

**STOCK PRICE DYNAMICS BEFORE CRASHES:
A COMPLEX NETWORK STUDY ON THE U.S. STOCK MARKET**

JIAJIA REN

Bachelor of Economics, Hunan University, 2013

A Thesis

Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfilment of the
Requirements for the Degree

MASTER OF SCIENCE IN MANAGEMENT

Faculty of Management
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

© Jiajia Ren, 2015

**STOCK PRICE DYNAMICS BEFORE CRASHES:
A COMPLEX NETWORK STUDY ON THE U.S. STOCK MARKET**

JIAJIA REN

Date of Defence: November 9, 2015

Dr. R. Yalamova, Ph.D.
Supervisor
Thesis Examination Committee

Dr. R. Benkoczi, Ph.D.
Reader
Thesis Examination Committee

Dr. S. Hossain, Ph.D.
Reader
Thesis Examination Committee

Dr. A. Swishchuk, D.Sc.
External Examiner
Thesis Examination Committee
University of Calgary
Calgary, Alberta

Dr. K. Williams-Whitt, Ph.D.
Chairperson, Thesis Examination Committee

Abstract

Historically, stock market crashes have caused trillions of dollars in losses and have dramatically destroyed investors' confidence in the stock market. Independent empirical studies have converged to prove the synchronization phenomenon as the trigger of stock market crashes (Tse, Liu, & Lau, 2010). As well, the Phase Transition Model explains the building-up mechanism and the critical point existing in stock market crash (Yalamova & McKelvey, 2011). In this study, we propose to add more empirical evidence to the current studies and provide an indicator to possibly predict the stock market crashes. We apply the Potential-based Hierarchical Agglomerative (PHA) Method, the Backbone Extraction Method, and the Dot Matrix Plot to extract and display the changing clusters' structure dynamics from the market equilibrium state to a bubble building-up state by applying the Standard & Poor 500 (S&P 500) index constituents' daily price correlation matrix.

Acknowledgements

This thesis would not have been possible without the support and assistance that I received from all the people who have helped me towards completing it.

Firstly, I would like to express my deepest gratitude to my supervisor, Dr. Rossitsa Yalamova, for her constant encouragement and guidance throughout my graduate study and research. She has always been there to provide me important comments and suggestions for my thesis. Besides that, she has offered me invaluable guidance to not only my thesis, but also my future career development.

Secondly, I would like to sincerely thank Dr. Robert Benkoczi, Dr. Shahadat Hossain, and Dr. Anatoliy Swishchuk for kindly serving as Readers and External Examiner on my Examination Committee. I also genuinely thank Dr. Kelly Williams-Whitt and Ms. Tammy Rogness for all their support and assistance during my studies in this program. In addition, I thank Ms. Rebekah Eckert from the Academic Writing Center of University of Lethbridge for her constant assistance on editing my thesis.

Finally, I wish to thank my beloved parents for their considerations and confidence in my study and life. I also thank the professors, my friends, and my fellow classmates for all their help in this program.

Table of Contents

1. Introduction.....	1
2. Literature Review	7
2.1 Stock Market Crash	7
2.2 Complex Network.....	9
2.2.1 Stock Market Complex Network.....	12
2.2.2 Scaling and Synchronization in the Stock Market Complex Network.....	13
2.3 Phase Transition Theory	18
3. Hypotheses Development	22
4. Data and Methodology.....	25
4.1 Data.....	25
4.2 Methodology.....	26
4.2.1 PHA Clustering	28
4.2.2 Backbone Extraction Method.....	32
4.2.3 Dot Matrix Plot	34
5. Results and Discussion.....	35
5.1 PHA Clustering.....	35
5.1.1 Market Equilibrium State Clusters.....	35
5.1.2 Market Disequilibrium Clustering	39
5.2 LANS & Dot Matrix Plot	40
5.3 Discussion.....	41
5.3.1 Market Equilibrium State	41
6. Contribution, Limitation, and Conclusion	47
6.1 Contribution.....	47
6.2 Limitation	49
6.3 Conclusion.....	50
References.....	54
Appendix.....	58
S&P 500 Constituents List from FactSet.....	58

List of Figures

Figure 4.1. The distribution of the data points.....	31
Figure 4.2. The dendrogram derived from the PHA method.....	32
Figure 5.1. PHA result for the time period of Jan 2nd 2002 to Dec 30th 2005.....	35
Figure 5.2. PHA result for the time period of Jan 2nd 2002 to Dec 31st 2002.....	37
Figure 5.3. PHA result for the time period of Jan 2nd 2003 to Dec 31st 2003.....	37
Figure 5.4. PHA result for the time period of Jan 2nd 2004 to Dec 31st 2004.....	38
Figure 5.5. PHA result for the time period of Jan 3rd 2005 to Dec 30th 2005.....	38
Figure 5.6. PHA result of the time Period of Jan 3rd 2006 to Sept 15th 2008.....	39
Figure 5.7. Dot Matrix Plot for the time period of Jan 2nd 2002 to Dec 30th 2005.....	40
Figure 5.8. Dot Matrix Plot for the time period of Jan 3rd 2006 to Sept 15th 2008.....	41

1. Introduction

In September 2008, the exposure of consumer defaults on subprime mortgages triggered the 2008 Global Financial Crisis and started another worldwide stock market crash. This crash is considered by many researchers to have been the worst stock market crash since the Wall Street Crash of 1929. Both of these two major crashes in stock market history have been accompanied by financial crises, such as the Great Depression, which have drawn researchers' attention from the area of finance, economics, psychology, complex networks, or even physics to study the mechanism or causes for the financial crises.

Back to the early 1990s, White (1990) compared one of the major crashes, the Wall Street Crash of 1929, with the 1987 Black Monday Crash. White has shown the high similarity of stock market indices between the 1929 crash and the 1987 crash, which indicates that crashes in the 20th century might have already shown a similar building-up mechanism (White, 1990). At that time, researchers were still concentrating on the policy makers whom they thought should be accused of causing the Great Depression. Cecchetti (1997) summarized that the central bank, deflation, and the gold standard should be considered the key factors that caused the stock market crash and the Great Depression (Cecchetti, 1997). Afterwards, researchers have shifted to quantitative analysis on the stock price movements. Farmer, Gillemot, Lillo, Mike, and Sen (2004) studied the

reasons for the highly volatile time periods on the London Stock Exchange. They found that liquidity, variations of less frequently traded stocks could cause the large fluctuations in stock market (Farmer, Gillemot, Lillo, Mike, & Sen, 2004). Additionally, Baker and Wurgler (2007) also found that sentimental investors could cause some younger, lower capitalization, higher volatility, and growth companies to fluctuate much more heavily during the market volatile time periods, such as market crashes (Baker & Wurgler, 2007). Similarly, Zouaoui, Nouyrigat, and Beer (2011) also found that investor sentiment had a strong positive relationship with the occurrence of a stock market crash within a one-year time period. However, how could we identify the occurrence of large-scale investor sentiment so that we could have indicators to predict and prevent a market crash? In order to identify this prevailing sentiment information, we propose an indicator of trading synchronization based on the clustering changes in the stock market complex network.

In complex network theory, Mantegna (1999) has been the first to reveal the hierarchical structure or complex network structure in financial markets by analyzing the correlation matrix of stock prices time series. Mantegna (1999) also confirmed the valuable information contained by time series of stock prices to predict the local structure movement for the stock market (Mantegna, 1999). Afterwards, more and more researchers in finance, economy, and physics areas have started to apply complex network theory into financial markets. Vandewalle, Brisbois, and Tordoir (2001) have

found the topological structure in stock markets by analyzing a cross correlation matrix of 6358 US stock prices time series. They also confirmed the existence of complex network within stock markets (Vandewalle, Brisbois, & Tordoir, 2001). Thereafter, Krause (2004) built a universal model of an evolving complex network and managed to predict the crashes by constructing a score function based on the eigenvalue of the correlation matrix. He has also concluded that his findings are consistent with the observations or homogeneous behaviors before financial market crashes (Krause, 2004). Later in 2008, shortly before the market crash, Leibon, Pauls, Rockmore, and Savell used the Partition Decoupling Method (PDM) to display the topological structure in the US stock market. They have also found that the network clusters coincide with industry classifications and represent the capital flows moving through different stages (Leibon, Pauls, Rockmore, & Savell, 2008). Then, Tse, Liu, and Lau (2010) developed a correlation matrix study on all the US stock prices and found a vital and strong relationship between the market variation and a small group of stocks (Tse et al., 2010).

Therefore, it is not difficult for us to relate these above studies to some of the key features of complex networks, that is, synchronization and scaling. Synchronization and scaling are the self-organizing characteristics rooted within most complex networks. Scaling is used to describe the self-organizing mechanism due to the individual participants' decisions in a scale-free network (Barabási & Albert, 1999). And

synchronization describes the phenomenon that adding some small new information to a network can significantly cause the network to oscillate into a similar movement (Watts & Strogatz, 1998). Actually, scaling and synchronization are the necessary steps to build up a market crash. Based on these two features, we wonder whether there is a way to identify the scaling and synchronization phenomenon before the market crash. As mentioned above, Leibon, Pauls, Rockmore, and Savell (2008) have developed a mathematical computation method to find the structure or clusters in the normal stock market (Leibon et al., 2008). We believe that changes in the clusters' structure will allow us to identify a precursor of the stock market crash. In addition, Grossman and Stiglitz (1980) showed that the stock market could not remain in an equilibrium state when the information becomes costly (Grossman & Stiglitz, 1980). Once the uninformed traders start to make investment decisions based on the informed traders' behaviors, both informed and uninformed traders would affect each other to move up the stock price regardless of rationality (LeBaron, 2001). Sornette, Johansen, and Bouchaud studied the time series of the S&P 500 index before and after the 1987 stock market crash and found the existence of a log-periodic oscillation price pattern and suggested a phase transition theory to explain the log-periodic pattern (Sornette, Johansen, & Bouchaud, 1996).

Based on Sornette et al.'s (1996) phase transition theory and characteristics of a stock market complex network, Yalamova and McKelvey (2011) built an innovative

Phase Transition Model analogical from physics to explain the homogeneous behaviors, such as herding behavior in the stock market (Yalamova & McKelvey, 2011). According to their model, the imitating behavior or herding behavior occurs at the tipping point, which eventually triggers a crash (Yalamova & McKelvey, 2011). This explanation also corresponds to the synchronization and scaling phenomenon found in complex networks. In addition, Yalamova and McKelvey have also illuminated the existence of a critical point at which the highest level of homogeneous trading behavior happens: that is, the market crash point (Yalamova & McKelvey, 2011). Besides this, they have pointed out that the building-up mechanism of homogeneous trading behavior is driven by the scaling and synchronization characteristics of a complex network with individual stocks as nodes and cash flows forming links (Yalamova & McKelvey, 2011). If most shareholders of one stock put in selling orders for this stock, there is no doubt that the stock price will drop dramatically. Similarly, if there is a large number of buyers on the opposite side, the stock price will go up significantly. Yalamova and McKelvey's (2011) model attempts to build a theoretical framework that accommodates the EMH in the market equilibrium state and bubble building-up stage in the market disequilibrium state.

With this study we would like to contribute to the empirical evidence of the clusters' structure before market crashes provided by a number of econophysics studies in support of Yalamova and McKelvey's (2011) theoretical framework of the bubble

building-up regime as a result of herding, imitating, and rule-based trading. In our study, changes in the stock market network clusters' structure is proposed as an indicator of the bubble building-up state. By analyzing a broadly-used US stock market index, the S&P 500 index, we can build a daily return correlation matrix by collecting the daily returns of all the S&P 500 constituent companies. We use the Potential-based Hierarchical Agglomerative (PHA) clustering method to capture the clusters' structure by building the dendrogram linkage trees (Lu & Wan, 2013). We also apply the LANS method to extract the significant edge backbone from the correlation matrix (Foti, Hughes, & Rockmore, 2011). And then we plot the significant edge backbone to a dot matrix to display the clusters' structures of both the market equilibrium state and the market disequilibrium state, such as the bubble building-up state (Newman & Girvan, 2004).

The rest of the study is organized as follows. Section 2 reviews the related theoretical and empirical literature. Section 3 develops the hypotheses for the clusters' changes in different market states. Section 4 presents the data and methodology utilized in our study. Section 5 summarizes the results and discussions of the clusters' computation for different market states. Lastly, section 6 shows the contributions, limitations of our study, and possible further research areas.

2. Literature Review

2.1 Stock Market Crash

The Stock market crash describes the sudden and dramatic prices drop across the stock market. We focus on the endogenous stock market crashes where there is no external bad news. In the global stock market history, there are two major endogenous crashes: the 1929 Wall Street Crash and the 2008-2009 Crash.

In 1929, the United States stock market experienced the most terrible market crash known as the Great Crash. During the two-day Black Tuesday crash, the U.S. stock market had generated a loss of over \$30 billion. Within the 1929 Great Crash, the Dow Jones Industrial Average had hit the bottom closing at 41.22, which was the lowest level during the 20th century from the very peak level at 381.2 from September 3th 1929 to July 8th 1932 ("Historical Prices, Dow Jones", n.d.).

After the Wall Street Crash of 1929, there was another smaller crash of 1987 that did not lead to a global bearish market. However, White (1990) compared the hypotheses to explain the 1929 stock market crash with the ones for the 1987 market crash. White (1990) pointed out that the emergence of many newly published companies and the subsequent difficulties to evaluate those companies were the beginning stage of the stock market bubble, which finally caused the large-scale panic selling in 1929. For both of these two crashes, it was the similar massive panic selling behaviors that triggered the

dramatic price decrease. Researchers at that time mainly focused on the monetary policy and economy policies. Cecchetti (1997) summarized three factors causing this financial crisis, that is, the influence of the central bank, deflation, and the gold standard (Cecchetti, 1997). Doyne Farmer et al. (2004) applied quantitative analysis to study the reasons for the large fluctuations on the London Stock market and found out that liquidity and variations could be the key factors (Farmer et al., 2004).

During the 2008 – 2009 Crash, investors in the stock market were negatively influenced by the exposure of consumer defaults on subprime mortgages and the resulting large-scale failures of financial institutions, such as the bankruptcy of Lehman Brothers. The S&P 500 index had experienced a huge 53.9% drop from the peak point of 1565.15 to the bottom of 676.53 during the October 9th 2007 - March 9th 2009 time period ("Historical Prices, S&P 500", n.d.). Even though the failure of financial institutions and the exposure of consumer defaults on subprime mortgages triggered the 2008-2009 market crash, investors' homogeneous trading decisions, here mainly selling orders, caused the market to drop suddenly and dramatically. Therefore, this encourages us to consider the impact of trading behaviors and the limit order book data inherent in the stock prices.

Based on the investor behavior standpoint, Baker and Wurgler (2007) found that the sentimental investors could drive the younger, lower capitalization, higher volatility,

and growth companies to fluctuate much more severely during the market volatile time periods, such as market crashes (Baker & Wurgler, 2007). Similarly, Zouaoui, Nouyrigat, and Beer (2011) also found that investor sentiment had a strongly positive relationship with the occurrence of the stock market crash within a one-year time period (Zouaoui, Nouyrigat, & Beer, 2011). Obviously, there is a common market crash point, the so-called “Minsky Moment”, for both of the two major market crashes. Right before the two major crashes, we can recall that the market was experiencing unsustainable growth and reached the peak level at that time. So investors were eager to put more money into the market during this unsustainable growth period. However, once the traders’ buying behaviors lead to the Minsky Moment, the market crashed down and into the global depression as the two major crashes had shown (Yalamova & McKelvey, 2011). What is more, the sequence of investors’ behavior is the simulator of the market phrase changes (Yalamova & McKelvey, 2011). Baker (2009) and Foster and Magdoff (2009) also mentioned that Wall Street, the Federal Reserve and other financial experts should have noticed the indisputable facts and cumulative risk of the derivatives, high leverage, and other subprime mortgages that were trading in the market.

2.2 Complex Network

Complexity science was founded in the 1980s. It uses non-linear mathematics to deal with problems in physics, chemistry, economics, society, biology and so on

(Prigogine, 1980). Complexity studies the interactions among the sub-systems and their properties, patterns, and mechanisms. And complexity theory can explain the evolution, emergence, and adaptability in complex networks. Complexity studies the whole complex network's properties that come from the interactions among the sub-systems. With the development of complexity theory, researchers have found that complex networks are an essential part of complexity theory. Complex networks promote the development of complexity science. All complex networks come from reality and exist around us all the time.

Watts and Strogatz (1998) published an article in *Nature* journal. They discussed the structure and dynamics of small world networks. They also found out that adding some small new information to a network can significantly cause the network to oscillate into a similar movement (Watts & Strogatz, 1998). This phenomenon is described as synchronization afterwards. In 1999, *Science* journal published Barabasi and Albert's (1999) article that showed us the scale-free complex network model. They have pointed out that scaling is to describe the self-organizing mechanism due to the individual participants' decisions in a scale-free network (Barabási & Albert, 1999). Over the next decades, scientists have devoted themselves to complex networks and have gained numerous meaningful results. With the rapid development of computer science, research on complex network has also developed quickly. The analysis of complex networks

changed from hundreds of nodes to millions of nodes. By analyzing different kinds of networks, researchers made significant research achievement. Firstly, scholars adopted new definitions and measurements to describe the topology of networks. Secondly, by simulating complex networks with the use of dynamic models, researchers were able to display the topology of real complex networks. The nodes in the network are abstracted out of the real interacted individuals. The lines between nodes represent the interactions. All the nodes and their connections form a network.

In networks, all the calculations are dependent on the adjacency matrix. The matrix has N^2 orders. We can use the average connection length to represent the relevance of nodes.

$$L = \frac{1}{\frac{1}{2}N(N+1)} \sum_{i \geq j} d_{ij} \tag{2.1}$$

In equation 2.2, N represents the number of nodes in a network. d_{ij} is the distance between node i and node j , representing the shortest distance. The maximum distance between two random nodes is the diameter of this network, represented by D :

$$D = \max d_{ij} \tag{2.2}$$

As the emergence of scaling in network, Barabási and Albert (1999) also found that the common feature of natural complex networks is the nodes' correlations following a scale-free power law distribution (Barabási & Albert, 1999). The scale-free power law

distribution is as follows:

$$P(k) \sim k^{-\gamma} \quad (2.3)$$

Here, $P(k)$ denotes the probability of one node having k number of edges with other nodes. While γ denotes the power of those edges, γ has a range of 2 to 3 in most networks (Barabási & Albert, 1999). The scale-free power law distribution has a long tail for larger k .

2.2.1 Stock Market Complex Network. In the stock market, stock prices are characterized by investors' opinions of the company and influenced by all the information exchanged with other investors or market participants (Onnela, Saramäki, Kaski, & Kertész, 2006). So, stock prices contain vital information regarding stock market volatility and movement. Mantegna (1999) was the first to build a network for the stock market and found the hierarchical structure within the stocks in the analyzed portfolio.

Furthermore, Vandewalle, Brisbois, and Tordoir (2001) analyzed the cross correlations of stock daily returns in the US market by building the Minimum Spanning Tree and confirmed the slow and local structure evolving in the stock market. Onnela et al. (2006) constructed a NYSE traded stock network in which the stocks represent nodes and interactions between stocks' exhibit edges. They also found that there were clusters or so-called orders within the stock market network. Leibon, Pauls, Rockmore, and Savell (2008) introduced a new method to study the topological structure and to display the

scale-dependent distribution within many complex networks. They analyzed the daily return correlation matrix built from the New York Stock Exchange (NYSE) and National Association of Securities Dealers Automated Quotation (NASDAQ) traded stocks. And they found the existence of scales corresponding with the movement inside the stock market and that the stock market is a classic complex network (Leibon et al., 2008). Tse, Liu and Lau (2010) analyzed the cross correlations of all the US stocks traded over a specific time period and reported the scale-free degree distribution in stock price returns and trading volumes based stock market networks. Tse et al. (2010) also concluded that the variation of the majority stock prices was strongly correlated with a relatively small number of highly connected stocks, which corresponded to the scale or cluster conclusions from previous researchers.

2.2.2 Scaling and Synchronization in the Stock Market Complex Network. As mentioned above, synchronization and scaling are the two key characteristics in the natural complex network (Watts & Strogatz, 1998; Barabási & Albert, 1999). What's more, researchers have worked on the demonstration on the scaling of stock market complex network as well.

Scaling is an essential feature in complex networks. It describes the self-organizing mechanism due to the individual participants' decisions in a scale-free network (Barabási & Albert, 1999). In addition, scaling is the mechanism for the

accelerating growth in a network once some connections are enhanced. It is also the growth engine within most of the common networks, such as genetic networks, the World Wide Web system, business networks, and social networks that describe individuals or organizations (Barabási & Albert, 1999). Barabási and Albert (1999) found evidence of a self-organization characteristic and the power law or scale-free distribution, $P(k) \sim k^{-\gamma}$, in complex networks. $P(k)$ is the probability that one individual interacts with k other individuals (Barabási & Albert, 1999). Barabási and Albert (1999) have also proved that growth and preferential attachment within natural networks are the key mechanisms for network evolution, including business networks, which explains the ‘richer-get-richer’ phenomenon (Barabási & Albert, 1999). In 2000, Albert and Barabási extended their research on the power law distribution in complex networks and developed a phase diagram theory to predict the scaling exponents. And they concluded in favor of the existence of scale-free phase and exponential phase (Albert & Barabási, 2000).

Later in 2002, H. Kim, Kim, Lee and Kahng analyzed the network composed of S&P 500 constituents and found the power law distribution in the sum of all the connected edge weight to each node (H. Kim, Kim, Lee, & Kahng, 2002). H. Kim et al. (2002) results have further proved the scale-free distribution existing within the connection strength of a stock market network (Kim et al., 2002). They also expected that pullback of one single stock among the most influential companies could lead to a crash

in the stock market due to the power-law distribution (Kim et al., 2002). They also found the exponent of the power-law distribution for the S&P 500 constituents network to be around 1.8 ($\gamma \approx 1.8$) (Kim et al., 2002). Afterwards in 2003, Guimerà, Danon, Díaz-Guilera, Giralt, and Arenas studied a social email network and found the scaling and self-organized feature within the network of human interactions (Guimerà, Danon, Díaz-Guilera, Giralt, & Arenas, 2003). In the same year of 2003, Ravasz and Barabási proved that the scaling and self-organization features of complex networks were due to the hierarchical structure of complex networks (Ravasz & Barabási, 2003). Then, Amaral and Ottino (2004) summarized the literature on the important areas for the study of complex networks. They supported the conclusion that scaling was vital to study the critical phenomenon that led to the structure changes in an evolving network (Amaral & Ottino, 2004). What is more, scaling and scale-free distribution can also explain the correlated volatility which often occurred in the stock market. For example, different companies' stock prices can drop together even though there's no information released for this, which differs with the Efficient Market Hypothesis. In summary, scaling and scale-free distribution have been proven by various researchers to be vital adaptive features and to be the growth engine for the exponential growth or decay and volatility evolution within a stock market network.

Synchronization is another vital characteristic existing in natural complex

networks. Synchronization describes the phenomenon that adding some small new information to a network can significantly cause the network to oscillate into a similar movement (Watts & Strogatz, 1998). We suggest market crashes in the stock market occur as a result of the expression of synchronization within the evolving and self-organized stock market complex network. Barahona and Pecora (2002) identified synchronization could lie within the phase diagram boundary, which might lead to the phase change of a complex network (Barahona & Pecora, 2002). Nishikawa, Motter, Lai, and Hoppensteadt (2003) further proved the synchronizability of networks especially those with a higher degree of homogeneity, such as neural networks (Nishikawa, Motter, Lai, & Hoppensteadt, 2003). Krause (2004) conducted an empirical study on the crashes of evolving complex networks that contain extinct individuals. Krause (2004) found a high degree of homogeneity in the investment choices before the stock market crashes. He also presented the figure that showed the variance of behaviors decreased significantly before a crash (Krause, 2004). In the stock market, synchronization describes the highly homogeneous traders' behaviors, such as herding, imitation in the bubble building-up stage in the stock market.

In order to reveal the relationship between synchronization and scaling, Arenas, Díaz-Guilera, and Pérez-Vicente (2006) studied the dynamic movement towards the synchronization of a complex system. They concluded that modular structure and nodes

emerged and evolved during the synchronization process. This shows us that it is important to pay attention to the structure change before and after crashes. As noted in Arenas, Díaz-Guilera, Kurths, Moreno, and Zhou's research (2008), they summarized the results of using the correlation return matrix to study the synchronization pattern in stock markets (Arenas, Díaz-Guilera, Kurths, Moreno, & Zhou, 2008). Arenas et al. (2008) concluded that stocks could synchronize and be strongly connected by some interactions in the market, such as money flows or sector correlations (Arenas et al., 2008). In 2011, Gómez-Gardeñes, Moreno, and Arenas further proved the synchronization patterns differ between homogeneous and heterogeneous complex networks. And they concluded that nodes and scaling clusters are the key drivers during the synchronization transition (Gómez-Gardeñes, Moreno, & Arenas, 2011). In 2013, Singh, Sreenivasan, Szymanski, and Korniss applied a threshold model to reveal the fact that individual opinion could become a threshold point once all the neighbors adopted the same opinion. They also concluded that the local clustering promoted the synchronization phenomenon in a high-school friendship network (Singh, Sreenivasan, Szymanski, & Korniss, 2013). In 2014, Brú, Alós, Nuño, and de Dios built a graph to show the growing scaling interface in dynamic networks. They concluded that graphs could also reveal the scaling property in complex networks and critical exponent existed in the network as well (Brú, Alós, Nuño, & de Dios, 2014).

All in all, scaling can be used to explain the market volatility and evolution and the bubble building-up mechanism in a stock market network. And synchronization describes the highly homogeneous behavior or stock price coincident movement in a stock market network.

2.3 Phase Transition Theory

According to the Efficient Market Hypothesis, the market equilibrium state should reflect all the available information in the market (Fama, 1970). And the equilibrium expected return is the expressed form of the market equilibrium state (Fama, 1970). Once the market information is not available to everyone, the market will step into a disequilibrium state. There will be uninformed and informed investors regarding some specific information in the market. Therefore, in order to explain the abnormal market movement or disequilibrium state in stock market, Grossman and Stiglitz (1980) studied the market disequilibrium state reflected by the stock prices and the degree of uninformed investors influenced by the informed investors. They also proved the impact of the price system on information spreading from informed traders to uninformed investors by building a mimic stock market model, which would be considered to be a reason for the 'herding behavior effect' in the stock market (Grossman & Stiglitz, 1980). In other words, the limit order book, such as bid orders or ask orders, is believed to contain information from informed investors. If the number of ask orders exceeds the bid orders, this would

show a good perspective for this stock. This means that the information here is not fully public and efficient to everyone. Information becomes costly here, which would influence the uninformed investors to imitate the informed ones (Grossman & Stiglitz, 1980).

In 1996, Sornette, Johansen, and Bouchaud studied the time series of S&P 500 index before and after the 1987 stock market crash and found the existence of a log-periodic oscillation price pattern with a dynamical critical point during the crash (Sornette et al., 1996). Sornette et al. (1996) also suggested a phase transition theory to explain the log-periodic pattern. Afterwards, Sornette (2006) has further proved the existence of critical events in stock market complex networks and other natural networks (Sornette, 2006). In addition, Sornette also fully explained the stock market crash by applying the critical point theory (Sornette, 2009). In a market disequilibrium state, once the uninformed traders start to make investment decisions based on other traders' behaviors, both informed and uninformed traders would affect each other to move the stock price regardless of rationality (LeBaron, 2001). LeBaron (2001) applied the agent based model to explain the similar herding effect above. LeBaron (2001) found that rational agents and non-rational agents would interact with each other and lead to higher volatility or large price jumps (LeBaron, 2001). The influence from rational agents on non-rational agents would cause the imitating behavior or herding behavior, similar to the effect of asymmetric information in a market disequilibrium state. This ultimately will

lead up to market crash if there is no market regulation or interfering. In this study, we can take the market crash building-up stage as an apparent market disequilibrium state.

Based on the empirical study and complex network theory above, Yalamova and McKelvey (2011) built an innovative Phase Transition Model analogical from physics theory to explain the homogeneous behaviors, such as herding behavior in stock market (Yalamova & McKelvey, 2011). According to their model, the imitating behavior or herding behavior occurs at the ‘tipping point’, which eventually triggers a crash (Yalamova & McKelvey, 2011). This explanation also corresponds to the synchronization and scaling phenomenon existing in complex networks. In addition, Yalamova and McKelvey (2011) have also illuminated the existence of a critical point at which the highest level of homogeneous trading behavior happens, that is, the market crash point. Besides this, they have pointed out that the building-up mechanism of homogeneous trading behavior is driven by the scaling and synchronization characteristics of complex network in the stock market (Yalamova & McKelvey, 2011).

The Phase Transition Theory and other empirical evidence have converged to provide us a solid theory to explain the mechanism of stock market crashes. Meanwhile, the studies from both stock market complex networks and other complex networks have also contributed a firm background to extract the structure of complex networks. To the best of our knowledge, we find no empirical research to extract the cluster structures

from the market equilibrium state to a bubble building-up state and to support the Phase Transition Theory. Therefore, it is worthwhile to apply the complex network clusters extraction method to study the structure changes during stock market crashes. What's more, this study will allow us to contribute both to the empirical analysis on the dynamics of the stock market network and on the growing literature of econophysics.

3. Hypotheses Development

As summarized in the Literature Review part, there are at least two different market states, the market equilibrium state and the market disequilibrium state. Under a market equilibrium state or an Efficient Market, the stock expected return should have fully revealed the available information in the market (Fama, 1970). According to the study on scaling and synchronization feature in complex networks and the Phase Transition Theory, we apply these to the stock market network and reveal the cluster movement to prove the existence of a critical point before stock market crashes by analyzing the stock market price correlations matrix (Onnela, Chakraborti, Kaski, Kertész, & Kanto, 2003). In this study, we expect to observe the number of clusters changing from a market equilibrium state to the critical point before market crashes by computing the stock daily return correlation matrixes in some specific ways. Therefore, our hypotheses are as follows:

In market equilibrium state, there are mainly sector clusters because of the high correlation among stock prices within the similar industries, such as financials or technologies (Leibon et al., 2008). Besides, based on Foti, Hughes, and Rockmore's (2011) results, there exist 22 sector clusters for the US stock market. Therefore, in the market equilibrium state, we also expect to observe the 22 sector clusters and develop the first hypothesis.

H1: In the market equilibrium state, there should be at least 22 clusters.

According to the Efficient Market Hypothesis, the market equilibrium state should reflect all the available information in the market (Fama, 1970). And the equilibrium expected return is the expression form of the market equilibrium state (Fama, 1970). Therefore, if the market is still in a market equilibrium state, there should always be sector clusters and there should exist a similar number of clusters during different time periods. In the market equilibrium state, we also expect to observe sector clusters and the number of clusters should be similar even during different time periods. Hence, we develop the second hypothesis here.

H2: In the market equilibrium state, there should be a similar number of clusters during different time periods.

However, if it is in a market disequilibrium state, the specific information is only available to the informed investors and will lead to the uninformed investors' herding behaviors in the market (Grossman & Stiglitz, 1980). Once this herding behavior becomes increasingly severe, it will reach a common market crash point, the so-called "Minsky Moment" or Critical Point (Yalamova & McKelvey, 2011) in a stock dynamic network. According to the study on the scaling and synchronization feature in complex networks and the Phase Transition Theory, the critical point represents the extreme synchronization phenomenon in the scaling process of dynamic complex networks. At the

critical point, the stocks are highly correlated despite the different sectors. Therefore, the critical point captures the patterns of stock market crashes (Yalamova & McKelvey, 2011). So, there will be fewer clusters because of the higher and wider correlation among stock prices within the whole market despite the variation in the sectors or industries. Therefore, we expect to observe fewer but larger clusters, sometimes even only one large cluster, during the critical point building-up time period in the market disequilibrium state. So, in the market disequilibrium state, we develop the third hypothesis.

H3: In an extremely evident market disequilibrium state, such as the pre-crash critical point, there should be fewer clusters or even only one cluster.

4. Data and Methodology

4.1 Data

In order to construct the daily return matrix, our primary data source is from the Center for Research in Security Prices (CRSP). The S&P 500, or the Standard & Poor's 500, is founded to include the 500 selected stocks traded in New York Stock Exchange (NYSE) and National Association of Securities Dealers Automated Quotation (NASDAQ). The S&P 500 index (Ticker: SPX) is the second largest US market index following the Dow Jones Industrial Average (DJIA). Compared with DJIA, the S&P 500 index contains a larger number of large capital public companies and better captures the movement of the US market. The S&P 500 index is considered to be the best market benchmark index and captures approximately 80% coverage of available market capitalization (S&P Dow Jones Indices, 2015). In order to best capture the US stock market movement and scaling transition, we decide to analyze the S&P 500 index constituents in our study. All the current constituent companies are listed in Appendix.

In order to observe the desired structure movement, we chose the 2008-2009 stock market crash as our crash event in this study. Therefore, we first decided to study the market daily returns during the time period from Jan 2nd 2002 to Dec 31th 2010. In order to capture all the useful price movement information for the 2008-2009 market crash, we included whatever constituent companies that have been in the S&P 500 into

our list during our chosen time period. We downloaded the valid constituents' historical daily returns from the CRSP data center for our chosen time period from Jan 2nd 2002 to Dec 31st 2010. Daily total return is the combination of intraday return and overnight return. The formula for daily return is as follows;

$$R_i = \frac{P_t - P_{t-1}}{P_{t-1}} \quad (5.1)$$

Here P_t is the current day close price for stock i , P_{t-1} is the previous day close price for the same stock i , and R_i represents the current daily return for stock i . Then, we removed all the stocks that missed over 30% daily returns, after which we had 581 stocks left (Leibon et al., 2008). Lastly, we constructed the spreadsheet containing all these 581 stock daily returns from Jan 2nd 2002 to Dec 31th 2010.

4.2 Methodology

In the history of study on the complex network structure, it is not difficult to find out a large number of studies applying correlation matrices. In 1999, Mantegna as the first researcher who found the complex network characteristics in financial markets applied the matrix of correlation coefficients for daily prices (Mantegna, 1999). Similarly, Nishikawa, Motter, Lai, and Hoppensteadt (2003) also applied a correlation matrix to compute the eigenvalue spectrum in order to get the connection topology of networks (Nishikawa et al., 2003). Moreover, Hong, Kim, Choi, and Park (2004) built a coupling matrix to construct a Watts-Strogatz small-world network in order to study the factors

that predict precise synchronizability. They also calculated the eigenvalues of the coupling matrix with different factors and concluded that betweenness centrality could be a better signal for the prediction of synchronization (Hong, Kim, Choi, & Park, 2004). Their results also illuminate that the correlation strength between every two nodes will influence the emergence of a critical point in complex networks.

Afterwards, Kim and Jeong (2005) introduced a maximum likelihood clustering method that also applied the correlation matrix to study the clusters in stock markets. They found the eigenvectors concentrated together belonged to one common industry (Kim & Jeong, 2005). What's more, Tumminello, Di Matteo, Aste, and Mantegna (2007) also concluded that clusters generated corresponding to the economic sectors from the analysis of correlation matrix and the application of a method called Planar Maximally Filtered Graph. Next, Leibon, Pauls, Rockmore, and Savell (2008) applied an innovative method to present the topological structure in the stock market network. They detected the clusters' structure by applying a four-year daily return correlation matrix to the hierarchical spectral clustering and decoupling methods package (Leibon et al., 2008).

In summary, in order to extract the information from a large correlation matrix, hierarchical clustering method has been often and widely used in most of the scientific areas, such as biology, physics, and economics (Omran, Engelbrecht, & Salman, 2007). The traditional hierarchical clustering method starts with one node in the network, and

then merges the next two most similar clusters by applying the eigenvalue and similarity linkage calculation and comparison (Omran et al., 2007). In addition, the hierarchical clustering method generates a cluster linkage tree which can be displayed by a dendrogram form (Omran et al., 2007).

4.2.1 PHA Clustering. In this study, we applied a Potential-based Hierarchical Agglomerative (PHA) clustering method (Lu & Wan, 2013). By applying this Potential-based Hierarchical Agglomerative (PHA) method, we built the dendrogram linkage trees to find the number of clusters under the different states of the stock market. The PHA method is a novel hierarchical clustering method based on the construction of a hypothetical potential field and the pattern recognition progress of hierarchical clustering metric (Lu & Wan, 2013).

Shi, Yang, and Wang (2002) applied the potential model to the hierarchical clustering process and pointed out the advantage of the potential based hierarchical clustering process over the traditional clustering processes. Afterwards, Yamachi, Kambayashi, and Tsujimura (2009) have presented a clustering method based on the potential field to optimize the correlated effects and to capture the most stable clustering. Based on the previous study on potential-based clustering, Lu and Wan (2013) have proposed this novel Potential-based Hierarchical Agglomerative (PHA) clustering method and have proved the effectiveness of it. Basically, a potential model is constructed by

defining the potential between node i and node j . If the edge between node i and node j is r_{ij} , then we can set the potential between them as follows (Lu & Wan, 2013).

$$\Phi_{ij}(r_{ij}) = \begin{cases} -\frac{1}{r_{ij}} & \text{if } r_{ij} \geq \delta \\ -\frac{1}{\delta} & \text{if } r_{ij} < \delta \end{cases} \quad (4.1)$$

Here, the parameter δ is determined from the correlation matrix of the data set by finding the average of the minimum edges between node i and all the other nodes for node i to node n (M_i) and dividing the mean by a scale factor S . The formula for parameter δ is as follows. The value of scale factor S is set to 10 in order to have a better balance between sensitivity and robustness (Lu & Wan, 2013).

$$\delta = \frac{\text{mean}(M_i)}{S} \quad (4.2)$$

The total potential value for node i is summed by all the potential value of nodes connected with node i .

$$\Phi_i = \sum_{j=1}^N \Phi_{ij}(r_{ij}) \quad (4.3)$$

Following the potential model, Lu and Wan (2013) have launched a new similarity metric combining the potential field information and the data set information.

The last step is to extract the clusters based on the edge-weighted tree of the data set by

applying another similarity computation method (Lu & Wan, 2013).

In order to illustrate the PHA clustering process, we present an example with a six-node data set here: N1, N2, N3, N4, N5, N6; and they are located at (0.4, 0.8), (0.6, 1.0), (1.4, 0.5), (2.0, 1.0), (2.3, 0.5), and (2.4, 0.7) respectively as shown in Fig 4.1. In this figure, the horizontal and vertical axes show the coordinates for all the six nodes. Besides this, the potential edges between any two different nodes are marked with the underlined numbers near the “edge” respectively in Fig 4.1. According to the potential model (4.1), we discover all the potential values respectively marked with the numbers in parentheses in Figure 4.1. Sorting all the calculated potential values, we find the sequence of all the six nodes to be: $N6 < N5 < N2 < N4 < N1 < N3$.

Therefore, the node containing the lowest potential value, N6, has been chosen as the first root. And the nodes containing the second lowest potential has been selected and is the nearest one connected with N6. Then, N2 is the next one chosen and is connected to N5 regarding the potential values. Similarly, N4 has been picked as the next one and is connect to N6 prior to N3, because the correlation between N4 and N6 is smaller than that between N4 and N3. And N1 is the next to be selected and connected with N2. Then, N3 is chosen and connected to N4.

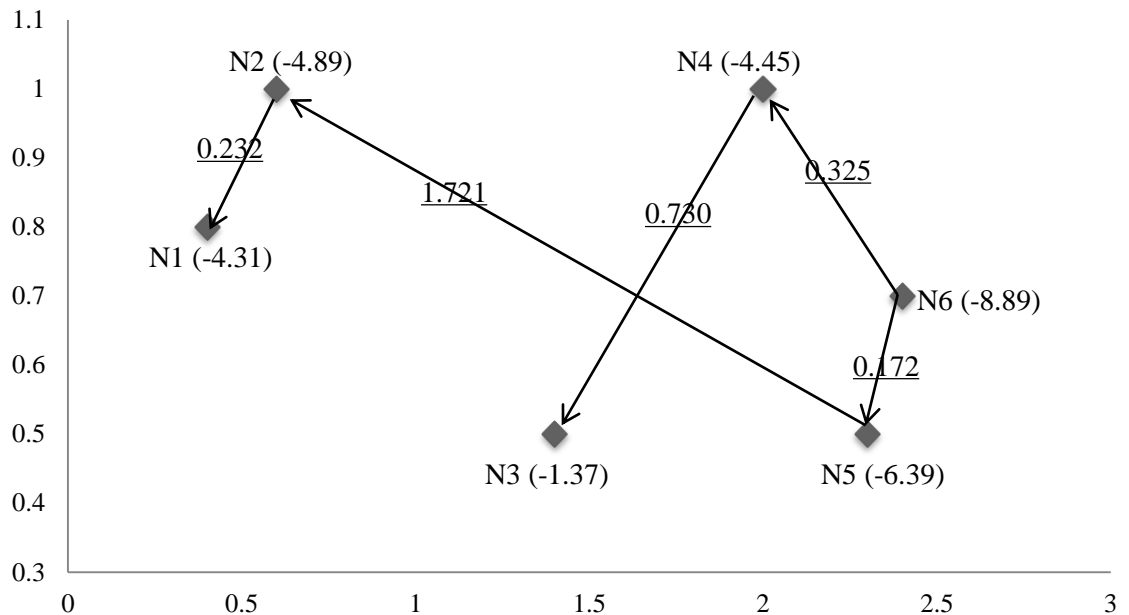


Figure 4.1. The distribution of the data points. Adapted from “PHA: A Fast

Potential-based Hierarchical Agglomerative Clustering Method,” by Y. Liu and Y. Wan, 2013, *Pattern Recognition*, 46, p. 1127-1239. Copyright 2012 by Elsevier Ltd.

So, summarizing the results above, we can tell that N6 and N5 are merged first as a new small cluster. According to the correlation or edge strength, N2 has merged with N1 to form (N2, N1) and is followed by the merge between N4 with the new small cluster (N6, N5). Then, N3 is merged with the newer cluster (N6, N5, N4) to form (N6, N5, N4, N3). And lastly, (N2, N1) has merged with (N6, N5, N4, N3). Finally, the dendrogram figure is built regarding the merging sequence mentioned and the respective correlation strength in Figure 4.2. In the dendrogram graph, the vertical axis shows the heights of all the different U-shapes representing the relative connection distance between any two small clusters or nodes.

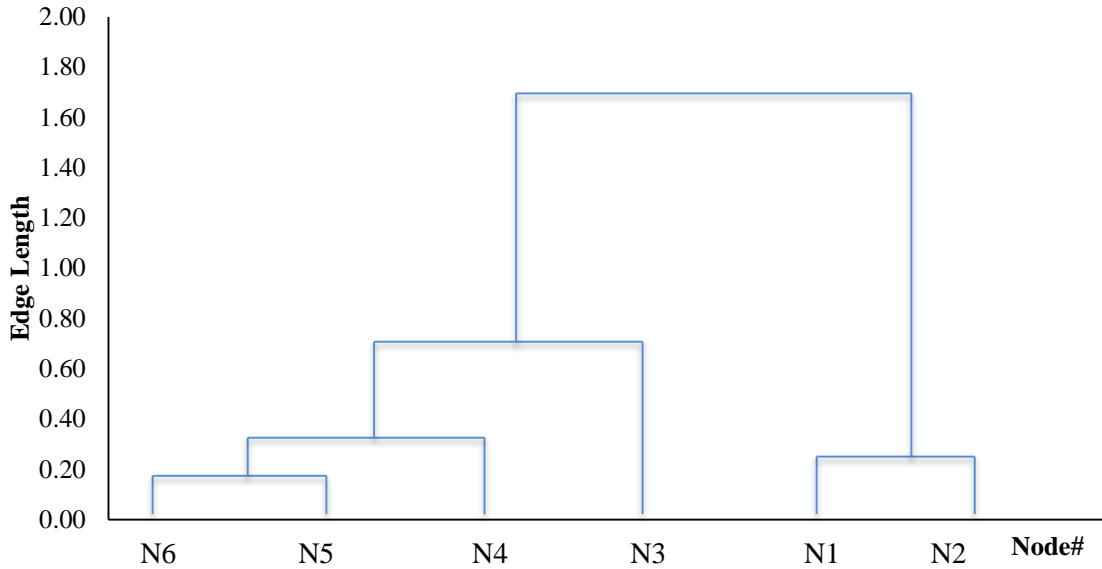


Figure 4.2. The dendrogram derived from the PHA method. Adapted from “PHA: A Fast Potential-based Hierarchical Agglomerative Clustering Method,” by Y. Liu and Y. Wan, 2013, *Pattern Recognition*, 46, p. 1127-1239. Copyright 2012 by Elsevier Ltd.

4.2.2 Backbone Extraction Method. Based on Leibon et al.’s method, Foti, Hughes, and Rockmore (2011) published a better method called *LANS* to extract the significant backbone of complex networks. They drew from the traditional “Disparity filter” method to get the fractional edge weight for every two nodes (Serrano, Boguñá, & Vespignani, 2009; Foti et al., 2011). Let

$$p_{ij} = \frac{w_{ij}}{\sum_{k=1}^{N_i} w_{ik}} \quad (4.4)$$

define the fractional edge weight from node i to node j . Here, N_i represents the number of all the neighbors of node i (for a weighted network, node j is a neighbor of node i if w_{ij} is larger than 0) (Foti et al., 2011). And w_{ij} denotes the edge weight

between node i and node j , which is the correlation from the daily return time series between stock i and stock j (Foti et al., 2011).

Next, in order to get the significant edge weight for all the nodes in a network, Foti et al. (2011) introduced an indicator function method to determine the selection of significant edge weight. Let

$$\hat{F}(p_{ij}) = \frac{1}{N_i} \sum_{k=1}^{N_i} 1\{p_{ik} \leq p_{ij}\} \quad (4.5)$$

Here, $1\{\}$ denote the indicator function, N_i represents the number of neighbors of node i in the network, and the sum is to count the number of node i that the edge weight is less or equal than p_{ij} (Foti et al., 2011). $\hat{F}(p_{ij})$ denotes the probability of all the fractional edge weights that are less than or equal to p_{ij} (Foti et al., 2011). If we can select a significance level α and that $1 - \hat{F}(p_{ij})$ is less than α , we can say that this edge weight between node i and node j is significant and extract it into the backbone network (Foti et al., 2011). According to the LANS pseudo code provided by Foti et al. (2011), we develop a MatLab programming to help us compute and select all the backbone components. By applying the *LANS* backbone extraction method, we construct a new correlation matrix with all statistically significant edge weights based on the normal correlation matrix. In this way, we can screen out the insignificant edge weight or outliers existing in the normal correlation matrix.

4.2.3 Dot Matrix Plot. Dot Plot is widely used in the bioinformatics industry. Here, we apply one of the Network Toolboxes from the MIT network research center, Dot Matrix Plot. The Dot Matrix Plot can draw the matrix as a column and row displayed square dot-plot pattern according to the given number of clusters and the correlation matrix. Based on our previous methodology, we can discover the number of clusters by applying the PHA clustering method (Lu & Wan, 2013). As well, we can get the highly significant backbone correlation matrix from the LANS method (Foti et al., 2011). By applying the Dot Matrix Plot, we get the image display of the clusters from the backbone correlation matrix. Within the Dot Matrix Plot, we apply one of the algorithms by Newman and Girvan to display the sparsity plot of the clusters structures in complex networks sorted by ‘betweenness’ (Newman & Girvan, 2004). Newman and Girvan (2004) introduced a set of measures to split the nodes into different by computing a number of ‘betweenness’ measures. They proposed their algorithm to effectively detect the community structures in real-world complex networks (Newman & Girvan, 2004). By choosing the number of clusters computed by the PHA clustering method, we are able to discover the clusters structures or the so-called modularity in the Newman- Girvan algorithm.

5. Results and Discussion

5.1 PHA Clustering

5.1.1. Market Equilibrium State Clusters. Since we choose the 2008-2009 stock market crash as our study event, we selected the time period from the beginning of 2002 to the end of 2005 in order to reflect the market equilibrium state. We found the daily return of the 582 available stocks during the time period from January the 2nd 2002 till December the 30th 2005. We ran the data with the PHA clustering method and record the results in Figure 5.1.

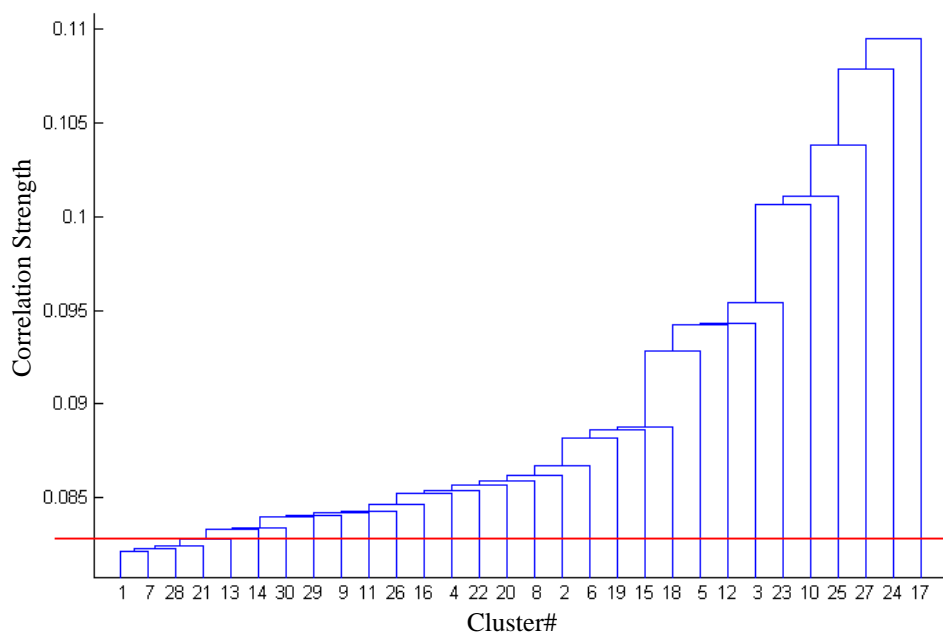


Figure 5.1. PHA result for the time period of Jan 2nd 2002 to Dec 30th 2005

As we have explained in Section 4, Methodology, we use this dendrogram figure to display the computation results of the clusters during 2002 and 2005. In Figure 5.1, the

dendrogram figure actually shows us the cluster tree from our data set. The height of the U shapes or the corresponding value on the vertical axis in the dendrogram represents the distance between the two nodes. Applied to our data set, the height of the cluster tree shows us the correlation strength between any two nodes or small clusters. For the time period during 2002 to 2005, in order to capture as many as clusters, we set the correlation threshold as 0.08 and we found 27 clusters as shown in Figure 5.1. Therefore, we can conclude that there are 27 clusters in the market equilibrium state.

Next, we also computed the number of clusters for several different time periods within the market equilibrium state to test our Hypothesis 2. We calculated the time periods of Jan 2nd 2002 to Dec 31st 2002, Jan 2nd 2003 to Dec 31st 2003, Jan 2nd 2004 to Dec 31st 2004, and Jan 3rd 2005 to Dec 30th 2005. We present the results for the four time periods as Figure 5.2, Figure 5.3, Figure 5.4, and Figure 5.5 respectively. If we continue to use the threshold of 0.08 from the time period of Jan 2nd 2002 to Dec 30th 2005, we found similar number of clusters during the four time periods. For the time period of Jan 2nd 2002 to Dec 31st 2002, we found 28 clusters as shown in Figure 5.2. For the time period of Jan 2nd 2003 to Dec 31st 2003, we found 26 clusters as shown in Figure 5.3. For the time period of Jan 2nd 2004 to Dec 31st 2004, we found 30 clusters as shown in Figure 5.4. For the time period of Jan 3rd 2005 to Dec 30th 2005, we found 30 clusters as shown in Figure 5.5.

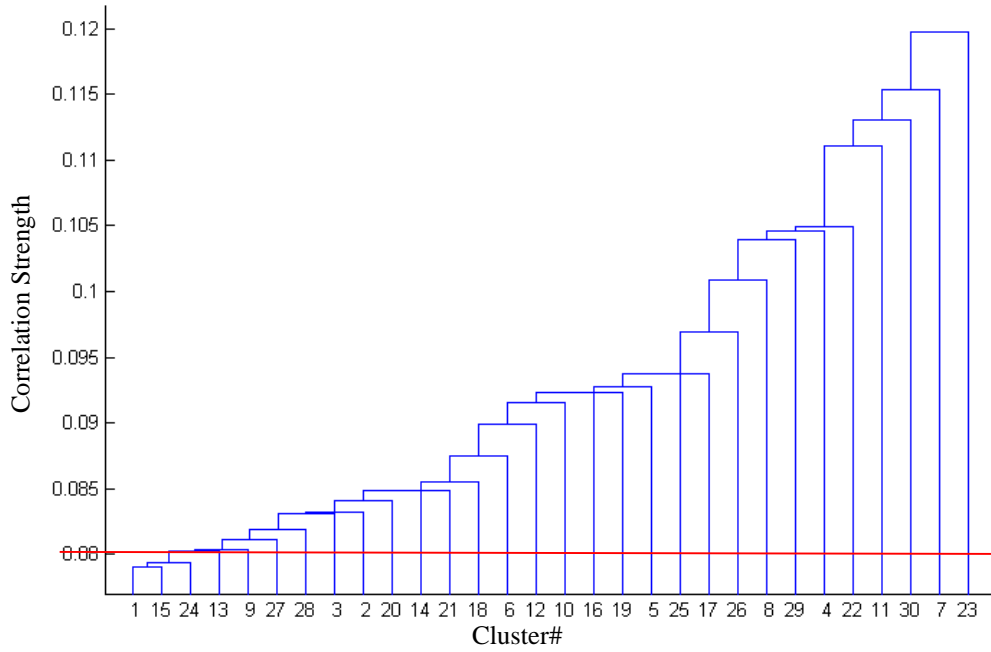


Figure 5.2. PHA result for the time period of Jan 2nd 2002 to Dec 31st 2002.

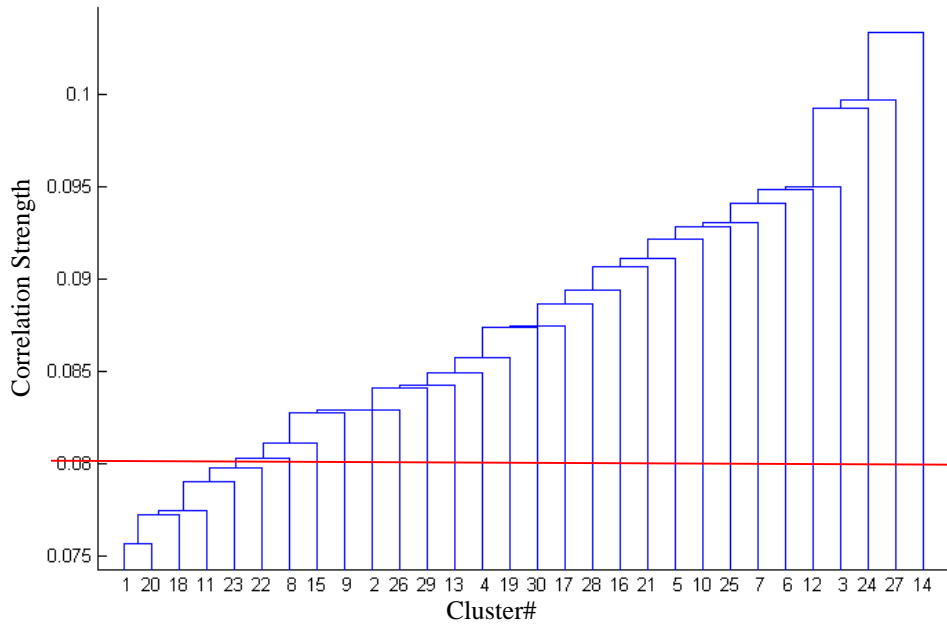


Figure 5.3. PHA result for the time period of Jan 2nd 2003 to Dec 31st 2003.

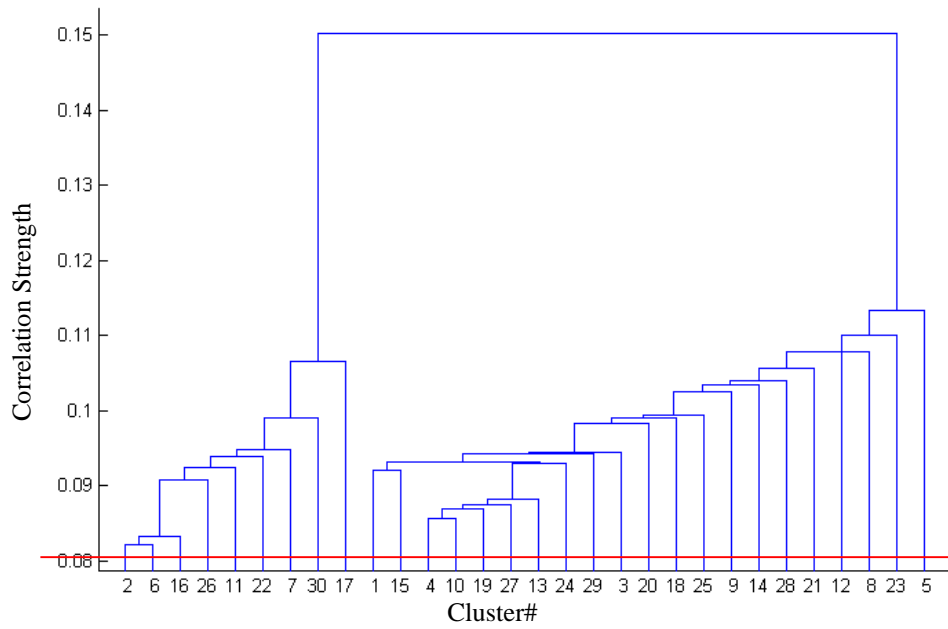


Figure 5.4. PHA result for the time period of Jan 2nd 2004 to Dec 31st 2004.

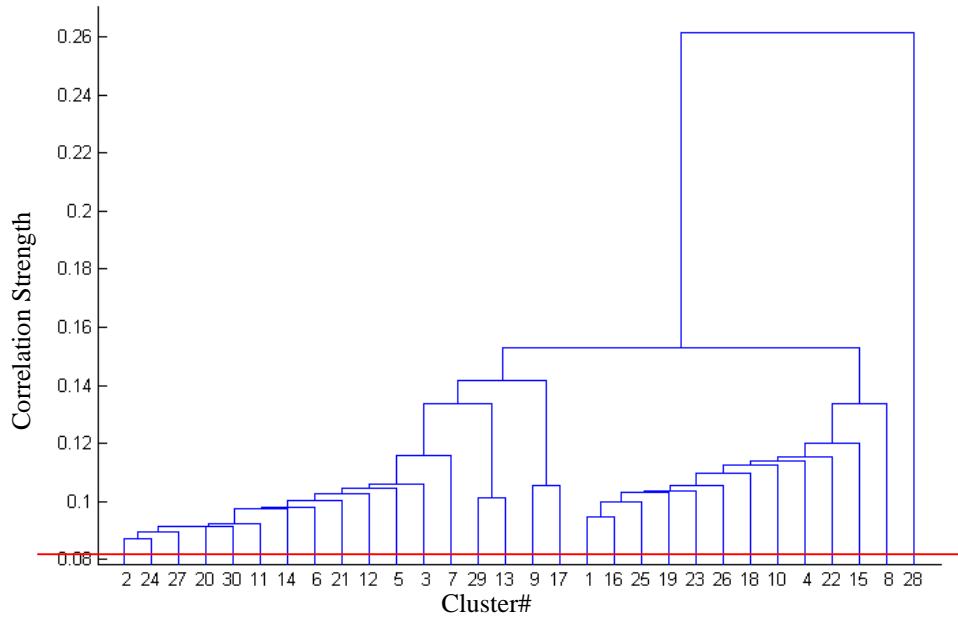


Figure 5.5. PHA result for the time period of Jan 3rd 2005 to Dec 30th 2005.

5.1.2 Market Disequilibrium Clustering. In order to capture the clustering building time period, we selected the time period from 2006 to 2009 as our Market Disequilibrium state. We assessed the daily return of the 582 available stocks during the time period from January the 3rd 2006 till September the 15th 2008 from the CRSP data center. We ran the data with the PHA clustering method and report the results in Figure 5.6.

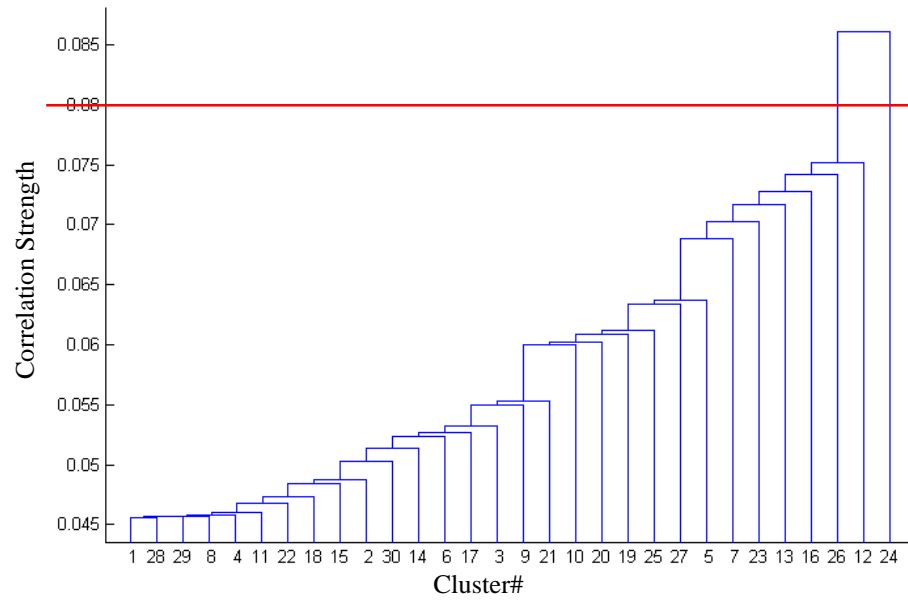


Figure 5.6. PHA result for the time Period of Jan 3rd 2006 to Sept 15th 2008.

For the 2006 to 2008 time period, we applied the threshold of 0.08 from the Market Equilibrium State. As shown in Figure 5.6, we can see the number of clusters changing from 27 clusters to only one cluster. Therefore, we can conclude that there is only one cluster in the Market Disequilibrium State.

Overall, the results from the PHA clustering method show us that the number of

clusters changed from 27 clusters of the market equilibrium state to 1 cluster of the market disequilibrium state. So, the results support our Hypothesis 1 and Hypothesis 2.

5.2 LANS & Dot Matrix Plot

In order to present a visual display of the clusters' movement, we applied the LANS method and Dot Matrix Plot method to show this difference between the Market Equilibrium State and Market Disequilibrium State. By applying the LANS backbone extraction method to our data set, we indicate the significant correlation matrix to be at a significant level of 0.05. After extracting the highly significant correlation matrix, we applied the Dot Matrix Plot and discovered the clusters structures for different market states. We chose one of the plot methods, Newman-Girvan algorithm to capture these clusters structures during different time periods. The results for the Newman-Girvan algorithm recorded as follows.

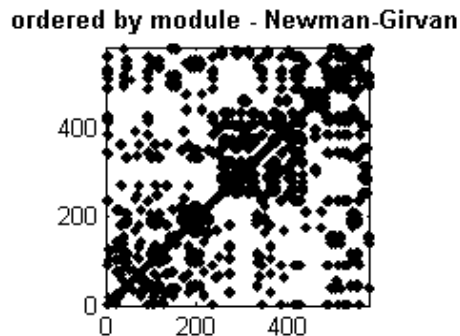


Figure 5.7. Dot Matrix Plot for the time period of Jan 2nd 2002 to Dec 30th 2005.

For the 2002-2005 market equilibrium state, we reported the results in Figure 5.7. In this dot matrix plot figure, both the horizontal and vertical axes represent the actual

stocks in our target stock market network. From the result for the market equilibrium state, we can tell there are both larger and smaller clusters in this network. However, for the market disequilibrium state, we found the clusters structure as follows.

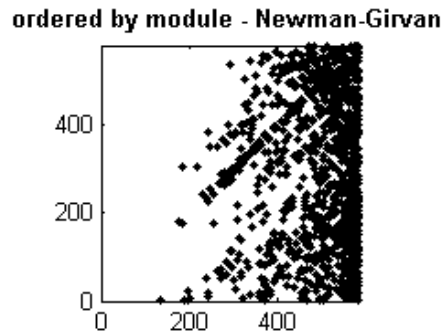


Figure 5.8. Dot Matrix Plot for the time period of Jan 3rd 2006 to Sept 15th 2008.

For the 2006-2009 market disequilibrium state, we found the results shown in Figure 5.8. From this result for the market disequilibrium state, it is obvious that the large and small clusters from the market equilibrium state have merged together to a larger cluster.

5.3 Discussion

5.3.1 Market Equilibrium State. Figure 5.1 and Figure 5.7 show us the number of clusters and visual dot plot display of the stock network in the market equilibrium state respectively. If we assume the market is in the market equilibrium state, we will get sector clusters according to Leibon (2008)'s US stock market structure study (Leibon et al., 2008). Furthermore, Foti and his group (2011) have proved that there are 22 clusters within the S&P 500 index constituents when presenting the extraction method of

significant correlation matrix. For the 2002 to 2005 time period or the assumed market equilibrium state, we found 27 clusters from the S&P 500 constituents by applying the PHA clustering method. In addition, the Dot Matrix Plot showed us the visual display of clusters structure in the market equilibrium state, which corresponded to the result of the PHA method. Therefore, the PHA clustering and the Dot Matrix Plot results support our Hypothesis 1, which indicates that there should be at least 22 clusters in the market equilibrium state.

In order to test Hypothesis 2, we applied the PHA clustering method to four shorter time periods within the selected time period of Jan 2nd 2002 to Dec 30th 2005 in the market equilibrium state. Then, we received the results as shown in section 5.1. We continued to use the threshold of 0.08 from the time period of Jan 2nd 2002 to Dec 30th 2005. And we can found similar number of clusters during the four time periods. For the time period of Jan 2nd 2002 to Dec 31st 2002, we found 28 clusters as shown in Figure 5.2. For the time period of Jan 2nd 2003 to Dec 31st 2003, we found 26 clusters as shown in Figure 5.3. For the time period of Jan 2nd 2004 to Dec 31st 2004, we found 30 clusters as shown in Figure 5.4. For the time period of Jan 3rd 2005 to Dec 30th 2005, we found 30 clusters as shown in Figure 5.5. Therefore, our results also support Hypothesis 2, which means there should exist similar number of clusters during different time periods in the market equilibrium state.

Further, we can tell that the number of clusters remains similar even though it is changing from time to time according to our results above. The reasons for these clusters structure can be presented as follows. Firstly, it is one of the key features of a complex network. It is dynamic and the participants are interacting with each other continuously within this network. Secondly, the clusters show us the interaction activities inside a network. For a stock market network, those activities can tell us the strength of the connections between any two clusters and the order flows among the stocks in this network. According to LeBaron's (2006) agent-based model in the stock market, the order flows collect information from the market participants' opinions on different stocks and provide feedback to the market by showing the price movements. Under the market equilibrium state, information is available for all the market participants, so the order flows in and out of a large scale of stocks. The order flowing into a stock will cause the stock price to move up and will be shown as a positive return. Otherwise, the order flowing out of a stock will cause a negative return. Only a sufficient amount of both out-flowing and in-flowing orders will provide a reasonable level of heterogeneity in the stock market according to the agent-based model by LeBaron (2006). And only when the information is available to all the market participants will there be adequate order flows in the market, which means the market is in an equilibrium state (Grossman & Stiglitz, 1980). Therefore, the clusters from our results provide an insight to the order flows

interactions and connections in the US stock market under the market equilibrium state. And we will compare the number of clusters in both market equilibrium state and market disequilibrium state in the next section, 5.3.2.

5.3.2 Market Disequilibrium State. Figure 5.6 and Figure 5.8 have provided us with the number of clusters and a visual dot plot of the stock market under the market disequilibrium state. Comparing this result with the results from the market equilibrium state, we can see that the number of clusters has changed from 27 clusters to 1 large cluster by the clustering result of the PHA method. In addition, the Dot Matrix Plot result for the market disequilibrium state has showed us the visual clusters structure, which also corresponds to the PHA clustering result. These results together support our Hypothesis 3 in which we think clusters from the market equilibrium state should converge into one or two large clusters during the pre-crash time period in the market disequilibrium state. According to the Efficient Market Hypothesis, the market should be in an equilibrium state while information is available to everyone (Fama, 1970). However, Grossman and Stiglitz (1980) have uncovered the truth that information could become unavailable to most of the participants in the market and become costly. Once valuable information becomes unavailable to some participants, the market will not fully follow the Efficient Market Rules and will be under a disequilibrium state. Grossman and Stiglitz (1980) have demonstrated the information-spreading process from informed investors to uninformed

investors by building a mock stock market system, and this process could be considered to explain the 'Herding Behavior' in the stock market. Once information becomes expensive, uninformed investors would start to observe and imitate informed investors' behaviors, and this would lead to homogeneous trading orders in the market (Grossman & Stiglitz, 1980).

In a market equilibrium state, once the uninformed traders start to make decisions by observing and imitating informed investors' behaviors, both informed and uninformed traders would affect each other and move the stock price regardless of rationality (LeBaron, 2001). The influence from informed investors on uninformed investors would cause the imitating behavior or herding behavior, similar to the effect of asymmetric information in a market disequilibrium state. According to LeBaron's (2006) agent-based model, the order flows capture the information spreading process. Once the market is not efficient, information becomes costly and investors start to mimic other investors' behaviors and order flows will synchronize to only on one side, buying or selling, instead of the heterogeneous trading decisions. And the stock market participants converge to one dominant rule in a market disequilibrium state (LeBaron, 2006).

This result could also be explained by another key feature of a complex network, synchronization. Synchronization describes the phenomenon that adding some small new information to a network can cause the network to significantly oscillate into a similar

movement (Watts & Strogatz, 1998). In natural networks, the normal state for the participants is to stay disordered and balanced. However, once new information is added to a network, it will start to synchronize and become ordered because of the interaction among all the participants, such as the ants' homogeneous reaction to the signal of upcoming rain. In a stock market situation, informed investors bring this new information to the stock market network and cause the network participants to move homogeneously just like other networks. In a stock market network, the synchronization shows similar trading behaviors or herding behaviors and similar order flows in the market. This synchronization process will ultimately lead up to a market crash if there is no market regulation or interference. We can take the market pre-crash building-up stage as an apparent market disequilibrium state. What's more, all these above findings have explained the pre-crash building-up process to some extent, especially in the area of network theory.

6. Contribution, Limitation, and Conclusion

6.1 Contribution

Based on the PHA clustering results and the Dot Matrix Plot results under the market equilibrium state and the market disequilibrium state, we have observed actual clusters' movement and clusters' converging from the market equilibrium state to the disequilibrium state. We have analyzed the whole process building up to the 2008 – 2009 market crash by applying the network computation to the US S&P 500 index network. And we have identified the significant change in the number of clusters from an equilibrium market to a pre-crash disequilibrium state. In this study, we propose an indicator of the bubble building-up state in the stock market. Imitation and herding behaviors can be detected as synchronization of a stock market complex network, which leads to fewer but larger clusters. We believe that we are the first to introduce this precursor of the stock market crashes to detect the bubble building-up state.

Therefore, we believe that we have uncovered some insight into the study of the stock market crashes and we have also contributed towards retaining the stock market equilibrium state by providing a detecting tool for market regulators and investors.

In the global stock market history, there have been two major crashes in the recent two centuries: the 1929 Wall Street Crash and the 2008-2009 Crash. Both of these two market crashes have caused billions in losses and have damaged investors' confidence

significantly afterwards. Market regulators could have been blamed for the lack of proper interference and regulation in the market. In order to help maintain the market equilibrium state, it is necessary to identify the disequilibrium state or the bubble building-up stage. Our study provides a new indicator to detect the bubble or pre-crash building-up state by computing the number of clusters. Therefore, we would like to provide another possible indicator to detect a pre-crash disequilibrium state in the market. Once we observe a convergence of clusters in the market, there could be a bubble building up in the market. Under this circumstance, we suggest the market regulators should interfere in the market and correct the possible overpricing with transparent free information in the market. According to our theory, we believe that information is not available to all the market participants during the market equilibrium state. So, we also think that market regulators should try to increase the information transparency by releasing some valuable information to the market. Taking interest rate for an example, market regulators should increase the interest rate once a pre-crash market disequilibrium state has been identified in order to calm the market down and bring the market back to equilibrium state instead of supporting the bubble to continue building up. In our view, it is essential to be able to detect the crash before it bursts, because after the crash is simply too late: everyone will try to sell but no investors will be able to sell their holdings.

However, investors should also have felt responsible for the market crashes. So,

from investors' standpoint, we believe our cluster-detecting process should also be helpful for investors when they are making decisions. For investors, it is necessary to always keep rational and make decisions based on as much as information instead of imitating or following other investors' behaviors. Once we observe a convergence of clusters in the market, there could be a bubble building up in the market. Then, investors should keep calm and try not to follow other investors. Under this circumstance, it is vital to keep applying the fundamental analysis and collect information as much as possible in order to make rational and long-term investment decisions. In this way, there will still be order flows in and out of the stocks; this will help in slowing down or even stopping the pre-crash building-up stage and will bring the market back to equilibrium state and reduce the volatility in the market. Therefore, in summary, our main contribution is to provide another indicator to be able to detect the crash before it actually happens.

6.2 Limitation

Even though we have identified clusters' movement under different market states for the 2008 – 2009 market crash, it is important to test our model for other market crashes and provide more evidence to support our model. Therefore, our model and method are still not mature enough in their current stage. The results we have in this study can only explain the bubble building-up process for the 2008 – 2009 market crash and can only capture the clusters movement within the S&P 500 index constituents. We

hope there will be further studies to test other market crashes and further evidence to support our study. Other than that, there have also been quite different crashes in the stock market history. There are minor market crashes, major crashes, long-term and slow crashes, and short-term and fast crashes. In our study, we have only studied the 2008-2009 crash that is considered to be a fast and major market crash. So, we still need to test other crashes and it is possible to adjust our method in order to fit other crashes. Besides of these study design limitations above, we also think that it is vital to question the difference between identifying the clusters movement for post crashes and forecasting a pre-crash building-up state. But we hope our study has shed some light on research in identifying clusters' patterns in the stock market network and could be useful for market regulators, stock investors, and any other market participants. Therefore, we believe it is also worthwhile for other researchers to look into our limitations and provide further studies in the near future.

6.3 Conclusion

In the recent two centuries, two major crashes have occurred in the global stock market: the 1929 Wall Street Crash and the 2008-2009 Crash. Both of these two market crashes have caused billions of losses and have damaged investors' confidence significantly afterwards. Researchers from multiple disciplines have tried to understand the mechanism of stock market crashes. Independent empirical evidence has converged to

prove the synchronization phenomenon as the trigger of stock market crashes (Tse et al., 2010). As well, the Phase Transition Model explains the building-up mechanism and the critical point theory in stock market crashes (Yalamova & McKelvey, 2011). In this study, we have proposed a possible method to identify the synchronization or pre-crash building-up stage before the 2008 – 2009 crash by applying the network theory and methods.

We have introduced a novel hierarchical clustering method called the Potential-based Hierarchical Agglomerative (PHA) clustering method used in biology and physics and applied this method to the US stock market network (Lu & Wan, 2013). In addition, we have applied another novel significant correlation matrix extraction method in order to build up the significant visual display of the clusters. What's more, we have adopted a Dot Matrix Plot method that is mainly used in bioinformatics to show a graphical display of the clusters under different market states. By applying these methods to our data set, we found the results that support our hypotheses. We found 27 clusters during the time period from the beginning of 2002 to the end of 2005 that is considered to be within the market equilibrium state. Our findings correspond to Leibon et al.'s (2008) study on the topological structure and the existence of sector clusters in the US stock market. As well, our results also further support Foti et al.'s (2011) conclusion that there have been 22 clusters within the S&P 500 index constituents. Furthermore, we have also

identified that there exists the similar number of clusters during different time periods under the market equilibrium state. We found 28 clusters, 26 clusters, 30 clusters, and 30 clusters during the time periods of Jan 2nd 2002 to Dec 31st 2002, Jan 2nd 2003 to Dec 31st 2003, Jan 2nd 2004 to Dec 31st 2004, and Jan 3rd 2005 to Dec 30th 2005 respectively. More importantly, we have also identified the clusters' convergence into only one large cluster in the market disequilibrium state, which further supports our hypotheses.

Based on the results we have found, we also provide some suggestions to market regulators, investors, and other market participants. We believe that our study has provided another indicator to detect the pre-crash building-up stage. Market regulators should interfere in the market once the pre-crash signal has been detected by possible ways such as interest rate adjustment or information transparency regulation actions. Additionally, stock market investors should remain rational and collect as much as possible information before making decisions, especially when a pre-crash disequilibrium state is identified.

However, our study is only based on the US stock market S&P 500 index constituents and is only built on the 2008 – 2009 market crash, which leads to the study's design limitations. Besides this, we also acknowledge ourselves that there is a difference between identifying the clusters' movement for post crashes and forecasting a pre-crash building-up state. But we do hope our study could shed some light on research in

identifying clusters' pattern in the stock market network. Therefore, we believe it is also worthwhile for other researchers to look into our limitations and provide further studies in the near future. And we are confident that future studies could be much more useful for market regulators, stock investors, and any other market participants in order to maintain the market equilibrium state and reduce market volatility.

References

- Albert, R., & Barabási, A.-L. (2000). Topology of Evolving Networks: Local Events and Universality. *Physical Review Letters*, 85(24), 5234-5237.
- Amaral, L. A. N., & Ottino, J. M. (2004). Complex networks. *The European Physical Journal B - Condensed Matter and Complex Systems*, 38(2), 147-162.
- Arenas, A., Díaz-Guilera, A., Kurths, J., Moreno, Y., & Zhou, C. (2008). Synchronization in Complex Networks. *Physics Reports*, 469(3), 93-153.
- Baker, M., & Wurgler, J. (2007). Investor Sentiment in the Stock Market. *National Bureau of Economic Research Working Paper Series, No. 13189*.
- Barabási, A.-L., & Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, 286(5439), 509-512.
- Barahona, M., & Pecora, L. M. (2002). Synchronization in Small-World Systems. *Physical Review Letters*, 89(5), 054101.
- Brú, A., Alós, E., Nuño, J. C., & de Dios, M. F. (2014). Scaling in Complex Systems: A Link between the Dynamics of Networks and Growing Interfaces. *Sci. Rep.*, 4.
- Cecchetti, S. G. (1997). Understanding the Great Depression: Lessons for Current Policy. *National Bureau of Economic Research Working Paper Series, No. 6015*.
- Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2), 383-417.
- Farmer, J. D., Gillemot, L., Lillo, F., Mike, S., & Sen, A. (2004). What Really Causes Large Price Changes? *Quantitative Finance*, 4(4), 383-397.
- Foti, N. J., Hughes, J. M., & Rockmore, D. N. (2011). Nonparametric Sparsification of Complex Multiscale Networks. *PLoS ONE*, 6(2), e16431.
- Gómez-Gardeñes, J., Moreno, Y., & Arenas, A. (2011). Evolution of Microscopic and Mesoscopic Synchronized Patterns in Complex Networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 21(1).

- Grossman, S. J., & Stiglitz, J. E. (1980). On the Impossibility of Informationally Efficient Markets. *The American Economic Review*, 70(3), 393-408.
- Guimerà, R., Danon, L., Díaz-Guilera, A., Giralt, F., & Arenas, A. (2003). Self-similar Community Structure in a Network of Human Interactions. *Physical Review E*, 68(6), 065103.
- Hong, H., Kim, B. J., Choi, M. Y., & Park, H. (2004). Factors that Predict Better Synchronizability on Complex Networks. *Physical Review E*, 69(6), 067105.
- Kim, H., Kim, I., Lee, Y., & Kahng, B. (2002). Scale-Free Network in Stock Markets. *Journal-Korean Physical Society*, 40, 1105-1108.
- Krause, A. (2004). Predicting Crashes in a Model of Evolving Networks. *Complexity*, 9(4), 24-30.
- LeBaron, B. (2001). Evolution and Time Horizons in an Agent-Based Stock Market. *Macroeconomic Dynamics*, 5(02), 225-254.
- LeBaron, B. (2006). Agent-based Computational Finance. In L. Tesfatsion & K. L. Judd (Eds.), *Handbook of Computational Economics* (Vol. Volume 2, pp. 1187-1233): Elsevier.
- Leibon, G., Pauls, S., Rockmore, D., & Savell, R. (2008). Topological Structures in the Equities Market Network. *Proceedings of the National Academy of Sciences*, 105(52), 20589-20594.
- Lu, Y., & Wan, Y. (2013). PHA: A Fast Potential-based Hierarchical Agglomerative Clustering Method. *Pattern Recognition*, 46(5), 1227-1239.
- Mantegna, R. N. (1999). Hierarchical Structure in Financial Markets. *The European Physical Journal B - Condensed Matter and Complex Systems*, 11(1), 193-197.
- Newman, M. E. J., & Girvan, M. (2004). Finding and Evaluating Community Structure in Networks. *Physical Review E*, 69(2), 026113.
- Nishikawa, T., Motter, A. E., Lai, Y.-C., & Hoppensteadt, F. C. (2003). Heterogeneity in Oscillator Networks: Are Smaller Worlds Easier to Synchronize? *Physical Review Letters*, 91(1), 014101.

- Omran, M. G. H., Engelbrecht, A. P., & Salman, A. (2007). An Overview of Clustering Methods. *Intell. Data Anal.*, 11(6), 583-605.
- Onnela, J. P., Saramäki, J., Kaski, K., & Kertész, J. (2006). Financial Market - A Network Perspective. *Practical Fruits of Econophysics* (pp. 302-306): Springer.
- Prigogine, I. (1980). *From Being to Becoming: Time and Complexity in the Physical Sciences* (Vol. 51): Monograph Collection (Matt - Pseudo).
- Ravasz, E., & Barabási, A.-L. (2003). Hierarchical Organization in Complex Networks. *Physical Review E*, 67(2), 026112.
- Singh, P., Sreenivasan, S., Szymanski, B. K., & Korniss, G. (2013). Threshold-limited Spreading in Social Networks with Multiple Initiators. *Scientific Reports*, 3.
- Sornette, D. (2006). *Critical Phenomena in Natural Sciences: Chaos, Fractals, Selforganization and Disorder: Concepts and Tools* (Springer Series in Synergetics): Springer.
- Sornette, D. (2009). *Why Stock Markets Crash: Critical Events in Complex Financial Systems*: Princeton University Press.
- Sornette, D., Johansen, A., & Bouchaud, J.-P. (1996). Stock Market Crashes, Precursors and Replicas. *Journal de Physique I*, 6(1), 167-175.
- Tse, C. K., Liu, J., & Lau, F. C. M. (2010). A Network Perspective of the Stock Market. *Journal of Empirical Finance*, 17(4), 659-667.
- Vandewalle, N., Brisbois, F., & Tordoir, X. (2001). Non-random Topology of Stock Markets. *Quantitative Finance*, 1(3), 372-374.
- Watts, D. J., & Strogatz, S. H. (1998). Collective Dynamics of 'Small-world' Networks. *Nature*, 393(6684), 440-442.
- White, E. N. (1990). The Stock Market Boom and Crash of 1929 Revisited. *The Journal of Economic Perspectives*, 4(2), 67-83.
- Yalamova, R., & McKelvey, B. (2011). Explaining What Leads Up to Stock Market

Crashes: A Phase Transition Model and Scalability Dynamics. *Journal of Behavioral Finance*, 12(3), 169-182.

Zouaoui, M., Nouyrigat, G., & Beer, F. (2011). How Does Investor Sentiment Affect Stock Market Crises? Evidence from Panel Data. *Financial Review*, 46(4), 723-747.

Appendix

S&P 500 Constituents List from FactSet.

Name	Symbol
S&P 500	SP50-SPX
10 Energy	
101010 Energy Equipment & Services	
Baker Hughes Incorporated	BHI-US
Cameron International Corporation	CAM-US
Diamond Offshore Drilling, Inc.	DO-US
EnSCO plc	ESV-US
FMC Technologies, Inc.	FTI-US
Halliburton Company	HAL-US
Helmerich & Payne, Inc.	HP-US
National Oilwell Varco, Inc.	NOV-US
Noble Corporation plc	NE-US
Schlumberger NV	SLB-US
Transocean Ltd.	RIG-US
101020 Oil Gas & Consumable Fuels	
Anadarko Petroleum Corporation	APC-US
Apache Corporation	APA-US
Cabot Oil & Gas Corporation	COG-US
Chesapeake Energy Corporation	CHK-US
Chevron Corporation	CVX-US
Cimarex Energy Co.	XEC-US
ConocoPhillips	COP-US
CONSOL Energy Inc.	CNX-US
Devon Energy Corporation	DVN-US
EOG Resources, Inc.	EOG-US
EQT Corporation	EQT-US
Exxon Mobil Corporation	XOM-US
Hess Corporation	HES-US
Kinder Morgan Inc. Class P	KMI-US
Marathon Oil Corporation	MRO-US
Marathon Petroleum Corporation	MPC-US
Murphy Oil Corporation	MUR-US
Newfield Exploration Company	NFX-US
Noble Energy, Inc.	NBL-US

Occidental Petroleum Corporation	OXY-US
ONEOK, Inc.	OKE-US
Phillips 66	PSX-US
Pioneer Natural Resources Company	PXD-US
QEP Resources, Inc.	QEP-US
Range Resources Corporation	RRC-US
Southwestern Energy Company	SWN-US
Spectra Energy Corp	SE-US
Tesoro Corporation	TSO-US
Valero Energy Corporation	VLO-US
Williams Companies, Inc.	WMB-US
15 Materials	
151010 Chemicals	
Air Products and Chemicals, Inc.	APD-US
Airgas, Inc.	ARG-US
CF Industries Holdings, Inc.	CF-US
Dow Chemical Company	DOW-US
E. I. du Pont de Nemours and Company	DD-US
Eastman Chemical Company	EMN-US
Ecolab Inc.	ECL-US
FMC Corporation	FMC-US
International Flavors & Fragrances Inc.	IFF-US
LyondellBasell Industries NV	LYB-US
Monsanto Company	MON-US
Mosaic Company	MOS-US
PPG Industries, Inc.	PPG-US
Praxair, Inc.	PX-US
Sherwin-Williams Company	SHW-US
Sigma-Aldrich Corporation	SIAL-US
151020 Construction Materials	
Martin Marietta Materials, Inc.	MLM-US
Vulcan Materials Company	VMC-US
151030 Containers & Packaging	
Avery Dennison Corporation	AVY-US
Ball Corporation	BLL-US
MeadWestvaco Corporation	MWV-US
Owens-Illinois, Inc.	OI-US
Sealed Air Corporation	SEE-US

151040 Metals & Mining	
Alcoa Inc.	AA-US
Allegheny Technologies Incorporated	ATI-US
Freeport-McMoRan, Inc.	FCX-US
Newmont Mining Corporation	NEM-US
Nucor Corporation	NUE-US
151050 Paper & Forest Products	
International Paper Company	IP-US
20 Industrials	
201010 Aerospace & Defense	
Boeing Company	BA-US
General Dynamics Corporation	GD-US
Honeywell International Inc.	HON-US
L-3 Communications Holdings, Inc.	LLL-US
Lockheed Martin Corporation	LMT-US
Northrop Grumman Corporation	NOC-US
Precision Castparts Corp.	PCP-US
Raytheon Company	RTN-US
Rockwell Collins, Inc.	COL-US
Textron Inc.	TXT-US
United Technologies Corporation	UTX-US
201020 Building Products	
Allegion PLC	ALLE-US
Masco Corporation	MAS-US
201030 Construction & Engineering	
Fluor Corporation	FLR-US
Jacobs Engineering Group Inc.	JEC-US
Quanta Services, Inc.	PWR-US
201040 Electrical Equipment	
AMETEK, Inc.	AME-US
Eaton Corp. Plc	ETN-US
Emerson Electric Co.	EMR-US
Rockwell Automation, Inc.	ROK-US
201050 Industrial Conglomerates	
3M Company	MMM-US
Danaher Corporation	DHR-US
General Electric Company	GE-US
Roper Industries, Inc.	ROP-US

201060 Machinery	
Caterpillar Inc.	CAT-US
Cummins Inc.	CMI-US
Deere & Company	DE-US
Dover Corporation	DOV-US
Flowserve Corporation	FLS-US
Illinois Tool Works Inc.	ITW-US
Ingersoll-Rand Plc	IR-US
Joy Global Inc.	JOY-US
PACCAR Inc.	PCAR-US
Pall Corporation	PLL-US
Parker-Hannifin Corporation	PH-US
Pentair plc	PNR-US
Snap-on Incorporated	SNA-US
Stanley Black & Decker, Inc.	SWK-US
Xylem Inc.	XYL-US
201070 Trading Companies & Distributors	
Fastenal Company	FAST-US
United Rentals, Inc.	URI-US
W.W. Grainger, Inc.	GWV-US
202010 Commercial Services & Supplies	
ADT Corporation	ADT-US
Cintas Corporation	CTAS-US
Pitney Bowes Inc.	PBI-US
Republic Services, Inc.	RSG-US
Stericycle, Inc.	SRCL-US
Tyco International PLC	TYC-US
Waste Management, Inc.	WM-US
202020 Professional Services	
Dun & Bradstreet Corporation	DNB-US
Equifax Inc.	EFX-US
Nielsen N.V.	NLSN-US
Robert Half International Inc.	RHI-US
203010 Air Freight & Logistics	
C.H. Robinson Worldwide, Inc.	CHRW-US
Expeditors International of Washington, Inc.	EXPD-US
FedEx Corporation	FDX-US
United Parcel Service, Inc. Class B	UPS-US

203020 Airlines	
American Airlines Group, Inc.	AAL-US
Delta Air Lines, Inc.	DAL-US
Southwest Airlines Co.	LUV-US
203040 Road & Rail	
CSX Corporation	CSX-US
Kansas City Southern	KSU-US
Norfolk Southern Corporation	NSC-US
Ryder System, Inc.	R-US
Union Pacific Corporation	UNP-US
25 Consumer Discretionary	
251010 Auto Components	
BorgWarner Inc.	BWA-US
Delphi Automotive PLC	DLPH-US
Goodyear Tire & Rubber Company	GT-US
Johnson Controls, Inc.	JCI-US
251020 Automobiles	
Ford Motor Company	F-US
General Motors Company	GM-US
Harley-Davidson, Inc.	HOG-US
252010 Household Durables	
D.R. Horton, Inc.	DHI-US
Garmin Ltd.	GRMN-US
Harman International Industries, Incorporated	HAR-US
Leggett & Platt, Incorporated	LEG-US
Lennar Corporation Class A	LEN-US
Mohawk Industries, Inc.	MHK-US
Newell Rubbermaid Inc.	NWL-US
PulteGroup, Inc.	PHM-US
Whirlpool Corporation	WHR-US
252020 Leisure Products	
Hasbro, Inc.	HAS-US
Mattel, Inc.	MAT-US
252030 Textiles Apparel & Luxury Goods	
Coach, Inc.	COH-US
Fossil Group, Inc.	FOSL-US
Hanesbrands Inc.	HBI-US
Michael Kors Holdings Ltd	KORS-US

NIKE, Inc. Class B	NKE-US
PVH Corp.	PVH-US
Ralph Lauren Corporation Class A	RL-US
Under Armour, Inc. Class A	UA-US
V.F. Corporation	VFC-US
253010 Hotels Restaurants & Leisure	
Carnival Corporation	CCL-US
Chipotle Mexican Grill, Inc.	CMG-US
Darden Restaurants, Inc.	DRI-US
Marriott International, Inc. Class A	MAR-US
McDonald's Corporation	MCD-US
Royal Caribbean Cruises Ltd.	RCL-US
Starbucks Corporation	SBUX-US
Starwood Hotels & Resorts Worldwide, Inc.	HOT-US
Wyndham Worldwide Corporation	WYN-US
Wynn Resorts, Limited	WYNN-US
YUM! Brands, Inc.	YUM-US
253020 Diversified Consumer Services	
H&R Block, Inc.	HRB-US
254010 Media	
Cablevision Systems Corporation Class A	CVC-US
CBS Corporation Class B	CBS-US
Comcast Corporation Class A	CMCSA-US
DIRECTV	DTV-US
Discovery Communications, Inc. Class A	DISCA-US
Discovery Communications, Inc. Class C	DISCK-US
Gannett Co., Inc.	GCI-US
Interpublic Group of Companies, Inc.	IPG-US
News Corporation Class A	NWSA-US
Omnicom Group Inc.	OMC-US
Scripps Networks Interactive, Inc. Class A	SNI-US
Time Warner Cable Inc.	TWC-US
Time Warner Inc.	TWX-US
Twenty-First Century Fox, Inc. Class A	FOXA-US
Viacom Inc. Class B	VIAB-US
Walt Disney Company	DIS-US
255010 Distributors	
Genuine Parts Company	GPC-US

255020 Internet & Catalog Retail	
Amazon.com, Inc.	AMZN-US
Expedia, Inc.	EXPE-US
Netflix, Inc.	NFLX-US
Priceline Group Inc.	PCLN-US
TripAdvisor, Inc.	TRIP-US
255030 Multiline Retail	
Dollar General Corporation	DG-US
Dollar Tree, Inc.	DLTR-US
Family Dollar Stores, Inc.	FDO-US
Kohl's Corporation	KSS-US
Macy's Inc.	M-US
Nordstrom, Inc.	JWN-US
Target Corporation	TGT-US
255040 Specialty Retail	
AutoNation, Inc.	AN-US
AutoZone, Inc.	AZO-US
Bed Bath & Beyond Inc.	BBBY-US
Best Buy Co., Inc.	BBY-US
CarMax, Inc.	KMX-US
GameStop Corp. Class A	GME-US
Gap, Inc.	GPS-US
Home Depot, Inc.	HD-US
L Brands, Inc.	LB-US
Lowe's Companies, Inc.	LOW-US
O'Reilly Automotive, Inc.	ORLY-US
Ross Stores, Inc.	ROST-US
Staples, Inc.	SPLS-US
Tiffany & Co.	TIF-US
TJX Companies, Inc.	TJX-US
Tractor Supply Company	TSCO-US
Urban Outfitters, Inc.	URBN-US
30 Consumer Staples	
301010 Food & Staples Retailing	
Costco Wholesale Corporation	COST-US
CVS Health Corporation	CVS-US
Kroger Co.	KR-US
Sysco Corporation	SYYS-US

Wal-Mart Stores, Inc.	WMT-US
Walgreens Boots Alliance Inc.	WBA-US
Whole Foods Market, Inc.	WFM-US
302010 Beverages	
Brown-Forman Corporation Class B	BF.B-US
Coca-Cola Company	KO-US
Coca-Cola Enterprises, Inc.	CCE-US
Constellation Brands, Inc. Class A	STZ-US
Dr Pepper Snapple Group, Inc.	DPS-US
Molson Coors Brewing Company Class B	TAP-US
Monster Beverage Corporation	MNST-US
PepsiCo, Inc.	PEP-US
302020 Food Products	
Archer-Daniels-Midland Company	ADM-US
Campbell Soup Company	CPB-US
ConAgra Foods, Inc.	CAG-US
General Mills, Inc.	GIS-US
Hershey Company	HSY-US
Hormel Foods Corporation	HRL-US
J. M. Smucker Company	SJM-US
Kellogg Company	K-US
Keurig Green Mountain, Inc.	GMCR-US
Kraft Foods Group, Inc.	KRFT-US
McCormick & Company, Incorporated	MKC-US
Mead Johnson Nutrition Company	MJN-US
Mondelez International, Inc. Class A	MDLZ-US
Tyson Foods, Inc. Class A	TSN-US
302030 Tobacco	
Altria Group, Inc.	MO-US
Lorillard, Inc.	LO-US
Philip Morris International Inc.	PM-US
Reynolds American Inc.	RAI-US
303010 Household Products	
Clorox Company	CLX-US
Colgate-Palmolive Company	CL-US
Kimberly-Clark Corporation	KMB-US
Procter & Gamble Company	PG-US
303020 Personal Products	

Estee Lauder Companies Inc. Class A	EL-US
35 Health Care	
351010 Health Care Equipment & Supplies	
Abbott Laboratories	ABT-US
Baxter International Inc.	BAX-US
Becton, Dickinson and Company	BDX-US
Boston Scientific Corporation	BSX-US
C. R. Bard, Inc.	BCR-US
DENTSPLY International Inc.	XRAY-US
Edwards Lifesciences Corporation	EW-US
Intuitive Surgical, Inc.	ISRG-US
Medtronic Plc	MDT-US
St. Jude Medical, Inc.	STJ-US
Stryker Corporation	SYK-US
Varian Medical Systems, Inc.	VAR-US
Zimmer Holdings, Inc.	ZMH-US
351020 Health Care Providers & Services	
Aetna Inc.	AET-US
AmerisourceBergen Corporation	ABC-US
Anthem, Inc.	ANTM-US
Cardinal Health, Inc.	CAH-US
Cigna Corporation	CI-US
DaVita HealthCare Partners Inc.	DVA-US
Express Scripts Holding Company	ESRX-US
HCA Holdings, Inc.	HCA-US
Henry Schein, Inc.	HSIC-US
Humana Inc.	HUM-US
Laboratory Corporation of America Holdings	LH-US
McKesson Corporation	MCK-US
Patterson Companies, Inc.	PDCO-US
Quest Diagnostics Incorporated	DGX-US
Tenet Healthcare Corporation	THC-US
UnitedHealth Group Incorporated	UNH-US
Universal Health Services, Inc. Class B	UHS-US
351030 Health Care Technology	
Cerner Corporation	CERN-US
352010 Biotechnology	
Alexion Pharmaceuticals, Inc.	ALXN-US

Amgen Inc.	AMGN-US
Biogen Inc.	BIIB-US
Celgene Corporation	CELG-US
Gilead Sciences, Inc.	GILD-US
Regeneron Pharmaceuticals, Inc.	REGN-US
Vertex Pharmaceuticals Incorporated	VRTX-US
352020 Pharmaceuticals	
AbbVie, Inc.	ABBV-US
Actavis Plc	ACT-US
Bristol-Myers Squibb Company	BMY-US
Eli Lilly and Company	LLY-US
Endo International Plc	ENDP-US
Hospira, Inc.	HSP-US
Johnson & Johnson	JNJ-US
Mallinckrodt Plc	MNK-US
Merck & Co., Inc.	MRK-US
Mylan N.V.	MYL-US
Perrigo Co. Plc	PRGO-US
Pfizer Inc.	PFE-US
Zoetis, Inc. Class A	ZTS-US
352030 Life Sciences Tools & Services	
Agilent Technologies, Inc.	A-US
PerkinElmer, Inc.	PKI-US
Thermo Fisher Scientific Inc.	TMO-US
Waters Corporation	WAT-US
40 Financials	
401010 Banks	
Bank of America Corporation	BAC-US
BB&T Corporation	BBT-US
Citigroup Inc.	C-US
Comerica Incorporated	CMA-US
Fifth Third Bancorp	FITB-US
Huntington Bancshares Incorporated	HBAN-US
JPMorgan Chase & Co.	JPM-US
KeyCorp	KEY-US
M&T Bank Corporation	MTB-US
PNC Financial Services Group, Inc.	PNC-US
Regions Financial Corporation	RF-US

SunTrust Banks, Inc.	STI-US
U.S. Bancorp	USB-US
Wells Fargo & Company	WFC-US
Zions Bancorporation	ZION-US
401020 Thrifts & Mortgage Finance	
Hudson City Bancorp, Inc.	HCBK-US
People's United Financial, Inc.	PBCT-US
402010 Diversified Financial Services	
Berkshire Hathaway Inc. Class B	BRK.B-US
CME Group Inc. Class A	CME-US
Intercontinental Exchange, Inc.	ICE-US
Leucadia National Corporation	LUK-US
McGraw Hill Financial, Inc.	MHFI-US
Moody's Corporation	MCO-US
NASDAQ OMX Group, Inc.	NDAQ-US
402020 Consumer Finance	
American Express Company	AXP-US
Capital One Financial Corporation	COF-US
Discover Financial Services	DFS-US
Navient Corp	NAVI-US
402030 Capital Markets	
Affiliated Managers Group, Inc.	AMG-US
Ameriprise Financial, Inc.	AMP-US
Bank of New York Mellon Corporation	BK-US
BlackRock, Inc.	BLK-US
Charles Schwab Corporation	SCHW-US
E*TRADE Financial Corporation	ETFC-US
Franklin Resources, Inc.	BEN-US
Goldman Sachs Group, Inc.	GS-US
Invesco Ltd.	IVZ-US
Legg Mason, Inc.	LM-US
Morgan Stanley	MS-US
Northern Trust Corporation	NTRS-US
State Street Corporation	STT-US
T. Rowe Price Group	TROW-US
403010 Insurance	
ACE Limited	ACE-US
Aflac Incorporated	AFL-US

Allstate Corporation	ALL-US
American International Group, Inc.	AIG-US
Aon plc	AON-US
Assurant, Inc.	AIZ-US
Chubb Corporation	CB-US
Cincinnati Financial Corporation	CINF-US
Genworth Financial, Inc. Class A	GNW-US
Hartford Financial Services Group, Inc.	HIG-US
Lincoln National Corporation	LNC-US
Loews Corporation	L-US
Marsh & McLennan Companies, Inc.	MMC-US
MetLife, Inc.	MET-US
Principal Financial Group, Inc.	PFG-US
Progressive Corporation	PGR-US
Prudential Financial, Inc.	PRU-US
Torchmark Corporation	TMK-US
Travelers Companies, Inc.	TRV-US
Unum Group	UNM-US
XL Group Plc	XL-US
404020 Real Estate Investment Trusts (REITs)	
American Tower Corporation	AMT-US
Apartment Investment and Management Company Class A	AIV-US
AvalonBay Communities, Inc.	AVB-US
Boston Properties, Inc.	BXP-US
Crown Castle International Corp	CCI-US
Equity Residential	EQR-US
Essex Property Trust, Inc.	ESS-US
General Growth Properties, Inc.	GGP-US
HCP, Inc.	HCP-US
Health Care REIT, Inc.	HCN-US
Host Hotels & Resorts, Inc.	HST-US
Iron Mountain, Inc.	IRM-US
Kimco Realty Corporation	KIM-US
Macerich Company	MAC-US
Plum Creek Timber Company, Inc.	PCL-US
Prologis, Inc.	PLD-US
Public Storage	PSA-US
Realty Income Corporation	O-US

Simon Property Group, Inc.	SPG-US
SL Green Realty Corp.	SLG-US
Ventas, Inc.	VTR-US
Vornado Realty Trust	VNO-US
Weyerhaeuser Company	WY-US
404030 Real Estate Management & Development	
CBRE Group, Inc. Class A	CBG-US
45 Information Technology	
451010 Internet Software & Services	
Akamai Technologies, Inc.	AKAM-US
eBay Inc.	EBAY-US
Equinix Inc.	EQIX-US
Facebook, Inc. Class A	FB-US
Google Inc. Class A	GOOGL-US
Google Inc. Class C	GOOG-US
VeriSign, Inc.	VRSN-US
Yahoo! Inc.	YHOO-US
451020 IT Services	
Accenture Plc	ACN-US
Alliance Data Systems Corporation	ADS-US
Automatic Data Processing, Inc.	ADP-US
Cognizant Technology Solutions Corporation Class A	CTSH-US
Computer Sciences Corporation	CSC-US
Fidelity National Information Services, Inc.	FIS-US
Fiserv, Inc.	FISV-US
International Business Machines Corporation	IBM-US
MasterCard Incorporated Class A	MA-US
Paychex, Inc.	PAYX-US
Teradata Corporation	TDC-US
Total System Services, Inc.	TSS-US
Visa Inc. Class A	V-US
Western Union Company	WU-US
Xerox Corporation	XRX-US
451030 Software	
Adobe Systems Incorporated	ADBE-US
Autodesk, Inc.	ADSK-US
CA, Inc.	CA-US
Citrix Systems, Inc.	CTXS-US

Electronic Arts Inc.	EA-US
Intuit Inc.	INTU-US
Microsoft Corporation	MSFT-US
Oracle Corporation	ORCL-US
Red Hat, Inc.	RHT-US
salesforce.com, Inc.	CRM-US
Symantec Corporation	SYMC-US
452010 Communications Equipment	
Cisco Systems, Inc.	CSCO-US
F5 Networks, Inc.	FFIV-US
Harris Corporation	HRS-US
Juniper Networks, Inc.	JNPR-US
Motorola Solutions, Inc.	MSI-US
QUALCOMM Incorporated	QCOM-US
452020 Technology Hardware Storage & Peripherals	
Apple Inc.	AAPL-US
EMC Corporation	EMC-US
Hewlett-Packard Company	HPQ-US
NetApp, Inc.	NTAP-US
SanDisk Corporation	SNDK-US
Seagate Technology PLC	STX-US
Western Digital Corporation	WDC-US
452030 Electronic Equipment Instruments & Components	
Amphenol Corporation Class A	APH-US
Corning Incorporated	GLW-US
FLIR Systems, Inc.	FLIR-US
TE Connectivity Ltd.	TEL-US
453010 Semiconductors & Semiconductor Equipment	
Altera Corporation	ALTR-US
Analog Devices, Inc.	ADI-US
Applied Materials, Inc.	AMAT-US
Avago Technologies Limited	AVGO-US
Broadcom Corporation Class A	BRCM-US
First Solar, Inc.	FSLR-US
Intel Corporation	INTC-US
KLA-Tencor Corporation	KLAC-US
Lam Research Corporation	LRCX-US
Linear Technology Corporation	LLTC-US

Microchip Technology Incorporated	MCHP-US
Micron Technology, Inc.	MU-US
NVIDIA Corporation	NVDA-US
Skyworks Solutions, Inc.	SWKS-US
Texas Instruments Incorporated	TXN-US
Xilinx, Inc.	XLNX-US
50 Telecommunication Services	
501010 Diversified Telecommunication Services	
AT&T Inc.	T-US
CenturyLink, Inc.	CTL-US
Frontier Communications Corporation Class B	FTR-US
Level 3 Communications, Inc.	LVLT-US
Verizon Communications Inc.	VZ-US
55 Utilities	
551010 Electric Utilities	
American Electric Power Company, Inc.	AEP-US
Duke Energy Corporation	DUK-US
Edison International	EIX-US
Entergy Corporation	ETR-US
Eversource Energy	ES-US
Exelon Corporation	EXC-US
FirstEnergy Corp.	FE-US
NextEra Energy, Inc.	NEE-US
Pepco Holdings, Inc.	POM-US
Pinnacle West Capital Corporation	PNW-US
PPL Corporation	PPL-US
Southern Company	SO-US
Xcel Energy Inc.	XEL-US
551020 Gas Utilities	
AGL Resources, Inc.	GAS-US
551030 Multi-Utilities	
Ameren Corporation	AEE-US
CenterPoint Energy, Inc.	CNP-US
CMS Energy Corporation	CMS-US
Consolidated Edison, Inc.	ED-US
Dominion Resources, Inc.	D-US
DTE Energy Company	DTE-US
Integrus Energy Group, Inc.	TEG-US

NiSource Inc.	NI-US
PG&E Corporation	PCG-US
Public Service Enterprise Group Incorporated	PEG-US
SCANA Corporation	SCG-US
Sempra Energy	SRE-US
TECO Energy, Inc.	TE-US
Wisconsin Energy Corporation	WEC-US
551050 Independent Power and Renewable Electricity Producers	
AES Corporation	AES-US
NRG Energy, Inc.	NRG-US