Potential-based Hierarchical Agglomerative Method in Stock Price Topological Clustering

Rossitsa Yalamova and Jiajia Ren

Dhillon School of Business, University of Lethbridge, Canada

rossitsa.yalamova@uleth.ca

Abstract

Methods from multiple disciplines have provided indicators of price bubble patterns preceding stock market crashes, e.g.log-periodic oscillation, synchronization in trading (herding and imitation), partitioning and clusteringetc. In this study, we propose to add more empirical evidence on trading synchronization dynamics analyzing the Standard & Poor 500 (S&P 500) index companies' daily returnsin the period leading to 2008 market crash. We apply the Potential-based Hierarchical Agglomerative (PHA) Method to extract and displayclusters' dynamics from equilibrium state to a bubble build-up regime leading to a 'tipping point'.

Keywords: *Complex networks, Clusters, Potential-based Hierarchical Agglomerative model*

1. Introduction

In 2008, borrowers' defaults on subprime mortgages precipitated the Global Financial Crisis and worldwide stock markets experienced prolonged drawdown. Patterns before market crashes has been shown in various empirical studies most prolific among many are the log periodic oscillations patterns detected by geophysicists.[1]

Hierarchical arrangement of stocks and the minimum spanning tree of the DJIA stocks was shown in [2]. Physicistsuse concepts from statistical physics in the description of financial systems, e.g., the scaling concepts used in probability theory, critical phenomena, and fully developed turbulent fluids [3]. Cross correlations of daily fluctuations for N= 6358 US stock prices provide information for the minimum spanning tree to be built[4]. A universal model of an evolving complex network predicts crashes by constructing a score function based on the eigenvalue of the correlation matrix [5]. Such findings are consistent with the observations or homogeneous behaviors before financial market crashes. A Partition Decoupling Method (PDM) also is used to display the topological structure in US stocks [6] market. They have also found that the network clusters coincide with industry classifications and represent the capital flows among sectors.In times of financial turmoil, the stock network changes its composition and the disassortative structure of prosperous markets transforms into a more centralized topology [7].A recent new method for constructing the MST-Partial from the correlation matrix of stock prices is presented in [8].

Scale-free networks are characterized by a power-law degree distribution, a topology arising from preferential attachment phenomenon.Complex network synchronization may occuras a self-organizing dynamics[9]. Small perturbation to a complex network can cause synchronized oscillations[10].

Stock marketsmove out of equilibrium when information becomes costly,then imitation, herding and rule-based trading prevails [11].Therefore, one may explain bubble build up withsuch imitation and herding among traders when information becomes complex and costly to analyze [12]. Those imitating behavior among market participants create groups of synchronized trading which gradually merge into bigger groups leading to one synchronized trading cluster causing massive sell off.

In our study, changes in the stock market network clusters' structure arehypothesized as anindicator of the bubble building-up regime caused by imitation and herding among traders. We use the Potential-based Hierarchical Agglomerative clustering method to capture the clusters' structure by building the dendrogram linkage trees [13].

The rest of the study is organized as follows. Section 2 presents the data and methodology utilized in our study. Section 3 develops the hypotheses for the clusters' changes in different market states. Section 4 summarizes the results and (5) discussion of the clusters' formation explaining market dynamics and traders' strategies.

2. Methodology and Data

2.1 Potential-based Hierarchical Agglomerative (PHA) method

Applying this Potential-based Hierarchical Agglomerative (PHA) method, we built the dendrogram linkage trees to find the number of clusters inthe period around 2008 market crash. 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 [13].

Two potential-based similarity metrics, APES and AMAPES, inspired by the concept of electric potential in physics,one can find clusters of complex irregular shapes [14].Objects with imaginary potential such as gravity or electromagnetic field, may show potentials overlap and produce the aggregate effects on the entire potential field. Objects sense the potential and move toward the higher potential directions. The more the objects get together, the higher the potential becomes, and they attract other objects similar to a planet with large gravity attracts asteroids in the space.

The method we apply follows testedandimproved Potential-based Hierarchical Agglomerative (PHA) clustering method [13].To simplify apotential-based clustering method,considering the potential at infinity is zero and the gravitational constantis set to 1;we calculate the corresponding potential at point$i$from point$j$as:

$$\Phi_{ij}(r_{ij}) = \begin{cases} -\frac{1}{r_{ij}} & if\ r_{ij} \geq \delta \\ -\frac{1}{\delta} & if\ r_{ij} < \delta \end{cases} \tag{2.1}$$

Where the edge between node i and node j is r_{ij}, 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. 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.

$$\delta = \frac{mean\ (D_i)}{S} \tag{2.2}$$

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

$$\Phi_i = \sum_{j=1}^{N} \Phi_{ij}(r_{ij}) \tag{2.3}$$

A new similarity metric combining the potential field information and the data set information is constructed. The last step is to extract the clusters based on the edge-weighted tree of the data set by applying another similarity computation method,based on an edge weighted tree of all the data points. This leads to a fast agglomerative clustering algorithm with time complexity.

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 2.1. The potential edges between any two different nodes are marked with the underlined numbers near the "edge" respectively.We find all the potential values marked with numbers in parentheses. 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.

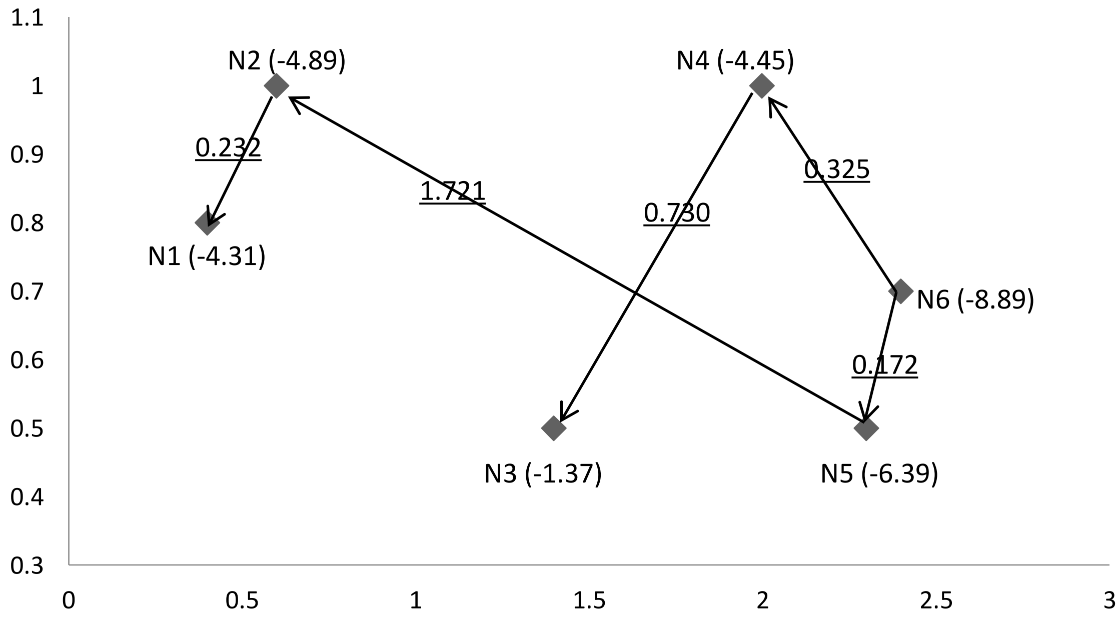

Figure 2.1 Six node data set example

So, summarizing the results above, we can assert 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 is built regarding the merging sequence mentioned and the respective correlation strength in Fig 2.2. In the dendrogram graph, the heights of all the different U-shapes show the relative connection distance between any two small clusters or nodes.

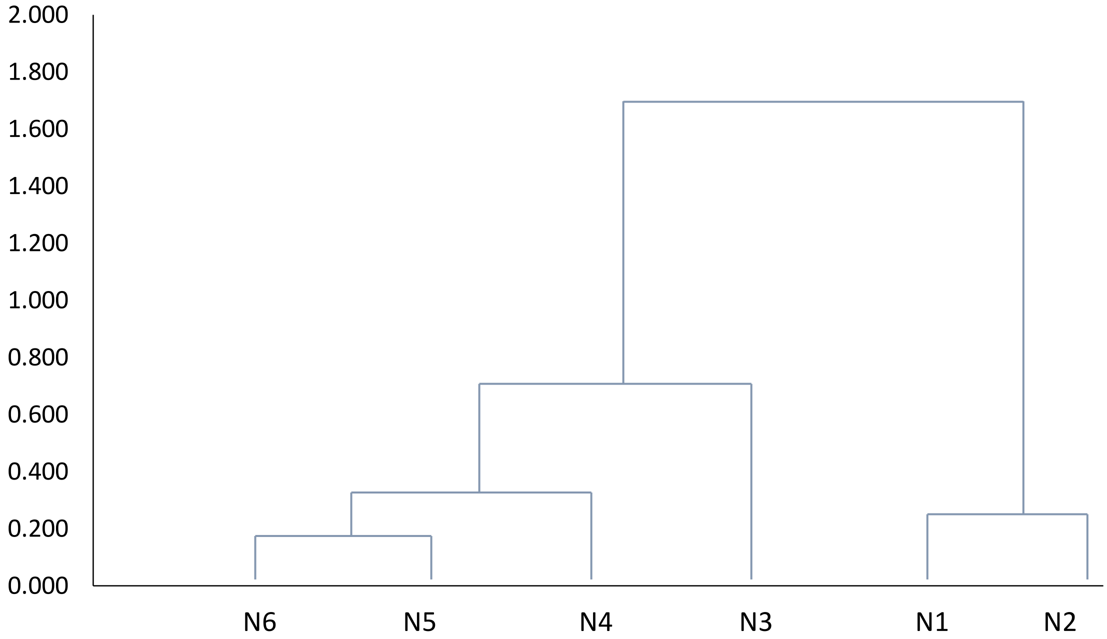

Figure 2.2 Dendrogram based on six node example

2.2 Data

In order to construct the daily return matrix, we extract daily stock pricesfrom the Center for Research in Security Prices (CRSP) database. The S&P 500, or the Standard & Poor's 500, is 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). The S&P 500 index is considered to be the best market benchmark index and captures approximately 80% of available market capitalization.

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 use stock daily returns from Jan 2nd,2002 to Dec 31th, 2010. The formula for daily return is:

$$R_i = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{2.2.1}$$

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 with daily returns from Jan 2nd, 2002 to Dec 31th 2010.

3. Hypothesis Development

We hypothesize that we should observe efficient market in equilibrium and a bubble build up regime with synchronization of trading. Under a market equilibrium state based on the Efficient Market Hypothesis, information arrives to the market in a random fashion, stock prices incorporate all available information in the market. Clusters should reveal the well-known pattern of capital flow called sector rotation illustrated in [6] as 22 sector clusters for the S&P 500 index stocks. In out of equilibrium market, we expect to observe ‘agglomeration’ of clusters as a result of synchronized trading dynamics stemming from rule based, imitation and herding behavior.

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

However, if it is in a market away from equilibrium, information becomes complex and costly leading to imitation and herding in trading.Once this herding behavior becomes increasingly severe, it will reach a common market crash point, the so-called “Minsky Moment” or Critical Point [12]. At the critical point (in extreme events), all stock movements are highly correlated. So, there will be fewer clusters because of the higher and wider correlation among stock prices. Therefore, we expect to observe fewer but larger clusters.

H2: In a price bubble build up regime we should observe fewer clusters.

Based on synchronization of trading evidence from previous research we expect agglomeration of clusters based on herding and imitating traders’ behavior. When all traders follow the same panic selling strategy a pattern of synchronizationcould be visualized in only one cluster identified by our method.

4. Results

4.1 Market Equilibrium State Clusters

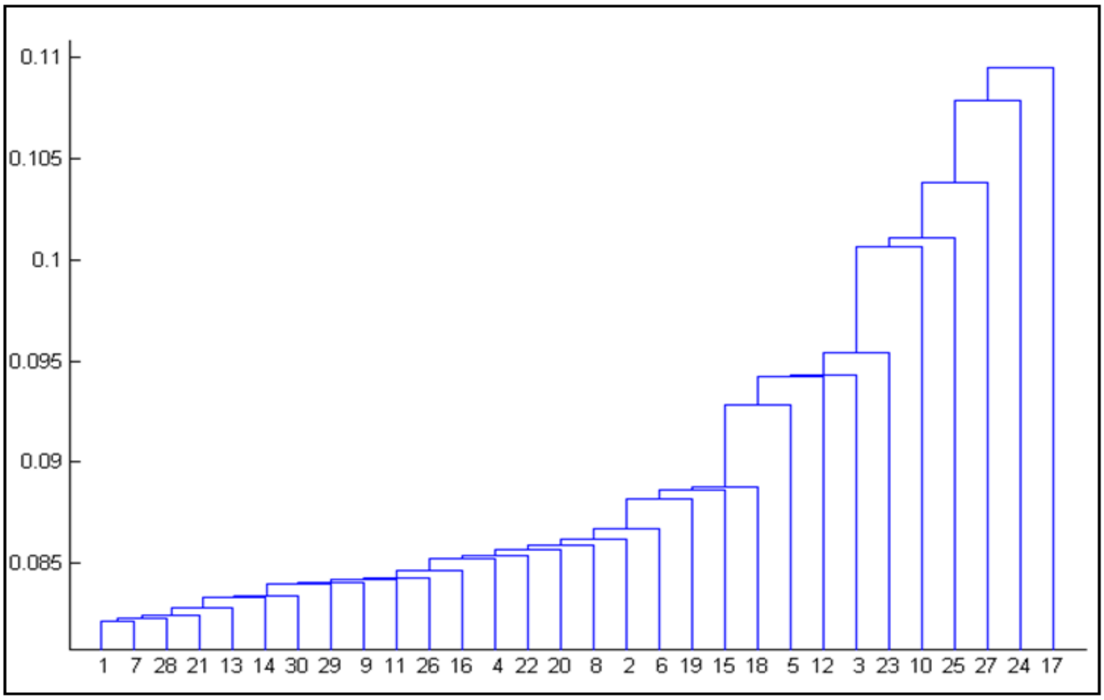

Fig 4.1 Time period: Jan 2nd 2002 to Dec 30th 2005

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 Fig 4.1.

As we have explained in in Methodology, we use this dendrogram figure to display the computation results of the clusters during 2002 and 2005. In Fig 4.1, the dendrogram figure actually shows us the cluster tree from our data set. The height of the U shapes in the dendrogram represent 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 Fig 4.1. 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. 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 Fig 4.2, Fig 4.3, Fig 4.4, Fig 4.5 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 Fig 4.2. For the time period of Jan 2nd 2003 to Dec 31st 2003, we found 26 clusters (as shown in Fig 4.3). For the time period of Jan 2nd 2004 to Dec 31st 2004, we found 30 clusters as shown in Fig 4.4. For the time period of Jan 3rd 2005 to Dec 30th 2005, we found 30 clusters as shown in Fig 4.5.

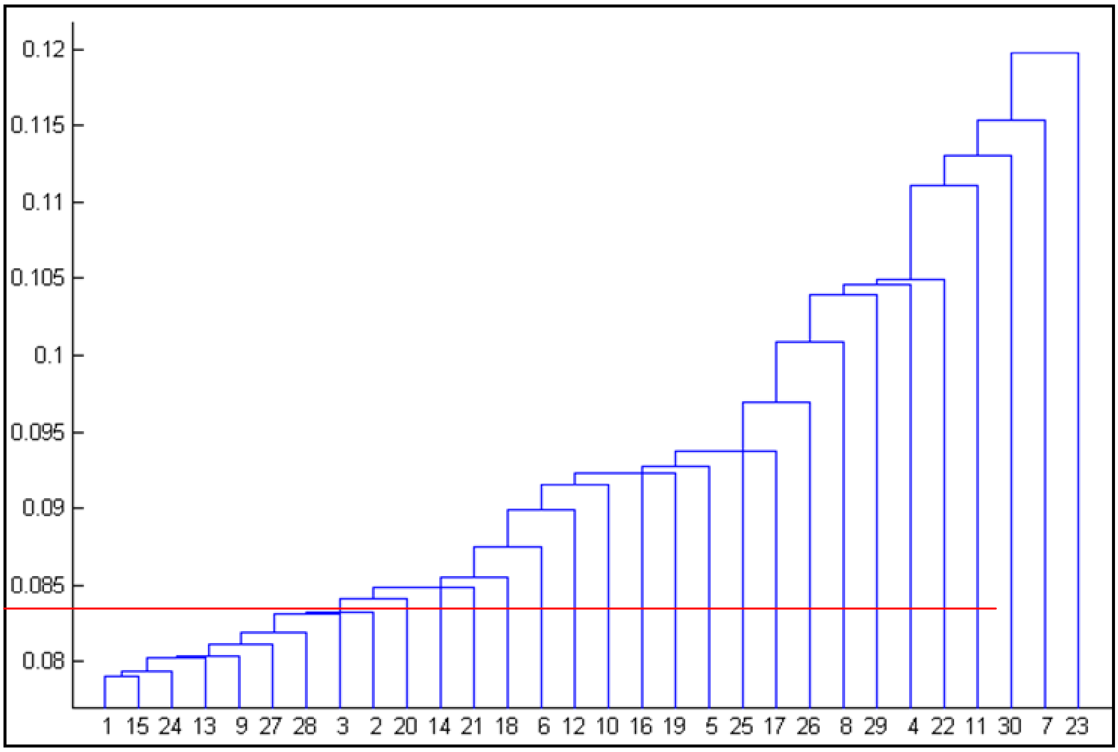

Fig 4.2 Time period: Jan 2nd 2002 to Dec 31st 2002

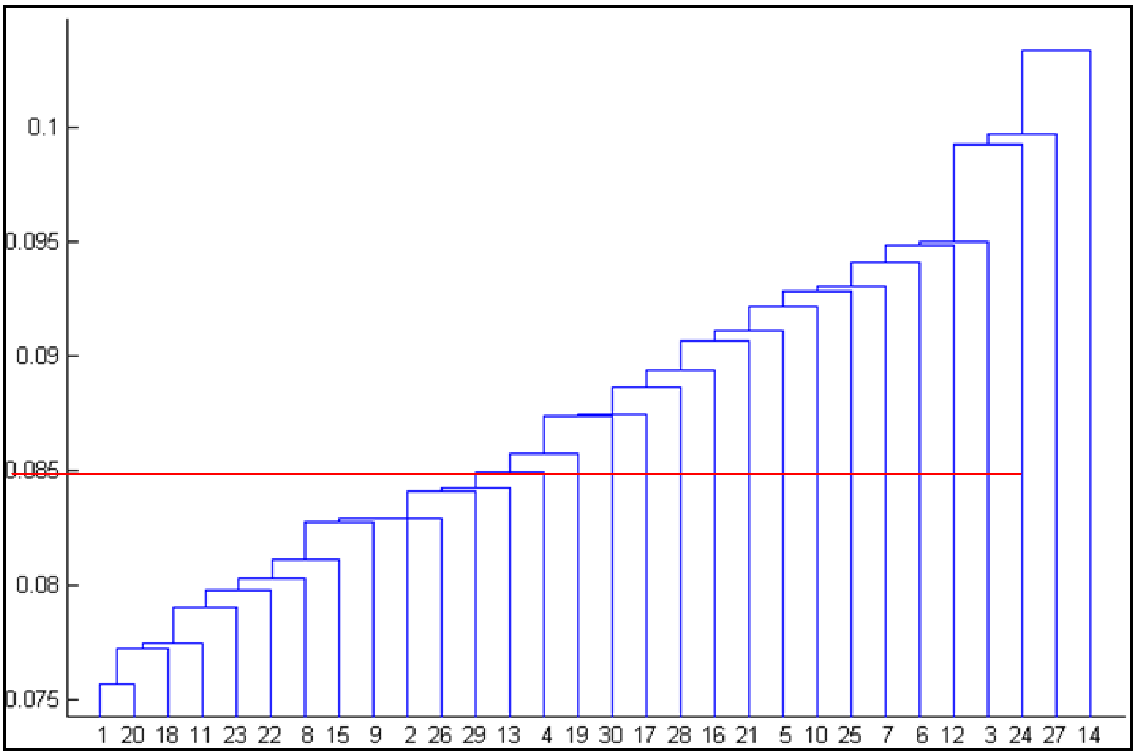

Fig 4.3 Time period: Jan 2nd 2003 to Dec 31st 2003

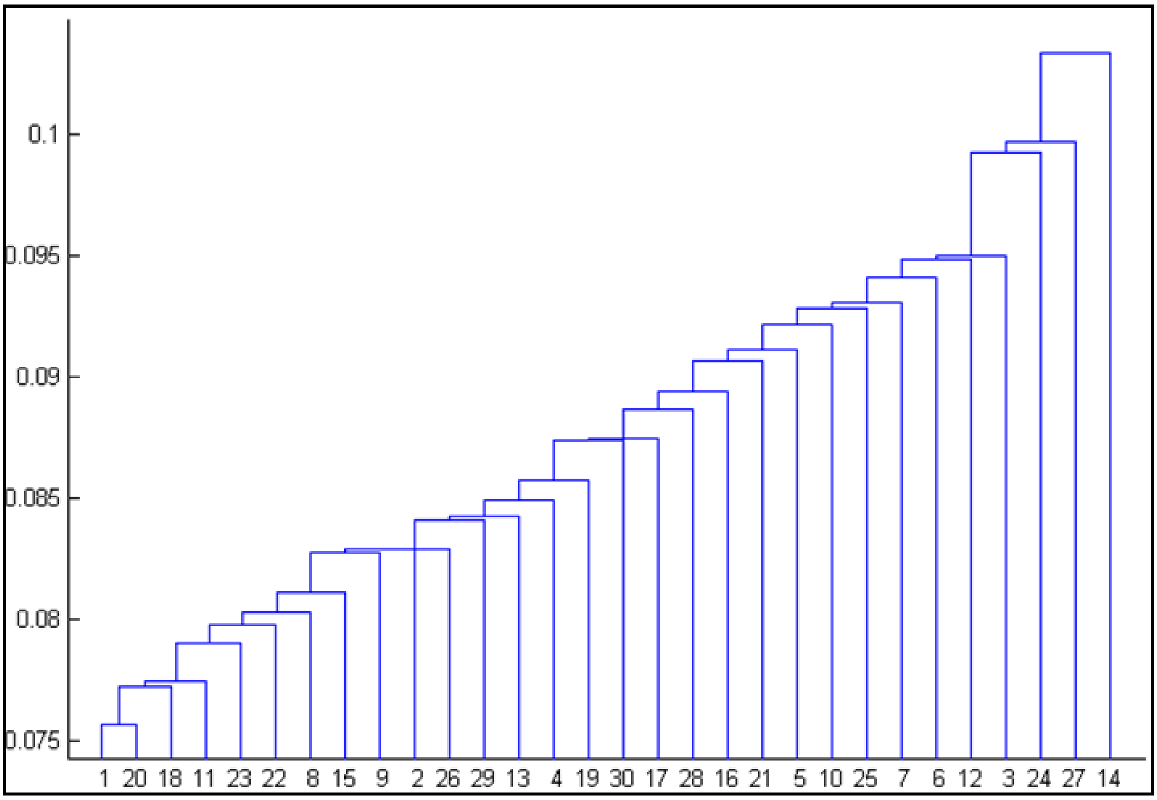

Fig 4.4 Time period: Jan 2nd 2004 to Dec 31st 2004

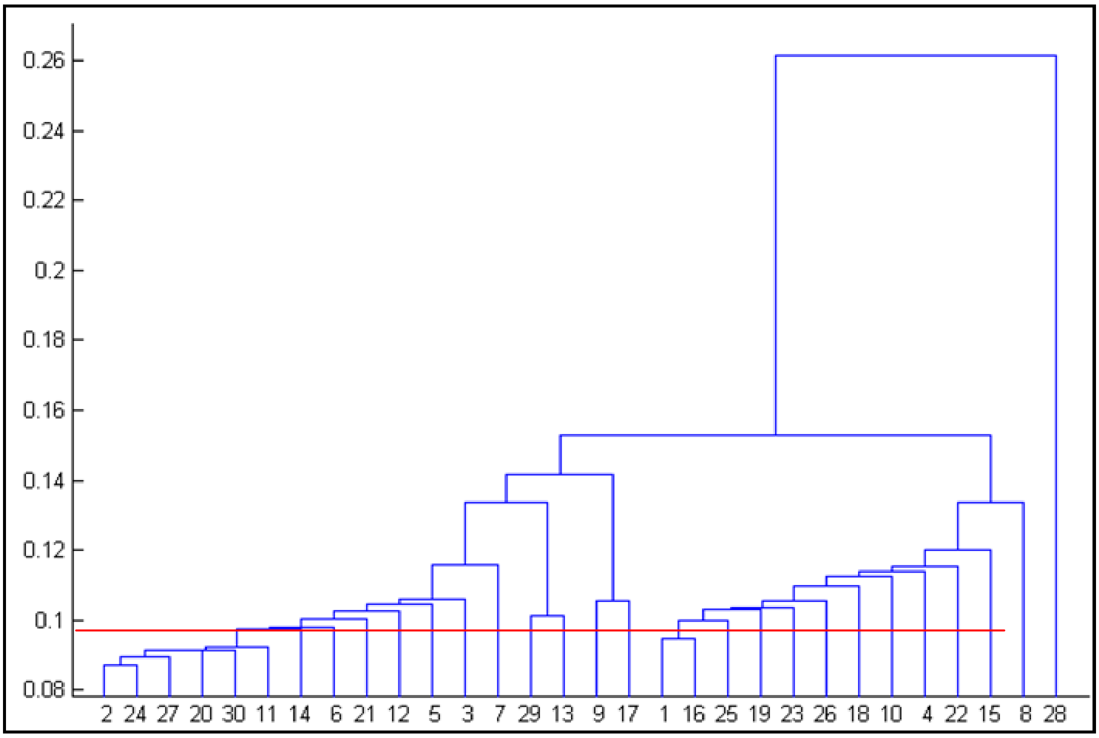

Fig 4.5 Time period: Jan 3rd 2005 to Dec 30th 2005

4.2 Market Disequilibrium State Clusters

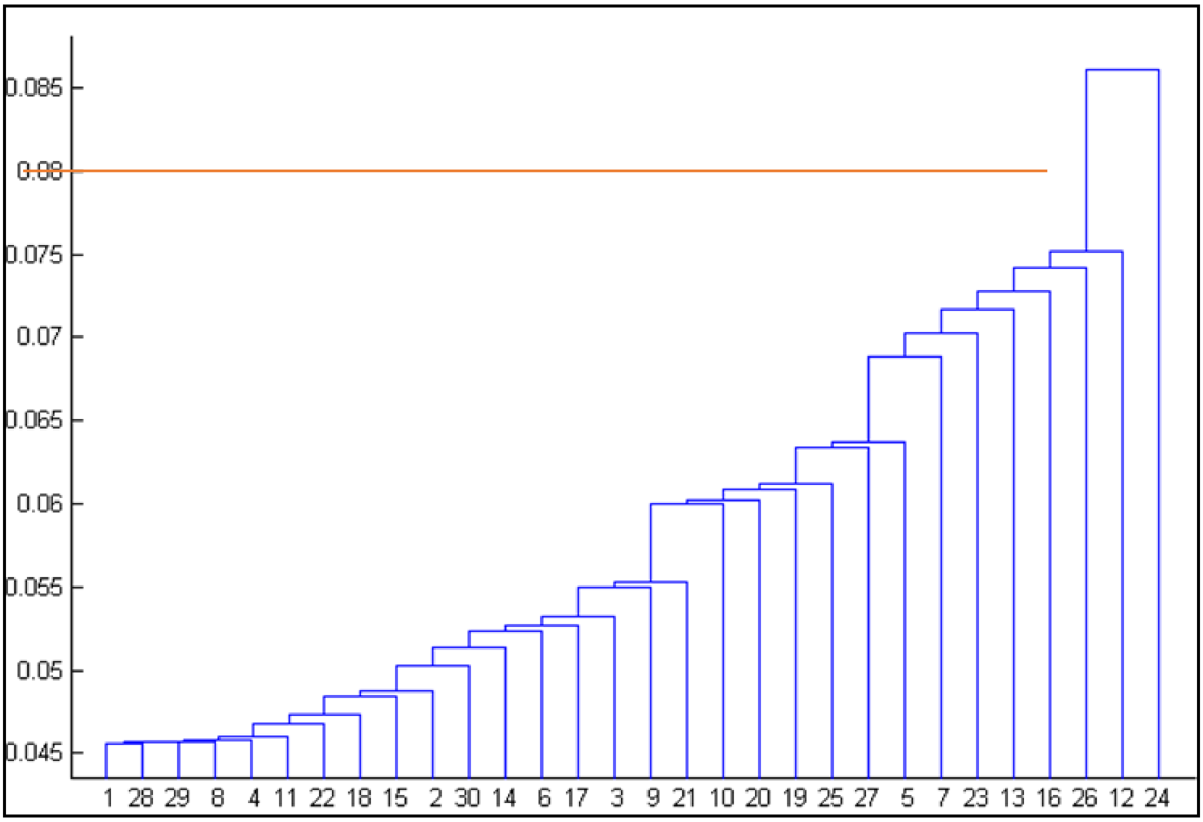

Fig 4.6 Time Period: Jan 3rd 2006 to Sept 15th 2008

In order to capture the cluster ‘agglomeration’ dynamics, we selected the time period from 2006 to 2009. We assessed the daily return of the 582 available stocks during the time period from January the 3rd 2006 till December the 31st 2009 from the CRSP data center. We ran the data with the PHA clustering method and report the results in Fig 4.6.

For the 2006 to 2009 time period, we applied the threshold of 0.08 from the Market Equilibrium State to the 2006 to 2009 time period. As shown in Fig 4.6, we can see the number of clusters changing from 27 clusters to only one cluster.

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 at the critical point in the market. So, the results support our Hypothesis 1 and Hypothesis 2.

5. Discussion

5.1 Market Equilibrium State

Figure 4.1 to Figure 4.5 show the number of clusters during periods of normal/equilibrium market dynamics. We assume thisis a topological manifestation of a well-known pattern of capital flow called “sector rotation.[6].On S&P 500 index data a new method for backbone extraction that does not rely on any particular null model, but instead uses the empirical distribution of similarity weight to determine and then retain statistically significant edges, identify22 clusters [15].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. Therefore, the PHA clustering results support our Hypothesis 1, we expected at least 22 clusters.

We have also applied the PHA clustering method to four shorter time periods (subperiods) within the selected time period of Jan 2nd, 2002, to Dec 30th, 2005, in the market equilibrium state. We continued to use the threshold of 0.08 from the time period of Jan 2nd 2002 to Dec 30th 2005. The number of clusters during the four time periods are slightly varying but still in a range from 26 to 30. For the time period of Jan 2nd 2002 to Dec 31st 2002, we found 28 clusters as shown in Fig 4.2. For the time period of Jan 2nd 2003 to Dec 31st 2003, we found 26 clusters as shown in Fig 4.3. For the time period of Jan 2nd 2004 to Dec 31st 2004, we found 30 clusters as shown in Fig 4.4. For the time period of Jan 3rd 2005 to Dec 30th 2005, we found 30 clusters as shown in Fig 4.5. We found strong support for our hypothesis of cluster dynamics during normal market.

The reasons for these clusters structure can be explained as follows. Firstly, it is one of the key features of a complex network with freely available transparent information to adapt and function in a stable range. Secondly, the clusters show us the interaction and exchange inside a network with co-operation/competition balance. For a stock market network, those activities can tell us the strength of the connections between any two clusters and the cash flows among the stocks in this network. According to agent-based models in the stock market, the order flows reveals information about market participants’ valuationsfor different stocks and impacts price movements in a feedback process. Under the market equilibrium state, information is available for all the market participants, so the order flows in and out of stocks as industry rotation occurs. Prevailingbuy order flow into a stock will cause the stock price to move up and will be shown as a positive return. Sell orders direct cash flows out of the stock and move the price down.In normal market the balance between buy and sell orders sustain around the “fair price”reflecting heterogeneity of traders’ intentions/valuations, while homogeneity leads to imbalance in the order flow.We suggest that clusters provide an insight to the order/cash flows among stocks.

5.2 Market out of Equilibrium in a bubble build up

Fig 4.6 provides the number of clusters in the period preceding 2008 market crash. 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. We expectedindustry/sector clusters to aggregate to larger fewer leading to one or two large clusters during the pre-crash time period. Once free transparent information becomes unavailable market can not find equilibrium [11]. Following rule based trading, imitation and herding, traders synchronize their activity, which leads to high correlation in stock price movements.As a result correlations between clusters lead to agglomeration of clusters.Agent based models demonstrate how stock market participants converge to one dominant rule and that showed as one cluster in our empirical investigation. 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 [10]. 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 in the form of free transparent information. 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. Conclusion and Future Work

Researchers from multiple disciplines have tried to identify patterns in stock price movements before drawdowns. Independent empirical evidence has converged to prove trading strategies synchronization phenomenon as the trigger of stock market crashes[16]. A Phase Transition Model explains bubble build up building-up mechanism leading to a critical point followed by stock market crashes [12]. In this study, we have proposed another method to identify the synchronization dynamics of stock price movements before the 2008 – 2009 crash. We apply the Potential-based Hierarchical Agglomerative (PHA) clustering method used in biology and physics to reveal S&P 500 stocks' complex network topological structure. Results support our expectation of agglomeration in the cluster structure culminating in one cluster topology. Our resultssupport findings revealing sector clusters topology in the US stock market [6]. In agreement with [15] we identify 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 in a market equilibrium state. More importantly, we have also identified the clusters' convergence into only one large cluster in the market bubble build up state, which further supports our hypotheses.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 other market participants.

Our empirical results cover only stock price dynamics before 2008 – 2009 market crash, more empirical evidence using different methods is needed to support patterns of trading synchronization before crashes. We hope there will be further studies to test price patterns in other periods of market turmoil. Moreover, we will add two new methods of identification of cluster in a search for further evidence to support our study, namely backbone extraction method and dot matrix plot.

References

[1]. D. Sornette ‘Why Stock Markets Crash: Critical Events in Complex Financial Systems, Princeton Science Library, 2017, pp.448

[2]. R. N. Mantegna,“Hierarchical structure in financial markets”,The European Physical Journal B - Condensed Matter and Complex Systems, *11*(1), 1999, pp.193-197.

[3]. R. N. Mantegna and E. Stanley,“Introduction to Econophysics: Correlations and Complexity in Finance”, Cambridge University Press, 1999.

[4]. N. Vandewalle, F. Brisbois, and X. Tordoir, “Non-random topology of stock markets”,Quantitative Finance, vol. 1(3), 2001, pp. 372-374.

[5]. A. Krause, “Predicting crashes in a model of evolving networks”,Complexity, vol. 9(4), 2004, pp. 24-30.

[6]. G. Leibon, S. Pauls, D. Rockmore and R. Savell,“Topological structures in the equities market network”,Proceedings of the National Academy of Sciences, vol. 105(52), 2008, pp. 20589-20594.

[7]. R. Heiberger, “Stock network stability in times of crisis”,Physica A: Statistical Mechanics and its Applications, v. 393, 2014, pp. 376-381, https://doi.org/10.1016/j.physa.2013.08.053.

[8]. G. J. Wang, C. Xieand H. E. Stanley,“Correlation Structure and Evolution of World Stock Markets: Evidence from Pearson and Partial Correlation-Based Networks”, Computational Econom.,vol. 51, 2018, pp. 607–635. https://doi.org/10.1007/s10614-016-9627-7

[9]. R. Albert. andA.-L. Barabási, “Topology of Evolving Networks: Local Events and Universality”,Physical Review Letters, vol. 85(24), 2000, pp. 5234-5237.

[10]. D. J. Watts andS.H. Strogatz, “Collective dynamics of 'small-world' networks”,Nature, 393, (6684), 1998, pp. 440-442.

[11]. S. J. Grossman and J. E. Stiglitz,“On the Impossibility of Informationally Efficient Markets”,The American Economic Review, vol. 70(3), 1980, pp. 393-408.

[12]. R. Yalamova and B. McKelvey, “Explaining What Leads Up to Stock Market Crashes: A Phase Transition Model and Scalability Dynamics”,Journal of Behavioral Finance, vol. 12(3), 2011,pp. 169-182.

[13]. Y. Lu and Y. Wan, “PHA: A fast potential-based hierarchical agglomerative clustering method”,Pattern Recognition, vol. 46(5), 2013, pp. 1227-1239.

[14]. S. Shuming,Y. Guangwen, W. Dingxing and Z. Weimin, "Potential-based hierarchical clustering,2002 International Conference on Pattern Recognition, vol.4, 2002, pp. 272-275. doi: 10.1109/ICPR.2002.1047449.

[15]. N. J. Foti, J. M. Hughes and D. N. Rockmore, “Nonparametric Sparsification of Complex Multiscale Networks”, PLoS ONE,2011, 6(2): e16431.

[16]. C. K. Tse, J. Liu and F. C. M. Lau, “A network perspective of the stock market”,Journal of Empirical Finance, vol. 17(4), 2010, pp. 659-667. doi: http://dx.doi.org/10.1016/j.jempfin.2010.04.008