The Quest for Living Beta: Investigating the Implication of Shareholder Networks

Introduction

The behavior of financial markets has, and continues, to frustrate investors and academics. With the advent of new approaches, including complex systems and network analysis, the search for an explanation as to how and why markets behavior as they do has greatly expanded. The complex system approach is consistent with the thoughts of Sornette (2014), who concluded that after 20 years of research, the key concepts required to understand stock market returns are; imitation, herding, self-organized co-operativity, and positive feedbacks, with many of these features captured by network analysis. This point is confirmed by Newman (2010), who suggested that networks are a "powerful means of representing patterns of connections or interactions between the parts of a system". In short the rationale for utilizing a network is that the behavior of a system can vary greatly depending on the network structure (the topology) of a system.

This paper utilizes an extensive data set provided from the Thomson Reuters 13f database, to undertake a temporal analysis of the networks formed between US institutional investors and the stocks in the S&P 500. The analysis makes use of both projected and bipartite networks and uncovers numerous insights regarding relationships between the market in general, stocks and their shareholders.

The remainder of this paper is set out in the following manner: Section 2 provides a brief literature review justifying the approach taken in this paper; Section 3 details the approaches taken and the data used to undertake the analysis; and Section 4 provides the results of the analysis. Both Section 3 and 4 are subdivided into separate sections covering the bipartite, the investor-by-investor, and stock-by-stock networks. A conclusion and summary is then provided in Section 5.

Background

Financial markets are characterized by periods where price movements and trading volumes have become much more volatile than expected. One of the most recent occurrences of such an event saw the Dow Jones Industrial Average fall 21% in the first nine days of October 2008, and the world subsequently plunged into the Global Financial Crisis (GFC). In response to the failure of traditional economic and finance theories to explain such an event, Schweitzer et al.(2009) argue that the analysis of economic networks has become essential. Their rationale is that existing theories, and the policies associated with them, are inadequate in understanding and analyzing the growing interdependencies that had and continue to be formed across global trade, supply chains, and investments networks.

Since the existence of a network between investors was formalized by Shiller and Pound (1989), there has been an extensive body of work developed to understand the ramifications of opinions and the actions of investors flowing across these networks. This includes the ability of networks to explain investor trading decisions and portfolio performance (Ozsoylev & Walden, 2011). The importance of networks to financial markets is further emphasized with the excess volatility seen in stock markets being partially explained by investors herding as they mimic investors in their network, with Cont & Bouchaud (2000) first providing evidence of this behavior in financial markets.

The existing work relating to network structures associated with stock markets has tended to focus on forming networks based the correlation between individual stocks (see Preis et al.(2012), Bonanno et al.(2004), Boginski et al.(2006) and Kenett et al.(2010)), between investors by implying an investor network (Ozsoylev et al.(2014)) or finding an actual network (Shiller and Pound (1989), or Hong et al.(2005)). A characteristic of many of these studies is that the analysis has been performed on a static network, where ultimately a temporal network will provide greater insights. However, in an approach relevant to this paper, utilizing the correlation –between-stocks approach, Boginski et al.(2006) assessed the temporal dynamics of the US stock market network between 1998 and 2002 and found that the market network increased in density over time, a fact Boginski et al.(2006) attributed to increased globalization. The size of the maximum clique was also reported to increase over time, suggesting stocks had an increasing tendency to move together.

The approach taken in this paper extends the use of a bipartite network, a network between stocks and investors, as seen in Caldarelli et al.(2004). A benefit of this approach is that the relationship between a manager and a stock is directly supported by the data, and so does not need it to be implied through an alternate approach. The point of difference for this paper is that the network will be assessed over 16 quarters, as opposed to a static network. A key finding of

Caldarelli et al.(2004) was that their resulting market graph was characterized by a scale-free topology, where the degree distribution of the stocks matched a power-law distribution. This result provides evidence that financial markets operate as a complex adaptive system (CAS), meaning that traditional financial theories based on equilibrium conditions are likely to be insufficient in explaining the dynamics of financial markets. To emphasis the point made by Schweitzer et al.(2009) that traditional method are inadequate in understanding the interdependencies seen in networks, the aforementioned findings regarding the degree distribution of the network are inconsistent with the Capital Asset Pricing Model CAPM (Sharpe, 1964). If the CAPM held, then Caldarelli et al.(2004) suggest that using their framework stocks would record a constant in-degree, something that was not observed.

Method

1.1 Approach

The research undertaken in this paper explores the temporal aspects of the bipartite networks (and subsequent stock-by-stock, and investor-by-investor networks) formed for each quarter between 2007 and 2010, of the stocks in the S&P 500 and the US institutional investment managers that held them (note that the term investment manager, or manager, is interchangeable with investor, and has been used as such in this paper). This period was selected as it covers the period including the GFC, therefore providing a rich source of dynamics as investors reacted to the falling and then recovering market. The main contribution of this approach is an attempt to understand the network dynamics across multiple periods rather than interpreting a static position of the network.

1.2 Data

The data used for the research was collected from two sources. The shareholding data was collected from the Thomson Reuters 13f database, which records the stock ownership of all US institutional investment managers, and was accessed via the Wharton Research Data Services website. A US institutional investment investor is defined as an investment adviser, bank, insurance company, broker-dealer, pension fund, or corporation who uses the U.S. mail during its business, and exercises investment discretion over at least $100 million in assets. Therefore, while not capturing all investors, the data does encompass approximately 65% of investors[1] . The database provided the option of forming the network based on a manager’s ID number or by their

[1] Based on http://www.heritage.org/taxes/report/most-stocks-are-held-private-investors.

name. The manager ID number was chosen because there was evidence of trivial differences in manager name from period to period, causing the formation of unnecessary nodes across the sample period.

In terms of the filing requirements for investment managers, Rule 13f-1(a)(1) requires managers submit four Form 13F filings once they meet the $100 million filing threshold on the last trading day of any month during any calendar year. The rule also requires the managers file four Form 13F filings, even if after meeting the $100 million filing threshold they fall below the threshold in the subsequent periods. This rule will limit the influence of any survivorship bias in the sample set, because managers will still be required to report regardless of whether their funds under management took a temporary hit during the GFC.

The shareholding data did require considerable pre-processing including the identification and removal of duplicated records. In addition, stocks that did not have records for each of the 16 quarters, because of a takeover or bankruptcy or the like, were removed. In summary, the data comprised over 3.6 million records, with approximately 3,600-plus distinct investors identified for 477 stocks. It is evident from Figure 1 that the number of managers who reported each period varied, with the direct implications on the network being unclear. However, it was assumed that if a manager no longer reported it was due to merging with another manager, falling below the $100m threshold for an extended period, or no longer holding the S&P 500 stocks in the sample.

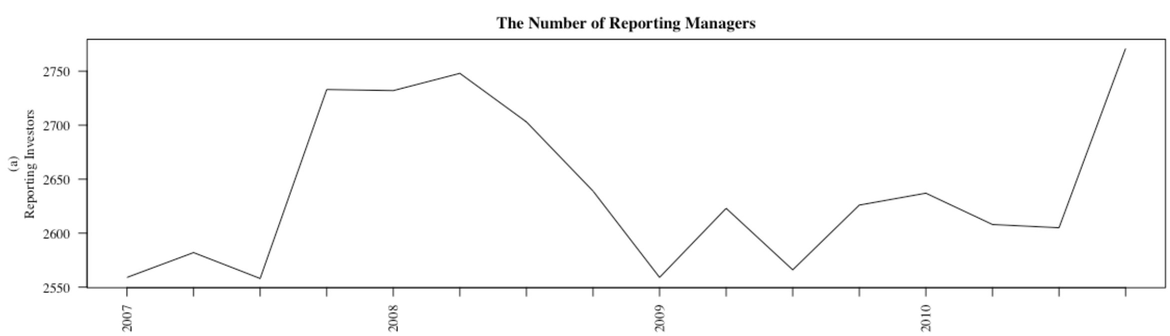

Figure 1: The number of investors reporting in each period

The shareholding data provides the benefit of being able to generate two network frameworks. The first is a binary network, where a link is formed and has a weight of 1 if an investment manager holds a given stock. The second is where the weight of the link is determined by the manager’s proportional holding in a given stock, as per Equation 1. It should be noted that the denominator in Equation 1 is the sum of the shares from the data rather than the total shares on offer for the stock. Future iterations of this analytical framework may look to utilize the alternate approach.

$$w_{ij} = {S_{ij}}\Big/{\sum_j^J S_{ij}} \tag{1}$$

By way of definitions for Equation 1, w_{ij} refers to the weight of the link between the ith stock and the jth manager, S_{ij} is the number of shares the jth manager holds in the ith stock, and $\sum_j^J S_{ij}$ is the sum of all the managers (J) holdings in the ith stock.

The financial data – price and financial metrics - was collected from the Thomson Reuters' Datastream system. The pricing data was used to create a return series for each of the stocks, with its use seen in Section 1.6, and to calculate the betas, as per the CAPM, for each of the stocks for each of the 48 months in the sample. The betas were calculated as per Yahoo Finance's definition, which is to regress 36 monthly returns for the stock in question against the returns of the S&P 500 index for the comparable period. The financial metrics were collected to assist in identifying the characteristics of the communities, as identified in Section 1.6. The metrics, and the rationale for their use were:

- Price-to-book ratio (PB): This metric compares the market value of a stock to its book value, as provided in the stock's financial statements. It is calculated by dividing the current closing price of the stock by the latest quarter's book value per share. A lower PB is seen to indicate that a stock is undervalued and therefore a candidate for a value manager.
- Price-earnings ratio (PE): This is the measure of a stock's current share price relative to its per-share earnings. Generally, a high PE ratio is an indication that investors are anticipating higher growth in the future.
- Market capitalization (Market Cap.): Refers to the total dollar market value of a company's outstanding shares. It is calculated by multiplying a company's shares outstanding by the current market price of one share. The metric is used to determine a company's size, and categorize a stock as a large or small cap stock.

- Dividend Yield (DY): The metric is calculated by dividing the dollar value of a stock's dividend per share divided by the stock's share price. In general, a higher DY indicates that a stock is ex-growth as the company does not need the funds for future growth.

1.3 Network Analysis

The collected dataset allows for analysis to be undertaken at both the bipartite level and single node networks. The justification for undertaking analysis at the bipartite level is based on Opsahl (2013), where the shortfalls of projecting a two-node network onto a single node network are discussed. The main concerns put forward are: the creation of links for the projected network are not consistent with how they would be formed in a single node network, with the contention being that the links are no longer independent, nor were they formed in a random manner; and the projected network will have larger and more fully connected cliques.

The downside to the bipartite approach is that many of the traditional network analysis metrics and techniques, such as betweenness and closeness centrality, are not capable of being used. To overcome this the single mode networks (either investor-by-investor or stock-by-stock networks) are formed utilizing matrix algebra. The traditional analytical tools were then applied to these single mode networks, with the line of investigation detailed in Section 1.3.2 and the results in Section 1.5 and 1.6.

1.3.1 Top Level

To undertake the analysis at the bipartite level, the R (R Core Team, 2017) bipartite package (Dormann, Gruber, & Fruend, 2008) was utilized. Using the network level function, the following variables were used to investigate the characteristics of the network:

- Cluster coefficient: Using the approach outlined in Opsahl (2013), the clustering coefficient is returned for the overall network, and the upper and lower levels. The coefficient for a level is the (weighted) average cluster coefficients of its members.

- Connectance: The variable is defined as the sum of links divided by the number of companies multiplied by the number of investors, effectively the density of the network.

- Links per species: This is the average number of links for the companies and investors.

An important point relating to these variables is that they will not be directly affected by the variations in the number of manager reporting in each period, as illustrated in Figure 1. The rationale is that the metrics are measures of actual links compared to possible links for each independent period or, in other words, they describe how connected the current set of managers are with the current set of stocks.

1.3.2 Single Mode Analysis

The single mode networks (stock x stock (SxS) and investor x investor(IxI)) were created using both the binary and weighted bipartite networks (as described in Section 1.2). The relevance of the SxS network is that it generates a network where the weight of the link between any two stocks represents the number of managers that have a common holding in the two stocks. The creation of the SxS networks resulted in a fully connected graph because amongst the 2,500-plus managers in any one period, at least two managers had a common shareholding in all the stocks in the dataset. Given the extreme density of the network the use of the traditional network metrics were somewhat hindered, and visualization of the network was pointless. However, as detailed in Section 1.6, the networks did provide some novel and useful insights.

In contrast to the SxS network, the IxI network forms a link between two managers if they have a common share holding in a given stock. As illustrated in Figure 2, while high, the density of this network did not reach the saturation levels of the SxS network. The density of the network follows a distinct pattern which is in--line with the findings relating to Figure 3; that is, the density leads the market decline, as the market descended into the GFC, before recovering ahead of the market. The density also declines during the market correction in mid 2010.

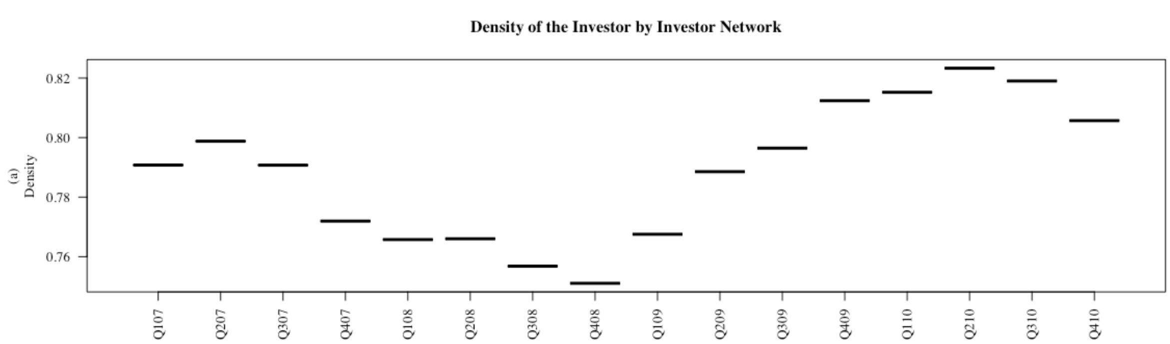

Figure 2: The density as returned by the investor-by-investor network. The trend is consistent with what is seen and reported below.

Findings

1.4 Top Level

Illustrated in Figure 3, graph (a), which plots the S&P 500 index and the cluster coefficient of the bipartite network, is the fact that the connectedness of the market system declines ahead of the market, before recovering as the market recovers. The linkage density and connectance coefficients of the bipartite network also show similar characteristics. The most feasible explanation for this result is that initially investors reduce their holdings by moving to cash and/or focusing on quality stocks as the market falls. The cluster coefficient reaches its minimum point in the fourth quarter of 2008, before recovering as investors return to the market with increased confidence following the various policy initiatives from governments and central banks. The increased clustering occurs as investors spread their bets across the market, a point explored farther in the following section. Exiting and reentering the market is a classic example of herding behavior, and has been captured clearly by movement in the network's clustering coefficient, something not reported before. Consistent with behavior of the markets during 2008 and 2009, at the start of 2010, there was a correction in the market. with concurrent declines in the bipartite clustering coefficient, and the network density from the IxI network.

The question of what is driving this process is partially explained in Figure 3 graphs b and c. The clustering coefficients for the stock and investor networks behave in very different manners. The stock network more closely resembles the overall network, with there being a decline in the clustering that occurs ahead of the collapse in the S&P index. As the clustering coefficient is the actual number of links divided by the possible links, this finding tells us that investors dropped their holdings in certain stocks, before picking them up when the market recovered. This process is further investigated in Section 1.6. In contrast to the overall market network, the clustering of the investors grows almost unabated through the sample period. This is suggestive of the fact that over the period the trend for the investor population was to "buy the market"-- that is managers diversify across a greater number of stocks, or at least buy a common set of safe stocks, rather than attempt to identify stock outside of the "herd."

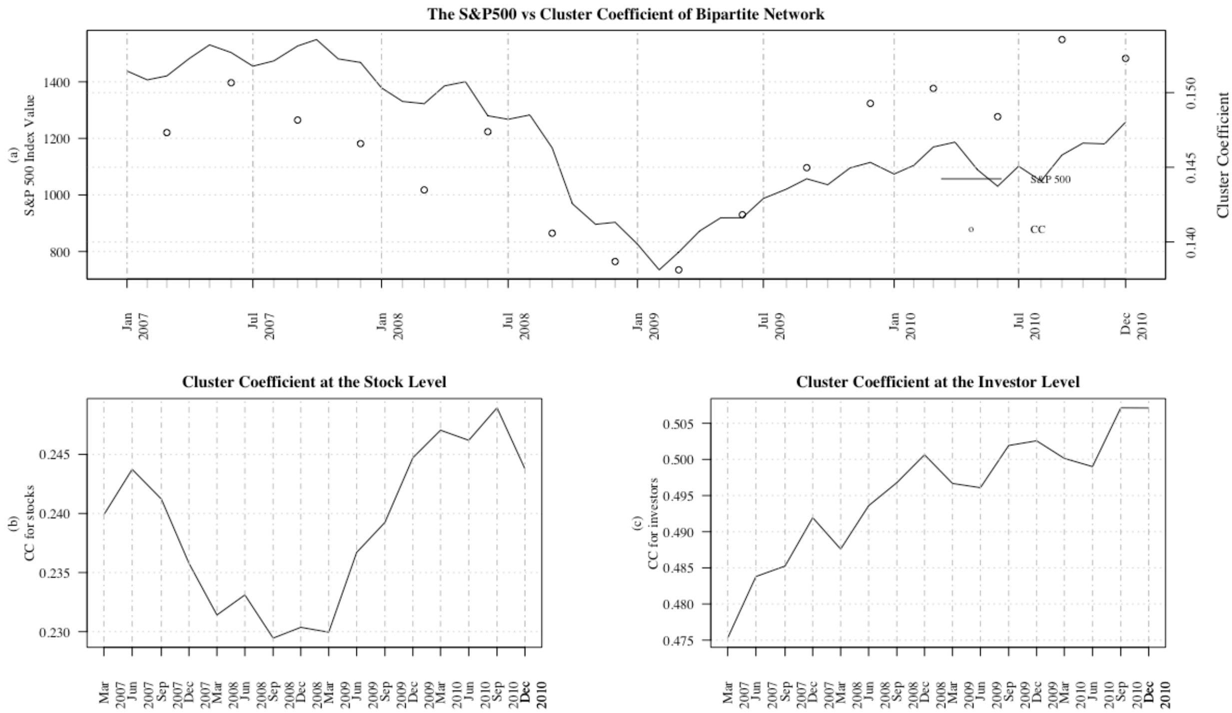

Figure 3: Graph (a) plots the movement of the S&P 500 and the cluster coefficient of the bipartite network from 2007 to 2010. Graph (b) plots the weighted cluster coefficient for the stock level network, with Graph (c) plotting the weighted cluster coefficient for the investor level network.

A possible explanation of how and why the market responded in the fashion it did comes from ecology, and the model of the adaptive cycle (Holling, 2001). The appeal of the theory is that it focuses upon the processes of destruction and reorganization, thus providing a more complete view of system dynamics as it links system organization, resilience, and dynamics. A key implication of the model is that as a system grows the components within the system will become more connected, increasing the output of the system. For the stock market, this means investors and companies become more connected with a resulting increase in market capitalization. However, as the system becomes more connected it becomes less resilient to any external stocks and may fail following a shock. A collapse will see the connectedness and output of the system decline, before it eventually reconnects in the recovery phase, a phenomenon illustrated in Figure 3, where the clustering declines in unison with the value of the index.

The read-through in general for an investment market network is that investors and market observers should expect to see the density of the bipartite investor stock network increase to a point at which time it will become susceptible to a shock. If the shock is sufficient, then investors will dissipate as they reduce their holdings before reconnecting during the recovery. This approach leaves open the question of how large the shock needs to be to fracture the network. The level of "connectedness" could help explain why the markets capitulated with the collapse of Lehman brothers (and not Bear Stearns, where the market was not as connected), recovered and then experienced further corrections, once the market became more connected.

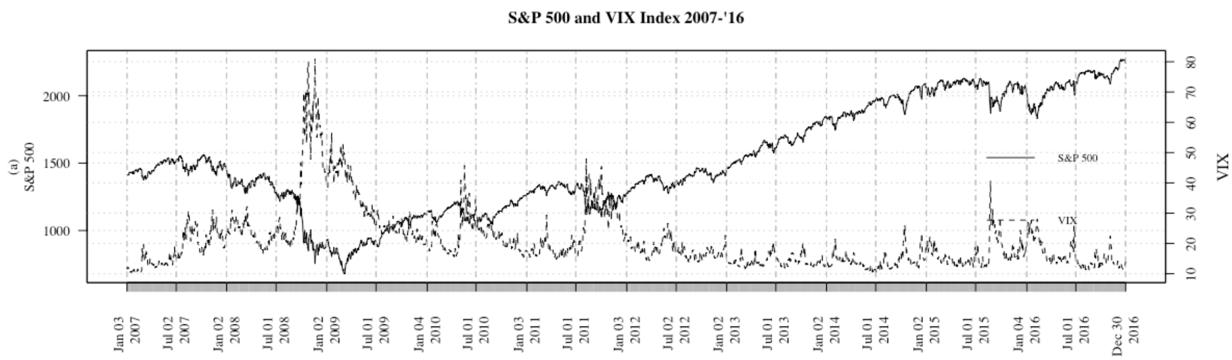

Figure 4: The S&P 500 index from 2007 until December 2016 ploted against the VIX volatility index.

The question of whether the findings of Figure 3 are pure coincidence or the system did behave in accordance with the adaptive cycle is difficult to answer definitively. In support of the argument that the financial markets do function in accordance with the adaptive cycle, Figure 4, which plots the S&P 500 and VIX indexes from 2007 to 2015, is provided. The figure illustrates that the market has experienced several periods of increased volatility associated with market corrections since the GFC. These corrections were generally the result of panic relating to either a change in central bank policies, or concerns around sovereign debt issues in Europe. As previously noted, the first post-crash correction (mid 2010) saw concurrent movements in the relevant network structure, so it would be expected that the subsequent correction would experience similar behavior.

The bipartite package produces cumulative degree distribution charts for each network level, stocks and investors, and fits a power-law function to them. The relevance of testing for the presence of a power-law comes from Boginski et al.(2006) and Caldarelli et al.(2004), who both found that the degree distribution in their networks followed a power-law, with the ramifications explained in Section 0. In general, across the sample period the distribution was heavily skewed (as seen in Figure 5), but did not meet the definition of a power-law, something not helped by the presence of some outlying stocks (see Figure 5(b)). A more detailed analysis of the degree distributions and how they change over--time is contained in the following sections.

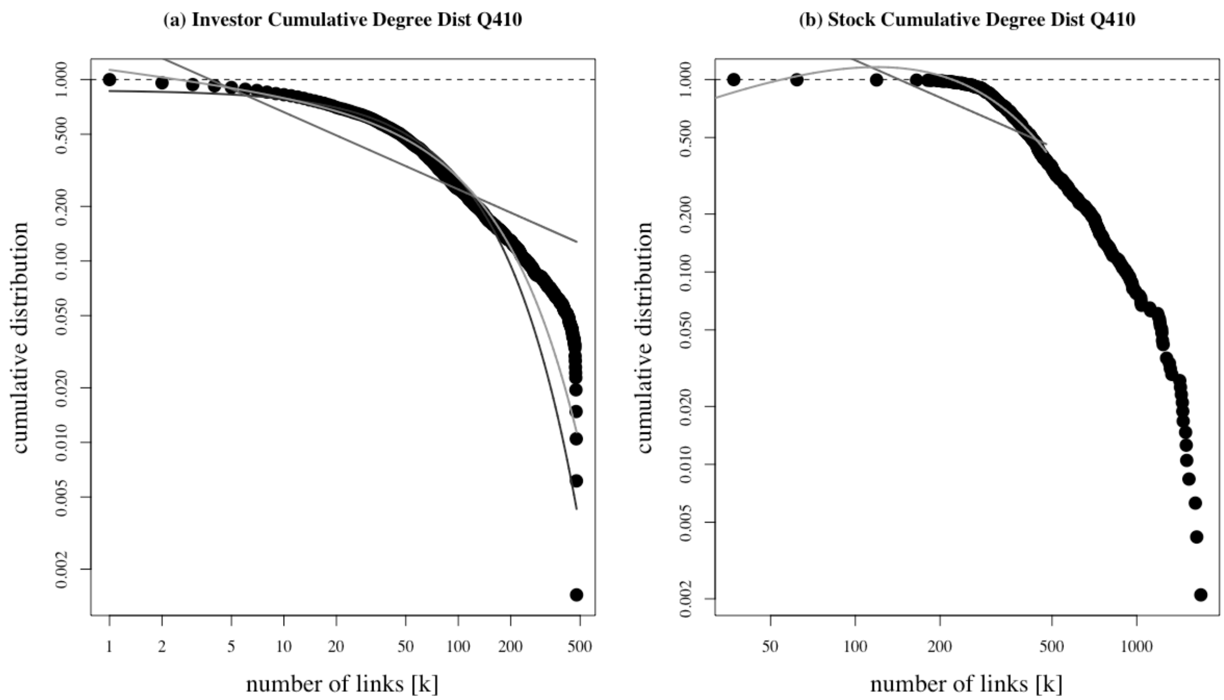

Figure 5: The cumualative degree distribution for the investor level network (Graph (a)) and the stock level network (Graph (b)).

1.5 Investor Analysis

An alternate interpretation of the degree centrality distribution of the investors, network is provided by Figure 6. Via boxplots for each of the 16 periods in the sample, the figure presents the degree centrality distribution calculated from IxI network and shows that, consistent with the skewed distribution, as seen in Figure 5(a), the distribution is heavily skewed. The interpretation is slightly different, in that most managers are linked to many other managers, while there is a smaller number of managers linked to few managers. The key observation is that the range of the degree distribution metric contracts post the bottom of the market, suggesting that the institutional managers were more likely to share common holdings. This outcome is consistent with the managers herding, as they rotated into a common set of less risky stocks. The median declines, before recovering, in a manner consistent with what was seen in Figure 2.

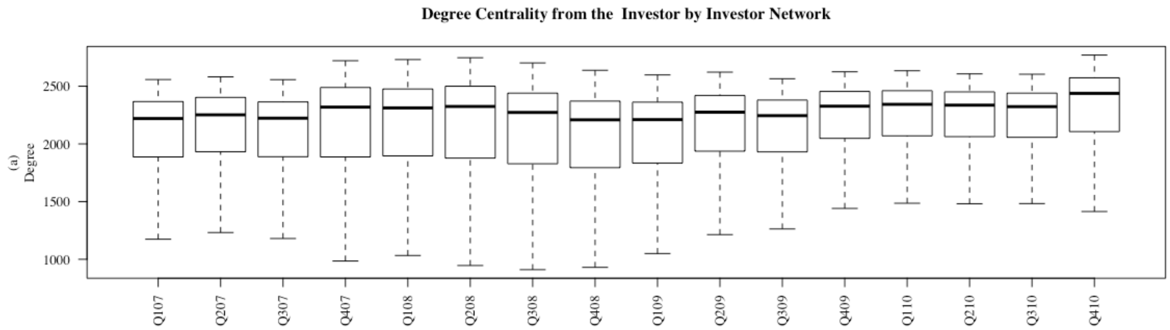

Figure 6: Boxplots illustrating the variation in the degree centrality as calculated from the IxI network.

An alternate interpretation of degree centrality is illustrated in Figure 7, where the centrality measure is calculated in accordance with Equation 2. The variable is measuring the number of stocks each manager held in each period, or in other words the in-degree from the stock level in the bipartite network. As such it is a closer representation of Figure 5(a).

$$d_{jt} = \sum_i^I S_{ijt} \Big/ I_t \qquad (2)$$

By way of definitions for Equation 2, d_{jt} refers to in-degree for the j^th manager in period t, S_{ijt} is for 1 or 0 depending on whether the j^th manager holds the i^th stock in period t, and $\sum_i^I S_{ijt}$ is the sum of stocks the j^th manager holds in period t. The denominator normalizes the degree count such that if a manager held all 477 stocks in the sample for period t, d_{jt} would be 1.

While the metric does not show the same variability as seen in Figure 6, it shows a skewed distribution indicating that most managers held fewer stocks but there are some managers that hold many (noting that outliers are not shown in the chart). This outcome, as expected, is consistent with Figure 5(a). To uncover the dynamics of how managers varied their holdings through the period, Figure 7(b), which shows the standard deviation of the degree distribution, and Figure 7(c), which shows the mean degree distribution for the managers, are provided. From Figure 7(b) it appears across the sample that most managers did not change the number of stocks they held materially. In terms of the average number of stocks, held Figure 7(c) highlights the presence of index funds which would have held all the stocks in the index, with most managers holding 10% of the possible stocks.

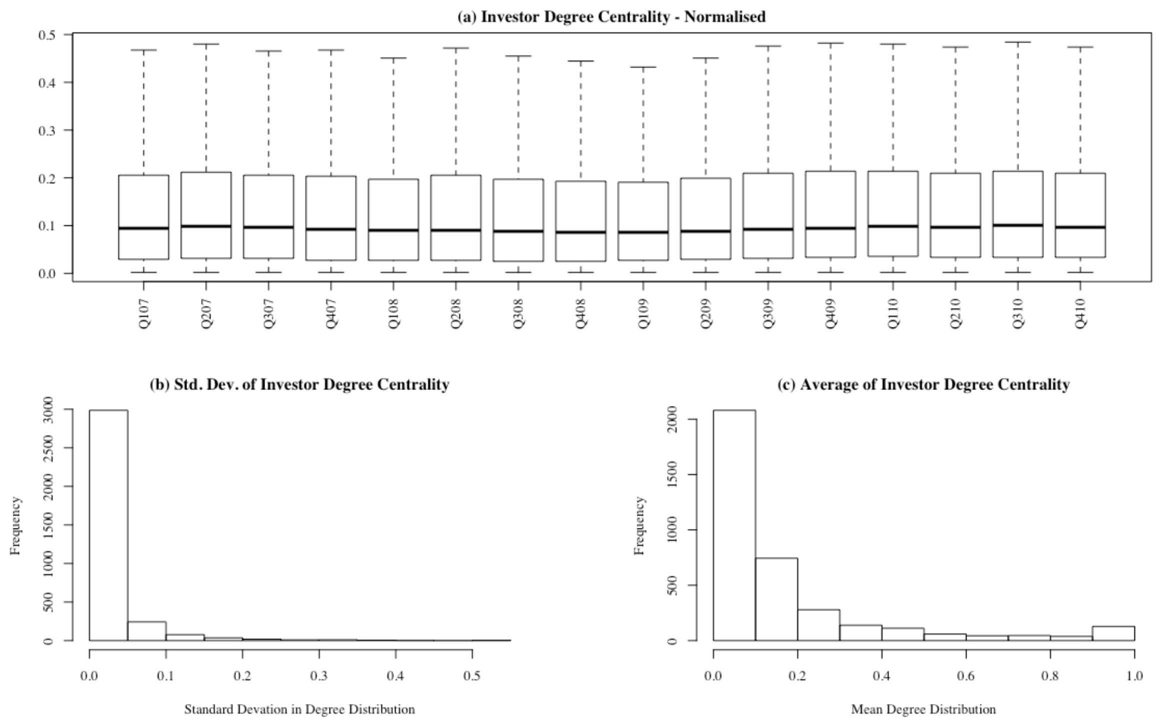

Figure 7: Boxplots (graph (a)) illustrating the variation in the degree centrality for investors as calculated by equation 1, from the bipartite network. Graphs (b) and (c) are histograms showing standard deviations and mean of the centrality measure across the sample.

The initial impression from the Figure 7(a) is that the change in the degree centrality measure across the 16 periods does not match that seen in Figure 6. A further investigation is provided in Figure 8(a), which shows the change in the mean number of stocks held by all managers. Consistent with the other findings in this paper, the average declines in periods of increased market volatility before recovering as the market improves. Figure 8(b) provides the standard deviation in terms of the number of stocks held by each manager. Consistent with what was seen in Figure 3(c), this metric increases effectively in a monotonic fashion. The interpretation is that the investor population is diverging in its approach, with one side either maintaining or decreasing its number of holdings, while the other increases their holdings, as they adopt the strategy of "buying" the market. It may also be an indication that the underlying investors are moving towards investment managers that are tracking the market.

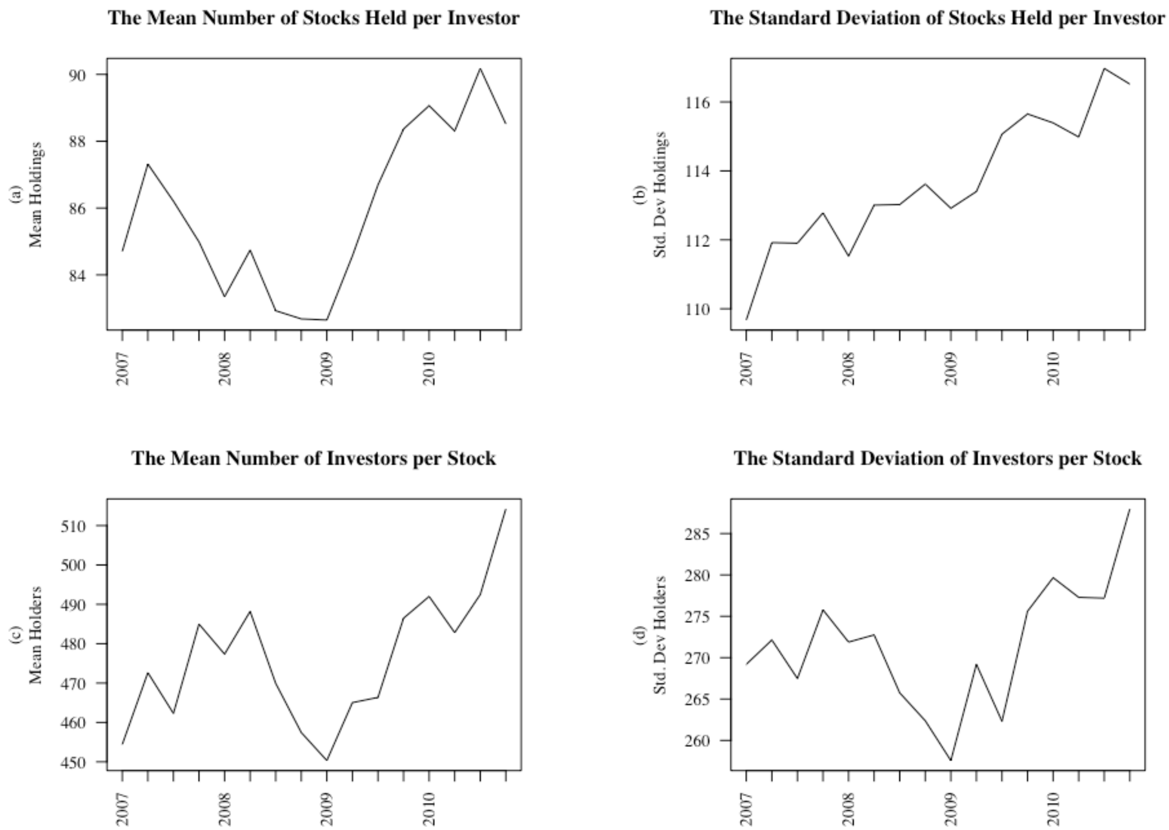

Figure 8: The dynamics of the number of stocks held per investor (graphs (a) and (b)), and the the number of investors per stock (graphs (c) and (d)).

1.6 Stock Analysis

This section investigates more specifically how the network(s) pertaining to the stocks varied over time. In a similar manner to Figure 7, Figure 9 provides information as to how the in-degree (a link from a manager) for each stock from the bipartite network varied. The metric was calculated as per Equation 3, and as such is in line with Figure 5(c).

$$d_{it} = \frac{\sum_j^J S_{ijt}}{J_t} \tag{3}$$

By way of definitions for Equation 3, d_{it} refers to in-degree for the ith stock in period t, S_{ijt} is for 1 or 0 depending on whether the ith asset is held by the jth manager in period t, and $\sum_j^J S_{ijt}$ is the sum of managers holding the ith asset in period t. The denominator normalizes the degree count such that if a stock is held by all the managers in the sample in period t, d_{it} would be 1.

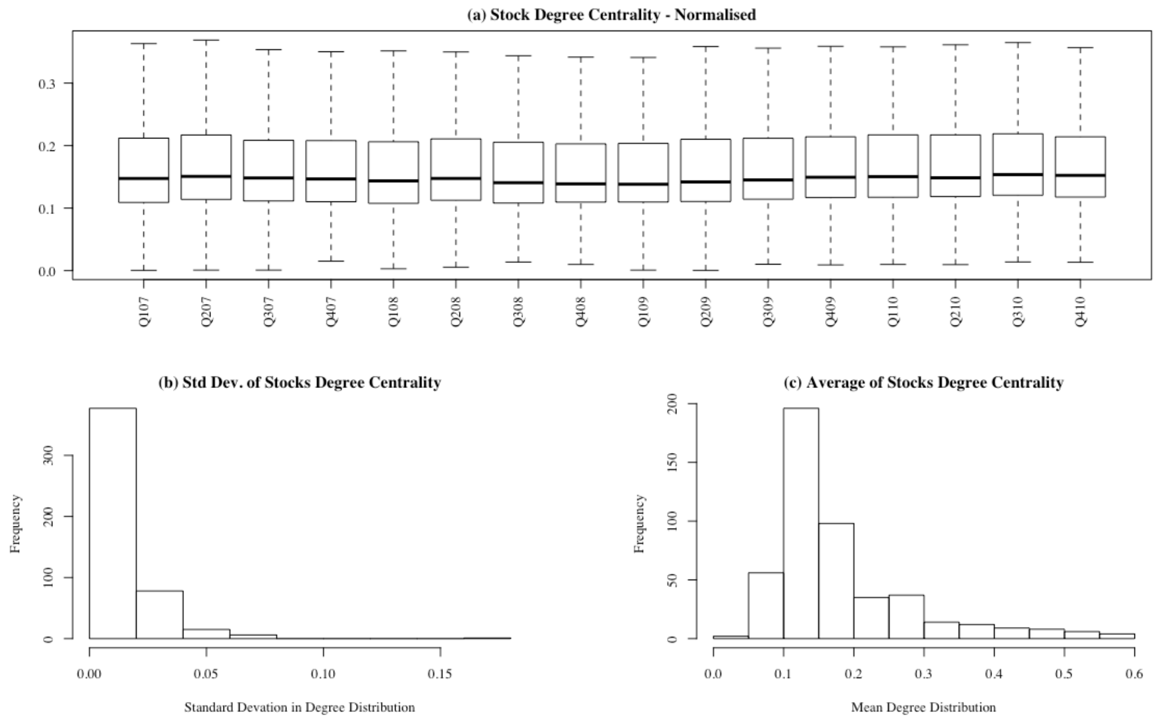

Figure 9: Boxplots (graph (a)) illustrating the variation in the degree centrality for stocks as calculated by equation 2, from the bipartite network. Graph (b) and (c) are histograms showing standard deviations and mean of the centrality measure across the sample.

From Figure 9(a), it appears that there was little movement in the number of managers who held each stock, with Figure 9(c) showing the average in-degree centrality for the 477 stocks. It does appear that they are some heavily owned stocks, that is those stocks with a high degree distribution, in the index. The lack of volatility is supported by Figure 9(b) which shows that the standard deviation in the in-degree distribution for the 477 stocks was not large. The lack of variation is consistent with what was seen in Section 1.5, and a clearer picture of what is occurring is seen in Figure 8(c). The chart illustrates that stocks during the GFC, on average, had a reduced number of investors, as investors reduced their holdings as they headed for quality stocks or cash. This finding is ultimately the same as what was seen in Figure 8(a). However, there is a contrasting finding in terms of the standard deviations. Figure 8(d) shows that variation in the number of investors in a stock declined during the GFC before recovering, something not seen in Figure 8(b). Interestingly, towards the end of the sample period the standard deviation grows strongly, again suggesting that there were some heavily favored stocks.

Having established the SxS networks, an attempt was made to detect a community structure in the these networks. Ideally, the network would be split into clear communities where a stock would be allocated to a community based on one or more of the following characteristics: industry classification, growth profile, dividend yield, and/or volatility. Alas, the outcomes as seen in Figure 10 and Figure 11 were not as clear cut. An issue that proved problematic in the community detection was the density of each of the 16 SxS networks, which meant that the igraph (2006), cluster_fast_greedy function, which attempts to define communities via directly optimizing a modularity score, had little variation on which to split the stock. Greater success, in terms of defining a greater number of communities, was achieved by using the SxS networks that were weighted by the proportion of shares held by each manager.

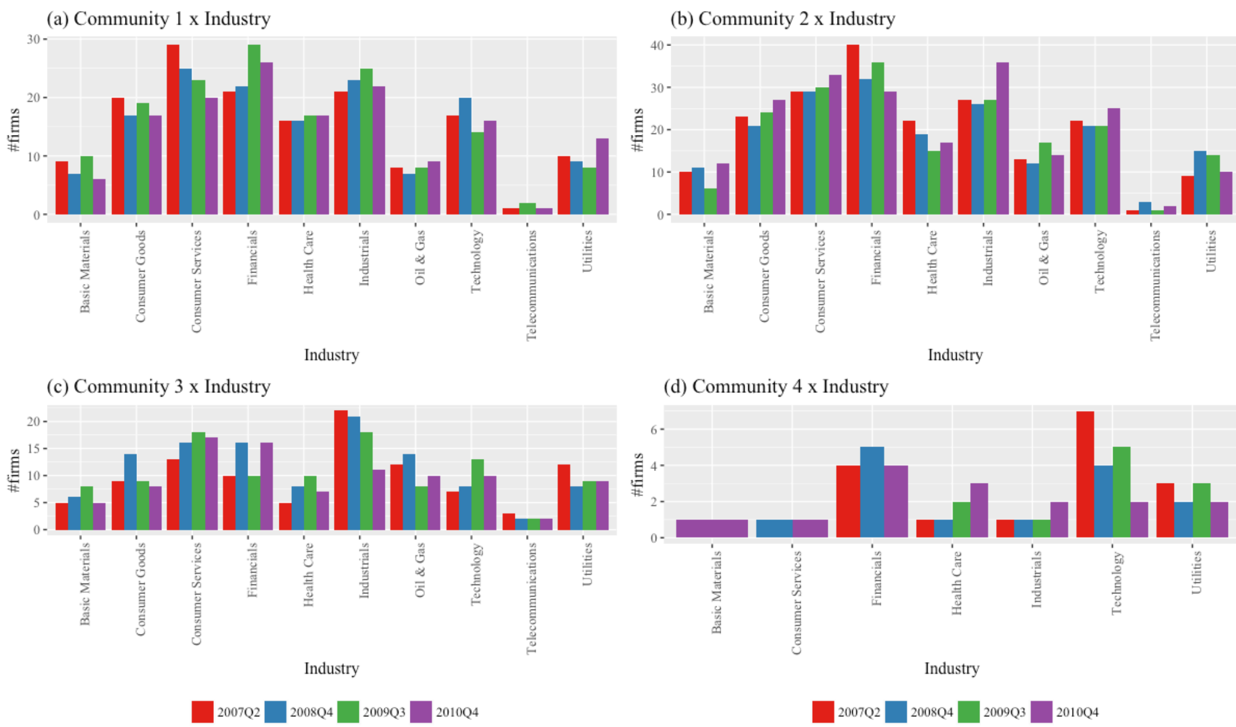

Figure 10: The results of the community detection and divided into industry and time segments.

The result of the community detect was that 4 communities were found in each of the 16 periods. Community 2 was the dominant community, with approximately 42% of the stocks classified to it. Community 4 was the smallest with approximately 2.5% of stocks classified to it, with several periods over 3%. Figure 10 illustrates for, 4 specific periods, the distribution of stocks within the communities and their industries. In terms of intercommunity movement, it appears that financial and industrial stocks joined community 1, while consumer services and technology stocks moved to community 3. The exact dynamics behind these changes are unclear but to understand the classification a further interpretation of the community structure is provided for Q4, 2010 (the other three charts are held in the appendix), in Figure 11. Here the communities are subdivided by the financial variables as detailed in Section 1.2.

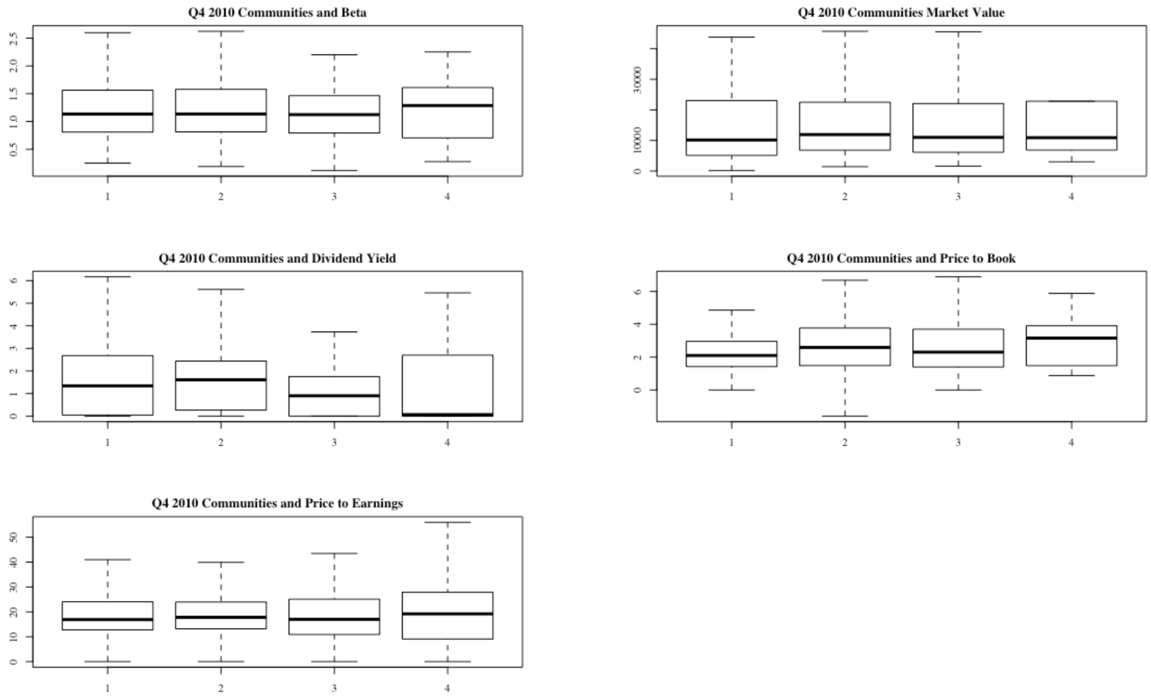

Figure 11: The results of the community detection for the fourth quarter of 2010 divided by various financial metrics for the relevant stocks.

The charts suggest that Community 4 may be a large cap growth community, with the justification being a higher average beta, market value, PE ratio, and PB ratio, and a lower dividend yield. It also appears that Community 1 may be value oriented in comparison to Community 3, given the lower PB ratio and higher dividend yield. Further work in the area may look to extend this analysis as the identification of clear set communities will aid portfolio diversification.

1.7 Stock Pricing Behavior

Having seen the dynamics regarding the degree distribution of the stocks and investors, this section looks to uncover any relationship between the price movements of the stocks and their position in the network. The first step in this process was to gauge how a stock should have performed, which was achieved by calculating the betas, as per the CAPM, for each of the stocks, for each of the 48 months in the sample. While the metric is not without issues, it does provide a meaningful measure of how a stock’s returns compared to the market, with a beta greater (lesser) than 1 suggesting returns (positive or negative) greater (less) than the markets, but in the same direction.

Figure 12(a) provides the boxplots for each of the 48 months and in combination with Figure 12(b) and Figure 12(c) provides a significant insight. What is seen is that the range of betas contracts as the market heads into the GFC. However, it is during the recovery period that the analysis uncovers the most significant insights. The first point of note is that the median beta effectively remains constant from March 2009 with the range tapering from that point. The range of returns for the stocks (Figure 12(b)) also contracts over this period. A possible mechanism for this outcome is given by Figure 12(c), where we see the cluster coefficient at the investor level increases in close to a monotonic manner. This suggests that the managers on a whole were broadening their portfolios, in effect "buying the market" as the market rallied, thereby reducing the heterogeneity among holdings and returns.

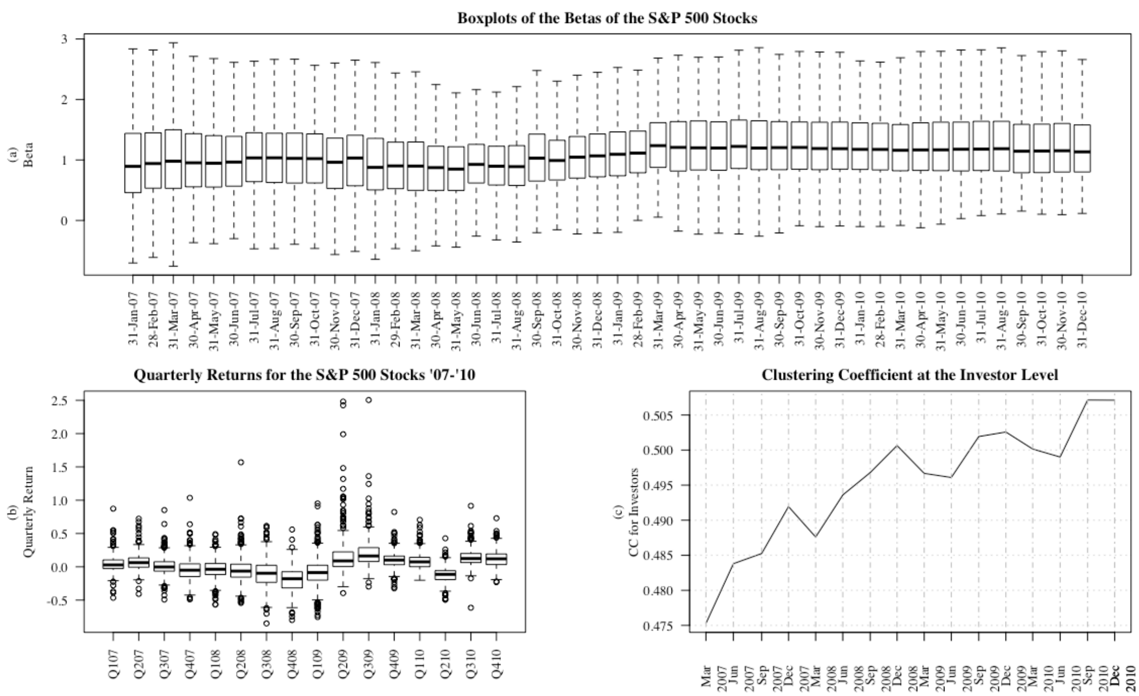

Figure 12: Boxplots (graph (a)) illustrating the variation in stock betas calculated as per the CAPM. Graph (b) illustrates variation in stock returns for the sample period. Graph (c) plots the weighted cluster coefficient for the investor level network.

To understand whether network metrics are capable of explaining stock returns several exploratory steps were taken. The first was to see whether a stock's mean Eigenvector centrality (from the SxS networks) across the sample period was related to its mean returns (Figure 13(a)).

The standard deviations were also compared (Figure 13(b)). The rationale for this approach was that the interpretation of a high mean Eigenvector centrality measure is that managers that have a co-holding in the given stock have co-holdings in other well-held stocks. This means the stock is being used to balance a portfolio, and is likely to be held by "core" managers; that is, ones that track the market. Therefore, the stock should show less volatility as it is held with other commonly held stocks. Unfortunately, there appears to be little in the way of a meaningful relationship.

The second approach was to see whether a stock's mean degree centrality (as defined by Equation 3) across the sample period was related to its mean returns (Figure 13(c)). The rationale for this approach was that the interpretation of the degree centrality in this framework is that a stock with a consistently high value is one that many managers always tend to hold, thereby identifying it as a core holding, or market darling. Therefore, a stock with a high mean value should show less volatility as managers will always hold the stock. Alternatively, a stock that fluctuates from a low to high degree (or vice versa) could be classified as a fade stock, with a dramatic increase suggesting that trade in the stock has become "crowded" as the market herds.

There does appear to be a stronger relationship with stocks with a higher mean degree experiencing lower average returns. This warranted further investigations to see if the lower average returns are the result of larger variations in their returns or more consistent returns, albeit lower. The result of this analysis is illustrated in Figure 13(d), where there is a negative relationship (albeit with an R squared of only 5%) in terms of the price volatility of a stock, as given by the standard deviation of its returns (the y-axis in Figure 13(b) & (d)), and the average number of investors. The justification is that stocks that are held by more managers are likely to see less volatility as a single manager in general does not have sufficient power to move the price of the stock.

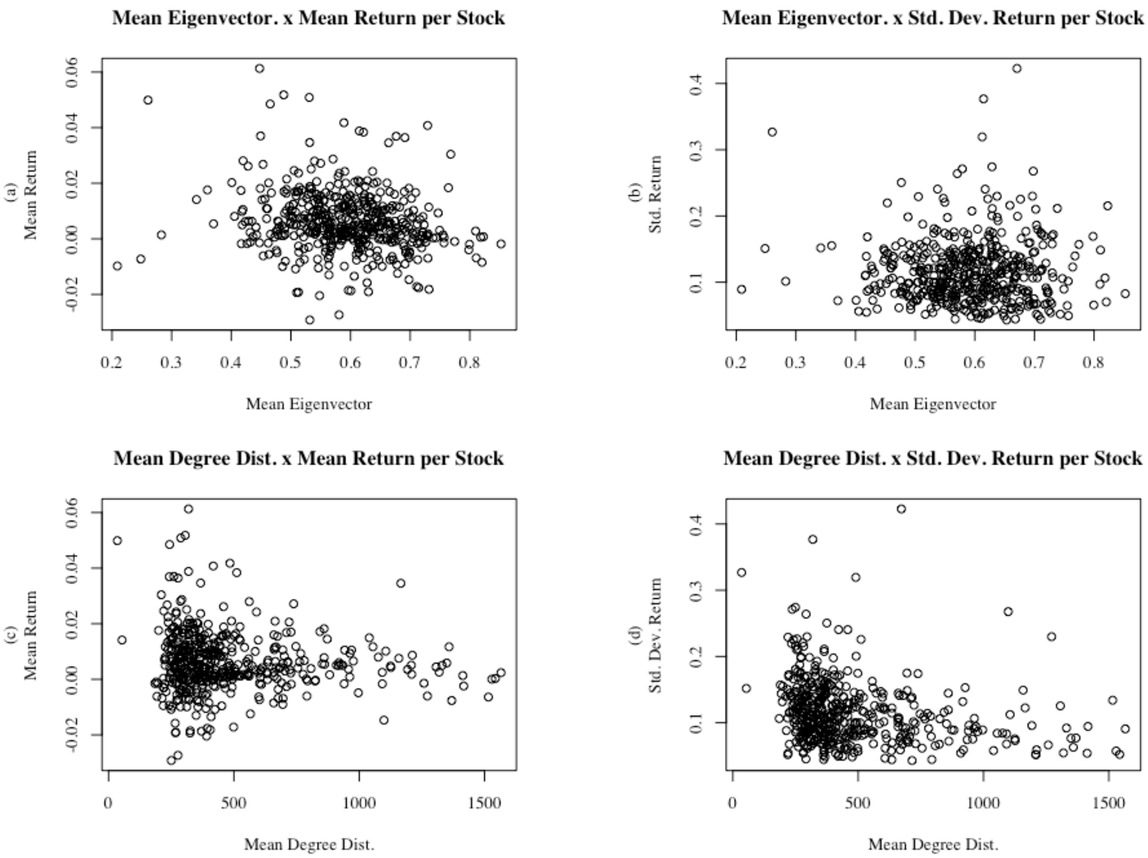

Figure 13: Scatter plots detailing the relationships between the average (and standard deviation) stock returns and their mean degree distribution and Eigenvector centrality.

To further investigate the findings from Figure 13, linear regressions comparing a stock's quarterly change in price (their return) and degree distribution were run for each of the 477 stocks, with the interpretation of the slope being the "living beta" for the stock. The results are summarized in Figure 14(b) and (c), where in general there is a positive relationship between the change in a stock's price and its change in its degree distribution. This is a relatively straightforward interpretation; that is, as more investors hold a stock the price goes up, and vice versa. Unfortunately, as seen in Figure 14(c) the "living betas" do not provide a sufficiently strong explanation of a stock's price movement.

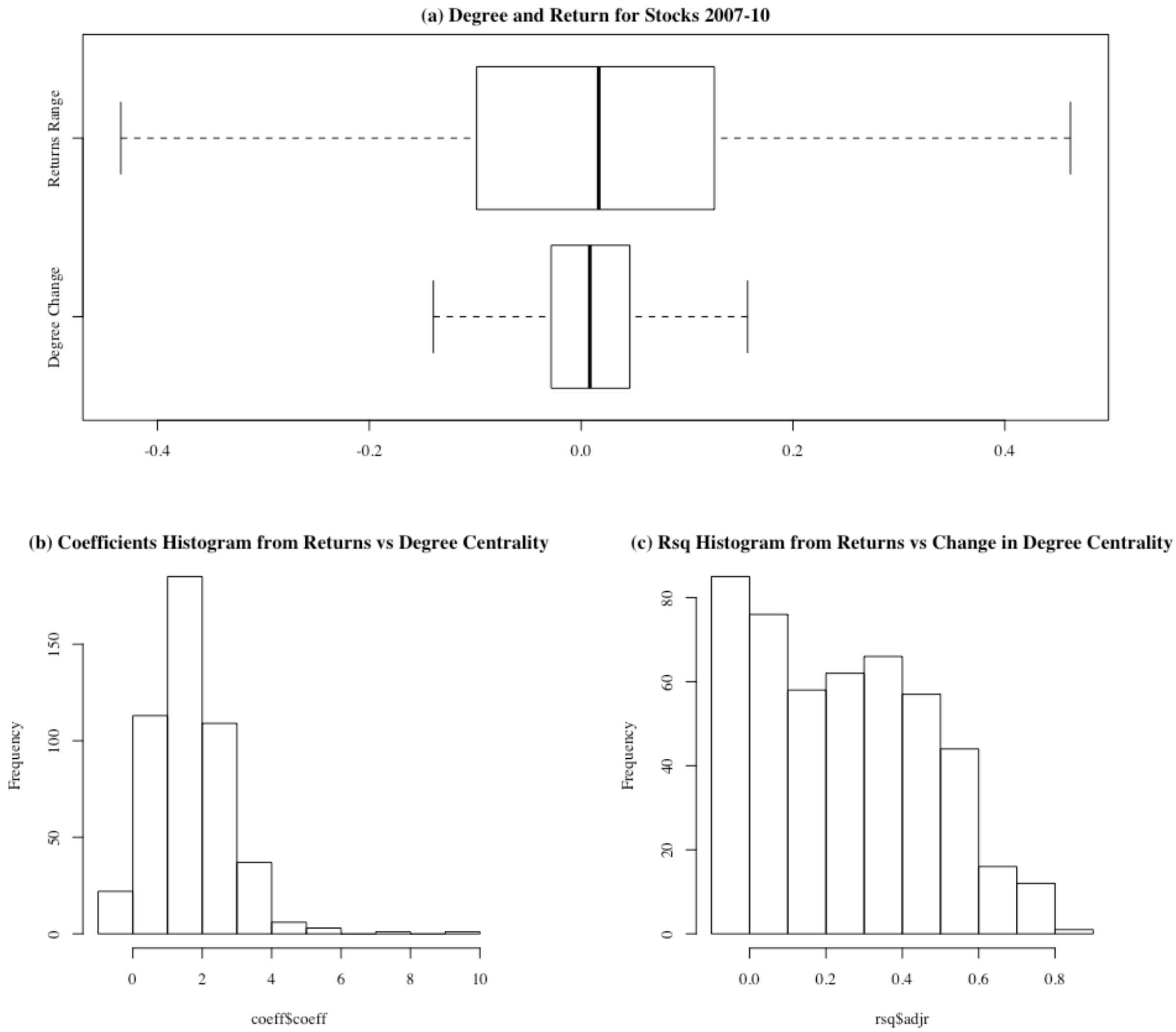

Figure 14: Boxplots illustrating the range of changes in the quarterly degree distribution and share price for the stocks in the sample is provided in Graph (a). Graph (b) is a histogram plotting the coefficients from regressing quarterly changes in degree distrbution against quarterly changes in price. Graph (c) is a histogram of the R-squares returned from those regressions.

Figure 14(a) provides a summary of the variation in prices compared to the variation in the degree distribution. Consistent with the "living beta" coefficients being at least 1, there was far more variation in price than the degree distribution. Interesting questions stemming from this figure are why is the variation so much larger for price, and what is occurring outside the US institutional investor network? Regardless, the analysis shown in this section can identify fade stocks and when investors act with a herding mentality.

Summary and Conclusion

There is little doubt that the traditional economic/finance equilibrium solutions developed in the 1950's and 1960's have been unable to explain periods of excess volatility in global financial markets. This has led to numerous other approaches gaining prominence, as they have provided greater insight. One such approach is to consider stock markets as complex adaptive systems (CAS), (see for example Johnson, Jefferies & Hui (2003) or Farmer et al. (2012)). Within this framework the importance of networks has been identified and has led to extensive research utilizing network science. This field of research has considered, and found numerous factors, that can explain the behavior of financial markets, with these factors being outside the realm of standard equilibrium analysis.

The exploratory steps undertaken in this paper have confirmed the utility of utilizing network data in uncovering the dynamics of the stock market. This paper presents a novel result, which was to capture the parallel movement in the density of the investment network, and the volatility and value of S&P 500 index. This evidence suggests that the market may indeed function in a similar fashion to an ecosystem, with the adaptive cycle theory (Holling, 2001) providing a feasible framework. Further work in this area should look to make better use of the weighted networks, and to integrate trading volumes, which would provide further insight into how rapidly the network is changing. While the results in terms of finding the "living beta" of stocks were not sufficient to challenge the established finance theories, the analysis did provide useful insights, and avenues for future research. The first priority in extending this line of research will be to extend the dataset, and if possible increase the granularity of the data, because as it stands the quarterly data may be missing some of dynamics. Another extension will be to try and understand why some stocks experience greater fluctuations in their degree distribution, and to understand the tipping points for when a trade in a stock has become too crowded.

References

Boginski, V., Butenko, S., & Pardalos, P. M. (2006). Mining Market Data: A Network Approach. *Computers & Operations Research*, *33*(11), 3171–3184. https://doi.org/10.1016/j.cor.2005.01.027

Bonanno, G., Caldarelli, G., Lillo, F., Micciche, S., Vandewalle, N., & Mantegna, R. N. (2004). Networks of Equities in Financial Markets. *The European Physical Journal B - Condensed Matter*, *38*(2), 363–371. https://doi.org/10.1140/epjb/e2004-00129-6

Caldarelli, G., Battiston, S., Garlaschelli, D., & Catanzaro, M. (2004). Emergence of Complexity in Financial Networks. In E. Ben-Naim, H. Frauenfelder, & Z. Toroczkai (Eds.), *Complex Networks* (Vol. 650, pp. 399–423). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-44485-5_18

Cont, R., & Bouchaud, J.-P. (2000). Herd Behavior and Aggregate Fluctuations in Financial Markets. *Macroeconomic Dynamics*, *4*(2). https://doi.org/10.1017/S1365100500015029

Csardi, G., & Nepusz, T. (2006). The igraph Software Package for Complex Network Research. *InterJournal*, *Complex Systems*, 1695.

Dormann, C. F., Gruber, B., & Fruend, J. (2008). Introducing the bipartite Package: Analysing Ecological Networks. *R News*, *8*(2), 8–11.

Farmer, J. D., Gallegati, M., Hommes, C., Kirman, A., Ormerod, P., Cincotti, S., … Helbing, D. (2012). A complex systems approach to constructing better models for managing financial markets and the economy. *The European Physical Journal Special Topics*, *214*(1), 295–324. https://doi.org/10.1140/epjst/e2012-01696-9

Holling, C. S. (2001). Understanding the Complexity of Economic, Ecological, and Social Systems. *Ecosystems*, *4*(5), 390–405. https://doi.org/10.1007/s10021-001-0101-5

Hong, H., Kubik, J. D., & Stein, J. C. (2005). Thy Neighbor's Portfolio: Word-of-Mouth Effects in the Holdings and Trades of Money Managers. *The Journal of Finance*, *60*(6), 2801–2824. https://doi.org/10.1111/j.1540-6261.2005.00817.x

Johnson, N. F., Jefferies, P., & Hui, P. M. (2003). *Financial market complexity*. Oxford ; New York: Oxford University Press.

Kenett, D. Y., Tumminello, M., Madi, A., Gur-Gershgoren, G., Mantegna, R. N., & Ben-Jacob, E. (2010). Dominating Clasp of the Financial Sector Revealed by Partial Correlation Analysis of the Stock Market. *PLoS ONE*, *5*(12), e15032. https://doi.org/10.1371/journal.pone.0015032

Newman, M. E. J. (2010). *Networks: An Introduction*. Oxford ; New York: Oxford University Press.

Opsahl, T. (2013). Triadic Closure in Two-Mode Networks: Redefining the Global and Local Clustering Coefficients. *Social Networks*, *35*(2), 159–167. https://doi.org/10.1016/j.socnet.2011.07.001

Ozsoylev, H. N., & Walden, J. (2011). Asset Pricing in Large Information Networks. *Journal of Economic Theory*, *146*(6), 2252–2280. https://doi.org/10.1016/j.jet.2011.10.003

Ozsoylev, H. N., Walden, J., Yavuz, M. D., & Bildik, R. (2014). Investor Networks in the Stock Market. *Review of Financial Studies*, *27*(5), 1323–1366. https://doi.org/10.1093/rfs/hht065

Preis, T., Kenett, D. Y., Stanley, H. E., Helbing, D., & Ben-Jacob, E. (2012). Quantifying the Behavior of Stock Correlations Under Market Stress. *Scientific Reports*, *2*. https://doi.org/10.1038/srep00752

R Core Team. (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Schweitzer, F., Fagiolo, G., Sornette, D., Vega-Redondo, F., Vespignani, A., & White, D. R. (2009). Economic Networks: The New Challenges. *Science*, *325*(5939), 422–425. https://doi.org/10.1126/science.1173644

Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk. *The Journal of Finance*, *19*(3), 425–442. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x

Shiller, R. J., & Pound, J. (1989). Survey Evidence on Diffusion of Interest and Information Among Investors. *Journal of Economic Behavior & Organization*, *12*(1), 47–66. https://doi.org/10.1016/0167-2681(89)90076-0

Sornette, D. (2014). Physics and financial economics (1776–2014): puzzles, Ising and agent-based models. *Reports on Progress in Physics*, *77*(6), 62001. https://doi.org/10.1088/0034-4885/77/6/062001

Appendix

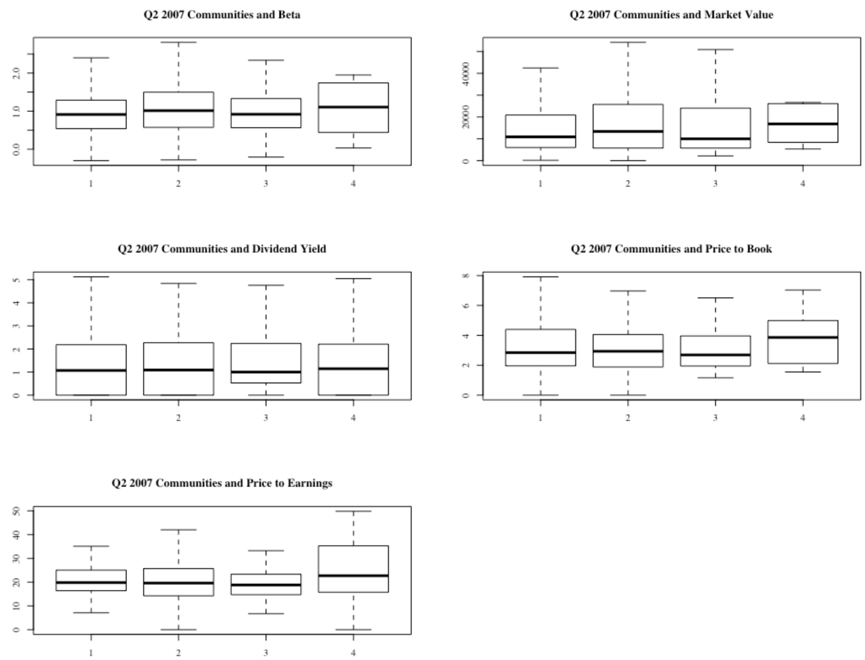

Figure 15: The results of the community detection for the second quarter of 2007 divided by various financial metrics for the relevant stocks.

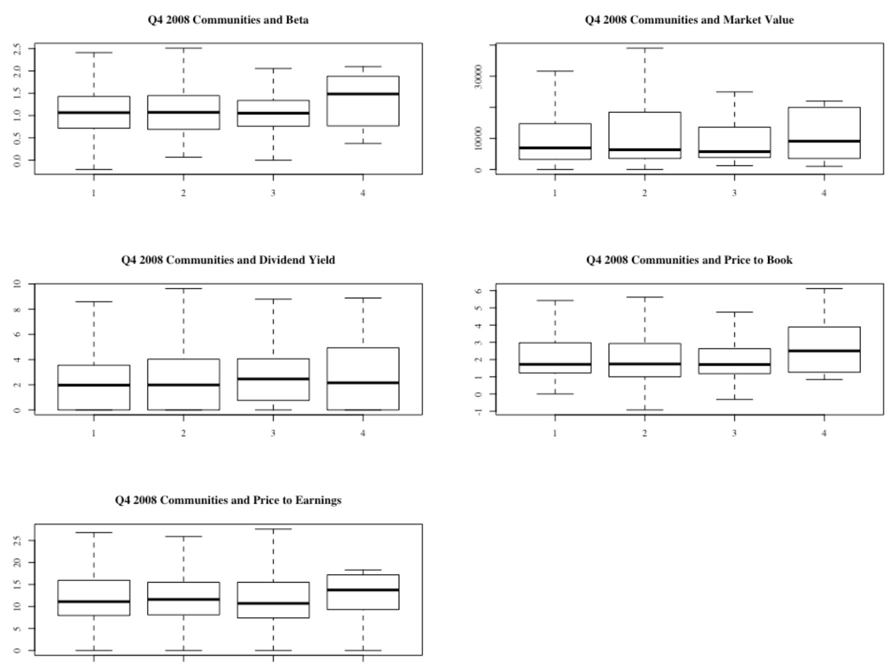

Figure 16: The results of the community detection for the fourth quarter of 2008 divided by various financial metrics for the relevant stocks.

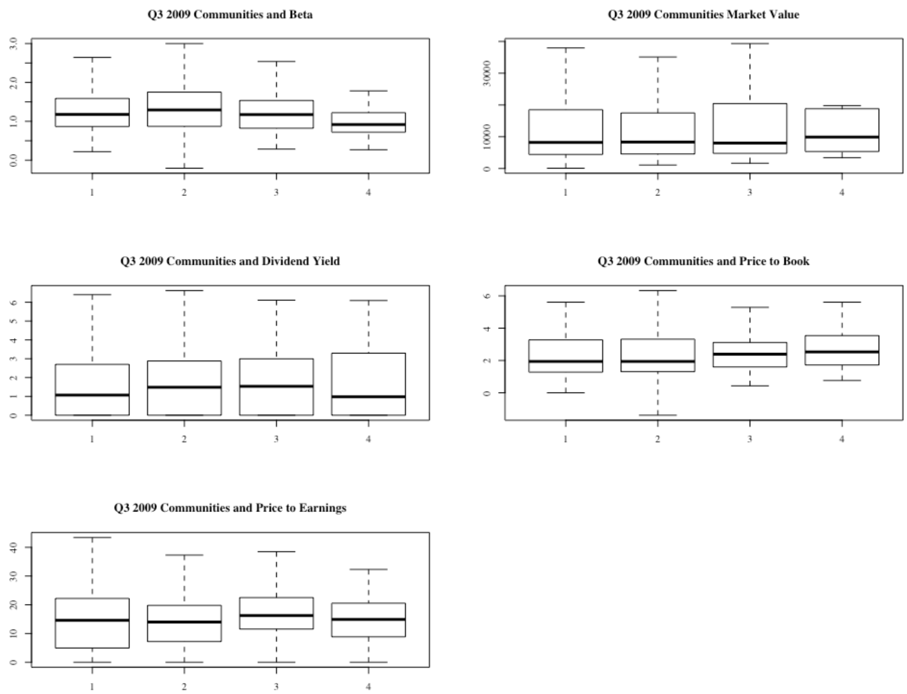

Figure 17: The results of the community detection for the third quarter of 2009 divided by various financial metrics for the relevant stocks.