2013

Statistical physics approaches to complex systems

Li, Wei

http://hdl.handle.net/2144/15162

Boston University
BOSTON UNIVERSITY
GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

STATISTICAL PHYSICS APPROACHES TO COMPLEX SYSTEMS

by

WEI LI
B.S., 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
I dedicate this dissertation to my parents, Xiue Ling and Wenqiong Li
The thesis is one of the greatest rewards on the journey of pursuing my Ph.D. degree. I have traveled with support and encouragement from many people to complete this challenge.

I would like to thank my parents for bringing me to the world, raising me up and supporting me whatever I am hunting for and wherever I am heading to.

I would like to express my sincere gratitude to my thesis advisor, Professor H. Eugene Stanley. This work would not have been possible without his guidance and support. Gene introduced me to this fantastic field of research as well as his collaboration networks. I benefited a great mount from his inspiration to science, dedication to work and commitment to the whole research group. I am grateful for being one of his students and being exposure to the frontier of world class research.

I would like to thank Professor Shlomo Havlin for guiding and encouraging me throughout my thesis. I feel lucky and honored to collaborate with him.

I am thankful to other collaborators, Professor Sergey V. Buldyrev, Professor Kazuko Yamasaki, Fengzhong Wang, Amir Bashan, Dror Kenett for sharing knowledge and experience.

I am grateful to Professor William Skocpol for his enormous patience and critical comments on this dissertation.

I would like to acknowledge my thesis committee, Professor H. Eugene Stanley, Professor William Skocpol, Professor Karl Ludwig, Professor Tulika Bose. I also thank Professor Richard Averitt, and Professor Bennett Goldberg for reading the
abstract of this dissertation.

I would like to thank all the friends in the group who made my years at Boston University cheerful and cherishable. To mention a few, Jia Shao, Guanliang Li, Alex Petersen, Jiayuan Luo, Joel Tenenbaum, Elena Strekalova, Erik Lascaris, Xuqing Huang, Zhiqiang Su, Qian Li, Duan Wang, Di Zhou, Ling Feng, Shuai Shao, Xin Zhang, Wen Fang, MeiChu Chang and many others.

I also would like to thank the staff at Center for Polymer Studies and Physics Department, especially Bob Tomposki and Mirtha Cabello, for helping me all these years.
STATISTICAL PHYSICS APPROACHES TO COMPLEX SYSTEMS

( Order No. )

WEI LI

Boston University Graduate School of Arts and Sciences, 2013

Major Professor: H. Eugene Stanley, Professor of Physics

ABSTRACT

This thesis utilizes statistical physics concepts and mathematical modeling to study complex systems. I investigate the emergent complexities in two systems: (i) the stock volume volatility in the United States stock market system; (ii) the robustness of networks in an interdependent lattice network system.

In Part I, I analyze the United States stock market data to identify how several financial factors significantly affect scaling properties of volume volatility time intervals. I study the daily trading volume volatility time intervals \( \tau \) between two successive volume volatilities above a certain threshold \( q \), and find a range of power law distributions. I also study the relations between the form of these distribution functions and several financial factors: stock lifetimes, market capitalization, volume, and trading value. I find that volume volatility time intervals are short-term correlated. I also find that the daily volume volatility shows a stronger long-term correlation for sequences of longer lifetimes.

In Part II, I apply percolation theory to interacting complex networks. The dependency links between the two square lattice networks have a typical length \( r \) lattice units. For two nodes connecting by a dependency link, one node fails once the node on which it depends in the other network fails. I show that rich phase transition phenomena exist when the length of the dependency links \( r \) changes. The results suggest that percolation for small \( r \) is a second-order transition, and for larger \( r \) is a first-
order transition. The study suggests that interdependent infrastructures embedded in two-dimensional space become most vulnerable when the interdependent distance is in the intermediate range, which is much smaller than the size of the system.
# Contents

1 Introduction ........................................ 1
   1.1 Motivation of Thesis .......................... 1
   1.2 Organization of Thesis ........................ 4

2 Scaling and Memory of Trading Volume in Stock Markets ............. 6
   2.1 Background .................................... 6
   2.2 Data .......................................... 8
   2.3 Notation ..................................... 9
   2.4 Definition of Volume Volatility and Volume Return .......... 9
   2.5 Definition of Volume Return Interval ................. 10
   2.6 Distribution of Volume Volatility Time Intervals .......... 11
   2.7 Financial Factor Influences ...................... 15
   2.8 Short-term Memory Effects ...................... 19
   2.9 Detrended Fluctuation Analysis Method ................. 22
   2.10 Long-term Memory Effects ..................... 24
   2.11 Conclusions .................................. 26

3 Cascading Failures in Complex Networks ..................................... 27
   3.1 Background ................................... 27
   3.2 Overview of Complex Network Studies .................... 28
      3.2.1 Erdös-Rényi Networks .................... 28
3.2.2 Scale-Free Networks ........................................ 29
3.3 Percolation Theory ........................................... 29
3.4 Interdependent Complex Networks ......................... 30
3.5 Interdependent Lattice Networks .......................... 31
3.6 Cascading Failures .......................................... 33
3.7 Cascading Failures of Various Lengths of Dependency . 37
3.8 Propagation of the Flat Interface .......................... 43
3.9 Critical $p_c$ and Dependency Length $r$ .................. 46
3.10 Conclusions ................................................. 48

4 Summary .......................................................... 52

Bibliography ....................................................... 54

Curriculum Vitae ................................................... 61
## List of Figures

2.1 Illustration of volume volatility time intervals .......................... 12
2.2 Probability distributions of volume volatility time intervals .......... 13
2.3 Four financial factors influence ............................................. 16
2.4 Power-law tail exponents ................................................... 18
2.5 Scatter plots of the relations between financial factors ................. 20
2.6 Short-term memory of volume volatility time intervals............... 21
2.7 Long-term memory of volume volatility return intervals ............... 25

3.1 A diagram of two interdependent lattice networks ..................... 32
3.2 Giant component size as a function of step ............................. 35
3.3 Schematic graphical solution of percolation giant component size ..... 36
3.4 Giant component as a function of remaining fraction .................. 38
3.5 Largest and second largest cluster sizes as dependent distance changes 39
3.6 Complementary cumulative distribution of cluster size per system .... 41
3.7 Three typical behaviors of interdependent lattices ..................... 42
3.8 Pictures of flat interface propagation ................................. 44
3.9 A diagram of two lattice networks with a gap .......................... 45
3.10 Giant component as dependent distance changes ..................... 47
3.11 Diameter of the hole size $\xi_h$ as a function of $r$ .................... 49
3.12 The critical $p_c$ vs. interdependent distance $r$ ..................... 50
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMEX</td>
<td>American Stock Exchange</td>
</tr>
<tr>
<td>CDF</td>
<td>Cumulative Distribution Function</td>
</tr>
<tr>
<td>CRSP</td>
<td>Center for Research in Security Prices</td>
</tr>
<tr>
<td>DFA</td>
<td>Detrended Fluctuation Analysis</td>
</tr>
<tr>
<td>DJIA</td>
<td>Dow Jones Industrials Average</td>
</tr>
<tr>
<td>ER</td>
<td>Erdős-Rényi</td>
</tr>
<tr>
<td>NASDAQ</td>
<td>National Association of Securities Dealers Automated Quotations</td>
</tr>
<tr>
<td>NYSE</td>
<td>New York Stock Exchange</td>
</tr>
<tr>
<td>PDF</td>
<td>Probability Density Function</td>
</tr>
<tr>
<td>SF</td>
<td>Scale-Free</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>Standard &amp; Pool’s 500 Index</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Motivation of Thesis

Interdisciplinary science is a growing field, which applies methods across traditional boundaries between academic disciplines to address research topics that have traditionally been analyzed within a specific field. The dynamics of complex systems is of great importance with the emerging interest in interdisciplinary science. A complex system is a system composed of many interconnected parts that as a whole exhibit one or more properties which are not obvious from the properties of the individual parts. Examples of complex systems include human economies, climate, nervous systems, and modern energy infrastructures, as well as social network structures. Fields that specialize in the study of complex systems include natural science, mathematics and social science. Statistical physics approaches are widely applied to study complex systems.

Statistical physicists study systems consisting of a collection of interacting units. The collection is complex and everything depends on everything else. It is impossible to predict the exact behavior of the individual unit. Hence, the interesting question is: how to make statistical predictions regarding the collective behavior of the units. Physicists are looking for universal laws that will help us to understand this collective
behavior. Universal laws can be described by scaling property. A system obeys a scaling law if its behavior is characterized by the same functional form over a certain range of scales. Recently, many systems comprising of a large number of interacting units are reported to obey universal laws. In physical systems, these universal properties do not depend on the specific form of the interactions. One conjecture will be that universal laws also exist in economic and social systems.

 Scaling is an important property to describe the collective behavior of interacting units. However, the underlying time organization of the data is not fully characterized by the scaling laws. If the time series is uncorrelated, the data set is fully described by the distribution function. If the time series is auto-correlated, the order of the data points is of great interest because it reveals how the sequence is constructed.

 Financial markets are typical complex systems consisting of a large number of agents with various anticipations, risk tolerances, skills and accessible information. Moreover, financial markets are affected, in a large degree, by external information. A small perturbation may trigger a large shock on the market and a large shock may lead to a series of successive shocks. Hence, financial markets are ideal systems for studying complexity.

 In particular, the dynamics of stock price and trading volume have been studied for decades in order to understand financial markets and to develop investment strategies. Econophysics research has found that the distribution of stock price returns has a power law tail and that the price volatility time series has a long-term power law correlation. The distribution of the scaled price time interval $\tau/\langle\tau\rangle$ can be approximated by a stretched exponential function. Historically, large price movements are typically associated with high trading volume. Understanding the precise relationship between price and volume fluctuations has been a topic of great interest in recent research.

 Another important approach is to study complex systems by modeling them as complex networks. The field of complex network systems has attracted growing inter-
est in the past twenty years. Prompted by enormous accumulations of data collected in economic, biological, and technological systems, complex networks become an important tool to understand the behaviors of a large system as a whole, and to mimic the interactions between the components.

Besides the fundamental need to understand the complex systems, the political and social environment today start to draw more attention on networks. The war on terror is largely about destroying the terrorist message-spreading networks, for example, internet chat room. The prevention of epidemic explosion, such as Influenza A (H1N1) virus and Severe acute respiratory syndrome (SARS), demand understanding of epidemiological networks.

Today, networks are becoming increasingly dependent on one another in order to provide proper functionality. There is a growing interest in studying the robustness of interdependent networks subject to cascading failure. In interdependent networks, nodes from one network depend on nodes from another network and vice versa. When a node from one network fails, it causes the corresponding node in the other network to fail as well. Thus, when some initial failures of nodes happen, it may trigger an iterative process of cascading failures and may destroy all the networks completely.

However, most previous studies of the robustness of interdependent networks focused on random networks without considering spatial restrictions. Most real networks, indeed, are embedded either in two-dimensional or in three-dimensional space. The nodes in each network are interdependent with nodes in other networks. Spatial constraints, such as the network dimensionality, influence the interdependent network properties dramatically. The question about the resilience of interdependent spatial networks becomes of much interest. In this thesis, a more realistic coupled network system is developed to address this question.
1.2 Organization of Thesis

The following parts of this thesis are organized as below.

In Chapter 2, we analyze the trading volume data in U.S. stock markets. We study the daily trading volume volatility time intervals and find a range of power law distributions. We find a unique scaling of the probability density function (PDF) $P_q(\tau)$ for different thresholds $q$. We also perform a detailed analysis of the relations between volume volatility time intervals and four financial stock factors: (i) stock lifetimes, (ii) market capitalization, (iii) average trading volume, and (iv) average trading value. We find systematically different power law exponents for the PDF $P_q(\tau)$ when binning stocks according to these four financial factors. We analyze the conditional probability distribution of volume volatility time interval, $P_q(\tau|\tau_0)$ for $\tau$ following a certain interval $\tau_0$. We find that, immediately following a short (long) time interval, a second short (long) time interval tends to occur, which demonstrates that volume return intervals are short-term correlated. We also find that there is a long-term correlation in the daily volume volatility.

In Chapter 3, we study the cascading failures in a system composed of two interdependent square lattice networks. Most real networks are embedded in space and the nodes in different networks are in many cases interdependent on each other. However, the spatial restrictions are mostly not considered in the theory of interdependent networks. In order to take into account the spatial restrictions, we study a system composed of two interdependent square lattice networks A and B, where nodes in network A depend on randomly chosen nodes within a certain distance $r$ lattice units of its corresponding node in network B and vice versa. For two nodes connecting by a dependency link, one node fails once the node on which it depends in the other network fails.

We find that mutual percolation in two interdependent lattice networks system for small $r$ is a second order phase transition, while for larger $r$ the percolation transition
is first order. The change from second to first order transition happens at $r_{\text{max}} \approx 8$. The critical $p$ of mutual percolation linearly increases with $r$ for $r < r_{\text{max}}$ and then gradually decreases to $p_{\mu}^{(c)} = 0.68$ for $r \to \infty$. Our analytical considerations are in good agreement with simulations of mutual percolation on interdependent lattices system. Our study suggests that the interdependent infrastructures embedded in space become most vulnerable when the interdependent distance reaches intermediate range, which provides general guidelines in designing complex infrastructures.
Chapter 2

Scaling and Memory of Trading Volume in Stock Markets

2.1 Background

The dynamics of financial markets are of great interest for economic and econophysics researchers. With the development of the society, the fluctuations of the financial markets attract more and more attention and affect our daily life in many aspects. For instance, in 2008, the financial crisis initiated by subprime mortgage default caused huge financial losses for many investors and a recession in the whole economy. Recently, quantitative easing, an unconventional monetary policy, was used by central banks to stimulate the national economy. The S&P 500 Index climbed to an all-time-high owing to the recovery of the markets. Are we over-optimistic by saying the crisis has walked away now? As an investor, one would have to properly control the risk of a portfolio and prepare for possible financial shocks all the time. Banks are required by the Federal Reserve to maintain sufficient amount of capital in order to survive in a crisis. The question about the mechanisms of financial fluctuations becomes of great importance nowadays.

The dynamics of stock price and trading volume have been studied [1–9] in the
field of technical analysis for years as a prerequisite to designing good investment strategies. Econophysics research has found that the distribution of stock price returns exhibits power law tails and that the price volatility time series has long-term power law correlations [10–21]. A power law distribution characterizes the frequency of occurrences of extreme events. Extreme events do not only occur in financial markets, but also appear in very many other fields. The scaling and memory properties of financial records are reported to be similar to those found in climate and earthquake data [25–30]. If one wants to make preparation for earthquake, it might be very important to know when the next shock is likely to occur. In this scenario, a practical approach is to study the time intervals between pairs of successive shocks which are larger than certain threshold levels based on historical records. This way one can gather information on the temporal structure of the fluctuations.

Analogous to earthquakes analysis, Yamasaki et al. [22] and Wang et al. [23, 24] studied the behavior of price return intervals $\tau$ between volatilities occurring above a given threshold $q$. They found that, for both daily and intraday financial records, (i) the distribution of the scaled price interval $\tau/\langle\tau\rangle$ can be approximated by a stretched exponential function; (ii) the sequence of the price return intervals has a long term memory related to the original volatility sequence.

A feature of the recent history of the stock market is large price movements associated with high trading volume. In the Black Monday stock market crash of 1987, the Dow Jones Industrials Average (DJIA) plummeted 508 points, losing 22.6 percent of its value in one day, which led to the pathological situation in which the bid price for a stock actually exceeded the ask price. In this financial crash approximately $6 \times 10^8$ shares traded, a one-day trading volume three times that of the entire previous week. Understanding the precise relationship between price and volume fluctuations has thus been a topic of considerable interest in recent research [31–33]. Trading volume data in itself contains much information about market dynamics, e.g., the distribution of the daily traded volume displays power law tails with an exponent within the Lévy
stable domain [34,35]. Recently, Ren and Zhou [36] studied the intraday database of two composite indices and 20 individual indices in the Chinese stock markets. They found that the intraday volume recurrence intervals show a power-law scaling, short-term correlations and long-term correlations in each stock index.

In this study we analyze the U.S. stock market data over a sufficiently broad range of time scale to identify how several financial factors significantly affect scaling properties. To better understand these scaling features and correlations, we study the daily trading volume volatility time intervals $\tau$ between two successive volume volatilities above a certain threshold $q$, and find a range of power law distributions. We also perform a detailed analysis of the relations between volume volatility time intervals and four financial stock factors: (i) stock lifetimes, (ii) market capitalization, (iii) average trading volume, and (iv) average trading value. We find systematically different power law exponents for $P_q(\tau)$ when binning stocks according to these four financial factors. We find that in the U.S. stock market the conditional probability distribution, $P_q(\tau|\tau_0)$ for $\tau$ following a certain interval $\tau_0$, demonstrates that volume intervals are short-term correlated. We also find that the daily volume volatility shows a stronger long-term correlation for sequences of longer lifetimes but no clear changes in long-term correlations for different stock size factors such as capitalization, volume, and trading value.

2.2 Data

With the rapid development of technology in the past decade, a huge amount of financial data is available for researchers to explicitly study the behaviors of the markets. We obtain our data from the Center for Research in Security Prices (CRSP) U.S. stock database. CRSP records all the daily prices of all listed New York Stock Exchange (NYSE), American Stock Exchange (Amex), and National Association of Securities Dealers Automated Quotations (NASDAQ) common stocks, along with
basic market indices starting from January 2, 1926. In order to obtain a sufficiently long time series and reliable statistics, we analyze the daily trading volume volatility of 17,197 stocks listed in the U.S. stock markets for at least 350 days. The period we study extends from January 1st, 1989 to December 31st, 2008, a total of 5042 trading days.

2.3 Notation

In this section, we would like to describe the basic notations. For a time series $x_i$, where $i = 1, ..., N$, where $N$ is the size of the data set.

- Mean value

$$\langle x \rangle \equiv \frac{1}{N} \sum_{i=1}^{N} x_i$$

for variable $x$. $\langle \cdot \rangle$ is the average of a time series.

- Standard deviation

$$\sigma(x) \equiv \sqrt{\langle x^2 \rangle - \langle x \rangle^2}$$

- Probability density function (PDF) is denoted as $P(x)$ for variable $x$.

- Conditional distribution function is denoted as $P(x|x_0)$, which measures the distribution on the condition that the previous value belongs to subset $x_0$.

2.4 Definition of Volume Volatility and Volume Return

Return is a basic quantity to measure financial fluctuation. For a stock trading volume time series, in a manner similar to stock price analysis [20, 21, 23], we define a basic measure: volume return $R$. The volume return $R$ is defined as the logarithmic change
in the successive daily trading volume for each stock,

\[ R(t) \equiv \ln \left( \frac{V(t)}{V(t-1)} \right), \tag{2.3} \]

where \( V(t) \) is the daily trading volume at time \( t \). \( R(t) \) is approximately the fractional change for small percentage change.

Another basic quantity for financial fluctuations is volatility. We define volume volatility \( v \) to be the absolute value of the volume return, i.e.,

\[ v(t) \equiv |R(t)|. \tag{2.4} \]

Note that in much of the literature, the volatility is defined as the standard deviation of returns. However, Eq. (2.4) has one volume volatility data point for every point of volume return. For example, a weekly volatility time series might be constructed from the the standard deviations of daily returns during each week. In our case, the volume volatility time series is based directly on single day-to-day volume changes.

In order to obtain significant statistics from sufficient data points, in this thesis, we choose Eq. (2.4) as the volatility definition.

In order to compare different stocks, we determine the volume volatility \( \nu(t) \) by dividing the absolute returns \( |R(t)| \) by their standard deviations over the entire time series for each stock,

\[ \nu(t) \equiv \frac{|R(t)|}{(\langle |R(t)|^2 \rangle - \langle |R(t)| \rangle^2)^{1/2}}. \tag{2.5} \]

### 2.5 Definition of Volume Return Interval

In this thesis, we use a new approach to analyze the volume volatility time series. In financial markets, compared to small volatilities, large volatilities are usually more meaningful since they correspond to large volume transactions of the equity, suggesting a big risk for investors. To analyze them, we choose a volume volatility value as threshold \( q \) and collect volatilities that are above \( q \). The collected volatilities represent "events" of the large volume movements. The volume volatility time intervals
are defined as time intervals $\tau$ between two successive volume volatilities above a certain threshold $q$. In this definition, volatility time intervals are strongly related to the structure of the volatility time series. Therefore, we can understand large volume movements much further by studying time intervals of their occurrences. Note: volume volatility (in Eq. (2.5)) is in units of standard deviation. Thus the threshold $q$ is also in units of the volatility standard deviation. To illustrate this process, Fig. 2.1 shows the collections of volume volatility time intervals $\tau_q$ with threshold $q = 1$ and $q = 2$ for 50 daily records of General Electric stock.

2.6 Distribution of Volume Volatility Time Intervals

We study the daily trading volume volatility time intervals of CRSP data sets described above. For a volume volatility time series, we collect the time intervals $\tau$ between consecutive volatilities $\nu(t)$ above a chosen threshold $q$ and construct a new time series of volume return intervals $\{\tau(q)\}$. Fig. 2.2(a) shows the dependence of $P_q(\tau)$ on $q$, where $P_q(\tau)$ is the PDF of the volume volatility return interval time series $\{\tau(q)\}$. Obviously, $P_q(\tau)$ decays more slowly for large $q$ than for small $q$. For large $q$, $P_q(\tau)$ has a higher probability of having large interval values because extreme events are rare in a high threshold series.

We next determine whether there is any scaling in the distribution by plotting the PDFs of the volatility time intervals $P_q(\tau)$, scaled with the mean volume return interval $\langle \tau(q) \rangle$, for different thresholds.

If we have the following scale transformation,

- return interval $\tau \Rightarrow$ scaled return interval $\tau / \langle \tau \rangle$
- accordingly, PDF $P_q(\tau) \Rightarrow$ scaled PDF $P_q(\tau) \langle \tau \rangle$

we can see that all five threshold values $q$ curves (full symbols) collapse onto a single
Figure 2.1: Illustration of volume volatility time intervals. The volatility is in units of standard deviation. The solid circles are daily record of volume volatility values of the General Electric stock starting from January 1st, 1989. Volume volatility time intervals $\tau_{q=1}$ and $\tau_{q=2}$ for two typical thresholds $q$ are displayed. Note the tendency for two or more successive days to have high volumes. This qualitatively suggests a short-range correlation in the volume volatility data (see Section 2.8).
Figure 2.2: Probability distributions of volume volatility time intervals and price volatility return intervals for 17197 stocks. Full symbols with different shapes represent different thresholds $q$ varying from 2.0 to 4.0. (a) Distribution of volume volatility return intervals, $P_q(\tau)$ versus $\tau$. (b) Scaled distribution of volume return intervals (full symbols) $P_q(\tau)/\langle \tau \rangle$ versus $\tau/\langle \tau \rangle$, and distribution of volume return intervals for shuffled volatility records (open symbols). The four curves with full symbols collapse onto one single curve, indicating a universal scaling function. The tail of the scaling function is approximately power-law distribution, $f(x) \sim x^{-\gamma}$, with $\gamma \approx 3.2$, while the curve fitting the shuffled records is exponential function, $f(x) = e^{-ax}$, from Poisson distribution. A Poisson distribution indicates no correlation in shuffled volatility data, but the original data set suggests strong correlation in the volatilities. For comparison, (c) and (d) show the distribution and scaled distribution of price volatility return intervals respectively. Note the narrow range of power-law compared to (a).
curve, suggesting the existence of a scaling relation,

\[ P_q(\tau) = \frac{1}{\langle \tau \rangle} f \left( \frac{\tau}{\langle \tau \rangle} \right). \]  \hspace{1cm} (2.6)

As the threshold \( q \) increases, the curve (extreme rare events) tends to be truncated due to the limited size of the data sets. The function \( f \) in Eq. 2.6 represents the scaling feature of the return interval distribution. The scaling functional form does not explicitly depend on \( q \), but only through the mean interval \( \langle \tau \rangle \). To simplify the notation, we neglect the subscript \( q \) for \( P \) in the following text. If \( P(\tau) \) is known for one value of \( q \), we can predict for other values of \( q \) by Eq. 2.6 when \( q \) is very large and the data points are rare due to the lack of statistics.

The tails of the scaling function can be approximated by a power law function as shown by the dashed line in Fig. 2.2(b),

\[ f \left( \frac{\tau}{\langle \tau \rangle} \right) \sim \left( \frac{\tau}{\langle \tau \rangle} \right)^{-\gamma}, \]  \hspace{1cm} (2.7)

where the tail exponent is \( \gamma \). The exponent of the scaled PDFs for \( q = 2 \) is \( \gamma \cong 3.2 \) by the least square method, which is the same as the unscaled PDF exponent \( \gamma \cong 3.2 \) as shown in Fig. 2.2(a). The power-law exponents for intraday volume recurrence intervals of several Chinese stock indices are from \( \gamma = 1.71 \) to \( \gamma = 3.27 \) [36]. Our exponents \( \gamma \) are larger than those in the Chinese stock markets. This might be due to differing definitions of volume volatility. In Ref. [36], the volume volatility is defined as intra-day volume divided by the average volume at one specific minute of the trading day averaged over all trading days. Here we define the volume volatility to be the absolute value of logarithmic change in the successive daily volumes [Eqs. 2.3 and 2.5]. For comparison, and using the same approach, Fig. 2.2(c) and Fig. 2.2(d) show the analogous results for the price volatilities of our time series (see also the studies in Refs. [22, 23, 37]). Note that it is not easy to distinguish between a stretched exponential and a power-law when studying price volatilities [22], i.e., the power-law range is small and a stretched exponential could also provide a good fit. In contrast,
the PDFs of the volume volatility time intervals display a wide range of power law tails, which differs from the stretched exponential tail apparent in the price return intervals [23]. Our results for volume volatility may suggest that $P_q(\tau)$ for price volatility is also a power law, but this could not be verified because the range of the observed power law regime [see Figs. 2.2(c) and 2.2(d)] is more limited than the broad range of scales seen in the volume volatility [Figs. 2.2(a) and 2.2(b)]. The difference between the power law and stretched exponential behavior of $P_q(\tau)$ may be related to the existence or non-existence respectively of non-linearity represented in the multi-fractality of the time series. When non-linear correlations appear in a time record, Bugachev et al. [38] showed that $P_q(\tau)$ is a power-law. On the other hand, when non-linear correlations do not exist and only linear correlation exists, Bunde et al. [26] found stretched exponential behavior.

A comparison with the shuffled records allows us to see how the empirical records differ from randomized records. We shuffle the volume volatility time series to make a new uncorrelated sequence of volatility, and then collect the time intervals above a given threshold $q$ to obtain synthetic random control records. The curve that fits the shuffled records [the open symbols in Fig. 2.2(b)] is an exponential function, $f(x) = e^{-ax}$, and forms a Poisson distribution. A Poisson distribution indicates no correlation in shuffled volatility data, but the empirical records suggest strong correlations in the volatility.

2.7 Financial Factor Influences

We study the relations between the scaled PDFs $P_q(\tau)/\langle \tau \rangle$ of our volume volatility series and four financial factors respectively: (a) stock lifetimes, (b) market capitalization, (c) mean volume, and (d) mean trading value for threshold $q = 2.0$. For higher $q$ values, we do not have sufficient data for conclusive results [38].

In Fig. 2.3, we plot the scaled PDFs for these four factors. The volume return
Figure 2.3: Relations between distribution function $P_q^2(\tau)\langle \tau \rangle$ of volume volatility return intervals and four financial factors: (a) lifetimes, (b) market capitalization, (c) average daily trading volume, (d) average daily trading value, for the threshold $q = 2.0$. The distribution functions decay with various exponents $\gamma$ and show similar systematic tendency for four financial factors.
intervals characterize the distribution of large volume movements. A high probability of having a large volume return interval $\tau$ suggests a correlation in volume volatility. Because small volatilities are tend to be followed by small volatilities and the time interval between the two large volatilities becomes relatively longer than those of random records. In order to characterize how these four factors affect the distribution of volume return intervals, we divide all stocks into four subsets for each factor. In Fig. 2.3(a), the probability that $\tau$ will be large is greater in the subset with 15~20 year old stocks (triangles) than in the subsets of younger stock. This indicates that small volatilities (below the threshold) tend to follow small volatilities and that the time intervals between large volatilities in the subset of 15~20 year-old stocks are larger than the time intervals in the subset of 5 years old stocks (dots). The decaying parameters represented by the power-law exponents are quite different: $\gamma \approx 4.2$ for the shortest lifetimes subset and $\gamma \approx 2.8$ for the longest lifetimes subset. This significant difference might be caused by differences in autocorrelation in these series (see Section 2.10).

In Figs. 2.3(b), 2.3(c), and 2.3(d), we use the same approach for stock subsets with different capitalizations, mean volumes, and mean trading values. Trading value is defined as stock price multiplied by transaction volume. For each stock, we designate the lifetimes average of capitalization, volume, and trading value as performance indices. For example, the power-law exponents of the PDFs, $P_q(\tau)\langle \tau \rangle$, decrease as the capitalization becomes larger [see Fig. 2.3(b)]. Because the negative sign in the power law means that the tail has a longer time scale for larger companies.

To clarify the picture, we divide all stocks into different subsets and study the behavior of the power-law exponent $\gamma$ with regard to these four factors. In Fig. 2.4(a), stocks are sorted into 10 subsets, from 508 days (2 years) to 5080 days (10 years). We fit the power-law tails of the volume return intervals for each subset and plot the exponent $\gamma$ versus the lifetimes of the stocks. In Fig. 2.4(a), we can observe a systematic trend with stock lifetimes. It is seen that long lifetime stock subsets have
Figure 2.4: The power-law tail exponent $\gamma$ for different subsets of stocks. (a) Stocks are sorted into 10 subsets of different lifetimes. Exponent $\gamma$ are obtained by fitting the PDF of volume volatility return intervals for each subset; (b) stocks are sorted into 8 subsets for different capitalization; (c) stocks are sorted into 11 subsets for different mean volume; (d) stocks are sorted into 9 subsets for different trading value. Long time series stock subsets have smaller exponent $\gamma$, indicating a broader power-law tail in the distribution of normalized volume volatility return intervals. One possible explanation is that long time series stocks have larger volume volatility autocorrelations. The trends are not as obvious when we group stocks by these three factors (Note the large constant range because of logarithmic scales for the three factors) because $\gamma$ decreases with the increasing of these factors but seems to become constant for large values of capitalizations, mean volumes and mean trading values.
smaller exponent $\gamma$, indicating a broader power-law tail in the distribution of normalized volume volatility return intervals. One possible explanation is that long lifetime stocks have more stable expectations and thus larger volume volatility autocorrelations. Similarly, we sort the stocks by capitalization, mean volume, and mean trading value, as shown in Figs. 2.4(b), 2.4(c), and 2.4(d). The trends are not as obvious when we group stocks by these three factors because $\gamma$ decreases with the increasing of these factors but seems to become constant for large values of capitalizations, mean volumes and mean trading values.

Since all factors similarly affect the scaling of the PDF, $P_q(\tau)\langle \tau \rangle$, we now determine how much these factors are correlated. To study the relations between different stock subsets, we plot, as shown in Fig. 2.5, the correlation between trading value and capitalization, mean volume and capitalization, mean trading value and mean volume respectively for all the stocks. The correlation coefficients between trading value and capitalization, mean volume and capitalization, and trading value and volume are 0.62, and 0.55, and 0.78, respectively. The correlation coefficients are high because these capitalization, volume, and trading value factors are all affected by firm size. Our analysis do not, however, show a significant relationship between stock lifetime and its trading value, capitalization, and mean volume, and the correlation coefficients are all smaller than 0.20.

2.8 Short-term Memory Effects

We characterize a sequence of volume return intervals in terms of the autocorrelations in the time series. If the volume return intervals series are uncorrelated and independent of each other, their sequences can be determined only by the probability distribution. On the other hand, if the series is auto-correlated, the preceding value will have a memory effect on the values following in the sequence of volume volatility return intervals.
Figure 2.5: Scatter plots for the relations in stocks between trading value and capitalization, mean volume and capitalization, trading value and mean volume for all stocks. For example, a point on panel (a) represents a stock, which has $10^8$ capitalization and $10^6$ average trading value. The correlation coefficients between trading value and capitalization, mean volume and capitalization, trading value and volume are 0.62, and 0.55, and 0.78 respectively. However, our analysis do not show a significant relationship between stock lifetime and its trading value, capitalization, and mean volume, and the correlation coefficients are all smaller than 0.20.
Figure 2.6: Conditional PDF $P_q(\tau|\tau_0)$ of volume volatility time intervals $\tau$ for different thresholds $q = 2.0, 2.5, 3.0$, as a function of $\tau/\langle \tau \rangle$ for different $\tau_0/\langle \tau \rangle$ subsets. A small $\tau_0$ subset $Q_2$ (full symbols) and a large $\tau_0$ subset $Q_6$ (open symbols) are displayed in (a). For example, subset $Q_6$ contains events of finding $\tau$ after large interval $3.2 < \tau_0/\langle \tau \rangle < 6.4$. In contrast to subset $Q_6$, subset $Q_2$ has larger probability to be followed by small $\tau/\langle \tau \rangle$ and smaller probability to be followed by large $\tau/\langle \tau \rangle$, which indicates short term correlation in the records: small intervals are followed by small intervals and large intervals are followed by large intervals. There is no memory effect in shuffle records as seen in (b) that the PDFs of all the subsets collapse onto one curve.
In order to investigate whether short-term memory is present, we study the conditional PDF, $P_q(\tau|\tau_0)$, which is the probability of finding a volume return interval $\tau$ immediately after an interval of size $\tau_0$. In records without memory, $P_q(\tau|\tau_0)$ should be identical to $P_q(\tau)$ and independent of $\tau_0$. Otherwise, $P_q(\tau|\tau_0)$ should depend on $\tau_0$. Because the statistics for $\tau_0$ of a single stock are of poor quality, we study $P_q(\tau|\tau_0)$ for a range of $\tau_0/\langle \tau \rangle$. The entire data set is partitioned into eight equal-sized subsets, $Q_1, Q_2, ... Q_8$, with intervals of increasing size $\tau_0/\tau$.

Figure 2.6 shows the PDFs $P_q(\tau|\tau_0)$ for $Q_2$, i.e., small interval size $0.2 < \tau_0/\langle \tau \rangle < 0.4$ and $Q_6$ large interval size $3.2 < \tau_0/\langle \tau \rangle < 6.4$ for different $q$. The probability of finding large $\tau/\langle \tau \rangle$ is larger in $Q_6$ (open symbols) than in $Q_2$ (full symbols), while the probability of finding small $\tau/\langle \tau \rangle$ is larger in $Q_2$ than that in $Q_6$. Thus large $\tau_0$ tends to be followed by large $\tau$, and vice versa, which indicates short-term memory in the volume return intervals sequence. Moreover, note that $P_q(\tau|\tau_0)$ in the same subset for different thresholds $q$ fall onto a single curve, which indicates the existence of a unique scaling for the conditional PDFs as well. Similar results were found for the volume volatility of the Chinese markets [36] and for price volatilities [22, 23].

### 2.9 Detrended Fluctuation Analysis Method

In previous studies, the price return volatility series was shown to have long-term correlations. Using a similar approach, we test whether the volume volatility sequence also possesses long-term correlations. To answer this question, we employ the detrended fluctuation analysis (DFA) method [39–41] to further reveal memory effects in the volume volatility series.

The DFA method has proven useful in revealing the extent of long-range correlations in time series. The DFA method is developed to determine the statistical self-similarity of a signal. The idea is based on the observation that a correlated time series can be mapped to a self-similar process by integration. Therefore, the
self-similarity of a process can represent the correlation properties. The obtained exponent in DFA is similar to the Hurst exponent [42]. The advantage of the DFA method over conventional methods such as Hurst exponent is that DFA can apply to a non-stationary time series and avoid the detection of spurious correlation. The method was first introduced by Peng et al. in 1994 and has been cited for more than 2000 times as of 2013.

Briefly, we implement the DFA method as in following steps:

(i) first, integrate the time series \( x(t) \) with \( N \) data points into a new series \( y(t) \),

\[
y(t) = \sum_{i=1}^{t} x(i),
\]

(ii) second, the integrated time series \( y(t) \) is divided into boxes of equal length \( n \). The data in each box is fitted using a least square line.

(iii) third, detrend the integrated time series by subtracting the local trend in each box. The \( y \) coordinate of the straight line segments is denoted by \( y_n(k) \).

(iv) fourth, compute the root-mean-square fluctuation \( F(n) \) of this integrated and detrended time series,

\[
F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^{N} [y(k) - y_n(k)]^2},
\]

(iv) fifth, determine the correlation exponent \( \alpha \) by fitting the fluctuation function \( F(n) \) with the following function. The exponent \( \alpha \) represents the correlations in the \( x(t) \) time series.

\[
F(n) \sim n^{\alpha},
\]

This fluctuation measurement process is repeated over the whole time series at a range of different box sizes \( n \). Typically, \( F(n) \) will increase with box size. If we plot \( n \) against \( F(n) \) on a log – log graph, the fluctuations can be characterized by a scaling exponent, which is the slope of the line.

This exponent \( \alpha \) is a generalized case of the Hurst exponent. Because the expected displacement grows like \( \sqrt{n} \) in an uncorrelated random walk of length \( n \). The
exponent is $\frac{1}{2}$ for white noise. When the exponent is between 0 and 1, the result is called Fractional Brownian motion.

The correlation type of the exponent $\alpha$ is summarized as the following:

- $\alpha < \frac{1}{2}$: anti-correlated
- $\alpha = \frac{1}{2}$: uncorrelated, white noise
- $\alpha > \frac{1}{2}$: correlated
- $\alpha \geq 1$: non-stationary, unbounded

### 2.10 Long-term Memory Effects

Using the DFA method, we analyze the price volatility and volume volatility time series by plotting in subsets the relation between correlation exponent $\alpha$ and the four financial factors, including stock lifetimes, market capitalization, mean trading volume, and mean trading value. All the price volatility and volume volatility correlation exponents are significantly larger than 0.5, suggesting the presence of long-term memory in both price volatility sequences and volume volatility sequences. In all of the plots, the price volatility series shows a stronger long-term correlation than the volume volatility series.

Moreover, as shown in Fig. 2.7(a), $\alpha$ on average increases for the stocks with a lifetime ranging from 350 days to 3800 days (about 15 years), and then shows a slight decrease, suggesting that longer-lasting stocks tend to have a more persistent price and volume movement. We note that Ren and Zhou [36] also found long range correlations in the volume records consistent with our findings. The increasing exponent $\alpha$ indicates that the volume volatility of long lifetime stocks is more auto-correlated than that of younger stocks. This is consistent with the indication in Fig. 2.3(a) that the volume volatility of long lifetime stocks are more auto-correlated. Figures 2.7(b),
Figure 2.7: Correlation exponent $\alpha$ obtained from detrended fluctuation analysis (DFA) of volume volatility (square) and price volatility (triangle). The plot shows the relation between $\alpha$ and four factors: (a) lifetimes, (b) market capitalization, (c) average daily trading volume, (d) average daily trading value, for the threshold $q = 2.0$. In (a), the increasing $\alpha$ versus lifetime suggests that longer time series stocks tend to have a more persistent price and volume movement. (b), (c), and (d) show that there is no systematic tendency relation between $\alpha$ and market capitalization, trading volume, and trading value.
2.7(c) and 2.7(d) show that there is no systematic tendency relation between $\alpha$ and market capitalization, trading volume, and trading value.

### 2.11 Conclusions

In summary, we have shown the scaling properties and memory effects of volume volatility time intervals in large stock records of the U.S. market. The scaled distribution of volume volatility intervals displays unique power law tails for different thresholds $q$. We also find different power law exponents $\gamma$ of $P_q(\tau)$ for the four essential stock factors: stock lifetimes, market capitalization, average trading volume, and average trading value. These different exponents may be related to long-term correlations in the interval series. Significantly, the daily volume volatility exhibits long-term correlations, similar to that found for price volatility. The conditional probability, $P_q(\tau|\tau_0)$ for $\tau$ following a certain interval $\tau_0$, indicates that volume return intervals are short-term correlated. Using the DFA method, we also find that the daily volume volatility shows a stronger long-term correlation for sequences of longer lifetimes.
Chapter 3

Cascading Failures in Complex Networks

3.1 Background

Network Science is an interdisciplinary academic field bringing together useful tools from Statistical Physics, Mathematics, Computer Science, Epidemiology, Sociology, and Geology to study complex systems such as telecommunication networks, power grid networks, computer networks, biological networks, and social networks in order to predict the collective behaviors of these systems. More specifically, network science aims to develop theoretical frameworks and practical approaches to increase our understanding of natural and artificial networks.

One would ask "Why should physicists study networks". My answer is that, the mission of physics is to understand basic principles of how the universe behaves. Generally speaking, networks are one of the simplifications of the universe.

At any moment in time, we are interacting with an integral of many dynamically changing networks. Our organs work together to provide proper function to our bodies. In a complex food web, we learn the feeding connections to maintain population growth and convolution. Moreover, we create advanced infrastructures of water sup-
ply and power grid systems, street and airline systems, even virtual systems like the Internet. The more systems we discover and create, the more challenges we would encounter in order to understand major connections, trends, and patterns of these systems.

### 3.2 Overview of Complex Network Studies

The field of complex networks has been studied intensively in the past decade [43–55]. This development is prompted by several significant advances. First, the enormous amount of data collection from all fields provides a large database of real networks with intriguing interest. Second, the increasing power of super computers enable us to simulate networks with millions of nodes and to analyze gigabit data in only a few minutes. Last but not the least, in this era of big data, complex networks become a powerful tool to understand the behaviors and to capture the properties of the system as a whole.

#### 3.2.1 Erdős-Rényi Networks

Traditionally, the study of complex networks was in the field of graph theory mainly focusing on regular graphs. Random graph theory was originally developed in the 1960s by Erdős and Rényi [56, 57] to study large-scale networks without specific design pattern. The Erdős-Rényi (ER) model assumes that each pair of nodes is randomly connected with the same probability, leading to a Poisson degree distribution,

\[
P(k) = \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!},
\]

with \( \langle k \rangle = Np \). \( N \) is the total number of nodes in the network and \( p \) is the probability that every pair of nodes is randomly connected.

The ER model was the primary model of networks for decades. Although graph theory is a well-established mathematical tool, it cannot represent many if not most real-life networks.
3.2.2 Scale-Free Networks

In 1999, Barabasi and Albert [58] observed that many real networks do not follow the ER model, but display some organizing principles. For instance, it was found that the degree distribution of the World Wide Web (WWW) indeed did not follow a Poisson distribution but rather, for a large range of \( k \), a power law distribution.

\[
P(k) \sim k^{-\lambda}, \tag{3.2}
\]

Later, many other networks were found to be scale-free (SF). Ever since then, an overwhelming literatures of supporting data and advanced mathematical tools were published to measure to the underlying principles of complex networks.

3.3 Percolation Theory

Percolation theory is widely used in complex network study. A representative question in percolation theory is the following. Assume that a liquid is poured on top of a porous material, such as a sponge or foam rubber. Will that liquid be able to make its way from pore to pore and finally reach the bottom of the porous material? This question can be modeled by a network with pores as nodes and channels between pores as links. Each link between two neighboring nodes may be passable with probability \( p \). Otherwise the link is blocked with probability \( 1 - p \). Therefore, for a given \( p \), what is the probability that an open path exists from the top to the bottom? This question is now called ”bond percolation”. In 1957, Broadbent and Hammersley [59] first introduced discussion of this question in the mathematics literature.

Similarly, in a random graph, the site is ”occupied” with probability \( p \) or ”empty” with probability \( 1 - p \) (the links to the empty site are removed). The question is the same: for a given \( p \), what is the probability that a path exists between top and bottom? This corresponding problem is called ”site percolation”. For infinite systems, there is a critical \( p = p_c \) that below \( p_c \), the largest cluster consists of only a finite
number of nodes, and percolation occurs with probability 0. While \( p \) is above \( p_c \), the largest cluster consists of infinite nodes, and percolation occurs with probability 1.

In this thesis, we only focus on site percolation problem.

### 3.4 Interdependent Complex Networks

With the advance of technology, modern network systems tend to become mutually inter-connected and dependent on each other in order provide proper functionality. Diverse infrastructures such as transportation, fuel, water supply, and power stations are strongly coupled together. Failures in one network can cause dramatic damage to other networks and finally destroy the whole coupled system. There is an increasing interest in studying the robustness of interdependent networks very recently [60–77]. In interdependent networks, nodes from one network depend on nodes from another network and vice versa. Consequently, when a node from one network fails, it causes the corresponding node in the other network to fail, too. When some initial failure of nodes happens, this may trigger an iterative process of cascading failures that may completely destroy both networks.

The previous studies of the robustness of interdependent networks focused on random networks in which space restrictions are not considered. However, most real networks are embedded either in two-dimensional or in three-dimensional space, and the nodes in each network might be interdependent with nodes in other networks. One example is a computer in a computer network is dependent on power from a local power grid network where both networks are spatially embedded. Another example is the way the world-wide network of seaports embedded in the two-dimensional surface of the earth is interdependent with power grid networks embedded on the same surface. A seaport needs electricity from a nearby power station to operate and a power station needs fuel supplied through a nearby seaport to operate. Thus the failure of a power station in a power grid network will cause a failure in a nearby
seaport and vice versa. Spatial constraints, such as the network dimensionality [78], influence the network properties dramatically. Thus the question about the resilience of spatially interdependent networks is of much interest.

### 3.5 Interdependent Lattice Networks

The case of interdependent spatially embedded networks is significantly different from interdependent random networks in two ways: (i) within each network, nodes are connected only to the nodes in their spatial vicinity, while in the randomly connected networks, the concept of spatial vicinity is not defined; (ii) the dependency links establishing the interdependence between the networks are not random but have a typical length $r$. To understand how these spatial constraints affect the resiliency of interdependent networks, we study the mutual percolation of a system composed of two interdependent two-dimensional lattices $A$ and $B$, where a node $A_i$ can connect to its dependent node $B_j$ only within distance $r$ from $A_i$ (see Fig. 3.1). Since a node can be functional only if it is connected to the network, the resilience can be measured, using percolation theory, as the size of the remaining giant component after an attack on the network.

Our model consists of two identical square lattices $A$ and $B$ of linear size $L$ and $N = L^2$ nodes with periodic boundary conditions. In each lattice, each node has two types of links: connectivity links and dependency links. Each node is connected to its four nearest neighbors within the same lattice via connectivity links. Also, a node $A_i$ located at $(x_i, y_i)$ in lattice $A$ is connected with one and only one node $B_j$ located at $(x_j, y_j)$ in lattice $B$ via a dependency link, with the only constraint that $|x_i - x_j| \leq r$ and $|y_i - y_j| \leq r$ (Fig. 3.1). The parameter $r$ is related to the maximum distance a node in one network gets support from a node in another network.

Although real networks embedded in two-dimensional space may have more complex structures than the square lattice, our model can serve as a benchmark for more
Figure 3.1: Two square lattices A and B where in each lattice every node has two types of links: connectivity links and dependency links. Every node is initially connected to its four nearest neighbors within the same lattice via connectivity links. Also, each node $A_i$ in lattice A depends on one and only one node $B_j$ in lattice B via a dependency link (and vice versa), with the only constraint that $|x_i - x_j| \leq r$ and $|y_i - y_j| \leq r$. If node $A_i$ fails, then node $B_j$ fails. If node $B_j$ fails, then node $A_i$ fails. Network A is shifted vertically for clarity.
complex situations. Moreover, it is known that the percolation transition in two di-
mensions has universal scaling behavior which does not depend on the coordination
number and is the same for lattice and off-lattice models, as long as the links have a
finite characteristic length. Hence mutual percolation in two dimensions should not
depend on the particular realization of the model.

3.6 Cascading Failures

The difference between connectivity and dependency links is that for connectivity
links, a node fails only when it does not belong to the giant cluster of its network,
while for dependency links, a node fails once the node on which it depends in the
other network (connected via a dependency link) fails. An initial random attack
destroyed a fraction $1 - p$ of nodes in network A. This causes a certain number of
nodes to disconnect from the giant component of network A so that only a fraction
of nodes $p_1 = P_\infty(p)$ remains functional. Here $P_\infty(p)$ is the order parameter of
conventional percolation in a square lattice [79]. The removal of nodes in network A
causes the removal of the dependent nodes in network B. As a result, only a fraction
$P_\infty(p_1)$ of nodes in network B remains functional. This produces additional damage
in network A and so on. The cascading failure process stops when no further damage
propagates between the lattices. If the length of dependency links is totally random
($r = L$), the formalism developed in Ref. [60] can be applied. At the $i$-th stage of the
cascade the resulting giant component $P_\infty(p_i)$ is the order parameter of conventional
percolation computed for a random fraction of nodes $p_i$ surviving after all the nodes in
network A that depend on the nonfunctional nodes of the other network are removed.
Accordingly we can represent the cascading failure by the recursive equations for the
survived fraction $p_i$, 

$$p_0 = p, \quad (3.3)$$

$$p_1 = \frac{p}{p_0} P_\infty(p_0) = P_\infty(p), \quad (3.4)$$

$$\vdots$$

$$p_i = \frac{p}{p_{i-1}} P_\infty(p_{i-1}). \quad (3.5)$$

The recursive steps of Eq. (3.7), representing the cascading failures in the giant component shown in Fig. 3.2, are in good agreement with simulations. In the limit $i \to \infty$, Eq. (3.7) yields the equation for the mutual giant component at steady state, $\mu \equiv P_\infty(p_\infty)$,

$$x = \sqrt{pP_\infty(x)}, \quad (3.8)$$

where $x \equiv p_\infty$.

Using the form of $P_\infty(x)$ for conventional percolation obtained from numerical simulations, Eq. (3.8) can be solved graphically as shown in Fig. 3.3. Due to the specific shape of the function $P_\infty(p)$ [see Fig. 3.4], $(P_\infty(p) < p$, $\lim_{p \to 1} P_\infty/p = 1$, $\lim_{p \to p_c} P_\infty(p) = 0$, and $p_c = 0.5927$ for square lattice), it does not have solutions for a small $p$ except for the trivial case $x = 0$.

Figure 3.4 shows the numerical solution of Eq. (3.8) which is in good agreement with simulations and compares it with $P_\infty(p)$ of a single network. The critical $p$ for which the nontrivial solution ceases to exist, $p \equiv p^\mu_c$, corresponds to the case when the r.h.s. of Eq. (3.8) becomes tangential at the point of their intersection $x = x_c$ to its l.h.s. (Fig. 3.3). Hence

$$P'_\infty(x_c)x_c = 2P_\infty(x_c), \quad (3.9)$$

from which the critical $p$ for mutual percolation is

$$p^\mu_c = \frac{x_c^2}{P_\infty(x_c)}. \quad (3.10)$$
Figure 3.2: Giant component size $P_{\infty}$ as a function of step $i$ at the first-order transition regime at $p = 0.6825$ for $r = L = 1000$. The simulation results (solid lines) are in good agreement with the theoretical results (dots). The value of $p$ is close to the percolation threshold $p^c = 0.6827$. 
Figure 3.3: A schematic graphical solution of Eq. (3.8) is shown. The curves are $\sqrt{pP_\infty(x)}$ for different $p$ and the solution of Eq. (3.8) is given by the intersection of the solid curves and the straight line $y = x$. The critical $p = p_c^\mu$ corresponds to the case when the solid curve is tangential to the straight line $y = x$. Numerical solutions of Eqs. (3.9) and (3.10) yield $x_c = 0.641$, $P_\infty(x_c) = 0.602$, and $p_c^\mu = 0.683$. 
Numerical solutions of Eqs. (3.9) and (3.10) yield \( p^\mu_c = 0.683 \), \( x_c = 0.641 \), and \( P_\infty(x_c) = 0.602 \), in good agreement with simulations of the mutual percolation on lattices for \( r = L \) as seen in Fig. 3.4. The Fig. 3.3 shows a discontinuity in the order parameter of mutual percolation \( \mu(p) = P_\infty(p) \) at \( p = p^\mu_c \), which drops from \( \mu(p) = 0.602 \) to zero for \( p > p^\mu_c \), characteristic of a first-order transition. This transition is based on the theoretical construction for solving the implicit equation (3.8).

### 3.7 Cascading Failures of Various Lengths of Dependency

Next, we study the mutual percolation for different dependency lengths \( r \). An infinite coupling distance \( r = \infty \) corresponds to the scenario of random dependency links between the lattices discussed above. For \( r = 0 \), every failed node in network A leads to removal of a node in network B in the same location. Thus, the percolation clusters in the two lattices are identical and there is no feedback failure in network A. Therefore, the case of \( r = 0 \) is identical to the case of conventional percolation in non-coupled lattices.

Fig. 3.5 shows, when \( r \) is larger than zero, the largest and second largest cluster sizes as a function of \( p \). When \( r \leq 6 \), the size of the second largest cluster has a peak, implying a second order phase transition. This is similar to the scenario of \( r = 0 \), a conventional percolation on one square lattice. When \( r > 6 \), the size of the second largest cluster is close to zero, suggesting a first order phase transition.

Cluster size distribution is another measurement of percolation. Cluster size distribution counts all the remaining cluster size except the largest cluster. As shown in Fig. 3.6, when \( r \) is small, the complementary cumulative distribution of cluster size per system is linearly related to the cluster size on a \( \log - \log \) plot. The linear relation changes as \( r \) increases. We are interested in the question that how often the random
Figure 3.4: The giant component size $P_\infty$ as a function of remaining fraction of nodes $p$. The solid curve is for conventional percolation on a single square lattice, which describes the limiting case of $r = 0$. The solid curve is obtained by numerical simulations on $N = 4000 \times 4000$ lattice sites with periodic boundary conditions and averaged over 100 realizations. The dash curve represents the theoretical result for two interdependent lattice networks with $r = L$ given by Eq. (3.8). The simulation results (dots) are for two interdependent lattice networks with $N = 1000 \times 1000$ and $r = L$. 
Figure 3.5: Largest and second largest cluster sizes as a function of nodes survived $p$ after the initial attack. The simulations shows there is a dramatic increase in the size of largest cluster and the size of second largest cluster is almost zero for $r = 7$ and $r = 8$. (discussion in next section). When $r \leq 6$, the second largest cluster displays a peak as $p$ changes respectively, suggesting a second order transition.
cluster size is above a particular level. The complementary cumulative distribution function is defined as,

\[ C(s) = P(S > s) = 1 - F(s). \] (3.11)

where \( F(s) \) is the cumulative distribution function of cluster size.

We perform percolation at critical \( p_c \) on 100 systems for each dependency distance \( r \). Each system is consist of two lattices with total \( N = 1000 \times 1000 \) nodes respectively. Fig. 3.6 shows the number of clusters whose size \( S \) are larger than \( s \) per system. When \( r = 8 \), besides the largest cluster, only very small clusters are left at critical point, which are significantly different from the case when \( r \leq 7 \). Note: The largest cluster is excluded in calculating the cluster size distribution.

Figures 3.7(a), 3.7(b) show the structure of the giant component just above \( p_c^r \) for very small \( r \) (few lattice units) and for \( r = L \) respectively. For small \( r \) the structure is similar to the heterogeneous fractal-like giant component of single network [79]. In contrast for \( r \) of the order of \( L \), the giant component is homogeneous and almost compact on the verge of a sudden collapse as a first-order transition.

For intermediate values of \( r \), the collapse occurs in a very different way. Figures 3.7(c), 3.7(d), and 3.7(e) show for intermediate values of \( r \) (discussed below) that the initial cascade of failures is localized to a region of size \( r \). Because of local density fluctuations, the effective fraction of nodes \( p \) in one region can be smaller than the overall average, and therefore small clusters at this region become isolated from the giant component and fail even when the entire lattice is still connected. As soon as a region of size \( r \) fails, the system becomes unstable: the interface of this bubble starts to expand and soon engulfs the entire system [Fig. 3.7(c)–(e)]. This local effect of a propagating interface owing to finite dependency links increases the system vulnerability compared to the case of random dependency links. Thus we expect, \( p_c^r(r) > p_c^r(\infty) \) found for random dependency links. The process of formation of the critical bubble is similar to nucleation near the gas-liquid spinodal [81]. Thus, it is important to understand the propagation of a flat interface, which would correspond
Figure 3.6: Complementary cumulative distribution of cluster size per system. For each dependency distance $r$, we perform percolation at critical $p_c$ on 100 systems. Each system is consist of two lattices of size $N = 1000 \times 1000$ respectively. The $x$-axis is the cluster size $s$ of the cluster and the $y$-axis shows the number of clusters whose size $S$ are larger than $s$ per system. The plots shows that when $r = 8$, besides the largest cluster, only very small clusters are left at critical point, which are significantly different from the case when $r \leq 7$. Note: The largest cluster is excluded in calculating the cluster size distribution.
Figure 3.7: Three different typical behaviors of interdependent lattices near criticality. Pictures of stable mutual giant component at criticality of two interdependent lattices ($N = 1000 \times 1000$) after cascading failures initiated by a random removal of $1-p$ of the nodes for (a) $r = 4$ and $p = 0.680$ and for (b) $r = 1000$ and $p = 0.683$. The dynamics of a growing bubble (explained in the text) for $r = 20$ is demonstrated by three snapshots, (c), (d) and (e), of the non-stable giant component of the interdependent lattices ($N = 500 \times 500$) during the cascading process initiated with $p = 0.700$. 
3.8 Propagation of the Flat Interface

In order to systematically study the conditions for propagation of a flat interface, we study the two interdependent networks with an empty gap (size of $r \times L$) on one edge in lattice A (see Fig. 3.9). We construct the two networks with the length of interdependent links smaller than $r$. The only difference from our original system is that after random removal of a certain fraction of nodes $1 - p$, we eliminate the nodes in lattice A with coordinate distance $y_i \leq r$ to create an artificial flat interface.

Simulations show that the flat interface freely propagates and that the interdependent lattices system totally collapses if $p < p_f^I(r)$, where $p_f^I(r)$ is approximately a linear function of $r$ with $p_f^I(0) = p_c = 0.5927$, $p_f^I(r_f) = 1$, and $r_f \cong 38$. For $r > r_f$, as shown in Fig. 3.8 (a) and Fig. 3.8 (b), the interface freely propagates through the system even when the lattice is completely intact. This happens because the removed nodes of lattice A above the interface eliminate half of the nodes in lattice B with $y_j \leq r$. Thus the effective concentration of nodes in the lattice B linearly changes from $p$ at distance $r$ from the interface to $p/2$ right at the interface. This system is analogous to percolation in diffusion fronts studied by Sapoval et al. [80]. There is thus a certain distance from the interface $r_c = r(2p_c - p)/p$ that corresponds to the critical threshold of conventional percolation. If $r_c$ is much larger than the typical cluster size in the range between $p_c$ and $p/2$, all the nodes in lattice B in this layer will be disconnected and hence the interface will propagate freely. The interface can stop if $r_c = \xi(p/2)$, i.e., the connectedness correlation length [79] when $p/2$ is less than $p_c$.

We estimate the critical concentration $p_f^I$ from the equation $\xi(p_f^I/2) = r(2p_c - p_f^I)/p_f^I$, which yields $r_f = \xi(1/2)/(2p_c - 1) = 41$ for the case $p = 1$, where $\xi(1/2) = 7.6$ obtained by numerical simulations of conventional percolation on a single lattice. This prediction agrees well with simulations ($r_f \cong 38$). The propagation of the flat inter-
Figure 3.8: Pictures of the largest cluster in the process of cascading failure for \( r = 40 \) and \( p = 1 \) in two lattices starting a gap in lattice A. The nodes in lattice A with coordinate distance \( y_i \leq r \) are eliminated to create a flat interface before random removal. (a) shows the upper flat interface stops and the lower flat interface is still growing. (b) shows more clearly the edge of the nodes in the process of cascading failures. We can observe the emergence of large holes inside the boundary of the interface, which will eliminate nodes in the other network in the next step.
Figure 3.9: Two square lattices A and B with an empty gap (size of \( r \times L \)) on one edge in lattice A. The only difference from the original complete lattice system is that after random removal of a certain fraction of nodes \( 1 - p \), we eliminate the nodes in lattice A with coordinate distance \( y_i \leq r \) to create an artificial gap.
face close to $p_c^f(r)$ is similar to invasion percolation, which is a fractal process with vanishing number of active sites, and the average interface velocity approaches zero at $p_c^f(r)$, a characteristic of a second-order transition. Thus, the system completely collapses when (1) a flat interface exists and (2) $p < p_c^f$. The conditions for flat interface propagation, $p_c^f(r)$ were obtained for the artificial model where the flat interface is initially created. However, when the system is initiated by a spatially random removal, a flat interface may be created by random fluctuations over the lattice.

3.9 Critical $p_c$ and Dependency Length $r$

What can we learn from the flat interface behavior on our original system with only initial random failures? When $r$ is large, the system begins to locally disintegrate and, at $p < p_c^f$, a local cascade of failures is initiated. As soon as a hole of size $r$ is formed, an interface appears in a low $p$ regime, and freely propagates through the system—because $p$ is already below the critical point $p_c^f$ of the interface propagation. As a result, the interface will completely wipe out the remaining giant component (see Fig. 3.7(c)–(e)). Thus for large $r$, the transition is first order, meaning it is all or nothing, a transition similar to spontaneous nucleation. At these conditions, the removal of even a single additional node may cause the disintegration of the entire system (Fig. 3.10).

The dynamics of the system becomes completely different for small $r$. In this case, when $p_c^f$ is small, the characteristic size of the holes $\xi_h$ in the percolation cluster is sufficiently large and there are many holes of size $\xi_h(p_c^f) > r$. Thus, the flat interface is formed before it begins to propagate. Once $p$ approaches $p_c^f$ from above, the interface begins to propagate simultaneously from all large holes in the system. It can spontaneously stop at any stage of the cascade, leaving any number of sites in the mutual giant component (Fig. 3.10). The average number of sites in the giant component will approach zero as $p$ approaches $p_c^f$, subject to strong finite-size effects
Figure 3.10: The fraction of nodes in the giant component as a function of nodes survived after the initial attack. We perform the simulations by gradually removing additional nodes. For $r = 6$ the decrease of giant component occurs in multiple steps, characteristic of a second-order transition. For $r = 8$ and $r = 16$, the giant component may completely collapse by removal of even a single additional node, characteristic of a first-order transition.
as in conventional percolation. So for small \( r \), the transition is a second-order, and 
\[
p^c_c(r) = p^f_c \text{ linearly increases with } p \text{ (Fig. 3.12).}
\]

The Fig. 3.11 shows that at \( r = r_{\text{max}} \), \( \xi_h(p^f_c(r)) = r \approx 8 \), and a flat interface will not spontaneously form. Thus \( p \) must be below \( p^f_c(r) \) in order for the hole of size \( r \) to appear in the system. Once a single hole of such size appears, the flat interface will freely propagate below its critical threshold wiping out the entire coupled network system, as in a first-order transition. Note that \( p^f_c(r_{\text{max}}) \approx 0.738 > p^c_c = 0.6827 \). Thus as \( r \) increases, \( p^c_c(r) \) gradually decreases (Fig. 3.12).

This gradual decrease is caused by two factors. When \( r \) increases in the vicinity of \( r_{\text{max}} \), smaller and smaller \( p \) is needed in order to create holes of size \( r \). When \( p \) becomes close to \( p^c_c \), the system begins to undergo local cascades of failures if the average density in the region of size \( r \) falls below \( p^c_c \). The average over \( r^2 \) nodes of this region can deviate from the mean \( p \) on the order of a standard deviation \( \sqrt{p(1-p)/r} \), thus making the disintegration possible if \( p = p^c_c(r) \approx p^c_c + C/r \), where \( C \) is a constant. Note that \( p^c_c(r) \) has a tendency to increase with the system size. The larger is the system, the more likely a sufficiently large hole or a sufficiently large fluctuation in local density will lead to a local cascade of failures.

### 3.10 Conclusions

We study the cascading failures in a system composed of two interdependent square lattice networks A and B placed on the same Cartesian plane. Our analysis suggests that the change from a second-order to a first-order transition occurs at \( r_{\text{max}} \approx 8 \). Note that Ref. [82] found a second-order transition for \( r = 0 \) on two interdependent lattice networks. Our studies show rich phase transition phenomena when the length of the dependency links \( r \) changes. The critical \( p \) of mutual percolation increases linearly with \( r \) in the range of \( r < r_{\text{max}} \), and is characterized by a second-order transition. For \( r \geq r_{\text{max}} \), the cascading failures suggest a first-order transition and the critical
Figure 3.11: Diameter of the hole size $\xi_h$ as a function of $r$ on conventional percolation on a single lattice network. The connectedness correlation length $\xi_h \approx r_{max} = 8$ at $p = 0.744$ which is in good agreement with the simulation.
Figure 3.12: The critical $p_c$ as a function of interdependent distance $r$. The black circle shows the mutual percolation criticality $p^\mu_c$ for two intact lattices initially, while the red square shows percolation criticality $p^f_c$ for two lattices with a $r$ width empty gap on the edge of lattice A initially. i) For the intact two lattices percolation (black circle), the change from second to first order transition occurs at $r_{max} \approx 8$. The critical $p^\mu_c$ of mutual percolation linearly increases for $r < r_{max}$ following the percolation threshold for flat interface and then gradually decreases to $p^\mu_c = 0.683$ at $r = \infty$, which is in good agreement with the theoretical results. ii) For the two lattices percolation starting with an empty gap (red square), the critical $p^f_c$ is similar to two intact lattices when $r < r_{max}$. However, when $r > r_{max}$, $p^\mu_c < p^f_c$ due to the propagation of the empty gap. When $r = 38$, the empty gap can wipe out the system even if no node are removed after creating the empty gap in lattice A.
$p$ gradually decreases to $p_c^\mu = 0.683$ for $r \to \infty$. Our analytical considerations are in good agreement with simulations. Our study suggests that interdependent infrastructures embedded in Euclidean space become most vulnerable when the distance between interdependent nodes is in the intermediate range, which is much smaller than the size of the system.
Chapter 4

Summary

This dissertation covers my work in the applications of statistical physics to two complex systems, stock trading volume in financial markets and interdependent square lattice networks. A complex system exhibits more properties which are not obvious from the properties of the individual components. Our statistical physics concepts and models help to observe the emergent behaviors of complex systems and unveil general characteristics of these phenomena.

In the first project, we study the daily trading volume volatility of more than ten thousand stocks in the U.S. stock markets during the period 1989–2008 and analyze the time intervals $\tau$ between volume volatilities above a given threshold $q$. For different thresholds $q$, the probability density function $P_q(\tau)$ scales with mean interval $\langle \tau \rangle$ as $P_q(\tau) = \langle \tau \rangle^{-1} f(\tau/\langle \tau \rangle)$ and the tails of the scaling function can be well approximated by a power-law $f(x) \sim x^{-\gamma}$. We also study the relations between the form of the distribution function $P_q(\tau)$ and several financial factors: stock lifetimes, market capitalization, volume, and trading value. We find a systematic tendency of $P_q(\tau)$ associated with these factors, suggesting a multi-scaling feature in the volume return intervals. We analyze the conditional probability $P_q(\tau|\tau_0)$ for $\tau$ following a certain interval $\tau_0$, and find that $P_q(\tau|\tau_0)$ depends on $\tau_0$ such that immediately following a short/long return interval a second short/long return interval tends to occur. We also
find indications that there is a long-term correlation in the daily volume volatility. We compare our results to those found earlier for price volatility.

In the second project, we study the cascading failures in a system composed of two interdependent square lattice networks A and B placed on the same Cartesian plane, where each node in network A depends on a node in network B randomly chosen within a certain distance $r$ from the corresponding node in network A and vice versa. Our results suggest that percolation for small $r$ below $r_{\text{max}} \approx 8$ (lattice units) is a second-order transition, and for larger $r$ is a first-order transition. For $r < r_{\text{max}}$, the critical threshold increases linearly with $r$ from 0.593 at $r = 0$ and reaches a maximum, 0.738 for $r = r_{\text{max}}$ and then gradually decreases to 0.683 for $r = \infty$. Our analytical considerations are in good agreement with simulations. Our study suggests that interdependent infrastructures embedded in Euclidean space become most vulnerable when the distance between interdependent nodes is in the intermediate range, which is much smaller than the size of the system.
Bibliography


Curriculum Vitae

Wei Li

Boston University, Physics Department
Boston, Massachusetts 02215 USA

Telephone: 617-353-2701
E-mail: liw@buphy.bu.edu

EDUCATION

• 2013, Ph.D. Physics, Boston University, Boston, MA, USA
  Advisor: H. Eugene Stanley
  Dissertation: Statistical Physics Approaches to Complex Systems
• 2010, M.A. Physics, Boston University, Boston, MA, USA
• 2007, B.S. Physics, University of Science and Technology of China, Hefei, China

PUBLICATIONS


CONFERENCES


Referee service

• Physica A

SKILLS

• Programming: C++, Matlab, R, Python, SQL, VBA, Stata

• Database: Access, SQL Server, Oracle