer, Giorgio Fagiolo, Didier Sornette, Fernando Vega-Redondo, Alessandro Vespignani, and Douglas R White. Economic networks: The new challenges. *science*, 325(5939):422, 2009.
- [11] Antonios Garas, Panos Argyrakis, Céline Rozenblat, Marco Tomassini, and Shlomo Havlin. Worldwide spreading of economic crisis. *New journal of Physics*, 12(11):113043, 2010.
- [12] Neil Johnson and Thomas Lux. Financial systems: Ecology and economics. *Nature*, 469(7330):302–303, 2011.
- [13] Andrew G Haldane and Robert M May. Systemic risk in banking ecosystems. *Nature*, 469(7330):351–355, 2011.
- [14] Stefano Battiston, Michelangelo Puliga, Rahul Kaushik, Paolo Tasca, and Guido Caldarelli. Debtrank: Too central to fail? financial networks, the fed and systemic risk. *Scientific Reports*, 2, 2012.
- [15] P. Erdős and A. Rényi. On random graphs i. *Publicationes Mathematicae (Debrecen)*, 6:290, 1959.
- [16] P. Erdős and A. Rényi. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17, 1960.
- [17] P. Erdős and A. Rényi. On the strength of connectedness of a random graph. *Acta Mathematica Scientia Hungary*, 12:261–267, 1961.

- [18] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47, 2002.
- [19] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. *ACM SIGCOMM Computer Communication Review*, 29(4):251–262, 1999.
- [20] Albert-László Barabási, Réka Albert, and Hawoong Jeong. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A: Statistical Mechanics and its Applications*, 281(1):69–77, 2000.
- [21] Holger Ebel, Lutz-Ingo Mielsch, and Stefan Bornholdt. Scale-free topology of e-mail networks. *Physical Review E*, 66:035103, 2002.
- [22] Sidney Redner. How popular is your paper? an empirical study of the citation distribution. *The European Physical Journal B-Condensed Matter and Complex Systems*, 4(2):131–134, 1998.
- [23] Hawoong Jeong, Bálint Tombor, Réka Albert, Zoltan N Oltvai, and A-L Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, 2000.
- [24] Alain Barrat, Marc Barthélemy, Romualdo Pastor-Satorras, and Alessandro Vespignani. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(11):3747–3752, 2004.
- [25] Victor M Eguiluz, Dante R Chialvo, Guillermo A Cecchi, Marwan Baliki, and A Vania Apkarian. Scale-free brain functional networks. *Physical review letters*, 94(1):18102, 2005.
- [26] Mark EJ Newman. A measure of betweenness centrality based on random walks. *Social networks*, 27(1):39–54, 2005.
- [27] Duncan Watts and S Strogatz. Collective dynamics of small-world networks. *Nature*, 393:440–442, 1998.

- [28] Mark EJ Newman and Juyong Park. Why social networks are different from other types of networks. *Physical Review E*, 68(3):036122, 2003.
- [29] M Ángeles Serrano and Marian Boguna. Clustering in complex networks. i. general formalism. *Physical Review E*, 74(5):056114, 2006.
- [30] Phillip Bonacich. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2(1):113–120, 1972.
- [31] Phillip Bonacich. Power and centrality: A family of measures. *American journal of sociology*, pages 1170–1182, 1987.
- [32] Mark EJ Newman, Steven H Strogatz, and Duncan J Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64(2):026118, 2001.
- [33] Mark EJ Newman. Spread of epidemic disease on networks. *Physical Review E*, 66(1):016128, 2002.
- [34] Adilson E Motter and Ying-Cheng Lai. Cascade-based attacks on complex networks. *Physical Review E*, 66(6):065102, 2002.
- [35] Ingve Simonsen, Lubos Buzna, Karsten Peters, Stefan Bornholdt, and Dirk Helbing. Transient dynamics increasing network vulnerability to cascading failures. *Physical Review Letters*, 100(21):218701, 2008.
- [36] SE Chang. Infrastructure resilience to disasters. *The Bridge*, 39(4):36–41, 2009.
- [37] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, 2000.
- [38] Duncan S Callaway, Mark EJ Newman, Steven H Strogatz, and Duncan J Watts. Network robustness and fragility: Percolation on random graphs. *Physical Review Letters*, 85(25):5468–5471, 2000.

- [39] Reuven Cohen, Keren Erez, Daniel Ben-Avraham, and Shlomo Havlin. Breakdown of the internet under intentional attack. *Physical Review Letters*, 86(16):3682–3685, 2001.
- [40] Lazaros K Gallos, Reuven Cohen, Panos Argyrakis, Armin Bunde, and Shlomo Havlin. Stability and topology of scale-free networks under attack and defense strategies. *Physical Review Letters*, 94(18):188701, 2005.
- [41] Andre A Moreira, José S Andrade Jr, Hans J Herrmann, and Joseph O Indekeu. How to make a fragile network robust and vice versa. *Physical Review Letters*, 102(1):18701, 2009.
- [42] Alessia Annibale, ACC Coolen, and Ginestra Bianconi. Network resilience against intelligent attacks constrained by the degree-dependent node removal cost. *Journal of Physics A: Mathematical and Theoretical*, 43(39):395001, 2010.
- [43] Petter Holme, Beom Jun Kim, Chang No Yoon, and Seung Kee Han. Attack vulnerability of complex networks. *Physical Review E*, 65(5):056109, 2002.
- [44] Reuven Cohen, Shlomo Havlin, and Daniel Ben-Avraham. Efficient immunization strategies for computer networks and populations. *Physical Review Letters*, 91(24):247901, 2003.
- [45] Jia Shao, Sergey V Buldyrev, Lidia A Braunstein, Shlomo Havlin, and H Eugene Stanley. Structure of shells in complex networks. *Physical Review E*, 80(3):036105, 2009.
- [46] Roni Parshani, Sergey V Buldyrev, and Shlomo Havlin. Critical effect of dependency groups on the function of networks. *Proceedings of the National Academy of Sciences*, 108(3):1007–1010, 2011.
- [47] Béla Bollobás. *Random graphs*, volume 73. Cambridge University press, 2001.

- [48] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- [49] Reuven Cohen and Shlomo Havlin. *Complex networks: structure, robustness and function*. Cambridge University Press, 2010.
- [50] US-Canada Power System Outage Task Force, Spencer Abraham, Herb Dhaliwal, R John Efford, Linda J Keen, Anne McLellan, John Manley, Kenneth Vollman, Nils J Diaz, Tom Ridge, et al. *Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations*. US-Canada Power System Outage Task Force, 2004.
- [51] James Peerenboom, R Fischer, and Ronald Whitfield. Recovering from disruptions of interdependent critical infrastructures. In *Proc. CRIS/DRM/IIIT/NSF Workshop Mitigat. Vulnerab. Crit. Infrastruct. Catastr. Failures*, 2001.
- [52] Osman Yagan, Dajun Qian, Junshan Zhang, and Douglas Cochran. Optimal allocation of interconnecting links in cyber-physical systems: Interdependence, cascading failures, and robustness. *Parallel and Distributed Systems, IEEE Transactions on*, 23(9):1708–1720, 2012.
- [53] Amir Bashan, Ronny P Bartsch, Jan W Kantelhardt, Shlomo Havlin, and Plamen Ch Ivanov. Network physiology reveals relations between network topology and physiological function. *Nature Communications*, 3:702, 2012.
- [54] Rae Zimmerman. Decision-making and the vulnerability of interdependent critical infrastructure. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, volume 5, pages 4059–4063. IEEE, 2004.
- [55] David Mendonça and William A Wallace. Impacts of the 2001 world trade center attack on new york city critical infrastructures. *Journal of Infrastructure Systems*, 12(4):260–270, 2006.

- [56] Benoit Robert and Luciano Morabito. The operational tools for managing physical interdependencies among critical infrastructures. *International Journal of Critical Infrastructures*, 4(4):353–367, 2008.
- [57] Dorothy A Reed, Kailash C Kapur, and Richard D Christie. Methodology for assessing the resilience of networked infrastructure. *Systems Journal, IEEE*, 3(2):174–180, 2009.
- [58] Ebrahim Bagheri and Ali A Ghorbani. Uml-ci: A reference model for profiling critical infrastructure systems. *Information Systems Frontiers*, 12(2):115–139, 2010.
- [59] Daniel Mansson, Rajeev Thottappillil, and M Backstrom. Methodology for classifying facilities with respect to intentional emi. *Electromagnetic Compatibility, IEEE Transactions on*, 51(1):46–52, 2009.
- [60] Jonas Johansson and Henrik Hassel. An approach for modelling interdependent infrastructures in the context of vulnerability analysis. *Reliability Engineering & System Safety*, 95(12):1335–1344, 2010.
- [61] Amir Bashan, Yehiel Berezin, Sergey V Buldyrev, and Shlomo Havlin. The extreme vulnerability of interdependent spatially embedded networks. *arXiv preprint arXiv:1206.2062*, 2012.
- [62] Michael JO Pocock, Darren M Evans, and Jane Memmott. The robustness and restoration of a network of ecological networks. *Science*, 335(6071):973–977, 2012.
- [63] Xuqing Huang, Jianxi Gao, Sergey V Buldyrev, Shlomo Havlin, and H Eugene Stanley. Robustness of interdependent networks under targeted attack. *Physical Review E*, 83(6):065101, 2011.
- [64] Jia Shao, Sergey V Buldyrev, Shlomo Havlin, and H Eugene Stanley. Cascade of failures in coupled network systems with multiple support-dependence relations. *Physical Review E*, 83(3):036116, 2011.

- [65] Jianxi Gao, Sergey V Buldyrev, H Eugene Stanley, and Shlomo Havlin. Networks formed from interdependent networks. *Nature Physics*, 8(1):40–48, 2011.
- [66] Jianxi Gao, Sergey V Buldyrev, Shlomo Havlin, and H Eugene Stanley. Robustness of a network of networks. *Physical Review Letters*, 107(19):195701, 2011.
- [67] Mark EJ Newman. Random graphs with clustering. *Physical review letters*, 103(5):58701, 2009.
- [68] Joel C Miller. Percolation and epidemics in random clustered networks. *Physical Review E*, 80(2):020901, 2009.
- [69] James P. Gleeson. Bond percolation on a class of clustered random networks. *Phys. Rev. E*, 80:036107, Sep 2009.
- [70] James P Gleeson, Sergey Melnik, and Adam Hackett. How clustering affects the bond percolation threshold in complex networks. *Physical Review E*, 81(6):066114, 2010.
- [71] Adam Hackett, Sergey Melnik, and James P Gleeson. Cascades on a class of clustered random networks. *Physical Review E*, 83(5):056107, 2011.
- [72] Chai Molina and Lewi Stone. Modelling the spread of diseases in clustered networks. *Journal of Theoretical Biology*, 2012.
- [73] V Zlatić, Diego Garlaschelli, and Guido Caldarelli. Networks with arbitrary edge multiplicities. *EPL (Europhysics Letters)*, 97(2):28005, 2012.
- [74] Piet Van Mieghem, Huijuan Wang, Xin Ge, Siyu Tang, and FA Kuipers. Influence of assortativity and degree-preserving rewiring on the spectra of networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 76(4):643–652, 2010.
- [75] Di Zhou, Gregorio D’Agostino, Antonio Scala, and H Eugene Stanley. Assortativity decreases the robustness of interdependent networks. *arXiv preprint arXiv:1203.0029*, 2012.

- [76] Brian Karrer and MEJ Newman. Random graphs containing arbitrary distributions of subgraphs. *Physical Review E*, 82(6):066118, 2010.
- [77] Michael Useem. *The inner circle: Large corporations and the rise of business political activity in the US and UK*. Oxford University Press, USA, 1986.
- [78] Gerald F Davis. Agents without principles? the spread of the poison pill through the intercorporate network. *Administrative Science Quarterly*, pages 583–613, 1991.
- [79] Mark S Mizruchi. What do interlocks do? an analysis, critique, and assessment of research on interlocking directorates. *Annual review of sociology*, pages 271–298, 1996.
- [80] Stefano Battiston, Gérard Weisbuch, and Eric Bonabeau. Decision spread in the corporate board network. *Advances in Complex Systems*, 6(04):631–644, 2003.
- [81] Pamela R Haunschild and Christine M Beckman. When do interlocks matter?: Alternate sources of information and interlock influence. *Administrative Science Quarterly*, pages 815–844, 1998.
- [82] Ulrik Brandes. A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25(2):163–177, 2001.
- [83] Stanley Wasserman and Katherine Faust. *Social network analysis: Methods and applications*, volume 8. Cambridge university press, 1994.
- [84] Shai Carmi, Shlomo Havlin, Scott Kirkpatrick, Yuval Shavitt, and Eran Shir. A model of internet topology using k-shell decomposition. *Proceedings of the National Academy of Sciences*, 104(27):11150–11154, 2007.
- [85] Maksim Kitsak, Massimo Riccaboni, Shlomo Havlin, Fabio Pammolli, and H Eugene Stanley. Scale-free models for the structure of business firm networks. *Physical Review E*, 81(3):036117, 2010.

- [86] Maksim Kitsak, Lazaros K Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H Eugene Stanley, and Hernán A Makse. Identification of influential spreaders in complex networks. *Nature Physics*, 6(11):888–893, 2010.
- [87] Phillip Bonacich. Technique for analyzing overlapping memberships. *Sociological methodology*, 4:176–185, 1972.
- [88] Peter Mariolis. Interlocking directorates and control of corporations: The theory of bank control. *Social Science Quarterly*, 56(3):425–439, 1975.
- [89] Mark S Mizruchi and G William Domhoff. *The American corporate network, 1904-1974*. Sage Beverly Hills, CA, 1982.
- [90] Beth A Mintz and Michael Schwartz. *The power structure of American business*. University of Chicago Press, 1985.
- [91] Riskmetrics (iss) - directors data. <https://wrds-web.wharton.upenn.edu/wrds/>, [Online; accessed 01-Jun-2010].
- [92] Stefano Battiston and Michele Catanzaro. Statistical properties of corporate board and director networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 38(2):345–352, 2004.
- [93] Mark EJ Newman. Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. *Physical review E*, 64(1):016132, 2001.
- [94] Vidhi Chhaochharia and Yaniv Grinstein. The changing structure of us corporate boards: 1997–2003. *Corporate Governance: An International Review*, 15(6):1215–1223, 2007.
- [95] Mark Newman, Albert-Laszlo Barabasi, and Duncan J Watts. *The structure and dynamics of networks*. Princeton University Press, 2011.

- [96] Ashley G Smart, Luis AN Amaral, and Julio M Ottino. Cascading failure and robustness in metabolic networks. *Proceedings of the National Academy of Sciences*, 105(36):13223–13228, 2008.
- [97] Simon Wells. Uk interbank exposures: systemic risk implications. *Financial Stability Review*, 13:175–182, 2002.
- [98] Craig H Furfine. Interbank exposures: Quantifying the risk of contagion. *Journal of Money, Credit and Banking*, pages 111–128, 2003.
- [99] Christian Upper and Andreas Worms. Estimating bilateral exposures in the german interbank market: Is there a danger of contagion? *European Economic Review*, 48(4):827–849, 2004.
- [100] Helmut Elsinger, Alfred Lehar, and Martin Summer. Risk assessment for banking systems. *Management science*, 52(9):1301–1314, 2006.
- [101] Erlend Nier, Jing Yang, Tanju Yorulmazer, and Amadeo Alentorn. Network models and financial stability. *Journal of Economic Dynamics and Control*, 31(6):2033–2060, 2007.
- [102] Rodrigo Cifuentes, Gianluigi Ferrucci, and Hyun Song Shin. Liquidity risk and contagion. *Journal of the European Economic Association*, 3(2-3):556–566, 2005.
- [103] Igor Tsatskis. Systemic losses in banking networks: indirect interaction of nodes via asset prices. *Available at SSRN 2062174*, 2012.
- [104] Rebel A Cole and Lawrence J White. Déjà vu all over again: The causes of us commercial bank failures this time around. *Journal of Financial Services Research*, 42(1):5–29, 2012.
- [105] Gary S Corner. Agriculture banks are outperforming their peers, but how long will it last? *Central Banker*, (Fall), 2011.

- [106] Yadav Gopalan. Earliest indicator of bank failure is deterioration in earnings. *Central Banker*, (Spring), 2010.
- [107] Robert C Merton. On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, 29(2):449–470, 1974.
- [108] Peter Crosbie and Jeff Bohn. Modeling default risk. 2003.
- [109] Dale F Gray, Robert C Merton, and Zvi Bodie. Contingent claims approach to measuring and managing sovereign credit risk. *Journal of Investment Management*, 5(4):5, 2007.
- [110] John A Swets. *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers*. Lawrence Erlbaum Associates, Inc, 1996.
- [111] Enrico Perotti and Javier Suarez. Liquidity insurance for systemic crises. 2009.
- [112] Fabio Caccioli, Munik Shrestha, Cristopher Moore, and J Farmer. Stability analysis of financial contagion due to overlapping portfolios. *Available at SSRN: <http://ssrn.com/abstract=2176080> or <http://dx.doi.org/10.2139/ssrn.2176080>*, 2012.
- [113] Commercial banks balance sheet data. <https://wrds-web.wharton.upenn.edu/wrds/>, [Online; accessed 01-Jul-2011].
- [114] Fdic failed banks list. <http://www.fdic.gov/bank/individual/failed/banklist.html>, [Online; accessed 01-Jul-2011].

# Curriculum Vitae

## Xuqing Huang

Boston University, Physics Department  
590 Commonwealth Avenue  
Boston, Massachusetts 02215 USA

Telephone: 857-654-7486

E-mail: eqing@buphy.bu.edu

### EDUCATION

- 2013, Ph.D. Physics, Boston University, Boston, MA, USA

Advisor: H. Eugene Stanley

Thesis: *Network Theory and its Application in Economics Systems*

- 2010, M.A. Physics, Boston University, Boston, MA, USA
- 2007, B.A. Physics, University of Science and Technology of China, Hefei, Anhui, China.

### EXPERIENCE

- May '12 – Aug '12, Quantitative Strategist Summer Associate, Goldman Sachs Group.
- Sep '08 – Present, Research assistant, Boston Univ..
- Sep '07 – Aug '08, Teaching assistant, Boston Univ..

### TALKS

- International Conference on Econophysics (2011), Shanghai, China.
- International School and Conference in Networks Sciences (2010), MIT, Boston, USA.

## PUBLICATIONS

1. S. Shao, X. Huang, J. Shao, H. E. Stanley and S. Havlin, *Percolation of local attacks on random networks*, [working paper].
2. X. Huang, Q. Li, J. Shie, S. Havlin and H. E. Stanley, *Bipartite networks: increasing survival threshold leads to a change of from second order to first order phase transition*, [working paper].
3. X. Huang, D. Kenett, I. Vodenska, S. Havlin and H. E. Stanley, *Partial Correlation Analysis of Global Stock Markets*, [working paper].
4. S. Shao, X. Huang, H. E. Stanley and S. Havlin, *Robustness of Partially Interdependent Network of Networks*, [preprint].
5. X. Huang, S. Shao, H. Wang, S. V. Buldyrev, H. E. Stanley, and S. Havlin, *The Robustness of Interdependent Clustered Networks*, Europhysics Letters (EPL) bf 101, 18002 (2013).
6. X. Huang, I. Vodenska, S. Havlin, and H. E. Stanley, *Cascading Failures in Bi-partite Graphs: Model for Systemic Risk Propagation*, Scientific Reports **3**, 1219 (2013).
7. X. Huang, J. Gao, S. V. Buldyrev, S. Havlin, H. E. Stanley, *Robustness of Interdependent Networks Under Targeted Attack*, Physical Review E **83**, 065101(R) (2011).
8. X. Huang, I. Vodenska, F. Wang, S. Havlin, H. E. Stanley, *Identify the Most Influential Directors in US Corporate Governance Network*, Physical Review E **84**, 046101 (2011).