INTERNATIONAL JOURNAL OF ECONOMICS AND STATISTICS
DOI: 10.46300/9103.2022.10.1 Volume 10, 2022

A Complex Network Clustering and Phase Transition Models for Stock Price Dynamics before Crashes

Jiajia Ren, Rossitsa Yalamova
Dhillon School of Business
University of Lethbridge
Canada
rossitsa.yalamova@uleth.ca

Received: May 9, 2021. Revised: November 26, 2021. Accepted: December 17, 2021. Published: January 25, 2022.

Abstract - Researchers from multiple disciplines have tried to understand the mechanism of stock market crashes. Precursory patterns before crashes agree with various empirical studies published by econophysicists, namely the prolific work of Didier Sornette. We intend to add more empirical evidence of synchronization of trading and demonstrate the prospect of predicting stock market crashes by analyzing clusters' dynamics in the period of bubble build-up leading to a crash. We apply the Potential-based Hierarchical Agglomerative (PHA) Method, the Backbone Extraction Method, and the Dot Matrix Plot on S&P500 companies daily returns. Our innovative approach is proposed in this paper, empirical results and discussion presented in another publication.

Key-words – complex networks, clustering, dot matrix, synchronization, crash

I. INTRODUCTION

The 2008 Global Financial Crisis 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

DOI: 10.46300/9103.2022.10.1 Volume 10, 2022

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 in financial markets by analyzing the correlation matrix of stock price 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). 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). As mentioned above, Leibon, Pauls, Rockmore, and Savell (2008) have developed a mathematical computation method to identify clusters of stocks (Leibon et al., 2008). We expect that changes in the clusters' structure will allow us to identify a precursor of the stock market crash.

Grossman and Stiglitz (1980) showed that the stock market could not remain in an equilibrium state when the information becomes costly. When information becomes costly (complex) traders resort to rule based trading and imitation, both informed and uninformed traders would affect each other to move up the stock price regardless of rationality (LeBaron, 2001). Sornette, Johansen, and Bouchaud (1996) studied the time series of the S&P 500 index and found a log-periodic oscillation price pattern from 2 to four years before the 1987 stock market crash. Numerous studies on such pattern followed in Econophysics but results were met with skepticism as data mining lacking theoretical underpinnings.

Yalamova and McKelvey (2011) attempt to bring the Phase Transition Model closer to finance theory, proposing a dynamical system that moves between equilibrium market and bubble build up regime. They suggest the existence of a first critical point (tipping point) when the information complexity brings additional cost and based on Grossman and Stiglitz (1980) assertion market can not find equilibrium, traders herding and imitation creates a bubble build up. This 'tipping point' corresponds to the origination of the log-periodic oscillation pattern detected by econophysicists . Our contribution is aimed at supporting evidence of herding and imitation among traders that will be observed in synchronization of trading. Throughout his numerous publications Sornette equates the stock market to a complex dynamical system in disorder at equilibrium and fully ordered during crash (all traders sell). Yalamova and McKelvey (2011) relate the dynamics of synchronization to herding and imitation that produces agents' homogeneity causing crash, which is the second critical point in the phase transition model.

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 summarizes the expected contributions, possible limitations of our study, and suggests further research directions.

II. LITERATURE REVIEW

A. 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 three major endogenous crashes: the 1929, the Black Monday 1987and 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.

B. 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 theory breaks from Sir Isaac Newton's world outlook. 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:

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

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.

B.1 Stock Market Complex Network

Stock prices reflect investors' valuation of the company incorporating all the information public or private. Historical stock prices are used to analyze returns, volatility, movements in i.i.d fashion . Mantegna (1999) find a hierarchical arrangement of stocks traded in a financial market by investigating the daily time series of the logarithm of stock price. Vandewalle, Brisbois, and Tordoir (2001) studied the topological correlations for neighbouring nodes of the Minimum Spinning Tree and confirmed local structure evolving. Onnela et al. (2006) construct a weighted financial network for a subset of NYSE traded stocks, in which the nodes correspond to stocks and edges to interactions between them. 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.

B.2 Scaling and Synchronization

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 edge absolute magnitude (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.

DOI: 10.46300/9103.2022.10.1 Volume 10, 2022

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.

C. 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, Johansen, & Bouchaud, 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 explaining with 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.

III. HYPOTHESES

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 S&P 500 index constituents. So, in the market equilibrium state, we expect to observe 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

DOI: 10.46300/9103.2022.10.1 Volume 10, 2022

(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.

IV. CONTRIBUTION AND LIMITATION

A. 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 stock market crash dynamics, and we provide an 'destabilizing dynamics' precursor to benefit regulators and market participants.

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 find bubble build-up signs. Our study provides a new indicator to detect the bubble or pre-crash dynamics 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. A convergence of clusters may indicate synchronization of trading strategies, imitation and herding leading to a crash.

B. Limitation

Our empirical results cover only stock price dynamics before 2008 – 2009 market crash, more empirical evidence using different methods is needed to support the trading synchronization theory. We hope there will be further studies to test other market crashes and further evidence to support our study. 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.

V. CONCLUSION

Researchers from multiple disciplines have tried to model price patterns before 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 introducing a novel hierarchical clustering method - 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. 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. 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.

REFERENCE:

[1]. Albert, R., & Barabási, A.-L. (2000). Topology of Evolving Networks: Local Events and Universality. *Physical Review Letters, 85*(24), 5234-5237.

[2]. Amaral, L. A. N., & Ottino, J. M. (2004). Complex networks. *The European Physical Journal B - Condensed Matter and Complex Systems, 38*(2), 147-162. doi: 10.1140/epjb/e2004-00110-5

[3]. Arenas, A., Díaz-Guilera, A., Kurths, J., Moreno, Y., & Zhou, C. (2008). Synchronization in complex networks. *Physics Reports, 469*(3), 93-153.

[4]. Baker, M., & Wurgler, J. (2007). Investor Sentiment in the Stock Market. *National Bureau of Economic Research Working Paper Series, No. 13189*.

[5]. Barabási, A.-L., & Albert, R. (1999). Emergence of Scaling in Random Networks. *Science, 286*(5439), 509-512.

[6]. Barahona, M., & Pecora, L. M. (2002). Synchronization in Small-World Systems. *Physical Review Letters, 89*(5), 054101.

[7]. 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*. doi: 10.1038/srep07550

[8]. Cecchetti, S. G. (1997). Understanding the Great Depression: Lessons for Current Policy. *National Bureau of Economic Research Working Paper Series, No. 6015*. doi: 10.3386/w6015

[9]. Fama, E. F. (1970). EFFICIENT CAPITAL MARKETS: A REVIEW OF THEORY AND EMPIRICAL WORK*. *The Journal of Finance, 25*(2), 383-417. doi: 10.1111/j.1540-6261.1970.tb00518.x

[10]. Farmer, J. D., Gillemot, L., Lillo, F., Mike, S., & Sen, A. (2004). What really causes large price changes? *Quantitative Finance, 4*(4), 383-397. doi: 10.1080/14697680400008627

[11]. Foti, N. J., Hughes, J. M., & Rockmore, D. N. (2011). Nonparametric Sparsification of Complex Multiscale Networks. *PLoS ONE, 6*(2), e16431. doi: 10.1371/journal.pone.0016431

[12]. 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). doi: doi:http://dx.doi.org/10.1063/1.3532801
[13]. Grossman, S. J., & Stiglitz, J. E. (1980). On the Impossibility of Informationally Efficient Markets. *The American Economic Review, 70*(3), 393-408. doi: 10.2307/1805228
[14]. 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.
[15]. 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.
[16]. Kim, H., Kim, I., Lee, Y., & Kahng, B. (2002). Scale-Free Network in Stock Markets. *Journal-Korean Physical Society, 40*, 1105-1108.
[17]. Krause, A. (2004). Predicting crashes in a model of evolving networks. *Complexity, 9*(4), 24-30. doi: 10.1002/cplx.20017
[18]. LeBaron, B. (2001). Evolution and Time Horizons in an Agent-Based Stcok Market. *Macroeconomic Dynamics, 5*(02), 225-254. doi: doi:null
[19]. 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.
[20]. 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. doi: 10.1073/pnas.0802806106
[21]. Lu, Y., & Wan, Y. (2013). PHA: A fast potential-based hierarchical agglomerative clustering method. *Pattern Recognition, 46*(5), 1227-1239. doi: http://dx.doi.org/10.1016/j.patcog.2012.11.017
[22]. Mantegna, R. N. (1999). Hierarchical structure in financial markets. *The European Physical Journal B - Condensed Matter and Complex Systems, 11*(1), 193-197. doi: 10.1007/s100510050929
[23]. Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E, 69*(2), 026113.
[24]. 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.
[25]. Omran, M. G. H., Engelbrecht, A. P., & Salman, A. (2007). An overview of clustering methods. *Intell. Data Anal., 11*(6), 583-605.
[26]. Prigogine, I. (1980). *From Being to Becoming: Time and Complexity in the Physical Sciences* (Vol. 51): Monograph Collection (Matt - Pseudo).
[27]. Ravasz, E., & Barabási, A.-L. (2003). Hierarchical organization in complex networks. *Physical Review E, 67*(2), 026112.
[28]. Singh, P., Sreenivasan, S., Szymanski, B. K., & Korniss, G. (2013). Threshold-limited spreading in social networks with multiple initiators. *Sci. Rep., 3*. doi: 10.1038/srep02330
[29]. http://www.nature.com/srep/2013/130731/srep02330/abs/srep02330.html#supplementary-information
[30]. Sornette, D. (2006). *Critical Phenomena in Natural Sciences: Chaos, Fr*